Fifth International Conference on Computer Information Science and Artificial Intelligence (CISAI 2022)

Yuanchang Zhong
Editor

16–18 September 2022
Chongqing, China

Organized by
Chongqing University of Technology (China)
Chongqing University (China)

Sponsored by
Chongqing Engineering Research Centre for Computer Vision Intelligence Technology (China)
School of Artificial Intelligence, Liangjiang, Chongqing (China)
AEIC Academic Exchange Information Centre (China)

Published by
SPIE

Volume 12566
Part One of Two Parts
## Contents

xiii  Conference Committees

### Part One

**COMPUTER SYSTEMS AND INTERNET APPLICATIONS**

<table>
<thead>
<tr>
<th>Proc. of SPIE Vol. 12566 1256601-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>12566 02</td>
</tr>
<tr>
<td>12566 03</td>
</tr>
<tr>
<td>12566 04</td>
</tr>
<tr>
<td>12566 05</td>
</tr>
<tr>
<td>12566 06</td>
</tr>
<tr>
<td>12566 07</td>
</tr>
<tr>
<td>12566 08</td>
</tr>
<tr>
<td>12566 09</td>
</tr>
<tr>
<td>12566 0A</td>
</tr>
<tr>
<td>12566 0B</td>
</tr>
<tr>
<td>12566 0C</td>
</tr>
<tr>
<td>12566 0D</td>
</tr>
<tr>
<td>12566 0E</td>
</tr>
<tr>
<td>12566 0F</td>
</tr>
</tbody>
</table>
Design and application of smart industrial park based on Internet of Things technology [12566-61]

A protection framework based on dynamic heterogeneous redundancy architecture [12566-27]

Taxi destination prediction based on LSTM with tree memory module [12566-66]

Research advanced in human pose estimation based on deep learning [12566-155]

Research on the main elements of mimic platforms [12566-55]

Fire detection system based on improved multi-sensor information fusion [12566-71]

Logit-based stock prediction network [12566-99]

Application analysis of machine learning in cyberspace security research [12566-204]

Multi-hop question answering for SRLGRN augmented by textual relationship modelling [12566-127]

Multi-level fast personalized recommendation system [12566-148]

A novel hash scheme for high elastic grid blockchain [12566-67]

Ship scene vehicle location method based on multi-source information fusion [12566-137]

Mimic adjudication scheme based on homomorphic encryption [12566-108]

Orientational information matters in trajectory prediction [12566-141]

Database design of load spectrum calculation software for wheeled crane [12566-58]

Dialog history management for end-to-end task-oriented dialog systems [12566-117]

Transformer and long short-term memory networks for long sequence time sequence forecasting problem [12566-151]

Multi-master and multi-slave oriented task offloading strategy for real time and low power Internet of Vehicles [12566-119]

Gateway malicious command detecting for intelligent fishery sensor networks [12566-80]

Calculation and analysis method for distribution network fault location based on improved differential evolution algorithm [12566-130]
| 12566 10 | Research on numerical simulation of deep seabed blowout and oil spill range [12566-54] |
| 12566 11 | Numerical simulation on compressive strength of waste fiber recycled concrete based on Python. [12566-48] |
| 12566 12 | IoT and big data analysis based prevention and intervention network system for breast cancer susceptible people [12566-186] |
| 12566 13 | Numerical simulation study on impacts of the Xinhengsha reclamation project on salinity in the Yangtze Estuary [12566-93] |
| 12566 14 | Design and implementation of log audit system based on spark big data platform [12566-171] |
| 12566 15 | Long-term stock price forecast based on PSO-informer model [12566-112] |
| 12566 16 | Learning anisotropy and asymmetry geometric features for medical image segmentation [12566-24] |
| 12566 17 | Influence of images generated by CGAN on the performance of image classification based on CNN [12566-215] |
| 12566 18 | Medical data sharing scheme based on hybrid encryption with revocable attribute [12566-19] |
| 12566 19 | Data collection and talent demand analysis of recruitment website based on principal component analysis [12566-102] |
| 12566 1A | Research on remote sensing image classification based on RA-UNet [12566-120] |
| 12566 1B | Analysis of heterogeneous data model based on federated learning [12566-196] |
| 12566 1C | Stock market trend prediction using CBAM and CNN [12566-41] |
| 12566 1D | Research on AOD furnace end-point temperature prediction based on sparrow search algorithm optimization ELM [12566-64] |
| 12566 1E | Transformer in galaxy: a study of galaxy classification based on the Swin transformer [12566-146] |
| 12566 1F | iDHS-DPPE: a method based on dual-path parallel ensemble decision for DNase I hypersensitive sites prediction [12566-56] |
| 12566 1G | Exploration of profit model of beach sports based on data drive [12566-74] |
| 12566 1H | Logistics data sharing method based on federated learning [12566-20] |
| 12566 1I | Application and prospect of big data technology in medical image enhancement [12566-140] |
FSC-based data fusion sensitivity assessment scheme [12566-69]

Theoretical analysis of vibration conditions in high voltage power equipment and on-line processing of field data [12566-205]

Application of attention mechanism in image search [12566-123]

Correlation analysis and prediction between bitcoin price and its influencing factors [12566-145]

Multi-source data fusion evaluation method based on uncertainty quantification and Bayes method for success-failure type test [12566-95]

Database management and data mining [12566-172]

Design and implementation of art grading line dance scoring system based on Android and SQLite [12566-100]

A high performance bitcoin trading strategy prediction model [12566-50]

Discussion on cooperative governance strategy and security storage technology of power data based on blockchain [12566-62]

Research on adaptive model of English translation based on data fusion [12566-72]

FL-Lightgbm prediction method of unbalanced small sample anti-breast cancer drugs [12566-44]

Cointegration identification with metric learning [12566-76]

Tracking pedestrians from a moving camera based on Kalman filter [12566-142]

A prediction of human body key-points based on Kalman filter in transient visual masking [12566-2]

Research on vehicle tracking technology in expressway cross-monitoring area based on radar and video fusion [12566-39]

An image analysis tool for cell migration assay with a large sample size [12566-63]

Continuous sign language recognition based on 3DCNN and BLSTM [12566-83]

Blind image super-resolution reconstruction based on dual regression network [12566-153]
Remote sensing image segmentation network based on attention mechanism feature fusion [12566-194]

Agricultural text named entity recognition based on the BiLSTM-CRF model [12566-125]

Multi-task and feature attention for semi supervised medical images segmentation [12566-79]

The method of grounding fault location of distribution line based on transient signal [12566-163]

Embedding diverse features in latent space for face attribute editing [12566-122]

Image harmonization with spatial feature interaction and back-projection upsample [12566-87]

Research and implementation of dynamic gesture recognition system based on ZYNQ [12566-110]

VisMole: a molecular representation based on voxel for molecular property prediction [12566-94]

Weld defect recognition method based on improved DenseNet [12566-115]

Design of vulnerability scanning method for video surveillance system [12566-12]

Entity extraction based on the parts of speech attention mechanism [12566-70]

A fast point clouds registration method with ISS-shot feature descriptors [12566-57]

Accuracy analysis and development of multi-channel acquisition device for domestic intelligent programmable logic controller [12566-40]

Research on new recognition method of web component based on machine learning [12566-42]

Commodity knowledge graph-based TransD-KGAT method for recommendation [12566-97]

Research advanced in occluded face recognition [12566-159]

Face-based gender recognition with small samples generated by DCGAN using CNN [12566-174]

Wild animal recognition based on effective-class-balanced softmax loss [12566-36]

Adversarial attacks on cross-resolution person re-identification [12566-157]

A method for wireless communication interference signal identification based on extreme learning machine [12566-106]
Knowledge graph construction of component quality management [12566-53]

Remote sensing landslide recognition method based on LinkNet and attention mechanism [12566-82]

Part Two

Multi-scale context-aware segmentation network for medical images [12566-91]

A fault detection method for AUV based on multi-scale spatiotemporal feature fusion [12566-14]

Application of data-driven feature extraction methods in biometrics [12566-128]

Analysis of cigarette quality and its influencing factors based on XGBoost model [12566-22]

An obstacle detection method for LiDAR after removing water surface clutter [12566-88]

Maize pests identification based on improved YOLOv4-Tiny [12566-169]

Real-time detection of railway cracks based on improved YOLOX-Nano [12566-78]

A method for patient gait real-time monitoring based on powered exoskeleton and digital twin [12566-136]

ROBOT MODELING AND AUTOMATION CONTROL

An automated storage and retrieval system optimization with MILP methods [12566-129]

TransCGan-based human motion generator [12566-188]

Internet of Things real-time data remote monitoring system based on Wi-Fi technology [12566-165]

Research on collaborative knowledge innovation mode mining based on user behavior in open source community [12566-114]

An intelligent blind people guide cane design based on Arduino [12566-52]

A framework for automatic identification of neural network structural redundancy based on reinforcement learning [12566-185]
Ranking influential nodes by combining normalized degree centrality and fine-grained K-Shell

Study for CT to MRI translation based on cycle-GAN and UNet-GAN

Electronic insulator detection based on DN-DETR

An adaptive scheduling approach with agent and Q-learning for aircraft overhaul shop-floor

Substation priority maintenance planning based on genetic ant colony algorithm

AI-based substation construction remote management and violation detection system

Autonomous driving based on deep neural network

Activity recognition based on adaptive window and broad learning

Research on LSTM multi-factor quantitative stock selection strategy based on attention mechanism

Utilize multichannel attention amplification fusion for skin disease diagnosis

Automatic driving model based on machine learning agents

Analysis of location selection of urban public charging stations based on machine learning

Automatic picking method of microseismic first arrival based on support vector machine based on particle swarm optimization

Application of Matlab in Yinchuan climate change characteristics

Ti-SSE: a time interval based sequential recommendation model via stochastic shared embeddings

Spatial location-based outdoor mobile augmented reality 3D registration technology

Prediction of Bitcoin prices' trends with ensemble learning models

Investment capacity measurement model of grid enterprises under transmission and distribution cost supervision

Schedule control of substation engineering based on artificial intelligence

Detection of violent crowd behavior based on mean kinetic streak flow
<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Title</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>12566 3L</td>
<td>A Seq2Seq based model for generating diverse sentiment dialogues</td>
<td>[12566-21]</td>
</tr>
<tr>
<td>12566 3M</td>
<td>Simulation study on explosion damage of reinforced concrete column</td>
<td>[12566-92]</td>
</tr>
<tr>
<td>12566 3N</td>
<td>Human resource scheduling technology based on improved genetic algorithm for pulse assembly beat balancing</td>
<td>[12566-9]</td>
</tr>
<tr>
<td>12566 3O</td>
<td>Research on digital watermarking algorithm based on discrete cosine transform</td>
<td>[12566-182]</td>
</tr>
<tr>
<td>12566 3P</td>
<td>Face tracking using Kanade-Lucas-Tomasi algorithm</td>
<td>[12566-132]</td>
</tr>
<tr>
<td>12566 3Q</td>
<td>Hybrid forecasting model in used car price forecasting based on stochastic algorithm</td>
<td>[12566-121]</td>
</tr>
<tr>
<td>12566 3R</td>
<td>Research on application of artificial intelligence on translation applications: a case study of Xun-Translation application</td>
<td>[12566-7]</td>
</tr>
<tr>
<td>12566 3S</td>
<td>DJ-Agent: music theory directed a cappella accompaniment generation using deep reinforcement learning</td>
<td>[12566-154]</td>
</tr>
<tr>
<td>12566 3T</td>
<td>A novel genetic algorithm for resource allocation and task scheduling in MEC scenario</td>
<td>[12566-189]</td>
</tr>
<tr>
<td>12566 3U</td>
<td>Human fall detection based on multi-algorithm fusion</td>
<td>[12566-190]</td>
</tr>
<tr>
<td>12566 3V</td>
<td>Research on multi-label long text classification algorithm based on transformer-LDA</td>
<td>[12566-138]</td>
</tr>
<tr>
<td>12566 3W</td>
<td>Neighborhood node selection strategy for blockchain networks based on Boltzmann method</td>
<td>[12566-13]</td>
</tr>
<tr>
<td>12566 3X</td>
<td>Classic Augmentation based Classifier Generative Adversarial Network (CACGAN) for COVID-19 diagnosis</td>
<td>[12566-180]</td>
</tr>
<tr>
<td>12566 3Y</td>
<td>Application of single hidden layer fuzzy recurrent wavelet neural network to multi-physical parameter monitoring and on-line analysis of high voltage power equipment</td>
<td>[12566-206]</td>
</tr>
<tr>
<td>12566 3Z</td>
<td>ProGAN-based convolutional neural network for improving gender classification</td>
<td>[12566-176]</td>
</tr>
<tr>
<td>12566 40</td>
<td>Research on blockchain scalability based on sharding strategy</td>
<td>[12566-32]</td>
</tr>
<tr>
<td>12566 41</td>
<td>A hybrid optimized multi-population flower pollination algorithm for web service composition problem</td>
<td>[12566-68]</td>
</tr>
<tr>
<td>Page</td>
<td>Title</td>
<td>Reference</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>42</td>
<td>An improved algorithm for elliptic curve digital signature</td>
<td>12566-166</td>
</tr>
<tr>
<td>43</td>
<td>Research on multi-target tracking method based on Kalman filter and deep learning</td>
<td>12566-147</td>
</tr>
<tr>
<td>44</td>
<td>Banks customer churn analysis and prediction using random forest</td>
<td>12566-152</td>
</tr>
<tr>
<td>45</td>
<td>JSVuExplorer: a JavaScript vulnerability detection model based on transfer learning</td>
<td>12566-28</td>
</tr>
<tr>
<td>46</td>
<td>Airfoil inverse design based on particle swarm optimization algorithm</td>
<td>12566-11</td>
</tr>
<tr>
<td>47</td>
<td>An improved deep-learning monocular visual SLAM method based on local features</td>
<td>12566-104</td>
</tr>
<tr>
<td>48</td>
<td>Construction of agricultural pest killing model based on artificial intelligence: Take phototactic pests as an example</td>
<td>12566-59</td>
</tr>
<tr>
<td>49</td>
<td>Research on stock trading prediction based on MAD and Q-learning</td>
<td>12566-183</td>
</tr>
<tr>
<td>4A</td>
<td>Multi-site air quality prediction based on graph convolutional neural network-bi-directional LSTM model</td>
<td>12566-101</td>
</tr>
<tr>
<td>4B</td>
<td>Research on pedestrian target detection based on deep learning</td>
<td>12566-37</td>
</tr>
<tr>
<td>4C</td>
<td>Reference-based super-resolution with texture transformer and residual network</td>
<td>12566-181</td>
</tr>
<tr>
<td>4D</td>
<td>Ozone concentration prediction based on deep neural networks</td>
<td>12566-164</td>
</tr>
<tr>
<td>4E</td>
<td>A reinforcement learning-based network load balancing mechanism</td>
<td>12566-162</td>
</tr>
<tr>
<td>4F</td>
<td>Research on optimization of charging pile siting based on immune algorithm</td>
<td>12566-203</td>
</tr>
<tr>
<td>4G</td>
<td>Personalized recommendation algorithm for enterprise science and technology talents based on labeling</td>
<td>12566-201</td>
</tr>
<tr>
<td>4H</td>
<td>AIACT-GAN: CT reconstruction based on dynamic attention and generative adversarial networks</td>
<td>12566-65</td>
</tr>
<tr>
<td>4I</td>
<td>Abnormal trajectory recognition algorithm of networked vehicles based on improved particle swarm optimization</td>
<td>12566-170</td>
</tr>
<tr>
<td>4J</td>
<td>Forecasting method of rural science and technology talents demand based on coupling algorithm</td>
<td>12566-173</td>
</tr>
<tr>
<td>4K</td>
<td>Research on fast cable fault location method based on artificial intelligence</td>
<td>12566-192</td>
</tr>
<tr>
<td>4L</td>
<td>Research on tomato leaf disease identification based on deep learning</td>
<td>12566-43</td>
</tr>
</tbody>
</table>
12566 4M  Safety status identification model of civil construction project based on deep learning  
[12566-175] 

12566 4N  Performance diagnosis and optimization of building system based on building information  
model  [12566-89] 

12566 4O  Design and implementation of drug pick-and-place robot for elderly service  [12566-179] 

12566 4P  Intelligent system design of storage materials in automatic stereo warehouse  [12566-86] 

12566 4Q  Research on interactive product design of procrastination patients  [12566-187] 

12566 4R  Sequential recommendations based on time series and dynamic interest for TV program  
[12566-8] 

12566 4S  Non-cooperative game power control algorithm based on Sigmoid effective function in  
cognitive radio  [12566-34] 

12566 4T  Application of the model in the stock market  [12566-208] 

12566 4U  Emotional visual processing of Chinese information based on ERPs  [12566-214] 

12566 4V  Neural mechanism of visual paralanguage based on computer network  [12566-213] 

12566 4W  Determining efficient placement of electric vehicles charging stations using integer linear  
programming  [12566-210] 

12566 4X  Extension of the Traveling Salesman Problem on Tourism in California  [12566-212] 

12566 4Y  Application of C4.5 algorithm in customer group classification of business websites  [12566-216] 

12566 4Z  Galaxy morphology classification based on ResNeXt  [12566-60] 

12566 50  Prediction model of students' learning behavior on learning effect in online live class based on  
machine learning algorithm  [12566-209] 

12566 51  Research on visual question answering system based on deep neural network model  
[12566-217]
Conference Committees

Conference General Chairs

He Yan, Chongqing University of Technology (China)
Yuanchang Zhong, Chongqing University (China)

Publication Chair

Ning Sun, Nankai University (China)

Technical Program Committee Chair

Xiaochuan Zhang, Chongqing University of Technology (China)

Organizing Committee Chair

Bin Liu, Chongqing University of Technology (China)

Organizing Committees

Michal Feckan, Comenius University in Bratislava (Slovakia)
Shuanghua Yang, Southern University of Science and Technology (China)
Ayush Dogra, Council of Scientific and Industrial Research (India)
Nada M. Al Hakak, Baghdad College for Economic Science University (Iraq)

Academic Committees

Julia Qing Zheng, California Baptist University (United States)
Sahil Verma, Lovely Professional University (India)
Babu George, Fort Hays State University (United States)
Tao Wang, Sun Yat-Sen University (China)
Kejun Li, Shandong University (China)
Guohua Qu, Dalian University of Technology (China)
Badrul Hisham bin Ahmad, Universiti Teknikal Malaysia Melaka (Malaysia)
Radu Rahul Vishwanath Dandage, Rajendra Mane College of Engineering & Technology (India)
Mohamed Said Mahmoud, China National Pulp and Paper Research Institute (China)
Surej Rajan C, Toc H Institute of Science and Technology (India)
Shivani Dhall, DAV College (India)
Emotion Analysis Method based on Emotion Intensity Fusion and BiGRU

Haoyang Zhang¹,a*, Changming Zhu¹,b
¹College of Information Engineering, Shanghai Maritime University, Shanghai, China
Corresponding author’s e-mail: 15137312774@163.com

Abstract

In Chinese sentiment analysis, sentiment words are just a drop in the ocean compared with the whole corpus. In order to solve the problem of insufficient emotion lexicon and prior knowledge, proposes a method to predict the emotion intensity of target words based on neural network model(Neural Network Emebedding Score, NNES). By training a small number of labeled samples, using clustering algorithm to find the seed words, calculate the similarity between the target words and the seed words, and using it as the input of neural network to predict the emotional intensity of the unlabeled words. Compared with the traditional machine learning regression models, it has smaller mean square error. Meanwhile, a BiGRU model based on attention mechanism and convolution is proposed by integrating the predicted emotion intensity with word vector(Neural Network Emebedding Score with CNN and Attention-BiGRU, NNESC-Att-BiGRU). To compare several popular models on product and hotel review data sets, and the proposed model has better classification effect on Chinese sentiment classification task.

Keywords-Emotional analysis, Emotional intensity fusion, BiGRU, Convolution network, Regression model.

1. Introduction

Text sentiment analysis refers to the process of analyzing, processing and extracting subjective texts with personal emotions by using techniques such as text mining and natural language processing[1]. In recent years, text sentiment analysis has become one of the hot topics in natural language processing and text mining due to its wide application in many fields.

In literature [2], Calvo and Kim put forward the text dimension and classification model: discrete category representation and continuous dimension representation. Text sentiment analysis can be divided into five categories according to different levels[3]: (1)Vocabulary level:Usually, the emotion dictionary is constructed in advance to classify the emotion categories by calculating the score of the target text in the emotion dictionary[4]. (2)Phrase level:Compared with lexical level, phrase-based text sentiment analysis considers the influence of different degrees of modifier on emotional intensity. For example, "very good" and "not good", if the single word "good" is considered, then the emotional score of both is the same, which obviously leads to some error. However, phrase-level emotion analysis requires the establishment of a large enough adverb emotion dictionary in advance and the method of domain rules[5] for emotion analysis, while the fixed modifier weight cannot adapt to a more complex language expression model. (3)Sentence level: The task at sentence level was to excite the target with VA(Valence-Arousa) value and to find sentences with subjective color in the text so as to predict the emotional tendency of the target sentence. (4)Text level: In the same way as the sentence level, it only deals with the difference of sentence length, divides the text into sentences so as to mark the VA value and analyze the emotional tendency of the text. (5) Attribute level, also known as feature level: It is no longer limited to the grammar, vocabulary and structure of the sentence itself, but pays more attention to the sentence itself, and can carry out more complex emotional analysis. However, the existing emotion dictionary cannot completely cover most of the emotion vocabulary. Dalian University of Technology Chinese Emotion Vocabulary Noumenal database provides 27,466 words annotation, but only 974 words are included in the hotel data set.

In recent years, with the continuous development of neural network models, emotion analysis based on neural network and deep learning has made great progress. In the emotion analysis based on statistical machine learning, literature [6] adopts the minimum learning framework, combined with machine learning, and experiments prove that the application of text classification to subjective analysis of documents improves the performance of emotion analysis. CNN's text representation model has achieved good results in the work of local feature information. RNN can process time series data input, and the output of neuron can be directly applied to itself in the next time period, but gradient explosion will...
occur as the gradient propagates. Long Short-Term Memory network\textsuperscript{7} (LSTM) and Gated Recurrent Unit\textsuperscript{8} (GRU) solve this problem. Input gate, output gate and forgetting gate are added to the network structure to save and control the transmission of sequence information, so that sequence information can be better extracted from the text and long-term dependence relationship can be learned. A series of variants based on LSTM and GRU have also achieved good results in Chinese emotion analysis tasks\textsuperscript{9}\textsuperscript{13}. In sentiment analysis tasks, text sentences are often segmented and word vectors are constructed as input. However, due to the reason of sentence structure, the training model of word vector has errors for some words with similar syntactic structure but completely opposite semantics. For example, the similarity of "Heaven" and "Hell" in Word2vec model reaches 0.695.

In order to solve the problem that the construction of sentiment dictionary is very complicated and cannot cover a large number of words in the target sentence, this paper proposes a neural network-based sentiment analysis method for content words to obtain the sentiment score of the target words. Compared with the machine learning regression method, the model achieves better results. In order to reduce the errors generated by Word2vec model mentioned above and apply the emotional information of words themselves, this paper proposes a model based on attention mechanism and convolutional neural network for emotion analysis by combining emotional intensity with word vector. The pre-trained emotion score is applied to a Bidirectional Gated Recurrent Unit (BiGRU) based network, emotional vocabulary score is added in the embedding layer to better obtain prior knowledge, and attention mechanism and convolution operation are combined to obtain local and global semantic features of the text. Compared with the current model, good results are achieved.

2. Materials and Methods

2.1 Emotion intensity prediction method based on neural network

This paper proposes a neural network-based emotion intensity prediction method (NNES) to predict the emotion value of the target vocabulary. The training data comes from the Chinese Emotional Vocabulary Noumenal Database of Dalian University of Technology, which is divided into 7 categories and 21 sub-categories, and contains a total of 27,466 words. On this basis, only affective intensity and polarity data are selected in this paper. Eliminate words with a polarity of 0 (neutral); normalize emotional words with a polarity of 1 (positive) and an intensity of 1-9 to 0.5-1. The greater the intensity, the greater the normalized value, that is 9 is 1 after normalization; Emotion words with a polarity of 2 (negative) and an intensity of 1-9 are normalized to 0-0.5, with a bias of 0.001 to prevent 0 from appearing. In negative words, the higher the emotional intensity, the lower the normalized value, that is, the normalized value of 9 is 0.001. After that, on this basis, 2000 positive words (0.5-1) and 2000 negative words (0.001-0.5) were randomly selected for the experiment.

Through Word2vec model, word vectors of 4000 seed words are obtained, and k-means method is used to try to find clustering centers of 4000 seed words. Experiments were conducted within the range of 10-200 cluster centers, and the experimental results are shown in Figure 1. It can be seen that the sum of the squares of error in the cluster (the sum of the distance between the sample and the nearest cluster center) decreases sharply when the number of cluster centers is 30-50, and the sum of the squares of error in the cluster keeps decreasing with the increase of the number.

![Fig.1 Relationship between square error in cluster and the number of cluster centers](image)
In order to retain the semantic information of seed words as much as possible and give consideration to the clustering effect and the input layer of the neural network model, 128 clustering centers were finally adopted. Through Euclidean distance, 128 words closest to the cluster center were calculated as seed words. The similarity between each word and seed word in 4000 words was calculated as the input of neural network. Emotional vocabulary scores were used as regression targets. The flow chart of the entire algorithm is shown in Figure 2. The neural network model is shown in Figure 3.

In the network model, the Input layer is a 128-dimensional vector of similarity between target words and seed words. The output dimensions of the three-layer Dense are: 7,13,1. The activation function was set as tanh, relu and tanh after some test. Dropout was added between the first and second layers of Dense and the parameter was set to 0.2. MSE was used as the loss function and Adam was used as the optimizer.

### 2.2 BiGRU model based on attention mechanism and convolution

CNN network generally consists of four layers: convolution layer, linear rectification layer, pooling layer and full connection layer[16]. The convolution layer is used to extract local features of input data. The characteristic information of the whole matrix is obtained by scanning the convolution kernel step by step. The one-layer convolutional operation can only extract some simple edge features, while the multi-layer convolutional operation can iterate more complex network models. After previous operations such as convolutional pooling, the model will learn feature information with higher dimensions and more abstract.

This paper integrates emotional vocabulary scores and proposes a neural network model based on attention mechanism and convolution (Neural Network Embedding Score with CNN and Attention-BiGRU, NNESC-Att-BiGRU) to complete the work of sentiment analysis. The model structure is shown in Figure 4.
In this model, word vector is trained by Word2vec model. Firstly, for the input text sequence \( \{w_1, w_2, w_3, ..., w_n\} \), input to Word_embedding and NNES at the same time, by Word_embedding, each word gets 128-dimensional word vector \( E = \{e_1, e_2, e_3, ..., e_n\} \). By passing through NNES network, each word gets 128-dimensional emotional feature \( T = \{t_1, t_2, t_3, ..., t_n\} \).

In the CNN, the input is \( E \) and \( T \). Through the convolution operation, the word vector expression \( H = \{h_1, h_2, h_3, ..., h_n\} \) integrating emotional information is obtained by multiplying the para elements. The word vector model \( H \) obtained contains the emotional tendency of the vocabulary. For negative vocabulary, the final word vector is smaller, while positive vocabulary is just the opposite.

In the BiGRU layer, the current hidden state of BiGRU is determined by the current input \( H \), the forward hidden layer state input \( M_{t+1}^+ \) and the reverse hidden layer state input \( M_{t+1}^- \) at \( t+1 \) moment. Therefore, the hidden layer state of BiGRU at time \( T \) can be obtained by the weighted sum of forward and reverse hidden layer states, as shown in Formula (1) - (3).

\[
M_t^+ = GRU(H, M_{t+1}^-) \\
M_t^- = GRU(H, M_{t+1}^+) \\
M_t = w_t M_t^+ + v_t M_t^- + b
\]

In the formula, \( w_t \) and \( v_t \) represent the weight of the state of the forward hidden layer and the state of the reverse hidden layer at time \( t \), and \( b \) is the bias.

The output \( M \) after BiGRU is input to the Attention layer, and the product of different probability weights and states of each hidden layer is calculated and then added up. After Softmax function is normalized, the Attention matrix is obtained. Finally, the predicted probability is output through the two-layer fully connected layer and Softmax layer.

3. Results & Discussion

There are two experiments in this chapter. Experiment 1 is the prediction method of emotion intensity based on neural network proposed in section 2.1. Experiment 2 is the BiGRU model based on attention mechanism and convolution proposed in section 2.2.
3.1 Experiment 1

3.1.1 Experimental Setup

In order to verify the effectiveness of the neural network model proposed in this paper, seven regression models in machine learning were used for comparison. There are SVR, KNN, RandomForest, Adaboost, GBRT, Bagging, ExtraTree. At the last is NNES which proposed of this paper.

Experimental data: Based on the Noumenal database of Chinese Emotion words of Dalian University of Technology, 2000 positive and negative words were randomly selected from a total of 4000 sample data after data pretreatment. Among them, 3000 are used as the training set and 1000 are used as the test set.

Evaluation indicator: Mean Squared Error (MSE). MSE value measures the average difference between the real value of the data sample and the predicted value of the model. The smaller the value is, the closer the predicted value is to the true value.

Network parameter setting: Dropout=0.2, Epoch=200, Batch_Size=200.

3.1.2 Experimental results and analysis

The experimental results are shown in Table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>Model</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>0.041</td>
<td>GBRT</td>
<td>0.042</td>
</tr>
<tr>
<td>KNN</td>
<td>0.049</td>
<td>Bagging</td>
<td>0.047</td>
</tr>
<tr>
<td>RandomForest</td>
<td>0.042</td>
<td>ExtraTree</td>
<td>0.090</td>
</tr>
<tr>
<td>Adaboost</td>
<td>0.051</td>
<td>NNES</td>
<td>0.037</td>
</tr>
</tbody>
</table>

It can be seen that the ExtraTree regression model has the largest error, and its MSE value is much higher than other models, reaching 0.09. Excluding the maximum value of 0.09, the mean MSE in the machine learning regression model was 0.045, while the lowest MSE in the SVR model was 0.041. The MSE of the neural network model proposed in this paper is 0.037, which is significantly improved compared with the machine learning regression algorithm.

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>real</th>
<th>predict</th>
<th>Vocabulary</th>
<th>real</th>
<th>predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>罪状</td>
<td>0.01</td>
<td>0.02</td>
<td>易如反掌</td>
<td>1.0</td>
<td>0.899</td>
</tr>
<tr>
<td>不干不净</td>
<td>0.343</td>
<td>0.289</td>
<td>求索</td>
<td>0.667</td>
<td>0.730</td>
</tr>
<tr>
<td>� שש</td>
<td>0.232</td>
<td>0.259</td>
<td>恶魔</td>
<td>0.778</td>
<td>0.539</td>
</tr>
<tr>
<td>孤单</td>
<td>0.232</td>
<td>0.217</td>
<td>棋名</td>
<td>0.889</td>
<td>0.783</td>
</tr>
<tr>
<td>战祸</td>
<td>0.232</td>
<td>0.222</td>
<td>诗仙</td>
<td>0.889</td>
<td>0.805</td>
</tr>
<tr>
<td>素洁</td>
<td>0.343</td>
<td>0.289</td>
<td>素洁</td>
<td>0.667</td>
<td>0.697</td>
</tr>
<tr>
<td>脏乱</td>
<td>0.121</td>
<td>0.279</td>
<td>水蒸</td>
<td>0.667</td>
<td>0.730</td>
</tr>
</tbody>
</table>

Table 2 shows the affective predicted values of some positive and negative words. It can be seen that the predicted values are completely consistent with the affective tendency of words, which is in line with people’s emotional cognition. Moreover, the error between the affective intensity and the true value is relatively low. For example: the predicted value of “罪状” is 0.02, and the true value is 0.01; The predicted value of “易如反掌” was 0.899; the real value was 1.0. Although there is some error, the emotional orientation of the target words is completely correct. The model could not distinguish the emotional score of polysemy. In order to alleviate this problem, 128 emotional words containing various emotional tendencies were used to calculate the similarity of target words from different perspectives at the beginning of the model design, and different directions were comprehensively considered in the emotional score, so as to alleviate the polysemy problem and provide more objective data for the model.

3.2 Experiment 2

3.2.1 Experimental Setup

In this paper, word vector was trained by Word2vec model, python version is 3.6, framework is TensorFlow=1.14.0, GPU is 14G, CUDA version is 10.0. We use the hotel and product reviews data sets come from https://github.com/SophonPlus/ChineseNLPCorpus, to finish the experiment. Information of data sets as shown in table 3. Network model parameters are shown in Table 4.
Table 3  The details of dataset

<table>
<thead>
<tr>
<th>dataset</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
<th>Average sentence length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotel reviews</td>
<td>5322</td>
<td>2444</td>
<td>7766</td>
<td>40.16</td>
</tr>
<tr>
<td>Product reviews</td>
<td>31728</td>
<td>31046</td>
<td>62774</td>
<td>18.34</td>
</tr>
</tbody>
</table>

Table 4  Network Model parameter setting

<table>
<thead>
<tr>
<th>parameter</th>
<th>Set up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max_sentence_length</td>
<td>100</td>
</tr>
<tr>
<td>Batch_size</td>
<td>64</td>
</tr>
<tr>
<td>Embedding_dims</td>
<td>128</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.2</td>
</tr>
<tr>
<td>Num_class</td>
<td>2</td>
</tr>
<tr>
<td>Epoch</td>
<td>250</td>
</tr>
<tr>
<td>Kernel_size</td>
<td>3</td>
</tr>
</tbody>
</table>

A total of 7 groups of comparison experiments were set:

(1) LSTM: Long and Short Term Memory network;
(2) CNN-LSTM: join CNN network in LSTM;
(3) BiLSTM: bidirectional LSTM;
(4) BiGRU: Bidirectional gated recurrent unit network;
(5) Att-BiGRU: Added attention layer behind BiGRU network;
(6) ES-BiGRU: Embedding emotional score in BiGRU;
(7) NNSC-Att-BiGRU: The final model was obtained by introducing CNN and Attention mechanism into the model ES-BiGRU.

Three evaluation indicator were set up in the experiment: Accuracy, Recall and F1-Score.

3.2.2 Experimental results and analysis

The experimental results are shown in Table 5 and Table 6.

Table 5  Result on hotel reviews

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>85.641%</td>
<td>83.983%</td>
<td>83.933%</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>85.741%</td>
<td>82.943%</td>
<td>83.585%</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>86.349%</td>
<td>84.611%</td>
<td>84.686%</td>
</tr>
<tr>
<td>BiGRU</td>
<td>85.705%</td>
<td>82.424%</td>
<td>83.434%</td>
</tr>
<tr>
<td>Att-BiGRU</td>
<td>86.027%</td>
<td>82.998%</td>
<td>83.899%</td>
</tr>
<tr>
<td>ES-BiGRU</td>
<td>86.541%</td>
<td>84.420%</td>
<td>84.777%</td>
</tr>
<tr>
<td>NNSC-Att-BiGRU</td>
<td>88.474%</td>
<td>86.637%</td>
<td>86.967%</td>
</tr>
</tbody>
</table>

From the experimental results, the experimental effect of the product review data set is significantly better than the hotel data set. Analysis of its possible existence of three reasons:1. In terms of the number of data sets, the product review data set is nearly 8 times that of the hotel data set. A large number of labeled training will have a good impact on the model. 2. The average sentence length of the product review dataset is only 18.34, which is much smaller than the hotel dataset. The sentence structure is simple, and the network model is more intuitive and easy to obtain in-depth information. 3. There are some mislabels in the hotel data set. For example, there is such a sentence in the article, "The breakfast is so bad, no matter how many people go, there is no food. It's time for hotels to take this issue seriously.". It should be an expression of negative emotion, but the data set is labeled positive. The error marks in data sets effect model training and the accuracy of model prediction.
According to the experimental results of commodity review data set, the model with poor performance is LSTM, with an accuracy of only 88.905%. After adding CNN network structure, the prediction accuracy of the model is improved by 0.317%. It can be seen that CNN is more conducive to the model capturing semantic information, thus improving the prediction accuracy. Comparing the experimental results of BiLSTM and BiGRU, the accuracy of BiLSTM is improved by 0.693%, because the BiLSTM model contains more complex network structure and larger network parameters, and BiLSTM has better performance in the test set of this experiment. The prediction accuracy of BiGRU is 89.198%, the accuracy of Att-BiGRU is 90.384%, and the accuracy of ES-BiGRU is 91.365%, which verify the effectiveness of the proposed model. Compared with the original BiGRU, the accuracy was improved by 2.167% after the emotion intensity of words was integrated. The addition of attention mechanism also improved the accuracy of the model by 1.186%.

Finally, the proposed bidirectional gated recurrent unit model (NNESC-ATT-BiGRU) with emotion score has 92.034% accurate, which is 2.863% and 0.669% higher than model BiGRU and Att-BiGRU, respectively. Compared with BiGRU and Att-BiGRU, the proposed model improves 2.769% and 2.447% respectively, according to the hotel review data set. The validity and significance of the model are verified.

### 4. Conclusion

This paper proposes a neural network model based on attention mechanism and convolution to solve the problem of neglecting the emotional intensity of words in Chinese sentiment analysis. By integrating the emotion score, the model can obtain more prior knowledge, so that words with similar word vectors but opposite meanings can be better distinguished. Compared with BiGRU and Att-BiGRU with attentional mechanism, the prediction accuracy of our model is improved by 2.863% and 0.669% on the product review data set respectively. On the hotel review data set, the prediction accuracy of the model in this paper is improved by 2.769% and 2.447% respectively. At the same time, a neural network-based affective intensity prediction method is proposed to predict the affective intensity of a large number of unlabeled target words, which solves the problems of tedious labeling and too few words in the affective dictionary. Compared with the regression model in machine learning, the proposed model has better mean square error and performance. It has achieved good results on the test set, and can correctly predict the emotional tendency and emotional intensity.

In the future work, based on the Dalian University of Technology Noumenal vocabulary and the work of this model, comprehensively consider the influence of degree adverbs and adjectives in text sentences on emotional vocabulary to further improve the accuracy of sentiment analysis. In Word2vec, for the same word in different sentences, there is the same word vector expression, without considering the polysemy of a word. In the future emotion analysis task, the influence of different word vector models on the emotion analysis task will be explored.

### References


Intelligent Classification and Identification of Radar Jamming Signals

Dongxia LI¹,a, Yahui SHI¹, Yangdong SUN¹, Bin ZHANG¹
Unit:¹Beijing Research Institute of Telemetry, Beijing, China
* e-mail:2636361054@qq.com

Abstract

Aiming at the problem of intelligent classification and recognition of radar jamming signals, the convolutional neural network structure is studied. By optimizing the basic network, the normalization layer, activation layer are added to the LENET-5 structure. Improve the accuracy of recognition results. The linear frequency modulation signal and amplitude modulation interference, frequency modulation interference, comb spectrum interference, slice reconstruction interference, intermittent sampling and forwarding interference are analyzed. Six signal models are used to generate data sets, and intelligent methods are adopted to realize classification and recognition.

Keywords-Radar jamming; Interference recognition; Convolutional neural network;

1. Introduction

With the development of electronic information technology, the international war has gradually evolved into a high-tech war dominated by electronic warfare. At present, radar, as the "eye" of complex battlefield, is the key to obtain battlefield information. Therefore, whether the radar can function properly is very important to the outcome of the war. At present, a variety of radar jamming technologies have been produced to prevent enemy radar from completing target detection, tracking, warning or navigation. At the same time, to deal with all kinds of interference anti-jamming technology emerged. Radar jamming identification is the prerequisite of anti-jamming technology, only understand the enemy jamming technology, can effectively anti-jamming.

2. Radar Active Jamming Signal

2.1 Linear frequency modulation signal

Linear frequency modulated signal (LFM) refers to the signal with continuous linear frequency change during the existence time. It increases the transmission bandwidth of the signal by modulating the carrier frequency, and achieves high range resolution under wide pulse by performing pulse compression at the receiver.

The mathematical expression of LFM signal (also known as Chirp signal) is as follows:

\[ S(t) = \text{rect}\left(\frac{t}{T}\right) e^{j2\pi\left(f_0 + \frac{K}{2}t^2\right)} \]  \hspace{1cm} (1)

Where \( f_0 \) is the initial frequency, \( t \) is the time variable, \( K = B/t \) is the FM slope, \( B \) is the bandwidth, and \( t \) is the time width for which the signal exists.

2.2 Amplitude modulation interference

The mathematical model of noise amplitude modulated interference (AM) is expressed as:

\[ J(t) = [U_0 + K_{AM}U_n(t)] \exp(j\omega_j + \varphi) \]  \hspace{1cm} (2)

The expression of this function is a generalized stationary stochastic process, where \( \exp() \) is an exponential function based on \( E \), \( U_n(t) \) is zero-mean Gaussian white noise, \( U_0 \) is the carrier voltage, and \( K_{AM} \), the modulation coefficient of noise, controls the power of \( U_n(t) \). \( \omega_j \) is the carrier frequency phase \( \varphi \) of interference uniformly distributed on \([0, 2\pi]\). And \( U_n(t) \) are independent random variables.
2.3 Frequency modulation interference

The mathematical model of noise frequency modulation interference (FM) is

\[ J(t) = U_j \exp \left( j\left( 2\pi f_c + 2\pi K_{AM} \int_0^t u(t') dt' + \varphi \right) \right) \]  \hspace{1cm} (3)

Noise FM interference is a generalized stationary stochastic process, in which \( \exp() \) is an exponential function based on \( E \), and the modulation noise \( \varphi \) is zero-mean white Gaussian noise uniformly distributed within \([0, 2\pi]\) and independently distributed with \( u(t) \). \( f_c \) is the central frequency of noise FM signal, \( K_{FM} \) is the frequency modulation slope, \( U_j \) is the amplitude of interference. Used to control the increase or decrease of frequency caused by a unit of modulated signal strength.

2.4 Comb spectrum interference

The COMB spectrum interference of LFM radar is mainly generated by the product modulation of COMB spectrum signal and linear frequency modulation signal. The generated interference can have the effect of deception or suppression. The expression of COMB spectrum signal is as follows:

\[ \text{comb}(t) = \sum_{i=1}^{M} a_i e^{j2\pi f_i t} \]  \hspace{1cm} (4)

Where, \( f_i \) corresponds to the frequency point of each comb tooth, and \( a_i \) is the amplitude at the corresponding \( i \) frequency point.

Then the mathematical model of comb spectrum interference of LFM radar is as follows:

\[ J(t) = s(t) \cdot \text{comb}(t) = \text{rect} \left( \frac{t}{T} \right) M \sum_{i=1}^{M} a_i e^{j2\pi \left( f_i + f_0 \right) \frac{t}{2} \left( K_{M} t \right)^2} \]  \hspace{1cm} (5).

2.5 Slice reconstruction interference

Slice reconstruction (C&I) jamming is a kind of dense range false target jamming developed for pulse compression radar. The generation principle of this kind of interference is shown in Figure 1:

![Figure 1. Schematic diagram of interference generation in slice reconstruction](image)

Assuming that the radar signal is \( S(t) \), the C&I interference signal can be expressed as the following mathematical model:

\[ J(t) = \sum_{k=0}^{n-1} p(t - kT / mn) \]  \hspace{1cm} (6)

Where, \( p(t) \) is expressed as:

\[ p(t) = s(t) \left[ \text{rect} \left( \frac{t - \tau_a}{\tau_a} \right) * \sum_{i=0}^{m-1} \delta(t - iT_a) \right] \]  \hspace{1cm} (7)
2.6 Intermittent sampling and forwarding interference

Intermittent sampling and forwarding (IS) jamming IS formed by using DRFM jammer to sample radar signals intermittently and then forward them in turn. It IS jamming IS formed by sampling and forwarding radar signals several times in a sampling period. The mathematical expression of such interference is:

\[ x_s(t) = \sum_{n=1}^{N} \frac{1}{\sqrt{T}} \cdot \text{rect}\left(\frac{t - \frac{\tau}{2} - \left(n - 1\right)T_s}{T}\right) e^{j2\pi\left(\frac{f_d + E}{2}t + \tau\right)} \]  \hspace{1cm} (8)

Where \text{rect} is the rectangular function, \( \tau\) is the width of the intermittent sampling pulse, \( T\) is the pulse width of the radar signal, \( T_s\) is the sampling period, then \( \tau/T_s\) represents the duty cycle of intermittent sampling.

3. Build The Data Set

3.1 The time-frequency analysis

3.1.1 Short Time Fourier Transform

STFT (Short Time Fourier Transform) is similar to the main idea of Fourier Transform\(^3\). The Fourier Transform of Time domain signal \( F(t) \) can be expressed as follows

\[ F(\omega) = F[f(t)] = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \]  \hspace{1cm} (9)

However, in the short-time Fourier transform, the time domain part of the Fourier transform is localized, and its mathematical expression is as follows:

\[ STFT(n,k) = \sum_{m=-(N-1)/2}^{(N-1)/2} f(m) \otimes (m-n) e^{-2j\pi mk} \]  \hspace{1cm} (11)

The discrete variables corresponding to continuous variables \( t, f \) and \( u \) are \( n, m \) and \( k \), respectively.

3.1.2 A simulated image of the interference signal

Figure 2. Time-frequency diagram of six kinds of signals
As can be seen from Figure 2, there are obvious differences in the time-frequency graphs of the six signals, and it is theoretically possible to use STFT for time-frequency analysis for signal recognition.

3.2 Generate the data set

The interference identification tasks set in this paper are divided into six categories: linear frequency modulation signal and amplitude modulation interference, frequency modulation interference, comb spectrum interference, slice reconstruction interference and intermittent sampling and forwarding interference.

Since it is difficult to acquire radar real jamming signals and a large number of sample signals are required for network training, this paper uses Matlab simulation method to generate database and adds random noise to the jamming signals generated by simulation to simulate the complex battlefield environment. The simulation parameters are as follows:

The center frequency of the interference signal is 200 MHz, the sampling frequency is 2000 MHz, and the bandwidth is 300 MHz. The additional noise of the interference signal was white Gaussian noise with SNR of -10 ~ 10 d B interval. 200 random samples were generated at each SNR, and a total of 6 × 200 × 11 samples were generated. 70% of the samples were the training set and the rest were the test set. STFT transformation was performed on each sample, and the time-frequency matrix obtained was stored as pictures to form the data set. The size of the generated image is normalized to match the input size of the model to reduce the computational burden.

4. Convolutional Neural Network

4.1 LeNet-5

![Network structure of LeNet-5](image)

Figure 3. Network structure of LENET-5

Although LENET-5 is a small network proposed earlier, it contains the basic modules of deep learning convolutional neural network: convolutional layer, pooling layer and fully connected layer. As shown in FIG.3, LENET-5 contains a total of 7 layers (the input layer is not used as the network structure), consisting of 2 convolution layers, 2 pooling layers and 3 connection layers respectively.[4] The kernel size of the down sampling layer and the fully connected layer represent the sampling range and the size of the connection matrix respectively.

4.2 BN-CNN

For a long time since the birth of CNN, LENET-5 has been the standard configuration structure of CNN classification network, in which the convolution-pooling layer is usually followed by one or more fully connected layers, and finally the classification function is used to get the probability prediction.

4.2.1 Add batch normalization layer

The BN (Batch Normalization) algorithm is input into one layer as well as the convolutional layer, pooling layer, and
activation layer. The BN layer is added after the convolutional layer and before the activation function, the input of the input activation function is normalized. In this way, the influence of offset and increase of input data is solved.

4.2.2 Nonlinear activation function ReLU.

The Sigmoid activation function in the Lenet-5 model was changed into a simpler ReLU activation function. On the one hand, the calculation of ReLU activation function is simpler, for example, it does not have the exponentiation of Sigmoid activation function. On the other hand, the ReLU activation function makes the model easier to train under different parameter initialization methods. This is because when the output pole of sigmoid activation function is close to 0 or 1, the gradient of these regions is almost 0, which causes that the backpropagation cannot continue to update some model parameters. However, the gradient of ReLU activation function in the positive interval is always 1. Therefore, if the model parameters are not properly initialized, the Sigmoid function may get a gradient of almost zero in the positive interval, which makes the model unable to be effectively trained.

5. Simulation Results

5.1 Simulation flow chart

![Simulation flow chart](image)

Figure 4. Simulation flow chart

5.2 Simulation experiment of interference recognition based on BN-CNN neural network

According to the generated interference signals and the constructed network, the classification and recognition algorithm is verified by simulation, and the simulation results are shown in the figure.

(a) The training set loss varies with the number of iterations
The recognition rate of trial set varies with the number of iterations, indicating that the network is well fitted. As the number of iterations continues to increase, the loss function no longer decreases significantly, but fluctuates normally within a certain range, indicating that the network has become stable at this time. Figure 5(b) is the curve of the recognition rate of the test set after smooth processing, and the recognition rate of the final test set is stable at 99.26%.

![Figure 5. Recognition results of convolutional neural network classifier](image)

The network’s recognition rate was 99.26 percent. As can be seen from FIG. 5(a), the loss of the network decreases rapidly with the number of iterations, indicating that the network is well fitted. As the number of iterations continues to increase, the loss function no longer decreases significantly, but fluctuates normally within a certain range, indicating that the network has become stable at this time. Figure 5(b) is the curve of the recognition rate of the test set after smooth processing, and the recognition rate of the final test set is stable at 99.26%.

![Figure 6. Confusion matrix of recognition rate of test set](image)

According to the test results of the test set, the confusion matrix of the recognition rate is drawn as shown in Figure 6. Among them, 6 kinds of signals correspond to labels 1, 2, 3, 4, 5, and 6 respectively. Confusion matrix gives the comparison between the output label of each interference signal and the actual label of the interference signal, which is more intuitive to represent the correct rate of the network.

As can be seen from FIG. 6, the recognition rate of each interference is above 97%, and the recognition rate of 5 kinds of interference is above 99%, while the recognition rate of intermittent sampling and forwarding and slice reconstruction interference is relatively low, because the generation mechanism of these two kinds of interference is similar, and the generated time-frequency map is slightly different, which makes the discrimination difficult.
5.3 The relationship between recognition rate and signal-to-noise ratio

Figure 7 shows the network recognition rate under different SNR. When it is -10~ -6db, the network recognition rate quickly gets better, then changes slowly, and the recognition rate is almost 100% when it is above 0 Db. It can be seen from Figure 6 and Figure 8 that the recognition accuracy of the network in the full SNR range (-10 ~ 10d B) is 99.26% and 95.9%, and when SNR is greater than 0d B, the recognition rate of the network is close to 100%, exceeding the total recognition rate. Combining with the curve change trend in Figure 7, it can be concluded that, The recognition rate of the convolutional neural network interference recognition algorithm is approximately 0 when the SNR is very low, and then increases rapidly with the increase of SNR and approaches 1 when the SNR is large, that is, all interference types can be correctly identified.

6. Conclusion

This paper studies the radar jamming signal type recognition algorithm based on the convolutional neural network model of STFT. The algorithm combines the time-frequency image based on the jamming signal and the convolutional neural network, and analyzes the recognition rate under different SNR. The simulation results show that the unique model structure of LENet-5 model and BN-CNN model can effectively extract the features from the low level to the high level of the image. In the range of -10 ~ 0 d B, the recognition rate of the network increases rapidly with the increase of the SNR, and the recognition rate is almost 100% when the SNR is above 0 d B.

In the full SNR range, the recognition accuracy of the network is high, and the recognition effect is good, but the calculation is large, which is not conducive to the engineering implementation. How to extract more efficient feature parameters needs to be further studied. The recognition of radar active jamming signals cannot stay in the theoretical research, and further engineering is needed to meet the demand of radar anti-jamming.

References

Research on parking lot recommendation System based on collaborative Filtering

Jingjing Yin¹,a /per 1st, Ying Liao²,b /per 2nd, Lizhu Yang³,c / per 3rd, Juan Zhou⁴,d /per 4th
¹Traffic Engineering College, Shandong Jianzhu University, Jinan, China
²Shandong Hi-speed Transportation Construction Group Co.,Ltd. Jinan, China
³Yunnan Design Institute Group Co.,Ltd. Kunming, China
⁴CHELBI Engineering Consultants,Inc. CCCC Highway consultants co.,Ltd. Beijing,China
 a-mail: yinjingjing0323@163.com, b-mail: SDEng_bit@163.com, c-mail: 546232127@qq.com,
d-mail: 727451599@qq.com

Abstract

Parking lot selection is an important part of drivers' parking behavior. Building a good parking lot recommendation system can reduce the time of parking lot selection and effectively alleviate parking problems. Firstly, the process of parking lot selection and the principle of collaborative filtering are analyzed. Then, the data of drivers' preferences are collected, and the cosine similarity method is used to find similar drivers and parking lots. Finally, based on the collaborative filtering theory, using MATLAB to realize the construction of parking lot recommendation system. The results show that the recommendation system can better recommend parking lots for drivers and facilitate drivers to make decisions quickly.

Key words -- parking choice; Cosine similarity; Collaborative filtering; Recommendation system

1. INTRODUCTION

According to the different information obtained by the parkers, the basic process of parking lot searching and choosing can be represented by Fig.1[1]. In the parking search process shown in the figure, a series of decisions are made from the initial search for parking space to the final parking space and then to the destination[2]: search for parking space, select the type of parking facilities (such as roadside parking, flat parking, three-dimensional parking, etc.), evaluate the parking space, and decide the route to the next parking space. The process of evaluating the parking lot is the most important decision-making process, which determines whether the search process should be terminated. If the current parking lot is accepted and available, the search will be terminated. The process becomes much easier when complete information is available before departure or on the road. It can be concluded that providing parking lot information can not only simplify the process of parking lot searching, but also reduce the amount of traffic searching for parking lot and the waiting time for entering the parking lot, which is conducive to promoting the improvement of traffic order.

Figure 1. Parking lot searching process
When choosing a parking place, the main factors to be considered are: the walking distance to the destination after parking, parking cost, waiting time, parking search time, parking safety and so on. Obviously, the parking place choice is based on weighing the above factors.

Parking lot choice often costs drivers a lot of time, and when drivers hesitate to park, it will also cause parking congestion to a certain extent, affecting traffic order and causing traffic problems.

At present, domestic and foreign experts [3-4] mainly consider the individuation of parking and study the parking recommendation system based on the subjective decision of parking and the concept of resource agent. Common recommendation algorithms include collaborative filtering recommendation algorithm[5], association rules-based recommendation algorithm[6], content-based recommendation algorithm[7] and hybrid recommendation algorithm[8-9]. Collaborative filtering recommendation algorithms are the most mature, including user-based and object-based[10].

This article first to the chosen area for similarity analysis, it is concluded that the development degree of similar area, parking lot parking index and speculation in different areas, the berth utilization, parking turnover, parking index such as time, determine the area of parking congestion and convenience, combined with the driver acceptable walking distance, parking purpose, etc., On the basis of comprehensive consideration of the sensitivity of parking choice factors in different areas, determine the probability of drivers choosing parking lots, and scientifically recommend parking lots to drivers.

2. THE BASIC PRINCIPLE OF COLLABORATIVE FILTERING

A complete recommendation system consists of three modules: behavior record module for collecting user information, model analysis module for analyzing user preferences and recommendation algorithm module. Among them, the recommendation algorithm module is the most core part of the recommendation system. The recommendation algorithm of collaborative filtering is currently the most widely used[11], which includes user-based collaborative filtering and item-based collaborative filtering. The realization process of collaborative filtering is mainly divided into three parts, and the specific steps are as follows:

2.1 Collect user preferences and standardize processing

2.1.1 Collecting user preferences

Find patterns in users' behaviors and preferences and make recommendations based on them. Therefore, how to accurately collect user preference information becomes a decisive factor affecting the system recommendation effect. There are many ways for users to provide information about their preferences to the system, and it can vary widely from application to application.

2.1.2 Data filtering and standardization

After collecting user behavior data, we also need to preprocess the data to some extent, among which the most core work is data screening and standardization.

Data filtering: The user behavior data is generated by the user in the process of use, which may have a large number of user misoperations. The invalid or incorrect data in the behavior data can be filtered through the classical data filtering method, which can make the analysis more accurate.

Standardization: It may be necessary to weight different behavioral data when calculating users' preferences for items. However, the data values of different behaviors may vary greatly. In order to unify the data of each behavior in a same value range and make the overall preference obtained by weighted sum more accurate, data standardization is needed. The simplest normalization process is the "extreme value method", which ensures that the data value is in the range of [0,1].

After pre-processing, groups or weighted processing can be selected according to different behavior analysis methods. After that, a two-dimensional matrix of user preferences can be obtained. One dimension is the list of users, the other dimension is the list of items, and the value is the user's preference for items.

2.2 Find similar users or items

After analyzing user behaviors and obtaining user preferences, we can calculate similar users and items according to user preferences, and then make recommendations based on similar users or items, which are the two branches of the most typical collaborative filtering: user-based collaborative filtering and object-based collaborative filtering. Both methods
need to calculate the similarity. As for the calculation of similarity, the existing basic methods are all based on vectors, which is to calculate the distance between two vectors. The closer the distance is, the greater the similarity is. In the recommended scenario, in the two-dimensional matrix of user-item preference, we can take one user's preference for all items as a vector to calculate the similarity between users, or take all users' preference for a certain item as a vector to calculate the similarity between items.

The calculation of similar neighbors includes a fixed number of neighbors and neighbors with a fixed similarity threshold.

A fixed number of neighbors (K-neighborhoods): no matter the neighbor's "distance", only in recent K, as its neighbours. As shown in A in Figure 2, assuming that five neighbors of point 1 are to be calculated, the nearest five points are selected according to the distance between points, namely point 2, point 3, point 4, point 7, and point 5. But obviously you can see, this kind of method for the calculation of isolated point effect is bad, because the neighbors of the fixed number, when it is not enough the similar point nearby, was forced to take some not very similar points as neighbors, thus influence the neighbor similar degree, can be seen from the graph, the points 1 and 5 are not very similar.

Neighbor calculation based on the similarity threshold (Threshold-based neighborhoods): The principle of neighbors with a fixed similarity threshold is different from that of neighbors with a fixed number. Neighbor calculation based on the similarity threshold is to the maximum limit of neighbor distance, fall in the current point as the center, a distance of K all points as the current point in the area of neighbor, this kind of method to calculate the number of neighbors are not sure, but the similarity of won't appear larger error. As shown in B in Fig 2, starting from point 1, neighbors with similarity within K are calculated, and points 2, 3, 4, and 7 are obtained. The similarity degree of neighbors calculated by this method is higher than that of the previous method, especially for the processing of isolated points.

<table>
<thead>
<tr>
<th>A: K-neighbourhoods</th>
<th>B: Threshold-based neighborhoods</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diagram A" /></td>
<td><img src="image" alt="Diagram B" /></td>
</tr>
</tbody>
</table>

2.3 To recommend

Item-based Collaborative Filtering (ItemCF): The basic principle is to discover the similarity between items based on the preferences of all users for items or information, and then recommend similar items to users based on the preference information of users. In short, suggest items that are similar to items your target audience has previously liked. In calculation, the preference of all users for an item is used as a vector to calculate the similarity between items. After obtaining the similar items of the item, the items that the current user has not expressed preference are predicted according to the user's preference, and a sorted item list is calculated as a recommendation. As shown in Fig 3, the user A like item A and C, users like items A and B item B and item C, user C items like A, from the history of these user preferences can analyze the items A and C items are relatively similar, like A people all like items C, based on this data can be concluded that the user is likely to also like C, C Therefore, the system will recommend item C to user C.
Object-based collaborative filtering does not involve content and has two major advantages: ① it does not need a large number of users to achieve moderate recommendation accuracy; ② it can immediately recommend new items once the attributes of the items are obtained.

The formula of user-based collaborative filtering recommendation algorithm is as follows:

$$p_{ui} = \sum_{j \in N(u) \cap S(i,K)} w_{ji} r_{ui} \quad (1)$$

Among them, \( p_{ui} \) represents user \( u \) interest in the goods \( j \), \( N(u) \) represents users like items collection (\( i \) was the one of the users like items), and \( S(i,K) \) represents \( i \) most similar \( K \) items collection (\( j \) is the one of a set of items), \( w_{ji} \) represents the similarity between item \( j \) and item \( i \), \( r_{ui} \) represents user \( u \) interest in item \( i \), if user \( u \) interested in item \( i \), \( r_{ui} \) is 1, otherwise \( r_{ui} \) is 0.

### 2.4 Parking lot recommended method calculation case

Since each driver's preference for parking lot selection varies greatly, parking lot recommendation should be based on the driver's preference to find parking lot similar to his preference. The recommendation method is object-based collaborative filtering recommendation. In this case, the driver is the user, and the parking lot is the item. As shown in Figure 4, the drivers are A, B, C, D, E, and the parking lot is a, b, c, and d. Parking lot a has been selected by the driver \{A,B,D\}, and parking lot b has been selected by the driver \{A,C\}, then the cosine similarity calculation formula can be used to obtain the similarity between parking lot a and b:

$$w_{ab} = \frac{\|A,B,D\cap A,C\|}{\|A,B,D\|\|A,C\|} = \frac{1}{\sqrt{6}} \quad (2)$$

Similarly, the similarity between parking lot a and c and between parking lot a and d can also be calculated:

$$w_{ac} = \frac{\|A,B,D\cap B,E\|}{\|A,B,D\|\|B,E\|} = \frac{1}{\sqrt{6}} \quad (3)$$

![Figure 3 Item-based collaborative filtering](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)
In order to calculate the cosine similarity formula section, each divided by its respective denominator, get the parking lot of similar matrix.

\[
W_{ad} = \frac{||A, B, D\| \cap \{C, D, E\}}{\sqrt{||A, B, D\||\{C, D, E\}}} = \frac{1}{3} \quad (4)
\]

If the cosine formula were used to calculate the similarity for any two parking lots, this would be time-consuming and a lot of work. In fact, the similarity between many parking lots is 0 because they have never been selected by the same driver, and \( |N(u) \cap N(v)| = 0 \), the above calculation takes a lot of time to calculate the similarity of such parking lots with the result of 0, which is meaningless. Therefore need to first screen the parking lot of \( |N(u) \cap N(v)| \neq 0 \), in order to achieve this goal, there need to be a parking lot - driver inversion schedule, as shown in Fig 4, gets a parking lot of 4 * 4 matrix \( w \), the molecules in this matrix is the cosine similarity formula section, each divided by its respective denominator, get the parking lot of similar matrix.

After obtaining the similarity information between different parking lots and determining the neighbor set of the current parking lot, formula (1) can be used to predict the degree of preference of drivers to parking lots. In this case, drivers C and E can be recommended. Drivers C and E have not chosen parking lot a, so this parking lot can be recommended to drivers C and E. According to UserCF algorithm, the degree of interest of drivers C and E in parking lot a is:

\[
p(C, a) = w_{ab} + w_{ad} = 0.7416 \quad (5)
\]

\[
p(E, a) = w_{ac} + w_{ad} = 0.7416 \quad (6)
\]

3. CONSTRUCTION OF PARKING LOT RECOMMENDATION SYSTEM

In practice, a large number of parking lots and drivers are often involved in the recommendation process, and simple formula calculation cannot meet the needs. Therefore, it is necessary to write and run programs to complete the recommendation of parking lots, and then build a parking lot recommendation system.

3.1 Driver preference data collection and collation

3.1.1 Get parking behavior data

Data were obtained by questionnaire survey, on-site survey and on-site recovery, collecting 3475 questionnaires from 140 parking lots, mainly investigating the parking behavior of drivers. This paper aims to obtain drivers' evaluation of parking lots. Drivers' evaluation of parking lots generally takes parking costs, convenience and safety into consideration. Question 1 parking price refers to parking cost, Question 2/3/4 the distance between a parking lot and its destination, the presence or absence of guidance, and the time spent looking for a parking space reflect the convenience of the parking lot. Question 5 is the security of the parking lot. According to the principle that the higher the score is, the better the driver's evaluation of the parking lot is, the values of each survey option are assigned. The specific survey content and the assigned values are shown in Table 1.
TABLE 1  DRIVERS' PARKING BEHAVIOR AND PREFERENCES SURVEY

<table>
<thead>
<tr>
<th>survey content</th>
<th>options</th>
<th>Option value assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What do you think the parking price of the parking lot is</td>
<td>① Too high ② easonable ③ cheap</td>
<td>The values are respectively 1/2/3</td>
</tr>
<tr>
<td>2. Approximate distance of the parking lot from your destination</td>
<td>① ≤5min ② 5<del>10min ③ 10</del>20min/④ &gt;20min</td>
<td>The values are respectively 4/3/2/1</td>
</tr>
<tr>
<td>3. Do you have any guide for parking this time</td>
<td>① yes ② no</td>
<td>The values are respectively 2/1</td>
</tr>
<tr>
<td>4. The time between entering the parking lot and finding the parking space</td>
<td>① ≤5min ② 5<del>10min ③ 10</del>20min/④ &gt;20min</td>
<td>The values are respectively 4/3/2/1</td>
</tr>
<tr>
<td>5. Parking lot safety</td>
<td>① Unsafe ② safe ③ general</td>
<td>The values are respectively 1/2/3</td>
</tr>
</tbody>
</table>

3.1.2 Calculate the parking lot appraisal value

Drivers' evaluation value of parking lot is mainly through clustering analysis of parking behavior data and then assigning value to get. In this paper, k-mean clustering method is adopted to conduct clustering analysis on 3475 parking behavior data, which can be divided into 5 categories. The specific steps are as follows:

3.1.2.1 Standardization of parking behavior data

The "extreme value method" is used to standardize the data. After the data is standardized, the value is between 0 and 1, with the minimum value being 0 and the maximum value being 1. The normalized data is shown in Fig 5.

![Figure 5 Data standardization](source)

3.1.2.2 Cluster analysis

K-mean clustering method was used to divide the data into 5 categories, and the final clustering centers and clustering results were shown in Table 2 and Table 3.

TABLE 2 THE FINAL CLUSTERING CENTER

<table>
<thead>
<tr>
<th>clustering evaluation index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parking price</td>
<td>0.5246</td>
<td>0.1744</td>
<td>0.8381</td>
<td>0.1301</td>
<td>0.8306</td>
</tr>
<tr>
<td>Distance to destination</td>
<td>0.4696</td>
<td>0.4963</td>
<td>0.4994</td>
<td>0.5405</td>
<td>0.5066</td>
</tr>
<tr>
<td>With or without guide</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Parking time</td>
<td>0.5515</td>
<td>0.3641</td>
<td>0.6146</td>
<td>0.5987</td>
<td>0.3895</td>
</tr>
<tr>
<td>security</td>
<td>0.9707</td>
<td>0.5156</td>
<td>0.5154</td>
<td>0.2330</td>
<td>0.2446</td>
</tr>
</tbody>
</table>

TABLE 3 CASE NUMBER IN EACH CLUSTER

<table>
<thead>
<tr>
<th>clustering</th>
<th>Case number</th>
<th>number of effective</th>
<th>missing value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>877</td>
<td>3475</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>552</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.1.2.3 Determine the parking lot evaluation value

Combined with the actual situation of parking lots in each cluster, the five parking lots were assigned 1, 2, 3, 4 and 5 according to the evaluation value from low to high. The result of the assignment is the appraised value of the parking lot.

The data is divided into two parts, 2475 data as training set and 1000 data as test set. Base file database is established respectively, as shown in Fig 6, that is, data acquisition and collation are completed. The data in the three columns are driver number, parking lot number and parking lot appraisal value respectively.

![Figure 6 Evaluating value of the parking lot](image-url)

3.2 Recommendation system construction

The construction of parking lot recommendation system mainly includes three parts: ① collecting driver preference and standardized processing; ② Find a similar driver or parking lot; (3) Parking lot recommendation, collection of driver preferences and standardized processing have been completed in the early stage. This paper mainly studies the calculation of similar parking lots and the construction of recommendation system.

3.2.1 Calculate similar parking lots

To calculate the similarity of parking lots is to calculate the similarity between parking lots by taking the preference of all drivers to a certain parking lot as a vector, to get the similar parking lot of the parking lot, and make recommendations according to the historical choice of drivers. The similarity calculation method in this paper adopts the most commonly used cosine similarity method \(^{[14]}\), which is completed by editing program code in combination with Matlab software. The specific operation interface is shown in Fig 7.
By running the program, the similarity matrix of 140 parking lots is obtained, as shown in Fig 8. Parking lot names are replaced by numbers from 1 to 140. The closer the cosine is to 1, the closer the Angle is to 0, the more similar the two vectors are. It can be seen from the similarity matrix that the cosine similarity of parking lot 1 and parking lot 2 is 0.7523, indicating that the similarity of the two parking lots is high. If a driver has chosen parking lot 1, parking lot 2 can be recommended to him. When cosine is 0, parking lots are not similar. When the similarity of parking lot is high, the recommendation can be considered. However, whether to recommend the parking lot or not needs to be calculated later.

3.2.2 Construction of recommendation system based on parking lot

The similarity of parking lots has been studied before. On this basis, drivers' evaluation of parking lots is calculated and specific parking lot recommendations are made. By writing the program and using the code to establish the recommendation model, the specific process is as follows: loading the training set data, determining the model → loading the test set data, testing the model → determining the parking lot recommendation according to the driver's score on the parking lot.

3.2.2.1 To load the training set data and determine the model.

In this paper, there are 2475 training sets of data. The data files are imported into the program and run by Matlab to determine the parking lot recommendation model.
3.2.2.2 To load the test set data and test the model.

In this paper, there are 1000 training sets of data. The data files are imported into the program and run with Matlab to test the parking lot recommendation model and test and modify the recommendation ability of the model.

3.2.2.3 Parking lot recommendation is determined according to drivers' score on parking lot.

Grading procedures to write drivers for a parking lot, through the program is running, producing similar matrix (Fig 9) and the driver parking in the parking lot of 500×140 evaluation matrix (Fig 10), can be seen from the results, the number of 25, 128, 38, 11, 17 in the parking lot are similar, if a driver to choose one of them or a few parking lots, You can consider other parking lots for drivers to recommend. The recommendation shall be based on the score matrix shown in figure10. The one with a higher score can be recommended first.

![Figure 9 Matrix of the similar parking lot](image1)

![Figure 10 Score matrix of parking lot](image2)

As can be seen from the results, the number 124 drivers who choose parking 25, 38, 11, 128 and 17 on the parking lot can be recommended, according to the pilot rating matrix of the parking lot, pilot of the two parking lot score five points and three points, respectively, can according to the score of the high and low order, on the driver parking lot is recommended.

4. CONCLUSION

In this paper, collaborative filtering technology is used for in-depth analysis of parking behavior, parking recommendation procedures are written, and a parking recommendation system is established, which provides a basis for drivers to choose parking lots. The system can reduce the time of parking lot selection, thus alleviating parking congestion and solving parking problems. In future studies, the sample size can be increased to verify the model results.
REFERENCES

EFDet-SPP: Efficient anchor-free network for fine vehicle detection

Yongsheng Xie\textsuperscript{1,a}, Ming Ye\textsuperscript{*1,b}, Zhe Zhang\textsuperscript{1,c}, He Liu\textsuperscript{2,d}

\textsuperscript{1}College of Artificial Intelligence, Southwest University, Chongqing, China 400700; \textsuperscript{2}Chongqing Academy of Educational Science, Chongqing, China 400700
\textsuperscript{a}xieys0125@foxmail.com, \textsuperscript{b}Corresponding author: zmyxm@swu.edu.cn, \textsuperscript{c}zhangandzhe@foxmail.com, \textsuperscript{d}bjxiexintong@163.com

ABSTRACT

Existing vehicle detection methods lack the fine vehicle detection algorithm. In order to improve the accuracy and applicability of anchor-based object detection models, a novel and practical vehicle Fine-grained identification network (EFDet-SPP) based on the EfficientDet is proposed. The improved network adds a spatial pyramid pooling module (SPP) after the feature extraction network for concatenating features to enhance network learning capabilities, and multi-scale extraction of highly semantic features of images. Anchor-based predictions are converted to pixel-based predictions by combining FCOS's head network, eliminating the hyperparameters associated with anchor boxes. And with Mosaic, Copy-Paste data augmentation methods scale small object samples to achieve data sample balance. Experimental results show that the improved network has achieved 94.86% in the actual collected fine vehicle detection dataset, which is greatly improved compared with the EfficientDet network, and does not significantly increase the training parameters and calculation amount of the network.

Keywords: fine vehicle detection, computation efficiency, convolutional neural network

1. INTRODUCTION

The fine vehicle detection task, which aims to obtain the target fine vehicle components, plays a great role in the safety field of automobile manufacturers. The method of deep learning relies on the deep convolutional neural network to extract the high-level features of the image, which is often more accurate and effective. But being more accurate comes with the large size of the model and expensive computational cost. For example, the AmoebaNet-based NAS-FPN detector\textsuperscript{1} requires 167M (parameters) and 3045B (FLOPs) (30x more than RetinaNet\textsuperscript{2}) for state-of-the-art accuracy. More efficient detection models often sacrifice some accuracy and focus only on specific or small-scale resource requirements, but various real-world applications, from mobile devices to data centers, often require different resource constraints. Considering the balance of accuracy and efficiency, EfficientDet proposes a scalable detection architecture that seeks higher accuracy and efficiency in a wide range of resource constraints. While, it relies on a set of predefined anchor boxes like mainstream detectors such as Faster-RCNN\textsuperscript{3}, SSD\textsuperscript{4}, and the size, aspect ratio and number of anchor boxes have a huge impact on detection performance. FCOS\textsuperscript{5} has a relatively simple structure, which treats all pixels within an image as training samples and directly predicts the four distances of the bounding box. At the same time, it introduces a novel Center-ness score to suppress low-quality bounding boxes. Focusing on how to build an efficient and high-precision vehicle detection model, this paper has completed the key contributions of this work are as follows:

- Combining the EfficientDet and the FCOS modules, anchor-free Network (EFDet-SPP), is proposed to eliminate hyperparameters related to anchor boxes, improves the generalization ability and reduce computation enabling accurate efficient object detection.
- The spatial pyramid pooling module is added at the end of the backbone to perform multi-scale sampling on the feature map, concatenate features and effectively capture high semantic information.
- Establish two fine vehicle detection datasets (VLC, VDC) to ensure sufficient data, complex and universal scenarios, and use Mosaic and Copy-Paste data augmentation methods to balance the samples.
2. RELATED WORK

In the era of deep learning, object detection problems are usually modeled as the problem of classifying and regressing some candidate regions. YOLO\(^6\) was the first one-stage detector, these candidate regions are the anchors generated by the sliding window method. There is also EfficientDet\(^7\), a scalable and efficient one-stage detector. RCNN\(^8\) was the first two-stage detector, then Faster-RCNN introduced the Region Proposals Network (RPN), the candidate regions are the proposals generated by the RPN, but the RPN itself still classifies and regresses the anchors generated by the sliding window method. These methods all require anchor boxes, which lead to excessive hyperparameters and computing costs so that usually need to be carefully tuned to achieve good performance.

Nowadays anchor-free detectors attract more attention. These methods abandon anchor boxes and directly classify and localize objects using pixels as training samples. At the same time, it can be comparable to the anchor-based method in accuracy. The earliest YOLOv1\(^9\) treats object detection as a regression problem of spatially separated bounding boxes and associated class probabilities, which can directly predict bounding boxes and classification scores from the entire image. Then CenterNet\(^10\) directly predicts the center key-point of an object. RepPoints\(^11\) represented objects as a set of sample points and bounded them on a spatial extent for an object. Compared to the aforementioned detectors, FCOS had a relatively simple structure, which regarded all the pixels inside the images as training sample and predicted the four distances for a bounding box directly. Simultaneously, it introduced a novel "Center-ness" score to suppress the low-quality bounding boxes.

3. METHODOLOGY

The author proposes the fine detection network (EFDet-SPP) based on EfficientDet, which has been improved in data preprocessing, backbone network and head network, the overall network structure is shown in Figure 1.

3.1 Data pre-processing

Existing public datasets focus on vehicle detection, and there are almost no datasets for detecting fine parts of vehicles. Therefore, this paper collected a large number of images of vehicles and their parts from vehicle management offices and
inspection stations in different regions. According to the density and proportion of labels on the image, these images are divided into the Vehicle Large Component dataset (VLC) and the Vehicle Dense Component dataset (VDC). First, statistical observations are made for each type of label box data in the dataset as shown in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>area &lt; 32²</th>
<th>32² ≤ area ≤ 96²</th>
<th>area &gt; 96²</th>
<th>AP_{small}(Baseline)</th>
<th>mAP(Baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLC</td>
<td>1154</td>
<td>11562</td>
<td>48114</td>
<td>56.3</td>
<td>84.6</td>
</tr>
<tr>
<td>VDC</td>
<td>108579</td>
<td>82314</td>
<td>32141</td>
<td>43.5</td>
<td>74.9</td>
</tr>
</tbody>
</table>

It can be seen in Table 1 that the EfficientDet model does not perform well in the fine vehicle detection scene. The number of small object samples for VLC datasets is less than general datasets and most of the current models are for small objects is not good in detection effect. This results in a very small accuracy (AP_{small}) even model fails to converge. Therefore, this paper performed two methods of data augmentation, namely Mosaic and Copy-Paste (for small objects), on the basis of Mean-subtraction images to achieve the balance of data samples. In real-life application scenarios, the data is complex and changeable, and data augmentation can make the data samples more complex, thereby improving the robustness of the target detection model.

### 3.2 Spatial pyramid pooling

To obtain an efficient Fine vehicle detection network, this paper chooses a lower image resolution, but it will reduce the semantic information contained in each pixel, especially for small objects. This paper add the spatial pyramid pooling module (SPP) after backbone network from the EFDet-SPP. This performs multi-scale sampling for feature maps, concatenates features and effectively captures high semantic information. The feature map is spatially divided into \( m \times m \) fixed representations, where \( m \) can be 1, 3, 5, etc. respectively. Max pooling is then applied once per channel to form a spatially sized representation, which can then be further processed using BiFPN. The SPP module is designed to extract rich features at multiple scales, forming a feature map of the entire image in a single process. The SPP module is shown in Figure 2.

![SPP example](image)

Figure 2. Different Spatial Pyramid Pooling modules for feature maps.

The input image size of EFDet-SPP is 512×512, and five feature maps are obtained through a series of convolution pooling operations:

\[
P_1: 64 \times 64; P_2: 32 \times 32; P_3: 16 \times 16; P_4: 8 \times 8; P_5: 4 \times 4,
\]

As shown in Figure 2, the multiple convolution kernels used by the SPP module:

\[
P_4: 7 \times 7; P_5: 3 \times 3, 1 \times 1
\]

SPP module will eventually come from feature maps of different convolution kernel sizes are concatenated and output to the neck network (BiFPN).

### 3.3 Head network

It can be seen from Table 1 that EfficientDet’s mAP on VLC and VDC datasets is low, and the detection effect is not good. The reason is that it depends on pre-defined anchor boxes. The size, width ratio and quantity of the anchor boxes
will have a huge impact on the detection performance. And in different application scenarios, there are huge differences between datasets. This paper combines the EfficientDet module and the FCOS’s head network to eliminate hyperparameters related to anchor boxes and improve the generalization ability enabling accurate efficient object detection. It directly returns to each position \((x, y)\) corresponding to the original picture in the feature map. Each \((x, y)\) is used as a training sample. Similar to semantic segmentation, FCOS’s head network will train \(C\) binary classifiers and adds \(\text{Center-ness}\) loss to suppress the low-quality prediction bounding boxes due to far from the target center. \(\text{Center-ness}\) is defined as follows:

\[
\text{Center-ness} = \sqrt{\frac{\min(l^*, r^*) \times \min(t^*, b^*)}{\max(l^*, r^*) \times \max(t^*, b^*)}}
\]  

(3)

Where, \(l^*, r^*, t^*, b^*\) denote a pixel point to the vertical distance of the left boundary, right boundary, upper boundary and lower boundary of the bounding box, respectively.

4. EXPERIMENTS

All our experiments were done on a Nvidia RTX 2080Ti server using Pytorch. In addition, the processor is Intel Core i7-8700K, and the optimizer is Adamax. Finally, a Non-Maximum Suppression (NMS) with a threshold of 0.05 is applied to the results for post-processing. Our test indicators for accuracy are all mAP50.

4.1 Datasets and preprocessing

We established two datasets (VLC, VDC), a total of 89,092 pictures, of which 75,000 are used for training, 14092 is used for testing. Simultaneously, we ensure that the data is sufficient, the scene is rich and universal. Figure. 3 gives the two datasets with some band labels. It can be seen from Figure. 3 that the overall VLC dataset is a single label sample and the calibration box is large; the overall VDC dataset is multiple label samples and the calibration box is small.

![VLC and VDC datasets](image)

(a) VLC dataset  (b) VDC dataset

Figure. 3: Examples visualization on the datasets.

The author visualizes the data enhancement process of input images as shown in Figure. 4. For a three-channel image of \(512 \times 512\), the de-average value centralization is first performed. Then stitch it with other three images after random crop. In the end, the labels on the image are screened, and the small target sample label is copied and pasted to the random area that does not coincide with the original label area. In this way, the input image of EFDet-SPP is obtained.
4.2 Result analysis

The training and testing process of the EFDet-SPP network is carried out on the fine vehicle detection datasets (VLC and VDC). We randomly disrupted the two datasets, and then divided into training sets and test sets at a ratio of 8:2, respectively. Table 2 gives some important categories corresponding accuracy ($AP$) and average accuracy ($mAP$). On the whole, the accuracy of the identification of each category is high. The $mAP$ of the EFDet-SPP network on the two datasets is above 87%, and it can even be as high as 94.8% on the VLC dataset, which is greatly improved compared to the EfficientDet network. This paper studies a large number of vehicles and component sample data in different regions. The EFDet-SPP network proposed in this paper can accurately identify and judge vehicles and vehicle components, and efficiently carry out safety prevention and control work.

### Table 2. Performance with EFDet-SPP on the VLC and VDC datasets.

<table>
<thead>
<tr>
<th>Category</th>
<th>VLC</th>
<th>VDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seat belt</td>
<td>90.4</td>
<td>99.4</td>
</tr>
<tr>
<td>Driving recorder</td>
<td>1.0</td>
<td>93.1</td>
</tr>
<tr>
<td>Extinguisher</td>
<td>91.9</td>
<td>97.9</td>
</tr>
<tr>
<td>Frame number</td>
<td>95.8</td>
<td>99.3</td>
</tr>
<tr>
<td>Emergency hammer</td>
<td>95.8</td>
<td>81.8</td>
</tr>
<tr>
<td>mAP</td>
<td>94.8</td>
<td>87.5</td>
</tr>
</tbody>
</table>

For vehicle testing, at the same time of high accuracy, the size and calculation cost of the model are often what we need to consider. In order to further verify the performance of the EFDet-SPP network, under the same conditions, this paper calculates the model parameters, $FLOPs$, $mAP$ as shown in Table 3. It can be found that the parameter of the EfficientDet network (Baseline) EFDet and EFDet-SPP is 3.9M, 1.5M, 2.3M. Obviously, the improvement of the head network can greatly reduce the number of parameters and complex calculations of the network. The SPP module has significantly increased the parameters of the network. At the same time, the params and $FLOPs$ have not significantly improved, but the accuracy of the network is improved.

### Table 3. Performance comparison of proposed method with others on the VLC dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Params</th>
<th>$FLOPs$</th>
<th>$mAP$(VLC)</th>
<th>$mAP$(VDC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EfficientDet(baseline)</td>
<td>3.9M</td>
<td>2.5B</td>
<td>84.6</td>
<td>74.9</td>
</tr>
<tr>
<td>Vision Transformer($p=16$)</td>
<td>86M</td>
<td>-</td>
<td>83.3</td>
<td>72.8</td>
</tr>
<tr>
<td>RetinaNet-R50</td>
<td>34M</td>
<td>97B</td>
<td>80.6</td>
<td>70.1</td>
</tr>
<tr>
<td>EFDet (Ours)</td>
<td>1.5M</td>
<td>2.3B</td>
<td>89.2</td>
<td>83.6</td>
</tr>
<tr>
<td>EFDet-SPP (Ours)</td>
<td>2.3M</td>
<td>2.8B</td>
<td><strong>94.8</strong></td>
<td><strong>87.5</strong></td>
</tr>
</tbody>
</table>
4.3 Ablation experiments

In order to explore the effectiveness of data augmentation, the SPP module and head network on the fine vehicle detection network. At the same conditions, the author has experimented with EfficientDet, EFDet EFDet-SPP and EFDet-SPP (data augmentation) on the mAP and model parameters as shown in Figure. 5. It can be found that the SPP module has further improved the performance of the algorithm. Although EfficientDet has achieved excellent effects on the COCO dataset, because of the huge differences between the bounding box between the datasets, it cannot play a better performance in fine vehicle detection. FCOS's head network significantly increases the model's applicability in fine vehicle detection scenarios and at the same time, data augmentation slightly improves the detection accuracy. The results of Table 2, 3, and Figure. 5 show that EFDet-SPP is an effective module to distinguish the fine components of the vehicle.

5. CONCLUSION

In this paper, we propose an efficient anchor-free network framework called EFDet-SPP for high-accuracy and real-time fine vehicle detection. Specifically, we combine the EfficientDet module and the FCOS module, so that the anchor-based prediction is transformed into pixel-based prediction, eliminating the hyperparameter related to the anchor boxes. And EFDet-SPP add the spatial pyramid pooling module to solve the problem of insufficient information brought about by the smaller input size. At the same time, we collected a large number of vehicle component pictures in various regions and established VLC, VDC datasets. Considering the sample imbalance problem, we have made two data Augmentations: Mosaic and Copy-Paste. Extensive experiments conducted on VLC and VDC datasets indicated that EFDet-SPP can achieve the balance between accuracy and computation efficiency.

REFERENCES


Towards Building Long-Range Relationship for Super-Resolution

Zifan Wang
XD University, Xi’an, China
*zfwang_1@stu.xidian.edu.cn

ABSTRACT

A new network super-resolution algorithm is proposed in the light of multi-feature fusion to surmount: the problem of manual feature extraction as well as low information throughput in traditional super-resolution algorithms. The algorithm adopts a learning method of the indirect mapping from the hazing image to the clear super-resolution image, which calculates the similarity between the point and other points, normalizes the similarity to obtain the weight between each point, and then multiply it by the feature map value of the corresponding point and use the similarity. In the algorithm, utilizing multi-scale feature fusion enables to rebuild details of the image. Different depths of field combine the contour information obtained by shallow convolution with the detail information obtained by deep convolution to enhance the overall super-resolution’s effect. The experimental results and related data show that the channel throughput has been significantly improved, and compared with other algorithms, it has better detail information and contrast, which provides a new idea for dehazing methods.

Keywords- super-resolution; non-local; multi-scale feature fusion

1. Introduction

For every worker who studies the image, the super-resolution algorithm is used in some way such as figuring out some related research puzzles of image processing (especially when it comes to the partial enlargement of a specific part of the image in image processing).

When talking about super-resolution algorithms, we often conduct in-depth discussions and qualitative research in these two aspects. One is the effective fusion of traditional image algorithms and image sharpening techniques: the excellence of this method is that it stands on the shoulders of giants and looks into the distance, because the traditional resize method has been polished over a long period of time. It can be said to be very mature [4]. As long as the model with the best adaptability is selected for a specific depth of field, it can achieve twice the result with half the effort and achieve the goal efficiently. Unfortunately these methods often produce blurry results that lack high frequency detail. In contemporary times, deep learning is in full swing, especially the super-resolution based on neural network has gradually become the core of the field. Today, there are many examples of using neural networks to super-resolution images, and the results are amazing. For example, the papers published by Saeed Anwar, Whihao Wan, etc., provide us with examples of applying neural networks [1][2]. Upscale the input image using bilateral interpolation and then use a simple convolutional network to form a nonlinear mapping from LR input to super-resolution output [3]. Compared with the traditional convolution structure, this method achieves a higher PSNR value; The second is to combine extremely large data with deep learning to effectively train the algorithm. This paper conceive a radically different approach established on quantization coding technique as well as regression separation theory based on the concept of PSNR (a general indicator utilized to measure the reconstruction quality of lossy compression codecs).

2. Preliminary

2.1 Non-local

For measure effectively the pixels that have some similar structures to some extent, and then determining the weight coefficient [7]. The specific steps of this method are divided into the following: for the weight of each pixel, the image sub-block (7*7) or (9*9) centered on the pixel is used to The Gaussian weighted Euclidean distance between sub-blocks centered on the current pixel is calculated. Non-local network is a neutralized non-local mean filter, and its self-attention mechanism is embodied in that it is not only a graphical model used to represent the relationship between pixels at different locations or at different times, but also a feed-forward model for sequence analysis. The NON-LOCAL module can directly calculate the long-range dependence between two positions, and its formula is:
In the formula of this model, I assume that the input signal is $x$ and the output signal is $y$, where the weight coefficient is represented by $f(x_i, x_j)$, and $C(x)$ is the normalization parameter [5].

2.2 SRCNN

On the basis of the bicubic interpolation algorithm, it is worthwhile to scale the LR image to obtain the processed LR image, interpreted as a pre-processing process. For example: the input processed low-resolution image, the three convolutional layers are: convolutional layer 1: kernel Size is $8 \times 8$, convolutional layer 2: kernel Size is $2 \times 2$, convolutional layer 3: kernel Size is $5 \times 5$. In the training process of SRCNN, the MSE goes by the name of the loss function, and the image is not filled. During these process, in certain circumstances, those parameters are supposed to set artificially to 0.0001. Besides the learning rate parameter of the third layer can then be assumed to be an exact value of 0.00001 to distinguish it from the first two groups [6].

2.3 CONV module

In general, when using the convolutional neural network to obtain the global receptive field to capture image features hierarchically, if the convolutional layer and the nonlinear layer (an essential part of the Convolutional Neural Network) are missing, this CNN network can be said to be incomplete.

3. Test Method

In this paper, I combine the SE module and the CNN network. Let the network adaptively calibrate the channel importance. Along with the operation of the Squeeze operation, the GSI is compressed into the form of a channel descriptor, besides using the global average pool to count the channel information and then the two-dimensional feature information is turned into real numbers. Then a new statistic based on log-likelihood is proposed. By using the kernel function estimation method, any density function can be asymptotically converged under the condition of complete sampling, that is, the density of any distribution data can be estimated. The basic form of the vector of MEANSHIFT at point $x$ is defined as:

$$M_h = \frac{1}{K} \sum_{x_i \in S_k} (x_i - x)$$

As shown in Figure 1, we can see that the network mainly includes several parts.

First of all, in order to enable feature-dependent learning and feature representation to be better implemented in applications, we refer to SAN [9] to help us in this task. A new trainable second-order channel attention module is added to the code for relevant learning and the SE module is purposely added before multifarious operations such as data compression as well as GAP using deep convolution In the deep convolution module; the subsequent SE operation makes all the network parameters significantly improved, the total number of basic data obtained is increased, and the image’s accuracy and the target detection rate are enriched. In addition, the NON-Local module added to it makes the whole a Non-locally Enhanced Residual Group (NLRG). Based on this special structure, the convolutional neural network can run faster and obtain more...
accurate data within a limited time. SSRG consists of G-LSRAG and SSC [10]. After making the above improvements, we evaluate our model on a publicly available dataset.

4. Test Results and Discussions

4.1 Experimental configuration
The experiment is completed in the python environment, integrating non-local network, mean shift, super-resolution, SRCNN and other methods.

4.2 Experimental data
The experimental data in this paper use the foggy image library established by clear images and the foggy images collected in natural scenes. Three images are selected for the experiment

4.3 Analysis of the results of the dehazing algorithm

![Comparison of different resolutions of the same image.](image)

Fig. 2 Comparison of different resolutions of the same image.
With the continuous iterations of training in this paper, the loss function decreases and it appears to converge in the image, which is not difficult to prove that the model used in this paper has been trained.

Comparing the image effects, first of all, the pictures do not produce obvious Halo effect, but the subjective effect is that the colour of the SR picture is fuller and the visual sense is good, while the LR is overall blurry. The image after dehazing by the algorithm in this paper has no contrast, brightness shift phenomenon, and is relatively clear, the colour saturation is also improved, and the overall operating efficiency of the code is also higher.

5. Conclusion

The previous image dehazing algorithms are mainly divided into the following modules: establishing an atmospheric physical model, introducing numerical parameters, and refining the throw rate. However, the dehazing algorithm based on the traditional mathematical model is obviously difficult to apply to more variable and complex depth-of-field
environments, and the throughput of the algorithm network is difficult to break through its upper limit. This paper proposes a mode idea combining super-resolution and parallel communication serial port.

The results of experimental obviously show that this algorithm has a certain ability to improve the image processing speed by the algorithm network, and can be better applied to different depths of field.

For better dehazing effect, subsequent research directions will focus on improving the problem of local neighborhood construction or processing dynamic image control sequences more efficiently.

References

Design and Analysis of OAM-DM Underwater Wireless Optical Communication System

Jingyu WANG*a, Shaohua Zhou, Juan Li, Chengjin Wang
*aCollege of information and communication, National University of Defense Technology, Xi’an 710106 China.
*Corresponding author: wangjingyu@nudt.edu.cn

ABSTRACT

As the main way of underwater data transmission, acoustic communication is still limited by the low-level signal-to-noise ratio and channel capacity. The orbital angular momentum (OAM) based wireless optical communication provides a new dimension to data transmission with an expanded channel capacity. We introduce the basic OAM multiplexing communication system, analyze the basic principle of its transmission process, and verify the feasibility of underwater vortex optical communication by simulation. On this basis, the influence of transmission distance and underwater turbulence on the communication system is studied. Our results show that the transmission performance of the system is better in the transmission range of 30 m, and the influence of turbulence on the system performance increases gradually with the increase of signal-to-noise ratio.

Keywords: orbital angular momentum, oceanic turbulence, underwater wireless optical communication

1. INTRODUCTION

In recent years, with the continuous exploration and development of the ocean, effective communication means are extremely important tools. How to transmit information rapidly and reliably under the complex underwater communication channel has also become a hot topic of current research. Underwater wireless optical communication technology has many advantages such as low transmission delay, high transmission rate and point-to-point communication security, which reflects great application potential and is an important development direction of underwater wireless communication in the future. The existing research has found that visible light with blue-green wavelength has low transmission loss in such an underwater channel as the ocean, which can be applied to underwater communication, and underwater wireless optical communication has been gradually developed[1]. With the continuous development of economy and technology, there is a greater demand for underwater information communication. A new wireless optical communication system that can adapt to the complex underwater communication environment is the key to solve this problem[2]. Orbital Angular Momentum (OAM) is a basic physical property of light propagation, which is related to the spatial distribution of light phase and is a new available electromagnetic capacity resource[3]. The underwater wireless optical communication technology using orbital angular momentum for information transmission is a new type of optical communication system. It mainly transmits information through the underwater propagation of vortex beams with orbital angular momentum as the carrier of communication. The information it carries is mainly based on the degree of freedom of OAM in the spatial transmission of optical waves, which can greatly improve the transmission capacity and rate of the channel. At present, there are mainly three ways to use OAM for wireless optical communication: OAM keying (OAM-SK), OAM multiplexing (OAM-DM) and OAM multicast (OAM-MC)[4]. Compared with the traditional wireless optical communication technology, these three communication methods can better improve the spectral efficiency, transmission capacity and security performance of the system.

In 2010, Berkhout et al. proposed a separation scheme of vortex beams based on coordinate transformation for simultaneous separation and detection of multi-OAM modes[5]. In 2013, Mohammad Mirhosseini et al. improved the combination of coordinate transformation-based OAM separation with Berkhout fan-out and coordinate transformation systems[6]. In 2016, Ren et al. confirmed through experiments that under the weak turbulence of the actual atmosphere, the 400 Gbit / s free space optics (FSO) link of OAM multiplexing can be realized within 120m[7]. In 2016, Baghdady et al. used the air separation multiplexing technology to prove the possibility of realizing 2.5 Gbit / s OAM orthogonal multiplexing link with transmission distance of 2.96 m in different channel environment models from pure seawater to coastal ocean[8]. In the same year, they used the SDM technology to reuse the two OAM modes and transmit them in a 2.96 m non-turbulent flume.
realizing the underwater optical communication at the 3 Gbit/s information acquisition rate[9]. In 2017, Huang et al. used the Monte Carlo method to numerically simulate the diffusion of vortex beams caused by seawater turbulence, and compared the effects of seawater type, turbidity, beam width and divergence angle on the received light intensity[10]. In 2018, Yin et al. from Beijing University of Posts and Telecommunications proposed an ocean turbulence mitigation scheme for OAM multiplexing communication system based on spatial diversity technology[11]. The research results show that the error rate increases with the increase of transmission distance and turbulence intensity. At the same time, it proves that the diversity merging technology can effectively alleviate the influence of turbulence on signal transmission. In 2020, Yin Xiaoli and others from Beijing University of Posts and Telecommunications applied the space-time coding technology to the OAM multiplexing communication system[12]. The simulation proved that the use of this technology can effectively alleviate the impact of ocean turbulence on the system bit error rate.

To sum up, the development potential of underwater wireless optical communication using vortex beam with OAM as carrier is very great, but there are many limiting factors for communication in underwater channel, especially the influence of ocean turbulence on vortex beam propagation. The above research methods are mainly to design the OAM multiplexing communication scheme from the system level by building an experimental platform or simulating by software, but the research on the influence of OAM multiplexing mode on the performance of communication system is not deep enough. Therefore, the better application of OAM technology in underwater communication system has important theoretical and practical significance for solving the problems faced by underwater communication.

2. WIRELESS OPTICAL COMMUNICATION SCHEME BASED ON OAM MULTIPLEXING

In 2012, Wang Jian et al. proposed a high-speed communication mode based on orbital angular momentum reuse[13]. In this way, LG beams with different topological charge values are used as the carriers of multiplexing signals, and the required transmitted information is loaded onto the beam through modulation, and then transmitted to the receiving end after multiplexing. This method can greatly improve the transmission capacity of the system and the utilization rate of the frequency band, and achieve high-speed data transmission.

Figure 1 is the basic principle diagram of OAM multiplexing communication. In the figure, we can see that first, the laser at the transmitter generates a Gaussian beam, which is divided into N beams after passing through the optical beam splitter, and each beam is modulated with the information that needs to be transmitted. Then, the N beams are loaded by SLM with different holograms. At this time, the original Gaussian beam becomes an N-channel vortex beam loaded with information (LG beam is used), and then the N-channel LG beam with different modal values becomes a multiplexing beam after passing through the optical coupler and is sent to the transmission channel. At the receiving end, the received multiplexed state light with information is also divided into N-channel beams through the optical beam splitter. The N-channel beams are loaded with a spatial light modulator with a reverse hologram. Each spatial light modulator corresponds to a LG beam with a modal value. It can demultiplex the corresponding LG beam from the multiplexed state, and then recover the information carried by the demodulation detection.

Figure 1. Basic principle diagram of OAM multiplexing communication
Next, we analyze the basic principles and main processes of each component of the scheme.

2.1 Generation and modulation of vortex beams

In the OAM multiplexing optical communication system, the first problem is how to modulate the information to be transmitted to the multiplexing OAM beams. This process can be divided into two steps: first, the information to be transmitted is modulated to the Gaussian beam generated by the laser source; the second step is to convert the Gaussian light loaded with information into vortex beam by optical modulator. In this process, the information to be transmitted is the modulation signal, and the vortex beam is the carrier used in the communication. Optical waves have the general properties of electromagnetic waves, so we can process optical waves like amplitude modulation, frequency modulation and phase modulation in wireless communication, so that the information needed to be transmitted can be effectively loaded on the beam. By comparison, we found that in the OAM multiplexing communication system, the use of amplitude to modulate the vortex beam is easier to achieve and the transmission effect is better, which is more suitable.

Amplitude modulation refers to the relationship between the amplitude of the carrier and the modulation signal, and there is a regular change. The principle analysis of vortex beam modulation process is as follows:

The first information to be transmitted is modulated to the Gaussian beam generated by the light source. The electric field intensity of the Gaussian beam is expressed as:

\[ U(t) = A_c \cos(\omega_c t + \varphi_c) \]  

(1)

Where \( A_c \) represents the amplitude of the light wave, \( \omega_c \) represents the angular frequency, and \( \varphi_c \) represents the phase of the light wave.

Assuming that the modulation signal is \( s(t) \), the Gaussian light expression loaded with the modulation signal is:

\[ U'(t) = A_c s(t) \cos(\omega_c t + \varphi_c) \]  

(2)

Next, we convert Gaussian light into vortex light, which needs to be used in spatial light modulator. The vortex beam expression is:

\[ U_{OAM}(t) = s(t) U(t) \exp(i l \theta) \]  

(3)

Where \( \exp(i l \theta) \) is the phase factor formed by the converted Gaussian light and \( l \) is the topological charge of the vortex beam.

2.2 Multiplexing of vortex beams

By modulation we have loaded the information we need to transmit to the vortex beam, and the generated vortex beam needs to be multiplexed to send out. At this time, the expression of a certain vortex beam is

\[ U_{OAM}(r, t) = s_i(t) U(r) \exp(i l \theta) \]  

(4)

Because the topological charges of each vortex light are different, so they meet the orthogonal relationship. The beam expression after multiplexing is

\[ U_{MUX}(r, \theta, t) = \sum_{i=1}^{N} s_i(t) U(r) \exp(i l \theta) \]  

(5)

After passing through the underwater channel, the vortex beam of the multiplexing state receives interference, and the intensity and phase are disturbed, which can be expressed as:

\[ U_{MUX}^{Rx}(r, \theta, t) = \sum_{i=1}^{N} s'_i(t) U'(r) \exp(i l \theta) \]  

(6)

\( s'_i(t) \) and \( U'(r) \) are the expressions of signal and light field after \( s_i(t) \) and \( U(r) \) are affected by underwater turbulence, respectively.

2.3 Demultiplexing of vortex beam

When the receiving end receives the transmitted signal, it is first necessary to demultiplex it. Taking one of the signals as an example, the received multiplexing vortex beam can be obtained through the spatial light modulator opposite to the modulation hologram. The expression is as follows:
\begin{align}
x_p(t) &= U_{\text{det}}(r, \theta, t)U_p(r)\exp\left(-i\ell \theta \right)drd\theta \\
&= s_p(t)U'(r)U_p(r) + \int \sum_{p=1}^{\infty} s_p(t)U'(r)U_p(r)\exp(i\ell \theta - i\ell \theta) \\
&= s_p(t)U'(r)U_p(r)
\end{align}

Then the information can be recovered by demodulation and correlation detection of the demultiplexed signal. Finally, the information transmitted can be obtained by integrating the information detected by each channel.

3. SIMULATION RESULTS AND ANALYSIS

Without considering the influence of underwater communication channel noise and ocean turbulence, the simulation analysis of the OAM multiplexing communication system in the previous section is carried out. Three multiplexed LG beams with topological charge $l_1 = 1, l_2 = 3, l_3 = 5$ are set, and the information carried by the three beams is 1.

Fig. 2 is the simulation of spatial distribution of superposition beam in three-way multiplexed communication system. It can be seen that when each multiplexing beam is coherent, due to the existence of interference, the phase and light intensity distribution after superposition are changed compared with the single beam.

Fig. 3 is the comparison of signal energy distribution in receiver of OAM multiplexed communication system (adding noise and turbulence).
Fig. 4 shows the energy distribution comparison of the multiplexing branches at the receiving end after the underwater noise and turbulence are added, and the information at the sending end remains unchanged. At this time, the energy distribution of the receiving end has also changed. When the energy of the beam modulated with the information ‘0’ is not zero at the receiving end, it indicates that the underwater noise and turbulence affect the intensity and phase information of the beam transmission, resulting in errors in the information after demultiplexing and demodulation.

From the above simulation results, we can find that the scheme is feasible in theory and can be used as the basic scheme of underwater optical communication system design.

4. OAM-DM UNDERWATER WIRELESS OPTICAL COMMUNICATION SYSTEM SCHEME

4.1 Efficient separation method of orbital angular momentum states

In 2010, Martin P. J. Lavery et al. proposed a new method for efficiently separating OAM states [5]. This method can separate OAM beams with different modal values at the same time, and the difficulty of implementation is lower than other methods, which is suitable for OAM multiplexing communication system. Fig. 5 shows the schematic diagram of this orbital angular momentum state separation method based on coordinate transformation-position decision.

![Schematic diagram of OAM state separation based on coordinate transformation-position decision](image)

The SLM in Figure 5 is represented as a spatial light modulator, where the SLM _ 1 function is to generate LG light; the spatial light modulator SLM _ 2 is used for coordinate transformation of the light field. The SLM _ 3 spatial light modulator is used to correct the phase of the light field; L _ 1 is Fourier transform lens; L _ 2 is used to focus light; the function of CCD detector is to convert the received optical image into digital electrical signal which can be processed.

By studying the transmission process of a certain LG beam in the multiplexing communication system, a specific theoretical analysis is given.
(1) First, the laser source generates a Gaussian beam, and then passes through a spatial light modulator $SLM_1$ with a phase mask of $\exp(i\theta)$ to obtain a LG light with a topological charge of $l$, that is,

$$U^{LG}(x, y) = A(x, y) \cdot \exp (il\theta)$$

(8)

(2) Then, let LG light through a phase mask $\phi_1(x, y)$ on $SLM_2$, after the light wave will pass through the lens $L_1$, it is equivalent to a Fourier transform of the light wave processing. Among them.

$$\phi_1(x, y) = \frac{2\pi a}{\lambda f} \left[ \arctan \left( \frac{y}{x} \right) - x \ln \left( \frac{\sqrt{x^2+y^2}}{b} \right) + x \right]$$

(9)

$a$, $b$ represents the expansion (compression) factor of a transformation from $(x, y)$ to $(u, v)$.

(3) The light wave then passes through the phase mask which is $\phi_2(u, v)$ on $SLM_3$, thereby reducing the phase change caused by the aforementioned coordinate transformation. Among them.

$$\phi_2(u, v) = -\frac{2\pi ab}{\lambda f} \exp \left( -\frac{u}{a} \right) \cos \left( \frac{v}{a} \right)$$

(10)

Through Formula 9 and Formula 10, we complete the coordinate transformation from $(x, y)$ to $(u, v)$, that is:

$$u = -a \ln \left( \frac{\sqrt{x^2+y^2}}{b} \right), \quad v = -aa \arctan \left( \frac{y}{x} \right)$$

(11)

(4) Finally, the light wave passes through the lens $L_2$, and the light wave will be focused on the CCD detector. In the end, a rectangular light wave is generated, and its abscissa is

$$u' = \frac{\lambda l}{2\pi a}$$

(12)

That is to say, LG beams with different $l$ values display different positions on the CCD detection screen, and there is a one-to-one correspondence. Therefore, we can distinguish the modal value of the vortex beam by the displayed position, so as to achieve the separation effect.

4.2 OAM state multiplexing underwater optical communication scheme based on efficient separation method

Figure 6 is a schematic diagram of OAM multiplexing underwater wireless optical communication system based on efficient separation method. The laser at the transmitter first produces a path of Gaussian light, which becomes N paths after passing through the optical beam splitter, and the information that needs to be transmitted is also divided into N groups after processing, each group corresponds to a path of light, and then through OOK modulation, that is, amplitude modulation. The N sets of information are loaded onto the N paths of Gaussian light, and then the modulated N paths of Gaussian light pass through N spatial light modulators loaded with different modulus holograms. At this time, the beam becomes an N-path LG beam carrying information, and these beams are combined into one path through the optical coupler at the transmitter. After the transmission process in the underwater channel, the beam is affected by turbulence and other factors to change. At the receiving end, the LG beams with different modal values are separated by the method of coordinate change, and they are focused on different lateral positions of the detection screen, so that we can get the information carried by the LG beams with different modes.
At the receiving end of the signal, the multiplexed LG beams can be separated according to this characteristic, because the LG beams with different topological charges eventually show different abscissa positions on the detector. At the same time, the modulation method we use at the transmitter is OOK modulation, so the information sent will only affect the amplitude of the light wave, that is, the signal is only related to the light intensity, so the intensity of the light wave at different positions on the detector can represent the content of the starting signal, only through reasonable design and analysis, can effectively recover the information.

5. PERFORMANCE SIMULATION OF COMMUNICATION SYSTEM

After the multiplexing vortex beam passes through the underwater channel, due to the influence of turbulence, channel attenuation and noise, the amplitude and phase of the optical signal obtained at the receiving end are changed to some extent. At this point, the energy between each OAM mode will appear mutual ‘diffusion’ phenomenon, which is also called mode crosstalk. Mode crosstalk will affect the demultiplexing process of the signal at the receiver, which makes the final detected signal appear error code. At the same time, the change of amplitude will also affect the process of signal judgment at the receiving end, and also reduce the system performance. Considering the influence of underwater channel on OAM beam, we analyze the influence of underwater channel on the system from two aspects of propagation distance and turbulence intensity.

5.1 Influence of underwater transmission distance on system performance

The attenuation of optical signal in underwater channel is fast, and the power of received signal decreases to a low level after a short transmission distance, so the transmission distance is an important factor restricting the performance of underwater communication system. At the same time, the transmission distance itself is also an aspect of evaluating system performance. Therefore, it is necessary to study the relationship between transmission distance and system performance, and analyze the signal transmission range that the system can better communicate. In this section, the average bit error rate of the receiver is obtained by simulating the system under different transmission distances, the results are analyzed and concluded.

The system parameters are set as follows: the wavelength of light is \( \lambda = 632 \) nm, and three OAM multiplexing states are adopted. The topological charge values are \( l_1 = 1, l_2 = 3, l_3 = 5 \), respectively. The underwater channel adopts medium-intensity turbulence (kinetic energy dissipation rate \( \epsilon = 10^{-5} \)), Kolmogorov length size \( \eta = 10^{-4} \), temperature difference dissipation rate \( \gamma_T = 10^{-7} \), underwater power attenuation coefficient is 1.25 dB / m, input SNR is 12 dB, and the range of transmission distance is \( z = 2\text{~to~}50\text{~m} \). In the multiplexing communication system, due to certain differences among the signals, the transmission performance of each signal may also be not completely consistent. Therefore, we calculate the error rate of the overall transmission data, which can be regarded as the average error rate of each multiplexing signal. In this paper, the concept of average error rate is also used in the part of system performance evaluation. The bit error rate is defined as the
ratio of the number of error code elements in the recovered information demodulated by the receiver to the total number of code elements. Can be expressed as:

\[ P_e = \frac{N_e}{N} \]  

(8)

where \( N_e \) denotes the number of error elements, and \( N \) is the number of total code elements sent.

From the theoretical analysis, the influence of transmission distance on the bit error rate is mainly due to the strong attenuation of underwater. When the transmission distance increases, the optical power of the vortex beam becomes weak, and the influence of underwater noise and marine turbulence on it is also greater. The signal received by the receiver changes more greatly compared with the transmitter, and the bit error rate of the recovered information after demodulation is correspondingly greater.

Figure 7 shows the trend of average bit error rate with transmission distance. Obviously, with the increase of transmission distance, the average bit error rate of the system increases gradually. When the transmission distance increases from 2m to 50m, the average bit error rate of the system increases from below \( 10^{-3} \) to above \( 10^{-1} \), and the overall change is large. At the same time, it can be seen that when the transmission distance is small, the average bit error rate changes slowly, and it increases linearly in the logarithmic coordinate system. When the transmission distance reaches about 35m, with the increase of distance, the bit error rate increases faster, and the system performance has been at a low level. From the above analysis, it can be concluded that the transmission distance is an important factor affecting the performance of OAM multiplexing wireless optical communication system. With the increase of transmission distance, the system performance deteriorates faster. In the transmission range of 30m, the transmission performance of the system is good.

![Figure 7. The trend of average bit error rate with transmission distance](image-url)

### 5.2 Effect of Turbulence on System Performance

The crosstalk between vortex beam modes caused by ocean turbulence is one of the important factors of error code generation. By simulating the influence of turbulence intensity on system performance, the availability of the system under different channel conditions can be explored, and the limit conditions of system application can be studied, which provides some help for practical application.

Due to the large range of changes in the intensity of underwater turbulence, if the simulation is carried out for each case, the amount of data is extremely large. Therefore, we simulate by setting three levels of turbulence intensity as the focus. The system simulates the channel under these three different turbulent conditions, and obtains the average bit error rate size distribution in turn. The weak turbulence parameters are set as follows: kinetic energy dissipation rate \( \varepsilon = 10^{-5}m^2/s^3 \), Kolmogorov length size \( \eta = 10^{-4}m \), and temperature difference dissipation rate \( \chi_T = 10^{-7}K^2/s \); the medium turbulence parameters are set as follows: kinetic energy dissipation rate \( \varepsilon = 10^{-5}m^2/s^3 \), Kolmogorov length size \( \eta = 10^{-4}m \), and temperature difference dissipation rate \( \chi_T = 10^{-7} \); strong turbulence parameters are: kinetic energy dissipation rate \( \varepsilon = 10^{-5}m^2/s^3 \), Kolmogorov length size \( \eta = 10^{-4}m \), and temperature difference dissipation rate \( \chi_T = 10^{-7}K^2/s \).
The influence of turbulence on the system's performance is mainly due to its disturbance to the phase of the vortex beam propagation. When the beam is demultiplexed at the receiving end, the crosstalk between modes is caused. This makes the multiplexed signals interfere with each other, and bit errors occur during the final demodulation recovery. Therefore, when the turbulence becomes larger, the bit error rate of the system will further increase.

Figure 8. The curve of average bit error rate with signal-to-noise ratio under three turbulence intensities

Figure 8 shows the variation curve of the average bit error rate of the system with the input signal-to-noise ratio under three different intensities of turbulence.

From the impact of turbulence on the system, the greater the turbulence intensity, the greater the average bit error rate of the system. As the input SNR increases, the bit error rate gap between different turbulence levels also increases. Through the analysis, it can be concluded that underwater turbulence is an important factor affecting the performance of the system. With the increase of the signal-to-noise ratio, the influence of turbulence on the performance of the system also increases gradually. At the same time, it also shows that the simulation method we adopted can effectively simulate the influence of turbulence on the underwater wireless optical communication system, which is beneficial to further research and analysis.

From the overall performance of the system, the underwater channel has a large interference to the signal. When the signal-to-noise ratio is low, the average bit error rate is large, and it is at a high level and the variation range is small. When the signal-to-noise ratio (SNR) is greater than 9dB, the bit error rate (BER) decreases obviously. When \( BER = 10^{-3} \) is used as the acceptable level of BER, the required SNR is 15dB in strong turbulence, 14dB in moderate turbulence and 12dB in weak turbulence. Compared with the OAM communication system for signal transmission in the atmosphere\(^{[14]}\). On the premise of the same OOK modulation, the signal to noise ratio is only about 8dB when \( BER = 10^{-3} \). It can be seen that the signal-to-noise ratio required for underwater and atmospheric transmission is about 4~7dB. We can analyze that the underwater wireless optical communication system requires higher signal-to-noise ratio and requires greater transmission power to ensure system performance.

6. CONCLUSIONS

In this paper, we analyze the feasibility of underwater communication using OAM beams, and on this basis, construct the model of underwater optical communication system. Through the simulation and analysis of its communication performance, we demonstrate the influence of the transmission distance of OAM-based underwater optical communication and the communication performance of ocean turbulence, respectively. The results show that the transmission performance of the system is good within the transmission range of 30 m; the underwater wireless optical communication system has higher requirements for signal-to-noise ratio, and requires greater transmission power to ensure system performance. At the same time, with the increase of signal to noise ratio, the influence of turbulence on system performance is also increasing.
REFERENCES

MFCSA-CAT: A Multimodal Fusion Method for Cancer Survival Analysis Based on Cross-Attention Transformer

Shenyang Deng\textsuperscript{a}, Yuanchi Suo\textsuperscript{a}, Shicong Liu\textsuperscript{a}, Xin Ma\textsuperscript{a}, Hao Chen\textsuperscript{a}, Xiaoqi Liao\textsuperscript{a}, Jianjun Zhang\textsuperscript{b}, Wing W.Y. Ng\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a}School of Mathematics, South China University of Technology, Guangzhou, Guangdong, 510641
\textsuperscript{b}School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, 510006

* Corresponding author: wingng@scut.edu.cn

ABSTRACT

Cancer diagnosis, prognosis, and therapeutic response predictions are based on data from various modalities, such as histology slides and molecular profiles from genomics data. In cancer clinical treatment, the technology of intelligent diagnosis for cancer patients has become an essential research domain with the rapid growth of various pathological data. In this work, we propose a multimodal fusion method for cancer survival analysis based on Cross-Attention Transformer. Compared to similar bimodal work, our work greatly reduces the number of parameters in the feature fusion model (our fusion model has 7625 parameters), and achieves the State-of-the-Art effect (81.85\%) in bimodal cancer survival analysis task with histology images and genomic features data of Glioma cancer from TCGA database. (Previous bimodal Sota work in this task is Kronecker Product which achieves 81.40\% with 170130 parameters) In addition, our experiments show that Cross-Attention can not only increase the correlation between the two modalities but also offer a better bimodal feature representation for the final fusion.

Keywords: Multimodal Fusion, Cross-Attention, Transformer, Survival Analysis

1. INTRODUCTION

Cancer diagnosis is a difficult task for pathologists. Due to the fact that understanding patients' current condition as well as giving a better prognosis requires pathologists to go through hundreds of histopathological images, cancer diagnosis inevitably takes a considerable amount of time. Even among experts, this method still exits a lot of subjectivity. It is found that the result of qualitative inspection of human pathologists has significant differences intra- and interobservers due to this subjectivity [1]. Moreover, the genomic analysis of tissue biopsies is not able to precisely separate tumor-induced genotypic measures and changes from those of non-tumor entities [2]. Modern sequencing technologies, spatial transcriptomics and multiplexed immunofluorescence techniques help extract genetic information, but they lack clinical penetration.

Artificial intelligence (AI) has recently shown satisfying results in many applications, including the medical field. As modern medical imaging technology becomes more and more mature, researchers can digitize tissues on glass slides into an image form [3]. It creates a critical opportunity for AI approaches in joint image-omic assays.

However, analyzing image data is far from sufficient. The data in cancer diagnosis, prognosis and therapeutic response prediction are usually heterogeneous, such as clinical data, histologic slides and molecular profiles. What's more, oncologists need both qualitative information from histology and genomics data to predict clinical outcomes [4]. It is shown that information from spatial organizations and community structures of cells has diagnostic and prognostic relevance [5][6]. However, most histologic analyses do not provide genomic information or explicitly combine these kinds of information. Therefore, fusing morphological information from histology and molecular information from genomics is a sensible idea for quantifying the tumor micro-environment in analysis.

In recent years, many multimodal fusion strategies have been proposed to solve the problems above. Effective as they are, some of these strategies just simply concatenate heterogeneous data and then apply machine learning [3], and they potentially lack sufficient multimodal fusion to overcome the heterogeneity of pathological images and genomics data,
which causes excessive redundancy of parameters. In this work, we propose a novel fusion strategy named Cross-Attention Concatenation, which is better to combine these heterogeneous data. It mainly uses the Cross-Attention Transformer to conduct exchange learning on the feature information from various modalities in a cross-flow manner, and then sends the extracted features to the Transformer layer after concatenating the feature output of the Cross-Attention architecture.

During our experiment, we process data from two modalities: histologic images and genomic features. First, we conduct representation learning on the data from the two modalities separately and construct the feature extraction network for the two modalities. Then we train a feature encoder for histologic images and genomic features respectively. After that, we use Cross-Attention Concatenation to fuse multimodal medical data. Finally, the result of our experiment shows that the method we proposed achieves a higher accuracy rate compared to current works and also greatly improves parameter utilization.

**Contributions:** The contributions of this work are highlighted as follows:

- We propose a multimodal fusion method for cancer survival analysis based on histologic images and genomic features and achieve State-of-the-Art performance (81.85%) in similar bimodal fusion work. What's more, because our model has fewer parameters than other related work, our solutions offer intelligent Glioma cancer diagnosis a more efficient solution.
- We verify that for multimodal cancer features, strengthening the information correlation between modal features before the fusion stage is crucial to the quality of the final fusion features. And our ablation experiment proves that Cross Attention Transformer can do this. This provides a better idea of multimodal learning modeling for other intelligent clinical medical diagnosis research.

## 2. RELATED WORK

**Cancer Outcome Prediction via Survival Analysis:** Survival analysis is a standard method that is used for cancer prognosis [7]. As the data from next-generation sequencing become deeper and deeper, it can conduct convincing statistical survival analysis, such as Cox regression and Log-rank test, which are commonly used in retrospective studies. Moreover, many machine learning methods have been incorporated to conduct research for years. For example, Huang et al. [8] proposed a method that conducted weighted gene-expression network analysis for dimensionality reduction and learning eigen features from RNA-Seq and micro-RNA data for survival analysis in TCGA. However, these approaches do not incorporate the wealth of multimodal information from heterogeneous data sources that have prognostic value. In order to better utilize multimodal information, Richard J. Chen et al. [3] proposed a fusion strategy that used the Kronecker product to directly model pairwise feature interactions across modalities.

**Multimodal Learning:** Multimodal fusion is one of the basic problems in the field of multimodality. Multimodal learning [9][10], which is a general and practical method to solve this problem, creates corresponding communication perceptrons to combine and complement the extracted data information from multiple channels with different statistical characteristics. It aims to seek for a more stable representation of the heterogeneous information data to achieve the goal of fusion. In the biomedical field, a single modality provides very limited information about the research object. However, Multimodal learning can comprehensively consider various information from the research object and capture the relationship among different data for complex disease classification or prediction tasks. With the deepening of the research, multimodal fusion can fuse structured and unstructured data with different statistical properties, abiotic variation sources and various patterns of missing data [11], as well as make full use of complementary and redundant features from different modalities to solve problems [12]. For example, in a multimodal study of cancer patients, data from the genomic modality can recognize cancer driver genes, and data from histologic images provide a view of the tumor’s microenvironment. The fusion of these two modalities realizes the complementarity of information. But the fusion of transcriptomic and proteomic data exists in information redundancy which is especially important if there are missing values in the data of one modality. The aim of multimodal fusion is to fully exploit the complementarity and redundancy of different modalities.

**Multimodal Fusion method:** From the perspective of data fusion time, multimodal fusion methods can be divided into early, intermediate and late fusion [13]. In the field of biomedical, early fusion concatenates the medical data features from different modalities into a single feature vector and then feeds it into the corresponding model for task processing next. While in late fusion, each medical data is trained separately to obtain the prediction results first, and then the outputs of models are further fused according to some rules. And intermediate fusion combines the advantages of early fusion and late fusion to fuse the data of different modalities globally. Specifically, the methods used in multimodal fusion are closely
related to the information types of different modalities. The main methods are feature concatenation, Bayesian models, tree-based ensemble methods, multiple kernel learning, network-based methods, matrix factorizations, and deep neural networks [14]. In particular, Transformers [15] is a highly competitive architecture family among the methods of fusion, which bring plenty of opportunities and possibilities for multimodal learning.

Multimodal Transformers for Histology and Genomics: The proposal of Transformer model is a major milestone breakthrough and progress in the field of NLP. It strips away large network architectures such as RNN and CNN entirely and has achieved state-of-art on various tasks by relying solely on attention mechanisms [15]. The attention mechanism in Transformer applies human perception and attention behavior to the machine so that the machine can learn how to distinguish the important and unimportant parts of the data [16]. Because of this, Transformer has the ability of feature aggregation in different feature spaces with the perspective of global scope. In recent years, studies on the application of Transformer and its related derivative structures in multimodal fusion emerge endlessly. Related methods include Early Concatenation [17] [18], Hierarchical Attention [19], Cross-Attention [20][21], and so on. However, in the field of cancer genomics, most work mainly focuses on establishing the correspondence between histology and genome [22] while fusion strategies related to Transformer have not been explored. We believe that if the Transformer fusion strategies can be reasonably applied to the fusion problem of cancer histologic image modality and genomic modality, it is expected to give an accurate representation of histological features and genomic data to promote more effective and deeper research in the field of cancer diagnosis and prognosis.

3. METHODS

The dataset we used is the Glioma dataset from the TCGA (Cancer Genome Atlas). The Glioma dataset is a comprehensive dataset that contains nearly all the PLCO (The Prostate, Lung, Colorectal, and Ovarian Cancer) study data available for Glioma cancer incidence and mortality analyses, such as histologic images, genomic features, and survival labels of corresponding patients. Histologic images and genomic features are data of different modalities. In our work, we design two neural networks for representation learning in each modality and our multimodal learning paradigm.

3.1 Representation Learning

We used two modalities of data for the cancer survival analysis: Histologic images and genomic features. If we train the entire multimodal model with all modalities jointly, the hypothesis space of the whole multimodal model is so complex that we need a large number of data to train it. Therefore, by learning the same task for unimodal model, we train feature encoders to systematically learn the representation of each unimodal and use these feature encoders in multimodal model training to reduce the hypothesis space of the final multimodal model.

3.1.1 Representation Learning for Histologic Image

Histologic images can reflect the characteristic information of cancer, which is an important tool for cancer prognosis. Tumor microenvironment features contained in them are also of clinical significance in the treatment of tumors. Therefore, we expect to design a Convolutional Neural Network to capture the features in histologic images. Extensive research experience has shown that the pre-trained models on ImageNet can also provide effective pre-trained features in medical image processing work. Due to the limited number of histologic images, we choose EfficientNet-B0 [23] which is pretrained on ImageNet as the backbone of the initial feature encoder for our task. We remove the last fully connected layer of EfficientNet-B0 and replace it with an Adaptive Pooling layer and a fully connected layer. The above is our initial feature encoder, which outputs 32-dimensional vectors \( h_{\text{feature}} \in \mathbb{R}^{32 \times 1} \) as the feature we captured from histologic images and the part of inputs to fusion. The decoder also consists of a Fully Connected layer, which converts \( h_{\text{feature}} \in \mathbb{R}^{32 \times 1} \) into \( h_{\text{label}} \in \mathbb{R}^{1 \times 1} \).

In our task of cancer survival analysis, we use Cox Partial likelihood loss as the loss function and the C-index of survival analysis is introduced to evaluate the designed model. The above setting is also the set for representation learning on Genomic Features and multimodal networks. The corresponding network architecture is shown in Figure 1.
3.1.2 Representation Learning for Genomic Features

In recent studies of cancer survival analysis, genomic features such as transcript abundance (RNA-Seq) and copy number variation (CNV) are frequently used to analyze the survival time of glioma patients. For genomic features, we use Self-Normalizing Networks [24] (hereinafter referred to as SNN) for learning. Because the ordinary fully connected neural network is prone to over-learning the noise in the training set after many times of training, which results in overfitting. In order to control the overfitting of the fully connected neural network training on the geometric features, we use the ELU as the activation function to make the gradient closer to the unit gradient which means that the mean value is accelerated towards zero. At the same time, we add Alpha Dropout to maintain this self-normalization feature and reduce the overfitting effect, so that we can train a network with stronger generalization ability.

The encoder part of our SNN network consists of four fully connected layers. Each fully connected layer contains a linear layer, an ELU activation linear layer and an Alpha Dropout layer. The encoder part output a 32-dimensional vector \( h_{\text{feature}} \in \mathbb{R}^{32 \times 1} \), and \( h_{\text{feature}} \) is the feature of the SNN model in the learning of this modality which then is used as an input to the fusion module. The decoder part consists of a fully connected layer, which converts \( h_{\text{feature}} \in \mathbb{R}^{32 \times 1} \) to \( h_{\text{label}} \in \mathbb{R}^{1 \times 1} \). The network architecture is shown in Figure 2.

3.2 Multimodal Fusion via Transformer Encoder

Given data from both histologic and genomic modalities with known cancer outcomes, we hope to construct an appropriate model to obtain better representations than unimodal ones and thus improve the accuracy of prediction by considering the data information from both modalities. Richard J. Chen et al. [3] proposed a method that relied on a gated attention mechanism before using Kronecker Product to fuse feature data from different modalities. However, this method does not fully consider the intersections and complementarities in the fusion process of different modalities. In this section, we propose a multimodal pretraining fusion method based on Transformer architecture to capture the overall effect of multimodal features. Details are as follows.

**Fusion via Cross-Attention Concatenation:** For this survival analysis of cancer outcome prediction, the methods of extracting data features respectively from histologic image modality and genomic modality were introduced in detail above. Since the histologic image data of different patients are characterized by discrete pixels in 3D space while the genomic data are represented by one-dimensional covariates [25], it indicates that there is a heterogeneity gap in data representation between histologic image modality and genomic modality of cancer. So, performing fusion of the two modalities directly
cannot be compatible with consistency and complementarity, resulting in unreliable fusion results or even inferior to the representation of a single modality. Therefore, our goal is to explore the commonality of these two modalities and give full play to their complementary advantages with multimodal learning. Ultimately, we can achieve better discrimination and prediction of the different effects presented by the experimental group and the control group of cancer patients.

The multimodal fusion model we proposed is divided into three modules, its architecture can be shown in figure 3. In the first module, we use the feature encoder respectively to extract the feature representation of the corresponding modality. Then we carry out tokenization and position encoding on the features of these two modalities and finally we put features into the next module. The second module is to correlate features from histologic image modality and genomic modality through Cross-Attention, in an attempt to maximize the relation between the two modal representations. We aim to control the expression of the two modal features as well as correct the deviation and variation of the fusion process caused by heterogeneity. In the final output pair of features, the feature of histologic image modality is obtained under the condition of genomic feature, and the extraction of genomic feature is also obtained with the participation of histologic image feature. In the third module, we first concatenate paired features from the second module and send them to another Transformer layer for further fusion to achieve global and cross-modal attention. In the meanwhile, this module can fully extract the shared components and retain the unique components of a single modality. The output obtained from this module is the result of multimodal fusion. Finally, we put the result into the corresponding fully connected layer for Cox regression.

**Cross-Attention:** The Cross-Attention Transformer module is a dual-stream Transformer: this module has two parallel Transformer modules and the Query matrices between two modalities are exchanged in a cross-stream manner, as depicted in figure 4. This module puts a pair of features \((h_{\text{h}}^p, h_{\text{g}}^p)\) which are extracted from corresponding feature encoder and added positional embedding after being tokenized into \(N_1 \times 4\) Cross-Attention Transformer layers, and then generates a pair of final output features that contain mutual information.

Figure 3. Fusion via Cross-Attention Concatenation

Figure 4. Cross-Attention
Specifically, \((h_C^j, h_S^j)\) denotes the output of features \((h_C^0, h_S^0)\) passing through the dual-current Transformer layer at the \(j^{th}\) layer. In the \((j+1)^{th}\) layer, for input \(h_C^j\) and \(h_S^j\) we compute Query, Key and Value matrices by linear transformation respectively. The Key and Value from one modality together with the Query from the other modality are sent to the multi-head attention as input. Therefore, attention from each modality generates features based on the relevant information of the other modality. The above process can be expressed as follows:

\[
\begin{align*}
Q_m &= W_{Qm} \cdot h_m \\
K_m &= W_{Km} \cdot h_m \\
V_m &= W_{Vm} \cdot h_m \\
\hat{h}_C^{j+1} &= \text{MHSA} (Q_C^j, K^j_C, V^j_C) \\
\hat{h}_S^{j+1} &= \text{MHSA} (Q_S^j, K^j_S, V^j_S)
\end{align*}
\]

The purpose of this operation is to maximize the association and crossover between the histologic features and the genomic features to update the features of the histologic images. For the genomic features, it is also necessary to add the relevant information of the histologic images to update genomic features. After that, the operations are basically the same as the conventional Transformer Encoder processing, where the features from the Attention block are fed into a feedforward network block. The two blocks are connected by residual concatenation and Normalization Layer for the transition.

**Concatenation Transformer Fusion:** In this module, as depicted in Figure 5, the global fusion is required for the Cross-Attention feature \((h_C^j, h_S^j)\) obtained from the second module to obtain a better fusion effect. Specifically, the features obtained in the second module should be concatenated to form a sequence of fusion features \(h_{\text{mixed}}\) (sequence lengths are 8). And then we calculate the Key, Value and Query matrices of \(h_{\text{mixed}}\) and pass them to \(N_g=2\) Transformer layer to get the fused feature representation. The above process can be expressed as follows:

\[
\begin{align*}
h_{\text{mixed}} &= \text{Concatenate} (h_C^j, h_S^j) \\
h_{\text{fusion}} &= T^j_f (h_{\text{mixed}})
\end{align*}
\]

**Positional Embedding:** \(h_C\) and \(h_S\) are features extracted from histologic modality and genomic modality. Before fusing these two modal features, \(h_C\) and \(h_S\) should be tokenized as a token sequence and then added positional embedding. In detail, we transform a 32 dimension feature \(h_m\) into a token sequence \(h_m^0\) (where \(h_m^0 = (t_1^m, t_2^m, ..., t_{32}^m)\), \(t_i^m \in \mathbb{R}^8\), \(m = C, S\)). And positional embedding is to increase location information to the sequence data, as shown in the following formula:
where $i$ represents the dimension and $\text{pos}$ represents the location, which is the same as the positional embedding proposed by Vanilla Transformer [15].

4. EXPERIMENT SETUP

The data of our survival analysis comes from TCGA\_GBMLGG, which has histologic images and genomic features. When using a unimodal CNN network to learn the representation of histologic images, due to the small number of histologic images, there are only 1505 $1024 \times 1024$ size images without ignoring patients with missing molecular subtypes and missing histological subtypes, which is impossible to fully train deep CNN networks. What's more, the histologic images are fine-grained, so we can segment them and then train our model better. Considering that the image size on ImageNet is $224 \times 224$, we use a stride of 160 (stride = 160) to cut each original $1024 \times 1024$ image into $25 \times 224 \times 224$ sub-images (shown in Figure 6). Because it requires too much time to train a full multimodal model, it is only randomly cut a $224 \times 224$ sub-images from the original $1024 \times 1024$ size histologic image for training when we train a multimodal model.

In order to exploit the given data to complete the survival analysis and reduce the uncertainty of error estimation, our experiments are held in 15-fold cross-validation. The dataset is randomly divided into 15 non-coincident subsets. For the $k^{\text{th}}$ test, the $k^{\text{th}}$ subset of the dataset is used as the test set, and the rest is used as the train set. In the training of each modality, the model is saved under each split to prevent the problem of information leakage during the multimodal fusion.

![Figure 6. Data Processing of Histologic Images. On the left, the histologic image of size $1024 \times 1024$ is transformed into 25 $224 \times 224$ subgraphs. On the right, a block of size $224 \times 224$ subgraphs is randomly taken from the histology image of size $1024 \times 1024$.](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

The hyper-parameters and training strategies we set are shown in Table 1. The SNN and Efficientnet-B0 are trained as the unimodal model backbone with genomic features and histologic images at first. The backbone of each unimodal model is saved under each split, and all of them are used as the feature encoders of the multimodal fusion. Considering that if the pre-trained feature encoder updates its parameters together with the parameters of the fusion model, the noise gradient will be transmitted to the feature encoder during the early training process because the fusion model has not been trained with good parameters. Therefore, we do not update the parameters of these two unimodal feature encoders in the process of fusion training and only trained the fusion model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Optimizer</th>
<th>Fusion Method</th>
<th>LR</th>
<th>BS</th>
<th>Decay of LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>EfficientNet for Path</td>
<td>Adam</td>
<td>(None)</td>
<td>0.0004</td>
<td>80</td>
<td>Cosine</td>
</tr>
<tr>
<td>SNN for Omic</td>
<td>Adam</td>
<td>(None)</td>
<td>0.002</td>
<td>64</td>
<td>Linear</td>
</tr>
<tr>
<td>Kronecker Product</td>
<td>Adam</td>
<td>Kronecker Product Based on Gating</td>
<td>0.0001</td>
<td>32</td>
<td>Linear</td>
</tr>
<tr>
<td>Transformer Fusion</td>
<td>Adam</td>
<td>Transformer</td>
<td>0.0003</td>
<td>64</td>
<td>Cosine</td>
</tr>
<tr>
<td>Cross-Attention</td>
<td>Adam</td>
<td>Cross-Attention</td>
<td>0.0003</td>
<td>64</td>
<td>Cosine</td>
</tr>
<tr>
<td>Cross-Attention Concatenation (Ours)</td>
<td>Adam</td>
<td>Cross-Attention Concatenation (Ours)</td>
<td>0.0003</td>
<td>64</td>
<td>Cosine</td>
</tr>
</tbody>
</table>

We used C-Index to evaluate our models. C-index refers to the proportion of pairs among all patient pairs whose ranking of predicted survival time is consistent with the actual outcome. It estimates the probability that the predicted outcome is
consistent with the actual observed outcome. When the C-index is 1, it indicates that the predicted result is completely consistent with the real situation, and the closer to 1, the better the model effect.

5. EXPERIMENTAL RESULTS AND DISCUSSION

Ablation Experiment: In order to confirm whether the Cross-Attention part plays an indispensable role in feature fusion, we conduct an ablation experiment. Under the same experiment setting, we perform experiments on Cross-Attention Concatenation fusion and Transformer fusion with Cross-Attention removed. The results are shown in Table 2. Through the comparison, it can be seen that the effect of Cross-Attention Concatenation is significantly improved compared with the simple Transformer fusion, which shows that the Cross-Attention part plays a significant role in feature fusion.

Table 2. Comparison of the Result in the Ablation Experiment

<table>
<thead>
<tr>
<th>Model</th>
<th>C-Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer Fusion</td>
<td>0.7806*</td>
</tr>
<tr>
<td>Cross-Attention Concatenation (Ours)</td>
<td>0.8185*</td>
</tr>
</tbody>
</table>

* $p < 0.05$

Final Result: The result of our experiment is shown in Table 3, Figure 7 and Figure 8. It can be seen that the C-Index of the Cross-Attention Concatenation method reaches 0.8182. This is better than unimodal, Kronecker product multimodal fusion and the fusion method which only contains the Cross-Attention or only integrates data into the Transformer layer. We confirm that Cross-Attention can promote the fusion of two modalities. Furthermore, the last Transformer layer can extract data features to model globally and then achieve global cross-modal attention.

Table 3. The C-Index in Glioma Survival Prediction for Various Models

<table>
<thead>
<tr>
<th>Model</th>
<th>C-Index$^1$</th>
<th>Number of Parameters for Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>EfficientNet</td>
<td>0.7272</td>
<td>(None)</td>
</tr>
<tr>
<td>SNN</td>
<td>0.7983</td>
<td>(None)</td>
</tr>
<tr>
<td>Kronecker Product</td>
<td>0.8140</td>
<td>170130</td>
</tr>
<tr>
<td>Concatenation</td>
<td>0.7806</td>
<td>3425</td>
</tr>
<tr>
<td>Cross-Attention</td>
<td>0.8092</td>
<td>6785</td>
</tr>
<tr>
<td>Cross-Attention Concatenation (Ours)</td>
<td>0.8185$^*$</td>
<td>7625</td>
</tr>
</tbody>
</table>

* $p < 0.05$

$^1$ The result in this table is the average C-Index of 15 splits.
The result in Table 3 shows that the “Cross-Attention Concatenation” method is the most efficient in parameter utilization. Although the number of parameters in our model is only about 1/22 of that in Kronecker Product, it performs better than the latter. The essential reason is the multimodal learning method, which is based on the Cross-Attention and Transformer, can take better account of the intersection and complementarity in the fusion process of different modalities. As a result, it can learn the information from different modalities more efficiently, thus achieving excellent results with fewer parameters.

6. CONCLUSIONS

In this work, we verified that Cross Attention plays an essential role in enhancing the correlation between the different modal features of cancer pathological data. For the feature pairs after Cross Attention enhancement, its final fusion effect is often better than the normal feature pairs with the same decoder, (3.79% increase). Based on this result, we propose a multimodal cancer feature fusion method based on Cross Attention Transformer with fewer parameters and better effects. This also proves that multimodal Transformer also has great potential in clinical medical diagnosis. And this work points out a meaningful research direction: With the development of the deep neural network, we can use more and more modal data in the research of machine learning tasks in the field of clinical medicine by representation learning, we can fusion different model representation features through a variety of multimodal fusion methods. And it shows that the final regression or classification task can achieve better results by using the fusion feature. Future work should focus more and more on the representation learning methods of various kinds of medical data and the fusion methods of various model representation features.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61876066 and in part by the Special Fund for Scientific and Technological Innovation Strategy of Guangdong Province, China (cultivation of scientific and technological innovation of college students, project code: pdjh2022a0030).

REFERENCES

Power grid fault diagnosis method based on alarm information and PMU fusion

Liu Jiang* a, Zheng Ling a, Zhang Xu b, Guo Xing b

a School of Control & Computer Engineering, NCEPU, Changping District, Beijing 102206, China;
b School of Electrical & Electronic Engineering, NCEPU, Changping District, Beijing 102206, China

*Corresponding author: 18435223933@163.com

ABSTRACT

With the rapid development of smart grid technology, more and more attention has been paid to the stability of the power grid. When the power grid fails, the dispatch center will receive various types of data, including the PMU data collected by the PMU (Power Management Unit) device used by the Wide Area Monitoring System (WAMS) and the data collected by the Supervisory Control System for Data Acquisition and Control (SCADA). Alarm information data, they describe the same grid fault from different dimensions. At present, the power grid fault diagnosis based on PMU data and alarm information texts has been fully studied, but the information source is single, and it is difficult to ensure the accuracy of the diagnosis results when complex faults and multiple faults occur in the power grid. This paper starts with the information source of power grid fault diagnosis, and develops a power grid fault diagnosis method that integrates the dual data sources of PMU and alarm information: 1. effectively connect alarm message text and PMU based on fault events. 2. complete the fusion of PMU and alarm information to improve the scope and accuracy of power grid fault diagnosis. Finally, the simulation fault data is used to test. The experimental results show that the accuracy of fault diagnosis of dual data source fusion is higher than that of single data.

Keywords: PMU, alarm information, Grid fault diagnosis, GAN

1. INTRODUCTION

With the continuous expansion of the scale of my country's power system, many new technologies are applied in the power grid, and the power grid is developing towards intelligence and intensification. However, the complex power system will also generate more unstable factors, posing new challenges to the stability of the power grid. If faults cannot be quickly and accurately diagnosed and isolated, serious power outages can easily be caused, causing serious economic losses to the society. The traditional power grid fault diagnosis is based on the action alarm information data, but in the face of complex fault events, the entire diagnosis model becomes increasingly complex; the current power grid fault diagnosis uses PMU data, but it is difficult to fully describe the fault process, and complex faults only discrete results can be obtained.

The accident alarm information collected by the SCADA system and the PMU data collected by the synchrophasor measurement unit are both important sources for obtaining information on the operation status of the power grid. These two types of data information have different description angles and methods for the state of the power grid, and are complementary. It is very complicated to use alarm information data when analyzing complex faults, and the accuracy rate is not high, but PMU data can collect phase and frequency data, which is more advantageous for complex cascading failure analysis. Therefore, when the power grid fails, the fault diagnosis method integrating the alarm information data and the PMU data is studied in order to analyze the fault event more comprehensively. Based on this, the power grid should effectively use these two types of data, extract the side information of complex faults from different data sources, and restore the whole process of complex faults, so as to promote the whole system to be more complete and accurate.

The data fusion method has been widely used in the power grid. By fusing data from different sources, it is improved to form a richer database that is more reliable than a single data source. This is due to the complementary characteristics of data fusion methods. In [4], the discrete time data of the circuit breaker and the continuous time data of the recorder in the PMU are integrated to improve the fault diagnosis capability of the smart grid; Multi-source data is integrated and analyzed, and applied to the fault diagnosis model of metal oxide arrester; [5] The data sources scattered in different
regions are integrated on the distribution network system platform and analyzed in a unified manner, using multi-source data integration technology and The public information model of distribution network grid based on data synthesis can improve the fault repair rate of 10kV distribution network, shorten the time to restore power supply, and effectively improve the accuracy of fault diagnosis of distribution network; [6] established a fusion of external meteorological factors, equipment The real-time status evaluation model of 10kV distribution transformers based on internal health status and current operating conditions can significantly improve the prediction effect of power outage duration. dynamic characteristics. In [7], the switch state data of SCADA and the continuous time data of fault recorder and WAMS are integrated to diagnose the power failure of multiple components. The method of data fusion can improve the accuracy of diagnosis. [16] fuses the data of Micro-PMU and SCADA, and proposes a distribution network state estimation algorithm based on the least squares method of dynamic variable weights to achieve high-precision state estimation, and the effectiveness of the algorithm is verified by tests; [17] A hybrid dynamic estimation algorithm for PMU and SCADA measurements is proposed, and experimental studies show that the hybrid method can improve the estimation to a certain extent; in [18], the slow sampling rate SCADA data and the high sampling rate PMU are fused to the dynamic state of the power system In the estimator, the dynamic tracking of the power system is realized, and the test results prove its availability; [19] uses the fusion method to make the PMU fill the missing SCADA data, and establish a multi-time scale data set for multi-time scale state estimation.

The data fusion method has been widely used in the power grid. By fusing data from different sources, it is improved to form a richer database that is more reliable than a single data source. This is due to the complementary characteristics of data fusion methods. In [4], the discrete-time data of the circuit breaker and the continuous-time data of the recorder in the PMU are fused to improve the fault diagnosis capability of the smart grid.; In [5], multi-source data such as online monitoring, live detection, power failure maintenance, and on-site inspection are integrated and analyzed, and applied to the fault diagnosis model of metal oxide arresters.; [6] Integrate data sources scattered in different regions and conduct unified analysis on the distribution network system platform. Using multi-source data integration technology and a public information model of distribution network grid based on data integration can improve the 10kV distribution network. Fault repair rate, shorten the time to restore power supply, and effectively improve the accuracy of fault diagnosis of distribution network; [7] completed the integration of multi-source real-time data such as SCADA, PMU and safety and stability control system, which can accurately simulate the dynamic characteristics of wind farms; In [8], the switch state data of SCADA, the continuous time data of fault recorder and WAMS are integrated to diagnose the power failure of multiple components, the method of data fusion can improve the accuracy of diagnosis; [9] fused Micro-PMU and SCADA data, and proposed a distribution network state estimation algorithm based on the least squares method of dynamic variable weights to achieve high-precision state estimation, and the effectiveness of the algorithm was verified by tests; [10] proposed a hybrid dynamic estimation algorithm for PMU and SCADA measurements, and experimental studies show that the hybrid method can improve the estimation to some extent. In [11], SCADA data with slow sampling rate and PMU with high sampling rate are fused into the dynamic state estimator of the power system to realize the dynamic tracking of the power system, and the test results prove the availability; [12] adopted a fusion method to make PMU fill in the missing SCADA data, and established a multi-time scale data set for multi-time scale state estimation.

This paper is mainly based on the multi-source data fusion of alarm information and PMU for analysis and research. The alarm information and PMU are used as data sources to input Generative Adversarial Network (GAN) to learn the corresponding relationship between alarm information and PMU in each fault event, and in the discriminator A dual-data filtering mechanism is added to display different aspects of the fault through different data sources. Finally, a fault diagnosis model is constructed, and the type of grid fault is judged based on the fusion of dual data sources as the input of the model.

2. BACKGROUND

2.1 Dual data sources and PMU visualization

In this paper, the power grid fault simulation system is used to export data. The system uses DigSILENT and TS2000 simulation software as the underlying support software. By setting different types of faults of different devices in the power grid, the system can call the corresponding interface to generate the corresponding simulated alarm information text and PMU data. Table 2.1 and 2.2 are the corresponding alarm information when the Lei-Zhao 2 line fails. Text and PMU data.
### Table 1. Lei-Zhao 2 Line Fault alarm information

<table>
<thead>
<tr>
<th>Time</th>
<th>Plant station</th>
<th>Device description</th>
<th>Action description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022-08-01 00:00:00</td>
<td>Lei Station</td>
<td>Lei-Zhao 2 Line RCS931 differential protection</td>
<td>Operate</td>
</tr>
<tr>
<td>2022-08-01 00:00:01</td>
<td>Lei Station</td>
<td>Lei-Zhao 2 Line RCS931 phase-to-phase distance</td>
<td>Operate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022-08-01 00:00:17</td>
<td>Lei Station</td>
<td>Lei-Zhao 2 Line Phase differential</td>
<td>Operate</td>
</tr>
<tr>
<td>2022-08-01 00:00:17</td>
<td>Lei Station</td>
<td>Lei-Zhao 2 Line 2251 switch Total exit trip</td>
<td>Operate</td>
</tr>
<tr>
<td>2022-08-01 00:00:18</td>
<td>Lei Station</td>
<td>Lei-Zhao 2 Line 2251 switch First set of exit trip</td>
<td>Operate</td>
</tr>
<tr>
<td>2022-04-01 00:00:18</td>
<td>Zhao Station</td>
<td>Lei-Zhao 2 Line Longitudinal differential</td>
<td>Reset</td>
</tr>
</tbody>
</table>

### Table 2. Lei-Zhao 2 Line Fault PMU data

<table>
<thead>
<tr>
<th>Time</th>
<th>A phase Voltage Magnitude</th>
<th>B phase Voltage Magnitude</th>
<th>C phase Voltage Magnitude</th>
<th>A phase Current Magnitude</th>
<th>B phase Current Magnitude</th>
<th>C phase Current Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022-08-01 00:00:00</td>
<td>223.1982</td>
<td>223.1982</td>
<td>223.1982</td>
<td>0.26756</td>
<td>0.26756</td>
<td>0.26756</td>
</tr>
<tr>
<td>2022-08-01 00:00:01</td>
<td>223.1982</td>
<td>223.1982</td>
<td>223.1982</td>
<td>0.26756</td>
<td>0.26756</td>
<td>0.26756</td>
</tr>
<tr>
<td>2022-08-01 00:00:17</td>
<td>162.6409</td>
<td>209.3513</td>
<td>209.3032</td>
<td>13.08758</td>
<td>13.11966</td>
<td>0.266835</td>
</tr>
<tr>
<td>2022-08-01 00:00:18</td>
<td>162.2062</td>
<td>208.7917</td>
<td>208.7436</td>
<td>13.052343</td>
<td>13.08439</td>
<td>0.265926</td>
</tr>
<tr>
<td>2022-04-01 00:00:18</td>
<td>161.6036</td>
<td>207.3203</td>
<td>207.2725</td>
<td>12.959747</td>
<td>12.99167</td>
<td>0.263558</td>
</tr>
</tbody>
</table>

In the power grid fault simulation system used in this paper, the data collection of the PMU used is as follows: the three-phase voltage amplitude and phase angle at both ends of the bus, the three-phase current amplitude and phase angle at both ends of the line and transformer, a total of 36 dimensions. Because the PMU data is continuous and time series, when the electrical equipment fails, its voltage or current amplitude often fluctuates. Therefore, this paper extracts the three-phase voltage and current amplitude of the faulty equipment in the PMU data. A total of six dimensions It is drawn into polar graph and alarm information text for data source fusion, as shown in Figure 2.3. In addition, the fluctuation shapes of different fault voltages and currents are different. Compared with numbers, the images are more clear in representing fluctuations, and it can avoid the waste of resources caused by processing large-scale data and the influence of irrelevant data on the results of fault diagnosis, and improve the training speed of the model.

![Figure 1. PMU data to generate polar graph.](image-url)
2.2 GAN

A generative adversarial network (GAN) consists of a generator and a discriminator, as shown in Figure 2.1. the generator captures the latent distribution of real data samples and generates new data samples; the discriminator is a binary classifier, and the discriminant input is the real data is still a generated sample. Both the generator and the discriminator can use the commonly used deep neural network. At present, convolutional neural network (CNN) and recurrent neural network (RNN) are commonly used as generators and discriminators.

![GAN structure](image)

Figure 2.GAN structure.

The basic model of GAN is shown in the formula (1). Let Z be random noise and X be the real data. The generator and discriminator are represented by G and D respectively, where D can be regarded as a binary classifier, which is represented by cross entropy:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

(1)

Here, \(p_{data}(x)\) represents the true sample distribution, and \(p_z(z)\) represents the distribution of the generated samples.

The output of \(D(x)\) is a real number in the range 0-1 representing the probability of whether the picture is a real picture. Through such a Max-min game competition, G and D are optimized alternately back and forth until the loss function converges.

The goal of GAN is that the discriminator D learns a data distribution close to the real image, and the generator G generates an image that the discriminator D judges as a real image as much as possible under the guidance of the discriminator D. When their training reaches the Nash equilibrium, the generator G can generate realistic samples, and the discriminator D has the ability to perfectly distinguish the generated samples from the real samples.

2.3 GAN-CLS

Traditional GANs generate images from noise, and then hand them over to the discriminator to judge true and false. This free training method makes it impossible to make judgments on multi-class data. Therefore, we introduce the conditional generative adversarial network CGAN, the difference is that the conditional variable y is introduced into the input of the generative model and the discriminant model, and the directional generation of the image is guided by the conditional variable. The condition variable y can be a label or an image feature. CGAN is a generative adversarial network based on supervised learning as shown in Equation (2).

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x), y \sim p_y(y)}[\log D(x \mid y)] + E_{z \sim p_z(z)}[\log(1 - D(G(z \mid y)))]$$

(2)

Here, \(x \mid y\) represents the occurrence of \(x\) under the conditions that \(y\) occurs. GAN-CLS is an augmented CGAN network with a matching mechanism. Compared with ordinary GAN, GAN-CLS can not only judge the authenticity of the image, but also judge whether it matches. The model itself can generate images with better clarity and detail with the help of the image generation performance of DCGAN, and the GAN-CLS training algorithm can improve the degree of correlation between text description and generated image content to a certain extent, so that the final image generation not only The quality is high, and the degree of fit with the text description content is also high.

In order to achieve the matching of alarm information and images, this paper adopts the training method of image-text consistency. First, the text and image are used as input to train the discriminator, and whether the image and text match is judged by observing the output probability of the discriminator. In our generative adversarial network model, the discriminator receives three inputs: (real image, correct text), (wrong image, arbitrary text) and (correct image, wrong text). Different from the traditional generative adversarial network, the discriminator not only needs to judge whether the
image is true or false, but also needs to judge whether the text and the image match.

3. TEXT-IMAGE MATCHING LEARNING

To obtain visually discriminative vector representations of text descriptions, a deep convolution and a recurrent text encoder are used to learn the correspondence between images and text descriptions. To obtain visually discriminative vector representations of text descriptions, deep convolutions and recurrent text encoders are used to learn the correspondence between images and text descriptions. Equation (3) represents learning the matching function between text and image, and the closer the value is to 1, the higher the degree of matching between text and image.

$$\frac{1}{N} \sum_{n=1}^{N} \Delta(y_n, f_v(v_n)) + \Delta(y_n, f_t(t_n))$$

(3)

Here, $v_n$ is the image set, $t_n$ is the corresponding text description set, and $y_n$ is the class label. $f_v$ and $f_t$ are the image discriminator and the text discriminator respectively, the formulas are (4) and (5), and the output result is the predicted label.

$$f_v(v) = \arg \max \mathbb{E}_{t \sim \alpha}[\phi(v)^T \varphi(t)]$$

(4)

$$f_t(t) = \arg \max \mathbb{E}_{v \sim \beta}[\phi(v)^T \varphi(t)]$$

(5)

Here, $\phi$ is the image encoder, $\varphi$ is the text encoder, $\alpha$ is the text description set, and $\beta$ is the image set corresponding to the alarm information, $v$ and $t$ are the pmu images in the image set and the alarm information text in the text set, respectively. This paper defines $(v, t) = \phi(v)^T \varphi(t)$ as the fitness function, and the scores of images and matching texts should be significantly higher than other scores that cannot be matched.

At the beginning of training, the discriminator will ignore the condition information, because the pmu images generated by the generator are unreasonable, so the discriminator can easily judge the samples generated by the generator as false. After training, when the generator can generate matching alert information, the discriminator should also evaluate whether the generator satisfies this conditional constraint.

4. SIMULATION EXPERIMENTS AND RESULTS

Based on the Generative Adversarial Networks (GAN) alarm information and PMU fusion system, the visual PMU data generated by the simulation software and the alarm information generated by the simulation are used as the basic data sources. We train three generators, image discriminators and text discriminators. A module is used to realize:

1. effectively connect the alarm information text and PMU based on fault events;
2. complete the fusion of PMU and alarm information, and improve the scope and accuracy of power grid fault diagnosis.

The fusion diagnosis process of GAN-CLS-based alarm information and PMU diagnosis results is shown in the following Figure 3.

![Figure 3. Fusion process of alarm information and PMU based on GAN.](image-url)
The generator and discriminator models designed in this paper are shown in the figure, which are 23-layer and 13-layer convolutional neural networks respectively. A BN layer is added in the middle of the convolutional layer to ensure the stability of data distribution. Adding downsampling to the generator reduces the complexity of the model and preserves key information.

4.1 Connection of alarm information

The alarm information and the PMU image are encoded by the text encoder and the image encoder respectively, and the matching degree is calculated, and the matching degree is trained by the gradient descent algorithm. A high value is assigned to the alarm information and PMU image with a high degree of matching, and a low value is assigned otherwise. The operation process is as formula (6).

\[(v, t) = \phi(v)^T \varphi(t)\]  

(6)

Here, \(\phi\) is the image encoder, \(\varphi\) is the text encoder, \(v\) and \(t\) are the PMU images and the alarm information, respectively. The image encoder designed in this paper is a 5-layer convolutional neural network, and the output is a 512-dimensional one-dimensional vector. The text encoder uses LSTM word embeddings, and the output is a 128-dimensional vector.

4.2 GAN-CLS training process

<table>
<thead>
<tr>
<th>Table 3. GAN-CLS training process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: PMU images (x), matching alarm information (t), mismatching alarm information (\hat{t})</td>
</tr>
<tr>
<td>1. (h = \varphi(t)), (\hat{h} = \varphi(\hat{t}))</td>
</tr>
<tr>
<td>2. (z \sim N(0, 1)^{\hat{z}})</td>
</tr>
<tr>
<td>3. (\hat{x} = G(z, h))</td>
</tr>
<tr>
<td>4. (s_r = D(x, h), s_w = D(x, \hat{h}), s_f = D(\hat{x}, h))</td>
</tr>
<tr>
<td>5. (\log(s_r) + (\log(1 - s_w) + \log(1 - s_f)) / 2 \rightarrow D)</td>
</tr>
<tr>
<td>6. (\log(s_r) \rightarrow G)</td>
</tr>
</tbody>
</table>

The input is the PMU image, the corresponding alarm information text, and the non-corresponding alarm information text. First, use a text encoder to encode the warning information as a vector of the same dimension, and generate noise (lines 1-2). The PMU image is then generated from the noise and warning message text. The fourth line takes three inputs: (real image, corresponding to text), (generated image, corresponding to text) and (real image, not corresponding to text) through the score of the discriminator. Finally, the gradient descent algorithm is performed through this parameter to update the parameters of the generator and discriminator (lines 5-6).

4.3 Model training

By setting different types of faults, the PMU data and alarm information generated by the power grid fault simulation system at the same time correspond to each other and have the same file names. Perform data processing on the acquired double data.

First, we input the processed training set into gan-cls for training, first reduce the dimensionality of the PMU picture data through coding, split the warning message text into words and convert them into matrices. The input data of the three modules are as follows:

1. The input data for the generator is noise and warning message text.

2. There are four types of input data for the discriminator: generated PMU image, non-corresponding alarm information, real PMU image, and corresponding alarm information.

After passing through the hidden layer, the output is one-dimensional data, and the loss value is calculated. Finally, the model is optimized by the AdamOptimizer function, and the algorithm based on gradient descent is used to solve the
minimum loss. Through such a process, the matching process of the dual data sources is realized, and the Loss value is kept within the set threshold value of 0.01 after several cycles.

4.4 Model training

Using the power grid fault simulation system, set the power grid to have no faults and faults, and export the alarm information text and PMU data. The sample composition of the fault diagnosis model is shown in Table 4. And divided into training set and test set according to the ratio of 8:2.

<table>
<thead>
<tr>
<th>Table 4. Data sample composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>No faults/ Faults</td>
</tr>
<tr>
<td>No faults</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Faults</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The structure of the fault diagnosis model is shown in Figure 10. The text encoder is used to encode the alarm information, the image encoder is used to encode the PMU image, the two encoding results are fused, and input into the fault diagnosis model for training, and finally, a fault diagnosis model of dual data source fusion is formed.

![Figure 4. Fault Diagnosis Model Based on Dual Data Source Fusion.](image)

In this paper, the $A_{cc}$ value is used as an indicator to evaluate the performance of the model, indicating the proportion of correct predictions in all prediction results. The test set is input into the trained model to verify the training results of the model. As a comparison, three comparison cases are set up in this paper: only PMU images, only alarm information text, the fusion of alarm information and the PMU image generated by the Generator. The network models used in all cases are the same, and the comparison results are shown in the table 5. Experiments show that the fault diagnosis effect of dual data source fusion has higher accuracy than single data source.

<table>
<thead>
<tr>
<th>Table 5. Fault diagnosis result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fault/No fault</td>
</tr>
<tr>
<td>PMU image only</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Alarm information only</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>The fusion of alarm information and generated PMU image</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>The fusion of alarm information and PMU image</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
5. CONCLUSION

This paper attempts to develop a fault diagnosis method based on GAN-based dual data source fusion. First, the alarm information and PMU data are generated by the simulation system, and the fusion learning of the alarm information and the PMU image is established through the training of GAN, and then the result of the fusion of the dual data sources is applied to the fault diagnosis. Through case comparison, the fault diagnosis of dual data source fusion input is more accurate than that of single data. In the future, more in-depth research will be carried out on the process of alarm information and PMU image fusion.

Acknowledgement

This work is supported by National Natural Science Foundation of China Power grid fault diagnosis based on neural machine translation (No:51877079)

REFERENCES


[21] Qing-Sen C, Kei E, Tong C. A New Data Fusion Method for Multi Terminal Fault Location Based on Sensor Networks[J]. Itm Web of Conferences, 2018, 17:02007-.


The Study of Object Transformation Based on CycleGAN

Junqing Wang1a*
1Department of computing, Northeast Electric Power University, Changchun 130000, China
a*2019303030229@neepu.edu.cn

Abstract

At present, generative adaptive networks (GAN), has become a popular module in deep learning. GAN is very effective in image generation and image style migration. At present, the research on the migration of image style is focused on images such as oil painting and landscape painting, and lacks the research on the conversion between object images. This paper extends the style transfer technology to object recognition, uses CycleGAN method to learn the mapping relationship between zebra and horse, and realizes the transformation between zebra and horse. Viewing the generation effect of different learning methods by changing the learning times and learning rate policy. This work realizes the conversion between zebra and horse, and shows the generated pictures under different training times and different learning situations. Under the same training times, the conversion effect from horse to zebra will be better. After a certain number of trainings, the training effect will gradually decline. The conversion effect of the same type will be improved with the increase of training times. Different learning rate policies will bring different generation effects.

Keywords-GAN; CycleGAN; Style migration.

1. Introduction

Generative methods that generate new samples from large data distributions, like images, are widely used, for instance in image-to-image translation and the most prominent approaches are autoregressive models, variational autoencoders, and generative adversarial networks [1]. CycleGAN is an important tool for image migration. CycleGAN has the feature that datasets do not have to be one-to-one corresponding and has been widely developed in the business fields such as Artificial Intelligence (AI) face changing and beauty. CycleGAN consists of four parts, two generation models and two discrimination models. Two generators are used to generate style migrated images according to the mapping of two domains, and two discriminators are used to detect the difference between the resulting image and the original image. Through the result of the discriminator, the weights of the mapping relationship between domains can be updated to make the style migrated image clearer.

Deep learning is growing fast not only in other computer vision related tasks [2-5], but also in image generation. In the initial research, GAN, a framework for generating models through confrontation process estimation, was proposed [6]. GAN improves the prediction effect through the confrontation between the generator model and the discriminator model. Later, CGAN changed the original unsupervised GAN model into a supervised GAN model by introducing conditional variables into the generator model and the discriminant model, which effectively improved the generation effect of GAN [7]. In addition, another variant of GAN called DCGAN improves the network structure of GAN and improves the quality of GAN generation results by replacing the full connection layer with the convolution layer. Before CycleGAN was proposed, a network structure called pix2pix has been proposed [8]. However, the dataset of pix2pix must be paired, that is difficult to be satisfied in reality, just like the two pictures of day and night in the same scene, it is difficult to find a dataset containing the pictures of day and night in the same scene. Therefore, CycleGAN was proposed, it mainly solved the problem that datasets of pix2pix must be paired [9, 10]. CycleGAN only needs datasets in two fields, such as pictures of common horses and zebras, but it does not need to be mapped one by one. However, in article [9], it is more inclined to introduce the structure of CycleGAN and various parts. The impact of each part on the results is not shown in detail.

In this research, horses and zebras were used as datasets to study the impact of various parts of CycleGAN on the final results. In addition to a single factor, this research aims to study the impact of multiple factors on the results. Because the influence of multiple factors is more universal than that of a single factor. More other factors including training times, learning rate policy and dataset size. Different factors were replaced in this study to see different results and compare them.
2. Method

2.1 Dataset description and preprocessing

In this project, the horse2zebra dataset provided by article was used [9]. The initial dataset contained 400 pictures of regular horses and 400 pictures of zebras for training, and 120 pictures of common horses and 140 pictures of zebras for testing. In other words, there are 1,060 pictures with 256×256 pixels, of which 800 are used for training and 260 for testing.

The preprocessing includes four parts. First, in order to perform this process smoothly, image pixels were resized in the dataset to 286×286. Second, the pixel size was further cropped to 256×256. Too large or too small pixel size will affect the quality of the generated picture. If the pixel is too small, the generated picture will become unclear. If the pixel is too large, there will be too much data to be processed and the operation time will be too long. Third, after resize, there is a 50% chance to enhance the data through data inversion to avoid over fitting. Fourth, specify the value of data from -1 to 1 through normalization.

2.2 Proposed approach—CycleGAN

![CycleGAN Diagram]

Fig. 1 (a) The model has two areas, X and Y, two mapping functions, G and F, and two discriminators, D(X) and D(Y). (b) and (c) represent the same process in the opposite direction. The generator will generate a result image that cannot be distinguished by the discriminator through the corresponding mapping function, such as G(x)=y and F(y)=x. The cycle consistency loss means that we should obtain the original input image by passing the result image produced by the corresponding mapping feature, like F(G(x))≈x [7].

Figure 1 indicates the process of CycleGAN [9]. Adversarial loss for two mapping functions and their corresponding discriminators, they have their own adversarial loss. In the case of the mapping function G and its discriminator D(Y), the objective is:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))]$$

(1)

Generator G wants to generate a result image that looks similar to domain y, the discriminator D(Y) wants to distinguish the resulting image from the original image in domain y. And generator G strives to keep the difference between the generated image and the original image as small as possible, the discriminator D(Y) wants to maximize the gap between them. So the objective of Generator G and discriminator D(Y) is minimize maxD(Y) L_{GAN}(G, D_Y, X, Y). The same principle applies to the mapping function F and its discriminator D(X).

2.3 Cycle Consistency Loss

As shown in Figure 1 (b) (c), Cycle Consistency Loss means the distinction between the resultant image returned to the original domain and the original image entered according to the corresponding mapping function, and the function is:

$$L_{Cyc}(G, F) = E_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] + E_{y \sim p_{data}(y)}[||G(F(y)) - y||_1]$$

(2)

Adversarial Loss [6] and Cycle Consistency Loss [9] are both part of the general objective. And the general objective function is:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{Cyc}(G, F)$$

(3)

And the \( \lambda \) controls the relative importance of the two objectives. Through the objective function, we can get:

$$G^*, F^* = \arg \min_{G,F} \max_{D_X,D_Y} L(G, F, D_X, D_Y)$$

(4)

This study trained two generators and two discriminators to optimize the results of the tests.
2.4 Implementation Details

0.0002 was took as the initial learning rate for learning. In the first half of the epoch, the learning rate remains unchanged, and then gradually decreases to 0. Weights are started based on the Gaussian distribution N(0,0.02).

3. Test Results and Discussions

Fig. 2 Horse → Zebra, total epoch=100, learning rate policy="liner". “real” indicates the original picture, “fake” indicates the objective picture, “rec” indicates original picture generated from target picture. The same is true for the following.

Fig. 3 Horse → Zebra, total epoch=60, learning rate policy="liner".

Fig. 4 Horse → Zebra, total epoch=50, learning rate policy="liner".

Fig. 5 Zebra → horse, total epoch=50, learning rate policy="liner".
Fig. 6 Zebra → horse, total epoch=60, learning rate policy="liner".

Fig. 7 Zebra → horse, total epoch=100, learning rate policy="liner".

Fig. 8 Horse → Zebra, total epoch=60, learning rate policy="cosine"

Fig. 9 Horse → Zebra, total epoch=60, learning rate policy="plateau"

Fig. 10 Zebra → horse, total epoch=60, learning rate policy="cosine".
Under the same learning rate policy, in most cases, the more epochs, the better the resulting image will be generated. As shown in Figure 2 (a), Figure 3 (a) and Figure 4 (a). However, there are also a few pictures that have a good effect in training with a small number of epoch times. As shown in Figure 3 (b) and Figure 4 (b), the effect of Figure 4 (b) is obviously better.
For the conversion from zebra to horse, the effect is not as good as that from horse to zebra. When a zebra is converted into a horse, it still retains zebra stripes, which is more like changing the skin color of a horse to a zebra. However, with the increase of epoch times, the fringe sharpness decreased significantly. As shown in Figure 5, Figure 6 and Figure 7.

After changing the learning rate policy, the effect of generating pictures decreases significantly. It is more like changing the background color of the picture than the conversion between horse and zebra. As shown in Figure 8, 9, 10, 11.

The number of horses in the picture and the skin color of horses may affect the quality of the generated picture. As shown in Figure 12, the number of horses in the generated image appears to be one less than that in the original image. And as shown in Figure 13, the conversion effect from black horse to zebra is obviously not good.

From the Figure above, the conversion from horse to zebra is relatively good, the conversion from zebra to horse is not effective. In addition, there are still a few images whose conversion effect is not so perfect, like Figure 12. This may be due to the failure to recognize individual differences when converting multiple horses. And when the horse's color is dark, the learned zebra color cannot completely cover the horse's skin color. For the conversion from zebra to horse, it can be seen that the training effect is better with the increase of epoch times. Perhaps it is because the characteristics of zebra's body color have not been completely changed. With the increase of training times, the effect will be better. After changing the learning rate policy, the results of both conversion cases are significantly worse. This may be due to different learning rate policies during training and the change of learning rate reduction strategy during training.

4. Conclusion

In this work, using horse to zebra transitions to test CycleGAN performance. By changing the times of epoch and learning rate policy, different pictures are generated for comparison to see the impact of influencing factors on the results. The experimental results show that under the same learning rate policy, the more times of epoch, the better the effect of most of the generated images will be, but after a certain number of training, the training effect will gradually decrease. At the same time, with the same epoch times, different learning rate policies will make the effect of generating pictures very different. Perhaps it is because the training times and training samples are less, and the conversion effect of the model for horses with dark skin and multiple horses is poor. In the future, the model will be trained and improved more effectively by increasing data samples and training times to make the conversion more realistic.

References

TRFit: Learning 3D point cloud normal estimation with Transformer

Hongwen Liu*#, Yufeng Wang#, Zheng Ma#

aSchool of Electrical Engineering and Automation, Anhui University, Hefei, China
* Corresponding author: liuhw9809@163.com

ABSTRACT

In this study, we provide an approach named TRFit for unstructured 3D point cloud normal estimation. It handles noise and uneven densities point clouds well. Recently, learning-based normal estimation methods have significantly outperformed traditional methods on benchmark normal estimation datasets. In order to estimate normals, they frequently employed neural networks to learn point-wise weights for weighted least squares polynomial surfaces fitting. However, existing methods often ignore local geometric relationships, which will make the fitted surface significantly different from the real. To this end, we propose to use graph convolutional to learn local structural information. Meanwhile, we suggest the Geometric Relation Transformer (GRT), a transformer-based scale aggregation module, to fully utilize points from various neighborhood sizes. It can adaptively capture the relations between different regions. We achieve state-of-the-art results on the baseline normal estimation dataset, and experimental results show that TRFit obviously improves the accuracy of normal estimates, preserves their details. Moreover, it exhibits robustness to noise, density variations, and outliers. Besides, we demonstrate its application to surface reconstruction and denoising.

Keywords: Graph convolution, Normal estimation, Point Cloud, Transformer

1. INTRODUCTION

In point cloud processing, the surface normal information of point clouds has been widely used for classification1, segmentation2, surface reconstruction3, and denoising4. Meanwhile, with the development of consumer-grade 3D sensors, it make it easy to capture raw point clouds. However, due to the limitation of scanning technology, the acquired data is often corrupted by various noises. This will make the above tasks more difficult. Therefore, it is crucial to obtain the normal information of the point cloud accurately for the subsequent study.

In general, point normal information can represent point cloud geometric properties. It is strongly influenced by the geometric information in the neighborhood and interacts between different scales. Hence, it is reasonable to assume that enriching the features between point-wise and different neighborhood sizes will make the final estimated weights more accurate. However, by analyzing previous methods, they generally use a fixed neighborhood as input, use the network to learn point-wise weights, thus ignoring the position relationship of points in the same patch. Inspired by this, we propose to use graph structures to enrich the geometric information of point-wise. Benefit from AdaptConv's5 excellent performance in point cloud analysis, we propose a new approach which uses AdaptConv as a feature extraction module. Also, in order to learn more expressive local features, we combine the cross-Transformer module with AdaFit's6 cascaded scale aggregation (CSA) module, to enhance the expressiveness of local features by fusing local information from two different scales. On the benchmark point cloud normal estimate dataset, we compare this method to the most recent state-of-the-art methods, experimental evidence that suggested method can obtain more accurate estimation results, while our model still achieves the best under the influence of different levels of noises. We also demonstrated the use of TRFit in point cloud denoising as well as Poisson reconstruction.

2. RELATED WORK

2.1 Traditional methods

Principal Component Analysis (PCA)7 and Singular Value Decomposition (SVD)8 are the two most traditional techniques for estimating the normal of point cloud, the estimated normals are the tangent plane's normal vector in both cases, which choose a neighborhood at a specific scale and estimate it using PCA or SVD. As a result, the performance of these methods greatly depends on the neighborhood size that is chosen. Large neighborhoods will lead to over-smoothed results, especially at sharp edges, while small scales make the model less robust to noise. After this, some...
works were proposed to fit complex surfaces, such as several variants local spherical surfaces fitting⁹, moving least squares (MLS)¹⁰ and Jets¹¹. These methods improve the accuracy of normal estimation, but still do not get rid of the choice of neighborhood size, they prefer to use large neighborhoods to improve robustness, lead them to smooth sharp features. Mitra¹² analyzed the effect of neighborhood size, curvature and other factors to calculate the optimal neighborhood to reduce errors, but it requires a serious computational cost.

2.2 Learning-based methods

Since deep learning has achieved excellent performance in point cloud processing¹²,¹³. Meanwhile, some learning-based methods have been suggested for point cloud normal estimation. Two of the main methods are regression-based and surface-fitting-based methods.

Regression-based methods. Generally speaking, these methods are relatively simple. They use fully connected layer brute force regression normal vectors. As a pioneer, Boulch et al¹⁴, inspired by the application of 2D convolution on images, they establish the mapping of 2D grid to 3D points via a Hough transform, and use 2D convolutional neural network to regress the normal vector, but this method cannot be used directly with point data. With the success of PointNet², neural networks can directly process 3D coordinates, PCPNet¹⁵ designs a PointNet-based multiscale normal estimation scheme. Hashimoto et al¹⁶ use PointNet as a local feature extraction, and 3DCNN for feature fusion. Zhou et al¹⁷ introduce additional feature constraints and a new multiscale selection strategy to improve surface normal estimation. Nesti-Net¹⁸ proposes to use a mixture-of-experts (MoE)¹⁹ to improve the scale selection problem.

Surface-fitting-based methods. Such brute-force regression, as AdaFit⁶ notes, simply drives the network to memorize normal vectors and results in a subpar generalization of the network. To solve this, recent methods suggest estimating normals from fitted surfaces. IterNet²⁰ learns the anisotropy kernel by an iteration to fit the local tangent plane. DeepFit²¹ makes use of a neural network to predict point-wise weights for weighted least squares polynomial surface fitting. Zhou et al²² introduce a patch point updating technique for surface optimization together with a dynamic top-k selection strategy to concentrate on crucial locations. Zhang et al²³ propose a two-stage network to optimize the accuracy of the results. A recent study AdaFit⁶ suggests a point offset approach and a cascaded scale aggregation module to improve the reliability and accuracy of the output normals. Our method belongs to this, unlike previous methods, we focus on the local features learning, to achieve more accurate results by enriching local geometric features.

2.3 Transformers

Originally, transformer²⁴ was proposed to be used for NLP tasks. It proposed a novel attention mechanism that exchanges information between input and output using multiple layers of self and cross multi-head attention. Some subsequent works²⁵,²⁶ applied transformer to computer vision tasks and get great success. In the 3D domain, PCT²⁷ enriches the local neighborhood points features using positional embedding and offset attention. PoinTr²⁸ succeed in point cloud completion tasks using a standard transformer encoding and decoding structure. REGTR²⁹ achieves point cloud registration by estimating the rigid transformation directly from the predicted correspondence. In this work, we use the transformer structure containing self and cross multi-head attention to enhance the feature information between different scale sizes.

3. METHOD

3.1 Pre-processing

Let \( P = \{p_i\}_{i=1}^N \in \mathbb{R}^{N \times 3} \) represents a point cloud contains \( N \) points, we employ K-nearest-neighbor search to sample a patch \( P_t = \{p_{t,i} | p_{t,i} \in kNN(p_t), j = 1,\ldots,k\} \) for each query point \( p_t \), where \( k \) denotes the sample size, then for each sampled patch, a classic PCA is used to align the patch as initial input.

3.2 Normal Estimation

Given a local patch with PCA alignment, we aim to get the normal vector \( \mathbf{N}_i \) and \( \mathbf{N}_j \) of the query point \( p_i \) and its neighbors \( p_{i,j} \). By fitting a surface onto nearby points and deriving the normals from the fitted surface, this issue may be resolved. We use weighted least-squares surface fitting with the truncated Taylor expansion (n-jet), which is similar to
DeepFit21. This process can be written as a “height function” of height $z$ with respect to coordinates $(x, y)$, the term “height function” might mean:

$$z = f(x, y; \beta) = J_n(x, y; \beta) = \sum_{k=0}^{n} \sum_{j=0}^{k} \beta_{k-j,j} x^{k-j} y^{j}$$

(1)

where $\beta$ represents the coefficient vector of jet, it consists $N_n = (n+1)(n+2)/2$ terms. A Vandemont Matrix $M = (1, x_j, y_j, ..., x_j y_j^{n-1}, y_j^{n})_{j=1,...,k} \in R^{k \times N_n}$ is defined in order to obtain the surface of the local patch. The sampled points can be represented as a height function vector $B = (z_1, z_2, ..., z_k)^T \in R^k$. Thus, we can obtain a system of linear equation:

$$M\beta = B$$

(2)

Since the sampling points contain noise and outliers, we hope the noisy points and outliers contribute less to the fitted surface. Therefore, we use WLS append weights for sampled points. We construct a weight matrix $W = diag(w_1, w_2, ..., w_k) \in R^{k \times k}$. Then the optimization problem becomes:

$$\hat{\beta} = \arg \min_{\beta \in R^{N_n}} w_i \| M_{\beta} - B \|$$

(3)

Besides, we predict the offset $(\Delta x, \Delta y, \Delta z)$ for each point, the final optimization becomes:

$$\hat{\beta} = \arg \min_{\beta \in R^{N_n}} \sum_i w_i \| J_n(x_i + \Delta x, y_i + \Delta y, \beta) - (z_i - \Delta z) \|^2$$

(4)

The final equation can be expressed as a general formula:

$$\hat{\beta} = (M^TWM)^{-1} M^T WB$$

(5)

Given the n-jet coefficient $\beta$, let $\beta_1 = \beta_{1,0}, \beta_2 = \beta_{0,1}$, the center point’s normal $N_i$ could be solved by:

$$N_i = \frac{(-\beta_1, -\beta_2, 1)}{\|(-\beta_1, -\beta_2, 1)\|_2}$$

(6)

### 3.3 Network Architecture

We show the network framework in Figure 1. With a 3D point cloud $P$ and a query point $p$, we first extract the local subset $P_i$, the network then makes a point-wise weight and offset prediction. In particular, we put $P_i$ into a quaternion spatial transformer network (QSTN), to transform the input points into a standard pose $X_i$. Then, a shared multi-layer perceptron (MLP) and a feature spatial transformer network (FSTN) followed. After AdaptConv feature extraction module and GRT feature fusion module, we get local features $F_i = \{f_1, f_2, ..., f_k\}$, and a global feature representation $G_i = MaxPool\{F_i\}$ is obtained after maximum pooling. Finally, these two representations are combined, fed into a MLP $\phi(\cdot)$, and then a sigmoid activation function is used. The final weights could represent as:

$$w_i = \text{sigmoid}(\phi([G_i, F_i])) + \varepsilon$$

where $[\ldots]$ is concatenation operation. According to DeepFit21, the tiny constant $\varepsilon$ is used to prevent the degenerate scenario of a zero or improperly conditioned matrix.
3.4 Feature Extraction

We follow AdaptGCN\(^5\) to adopt the Adaptive Graph Convolution (AdaptConv) backbone for feature extraction. See Figure 2 for specific implementation details. For the standard pose \(X_i\) and the local features \(F_i\), we use AdaptConv to extract the feature of a point \(F_i\):

\[
e_{ijm} = g_m(\Delta f_{ij}), j \in P_i
\] (8)

Where \(g()\) is a mapping function using MLP, \(m\) represents one dimension of the output. We define \(\Delta f_{ij} = [f_i, f_j - f_i]\) to combine global and local information. The adaptive kernel in \(X_i\) can be represent as:

\[
h_{ijm} = \sigma(e_{ijm}, \Delta x_{ij})
\] (9)

where \(\langle \ldots \rangle\) is used for two vectors inner product, \(\sigma\) represents a activation function, \(\Delta x\) is equal to \([x_i, x_j - x_i]\). The final output for local features can be expressed as:

\[
f_i' = \text{MaxPool}(h_{ijm} + e_{ijm})
\] (10)

3.5 Feature Fusion

By integrating CSA, we propose Gometric Realition Transformer (GRT), it contains multi-headed cross-attention (MHCA) and multi-headed self-attention (MHSA), to adaptively utilize the correlation between point features for contextual enhancement. For features \(f_{i,j}\) and \(f_{i+1,j}\) from CSA, GRT firstly using cross-attention collects cross-context information between two point clouds. As shown in Figure 3. Specifically, we use the MHCA learning feature matrix \(F\) in the form of residuals:

\[
F = \text{Norm}(Q + \text{MultiHead}(Q,K,V))
\] (11)
where \( Q = f_{k+1,i}W_q \), \( K = f_{k,i}W_k \), \( V = f_{k,i}W_v \), \( W_q \) and \( W_k \) are linear transformation matrices, \( \text{MultiHead()} \) is a function similar to [35]. Then, a feedforward network (FNN) [35] is used to update the features of points:

\[
F_{\text{MHCA}} = \text{MHCA}(f_{k,i}, f_{k+1,i}) = \text{Concat}(F + \text{FFN}(f_{k,i}), Q)
\]

(12)

After this, each point in \( f_{k+1,i} \) can aggregate the features of similar points.

The MHSA is used to establish spatial relationships between points to improve the feature representation capability. Similar to the above Eq (11), but MHSA use the same features as Q, K, V. A FFN is also utilized here, thus the final output can be computed as:

\[
F_{\text{MHSA}} = \text{MHSA}(F_{\text{MHCA}}, F_{\text{MHCA}}) = F_{\text{MHCA}} + \text{FFN}(F_{\text{MHCA}})
\]

(13)

![Figure 3. TRFit’s geometric relation transformer layers. It is used for feature fusion.](image)

### 3.6 Loss Function

Similar to DeepFit [31], to train our proposed network, we use angle loss \( L_{\text{anger}} \), consistency loss \( L_{\text{con}} \). The angle loss measures the deviations of the ground truth normals and the estimated:

\[
L_{\text{anger}} = \left| N_{gt} \times N_i \right|
\]

(14)

while the consistency loss is used to constrain points on fitting surface:

\[
L_{\text{con}} = \frac{1}{N_p} \left[ \sum_{j=1}^{N_p} \log(w_j) + \sum_{j=1}^{N_p} w_j \left| N_{gt} \times N_j \right| \right]
\]

(15)

Then with the PointNet’s transformation matrix regularization loss \( L_{\text{reg}} = \left| I - AA^T \right| \). The final training loss is:

\[
L_{\text{total}} = L_{\text{anger}} + \alpha_1 L_{\text{con}} + \alpha_2 L_{\text{reg}}
\]

(16)

### 4. EXPERIMENTS

#### 4.1 Dataset and training details

We train and evaluate our suggested method using the PCPNet [15] dataset. Both CAD items and excellently scanned models are included in the file. Each shape has 100k densely sampled points overall. Adding noise and density for training data augmentation, we use PCPNet's methodology to assure fair comparison. We train the network across 600 epochs with the Adam optimizer [30], and the learning rate is \( 5 \times 10^{-4} \) with a batch size of 256.
4.2 Quantitative Evaluation

We compute the RMSE between the ground truth normals and the estimated normals using the PCPNet dataset to confirm the model’s validity. The comparison results of TRFit with others: AdaFit\textsuperscript{6}, DeepFit\textsuperscript{21}, IterNet\textsuperscript{20}, Nesti-Net\textsuperscript{18}, as well as classical methods: PCA\textsuperscript{7}, Jet\textsuperscript{11} in RMSE are shown in Table 1. As can be seen, TRFit performs better than all other methods at all noise levels, it achieves more robust surface fitting by acquiring more structural information.

<table>
<thead>
<tr>
<th>Method</th>
<th>PCA</th>
<th>Jet</th>
<th>PCPNet</th>
<th>Nesti-Net</th>
<th>IterNet</th>
<th>DeepFit</th>
<th>AdaFit</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>12.87</td>
<td>12.84</td>
<td>11.37</td>
<td>10.11</td>
<td>9.95</td>
<td>9.21</td>
<td>9.05</td>
<td>8.7</td>
</tr>
<tr>
<td>Med</td>
<td>18.38</td>
<td>18.33</td>
<td>18.87</td>
<td>17.63</td>
<td>17.18</td>
<td>16.72</td>
<td>16.44</td>
<td>16.05</td>
</tr>
<tr>
<td>High</td>
<td>27.5</td>
<td>27.68</td>
<td>23.28</td>
<td>22.28</td>
<td>21.96</td>
<td>23.12</td>
<td>21.94</td>
<td>21.55</td>
</tr>
<tr>
<td>Stripes</td>
<td>12.81</td>
<td>13.39</td>
<td>11.7</td>
<td>9.00</td>
<td>7.73</td>
<td>7.92</td>
<td>6.01</td>
<td>5.44</td>
</tr>
<tr>
<td>Gradients</td>
<td>13.66</td>
<td>13.13</td>
<td>11.18</td>
<td>8.47</td>
<td>7.51</td>
<td>7.31</td>
<td>5.90</td>
<td>5.18</td>
</tr>
<tr>
<td>Avg</td>
<td>16.25</td>
<td>16.29</td>
<td>14.34</td>
<td>12.41</td>
<td>11.84</td>
<td>11.8</td>
<td>10.76</td>
<td>10.23</td>
</tr>
</tbody>
</table>

Table 1. RMSE comparison of conventional and learning-based approaches for unoriented normal estimation on the PCPNet dataset.

We further assess the performance of normal estimate on PCPNet using the percentage of good points (PGP). The results are reported in Figure 4. It is illustrated in the figure that the proposed method, which has the best performance overall and can produce more precise normal estimation.

4.3 Qualitative Evaluation

Figure 5 visualizes the normals estimated by TRFit, the normal vector is mapped to the RGB cube. The right shows the visualization of angular errors on some shapes for TRFit and other methods, with colors corresponding to angular differences, being mapped to heat maps from 0 to 60 degrees. It should be aware that our strategy achieves a lower RMSE. Compared to the other approach, it is more able to keep crisp details of shapes.
4.4 Ablation study

Effect of the GRT module. Using models with or without GRT, default fixed order $n = 3$, and various neighborhoods $k = 256, 512,$ and $700$ points, we test on the PCPNet dataset. Table 2 reports the findings. From this table, it is clear that GRT layers can enhance normal estimate and eliminate the need to choose a certain neighborhood size.

<table>
<thead>
<tr>
<th>Scale</th>
<th>256</th>
<th>512</th>
<th>700</th>
<th>700</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdaptConv</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRT</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4.77</td>
<td>4.70</td>
<td>4.80</td>
<td>4.44</td>
</tr>
<tr>
<td>Low</td>
<td>8.83</td>
<td>8.72</td>
<td>8.59</td>
<td>8.70</td>
</tr>
<tr>
<td>Med</td>
<td>16.66</td>
<td>16.41</td>
<td>16.24</td>
<td>16.05</td>
</tr>
<tr>
<td>High</td>
<td>23.03</td>
<td>22.13</td>
<td>21.81</td>
<td>21.55</td>
</tr>
<tr>
<td>Stripes</td>
<td>5.69</td>
<td>5.66</td>
<td>5.78</td>
<td>5.44</td>
</tr>
<tr>
<td>Gradients</td>
<td>5.50</td>
<td>5.44</td>
<td>5.52</td>
<td>5.18</td>
</tr>
<tr>
<td>Avg</td>
<td>10.75</td>
<td>10.51</td>
<td>10.46</td>
<td>10.23</td>
</tr>
</tbody>
</table>

Table 2. Normal RMSE of models with or without GRT layers.

Table 3. RMSE comparison with different Jet order.

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdaptConv</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>GRT</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>5.36</td>
<td>5.32</td>
<td>5.05</td>
<td>4.68</td>
</tr>
<tr>
<td>Low</td>
<td>8.96</td>
<td>8.94</td>
<td>8.74</td>
<td>8.80</td>
</tr>
<tr>
<td>Stripes</td>
<td>6.45</td>
<td>6.56</td>
<td>6.02</td>
<td>5.71</td>
</tr>
<tr>
<td>Gradients</td>
<td>5.93</td>
<td>6.06</td>
<td>5.61</td>
<td>5.32</td>
</tr>
<tr>
<td>Avg</td>
<td>10.80</td>
<td>10.81</td>
<td>10.61</td>
<td>10.35</td>
</tr>
</tbody>
</table>

Different n-Jet orders. We set $n = 1, 2, 3, 4$ and the neighborhood size is fixed as 700 points to test the performance of normal estimate using RMSE for various n-jet orders. The RMSE of different orders is displayed in Table 3. The findings demonstrate that when the jet order rises, the suggested approach can estimate the normals more precisely.
4.5 Application

Surface reconstruction and denoising are important tasks in 3D vision. In Figure 6, we show visualization of Poisson reconstruction and Lowrank denoising using estimated normals. As observed, our method preserves the full structure of the object and provides clear edges.

![Figure 6. Visualization of Poisson reconstruction (left) and denoising (right).](image)

5. CONCLUSION

We present a deep learning method for unstructured point cloud normal estimation, this method focuses on point cloud geometry learning. We enhance the point features learning by adaptive graph convolution. At the same time, we introduce a transformer-based multi-scale feature fusion module to further enrich the geometric features and enhance the robustness. It is experimentally demonstrated that the proposed method achieves optimal results with different levels of noise interference on the synthetic dataset PCPNet. It shows a fairly stable generalization ability that it can retain clear features on the real scanned data.

REFERENCES


Government and Enterprise Data Security Sharing Method Based on Differential Privacy Protection

Xiaomin Xu, Zhenglei Zhu*, Jiange Liu, Xin Liu, Qingxuan Guo

1STATE GRID HUAIAN POWER SUPPORT COMPANY, Huaian, Jiangsu, China, Postal Code 223002
*diamondhero@163.com

Abstract

The security of government and enterprise data sharing is very important and critical. To increase the number of data sharing, create a more stable transmission and processing environment, and reduce network threats, this paper studies the security sharing method of government and enterprise data based on differential privacy protection. Firstly, the government and enterprise data encryption is described, and the Tendermint differential overlapping interactive sharing nodes are deployed in the preset area. Then, based on this, the interactive differential privacy protection data sharing model is designed, and the threat identification method is used to achieve data security sharing. The experimental results show that compared with the traditional proxy encryption data security sharing test group and the traditional CP-ABE data security sharing test group, the differential privacy protection sharing test group designed in this paper achieves relatively more times of one-way data security sharing, which indicates that the proposed method has a smaller error and fast speed in the actual data transmission process. The data processing in the region has less restrictive conditions and has practical application value.

Keywords-Differential privacy protection, Government and enterprise data, Secure sharing, Sharing method, Privacy protection, Data encryption

1. Introduction

Differential privacy is a new definition of privacy proposed by Dwork in 2006 to solve the problem of privacy disclosure of statistical databases. Under this definition, the calculation result of the data set is not sensitive to the change of a specific record, and a single record in or out of the data set has little effect on the calculation result[1]. Therefore, the risk of privacy disclosure caused by adding a record to the data set is controlled within a very small and acceptable range, and the attacker can not obtain accurate individual information by observing the calculation results.

Differential privacy can solve two defects of the traditional privacy protection model. First, the differential privacy protection model assumes that the attacker can obtain the information of all other records except the target record, and the sum of these information can be understood as the maximum background knowledge that the attacker can grasp. Under this maximum background knowledge assumption, the differential privacy protection does not need to consider any possible background knowledge possessed by the attacker, because this background knowledge cannot provide more information than the maximum background knowledge[12]. Secondly, it is based on a solid mathematical foundation, defines privacy protection strictly and provides a quantitative evaluation method, which makes the level of privacy protection provided by data sets with different parameters comparable. Therefore, it has gradually become a research hotspot in the field of privacy protection. In recent years, the combination of differential privacy and other areas of research has led to a large number of new results.

The essence of government-enterprise data sharing is the two-way flow of data between the two major market players of the government and enterprises, with the aim of releasing greater data value through the sharing and integration of data between the two sides to achieve the effect of $1 + 1 > 2$. At present, China's government-enterprise data fusion has been explored, and the direction, method and operation mode of data flow has changed. First, data flow is no longer a one-way flow from enterprises to the government but gradually transforms into a two-way flow between government and enterprises[3]. The most common way is to provide convenient conditions for enterprises to obtain government data through the open government data platform. Second, data sharing is undertaken by the platform. Due to the exponential growth of both government and enterprise data, promoting sharing and opening by opening data exchange interfaces on a unified platform has become the main method of data sharing. Third, the operation mode is no longer just government-led, government-enterprise cooperation has become the mainstream. The two-way flow of data between the
government platform and the power company is the basis for building a service platform, through the integration and sharing of data between the two sides, release greater data value.

In recent years, with the continuous development of the Internet and computer technology in China, the efficiency of data sharing has been greatly improved, and the security of network data has also changed. Usually, in the practical application scenarios of edge networks, data sharing is generally highly targeted. One-to-one docking is also required during transmission, especially for government-enterprise data, which has a higher degree of privacy, and some data will even affect the future development prospects of government and enterprises. Thus, the correctness and confidentiality of data must be guaranteed in the process of processing. Secondly, the sharing of government and enterprise data generally needs to be in a specific website or channel. On the premise of ensuring data integrity, it is also necessary to eliminate the security risks of data transmission and reduce the data distortion rate. Although the traditional multi-source heterogeneous data sharing method can achieve the preset goal of data sharing, it has many uncontrollable factors and is vulnerable to external factors, so it can not better ensure the security and stability of data [4].

Therefore, this paper combines the differential privacy protection technology to design a secure data sharing method between government and enterprises. Considering the stability and reliability of the final test results, it is necessary to select more authentic government and enterprise data as the main target of the test. Then, according to the transmission range of data, the corresponding shared nodes are deployed, and the differential privacy protection program is combined with the distributed data scheduling mode. Thirdly, we should further balance the contradiction between the privacy protection of government and enterprise data and the availability of data, strengthen the control of security threats and data leakage risks, and further avoid the omission of data transmission monitoring[5]. Finally, the association of government-enterprise data sharing mechanism is realized to provide a reference for subsequent technological innovation.

2. Design of government and enterprise data differential privacy protection sharing method

2.1 Description of encryption

Before the design of the government-enterprise data differential privacy protection sharing method, according to the degree of privacy of government-enterprise data, the application encryption level is set, and different types of data are encrypted and described to create a basic data sharing environment. The transmission space can be constructed on the government-enterprise data processing platform, and the ciphertext can be described at the same time [6].

Using the password mechanism to protect data, this part needs to set the access ciphertext program first, then carry out indirect data format conversion, and finally gradually form the attribute set of data [7]. Set the transmission position and receiving position of the data in the platform, authorize the attribute to access the ciphertext coverage area, produce the key of the data ciphertext, and calculate the one-way set element of the key, as shown in Formula 1 below:

\[ F = \frac{y_1 y_2 \times \int w \chi}{\chi(y_2 - 9R^2)} + y_1 \tag{1} \]

In Formula 1, \( F \) represents the key unidirectional set element, \( \chi \) represents the integration range, \( y_1 \) represents the key identification radius, \( y_2 \) represents the key overlap radius, \( R \) represents the set element difference, and \( w \) represents the ciphertext error. Through the above calculation, the actual one-way set element of the key can be finally obtained, the identifiable area of the data ciphertext is set in combination with the change of the one-way set element[8]. At the same time, the data sharing program is set on the receiving position corresponding to the platform to perform coverage scanning on the government and enterprise data in the platform to create a stable and safe data sharing environment, which lays the foundation for the execution of subsequent shared work.

2.2 Tendermint differential overlap sharing interactive node layout

After the encryption description of government and enterprise data is completed, according to the data conversion requirements and standards, combined with the differential privacy protection technology, a certain number of Tendermint differential overlapping sharing nodes are deployed in the calibrated network data transmission area [9]. The C/S architecture can be used to build a centralized interactive data sharing blockchain [10], and an auxiliary overlapping monitoring interactive node can be built on the periphery of the platform. In this part, it should be noted that the
overlapping monitoring nodes will not cause damage to the protected data, and the degree of restriction is not high. Therefore, in the actual acquisition process, auxiliary programs can be added appropriately to protect the data in the second order.

The set alliance data chain or data private chain is interactively bound to form a fixed data access area, and at the same time, a consensus data sharing structure is constructed by combining the differential privacy protection technology and the Tendemint overlapping processing method, as shown in Figure 1 below:

![Figure 1 Tendemint overlapping differential overlapping sharing architecture](image)

According to Figure 1, the design and adjustment of Tendemint overlapping differential overlapping sharing structure can be completed, and then on this basis, the corresponding nodes are set in the corresponding network program or block chain one by one. The interactive consensus efficiency of government and enterprise data is measured at this time, which can be controlled above 92.25%. Finally, the layout of the Tendemint differential overlapping interactive sharing nodes is completed.

### 2.3 Design of interactive differential privacy preserving data sharing model

After the completion of the Tendemint differential overlapping sharing node deployment, combined with the differential privacy protection technology, a more targeted and interactive data sharing model is built. Usually, for the safe sharing of government and enterprise data, it is necessary to design specific transmission and identification instruction protocols to enhance the reliability and authenticity of data[11].

However, due to the increasing data of the government and enterprises in recent years, the basic sharing program can not meet the needs of work. Therefore, we need to build a more flexible and powerful data sharing model. The key of the platform can be modified to the multi-creation mode first, and the key generation corresponds to the interactive control module, and then an interactive data sharing model structure is established by using the differential privacy protection technology according to the data and information collected by the nodes, as shown in Figure 2 below:
According to Figure 2, the design of the bidirectional differential privacy protection data sharing model structure can be completed. In combination with the above-arranged data sharing and monitoring nodes, the corresponding data interaction and sharing range is adjusted at all times, and a better interaction and sharing environment is created to improve the security of data transmission.

2.4 Threat identification method to achieve secure data sharing

After completing the design of the two-way differential privacy protection data sharing model, combined with the actual data sharing needs, the threat identification method is used to complete the data sharing. The command design platform can be used to construct a threat identification program, which is set in the shared model. Within the scope of calibration, the corresponding identification node is accessed. Through data and information, the transmission rules and characteristics of government and enterprise data are extracted, and the corresponding threat protection program is designed in combination with the characteristics.

It should be noted that the design of this kind of program is generally highly targeted, only for the data with obvious characteristics. The degree of security will be higher, and the damage and impact of data will be more controllable, which can help government enterprises to complete data transmission and conversion quickly and stably, and improve the overall work efficiency.

3. Experimental test and analysis

This paper mainly analyzes and studies the practical application effect of the government-enterprise data differential privacy protection sharing method. Considering the stability and authenticity of the final test results, the internal data platform of G enterprise is selected as the main target object of the test[12]. Set the traditional proxy encryption data security sharing test group, the traditional CP-ABE data security sharing test group and the differential privacy protection sharing test group designed in this paper. The test is analyzed in the form of comparison, and the final test results are verified by comparison. Next, set up the corresponding test environment.

3.1 Test preparation

Before analyzing and studying the practical application effect of the government-enterprise data differential privacy protection sharing method, it is necessary to build a test environment. Firstly, the basic data information is downloaded and imported into the government-enterprise platform, and after appropriate conversion, it is set within the control range in a standard format. Then, the data is adjusted to four data packets as the main target object of the test, and at this time, the calibration characteristic value of the data packet needs to be calculated, as shown in Formula 2 below:
\[ M = \kappa^3 + \left( \sum_{i=1}^n d^2 + \rho i \right) - 1 \]  

(2)

In Formula 2, \( M \) represents the calibration characteristic value of the data packet, \( \kappa \) represents the adjustment range, \( d \) represents the strain difference, \( \rho \) represents the differential distance, and \( i \) represents the number of sharing times. The actual characteristic value can be obtained by measurement, and the data packet is classified according to the characteristic for testing. Subsequently, on the G enterprise platform, monitoring nodes are set to facilitate the collection and summary of data, and the construction of the test environment is completed.

### 3.2 Test process and result analysis

In the above test environment, combined with the differential privacy protection technology, the data sharing effect of G enterprise is measured[13]. The four data packages are set in different areas of the platform, and 100 users are selected to complete the basic processing of the data and transmit it to the database preset by the enterprise.

At this time, the total number of users that can be shared by different test groups in the same environment can be measured, as shown in Formula 3 below:

\[ B = \sqrt{\mu_1 - (0.5 \tau^2 \times \mu_2)} + \mu_1 \mu_2^2 \]  

(3)

In Formula 3, \( B \) represents the number of shared users, \( \mu_1 \) represents the preset coverage, \( \mu_2 \) represents the measured coverage, and \( \tau \) represents the difference value. The actual test results can be obtained through the test.

The comparative analysis is shown in Table 1 below:

<table>
<thead>
<tr>
<th>Test packet</th>
<th>Traditional Agent Encrypted Data Secure Sharing Test Group Sharing Times (times)</th>
<th>Traditional CP-ABE data security sharing test group sharing times (times)</th>
<th>Differential privacy protection sharing test group sharing times (times)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Packet 1</td>
<td>76</td>
<td>81</td>
<td>90</td>
</tr>
<tr>
<td>Data Packet 2</td>
<td>85</td>
<td>74</td>
<td>96</td>
</tr>
<tr>
<td>Data Packet 3</td>
<td>73</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>Data packet 4</td>
<td>79</td>
<td>73</td>
<td>93</td>
</tr>
</tbody>
</table>

According to Table 1, the analysis of the test results can be completed. Compared with the traditional proxy encryption data security sharing test group and the traditional CP-ABE data security sharing test group, the differential privacy protection sharing test group designed in this paper finally realizes relatively more times of one-way data security sharing, which indicates that in the actual data transmission process of G enterprise. The error is small and the speed is fast. The data processing in the region has less restrictive conditions and has practical application value.

### 4. Conclusion

The above is the design and research of government and enterprise data security sharing methods based on differential privacy protection. Compared with the traditional data sharing method, this paper combines the differential privacy protection technology to gradually build a more flexible and changeable data sharing program. In the design, the privacy and security of government and enterprise data are taken into account, and the protection means in data sharing are increased. At the same time, in the calibrated transmission channel and sharing area, a stable data sharing stream is constructed in combination with differential protection, which reduces the distortion of data as a whole, further strengthens the security of sharing, and creates a more stable and reliable environment for government and enterprise office.
Reference

Research on Information Security Architecture under the New Generation Marketing System

Shuang Yang\textsuperscript{a}, Haomiao Wang\textsuperscript{a*a}, Tianbo Wang\textsuperscript{a}, Zhibin Wang\textsuperscript{a}, Naidi Kang\textsuperscript{b}

\textsuperscript{a}State Grid Liaoning Electric Power Co., LTD, Shenyang, Liaoning, CHINA 110006;
\textsuperscript{b}State Grid Liaoning Electric Power Co., LTD. Marketing Service Center., Shenyang, Liaoning CHINA 110168

* Corresponding author: whm0505@139.com

ABSTRACT

Under the background of today's era, the development of power grid intelligence is an inevitable trend in the development of power companies. In this process, what the power industry needs to do is to do a good job in the prevention of smart grid information security risks, so as to avoid power companies due to excessive power grid information security risks bring certain losses. The power marketing system is a vital part of the power system. Starting from the marketing system, combined with the current business situation, we will study the development trend and overall construction plan of the new generation of marketing business application systems, and further study the security architecture of the system. From the perspective of multi-faceted security management and control, to implement the relevant national and company-related personal information protection requirements, and to improve the information security of the new generation of marketing systems has become an urgent problem to be solved.

Keywords: Marketing system, information security, Electricity Marketing

1. INTRODUCTION

Electric power marketing is a business activity carried out by electric power enterprises with the goal of meeting customer needs [1]. Under the new situation, electric power marketing activities must provide customers with safe, reliable and economical services. At present, most of the power companies use the traditional power marketing model, which does not fully meet the customer's service requirements for information security. In this regard, electric power enterprises propose to optimize the power marketing business, and design a more secure and reliable power marketing business system by introducing advanced technology to provide consumers with more satisfactory services and create more benefits for enterprises [2]. At the same time, with the rapid economic growth, the society's demand for electricity is also increasing. In order to meet the demand for electricity in the whole society, we must formulate reasonable and effective marketing methods, use advanced information security technology to build a marketing system with complete functions and leading technology, and grasp the demand situation of the electricity market while keeping abreast of market changes in a timely manner.

2. NEW GENERATION MARKETING BUSINESS SYSTEM ARCHITECTURE

With the rapid development of smart grid construction, more and more smart power terminal devices have been introduced into the marketing business system. At the same time, in the development process of smart grid, information technology plays an important role in the construction, operation and management of smart grid as a supporting technology. The close connection between power network and communication network not only brings about technological innovation of power business system, but also provides Electricity network security brings great challenges.

2.1 Data Architecture

Nowadays, the power grid industry is gradually integrating new media information technology, and the power industry will generate a large amount of power information data in the actual operation process. The processing of power marketing data has become an important issue that needs to be properly handled. Economic and social development has brought about changes in technological revolution, which can effectively improve the power information data processing capability and improve user data caching services. It is an important form of cooperation between power companies and power users. According to the needs of users, enterprises provide power commodities that meet the needs of users and can ensure their safety and reliability, and provide guarantees for subsequent services.
The power marketing link is a very important link in the power enterprise. Solving the development and service of this link will greatly improve the operation efficiency of the power grid enterprise, and the service level will also be effectively improved. Data integration technology can uniformly process different types of scattered data sources, realize classification processing according to different processing requirements, and break the logical difference between data. Traditional data integration technology includes multi-database system and database system of Internet of Things technology[3]. With the development of technology, data integration technology has also improved, and this advanced technology has been widely used in power grid enterprises.

In order to realize the design of the system architecture and meet the compatibility requirements, the system hardware is designed with multiple modules, as shown in the figure.

![Figure 1. Hardware structure of power marketing data analysis system](image-url)

1) Data acquisition module

The data acquisition module mainly collects and mines data resources of various parts, and is used to realize functions such as data maintenance, selecting specific data sets, and viewing data attributes. In order to select specific target data, the staff has the management and maintenance authority for this part of the data, while the requirements for the data analysis module are more stringent, and only some staff with management and analysis authority can have this authority. In hardware systems, data preprocessing is very necessary, which can make data with different properties and more dispersed data. The structure of the data acquisition module is shown in the figure.

![Figure 2. Data acquisition module structure](image-url)

In the data acquisition module, the micro-control unit can distinguish the data from complex sources. When encountering duplicate data, abnormal data and intrusion data that may endanger the security of the system, it should be screened...
immediately, and the data that meets the requirements will be put into the database to ensure data collection accuracy, so that different types of data can be entered more clearly in the mining and results analysis modules. The asynchronous transceiver inside the data acquisition module retrieves, inquires, and outputs the collected data, so that the staff can query and output the data in various types, ensuring that the whole process is smoother and more convenient.

2) Data mining module

In the whole hardware system design, data mining module is an important part. The data mining module is an important step in the secondary processing of user data resources after the completion of the data acquisition module, and plays a linking role in the entire hardware system processing process. The data mining module can be further divided according to different categories of users, and analyzed according to the past statistical data to predict the high-risk users of electricity arrears. On the basis of data collection, secondary analysis and processing is the focus of the whole module. In data mining, the two parts of data analysis and management are connected, and users are segmented through cluster analysis technology. The database inside the data mining module can provide corresponding ports for other subprograms to access and read data, so as to ensure that these ports have sufficient hardware interfaces. The circuit diagram of the data mining module is shown in the figure.

![Data mining module circuit diagram](image)

Figure 3. Data mining module circuit diagram

3) Data result analysis module

In terms of data extraction, data verification, data fusion, comparative analysis, etc., the data result analysis module has various requirements. The data result analysis module mainly collects and evaluates the system data set, generates reports and compares it, conducts high-standard data analysis on the data generated in different situations in the hardware system, outputs and organizes the data obtained under different conditions, and presents complete information to the user. The structure of the data result analysis module is shown in the figure.
When the user enters the corresponding keywords and data, the system can integrate the data according to the user's needs, analyze the data in a more intuitive way, realize the output of statistical information, and display it to the user. In the current system experiment process, after entering the corresponding name and keywords, the system can generate the corresponding data distribution diagram, and then start the specific data analysis model, and display the analysis results to the user through the browser, and can display the analysis results through the histogram, bar graph, bar graph, etc. Graphs and other forms are used to compare and analyze the data and present the results to the user.

2.2 Application Architecture

The basic content of electricity marketing can be summarized into three parts, one is the statistics of users' electricity consumption, that is, billing management; The third is to charge various tariffs, such as electricity fees, installation of power supply equipment and other fees, that is, accounting management[4]. Accounting management is the final link to complete commodity conversion, and it is interrelated with expansion and billing management.

According to the current grouping of marketing work, business expansion reporting and installation can be divided into the application of electricity for new users, the installation of power supply equipment, the inquiry of electricity consumption, the change of electricity consumption of old users, and the acceptance of business applications; billing management can be divided into several parts. It consists of three parts: meter reading, billing management and storage of electricity fee information files; financial management is further divided into electricity fee statistics inquiry, bank collection of electricity fees, on-site charges in business halls and other payment methods. The overall business application architecture of power marketing is shown in the following figure.
3. NEW GENERATION MARKETING SYSTEM INFORMATION SECURITY PROTECTION SYSTEM

The information security protection of the new generation marketing system is jointly guaranteed by the technical protection system and the security management system. The overall structure diagram is shown in the figure. The technical protection system provides technical protection from the terminal layer, network layer, platform layer, and application layer; the safety management system puts forward requirements for safety responsibilities and management systems, and combines with the technical protection system to form a complete protection system.
Figure 6. Marketing system information security protection system

3.1 Technology System

The technical protection system of the information security protection of the new generation marketing system carries out technical protection from the terminal layer, the network layer, the platform layer and the application layer. In terms of terminal security, it is necessary to protect the physical environment of the terminal, pay attention to the selection of physical location, anti-theft, anti-damage, waterproof and moisture-proof, etc.; The system and edge IoT agents are protected; the transmission and access of terminal data are protected; the terminal itself and on-site traffic are monitored. In terms of network security, the focus is on secure access and access control, and clear channel-specific and secure partitions to ensure transmission isolation and confidentiality. In terms of platform security, focus on the security of the platform itself, security situation awareness, dynamic and active early warning capabilities, to achieve attack source traceability, and to desensitize and prevent platform data from leaking. In terms of application security, protect the security of system applications and mobile applications, and protect the integrity and confidentiality of application data[5].

3.2 Management System

The management system of information security protection of the new generation marketing system consists of three aspects: security responsibility, management system, and terminal layer construction and operation security. Security responsibilities follow the principle of "whoever is in charge is responsible, whoever runs is responsible, whoever uses is responsible, and business management must manage security", and strictly implement network security responsibilities and management responsibilities; management systems include personnel security management, system security management, and equipment security. Management and data security management are formulated to form a fully closed-loop management; relevant requirements are put forward for the construction and operation security of the terminal layer, and the construction and operation and maintenance of the technical protection system are supported.
The power enterprise network information system is a complex technical system, so the existence of security risks in the system is unavoidable. To strengthen the power enterprise network information security management, we must first assess the security risks of the information system. Enterprises must ensure the scientific nature of the information structure and control the network information security risks to a minimum. Secondly, enterprises should build a complete network virus defense system to prevent the impact of computer virus intrusion on network security management, and promote the more stable operation of the power enterprise network information security system[6].

It is also necessary to reasonably divide the network security area of the power grid enterprise, and scientifically and reasonably divide the network security area according to the security classification and security planning of the network information of each part of the power grid enterprise. Usually, the network security area of the power grid enterprise can be divided into three parts. That is to say, the key prevention area, the general prevention area and the completely open area can realize the security management of the network information of the power grid enterprise, so that the work of each network area can be carried out smoothly.

4. CONCLUSION

For the information security of the new generation marketing system, this paper expounds the system architecture of the new generation of marketing business system from the two dimensions of data architecture and application architecture. The overall information security architecture of the new generation marketing business system is discussed. The information security protection system architecture of the new generation marketing system is described from the technical system and management system of information security protection. The technical protection system provides technical protection from the terminal layer, network layer, platform layer, and application layer; the safety management system puts forward requirements for safety responsibilities and management systems, and combines with the technical protection system to form a complete protection system. The architecture proposed in this paper provides an important foundation for the future application of marketing business.

ACKNOWLEDGEMENTS

We would like to express our heartfelt thanks to all those who have helped this research work. This project is supported by the "Exploration and Research on Information Security Architecture Based on New Generation Marketing System" project.

REFERENCES

Design and Application of Rhythmic Gymnastics Auxiliary Training System Based on Kinect

Qi Zhuang
Qilu Normal University, Shandong, Jinan, 250000
80366647@qq.com

Abstract

Based on Microsoft Kinect equipment, this paper adopts motion capture technology to collect human motion data in the actual training process of rhythmic gymnastics, designs and develops the functional module of motion data analysis combined with ASP.NET development environment, and integrates and packages other auxiliary functional modules to form an auxiliary training system of rhythmic gymnastics. Aiming at all kinds of problems in the teaching practice of rhythmic gymnastics in colleges and universities, the system will put forward a comprehensive application solution, which takes the physical movements, equipment movements and self-selected complete movements of rhythmic gymnastics students in daily training in colleges and universities as the research objects, takes the transformation and analysis of movement information data as the method, and takes the visual interaction at the network end as the display form. It greatly facilitates the practical application of students and teachers, effectively improves the training effect of students, achieves the purpose of auxiliary training, optimizes the teaching process of rhythmic gymnastics education, and further promotes the process of information teaching construction in colleges and universities.

Keywords: Kinect; rhythmic gymnastics; ASP.NET; auxiliary training system; action evaluation

1. Introduction

In recent years, rhythmic gymnastics has shown a positive development trend. On the basis of the prescribed movements, individual events or collective events have added, changed or innovated technical movements in a guiding way, and paid attention to the standardization and promotion of techniques. In addition, the "2022-2024 Rhythmic Gymnastics Scoring Rules" promulgated by the International Sports Federation also redefined the changes of rhythmic gymnastics scoring rules and the latest technical trends in the sexual cycle. The newly added artistic score, the joint physical difficulty and the newly added movements pose new requirements for the current education and teaching of rhythmic gymnastics in colleges and universities. [1]

Colleges and universities should actively seek new changes, combine the current background of the national integration of sports and education and the construction of "four new" disciplines, make overall layout according to the current development situation and future development plan of rhythmic gymnastics in China, and focus on the reform and optimization of teaching methods, training concepts, theoretical knowledge, technical skills, evaluation and assessment, etc., highlight the cross-integration of disciplines, and strive to improve all kinds of shortcomings existing in the actual teaching process. So as to strengthen students' comprehensive ability, improve the effectiveness of course teaching, and further build a new teaching reform, new quality and new system of physical education in the new period [2]. In view of this, this paper believes that the innovative integration of network information, motion capture, visual interaction, Web application design and construction and other technical means, with Microsoft Kinect equipment as hardware support, combined with ASP.NET web application development framework, will complete the construction of rhythmic gymnastics auxiliary training system. According to the actual needs of students and teachers, the system will complete the virtual construction and structural optimization of the teaching process of rhythmic gymnastics and put forward comprehensive application solutions to effectively improve various shortcomings existing in the current teaching practice of rhythmic gymnastics, which will promote the reform of rhythmic gymnastics education mode and promote the development process of information construction of higher education.
2. Introduction of key technologies

2.1 Motion capture technology

Motion capture technology refers to a technology that uses special instruments or equipment to track, measure and record the motion trajectories of human bodies and other objects. Its core application lies in the collection of motion information, and the mathematical way to transform, express or output it. From a professional point of view, motion capture is a high-tech that can accurately measure and record all kinds of motion trajectories and postures of a moving object in the actual three-dimensional space in real time, and reconstruct the motion state of the object at every moment in the virtual three-dimensional space by combining data information such as size measurement, positioning and orientation determination of the object in the physical space [3]. In this paper, the rhythmic gymnastics auxiliary training system will adopt the 3D depth information recognition motion capture technology and rely on Microsoft Kinect equipment to complete the motion capture.

2.2 Kinect

Kinect is a somatosensory device designed and developed by Microsoft in 2010. Its core function is 3D somatosensory camera, which helps gamers to control the game through body movements or voice interaction by using functions such as dynamic capture, image recognition, audio input and voice recognition [4].

2.2.1 Equipment structure

Compared with the previous version, Kinect v2.0 is improved in configuration and running development environment. See Table 1 for details.

<table>
<thead>
<tr>
<th></th>
<th>Resolution</th>
<th>1920*1080</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color camera</td>
<td>fps</td>
<td>30fps</td>
</tr>
<tr>
<td>Deep camera</td>
<td>Resolution</td>
<td>512*424</td>
</tr>
<tr>
<td>Number of characters</td>
<td></td>
<td>Six people</td>
</tr>
<tr>
<td>Character posture</td>
<td></td>
<td>Six people</td>
</tr>
<tr>
<td>Joint</td>
<td></td>
<td>25 Joint / person</td>
</tr>
<tr>
<td>Scope of detection</td>
<td></td>
<td>0.5 ~ 4.5m</td>
</tr>
<tr>
<td>Angle</td>
<td></td>
<td>70 degree</td>
</tr>
<tr>
<td></td>
<td>Horizontal</td>
<td>60 degree</td>
</tr>
<tr>
<td></td>
<td>Perpendicular</td>
<td></td>
</tr>
</tbody>
</table>

2.2.2 Working principle

Depth detection is the core technology of Kinect, and developers can get the depth information of users and judge their position through Kinect. When the human body moves and displaces, it will automatically construct three-dimensional space coordinates under Kinect equipment, so as to obtain the position coordinates of the human body and the joint movement changes, and accurately evaluate the human body's movement state with sufficient data information to complete the human body movement capture.

2.3 Development process

According to the requirements of the above related application technologies, this paper completes the construction and deployment of the development environment of rhythmic gymnastics auxiliary training system. The system development environment is divided into two parts, one is the hardware configuration based on Kinect v2.0; Second, analyze Kinect v2.0 data, develop application function modules, and integrate other function modules to complete the package and release of the whole system. As for the hardware configuration of Kinect v2.0 device, the underlying operating system is Windows 10.0, 64-bit operating system, with 16 cores and 24 threads of 3.20 GHz Intel Core i9-12900K processor. Adapted to USB 3.0 bus driver, 4 GB RAM and DX11 graphics adapter.

The application of Kinect data analysis depends on the support of Kinect for Windows SDK, Visual Studio 2019,.NET Framework 4.7 and C# language. Under the computer system, first complete the download and installation of Kinect for Windows SDK 2.0 and OpenCV 3.4.1, so as to realize the connection between Kinect v2.0 devices and the computer system. Then, under Visual Studio 2019, create a new project console application and open the property option to complete the corresponding kinect20.lib and. html files, and complete the deployment of the development environment.
of Kinect v2.0 devices [5]. After the basic deployment is completed, users can test the depth and distance of the equipment by code to prove whether the development environment has been successfully configured. The key code of the test part is shown in Figure 1.

```c
DepthFrame* pFrame = nullptr;
if (pFrameReader->AcquireLatestFrame(&pFrame) == S_OK)
{
    int iWidth = 0;
    int iHeight = 0;
    IDriverDescription* pDriverDescription = nullptr;
    pFrame->get_FrameDescription(&pDriverDescription);
    pDriverDescription->get_Width(&iWidth);
    pDriverDescription->get_Height(&iHeight);
    pDriverDescription->Release();
    pFrameDescription = nullptr;
    UINT uBufferSize = 0;
    UINT16* pBuffer = nullptr;
    pBuffer->AccessUnderlyingBuffer(&uBufferSize, &pBuffer);
    int x = iWidth / 2;
    y = iHeight / 2;
    size_t idx = x * iHeight + y;
    std::cout << pBuffer[idx] << std::endl;
    // 4e release frame
    pFrame->Release();
    pFrame = nullptr;
    ++uFrameCount;
}
```

Figure 1: Key code for testing the depth distance of Kinect v2.0 equipment

For the rhythmic gymnastics auxiliary training system, we choose the C# language environment under the ASP.NET framework to complete the system construction, complete the corresponding Web application package, and publish it on the Windows IIS 10.0 server. In addition, the system will adopt SQL Server 2016 as the database server, providing necessary space support for the data storage and call of the whole system. Through the introduction of the above key technical theories, we have determined the overall environment of the system development, the configuration of related software and tools, and the technical feasibility of the overall project of the rhythmic gymnastics auxiliary training system.

4. Detailed function implementation

4.1 Teacher Client

When a teacher starts the rhythmic gymnastics auxiliary training system, it is necessary to arrange Kinect v2.0 equipment in the training ground in advance to determine the motion capture area. When the teacher clicks the test start in the system, the Kinectv2.0 device will start automatically, and complete the capture of students' body movements, equipment technical movements and complete sets of technical movements in rhythmic gymnastics training. Among them, body movements include long-span jump, deer jump, Basse turn, etc. Technical movements of instruments include rope, circle, ball, stick and belt. The complete sets of movements include 5 laps, 3 belts and 2 balls in the new periodic rules, etc. Teachers can timely extract the test data information in Kinectv2.0 and save it in the system database for subsequent analysis and evaluation.

Under the data processing module, when the teacher user clicks "Data Processing", the system will automatically filter and reduce noise, reduce data jitter and predict the three-dimensional position data of the bone joints of the students during rhythmic gymnastics training, and calculate the instantaneous speed and joint angle of each joint in each frame according to the position information to represent the technical movement characteristics of the students [6]. The technical movement characteristics of rhythmic gymnastics will cover 25 joints of the whole body, but only some joints will move and change. For example, in the long-span jump, the movements of the head and neck do not change much before and after the movement is completed, so it is necessary to select limbs, waist and crotch as the core features to complete the follow-up evaluation and examination.

Under the data analysis module, the teacher user compares the selected joint points with the corresponding data in the standard movement feature database to obtain the similarity calculation result between the student movement feature
data and the standard movement feature data. Synthesize the comparison data of all joints, and its calculation formula is shown in Formula 1, where $a_j$ represents the angle of each joint of the trainee, $\beta_j$ represents the angle of each joint of the standard movement, and $k$ represents the set of joints involved in the comparison of angle data [7].

$$
SCA = \sum_{j \in k} \frac{|a_j - \beta_j|}{180^\circ} \tag{1}
$$

The comparison results of the joint angle data are displayed visually, that is, the joint points are taken as the horizontal axis, and the normalized calculation results of the student's bone joint points and the standard movement bone joint points are drawn as the vertical axis, showing the curve of the similarity of the two joint angle data, as shown in Figure 2. Among them, the relatively gentle curve is the angle curve of each joint in the standard movement, while the curve with large slope change is the angle curve of each joint in the technical movement of students. The higher the coincidence degree of the two curves, the closer the angle of each joint of students is to that of the standard movement, and the higher the accuracy of the movement. Otherwise, the quality of movement completion needs to be adjusted and changed.

![Figure 2: Evaluation result of long-span jump action](image)

4.2 Student Client

Under the student client, students can consolidate and review the theoretical knowledge of rhythmic gymnastics through the knowledge learning module, and can also use various resources uploaded by teachers to improve their professional quality and their understanding and feelings of culture and art. For evaluation and assessment, students can not only see their technical problems in the training process more clearly and intuitively, but also get professional guidance from teachers in time. The scientific and data-based analysis and comparison results not only improve the training effect, but also can form the examination process into a video file, which greatly improves the shortcomings of the current teaching process and provides substantial help for the improvement of students' artistic gymnastics professional level.

5. Conclusion

Based on the actual needs, this paper aims at various problems in the teaching practice of rhythmic gymnastics in colleges and universities, and completes the construction of the auxiliary training system of rhythmic gymnastics. The system integrates the application advantages of various information technologies such as motion capture, network information, data analysis and application, visual interaction, etc., and effectively promotes the reform of teaching mode. It further improves the teaching system of rhythmic gymnastics and organically combines theoretical knowledge learning, practical training and assessment, which is not only conducive to the cultivation of national high-level athletes, but also conducive to the popularization of rhythmic gymnastics. In the follow-up research, we need to continue to expand the technical means of data collection, analysis and processing, reduce the error, and make the technical action analysis more accurate and comprehensive.
Reference


Stacked hourglass deep learning networks based on attention mechanism in multi-person pose estimation

Jiazhi Di1,a, Ben Wang1,b*, Hua Hu1,c
1School of Information Science and Engineering, Hangzhou Normal University, Hangzhou, Zhejiang, 311121, China
a*2020112011016@stu.hznu.edu.cn, b*20170056@hznu.edu.cn, c hh@hznu.edu.cn

Abstract

In order to solve the problem of small-scale key point positioning accuracy in multiplayer pose estimation, using top-down method and advanced yolov4-tiny, this paper proposes an improved multi-person pose estimation method based on stacked hourglass deep learning network. The coordinate attention mechanism is introduced in the original hourglass network residual module to perform feature enhancement, which suppresses useless features and improves useful features, thus improving the recognition rate of small-scale joint points. From experiments, the index PCK@0.5 reaches 88.9% on MPII dataset, which verifies the effectiveness of our proposed method.

Key words—stacked hourglass network; human pose estimation; human target detection; attention mechanism;

1. Introduction

With the development of deep learning, the basic theory of convolutional neural network, and the related software and hardware, the deep convolutional neural network began to be applied to the human pose estimation tasks. The proposed DeepPose by Alexander Toshnev and Christian Szegedy[1] is the first model to apply CNN (convolutional neural network) to human joint point detection. The proposed classical network structure by Newell class[2], the Stacking hourglass network (SHN), adopts a symmetrical structure with a resolution from high to low and from low to high, which can mix the local and global information very well. Finally, it is important to obtain the information of various joints in the attitude estimation task. Papandreuol[3] first used the faster R-CNN[4] to detect the area in the picture that may contain characters, and finally used ResNet[5] (full convolution residual network) to predict the key point coordinates of each person in the area, but the human target detection effect is not ideal.

In the process of multi-person pose estimation, detecting human joint points requires distinguishing multiple human gestures, which increases the complexity of the algorithm, and includes the following difficulties: (1) The number of people existing in images is uncertain. It requires the detector to traverse the number in the image, with efficiency and accuracy. (2) This is a small and medium-sized problem. In the image, the proportion of people who are far or near the lens will be different. To handle above problems, a stacked hourglass network model, with improved coordinate attention mechanism, has been proposed. The network combines the excellent human target detection model yolov4-tiny[6] to improve human detection accuracy, by adding coordinate attention mechanism to the residual module of hourglass network. It also increases the network's attention to small and medium-sized human joints, and solves the occlusion problem, thus improving the recognition effect of human pose estimation. Finally, the effectiveness of the new algorithm is verified on the MPII dataset.

2. Network Structure

The model diagram of a separate hourglass network consists of a pooling layer, an upsampling layer, and the residual modules in Fig 1. The image is input from C1, and the resolution of the feature map is gradually reduced. After reaching the lowest resolution C7, the network begins to gradually upsample the feature map and fuse the features across scales. The structure of the hourglass network is symmetrical, so the resolution of the final output is the same as that of the input. This bottom-up or top-down duplication can build a stacked hourglass network.
Fig. 1 An hourglass network model

The stacked hourglass network can achieve good results in the pose estimation and positioning tasks of single joint points, but the recognition effect of small and medium joint points is poor. Meanwhile, the residual module used in the original hourglass network passes all the previously identified features with a unified weight to the later network in the process of feature graph transfer. When summing the outputs of multiple residual modules, doubling the variance will result in a larger amount of computation.

The overall process of the network proposed is shown in Fig 2. First, the improved human target detector (yolov4-tiny-person) is used to cut out the area containing human targets, and then send the improved attitude estimation network for human key point detection. Finally, the detected image is mapped back to the original image to complete the multi-person pose estimation of the image by Gauge identification.

Fig. 2 Overall network flow

3. Improvement Methods

3.1 Residual module integrated into the coordinate attention mechanism

Kaiming[7] proved that deepening the network depth can improve the performance of the network to certain extent, because the residual module not only extracts higher-level features but also retains the original level feature information by using conventional roads and jumping roads. The hourglass network utilizes the residual module to effectively improve the recognition performance of the human pose estimation network.

However, not all feature channels are useful for feature graph learning in the process of network transmission. According to the importance of the feature channel, we should use the useful features to improve and suppress useless features, so as to improve the identification rate of small and medium-scale joint points. The proposed SE-Net by Shen[8] can
automatically acquire the importance of all feature channels, learn the feature weights according to the loss, and increase the effective feature mapping weight to obtain better training results. However, the SE-Net only considers the internal channel information and neglects the location information. The spatial structure of the target in vision is very important. With coordinate attention mechanism\cite{9}, large overhead is avoided by embedding location information into the channel attention, which enables the mobile network to obtain information from larger regions. At the same time, in order to avoid the loss of position information caused by the introduction of 2D global pooling, the decomposition channel attention is encoded for two parallel 1D features to efficiently integrate the spatial coordinate information efficiently. As shown in Fig 3, two 1D vectors are averaged to collect the horizontal and vertical directions respectively, the 1x1 convolutions are merged in the spatial dimension. And the BN, vertical and nonlinear are collected as well. Next, by segmentation, each of the same number of channels is mapped by 1x1 input features, and then normalized and weighted. After entering the attention mechanism, the residual module deepens the entire network, which will improve the accuracy of network identification to a certain extent. Jumping the road can make it possible to directly and smoothly reach the upper layer from the deep gradient, so that the shallow network layer parameters can be effectively trained, thus solving the problem of network gradient disappearance and better optimizing the network performance.

![Fig. 3 Coordinate attention](image)

### 3.2 Other improvements

The detection accuracy of top-down pose estimation algorithms depends on the effect of the first step of human target detection. Therefore, yolov4-tiny is implemented as the human target detection model. YOLOv4-tiny is multi-task, end-to-end and multi-scale. Multi-tasking means that the classification and regression of targets are done simultaneously to achieve parameter sharing and avoid overfitting; end-to-end means that the model receives the image data and gives the prediction information of classification and regression directly; multi-scale is characterized by fusing the downsampled and upsampled data with each other, and its effect is to be able to segment targets of multiple scale sizes. Mosaic data enhancement, label smoothing, and learning rate cosine annealing decay can be used to improve the training speed and detection accuracy of the model when training it. For the human target detection task, we combined the above features of YOLOv4-tiny with the following improvements to the YOLOv4-tiny algorithm, and the improved network is YOLOv4-tiny-person.

1. The size of the anchor frame is modified to adapt to human characteristics. The anchor frame size set in the original yolov4-tiny algorithm is adapted to the universal object detection VOC2007 dataset, containing 20 kinds of objects. This article only requires the person class and needs to redesign the anchor frame size for the VOC dataset. We obtain the anchor frame size satisfying the human target features through the k-means clustering algorithm.
An attention mechanism is added to improve the detection rate. The visualization results of detecting human targets in the VOC dataset trained by the original yolov4 micro algorithm are shown in Fig. 5-a. For the large-scale human targets in the image, the model can be effectively detected. However, the model shows the detection of small-scale human objects in the image. In order to solve this problem, this paper integrates an attention mechanism into the yolov4-tiny network to improve the detection performance of network models for detecting small and medium-scale human targets. Fig 5-b shows the results of adding the attention mechanism in the original yolov4-tiny network model. It can be seen that the modified model has significantly enhanced the detection performance of small and medium-scale human targets. The attention mechanism is mainly added to the FPN (the network that strengthens the feature extraction), rather than the backbone network, because the weights of the attention mechanism are randomly initialized. If added, the full value of the backbone network will be destroyed, making the extracted features untrue and affecting the prediction results. As a result, the network diagram of the YOLOv4-tiny with attention mechanism is shown in Fig.4.

![Fig.4 yolov4-tiny+attention model](image_url)

![Fig.5-a yolov4-tiny](image_url)

![Fig.5-b yolov4-tiny + attention](image_url)
4. Experimental Results

4.1 Evaluation dataset

The validity of the proposed hourglass model is verified on the MPII multi-person pose estimation data set. The data set consists of 3844 training groups and 1758 test groups. It marks 16 human key points and also contains more than 20000 single person posture estimation training samples. This paper uses the training network of all training samples and 80% of the multi person training data set in the single person posture estimation data set, leaving 20% of the data for verification. PCK (percentage of correct key point) is generally used to calculate the accuracy of predicted key points on MPII data sets. In the calculation process of PCK, the normalized distance between the predicted key point and its corresponding real tag key point is calculated first, that is, the Euclidean distance between the real and predicted key points divided by the area of the marked human body area. The closer the distance, the more accurate the prediction. Therefore, the proportion of all human body prediction instances less than the set threshold is PCK. The calculation process of PCK is shown in equation (1), where \( \text{Int} (\cdot) \) indicates that the output value is 1 when the condition is correct, otherwise the output value is 0, \( R \) is the threshold value, \( n \) is the total number of samples in the test data set, and \( j \) is the index of the key point, that is, the number of key points. In general, the threshold used by MPII is 0.5, that is, the calculation PCK@0.5. This article uses PCK@0.5 as the evaluation index.

\[
PCK(j) = \frac{1}{N} \sum_{i=1}^{N} \text{Int}(\frac{|y_i - \hat{y}_i|_1}{\max(w_j,h_j)} \leq R)
\]

(1)

4.2 Details of experimental training

In the experiments, YOLOv4-tiny-person is trained for the person in the VOC2007 dataset, and the rest of the categories are ignored. The training was performed by setting the batch size to 32, selecting the Adam algorithm for optimization, with an initial learning rate of 1e-2, and iteratively decreasing the learning rate at 150 and 250 rounds, for a total of 300 rounds. After that, in order to be able to extract the human target region, the image output from the yolov4-tiny-person model is resized to 256*256 pixels as the input of the improved hourglass network. In the improved stacked hourglass network, the adam algorithm is selected for optimization with an initial learning rate of 1e-3 and a batch size of 16, and the learning rate is reduced to 2e-4 after 80% completion for a total of 200 training rounds. This paper uses Python language, PyTorch deep learning framework for experiments, trained on a machine with NVIDIA 3060 GPU.

4.3 Experimental results

The effectiveness of this method was verified on the MPII data set. The experimental results are shown in Table 2. It can be seen from the table that for MPII data set, the overall performance of this method is good compared with some previous pose estimation methods and superimposed hourglass network. Compared with the original superimposed hourglass network, this method improves the index by 3.4% PCK@0.5, and also improves the detection effect of small-scale joint points such as ankle and knee, which proves that the integration of coordinate attention mechanism in the residual module can better focus on small and medium-sized joint points.

The effectiveness of the present method is validated on the MPII dataset, and the experimental results are shown in Table 2. It can be seen from the table that for the MPII dataset, the overall performance of this method is better than previous pose estimation methods and stacked hourglass networks. Compared with the original stacked hourglass network, this modified method improves by 3.4% in the index PCK@0.5, and improves the detection effect of small scale joint points such as ankle and knee at the same time. It also proves that the integration of coordinate attention mechanism in the residual module can better focus on small and medium scale joint points.

Table 2 Experimental Results of the MPII dataset (PCK@0.5)/%

<table>
<thead>
<tr>
<th>Network model</th>
<th>head</th>
<th>shoulder</th>
<th>elbow</th>
<th>knee</th>
<th>ankle</th>
<th>buttocks</th>
<th>artifice</th>
<th>ensemble</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iqbal et al.[10]</td>
<td>58.4</td>
<td>53.9</td>
<td>44.5</td>
<td>36.7</td>
<td>31.1</td>
<td>42.2</td>
<td>35.0</td>
<td>43.1</td>
</tr>
<tr>
<td>Insafutdinov et al.[11]</td>
<td>78.4</td>
<td>72.5</td>
<td>60.2</td>
<td>52.0</td>
<td>45.4</td>
<td>57.2</td>
<td>51.0</td>
<td>59.5</td>
</tr>
<tr>
<td>Newell et al.</td>
<td>96.2</td>
<td>93.8</td>
<td>85.3</td>
<td>80.2</td>
<td>76.2</td>
<td>85.3</td>
<td>79.9</td>
<td>85.4</td>
</tr>
<tr>
<td>Ours</td>
<td>95.5</td>
<td>94.4</td>
<td>88.4</td>
<td>83.2</td>
<td>78.6</td>
<td>87.6</td>
<td>83.0</td>
<td>88.8</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, a new and improved multi person pose estimation method is proposed, which uses the original superimposed hourglass network and the improved human target detection model of yolov4-tiny. By integrating the
coordinated attention mechanism into the residual module, the ability of the model to identify effective channels is improved. Finally, the results displayed in the MPII dataset show that the proposed method is superior to the original superimposed hourglass network, and the recognition rate of small-scale joint points such as ankle and wrist is increased by about 3%, and the recognition rate is improved the index PCK@0.5 by 3.4%. In short, it proves that the proposed method can be used for multi person pose detection and can effectively improve the recognition rate of key points.

Acknowledgments

This paper is funded by projects: Zhejiang Province (lgf19f020011,Y202044936), and Hangzhou (20191203b14).

References

Design and Application of Smart industrial park Based on Internet of Things Technology

YIXIN YANG*a
*School of Management, Zhengzhou University, Zhengzhou, China
*892363984@qq.com

ABSTRACT

Smart industrial parks are an important part of smart cities, and building an information platform using the Internet of Things and cloud computing has become the development direction of smart industrial parks. According to the current needs of related users in the smart industrial park, the architecture of the information system of the smart industrial park was proposed. The design scheme realized the functions of intelligent analysis, interconnection and optimal decision-making of the smart industrial park. And based on the Internet of Things, the application in the smart industrial park was introduced to provide guidance for the improvement of the intelligence level of the smart industrial park.

Keywords: Smart industrial parks, Internet of Things, Building application

1. INTRODUCTION

With the continuous in-depth development of Internet information technology, the attention of smart industrial parks has gradually increased, and it is considered to be an important way for parks to achieve sustainable development [1]. As a microcosm and an important part of a smart city, the construction of the smart industrial park will greatly promote the smart city. In the context of the rapid development of smart cities and information technology, traditional parks are exploring the road to intelligent transformation [2].

The intelligence of the smart industrial park is mainly reflected in the wisdom of services and management, as well as the establishment of a smart park information platform to improve the utilization of park resources. Therefore, the information in the park cannot be independent and needs to be interconnected to achieve data exchange. The Internet of Things (IoT) is characterized by the interconnection of everything and sensing everywhere. It is considered as an important tool to solve this problem [3]. The development of IoT technology has improved the quality and efficiency of information collection, transmission, and processing. Currently, IoT has been widely used in smart cities, smart transportation, smart energy, and other fields [4-6]. IOT technology collects a variety of scientific information technologies, and many of these key technologies play an important role in the construction of smart parks, for example, RF identification technology can effectively identify each individual in the park and has wide applications in access control management, electronic traceability, etc.; software technologies include databases, middleware, etc., which configure responsive software for each operating system in the park and can be provided according to the relevant needs personalized services.

In recent years, domestic and foreign scholars have conducted a lot of research on the information construction of smart parks. Tang proposed the operation mode of a smart logistics park based on IoT technology [3]. Wang proposed the design framework of an intelligent platform for agricultural science and technology parks [7]. Paul and other scholars used IoT technology to create an intelligent storage management platform for smart parks [8]. However, existing smart parks still suffer from problems such as information silos, and the degree of platforming and integration is not high. Therefore, this paper proposes a scheme design of intelligent platform for smart parks to promote the construction of IoT-based information system for smart parks, and tries to achieve information interconnection, intelligent analysis and decision making in all aspects of smart industrial park operations.

2. SMART INDUSTRIAL PARK DEMAND ANALYSIS

2.1 Smart industrial park user requirements

The construction of a smart industrial park requires extensive consideration of the demands of users at all levels in the park, continuously expanding the coverage of the smart industrial park information platform, and improving the
intelligence level of the smart industrial park. According to the survey and analysis, the users of the smart industrial park can be roughly divided into four categories: park operators, resident companies, users or visitors of the park. Their demands for smart industrial parks are also different, as shown in Figure 1.

Figure 1: Analysis of user needs of smart industrial parks

(1) The park management: It is hoped that through intelligent information technology, the management efficiency of the park will be improved and management costs will be reduced. Smart industrial parks can provide more satisfactory services for resident enterprises, thereby attracting more enterprises to settle in the park and increasing income. (2) Settled enterprises: enterprises can use the convenient services of the park to increase the company's business, improve the company's reputation, and attract and retain talents. In addition, enterprises also hope that the park can provide security protection to prevent corporate secrets from being leaked. (3) Users or visitors: convenient living, comfortable working environment, and various convenient services. (4) Government and regulatory agencies: While promoting the business growth of enterprises, the park pays attention to the impact of the park on the environment.

2.2 Smart industrial park functional requirements

At this stage, the construction purpose of the smart industrial park is to better apply advanced technologies such as the Internet of Things and cloud computing, so as to meet the needs of users at all levels. [3]. The data is analyzed through the platform to promote the smart transformation of the park. According to the needs of users, the functions of the smart industrial park should include: equipment management services, provision of information services, security services, environmental management, energy efficiency management, intelligent operation, and convenient transportation services.

3. OVERALL STRUCTURE DESIGN OF SMART INDUSTRIAL PARK

3.1 System design goals

Smart industrial parks aim to build a smart industrial park information platform with integrated management and control. Through cloud computing and other cutting-edge technologies to empower smart industrial parks and provide customers with convenient services. Import the data collected by the Internet of Things technology into the database, which can improve the intelligent information level of the park, and meet the information needs of different customers [7].

Each subsystem of the smart industrial park should be interconnected. Through intelligent technology, the circulation of information between various systems is realized, the efficiency of the integrated operation of the park is improved, the business management cost is reduced, and the waste of resources is reduced [9].
3.2 System architecture design

The smart industrial park obtains data from various ends such as surveillance cameras and gates to form a shared database, and analyzes the data through cloud computing to provide auxiliary decision-making support for the smart industrial park. The overall structure of the smart industrial park includes six parts: external data source layer, network transport layer, data layer, platform layer, application layer and experience layer [10]. The overall structure of the smart industrial park is shown in Figure 2.

(1) External data source layer. The external data source layer is the data source of the smart industrial park. Through monitoring, access control, lighting, gates and other terminal equipment, the information collection and monitoring of the infrastructure, environment, security and other aspects within the park is realized.

(2) Network transport layer. It is used for data transmission, including network transmission infrastructure such as campus private network, public communication network, edge nodes and communication equipment rooms.

(3) Data layer. Manage and mine data. After the data is processed, it is classified into different databases for storage management. The data layer also uses various advanced cloud computing processing technologies to dig deep into the data. The data layer can realize the centralized storage, sharing and exchange of various data resources in the park.

(4) Platform layer. The platform layer is the support core of the smart industrial park. Through the application of information and communication technology, it not only integrates, processes and analyzes the information systems of the lower-level departments, but also supports the expansion of upper-level intelligent applications and intelligent horizontal services, and supports the needs of different users at the application layer.

![Figure 2: The overall architecture of the smart industrial park](image-url)
5. Application layer. The application layer is a user-oriented service layer and should be equipped with various intelligent applications required for the daily operation of the smart industrial park. Based on the core data and services provided by the platform layer, artificial intelligence technology is used to provide information-based services for park managers and park enterprises.

6. Experience layer. The experience layer is to feed back the results of intelligent decision-making and analysis on the interactive terminal, so that users in the park can intuitively see the running status of the system. Users can experience the intelligent services of the smart industrial park through computer websites, mobile APPs, monitoring large screens and other multi-terminal devices.

4. IOT TECHNOLOGY APPLICATION SPECIFIC PRACTICE

4.1 Environmental protection for smart industrial parks
The environmental management of the smart industrial park is to use the Internet of Things technology to collect the environmental quality information of the environmental detection sensors in the park, and transmit the collected information to the data center through the park network to store, analyze and monitor the environmental quality information of the park. And released through data and reports, including air quality, water quality monitoring and other content. In addition, when the environmental quality is lower than the preset standard, the environmental protection system can make intelligent judgments and execute predefined linkage responses, so as to ensure the environmental quality of the park and maximize the comfort of the living environment.

4.2 Smart industrial park security services
The intelligent security system based on the Internet of Things technology integrates real-time sensors and data acquisition subsystems deployed throughout the park, and uses existing information technology to collaboratively analyze the collected information and make decisions. The smart industrial park security system can realize the collection of abnormal events in the park, and immediately transmit the alarm information to the alarm center, deal with dangerous situations in time, and form an intelligent system of public security, fire protection and other parties. In addition, the access control system in the park is often a one-card system, which has multiple functions such as access control, attendance control, and card swiping in the park canteen, which facilitates the daily life of users in the park, and also ensures the safety of people and property in the park.

4.3 Smart industrial park equipment management
The facility management of smart industrial parks can realize the integration of facility subsystems, and expand the scope of facility management to include business equipment and facilities. The park adopts IoT technology, edge computing and other information technologies, integrates fault prediction and health management technology, builds a smart industrial park equipment operation and maintenance platform based on equipment configuration management, integrates equipment warnings and data information uploaded by equipment subsystems, and enables the park users can view warning messages in real time. Automatically dispatch maintenance work orders for equipment operation failures. For equipment and facility warnings, it can automatically associate with other systems for disposal, and can query, count and analyze work orders on the user side.

4.4 Smart industrial park energy management
The energy consumption management of the smart industrial park is to use the Internet of Things technology to collect the information of smart devices such as smart water meters, smart electricity meters, and smart gas meters, monitor the usage of each energy module in real time, and use artificial intelligence algorithms to analyze the optimal energy efficiency point to achieve overall energy saving optimal. By recording energy usage, the trend of energy usage is displayed graphically, so as to guide energy conservation and consumption reduction. Real-time monitoring of the operating status and energy use indicators of energy-efficiency-related equipment in the park, and relying on BIM to analyze the statistics of energy consumption in different areas, and automatically warn areas with abnormal energy consumption, which can help administrators to check abnormal areas in a targeted manner, investigate possible hidden dangers of accidents, and ensure the normal operation of the business.
4.5 Smart industrial park smart operation center

The intelligent operation center of the smart industrial park can supervise the operation situation of the smart industrial park, and has the ability of business analysis, decision-making and analysis. The smart industrial park takes the digital twin construction achievements as the basic skeleton, uses the IoT perception system as the neural network, and uses technologies such as big data and artificial intelligence to form core service capabilities such as analysis and decision-making. GIS technology is used to realize visual and dynamic management of the park, and the operation status of the smart industrial park can be viewed at multiple terminals. The park uses intelligent audio and video to supervise the park. When an abnormal situation is found, it will start a preparatory plan independently to realize the joint command between various systems and realize the digital operation goal of the park.

4.6 Smart industrial park traffic management

The traffic management of the smart industrial park mainly uses the Internet of Things, big data and other technologies to solve the problems of traffic congestion and parking spaces in the park.

The Internet of Vehicle Control utilizes sensors, RFID and cameras to collect various information about the vehicle and its surrounding environment and the vehicle itself in the park, and transmit this information to the entire vehicle control system. Through functions such as big data analysis, it can plan reasonable routes for vehicles in the park, avoid traffic congestion in the park, and monitor the cruise status of each vehicle, such as vehicle speed.

For the parking problem of the smart industrial park, based on the intelligent parking space configuration resources, the automatic geomagnetic intensity sensor installed in the vehicle and the monitoring camera of the parking lot are used to identify the photo information of the vehicles in the park, recommend the most convenient parking space for the vehicle, and realize parking space information search and parking reservation. Use APP-based functions such as automatic collection of parking fees to improve the efficiency of traffic in the park and shorten the parking time of other vehicles.

5. SPECIFIC RISKS OF IOT APPLICATIONS

When applying IoT technologies, several security aspects need to be taken into account.

1) Terminal equipment security. The smart industrial park involves diverse sensing terminals. For their different security threats, the security issues in the design and use of terminal equipment should be considered.

2) Information transmission security. Information may be exposed to security risks such as eavesdropping and tampering in the process of transmission. Therefore, the park needs to take certain encryption measures to protect information security. Transmission encryption technology includes several aspects such as communication peer authentication, key negotiation, data compression and encryption.

3) Network security. The smart industrial park has networks with different functional services, which have different security requirements; therefore, the smart industrial park needs to divide different security domains on the network and carry out logical isolation and access control of the network.

4) Platform security. The operation of the wisdom park platform is vulnerable to attacks such as Trojan horses and viruses, which cause the platform to be paralyzed. Therefore, in the process of using the information platform of the wisdom park, security protection measures such as firewalls need to be set to prevent malicious attacks.

6. CONCLUSION

This paper analyzed the needs of each user in a smart industrial park and proposed an overall architecture for the construction of a smart industrial park. It used new technologies such as IoT and cloud computing for data collection and analysis to solve the problems of poor service experience, low operational efficiency and information silos in traditional parks. In addition, this paper analyzed the application of IoT technology in the smart park and provided a path for the information construction of the smart industrial park. In the future, it will further improve the reliability of IOT, cloud computing and other technologies applied in smart industrial parks, reduce operation and maintenance costs, and improve operational efficiency.
REFERENCES

A Protection Framework Based on Dynamic Heterogeneous Redundancy Architecture

Yunfan Wang1a*, Zesheng Xi1b*, Bo Zhang1, Chuan He1
1State Grid Smart Grid Research Institute co., Ltd., State Grid Key Laboratory of Information & Network, Nanjing, China

a*wangyunfan95@foxmail.com, b*4317045xi@163.com

Abstract

Security and verification applications are primarily developed in C or C++, so reverse engineering is prone to analyse and invade programs based on C or C++. As a result, protecting the source code of C or C++ as well as guaranteeing the safety of those security and verification applications are vital tasks in the network security area. This paper proposes a new framework enlightened by dynamic heterogeneous redundancy architecture, which originated from the mimic defence field and mixed with a state-of-the-art obfuscation method to not only protect the code property but also improve the security level of those applications. Three key characteristics of the framework are redundancy, heterogeneous and dynamic. This article firstly builds an LLVM optimizer to generate redundant binary code. Besides, we import some random perturbation in the optimizer to realize heterogeneity between generated redundant codes. Finally, exploiting these heterogeneous redundant codes constructs a dynamic architecture to detect and defend anomaly requests.

Keywords—dynamic heterogeneous redundancy architecture; mimic defence; code obfuscation; LLVM; bogus control flow; control flow flattening

1. Introduction

Ever since the appearance of the network, hackers sprang out with malicious intentions to invade our computers and became a constant threat to personal assets. At the early stage of cyber security, programs were much more vulnerable, and most programs contained many flaws which were often exploited easily by attackers. Given that those vulnerabilities were easy to fix, security engineers were exhausted by patching infinite flaws. Therefore, various defense tools and applications were invented, such as antivirus software, firewalls, and encryption techniques. This paper categorizes these techniques into two methods, which are intrusion prevention and intrusion detection. The two methods can be recognized as static defense [1] because they are static and can only wait for attacks to come. The invasion can be prevented if it can distinguish virus or attack instruction from routine visits. However, for those unknown attacks, the system is incapable of detecting any anomaly. As a result, the system can only wait to find and fix some flaws until those vulnerabilities have broken it, and the damage caused by them will not be remedied.

Nowadays, modern engineers have developed a dynamic defence technology called “moving target defence” [2], which keeps changing the status of the defence system to hide vulnerabilities and reduce the risk of successful attacks. Based on the principle of dynamic defence, Wu proposed an architecture called mimic defence [3]. The core concept of this architecture is dynamic, heterogeneous, and redundant, so the system can also be named as dynamic heterogeneous redundancy architecture (DHRA) [4]. The system consists of plenty of heterogeneous entities, which are made up of different hardware and software and have the same functionality. Although those entities share the same function, they have various vulnerabilities as they have different implementation processes. Furthermore, the system will examine the outputs of the entities, and the input will be considered as safe input only if all outcomes are the same. Consequently, the possibility of breaking by a single attack is subtle [5].

The framework proposed in the paper is based on the DHRA. However, different from the classic DHRA, which needs both hardware and software and usually can only be implemented on large systems. We generate heterogeneous redundant code from original C or C++ source code as our heterogeneous entities, which can be beneficial to small systems and is convenient for old devices to upgrade old systems. Chapter 2 illustrates related research on DHRA and code obfuscation technology. Afterward, Chapter 3 demonstrates an obfuscation generator based on an LLVM optimizer, which can comply the source C or C++ code into redundant binary code with a different flow at a rapid speed. Then, in Chapter 4,
the paper introduces the dynamic heterogeneous redundancy architecture based on binary codes generated in Chapter 3. Chapter 5 evaluates the framework and makes a comparison with other similar technologies such as OLLVM [6]. At last, we bring a summary of our research in Chapter 6.

2. Related Research

2.1. Obfuscation

Obfuscation technology occupies a significant role in software security to counter against reverse engineering. The main methods for binary code obfuscation can generally be categorized into three parts – data encryption, instruction substitution, and control flow transformation [7].

Data encryption only transforms variables into an encrypted string which is hard to understand from a human’s perspective with the control flow, and all instructions remain unchanged. As a result, this method can usually be decoded easily with some reverse tools. Therefore, data encryption is typically applied with other more advanced technologies to ensure safety [8].

Instruction substitution is a more sophisticated technology in the obfuscation field, which usually represents the replacement of original instructions with an alternative instruction set that is equivalent to the original version. Furthermore, instruction substitution can also refer to complicating comparing conditions. For example, logic operations $a = \&b\&c$ and $a = \neg(\neg\neg b\neg\neg c) \& (\neg\neg r)$ share the same result, while the second is more difficult to calculate the outcome. Such work reduces the readability of source code, and reverse tools fail to simplify those function sets and operations, but of course, it is not impossible for experienced reverse engineers [9][10].

The last method is the most potent obfuscation algorithm among the three transformations, and it has remained a pervasive implementation since it occurred. There are two main directions to realize control flow transformation: control flow flattening and bogus control flow [11]. To carry out these two algorithms, the source code needs to be divided into basic blocks, which consist of a series of instructions. Control flow flattening exploits switch instruction as a dispatcher and adds a jump command with an opaque predicate at the end of each basic block to ensure the sequence of the program so that the technique hides the real order of the program. Hence reverse engineering could be stuck with disordered control flow [12]. Another method is bogus control flow, which creates faux basic blocks into the program with opaque predicates as well [13]. Although these blocks will never be executed in the actual flow, reverse tools cannot distinguish them except by calculating all opaque predicates, which is an apparent time-consuming work. While control flow transformation still exists risks when it comes to reverse-engineering experts, it remains the most popular obfuscation choice, and many variants are also invented to improve code security.

2.2. Mimic Defence

Mimic Defence belongs to active defence technology, which The U.S. Moving Target Defences firstly mention (MTD), focusing on defending against external attacks through unknown vulnerabilities[2][14]. Based on this theory, Chinese academician J.X. Wu pioneered the Cyber Mimic Defence (CMD) technology and constructed dynamic heterogeneous redundancy (DHR) architecture which can defend against unknown vulnerability external attacks as well as defeat unknown backdoor attacks. Currently, a bunch of efforts has been put into improving the mimic defence theory [15]. CMD has been successfully realized in the web server system by Q. Tong [16]. According to H. Ma, a router mimic defence infrastructure has been implemented based on DHR architecture [17]. K. Song designed an ethernet switch based on the endogenous security architecture [18]. The essence of CMD is to transform the static system into a dynamic state through deploying and replacing functionality equivalent heterogeneous entities. Research has verified that when heterogeneous software is assigned to complete the same task, the probability of endogenous osmosis by unknown attacks is obviously reduced [19].

3. Methodology

3.1. Obfuscation Generation

In this chapter, we combine the bogus control flow technique with the control flow flattening method to generate an obfuscation optimizer in LLVM compiler infrastructure. The optimizer is capable of parsing the source code into LLVM IR (intermediate representation), which is between the source code and binary code. This process is designed to optimize the original code, which can improve efficiency, compress volume, and provide API to realize different tasks based on
actual requirements. According to the official documentation of LLVM, LLVM IR firstly defines an executable source file as a module, and functions in a module are consist of several basic blocks, which are a series of instructions as we mentioned before. Firstly, we refine the traditional bogus control flow method into a randomized bogus control flow algorithm in the chapter. Unlike the old way, which only creates fake basic blocks which are not really passed through during the execution procedure, the randomized method produces modified blocks that could actually be executed when the program is operated. Afterward, we split basic blocks into new smaller basic blocks and perform flattening to these blocks. Finally, generating heterogeneous redundant binary code by LLVM compiler with functionality equivalence.

3.1.1. Randomized Bogus Control Flow

To construct a bogus control flow, the first operation is to split an original block into three basic blocks (BB) – basic entry block, basic body block, and end basic block. LLVM provides a convenient structure to split by parsing the source code into LLVM IR (intermediate representation). The transmission from the original basic block to three new blocks is shown in Figure 1.

![Figure 1. RBCF Splitting](image)

There are two purposes for performing this splitting. The first goal is to ensure the robustness of the transmitted program by splitting all PHI instructions into the basic entry block. The PHI instruction is specified in LLVM documentation to handle variants by single static assignment (SSA), and this instruction must be included in the head of each block. Thus, we sort all PHI instructions into the entry basic block to avoid duplicated usage in a later procedure. The end basic block is relatively easy as a simple return or jump instruction to control the program flow into the next part. As shown in Figure 2, the first step is to create a cloned basic block on the basis of the basic body block. In the article, we remain this cloned block unchanged as the bogus block because the cloned basic block will appear differently from the primary body block according to the LLVM SSA principle. Then, branch instructions are added as the left part of Figure 2. By a random variant created by a function rdrand() in LLVM with an opaque predicate construct, we set a complicated branch condition to control the flow of the transformed blocks. When the condition is true, the flow passes only through the basic body block, which is the same as the old version. While the condition is false, the flow will jump to the cloned basic block from the entry and return to the body later. This pass apparently creates more variants and dynamic indeterminacy during actual execution, making static and dynamic reverse engineering more difficult and time-consuming.
3.1.2. Control Flow Flattening

Overall, the control flow flattening method proposed by the paper is as straightforward and simple as other similar flattening methods through altering the old control flow with the switch-case format, as shown in Figure 3.

To transform the original flow into a switch-case flattened flow, the foremost process is to construct three new blocks; the entry block could be part of the original basic block 1 with all PHI instructions and escape variables stored in this part. The escape variables refer to variables that are reused by different blocks. The task for the dispatching block is to guide the program going to the correct block according to the flag returned from the return block. Furthermore, the control flags have been inserted into each basic block except for a return instruction in basic block 4. After the flattening procedure, the program flow is hidden in the switch-case construct and controlled by the dispatching and return block with opaque predicate flags. Such structure can confuse reverse engineers with the disordered program sequence, especially in static analysis.

![Diagram of Original and Flattened Flow](Link_to_image)

(a) Original Flow  
(b) Flattened Flow  
Figure 3. Control Flow Flattening

3.1.3. Heterogeneous Entity

Through the last two obfuscation methods, we have built a redundant code generator, but, in order to generate heterogeneous codes, we need to import some hyperparameters to control the degree of obfuscation. More precisely, the first imported parameter, called bogus rate, can control the probability of obfuscation of each block, which not only guarantees the heterogeneity of each block but also improve the efficiency by reducing the density of bogus blocks. And the parameters -- split number and the split rate, which control the splitting numbers from an original basic block and overall splitting density, respectively.

![Table of Similarity Matrix](Link_to_table)

<table>
<thead>
<tr>
<th>Similarity</th>
<th>Entity 1</th>
<th>Entity 2</th>
<th>Entity 3</th>
<th>Entity 4</th>
<th>Entity 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity 1</td>
<td>1</td>
<td>0.78</td>
<td>0.65</td>
<td>0.54</td>
<td>0.56</td>
</tr>
<tr>
<td>Entity 2</td>
<td>0.78</td>
<td>1</td>
<td>0.65</td>
<td>0.63</td>
<td>0.52</td>
</tr>
<tr>
<td>Entity 3</td>
<td>0.65</td>
<td>0.65</td>
<td>1</td>
<td>0.57</td>
<td>0.61</td>
</tr>
<tr>
<td>Entity 4</td>
<td>0.54</td>
<td>0.63</td>
<td>0.57</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>Entity 5</td>
<td>0.56</td>
<td>0.52</td>
<td>0.61</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. An example of the similarity matrix

<table>
<thead>
<tr>
<th>Matching Rate</th>
<th>Functions</th>
<th>Calls</th>
<th>Basic Blocks</th>
<th>Jumps</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched</td>
<td>1</td>
<td>0.318</td>
<td>0.76</td>
<td>0.469</td>
<td>0.501</td>
</tr>
<tr>
<td>Unmatched in E1</td>
<td>0</td>
<td>0.364</td>
<td>0.13</td>
<td>0.275</td>
<td>0.273</td>
</tr>
<tr>
<td>Unmatched in E2</td>
<td>0</td>
<td>0.318</td>
<td>0.109</td>
<td>0.256</td>
<td>0.226</td>
</tr>
</tbody>
</table>

Table 2. Matching rate between entity 1 and entity 2

Besides, a similarity matrix based on an application called bindiff, which can compare the similarity of two binary programs, is introduced to filter out valuable entities. An example of a similarity matrix is shown in Table 1. The
similarity degree is mainly judged by the matching rate of functions, calls, basic blocks, jumps, and instructions. Table 2 demonstrates the detailed matching rate between Entity 1 and Entity 2. As the obfuscation method remains function names unchanged, the rate of functions is up to 100%. At the same time, after the transformation of program structure, most calls and jumps cannot be recognized by the code matching detection application. With those parameters and the similarity matrix, the generator diversifies the layout of different structures in case the similarity of those entities reduces the system reliability.

3.2. Dynamic Heterogeneous Redundancy Architecture

Since we have generated heterogeneous redundant codes in the last three parts, the previous module of this project is the DHR architecture. Different from the classical DHR architecture, which combines various hardware, operation systems, compilers, and programming languages. This DHR architecture aims to develop a lightweight code-focused algorithm which can be executed in one machine like a gateway or firewall. As a result, our algorithm is relatively simple to achieve. Figure 4 illustrates the framework with obfuscated codes stored at the heterogeneous entity set.

Theoretically, entities in the backup set can be replaced or updated at any time regardless of the status of the whole system, and they are regulated by the strategy module. To initialize the system, the strategic scheduling module will randomly activate a given number of entities to become executors. Considering the balance between security and efficiency, we recommend that the number of executors is 3, which is also the minimum number for multimodal voting. After the system is initialized, when the system receives an input signal, the input agent will distribute the signal to all executors simultaneously. If the multimodal voting process receives the same result, the system will treat the input signal as a regular visit and pass the output to actual applications. However, when the result is different, for example, two executors raised specific error reports. At the same time, the other one has an expected output; the system will recognize the input as a network infiltration operation and inform the strategy module to replace the malfunctioned executor with a backup entity which has the most enormous difference degree based on the similarity matrix to reduce the risk of the exposed entity. This strategy allows the architecture to protect the whole system dynamically from unknown attacks and actively distinguishes and substitutes the exposed executor. The architecture can also block the risky visit when detecting inconformity and the exposed executor can be later analysed and make contributions to upgrading the source code protection program can create more safety heterogeneous with the obfuscation generator mentioned in the last part of this article.

4. Evaluation

The framework is mainly built to protect verification programs such as SSL gateway or core firewall code, so the evaluation focuses on source code protection and efficiency analysis. Figure 5 presents a toy example in C to display the appearance after obfuscation.
Figure 5. A Toy Example in C

```c
#include <stdlib.h>
#include <stdio.h>

void f(int x)
{
    int i;
    for (i=0; i<x;i++)
    {
        printf("%d\n",i);
    }
}

void g(int x)
{
    int count = 0;
    int i = 0;
    for (i=0; i<x;i++)
    {
        count++;
        printf("count is :%d\n",count);
        printf("x is :%d\n",x);
    }
}

int main(int argc, char** argv)
{
    int a = atoi(argv[1]);
    int b = atoi(argv[2]);
    if ((a*b)+12 == 16){f(a|b);}
    else{g(a&b);}
    return 0;
}
```

Figure 6 compares the control flows between the original binary code and obfuscated binary code by our algorithm by IDA, a technical reverse-engineering tool. The obfuscation algorithm provided in this paper successfully transforms a simple code with only one if-else stance in (a) into a complex code with plenty of blocks split, duplicated, and flattened. Concerning the efficiency and memory occupancy, Table 3 indicates that obfuscated code takes around 50% more time to execute and needs 23% more hardware space compared to the original code, while the memory space remains the same. The main prize for obfuscation is the execution time, as the bogus control flow step inserts too many randomized blocks. In order to mitigate that problem, the system can control the bogus rate and split rate to make the system less complex. However, a large proportion of security applications are essentially used as a verification tool that usually confirms the token once per visit, so the speed of the verification process should not be a severe issue.

Table 3. Comparison between original and obfuscated code

<table>
<thead>
<tr>
<th>Code</th>
<th>Execution Time</th>
<th>Code Size</th>
<th>Memory Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.002s</td>
<td>16.8kB</td>
<td>9044B</td>
</tr>
<tr>
<td>Obfuscated</td>
<td>0.003s</td>
<td>20.8kB</td>
<td>9044B</td>
</tr>
</tbody>
</table>

Eventually, to detect the stability of the DHR architecture, this paper deploys a classic stack-based buffer overflow code into the architecture. By building a shellcode to overwrite the return address in the stack, this overwrite would lead to
arbitrary code execution. After obfuscation, the old shellcode failed to exploit the flaws as the return address has been changed. Given that one executor has been hacked, the other two executors still raise the segmentation fault report because they are heterogeneous structures and have a different return address. The experiment proves that the novel designed DHR architecture is capable of protecting the system in some cases that an original verification code would fail.

5. Conclusion

This paper proposes a novel DHR framework to protect both system security and code property. The base stone of the framework is a code obfuscation technique combined with randomized bogus blocks and flattened control flow. This technique is developed by an open-source project called LLVM, which provides a multi-platform compiler infrastructure to support different computer instruction set architectures such as X86, and ARM. Although our method is designed to obfuscate C or C++ languages, with the extension of LLVM front-end compilers, the technique can be implemented in any existing languages or brand-new invented languages complied by LLVM because the method is an optimizer implemented in the IR layer in LLVM and should not influence the front-end and back-end. On the other hand, the DHR architecture in this paper is more light-weighted compared to the classical version. Therefore, the new architecture is more suitable for a small system which cannot afford redundant hardware devices. Further research can include improving the system performance by optimizing the obfuscation process and enhancing the feedback module in DHR to analyze and evaluate potential vulnerabilities intelligently for binary programs.

Acknowledgments

This work is supported by the Science and Technology Project of State Grid Corporation of China (Research on endogenous immune enhancement technology for secure access gateway of power IoT, No. 5400-202158332A-0-0-00).

References


Taxi Destination Prediction based on LSTM with Tree Memory Module
Dan Song, Yadong Li, Meng Zhang, Ting Zhang
School of Information Science and Engineering, Zaozhuang University, Zaozhuang 315211, China
a*101541@uzz.edu.cn, b*liyadong@uzz.edu.cn

Abstract

Taxi destination prediction can grasp the flow direction of the taxi, facilitate the taxi dispatches. There is always been a long-term dependency problem in taxi trajectory prediction. Although LSTM can solve the long-term dependency problem to a certain extent, it does not have a good ability to deal with the deep correlation between long trajectory sequences. To address the above problem, we propose a taxi Destination Prediction method based on LSTM with Tree Memory Module (TMM-LSTM). TMM-LSTM stores the state of the input trajectory through an external memory structure. It uses a tree structure to process more historical information and better deal with the long-term relationship between trajectory points. TMM-LSTM can better solve the long-term dependency problem in the taxi trajectory sequence. Experiments demonstrate that the average error distance is 6% lower than traditional LSTM model.

Keywords-GPS trajectory; taxi destination prediction; Tree Memory Module; LSTM

1. Introduction

With the rapid development of communication technologies and GPS (Global Positioning Systems), the trajectory data generated by moving objects during the moving process has shown an explosive growth trend. Since the introduction of the intelligent transportation system (ITS) [1], it has always been a research hot spot of researchers' attention.

Taxi destination prediction methods are diverse. Early taxi destination prediction researches are mostly based on probability theory. Hazelon et al. [2] used the method of multivariate normal estimation based on the Bayesian framework to predict the taxi destination. Mathew et al. [3] have used a hidden Markov model for the next location prediction. These locations are then incorporated into a Markov model to predict destination. When the amount of trajectory data is increasing, how to reduce the training and prediction time while ensuring the accuracy of prediction is a major challenge.

To solve this problem, scholars have conducted relevant research and achieved certain results. Some researchers have applied deep learning to predict taxi destination. Liu et al. [4] combined RNN (Recurrent Neural Network) [5] with spatial-temporal features to conduct trajectory prediction. Xu J [6] added the date and time of the taxi as an influencing factor using LSTM (Long Short-Term Memory network) [7] for taxi demand prediction. Although the above methods can achieve good prediction accuracy, they are only suitable for short trajectory sequences, and do not have good processing ability for deep correlation between long trajectory sequences.

In order to solve the problem, we propose taxi Destination based on LSTM with Tree Memory Module (TMM-LSTM). Tree Memory Module [8] expands the linear storage structure into a tree-shaped structure. Storage module's parent node can reflect more child nodes than a linear structure. A node can reflect the historical information of multiple descendant nodes through a recursive process. The tree structure of the storage module can handle longer trajectory sequences and avoid large deviations during backpropagation caused by too long trajectories. Therefore, TMM-LSTM can capture and process more historical information, and can better solve the long-term dependency problem in trajectory sequence modelling.

2. Problem Definition

The trajectory data set contains multiple trajectory data, which can be expressed as:

\[ T = \{T_1, T_2, \ldots, T_i, \ldots, T_n\} \quad (1) \]
where, $n$ represents the number of trajectories in the data set, and $T_i$ denotes one trajectory. A trajectory is composed of many trajectory points in time series, which is expressed as:

$$T_i = \{x_1, x_2, \ldots, x_j, \ldots, x_m\}$$  \hspace{1cm} (2)

where, $x_j$ represents the number of track point of $T_i$. $x_m$ denotes the last position of the moving object. A track point, $x_j$, can be expressed as:

$$x_j = (x_{jt}, x_{jlt}, x_{jlt})$$  \hspace{1cm} (3)

where, $x_{jlt}$ and $x_{jlt}$ are the longitude and latitude of the track point, and $x_{jt}$ is the sampling time. The trajectory prediction takes the front $n-1$ position sequence of the trajectory $T_i = \{x_1, x_2, \ldots, x_j, \ldots, x_{m-1}\}$ as input and predicts the final position $x_m$.

We set some metrics to evaluate the performance of the experiments. The metrics are defined as follows:

1. **Prediction Accuracy Rate (PA)**: The ratio between the number of given locations correctly predicted and the total number of predicted locations.

2. **Average Error Distance (AED)**: The average value of the haversine distance between all predicted positions and the corresponding real position. The haversine distance is the straight-line distance between two GPS coordinate points on the earth. Haversine formula is defined as follows:

$$H(P_i, P_j) = 2\text{R} \arctan \left( \frac{a(x_{i}x_{j})}{a(x_{i}x_{j})} \right)$$  \hspace{1cm} (4)

where, $R$ represents the radius of the earth, and its value is 6371km. $a(P1, P2)$ is defined as follows:

$$a(x_{i}, x_{j}) = \sin^2 \left( \frac{x_{jlt} - x_{jlt}}{2} \right) + \cos(x_{i}) \cos(x_{j}) \sin^2 \left( \frac{x_{jlt} - x_{jlt}}{2} \right)$$  \hspace{1cm} (5)

### 3. Methodology

#### 3.1. The framework

The overall framework of FTP-AS model is shown in Fig 1.

The steps of the method are demonstrated as follows:

**First Step**: The trajectory sequence $T_i$ and the trajectory-related metadata information are combined as input data $A_i$. The details are shown in Equation 6. Metadata is other driving-related data except the trajectory data when the taxi is running, including the taxi number, the user number, the time of getting on the bus and other information. This paper selects the first $k$ points $\{x_1, x_2, \ldots, x_k\}$ and the last $k$ points $\{x_{m-k}, x_{m-k+1}, \ldots, x_{m-1}, x_m\}$, each point contains latitude and longitude.

$$A_i = \{X_i, ClientId, TaxiId, Staidld \ldots\}$$  \hspace{1cm} (6)
Second Step: Through an LSTM layer, the current input is generated into an intermediate variable \( c_i \) that can represent the current input, as shown in Equation 7:

\[
c_i = f_{LSTM}(A_i)
\]  

(7)

Third Step: Taking \( c_i \) as the input of the storage structure, the correlation score \( m_i \) between the current input and the storage structure is calculated through the \( f^{score} \) attention score function. Where \( M_{t-1} \in \mathbb{R}^{k \times (2^l-1)} \) is the adjacency matrix from the root node to the leaf node representing the state of the current storage structure, \( k \) is the hidden layer embedding dimension, and \( l \) is the depth of the tree memory structure (height of a binary tree). The scoring function is defined as Equations 8, 9:

\[
m_i = f^{score}(M_{t-1}, c_i)
\]

(8)

\[
f^{score}(M_{t-1}, c_i) = c_i^T M_{t-1}
\]

(9)

Fourth Step: In this paper, \( c_i \) and \( z_i \) are processed according to formula 10, fused according to weight, and the prediction destination \( D_{fi} \) is obtained through the ReLU function, where \( W_{out} \) is the weight.

\[
D_{fi} = \text{Relu}(W_{out} z_i + (1 - W_{out} c_i))
\]

(10)

3.2. The tree memory module

The tree memory module consists of multiple nodes. Each parent node has two child nodes, and the information of the next layer can be obtained through the child nodes. Because of the tree structure, the distance between the current node and the history node becomes less, and the parent node can better map the history information. The storage module is composed of S-LSTM, which expands the general sequential structure in LSTM into a top-down tree structure. In this paper, the S-LSTM structure is represented as a binary tree, which realizes the method of long-distance interaction between the input of the trajectory sequence, and optimizes the LSTM model to deal with the problem of long sequences. The structure of an S-LSTM is shown in Fig 2. In the Fig 2, \( f_l^n, f_r^n, i_n, o_n \) respectively represent the left forget gate, right forget gate, input gate and output gate. \( \otimes \) means multiplication.
During the training process, the memory update of the storage structure is very important. $h_{t-1}$ and $c_{t-1}$ are the output of the left subnode and the right subnode, and these two sub-nodes will be used as the input of the current node. The input gate $i_t$ has four inputs: the hidden vector ($h_{t-1}$, $c_{t-1}$) and the unit vector of the two child nodes ($c_{t-1}^L$, $c_{t-1}^R$). These four parameters are also the input of the left forgetting gate $f_{t-1}^L$ and the right forgetting gate $f_{t-1}^R$, which are used to generate gate signals. Left and right forget gates can be controlled separately, allowing information to be passed up from child nodes. The output gate $o_t$ needs to consider the child node vector of the hidden layer and the current memory cell vector. In this way, the memory unit can directly and indirectly influence multiple subunits by merging subnodes through the forget gate. Therefore, TMM can extract long-range interrelationships on the structure. At time series $t$, the update method of the unit information and gating information in each node in the storage structure is according to formulas 11 to 17:

$$i_t = \sigma(W_i^L h_{t-1} + W_i^R h_{t-1} + W_{ci}^L c_{t-1} + W_{ci}^R c_{t-1})$$  \hspace{1cm} (11)$$

$$f_t^L = \sigma(W_{hf}^L h_{t-1} + W_{hf}^R h_{t-1} + W_{cf}^L c_{t-1} + W_{cf}^R c_{t-1})$$  \hspace{1cm} (12)$$

$$f_t^R = \sigma(W_{hf}^L h_{t-1} + W_{hf}^R h_{t-1} + W_{cf}^L c_{t-1} + W_{cf}^R c_{t-1})$$  \hspace{1cm} (13)$$

$$x_t = W_{hx}^L h_{t-1} + W_{hx}^R h_{t-1}$$  \hspace{1cm} (14)$$

$$c_t = f_t^L \odot c_{t-1}^L + f_t^R \odot c_{t-1}^R + i_t \odot \tanh(x_t)$$  \hspace{1cm} (15)$$

$$o_t = \sigma(W_{ho}^L h_{t-1} + W_{ho}^R h_{t-1} + W_{co}^R c_t)$$  \hspace{1cm} (16)$$

$$c_t = f_t^L \odot c_{t-1}^L + f_t^R \odot c_{t-1}^R + i_t \odot \tanh(x_t)$$  \hspace{1cm} (17)$$

where $\sigma$ is the sigmoid function, which is used to limit the gate signal between $[0, 1]$; $f^L$ and $f^R$ represent the left and right forgetting gates respectively; $W$ is the network weight matrix; $\odot$ represents a Hadamard product.

In repetitive models like LSTM, when the sequence becomes too long, the output is biased towards the most recent historical state rather than considering the entire historical state equally [9]. In this paper, we represent memory with a tree-like structure, which maintains logical consistency between adjacent memory units at different levels of abstraction, and successfully transfers the important features of adjacent regions to the previous layer.

4. Experiment and Analysis

4.1. Dataset and Environment

The experiment takes the Porto taxi trajectory data. The dataset is the trajectory data of 442 taxis collected in Porto from 2013-07-01 to 2014-06-30. The data is stored in csv format, retaining the latitude and longitude information and timestamp information in the original data. The full training set includes a set of 1.7 million data points, each representing a complete taxi trajectory. Each complete taxi trajectory consists of a series of GPS points spaced 15...
seconds apart and the timestamps taken to the first point, the timestamps of subsequent sample points are added 15 seconds at a time, and the last GPS point represents the destination.

4.2. Result analysis

Fig.3 ADE comparisons of all method

Fig 3 shows the average error distance of each prediction method. The prediction results of the TMM-LSTM and LSTM models are much better than that of the RNN, which also shows that the long-term dependencies between trajectories have a great impact on the prediction results. Moreover, the prediction result of the TMM-LSTM model is also better than that of LSTM, indicating that the TMM-LSTM model can better solve the long-term dependency problem and is effective in taxi destination prediction.

Fig.4 PA comparisons of methods

As shown in Fig 4, the prediction accuracy of TMM-LSTM is lower than none. In terms of the convergence speed of prediction accuracy, the TMM-LSTM model performs better, indicating that the TMM-LSTM model can better capture and handle long-term dependencies. In terms of prediction accuracy, the TMM-LSTM model is better than other models, and the prediction accuracy is 1% higher than that of the LSTM model.

Fig.5 Haversine distance distribution comparisons of methods

Fig 5 shows the distribution comparison results of the haversine distance between the predicted destination and the real destination of each prediction model. The haversine distances of the predicted destination and the real destination of the TMM-LSTM and LSTM models are concentrated within 1.5km, accounting for about 70%. This also shows that both models have better performance in taxi destination prediction. Compared with other models, the ratio of the haversine
distance within 1.5km of TMM-LSTM is much higher, which also shows that the TMM-LSTM model performs better in
more accurate predictions.

Finally, the ADE of all comparison methods is shown in Table 1. The average distance error of the TMM-LSTM model
is 2.59km. Compared with RNN and LSTM, it dropped by 0.16 km and 0.42 km.

<table>
<thead>
<tr>
<th>Method</th>
<th>ADE (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNN</td>
<td>3.01</td>
</tr>
<tr>
<td>LSTM</td>
<td>2.75</td>
</tr>
<tr>
<td>TMM-LSTM</td>
<td>2.59</td>
</tr>
</tbody>
</table>

This article gives a detailed introduction to the TMM-LSTM model. The theoretical basis of TMM-LSTM is given, and
the tree memory module is introduced. This paper uses tree memory module to memorize the long-term relationship
between trajectory points to enhance the LSTM model. The TMM-LSTM model is applied to the prediction problem of
taxi travel purpose to better solve the long-term dependency problem in the trajectory. Through the comparison of
experimental results, it can be found that the TMM-LSTM model has a great improvement compared with RNN and
LSTM. prediction model reduces the AED of trajectory prediction compared to other methods and improves the
prediction accuracy.

Acknowledgments

This work was supported by the Shandong Provincial Natural Science Foundation (ZR2020QF110). Thanks very much.

References

transportation system: A review and bibliometric analysis." International Journal of Production Economics 231
(2021): 107868.
1780.
[9] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. Enhancing and combining sequential and tree
Research Advanced in Human Pose Estimation based on Deep Learning

Ye Xia¹,a*,†, Xinxue Jiang²,b†, Liaowenjin Yan³,c†
¹Beijing Royal School, Beijing, China
²Tiangong University, Tianjin, China
³Jiangxi Science & Technology Normal University, Jiangxi, China
a*xiaye@hrs.edu.cn, b1127175575@qq.com, cjcwl@jxstnu.edu.cn
†These authors contributed equally

Abstract

Human pose recognition is a current research hot spot, which aims to detect the information of human joints, orientations and scales from images or videos and predict the specific pose of the human body based on all the key point information. Benefiting from the rapid evolution of deep learning, target detection algorithms based on convolutional neural networks have achieved breakthroughs in both accuracy and efficiency. On the basis of detailed literature research and analysis, this paper provides a comprehensive evaluation of the research progress of human pose estimation. Specifically, we mainly introduce the classification of human action poses and the methods of human action pose recognition, including the design ideas, basic framework and advantages and disadvantages of representative recognition methods. Finally, we summarize the main challenges and give an outlook on the future research development of object detection.

Keywords—Pose Estimation; Deep Learning; Markov model; Dynamic Bayesian network; template matching

1. Introduction

Human-computer interaction is a discipline of interaction between systems and users. Since the end of the 1970s, with people's continuous understanding of it, a multi-modal human-computer interaction has gradually formed, which is to imitate the interaction methods such as gestures, touches, expressions, and voices in human life[1]. These interaction methods enable computers to obtain information about the outside world and perceive the world like humans to a certain extent. For example, the ability to acquire visual perception is to have one's own "eyes" like a human being. However, an unavoidable problem in achieving efficient human-computer interaction is how to enable machines to correctly understand human behavior. In this context, pose estimation has been proposed and has become one of the important technologies in academia and industry[2].

In computers, posture estimation is a very useful method to help people better understand their movements and actions. Posture estimation is generally related to posture recognition, but they are two different concepts. The purpose of motion recognition is to carry out special motion, such as walking, standing, etc., in the image or image. Postural assessment is based on the reconstruction of various parts and joints to identify certain parts of the body. A person's image will appear in the image image. In general, gesture-based estimation can be used for action identification. In the early stage of computer vision, people's expression and estimation of body posture has always been the focus of people, but due to the lack of intuitive technical support, the preliminary research on body posture is limited. In addition, body movements have more than 240 degrees of freedom, so it is difficult to describe them accurately in theory. In recent years, due to the continuous development of imaging technology and computer computing power, people have made some new progress in the performance and classification of sports[3]. The movement of the body is made up of many parts, including overall movement and partial movement, such as the movement of the hands, the movement of the legs, the movement of the head, and the human face and posture. With the rapid growth of information, more and more research has been done on this kind of image and image data. In practical work, the estimation method of body posture is of great use in many aspects, including:

(1) Human-computer interaction and motion analysis. Human pose estimation is mostly used in VR (Virtual Augmentation) or AR (Reality Augmentation), such as: some entertainment games, which are used for game interaction by detecting the pose of the human body. In addition, it is also used for sports analysis, such as analyzing the patient's...
walking posture to diagnose bone recovery; or the sports industry supervises the posture of athletes to create auxiliary training systems, which can analyze the movements of athletes at every moment and assisting athletes to find better postures.

(2) Pedestrian capture for autonomous driving. At present, autonomous driving is gradually entering our lives and many regions are testing driver-less cars. For driver-less cars, in order to avoid danger, it is not only necessary to accurately detect the human body in front of them, but also need to accurately predict and understand the next behavior of human posture.

(3) Content-based video indexing and retrieval. The research on pose estimation will be of great help to obtain certain features in the video, which can assist users to extract the desired content and features of the video according to their own preferences, such as the know the specific action of this frame of the video.

(4) Animal Tracking. For animal researchers, understanding some gesture behaviors of animals is also a key step. By tracking animal tracks and understanding behaviors, they can gain a deeper understanding of animal psychology so that can better get along with animals.

(5) Identification and intelligent monitoring. The research of human pose estimation can detect the walking posture and gait of the human body to identify the specific person. Effective monitoring of abnormal behavior and alerting can significantly reduce the labor of monitoring personnel.

Focusing on the main technical frameworks of deep learning based human pose estimation, in this paper, the representative human pose estimation methods will first be introduced and their advantages and disadvantages will be further analyzed. Finally, we compare the performance of these methods on common pose estimation datasets and prospect the future development trend of this research field.

2. Human Pose Estimation based on Deep Learning

Statistics-based, template-based, grammar-based, depth-based, depth-based, statistics-based, template-based, grammar-based, depth-based, in the next section.

2.1 Statistical-based methods

Statistics-based methods are the classic solutions in dynamic recognition systems. At present, there are mainly two methods: Hidden Markov Model (HMMo) and Dynamic Member Yeasian Network (DBN).

Hidden Markov Horizontal Markov model is one of the most widely used methods today, which are composed of states and in-view parts, and are a stochastic type based on transition rate and transmission probability[3]. The estimated abundance of the current state is only related to the state of the previous moment, and has nothing to do with other historical state conditions. In the process of HMMs recognition, the feature vector sequence is first proposed and then the model parameters are trained by the learning algorithm to finally recognize the unknown motion sequence. Yamato et al. The first use of HMM to recognize human action poses. Since then, various HMM-based improved models have flourished and become the main framework for human action pose recognition. Brand et al. proposed a coupled hidden Markov model (CHMM) and performed sign language recognition. NguyenP et al. further proposed a Hierarchical Hidden Markov Model (HHMMo). HHMMs have the structure of multilayer Hidden Markov Models, which can clearly express the behavioral details in human motion. Duong et al. proposed S-HSMM (Swrchling Hid -den-semi Markov Model) model, which can achieve a better recognition performance than S-HSMM while is limited due to its high complexity[2].

The above-mentioned various hidden Markov models have greatly improved the accuracy of human pose estimation, but there are still some potential shortcomings. First, since HMMs are only suitable for systems that describe a single moving object. With the increase of moving objects numbers, the number of states, the parameters of the state transition matrix and observation models in the HMM will increase exponentially, which greatly increases the computational complexity of the system and inhibits the application scope of the HMM model. Secondly, the HMMs model cannot effectively integrate information to supervise the parameter errors, which leads to the inability to guarantee the accuracy of model parameters.

Dynamic Bayesian neural network (DBN) is a statistical method that can learn from random and random processes. Expandable and interpretable. Multi-object motion system identification is an effective method and it will become a developing trend in the future. In dynamic identification, DBN can design the corresponding network architecture
according to the characteristics of a specific problem, set weights, and synthesize various information for inference. In recent years, researchers at home and abroad have identified DBN and obtained certain results to some extent. In addition, some scholars have organically combined HMM and Bayesian neural network to achieve accurate modeling and identification of human moving targets. However, there are few researches on the application of Bayesian neural networks in human movement and body language. The reasons are as follows: 1) Because of the large number of characteristic parameters required by The Bayesian neural network when identifying the target with relative motion, the computation of the Bayesian network is very large. 2) The Hm algorithm should be able to complete the identification of general human movements in the current popular simple scenes and obtain good results.

### 2.2 Template-based methods

Template matching method, dynamic time-space deformation method and dynamic planning method are commonly used template matching techniques. The method first trains the motion posture of the human body, and compares it with the template of the model to get the similarity with the model. This algorithm is simple to implement and requires little computation. It has the disadvantage of being more sensitive to changes in voice and body position. In this method, the temporal sequence of the recognized human movement pattern is linearly projected onto the temporal sequence of the training template to reduce the distance between them. The dynamical - spatial bending method can be used to deal with the uncertainty of human movement and posture in time range. In the identification of dynamic scheduling method, the matrix of each sample can correspond to any feature without time adjustment. However, the requirements must be completed in the order of [2]. The biggest drawback of this method is that the larger the number of training samples is, the larger the computation will be.

The initialization of shape, appearance and others is an important step in the process of automatically capturing and analyzing human motion. Automated solutions have been discussed for a long way, which focuses on how to initialize human kinematic structures and how to view images from a single view to obtain human pose. In monocular view tracking and pose estimation, how to initialize the appearance model is still an open problem. Some methods propose the appearance model, which is initialized with an instance image patch or a mixed color model. Recent work on body part detection uses supervised learning method to distinguish the appearance and background of various body parts. Only a few studies have discussed how model the relationship between the changes of the human body appearance and motion. The methods that fully automatically initialize human kinematic structure, shape and appearance from single-view image sequences model can be used for pose estimation remains for future research.

Some researchers also have proposed to initialize from multi-view image sequences with an intermediate volume reconstruction process. Methods of human movement structure are able to solve the automatic initialization of human kinematic structure models for human pose estimation problem. The learning method and anthropocentric model are used to initialize the joint angle range of the human kinematic structure, which is used to track the person and motion estimation. In the past 5 years, there has been a large amount of research on automatically initializing model shapes from multi-view images. Constructing a human body model connected by joints can approximate the shape of a specific person and improve tracking accuracy. Recent studies have also begin to address modeling changes in human shape in motion. Similarly, multi-perspective reconstruction techniques can also be used to automatically initialize the appearance of the mannequin.

### 2.3 Deep learning-based methods

The main problem facing AI is solving tasks that are easy for humans but hard to formally describe, such as recognizing an acquaintance in a crowd. These problems are intuitively solvable for humans, but impossible for computers. Deep learning has made great progress in the past decade since it was formally proposed in 2006 [3]. Deep learning focuses on building very deep graphs where each layer of the graph is relatively simple. Deep learning finally achieves high-quality description of scenes by combining information from different layers through complex nonlinear transformations. Deep learning has been applied to a variety of visual tasks. For human pose estimation, according to different design ideas, methods based on convolutional neural network can be divided into two categories: detection based and regression based.

On this basis, the probability distribution map, namely heat map, is used to express each important node, and estimate the corresponding probability of each pixel. The pixel is more likely to be away from the actual node, and vice versa. The algorithm uses the quadratic element to represent the position of each node, and uses the characteristics of the human body to correspond to the position of the human body to train the model and get the coordinates of each node. Tosef et al. Using AlexNet as the node network, coordinate recovery of nodes is realized. Firstly, the coordinate of each node is obtained by using the existing serial network, and then the position of each node is obtained by using the image.
data on the basis of the coordinate, so as to achieve higher precision. Cascade network improves the positioning accuracy of regression network to a certain extent, but it is not suitable for low resolution input images. At the same time, as cascade algorithm is adopted, CNN is needed to carry out multi-layer convolutional operation in each input image. Fan et al., as the estimation of DeepPose cannot fully take into account the local shape, it is difficult to estimate the complex human posture. Ds-cnn (DS-CNN) was used to estimate the motion posture of human body. By extending r-CNN from single source to DS-CNN, the dual-source model DS-CNN can provide a comprehensive Angle attitude estimation for researchers. The DS-CNN is trained using a set of different types of image segments detected by the input image, and can be applied at a variety of scales compared to the sliding Windows or holograms used in previous human posture assessment methods. Capture better body language. The local (body) local and local global images were imported into DS-CNN for fusion and joint examination respectively to determine the joint locations and parts in the image, as well as the central part of the image. A more accurate estimate of human posture can be obtained by combining detection/position analysis from various images. The puppy, etc. This paper introduces a new algorithm for deep convolutional networks based on optical flow. It benefits from context by using light flow to divide the problem into several frames.

Some works estimate human posture based on generative adversarial networks. In view of the lack of effective feature representation in human pose estimation, Chou et al. Proposed a solution using generative adversarial networks. It builds two stacked hourglass network frameworks with the same architecture, generator, and discriminator. Generator network is a full convolutional network with residual blocks. After the input image is fed forward through the generator, a set of heat maps representing the confidence score of each node at each position are obtained. The heat maps generated by the generator network and the real heat maps are fed into the discriminator network, and the two sets of heat maps are reconstructed. The losses between the discriminator output and the real heat maps and the heat maps generated by the generator, L_true and false, are calculated respectively. In order to solve the over-fitting caused by training differences between generators and discriminators, we use the variable k_t to control the balance between generators and discriminators. The adaptive term can be expressed as:

\[ k_{t+1} = k_t + \lambda_k (\gamma L_{real} - L_{fake}) \quad (1) \]

Where k_t represents the importance of L_{fake}, t is the number of iterations; \gamma and \lambda_k are hyperparameters. When L_{fake} < L_{real}, it means that the heat-map generated by the generator is real enough to fool the discriminator, so the k_t value will increase to make L_{fake} more important. At the same time, the discriminator is trained more to identify the generated heat-map, where the speed of training depends on the distance between the discriminator and the generator, i.e. the \gamma L_{real} - L_{fake}. Also, when the discriminator is better than the generator, i.e. L_{fake} > \gamma L_{real}, k_t will be reduced to slow down the training of the network so that the generator output can keep up with the discriminator output.

In the last decade, a composite model has been used to assess various human postures. The human body composition model describes the whole body as a layered structure composed of multiple parts with specific joint constraints. The structure of this layer can reflect the structure of different parts of the body. This method uses a series of discrete parameters to model the compatibility of components, which includes not only the orientation and size of components, but also interdisciplinary knowledge such as straight arm and curved arm. Due to the existence of a large number of different types of components and all their subclasses, the state space of higher level components will show an increase in geometric multiples, which is the key to operation and storage. Tang and others try to make things right. On the basis of studying the complex structural relations of human body, this paper presents a DLCM (DLCM) for attitude estimation using depth analysis method and an expression mode for spatial regional information summary (SLIS) based on human skeleton, which can accurately identify the scale, direction and shape of each part. Thus, error components can be effectively prevented. The 16,12 and 6 levels of this pattern are three different semantic levels. Compared with previous prediction methods, DLCM is a hierarchical model that spans multiple semantic levels, and its assessment of mood is similar to multi-user assessment of mood, showing a bottom-up/top-down networking process. None of the previous CNN-based human posture assessment models divided the entities into reusable meaningful levels and did not deduce the semantic levels.

### 2.4 Grammar-based methods

This paper uses a grammar-based language, which includes constrained state machines and context-independent syntax. In recent years, many scholars at home and abroad have carried out in-depth discussions on this, and believe that using natural language to describe the body movement and posture is the future direction of development. However, the use of a series of images to describe the movement and posture of the human body, will be subject to events, behavior, height
and other conditions. It is very difficult. In business investigations, the description of body movement and posture has been used for semantic interpretation. We are going to go a step further and illustrate the movement of the body from the movement of the body.

### 3. EXPERIMENTS

#### 3.1 Datasets

It is necessary to train a large number of motion parameters when carrying out deep movement. In recent years, a large number of human posture assessments have been published. Table 1 lists some of the internationally used human posture assessments.

<table>
<thead>
<tr>
<th>Datasets name</th>
<th>Combination</th>
<th>Number of joint points</th>
<th>Number of samples/10³</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSP[9]</td>
<td>single</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>PennAction[10]</td>
<td>single</td>
<td>13</td>
<td>≈2</td>
</tr>
<tr>
<td>JHMDB[11]</td>
<td>single</td>
<td>15</td>
<td>≈1</td>
</tr>
<tr>
<td>FLIC[12]</td>
<td>single</td>
<td>10</td>
<td>≈20</td>
</tr>
<tr>
<td>MSCOCO[14]</td>
<td>multiple</td>
<td>17</td>
<td>&gt;300</td>
</tr>
<tr>
<td>HKD[15]</td>
<td>multiple</td>
<td>14</td>
<td>≈300</td>
</tr>
</tbody>
</table>

#### 3.2 Evaluation indicators

According to different task requirements and data characteristics, an evaluation method based on human pose is proposed. The general evaluation methods are: similarity of target key points. OKs is a common evaluation method for human skeleton key point recognition. It is inspired by the IoU index in object recognition to find the similarity of key parts of the human body. The average precision (AP) was used to calculate the precision on the test set. The presentation key (policy key) specifies the percentage of nodes. The normalized distance of the detected critical point and its corresponding ground-truth is taken as the ratio of the normalized distance smaller than the set threshold.

#### 3.3 Performance analysis

As can be seen from Table II and Table III, this paper compares MPII data and MSCOCO's multi-person pose estimation method.

<table>
<thead>
<tr>
<th>Method</th>
<th>head</th>
<th>shoulder</th>
<th>wrist</th>
<th>elbow</th>
<th>hip</th>
<th>knee</th>
<th>ankle</th>
<th>Ave</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRNet</td>
<td>98.6</td>
<td>96.9</td>
<td>89.0</td>
<td>92.8</td>
<td>91.5</td>
<td>89.0</td>
<td>85.7</td>
<td>92.3</td>
</tr>
<tr>
<td>MSSA</td>
<td>98.5</td>
<td>96.8</td>
<td>88.4</td>
<td>92.7</td>
<td>90.6</td>
<td>89.3</td>
<td>86.3</td>
<td>92.1</td>
</tr>
<tr>
<td>FRMs</td>
<td>98.5</td>
<td>96.7</td>
<td>88.7</td>
<td>92.5</td>
<td>91.1</td>
<td>88.6</td>
<td>86.0</td>
<td>92.5</td>
</tr>
<tr>
<td>HRUs</td>
<td>98.5</td>
<td>96.3</td>
<td>88.1</td>
<td>91.9</td>
<td>90.6</td>
<td>88.0</td>
<td>85.0</td>
<td>91.5</td>
</tr>
</tbody>
</table>

TABLE III. AVERAGE PRECISION OF MULTI-PERSON POSE ESTIMATION ALGORITHMS ON MSCOCO

<table>
<thead>
<tr>
<th>Method</th>
<th>AP/%</th>
<th>APM/%</th>
<th>APL/%</th>
<th>Time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPN</td>
<td>73.0</td>
<td>69.5</td>
<td>78.1</td>
<td>-</td>
</tr>
<tr>
<td>OpenPose</td>
<td>61.8</td>
<td>57.1</td>
<td>68.2</td>
<td>0.600</td>
</tr>
<tr>
<td>Mask R-CNN</td>
<td>62.7</td>
<td>57.4</td>
<td>71.1</td>
<td>0.200</td>
</tr>
<tr>
<td>SPN</td>
<td>66.9</td>
<td>62.6</td>
<td>73.1</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Through the analysis of the table data, the following conclusions can be drawn: 1) In the pose estimation of a single target, the recognition accuracy of each key point is quite different. Taking HRNet as an example, the method is 98.6% correct, 96.9% correct for shoulders, and less than 90% correct for knees and ankles. The difficulty of identifying each key point of the human body significantly affects the final pose estimation accuracy, and some other methods can also reflect this trend. How to improve the measurement accuracy is an important topic in the future. In addition, in a single position estimation, the positioning results of each algorithm are very different, and the positioning accuracy distribution of each key point is relatively close, which also shows that the influence of the above key points on the individual posture is obvious. The result of the posture assessment. 
(2) The accuracy of multi-person pose estimation varies greatly. In terms of AP index, the peak value of CPN reached 73%, surpassing the traditional Open Pose and Mask RCNN algorithms. Our performance improvement is due to the following aspects: on the one hand, the CPN method adopts the
local and global information of fusion points; on the other hand, it prevents occlusion by mining hard key points on the network. The key points and difficulties will be solved centrally.

4. DISCUSSION

In recent years, although the recognition of human poses has gone an comprehensive and rapid development process in the past research and has been widely used in many fields, the existing algorithms still have many fluxes, and its future research trends are briefly described below from several aspects.

(1) Restricting the movement of people. At present, people's understanding of human motion posture is mostly limited to simple human motion posture, such as: standard body posture, and simple movements such as walking, running, squatting, etc.; simple standard movements, such as standing. Objects are generally the actions of a person. Therefore, in a complex environment, how to correctly identify the interaction of multiple people is the future development direction.

(2) It is difficult to select features. The selection of feature vector has a great influence on the recognition effect. If multiple features are selected, the dimension of the feature vector will be too large, which will increase the complexity of the operation. Conversely, if the number of selected features is too small, high identification accuracy cannot be achieved. Therefore, there is a trade-off between correct feature selection and improved recognition accuracy.

(3) Limitations of machine learning; the application of machine learning has made some progress in recognizing human behavior and behavior, but it is still in its infancy. In recognition, limiting factors must be introduced into the motion patterns of the human body to reduce ambiguity. This constraint is different from the prevailing reality. Often incompatible. In addition, the way of machine learning is also a very high training, that is to say, in order to recognize each action, you must go through a lot of training. Therefore, in the learning process, machine learning can only identify specific behaviors, but not new behaviors. Therefore, machine learning has always been a thorny problem in human action recognition.

(4) Evaluate the performance of the algorithm. Robustness, real-time and accuracy are the three main factors to measure the performance of this method. At present, people are most concerned about the accuracy rate, while the other two are not taken seriously, especially the robustness of the algorithm. However, in practical applications, the real-time and robustness of the method need to be fully considered. Therefore, when evaluating the performance of an algorithm, these three factors need to be combined. Therefore, in practical applications, it is necessary to develop a new method that can effectively improve the recognition effect and reduce the computational complexity of the algorithm.

(5) The variability and complexity of human postural behavior. In different movement behaviors, the amplitude of human body is obviously different, so the accuracy of the test results obtained by the same model algorithm is different, and often large movements, such as kicking, swimming and other sports can get a large deviation value. For such problems, you can choose node detection.

(6) The accuracy of the algorithmic results is influenced by the environment. Various factors such as camera, light, photographer angle, etc. can affect the accuracy of the algorithm. For common lighting problems in practical applications, how to perform dynamic compensation of lighting or brightness enhancement of image adaptation is the focus of research. Reference [2] provides a solution for compensating image processing in complex environments through RGB-D sensors. At present, in the traditional two-dimensional human body part detection, the map mode is used to establish connectivity, and the human body pose estimation is carried out in combination with human kinematics dynamics.

5. CONCLUSION

Following the development of technical design ideas, this paper summarizes the current representative algorithms in human pose estimation from the statistical-based methods, template-based methods, grammar-based methods and deep learning based methods. We also compare and analyze their performance on mainstream datasets. Finally, the current research difficulties and future development trends are detailed.
REFERENCES

[1] T.B.Moeslund, "A survey of advances in vision-based human motion capture and analysis", Computer Vision and Image Understanding, 2006/12
Research on the Main Elements of Mimic Platforms

Bo Zhang*ab, Zesheng Xib, Yunfan Wangab, Chuan Heab

a State Grid Smart Grid Research Institute co., Ltd., Nanjing, China
b State Grid Key Laboratory of Information & Network, Nanjing, China

* Corresponding author: zhangbo@geiri.sgcc.com.cn

ABSTRACT

Cyber Mimic Defense (CMD) is a new generation of active defense technology after firewall, intrusion and other traditional defense technology. It aims to deal with uncertain threats in the network environment with visual uncertainty. This paper briefly introduces the architecture and principle of CMD and defines four main elements of the current simulation platform: system architecture, heterogeneous policy, scheduling policy and voting policy. Combined with examples, the four elements are respectively summarized. The system architecture is divided into C mode and D mode, and the heterogeneous strategy includes implementation mode, implementation method and synchronization mode. Scheduling policies are classified into offline policies and online policies. Voting policies include voting algorithms, voting levels, and delay control.

Keywords: mimic platform; heterogeneous policy; scheduling policy; voting policy

1. INTRODUCTION

With the development of information technology, there are many unknown threats in current cyberspace, and various defense strategies come into being. How to build a secure and reliable system based on contaminated components, J. Wu et al. [1] proposed the idea of imitation defense, which uses the idea of dynamic heterogeneous redundancy to break the relatively static execution environment on which vulnerabilities depend. CMD, as a new-generation active defense technology is verified in 2013 and made public in 2016 [2]. It is proved to be an effective defense strategy for space security that aims to respond to uncertain threats with the visual uncertain network.

At present, the related work mainly focuses on the principle of CMD architecture and the implementation of a simulation platform [3]. The theoretical work is relatively clear, but the specific scheme for the implementation of the simulation platform still needs to be improved. Based on COTS-level products, self-owned platforms and open-source platforms, a mimic platform family containing more than 10 types of platforms with different functions has been formed. After simulation reasoning and physical network verification, the endogenous safety effects of the system can be improved to a certain extent. According to the DHR architecture and the principle of CMD, the main implementation elements can be summarized as system architecture, heterogeneous strategy, scheduling strategy and voting strategy. This article focuses on the first two elements.

This article mainly summarizes the system structure, heterogeneous strategy, scheduling strategy and voting strategy of existing mimic platforms. The system architecture is divided into C mode and D mode, and the heterogeneous strategy includes implementation mode, implementation method and synchronization mode. Scheduling policies are classified into offline policies and online policies. Voting policies include voting algorithms, voting levels, and delay control.

2. SYSTEM ARCHITECTURE OF CMD PLATFORMS

Most of the system architecture implementations of mimic system platforms, that do not subvert the original architecture, but use incremental deployment. The heterogeneous transformation is achieved by setting mimic brackets on the core parts of the original systems. According to the input and output interaction methods, current mimic system platforms can be divided into centralized architecture (C-mode architecture for short) and distributed architecture (D-mode architecture for short).
2.1 C-mode architecture

The C-mode architecture provides centralized mimic services based on the C/S like the "request-response" working mode, which can be divided into two types of sub-architectures, component-based type and layered-based type. Component-based sub-architecture, such as mimic Web server [4] and mimic DNS server [5], are mainly composed of different components such as scheduler, heterogeneous server pool, and request-distribution balance module; hierarchical-based sub-architectures, such as mimic cloud data center [6] and mimic cloud service architecture [7], among them, the former system platform consists of an infrastructure layer, a heterogeneous network switching layer, a mimic bracket layer, a mimic cloud manager/mimic SaaS (Software-as-a-Service) application layer, and operation-maintenance management layer, the latter system platform does not change the original cloud platform system structure, but provides cloud service node mimic services by constructing MSP (Mimic Service Package) in the service layer [7], which is shown in Fig. 1.

Therefore, in C-mode architecture, when the entity as the client interacts with the meta-function body as the server, the information flow is output in a single line and is processed in a chain. After receiving the service request of entities, the meta-function distributes input, multimode execution, output arbitration and scheduling feedback and other mimic services are all carried out in the meta-function body.

The I/O functions are located on the same side of the architecture, the I/O interface at each level is normalized, and the mimic structure is realized by directly introducing the mimicry module (input agent and output decision) in the I/O interface.

![Figure 1. C-mode architecture and its information flow diagram](image)

2.2 D-mode architecture

The D-mode architecture adopts the peer-to-peer model of P2P-like architecture to provide distributed mimic services. It can be divided into 2 types of sub-architectures, component-based type and layered-based type. Component-based sub-architectures, such as mimic industrial control processors [8], and mimic switches [9], are composed of heterogeneous executor pool, scheduler, circuit, interfaces and other components; hierarchical-based sub-architectures, such as mimic routers [10] and MNOS (Mimic Network Operating System) [11], the former system platform is composed of an application layer (that is, the service set), a control layer, and a device layer (that is, the standard Open flow switch), and the latter system platform added a mimicry layer between the SDN data layer and control layer [8], which is shown in Fig. 2.
Therefore, in the D-mode architecture, entity can interact with each other, and the information flow is multi-level and multi-chain parallel. After the system receives input information I, only the internal output information O is triggered. The input I of the external interface and the output information O are not logically independent, that is, the I/O function is located in the architecture on the opposite side; in order to facilitate the judgment, the internal output vectors need to be normalized, and the mimic construction is realized by directly introducing mimicry module (input agent and output decision) in the I/O interface.

![Diagram](image)

**Figure 2. C-mode architecture and its information flow diagram**

### 3. HETEROGENEOUS STRATEGIES OF CMD PLATFORMS

Heterogeneous strategies of mimic platforms can realize full-stack heterogeneity at multiple levels such as infrastructure, operating system, file system, application software, etc. It can also enhance heterogeneity through dynamic strategies such as virtualization. There are 4 key aspects of heterogeneous strategies, implementation mode, realize hierarchy, and component synchronization.

#### 3.1 Heterogeneous implementation mode

The heterogeneous implementation modes of mimic products can be divided into 4 types. The first type is single source closed heterogeneous implementation, that is, for components that need to ensure availability, such as resource allocation and data storage, by directly adding mechanisms such as distribution, adjudication, arbitration and scheduling to the target component I/O interface, mimicry processing for business protocols [6]; the second type is single-source open heterogeneous implementation, that is, mimic circles are set for the core part of open source COTS-level products, and the source code of single-source products can be diversified and compiled. The third type is the multi-source closed
heterogeneous implementation, that is, the core parts of the non-open-source COTS-level products are set with mimic input and adjudication components to construct mimic subsystems.

Online modules of different products can be directly connected as heterogeneous executors due to the application and the platform are tightly coupled; the fourth type is multi-source open heterogeneous implementation, that is, for open source systems with dynamic, heterogeneous, and redundant characteristics, by directly adding part of the mimicry mechanism and achieve heterogeneity by combining multi-source applications due to the application and the platform are loosely coupled and have a better heterogeneous foundation.

For example, the mimic router architecture [10] controls the open source and virtualizable Cisco and Juniper routers through interface transformation based on the Open-Flow protocol, which is a single-source closed heterogeneous implementation; By combining physical router equipment and SDN controllers, controlling the open source and non-virtualizable ZTE, Maipu, and Fiber home routers are multi-source closed heterogeneous implementations; the open source software Route flow [11] directly migrates and controls the open-source Quagga, XORP and other routers, which is a multi-source open heterogeneous implementation.

3.2 Heterogeneous implementation method

Heterogeneous realization methods can be divided into coding heterogeneity, software heterogeneity, hardware heterogeneity, software and hardware cooperative heterogeneity, and mixed heterogeneity [6] which are shown in Fig. 3-6. Heterogeneous implementation methods and heterogeneous implementation modes do not correspond one-to-one and can be combined flexibly according to the actual situation. For example, the multi-source open heterogeneous implementation mode can achieve only software heterogeneity, or software and hardware coordination heterogeneity.

Figure 3. single-source closed heterogeneous implementation

Figure 4. single-source open heterogeneous implementation

Figure 5. multi-source closed heterogeneous implementation
1) Coding heterogeneity. Coding heterogeneity, that is, using different data encodings to achieve heterogeneity. For example, mimic memory splits files into blocks, randomly selects erasure code encoding to obtain several coding blocks, and stores file blocks and coding blocks in distributed storage at the same time to realize storage coding heterogeneous.

2) Software heterogeneity and hardware heterogeneity. Implementation of software heterogeneity means can be achieved through the reverse stack, ISR, heap layout randomization, stack base address randomization, stack protection, system call number randomization, and library function name randomization; in terms of implementation links, software heterogeneity can be implemented in development, compilation, linking, installation, loading and other links. For example, the mimic honeypot sets heterogeneous settings for two software layers of Web middleware and operating system in the honeypot. Most heterogeneous platforms realize software heterogeneity in the development stage by selecting heterogeneous software or transforming scripts, instead of compiling, linking, installing, loading and other links, which has high operational feasibility and low cost of heterogeneity. In heterogeneous hardware, due to different interfaces, different control software needs to be introduced when the hardware is different. Therefore, there are few pure hardware heterogeneous implementations at present.

3) Software and hardware synergistic heterogeneity. For example, the mimic switch uses different {CPU+OS} hardware modules, and generates heterogeneous protocol stacks and management software families through diversified compilation; the mimic firewall carries out mimic transformation for business processing logic and other levels and realizes 3-layer software and hardware heterogeneity of {hardware platform + operating system + security engine}. Therefore, the mimic platforms can use the combination of open-source products, self-owned products and COTS-level products by setting mimic brackets in the data interface and other parts to realize the collaborative heterogeneity of software and hardware.

4) Mixed heterogeneity, Coding heterogeneity. That, heterogeneous network components and heterogeneous network exchange are used to realize multi-dimensional hybrid heterogeneity. For example, in the mimic cloud data center, in addition to deploying mimic components, it also introduces heterogeneous strategies such as mimic storage, mimic cloud management, and heterogeneous network switching.

3.3 Heterogeneous synchronization mode

There are 3 main types of synchronization modes, the first mode is direct connection mode, that is, heterogeneous executors directly communicate through the interface between the executive bodies, such as in order to achieve high-bandwidth data communication, components of the mimic computing server adopt FPGA full interconnection mode, and can expand performance through connecting other computing servers with switches; the second mode is mimic plugin mode, that is, heterogeneous executors communicate indirectly through the scheduler or other mimic plugins, such as the mimic processor schedule and manage user requests, respond to arbitration and executor fault isolation, cleaning reset based on processor state data through the scheduler; the third mode is the message queue mode, such as MNOS uses the Zero MQ lightweight heterogeneous sub-system based on message queue to complete communication between heterogeneous executors. Among them, the direct connection mode is the most efficient, but may face coordinated attacks; the mimic plugin mode is more secure, but it causes a certain performance loss; the performance of the message queue mode is somewhere in between.
4. SCHEDULING STRATEGY OF CMD PLATFORMS

The scheduling strategy can be divided into 2 parts: the offline sub-strategy for the old executive body and the online sub-strategy for the new executive body.

4.1 Offline sub-strategy

The offline sub-strategy is the sub-strategy for scheduling and selecting a specific old executor to leave the service set, mainly including offline operation timing and offline cleaning methods. The first important point is the timing of offline operations, which can be divided into timed offline, adjudicated offline, manual offline, etc., among them, scheduled offline, that is, according to the trust weight, the executor is selected for offline cleaning at regular intervals to prevent collaborative penetration and common mode defects and such attacks; ruling offline, that is, after each ruling, the executor with the highest trust value is offline and cleaned; manual offline, that is, the executive body is selected by the operator's subjective decision to go offline. For example, the mimic switch adopts the offline strategy of “timing offline + ruling offline”, the mimic router selects the executive to judge offline based on the credibility of the executors and the performance weight index. The second important point is the offline cleaning method, which can be initialized, cleared, restarted and other simple methods, but it is not suitable for attacks on the underlying software and hardware that cannot be cleared by a simple data reset, it should be reconstructed to a certain extent after the data is reset, such as mimic honeypot adopts a scheduling strategy based on state rollback and rotation.

4.2 Online sub-strategy

The online sub-strategy, that is, the sub-strategy for scheduling and selecting appropriate new executors to enter the service set, it can be based on the credibility of executors, load and other indicators, such as M-DNS implements selection and adjustment based on the credibility of executors and the load calculation selection coefficient strategy.

5. VOTING STRATEGY OF CMD PLATFORMS

When studying voting strategies, it should be focused on factors such as voting algorithms, voting levels, and voting delay control.

5.1 Voting algorithm

When choosing a voting algorithm, there are traditional strategies such as large number voting unanimous voting 2 out of n voting random voting, polling voting, majority voting and improved strategies such as self-testing majority unanimous voting algorithm, adaptive unanimous voting algorithm. Among them, majority voting has the highest credibility. Although absolute correctness cannot be guaranteed, the majority error probability decreases while the redundancy increases. There are 3 main categories of voting algorithms, The first category is the static strategy, that is, once the basic voting strategy is determined, it will not be changed. For example, M-DNS specifies the use of a large number of voting algorithms. The second category is the dynamic strategy, which can flexibly determine the basic voting strategy based on the actual situation. For example, the mimic router specifies multimode voting algorithms such as majority decision, weight decision, and random decision based on the security level. The third category is the compound strategy, that is, multiple basic voting strategies are used at the same time. that is, multiple basic voting strategies are used at the same time. For example, mimic honeypot adopts a voting strategy based on behavioral records and alarm data comparison; mimic firewall adopts hierarchical and phased adjudication; in response to time coordination and common mode defect attacks, mimic switches add trust evaluation and timing disturbance to most voting, and implement self-cleaning large-number voting based on trust weight.

5.2 Voting level

The voting level should be selected according to the voting scenario. The presentation format of the output of the heterogeneous executors and the field settings in the information, etc. are different. If the voting is at the semantic level, the granularity is difficult to grasp, usually supplemented by intelligent semantic analysis algorithms, and the false alarm rate is low, while the false report rate is higher, which is suitable for fuzzy voting; while the data level voting is more accurate and the false report rate is reduced, but the false report rate is higher, which is used for accurate voting.

5.3 Delay control

Usually, the delay of the mimic platforms mainly comes from the voting delay. In order to reduce the voting delay and improve the accuracy of voting, the output vectors of heterogeneous executors should be standardized and normalized.
For example, the mimic Web server proposes to form dedicated components to reduce cost by solidifying module hardware and decoupling non-critical functions.

6. CONCLUSION

The realization of a mimic defense platform mainly includes 4 elements, system architecture, heterogeneous strategy, scheduling strategy and voting strategy. This article summarizes the elements with examples. It can be seen that the current implementation of mimic defense platforms is mainly based on non-native endogenous security products such as COTS-level platforms and open-source products. It is difficult to achieve fine-grained heterogeneity. Therefore, next-generation mimic products with native endogenous security attributes have high research value and Broad development space.

ACKNOWLEDGMENTS

This work is supported by the Science and Technology Project of State Grid Corporation of China (Research on endogenous immune enhancement technology for secure access gateway of power IoT, No. 5400-202158332A-0-0-00).

REFERENCES

Fire detection system based on improved multi-sensor information fusion

YuLin Jiang*
College of Information Engineering, Zhongshan Polytechnic, Zhongshan, China
* Corresponding author: ghostjiang@sina.com

ABSTRACT

In the fire detection system, the evidences of multiple sensors often conflict with each other. If the classical DS(Dempster-Shafer) evidence theory is used for information fusion, it will lead to a high rate of misjudgment. In this paper, an improved DS evidence theory algorithm based on standard deviation is proposed and applied to the fire detection system to judge the fire state. This paper first describes the working principle of the system in detail, and then uses the overall standard deviation of sensor evidence to measure each sensor, judge the conflicting evidence items, then redefine the weight of each sensor according to the degree of conflict, modify the probability distribution of the conflict evidence. Finally, the DS combination rule calculation is performed to correct basic probability assignment (BPA) to obtain the fusion result. The experimental results show that the method proposed in this paper can effectively improve the evidence problems in the classical evidence theory, improves the fire identification ability and the speed of fire prediction, and greatly enhances the reliability of the system.

Keywords: Improved D-S evidence algorithm, Fire detection system, Standard deviation, Multi-sensor information fusion, Extended identity frameset

1. INTRODUCTION

Fire is a common disaster that can bring serious threat to people's lives. How to effectively monitor the fire and prevent the fire has become a problem that must be considered in production and life. The traditional fire detection methods are often aimed at the single characteristic signal of fire, but the characteristic signal of fire changes in different stages of fire occurrence. The single parameter fire detection method can not meet the response sensitivity requirements of the characteristic signal. For example, the traditional contact detection method uses sensors to detect the gas composition or temperature at the site. The principle is simple and the engineering implementation is convenient. However, when an open fire occurs, this method finds that the fire situation is obviously delayed, the anti-interference ability is poor, and there are often defects such as high false alarm rate\[1-3\].

Another commonly used detection method is infrared flame detection method, which uses infrared technology to detect the infrared spectral characteristics of the flame. However, the installation cost of this method is high and the detection accuracy of this method is not high. In recent years, with the improvement of computer processing ability, it has become a research hotspot to use machine vision to automatically judge whether a fire has occurred. This method uses the camera to collect the scene video, and through machine learning to automatically detect whether there is flame in the video, and send out an alarm signal when an abnormal situation occurs. This method has a low false alarm rate, but the installation cost is high, and the fire broadcast obviously lags behind\[4-6\]. These detection methods have their own advantages and disadvantages. They are all aimed at different fire characteristic signals. If a variety of sensors can be integrated for fire detection\[7-8\], the detection accuracy will be greatly improved and the false alarm rate will be reduced.

D-S (Dempster Shafer) evidence theory is an uncertain reasoning method, which can represent the uncertainty caused by information location or incomplete information, and has been widely used in the field of information fusion\[9-11\]. However, the classical D-S theory may get unreasonable results when the sensor information conflict is serious. In order to solve this problem, many scholars have studied this problem and put forward a series of solutions. Liu Shen\[12\] proposed an improved evidence fusion algorithm based on similarity measure. First, each piece of evidence was modified according to the similarity measure of evidence, and then the evidence was synthesized using DS synthesis formula. Wang Li et al.\[13\] introduced the concept of closeness to reflect the degree of evidence conflict, and on this basis, obtained the weight of evidence through a series of calculations, obtained a new reliability function, and then combined it using Dempster
combination rules. Ding Wenjing et al. [14] first defined the scale of evidence value, calculated the value of the evidence itself through Euclidean distance, correct the evidence source, and finally combined it according to the Dempster combination rule. These methods have studied D-S evidence theory from different angles and greatly improved the accuracy of D-S evidence theory.

Although the above methods can improve the accuracy of the fusion results, most of them are complicated in calculation, with many steps, and require a long time for fusion judgment, which is disadvantageous to the fire monitoring system with high real-time requirements. In order to solve this problem, this paper uses standard deviation to comprehensively judge the sensor evidence, find out the conflict items that cause the conflict of D-S evidence theory and correct them, and finally use the classical Dempster combination rule to calculate. This method has the advantages of simplicity, and can effectively reduce the non correlation between multi-sensor data. Experimental results show that this method has high reliability and strong scalability.

2. HARDWARE SYSTEM MODEL.

The fire monitoring system based on multi-sensor information fusion should have functions such as automatic information monitoring, intelligent alarm, background monitoring and management, etc. Its structure is shown in Figure 1.

![Figure 1. Schematic diagram of Fire Detection System](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

In this system, Lora communication protocol is used for data transmission, and terminal nodes include temperature sensor, smoke sensor, CO₂ sensor and CO sensor. Each sensor node is connected to the nearby sensors and wireless router, and the sensor node automatically selects the appropriate wireless router for data transmission, then the gateway connects to the Internet to communicate with the cloud monitoring platform in real time. The sensor automatically collects the experimental environment data. If the sensor data exceeds the set threshold, the node can automatically broadcast the information of the sensor node and send a warning signal.

3. MULTI-SENSOR INFORMATION FUSION BASED ON D-S EVIDENCE THEORY

3.1 Classical D-S evidence theory [8,15]

This paper intends to use D-S evidence theory to fuse temperature, CO₂ concentration and smoke concentration data, and infer whether the fire occurs according to the reliability function of each information to represent the occurrence of a fire.

Generally, there are three types of fire: no fire, uncertainty and fire, which are expressed by \( F_1, F_2, F_3 \). These three types
of fire are mutually exclusive, and the identification framework can be expressed as $\Theta = \{F, F_2, F_3\}$. Assuming that function $m: 2^\Theta \rightarrow [0,1]$ satisfies $m(\emptyset) = 0$ and $\sum_{F \in \Theta} m(F) = 1$, then $m(F)$ is the basic probability function value of this situation, which reflects the degree of support of the sensor for conclusion $F$, and $\emptyset$ is an empty set. According to the output data of each sensor and the probability function of fire occurrence, the current fire occurrence probability can be obtained, which is expressed as $m_{\text{temp}} = \{R_1, R_2, R_3\}$, $m_{\text{CO}} = \{S_1, S_2, S_3\}$ and $m_{\text{smo}} = \{T_1, T_2, T_3\}$, where $R_1, R_2, R_3$ represents the probability of judging fire, no fire and uncertainty through temperature, $S_1, S_2, S_3$ represents the probability of judging fire, no fire and uncertainty through CO$_2$, and $T_1, T_2, T_3$ represents the probability of judging fire, no fire and uncertainty through smoke sensor. Then the Dempster combination rule of multiple sensor fusion under the identification framework is as follows

$$m(F) = \left\{ \begin{array}{l} \sum_{R_i \cap S_j \cap T_k \in F} \frac{m_{\text{temp}}(R_i)m_{\text{CO}}(S_j)m_{\text{smo}}(T_k)}{1-k}, \forall F \subset \Theta, F \neq \emptyset \\ 0, F = \emptyset \end{array} \right. \quad (1)$$

Where

$$k = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{m_{\text{temp}}(R_i)m_{\text{CO}}(S_j)m_{\text{smo}}(T_k)}{R_i \cap S_j \cap T_k = \emptyset, i,j,k, i=1} \quad (2)$$

where $k$ is called collision probability, which indicates the degree of collision between the arguments of each sensor, which satisfies $k < 1$.

### 3.2 Improved D-S evidence Algorithm

The classical D-S evidence theory is beneficial to the distributed implementation of information fusion system when there is no conflict between the evidences of various sensors. But it is easy to get wrong judgment when the sensor evidence conflicts. Therefore, the conflicting sensor evidence must be corrected to reduce its impact on criterion $F$. In this paper, the conflict evidence is judged according to the standard deviation of each sensor evidence, the conflict evidence is modified and the identification frame set is expanded. Finally the Dempster rule is used for re-fusion calculation.

In this paper, the occurrence of fire is divided into three cases: no fire, uncertainty and fire. In these three cases, the standard deviation of the three sensors can be expressed as

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n} [m_i(F_j) - \bar{m}(F_j)]^2}{n-1}} \quad j = 1, 2, 3 \quad (3)$$

where $\bar{m}(F_j)$ represents the mean value of the probability distribution of each sensor when case $F_j$ occurs, and $n$ represents the length of the identification frame set. Standard deviation reflects the degree of dispersion of each sensor data, and also reflects the degree of conflict between sensor evidences. Then the weight of each sensor evidence to the criterion $F$ can be redefined as

$$w_j = \frac{\sigma_j}{\sum_{i=1}^{3} \sigma_i} \quad j = 1, 2, 3 \quad (4)$$
According to the principle of equality of sensor evidence, the average weight \( \overline{w} \) assigned to each sensor is equal. If \( w_j \geq \overline{w} \) is satisfied, the sensor evidence is considered as conflict evidence and needs to be corrected. If \( w_j < \overline{w} \) is satisfied, the sensor evidence is considered as non conflict evidence, and the data remains unchanged. After the conflict evidence is identified, it is corrected according to the following formula:

\[
m'_j(F) = \frac{\sigma_{\min} m_j(F)}{\sigma_j} \quad F \subseteq \Theta
\]

Obviously, since the probability distribution in the conflict evidence is modified, and the cumulative sum of its probability distribution satisfies \( \sum_{i=1}^{3} m'_i(F) < 1 \), an additional item is added to the identification frame set \( \Theta \), and its probability distribution value can be expressed as

\[
m'_j(\Theta) = 1 + \frac{\sigma_{\min} (m_j(\Theta) - 1)}{\sigma_j}
\]

The new additional item of non conflict evidence is zero. The Dempster combination rule is used to re-analyze the corrected sensor evidence and other sensor evidence to identify the fire occurrence.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental site is located on the semi-enclosed workshop with a size of 5m×4m×3.5m. The combustibles are waste wood, and placed on one side of the workshop. The monitoring node is placed on the top of the middle of the room. The ambient temperature is 32°C. DS18B20 temperature sensor is used for temperature detection, TGS-4160 electrochemical sensor is used for CO\(_2\) concentration detection, and MQ-2 gas sensor is used for smoke detection. The CPU of the hardware system is stm32f103c8t6. The algorithm proposed in this paper is implemented on keil uVision 5 software programming platform. The detection node is shown in Figure 2. Each sensor measures data every ten seconds, and the duration of each experiment is five minutes. After several experiments, the average value of the experimental data is taken as the final result. The experimental results are shown in Figure 3-5.
As the combustor material is waste wood, the flame is small and the smoke is large during the initial combustion, but the temperature rises slowly. After 120 seconds, the smoke sensor reaches the alarm threshold, and about 180 seconds, CO₂ reaches the alarm threshold. After the CO₂ reaches 6000ppm, due to the lack of indoor oxygen, some CO is generated by combustion, and the growth of CO₂ concentration slows down. Due to the influence of the environment, the temperature reaches the alarm threshold at 250 seconds. It can be seen that the change curves of the three sensors are not consistent. According to the characteristics of each sensor, set the thresholds of each sensor in different states, as shown in Table 1

<table>
<thead>
<tr>
<th>sensor</th>
<th>No fire threshold</th>
<th>Uncertainty threshold</th>
<th>Fire threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke</td>
<td>0-500ppm</td>
<td>500-1500ppm</td>
<td>1500-5000ppm</td>
</tr>
<tr>
<td>CO₂</td>
<td>0-1500ppm</td>
<td>1500-3000ppm</td>
<td>3000-9000ppm</td>
</tr>
<tr>
<td>temperature</td>
<td>0-40°</td>
<td>40°-55°</td>
<td>55°-80°</td>
</tr>
</tbody>
</table>

According to the measured sensor data and the set threshold, the fire probability distribution of each sensor evidence at 200 seconds is obtained. Information fusion is carried out according to the traditional D-S evidence theory, and the results are shown in Table 2

<table>
<thead>
<tr>
<th>sensor</th>
<th>No fire probability</th>
<th>Uncertain probability</th>
<th>Fire probability</th>
<th>Fire identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke</td>
<td>m(F₁) 0</td>
<td>m(F₂) 0.1</td>
<td>m(F₃) 0.9</td>
<td>Fire</td>
</tr>
<tr>
<td>Temperature</td>
<td>m₂ 0.68</td>
<td>m₂ 0.3</td>
<td>m₃ 0.02</td>
<td>No fire</td>
</tr>
<tr>
<td>CO₂</td>
<td>m₃ 0</td>
<td>m₃ 0.15</td>
<td>m₃ 0.85</td>
<td>Fire</td>
</tr>
<tr>
<td>Fused</td>
<td>m₄ 0.66</td>
<td>m₄ 0.32</td>
<td>m₄ 0.02</td>
<td>No fire</td>
</tr>
</tbody>
</table>

It can be seen from table 2 that in the development stage of the fire, the data of the temperature sensor indicates that there is no fire, while the data of the smoke sensor and the CO₂ sensor indicate that there is a fire. Therefore, it is difficult for a single sensor to accurately judge whether a fire has occurred, which has great inaccuracy. When the sensor evidence conflicts, the traditional D-S evidence theory is used to get a conclusion that is inconsistent with the facts. The method proposed in this paper is used for information fusion of sensor data, and the results are shown in Table 3

<table>
<thead>
<tr>
<th>sensor</th>
<th>No fire probability</th>
<th>Uncertain probability</th>
<th>Fire probability</th>
<th>New uncertainty</th>
<th>Fire identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoke</td>
<td>m₁ 0</td>
<td>m₁ 0.1</td>
<td>m₁ 0.9</td>
<td>0</td>
<td>Fire</td>
</tr>
<tr>
<td>Temperature</td>
<td>m₂ 0.32</td>
<td>m₂ 0.14</td>
<td>m₂ 0.01</td>
<td>0.53 uncertainty</td>
<td>Fire</td>
</tr>
<tr>
<td>CO₂</td>
<td>m₃ 0</td>
<td>m₃ 0.15</td>
<td>m₃ 0.85</td>
<td>0</td>
<td>Fire</td>
</tr>
<tr>
<td>Fused</td>
<td>m₄ 0</td>
<td>m₄ 0.12</td>
<td>m₄ 0.88</td>
<td>0</td>
<td>Fire</td>
</tr>
</tbody>
</table>

The method proposed in this paper is used to process the measurement results, and it is determined that the temperature sensor is a conflict evidence item. After revising the conflict evidence item, the probability of no fire is reduced from 0.68 to 0.32, which greatly reduces the impact of conflicting evidence items on the criterion. The fusion result is consistent with the actual situation, which overcomes the conclusion conflict caused by conflicting evidence and greatly improves the accuracy of judgment.

5. CONCLUSION

In this paper, a fire detection method based on multi-sensor information fusion is proposed. In this method, the standard deviation is used to identify the conflict evidence, and the conflict evidence is corrected by adding the uncertainty item, which effectively avoids the misjudgment caused by the evidence conflict. Compared with the traditional D-S evidence theory, it can more comprehensively perceive the fire state, detect the fire more quickly and accurately, reduce the false alarm rate, and improve the reliability of the system. At the same time, since the system uses Lora wireless Ad hoc network to transmit data, the system can arbitrarily expand the number and location of monitoring sensors, which can better enhance the monitoring ability of the fire system.
ACKNOWLEDGMENT

This work was supported by the Science and Technology Project of Zhongshan city, China(2019B2080), and Science and Technology Project of Zhongshan Polytechnic, China(2018G01).

REFERENCES

Logit-based stock prediction network
Ling Dong\textsuperscript{a}, Mei Hong\textsuperscript{a\*}, Maojun Huang\textsuperscript{a}, Xue Dong\textsuperscript{b}
\hspace{1em}\textsuperscript{a}Shandong Technology and Business University, Yantai, 264003, China;
\textsuperscript{b}Zaozhuang University, Zaozhuang, 277000, China;
\textsuperscript{\*} Corresponding author: hmccec@126.com

ABSTRACT

Stock investors have been making accurate predictions about the stock market in search of maximum profits. However, the stock market has a high degree of uncertainty, which makes it difficult to predict the development trend of the stock market. Existing stock prediction models generally improve the accuracy by changing the network structure and lack in-depth research on abnormal stock data. To solve this problem, we propose a logit-based stock prediction network LogNet, which uses the correctly predicted logits to measure the reliability of stock data, then calculate the confidence interval of the stock data, and use the credible data to make stock predictions. In addition, the model uses the theory of the Extremely Randomized Trees (ExtraTrees) theory to select the historical price data features of stocks. Experimental results show that LogNet has state-of-the-art performance on Twitter data and historical price datasets.

Keywords: logits, ExtraTrees, stock prediction, stock data

1. INTRODUCTION

The stock market forecast has always attracted many researchers and investors as a hot issue, but the stock market is a highly complex system, and the stock trend is affected by many factors such as technology, news, and public opinion\textsuperscript{(1)}. The high volatility and uncertainty of the stock market make forecasting stocks difficult. Most of the existing stock forecasting models cannot provide accurate forecasts. Even so, there are still a large number of researchers who have conducted in-depth research on it and have proposed a lot of methods to achieve this goal. Traditional stock forecasting methods are mainly based on time series models to study the historical price data of stocks\textsuperscript{(2)}. It is believed that all stock-related information is reflected in the stock price, so it is possible to predict stock market conditions using only stock price analysis. However, forecasting using only historical prices is inaccurate. With the development of Natural Language Processing (NLP) technology, people began to add social media and news to the influencing factors of stock prediction. However, the quality of historical price data and social media data related to the stock market is relatively low, especially since news comments are a large part of rumors with low credibility, which will have a negative impact on the validity of stock prediction models. Hu et al.\textsuperscript{(3)} processed the noisy and chaotic social media comment data through a two-layer attention mechanism, used the processed data to predict the trend of the stock market, and achieved good prediction results. However, the researchers only processed social media data and did not consider the factor of historical stock prices. Facing this challenge, we propose the LogNet stock prediction network with the following main contributions:

(1) To solve the redundancy problem in the price features and improve the accuracy of the prediction model, we use the ExtraTrees algorithm to process the price features of the stock and select the features that are highly correlated with the stock market to simplify the model.

(2) To solve the problem of low-quality data contained in the stock data, a method for determining reliable data based on logit is proposed, using the correctly predicted logit to measure the reliability of the data. Then calculating the confidence interval of the high-quality data, predict stock trends using data with confidence intervals.

(3) Through experiments, it is found that compared with existing stock prediction methods, the LogNet method significantly outperforms other models, indicating that the proposed method is more effective and practical.

The rest of this paper is organized as follows: Section 2 summarizes the existing work related to our method. Section 3 describes our proposed model in detail. Section 4 demonstrates the effectiveness of the LogNet model by comparing it with some baseline models. Finally, we present conclusions and future work in Section 5.
2. RELATED WORK

Stock forecasts are usually analyzed from fundamentals, news, and technicals. Fundamental analysis refers to judging the future trend of the market according to the macroeconomic development law of a country or region. It is the basic method affecting the stock market trend. This method is suitable for long time horizons, and the usability of short-term stock forecasts is not high. News analysis refers to policies, regulations, and newly released hotspots related to the stock market. It is an important factor that triggers short-term fluctuations. Technical analysis is the regular development trend of the stock market under the conditions that the above two aspects are determined or stable. The technical analysis mainly deals with historical data such as price and trading volume in time series. The researchers found that machine learning models can use algorithms to continuously refine the model to improve the performance of stock predictions. Nelson\textsuperscript{[6]} used a long short-term memory network (LSTM) to predict the future trend of stock prices and proved that the application of machine learning algorithms to stock market forecasts produced satisfactory results.

However, the prediction results of machine learning models are not stable due to their simple structure. Deep learning methods have made great strides in improving the accuracy of stock trend forecasting, etc. With the rapid development of Internet technology, public speech on social media is used by researchers to predict future stock market trends. In addition, the emotions expressed by media remarks are also an important factor affecting stock trends. Dong Li\textsuperscript{[5]} et al. performed sentiment analysis on the text information obtained from social media, used the support vector machine (SVM) model to predict the stock market, and obtained good prediction results. Tetlock\textsuperscript{[6]} explored the relationship between text sentiment and stock market volatility and concluded that pessimistic sentiment predicts the decline of the stock, and positive sentiment predicts the rise of the stock. In the field of stock forecasting research, some studies combine social media text information with historical stock prices to predict stock movements. Xu et al.\textsuperscript{[7]} used articles on Twitter to combine historical price data to predict the trend, and the experimental results showed that the prediction results of this method were better than that of stock prices or tweets alone. However, this research based on social media and stock prices still has room for improvement in terms of data availability. When the number of data sets is large enough, the value of the data should satisfy a fixed probability, and the stock data belonging to the same trend should have a similar logit\textsuperscript{[8]}. Based on this, we eliminate the low-quality data in the stock by judging the abnormal logit output, improving the credibility of the stock data, and reducing errors. In addition, selecting features in historical price data is also a major problem for researchers to study the stock market. Extremely Randomized Trees (ExtraTrees)\textsuperscript{[9]} is an ensemble method that can aggregate multiple random decision trees to fit data. The algorithm builds an ensemble of decision trees through a top-down process. The main difference between it and other decision trees is that it splits nodes by choosing the left and right branches completely randomly, and it uses all training samples to build the decision tree. Compared to the random forest, ExtraTrees will perform better, because, in the random test set, some features still have a high degree of discrimination, indicating that this feature is indeed important.

In this paper, we propose a new stock prediction network LogNet, which starts from the logit of stock-related data, first calculates the expectation and variance of the logit generated by the correctly predicted data and then calculates the credible interval. Only save the data within the confidence interval. In addition, we use the decision tree algorithm to perform feature selection on the historical price data of stocks, eliminate duplicate data features, and obtain efficient stock data. The network achieves effective and practical stock forecasting results by using high-confidence fused data for forecasting.

3. LOGNET MODEL

We combine stock-related historical price data and social media commentary data to forecast stock trends. Considering that stock information has time series characteristics, we use Bi-directional Long-Short Term Memory (Bi-LSTM) to learn the price data and text data of stocks respectively. Then the data containing time series information is fused, and the data in the period of $D$ is selected for prediction. Since we use Long-Short Term Memory (LSTM) as the classifier for stock prediction, we will input the classifier before, and capture the time series of the fusion data again. Thereby effectively obtaining the future trend of the stock forecast. We present the basic framework of our stock forecasting model in Figure 1.
3.1 ExtraTrees select price features

Different from the traditional decision tree, ExtraTrees selects the optimal eigenvalue division points, the division points of eigenvalues and thresholds are randomly selected during the splitting process of ExtraTrees. Randomly select $k$ features from the whole sample, and randomly select a split node for each feature, to obtain $k$ classification nodes, then calculate the importance of these $k$ split nodes and then select the one with the highest score as the split node. We use the ExtraTrees algorithm to get the feature importance comparison of stock prices in Figure 2.

![Figure 1. Stock prediction basic network](image1)

According to Figure 2, we can see that the features of high price and percentage change are more important than other price features. The three features of volume, closing price, and low price are very close in importance, and the opening price has the lowest importance score. So we choose the high price and percentage change features to predict stocks.

3.2 Logit-based data processing

The logit vectors output by the samples belonging to the same category through the neural network should have similar characteristics. If the logit value of the predicted class output is very different from the logit value of the known samples, the prediction effect is not ideal. When there is enough data, the logit value satisfies a certain probability distribution. That is to say, according to the logit of the correctly predicted data output, to obtain its expectation and variance, we can obtain the confidence interval of the data, using the Forecasting data will improve the forecasting effect.

Step:

(1) Select the data that predicts the stock trend correctly during training;
(2) Calculate the expectation and variance of the logit of the correctly predicted data output;
(3) Set the credible interval of normal fluctuation data as
\[ m\mu + n\sigma^2, m\mu - n\sigma^2 \]
\( m,n \) are the parameter values for the normal fluctuation range;
(4) Save data within the confidence interval.

4. EXPERIMENT AND RESULT ANALYSIS

4.1 Dataset description
We use a dataset\(^{(10)}\) containing Twitter comments and historical price data for trend prediction of stocks. The dataset includes Twitter comments for 88 stocks and price data obtained through Yahoo Finance. The time frame is from January 1, 2014, to January 1, 2016.

4.2 Comparison with the baseline model
We use the following five baseline models for performance comparisons with our model:

Multi-layer Perceptron (MLP): We use the MLP model to forecast stock trends.

Bi-LSTM: Bi-LSTM consists of two LSTMs in different directions, taking time series vectors as input.

Hybrid Attention Networks (HAN)\(^{(3)}\): Use news layer and time layer attention mechanisms to obtain effective information in news, and predict stock trends through the sequence of stock news.

LogitsNet: Compared with BasicNet, low-quality content of social media is processed based on logits.

<table>
<thead>
<tr>
<th>Baseline Methods</th>
<th>accuracy</th>
<th>precision</th>
<th>recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>53.39</td>
<td>0.73</td>
<td>0.53</td>
<td>0.38</td>
</tr>
<tr>
<td>LSTM</td>
<td>53.19</td>
<td>0.68</td>
<td>0.53</td>
<td>0.38</td>
</tr>
<tr>
<td>HAN(Hu et al.,2018)</td>
<td>58.34</td>
<td>0.71</td>
<td>0.58</td>
<td>0.53</td>
</tr>
<tr>
<td>LogNet</td>
<td>62.78</td>
<td>0.61</td>
<td>0.63</td>
<td>0.58</td>
</tr>
</tbody>
</table>

4.3 Analysis of experimental results
In the field of stock forecasting, even small performance improvements can lead to huge profits. To evaluate the performance of each prediction method, we choose accuracy, precision, recall, and F1-score as evaluation metrics, and the comparison results are shown in Table 1. According to Table 1, we can see that if the historical price data and media data of stocks are not processed, the accuracy of stock forecasting using raw data is not ideal. Compared with BasicNet, LogitsNet proves the effectiveness of processing stock text data, and BasicNet shows the importance of selecting stock price features. Our LogNet combines LogitsNet and BasicNet, and it can be seen from Table 1 that the performance is better than other existing models.

5. CONCLUSIONS AND FUTURE WORK
This paper proposes a new stock forecasting method that separately processes historical stock price data and related news media data to improve usability. First, the ExtraTrees algorithm is used for feature selection, and price features with high importance to stock forecasting are selected. Secondly, the mean and variance of the correct forecast data logit are used to calculate the credible interval of the data, and the low-quality data that is not within the credible interval is eliminated. The processed price data and text data are then fused and a discriminator is used to predict stock trends. Experiments show that our model significantly outperforms other existing stock prediction methods. Our next work will test...
improving stock prediction performance by adding uncertainty to the stock prediction network and removing high uncertainty data information.

REFERENCES

Application Analysis of Machine Learning in Cyberspace Security Research

*Wenzhuo Du1  Gerun Wang2  Xuanbo Ma3  Xintian Yu4
People’s Public Security University of China
*3116552079@qq.com

Abstract

With the rapid development of society and the continuous change of the times, my country’s Internet technology has achieved great development space. Cloud computing, Internet of Things, etc. have formed the current era of big data. At present, there are a lot of network access points, and the networked devices among them will generate a lot of data, so there are higher requirements for the security of the cyberspace, and it is also a golden opportunity. Traditional cyberspace security research is relatively inefficient in data processing, machine learning has strong adaptability, and strong learning capabilities also ensure the security of cyberspace. Under such a social background, it is necessary to give full play to the role of machine learning. After applying machine learning, it can effectively solve the problems existing in cyberspace security research. Based on this, this paper discusses the application of machine learning in cyberspace security research.

Keywords: machine learning; cyberspace; security research

1. Introduction

Since entering the 21st century, my country has initially entered the information age, including the development of technologies such as the Internet of Things, cloud computing, and big data, which has brought great development opportunities to my country's cyberspace. But in the face of opportunities, there are also various difficulties. While the hardware and software are generated, the efficiency of data processing can be improved. As a virtual space, cyberspace has been attacked a lot in recent years, which has caused certain damage to the security of network users, and even caused serious economic losses to users. After incomplete statistics, the economic losses caused by digital crimes in my country each year have reached tens of billions of yuan [1]. Therefore, the security of cyberspace is of vital significance to social development and economic development, and needs to be paid attention to. Under such a social background, traditional security research methods are difficult to meet the needs of society. For example, manual repair by network experts is difficult to fundamentally solve the problem, and the efficiency is extremely low. The emergence of machine learning has brought a guarantee for the security of cyberspace. It can perform speech recognition and natural language processing. Whether it is in classification or decision-making, it can bring greater security to cyberspace.

2. Machine Learning Overview

Machine learning is the core technology of artificial intelligence. It mainly simulates human behavior. In the learning method, the computer can learn more knowledge, and after becoming more intelligent, it can optimize its own internal organization to help other people. Gain the ability to continuously learn. It is the core of the neural network. The formula of sigmoid (Logistic-Sigmoid, Tanh-Sigmoid) is shown in formula 1 and formula 2.

\[
\sigma(z)=\frac{1}{1+e^{-z}}=1+\tanh(z/2)^2 \quad (1)
\]

\[
\tanh(z)=\frac{e^z-e^{-z}}{e^z+e^{-z}}=\frac{e^{2z}-1}{e^{2z}+1} \quad (2)
\]

3. The application process of machine learning in cyberspace security research

For machine learning, the performance of the system is generally improved through a set of data. In the process of calculation, machine learning can classify according to the attributes of the data. The flow chart for machine learning in cyberspace security research is shown in Figure 1. It includes abstract security issues, data collection, data preprocessing, feature extraction, model building, model verification, and effect evaluation [2]. In this process, each stage is
complementary to each other, without the support of other links, it is difficult to exist independently.

Fig 1. Application process of machine learning in cyberspace security research

4. The role of machine learning in cyberspace security research

4.1 Detecting Gateway Abnormalities

For the Internet, the gateway protocol can ensure the normal operation of the Internet, and is also one of the core contents of the Internet. It mainly affects the communication quality of the router. In the specific practice process, the network protocol is not complete in terms of the authentication mechanism, so the ability to identify routing information is relatively weak. If it is attacked by the network, it is easy to cause system paralysis. [3]. After applying machine learning technology, the gateway protocol will track and monitor during operation, analyze routing data, and use ELM algorithm to identify abnormal data. The training model is shown in Equation 3.

$$t_i = \sum_{l=1}^{L} \beta_l g_l(w_ix + b_l), \quad l=1, \ldots, P$$

4.2 Domain name security monitoring

Generally speaking, the domain name system is one of the most attacked systems. Because the frequency of attacks is relatively high, the system cannot be used normally in severe cases. The traditional solution is to let the system record the source of the attack. If a domain name appears The abnormality is to detect the source of the attack, so as to find the corresponding security risk. There are many problems with this detection method. If the attacker blocks the record table, the system cannot identify it, so the source of the attack is listed as a security source. However, after the application of machine learning technology, further analysis will be carried out according to the data characteristics of the attack source, and the extracted characteristics will be compared to quickly identify security risks [4].

4.3 Network security detection

Network security detection is to organize and collect the data of different access objects or network access points. In the process of analyzing the data, it is necessary to find security risks through manual repair. Relevant research shows that after the application of machine learning technology, it can effectively improve the efficiency of security prevention, and further detect the network, which has a positive effect on the security of network security.

5. The application of machine learning in cyberspace security research

Cyberspace mainly refers to systems with computing capabilities, such as mobile secondary schools, computers, etc. The application of machine learning in cyberspace security is discussed from the three levels of chip, hardware and software.
5.1 Chip security

For chip security, because its own distribution is more multi-dimensional, the steps are very complicated. Chips are closely related to hardware devices. If a security problem occurs in a certain link, it may have a greater impact on the operation of the system. Such as chip quality, Trojan hardware, etc[5]. Relevant studies have shown that the use of machine learning technology can help chips solve security problems, and can quickly identify inferior chips and hardware Trojans by performing variable signal analysis, image recognition, etc. on the chip.

Among the inferior chips, many chips with low specifications and low quality have flowed into the market. Due to the particularity of the chip, it is difficult to see the quality with the naked eye, and in the traditional detection method, the cost is high, and it needs to undergo structural testing and functional testing. In fact, there are differences in parameters between inferior chips and original chips, which are manifested in side channels. Secondly, inferior chips can also be seen on the shape of the chip, such as color, traces and so on. In the detection of hardware Trojans, malicious functions are generally implanted on the chip implantation, and the hardware Trojans are divided into physical Trojans and activated Trojans. When the chip is implanted with a Trojan horse, the heat and path delay will be affected. Therefore, the parameters of the side channel of the chip are analyzed, and the parameters of the chip are compared with those of the ordinary chip to know whether the chip has been implanted with a virus. PUF (Physical Unclonable Function, PUF) [6] attack is a circuit that generates a unique reward-response pair on the circuit according to the difference in the manufacturing process of the chip, which is also used as the identification of the chip. After the identification, the chip cannot be processed. copy.

5.2 System hardware security

System hardware is also a common security concern in cyberspace. For example, most of the traditional authentication methods are through MAC address authentication. This authentication method is difficult to guarantee the authenticity of the user, and many attackers can forge the identity. In addition, in the physical environment of cyberspace, information leakage is easy to occur. Machine learning techniques can prevent similar situations from happening. First of all, in identity authentication, the identity of the user can be quickly extracted, and the authenticity of the user can be guaranteed by means of fingerprint recognition. Fingerprints are unique. After combining signal analysis technology with machine learning technology, identity authentication can be performed faster and more accurately. For now, most of the research is carried out through transient signal, spectral response, sensor response, modulated signal. The fingerprint recognition process is shown in Figure 2.

![Fig 2. Machine Learning Authentication Process](image)

5.3 System software security

Different from the device authentication in the system hardware, the authentication of the system software studies the authentication between the user and the hardware device or the user and the system. After applying machine learning, machine learning can be used in traditional user authentication methods to further strengthen the security of cyberspace. The machine learning cyberspace security research on system software security is shown in Table 1. Because machine learning is more complicated for user identity authentication methods, an identity authentication mechanism can be initially constructed. However, there are still low authentication accuracy, involving user privacy and so on.

<table>
<thead>
<tr>
<th>safe question</th>
<th>security features</th>
<th>Machine Learning Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>user</td>
<td>CAPTCHA color</td>
<td>SVM</td>
</tr>
<tr>
<td>identity</td>
<td>this feature</td>
<td>-</td>
</tr>
<tr>
<td>Certification</td>
<td>computer vision</td>
<td>K-Means algorithm</td>
</tr>
<tr>
<td>attack</td>
<td>side channel signal</td>
<td>K neighbor, SVM, Neural Networks</td>
</tr>
<tr>
<td>user ID</td>
<td>touch screen features</td>
<td>One-Class Learning Algorithms</td>
</tr>
<tr>
<td>Certification</td>
<td>keystroke behavior</td>
<td>SVM, Naive Bayes</td>
</tr>
<tr>
<td>design</td>
<td></td>
<td>K neighbor</td>
</tr>
</tbody>
</table>

Table 1. Application of Machine Learning to User Authentication Technology
In addition, we must also pay attention to the virtualization security in the cyberspace, which is generally widely used in cloud computing. At present, the virtualization security problem in the cyberspace is mainly the security of virtual machines, and side-channel attacks are also a threat in the isolation of virtual machines. After applying machine learning techniques, the cache architecture can be shared, which is a common channel for cross-VM attacks. Relevant research shows that [7], in the case of cross-virtual machine variable signal attack, the key of the virtual machine open source encryption library can be obtained through the SVM algorithm with the help of the symmetric multiprocessing environment and the cache architecture as the attack channel.

6. Conclusion

To sum up, the Internet has entered people's lives and has brought a great impact on people's life and work. While improving work efficiency, it also faces the problem of network security. As the core technology of artificial intelligence, machine learning can play a security role in the current cyberspace. At present, there is a lot of data on the network, and traditional data processing methods are difficult to meet the needs of society. Machine learning can improve the efficiency of network operation to a certain extent, and it has also caused widespread heated discussions in the society. It is believed that in the near future, the application of machine learning technology will be more extensive and promote the harmonious development of society.

References

Multi-hop question answering for SRLGRN augmented by textual relationship modelling

Xuesong Zhang\textsuperscript{1}a, Guanjun Li\textsuperscript{2}b, Dawei Zhang\textsuperscript{2}c, Zhao Lv\textsuperscript{1}d, Jianhua Tao\textsuperscript{2}c*

\textsuperscript{1}College of Computer Science and Technology, Anhui University, Hefei 230000, China, \textsuperscript{2}Laboratory of Pattern Recognition (Institute of Automation, Chinese Academy of Sciences), Beijing 100000, China

\textsuperscript{a}xuesong.zhang@nlpr.ia.ac.cn, \textsuperscript{b}guanjun.li@nlpr.ia.ac.cn, \textsuperscript{c}dawei.zhang@nlpr.ia.ac.cn, \textsuperscript{d}kjiz@163.com, \textsuperscript{e}jhtao@nlpr.ia.ac.cn

Abstract

Multi-hop question answering aims to predict answers to questions and generate supporting facts for answers by reasoning over the content of multiple documents. The recently proposed Semantic Role Labeling Graph Reasoning Network (SRLGRN) has achieved excellent performance on multi-hop QA tasks. However, SRLGRN is lacking in modelling the textual relationships, which are import cues for reasoning. To this end, this paper proposes an enhanced SRLGRN multi-hop question answering approach by modelling textual relationships at different granularity (document relationships and sentence relationships). By modelling document relationships, a novel document filter based on document relationship threshold is designed for SRLGRN to dynamically select documents relevant to the question from multiple documents. By modelling sentence relationships, a sentence relationship-aware answer type prediction module is added to SRLGRN, which models sentences in documents as sentence graphs and then uses graph convolution network to predict answer type. The obtained answer type further guides the answer reasoning module of SRLGRN to obtain question answer with supporting facts. The experimental results show that the proposed scheme outperforms SRLGRN in terms of answer prediction and supporting fact prediction, with a 2\% improvement in answer F1 metrics and a 3.1\% improvement in joint F1 performance.

Keywords-component: Multi-hop question answering; document filter; sentence graphs; graph convolution network

1. Introduction

Question Answering (QA) is a popular topic in natural language processing\textsuperscript{[1]}. With the flourishing of deep learning, scholars have proposed multi-hop QA models\textsuperscript{[2]} and designed several multi-hop QA datasets specifically for evaluating multi-hop reasoning capabilities, such as WIKIHOP\textsuperscript{[3]} and HotpotQA\textsuperscript{[4]}. The QA tasks corresponding to these datasets are very challenging as they require QA models that can perform multi-hop reasoning to obtain answers to questions under the distractor of multiple documents and text noise. Especially in the HotpotQA dataset, the QA model needs to provide not only the prediction of the answer, but also the supporting facts of the answer. Fig.1 shows an example from the HotpotQA dataset. One question corresponds to N documents, in which documents containing supporting facts (blue text) are relevant documents, and the rest are distractor documents. The HotpotQA dataset asks scholars to predict the answer to a question and the supporting facts for answering that answer in relevant documents.
Current research on multi-hop QA models is divided into two main types: The first decomposes multi-hop questions into multiple simple sub-questions to perform multi-document QA tasks\[^{5-8}\]. The other is the use of Graph Neural Network (GNN)\[^{10-12}\] to implement multi-hop reasoning across documents. GNN transmits multi-hop information through graph propagation with its inherent message transmission mechanism, and has great potential in predicting the supporting facts and answers of multi-hop questions\[^{13}\]. Among the GNN-based multi-hop QA models\[^{14,15}\], the Semantic Role Labeling Graph Reasoning Network (SRLGRN), recently proposed by Zhang et al \[^{16}\], has achieved excellent multi-hop QA performance. However, SRLGRN fails to adequately consider different granularity of textual relationship modelling, such as document relationships and sentence relationships. In terms of document relationships, the SRLGRN filtering document module treats each document individually, without paying attention to the documents interactions and relationships, which are essential for downstream multi-hop reasoning tasks. In terms of sentence relationships, SRLGRN simply input the sentences into the model in order to predict the answer type but ignore the sentence-sentence relationship information. However, sentence-sentence relationship information is important for machine understanding of documents.

In order to solve the above problems of SRLGRN, this paper proposes a textual relationship aware SRLGRN, which takes the document relationships and sentence relationships into consideration. Specifically, to modelling the document relationships, we propose a two-by-two threshold-based learning ranking method, which converts the task of filtering documents from a classification problem to a ranking problem and enhances the interaction between documents. Besides, the proposed flexible threshold parameter allows the document filter to flexibly select documents. In terms of sentence relationship, a sentence relationship-aware answer type prediction module is proposed, where sentences in a document are modelled as a sentence graph. Multi-hop reasoning is performed on the sentence graph by Graph Convolution Network (GCN), and the weighted sum of sentence nodes is obtained. Then, a two-layer Multi-Layer Perceptron (MLP) with three-dimensional output is used to predict the answer type.

2. Background: the SRLGRN model

2.1 Filtering documents

The process of filtering documents in SRLGRN first feeds each document into the Bidirectional Encoder Representation from Transformers(BERT) model as ”[CLS] + question + [SEP] + document + [SEP]”. The [CLS] output of BERT is extracted as a summary vector for each question/document pair. Project the summary vector into a singular dimension and the binary cross-entropy loss is as follows:

\[
\text{loss} = - \sum_{i=0}^{n} t_i \log P(D_i) + (1 - t_i) \log (1 - P(D_i))
\]

(1)
Where $t_i$ is the label of $D_i$, $n$ is the number of documents, and $P(D_i)$ is the probability that document $i$ is under the label $t_i$. This approach deals with each document individually and fails to take full account of the interactions between documents.

To address this issue, in Section 3.1, we use a two-by-two threshold-based learning ranking method to filter documents.

### 2.2 Predicting answers

The answer type prediction of SRLGRN model takes question $Q$ and filtered document content $C$ as inputs:

$$BERT_{[CLS]} = BERT([CLS]; Q; [SEP]; C)$$  \hspace{1cm} (2)

$BERT_{[CLS]}$ is a representation of questions and relevant documents. Then $BERT$ is used to classify the answer type. Answer type are classified into 3 categories "yes", "no" and "span". Answer type can be calculated as:

$$y_{type} = MLP_{type}(BERT_{[CLS]})$$  \hspace{1cm} (3)

For answer prediction, SRLGRN additionally constructs a parameter-predicate subgraph $Sub$. The parameter-predicate subgraph embedding $G_{Sub}$ is obtained by introducing GCN. The nodes of the parameter-predicate subgraph are used to help identify the answers to the questions. SRLGRN compute the logit of every token to find the start position and end position for answer $y_{ans}$. The logit is calculated using $BERT$ as the input given to two fully connected layers. The input token representation is the concatenation of $BERT$ token representation $BERT_{tok}$ and graph embedding $G_{Sub}$.

One important reason for the poor precision of the SRLGRN model in predicting the answer type (see Table 3) is that SRLGRN only feeds the document content into the model when predicting the answer type (see Eq.(3)) and fails to adequately model the sentence-sentence relationship information in the document. To address this issue, in Section 3.2, we use a sentence-embedding-based GCN model that models the relationship information of sentences in a document to replace the answer type prediction module of SRLGRN in Eq.(3).

### 3. SRLGRN enhanced by textual relationship modelling

The overall framework of the proposed multi-hop QA model is shown in Fig.2. We model document relationships through a two-by-two threshold-based learning ranking method that enhances the interaction between documents and allows the document filtering module to dynamically select documents that are relevant to the question. We also model sentence relationships by constructing a sentence graph from the sentences in the documents, which allows the answer type module to predict answer type using GCN.


![Fig.3 Document filtering framework](image-url)
To address the shortcomings of SRLGRN filtering documents mentioned in Section 2.1, we design a threshold-based two-by-two learning document filter to extract documents that are relevant to the question. Inspired by literature [17], we add a Multi-Head Self-Attention (MHSA) layer on "CLS" to make an internal interaction between documents, and obtain the information between each pair of documents through a Bi-linear layer. This information is sent to the scorer SCORE to get the score of each document, and the document with the score greater than the threshold $\beta$ is regarded as the relevant document. The binary cross entropy loss is as follows:

$$\text{loss} = - \sum_{i=0}^{n-1} \sum_{j=0, j \neq i}^{n-1} l_{ij} \log \left(P(D_i, D_j) + (1 - l_{ij}) \log \left(1 - P(D_i, D_j)\right)\right)$$  \hspace{1cm} (4)$$

Where $n$ denotes the number of documents and $i,j$ denotes the $i,j$th document. $P(D_i, D_j)$ denotes the prediction probability that $D_i$ is more relevant than $D_j$.

In the literature[17], a fixed number of $k$ documents are selected for the follow-up model. However, literature[17] usually sets $k$ to 2, which will result in a poor recall rate of documents and will not capture all documents relevant to the question, thus failing to answer the correct answer. For example, Question 1 in Fig.4 predicts the wrong document "1960 Cleveland Browns season" and fails to accurately predict all the relevant documents for Question 1, resulting in the wrong answer "Cleveland Browns ".Based on the observation that if the Top-3 document is selected, all relevant documents for Question 1 will be predicted, resulting in the correct answer. If the value of $k$ is set to a large value, the recall rate of the document can be improved. However, the precision rate of the document will be reduced, and the distractor documents will be increased, and the wrong answer will be generated. For example, Question 2 in Fig.4 predicts three documents: "Kevin Henke", "RATE project" and "TaskForce Majella". One of these documents, "Taskforce Majella", is a distractor document, which leads to the generation of the incorrect answer "John Baumgardner".

| Question1: | which team was older when it had it's worst season , The Cleveland Browns or the Detroit Lions |
| Documents Score Sorting: | (1)2016 Cleveland Browns season; (2)1960 Cleveland Browns season; (3)2008 Detroit Lions season; (4)Clayton Beemford; (5)2008 Cleveland Browns season; (6)1982 Detroit Lions season; (7)Detroit Lions Television Network; (8)Cleveland Browns relocation controversy; (9)Lew Carpenter; (10)1952 Detroit Lions season |
| Top-k: 2 | Label document: 2016 Cleveland Browns season, 2008 Detroit Lions season |
| Predict document: 2016 Cleveland Browns season, 1960 Cleveland Browns season; | Label answer: Detroit Lions |
| Predict answer: Cleveland Browns |

| Question2: | which member of the research project conducted by the Creation Research Society and the Institute for Creation Research between 1997 and 2005 was Kevin R. Henke most critical of? |
| Documents Score Sorting: | (1)Kevin R. Henke; (2)RATE project; (3)TaskForceMajella; (4)Research; (5)Hendren v. Campbell; (6)Frank Lewis March; (7)Stanford MIPS; (8)Institute for Creation Research; (9)Henry M Morris; (10)Creation Research Society |
| Top-k: 3 | Label document: Kevin R. Henke, RATE project |
| Predict document: Kevin R. Henke, RATE project, TaskForceMajella; | Label answer: Russell Humphreys |
| Predict answer: John Baumgardner |

Fig.4 Example of failure to predict relevant documents

Therefore the Top-$k$ documents cannot be selected fixedly for each piece of data, and the documents relevant to the question are selected dynamically from multiple documents. We consider setting a threshold $\beta$ for flexible selection of documents. First we compare the documents in pairs to obtain the probability $P(D_i, D_j)$ for each pair of documents and obtain the document scores according to Eq.(5).

$$\text{Score}[D_i] = \sum_{j \neq i}^{n} P(D_i, D_j)$$  \hspace{1cm} (5)$$

Where $D$={$D_1,D_2...D_n$} and $n$ is the number of documents.

$$D_t = \{D_{11}, D_{12}, ..., D_{tm}\}$$  \hspace{1cm} (6)$$

$$\text{Score}[D_{tm}] > \beta, m = \{1, 2, ..., M\}$$  \hspace{1cm} (7)$$

We selected the set $D_t$ as the relevant documents, where $D_t$ is a subset of $D$ and each document in the set $D_t$ has a score greater than the threshold $\beta$. 

3.2 Sentence relationship modelling: answer type prediction based on the GCN model

Inspired by the literature [17], we use a GCN model based on sentence embedding to perform answer type prediction. The sentences in the document are first constructed into a sentence graph and then a GCN multi-relational model with a gating mechanism [18][19] is used to pass messages to the nodes in the sentence graph to better exploit the complex relational information. After passing messages on a graph with a predefined number of hops, each node has its final representation $S_{end}$. We use $f_{type}$ to obtain answer type predictions.

$$y_{type} = f_{type} \left( \sum \lambda S_{end} \right)$$  \hspace{1cm} (8)

Where $f_{type}$ is implemented in MLP and the attention weights $\lambda$ are derived from $\lambda = \sigma(S_{node})$. $S_{node}$ is the score of each sentence node. It is obtained by taking the $S_{end}$ of each node as input to the two-layer MLP. The model adopts the multi-task learning method, the joint objective function includes the sum of cross-entropy Loss for the answer type prediction $loss_{type}$ and answer prediction $loss_{ans}$. The loss function is computed as follows:

$$Loss = y_1 loss_{type} + y_2 loss_{ans} \hspace{1cm} (9)$$

$$ = y_1 y_{type} log y_{type} + y_2 y_{ans} log y_{ans}$$

Where $y_1$ and $y_2$ are weighting factors. The type of answer predicted by Eq.(8) is used in Eq.(10) to obtain the final answer. If the answer type predicted by Eq.(8) is "yes" and "no", it will be directly output; if it is "span", it will output $y_{ans}$ (mentioned in Section 2.2).

$$ans = \begin{cases} 
\text{yes}, & \text{if } y_{type} = \text{yes} \\
\text{no}, & \text{if } y_{type} = \text{no} \\
\text{span}, & \text{if } y_{type} = \text{span} 
\end{cases}$$  \hspace{1cm} (10)

4. Experimental procedures and analysis

4.1 Experimental preparation

We used the HotpotQA dataset for the main evaluation of the model, which consisted of two subtasks: answer prediction and support fact prediction. For each subtask, model performance was evaluated using exact matching (EM) and partial matching (F1), and the final performance was measured using the joint score of EM and F1. In the HotpotQA distractor setting, there are 2 relevant documents and 8 distractor documents for each question.

For the value of $\beta$, we set different $\beta$ to compare the precision and recall of relevant documents, so as to select a reasonable $\beta$.

Fig.5 Document recall and precision at different thresholds $\beta$

As shown in Fig.5, when $\beta=0.05$, relatively high precision and recall can be achieved simultaneously (precision rate is 95.4% and recall rate is 97.41%). So we set the threshold beta to 0.05.
Next, we compared the proposed scheme with the following methods: 1) Baseline: the self-contained method in the HotpotQA dataset\(^4\); 2) QFE\(^{20}\) which performed prediction by considering dependencies between supporting facts; 3) DFGN\(^{21}\) which constructed entity graphs from text, in addition to constructed a dynamic fusion layer to find supporting facts; 4) SAE\(^{17}\) which is a pipeline system that first selects relevant documents and then predicts answers and facts using the selected documents; 5) SRLGRN model.

### 4.2 Analysis of experimental results

Table 1 shows the results of the model on the dev set of HotpotQA in the distractor setting, and it can be seen that our model outperforms most of the published results. On correct answer prediction, our model scores 64.6% for EM and 78.1% for F1. Our method achieves an absolute improvement of 20.2% in answer prediction EM and 19.8% in F1 compared to the baseline. Compared to the published SRLGRN model, our model improved by 2.6% on joint EM metrics and 3.1% on joint F1 performance. This confirms that it is beneficial to add textual relationship modelling for SRLGRN.

Next, we compare the proposed scheme with SRLGRN in terms of filtering documents and predicting answer type.

#### 4.2.1 Filtering documents

We compare the proposed two-by-two threshold-based learning document filter with the filtering document module of SAE, SRLGRN. Where SRLGRN-2 in Table 2 indicates that SRLGRN's filtering document module selects two documents each piece of data, SRLGRN-3 indicates that SRLGRN's filtering document module selects three documents each piece of data. The same is true for SAE-2 and SAE-3. According to Table 2, our proposed scheme (β=0.05) scored 95.4% in filtered document precision and 97.41% in recall. Compared to SRLGRN-2, there is an improvement of approximately 12.09% in precision and a decrease of 0.09% in recall. Compared to SAE-2, there was a 1.45% increase in recall and a 0.6% decrease in precision. In summary, it can be seen that our method can provide relatively high precision and recall.

### Table 2 Performance comparison on filter document of the dev set of HotpotQA in the distractor setting

<table>
<thead>
<tr>
<th>Model</th>
<th>Ans(%)</th>
<th>Sup(%)</th>
<th>Joint(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EM F1</td>
<td>EM F1</td>
<td>EM F1</td>
</tr>
<tr>
<td>Baseline(^4)</td>
<td>44.4 58.3</td>
<td>22.0 66.7</td>
<td>11.6 40.9</td>
</tr>
<tr>
<td>QFE(^{24})</td>
<td>53.7 68.7</td>
<td>58.8 84.7</td>
<td>35.4 60.6</td>
</tr>
<tr>
<td>DFGN(^{19})</td>
<td>55.7 69.3</td>
<td>53.1 82.2</td>
<td>33.7 59.9</td>
</tr>
<tr>
<td>SAE(^{17})</td>
<td>61.3 74.8</td>
<td>58.1 85.3</td>
<td>39.9 66.5</td>
</tr>
<tr>
<td>SRLGRN(^{16})</td>
<td>62.6 76.1</td>
<td>57.3 83.8</td>
<td>38.4 65.4</td>
</tr>
<tr>
<td>Proposal presented</td>
<td>64.6 78.1</td>
<td>59.7 85.9</td>
<td>41.0 68.5</td>
</tr>
</tbody>
</table>

#### 4.2.2 Answer type prediction

The HotpotQA dataset can be divided into three categories: "yes", "no" and "span" according to the answer type. As can be seen from Table 3, although both models can accurately predict the "span" type, when the answer type is "yes" or "no", compared with SRLGRN model, our proposed scheme has a relative improvement of more than 11% in terms of F1 value. This demonstrates that modelling sentence-sentence relationships can help to improve the performance of answer type prediction.

### Table 3 Compare by answer type on the HotpotQA data set

<table>
<thead>
<tr>
<th>Model</th>
<th>answer type</th>
<th>precision(%)</th>
<th>recall(%)</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRLGRN(^{10})</td>
<td>yes</td>
<td>76.7</td>
<td>79.1</td>
<td>77.9</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>79.2</td>
<td>76.8</td>
<td>78.0</td>
</tr>
<tr>
<td></td>
<td>span</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Proposal presented</td>
<td>yes</td>
<td>87.3</td>
<td>85.8</td>
<td>86.5</td>
</tr>
</tbody>
</table>
4.2.3 Ablation experiments

Table 4 Results of ablation experiments

<table>
<thead>
<tr>
<th>Model</th>
<th>Ans F1(%)</th>
<th>Sup F1(%)</th>
<th>Joint F1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal presented</td>
<td>78.1</td>
<td>85.9</td>
<td>68.5</td>
</tr>
<tr>
<td>Removal of proposed answer type prediction</td>
<td>77.6</td>
<td>85.9</td>
<td>68.0</td>
</tr>
<tr>
<td>Remove the proposed document filter</td>
<td>76.1</td>
<td>83.8</td>
<td>65.4</td>
</tr>
</tbody>
</table>

According to the experimental results indicated: removing our proposed answer type prediction, Ans F1 decreased by 0.5%, demonstrating the effectiveness of the GCN model based answer type prediction. Removing the threshold-based two-by-two learning document filter, Sup F1 decreased by 2.1% and Joint F1 decreased by 2.6%. The analysis of the ablation experimental results proved the effectiveness of the scheme proposed in this paper.

5. Conclusion

In this paper, we improve the SRLGRN multi-hop QA model by modelling textual relationships at different granularity. In terms of the document relationships, we propose a two-by-two threshold-based learning ranking method to reduce the number of distractor documents. In terms of the sentence relationships, a sentence relation-aware answer type prediction module is added to model the sentences in a document as a sentence graph, which in turn uses a graph convolution network to predict the answer type. Experiments show that the proposed scheme outperforms SRLGRN in terms of answer prediction, supporting fact prediction. In the future, we intend to further improve multi-hop QA performance by incorporating documents into sentence graphs.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (NSFC) (No.61831022, No.U21B2010, No.61901473, No.62101553).

References

Multi-level fast personalized recommendation system
Fang Qiyuan¹a*
1GuangDong Experimental High School, Guangzhou 511400, China
a* 2390003663@qq.com

Abstract

Personalized recommendation systems have high speed requirements, and most of the previous methods rely on powerful hardware to meet the real-time requirements. We algorithmically propose a fast recommender system. Our proposed method abandons the previous global comparison method, but uses a local comparison method, which greatly reduces the computational complexity. The experimental results also show that the effect of the method is not decreased, but improved.

Keywords-component: Multi-level, fast personalized recommendation system

1. Introduction

In the aspect of enterprise marketing, mining information from massive data and set complex data to quantitative data for analysis are important aspects of enterprise digital marketing. Accurate analysis of data can achieve the best business decision.

The core value of personalized recommendation is to provide content and services more accurately to users. With such a connection, the company can better promote its products, users can better enjoy the services brought by the products and finally achieve a win-win result. The core value of personalized recommendation is to establish a connection between the item and the user. With such a connection, the company can better promote its products, and the user can better enjoy the services brought by the products, and finally achieve a win-win result. With the abundance of content and the divergence of user needs, recommender systems will become an essential and fundamental application in the Internet ecosystem. For the product provider, the recommendation system can improve their product operation ability from thousands, tens of thousands, to tens of millions. From this point of view, the consumer's life is faster and more convenient, and the business also makes a lot of full of money, bringing huge commercial profits. However, there are also negative effects. Users' privacy is stolen, which will cause losses if it is used by criminals [1,2,3]. Consumers are passively trapped in the information cocoon room. Some Internet enterprises have their own functional technical teams to build flexible welfare platforms by themselves, while more enterprises will directly find mature flexible welfare providers and directly use ready-made platforms. It is not enough to recommend only through algorithm modeling, how to make informed recommendations more intelligent, this time we need to consider machine learning, through feature mining, behavior analysis, preference learning [4,5], and constantly optimize the recommendation effect and intelligent recommendation. Recommending products is relatively simple, and each recommended product is implemented as a separate service. The new recommendation system is a systematic project, which depends on the organic combination of data, architecture, algorithm, human-computer interaction, and other links. The goal of the new version of the recommendation system is to change "one side of a thousand people" into "one side of a thousand people" through personalized data mining, machine learning, and other technologies, improve user loyalty and user experience, and improve the quality and efficiency of user shopping decisions. Improve the cross-selling ability of the website, shorten the user shopping path, and improve the traffic conversion rate (CVR).

The algorithmic recommendation can directly affect information presentation, search ranking, news popularity, and dissemination effect, resulting in these network platforms becoming relatively independent public opinion "source" and "fermentation pool", and the possibility and risk of public opinion operation are increased. According to Neumann's "spiral of silence" theory, once users believe that a value orientation is widely welcomed, the voices corresponding to it will rise, and the voices opposing it will fall silent. An algorithmic recommendation is easy for users to create the impression that "many people have this value orientation", and the silence of positive value opinions leads to the growth of negative value opinions, thus falling into a vicious circle, resulting in the distortion of the network public opinion field.
2. Related Work

Current graph convolution personalization algorithms and their problems. A more powerful recommendation algorithm, the Convolutional Neural network (CNN), brings more convenient capabilities. Because of its powerful feature representation ability, it has received great attention in the fields such as computer vision and natural language processing [6,7,8]. Text, images, and videos are all data defined on a regular grid, which can be viewed as distributed on one-, two-, and three-dimensional grid support sets [9,10,11,12,13]. Graph convolutional networks have attracted more and more attention by using their powerful feature representation ability to improve learning effects.

In the actual production environment, how to improve the graph convolution operation to adapt to large-scale graphs is also a problem worth exploring. Graph convolution in GCN4RS uses spectral domain convolution based on matrix multiplication, which can effectively mine the information in the graph domain, but it is difficult to perform operations when the memory is limited. The improvement direction that can be considered is to choose the convolution operation based on sampling-aggregation for batch and distributed operations. But despite its power, the performance of GCN tends to drop when running the model deeper. In general, with more GCN layers, node feature information can be propagated to further nodes. This helps to aggregate information from distant nodes and therefore improves the performance of GCNs. To design a more reasonable graph convolutional network algorithm for a more reasonable design, First according to the different ways of using information, will existing recommendation algorithm based on convolution networks classification algorithm for heterogeneous vertex interactive algorithm and homogenous interaction, and in the relationship between two kinds of method ignores the mutual aid and then by means of the experimental results show that the real data sets can be found that this method has better performance than the existing methods.

3. Methods

3.1. Form Sub-Graph to whole-graph

From the above figure, we can see that after running GNN successively with several different sub-graphs (each sub-graph has its own unique properties and the results are not the same), the results obtained are fused together to get the output we call whole-graph.

Fig.1 Construction and geometrical dimensions of specimens

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 01 Nov 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use
3.2. Whole-Graph to the Result

It can be seen from the Graph that we can obtain the Whole-graph through step 3.1. Since the whole graph is a combination of multiple sub-graphs, the whole graph will be presented with the unique properties of each sub-graph. Then we run the GNN again using the whole-graph combination and the result is our Result.

3.3. Advantages of using this method to speed up computation

Because our GNN can effectively mine the information in the graph domain, the data after mining will be strongly represented. Because of this, combining the representatives of each set of data and then using GNN again means that the components contained in it will include all of the previous data. In other words, we can use the whole-graph data and skip the sub-graph step and go straight to the result, which is why we can use this method to speed things up.

4. Experiments

4.1. The evaluation index

After the model is trained, for each user, we calculate the predicted rating for all the items that have not been interacted with in the training set and then sort these items in descending order of rating. The Top N items are selected as the predicted items of interest to the user, which is the top-n rating commonly used in recommendation systems. The specific evaluation index takes into account whether the items that the user is interested in in the test set appear in the top N digits, without considering the position and order of the items. N is taken to be 20 and 100 in the experiments. After calculating the metric for each user, the total metric is obtained by averaging over all users. Hyperparameter Settings All experiments are implemented on the PyTorch platform.

4.2. Data set

We used a dataset of consumer websites, which contains 156,432 consumers' 18M purchases from 2015 to 2019. All private information is eliminated. All input data are normalized in advance.

In the experiment, the data set was divided into a training set and a test set, and the proportion of 9:1 was randomly selected to divide the data, in which the number of the training set was 90%, and the number of the test set was 10%. In addition, the missing values of the features are supplemented with 0 operations, and the data with problems are filtered to prevent incorrect data from an experimental production.

4.3. Results

Table I Results of different methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Speed</th>
<th>Model size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM [3]</td>
<td>45.1</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td>DNN [4]</td>
<td>67.3</td>
<td>45.6</td>
<td>12.4</td>
</tr>
<tr>
<td>CNN</td>
<td>68.2</td>
<td>23.2</td>
<td>36.5</td>
</tr>
<tr>
<td>DenseCNN [1]</td>
<td>74.1</td>
<td>66.7</td>
<td>232.2</td>
</tr>
<tr>
<td>GL-method [3]</td>
<td>76.1</td>
<td>112.6</td>
<td>334.5</td>
</tr>
<tr>
<td>Attention</td>
<td>83.2</td>
<td>655.7</td>
<td>1034.7</td>
</tr>
<tr>
<td>Ours</td>
<td>86.6</td>
<td>83.2</td>
<td>213.7</td>
</tr>
</tbody>
</table>

The experimental results are shown in Table I, and it can be seen that all the methods can be roughly divided into three categories. SVM, DNN, and CNN perform poorly and are divided into one category. Other methods except ours can be divided into one class. It can be seen that although the Attention method works well, the amount of calculation is too large. Our method achieves the best results in terms of accuracy, speed, and model complexity.
5. Conclusion

Since personalized recommendation systems require high speed, we use a local comparison method, which greatly reduces the computational complexity. The experimental results also show that the effect of the method is not decreased, but improved. This is a good direction to meet the needs of social development.

References


A novel hash scheme for high elastic grid blockchain
Ying Yao\textsuperscript{a}, Xianke Zhou\textsuperscript{b} *, Zhifei Pang\textsuperscript{b}, Yong Yan\textsuperscript{a}, Liangyu Zha\textsuperscript{b}  
\textsuperscript{a} State Grid Zhejiang Electric Power Company Electric Power Research Institute, Hangzhou, China, 310014  
\textsuperscript{b} Institute of Computing Innovation, Zhejiang University, Hangzhou, China, 310008  
* Corresponding author: xkzhou@zjuici.com

ABSTRACT

The use of blockchain technology in high-elastic power grids often requires the deployment of blockchain nodes in charging piles, substations, new energy vehicles and other equipment, rather than in traditional servers. Limited by equipment performance and deployment conditions, the technical implementation of blockchain also needs to be optimized accordingly. In this paper, a novel hash scheme without multiple iterations which is suitable for high-elastic power grid is proposed. The hash scheme is constructed using large scale bool functions, which meet the requirements of balance, nonlinearity and SAC. The characteristics of the hash algorithm are also suitable for hardware implementation, which can build hardware micro services in the deployment architecture of high-elastic power grids.

Keywords: High-elastic power grids, micro service, Boolean function, hash algorithm

1. INTRODUCTION

Renewable energy such as solar, wind and heat developed quickly in the past decades\textsuperscript{1}. An intelligent infrastructure for energy utilities which integrate Internet, smart grid and automated technologies has brought a revolution to the whole process of power generation, distribution, and transmission\textsuperscript{2,3}. The features of high-elastic power grid are fully-automated, bi-directional and combinations of network, sensor, automatic control, and communication technology\textsuperscript{4}. Usually high-elastic power grid provides power management in the form of micro service. High-elastic power grid are changing the method of energy production since it can generate and sell their energy by themselves at end user level, which is really different with traditional top-down flow of energy transmission where a centralized utility dominates and provides all energy services\textsuperscript{5}.

However, the high-elastic power grids also face new issues such as management, energy trading, and scalability. For example, the bidirectional energy trading demand a new electricity charging mechanism for each node, which brings a series of problems such as data communication and information security\textsuperscript{1,6}. The security of blockchain technology can be integrated with highly elastic power grid to provide security guarantee for efficient power management and transmission. There are several advantages of introducing blockchain in high-elastic power grid, because decentralization, security and transparency are maintained\textsuperscript{7}. Transparency property means that data can be accessed in their public address. Security means that the blocks in a chain can’t be tampered later in anyway.

Although deploying blockchain in high-elastic power grid can provide security solution, there are still some obstacles to solve. The use of blockchain technology in high-elastic power grids often requires the deployment of blockchain nodes in charging piles, substations, energy storage batteries, electricity meters, new energy vehicles and other equipment, rather than in traditional servers\textsuperscript{8,9}. Considering the limited hardware computing ability in the high-elastic power grid and the importance of hash algorithm in blockchain, a novel hash algorithm is proposed in this paper which based on large scale Boolean function (LSBF) to accelerate the calculating speed. The discussed hash scheme is suitable for plug and play deployment of hash algorithm hardware and support multiple device access to improve blockchain performance and scalability.

2. HASH ALGORITHM BASED ON PERMUTATION

Confusion and diffusion are two concepts defined by Shannon in his paper Communication Theory of Secrecy Systems for cipher algorithm design. Diffusion is to let each bit in the plaintext affect many bits in the ciphertext, or let each bit in the ciphertext be affected by many bits in the plaintext. The role of diffusion is to conceal the statistical characteristics of the plaintext. Of course, the ideal situation is to let each bit in the plaintext affect all bits in the ciphertext, or let each bit
in the ciphertext be affected by all bits in the plaintext. Confusion refers to make the statistical relationship between ciphertext and key as complex as possible. Those two concepts indicate that the non-uniformity in the distribution of the letters in the plaintext should be redistributed efficiently enough in the ciphertext\textsuperscript{10,11}.

The purpose of confusion is to make it difficult for the attacker to find the key through statistical analysis, even if the attacker has a large number of plaintext-ciphertext pairs generated using the same key. Therefore, any bit of the ciphertext should depend on all bits of the key, and these bits depend on different parts of the key in different ways. In particular, any bit of the changed key should completely change the ciphertext.

The use of complex nonlinear substitution transformation can achieve better confusion effect, while the simple linear substitution transformation cannot achieve satisfactory confusion effect. However, complex nonlinear transformations often cost more computation. A substitution-permutation network is a common diffusion and diffusion mechanism in which the plaintext and the key often play a similar role in generating the output. Figure 1 shows a round of iteration of hash algorithm. In classical hash algorithm such as sha256, the iterations are 64. It requires relatively strong computing ability.

![Diagram](image)

Figure 1. One round iteration in classical hash algorithm (total 64 iterations)

In a high-elastic power grid environment, the computing power of microservices is limited. To solve the problem, a novel hash algorithm based on large scale Boolean function (LSBF) is proposed in this paper which meet the requirements of confusion and diffusion in one step way and it still meets the cryptography criteria.

### 3. BOOLEAN FUNCTION MEET CRYPTOGRAPHIC CRITERIA

In general, the following security criteria should be met when designing Boolean functions: balance, local and global avalanche characteristics, high nonlinearity, etc. Because these criteria are mutually restrictive, it is difficult to give consideration to and obtain the best compromise at the same time. The more criteria are used, the more difficult it is to construct Boolean functions that meet these criteria simultaneously using algebraic methods. Therefore, the selection of criteria and the construction of Boolean functions have received extensive attention.

A novel hash algorithm based on large scale Boolean function (LSBF) is designed to reduce the iteration steps which can reduce the requirement for computing power. Figure 2 shows the framework of the novel hash algorithm, which realizes the confusion and diffusion in one step. A method to construct large scale Boolean function will be verified. Suppose $f_1(x_1, \ldots, x_m)$ and $f_2(y_1, \ldots, y_n)$ are two Boolean functions defined on $Z_2^m$, $Z_2^n$, which are balance and the nonlinearity are $N_{f_1}$ and $N_{f_2}$ separately, and both of them meet SAC($k_1$), SAC($k_2$), here, $0 \leq k_1 \leq m-2$, $0 \leq k_2 \leq n-2$. 

---

1. Suppose $f_1(x_1, \ldots, x_m)$ and $f_2(y_1, \ldots, y_n)$ are two Boolean functions defined on $Z_2^m$, $Z_2^n$, which are balance and the nonlinearity are $N_{f_1}$ and $N_{f_2}$ separately, and both of them meet SAC($k_1$), SAC($k_2$), here, $0 \leq k_1 \leq m-2$, $0 \leq k_2 \leq n-2$. 

---

### Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 01 Nov 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use
3.1 Balance

The balance of a Boolean function \( f \) is defined as that the output of the function has the same probability as the values in the truth table. A Boolean function \( f : \mathbb{GF}(2^n) \rightarrow \mathbb{GF}(2) \) is said to be balanced, for all possible values of the input vector, if the probability that the value of \( f \) equals 1 is exactly half. That means:

\[
F : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2, \quad |f| = 2^{n-1}, \quad |f| \text{ is the minimizer value in Boolean function } f.
\]

**Theorem 1:** If both the Boolean function \( f_1(x_1, \ldots, x_m) \) and \( f_2(y_1, \ldots, y_n) \) satisfy the balance respectively, then function \( F(z) = F(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_1 \oplus f_2 \) also satisfies the balance rules. Here, \( \oplus \) stands for bitwise exclusive or operation.

**Proof:** Let's suppose \( V_1 \) and \( V_2 \) is the domain of Boolean function \( f_1 \) and \( f_2 \) separately, then

\[
F(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_1 \oplus f_2 = f_1 \overline{f_2} \oplus f_2 \overline{f_1}
\]

Here \( f_1 \overline{f_2} \) and \( f_2 \overline{f_1} \) are incomparable, \( V_1 \) and \( V_2 \) are independent, so we can get that

\[
|F(x_1, \ldots, x_m, y_1, \ldots, y_n)| = |f_1| \cdot |\overline{f_2}| + |f_2| \cdot |\overline{f_1}|
\]

\[
= 2^{m-1} \cdot 2^{n-1} + 2^{m-1} \cdot 2^{n-1}
\]

\[
= 2^{m+n-1}
\]

Equation (1)

\( F(x_1, \ldots, x_m, y_1, \ldots, y_n) \) is a Boolean function defined in \( \mathbb{Z}_2^{m+n} \) and it meets the balance criteria.

3.2 Strict avalanche criterion (SAC)

In a cryptographic algorithm with good diffusion performance, if any bit of the plaintext changes, the ciphertext should synchronously change completely in an unpredictable or pseudorandom manner. That is, diffusion ensures that the output bits depend on the input bits in a very complex way. In particular, for randomly selected inputs, if the \( i \)-th bit is flipped, then for any \( i \) and \( j \), the \( j \)-th output bit will be flipped with a 50% probability - this is called the strict avalanche criterion (SAC). More generally, it may be necessary to flip a set of fixed length bits, and then change each output bit with a 50% probability\(^{12,13} \).

If complementing any single bit will cause the probability of the output bit to change by half, then the Boolean function is said to satisfy SAC. That's:

\[
\overline{F}(x_i) = \overline{f}(x_1, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)
\]

\[
= f(x_1, \ldots, x_{j-1}, 1+x_j, x_{j+1}, \ldots, x_n)
\]

\[
= f^r f_j (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)
\]

**Theorem 2:** If both the Boolean function \( f_1(x_1, \ldots, x_m) \) and \( f_2(y_1, \ldots, y_n) \) satisfy the SAC(k) respectively, then the Boolean function \( F(z) = F(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_1 \oplus f_2 \) satisfy SAC(k), \( k = \min(k_1, k_2) \).

**Proof:** According to the meaning of SAC, we can make the following deduction:

\[ F_1(x, y) = F(x_1, \ldots, x_m, x_r, x_{r+1}, \ldots, x_m, y_1, \ldots, y_n) \]

\[ = f_1(x_1, \ldots, x_m, x_r, x_{r+1}, \ldots, x_m) + f_2(y_1, \ldots, y_n) \]
\begin{align}
&= f_1(x_1, \ldots, x_i-1, x_i, x_{i+1}, \ldots, x_m) + f_2(y_1, \ldots, y_n) \\
&= f_1(x_1, \ldots, x_m) + \ldots \quad (7)
\end{align}

Here, $f_i$ satisfy SAC($k_i$), $i \leq k \leq k_1$, so $f_i$ meets SAC($k$); $f_2$ satisfy SAC($k_2$), $j \leq k \leq k_2$, so $f_2$ meets SAC($k$); $r_i(x)$ here is a Boolean function on $V_1$, so $f_i$ meets SAC($k$), $i \leq k \leq k_1$, and $f_i$ satisfy SAC($i$), then $|r_i| = 2^{m-1}$; As a Boolean function in $V_1 \times V_2$, each minimizer value in $r_i$ can be regarded as the full function of $V_2$, so $|r_i| = 2^{m+n-1}$. Now that the half of minimizers of $F(x,y)$ changes due to the change of $x_i$.

Similarly, it can be proved that each of input $y_i \in V_2$, we get:

\begin{align}
F_2(x,y) &= F(x_1, \ldots, x_m, y_1, \ldots y_{i-1}, y_{i+1}, y_i) \\
&= f_1(x_1, \ldots, x_m) + f_2(y_1, \ldots, y_{i-1}, y_{i+1}, y_i) \\
&= f_1(x_1, \ldots, x_m) + f_2(y_1, \ldots, y_{i-1}, y_{i+1}, y_i) + r_i(y) \quad (3)
\end{align}

The half minimizers of $F(x,y)$ changes synchronously with $x_i$. That is, in the input with any length of $k$ bits, for any one bit $z_i \in V_1 \times V_2$ is reversed, the value of $F(z)$ will change by half. That means the new constructed large scale Boolean function meets SAC($k$).

### 3.3 Nonlinearity

The minimum Hamming distance is introduced to define the nonlinearity of a Boolean function. All the affine functions of Boolean function is defined on $GF(2^n) \rightarrow GF(2)$. The nonlinearity is a key criteria for Boolean functions, which can be described as follows:

\begin{align}
N_f &= \text{min} \{d(f, \phi_1), i = 0, 1, \ldots 2^n-1, \phi_1 \text{ here is an affine function in the collection } \Phi, \\
\sum_{\phi \in \Phi} (f \oplus \phi), V \text{ stands for the input domain of Boolean function } f.
\end{align}

The maximum value of nonlinearity of Boolean function $f$ is $2^{n-1} - 2^{0.5n-1}$.

**Theorem 3**: If the Boolean function $f_1(x_1, \ldots, x_m)$ and $f_2(y_1, \ldots, y_n)$ have the nonlinearity $N_{f_1}$, $N_{f_2}$ respectively, then the constructed Boolean function $F(z) = F(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_1 \oplus f_2$ has nonlinearity $N_f = N_{f_1} + N_{f_2} - 2^{m-n} N_{f_1} \cdot N_{f_2}$.

**Proof**: Suppose $\Phi(x \oplus y)$ is the collection of all affine function defined on $V_1 \times V_2$, here, $V_1$, $V_2$ is the definition domain of Boolean function $f_1$ and $f_2$ separately. Then, the new constructed Boolean function $\phi(x \oplus y)$ can be expressed as $\phi(x) \oplus \phi(y)$, $\phi_1$ is any affine function in the collection $\Phi$. The distance between $F(x,y)$ and $\Phi(x \oplus y)$ is expressed as $d(F, \Phi)$, then the nonlinearity of $F(x,y)$ is the minimum value of $d(F, \Phi)$ of all $\phi (x \oplus y)$.

\begin{align}
d(F, \Phi) &= | \{ (\xi_1, \xi_2) : (x_1, \ldots, x_m, y_1, \ldots, y_n) = 1 \} | \\
&= | \{ (\xi_1, \xi_2) : \xi_1 \in V_1 \land \xi_2 \in V_2, (f_1(\xi_1) \oplus f_2(\xi_2)) \oplus (\phi_1(\xi_1) \oplus \phi_2(\xi_2)) = 1 \} | \\
\end{align}

$| \Phi |$ stands for the base of collection $\Phi$.

\begin{align}
&= \left( f_1(\xi_1) \oplus \phi_1(\xi_1) \right) \oplus \left( f_2(\xi_2) \oplus \phi_2(\xi_2) \right) \\
&= \left( f_1(\xi_1) \oplus \phi_1(\xi_1) \right) \oplus \left( f_2(\xi_2) \oplus \phi_2(\xi_2) \right)
\end{align}

Here, $\xi_1$ and $\xi_2$ is independent, and both sides of $\wedge$ are incompatibility.

\begin{align}
d(F, \Phi) &= | \{ \xi_1 : (f_1(\xi_1) \oplus \phi_1(\xi_1)) = 1, \xi_1 \in V_1 \} | \cdot | \{ \xi_2 : (f_2(\xi_2) \oplus \phi_2(\xi_2)) = 0, \xi_2 \in V_2 \} | \\
&+ | \{ \xi_1 : (f_1(\xi_1) \oplus \phi_1(\xi_1)) = 0, \xi_1 \in V_1 \} | \cdot | \{ \xi_2 : (f_2(\xi_2) \oplus \phi_2(\xi_2)) = 1, \xi_2 \in V_2 \} |
\end{align}

Suppose, $M_1$ is the distance of $f_1$ and $\phi_1$, $M_2$ is the distance of $f_2$ and $\phi_2$, then

\begin{align}
d(F, \Phi) &= M_1 \cdot (2^n- M_2)+ M_2 \cdot (2^n- M_1) = M_1 \cdot 2^n+ M_2 \cdot 2^n-2 \cdot M_1 \cdot M_2
\end{align}
\[ \frac{\partial d(F, \Phi)}{\partial M_1} = 2^{n-2} \cdot M_2, \quad (8) \]

When \( M_2 < 2^{n-1} \), \( d(F, \Phi) \) is a monotone increasing function of \( M_1 \), and its value is kept at \( M_2 \leq 2^{n-1} - 2 \) according to the nonlinear property.

In the same way,

\[ \frac{\partial d(F, \Phi)}{\partial M_2} = 2^{m-2} \cdot M_1, \quad (9) \]

When \( M_1 < 2^{m-1} \), \( d(F, \Phi) \) is monotone increasing function of \( M_2 \). Since the minimum value of \( M_1, M_2 \) is \( N_{f1} \) and \( N_{f2} \), so the minimum value of \( d(F, \Phi) \) can be got when \( M_1 \) and \( M_2 \) are equaled to \( N_{f1}, N_{f2}, \) we can get the result as follow:

\[ N_f = \min(d(F, \Phi)) = \min( M_1 \cdot 2^{n+} M_2 \cdot 2^{m-2} \cdot M_1 \cdot M_2 ) = N_{f1} \cdot 2^{n+} N_{f2} \cdot 2^{m-2} \cdot N_{f1} \cdot N_{f2} \quad (10) \]

Now, a way to construct large scale Boolean function (LSBF) is proved to be safe which can be used to design the novel hash algorithm.

4. NOVEL HASH ALGORITHM

According to the above theoretical proof, different hash algorithm affects overall performance of blockchain. The proposed hash scheme is about 3 times fast than traditional SHA256 hash algorithm. It is meaningless to give specific and exact data, because it is closely related to the actual environment of the system. Our algorithm exchanges the storage cost for the reduction of the calculation iteration rounds. Figure 3 shows the frame work of hash algorithm in the blockchain. It is sure that our scheme is much more quickly than traditional 64 round iteration hash scheme since it is one round confusion and diffusion calculation.

![Figure 3. LSBF based hash algorithm applied in blockchain](image)

Since the novel hash scheme is only one round of computation to realize confusion and diffusion operation, it is more suitable to for hardware implementation and be deployed in edge end where the computing ability is restricted.

5. CONCLUSIONS AND PROSPECTS

Hashing algorithm is an important component of blockchain. In high-elastic power grids environment, the computation capability is restricted. We have presented a highly efficient hash algorithm which is suitable for blockchain in high-elastic power grid in this paper. The paper proves theoretically that the novel hash scheme meets cryptographic criteria and it is security. The hash scheme in this paper provides a hash alternative in the application environment with limited computing capacity such as the micro service in high-elastic power grid.

In the future, the security of this scheme will be more comprehensively demonstrated in theory and practice. The research team will give full play to the advantages of the algorithm, commit to transplanting the algorithm to a dedicated...
ASIC chip for implementation, implement a complete network and micro service protocol stack in the hardware, implement plug and play scalable distributed hash capabilities in combination with switches, reduce the dependence of blockchain nodes on CPU computing ability, and try the deployment of high-performance blockchain nodes at the edge end.

ACKNOWLEDGMENT

This work was supported by Science and Technology Foundation of State Grid Zhejiang Electric Power Co. Ltd “research and application of state key technology of blockchain supporting high elastic power grid balance (B311DS21000H)”.

REFERENCES

Ship Scene Vehicle Location Method based on Multi-source Information Fusion

Wentao Pan\textsuperscript{1a*} and Jianpo Guo\textsuperscript{1b*} and Zongmiao Dai\textsuperscript{1c} and Weiru Wang\textsuperscript{1d}

\textsuperscript{1,2,3,4,9th laboratory,713th Research Institute of China Shipbuilding Corporation Limited, Zhengzhou, Henan, 450000, China}
\textsuperscript{a*649463239@qq.com, b*guojianpowork@126.com, c*daizongmiao@126.com, d*2415184918@qq.com}

Abstract

This paper proposes a real-time vehicle location method based on multi-source information fusion of vehicle wheel speed features and visual images for the problems of weak texture features, high similarity of regional features, low accuracy of feature recognition and detection, and jump of vehicle location results in ship scene. Firstly, in order to solve the problem of inertial navigation sensor failed and get accurate pose estimation of vehicle motion, the article builds a vehicle motion prediction model based on vehicle wheel speed features. Secondly, according to the significant geometric features of the actual scene, Hough transform, edge extraction, template matching and inverse projection transform algorithms are used to identify and detect visual features of the scene. Thirdly, an extended Kalman filter is used to fuse vehicle wheel speed pose estimation and visual features to identify and detect the results to obtain reliable and high-precision real-time location of the vehicle. Finally, experiments are employed to verify the real-time vehicle location method based on multi-source information fusion of vehicle wheel speed features and visual images. The experimental results show that this method can effectively achieve vehicle localization in weak texture, similar regional features, and non-stationary road scenes.

Keywords-Pose Estimation; Information Fusion; Feature Detection; Vehicle Real-time localization

1. Introduction

With the rapid development of artificial intelligence and wireless communication technology, machines gradually becoming a reality to replace humans in vehicle control tasks. Self-driving technology can greatly improve vehicle safety and traffic efficiency with high economic and social benefits, which is an important direction for future development of vehicle technology [1][2].

Self-driving vehicle perceives the real-time information of vehicle status and driving scene through the on-board sensor equipment and controls the vehicle to arrive at the destination by following the trajectory given by autopilot software according to the real-time information. Real-time vehicle location is a prerequisite for the implementation of self-driving technology. Self-driving vehicles need the real-time vehicle location to obtain their position in the scene of self-driving vehicle, and to plan the self-driving path of vehicle according to the mission.

Currently, majority scholars' research on real-time vehicle location methods focuses on satellite navigation combined with inertial device location, wireless signal location, SLAM, or the form of fusion of multiple location methods [3][4][5]. Reference [6] proposed a real-time location method that integrates UWB base stations and inertial measurement devices to improve the reliability and environmental adaptability of the location system. Reference [7] proposed a SLAM algorithm based on the fusion of vision and lidar information, which integrates visual images into the lidar SLAM framework to achieve accurate mapping and localization of dynamic, feature-sparse underground garage scene. Reference [8] proposed a point-line feature fusion lidar monocular inertial navigation tightly coupled SLAM system, which can better solve the location problem of weak texture or degradation.

Real-time location technology has a high correlation with the system application scene. Compare with the land-based scene, the ship scene has features of complex electromagnetic environment, weak texture features, high similarity of regional features, narrow work space, lots of moving targets, and non-stationary etc. The available location methods are difficult to solve vehicle location needs in the complex ship scene. For the features of ship scene, this paper proposes a real-time vehicle location method based on multi-source information fusion of vehicle wheel speed features and visual...
images. This method can obtain reliable and high-precision real-time vehicle location by building and calibrating an accurate vehicle kinematic model and fusing visual feature detection information.

2. System Frame

The location system includes visual feature information detection, and real-time pose estimation fused with visual feature detection and vehicle wheel speed features. Firstly, the system obtains the visual information of the scene via the in-vehicle visual sensor. Secondly, the system identifies and detects the visual features of scene through grayscale transformation, image noise reduction, edge extraction, HOG feature extraction and Classification filter etc. Finally, reliable and high-precision real-time vehicle location estimation results are obtained through extended Kalman filter.

![Flow chart of multi-source fusion location system](image)

3. Vehicle Pose Estimation Based on Vehicle Wheel Speed Features

The wheel speed of the vehicle is defined as the number of turns of the four wheels of the vehicle per unit time. For the Omni-directional vehicle with mecanum wheels, the real-time speed of the vehicle can be calculated through the detection of the vehicle wheel speed and the vehicle kinematics model. Then, the odometer information of the vehicle can be obtained by time integration of the real-time speed of the vehicle. Thus, the pose estimation of the vehicle can be realized when the initial pose of the vehicle is known. The relationship between the wheel speed of the mecanum wheel and the vehicle speed is shown in Fig. 2.

![Vehicle kinematics model of mecanum](image)
The vehicle kinematics model of mecanum is shown in formula (1). 

\[
\begin{bmatrix}
\omega_1 \\
\omega_2 \\
\omega_3 \\
\omega_4
\end{bmatrix} = \frac{1}{R} \begin{bmatrix}
1 & 1 & -\tan(a \tan \alpha + b) \\
\tan \alpha & -1 & \tan \alpha \\
\tan \alpha & 1 & -\tan(a \tan \alpha + b) \\
\tan \alpha & 1 & \tan \alpha
\end{bmatrix} \begin{bmatrix}
v_x \\
v_y \\
v_\gamma \\
\omega_z
\end{bmatrix}
\]

(1)

The real-time motion speed feedback of the vehicle can be obtained through the inverse kinematics model of the vehicle, when the speed of the four wheels of the vehicle is known. As shown in formula (2), where \( \alpha = 45^\circ \) is generally taken.

\[
\begin{bmatrix}
v_x \\
v_y \\
\omega_z
\end{bmatrix} = R \begin{bmatrix}
1 & 1 & -(a + b) \\
-1 & 1 & (a + b) \\
-1 & 1 & -(a + b) \\
1 & 1 & (a + b)
\end{bmatrix} \begin{bmatrix}
\omega_1 \\
\omega_2 \\
\omega_3 \\
\omega_4
\end{bmatrix}
\]

(2)

\([v_x, v_y, \omega_z]^T\) is the vehicle speed, where \(v_x\) and \(v_y\) represent the two components of the vehicle speed in the horizontal direction, respectively, and \(\omega_z\) represents the heading angular velocity as the four-wheel angular velocity, \(\omega_1, \omega_2, \omega_3, \omega_4\). The vehicle body pose prediction result \([x, y, \theta]^T\) is obtained by integrating the vehicle body speed \([v_x, v_y, \omega_z]^T\) through formulas (3), (4) and (5).

\[
x_t = x_{t-1} + \cos(\theta_{t-1}) v_x \Delta t - \sin(\theta_{t-1}) v_y \Delta t \\
y_t = y_{t-1} + \sin(\theta_{t-1}) v_x \Delta t + \cos(\theta_{t-1}) v_y \Delta t \\
\theta_t = \theta_{t-1} + \omega_z \Delta t
\]

(3, 4, 5)

Where \([x, y, \theta]^T\) represents the vehicle body pose, \(x, y\) is the vehicle body position, and \(\theta\) is the heading angle. \([x_t, y_t, \theta_t]\) represents the pose of the car body at time \(t\).

4. Pose Estimation of Visual Images

4.1. Visual Feature Map Creation

Visual map creation is mainly based on the detection results of visual features, given spatial coordinates to each visual feature. The map is used to estimate the spatial pose of the vehicle through visual feature detection. The visual feature map requires that the feature points in the map are different from each other. Each visual feature point in the map has its only ID number to distinguish the difference from each other. The map is composed of a series of feature points \(mt = \{m_1, m_2, m_3, \ldots\}\). A feature includes feature descriptor and its coordinates in the image \((m_{tx}, m_{ty})\).

Actually, the visual feature map creation problem can be expressed as solving the equation \(p(x_t, m|u_t, z_t)\). Where \(m\) represents the most generated map, \(x_t\) represents the pose of each frame of observation, \(u_t\) is the wheel speed feedback information, and \(z_t\) is a frame of observation data. The map is obtained by estimating each frame observed pose.

![Fig.3 The flowchart of visual feature map creates](image-url)
by extracting the features of the visual image, and then the accurate feature map is obtained by using the pose estimation between the two frames before and after for back-end optimization.

4.2. Visual Feature Detection

The texture features of the ship scene are poor, while the circular tethered base on the ground can be used as visual features to be detected for significant geometric features. Therefore, in this paper Hough transform, which is more suitable to geometric shape feature detection, is used for visual feature recognition. The method transforms problem of detecting a given curve in the original image to the problem of finding peaks in the parameter space by transforming the given curve in the original image space to a point in the parameter space through the curve expression according to the duality of point and line.

The principle of identification and detection is shown in Fig. 4. Figure (a) is a straight line in the parameter \((x, y)\) space, and Figure (b) is a straight line in the parameter \((m, b)\) space. The parameter \((m_0, b_0)\) of straight line in Figure (a) is a point in the parameter space \((m, b)\) shown in Figure (b), correspondingly, the line parameter in Figure (b) space corresponds to a point in Figure (a) space. By analogy, when two points in the \((x, y)\) space correspond to two straight lines in the parameter space \((m, b)\). The intersection of these two straight lines means that the two points in the original space are collinear. The coefficient of the line equation is the coordinate of the intersection of two straight lines in the parameter space. When multiple points are collinear, the coefficient of the line is the coordinate of the intersection point of the multiple lines in the parameter space. Similarly, for other geometric shapes such as circles and ellipses, the accurate parameters of their edge curves can also be determined through multi-point detection.

The process of visual feature detection is shown in Fig. 5.

Step 1: segment the image and intercept the region of interest to reduce the search range for detection.
Step 2: convert the captured original image into a grayscale image and filtered, filter the grayscale image. The color of feature in the image will change according to environmental changes, and the brightness information contained in the grayscale image is sufficient to express the feature. Therefore, feature detection should be based on the grayscale image.

Step 3: target detect. The brightness is significant difference between ground and visual signees, since their material is different. The candidate boxes containing visual signees can be obtained by edge detection and contour extraction. Edges are areas of an image with dramatic changes in brightness. This method adopts canny edge detection algorithm, which includes three steps of calculating gradient value, gradient direction, non-maximum value suppression and upper and lower threshold detection. Since the size and shape of visual signees are fixed, the size and shape of visual signees in the image, and the pose between visual signees and vehicle can be obtained by statistical methods.

Step 4: calculate the actual relative vehicle pose by the visual signees coordinates in the image according to the inverse projective transformation (IPM).

5. Vehicle Location Based on Information Fusion

Real-time vehicle location calculation by fusing wheel speed with visual pose detection and feature maps can be expressed as maximum posterior probability Bayesian estimation calculate, when vehicle wheel speed information, visual pose detection results and historical vehicle location are known.

\[
p(x_t|u_t, z_t, M) \quad (6)
\]

Where \( x_t \) is the pose of the vehicle at time \( t \), \( u_t \) is the control information of the vehicle at time \( t \), \( z_t \) is the current visual pose detection result of the vehicle at time \( t \), \( M \) is the visual feature map of scene.

\[
\begin{align*}
\mathcal{N}(x_t; \mu_t, \Sigma_t) \\
\end{align*}
\]

Fig.6 Bayesian Pose Estimation Model

The vehicle fusion location is realized by using the extended Kalman filter method, as shown in formula (7).

\[
p(x_t|u_t, z_t, M) = \eta p(z_t|x_t, M) \int_{x_{t-1}} p(x_t|u_t, x_{t-1})p(x_{t-1}|u_{t-1}, z_{t-1}, M)dx_{t-1} \\
\quad (7)
\]

Where \( p(x_{t-1}|u_{t-1}, z_{t-1}, M)dx_{t-1} \) is the posterior probability of the previous state. \( p(x_t|u_t, x_{t-1}) \) is the prediction model, that can also be expressed as the vehicle kinematics model. \( p(z_t|x_t, M) \) is the pose estimation model of visual, and \( \eta \) is a scale factor.

The prediction of the vehicle location state can be simplified as formula (8).

\[
\overline{bel}(x_t) = \int p(x_t|x_{t-1}, u_t)bel(x_{t-1})dx_{t-1} \quad (8)
\]

Where \( bel(x_{t-1}) \sim \mathcal{N}(x_{t-1}; \mu_{t-1}, \Sigma_{t-1}) \) is the location of the vehicle at time \( t - 1 \), that satisfies the Gaussian distribution with mean \( \mu_{t-1} \) and covariance \( \Sigma_{t-1} \). \( p(x_t|x_{t-1}, u_t) \sim \mathcal{N}(x_t; g(u_t, \mu_{t-1}) + G_t(x_{t-1} - \mu_{t-1}), R_t) \) is the vehicle motion control model. The vehicle pose can be obtained by vehicle control information according to the model. Similarly, the vehicle kinematics model satisfies the Gaussian distribution with mean \( g(u_t, \mu_{t-1}) + G_t(x_{t-1} - \mu_{t-1}) \) and covariance \( R_t \), where \( g(\cdot) \) is the vehicle kinematics model, \( G_t(\cdot) \) is the first derivative of \( g(\cdot) \). \( \overline{bel}(x_t) \sim \mathcal{N}(x_t; \mu_t, \Sigma_t) \) is the pose state of the vehicle at the current time \( t \), which is calculated according to the vehicle state at time \( t - 1 \) and the motion model of the vehicle.

The vehicle kinematics model is described in the wheel speed feature analysis chapter.
$$g(\cdot) = \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ \theta_{t-1} \end{pmatrix} + \begin{pmatrix} \cos(\theta_{t-1}) v_x \Delta t - \sin(\theta_{t-1}) v_y \Delta t \\ \sin(\theta_{t-1}) v_x \Delta t + \cos(\theta_{t-1}) v_y \Delta t \\ \omega \Delta t \end{pmatrix}$$

$$G_t = \begin{pmatrix} 1 & 0 & -\sin(\theta_{t-1}) v_x \Delta t - \cos(\theta_{t-1}) v_y \Delta t \\ 0 & 1 & \cos(\theta_{t-1}) v_x \Delta t - \sin(\theta_{t-1}) v_y \Delta t \\ 0 & 0 & 1 \end{pmatrix}$$

The prediction of the vehicle location state could be updated based on the visual perception information according to formula (11).

$$\text{bel}(x_t) = \eta p(z_t|x_t) \overline{\text{bel}}(x_t)$$

Where is $\text{bel}(x_t)$ the updated prediction of the vehicle location state, $p(z_t|x_t)$ is the perception information at the current moment.

Where $z_t = h(x_t, M) + \delta_t$, $h(x_t, M)$ is the pose estimation of visual according to the feature map and prediction of the vehicle location state, which reflects the conversion relationship between the pose estimation of visual and the prediction of the vehicle location state. The $\delta_t \sim N(0, Q_t)$ is noise of pose estimation of visual, which mean value is $h(\mu_t) + H_t(x_t - \mu_t)$ and covariance is $Q_t$. $h(\cdot)$ is the pose estimation model of visual, $H_t(\cdot)$ is the first derivative of $h(\cdot)$.

The position of the visual feature in the vehicle body coordinates are defined as $z = (x, y)^T$. $\hat{z}_t^i$ is the model output of $h'(x_t, M)$ according to the feature map and prediction of the vehicle location state. The results of pose estimation based on visual can be obtain through Comparison of $\hat{z}_t^i$ and $z = (x, y)^T$, which is shown as formula (13), and the first derivative of $\hat{z}_t^i$ is shown as formula (14). When comparing, it is necessary to convert the map feature points to the unified coordinate system of the vehicle body according to the formula (12).

Here, $R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$, $\theta = x_t(3)$, $T = \begin{pmatrix} x_t(1) \\ x_t(2) \end{pmatrix}$.

$$\hat{z}_t^i = h'(x_t, M) = R^{-1}(z_m - T)$$

$$\begin{pmatrix} x_{t+1}^i \\ y_{t+1}^i \end{pmatrix} = \begin{pmatrix} (x_m^i - x_{tx}) \cos(\theta) + (y_m^i - x_{ty}) \sin(\theta) \\ -(x_m^i - x_{tx}) \sin(\theta) + (y_m^i - x_{ty}) \cos(\theta) \end{pmatrix}$$

$$H_t' = \begin{pmatrix} -\cos(\theta) & -\sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$$

The algorithm flow of the vehicle location software that fuses the vehicle wheel speed features and visual features is as follows:

- Step 1: EKF ($\mu_{t-1}$, $\Sigma_{t-1}$, $\mu_t$, $z_t$, $M$);
- Step 2: $\mu_t = g(u_0, \mu_{t-1})$;
- Step 3: $\Sigma_t = G_t \Sigma_{t-1} G_t^T + R_t$;
- Step 4: $K_t = \Sigma_t H_t' (H_t \Sigma_t H_t' + Q_t)^{-1}$;
- Step 5: $\mu_t = \mu_t + K_t(z_t - h(\mu_t))^{-1}$;
- Step 6: $\Sigma_t = (I - K_t H_t) \Sigma_t$;
- Step 7: Output $\mu_t$ and $\Sigma_t$.

6. Experiments and Results Analysis

6.1. Experimental Conditions

In order to verify the effectiveness of the algorithm, a relatively open scene is selected in this paper and a stainless steel sheets is used to simulate the tethered seat features of the ship scene, which is arranged in the form of a 1.5m grid as shown in Fig. 7. The vehicle platform is a mecanum wheel chassis, whose dimensions is 4000mm*1400mm*800mm (length*width*height). The vision sensor is ZED binocular camera, which is installed and fixed on the vehicle platform as shown in Fig. 8. The computing unit is MXC-7410D, that processor is i7-6700, memory is 16G, and operating system is ubuntu16.04.
6.2. Experimental Results

In order to verify the effectiveness of the algorithm according to the actual experimental data, this experiment sets a stable driving trajectory to the vehicle, and records real-time location data of the vehicle during movement process. As shown in Figure 9, the blue curve is the given driving trajectory of the vehicle, and the red curve is the real-time location result of the vehicle. The location results obtained by the method in this paper are stable, reliable and highly accurate. The error of location results is larger and the maximum value of the error less than 17cm in the whole process when the vehicle is turning. The experimental results of vehicle location based on the visual feature are shown in Figure 10. This method realizes vehicle location by identifying and detecting ground visual features. The location results based on visual are discontinuous, greatly volatility, unstable, and low accuracy due to the discontinuity in visual feature detection. Moreover, the visual recognition and detection methods fail to distinguish visual features due to the high similarity of the visual features of the ship scene. When a visual feature is missed, vehicle location errors based on visual will be more than distance between adjacent visual features, which is 1.5m. The effectiveness of the real-time location algorithm fused with wheel speed features and visual information has been proved through experimental analysis and comparison.
7. Conclusions

In this paper, a real-time vehicle location method that fuses vehicle wheel speed features and visual information is proposed for the real-time vehicle location in ship scene. This algorithm effectively solves the problem of inertial device failure under moving scene conditions through building and calibrating the model of vehicle wheel speed characteristics. Considering the problems of weak texture, high similarity of regional features and visual feature pose estimation unreliable, the method has been achieved reliable and high-precision real-time vehicle location through fusing visual information and vehicle wheel features and verified by experiments.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (2021-62036008).

References

Mimic Adjudication Scheme Based on Homomorphic Encryption

Chuan He*ab, Bo Zhangab, Zesheng Xiab, Yunfan Wangab

aState Grid Smart Grid Research Institute co., Ltd., Nanjing, China; bState Grid Key Laboratory of Information & Network, Nanjing, China; *Corresponding author: 1255065000@qq.com

ABSTRACT

In the information times, with the rapid development of network technology, emerging technologies such as big data and artificial intelligence have gradually entered daily life, but various network security risks have also followed. In recent years, as a new security strategy, mimic defense technology has been a new force suddenly rises, providing a new idea and direction for protecting network security. In the actual application scenario, the resource consumption of the mimic defense service is large, and it is generally deployed and provided by the cloud service provider. Although the mimic defense can play a good role in defending against external attacks, the data transmission in the current system is based on plaintext, which cannot guarantee the privacy of user data. Homomorphic encryption technology, as an encryption supporting data ciphertext operation, can protect data privacy on the premise of completing data operation requirements. As an important field of homomorphic technology, homomorphic hash technology has been widely used in cloud storage scenarios. Aiming at the problem that the cloud service mimic defense system cannot protect data privacy, this paper introduces homomorphic encryption technology into the mimic defense system to protect the privacy of user data. The scheme proposed in this paper is based on the mimic defense architecture, which optimizes the adjudication process of the adjudication module through the homomorphic encryption technology, make full use of the characteristics of homomorphic operation, protects the privacy of user data, and improves the system security. Moreover, this paper builds a mimic routing platform for experimental analysis. The experimental results show that the malicious attacker can see the user’s private information under the traditional defense mechanism. However, after homomorphic encryption optimization, each user authentication is still less than 1 millisecond, and the user information data become invisible to malicious attackers.

Keywords: mimic defense. Homomorphic encryption, dynamic heterogeneous redundancy, privacy protection

1. INTRODUCTION

With the development of society, Internet technology and various related emerging technologies such as big data [1] and artificial intelligence [2] have rapidly penetrated into all aspects of society and are closely related to our daily work and life. The application of new technologies is bound to be accompanied by new risks. While enjoying the convenience brought by new technologies, people are also facing the risk of personal privacy, work data and sensitive information leakage. These security issues are unavoidable obstacles for the development of Internet technology [3]. In recent years, with the concept of "Internet of everything" and the rise of various intelligent devices, Internet technology and data are ubiquitous, and the form of data security is becoming more and more severe. Data security is the basis for maintaining social stability and ensuring national security. Attacks against important data will not only cause losses to individuals and companies, but also may harm important national data and endanger national security.

Among all the network attacks, the attack against data is the main form of hacker attack, which is highly harmful and involves a wide range. It has increasingly become the core security risk of the Internet system. Therefore, protecting data security has always been an important direction of network security research at home and abroad. Data security threat refers to that the security of all nodes cannot be guaranteed during data transmission. Hackers do not need to directly attack the database server. As long as they attack some key nodes, they can monitor, steal or even tamper with a large amount of data. The security threat of the system itself mainly refers to some security vulnerabilities that may exist in the design of the database server, such as security vulnerabilities in the operating system, hardware equipment and database application. In addition, internal personnel of the company directly delete, tamper with or steal data by using personal authority is also a typical internal security threat.
In the solution of data security, Mimic Defense [4] emerged as a new concept. With the characteristics of heterogeneous redundancy and active defense, it has repeatedly achieved good results in various network security performance tests. Based on a large number of isomers, the Mimic Defense technology is realized through the scheduling module and the arbitration module. It is adjustable and controllable on the premise that the security is higher than the traditional defense mode, and can achieve the unity of performance and security. It is a network security defense system that "changes the rules of the game".

In the actual application scenario, the resource consumption of the Mimic Defense server is large, and it is generally deployed and provided by the cloud service provider. Although the Mimic Defense plays a good defense role against external attacks, it cannot guarantee the privacy of user data. In the cloud security system, blockchain technology and searchable encryption technology can respectively solve the problems of data tampering and partial search of ciphertext. However, with the concept of Internet of things, the demand for data in cloud computing is increasing. However, not all data can be transmitted to cloud service providers in plaintext. Homomorphic encryption technology [5] is an encryption technology supporting data ciphertext operation, which can complete the data operation required by the user when the third party is not trusted, which provides a new idea for solving the above problems. With the partial homomorphic encryption technology, ciphertext can be added or multiplied in the cloud. In addition, with the full homomorphic encryption technology, the cloud can perform arbitrary calculation on ciphertext.

For the protection of user data privacy, this paper will study the application of homomorphic encryption technology in Mimic Defense, ensure data security on the basis of cloud computing. In addition, this paper will conduct in-depth research on the security, efficiency and usage scenarios of homomorphic encryption technology applied to the mimic defense system.

2. RELATED RESEARCH

The current cyberspace security technology mainly conducts precise defense by perceiving threat characteristics. This defense method belongs to passive defense, and requires a priori knowledge such as the source, path, behavior and characteristics of the attacker as support, such as authentication or encryption function. This traditional defense system is very fragile in the face of unknown loopholes. When the system is not completely credible, there is little to do in the face of threats without prior knowledge. Even if the system has authentication function or encryption function, the attacker can still attack by various means such as bypass attack. In addition, the current network system architecture is static, which makes it possible to identify, detect and evaluate hacker attacks. At the same time, the storage, processing and sharing of a single database is also a serious vulnerability of the current network information system. The attacker does not need to attack multiple targets. Once the database is broken, the hacker can perform almost any operation under the condition of resource sharing. Now, most network attack principles are born under this premise.

At present, the protection means for system security, network security and data security are mainly divided into two categories, namely, active security technology and passive security technology [6]. The active defense technology mainly deals with high persistent and high-intensity network attacks, and uses endogenous security mechanisms to resist unknown security threats, including trusted computing technology, mimic defense technology and mobile target defense technology [7]. Passive security defense technology mainly includes firewall technology, intrusion detection technology and vulnerability detection technology to strengthen the defense of the server system; Encrypting network data using cryptography encryption technology; Data backup technology is used to avoid data unavailability caused by hacker attacks.

2.1 Mobile target defense

Mobile target defense [8] has a symbiotic mechanism, which is characterized by randomization, diversification and dynamics to create a dynamic and constantly changing target. Current network attacks usually need to be attacked according to the certainty and staticity of the target. Mobile target defense technology constantly changes the resource allocation relationship or the system environment on which it is based to destroy the attack of the attacker. This technology sets a precedent for changing the internal environment to resist network attacks. It is a good idea, but it is still a shell defense method, which only gives the original environment a random and dynamic appearance, but it is still powerless for the back door.

2.2 Mimic defense

Mimic defense technology solves the problems encountered by the MTD technology. The mimic defense technology is also dynamically changing. However, it does not simply shell the existing objects, but uses redundant resources to
change the internal structure, operation mechanism and even abnormal performance without changing the functions of the original objects and affecting the performance of the original objects. By using space-time changes, the mimic defense technology can show a "specious" situation when facing the attacker, causing serious interference to the attacker and greatly reducing the success rate of the attack. Mimic defense is an inclusive, open and active defense technology system. Mimic defense no longer pursues to establish a defense system without loopholes, backdoors and defects to combat cyberspace security threats. Instead, it adopts various and constantly changing evaluation and deployment mechanisms and strategies to build a dynamic, heterogeneous, redundant and uncertain system architecture, forming a protection architecture that is difficult to detect, penetrate, motivate and utilize attack results, Breaking the static nature, certainty and similarity of the network architecture formed by the attack chain greatly increases the attack cost of the attacker. Specifically, Mimic defense expects to reduce the detectability of the system by increasing the dynamics of the system, increase the randomness of the system and reduce the permeability of the system. It uses the dynamic heterogeneous redundancy architecture of the system to force the attacker to face the problem of collaborative attacks, and comprehensively utilizes the dynamics, randomness and diversity of the system to damage the stability or effective utilization of the attack chain. When deploying and operating the network and system Similarity and statics can significantly improve the difficulty of attacks based on unknown vulnerabilities and backdoors, as well as the availability of attack results. In the decision scheme of the mimic defense system, the majority consensus decision is the most commonly used strategy. The advantage of this adjudication scheme is that it is simple to implement and can handle most abnormal output cases, but the disadvantage is that it cannot give further judgment when it cannot be reached due to the majority inconsistency.

2.3 Data encryption technology

Data encryption technology has a long history. The main principle is to encrypt the original data to make it hidden. The early cryptography encryption technology was simple, mainly based on letter replacement and sequence transformation. The modern cryptography encryption technology is mainly divided into symmetric encryption and asymmetric encryption. The symmetric encryption algorithm is more efficient and faster than the asymmetric encryption algorithm. It is suitable for encrypting a large number of data or files. At present, the most commonly used schemes for encrypting local data are des and AES. There are two keys for asymmetric encryption technology, one for encryption and the other for decryption. The security of asymmetric encryption scheme [35] is often based on mathematical problems, which is also known as public key encryption scheme. In addition, data encryption technology has different branches according to its functions. For example, searchable encryption technology provides ciphertext search function, homomorphic encryption technology can directly calculate ciphertext, and ciphertext access control technology [36]. Among them, homomorphic encryption technology is divided into two categories: partial homomorphic encryption and full homomorphic encryption. Homomorphic encryption refers to the arithmetic operation of ciphertext. After the key is decrypted, the operation result of the corresponding plaintext operation is obtained. Homomorphic encryption, as its name implies, can perform arbitrary operations on the ciphertext for any number of times. It is generally believed that any calculation can be completed by completing addition and multiplication operations. Partial homomorphic encryption means that only a limited number of operations can be performed on ciphertext. For example, RSA encryption algorithm and ECC encryption algorithm are typical multiplication homomorphic encryption. The results obtained by homomorphic multiplication of ciphertext data by the server will be obtained by plaintext multiplication after decryption with the same key, but the addition operation cannot be completed. The homomorphic encryption technology can be said to be a major innovation in cryptography technology. It can be compatible with many other functions through homomorphic operation, such as ciphertext access control technology. However, the efficiency of homomorphic encryption is not high and the difficulty of implementation is an important factor restricting the development of homomorphic encryption technology.

3. SYSTEM MODEL

This section introduces the system model. In this paper, homomorphic encryption technology is applied to the mimic defense system. The specific scheme includes system initialization, data input, policy scheduling, homomorphic attachment, input agent, execution operation, multi-mode/strategy-adjudication, feedback mechanism and output module. The system model is shown in Fig.1.
3.1 System Initialization

Execute the key generation function according to the selected homomorphic encryption scheme. \( \text{SysIn} = str \) is the system input. \( K_e \) is the public key. \( K_r \) is the private key. \( K_e \) is sent to the homomorphic add-on module through the key negotiation mechanism and \( \text{SysIn} \) is input to the input module of the system. Then, Each module is initialized in the system to detect the functional integrity.

3.2 Data Input

After receiving the input information \( \text{SysIn} \), the input module analyzes whether its format meets the system input requirements. If it does not meet the system input format or requirements, it discards the information and returns an error value; If the information meets the requirements, it is transmitted to the policy scheduling module and the homomorphic add-on module respectively, and the input agent module is notified to open the receiving interface to prepare to receive the information of the policy scheduling module and the homomorphic add-on module.

3.3 Policy Scheduling

After receiving the information \( \text{SysIn} \) from input module, the policy scheduling module analyzes the operations to be completed according to the information carried in the information, and determines the number of heterogeneous executors to be selected according to the system security requirements. The heterogeneous executor sets with the above number (set as \( n \)) are called from the heterogeneous pool to form the heterogeneous executor set \( E = \{e_1, e_2, \ldots, e_n\} \) after the scheme is determined. The policy scheduling module send the data \( m_0 \) to be processed obtained after \( \text{SysIn} \) is processed to the input agent module, and then informs the homomorphic add-on modules of the obtained user identity information and the number \( n \) of executors in the heterogeneous set \( E \), and informs the Multi-module/Strategy-adjudication module of the determined decision scheme.

3.4 Homomorphic Add-on

the homomorphic add-on module generates \( n \) random numbers according to the number of executors in the heterogeneous executor set \( E \) sent by the policy scheduling module, and the random numbers are denoted as...
At the same time, the public key $K_u$ is obtained from the user public key pool according to the user information. The homomorphic add-on module packages the information and sends it to the input agent module.

### 3.5 Input Agent

After receiving all the information from the policy scheduling module and the homomorphic add-on module, the input agent module sorts the information into $n(m_0, r, K_u)$ packets. Each packet contains the data $m_0$ to be processed, the allocated random number $r$, and the public key information $K_u$, which are sent to each heterogeneous executor in the heterogeneous set $E = \{e_1, e_2, \ldots, e_n\}$. The input agent module is solely responsible for the input of the executor.

### 3.6 Execution Operation:

After receiving the data $m_0, K_u, r$, the execution module $e_i$ starts to execute the operation on $m_0$, and encrypts the random $r$ with the key $K_u$. The encryption result is recorded as $R_i$. The operation result obtained after $m_0$ operation is recorded as $m_i$. The result of homomorphic operation on $m_i$ and $r$ is recorded as $c_i$, and the operations in the execution in the execution module are arranged as follows:

\[
\begin{align*}
    m_i &= H(m_0) \\
    R_i &= E_{K_u}(r) \\
    c_i &= E_{K_u}(r \otimes m_0)
\end{align*}
\]

$E(\cdot)$ refers to the encryption operation, and the above expression indicates that the encryption operation is performed on $r$ using $K_u$.

The above formula $\otimes$ represents homomorphic operation.

After completing the above operations, the execution module packages the $R_i$ and $c_i$ into a data package $(c_i, R_i)$ and sends it to the multi-module/strategy-adjudication module.

### 3.7 Multi-module/Strategy-adjudication

After obtaining the data package $(c_i, R_i)$ sent by each execution module and performs reverse deconstruction, the multi-module/strategy-adjudication module obtains the random number encryption result $R_i$ and the operation result $c_i$. The multi-module/strategy-adjudication module performs inverse homomorphic operation on the two results, and theoretically obtains the execution result. The result encrypted by the encryption key $K_u$ is recorded as $M_i$. The operation is as follows:

\[
M_i = c_i \oslash R_i = E_{K_u}^{-1}(m_i)
\]

The above formula $\oslash$ represents inverse homomorphic operation.

The above calculation obtains the executor operation result, which is the encrypted expression of execution result. According to the idea of mimic defense, the execution results of each executor should be consistent. The encryption results will be consistent if the execution result is encrypted with the same key and the same method. Based on this consistency, the results can be adjudicated.
Before receiving the result, the multi-module/strategy-adjudication module has obtained the adjudication scheme determined by the policy scheduling module according to the operation requirements and system security requirements. In this system, there is no difference between the usage of the adjudication module and the usage of the adjudication module in the common mimic defense system, except that the adjudication object has changed from the execution result of the executor to the encrypted result of the execution result of the execution body. The common adjudication schemes are as follows:

First, the output result $M_i$ is classified, and the same results in $M_i$ are classified into one class, and finally the classification result $M = \{ \{M_{11}, M_{12}, \ldots, M_{1a}\}, \{M_{21}, M_{22}, \ldots, M_{2b}\}, \ldots, \{M_{p1}, M_{p2}, \ldots, M_{pq}\}\}$ is obtained.

When the consistency judgment is used, the multi-mode/strategy-adjudication module outputs $M_i$ to the output module only the output results $M_i$ of all executors are consistent, that is, when the above classification is $M = \{M_{11}, M_{12}, \ldots, M_{1a}\}$. In other cases, the multi-mode/strategy-adjudication module outputs error information to the other module, and collates the error information and sends it to the scheduling module.

When the majority consensus decision is used, the arbiter finds the one with the largest number among all output result classifications, that is, $m = \text{Max}(a, b, \ldots, q)$ is taken in the classification $M = \{\{M_{11}, M_{12}, \ldots, M_{1a}\}, \{M_{21}, M_{22}\}, \ldots, \{M_{p1}, M_{p2}, \ldots, M_{pq}\}\}$. The multi-mode/strategy-adjudication module sends the result corresponding to the maximum value $m$ to the output module. It is necessary to introduce other priority methods to determine the final result if there are the same number of results and they are the largest. The error information is sorted and sent to the scheduling module after output.

When $k$ out of $n$ is used for decision, a safety factor $k$ is determined according to the safety requirements. Only when the number of the same result classification exceeds $k$, that is, there is $q > k$ in one class $\{M_{p1}, M_{p2}, \ldots, M_{pq}\}$ in the classification, the multi-mode/strategy-adjudication module outputs $M_p$ to the output module. If the classification number of each classification result is less than $k$, the multi-mode/strategy-adjudication module outputs error information to the output module, and collates the error information and sends it to the scheduling module.

3.8 Feedback Mechanism

After the judgment is completed, the multi-mode/strategy-adjudication module sends the random number encryption result $R_e$ corresponding to the obviously wrong result back to the policy scheduling module. The policy scheduling module locates the error executor and executes the corresponding processing measures. This feedback mechanism ensures the confidentiality of the wrong executor and prevents the attacker from discovering the weak links in the execution module.

3.9 Output module

The output module outputs the output information in the multi-mode/strategy-adjudication module to the user after obtaining it.

4. EXPERIMENT VERIFICATION

This chapter gives the specific experimental environment, as shown in Tab 1.
Table 1. Experiment environment

<table>
<thead>
<tr>
<th>Software and hardware specifications</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server CPU</td>
<td>R9 5900hx</td>
</tr>
<tr>
<td>Server RAM</td>
<td>32GB</td>
</tr>
<tr>
<td>Sever System</td>
<td>Windows10 64bit operating system</td>
</tr>
<tr>
<td>SSL/TLS Library</td>
<td>openSSL</td>
</tr>
<tr>
<td>Certification scheme</td>
<td>Wifidog+MySQL</td>
</tr>
<tr>
<td>Router CPU</td>
<td>MT6721A, MIPS dual core 880nhz processor</td>
</tr>
<tr>
<td>Router RAM</td>
<td>256MB</td>
</tr>
<tr>
<td>Router Operating System</td>
<td>OpenWrt</td>
</tr>
<tr>
<td>Homomorphic Encryption Algorithm</td>
<td>RSA algorithm, Elgamal algorithm, paillier algorithm</td>
</tr>
<tr>
<td>Key Length</td>
<td>1024 bits</td>
</tr>
<tr>
<td>Hash Function</td>
<td>SHA-256</td>
</tr>
</tbody>
</table>

The number of heterogeneous executors which perform hash calculation on user data in the experimental simulation system is 5. The function analysis of the scheme is to analyze whether the user information can be accurately hashed and encrypted, and whether the output of the executor and the output module has privacy.

The result comparison diagram of different adjudication schemes is shown in Fig 2. The function comparison of different adjudication schemes is shown in the following Table 2.

<table>
<thead>
<tr>
<th>Randomly generate 18 bit password:</th>
<th>dp4bgWUDZyJuQLSWD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>The output of other methods is:</td>
<td>Transmission scheme plaintext output result</td>
</tr>
<tr>
<td>Output of executor 1 of this scheme:</td>
<td>263900829473589522522854473153272007833754071985504002471590549696720235679025224958453855884898925432723101892969881136445551843738320591</td>
</tr>
<tr>
<td>Output of executor 2 of this scheme:</td>
<td>151023697967383015304633392920559223270368079491455294720945984209173956163047037937194268155976185333371731612315961811353328835300198311622323424307273477365205566272807670839569158755229160721095456320723984900384194569474897324897665979458441518652661565211966928730</td>
</tr>
<tr>
<td>Output of executor 2 of this scheme:</td>
<td>737384963718084091459292760286866635312649773769199172190057580474873802209881679929705435228993411838138738095288499345113895866854051771534198201193597816101140783350592861385988973803912354419866401695999886201300557406848826281236265204155924278400</td>
</tr>
<tr>
<td>Adjudication result of Adjudicator:</td>
<td>12432552900851656198269178544716135339805146164809479996730677473521399221612676523252438283333538310795746779173514833530061962055283839243</td>
</tr>
<tr>
<td>Declassify the verdict:</td>
<td>449d4d9ce7ff165d2653db9306262f0c03e852d0fced6392f534ca3b369c42</td>
</tr>
</tbody>
</table>

Figure 2 Comparison of results of different adjudication schemes
Table 2 Functional comparison of different adjudication schemes

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Number of actuators</th>
<th>Whether the weight is changed</th>
<th>Output results</th>
<th>Adjudication data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static scheduling</td>
<td>5</td>
<td>No</td>
<td>Identical</td>
<td>Plaintext</td>
</tr>
<tr>
<td>Random schedule</td>
<td>3</td>
<td>No</td>
<td>Identical</td>
<td>Plaintext</td>
</tr>
<tr>
<td>Historical confidence</td>
<td>3</td>
<td>Yes</td>
<td>Identical</td>
<td>Plaintext</td>
</tr>
<tr>
<td>degree of isomerism</td>
<td>3</td>
<td>Yes</td>
<td>Identical</td>
<td>Plaintext</td>
</tr>
<tr>
<td>Scheme of this paper</td>
<td>3</td>
<td>No</td>
<td>Different</td>
<td>Ciphertext</td>
</tr>
</tbody>
</table>

As can be seen from table 3-3, other schemes and this scheme all perform hash operation on the same data. The static scheduling scheme calls all the executors every time, and the system form does not cover, so it does not belong to a mimic defense scheme. In this scheme, all heterogeneous executors are used each time, which is the highest hardware cost among all schemes, and it can not place all kinds of attacks. The random schedule scheme extracts 3 executors from the heterogeneous pool each time, which is the most commonly used scheme at present. However, as malicious attackers become more familiar with the system, they gradually master the characteristics of each executor. The failure rate of this scheme will be higher and higher. The adjudication scheme based on historical confidence continuously improves the weight of the executors that output the correct results, while reducing the weight of the executors with high error rate. The decision scheme based on the heterogeneity of executors and the previous scheme can restrain the failure rate of the scheme to a certain extent, but the output result is still plaintext. After obtaining the permission, the internal malicious attacker can still calculate the highly reliable executors according to the statistics of historical output for a period of time to carry out targeted attacks, or directly steal the output data. Compared with the above scheme, the scheme proposed in this paper is also an optimization of the randomly selected scheme. However, the output result of this scheme is ciphertext each time, and the output results of all executors are different, and the authorized internal personnel can not determine which executor is vulnerable or highly trusted according to the historical output. Moreover, the final output result of this scheme is ciphertext. It is meaningless for a malicious attacker to obtain the data without the user's private key, thus protecting the user's data privacy.

5. SUMMARY

As a new idea for the development of network security, mimic defense has been widely concerned and studied by the academic community. Mimic defense introduces heterogeneous and redundant executors, which greatly increases the performance requirements. At present, cloud services are an important carrier of mimic defense. However, when using cloud services, we will inevitably face such problems as the service provider is not trusted, the authority is not controllable, and the user's privacy information is leaked. Homomorphic encryption, as an encryption technology supporting data ciphertext operation, can protect data privacy on the premise of completing data operation requirements. Homomorphic hash technology is an important branch of homomorphic technology and has been widely used in cloud storage scenarios. Aiming at the problem that the cloud service mimic defense system can not protect the privacy of user data, this paper introduces homomorphic encryption technology into the mimic defense system to protect the privacy of user data. This paper presents a homomorphic encryption scheme that only needs to add an additional module, and gives the function and operation mode of all modules in the whole system after adding this module. In terms of security, the proposed scheme can resist external malicious attacks and untrusted server attacks. Experimental results show that the proposed scheme can ensure the privacy of data in the transmission process under the premise of correct execution and correct result. However, homomorphic encryption and mimic defense technologies consume a lot of computing resources. When the scheme proposed in this paper is applied to the actual scenario, it still needs to optimize the operation efficiency.

ACKNOWLEDGEMENT

This work is supported by the Science and Technology Project of State Grid Corporation of China (Research on endogenous immune enhancement technology for secure access gateway of power IoT, No. 5400-202158332A-0-0-00).
REFERENCES

Orientational Information Matters in Trajectory Prediction

Zhenyu Tong\textsuperscript{1a*}, Yue Zhou\textsuperscript{1b*}

\textsuperscript{1}School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
\textsuperscript{a*}\texttt{zachytong@sjtu.edu.cn}, \textsuperscript{b*}\texttt{zhouyue@sjtu.edu.cn}

Abstract

Trajectory prediction is crucial for collision avoidance and path planning in the autonomous driving task stacks. Numerous models have been proposed to learn the complicated spatiotemporal and social interactions among traffic agents. However, we find that some models perform badly in some corner scenarios, and it is related to insufficient supervision using single mean squared error (MSE) loss. Therefore, we propose the orientational MSE (OMSE) loss for better trajectory prediction. We introduce the global proxy matching strategy in OMSE loss to enable model to optimize with supervision from both Euclidean distance and orientation information. We evaluate OMSE loss on two public trajectory datasets for pedestrians and vehicles. Both quantitative and qualitative results illustrate the improvement on forecasting performance on cases of interest. The proposed method works in a learnable manner and surpasses traditional post-processing methods with more flexibility.

Keywords-trajectory prediction; mean squared error; cosine distance; autonomous driving

1. Introduction

Trajectory prediction is an important task in autonomous driving system. It bridges the gap between downstream planning and upstream perception tasks. It can be used for collision avoidance and pedestrian protection in safety-critical scenarios.

Previous works focus on designing new architectures to model spatiotemporal relations and social interactions. The forecasting abilities of different networks, like CNNs, LSTMs, and GCNs, have been explored in [1], [2], [3], with the Transformer [4] becoming the favorites of researchers [5], [6]. Also, multiple generative models including CVAEs and GANs have been studied to incorporate the innate multi-modality of the task [7], [8]. Besides, numerous mechanisms have been proposed to account for sophisticated social interactions [9], [10], [11].

However, less attention has been paid to loss functions for better trajectory distance measurement. The quality of the predicted trajectory is evaluated mainly by trajectory distance loss, which is dominated by Mean Squared Error (MSE). Several variants of MSE have also been applied, such as mean absolute error [12] and Huber loss [13]. Surprisingly, we found that models make worse predictions over simple samples. As shown in Fig. 1, $\mathbf{Y}_{n}^{(1)}$ has the lowest MSE value $d_1$ among the three predicted points. But $\mathbf{Y}_{n}^{(1)}$ is not necessarily a better prediction than $\mathbf{Y}_{n}^{(2)}$ given the orientation of the trajectory. The red dotted circle marks the positions with same MSE to $\mathbf{Y}_{n}^{(1)}$ in Fig. 1.

To this end, we propose a novel loss called orientational MSE (OMSE) loss to encourage the model to cover corner cases mentioned above during training. The OMSE loss includes an orientational term based on cosine distance. Compared with pure Euclidean distance, cosine distance is more sensitive about the orientational difference. Therefore, we include cosine distance to measure pairwise distance between predicted and ground truth positional vectors with respect to trajectory centroid.
To summarize, the main contributions of this paper are:

1. We propose the orientational mean squared error (OMSE) loss for better trajectory predictions, especially over corner cases where orientation of trajectory matters more than spatial distance.

2. The OMSE loss is model-agnostic and can be plugged to different models without modifying the model structure.

3. The proposed method works in a learnable manner and is more flexible and extensible than handcrafted post-processing methods.

2. Orientational Mean Squared Error Loss

2.1. Problem Formulation

By convention, we formulate trajectory prediction as a problem of predicting future trajectories given history trajectories. The trajectory of the agent \( n \) over past \( T_b \) timesteps (current moment included) is denoted as \( \mathbf{X}_n = (\mathbf{x}^{-T_b}, \mathbf{x}^{-T_b+1}, \ldots, \mathbf{x}^0) \), where \( \mathbf{x}^0 \) is the position of the agent \( n \) at the current moment. Similarly, denote the trajectory in future \( T_f \) timesteps of agent \( n \) as \( \mathbf{Y}_n = (\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^{T_f}) \). For agent \( n \) at time \( t \), we have \( \mathbf{x}^t_n \in \mathbb{R}^d \) when \( t < 0 \) and \( \mathbf{y}^t_n \in \mathbb{R}^d \) when \( t > 0 \). In some works, input and output of prediction model may contain more information like angle besides the two-dimensional positions. In this paper, we set \( d_p = d_f = 2 \) for simplicity. The predicted future trajectory of the agent \( n \) is defined as \( \tilde{\mathbf{Y}}_n = (\tilde{\mathbf{y}}^1_n, \tilde{\mathbf{y}}^2_n, \ldots, \tilde{\mathbf{y}}^{T_f}_n) \). Naturally, predictions and true trajectories of all agents can be expressed as \( \tilde{\mathbf{Y}} \) and \( \mathbf{Y} \) respectively.

2.2. Global Proxy Match and Orientational Mean Squared Error

An ideal point match method is assumed to contain as much structure information as possible and have lower match complexity at the same time. The larger the paired number, the higher the match complexity. Pairwise match tries to make a trade-off between the two aspects. Inspired by the pooling operations, we propose the global proxy match (GPM) strategy that matches a predicted point with both its reference point and the centroid, as shown in Fig.2.

GPM can be formulated as

\[
L_{gp}(\tilde{\mathbf{Y}}^n_n, \mathbf{Y}_n) = M(\tilde{\mathbf{Y}}^n_n, \mathbf{Y}_n) + M(\tilde{\mathbf{Y}}^n_n, \mathbf{Y}_n_{\text{th}}),
\]

where \( M(\cdot) \) denotes a matched trajectory point pair and \( \mathbf{Y}_n_{\text{th}} \) is the global proxy of \( \mathbf{Y}_n. \mathbf{Y}_n_{\text{th}} \) is the centroid of the trajectory \( \mathbf{Y}_n. \tilde{\mathbf{Y}}_n \) is computed as
GPM will pair points locally and globally, with local pairs $M(\mathbf{y}_n^t, \mathbf{y}_n^t)$ and global pairs $M(\mathbf{v}_n^t, \mathbf{v}_n^t)$. One predicted point is matched with the centroid as an indirect yet economical way to supplement global orientation information.

GPM computes orientational information via centroid with a match complexity of $2T_f$, which is acceptable compared with $T_f$ of pairwise match and $T_f^2$ of one-to-all match.

The orientational mean squared error (OMSE) loss is defined in the following formula:

$$L_{omse}(\mathbf{y}_n, \mathbf{y}_n) = L_{mse}(\mathbf{y}_n, \mathbf{y}_n) + \lambda L_o(\mathbf{y}_n, \mathbf{y}_n).$$

$$L_{mse}(\mathbf{y}_n, \mathbf{y}_n) = \frac{1}{T_f} \sum_{t=1}^{T_f} | \mathbf{y}_n^t - \mathbf{y}_n^t |^2,$$

$$L_o = \frac{1}{T_f} \left( 1 - \frac{\langle \mathbf{y}_n^t - \mathbf{y}_n^t \rangle \cdot (\mathbf{v}_n^t - \mathbf{v}_n^t)}{| \mathbf{y}_n^t - \mathbf{v}_n^t | \cdot | \mathbf{v}_n^t - \mathbf{v}_n^t |} \right).$$

In (5), $\lambda$ is a hyperparameter balancing the local term and the orientational term.

The local distance refers to the MSE between predicted and real points. It acts as the main force to supervise the model at the beginning phase. With training process going on, the model is able to make better predictions and yield smaller MSE loss in general. The gap between good and bad prediction is narrowed from the perspective of MSE. This is when the orientational distance cut in and work. For example, the predicted trajectory is close to the true trajectory measured in MSE but have different orientation. The orientational term can provide enough supervision signals for further learning.

### 3. Experiments and Discussions

#### 3.1. Datasets and Metrics

The **ETH-UCY** dataset is composed of ETH and UCY [14], [15]. It is typically used as the major benchmark for pedestrian trajectory prediction. There are five scenes filmed from bird's-eye view, and the trajectories were sampled with a rate of 2.5Hz in crowded scenarios. We predict trajectories of 12 timesteps (4.8s) in the future with given history trajectory of 8 timesteps (3.2s). **nuScenes** [16] is a popular dataset of trajectory prediction for autonomous driving. It includes 1000 scenes and each of them is 20 seconds long. For nuScenes dataset, we forecast future 12 timesteps (6s) based on observed 4 timesteps (2s) with the same setting from [5].

We use $\min ADE_K$ and $\min FDE_K$ of $K$ possible trajectories as metrics. The two metrics are defined as

$$\min ADE_K = \frac{1}{T_f} \min_K \sum_{t=1}^{T_f} | \mathbf{v}_n^{(k)} - \mathbf{y}_n^t |^2,$$

$$\min FDE_K = \min_K \sum_{t=1}^{T_f} | \mathbf{v}_n^{(k)} - \mathbf{y}_n^t |^2.$$

The notations are coherent with definitions in problem formulation. We set $K = 20$ for ETH-UCY dataset and $K = 5, 10$ for nuScenes dataset.

#### 3.2. Experiments

We choose AgentFormer as the baseline model to evaluate OMSE loss. AgentFormer is a Transformer-based CVAE model for trajectory prediction. It designs and utilizes novel attention mechanism for social and temporal modelling of trajectory sequence. There is no scenario-specific consideration involved. It is suitable for trajectory prediction of both pedestrian and vehicles. We replace the MSE loss in its Future Decoder module with OMSE loss and retrain the model with the same setting. We do not change the loss in its trajectory sampler module. More details about the model and training process can be found in the original paper. Different numbers of $\lambda$ have been tried, and the results are achieved when $\lambda = 0.1$. 

\[ \mathbf{y}_n = \frac{1}{T_f} \sum_{t=1}^{T_f} \mathbf{y}_n^t. \]
We compare the retrained AgentFormer, which is denoted as AF (OMSE), against common baselines in two different tracks: pedestrian trajectory prediction and vehicle trajectory prediction respectively. On the nuScenes dataset, we evaluate our loss against the AF (original), Trajectron++ [17] and CoverNet [18]. All the results are reported in Table 1. The performance on average displacement error and final displacement error are improved for 5 samples setting.

Table 1: Baseline comparisons on the nuScenes

<table>
<thead>
<tr>
<th>Method</th>
<th>K = 5</th>
<th></th>
<th>K = 10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADE_FDE</td>
<td></td>
<td>ADE_FDE</td>
<td></td>
</tr>
<tr>
<td>CoverNet</td>
<td>1.96</td>
<td>-</td>
<td>1.48</td>
<td>-</td>
</tr>
<tr>
<td>Trajectron++</td>
<td>1.88</td>
<td>-</td>
<td>1.51</td>
<td>-</td>
</tr>
<tr>
<td>AF (original)</td>
<td>1.86</td>
<td>3.89</td>
<td>1.45</td>
<td>2.86</td>
</tr>
<tr>
<td>AF (OMSE)</td>
<td>1.83</td>
<td>3.83</td>
<td>1.50</td>
<td>2.86</td>
</tr>
</tbody>
</table>

On the ETH-UCY dataset, we compare the AF (OMSE) with Trajectron++, TF [6], PECNet [19], and the original AgentFormer. The performance of baselines is summarized in Table 2. Compared with the AF (original), performance of AF (OMSE) improves in the UCY and drops in the ETH dataset. The performance variance on other sub-datasets is not obvious.

Table 2: Baseline comparisons on the ETH-UCY

<table>
<thead>
<tr>
<th>Method</th>
<th>ADE/FDE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ETH</td>
</tr>
<tr>
<td>TF</td>
<td>0.61/1.12</td>
</tr>
<tr>
<td>PECNet</td>
<td>0.54/0.87</td>
</tr>
<tr>
<td>AF (original)</td>
<td><strong>0.45/0.75</strong></td>
</tr>
<tr>
<td>AF (OMSE)</td>
<td>0.48/0.76</td>
</tr>
</tbody>
</table>

3.3. Discussions

The quantitative result shows that the OMSE loss can improve the overall performance slightly. The qualitative result in Fig. 3. shows significant improvement on scenarios of interest. The trajectories in the first and the second row are framed in the x-y axis from the camera image view and the bird-eye view respectively. These trajectories often contain subtle orientational information due to movement like turning.

![Fig. 3 Comparisons of typical scenarios from ETH-UCY (top row) and nuScenes (bottom row)](image)

To be more specific, performance improvement on two datasets is not the same due to the different patterns in pedestrian and vehicle motions. Besides, the hyperparameter will affect the performance. We show the variance on ADE_F and FDE_F with different $\lambda$ in Fig. 4. The model achieves the best performance when $\lambda = 1$. The performance drops when value of $\lambda$ is too large. We think this may be caused by the actual range of trajectories in driving environment.
4. Conclusion

In this paper, we find models make bad predictions over corner cases due to orientation information loss in optimization using MSE. We propose the orientational mean squared error loss, where we use a new matching strategy called global proxy match and a new distance to incorporate orientational information of trajectory. We conduct experiments on two mainstream datasets. The results demonstrate that OMSE loss can improve model’s performance on cases of interest greatly in a learnable way. Compared with traditional complex post-processing methods, the proposed method shows more flexibility and extensibility.

References


Database design of load spectrum calculation software for wheeled crane

Changjian Zhu¹,a, Xingchen Zhang¹,b, Zhengde Zhang²,c, Yuguo Wang¹,d, Ruwen Chen¹,e, Kang Wu¹,f
¹School of Automotive and Rail Transit, Nanjing Institute of Technology, Jiangsu, China
²Xuzhou Heavy Machinery Co., Ltd. Jiangsu, China
azhuchangjian@njit.edu.cn, bzhangxc1225@163.com, cqzjzzd@163.com,
dwangyugo@njit.edu.cn, ejscrw@njit.edu.cn, fwnjicit163.com

Abstract

The software development of wheeled crane load spectrum calculation will involve a series of database design problems, especially the characteristics of many data table types, and large number of fields. Based on the above situation, a method of batch and fast database creation based on high-level programming language is proposed. Firstly, the requirement analysis and detailed design of the wheel crane load spectrum database are carried out, then the different database creation methods are analyzed and compared, the specific database creation algorithm flow is given, and finally the database creation software is developed in C++ language and a running example is given. The proposed database design method is efficient and labor-saving, and is conducive to the rapid construction of data tables for large number of fields in the crane load spectrum.

Keywords—Wheeled crane, Load spectrum, Database, Design

1. INTRODUCTION

At present, wheeled cranes basically adopt the design method with the maximum load in the design reference period as the standard load design value, which can ensure reliable product work during the life cycle, but the design method is prone to over-design and increase manufacturing and use costs[1-3]. Therefore, it is necessary to master the wheeled crane load spectrum, which is the key to achieve fatigue residual life assessment and reliability analysis of wheeled cranes[4].

Wheeled crane load spectrum refers to the relationship between the load and its cumulative frequency of occurrence of each component under various operating conditions, which is used to reflect the load-time course of the actual working process of the product, which is the prerequisite and basic data for fatigue residual life prediction, fatigue strength analysis and calculation, reliability design, and simulation test loading[5].

To compile a perfect and reliable load spectrum, it is necessary to calculate and analyze through the massive wheeled crane IoT data, which cannot be done without the application of database[6]. In the face of massive amounts of different types of IoT data, the general data storage method will lead to inconvenient data usage[7], so how to build an efficient wheeled crane load spectrum database that provides data storage, fast query, data modification and other functions is the key issue[8].

Database technology from the late 1960s, through the hierarchical database, mesh database and relational database and into the database management system (DBMS) stage so far, the database technology research has also continued to make progress[9]. In recent years, with the development of Web 2.0 technology on the Internet, some new changes have emerged in network database applications, which have increasingly failed to meet the needs of the Internet for data expansion, read and write speed, support capacity, and construction and operation costs[10], and new requirements have been put forward for database technology indicators such as data model, distributed architecture, and data storage related to this. no SQL is This new changes, new requirements under the output of a non-relational database products in general, with easy to scale, large data volume, high performance, flexible data model, high availability and other characteristics[11], can be divided into key-value storage database, column storage database, document-based storage database, graph database, etc.

In view of the stable table structure of the wheel crane load spectrum database, simple data model, infrequent demand changes, and the current mainstream database software MySQL, Oracle, etc. to create tables and fields in bulk is more cumbersome, so for the calculation of the wheel crane load spectrum needs, the wheel crane load spectrum database requirements analysis, modeling design and software development implementation, etc., proposed a The method of batch
and rapid creation of database based on high-level programming language solves the database design problems involved in the software development of wheel crane load spectrum calculation, such as many data table types and large amount of fields.

2. Requirement Analysis of Load Spectrum Calculation Software for Wheeled Crane

Using the functional decomposition method and object-oriented analysis method, the functional requirements, non-functional requirements and design constraints of the crane load spectrum calculation software are studied to establish the requirement model of the calculation software, including logical view, physical view and data view, and the database relationship diagram is shown in Figure 1.

There are many types of data to be processed by the load spectrum of wheeled crane, and the data volume is huge. There are several difficulties to store the data:

Firstly, according to the load spectrum calculation requirements of different crane models, the number of fields and data types are different, so it is difficult to use the data table with the same structure for storage.

Secondly, each table has a large number of fields and complex relationships. Complex tables often involve hundreds of fields. For example, there are dozens of tables and thousands of fields in the 25t crane airborne load spectrum database alone. The amount of data is very large, as shown in table 1 (x in the table represents the tonnage of cranes, * refers to the system sequence of the table, the same below).

Thirdly, it must be able to quickly read and store a large amount of original data, and the calculation results should be easy to view and analyze.

In this case, the data architecture should be clear and organized, and the data storage should be classified and placed with traces to follow. It is necessary to use database software for data storage to achieve the following purposes: data sharing, reducing data redundancy, maintaining data independence, data centralized control, data consistency and maintainability, and fault recovery.

<table>
<thead>
<tr>
<th>No.</th>
<th>Type</th>
<th>Number</th>
<th>Fields</th>
<th>Volume</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Model parameter</td>
<td>1</td>
<td>8</td>
<td>small</td>
<td>a_Xt_parameter</td>
</tr>
<tr>
<td>2</td>
<td>System configuration</td>
<td>1</td>
<td>4</td>
<td>small</td>
<td>b_configuration</td>
</tr>
<tr>
<td>3</td>
<td>Original data</td>
<td>1</td>
<td>23</td>
<td>big</td>
<td>c_Xt_data</td>
</tr>
<tr>
<td>4</td>
<td>Calculation results</td>
<td>66</td>
<td>2001</td>
<td>big</td>
<td>d_Xt_*</td>
</tr>
<tr>
<td>5</td>
<td>User authority</td>
<td>1</td>
<td>6</td>
<td>medium</td>
<td>e_role</td>
</tr>
<tr>
<td>6</td>
<td>Operation log</td>
<td>1</td>
<td>9</td>
<td>medium</td>
<td>f_log</td>
</tr>
</tbody>
</table>

TABLE 1. 25T CRANE DATABASE
3. Database Modeling of Load Spectrum Calculation Software for Wheeled Crane

Based on the demand model of crane load spectrum calculation software, a reasonable software architecture is studied and determined by using the data flow oriented design method, including the calling relationship between modules, the interface relationship between modules, the database structure, etc.

In this paper, the load spectrum database of wheeled crane is divided into three categories: system configuration table, calculation result table and original data table. The system configuration table is composed of crane model parameter table, software system configuration table, user authority table and operation log table; The original data sheet consists of the original data sheet of 25t crane, 300t crane, etc; The calculation result table consists of the calculation result table of the whole vehicle, the calculation result table of the telescopic system, the calculation result table of the luffing system, the calculation result table of the hoisting system and the calculation result table of the slewing system. The specific structure is shown in Figure 2.

After obtaining the relevant data stored in the original data table, such as VIN code, vehicle model and time, through the relevant parameters of the system configuration table, such as crane model parameters and software system configuration information, the load spectrum calculation software of the wheel crane will calculate the load spectrum and finally store it in the calculation result table.

![Database hierarchy diagram](image)

3.1 Design of System Configuration Table

Taking into account the difference in parameters brought about by different crane models, calculation software versions, user operation rights and operation logs and other requirements, the system configuration table therefore contains a crane model parameter table, software system configuration table, user rights table and operation log table.

Because of the need to store the crane parameters of each model, the table name of the crane parameters table is distinguished by a_Xt_parameter, and the table stores information such as ID, parameter name, value, unit, parameter system, parameter type, correspondence and remarks.

Software system configuration table is distinguished by b_configuration, storing information of each software system configuration: such as ID, parameter name, parameter value and comments.

The table name of the user rights table is distinguished by e_role, storing user information: such as ID, user name, user password, user rights, recent login IP and recent login time, etc.
The table name of operation log table is distinguished by f_log, storing operation log information: such as operation time, user name, user type, IP address, operation command, operation object, information before modification, information after modification and comments, etc.

3.2 Design of Original Data Table

The raw data table is distinguished according to vehicle models, including 25t crane raw data table, 300t crane raw data table, 1600t crane raw data table, etc.

The table name of raw data table is distinguished by c_Xt_data, storing IOT raw data information: such as VIN code, vehicle model, time, area, ACC status, oil pressure, water temperature, engine speed, running time, vehicle speed, actual lifting weight, rated lifting weight, length, angle, working range, working condition code, multiplier, torque percentage, force limiter fault code, control class fault code, Engine fault code, binding status, locking status, etc.

3.3 Design of Calculation Result Table

According to the working characteristics of wheeled cranes, the calculation result table is divided into five types of tables according to the whole vehicle, telescopic system, luffing system, hoisting system and slewing system, and the tables are divided into different types of tables according to the information stored in the calculation results, such as crane condition table, telescopic system condition table, luffing system condition table, hoisting system condition table and slewing system condition table. Each table is named according to the way of d_Xt_* , storing information such as VIN code, vehicle model, date and calculation result.

4. Establishment of Load Spectrum Calculation Software Database for Wheeled Crane

4.1 Research on the Method of Database Programming

According to the database model of the above wheeled crane load spectrum calculation software, MySQL 5.7 is finally selected as the load spectrum database software considering the amount of data of calculation results, user installation and operation, software development cost, etc. At present, there are two main methods to create MySQL database:

Firstly, created by running SQL statements directly through MySQL software. Secondly, created by visualization software such as Navicat, SQLyog, etc. to manipulate MySQL.

In view of the crane load spectrum database calculation results table has a large number of tables, field names complex and other difficulties and post-maintenance problems, this paper proposes a method to create a database using a high-level programming language, the method can create a more special field table, but also a large number of similar field tables, high efficiency, and not easy to make mistakes, easy to maintain later.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>MySQL</td>
<td>Simple and direct</td>
<td>Hard to create many tables</td>
</tr>
<tr>
<td></td>
<td>Reduce data transfer</td>
<td>More manual errors</td>
</tr>
<tr>
<td></td>
<td>High database security</td>
<td>Difficult to develop and maintain</td>
</tr>
<tr>
<td></td>
<td>Visual operation, convenient</td>
<td>Inefficient to create many tables</td>
</tr>
<tr>
<td></td>
<td>Import, export other formats</td>
<td>More manual errors</td>
</tr>
<tr>
<td></td>
<td>Backup SQL files for restore</td>
<td>Database interface expose</td>
</tr>
<tr>
<td></td>
<td>Easy to create similar tables</td>
<td>Creating a single form is hard</td>
</tr>
<tr>
<td>Programming</td>
<td>Fewer manual errors</td>
<td>Database interface will expose</td>
</tr>
<tr>
<td></td>
<td>Facilitate database restore</td>
<td>Cannot migrate data</td>
</tr>
</tbody>
</table>

4.2 Database Creation Algorithm Flow

Based on the research on the design results and creation methods of the load spectrum database model of wheeled cranes, after selecting the method of creating the database using high-level programming language, first find the common points of each table and its fields, and uniformly configure them, such as creation statements, engine statements, etc; Then find the differences between each table and its fields, and parameterize them, including field name, field type, default value, remarks, primary key and other information; Finally, according to the design of each table model, the parameterized information is input and configured to complete the software development.

The algorithm flow chart of database creation software is shown in Figure 3 (where n is the number of tables and K is the
number of fields). The specific algorithm flow is as follows:

Firstly, as shown in Figure 3(a), after clicking the create table command, create a file with different configuration information written according to the number and type of the selected tables. The configuration information is shown in Figure 3(b), save the file to the specified location, and provide the function of opening the file.

Secondly, as shown in Figure 3(b), write the table name, configuration information of each field, judge the existence statement, create statement, and write different create statements of each field into the file, as shown in Figure 3(c).

Thirdly, as shown in Figure 3(c), according to the configuration information of each field, write different creation statements of each field into the file, including field name, field type, default value, remarks, primary key and other information.

4.3 Database Creation Software and Examples

Based on the database creation software algorithm process, the database creation software is developed on Visual Studio 2019 platform using C++ language, and its interface is shown in Figure 4, and the SQL script file can be generated automatically by clicking "Create" after selecting the corresponding form. The SQL script file is shown in Figure 5.

The original data table stores the original data of all models, and the data is summarized into one table every half month, the number of records in each table ranges from tens of millions to billions. The time required to retrieve a specific field of a specific row on a laptop computer is less than 0.1 second, as shown in Figure 6.

This software solves the problem of needing to enter duplicate names during field creation, such as field names with the same English part and a regular distribution of numeric parts; it also facilitates developers to make changes at different places and adapt to fields with large differences, which improves work efficiency, reduces workload, and lowers the possibility of manual errors.

![Fig.3 Algorithm flow chart of database creation software](image-url)
In this paper, according to the calculation requirements of the load spectrum of wheeled crane, the database is modeled and designed, and a database creation method using high-level programming language is proposed, which can create tables with similar fields in large quantities, with high efficiency and not easy to make mistakes. It is especially suitable for software development scenarios with frequent demand changes.

Firstly, in the aspect of database modeling design, based on the actual requirements of the wheel crane load spectrum database, the hierarchical division, naming rules and field information of the database table are designed according to the system structure of the whole machine, telescopic, slewing, luffing, hoisting, etc.

Secondly, in the research of database model creation method, by analyzing and comparing different database creation methods, this paper puts forward a method of creating database using high-level programming language, and gives the specific algorithm flow of database creation, including unified configuration of the similarities of tables and fields, parametric programming of the differences, etc.

Thirdly, in the aspect of database model implementation, the database creation software is developed on the visual studio 2019 platform using c++ language, which verifies the feasibility of the database creation method using high-level programming language.

ACKNOWLEDGMENT

This work is supported by School-enterprise cooperation projects (3612411220238).
REFERENCES

Dialog History Management for End-to-End Task-oriented Dialog Systems

Shijia Nie 1a, Guanjun Li2b, Xuesong Zhang1c, Dawei Zhang2d, Jianhua Tao2e*, Zhao Lv1f*
1 School of Computer Science and Technology, Anhui University, Hefei 230601, China
2 NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing 100098, China
ashijia.nie@nlpr.ia.ac.cn, bguanjun.li@nlpr.ia.ac.cn, cxuesong.zhang@nlpr.ia.ac.cn, 
dawei.zhang@ia.ac.cn, ejhtao@nlpr.ia.ac.cn, fjkjlz@ahu.edu.cn

Abstract

End-to-end task-oriented dialogue systems rely heavily on an understanding of the dialog history. This often faces the challenge of inferring which dialog history information is critical to generating responses. In this paper, we address this challenge by leveraging a dialog history manager component that dynamically focuses on dialog history memory. It performs multiple add and forget operations by fusing an enhanced entity representation of dialog history and knowledge base (KB) information as queries, remembering entities relevant to responses and filtering out unimportant information. Experimental results on an open task-oriented dialogue dataset show that our model outperforms the baseline system in terms of effectiveness and produces contextually consistent responses.

Keywords: task-oriented dialogue systems; attention; multi-hop.

1. Introduction

Task-oriented dialog systems use natural language interaction to obtain user requirements and fulfill user-specific goals. It has received increasing attention due to its wide range of applications, such as restaurant reservations, calendar scheduling, navigation queries, etc. Traditional task-oriented dialog systems are implemented using a pipeline-based [1] which consists of modularly connected components for natural language understanding, dialogue management and natural language generation [2-3]. This construction method suffers from the problem of error transfer caused by module cascading and requires a large number of hand-crafted features and labels.

To improve the limitations of pipeline-based task-oriented dialog systems, related scholars have proposed end-to-end approaches based on deep learning [4-6], which directly map dialogue histories to output responses, effectively alleviating the drawbacks of pipeline-based approaches and attracting widespread attention from the research community. End-to-end approaches in the literature typically employ sequence-to-sequence (Seq2Seq) models to generate responses from dialogue histories and knowledge bases (KBs) [7-10]. Taking the dialog in as an example, the end-to-end dialogue system gives the dialog history and the corresponding KB in order to answer the driver's query about "shopping mall", and generates the system response directly. Task-oriented dialog systems rely heavily on dialogue histories, in order to fully understand dialogue histories, and perform knowledge-based reasoning, The Global-to-local Memory Pointer Networks for Task-oriented Dialogue (GLMP) [11] method has been proposed a combination of global, local memory pointers and replication mechanisms to extract KBs. Besides, the Effectively encoding knowledge for end-to-end task-oriented dialog (Fg2seq) method [12] has also been proposed a 'flow-to-graph' sequence model to efficiently encode knowledge by taking into account the inherent structural information of the knowledge graph and the underlying semantic information in the dialog history.
Despite the validity of previous research, there are a number of technical challenges in modelling multi-turn task-oriented
dialog systems. Existing end-to-end approaches string together the current turn of user utterance and all previous turns of
dialogue utterance as a dialog history, which introduces a lot of noise due to lengthy and redundant contexts [13], making
the dialog system difficult to effectively select the right entities for response generation, resulting in unstable response
generation. Unlike the open domain dialog system, this problem is particularly prominent in the task-oriented scenario, as
the entities in the dialog history are usually the ones expected in the response. In the case of the dialogue in , for example,
the dialog history contains a large amount of information that is not relevant to the generation of responses when answering
the user's question in turn 5, it is difficult for the model to select "Stanford shop

To alleviate the above limitations, we propose a dialog history manager that uses a multi-hop mechanism, with each hop
employing add and forget approach to dialog history information. Specifically, we use the dialog history manager to
efficiently maintain the historical utterance in the dialog history memory, which is kept up to date at each time step within
the dialog history, tracking the flow of historical information and capturing the entities needed to generate responses in the
current turn of the dialog history.

2. Proposed Model

The dialog context is represented by \( C = \{c_1, c_2, ..., c_T\} \), where \( T \) is the number of dialogue turns and each \( c_i \) consists of
one user utterance \( u_i \) and one system response \( s_i \). The KB is represented by a knowledge graph \( G \), which consists of a
series of triples \( \{\text{Subject}, \text{Relation}, \text{Object}\} \), where \( \text{Subject}, \text{Object} \in E \), \( \text{Relation} \in R \). \( E \) and \( R \) are represented as the set
of entities and the set of relations in the KB respectively. Dialog responses at time step \( t \) is represented as \( Y = \{y_1, y_2, ..., y_T\} \), \( t \in [1, n] \), where \( n \) is the number of words to generate responses. The model uses an encoder-decoder
architecture overall, with the dialog history \( D = \{c_1, c_2, ..., c_{t-1}, u_t\} \) and the knowledge graph \( G \) as input \( X = \{D, G\} \) in t-
th turn. The goal of the task-oriented dialog system is to optimize the probability of generating $Y$ conditional on $X$, i.e., $P(Y|X)$.

Our model is composed of an encoder module, a dialogue history manager module, and a decoder module. First, the encoder module consists of a context encoder and a KB encoder, the context encoder module encodes the dialog history at word-level and sentence-level to learn fixed length vectors as contextual dialog representations. Then, the contextual dialog representation is input to the knowledge base encoder module to obtain a representation of each entity in the knowledge graph. The contextual dialogue representation is then written to the dialogue history manager module to update the contextual dialogue representation and filter irrelevant dialogue histories. Finally, the decoder module connects the context and knowledge representation to generate the final system response. Fig. 2 illustrates the general architecture of the model, and in the following sections we describe each component in detail.

2.1. Encoder module

As shown in Fig. 3, the encoder is divided into a context encoder and a KB encoder, which extract features from the dialog history and KB entities respectively.

**Context Encoder:** Each user utterance $u_i$ and system response $s_i$ is composed of a series of tokens, and we denote the j-th token in the i-th sentence as $w_{ij}$ $i \in [1, L], j \in [1, m_i]$ , $L$ being the number of sentences, $m_i$ being the number of tokens in the i-th sentence. Each input token in the dialogue context is converted into a fixed-length vector by the embedding layer. The input embedding sequence is then entered into a hierarchical attention network \[15\] to encode the dialog history and learn the contextual representation of the dialog. A single layer Bi-GRU \[16\] is passed through in order of the words in the dialog history:

$$h_{ij} = BiGRU(e(w_{ij}), h_{ij-1})$$ \hspace{1cm} (1)

where $e(w_{ij})$ is the embedding representation of the token $w_{ij}$, we denote the hidden state of the i-th sentence tokens as $H_i = \{h_{i,1}, h_{i,2}, ..., h_{i,m_i}\}$, and we denote the hidden state of the all tokens as $H_s = \{H_1, H_2, ..., H_L\}$. Not all tokens play an equal role in expressing the meaning of the sentence, so we introduce an attention mechanism to extract the words important to the meaning of the sentence and aggregate the representations of these informative words to form the sentence vector $s_i$:

$$a_{ij} = \text{softmax} \left( V_a^T \tanh (W_a h_{ij} + b_a) \right)$$ \hspace{1cm} (2)

$$s_i = \sum_j a_{ij} h_{ij}$$ \hspace{1cm} (3)

where $V_a, W_a, b_a$ are learnable parameters, and then the sentence vector $s_i$ is encoded using Bi-GRU to obtain the context vector.

$$h_i = BiGRU(s_i, h_{i-1})$$ \hspace{1cm} (4)
where $h_{ij}$ is the stitching of the forward and backward hidden states of the encoder's output, and the $h_i$ concatenation is represented as the final context vector, denoted as $H_c = \{h_t, h_2, ..., h_L\}$

**KB Encoder:** We use a "flow to graph" (FG2Seq) approach proposed by Zhenhao He et al [12] to encode the KB, because the use of knowledge graphs to represent the KB is beneficial for knowledge inference, considering the rich structural information between entities in the KB. The representation of the KB is obtained in conjunction with the dialog history, and a flow-to-graph operation is used to facilitate the understanding of the dialog history.

**2.2. Dialog History Manager module**

Inspired by Jian Wang et al [17], the dialog history manager is initialized using a context vector generated by the context encoder. The dialog history manager is used to maintain the dialog history memory, which is kept up to date at each decoding time step. Capture the important information needed at the current time step by using the dialog history manager to dynamically track the flow of dialog history information.

We use a standard gated recurrent unit (GRU) [18] as the decoder to generate system responses word-by-word. At each decoder time step $t$, the GRU takes as input the previously generated word $y_{t-1}$ and the hidden state $h_{t-1}$ from the previous time step and generates a new hidden state $h_t$ using the following method:

$$h_t = GRU(y_{t-1}, h_{t-1})$$ (5)

The model cycles through the dialog history memory $K$ times, using $h_t$ as an initial query vector $q_t^k$ for attention to the dialog history memory at each hop $k$, and reads the weighted context representation $o^k_t$ from the dialog history memory, and then updates the dialog history memory using the query vector $q_t$ and the dialog history memory reads $o^k$, formally, $C = \{C^0, C^1, ..., C^K\}$, where $C^0$ is the context encoder output $h_L$, $C^1 \in \mathbb{R}^{|V| \times d_{emb}}$, $|V|$ is the number of tokens, and $d_{emb}$ is the dimension of each word vector. The specific dialogue history memory read and update operations are described below.

**Dialog History Memory Reading:** At time step $t$, we use the attention mechanism to obtain the correlation between the query vector $q_t^k$ and the memory $C^k$, i.e., the soft attention weight $a^k_t$. The soft attention weight $a^k_t$ for the $j$-th value is calculated as follows:

$$a^k_{ij} = \text{softmax}(V^k_t \tanh(W^k_t[q^k_t, C^k_{j,t}]))$$ (6)
where $V^K$ and $W^K$ are learnable parameters, the model then weights the dialog history memory $C_t^k$ by the soft attention weights $o_{j,t}^k$ and reads the memory $e_{j,t}^k$:

$$o_{j,t}^k = \sum_{j}^{[v]} a_{j,t}^k C_{j,t}$$  \hspace{1cm} (7)

**Dialog History Memory Update:** Dialog history memory update is inspired by Fandong Meng et al [19-20], where we use both adding and forgetting to update the dialog history memory. The forget operation determines how much of the current message should be removed from the dialog history memory, and similarly, the add operation determines how much of the current message should be written to the dialog history memory as an addition.

We first use another GRU to mimic the k-hop decoder and obtain the intermediate hidden history $\tilde{h}_t^k$. The calculation is as follows:

$$\tilde{h}_t^k = GRU(q_t^k, o_{j,t}^k)$$  \hspace{1cm} (8)

where $\tilde{h}_t^k$ is used to update the dialogue state memory and the forgetting operation is calculated as follows:

$$F_t^k = \text{Sigmoid}(W^c_f\tilde{h}_t^k), \quad \tilde{g}_{t,i}^k = g_{t,i}^{k-1}(1 - a_{t,i}^k F_t^k)$$  \hspace{1cm} (9)

where $W_c^f$ is the learnable parameter and $a_{t,i}^k$ is calculated in the same way as in equation (6), with the added operation calculated as follows:

$$A_t^k = \text{Sigmoid}(W^c_a\tilde{h}_t^k), \quad \tilde{g}_{t,i}^k = \tilde{g}_{t,i}^k (1 + a_{t,i}^k A_t^k)$$  \hspace{1cm} (10)

Modification of the dialog history memory by forget and add operations after a K-hop update. At different time steps, the Dialog Memory Manager is able to dynamically keep an eye on the entities in the dialog history memory.

### 2.3. Decoder

The decoder module is used to generate responses word-by-word. First, we follow the approach of Chien-Sheng Wu et al [11] and generate a sketch response using the GRU network learning defined in Equation (5), where the entities in the response are replaced using specific labels obtained from the entities provided in the training data. If the generated word is a label, we select the appropriate entity as the output word by querying the entity representation $h_g$, otherwise the output word is the word generated by GRU. Next, the dialog history and KB information are fused for interaction, then the augmented entity representation fused with the dialog history and KB information is used as the initial hidden state $h_0$, consisting of two parts: the contextual encoder output $h_L$ and the KB encoder output $h_g$. $h_0$ is computed using the following:

$$h_0 = [h_L; h_g]$$  \hspace{1cm} (11)

At time step $t$, the hidden state $h_t$ acts in two respects, the first, as input to the dialog history manager, using the dialog history manager to determine the importance of each word in the dialog history, using the last hop of read memory as the contextual distribution $P_c$, and the second, querying the knowledge graph to generate the graph distribution of all nodes in the graph $P_g$. We use a merge to calculate the final distribution of weights, and the final word distribution and entity word distribution are each calculated as follows:

$$W_{fin} = [h_t; P_c; P_g]$$  \hspace{1cm} (12)

$$P_{voc}^{t} = \text{softmax}(W_0 W_{fin})$$  \hspace{1cm} (13)

$$P_{ent}^{t} = \text{softmax}(V_0 \tanh(W_1 W_{fin} + W_2 h_g))$$  \hspace{1cm} (14)

where $W_0, W_1, W_2, V_0$ are learnable parameters.

### 2.4. Loss

At each decoding time step, we generate sketch responses using $P_{voc}$, and the word with the maximum posterior probability in $P_{voc}$ is used as the generated word, and if the word is a label, the graph entity with the maximum attention...
weight is copied according to $P_{ent}$ to replace the label. Otherwise, the word generated by $P_{vocab}$ are used as the generated word. In the training process, the model training is guided by minimizing the sum of two cross-entropy losses: one between $P^t_{vocab}$ and $y_t$(ground truth), and the other between $P^t_{ent}$ and $P^t_{index}$, where $P^t_{index}$ is the id of $y_t$ in the entity set $E$. Loss is calculated as follows:

$$\begin{align*}
Loss_v &= \sum_{t=1}^{n} -\log(P^t_{vocab}(y_t)) \\
Loss_e &= \sum_{t=1}^{n} -\log(P^t_{ent}(P^t_{index})) \\
Loss &= Loss_v + Loss_e
\end{align*}$$

3. Experiments

3.1. Experimental Setup

Dataset: We evaluate our model using a public multi-turn task-oriented dialogue dataset: the In-Car Assistant dataset [21]. The In-Car Assistant dataset consists of three different domains: calendar scheduling, weather information retrieval and point-of-interest navigation. The In-Car Assistant dataset contains 3031 multi-turn dialogues, divided into 2425/302/304 dialogues, for training/validation/testing, respectively. Each dialogue has an average of 2.6 turns and each dialogue provides its own KB information, which is complex.

Training details: The model is trained end-to-end using the Adam optimizer [22] with the learning rate set to $1 \times 10^{-3}$, the dropout rate set to 0.2. Following the previous work [11,18,26], the number of hops $K$ set to 3 the In-Car Assistant dataset. The model adopted the pytorch deep learning framework, trained on a GeForce RTX 2080 Ti GPU. All embeddings were initialised randomly and a simple greedy strategy was used in the decoding phase. Embedding size and GRU hidden units set to 128 and the batch size used for training set to 16. We use a grid search on the validation set and select the model with the best BLEU.

Evaluation methods: We used two methods for automatic evaluation of metrics: the BLEU [23] and Entity F1\cite{11} scores. BLEU calculates the n-gram overlap between generated responses and gold responses, which measures the ability of the model to learn dialogue patterns. Entity F1 is calculated by micro-averaging the precision and recall of KB entities across the system’s response set, assessing the system’s ability to generate relevant entities to complete a specific task by querying the KB.

Baseline: We compared our model with several end-to-end task-oriented dialogue systems, including

Seq2Seq/Attn\cite{24}, Seq2Seq model with copy mechanism (Ptr-Unk)\cite{25}, memory network-based seq2seq (Mem2Seq)\cite{6}, global-to-local memory pointer network (GLMP)\cite{11}, entity-consistent network with KB retriever (KB-retriever)\cite{10}, Flow-to-Graph seq2seq model (FG2Seq)\cite{12}, Dual Dynamic Memory Network (DDMN)\cite{18}.

3.2. Experimental results and discussion

Table 1 shows the performance of our model relative to the baseline. We can observe that in the In-Car dataset, our model achieves competitive results in both BLEU scores and entity F1 scores compared to most baseline methods, 17.0, 61.7\% respectively, indicating that our model can produce smoother, more appropriate responses. Our model outperforms the FG2Seq model, illustrating the effectiveness of tracking of important information with our Dialog History Manager. On this dataset, we show entity F1 scores for calendar scheduling, weather information retrieval and point-of-interest navigation, with our model scoring highest on the F1 score for the navigation domain. DDMN+ Self-Critical Sequence Training (SCST) scored the highest in the weather information retrieval domain, which may be due to the KB size corresponding to the weather information retrieval domain is relatively large, and DDMN+SCST uses a reinforcement learning approach to train the model with a reward function that can better generate entities related to dialog history in the weather information retrieval domain. FG2Seq scored the highest in the calendar scheduling domain, which may be due to the KB size corresponding to the calendar scheduling domain is relatively small, and FG2Seq uses a graph-based representation of the KB to better capture the importance of each entity node. In the future, we will use a multi-hop mechanism for graph-based knowledge inference to retrieve entities in depth to address the problem of inaccurate entity
generation. That will enable the model perform better in different domains and improve the generalization ability of the model.

### Table 1: Comparison of our model with baselines

<table>
<thead>
<tr>
<th>Model</th>
<th>In-Car BLEU</th>
<th>In-Car F1</th>
<th>Navigation In-Car F1</th>
<th>Weather In-Car F1</th>
<th>Schedule In-Car F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2seq+ Attn[24]</td>
<td>9.3</td>
<td>19.9</td>
<td>10.8</td>
<td>25.6</td>
<td>23.4</td>
</tr>
<tr>
<td>Ptr-UNK[25]</td>
<td>8.3</td>
<td>22.7</td>
<td>14.9</td>
<td>26.7</td>
<td>26.9</td>
</tr>
<tr>
<td>Mem2Seq[6]</td>
<td>12.6</td>
<td>33.4</td>
<td>20.0</td>
<td>32.8</td>
<td>49.3</td>
</tr>
<tr>
<td>KB-retriever[10]</td>
<td>13.9</td>
<td>53.7</td>
<td>54.5</td>
<td>52.2</td>
<td>55.6</td>
</tr>
<tr>
<td>GLMP[11]</td>
<td>14.7</td>
<td>59.9</td>
<td>52.9</td>
<td>62.5</td>
<td>69.5</td>
</tr>
<tr>
<td>DDMN+ SCST[19]</td>
<td>15.8</td>
<td>60.7</td>
<td>53.2</td>
<td>64.7</td>
<td>69.3</td>
</tr>
<tr>
<td>FG2Seq[12]</td>
<td>16.8</td>
<td>61.1</td>
<td>56.1</td>
<td>57.4</td>
<td>73.3</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>17.0</strong></td>
<td><strong>61.7</strong></td>
<td><strong>58.2</strong></td>
<td><strong>59.1</strong></td>
<td><strong>70.0</strong></td>
</tr>
</tbody>
</table>

In terms of evaluation metrics, it can be seen that KB-retriever receives a higher F1 score than previous models, which validates the effectiveness of the KB retrieval component in merging KB information into the seq2seq model to produce better responses. GLMP consistently outperforms Seq2Seq+Attn, Ptr-UNK and Mem2Seq, this is made possible by the ability of global and local pointers to enhance the replication mechanism and improve the accuracy of replicated entity words. FG2Seq is better than GLMP, which validates the effectiveness of using a 'flow-to-graph' approach to facilitate understanding of dialog history and learning information between entities. In order to improve F1 score, DDMN uses the self-critical sequence training (SCST) method to optimally combine discrete reward metrics. DDMN+ SCST overly pursues the generation of dialogue-related entity words, but DDMN+ SCST neglects the ability to learn dialogue patterns. The experiment results show that DDMN+ SCST cannot guarantee the n-gram overlap between the generated responses and the golden responses. Our model obtains the best BLEU and F1 score compare to the baseline. Therefore, the model dynamic attention on important dialog history and automatic filtering of unimportant dialog history information is an indispensable step in improving the accuracy of the system's responses.

#### 3.3. Ablation Study

To further understand the effect of the Dialog History Manager, we performed ablation tests (as shown in Table 3). Our removal of the Dialog History Manager (w/o DHM) resulted in a degradation of BLEU and F1 performance on the In-Car Assistant dataset, by 0.7% and 1.7% respectively. This suggests that updating the dialog context is a necessary step in extracting important information when generating responses. When not using the Dialog History Manager, it makes decoding difficult for the model, and the inclusion of the Dialog History Manager better improves the overall quality of the generated sentences.

### Table 2: Ablation results of our model on In-Car datasets.

<table>
<thead>
<tr>
<th>model</th>
<th>BLEU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td><strong>17.0</strong></td>
<td><strong>61.7</strong></td>
</tr>
<tr>
<td>w/o DHM</td>
<td>16.3</td>
<td>60.0</td>
</tr>
</tbody>
</table>

#### 4. Conclusion

In this paper, we propose an end-to-end trainable dialogue history management architecture for task-oriented dialogue, aiming to improve the performance of task-oriented dialogue systems in discarding unimportant information from the dialogue history and simultaneously generating appropriate responses. The model combines a multi-hop mechanism with an attention mechanism to better focus attention on important information by performing multiple forget and add operations to filter out dialog history information that is irrelevant to responses. Experimental results demonstrate the effectiveness of the approach and reveal a correlation between the success of attending to important information in the dialog history and the success of task-oriented dialog generation.
Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (NSFC) (No.61831022, No.U21B2010, No.61901473, No.62101553).

References

Transformer and Long short term memory networks for long sequence time sequence forecasting problem

Wei Fang¹a*
¹School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau 999078, China
a*1909853ji011001@student.must.edu.mo

Abstract

The long sequence time-sequence forecasting problem attracts a lot of organizations. Many prediction application scenes are about long sequence time-sequence forecasting problems. Under such circumstances, many researchers have tried to solve these problems by employing some models that have proved efficient in the Natural Language Processing field, like long short term memory networks and Transformers, etc. And there are a lot of improvements based on the primary recurrent neural network, and Transformer. Recently, a model called informer which is made for the LSTF was proposed. This model claimed that it improves prediction performance on the long sequence time-series forecasting problem. But in the later experiments, more and more researchers found that informers still can not handle all the long sequence time-sequence forecasting problems. This paper is going to look at how datasets effect the performance of different models. The experiment is carried out on the Bitcoin dataset with four features and one output. The result shows that the Informer (transformer-like model) cannot always perform well so that sometimes choosing models with simple architecture may gain better results.

Keywords- Long sequence time-sequence forecasting; Informers; Long short term memory networks; Deep learning;

1. Introduction

Many fields will face time series problems such as finance, energy, medical, etc [1, 2]. In these scenarios, the problems are not the simple time-series problem, in which long sequence time-series data is required to feed with outputting forecasting based on the sequence data, namely long sequence time-series forecasting (LSTF). Therefore, it is necessary to solve the problem of long sequence time-sequence forecasting because in these fields, if effective forecasting of time series can be carried out, it will significantly improve energy saving, higher return on revenue, and more accurate medical diagnosis.

The significant difficulty is that capacity to process long data is required for the LSTF. And in fact, the basic component of LSTF is the series problem. In other words, it assembles the natural language processing (NLP). So a lot of researchers have employed some models which are mature in NLP. The early try was based on the RNN, like DeepAR and the Deep State Space Model for the Time-series problem [3-5]. Both DeepAR and Deep State Space Model are one-horizon forecast models. They can only predict the value of one moment in the future at a time. To solve the problem, Multi-Horizon Quantile Recurrent Forecaster proposes a multi-horizon forecast model MQRNN that simultaneously predicts values for multiple time steps in the future [6]. MQRNN uses the Encoder-Decoder structure, which is optimized based on the classic Encoder-Decoder structure in NLP. But the "long" attribute of the LSTF problems mentioned above makes it difficult for the RNN in that RNN needs to process too much data and make many meaningless operations. Recently, Transformer has shown tremendous potential in the LSTF benefit from the self-attention mechanism, which can reduce the maximum length of network signals traveling path into $O(1)$ and prevent the recurrent structure which costs a lot [7]. However, more and more papers found that the self-attention mechanism and Transformer framework still cannot handle the LSTF perfectly because it costs a lot in the computation resource. Besides, more and more researchers are also trying to find feasible solutions in other aspects, such as introducing the concepts of reinforcement learning [8, 9]. Although the performance is also quite impressive, it is not the direction of discussion in this paper.

Under such circumstances, a new transformer-like model called Informer was proposed in 2020, which was used for the LSTF problems [10]. The model modifies the self-attention into ProbSparse self-attention. What's more, it also proposed different structures of encoder and decoder. It allows the encoder to process longer sequential inputs and the decoder to generate long sequential outputs through one forward procedure. Fortunately, Informer makes up for the current model...
gap on LSTF problems. According to the experiments based on five training sets shown by the author, the Informer performs far better than other mature methods like Transformer, Logtrans, long short term memory networks (LSTM), and Refor mer in solving related long sequence prediction problems. At the same time, the performance on MSE and MAE is still outstanding. The most important thing is that the Informer reduces both the time and space complexity of training to $O(L \log L)$ and the complexity of testing to $O(1)$. This greatly improves the equipment dependence on solving LSTF problems, improving space and time utilization, and preserving the extraction of sequence problem features as much as possible, effectively solving related long-series time series forecasting problems.

However, this paper finds that the performance of the Informer depends on the dataset. This paper collects the bitcoin dataset, including closing price, transaction fees, reward, difference, and hash rate from 2014/11/4 to 2022/1/10. The Informer cannot predict the hash rate based on the leaving four features well. At the same time, this paper uses the LSTM model to conduct experiments when the relevant parameters are consistent with the Informer. Although the experimental results of LSTM are not very good, they still have good performance compared to the Informer. Therefore, this paper aims to analyze the data sensitivity of the Informer, try to find which type of data is not suitable for the transformer-like models, and compare the relevant content and results of the Informer with the LSTM model.

2. Methods

2.1 Data processing

This part is divided into two parts, the first part is the processing of Informer-related data, and the second part is the processing of LSTM-related data. But before that, what is needed to clarify is that the data has been arranged by imitating the five datasets provided by the Informer's author. The required data is a total of 6 columns. The first column is the date, the second column is the closing price, the third column is the transaction fees, the fourth column is the reward, the fifth column is the difference, and the sixth column is the hash rate. The time series is 2014/11/4-2022/1/10 with 2625 pieces of data totally.

Regarding the data processing of the Informer, in this experiment, to ensure that the relevant variables are not modified, this paper directly used the prototype code provided by the original author of the Informer on Github, without any modification to the hyperparameters, only the parameters related to the data format were adjusted. For the LSTM, this paper performs the most basic normalization of the data and then staggered the data time. The data of the first seven days is used to predict the data for the next seven days, so the input set $x$ and the input set $y$ are staggered by seven days. Divide 80% of the data into the training set and the remaining 20% as the test set. Because this experiment is only used for effect comparison, the author thinks it is unnecessary to set the validation set. In the Informer, the training set, test set, and validation set are divided. But to make the experiments consistent, the proportions are set to be 80% and 20% respectively with no validation sets.

At the same time, this paper adds a set of experiments in the related experiments of LSTF. That is, the data set is randomly arranged, which will disrupt the typical characteristics of the data and increase the difficulty of prediction work because the changing trend of the data will be determined immediately. Changes in trends can be more challenging to predict. Suppose the LSTM model can still perform relatively well in such experimental settings. In that case, this paper can conclude that LSTM still has reasonably good advantages in processing such datasets like the bitcoin dataset this paper mentioned and the randomly arranged version of the bitcoin dataset.

2.2 Model details

The neural network model structures of Informer and LSTM can be changed by users according to different needs. This section mainly explains the detailed structure of the two models used in this paper. Informer structure used in this article is consistent with the sample code provided by Zhou et.al. The specific structure is as follows. The number of layers of the encoder is two, and the number of layers of the decoder is 1. Dimension of the full connect layer in the model is 2048. The attention mechanism used in the encoder is “prob” and employs the gelu activation function. The loss function is MSE. This is the specific structure of the Informer used in this paper. Of course, as mentioned above, according to the characteristics of the data set, this paper modifies the relevant processing parameters of the data set, which not affect the neural network model structure, nor will it affects the results of model training. Similarly, this paper also employs the classic LSTM model for comparison operations. The main structure of LSTM is as follows, accepting five input parameters, the number of hidden units is 16, the number of layers is 1, and the loss function is the same as the Informer, using Mean Square Error. The optimizer's choice is the Adam optimizer. The learning rate is set to be 1e-2 [11].
3. Experimental results

This paper trained the model using the method mentioned in the previous question. The goal is to use the four features, closing price, transaction fees, reward, and difference to predict the hash rate. This paper hopes this can effectively predict the current global virtual currency. The Fig. 1 and Fig. 2 show the performance of the informer and LSTM in predicting the hash rate.

![The prediction result of informer](image1)

![The prediction result of LSTM](image2)

The predicted value of informer and ground truth is almost two disjoint curves, and the differences are very large. It is worth noting that the data depicted in the two figures are displayed after normalization, so if the data is restored, the difference between the predicted value and the actual value will be more significant if the original data range is used. And from the secondary results, it can be found that the prediction curve cannot describe the trend of the real value change well. In contrast to the simple LSTM model, although the predicted value of LSTM cannot accurately represent the change curve of the actual value, most of the expected value is in the same numerical range as the actual value. The two-line segments intersect, which means that at least a relatively good forecast can be made in the field from 2400 - 2500. The reason of this situation may be that LSTM can use its RNN-like structure to learn more about the values of the datasets, and this features can help to improve the performance on the small bitcoin dataset employed by this paper.

Another experiment about LSTM is also carried out, employing the randomly arranged datasets. The randomly arranged data is shown in Fig.3, and the prediction result of the LSTM, which was trained by the randomly placed training dataset, is shown in Fig.4. As shown in the Fig.5, which gathers the value of the test dataset and the prediction results, this can help us compare the result. It can be easily detected that the data is arranged randomly to destroy the regular pattern of the value, making the difficulty of the prediction higher. However, the LSTM is not bad. It can still predict the general trend of data changes, although there are still some deviations.
After completing the prediction of the test set, the value of the loss function can also be obtained intuitively. As shown in the Table 1. The loss function used in the two experiments in this paper is the mean square error (MSE).

Table 1 The MSE of Informer and LSTM

<table>
<thead>
<tr>
<th>Model</th>
<th>Informer</th>
<th>LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>22.737</td>
<td>2.815</td>
</tr>
</tbody>
</table>

The MSE value of the informer is 22.737, which calculate the mean square error between the groudtruth and the prediction value, while the MSE result of the LSTM is 2.815, and the performance of the LSTM is much better than that of the informer in terms of mean square error. Therefore, it can be preliminarily considered that the performance of LSTM in processing this data set is far better than that of the informer. But it is worth noting that LSTM does not have the same complex structure as the informer, nor does it introduce a more advanced self-attention mechanism. Hence, the author initially thought this was not reasonable. Similarly, for problems in machine learning, it is not standardized to only focus on the value of the loss function because blindly pursuing the minimization of the loss function is likely to bring about problems related to overfitting.

Therefore, as mentioned above, it is necessary to pay attention to the measurement standard other than the loss function. It can be seen from the two prediction results that although the results of the informer and LSTM models in describing the development trend of the data are not satisfactory, In contrast, the results of the informer can better explain the changing direction of the original data more closely. In addition, looking at the prediction results of LSTM alone, it can be found
that the prediction range of LSTM is huge, which means that the variance is large, so the prediction results of LSTM are not credible and representative.

Fig. 7 The distilling result of layer:0, head:1

It is worth mentioning that there is a distillation mechanism in the structure of the informer. Its primary function is to extract the features in the sequence as much as possible and then operate on the features. From a macro perspective, this idea can be more helpful for predicting time series problems and judging trends. The distillation results of the informer model on the Bitcoin-related data set used in this paper are shown in the Fig.6, Fig.7, Fig.8, Fig.9, which shows the distilling result of head:0 and head:1 in the layer 0 and 1 respectively. The results are drawn by matplotlib packet. The performance of the distillation results of both layer0 and layer1 is the same as the experiment conducted in the informer article. The results are not very good by comparison. It is easy to see that the distilled information is insufficient to extract the features of the whole trend of the sequence.

From the points mentioned above, although it cannot be proved that the performance of LSTM must be better, it can be intuitively found that the results of the informer are not satisfactory and even inferior to LSTM in many aspects.

Fig. 8 The distilling result of layer:1, head:0

Fig. 9 The distilling result of layer:1, head:1
4. Discussion

This paper argues that the efforts made by the informer in the long-series time series forecasting problem are worthy of affirmation and well-founded. The optimization and simplification of the structure of the informer for the deficiencies of the transformer are recognized by most researchers. Therefore, when encountering such experimental results, this paper is also very puzzled. This paper also makes some explorations and guesses about the reasons after this experiment, mostly comparing models and data.

First, the dataset used in this paper is small, and the amount of data is far less than the demo experiments provided by the informer. Therefore, this article believes that part of the reason is that the lack of data leads to insufficient training in the model. Second, the trend of Bitcoin is different from the long-series time series in many other fields. The direction of Bitcoin has too many contingencies, and the development time is short. In the early days of the Bitcoin market, many concepts in the Bitcoin market did not establish a complete system, resulting in the loss of a lot of related data so that only 0 or other associated data can be used to make up. Such data filling will bring a lot of noise, so it is not surprising that the trained model is difficult to predict and judge.

In addition to the reasons mentioned above for the dataset, there are many differences between the informer and the LSTM. First, the informer has a more complex structure than the LSTM model. The LSTM model modifies the RNN, adding many gate structures and a memory function to the RNN. The model structure of the informer is more complex. As a Transformer-like model, it must include the structure of the encoder and the decoder. In addition, the convolution and pooling operations and the distillation operation are used. An improved model is introduced—the prob-sparse self-attention mechanism. The academic community also generally recognizes that such feature extraction is logical and can optimize the output results. However, this paper believes that such a structure dramatically increases the parameter amount of the model. Therefore, in many cases, insufficient data is challenging to meet the training requirements of more parameters. Instead, LSTM smaller amounts of parameters can be obtained. In addition, this paper also believes that the feature quantity of the Bitcoin data set used is not large enough. In practice, using such a complex model is not necessary, and a model with a relatively simpler structure can be used to solve such problems.

At the same time, it is necessary to review the overall structure of the informer again. The overall structure of the informer is mainly divided into the following parts: encoder, ProbSparse Self-attention, self-attention distilling layer, decoder, and projection. There are multiple layers of ProbSparse Self-attention layer, Self-attention Distilling layer, and a normalization layer. The ProbSparse self-attention mechanism makes some changes to the original self-attention. The traditional self-attention mechanism will become a long tail shape, with active queries in the "head" and "lazy" questions in the tail. The author designed ProbSparse Attention to select active "queries" and discard the part of "lazy" queries. The distillation layer is because mapping the encoder will bring redundancy, and this layer distills an essential part. In this layer, this paper believes that because the amount of data and the number of features is insufficient, after the operation of encoder and distillation, too many features will be discarded, leading to prediction ability. The decoder is of generative style. First, go through 2 layers of multi-head attention layers, and then do the norm.

Compared with the informer, the structure of LSTM is necessarily more straightforward. Although it has a more considerable input and output in many aspects, it is challenging to perform ad hoc extraction of data. Still, the simpler and more efficient structure also brings higher robustness to LSTM. In many cases, higher robustness in time series problems will result in more stable model predictions [12].

5. Conclusion

Through the methods and experiments mentioned above, this paper can intuitively conclude that although the informer has made contributions to the LSTF problem and has an excellent performance in many experiments, it still shows sensitivity to the dataset. The performance in the experiments provided in this paper is not very good. In the comparison experiment with a simple LSTM model, the performance is not superior in general. And this paper does the second experiment carried out on the random datasets. The classical LSTM model can also perform well, so sometimes, researchers can employ LSTM or other models with simple architecture to solve the LSTF problem. Therefore, this paper believes that when using the informer to solve the LSTF problem, it is still necessary to consider the characteristics of the data set and the structure of the informer to judge whether this method is feasible. This paper discusses and speculates the reasons that affect the predictor performance of the informer. The reasons discussed in this paper can be divided into two aspects. The first aspect is the characteristics of different datasets. The second aspect is the structure of the model itself. As mentioned in the
previous article, it may be the characteristics of the dataset that make the prediction work difficult, so in the future, this article will continue to seek to conduct more experiments on more different types and fields of datasets. Similarly, the previous article mentioned that the informer is more complex as a model. The author will continue to study the sensitivity of models of different sizes, structures, and complexity to the data set.

References


Multi-Master and Multi-Slave Oriented Task Offloading Strategy for Real Time and Low Power Internet of Vehicles

Jie Yang*

*aGuoneng Xinshuo Railway Co.Ltd. Zhungeer Banner, Ordos City, Inner Mongolia Autonomous Region 010300

* Corresponding author: 54290725@qq.com

ABSTRACT

With the rapid development of intelligent driving and on-board intelligent applications, the computing power of on-board units is gradually inadequate. Intelligent networked vehicles offloading tasks to cloud servers through the Internet of vehicles is considered to be a promising method. However, long distance deployment of cloud servers and the instability of return links also bring high time delay. While Mobile Edge Computing (MEC) effectively solves this problem by deploying computing resources to the network edge. Therefore, based on the idea of mobile edge computing, this paper first constructs the local edge collaborative computing model. By comprehensively considering the factors such as user psychology, vehicle speed, acceleration, location, communication ability and computing ability, the utility function of task vehicle and service vehicle is established. Then, according to the Stackelberg game strategy, the interaction behavior between task vehicle and service vehicle is modeled, the Stackelberg cyclic iterative task offloading algorithm in the Internet of vehicles environment is proposed. It is proved that there is a Nash equilibrium point between service vehicle and task vehicle. Finally, the simulation results show that the algorithm has achieved a balance between task delay and expense, task vehicle utility and service vehicle utility, and has higher performance than other algorithms.

Keywords: Stackelberg game, mobile edge computing, Internet of vehicles, Multi-Master and Multi-Slave

1. INTRODUCTION

As the rapid popularization of intelligent networked vehicles and the increase in road traffic density, the vehicle’s own computing power has been unable to meet the growing computing requirements [1]. The current solution is to offload the calculating task to the cloud [2], [3]. However, the task’s service delay will be greatly increased due to insufficient computing power, the long distance and the unstable return link, which will not match the delay sensitive application requirements of intelligent networked vehicles. In order to solve the problems of insufficient processing power and limited resources of terminal equipment, researchers have introduced edge computing technology [4]. Task offloading is the core problem of mobile edge computing, which includes offloading decision-making and resource allocation. In vehicle mobile edge computing, task offloading is also an urgent problem to be solved. The task offloading process is mainly composed of six steps: node discovery, task cutting, offloading decision-making, task transmission, task execution by edge server and task result reception. In [5], Raza S et al formulate the architecture of vehicle edge computing (VEC). To solve the congestion problem in other data networks, a solution is developed for effectively utilizing the potential resources of automobiles in the network [6]. An architecture for vehicular edge computing offloading is proposed. To handle the problem of the edge servers’ limited computing resource, a backup server is introduced to increase computing capacity [7]. In [8], Hu B et al designed a dynamic service allocation algorithm based on Pareto optimization, which solved the problem of edge node allocation. To increase the number of cloud services available for vehicle mobility applications. An architecture combining remote cloud servers, edge computing nodes, and vehicular cloud has also been proposed [9]. In [10], a vehicular network based reputation management system is proposed. The resource provider determines the allocation of computing resources based on the users’ reputation. In [11], Game theory is introduced in vehicle edge computing, and the task offloading problem is described as a multi-user non-cooperative game problem with a Nash equilibrium solution. Compared with binary offloading, it is more reasonable to partially offload to the mobile edge computing of vehicles, because the wireless channel resources are limited [12]–[14].

In this paper, we model the movement, communication and tasks of the vehicle. Factors such as user psychology, vehicle speed, acceleration, location, communication power, and computing power are taken into consideration. Then we establish
the utility function of the task vehicle and the service vehicle according to the task delay, expense and user sensitivity. The Stackelberg game model is used to modeling the offloading process, and the task offloading algorithm of Stackelberg cyclic iterative algorithm is proposed. Simulation results show that the algorithm achieves a balance between task vehicle delay and expense, task vehicle utility and service vehicle utility. Compared with other algorithms, the algorithm has higher performance and faster execution speed.

2. SYSTEM MODEL

In the Internet of Vehicles environment, we use $L_{road}$ to denote the length of the road and $L_{width}$ to denote the width. At time $t$, the vehicle enters the road from the left side, drives to the right, and turns around when it reaches the end. During driving, each vehicle has a chance of generating a computing task in each time slot, and the probability of occurrence is $\zeta$. Vehicles can execute the task using local computing resource or rent from other vehicles with free computing resources to execute local-edge collaborative computing to complete task within the time allowed [15].

Assuming that there are $M$ service vehicles in the system that have sufficient computing power and are capable of offering services. Task vehicles have limited computing power that need to offload tasks, and number of task vehicles is $N$. We use set $V_s = \{v^1_s, v^2_s, ..., v^m_s\}$ to represent service vehicles and set $V_t = \{v^1_t, v^2_t, ..., v^n_t\}$ to represent task vehicles. When the service vehicle has no task execution, it can provide computing power support to the task vehicle. $f_{i, max}$ denotes the upper limit of the computing power that the service vehicle can provide. $m$ represents the upper limit of the computing power that the service vehicle $s$ can provide. $N$ task vehicles compete with each other for $M$ service vehicles’ computing resources.

2.1 Communication model

We use $R_{v2v}$ to denote the maximum communication range of V2V communication. V2V communication uses independently and identically distributed channels, and the path loss can be calculated by Eq. (1):

$$PL = 63.3 + 17.7 \log(d_{i,j}) \quad 0 < d_{i,j} < R_{v2v}$$

Where $d_{i,j}$ denotes the distance between vehicle $i$ and vehicle $j$, which can be calculated by Eq. (2):

$$d_{i,j}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$$

Where $x$ is the abscissa of the vehicle position, and $y$ is the ordinate of the vehicle position. At time $t$, the relative speed between vehicle $i$ and vehicle $j$ can be calculated by Eq. (3):

$$v_{i,j}(t) = |v_i(t) + a_it - v_j(t) - a_jt|$$

We assume that the noise is 0, $B$ denotes the V2V communication bandwidth, $h$ denotes the decline factor the transmission link channel, $N$ denotes the power of Gaussian white noise and $P_{v2v}$ denotes the transmission power of vehicle-mounted communication equipment. Therefore, according to Shannon formula, the data transmission rate between any two vehicles $i$ and $j$ can be calculated by Eq. (4):

$$R(t) = B \log_2 \left( 1 + \frac{P_{v2v} b(t) h^2}{N} \right)$$

If the transmission distance is not exceeded and the transmission time is not exceeded, the average transmission rate between vehicle $i$ and vehicle $j$ can be calculated by Eq. (5):

$$R = \frac{\int_0^{t_{trans}} R(t) dt}{t_{trans}}$$

Where $t_{start}$ denotes the time to start the task upload, and $t_{trans}$ denotes the data transmission time.

2.2 Task Model

We denote $T$ as the offloading task. We assume that the task can be divided and assigned to different computers for parallel computing. The data size of the computing task $T$ is $D$ bit. Usually, the number of CPU cycles is used to judge how much computing power a task requires. The computing resources required for the task can be calculated by $c = \alpha D$. Where $\alpha$ denotes the number of CPU cycles computing 1 bit data. We assume that the task’s return data size is proportional to the task data size, which satisfies $c = \beta D$, where $\beta$ denotes the compression ratio. Usually, the return data size is much smaller...
than the input data in this task, so $\beta$ tends to 0. Hence, the task $T$ can be represented by a function $f\{D, \alpha, t^{\text{max}}, \beta\}$. The priority of entertainment tasks is usually lower than safety driving tasks, and the task priority is calculated by the following Eq. (6):

$$P_T = \frac{d}{t^{\text{max}}}$$

(6)

Where $t^{\text{max}}$ denotes the maximum execution time allowed for the task.

### 2.3 Utility Function

We use a function to represent the vehicle $f\{x_0, y_0, v_0, a, t, f_{\text{cpu}}, f_{\text{time}}, f_{\text{pay}}\}$, $x_0, y_0$ respectively denotes the position of the vehicle at the initial time, $v_0$ denotes the initial speed, $a$ is the acceleration of the vehicle, $t$ denotes the running time of the vehicle, we can get the current speed of the vehicle at time $t$ by formula $v = v_0 + at$. $f_{\text{cpu}}$ is the computing power of the in-vehicle CPU, $i$ is a flag that can be 0 or 1, where $i = 0$ means that the vehicle is task vehicle, $i = 1$ means that the vehicle is server vehicle. $f_{\text{time}}$ is the time sensitive factor of vehicle users, $f_{\text{pay}}$ is the price sensitive factor, they denote the sensitivity of the mission vehicle to time and expenditure costs. When $i = 1$, $f_{\text{time}}$ and $f_{\text{pay}}$ can be set to 0. According to the basic information of task $T$ and the in-vehicle computing power $f_{\text{local}}$, the time required for the task completely on its own can be calculated by Eq. (7).

$$t^{\text{all, local}} = \frac{ab}{f_{\text{local}}}$$

(7)

The revenue for the service vehicle is equal to the task vehicle’s cost of purchasing computing resources minus the cost of computing electricity. For the service vehicle $v$, the computing resources it sells are: $(f_{1,v}, f_{2,v}, \ldots, f_{n,v})$. Hence, its utility function is defined by Eq. (8):

$$U_{\text{service}}^v = p_v \sum_{i=1}^{n} f_{i,v} - ek \sum_{i=1}^{n} f_{i,v}^2 \frac{ad_i}{f_{\text{total}}}$$

(8)

The task vehicle must be set at a suitable price to obtain satisfactory income and satisfy $P_{\text{min}} \leq p_v \leq P_{\text{max}}$. The best price is calculated by Eq. (9):

$$p_v^* = \text{argmax} U_{\text{service}}^v (p_v, p_{\text{v}}^*, F^*)$$

(9)

Where $p_{\text{v}}^*$ denotes the optimal strategy for other service vehicles except the service vehicle $v$, $F^*$ denotes the optimal computing power purchase strategy for each task vehicle. The goal of any service vehicle is to maximize its utility function $\max U_{\text{service}}^v, v \in V$.

The utility function of the task vehicle is defined as the time utility function minus the cost utility function. $t^{\text{save}}$ denotes the user satisfaction by time, $t^{\text{all, local}}$ denotes the completion time of the task when it is computed locally. $t^{\text{me}}$ denotes the completion time that the task offloaded from task vehicle $u$ to service vehicle $v$. $t^{\text{save}}$ denotes the average time saved after offloading, it can be calculated by Eq. (10):

$$t^{\text{save}} = t^{\text{all, local}} - \frac{t^{\text{local}} + t^{\text{me}}}{n + 1}$$

(10)

$U_{\text{time}}$ is an increasing function with respect to $t^{\text{save}}$, and because of the user’s psychological threshold, the task vehicle is continuously purchasing computing power to increase $t^{\text{save}}$. But function $U_{\text{time}} = \gamma_1 \ln(1 + t^{\text{save}})$ is diminishing marginal utility. After reaching the delay threshold in the mind of the user, even if more time is saved, $U_{\text{time}}$ will not increase much. The fee that users need to pay to the service vehicle to purchase computing resources can be calculated by Eq. (11):

$$U_{\text{pay}} = \gamma_2 (p_{\text{pay}} + E_{\text{local}})^y$$

(11)

Which satisfies the function $p_{\text{pay}} = \sum_{i=1}^{n} f_{u,i} p_{i,\text{local}}$ denotes the electricity cost consumed by local calculation in local edge collaborative computing. Therefore, the utility function of task vehicle $u$ can be expressed by Eq. (12):

$$U_{\text{task}}^u = aU_{\text{time}}^u - bt_{\text{pay}}^u + C = a(\gamma_1 \ln(1 + t^{\text{all, local}} - \frac{t^{\text{local}} + \sum_{i=1}^{m} t^{\text{me}}}{m + 1}))$$

(12)
2.4 Solution to the Stackelberg game problem

We propose a task offloading algorithm based on the multi-master and multi-slave Stackelberg game model in IoV. At time $t$, the price strategy for the service vehicle to broadcast unit computing power to all task vehicles is $P(t) = (p_1, p_2, \ldots, p_m)$. The task vehicle quickly adjusts the computing power purchase strategy of each task vehicle in combination with the task information and its own sensitivity to price and time to maximize its own utility based on the price strategy received. The equilibrium of the service vehicle needs to be adjusted through multiple iterations to approach the Stackelberg equilibrium solution. The change rate of task vehicle $u$’s computing power demand is proportional to the gradient of its own utility function:

$$\frac{df_u}{dt} = \nabla f_u = \frac{\partial u_{\text{task}}}{\partial f_u} = \left(\frac{\partial u_{\text{task}}}{\partial f_{u,1}}, \frac{\partial u_{\text{task}}}{\partial f_{u,2}}, \ldots, \frac{\partial u_{\text{task}}}{\partial f_{u,m}}\right), u \in V_t$$

(13)

Where $\tau$ is the time variable, the task vehicle utility function’s concave function ensures that after a finite number of iterations, the iterative algorithm in this chapter can stably converge to the Nash equilibrium point. The time interval for each iteration of offloading is called $\Delta \tau$, In the iteration cycle $\Delta \tau$, the demand for computing power at time $\tau + 1$. The iterative formula is Eq. (14):

$$f_u(\tau + 1) = f_u(\tau) + L_{\text{task}} \nabla f_u$$

(14)

Where $L_{\text{task}} > 0$ denotes the adjustment step size of the task vehicle’s computing power purchase decision.

After all task vehicles have reached Nash equilibrium, the service vehicle adjusts its price strategy using an iterative formula based on each task vehicle’s computing power purchase strategy at time $t + 1$, the formula Eq. (15) is as follows:

$$p_v(t + 1) = p_v(t) + L_{\text{service}} \nabla p_v$$

(15)

where $L_{\text{service}}$ denotes the price strategy adjustment step size of the service vehicle. When the service vehicle is in iteration cycle $\Delta t$, The partial derivative Eq. (16) of the utility function of the service vehicle with respect to the price can be calculated by the change $\theta$ that tends to 0.

$$\frac{dp_v}{dt} = \nabla p_v = \frac{\partial u_{\text{service}}}{\partial p_v} = \lim_{\theta \to 0} \frac{u_{\text{service}}(p_v+\theta \ldots) - u_{\text{service}}(p_v-\theta \ldots)}{2\theta}, v \in V_s$$

(16)

Until the task vehicle’s computing power purchase strategy reaches the Nash equilibrium, service vehicle's price strategy should remain unchanged. This iterative process is $\Delta t$, One $\Delta t$ contains several $\Delta \tau$. After many dynamic adjustments, the service vehicle gets the optimal price strategy $P^*$, and the task vehicle gets the optimal purchase strategy $F^*$, which maximizes the utility.

3. SIMULATION ANALYSIS

The Stackelberg loop iteration algorithm proposed in this paper is suitable for task offloading when the number of task vehicles is small and the task priority is low when the vehicles are crowded and low-speed. Because the results obtained by the algorithm in this paper are relatively accurate, in addition to being used for task offloading decisions, this algorithm can also be used to generate a training set, and then train a neural network to quickly solve the task offloading problem of IoV.

Fig. 1 shows the change curve of the average utility value of the task vehicles when there are 5 service vehicles in the system, and the number of task vehicles ranges from 10 to 40 in different environments using various algorithms for offloading. In order to maximize the user experience, the QoS priority algorithm does not consider the price of the service vehicle and purchases as much computing resources as possible to reduce the completion time, but the cost is too high. Hence, its average utility value is not as good as the All-Local algorithm. The Stackelberg loop iteration algorithm proposed in this paper is better than other algorithms in the environment of different task vehicles. The performance of the two NSGA-III algorithms is similar. In the environment of 10, 15 and 20 task vehicles, the NSGA-III-DE algorithm performs better than the NSGA-III algorithm, but after that, the utility value of the two genetic algorithms declines rapidly. After analysis, it is found that as the number of task vehicles increases, the number of objective functions and population size of NSGA-III increase, resulting in the algorithm still not finding the optimal solution after 5000 evolutions. The Stackelberg loop iteration algorithm proposed in this paper, although with the increase of task vehicles, the utility of the task vehicles has slightly decreased, but the overall trend is stable. And the utility value of the task vehicle using the algorithm in this
paper is 12% higher than that using the NSGA-III algorithm; the performance is 25% better than the NSGA-III-DE algorithm.

Fig. 1 Average utility curve of task vehicles based on different algorithms.

Fig. 2 shows the comparison of the unit computing power revenue of the algorithm in this paper and other algorithms as the number of task vehicles increases. The unit computing power cost benefit is defined as the utility function value obtained for each 1GHz computing power purchased by the task vehicle. With the increase in the number of task vehicles, the unit computing power cost benefit of offloaded task vehicles using this method has been fluctuating slightly, and other algorithms have been lower than the algorithm in this paper. The QoS algorithm purchases too many computing resources but can not obtain a high utility value, resulting in the revenue per unit of computing power continues to decline. Except for the QoS algorithm, with the increase of the number of task vehicles, it shows a trend of first increasing and then decreasing. The algorithm in this paper has a small and steady increase before 30 task vehicles, and then a small decrease, showing a stable trend overall. The NSGA-III and NSGA-IIIDE algorithms maintained a rapid increase before the 15 and 20 task vehicles, respectively, but began to decline rapidly after the highest point, with large fluctuations. Considering the environment of the number of vehicles for each task, the average performance of the Stackelberg loop iteration algorithm is 40% higher than that of the NSGA-III algorithm and 12% better than the NSGA-III-DE algorithm.

Fig. 2 Unit cost benefit curve of different algorithms.

4. CONCLUSION

This paper studies a multi-user mobile edge computing system. Compared with the traditional method, this paper introduces economic factors into the computing offloading, and converts the traditional method of optimizing delay and energy consumption into optimizing delay and cost, which is more practical. The multi-master and multi-slave Stackelberg game model is introduced to model, and then the utility function of the computing power user task vehicle and the computing power provider service vehicle in IoV is proposed, the interaction constraints between the task vehicle and the service vehicle are also considered. We prove that the utility function of the task vehicles satisfies the concave function
property, which ensures that there is a Nash equilibrium point in the non-cooperative game between the task vehicles. According to the optimal response dynamics in EGT, a distributed execution of Stackelberg loop iteration algorithm is proposed to solve the optimal purchase strategy of task vehicles and the optimal price strategy of service vehicles. The two sides of the game interact and evolve each other, so that the task offloading of the task vehicle achieves a balance between delay and expenditure, realizes the Nash equilibrium between the task vehicle and the service vehicle, and realizes the utility optimization of both sides of the game. This paper describes and simulates the proposed algorithm in detail. The results show that the proposed algorithm is practical for the edge computing system of IoV, and obtains higher average utility function value and unit cost benefit than other algorithms, and the execution time is relatively short.

REFERENCES


Gateway malicious command detecting for intelligent fishery sensor networks

Malicious instruction detection method based on deep learning

FengWei Zhang\textsuperscript{a}, WenXin Tao\textsuperscript{b*}, YingZe Sun\textsuperscript{a}

\textsuperscript{a}Information Technology Research Center, Chinese Academy of Fishery Sciences, Beijing, China;
\textsuperscript{b}Faculty of Computing, Harbin Institute of Technology, Harbin, China

*Corresponding author: 21B903092@stu.hit.edu.cn

ABSTRACT

Fisheries IoT realizes real-time online detecting and precise modulation of aquaculture process, which greatly enhances the informationization, automation and intelligence level of farms, reduces production costs, promotes the scale of the fishery industry, and improves the quality level and market competitiveness of agricultural products. However, the key to large-scale fishery farming lies in accurate water environment modulation, the IoT gateway system is the key to realize the ability of intelligent fishery to achieve accurate water environment modulation. Once the aquatic product networking is attacked by malicious IoT commands, it may cause rapid disintegration of the farming environment ecosystem at any time, resulting in huge economic losses. Due to the size and power consumption of fishery IoT gateways, traditional security protection means are often difficult to be applied in such scenarios. Therefore, this paper proposes an improved malicious command detection method for smart fisheries sensor gateways based on deep learning techniques, aiming to accurately and efficiently identify malicious control commands of smart fisheries sensor networks on IoT gateways. The main contributions of this paper are three: (1) Transforming various types of signals from gateways into feature texts, which are used as inputs for deep learning for command classification. (2) Based on the BiLSTM model, it is improved by adding a multi-headed self-attention mechanism to obtain a smaller size and higher accuracy malicious command detection model applicable to smart fisheries gateways (3) Experimental results show that compared with other models, highly confusing malicious commands elaborated for the command characteristics of short, small size and high change rate of smart fisheries sensor commands The improved BiLSTM model has a higher recognition rate.

Keywords: malicious command detecting, multi-head-attention, command classification.

1. INTRODUCTION

With the continuous evolution of wired and wireless sensor network technology, the development of fishery aquaculture industry has also entered the fast track of intelligence. Fishery aquatic networking has achieved real-time online detecting and precise regulation of the farming process, greatly enhancing the level of information, automation and intelligence of the farm, reducing production costs, improving labor productivity, resource utilization and management efficiency, and improving the quality level and market competitiveness of agricultural products. However, along with these conveniences, malicious attacks and extortion incidents for intelligent fishery farming IoT devices have gradually become frequent. Limited by the structure and performance of sensor gateways, gateways are often unable to implement more complex self-verification means. Therefore, a large number of gateways use direct transmission of commands based on URL and TCP commands, which makes the security of sensor networks very vulnerable. With the sophistication of network attacks, the traditional network security protection based on the blacklist keyword defense method, because of the need to maintain a large volume of keyword list and not suitable for such as the IoT gateway such a limited storage space environment. Therefore, how to have IoT gateway limited physical storage space and energy constraints to achieve efficient and complex malicious command detection and interception has become a pressing problem. To solve this challenge and defend against malicious attacks against IoT gateways, researchers have proposed a series of deep learning-based malicious command detection algorithms.

The key to fishery scale farming lies in precise water environment modulation, and these modulation commands are often real-time and fickle. The IoT gateway system is the key to realize the ability of smart fisheries to achieve precise water environment modulation. Wrong IoT commands can cause dramatic changes in the water environment at any time and cause rapid disintegration of the farming environment ecosystem. With the large-scale application of sensor systems in the fisheries industry, the characteristics of short, small and large change rate of smart fisheries sensor commands...
make the damage caused by malicious cyber-attack sensors on the large-scale production of fisheries farming also increases. Therefore, the detecting of malicious commands at the gateway of smart fisheries sensor networks is therefore becoming more and more important.

Currently, the detection methods based on deep learning for gateway commands are mainly concentrated as follows:

I. Detection using machine learning-based methods. In response to the limitations and low efficiency of blacklisting technology, and to improve the generality of malicious commands detection algorithms, researchers have proposed a series of malicious commands link detection algorithms based on machine learning. Machine learning actually treats malicious commands detection as a text classification problem[1], and common algorithms for machine learning include support vector machine (SVM), Bayesian network, Naïve Bayes model, Random forest, and so on. However, most of the malicious commands detection methods based on machine learning rely on manually selected URL features, which have the following two drawbacks: (1) low efficiency, (2) different feature selection is highly subjective, which will affect the malicious command detection results to a certain extent and lead to poor robustness of the detection model. Thus, it seems that machine learning is also unable to meet the requirements for efficient malicious data-body detection in real environments.

II. Detection using neural network model: With the growing maturity and wide dissemination of deep learning techniques, researchers have started to apply deep learning techniques to the detection of malicious network requests. Like machine learning, deep learning also considers malicious command detection as a text classification problem. Ren F[2]. et al. used Word2Vec word embedding method to vectorize the URLs of network requests, and then used bi-directional long-term memory recurrent neural network (BiLSTM) to binary classify the URLs of network requests. Kang Chen and Huazheng Fu [3]et al. first analyzed the lexical features of URLs, trained recurrent neural networks (RNN) using benign commands in the dataset to generate embedding representations of URL characters, then converted URL links of commands into feature images, and finally used convolutional neural networks (CNN) to bifurcate URLs of commands, and the accuracy and recall of the model were able to reach 96.2% and 87.9%. Zhang Qiao [4] proposed a CNN-BiLSTM based counterfeit commands detection algorithm, which uses convolutional neural network (CNN) to obtain the spatial features of commands and bi-directional long and short-term memory recurrent neural network (BiLSTM) to obtain the temporal features of commands with good accuracy, precision and F1 value.

In this paper, we propose a new self-attentive method multi-head-attention (MHA) for malicious commands detecting based on the previously described BiLSTM detecting method.

The rest of this paper is organized as follows. The next section briefly describes the data acquisition and pre-processing process. Subsequently, a detailed optimized method, including the improvement mechanism and training configuration, is given in Section 2.4. In addition, the results of the traditional SVM, Text-CNN and BiLSTM detecting methods are compared with the computational results of this method, is given in Section 3. Finally, the fourth section concludes the paper.

2. METHOD

2.1 Data preparation

First, this paper extracts 1748822 attack data and 10287239 normal communication data from the attack events archived in the firewall of the core network of China Academy of Fisheries Science against the IoT sensor gateway in the fishery industry laboratory in the network as the original data samples, and converts the URL, Data-body and other information into json format. Then the data are pre-processed and marked as "malicious data" and "normal data" respectively.

One of the pre-processing steps is as follows:

A.URL address segmentation

A.1 For the protocol part, because HTTP, HTTPS, FTP and other protocols have great differences in data encryption and security performance, the protocol part cannot be discarded and needs to be separated out. In this paper, the protocol name is combined with the string "/!", for example, "http://", "https://".

A.2 For the host name part, since the URL contains multiple levels of domain names, although at the top-level and first-level domains, the non-facilitators usually cannot take measures to obfuscate, but they can adopt obfuscation strategies by adopting the second or higher level domains. In this paper, the domain name parts are separated by the separator
symbol ". " to separate them level by level, with each separator ". " follows the next level domain name. For example, for the host name part of the string "www.cafscfe.com", the result of this paper is :("www", ".cafscfe", ".com")

A.3 For GET query and POST data body part, the "key=value" information key-value pairs often contain sensitive information, this paper firstly separates multiple "key=value" information key-value pairs with '& amp;' as the separator, and then separate the key-value pairs with "+" as the separator, so as to extract the key and value values of the key-value pairs. Since the URL splitting process needs to preserve the connection between words and special symbols (separators), the splitting symbols (such as '& ', '=&', etc.) are not discarded after the splitting operation in this paper. For example, for the query part of "?username=cafs\password=b8151fa " of the query string, the result of this paper is :("?", "username","="","cafs","","\password","="","b8151fa"")

A.4 For the path part, each level of the path is separated by '\ ' as a separator, but in order to distinguish it from the string of the reference part, each separator '\ ' is combined with the path after it. For example, for the path string of "perm_ce/ysfi/ici", the result of this paper is as follows:("perm_ce\"yfs\"i\"c\"")

A.5 Since the parameter part and anchor part have almost no effect on the data-body detection results, and the dataset used in this paper contains almost no such samples, the parameter part and anchor part of the data-body URL are not specially treated in this paper.

B. Word2Vec based URL word vector modeling

Word2Vec is used as a tool for splitting commands into a vector of fixed dimensions, so that the semantic similarity of each word in the lexicon can be calculated. In this paper, the Word2Vec model is used to represent the word separation results of command datasets as word vectors in a high-dimensional vector space, and then the text representation model is applied to malicious detection. A sample word vector is shown in Figure 1.

![Figure 1. Word vector representation of partial vocabulary (128 dimensions)](image)

2.2 Multi-head-attention method

Multi-head-attention is a new learning mechanism. It used to calculate \( d_q, d_k, d_v \) dimensions respectively instead to single attention. Figure 2 shows the multi-head-attention’s training flow chart.
2.3 Equations

MHA allows the model to jointly attend to information from different representation subspaces at different positions. With a single attention head, averaging inhibits this. As shown in equation 1.

\[
MH(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O
\]

\[
\text{where } \text{head}_i = \text{Attention}(QW^Q_i, KW^K_i, VW^V_i)
\]

(1)

In the computation of the improved BiLSTM model, MHA is introduced to assist its learning of the relationship between tokens and commands. In the learning process, the commands are used as the learned Key and the values are the tokens associated with them, combined into a command-token sequence as the computational queue.

The following classification outcome measures are introduced in this paper[5]:

a. Defines(TP, TN, FP, FN)

TP (True Positive): The command is malicious and the classification result is malicious.

TN (True Negative): The command is benign and the classification result is benign.

FN (False Negative): The command is malicious and the classification result is benign.

FP (False Positive): The command is benign and the classification result is malicious.

b. Accuracy (accuracy): refers to the proportion of correctly classified command samples to the data set. As shown in equation 2.

\[
\text{acc} = \frac{TP + TN}{TP + TN + FN + FP}
\]

(2)

c. Recall (recall): refers to the proportion of all malicious command data sets that are correctly classified. As shown in equation 3.

\[
\text{recall} = \frac{TP}{TP + FN}
\]

(3)
d. Accuracy (precision): refers to the proportion of pages that are actually malicious in the set of all data predicted to be malicious samples. As shown in equation 4.
\[
\text{pre} = \frac{TP}{TP + FN}
\]  
(4)

e. False positive rate (FPR): the proportion of classification errors in the set of all benign data-bodys. As shown in equation 5.
\[
\text{FPR} = \frac{FP}{FP + TN}
\]  
(5)

f. False Negative Rate (FNR): The percentage of classification errors in the set of all malicious data-bodys. As shown in equation 6.
\[
\text{FNR} = \frac{FN}{TP + FN}
\]  
(6)

g. F1(F-Measure): the summed average of precision and recall. As shown in Equation 7.
\[
\text{F1} = \frac{2TP}{2TP + FP + FN}
\]  
(7)

2.4 Construction and training of improved BiLSTM models

BiLSTM (Bidirectional Long Short-Term Memory) is a type of RNN, the traditional RNN model is not good at processing long text data and has a poor memory retention ability for long text. BiLSTM adds memory cells to solve this problem. Meanwhile, BiLSTM uses two layers, forward and backward, to solve the problem of the lack of dependence of RNN below layer on the above layer, and uses the information of the context to determine the prediction result of the model at the same time, thus improving the accuracy of the model detection.

Cause the command text is a linear sequence and based on the property that the higher level of the URL is in the more backward position, we use the BiLSTM-based implementation to capture longer distance dependencies by considering the backward and forward order of words in a sentence, and to extract certain feature information in the command body from backward and forward encoding.

![Network structure of BiLSTM](image)

Figure 3. Network structure of BiLSTM

In order to improve the extraction efficiency of commands feature information and increase the accuracy of commands classification, Attention (attention mechanism) is introduced in this paper. In this paper, the Attention mechanism of
Transform model is used to calculate the weighted sum of the last time-series feature vector y output from the BiLSTM model proposed in Section 2.4 as the final feature vector output of the model, instead of directly outputting the last time-series output vector of BiLSTM into Softmax. The model algorithm is shown in Figure 4.

Figure 4. Improved BiLSTM based command detection model algorithm chart

3. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the accuracy and usability of the model, this paper mixes attack data and normal data from the attack events against the IoT sensor gateway in the in-network fisheries industry laboratory archived in the core network firewall of the China Academy of Fisheries Science, and extracts an additional 1 million new control command as the validation dataset. The dataset contains various types of normal requests, malicious permission forgery, DDOS attacks, injection attacks, etc., with benign and malicious data-bodys. The ratio of benign pages to malicious pages is close to 5:1.

To simulate a constrained physical environment of gateway, the experimental uses a RaspberryPi as the test environment for the trained model, using 8GB of memory and 64GB of storage.

The following validation conclusions were obtained in the experimental:

3.1 Experimental Results

The experiments in this section compare the malicious command detecting performance between SVM, TextCNN, BiLSTM, Improved BiLSTM models for the network requests using the same embedding method (Word2Vec). Details are shown in Table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy rate (%)</th>
<th>Accuracy rate (%)</th>
<th>Recall rate (%)</th>
<th>False positive rate (%)</th>
<th>False negative rate (%)</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>86.38</td>
<td>75.19</td>
<td>73.47</td>
<td>8.51</td>
<td>22.51</td>
<td>75.26</td>
</tr>
<tr>
<td>TextCNN</td>
<td>93.09</td>
<td>94.00</td>
<td>73.13</td>
<td>1.68</td>
<td>27.91</td>
<td>81.23</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>94.41</td>
<td>92.35</td>
<td>79.49</td>
<td>1.67</td>
<td>21.24</td>
<td>83.46</td>
</tr>
<tr>
<td>Improved BiLSTM</td>
<td>94.77</td>
<td>94.21</td>
<td>79.49</td>
<td>1.14</td>
<td>21.30</td>
<td>86.21</td>
</tr>
</tbody>
</table>
3.2 Analysis

The improved algorithms based on deep learning proposed in this paper perform significantly better than others (such as SVM). Among the three deep learning models in experimental, the improved BiLSTM has the best result with an F1 value of 86.21%. Although the TextCNN model is better than the BiLSTM model for the short text classification task, the experimental results show that the BiLSTM outperforms the TextCNN model in terms of accuracy and recall rate. Compared with the unimproved BiLSTM algorithm, the improved BiLSTM can be more effective in discriminating when dealing with elaborate confused commands.

4. CONCLUSIONS

In this study, we propose an improved BiLSTM model based on the MHA method for the application scenario of fisheries IoT gateways. The main advantage of this method is that the complex data with short instructions, small size and high change rate unique to fisheries IoT gateways and sensors or operation terminals are uniformly considered as special text for sampling, and targeted feature recognition of different data blocks with the support of MHA. Under the limited physical conditions of IoT gateways, efficient and stable malicious instruction detection is achieved. Compared with the traditional machine recognition model, the improved BiLSTM model has better detection results in smart fisheries scenarios.

Considering the similarity of signal features, the model is also expected to have potential applications in embedded and wearable sensor network scenarios beyond fisheries.

In addition to fisheries IoT autonomous systems, fisheries IoT gateways often work with data centers in practical application scenarios. These co-working signal's tend to wrap the control commands more deeply. This brings new challenges to the detection of malicious commands. In the next studies, it will be a new challenge to optimize the defense and detection capability of big data information forgery after establishing communication tunnels between controller gateways and between gateways and Internet servers under limited hardware conditions.

5. ACKNOWLEDGMENTS

This work was supported by "Central Public-interest Scientific Institution Basal Research Fund CAFS (NO.2020TD85)" "Central Public-interest Scientific Institution Basal Research Fund CAFS (NO.2022B001)" and "China Knowledge Centre for Engineering Sciences and Technology (CKCEST)" (Project Number: CKCEST-2022-1-21).

REFERENCES

Calculation and Analysis Method for Distribution Network Fault Location Based on Improved Differential Evolution Algorithm

Yangjun Zhou\textsuperscript{1a}, Chenying Yi\textsuperscript{1b}, Like Gao\textsuperscript{2a}, Jianna Ouyang\textsuperscript{1c}, Wei Zhang\textsuperscript{1d*}

\textsuperscript{1}Guangxi Key Laboratory of Intelligent Control and Maintenance of Power Equipment, Electric Power Research Institute of Guangxi Power Grid Co., Ltd, Nanning 530023, China
\textsuperscript{2}Guangxi Power Grid Co., Ltd, Nanning 530023, China

\textsuperscript{1}a1193736858@qq.com, \textsuperscript{1b}yichy31@163.com, \textsuperscript{2a}625703412@qq.com, \textsuperscript{1c}438902132@qq.com, \textsuperscript{1d*}2387459203@qq.com

Abstract

Distribution network has complicated grid structure and various kinds of faults, which make the current measures hard to accurately locate fault zones in field operation. This paper proposes a differential evolutionary algorithm to locate the fault based on the transient wave record data from the distribution network, using the mechanism of cooperative coevolution with penalty factor to optimize the solution set and punishment factor. The simulation result shows that the proposed method has good convergence and fault-tolerant ability when single point or multi-points fault occurs in the distribution network. In addition, it can provide important technical support for the fault position in the distribution network based on the transient wave data.

Keywords- power distribution network; fault location; differential evolution algorithm; penalty factor; transient wave record

1. Introduction

The operating conditions of the distribution network are diverse and the structure is complex. When a fault occurs, it is difficult for the dispatcher to accurately find the point of the fault, thereby reducing the reliability of the power supply. The power distribution automation system needs to accurately locate the fault point, and troubleshoot as soon as possible to reduce economic losses. At present, the widely used fault indicators and automatic terminals such as FTU/DTU are mostly used to judge the fault by the steady-state quantity. For the medium-voltage distribution network with 10 kV neutral point ungrounded or non-effectively grounded operation, the fault characteristic quantity is not obvious, so that the actual fault identification and segment location effect is limited.

At present, in the development process of intelligent distribution network, the fault detection of distribution network is also becoming more intelligent\textsuperscript{[1-2]}. For example, transient wave recording devices are installed at key points of distribution network to collect fault information in real time, so as to provide analysis data for subsequent fault location. Domestic and foreign research methods for distribution network fault location are roughly divided into two categories: algorithms based on graph theory and algorithms based on artificial intelligence\textsuperscript{[3-6]}. The matrix method is the most representative algorithm in the graph theory algorithm. It describes the structural characteristics of the corresponding distribution network, expresses its information data in the form of a matrix, and then compare with the information obtained by the detection device to generate a fault matrix for fault location. The typical algorithms include unified matrix algorithm\textsuperscript{[7]} and overheat arc search method\textsuperscript{[8]}, but such algorithms have poor fault tolerance and inaccurate results in the event of interference. Artificial intelligence algorithms are also known as indirect optimization algorithms. The typical algorithms include expert systems\textsuperscript{[9]}, ant colony algorithms\textsuperscript{[10]}, artificial neural networks\textsuperscript{[11]}, fuzzy theory\textsuperscript{[12]}, genetic algorithms\textsuperscript{[13]}, etc. Compared with the matrix method, the artificial intelligence algorithm has better fault tolerance, but it is not easy to deal with the fault location of the complex distribution network due to the influence of the calculation speed. Therefore, these two types of algorithms are not well compatible in terms of accuracy and fault tolerance. In addition, the distribution network is affected by conditions such as operating environment, device power supply and communication, and the data quality of fault information varies greatly, which leads to a decrease in the convergence of these two types of algorithms in practical applications, and the application effect is not ideal.

As a heuristic optimization algorithm, differential evolution (DE) was first used to deal with the problem of Chebyshev polynomials \textsuperscript{[14]}. Based on the evolution of species in bionics, the algorithm regards the search range as a population, and
realizes selection and optimization through variation and crossover within the population. Due to the good effect of the DE algorithm on complex optimization problems\cite{15}, this paper proposes an improved method of the algorithm, the co-evolutionary DE algorithm, which is used in fault location of distribution networks. The algorithm uses the coevolution mechanism of penalty factors to optimize the solution set and penalty factors synchronously, so as to improve the fault tolerance of the algorithm to the transient data of distribution network. An example is given to verify the accuracy of the algorithm in the fault location of distribution network.

2. Principle of Co-evolutionary DE Algorithm

2.1. DE algorithm

The evolution rules of DE algorithm are as follows:

Step 1: Set the initialization group as \( X \), then each individual is solved by formula (1)

\[
X_i = X_i^U + R \cdot (X_i^U - X_i^L)
\]

where \( i = 1, 2, ..., N \), \( N \) is the population size, \( R \) is a uniformly distributed random number between \([0,1]\); \( X_i^U \) and \( X_i^L \) are the upper and lower bounds of the \( i \)-th individual, respectively.

Step 2: The mutation operation of the differential evolution algorithm generates a mutation individual \( V_{i,j} \) according to the differential genes between randomly sampled individuals, and the mutation mechanism is as follows:

\[
V_{i,j} = X_{Best,j} + U \cdot (X_{t_1,j} - X_{t_2,j}) + U \cdot (X_{t_3,j} - X_{t_4,j})
\]

where \( X_{Best,j} \) is the optimal individual in the evolution of the \( t \)-th generation population, and together with the 4 individuals \( X_{t_1,j} \), \( X_{t_2,j} \), \( X_{t_3,j} \) and \( X_{t_4,j} \) randomly selected from the \( t \)-th generation population, it is used as the determinant for generating the variant individual \( V_{i,j} \), and the variation factor \( U \) is taken as 0.5.

Step 3: In order to increase the diversity of the population, use the random vector \( V_{i,j} \) and the target individual \( X_{i,j} \) in the mutation operation to perform a hybrid operation to generate a new individual \( X'_{i,j} \). The \( j \)-th element \( X'_{i,j} \) in \( X'_{i,j} \) satisfies the following formula:

\[
x'_{ij} = \begin{cases} x_{ij}, & R \leq C_R \text{ or } j = j_r \\ V_{ij}, & \text{others} \end{cases}
\]

where \( j \in [1, 2, ..., d] \) is the element subscript of the individual population, \( d \) is the problem dimension; \( V_{ij} \) is the \( j \)-th element of the \( t \)-th generation mutant individual \( V_{i,j} \); \( x_{ij} \) is the \( j \)-th element of the \( t \)-th generation target individual \( X_{i} \); \( R \in [0,1] \) is a random number; \( C_R \) is the crossover factor, which can be 0.1; \( j_r \) is an integer randomly selected in \([1, 2, ..., d]\) to ensure that there is at least 1 element variation in the target individual.

Step 4: Greedy selection of the newly generated individual \( X'_{i,j} \) and the current individual \( X_{i,j} \):

\[
X_{i,j+1} = \begin{cases} X'_{i,j}, & F(X'_{i,j}) < F(X_{i,j}) \\ X_{i,j}, & F(X'_{i,j}) \geq F(X_{i,j}) \end{cases}
\]

where \( F(X) \) is the fitness function.

2.2. Co-evolutionary mechanism

In order to further improve the search speed of the algorithm, this paper introduces a new mechanism, the co-evolutionary mechanism. The penalty factor population and the population to be optimized co-evolve, and combined with the DE algorithm, the solution population of the objective function and the solution population of the penalty factor are alternately evolved, and the algorithm terminates when the termination condition is satisfied.
In Fig.1, $W$ is the penalty factor population, including $nw$ individuals; $T$ is the population to be optimized, which consists of $n$ individuals, and corresponds to the solution of the objective function one-to-one. When co-evolution is carried out, the evolutionary objects are the population individuals to be optimized and their corresponding penalty factor individuals, and the DE algorithm is used to evolve to the $G$ generation; all individuals in the population $T$ have completed the evolution, and synchronization will obtain the evaluation value for evaluating the penalty factor. Then the evolution operation is performed on the penalty factor population to obtain the reconstituted penalty factor population $W_{i,j+1}$. For the new penalty factor population $W_{i,j+1}$, the solution population evolution process of $G$ generation is implemented.

2.2.1. Co-evolutionary mechanism

According to the distribution network fault location problem, assuming its constraints $g_i(s)$ to be:

$$g_i(s) \leq 0 \quad i \in [1,m]$$

(5)

The penalty function design method is shown in [16], and the criterion for its evaluation is that the smaller the value is, the better the individual is. The penalty function has the form:

$$F_f(X_{i,g}) = F(X_{i,g}) + (\alpha \times \text{sum} \times \eta_1 + \alpha \times \text{num} \times \eta_2)$$

(6)

where $F(X_{i,g})$ is the original objective function; $\text{sum}$ is the total amount of violations of constraints; $\text{num}$ is the number of violations of constraints; $\eta_1$, $\eta_2$ are penalty factors; $\alpha$ is the relationship between the objective value and the penalty function, so that The order of magnitude of the two is the same, and the variation interval of the penalty factor is set to $[1,100]$.

The calculation method of $\text{sum}$ satisfies the following formula:

$$\text{sum} = \sum_{j=1}^{k} g_j(s) \quad \forall g_j(s) > 0$$

(7)

2.2.2. Evaluation of penalty factor populations

When the solution population of the desired objective function evolves for $G$ generations, the criterion for evaluating the obtained value is that the smaller the value, the better the penalty factor. Set $F(L)$ as the penalty factor individual evaluation function.
(1) If there is at least one feasible solution in the solution population corresponding to the penalty factor $L_i$, the corresponding evaluation value can be expressed as:

$$F(L_i) = \sum_{j} F_j - \text{num} F_j$$

where $\sum F_j$ is the sum of all feasible solutions; $\text{num} F_j$ is the set of all solution populations corresponding to the penalty factor $L_i$.

(2) If there is no feasible solution in the solution population corresponding to the penalty factor $L_i$, the corresponding evaluation value can be expressed as:

$$F(L_i) = \sum \text{num} + \sum \text{sum} + \max(L)$$

where $\sum \text{num}$ is the sum of all the violations of constraints; $\sum \text{sum}$ is the sum of the sizes of violations of constraints; $\max(L)$ is the worst value of the evaluation value among the individuals of the penalty factor.

According to the evaluation value judgment standard, when an individual violates the constraints, its value will be greater than the evaluation value corresponding to the feasible solution; and the fewer individuals that violate the constraints, the better, and the search area can be narrowed.

### 3. Fault Location Algorithm for Distribution Network Based on Improved DE Algorithm

Use the transient wave recorder to monitor the current in the distribution network, and compare it with the set setting current to generate discrete data consisting of 0 and 1, which are stored in the array $X$ as fault information. 0 indicates that the monitoring equipment is in normal working state, and 1 indicates that the fault current is detected there, indicating that the feeder is faulty.

For a simple distribution network, the constructed objective function $F$ is:

$$F = \sum_{j=1}^{N} |X(j) - Y(j)| + \mu \sum_{j=1}^{N} |Z(j)|$$

where $X(j)$ is the $j$-th element stored in the information array $X$, reflecting the actual state value; $Y(j)$ is the $j$-th element stored in the information array $Y$, reflecting the expected state value; and in order to improve the accuracy of fault diagnosis, the weight coefficient $\mu$ is introduced, and its value range is $[0,1]$; $\sum_{j=1}^{N} |Z(j)|$ is the total number of faulty devices.

For the distribution network with higher complexity (as shown in Fig. 2), the constructed objective function $F$ is different from the simple distribution network, where $S$ is the network node equipped with the transient wave recording device. If a fault occurs at $L_{12}$ in Fig.2, the fault current will be detected between the fault point and power source $a$ and between the fault point and power source $b$ at the same time. When the distribution network is running in closed loop, the direction of the current in the network needs to be set in advance. When the power supply is at both ends of the line, set the power inflow direction as forward. If the monitoring current flows in the same direction, define the status value as 1, otherwise it is 0. For a complex distribution network, the objective function $F_S$ can be described as:

$$F_S = p_1 \cdot (F_{(n\rightarrow m)} + F_{(m\rightarrow n)}) + p_2 \cdot F_2$$

where $F_{(n\rightarrow m)}$ is the objective function when the power supply port of area 1 is $n$, and the positive direction of the network is set as the objective function from $n$ to $m$; $F_{(m\rightarrow n)}$ is the positive direction of the network when the power
supply port of area 1 is \( m \), and the positive direction of the network is set as the objective function from \( m \) to \( n \); \( F_2 \) is the objective function of region 2; the value of \( p_1 \) is 1 when a node in area 1 is detected to be faulty, and 0 when no fault occurs; the value direction of \( p_2 \) is the same as that of \( p_1 \).

![Fig. 2 Complex distribution network structure](image)

4. Case Study

Taking the distribution network shown in Fig. 2 as a simulation example, the MATLAB software is used to compile the distribution network fault location program based on the differential evolution algorithm. The parameters are set as follows: the population size of the algorithm \( q = 20 \), the maximum number of iterations \( G_{\text{max}} = 100 \), the penalty factor \( \eta_1 = \eta_2 = 2 \).

In order to verify the validity and fault tolerance of the distribution network fault location based on DE, several typical fault conditions are used for simulation tests, which are divided into two types: single point fault and multi-point fault. The corresponding simulation results are shown in Tab. 1 and Tab. 2.

Tab. 1 and Tab. 2 include two parts of input quantity and output quantity, and the input information is fault upload information composed of 0 and 1 signals. The fault information includes all monitoring points and line segments in area 1 and area 2 in the complex distribution network in Fig. 2. These are discussed separately below.

<table>
<thead>
<tr>
<th>Upload fault information</th>
<th>Fault line</th>
<th>Distortion switch</th>
<th>Evaluation value</th>
</tr>
</thead>
<tbody>
<tr>
<td>([1,1,1,1,0,0,1,1,1,1,0,0,0,1,1,1,0,1,0,1,0,0,0,1,1,1,0,0,1])</td>
<td>L16</td>
<td>/</td>
<td>0.5</td>
</tr>
<tr>
<td>([1,0,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1,0,0,0,1,1,1,0,0,1])</td>
<td>L16</td>
<td>S12, S14</td>
<td>2.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Upload fault information</th>
<th>Fault line</th>
<th>Distortion switch</th>
<th>Evaluation value</th>
</tr>
</thead>
<tbody>
<tr>
<td>([1,1,1,1,1,1,1,1,1,1,0,0,0,1,1,1,0,1,0,1,0,0,0,1,1,1,0,0,1])</td>
<td>L16, L17</td>
<td>/</td>
<td>2.5</td>
</tr>
<tr>
<td>([1,1,1,1,1,1,1,1,1,1,0,0,0,1,1,1,0,1,0,1,0,0,0,1,1,1,0,0,0])</td>
<td>L16, L17</td>
<td>S23, S27</td>
<td>5.7</td>
</tr>
</tbody>
</table>
In the first fault situation in Tab. 1, it is assumed that there is only a single point of fault in the grid structure, and DE is used to screen the known input information, and it is judged that the faulty line is L16, and no switch has information distortion. The evaluation value of its fitness function is 0.5. This example describes the evaluation value of a minor fault in the distribution network, indicating that the algorithm proposed in this paper can locate the general single-point fault.

In the second fault situation in Tab. 1, the DE algorithm is also used to diagnose the uploaded fault information, and a single-point fault is detected, the fault line segment is L16, and the information of the area 1 switches S12 and S14 corresponding to the information bits is also monitored. Distortion occurs, and the corresponding fitness evaluation value is 2.7, which is higher than that of pure line segment single-point fault, so it is diagnosed as an intermediate fault condition.

When multiple faults occur in the distribution network, in order to comprehensively monitor the fault points, the accuracy of the monitoring technology needs to be improved. The multi-point fault detection and verification are carried out through the example in this paper, and the results are shown in Tab. 2.

In the first fault situation in Tab. 2, the fault information is monitored by the monitoring method in this paper, and the final diagnosis is that the fault points are the line segments L16 and L17 in area 1. At this time, there is no switch distortion, and the evaluation value is 2.5, and the phase is 2.5. Compared with the single point fault without switch distortion, it belongs to a complex fault type, but compared with the single point fault with switch distortion, it belongs to a lighter fault type. It shows that the processing of switch distortion needs to be more real-time, because its impact on the distribution network is higher than that of the line.

In the second fault situation in Table 2, it is also a multi-point fault, but there are two switches that are distorted, namely S23 and S27 in area 2. The evaluation value at this time is 5.7. Compared with the multi-point fault without switching distortion, the evaluation value increases sharply, which fully proves the influence strength of the switching distortion fault.

From the test results in Tab. 1 and Tab. 2, it can be seen that the fault location algorithm proposed in this paper can perform accurate fault location in the case of single-point fault, multi-point fault and information distortion.

5. Conclusion

This paper proposes to use the improved DE algorithm for more accurate fault location in the process of using transient recording data for fault location in distribution network. The algorithm uses the co-evolution mechanism of penalty factor, and optimizes the solution set and penalty factor at the same time. It provides important technical support, and also provides an implementation method for the advanced intelligent analysis function of the main station of the distribution network. At the same time, through the simulation example, it is further demonstrated that the algorithm can effectively locate single-point or multi-point distribution network faults, and has certain engineering application value. The calculation example uses the number of fault points in the distribution network and the presence or absence of switch distortion as the scene classification, and analyzes four types of fault types. The degree of influence of distortion is greater, and the distribution network needs to strengthen the monitoring of each line switch to ensure the safe and stable operation of the entire network.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51277029).

References


Research on Numerical Simulation of Deep Seabed Blowout and Oil Spill Range

Yi Huang¹,a, YuHang Zhao²b*, YuLong Han¹c, Bin Zhong¹d, YangFan Lao¹c, Di Wu²f
¹Cnooc (China) Co., LTD Hainan Branch, Hainan, China
²China University of Petroleum (Beijing), Beijing, China

Abstract

Exploration and development in the far-reaching sea of the South China Sea faces many challenges, such as great difficulty in well control. Once an oil spill accident occurs, it is necessary to deal with the oil spill from submarine wells in time to avoid major environmental pollution. In order to study the spreading range of submarine blowout oil and its influencing factors, this paper uses the Fluent platform, combining with the standard $k-\varepsilon$ turbulence model and the multiphase flow VOF model, to preliminarily establish a submarine blowout oil spill model. This paper also analyzes the influence of ocean current velocity, blowout velocity, oil spill density on underwater migration trajectory and oil spill diffusion range. Numerical simulation results show that under the conditions of higher ocean current velocity, lower blowout velocity and greater oil spill density, the oil has a long underwater migration time, and the position where the oil first floats to the sea surface is far away from the wellhead level and will cause a greater range of oil spill pollution. This research is of practical significance for the rescue and rescue of blowouts and the emergency prediction and disposal of oil spills on deep water offshore platforms.

Keywords-Deepwater well control, Numerical Simulation; Submarine blowout, Oil spill range, Blowout velocity

1. Introduction

The deep water area in the South China Sea is rich in oil and gas resources. Water depth sometimes exceeds 1500m and drilling safety faces huge challenges. Due to the ultra-high formation pressure and the cumbersome well control technology, a little carelessness will cause a blowout, and a large amount of crude oil will overflow the wellhead. The 2010 Deepwater Horizon accident [1-3] in the Gulf of Mexico reflected the importance of accurately predicting the scope of deep-sea well blowout and oil spills for emergency response. In order to provide sufficient help for blowout rescue, this article will simulate deep sea blowout and analyze the spread of oil spill under different sea conditions and oil spill spout conditions.

At present, people's researches on oil spills mainly focus on the behavior and fate of oil spilled on the sea surface, and there are few studies on the migration, diffusion and fate of oil spilled under deep sea water. Foreign scholars' research on underwater oil spill simulation can be traced back to the 1970s. Winiarski et al. [4] proposed to use the Lagrange method to simulate water plumes in the 1970s; Yapa et al. [5-6] and Johansen et al. [7-8] established the CDOG model and the DEEP FLOW underwater oil spill model and these two oil spill models have achieved good results in deep-water underwater oil spill simulation.

Domestic scholars have also carried out in-depth research on underwater oil spill simulation. Wang Shoudong et al. [9] established a mathematical model of underwater pipeline oil spill to simulate the radiation flow formed by underwater oil spill; Wang Jing et al. [10] proposed a prediction model for oil spill pollution of underwater pipelines in the Bohai Sea, which is used to predict the floating behavior of small oil droplets after worm oil spills in underwater pipelines; Liao Guoxiang et al. [11] established an underwater oil and gas spill pollution source transport prediction model, which can simulate the common transport and separate transport process of oil and gas mixture in the near underwater area.

In view of the fact that current underwater oil spill models are mostly used in the simulation of oil spills from submarine pipelines, there are few domestic and foreign studies on applying oil spill models to the simulation of deep sea submarine blowout oil spills. This paper will use fluent software to preliminarily establish a practical oil spill model of deep-sea wells.
The model comprehensively considers the effects of ocean current velocity, blowout velocity, and oil spill density, simulates oil spill diffusion under different sea conditions.

2. Numerical Simulation Method of Deep Sea Blowout and Oil Spill

2.1 Mathematical model and boundary conditions

In order to facilitate the description of the oil spill mechanism on the seabed, this paper simplified the actual three-dimensional ocean model to two-dimensional, and used basic governing equations such as fluid continuity equation, momentum equation and turbulence equation in the model to ensure the reliability of the model.

The plan to build a 500m*500m two-dimensional numerical water tank is shown in Figure 1. The calculated sea area length is \( L = 500 \text{m} \), and the sea water depth is \( H = 500 \text{m} \) \((O-y \text{ direction})\). The blowout oil spill point is set on the seafloor plane \((x\text{-axis})\), and a hole with a diameter of \(d=0.244 \text{m} \) is set on the seafloor as the oil spill location. The sea current velocity direction is parallel to the seafloor plane, and the angle between the initial oil spill velocity and the seafloor plane is 90°. In order to simulate the real sea state more accurately, the ocean current velocity is calculated by the following formula:

\[
v_{\text{current}} = v_{\text{max}} \left[1 - \left(1 - \frac{y}{H}\right)^2\right]
\]

(1)

Where \( v_{\text{max}} \) is the maximum velocity of the ocean current.

In Figure 1, the velocity inlet boundary is adopted for the current inlet and oil spill nozzle. The wall boundary is adopted for the seabed plane. The free outflow boundary is adopted for the right boundary (BC), and the symmetrical boundary is adopted for the upper boundary (AB).

2.2 Basic equation

(1) Continuity equation. According to the law of conservation of mass in fluid mechanics, the mass of the fluid added inside the control body should be equal to the mass of the fluid flowing into the control body minus the mass of the fluid flowing out of the control body. The expression is:

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0
\]

(2)

Where \( \rho \) is the oil spill density. \( \vec{v} \) is the velocity vector of the oil spill at the point \((x, y, z)\) at time \(t\), and the velocities in the \(x, y, z\) directions are \(u, v, w\) respectively. This article regards spilled oil as a homogeneous incompressible fluid with a constant density, so:

\[
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0
\]

(3)
(2) Momentum equation. The total external force acting on the control volume is equal to the difference between the momentum of the liquid flowing out of the control surface and the liquid flowing out of the control surface in time. Also known as the N-S equation, the expression is:

\[
\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot [(\rho \vec{v}) \vec{v}] = \vec{F} - \text{grad}P + \mu \Delta \vec{v}
\]  

(4)

Where \(\rho\) is the oil spill density, \(P\) is the pressure, \(\vec{v}\) is the velocity vector of the oil spill at the point \((x, y, z)\) at time \(t\), and the velocities in the \(x, y, z\) directions are \(u, v, w\) respectively. \(\vec{F}\) is the external force per unit volume of fluid. \(\mu\) is dynamic viscosity.

(3) Turbulence equation. In this paper, we consider that the state of drained oil flowing in seawater is turbulent. Using the standard \(k-\epsilon\) model provided by Fluent, the equations of turbulent kinetic energy \(k\) and turbulent dissipation rate \(\epsilon\) are:

\[
\frac{\partial (\rho k)}{\partial t} + \frac{\partial (\rho u_i k)}{\partial x_i} = \frac{\partial}{\partial x_i} \left[ \left( u_i + \frac{u_i u_i}{\sigma_k} \right) \frac{\partial k}{\partial x_i} \right] + G_k + G_B + \rho \epsilon - Y_M + S_k
\]  

(5)

\[
\frac{\partial (\rho \epsilon)}{\partial t} + \frac{\partial (\rho u_i \epsilon)}{\partial x_i} = \frac{\partial}{\partial x_i} \left[ \left( u_i + \frac{u_i u_i}{\sigma_{\epsilon}} \right) \frac{\partial \epsilon}{\partial x_i} \right] + C_{1\epsilon} \frac{\epsilon}{k} (G_k + C_{3\epsilon} G_B) - C_{2\epsilon} \rho \frac{\epsilon^2}{k} + S_{\epsilon}
\]  

(6)

Where \(G_k\) is the turbulent kinetic energy produced by the laminar flow velocity gradient. \(G_B\) is the turbulent kinetic energy produced by buoyancy. \(Y_M\) is the fluctuation due to transitional diffusion in compressible turbulence. \(C_{1\epsilon}, C_{2\epsilon}, C_{3\epsilon}\) are constants. \(\sigma_k, \sigma_{\epsilon}\) is the turbulence Prandtl number of \(k\) equation and \(\epsilon\) equation. \(S_k, S_{\epsilon}\) are user-defined data.

2.3 Discrete method and meshing

The numerical study of oil spill from deep-sea subsea wells is carried out by fluent. The pressure based solver is adopted and the PISO method is selected as the coupling method. This method has a better solution to unsteady problems and solves the continuity equation and momentum equation at the same time, which speeds up the convergence speed. The convection term interpolation method adopts the QUICK mode, and the other settings adopt the default settings.

This model uses a triangular grid to partially refine the grid near the oil spill well head, so as to more accurately obtain fluid movement near the oil spill. The grid division is shown in Figure 2.

![Fig 2 Schematic diagram of meshing results](image)

3. Numerical simulation results and analysis

In this paper, the effects of different ocean current velocity, blowout velocity and oil spill density on the final oil spill range are considered comprehensively. First, at the same maximum current velocity of 0.2m/s, the velocity and range of oil spill passing through the monitoring time under different blowout velocities of 15m/s, 20m/s and 25m/s were simulated. Second, under the condition of the same blowout velocity of 20m/s, the influence of different current maximum velocity of 0.2m /s, 0.5m/s and 1.0m/s on the oil spill scope is studied. Thirdly, the time and range of oil spill to the surface with different oil densities of 0.75g/cm³, 0.80g/cm³ and 0.85g/cm³ are simulated under the same well blowout velocity of 20m/s and the same current velocity of 0.5m/s.
3.1 The influence of different blowout velocities

The blowout velocity represents the initial momentum of the oil spill. The approximate process of the spilled oil floating from the seabed to the sea surface under different water current speeds is shown in the figure below.

<table>
<thead>
<tr>
<th>Time</th>
<th>15m/s</th>
<th>20m/s</th>
<th>25m/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>20s</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>30s</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>60s</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>120s</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

As the oil spill rate increases, the jets of the oil spill become more and more obvious. The continuity of the oil spill has also become stronger. The time to reach the sea surface is shorter and shorter, and the state of the spilled oil reaching the sea surface has also become scattered. The oil droplets slowly become continuous oil masses or bands. It can be seen that when the oil spill occurs, although the higher oil spill rate is easy to be found, the spread of the oil on the sea is larger in this case.
3.2 The influence of different ocean current speeds

The ocean current plays a key role in the horizontal drift distance of the oil flow. The approximate process of the spilled oil floating from the seabed to the sea surface under different water current speeds is shown in the figure below.

<table>
<thead>
<tr>
<th>Time</th>
<th>0.2m/s</th>
<th>0.5m/s</th>
<th>1.0m/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>20s</td>
<td><img src="https://example.com/figure1.png" alt="Image" /></td>
<td><img src="https://example.com/figure2.png" alt="Image" /></td>
<td><img src="https://example.com/figure3.png" alt="Image" /></td>
</tr>
<tr>
<td>30s</td>
<td><img src="https://example.com/figure4.png" alt="Image" /></td>
<td><img src="https://example.com/figure5.png" alt="Image" /></td>
<td><img src="https://example.com/figure6.png" alt="Image" /></td>
</tr>
<tr>
<td>60s</td>
<td><img src="https://example.com/figure7.png" alt="Image" /></td>
<td><img src="https://example.com/figure8.png" alt="Image" /></td>
<td><img src="https://example.com/figure9.png" alt="Image" /></td>
</tr>
<tr>
<td>120s</td>
<td><img src="https://example.com/figure10.png" alt="Image" /></td>
<td><img src="https://example.com/figure11.png" alt="Image" /></td>
<td><img src="https://example.com/figure12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

It can be seen that when an oil spill occurs, the higher ocean current speed will cause the spilled oil to drift farther in the sea. The pollution range is larger, and the time for the spilled oil to float to the sea surface is longer, and the spilled oil is more difficult to be detected.
3.3 The influence of the density of different spilled oil products

The density of the spilled oil represents the effect of buoyancy on the results. The approximate process of the spilled oil floating from the seabed to the sea surface under different oil spilled density is shown in the figure below.

<table>
<thead>
<tr>
<th>Time</th>
<th>0.75 kg/cm³</th>
<th>0.80 kg/cm³</th>
<th>0.85 kg/cm³</th>
</tr>
</thead>
<tbody>
<tr>
<td>20s</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>30s</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>60s</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>120s</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

As the density of the spilled oil increases, the time it takes for the spilled oil to float to the surface will become longer and longer, and the distance that the spilled oil drifts horizontally will become further and further. It can be seen that when the oil spill occurs, the dense oil diffuses farther in the sea, rises slowly, and causes more pollution.
4. Conclusion

In this paper, through the numerical simulation of oil spill from submarine wells, the possibility of applying the oil spill model to the control of oil spill range of submarine wells is verified, which provides a theoretical basis for the emergency treatment of oil spill accidents in far-reaching offshore wells. Numerical simulation results show that the blowout first relies on the eruption velocity to rise, and then relies on the buoyancy of the oil to float to the sea surface. As the eruption speed increases, the time for the spilled oil to float to the sea surface is gradually shortened. The maximum eruption velocity and the minimum eruption velocity to the sea surface time are shortened by about 7.6%, and the oil spill radius is reduced by 5.8%. As the ocean current speed increases, the time for oil to float to the surface of the sea is almost the same, but the significant increase in the radius of the impact is almost proportional to the sea current speed. The greater the density of oil, the higher the distance to rise by eruption, and the distance between the highest and lowest density oil rises is about 2.4%. However, as the speed becomes 0, relying on buoyancy to continue to rise, the rising speed of low-density oil is significantly higher than that of high-density oil. On the whole, the rising time of low-density oil is 10.9% faster than that of high-density oil.

This paper only simulates the oil spill in 500m water depth. In future research, the water depth can be used as a factor to analyze the oil spill characteristics under different water depths, refine the boundary conditions of the model, and consider establishing an oil spill model of submarine wells under the joint action of multiple factors.

Fund Items

The research on tube-soil interaction mechanism and drilling parameter control method in the process of pipe jetting in deep water Drilling, No.2020M670581; Analysis and control method of time-varying Bearing capacity of deep water subsea Wellhead coupled with dynamic Action, No.2021T140736; Study on the mechanism of gas hydrate formation decomposition and weakening during drilling process (2462021JYRC006)

References


Numerical simulation on compressive strength of waste fiber recycled concrete based on Python

Jinghai Zhou$^{1,*}$, Hanglin Yu$^{2,b}$, Junwen Lu$^{3,c}$, Tianbei Kang$^{4,d}$, Yue Qi$^{5,e}$

$^1$School of Civil Engineering, Shenyang Jianzhu University Shenyang, China
$^2$School of Civil Engineering, Shenyang Jianzh University Shenyang, China
$^3$School of Civil Engineering, Shenyang Jianzhu University Shenyang, China
$^4$School of Civil Engineering, Shenyang Jianzhu University Shenyang, China
$^5$School of Civil Engineering, Shenyang Urban Construction University, Shenyang, China

$^*{\text{Corresponding author: zhoujinghai@sjzu.edu.cn}}$
$^b{\text{e-mail: 1461216215@qq.com}}$
$^c{\text{e-mail: 656360596@qq.com}}$
$^d{\text{e-mail: kangtianbei@sjzu.edu.cn}}$
$^e{\text{e-mail: 821447504@qq.com}}$

Abstract

At present, the mesoscale simulation of waste fiber recycled concrete is regarded as a two-phase medium composed of recycled concrete and waste fiber. It is only studied in two dimensions, the three-dimensional fine and microscopic model of waste fiber recycled concrete is not established. In this paper, mesoscale numerical simulations on the compressive strength of waste fiber recycled concrete are performed based on Python. The recycled concrete is considered as a multiphase body consisting of aggregates, old and new mortar, and interface. A random distribution model of recycled aggregates based on the Monte Carlo method is written in Python language. Thus, the mesoscale three-dimensional (3D) numerical model of waste fiber recycled concrete is established. By comparing the results of the compressive strength test with the simulation, it can be seen that the model based on Python can more accurately reflect the damage pattern and compressive strength of waste fiber recycled concrete under pressure.

Keywords—recycled aggregates; Waste fiber; Python; Numerical simulation

1. INTRODUCTION

The popularization and application of recycled concrete has become an upsurge. The traditional test is limited by human factors, test conditions, temperature differences and so on, which leads to the stagnation of fiber recycled concrete. The rational selection of computer model based on Monte Carlo method and the preparation of Python script program can not only solve the defects of the traditional test, but also effectively connect the micro and macro structures.

The emergence of large-scale electronic computers and the rapid development of computer technology provide effective calculation guarantee for the study of the mesoscale morphology of recycled concrete. Sadouki [1] and Cusatis [2] et al. assumed that in the study of aggregate shape in mesoscale simulation, the aggregate particles are assumed to be circular. Wriggers [3] et al. used the finite element method to model the concrete and randomly generated spherical coarse aggregate with different particle sizes and positions based on Monte Carlo method, and predicted the crack propagation through the plastic damage constitutive relationship. The research showed that the crack first occurred around the aggregate and then spread in the mortar. Bazant [4] et al. proposed that the random aggregate model is a numerical model established by defining different mechanical parameters of various components of concrete through finite element software. In order to improve the calculation efficiency, Liu J [5] proposed an efficient calculation method of mesoscale element equivalent model based on the random aggregate model, and subdivided the materials that make up the concrete including cement mortar, interface transition zone, aggregate, pores and microcracks. Dang N [6] et al. applied the nonlinear finite element method to simulate the damage, and the results showed that the uniaxial compressive strength of the specimen was in good agreement with the test results. Liu GT [7] et al. proposed that concrete could be regarded as a three-phase composite composed of aggregate, mortar matrix and the bonding surface between aggregate and mortar matrix.
In this paper, a 3D fiber reinforced concrete model is established based on the python software driven ABAQUS finite element platform. On this basis, it is compared with the compressive performance test to fully verify the applicability of the established waste fiber recycled concrete model.

2. NUMERICAL MODE OF RECYCLED CONCRETE WITH WASTE FIBER

2.1 Constitutive relation

The plastic damage model is adopted, and the compression constitutive model of recycled concrete adopts the fitting formula proposed in document [8].

\[
y = \begin{cases} 
  a_x x + (3 - 2a_x)x^2 + (a_x - 2)x^3 & 0 \leq x \leq 1 \\
  \frac{x}{a_x(x-1)^2 + x} & x > 1 
\end{cases}
\]  

(1)

Where: \( x = \varepsilon / \varepsilon_c \), \( y = \sigma / f_c \). In formula, \( a_x \) and \( a_y \) are the parameter values on the stress-strain curve under uniaxial compression; \( \varepsilon, \varepsilon_c \) is the strain and peak strain, respectively; \( \sigma \) by \( \varepsilon \) corresponding stress; \( f_c \) is the axial compressive strength.

The compression constitutive relation of fiber reinforced recycled concrete adopts the fitting formula proposed in literature [9].

\[
y = \begin{cases} 
  \frac{kx - x^2}{1 + (k - 2)x} & 0 \leq x \leq 1 \\
  \frac{x}{b(x-1)^2 + x} & x > 1 
\end{cases}
\]  

(2)

Where: \( k \) and \( b \) are the dimension parameters of the test piece.

2.2 Model creation method

This paper establishes a fine observation model of waste fiber recycled concrete based on the random aggregate model. The strength of waste fiber recycled concrete is C30, the replacement rate of recycled aggregate is 50 %, and the volume incorporation of waste fiber is 0 and 0.12 %, respectively. The model size is 100 mm × 100 mm × 100 mm. Continuous cascade is used to calculate the distribution probability of a certain particle size range by equation (3) proposed in reference [10]. In this model, the volume of coarse aggregate accounts for 49.7 %. The distribution probabilities of each particle size are shown in Table 1.

\[
p_D(D<D_j) = p [0.06 \cdot \frac{P_j}{D_{max}} - 0.03 \cdot \frac{P_j}{D_{max}} - 0.008 + 0.002 \cdot \frac{P_j}{D_{max}} - 0.0004 \cdot \frac{P_j}{D_{max}} + 0.0008 \cdot \frac{P_j}{D_{max}}]^2
\]  

(3)

<table>
<thead>
<tr>
<th>Particle size</th>
<th>40mm</th>
<th>20mm</th>
<th>10mm</th>
<th>5mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_0/D_{max} )</td>
<td>1</td>
<td>0.5</td>
<td>0.52</td>
<td>0.125</td>
</tr>
<tr>
<td>( P_c )</td>
<td>0.497</td>
<td>0.378</td>
<td>0.269</td>
<td>0.189</td>
</tr>
</tbody>
</table>

The aggregate is randomly distributed, using the spherical model. The Monte Carlo method is used to determine the position of the ball model. The distribution of the ball model in 3D space is replaced by \((x, y, z, r_c)\), where \((x, y, z)\) represents the center coordinates of the ball, and \(r_c\) is the ball radius. Intersection method is adopted for aggregate interference. The waste fiber is represented by a slender cylinder. In Python, the cylinder parameters representing the waste fiber can be composed of two endpoints plus a radius: \((point1, point2, r_fiber)\), where \(point1\) and \(point2\) represent the coordinates of the two endpoints and \(r_fiber\) represents the radius of the fiber. Interferometric judgments of each entity are required in the process of obtaining fiber parameters. The mesoscale model of waste fiber recycled concrete is shown in Figure 1.
3. NUMERICAL SIMULATION AND ANALYSIS OF COMPRESSIVE STRENGTH

3.1 Destructive phenomena

The final failure modes of 9 groups of specimens with 50% recycled aggregate and 0.12% waste fiber were analyzed. The failure form of waste fiber recycled concrete is different from that of recycled concrete. In the loading process, the crack development pattern varies greatly and the crack width decreases significantly. In the early loading stage, the crack is produced in the middle side position and develops along the vertical direction. The crack shape is almost parallel to the lateral edges of the specimen. As can be seen from the figure, the crack development condition of the random aggregate model of waste fiber recycled concrete is not much different from the test phenomenon of waste fiber recycled concrete, and the crack production position starts from the middle side and develops along the direction of loading until the final destruction. Results We verify that the model is highly consistent with the test results and is reliable.
(c) $\alpha_{ag} = 0.242$

(d) $\alpha_{ag} = 0.249$

(e) $\alpha_{ag} = 0.251$

(f) $\alpha_{ag} = 0.262$

(g) $\alpha_{ag} = 0.273$
3.2 Effect of aggregate on strength of recycled concrete

If the uniformity distribution coefficient is consistent, the relationship between the uniformity distribution coefficient of aggregate and the compressive strength is shown in Figure 3. The relationship between the recycled aggregate replacement ratio and the compressive strength is shown in Figure 4.

Figure 2. Failure mode of waste fiber recycled concrete

Figure 3. Relationship between aggregate uniformity distribution coefficient and compressive strength

Figure 4. Relationship between recycled aggregate replacement ratio and compressive strength
It can be clearly seen from Figure 3 that the compressive strength curve of recycled concrete shows a downward trend as a whole. The larger the $\alpha_{ag}$, the more uneven the aggregate distribution. When the aggregate is concentrated in the recycled concrete model, the amount of aggregate in other areas is relatively reduced, and the strength is lower than that in the area where the aggregate is concentrated. So it is relatively weak. With the application of pressure, cracks will first appear in this area and expand rapidly, which will reduce the compressive strength of recycled concrete model. However, when the aggregate is evenly distributed in the recycled concrete model, the weak area decreases and the compressive capacity of recycled concrete increases.

It can be seen from Figure 4 that the compressive strength values of 9 groups of ordinary concrete models are greater than the compressive strength values of recycled concrete under the same aggregate uniformity distribution coefficient. The strength of concrete decreases continuously with the addition of recycled aggregate. The crushing process will inevitably cause damage around or inside the recycled aggregate, and the old hardened mortar attached around the recycled aggregate will produce pores and micro cracks, which will reduce the strength and form a weak area. In case of compression failure, this area will be damaged first and the compressive strength will be reduced.

### 3.3 Effect of waste fiber on compressive strength

The recycled aggregate replacement ratio is 50 %, and the volume of waste fiber was 0, 0.12 % and 0.24 %, respectively. The relationship between the amount of waste fiber and the compressive strength of waste fiber recycled concrete is shown in Figure 5. The relationship between the uniformity distribution coefficient of waste fiber and the compressive strength is shown in Figure 6.

![Figure 5. Relationship between the amount of waste fiber and the compressive strength](image)

![Figure 6. Relationship between the uniformity distribution coefficient of waste fiber and the compressive strength](image)

It is obvious from Figure 5 that when the volume content of waste fiber changes from 0 to 0.12 %, the compressive strength of recycled concrete with waste fiber increases steadily with the increase of its content. When the volume content changes from 0.12 % to 0.24 %, the compressive strength decreases gradually due to the concentrated distribution of fibers, resulting in damage and continuous reduction of the overall compressive strength. The trend of simulation data is consistent with the test data, which fully verifies the applicability of the model.

As can be seen from Figure 6, when the uniformity distribution coefficient of waste fiber $\alpha_{f} \leq 0.524$, the compressive...
strength of waste fiber recycled concrete is more than 31.9 MPa. When the waste fiber uniformity distribution coefficient of waste fiber $\alpha > 0.524$, the higher the coefficient, the more uneven and concentrated the distribution in the waste fiber recycled concrete model. The occurrence of "huddle together" phenomenon increases the number and area of weak areas, and the compressive strength of waste fiber recycled concrete has an obvious downward trend.

4. CONCLUSION

In this paper, the structure of waste fiber recycled concrete is analyzed on the meso-level. The number of aggregates in the model was calculated using the fuller gradation curve, and the aggregate coordinates, particle size and central point coordinates of waste fibers were generated by Monte Carlo Method randomization. Finally, Python Language and ABAQUS finite element software are used to establish the waste fiber recycled concrete model and carry out mesoscale numerical simulation. The specific research results are as follows:

1) The crack development process of the recycled concrete model with waste fiber under compression is similar to the experimental phenomenon. The position of the crack is from the middle side along the direction of loading until the final failure. The results fully verify the applicability of the waste fiber recycled concrete model.

2) The failure mode of recycled concrete test block with waste fiber is different from that of the test block without waste fiber, and the crack width of recycled concrete test block with waste fiber is significantly smaller than that of the test block without waste fiber during loading.

3) During the loading process, the crack development characteristics are also different. At the initial stage of loading, the crack vertically occurs at the middle side position and is approximately parallel to the side edge. With the increase of load until failure, the boundary area of the concrete specimen begins to peel off, the cracks are relatively less obvious, and the brittleness of the concrete is weakened.

In the actual test, the detailed composition of waste fiber recycled concrete is extremely complex. I still have many improper points. Further research is still needed. In actual trials, the shape of the aggregate is often very irregular and cannot be simply regarded as a circular sphere. With the current technology, CT X-ray's technology can realize the recycled concrete cross-section map under the real material, and It is more accurate. However, under the 3D conditions, due to the influence of computer ability, the grid division will be very complex. How to effectively divide the waste fiber recycled concrete model under real materials is the next in-depth research problem.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 52108235 and 51678374), Fundamental Research Funds for the Central Universities (Nos. 300102341511) and Liaoning Provincial Department of Education Fund (Nos.lnqn2020003; LT2019011). The financial support is gratefully acknowledged.

References


IoT and big data analysis based prevention and intervention network system for breast cancer susceptible people

Yi Qin¹, LeYin Liu²*, Bai Yang¹, Juan Luo¹, Yan Zhang¹, YiHan Liao³ and ZiHan Wang⁴

¹International Sakharov environmental Institute, Belarus state university, Lenin District 220071, Minsk City, Minsk Region, Belarus
²Foreign Languages Institute, Chifeng Institute, Inner Mongolia Autonomous Region 024000, China
³Department of Cytogenetics, Southern Medical University, Guangzhou 510000, Guangdong Province, China
⁴College of Chinese Medicine, Macao University of Science and Technology, Macao Special Administrative Region 999078, Macao, China

*Corresponding author’s e-mail: lydiagagliu@163.com

Abstract

The medical intervention after a disaster event is the focus of government and citizen’s attention in the public health and clinical medicine industries. For breast cancer, which has a high incidence in Russia and Belarus, the author combines the IoT management system and big data analysis to build a model of the framework and a preventive medicine system for a framework analysis and outlook.

Keywords: Big data; Breast cancer; Intervention Network System; IoT and big data

1. Introduction

(I). Research Task.

Health intervention: the concept is based on the constructive layer of social Internet of Things. Establish a multi-functional perception layer that can collect information about high-risk patients and set up functions: 1. breast lump detection in the monitoring area 2. radiation measurement in the range of patient's residence 3. barcode device location 4. life services.

Apply big data analysis technology to construct the collection plug-in that can collect the exposure or presence of morbidity factors related to the high-risk group and popularize it to the high-risk group. According to this client, construct big data analysis system with log collection function, storage engine, data cleaning, data mining and data visualization, which can be used for disease policy regulation conducted by the national health department [1]. Analyze the feasibility analysis of each industrial production chain in the context of international trade, make an outlook in the context of perceptual layer component production, and give a feasibility study of the cooperation model guided by national trade policy.

(II) Existing methods and problems.

Wearable devices already exist under IoT medicine, but they are not yet popularized on a large scale. Diagnostic methods for high-risk groups of breast cancer after daring. Clinical breast screening: It does not need to rely on radioactive elements and is performed by physicians. It is less sensitive in obese patients, but has a specificity of ninety-four percent. MAM: This method requires invasive operations and is not promoted in many countries. Breast ultrasonography: simplified, quick and reproducible, but the diagnosis of ductal infiltrating carcinoma is poor and relies more on the subjectivity of the physician [2]. Breast magnetic resonance phenomenon (MRI): it can be used for many patients, especially those who are difficult to be diagnosed clearly. Computer-aided technology, digital breast tomosynthesis; it is less used at present [3], and there is a lack of clinical experience as well as a high rate of false positives. Breast cancer ultrasonography robot: capable of providing highly accurate image navigation for breast biopsy procedures, but the internal high magnetic field environment and limited space restrict the application of the robot. 4G technology: a means of data transmission, one of the basic technologies to guarantee the country's stable development on the road to development. Radiation, ultraviolet detectors: Japan is studying the plastic production material detectors, and the current portable detectors are also promising transformation.
2. Main Challenges

Although IoT has made some progress in screening for breast cancer, there is no strong opportunity for transformation in the face of the slower economic development of Russia and Belarus regions, where medical resources are more concentrated in Moscow, St. Petersburg, Minsk and Novosibirsk regions, and other regions are not densely populated with medical resources, but the post-Chernobyl nuclear radiation areas are mainly concentrated in Vyjebuzhk region [4]. These cities are small cities that are not political and economic centers of the country. Without medical intervention through the Internet of Things, patients with high incidence of breast cancer in these areas have difficulty in knowing their condition, and many of them are diagnosed in the central city hospitals only at a late stage of the disease.

Our intervention approach needs to address several major challenges: low cost, simplicity, instant information, and large data collection. Collecting factors that affect genetic variation at the same time as nuclear radiation, establishing a basic big data network architecture for preventive medicine, setting up a separate biological big data information base, and collecting information including smoking, mood, and mental status in the simplest way and at the lowest cost.

3. Innovative Ideas

(1). The current way for breast cancer screening is more stereotyped in the city, mainly concentrated in the tertiary hospitals above the big cities, and the workload is larger and there are more conflicts between doctors and patients, we try to use IoT technology for the initial screening of breast cancer disease, which is beneficial with both patients and hospitals [5]. The innovation of this design can relieve the pressure of both sides.

(2). Establish a theoretical idea and conduct a simple social survey experiment according to this theoretical model. Explore whether this perceptual layer of devices attracts users with innovative features in terms of popular demand, and the degree of cooperation with the biomedical big data collection network. And to understand the degree of demand for setting up medical API third-party services.

(3). API third-party services are: joint intervention of Russian and Chinese medical experts for susceptible people [6]. This model is not available in this region. The intervention of API third-party services can save the energy of both sides, and Chinese experts can play a meaningful role in the academic exchange of medical science in the project of protecting the susceptible population in the neighboring countries, and jointly improve the understanding of the factors of breast cancer development and clinical treatment, which is conducive to the promotion of Chinese medicine globally with the support of the Belt and Road policy.

4. Specific Steps

(1). Randomly select 100 females in the region for the acceptance of the sensory layer devices.

(2). To construct and assign functional definitions to the simulated IoT signal transduction model of the sensory layer-transmission layer-application layer, to conduct theoretical exploration in this direction, as well as model optimization.

(3). to create a separate large database for predicting the incidence of breast cancer in the region, taking into account the cell phone HF software applications in Russia and Belarus.

5. Main Experimental Results

(1). The questionnaire was distributed in Russian to 100 randomly selected women in Russia and Belarus.

(2). There is a high level of recognition of the establishment of a bio-big data information base.

(3). The demand for API services is relatively high.

6. Design Methodology

6.1. Data collector of perception layer.

6.1.1 Device construction is simple, with an insulator as the collector appearance, the shape is similar to a brassiere, 4 quadrants are set on both sides, a total of 8 and can be, located on the left and right sides, the brassiere can be contracted...
after the power is turned on, the time is 10-30 seconds, but also can be controlled by itself. In the quadrant where the resistance is felt can transmit the signal model can be seen in Figure (1)
6.4. **Back office management center.**

6.4.1 Receiving the signals transmitted by the patient to store, retain and keep the data confidential, receiving the signals of the amount of radiation during solar charging and storing them by region.

6.4.2 Receive the information obtained from the breast cancer screening robot in each region and store it by region.

6.5. **Information processing center.**

6.5.1. Classify and process the information transmitted by the patients in the susceptible population determined by the medical institution and divide it into resistance signals that do not receive lump interference, with me resistance signals that receive lump interference, etc. and other signal abnormal signals, such as too large breast, too small breast, etc. The large data of each patient will be summarized as well as statistically analyzed, and if there are patients at risk after the analysis will be handled by the application prompt center.

6.5.2 Accept the information reported by patients in the breast cancer screening robot and process the information and transfer it to the API.

6.5.3 Accept information input from the API and transfer the information to the application prompt center after processing and classifying.

6.5.4 The signal processing center generates files to store in the backend management center after each calculation.

6.6 **The functions of the Application Cueing Center are divided into**

6.6.1 Connecting to API, sending application SMS and making smart phone calls: informing patients at risk and instructing them to first pay the smart robot breast cancer detection center to get medical reports and log in to the website to upload reports by SMS reminder, as well as information delivery from API services.

6.6.2 The application prompt center stores the medical reports transmitted by the API in the background management center, and sets up a query website based on each user ID in the application prompt center, and publishes the medical reports imported by the API for each ID, which are generated by experts for "alarm" pushing. Most of the medical reports are complicated and private, and the work with guidance should not be notified by the third party.

6.7 **API third-party services are**

6.7.1. Medical intervention in case of voluntary uploading by the patient, including the application of client programs, submission of data and informing the patient about the time of medical advice queries on the functional level: web interface
6.7.2. Consulting services are provided with a team of medical experts from China and Russia, who set medical advice and reply to it. After the reply, the third part API is connected, and the API pours the signal into the application center at the same time for backstage management center information storage, data processing center processing and then connected to the application service center for information publication.

6.7.3 After receiving the message "lump exists" from the sensory layer, the API imports the signal into the signal processing center, which is stored in the backend management center, and then the API provides services such as smart SMS and smart phone calls to the subjects located in the sensory layer, indicating the patient's risk value, and other service options.

6.8 Big data analysis structure for independent settings

In the Russian and Belarusian regions, the download rate of Yangdex go in the population is relatively high, which can be hypothesized as a functional model based on yangdex go, which can be seen in Figure (3).

![Figure 3. Big Data Analytics Structure Flowchart](image-url)

7. Combining the Findings of Health Economics

The advantages of this idea are: reduction of labor costs, saving of time for those who need to be tested, competitive low production prices in the process, the trend of medical cooperation between Russia and China, and the high support rate of people in the region in the context of experimental surveys, which is worthy of translation.

However, there are still some technical problems with this idea, such as pressurization equipment, algorithmic research of various procedures, etc., but based on the fact that this idea has a translational value, it should strengthen multidisciplinary cooperation and serve the health of people worldwide.

References

Numerical simulation study on impacts of the Xinhengsha Reclamation Project on salinity in the Yangtze Estuary

Hanghang Lyu, Junjie Bian, Yuchi Hao*, Runli Tao
CCCC National Engineering Research Center of Dredging Technology and Equipment Co., Ltd., Shanghai 200082, China
*Corresponding author’s e-mail: haoyuchi@cccc-drc.com

Abstract

The impacts of the Xinhengsha Reclamation Project on saltwater intrusion and freshwater resources in the Yangtze Estuary is simulated by Ecom-si. It is found that the salinity in the North Branch decreases slightly, the salinity in the North Channel and South Channel decreases obviously after the implementation of the Xinhengsha Reclamation Project. In addition, the implementation of the Xinhengsha Reclamation Project has a significant impact on the water intaking of the three reservoirs (Dongfengxisha Reservoir, Chenhang Reservoir, Qingcaosha Reservoir). In a spring-neap tide cycle, after the implementation of the Xinhengsha Reclamation Project, Dongfengxisha Reservoir is shortened from a maximum of 7 days before the project to 6 days, Chenhang Reservoir is shortened from 3.5 days to 0 days, Qingcaosha Reservoir is shortened from 4 days to 2.5 days. The above research results show that the implementation of the Xinhengsha Reclamation Project is beneficial to the water intaking of the three reservoirs.

Key words: reclamation project, saltwater intrusion, numerical model, freshwater resources, Yangtze Estuary

1. Introduction

Coastal cities have witnessed rapid economic development in recent years, and their demands for land increasing. Shanghai is the largest city in China, which is located in the Yangtze Estuary. Yangtze Estuary is a typical bifurcate estuary with four inlet channels, namely, North Branch, North Channel, North Passage and South Passage. There are many tidal flats inside the braided channel so it has the innate advantage to carry out the reclamation projects (Fig. 1). Shanghai has implemented some large-scale reclamation projects recently, such as Qingcaosha Reclamation Project, Hengsha Reclamation Project and Nanhui Reclamation Project [1-3]. It provides a certain spatial basis for urban construction and economic development. However, the contradiction of demand for land still exists in the future economic development of Shanghai according to the research and analysis of relevant scholars [4]. Therefore, the Xinhengsha Reclamation Project has been planned to be implemented at this stage [5], which will lay a solid foundation for the next step of urban development and the improvement of land use efficiency in Shanghai (Figure 1).

Saltwater intrusion occurs frequently in the Yangtze Estuary, with frontal saltwater intrusion along the North Channel, North Passage and South Passage and lateral saltwater intrusion along the North Branch into the South Branch [6]. These two forms of saltwater intrusion seriously affect the water supply safety of the Yangtze Estuary reservoirs in the dry season. Dongfengxisha Reservoir, Chenhang Reservoir and Qingcaosha Reservoir are often threatened by saltwater intrusion in the dry season, resulting in the salinity at the water intakes of the reservoirs exceeds the 0.45 (the maximum salinity threshold of drinking water), and no fresh water can be obtained. From 1994 to 2015, due to natural disasters such as drought or saltwater intrusion, the maximum number of consecutive days without access to water was 38 days a year. Previous studies have shown that the reclamation projects can change the local river regime and affect the saltwater intrusion. Therefore, the implementation of the Xinhengsha Reclamation Project is bound to have a certain impact on the pattern of saltwater intrusion in the Yangtze Estuary. Therefore, this study uses the three-dimensional hydrodynamic model to study the impact of the Xinhengsha Reclamation Project on the saltwater intrusion and freshwater resources in the Yangtze Estuary.
2. Materials and methods

This paper adopts the ECOM-si model, which has been fully validated in studies of hydrodynamic processes and saltwater intrusion in the Yangtze Estuary. Due to the limited space, the initial boundary of the model, the model verification, the calculation method of the net transection water flux and split ration can be referred to the published literature [6,7].

The model is calculated from December 1, 2013 to February 28, 2014. Then the results for February were analyzed. The given time unit of runoff in the model is hour. Here, the runoff in December, January and February adopts the monthly average from 1950-2020, so the runoff in 0-744 h is set as 13600 m$^3$/s, and the runoff in 745-1488 h is set as 11300 m$^3$/s. The runoff of 1489-2160 h was 12000 m$^3$/s. The atmosphere boundary uses the semilunar average wind of the European Centre for Medium-Range Weather Forecasts for the past ten years.

Two control experiments were set up. Experiment 0 (Exp0): The Xinhengsha Reclamation Project is not considered (Figure 2b). Experiment 1 (Exp1): The Xinhengsha Reclamation Project is added on the basis of experiment 0 (Figure 2c).

Figure 1. Yangtze Estuary situation map before (a) and after (b) the Xinhengsha Reclamation Project, sec1 and sec2 are the sections of the North Branch and South Branch, and sec3 and sec4 are the sections of the North Channel and South Channel, respectively.

Figure 2. The computational grid (a) used in the numerical model and the local enlargement of the grid. (b): The Xinhengsha Reclamation Project is not considered, (c): The Xinhengsha Reclamation Project is considered.
3. Results and discussion

3.1 Before the Xinhengsha Reclamation Project

Before the implementation of the Xinhengsha Reclamation Project, during the spring tide, the high salt water mainly occupies the North Branch, while the salinity of the South Branch is low (Figure 3). As can be seen from Table 1, the net transection water flux in the South Branch and North Branch are 11775.86 m³/s and -306.48 m³/s, and the corresponding split ratio are 102.67% and -2.67%, respectively. Therefore, the upstream fresh water mainly enters the sea through the South Branch, while the saltwater in the North Branch spillover into the South Branch. It can be seen from the distribution of surface and bottom salinity that there is an obvious saltwater wedge in the bifurcation of the North Branch and South Branch, which is lower in the surface layer and higher in the bottom layer, and the isohaline 0.45 can affect the vicinity of Dongfengxisha Reservoir in the bottom. Due to the strong tidal force during spring tide, the mixing in the mouth area of the Yangtze Estuary is more uniform. As can be seen from Table 1, the net transection water flux in the South Channel and North Channel are 3655.42 m³/s and 7976.05 m³/s, and the corresponding split ratio are 31.43% and 68.57%, respectively.

As a result, fresh water enters the sea mainly through the North Channel, so the salinity of the North Channel is lower than that of the South Channel. In the North Channel, the isohaline 0.45 in the surface bottom layer invading along the open sea can reach the lower waters of Qing cao sha Reservoir, while the isohaline 0.45 of the surface bottom layer invading along the open sea can reach the front waters of Qing cao sha Reservoir in the South Channel.

Table 1. Changes of the net transection water flux and split ratio in the South Branch and North Branch, South Channel and North Channel before and after the project the Xinhengsha Reclamation Project

<table>
<thead>
<tr>
<th></th>
<th>Net transection water flux (m³/s)</th>
<th>Split ration (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spring N-B</td>
<td>Neap N-B</td>
</tr>
<tr>
<td>Exp0</td>
<td>-306.48</td>
<td>11775.86</td>
</tr>
<tr>
<td>Exp1</td>
<td>-289.59</td>
<td>11735.02</td>
</tr>
<tr>
<td>△Exp1-Exp0</td>
<td>16.88</td>
<td>-40.84</td>
</tr>
</tbody>
</table>

During neap tide, the salinity in the upper North Branch is significantly lower than that during spring tide (Figure 4). As can be seen from Table 1, the net transection water flux of the South Branch and North Branch are 12668.28 m³/s and 210.59 m³/s, respectively, and the corresponding diversion ratios are 97.59% and 2.41%, respectively. The phenomenon of saltwater spillover from the North Branch into South Branch disappears. In the middle region of the South Branch, the saltwater mass spillover from the North Branch during spring tide is oscillating and discharging under the action of runoff and tide. The split ratio between North Branch and South Branch during neap tide is similar to that of spring tide (Table 1), the North Channel is still the main freshwater outlet to the sea. Therefore, the salinity in the North Channel is low, while the salinity in the South Channel is high. Due to the weakening of tidal dynamics during neap tide, the baroclinic effect at the mouth of the Yangtze Estuary is significantly enhanced, and the salinity of the bottom layer is much higher than that in the surface layer. In the North Channel, the isohaline 0.45 in the surface layer invading along the open sea can reach the front waters of the Hengsha Island, and the isohaline 0.45 in the bottom layer can reach the lower waters of Qing cao sha Reservoir. In the South Channel, the southern end of isohaline 0.45 in the surface layer invading along the open sea can reach the central waters near the Chang xing Island, and the bottom layer can reach the front waters of the Qing cao sha Reservoir.
3.2 After the Xinhengsha Reclamation Project

The variation of the tidal mean salinity during spring tide after the implementation of the Xinhengsha Reclamation Project (Exp1-Exp0) as shown in Figure 5. During spring tide, the salinity in the North Branch decreases slightly and the salinity in the North Channel decreases significantly while the salinity in the South Channel increases obviously. As can be seen from Table 1, the net transection water flux from the North Branch spillover into the South Branch decrease by 16.88 m$^3$/s after the implementation of the Xinhengsha Reclamation Project. For the split ratio, the water flowing into the North Branch increases by 0.14%, the salinity in the North Branch decreases while the decrease was not obvious. In the North Channel, the surface salinity near the project area decreases by about 2 and the bottom salinity decreases by about 4 which showing a significant reduction. The main cause is that the implementation of the Xinhengsha Reclamation Project can lead to an increase of 591.53 m$^3$/s in the net transection water flux and an increase of 5.45% in the split ratio in the North Channel. As a consequence, the salinity in the northern of the Xinhengsha Reclamation Project decreases significantly. Accordingly, in the South Channel, the surface salinity increases by about 1.5 and the bottom salinity increases by about 2.

Note: N-B is the North Branch, S-B is the South Branch, N-C is the North Channel, S-C is the South Channel; minus denote the direction of flood tide, plus is the opposite.
During neap tide, the overall variation trend of salinity field is similar to spring tide (Figure 6). The main difference is that the salinity decreasing area is mainly located in the northeast waters of the Xinhengsha Reclamation Project and the salinity decreases by more than 4. This is mainly due to the tidal force is weakened during neap tide so that the runoff is dominant and the runoff deviates to the right under the influence of Coriolis force after discharge along the North Channel. The degree of deviated to the right is significantly stronger than spring tide near the Xinhengsha Reclamation Project area. Therefore, during neap tide, the salinity of the northeastern waters of the Xinhengsha Reclamation Project decreases obviously. Correspondingly, the salinity in the North Passage and South Passage increases by about 2.

3.3 Impacts on reservoirs

The freshwater resources in the Yangtze Estuary are often affected by saltwater intrusion. Previous studies have shown that Dongfengxisha Reservoir and Chenhang Reservoir are mainly affected by the saltwater intrusion from the North Branch spillover from the South Branch, while Qingcaosha Reservoir is not only affected by the saltwater intrusion from the North Branch spillover from the South Branch, but also affected by the saltwater intrusion along the North Channel [6,7].

The comparison model results of salinity changes with time at the water inlets of Dongfengxisha Reservoir, Chenhang Reservoir and Qingcaosha Reservoir before and after the Xinhengsha Reclamation Project are shown as Figure 7. The change of salinity in water inlets with time show obvious characteristics of semi-diurnal and semi-lunar tide. Previous studies have shown that it is considered that water cannot be taken during this period when the salinity of the reservoir water intake
is higher than 0.45 more than 4 hours [7]. As mentioned above, the implementation of the Xinhengsha Reclamation Project can lead to the weakening of saltwater intrusion in the North Branch, so the salinity of the inlets of three reservoirs will all decrease after the implementation of the Xinhengsha Reclamation Project. For the continuous time that cannot take water for reservoirs, Dongfengxisha Reservoir is shortened from 7 days without access to take water in a spring-neap cycle before the project to 6 days, Chenhang Reservoir is shortened from 3.5 days to 0 days, and Qingcaosha Reservoir is shortened from 4 days to 2.5 days.

Figure 7. The time series plot of the water level in the water inlet of the Qingcaosha Reservoir (a), salinity in the water inlet of the Dongfengxisha Reservoir (b), salinity in the water inlet of the Chenhang Reservoir (c) and salinity in the water inlet of the Qingcaosha Reservoir (d) from January 29 to February 19. Full line: before the Xinhengsha Reclamation Project, Dashed line: after the Xinhengsha Reclamation Project

4. Conclusion

Using ECOM-si model, this paper simulates and studies the impacts of the Xinhengsha Reclamation Project on saltwater intrusion pattern and water resources acquisition in the Yangtze Estuary. It is found that the tidal mean salinity changes significantly after the implementation of the Xinhengsha Reclamation Project. During spring tide, the implementation of the Xinhengsha Reclamation Project will lead to the increase of split ratio in the North Branch and North Channel by 0.14% and 5.15%, respectively. Therefore, the implementation of the Xinhengsha Reclamation Project will reduce the salinity in the North Branch slightly, and the salinity in the North Channel is significantly reduced. The surface salinity near the project area in the North Channel can be reduced by about 2, and the bottom salinity is reduced by about 4. Accordingly, in the South Channel, the surface salinity increases by about 1.5 and the bottom salinity increases by about 2. During neap tide, the overall variation trend of salinity field is similar to that during spring tide. The main difference is that the decreasing area of salinity is mainly located in the northeast waters of Xinhengsha Reclamation Project and the salinity decreases by more than 4. Correspondingly, the salinity in the North Passage and South Passage increases by about 2.

The implementation of the Xinhengsha Reclamation Project can lead to the weakening of the saltwater intrusion from the North Branch spillover from the South Branch, so the salinity of the inlet of the Qingcaosha Reservoir, Chenhang Reservoir and Dongfengxisha Reservoir all decreases after the implementation of the Xinhengsha Reclamation Project. Dongfengxisha Reservoir is shortened from the continuous 7 days without access to water in spring-neap cycle before the project to 6 days, Chenhang Reservoir is shortened from 3.5 days to 0 days, and Qingcaosha Reservoir is shortened from 4 days to 2.5 days. The above research results show that the implementation of the Xinhengsha Reclamation Project is beneficial to the usage of freshwater in the Yangtze Estuary.
Reference

Design and Implementation of Log Audit System Based on Spark Big Data Platform

Jian Yang*
Chongqing Three Gorges Vocational College, Chongqing Wanzhou 404155, China.
39577932@qq.com

Abstract

This paper designs a log analysis platform based on big data technology, which uses Flume log collection in Hadoop, and combines with HDFS for distributed storage, uses Spark Streaming to process real-time data, and finally stores the results into MySQL database, and Java Web technology to provide query and visualization functions. In view of the huge amount of data, the diversity of data storage formats and collection methods of network security logs in the current era, the log audit system based on the background of big data provides an integrated solution from collection, analysis, storage to presentation.

Keywords: Spark; Big data; Audit; system design

1. Introduction

Since entering the 21st century, China's information technology industry has developed rapidly, followed by more and more network applications, and the network environment is more complicated. More and more people and major enterprises are beginning to use advanced Internet technology, but the awareness of network security has not been popularized, which leads to endless network security problems. Therefore, all walks of life gradually begin to pay attention to the construction of network security. Acts endangering network security mainly include forgery, modification, eavesdropping and interruption, etc. To eliminate the hidden security threats of the system, security measures such as encryption, identity authentication, authorization and auditing are usually adopted to improve the security of the network. Audit is to track the contents and ways of customers' access to find out the potential security risks in the system, so as to formulate corresponding security measures to make up for the security loopholes in the system [1].

One of the important ways to maintain network security is log audit. Network operators obtain the records of network operation status through monitoring, manage these records in a unified way, and keep complete network logs for more than six months. Through these logs, the network security operation and maintenance personnel can know the user's access situation and find out the network security problems, thus improving the quality and efficiency of network security operation and maintenance. The audit system is an audit tool widely used in computers and networks. Therefore, in recent years, all industries have started to focus on building network security management projects for log audit analysis [2].

However, at present, most of the existing book log audit systems can only analyze the log data of a single machine, but there is no suitable access method and management platform for millions of logs generated by other network hardware devices such as servers, storage devices, switches, firewalls and security devices. Therefore, the network security operation and maintenance administrators urgently need new technical systems to assist their operation and maintenance work. [3] According to the above analysis, the author believes that a log audit system should be developed, which is based on Spark big data ecosystem and combines HDFS distributed storage with Hadoop technology and Flume log collection component. This paper describes the design and implementation of a series of functions such as the collection, storage, query and visualization of the underlying log data of the system. The system can not only analyze the log information accurately and in real time, analyze the data and give an alarm to the crisis situation, but also store the historical log data in real time and calculate and process it offline. The system has the advantages of high speed, good scalability, large amount of data processing, etc. It has achieved good results in actual operation, and made a useful attempt for the analysis and research of log audit system based on big data technology.
2. Key Technology

2.1 Spark

Spark is a big data technology ecosystem developed by AMP Lab of University of California. Spark big data ecosystem has five main functional components, which are SparkSQL, GraphX, Sparkstreaming, MLib and BLbase. SparkSQL refers to the processing engine for adding, deleting, modifying and searching SQL structured data. MLib is a distributed machine algorithm library used to handle the underlying logic. MLBase is responsible for the main calculation and processing of data. GraphX is a new computing framework for dealing with parallel graphs. The key component used in the development of this system is SparkStreaming, which is a software specially responsible for processing flow calculation. It can effectively process the accumulated stream data, and even convert these data into RDD for further micro-batch processing by specifying time segments [4].

2.2 Hadoop

Hadoop is very similar to Spark, and it is also an ecosystem based on big data and containing many components. The difference is that the development company of Hadoop is the Apache Software Foundation, which is more powerful. It is also because of relying on Apache platform that Hadoop is developed in a wider range of Java languages, and it is more widely used by developers in various fields. With the passage of time, Hadoop's project structure has been continuously enriched, gradually forming a big data ecosystem with the most complete functions at present.

2.3 Kafka

Kafka is one of the important components of Hadoop ecosystem, and also belongs to the open source project of Apache Software Foundation. In the development process of this system, kafka needs to complete a lot of data transmission work, so this paper will focus on kafka component technology. Kafka architecture is mainly composed of topic topic, byte payload, producer producer, proxy Broker, and client Consumer that reads messages. The producer data source objects usually include the page browsing data generated by the system web client, the log records of the server and the published message broker in a server. However, the throughput of the established cluster often becomes more and more because of the continuous growth of broker. The whole cluster is divided into three parts, the producer is responsible for processing the front-end messages, the broker is responsible for the information push of kafka service nodes, and the consumer is responsible for the back-end consumer node processing, in which another zookeeper component is needed to coordinate the whole process [5].

2.4 Exploitation environment

This paper briefly introduces the related technologies used in platform development. The log audit system of big data platform uses the big data server cluster with Spark as the main part and Hadoop as the auxiliary part to process data and store the data in MySQL database, and develops the corresponding application platform by using JavaWeb technology.

According to the data volume and overall operation requirements of the system, this paper chooses to build a Hadoop cluster with five nodes, including two master nodes and three slave nodes. The primary node is named namenode, and the secondary nodes are named datanode and numbered respectively. After that, all functional components such as zookeeper-3.5.5, HDFS and flume1.9.0 are installed and deployed in these five nodes synchronously, thus completing the initial construction of hadoop cluster. After that, the nodes of spark cluster are established on the basis of hadoop, which are also deployed by two primary nodes and three secondary nodes, with different names. The primary node is master and the secondary node is slaves. Spark cluster chooses version 2.3.4. The cluster will be developed under the Linux system. This paper selects Centos7.8 Server release version of Linux operating system.

The Java development tool used in the JavaWeb application of this system is IDEA 2021.1.3 (Ultimate Edition), the development environment is JDK 1.8, the development language is Java, and Apache Tomcat 9.0 is selected for server building. The system development is based on MVC pattern, using JSP+Servlet+Javabean framework, and MySQL 8.0.28 to help manage data. Through the introduction of the above key technical theories, the overall environment of system development, the configuration of related software and tools are determined, and the technical feasibility of the overall project is also defined.
3. Global design

The overall architecture of the log audit system based on Spark big data platform is shown in Figure 1. The data source of the log comes from servers, storage devices, switches, firewalls, security devices and other network access devices. These data will be sent to the target server through the designated port for unified configuration, so as to standardize the reserve. The preliminarily processed log data will be sent to Flume cluster for caching, and then Flume will distribute it to different platforms according to the different requirements of online and offline processing and analysis. Offline processing is to persist the data cached by FLume and save it in HDFS, then Spark will analyze and process the offline log, and finally output the information results. On-line processing is to transfer the data cached by Flume to Kafka message queue system, and then the Spark Streaming real-time computing program will process the data in real time, output it and save it in HBase. The final output data information will be exported by Sqoop to mysql, a relational database, for standardization and persistence, which is convenient for users to browse these data results through the web system. The web part of this system uses javaweb technology, adopts MVC pattern with front-end separation, and the selected framework is JSP+servlet+javabean [6].

![Figure 1 overall architecture of the system](image)

4. Function realization

The log audit system based on Spark big data platform has developed three functional modules: log acquisition, log acquisition module, log analysis module, storage module and visualization module to ensure the cyclic execution of the system in order to show the complete process of log audit collection, standardization, analysis and result display and meet the needs of network security operation and maintenance managers.

4.1 Log collection

After logging into this system for the first time, the network operation and maintenance administrator user needs to determine the range of the collected log data. The data comes from network access devices such as servers, storage devices, switches, firewalls, security devices, etc. The data types are divided into two main categories: traffic logs and threat logs. The core software tool used in the log collection function module is Flume distributed log collection system. Click to enter the log collection function module in this system, and users can click to import configuration files to start editing. In the configuration file, the user can specify whether the source of the collected logs is a server, a storage device, a switch or a firewall, and can also choose which node of hadoop cluster is the main data transmission channel. During the transmission process, the user can see the list of files that have been successfully transmitted. The user can interrupt the transmission manually, delete the files that have been transmitted, or choose to re-select the files for reading and collecting. After that, Flume carries out distribution processing on different platforms according to different requirements of online and offline processing analysis. Flume is using Syslog to collect the log code as shown in Figure 2.
4.2 Log processing analysis

After the user clicks into the log data processing and analysis module, the system will ask the user to select the processing method. Users can choose to calculate and analyze the real-time data, or make statistical analysis of all the log data stored in the offline system history, which can improve the efficiency of data analysis and make the data display more targeted. Of course, users can also choose two data analysis modes to perform simultaneously. The real-time computing function mainly combines Kafka, Spark Streaming and other technical components to realize multi-dimensional dynamic analysis of real-time data. The key tool used in real-time computing is Spark Streaming framework. Spark Streaming is not only an extension of Spark API, but also supports secondary development and expansion, which is convenient for reference. More importantly, developers can use it to perform advanced operator processing on complex stream data. After creating the constructor of StreamingContext, you can set up several micro-batches of new data to receive data at a rate of per second. The received text data is called DStream. Essentially, the operation of DStream will eventually be converted into the operation of the underlying RDD, and the Spark Streaming flow calculation program can realize real-time calculation by operating the high-level API interface provided by DStream. In the process of log analysis, the alarm function is set in the system, and the unit time frequency analysis method is used. For example, in the process of real-time log monitoring, the system finds that a user has entered the wrong password more than 5 times in the process of logging in within 10 seconds, and the system gives an alarm of violent cracking to remind the user. The time-frequency analysis code is shown in Figure 3.

```java
#Configure sources, and configure different types for dirSources and SyslogSources, the former is file directory and the latter is syslog.
agent.sources.dirSources.type = spooldir
agent.sources.dirSources.spoolDir = /home/spark/log
agent.sources.dirSources.fileHeader = true
agent.sources.syslogSources.type = syslogtcp
agent.sources.syslogSources.port = 5140
agent.sources.syslogSources.host = localhost
agent.sources.syslogSources.fileHeader = false
agent.sources.syslogSources.interceptors.ii.type = timestamp

#Configure the type of channels as memory.
agent.channels.memoryChannel.type = memory

#Configure sink, configure different types for loggerSink, hdfsSink, the former flows to SparkStreaming and the latter flows to HDFS.
agent.sink.loggerSink.type = avro
agent.sink.loggerSink.hostname = worker1
agent.sink.loggerSink.port = 1111
agent.sink.hdfsSink.type = hdfs
agent.sink.hdfsSink.path = hdfs://Master:9000/hadoop/orgLog
agent.sink.hdfsSink.fileType = datastream
agent.sink.hdfsSink.filePrefix = data
agent.sink.hdfsSink.fileRollInterval = 60
agent.sink.hdfsSink.writeFormat = TEXT
agent.sink.hdfsSink.rollSize = 1073741824

Figure 2 Syslog get log code
```
### 4.3 Data storage and visualization

Click on the log storage function module, and users can see that the distributed file storage system HDFS and database MySQL are used in the data storage part. HDFS, as the underlying storage device, is used to store the original log data. In this system, all the original data are imported into HDFS and saved. While browsing these two categories of data, users can query the data, and the query function is realized by Hive.

### 5. Conclusion

The function of the log audit system based on Spark big data platform is still slightly single compared with the requirements of the current severe network security era. It can only provide query and visualization of log data of the network access system, and lacks the application functions of various algorithms of association analysis. In the future, we should improve this deficiency and add more big data technologies and methods of feature analysis and machine learning. In the future improvement work, we can obtain more threat information, further improve the sensitivity of detecting hidden dangers and threats, achieve better monitoring and early warning effects, and make contributions to improving the comprehensive management level of network security in China.

### Fund project

Chongqing Three Gorges Vocational College's 2021 General Project of Natural Science Research "Design of Peach Bad Fruit Identification and Classification System Based on Computer Vision" (No.cqxs2021015), Moderator: Yang Jian.

### Reference

[1] Li Guangming, Li Zaizhou, Li Qi. Design of Log Cleaning System Based on Spark Platform. Computer Engineering and Design. 2020, 41(12): 3580-3586


Long-term stock price forecast based on PSO-Informer model

Hailun Liu1a, Deng Chen1b*, Wei Wei1c, Ziqiang Wei1d

1Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China,
a1015567806@qq.com, b2dchen@wit.edu.cn, c18933757@qq.com, d741705682@qq.com

* Corresponding author: dchen@wit.edu.cn

ABSTRACT

The long-term prediction of stock prices provides an important reference for quantitative investment decisions. Aiming at the problem of insufficient accuracy of long-term series prediction in existing stock forecasting models, this paper proposes a long-term stock price series forecasting method based on PSO-Informer. Methods First, 43 kinds of technical indicator factors and K-line data were selected to construct the input data, and then the PSO-Informer model was used to predict the future 60 time points of the stock closing price. In the model training process, the particle swarm algorithm is used to optimize the parameters of the Informer network. Based on the 5-minute K-line data of the SSE 50 stock index and the CSI 300 stock index, experimental research was conducted respectively. Taking the accuracy of the long-term stock price prediction overall trend as the evaluation index, and the prediction accuracy reaches 68.2% and 67.5% respectively. The comparison experiments with ARIMA, Prophet, PSO-LSTM and Informer prediction models show that the model has the best performance and is practical.

Keywords: Informer, particle swarm algorithm (PSO), technical indicator factors, stock price forecast

1. INTRODUCTION

The stock market has become an indispensable part of China’s economic development due to its high rate of return, and has long been one of the most popular investments [1]. The stock price forecast result is also an important reference for profit and risk avoidance [2]. Therefore, more and more scholars conduct research on stock price prediction and have achieved a series of results.

Traditional stock forecasting methods include auto-regressive integrated moving average (ARIMA) mode and special forms of ARIMA, time series models such as Prophet algorithm. The ARIMA [3] model and the Prophet [4] algorithm are suitable for single variable time series forecasting, and the forecast data needs to have stability and obvious periodicity. However, stock data has the characteristics of high noise, non-stationary, non-linear, etc, and there are many factors affecting stock price volatility [5]. Traditional time series models cannot achieve ideal results in stock data forecasting. In recent years, with the development of artificial intelligence technology, machine learning algorithms have gradually shown their advantages in stock price prediction and have been widely used. Kumar used the LSTM network to predict the stock price of Nasdaq, applied the Adam optimizer for training, and tested the performance of three commonly used output activation layers, verifying the applicability of the algorithm in short-term stock price prediction [6]. Kim combined the LSTM and CNN models, the CNN-LSTM neural network extracted the spatial structure features of the data through CNN, and then extracted the input time series features by using LSTM, and finally effectively predicted the highest stock price of the day in a short period of time[7], the experimental results show that the LSTM-CNN joint model can obtain higher prediction accuracy than a single model. Compared with methods based on traditional time series models, methods based on deep learning models can achieve higher robustness and prediction accuracy. However, most studies are limited to short-term forecasts of stock prices. Short-term forecasts of stock prices have strong stochastic volatility, and it is difficult to reasonably reflect the trend of stock prices, which makes such forecast results less practical in quantitative investment.

In view of the above problems, this paper proposes a forecasting model PSO-Informer for long-term stock price forecasting. Method First, 43 index factors [8][9] are combined with K-line data to construct input data. The input data is converted into vector form and sent to the PSO-Informer model for long-term series prediction of stock closing price 60 time points in the future. The PSO-Informer model uses particle swarm algorithm (PSO) to optimize the encoder input window size, learning rate, and batch size , which are highly subjective in the Informer model, to improve the accuracy.
of stock price prediction. At the same time, the Informer model [10] is based on the Transformer model, uses the Prob Sparse Self-attention mechanism in the process of encoding and decoding, only considers the part that contributes the most to the attention mechanism, and solves the problem of high computational complexity and large memory usage in the long sequence prediction process.

This paper selects the five-minute K-line data of the SSE 50 stock index and the CSI 300 stock index to make a long-term series prediction of the closing price at 60 time points. The comparison experiments with ARIMA, Prophet, PSO-LSTM, and Informer prediction models verify the strong prediction ability of the PSO-Informer model.

2. RESEARCH METHOD

2.1 Overall framework

The overall method structure of the long-term stock closing price forecasting based on PSO informer model proposed in this paper is shown in Figure 1. First, obtain the historical K-line data of stock prices, obtain 43 technical indicator factors at each time point based on the tablib library, and add the technical indicator factors to the K-line data to form training data. Then, the PSO-Informer model is used to forecast the stock closing price in a long-term series at 60 time points. During model training, the PSO algorithm is used to optimize the encoder input window size, learning rate, and batch_size of the Informer model. Finally, the 60 time points of the closing price predicted by the PSO-Informer model are linearly fitted, and the fitted trend is compared with the trend of the original data, and the prediction effect of the model is judged according to the prediction accuracy of the overall trend of the 60 time points of the closing price.

2.2 Informer model for long-term stock price prediction

The Informer algorithm uses the ProbSparse Self-attention mechanism in the process of encoding and decoding, and only considers the part that contributes the most to the attention mechanism. Compared with LSTM [11] and Transformer [12], the calculation amount is smaller and the memory usage is low.

First, the input of the model is 48 standardized technical index factors and K-line data at the first t time, $X = (X_{1}, X_{2}, \cdots, X_{t})^{T} \in R^{M \times t}$ where $M$ is the input vector dimension, and the input vector is input into the model after the dimension expansion transformation from $M$ to $d_{model}$ is performed. Since the model cannot grasp the contextual relationship of data input, the Informer encodes the position information of each input data as shown in the following formulas (1) and (2):

$$PE_{(pos, 2j)} = \sin(pos/(2L_{x})^{2j/d_{model}})$$

(1)

$$PE_{(pos, 2j+1)} = \cos(pos/(2L_{x})^{2j/d_{model}})$$

(2)
Among them, \( pos \) is the absolute coordinate of the data in the entire time series, \( L_x \) is the length of the input, and \( j = 1, 2, \cdots d_{\text{model}} / 2 \) is the dimension. The location information is added to the 48 technical index factors and K-line data after dimension expansion and then entered into the model together.

In order to improve the feature learning efficiency of the input data, the model adopts the Multi-head ProbSparse self-attention to calculate the feature correlation. The formula of ProbSparse self-attention is shown in formula (3) and formula (4):

\[
A(Q, K, V) = \text{Soft max} \left( \frac{QK^T}{\sqrt{d}} \right)V
\]

\[
Q = R^{L_q \times d}, \quad K = R^{L_K \times d}, \quad V = R^{L_v \times d}
\]

Among them, \( d \) is the input dimension, \( Q \) is a sparse matrix with the same shape as \( Q \), it only contains \( q \) calculated by sparse evaluation of the first \( U \) tours, and the size of \( U \) is determined by a sampling parameter. The multi-head probabilistic self-attention mechanism can reduce the computational complexity of each key-value pair and reduce the time cost.

The information input by the encoder predicts the part covered with 0 in the decoder input through the multi-head self-attention mechanism, and finally adjusts the dimension of the output through a fully connected layer, so as to output all the predicted closing prices.

### 2.3 PSO-Informer model establishment

The fitting effect during Informer model training is highly correlated with the input window size, learning rate, and batch size. In this paper, the PSO algorithm with fast convergence speed is used to optimize the three parameters of the Informer model, and the optimal stock price prediction model PSO-Informer is obtained.

First initialize the input window size, learning rate, and batch_size. The parameter ranges of the three parameters are [90, 180], [0.00005, 0.0001], [30, 50] respectively. The mean square error during model training is used as a fitness function[13], and its expression is shown in formula (5):

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i^\wedge - y_i)^2
\]

where \( y_i \) is the actual closing price, \( y_i^\wedge \) is the predicted closing price, and \( n \) is the number of samples.

The particle velocity and position are updated at each iteration as shown in the following formula (6):

\[
v_i = v_i + c_1 \times r \times (pbest_i - x_i) + c_2 \times r \times (gbest_i - x_i),
\]

\[
x_i = x_i + v_i.
\]

Among them, \( i = 1, 2, \cdots n \) is the number of particles, \( v_i \) represents the particle velocity, \( x_i \) represents the particle position at the moment, \( r \) is a random number between 0-1, \( c_1, c_2 \) represents the learning factor, and the value in this paper is 2.
2.4 Ups and downs decision module

Linearly fit the predicted closing price of a predicted time window size (60 time points) to the actual value. Given a set of data set \((x_i, y_i)\), \(x_i = 1, 2 \cdots 60\) represents 60 time points, and fit curve \(f(x) = a + bx\). The objective function is the mean square error between \(f(x)\) and the value \(y\). According to the minimum value of the objective function, the linear fitting values \(b : b_1, b_2\), of the predicted data and the actual data are obtained respectively. According to the linear fitting results, the ups and downs decisions are as follows:

\[
T = \begin{cases} 
1; b_1 > 0, b_2 > 0 \\
-1; b_1 < 0, b_2 < 0 
\end{cases} \tag{7}
\]

When \(T = 1, -1\) indicates that the trend forecast in this time window is accurate and the trend is up or down, in other cases, it is recorded as 0, indicating that the forecast is wrong.

3. EXPERIMENTAL EVALUATION

3.1 data set

A total of 23,327 sets of five-minute K-line data from September 23, 2019 to September 22, 2021 were obtained through python's open source financial data interface package TuShare. The 23327 sets of sample data are divided into three cycles to train the model respectively, each cycle is divided into 5 months training set to train the model, and 2 months test set to verify the model, a total of about 7775 sets of five-minute K-line data.

3.2 Analysis of results

3.2.1 Evaluation indicators

According to the ups and downs decision introduced in Section 2.4, four evaluation indicators \([15]\), namely Accuracy, Precision, Recall and F-measure, are used to evaluate the ups and downs. Its calculation formula is as follows.

\[
A = \frac{TP + TN}{TP + FP + TN + FN}, P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F1 = 2 \frac{P \times R}{P + R} \tag{8}
\]

Where \(TP\) represents true positive, \(TN\) represents true negative, \(FP\) represents false positive, and \(FN\) represents false negative.

3.2.2 Analysis of the results of PSO-Informer model on SSE 50 and CSI 300

The PSO-Informer model is used to predict the closing price of the SSE 50 stock Index and the CSI 300 stock Index at 60 time points according to the above data sets. In the third cycle of training, the input window sizes optimized by the PSO algorithm are 92 and 116, respectively. The test results are taken as a prediction window every 60 time points. The predicted and actual values are shown in Figures 2 and 3:
From Figure 2 and Figure 3, we can see that the curves of the predicted value and the actual value are generally consistent, and we can preliminarily judge that the use of this model will play a certain role in predicting the overall trend of the stock price in the next 60 time points. We further evaluate the prediction results according to the evaluation indicators, and the evaluation results are shown in Table 1 below:

<table>
<thead>
<tr>
<th>Index</th>
<th>Category</th>
<th>SSE 50 stock Index</th>
<th>CSI 300 stock index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td>68.2</td>
<td>69.5</td>
</tr>
<tr>
<td>Precision</td>
<td>Up</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Down</td>
<td>66.7</td>
<td>63.6</td>
</tr>
<tr>
<td>Recall</td>
<td>Up</td>
<td>69.2</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td>Down</td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td>F-measure</td>
<td>Up</td>
<td>71.9</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Down</td>
<td>63.2</td>
<td>66.6</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, the accuracy of the model’s predictions on the SSE 50 Index and the CSI 300 Index is above 65%. In the stock price trend prediction, the researchers believe that the accuracy rate of the upward trend has higher use value. It can be seen from the table that the accuracy rate of the upward trend can reach 75%. In summary, the model has a high reference value in the long-term series prediction of stock prices.

3.2.3 Comparative experiment

In order to further evaluate the performance of the PSO-Informer model in the long-term trend prediction of stock prices, this paper selects the time series models ARIMA, Prophet, PSO-LSTM, Informer, and PSO-Informer to conduct time
series prediction experiments on the SSE 50 Index. The evaluation results of each model according to the ups and downs decisions and evaluation indicators are shown in Table 2 below:

<table>
<thead>
<tr>
<th>Category</th>
<th>PSO-Informer</th>
<th>Informer</th>
<th>PSO-LSTM</th>
<th>ARIMA</th>
<th>Prophet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>68.2</td>
<td>57.6</td>
<td>49</td>
<td>42.8</td>
<td>45.4</td>
</tr>
<tr>
<td>Precision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>75</td>
<td>58.3</td>
<td>48.9</td>
<td>44.1</td>
<td>41.6</td>
</tr>
<tr>
<td>Down</td>
<td>66.7</td>
<td>57.1</td>
<td>49</td>
<td>36.3</td>
<td>50</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>69.2</td>
<td>53.8</td>
<td>48</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>Down</td>
<td>60</td>
<td>61.5</td>
<td>50</td>
<td>40</td>
<td>41.6</td>
</tr>
<tr>
<td>F-measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>71.9</td>
<td>55.9</td>
<td>48.4</td>
<td>46.2</td>
<td>45.4</td>
</tr>
<tr>
<td>Down</td>
<td>63.2</td>
<td>59.2</td>
<td>49.5</td>
<td>38</td>
<td>45.4</td>
</tr>
</tbody>
</table>

As can be seen from Table 2, the traditional time series models ARIMA and Prophet have poor forecast performance in different forecast window sizes, and are not suitable for stock price forecasting. For Informer and LSTM models, the accuracy can reach about 50%, which verifies that deep learning has reference value in long-term stock price series prediction. Meantime, Informer has improved the accuracy rate and the accuracy rate of upward trend compared with PSO-LSTM, indicating that Informer has certain advantages in long-term trend prediction of stock prices. After using the PSO algorithm to optimize the parameters, the accuracy of the PSO-Informer is further improved by 10.3% on the basis of the Informer, and the accuracy of the upward trend is further improved by 16.7%.

### 4. CONCLUSION

This paper proposes a long-term stock price forecasting model based on PSO-Informer. Compared with the previous short-term forecasting of stock prices, long-term forecasting has more practical value. Use PSO-Informer to forecast the closing prices of the Shanghai 50 stock index and the CSI 300 stock index in a long-term series. The overall ups and downs of the predicted value can be seen to have a high accuracy rate. The model is suitable for long time series forecasting of stock prices. Further, different prediction models are used to conduct long-term series prediction experiments on the Shanghai 50 stock index. The PSO-Informer model exhibits obvious superiority by the experimental results. Compared with traditional time series forecasting models and other machine learning methods, this model has higher forecasting accuracy.

### REFERENCES


Learning Anisotropy and Asymmetry Geometric Features for Medical Image Segmentation

Ankun Li*, Li Liu

Qilu University of Technology (Shandong Academy of Sciences)/Shandong Artificial Intelligence Institute, 19 Academy of Sciences Road, Jinan, Shandong Province, China 250014;

*Corresponding author: ansirlee@163.com

ABSTRACT

Finding contours of interest from medical images is an important task in the field of medical image analysis. The current deep learning-based image segmentation approaches have obtained promising results. However, most of these models do not take into account the anisotropy and asymmetric features which play an important role in describing the target contours. In order to address this issue, we propose a new loss-function applied to the deep learning model with dense distance regression, which can benefit the edge-based features, thus able to improve the stability of the segmentation procedure and to reduce the probability of outliers in the segmentation results. The introduced loss function is embedded into the deep learning model, which can perform an end-to-end image segmentation procedure for medical images. Ablation experiments were done with other loss functions and three datasets were used to verify whether this loss function is effective. SOTA results were obtained for the proposed loss function in this paper compared to the recently designed method for reducing the boundary error.

Keywords: V-Net, Deep, Geometric Enhancement, Segmentation, Medical Image, Anisotropy and Asymmetry, Active contour

1. INTRODUCTION

In the past, medical image segmentation plays a fundamental role in medical imaging and computer vision. There are two main difficulties in segmentation. Specifically, the first difficulty lies at automatically finding the location of the anatomical architectures, since the organs in the abdomen usually have large differences in location, shape, and size between individuals. The second one is the negative effects to the target boundaries from adjacent organs [1]. To solve this issue, many efforts have been devoted to develop automatic segmentation procedure. For medical image segmentation, semantic information is important [2]. It is known that variational or graph-based segmentation approaches [3,4,5,6,7,8] usually lack the semantic information of objects, while deep learning-based methods overcome the limitations of traditional segmentation methods and show good performance in detecting these information. However, although the deep learning-based image segmentation methods have yields promising segmentation performance, mage details may be ignored during the pooling and downsampling operations, thus reducing the accuracy of image segmentation. Moreover, deep learning-based segmentation approaches often lose the geometry features of the target boundary. Although good performance has been achieved by virtue of the CNN-based approach, it is still difficult to meet the requirements of medical images for segmentation accuracy.

In deep learning-based segmentation models, the loss function is regarded as an important component, and the parameters of the neural network need to be measured by the loss function to measure the error of the segmented image and ground truth. These parameters are optimized by the back propagation algorithm. However, medical images usually contain a lot of contextual information rather than point samples [9]. The dice loss, which is commonly used loss function in image segmentation tasks, considers only the information from each individual pixel instead of the relationship between the them, thus leading to unexpected results in some complicated scenarios. In order to solve this problem, we take into account the anisotropy of asymmetry of the target boundary contours for the construction of the loss function. This is done by adding the image gradients, which carry both the directional features and the saliency features, to the loss function used in our model.

In this paper, we propose a new CNN-based image segmentation model. Its main contribution is the introduction of the loss function, which is able to integrate the anisotropic and asymmetric geometric features of the boundaries, as well as the features extracted from the CNN, which is, to the best of our knowledge, an original work.
The organization of this paper is listed as follows. In Section 2, we present the related work on traditional segmentation models and the deep learning-based models. Section 3 introduces the main idea of this manuscript. The experimental results and conclusion are presented in Sections 4 and 5.

2. RELATED WORK

2.1 Traditional Image Segmentation Methods

In the past decades, various methods have been proposed for detecting the image boundaries [10,11,12,13], such as edge detection operators, region growing, graph-based models and variational segmentation methods.

Among them, the variational segmentation methods mainly refer to the active contours [14] and the approaches developed on its basis in conjunction with the level set formula [15]. The basic idea of active contours is to evolve continuous curves to delineate the target boundaries. Active contour models have been applied in medical image segmentation tasks [16] for a long time. Depending on the definition of the energy functionals, the active contour models can be divided into two categories. The first type is the region-based approaches which are established over the statistical models of the image gray levels or colors. The second type is the geometric active contour models, whose energy functionals are constructed by measuring the weighted curve length [17,18]. In some complex segmentation scenarios, it is difficult to model the image intensity distributions with simple statistical models, thus sometimes leading to unexpected results.

The geometric contour models make use of the local image gradients as edge descriptor. These features are insensitive to image intensity variations, thus able to overcome the drawback of the region-based active contours. However, image gradients are local features which are known to be sensitive to image noise and spurious edges. This also results in the demanding requirement on the initialization. In other words, these geometric models may require the user to place the initial contours close to the target boundaries. Moreover, they also suffer from a high computation burden to find the minimization of the energy functionals [19], which requires a special design to remove the numerical instability in the evolution. Moreover, the features of these models are selected manually, depending on the skills and experience of the researchers, and medical images usually contain noise with blurred boundaries, so their applicability and quality of results are limited.

2.2 Deep Learning Segmentation Methods

Since the outstanding performance of convolutional neural networks (CNNs) on the ImageNet image segmentation challenge, CNN have been one of the major tools in semantic and instance image segmentation tasks [20,21,22,23,24,25]. The performance of the CNNs can be improved by designing more complex network architectures, such as designing the network to be larger, more widely connected, and more complex forms of convolution. Because of its demonstrated powerful performance [26], CNNs have also been applied in the field of medical image segmentation [27,28]. The popular deep learning models for medical image segmentation are adversarial neural networks [29], residual neural networks [30], U-Net [31] architecture, and transformer architecture [20], which are applied to different medical image segmentation tasks [32] and obtained better results. After the U-Net network architecture was proposed, a large number of U-Net variants of the network were proposed [33] because of the good performance of U-Net and its simplicity, making this encoder-decoder architecture network to be is a popular network model today, where the encoder is used to extract the image features and the decoder is used to recover the extracted features to the size of the original image and output the final segmentation result. The accuracy of segmentation is improved by combining high-resolution images and low-resolution feature maps through hopping connections. So the U-Net architecture becomes the benchmark solution for many image segmentation tasks. However, the model of this architecture lacks the consideration of the architecture beyond the segmentation target, and it is easy to make the segmentation result produce leakage, or have outliers and other problems. And the existing models of deep learning only use the local features of the image and the abstract convolutional layer features, without considering the intrinsic geometric constraints of the segmented object and the global information of the image [34].

2.3 Combination of Geometric Enhancement and Deep Learning Segmentation Mode

The geometric active contour models are sensitive to the setting of initializations, which may yield poor segmentation results especially when the initial contours are far from the target boundaries. Furthermore, the energy minimization of these models is often carried out by the level set formulation [15]. This provides the avenue to the minimization of the geometric models using a gradient descent scheme carried by a CNN [35]. In the other hand, several approaches have
shown that using different loss functions can improve the performance of U-Net network architectures for specific datasets [36]. In [2], the authors propose a level set-based loss function, where the main idea is to use the area term of the energy functional of the Chan-Vese Model [5] as the loss function. The main idea of [36] is to consider the area inside and outside the object as well as the size of the boundary when learning image features. Ma et al. suggested to combine the advantages of the classical geometric active contour models and the CNNs for image segmentation [34]. In that model [34], the gap between both types of approaches are linked using the level set-based contour representation way. The loss function used in [34] are constructed using level set function such that the image gradient magnitudes and the features extracted by CNNs. However, in this model, the anisotropy and asymmetry features are lost. In our model, we propose to revisit these important geometric features and exploit them for image segmentation, and the image features of anisotropy and asymmetry are added to the CNN training, so the image segmentation method in this paper is proposed.

2.4 Dice Loss and MSE Loss

In CNN model training, the loss function is an important part, which is used to calculate the function of the gap between prediction and true segmentation. The backpropagation algorithm is used to update the weights of the network, thus improving the prediction ability of the network for the region of interest of the image.

Dice Coefficient (DC) Loss: Dice loss from the article V-Net[11]. Dice coefficient is a set similarity measure function, is usually used to calculate the measure of similarity of pixel points between two samples, the larger the value means the set is more similar (the value range is [0, 1], where 1 represents identical). Dice coefficient is defined as follows:

\[
DC(X, Y) = \frac{\sum_{i=1}^{N} (X_i \cap Y_i)}{\sum_{i=1}^{N} (X_i + Y_i)}
\]  

(1)

where X represents the predicted segmentation, Y represents the ground truth, \((X_i \cap Y_i)\) denotes the dot product of the corresponding elements of the two sets. Dice loss can be written as:

\[
L_{Dice}(X, Y) = 1 - DC(X, Y)
\]

(2)

he larger the dice coefficient is, the more similar the set is, the smaller the loss is, and vice versa.

\(L_1\) Loss & \(L_2\) Loss: The \(L_1\) parametric loss function, also known as least absolute deviation (LAD), and least absolute error (LAE). In general, it minimizes the sum of the absolute differences between the target value \(Y_i\) and the estimated value \(X_i\). The \(L_2\) parametric loss function, also known as least squared error (LSE). In general, it minimizes the sum of squares S of the differences between the target value \(Y_i\) and the estimated value \(X_i\). Since the \(L_1\) loss function is considered more stable than the \(L_2\) loss function, the \(L_1\) loss function is chosen as a branch of the overall loss function in this paper. \(L_1\) and \(L_2\) are defined as follows:

\[
L_1(X, Y) = \frac{1}{n} \sum_{i=1}^{N} |X_i - Y_i|
\]

(3)

\[
L_2(X, Y) = \frac{1}{n} \sum_{i=1}^{N} (X_i - Y_i)^2
\]

(4)

where Xi represents the predicted value and Yi represents the ground truth.

3. METHOD

In this section, the architecture of the network and the proposed loss function embedding with boundary directional features are introduced.
3.1 CNN Architecture

In this section, the classical V-Net [11] architecture is used as our basic segmentation framework to evaluate our proposed loss function. V-Net is a network model for 3D image segmentation, which is based on the encoder-decoder architecture network architecture, where the left path is the encoding path and the right path is the decoding path, with jump connections connecting them in the middle, which is a way to combine This is a method of combining high resolution images and low level feature images, which is used to reduce the loss of image accuracy and improve the stability of training. And the network architecture used in this paper does not add architecture such as residual networks to increase the complexity of the network. The left coding path layer i (i=1,2,3) has i 3×3 convolutional layers of step 1, one rectified linear unit (ReLU), and i (i=4,5) has three 3×3 convolutional layers of step 1, one rectified linear unit(ReLU), and the downsampling of each layer One 3×3 convolutional layer with a step size of 2 and one rectified linear unit(ReLU) are used, for a total of four downsampling layers. The convolutional layers of the decoding path are the same as the coding path, and the upsampling uses one 3×3 deconvolutional layer with a step size of 2. There are four upsampling layers in total, and the last layer of the network with upsampling convolutional layers outputs the predicted probability map of the 3D image. The obtained probability map is processed in Equation 1 to obtain the segmentation contour of the network, the segmentation threshold is set to 0.5, greater than 0.5 is the region of interest, and less than 0.5 is the background region. Figure 1 shows the network architecture.

![Network Architecture Diagram](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

Figure 1. A review of the network architecture of the proposed loss function. The network architecture uses the classical V-Net architecture and the final output is a probability map of the image prediction and uses a sigmoid function to map the values in the probability map to 0-1, setting a threshold of 0.5, with regions of interest greater than 0.5 and background regions less than 0.5. The first branch is to calculate the relationship between individual pixels and supervise the difference between Ground Truth and probability map by $L_{\text{Dice}}$; the second branch is $L_{\text{AC}}$, which combines the edge information of the image and drives the predicted level set function towards the segmentation boundary when the segmentation is far from the boundary of interest, so as to obtain lower energy.

3.2 Loss Function

Minimize the energy generalization containing image boundaries and geometric features. The energy functional is expressed using the level set function which is defined a Euclidean distance function. Given a region $S$ with boundary $\partial S$, the level set function is defined as follows:

$$
\phi(x) = \begin{cases} 
-d(x, \partial S), & \forall x \in S \\
0, & \forall x \in \partial S \\
d(x, \partial S), & \forall x \not\in S,
\end{cases}
$$

(5)

where $d(x, \partial S)$ is the Euclidean distance from the point x to the boundary $\partial S$, $-d(x, \partial S)$ represents the Eulerian distance for a point x inside the region of interest. $\phi(x) = 0$ represents the segmentation boundary.
Define an image edge indicator, which is used to extract the features at the edge of the image and has a lower value at the edge, which is defined as follows:

\[ g_I = \frac{1}{1 + b |\nabla(G_\sigma * I)|^2} \]  

(6)

where \( b > 0 \) is a constant, \( \nabla \) is the Euclidean gradient operator, * \( \) is the convolutional operator, and \( \sigma \) is the deviation of the Gaussian kernel.

With these definitions in hands, we consider the following weighted curve length measured using anisotropic and asymmetric geometry features:

\[ L_{ac}(\phi) = \int_\Omega \delta_\varepsilon(x)(\langle N(x), V(x) \rangle + g_\infty)dx \]  

(7)

where \( g_\infty \equiv \max_x \), \( g_\infty \) is the maximum value of \( g_I \). The function \( \delta_\varepsilon \) is smooth approximation of the Dirac function[5],

\[ \delta_\varepsilon(x) = \frac{\varepsilon}{\pi(\phi(x)^2 + \varepsilon^2)} \]  

(8)

where \( \varepsilon > 0 \) is a sufficient small parameter, \( \pi \) is the circumference of a circle. The vector field \( V \) contains the directional characteristics of the image gradients,

\[ V(x) = \begin{cases} 
\frac{\nabla I(x)}{|\nabla I(x)|}, & \text{if } |\nabla I(x)| \neq 0 \\
0, & \text{otherwise}
\end{cases} \]  

(9)

where \( \nabla I \) represents the image gradients of \( I \). The anisotropy and the asymmetry features of the gradient gradients are involved by the vector field \( V \).

The vector field \( N \) carries out the normals to the target contours. In the level set formulation, it can be defined as:

\[ N(x) = \begin{cases} 
\nabla \phi(x), & \text{if } |\nabla \phi| \neq 0 \\
0, & \text{otherwise},
\end{cases} \]  

(10)

where \( \nabla \phi \) is the gradient of the predicted level set function. Since \( \phi \) is set as a Euclidean distance, one has \( \|\nabla \phi\| \equiv 1 \). Note that for a point \( x \) such that the direction of the vector \( V \) and the normal vector to the target contour have opposite directions, the value of \( \langle N(x), V(x) \rangle + g_\infty \) is low, and high, otherwise. This allows our model to take full advantage of

Figure 2. An example of an image where ground truth is processed by the level set function. (a) is the original image, (b) is the ground truth, and (c) is the level set of the ground Truth.
the following assumption of gradient consistency: along the boundary of the region to be segmented, the gradient direction of the image points to the interior of the region to be segmented.

Image processing of the network architecture output using the smooth Heaviside function:

\[ \zeta(x) = \frac{1}{1 + e^{-kx}}, \]  

(11)

where \( k \) is a constant and controls the steepness of \( \zeta(x) \), \( \Omega \) represents the set of point values on the probability map. \( \zeta(x) \) tends to 1 when \( x < 0 \), and \( \zeta(x) \) tends to 0 when \( x > 0 \). The probability map obtained by setting \( k \) to 1500 in the experiment is 99% similar to the original probability map[23]. After \( \zeta(x) \) function processing, the places where the threshold value is greater than 0.5 are set as the region of interest. Then using the level set function \( \phi(x) \), the level set function is obtained. Finally, the original image \( I \) and the level set function \( \phi \) are substituted into the loss function and the loss function value is calculated. Dice loss function is added during training to evaluate the overlap between the ground truth and the predicted segmentation.

\[ L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^{N} s_i g_i}{\sum_{i=1}^{N} s_i^2 + \sum_{i=1}^{N} g_i^2}. \]  

(12)

Finally, our loss function \( L_{\text{total}} \) is defined as:

\[ L_{\text{total}} = L_{\text{Dice}} + L_{\text{AC}}. \]  

(13)

Our loss function has many advantages: When only the Dice loss function is available, only the relationship between voxels is considered, as in Eq. 12, and the calculation is done only between voxels without considering the geometric properties of the image, and with the addition of \( L_{\text{AC}} \), the anisotropy and asymmetry of the image are considered in the calculation.

4. EXPERIMENTS

4.1 Data Pre-processing

**Resampling.** 3D medical images are composed of several voxels, and the voxels in medical images are often different, so the data needs to be processed into voxels of the same size using resampling. The resampling is done by linear interpolation, and the SimpleITK tool is used to implement the image resampling.

**Z-score Standardization.** The data are normalized so that the data are in the same order of magnitude. The common methods of data normalization are min-max normalization, Z-score normalization (also called mean normalization), and function transformation (e.g., log function, atan function, Sigmoid function). The normalization method used for the experiments in this paper is Z-score normalization. Its formula is as follows:

\[ z = \frac{x - \mu}{\sigma} \]  

(14)

Where \( \mu \) represents the mean of the data, \( \sigma \) represents the standard deviation of the data, and \( x \) represents the image.

4.2 Dataset

To validate our proposed method, we conduct experiments on multiple datasets in which 3D datasets with different storage methods, sizes, and categories are used. A dataset of the left atrium was used, which contains 100 Gadoliniumenhanced magnetic resonance imaging (GE-MRI) 3D MRI images, with each original image having a samesex resolution of 0.625 × 0.625 × 0.625 mm\(^3\). preprocessing preprocessing was performed on the images, and in this paper the image preprocessing using the processed data in [37]. In which 80 data were randomly selected as training set and the remaining 20 data were used as test set.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Images</th>
<th>Input Size</th>
<th>Modality</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial</td>
<td>80</td>
<td>128 × 128 × 80</td>
<td>MRI</td>
<td>2018 Atrial Segmentation Challenge</td>
</tr>
<tr>
<td>Liver</td>
<td>131</td>
<td>80 × 160 × 160</td>
<td>CT</td>
<td>The Medical Segmentation Decathlon</td>
</tr>
<tr>
<td>Prostate</td>
<td>50</td>
<td>80 × 160 × 160</td>
<td>MRI</td>
<td>PROMISE12</td>
</tr>
</tbody>
</table>

The 3D Liver dataset was obtained from The Medical Segmentation Decathlon, which includes 201 CT images from patients with liver disease. The preprocessing used the abdominal CT values window width range [-160, 240] in [34], and the images were resampled to an isotropic resolution of 3.0 × 2.2 × 2.2 mm³ due to the source of the data and individual differences in the voxel size of the images. The data after sampling were cropped from the center of the images to 80 × 160 × 160 and normalized using the z-score method. The 3D Prostate dataset from 1, which contains 50 prostate MRI images, each of which is a transverse T2-weighted image. The images were resampled to an isotropic resolution of 1 × 1 × 1 during preprocessing, and finally cropped to a size of 80 × 128 × 128 according to the center of the image, and the voxel values of the images were normalized using the z-score method. For all datasets we use 80% as training set and 20% as test set.

![Images of datasets](image-url)

Figure 3. Showing a slice of one case on each dataset demonstrates the advantages of our approach. (a)(d) represents the original images of the liver and left atrium, respectively; (b)(e) represents the segmentation results using V-Net; (c)(f) represents the segmentation results using $L_{total}$.

### 4.3 Metric

We use four evaluation metrics to evaluate our model:

**Dice coefficient**: It is an ensemble similarity measure function, which is usually used to calculate the similarity of two samples and takes values in the range of [0,1].
**Jaccard coefficient:** It is defined as the ratio of the size of the intersection of set A and set B to the size of the union of sets A and B. When sets A and B are empty, \( J(A, B) \) is defined as:

\[
J(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}
\]  \((15)\)

**95% Hausdorff distance (95HD):** It is a measure to describe the degree of similarity between two sets of points, and calculates the distance between two sets, the smaller the value, the higher the similarity of the two sets, 0 mm means perfect segmentation. The 95% aim is to eliminate the effect of outliers.

**Average surface distance (ASD):** Also called Mean surface distance, its definition is as follows:

\[
ASD(A, B) = \frac{1}{|S(A)| + |S(B)|} \left( \sum_{x \in S(A)} d(x, S(B)) + \sum_{y \in S(B)} d(y, S(A)) \right)
\]  \((16)\)

where \( d(x, S(B)) \) represents the shortest Euclidean distance from any point \( x \) in the set \( S(A) \) to set \( S(B) \).

**4.4 Experimental design**

**Experiment 1:** Ablation experiments

To verify the proposed loss function in this paper, we did ablation experiments on three datasets (left atrium, liver, and prostate). Table 2 shows the data differences between using baseline (V-Net) and the proposed loss function in this paper. From the data obtained from the experiments, the accuracy obtained from the segmented images of the proposed loss function in this paper is better than that of V-Net. Figure 4 shows the folded images of the loss function when trained on the left atrium and kidney datasets, and it can be seen from the images that the proposed loss function can converge faster and eventually to a lower position than V-Net. In terms of prediction, using the loss function proposed in this paper on left atrium, leading to increase 1.71% in Dice, 2.85% in Jaccard, decrease 2.01 voxels in ASD and 0.27 voxels in 95HD. Figure 3 shows the visual segmentation results for some examples with different contrasts on the liver and left atrium datasets. It can be seen from the figure that the method proposed in this Chinese is closer to the segmentation boundary and is closer to Ground Truth.

![Figure 4. The figure shows the images of the loss function on the left atrium (a) and the convergence of the loss function on the liver image (b). The image clearly shows that \( L_{total} \) has the advantage of faster descent and lower convergence time.](image)

Figure 4. The figure shows the images of the loss function on the left atrium (a) and the convergence of the loss function on the liver image (b). The image clearly shows that \( L_{total} \) has the advantage of faster descent and lower convergence time.
Table 2. Ablation experiments were done on the left atrium, liver, and prostate datasets, comparing the results of baseline (naive V-Net) and the proposed loss function model. The results in the table on all three datasets show that the proposed loss function results outperform naive V-Net. The rating metrics include Dice(%), Jaccard(%), ASD(voxels) and 95HD(voxels).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dice</td>
</tr>
<tr>
<td>Left Atrial</td>
<td>$L_{dice}$</td>
<td>90.87</td>
</tr>
<tr>
<td></td>
<td>$L_{total}$</td>
<td>92.58</td>
</tr>
<tr>
<td>Prostate</td>
<td>$L_{dice}$</td>
<td>95.30</td>
</tr>
<tr>
<td></td>
<td>$L_{total}$</td>
<td>95.51</td>
</tr>
<tr>
<td>Prostate</td>
<td>$L_{dice}$</td>
<td>83.83</td>
</tr>
<tr>
<td></td>
<td>$L_{total}$</td>
<td>84.25</td>
</tr>
</tbody>
</table>

Experiment 2: $L_{total}$ Comparison with $L_1$ loss function

Comparing $L_{total}$ with $L_1$ on the left atrial dataset and the liver dataset, after experiments, the result was that $L_{total}$ obtained a better segmentation prediction than that of $L_1$. Table 3 shows the results of the comparison. From the table, we can conclude that $L_{total}$ has superior results, where the Dice score $L_{AC}$ is 1.20 higher than $L_1$ for the experiments on the left atrium and 0.13 higher on the liver, and $L_{AC}$ has higher ability than $L_1$ for image segmentation.

Table 3. Segmentation results of the proposed loss function for ablation studies on the left atrial and liver datasets. Where Dice and Jaccard units are percentages(%) and ASD and 95HD units are voxels.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dice</td>
</tr>
<tr>
<td>Left Atrial</td>
<td>$L_{dice}$</td>
<td>90.87</td>
</tr>
<tr>
<td></td>
<td>$L_{total}$</td>
<td>92.58</td>
</tr>
<tr>
<td>Prostate</td>
<td>$L_{dice}$</td>
<td>95.38</td>
</tr>
<tr>
<td></td>
<td>$L_{total}$</td>
<td>95.51</td>
</tr>
</tbody>
</table>

Experiment 3: Comparison with state-of-the-art active contour loss functions (level set loss [2] and Learning GAC[34])

In addition to comparing the widely used Dice loss, we also compared level set loss and Learning GAC on the left atrium and liver datasets, and from the experimental results, using the loss function proposed in this paper to segment medical images gives better results than both of them, as can be seen from Table 4 where $L_{total}$ achieves the best results. The segmentation accuracy of $L_{total}$ improves (Dice) 2.07% over Dice loss, (Dice) 0.26% over level set loss, and (Dice) 0.3% over Learning GAC on the left atrial dataset.

Table 4. The loss function proposed in this paper is compared with other sota activity profile loss functions, including Level set loss, Learning GAC on the left atrium and liver datasets, where the optimal results are bolded. The evaluation metrics include Dice(%), Jaccard(%), ASD(voxels) and 95HD(voxels).

<table>
<thead>
<tr>
<th>Method</th>
<th>Left Atrial</th>
<th>Liver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dice</td>
<td>Jaccard</td>
</tr>
<tr>
<td>Dice loss[38]</td>
<td>90.87</td>
<td>83.38</td>
</tr>
<tr>
<td>Level set loss[2]</td>
<td>92.68</td>
<td>86.40</td>
</tr>
<tr>
<td>Learning GAC [34]</td>
<td>92.64</td>
<td>86.34</td>
</tr>
<tr>
<td>$L_{total}$</td>
<td>92.94</td>
<td>86.86</td>
</tr>
</tbody>
</table>

4.5 Detail

Experiment 1 and Experiment 2 used the results of four experiments, taking the average, and Experiment 3 obtained the highest score. The system used for the experiments is Ubuntu 18.04, and the full program implementation uses Python 3.8.8 and Pytorch [39] 1.8.1, with a NVIDIA A100 GPU with 40G of memory, cuda version 11.1, and cudnn version 8.0.5. The optimization algorithm used is SGD, and the experiments use the learning rate change strategy is the cosine
annealing algorithm [40], and standard data enhancement methods are taken to prevent data overfitting, including arbitrary noise, and arbitrary cropping. Each experimental procedure was run for 600 epochs, training the network architecture on the left atrium consumed approximately 12 hours per experiment, and training on the liver dataset took approximately 30 hours.

5. CONCLUSION

In summary, a novel loss function is proposed in this paper. It takes into account the image anisotropy, incorporates the edge information of the image, and improves the accuracy of image segmentation. Through experiments on the left atrium, liver and prostate datasets, the proposed loss function model in this paper outperforms the baseline (naive V-Net) well. The segmentation results obtained on the left atrium and liver datasets compared with the $L_1$ loss function are better than adding the $L_1$ loss function; on the left atrium and liver datasets, compared with other state-of-the-art active contour loss function, the segmentation results obtained with the loss function proposed in this paper are better compared to the other recent loss functions.

Because this loss function considers the global information of the image, the relationship between pixels, while Dice only considers the differences between individual pixels, this loss function improves the accuracy and robustness of segmentation. However, when the image features of the tissue structure are mainly distinguished by texture, this method has a large limitation. It is very useful to improve the accuracy of segmented images of medical images, because large errors in segmentation or large errors in borders are very harmful. It is believed that this loss function can be used for other challenging segmentation tasks. But this loss function also has disadvantage that the computation is only layer by layer only the edge information of each slice is computed and the 3D information between image slices is not utilized. In terms of time spent, adding $L_{AC}$ when training the network increased the time by about 11% compared to not adding it. $L_{AC}$ took 8% more time compared to $L_1$. This method improves the time spent during training. While in segmentation, with the same inference time as other models, but $L_{total}$ improves the accuracy of segmentation, we think this time spent is worthwhile.

REFERENCES


Influence of Images Generated by CGAN on the Performance of Image Classification Based on CNN

Yushu Hu1a*
1Faculty of Innovation Engineering, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau
a*19098533i011007@student.must.edu.mo

Abstract

Conditional Generative adversarial net (CGAN) can generate data according to the given constraints. Convolutional neural network (CNN) is a neural network that can be used to process images providing learnable parameters for efficient classification. However, there are few researches on whether it is more beneficial for CNN to classify from the images generated by GANs as the training set in the past studies. To solve the issue, this paper compares the training effect of CNN classification of Handwritten-number images using real images, images generated entirely by CGAN that are results through CGAN’s trained from real handwritten-number images, and images mixed with them as training data sets. The dataset used is a famous dataset called MNIST which consists of 10 categories. Based on it, the corresponding experiments are carried out and the final experimental result indicated that the images generated by CGAN can further improve the performance of the CNN.

Keywords-component: CGAN, Image Classification, Convolutional Neural Network

1. Introduction

Generative adversarial network (GAN) is a kind of neural network used in unsupervised learning, which can help to solve the tasks such as generating images from text, improving image resolution, drug matching, and retrieving images with specific patterns. The original GAN is only concerned with whether the generated image is realistic, and its shortcomings are obvious. It is needed to generate images that are predictable and realistic enough on demand [1]. Generative adversarial nets can be extended to a conditional model, called conditional Generative adversarial net (CGAN), if both the generator and discriminator are conditioned on some extra information \( y \). CGAN can generate data according to our given constraints and besides, can do multimodal learning to generate description labels related to the input image.

In 2014, Ian Goodfellow et al. proposed the initial version of generative adversarial networks (GAN) [2]. With the popularity of GANs, at least hundreds of related derivative models have emerged, such as CGAN, DCGAN, BigGAN and so on in computer vision. After 2014, DCGAN was the first GAN variant that uses convolutional neural networks for image generation. DCGAN replaces Generator and Discriminator to two CNNs and makes some changes to the structure of the convolutional neural network in order to improve the sample quality and the speed of convergence. Before BIGGAN was proposed, GANs network were difficult to generate images with great precision, clarity and quality for complex data sets such as ImageNet [3]. BigGAN balances the scale and stability of GANs, staggeringly raising IS score of a generated image to 166. CGAN, that were published with the original GAN by Mehdi Mirza and Simon Osindero in NIPS in the same year 2014, achieved great success and became one of the most basic GAN algorithms. CGAN does not improve the internal network structure of generator and discriminator like DCGAN and BigGAN, but improves the unsupervised system with supervised information, making CGAN a supervised learning. If there are multiple classes of training data and they differ widely, the shortcomings of the original GAN, that the resulting data may be a mixture of more than two classes, will be exposed. Supervised learning CGAN perfectly solves the problem by making the generator and discriminator focus on whether the generated images conform to the label and meanwhile are real.

However, there are few researches on Whether it is more beneficial for CNN to classify from the images generated by GANs as the training set. CGAN is able to be chosen as a representative of GANs to test the ability, because it generates images that can be clearly labeled for classification and can that be easily controlled to generate certain directed data.
This paper compares the training effect of CNN classification of Handwritten-number images using real images, images generated entirely by CGAN that are results through CGAN’s trained from real handwritten-number images, and images mixed with them as training data sets.

2. Method

2.1. General Method

Firstly, this paper trained a conditional adversarial net on MNIST Handwritten-number images conditioned on their class labels like Mehdi Mirza and Simon Osindero. The result images, five hundred pictures for each number (0–9), are generated after 2,000,000 iterations. Each image is stored as grayscale image in a pixel size of 389 by 389 with tight border. Output five hundred pictures for each number from the MNIST handwritten number set in the form of feature vector and transform them to the same size and the same style. Then this paper lists three sets of data: real handwritten-number images from MNIST (500 pictures for each number), CGAN-generated images (500 pictures for each number), and mixture (250 real pictures and 250 generated pictures for each number). Each set of data is sent to be classified for CNN and the accuracy and loss rate for each set will be compared in order to find whether CGAN’s generated images are beneficial to CNN’s classification. Flowchart 2.1.1 shows the process of the whole method.

![Flowchart 2.1.1 showing the process of the whole method.](image)

2.2. Data set Preparation and Preprocessing

MNIST is a classic handwritten digits image data set. There are 55000 images in the training set, and the size of each image is 784(28*28); The training set has 55000 labels, and each label is a one-dimensional array of length 10. Each image represents a handwritten number, and the number below each box is the label corresponding to this image. Fig. 2 shows the first samples of the MNIST training set. MNIST data set is chosen to train CGAN and CNN in this article.

Here is the preprocessing. The size and the format of the real images group should be the same as that of the generated images, so that the same number of images from MNIST(500 pictures for each number) with the the generated images group should be transformed to grayscale images in a pixel size of 389 by 389 with tight border. Fig. 2 shows an example in real images group.
2.3. CGAN for generation

2.3.1 Introduction for the CGAN

The original GAN just cares about whether the generated pictures are real, so that the resulting data from it may be a mixture of more than two classes, if there are multiple classes of training data and they differ widely. For example, the train, there is a side to see the train, and there is also a front position to see the oncoming train as Fig. 3.

If one just tell G to generate an image of a train, it will probably generate a mixture of a front train and a side train, and the result will be strange. In order to solve the above problem, it is need to take condition C as the common input of G and D, so that the final judgment result of D will not only consider the authenticity of the image, but also consider whether it meets the description of the condition, which realizes the basic idea of CGAN. Fig. 4 shows G and D in the example.

Therefore, CGAN can generate data according to our given constraints.
2.3.2 Details of generating numbers for CGAN

The paper trains a conditional adversarial net on MNIST images conditioned on their class labels, and it generates five hundred pictures as grayscale images in a pixel size of 389 by 389 with tight border for each number (0~9) after 2,000,000 iterations.

The implementation of the CGAN model refers to Agustinus Kristiadi’s generative-models [4]. In the generator net, a noise prior z with dimensionality 100 was drawn from a uniform distribution within the unit hypercube. Both generator and discriminator contain two-layers-network and Fig. 5 shows their structures.

![Fig. 5 The structure of generator and discriminator.](image)

2.4. CNN for classification

Convolutional neural network (CNN) is a neural network that can be used to process images, which is widely used in past studies [5-10]. This type of neural network takes inputs from images and extracts features from images, and provides learnable parameters for efficient classification, detection, and more tasks. In this paper, CNN is trained to classify Handwritten-number images into ten categories from three groups data and the accuracy and loss rate will be tested and compared to find which group the data of are beneficial to CNN’s classification.

Before training, this paper carries out Data Augmentation to the training set, such as stretching, normalization, horizontal flip and so on. It's worth noting that it's better to only do the horizontal transformation, because the numbers 6 and 9 are centrally symmetric. (Totally 5000 images containing 0~9 in each group data) 80% of data as a training set are used for training the CNN module and the remaining 20% as a test set are used for testing the accuracy of classification. The CNN model in this paper has three convolution layers and pooling layers connected sequentially, and finally connected to the fully connected layer. Table 1 shows details of the model.

<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output shape</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv2D</td>
<td>(None,389,389,16)</td>
<td>448</td>
</tr>
<tr>
<td>Max Pooling</td>
<td>(None, 194,194,16)</td>
<td>0</td>
</tr>
<tr>
<td>dropout</td>
<td>(None, 194,194,16)</td>
<td>0</td>
</tr>
<tr>
<td>Conv2D</td>
<td>(None, 194,194,32)</td>
<td>4640</td>
</tr>
<tr>
<td>Max Pooling</td>
<td>(None, 97,97,32)</td>
<td>0</td>
</tr>
<tr>
<td>Conv2D</td>
<td>(None, 97,97,64)</td>
<td>18496</td>
</tr>
<tr>
<td>Max Pooling</td>
<td>(None, 48,48,64)</td>
<td>0</td>
</tr>
<tr>
<td>dropout</td>
<td>(None, 48,48,64)</td>
<td>0</td>
</tr>
<tr>
<td>Flatten</td>
<td>(None,147456)</td>
<td>0</td>
</tr>
<tr>
<td>Dense</td>
<td>(None,512)</td>
<td>75497984</td>
</tr>
<tr>
<td>Dense</td>
<td>(None,10)</td>
<td>5130</td>
</tr>
</tbody>
</table>

3. Result and Discussion

3.1 CGAN for generation

After 2,000,000 iterations, the discriminators loss is 0.571 and the generated loss is 2.238. Fig. 6 shows ten examples of generated images (number 0~9).
Human eyes can tell that the resulting images have almost the same shape as real handwritten numbers. It demonstrated that the CGAN model can generate realistic images in this case.

### 3.2 CNN for classification

Fig. 7, 8 and 9 shows the accuracy and loss rate of classification for 20 times from each group.

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Loss</th>
<th>Accuracy</th>
<th>Val_Loss</th>
<th>Val_Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/20</td>
<td>0.0218</td>
<td>0.7890</td>
<td>0.0850</td>
<td>0.9318</td>
</tr>
<tr>
<td>0.0214</td>
<td>0.5273</td>
<td>0.1310</td>
<td>0.9500</td>
<td>0.9500</td>
</tr>
<tr>
<td>0.0279</td>
<td>0.9557</td>
<td>0.0806</td>
<td>0.9750</td>
<td>0.9750</td>
</tr>
<tr>
<td>0.0296</td>
<td>0.9725</td>
<td>0.1265</td>
<td>0.9620</td>
<td>0.9620</td>
</tr>
<tr>
<td>0.0335</td>
<td>0.9703</td>
<td>0.0651</td>
<td>0.9820</td>
<td>0.9820</td>
</tr>
<tr>
<td>0.0474</td>
<td>0.9743</td>
<td>0.0563</td>
<td>0.9850</td>
<td>0.9850</td>
</tr>
<tr>
<td>0.0490</td>
<td>0.9845</td>
<td>0.0872</td>
<td>0.9870</td>
<td>0.9870</td>
</tr>
<tr>
<td>0.0623</td>
<td>0.9810</td>
<td>0.0580</td>
<td>0.9780</td>
<td>0.9780</td>
</tr>
<tr>
<td>0.0294</td>
<td>0.9902</td>
<td>0.0527</td>
<td>0.9870</td>
<td>0.9870</td>
</tr>
<tr>
<td>0.0342</td>
<td>0.9868</td>
<td>0.0418</td>
<td>0.9888</td>
<td>0.9888</td>
</tr>
<tr>
<td>0.0328</td>
<td>0.9900</td>
<td>0.0926</td>
<td>0.9668</td>
<td>0.9668</td>
</tr>
<tr>
<td>0.0331</td>
<td>0.9893</td>
<td>0.0565</td>
<td>0.9830</td>
<td>0.9830</td>
</tr>
<tr>
<td>0.0479</td>
<td>0.9847</td>
<td>0.0864</td>
<td>0.9700</td>
<td>0.9700</td>
</tr>
<tr>
<td>0.0288</td>
<td>0.9940</td>
<td>0.0567</td>
<td>0.9840</td>
<td>0.9840</td>
</tr>
<tr>
<td>0.0425</td>
<td>0.9865</td>
<td>0.0285</td>
<td>0.9800</td>
<td>0.9800</td>
</tr>
<tr>
<td>0.0292</td>
<td>0.9900</td>
<td>0.0481</td>
<td>0.9840</td>
<td>0.9840</td>
</tr>
<tr>
<td>0.0241</td>
<td>0.9912</td>
<td>0.0934</td>
<td>0.9770</td>
<td>0.9770</td>
</tr>
<tr>
<td>0.0213</td>
<td>0.9935</td>
<td>0.2327</td>
<td>0.9420</td>
<td>0.9420</td>
</tr>
<tr>
<td>0.0747</td>
<td>0.9868</td>
<td>0.0493</td>
<td>0.9890</td>
<td>0.9890</td>
</tr>
<tr>
<td>0.0225</td>
<td>0.9930</td>
<td>0.0372</td>
<td>0.9900</td>
<td>0.9900</td>
</tr>
</tbody>
</table>

**Fig. 8** The performance of Mixed images.
It can be found that the accuracy of the group completely consist of CGAN’s generated images is highest and that of the group completely from MNIST is lowest. Therefore, it can be concluded that CGAN’s generated images are beneficial to CNN’s classification.

The possible reason for this result is that the generated images can be considered as the data augmentation to further enhance the amount of data during the model training process. Based on it, the CNN model can see more variants of the data and achieve better results.

4. Conclusion

This work is to explore the influence of images generated by CGAN on the accuracy of image classification by CNN. The paper lists three sets of data: real handwritten-number images from MNIST, CGAN-generated images, and mixture. Each set of data is sent to be classified for CNN and the accuracy and loss rate for each set will be compared and this study proves that CGAN’s generated images are beneficial to CNN’s classification. In the future, further study can attempt to train some nets from the data generated by GANs to improve accuracy and to find more benefits.

References

Medical Data Sharing Scheme Based on Hybrid Encryption with Revocable Attribute

Guo-Shuai Liu¹,a, Feng Guo¹,b, Chuan-Kun Wu¹,c
¹School of information science and engineering, Linyi University, Shandong 276000, China
a liu779584750@163.com, b guofeng-g@lyu.edu.cn, c chuanjunwu@lyu.edu.cn

Abstract

With the popularity of electronic medical record, internet medical care, AI medical image and other applications, the wave of medical digitization is coming. The electronization of medical data has brought convenience to the sharing of medical data, but it also makes the protection of personal sensitive data in medical data face unprecedented challenges. How to effectively protect personal sensitive information while sharing medical data. In this paper, a hybrid encryption algorithm based on SM4 encryption algorithm and Ciphertext-Policy Attribute-Based Encryption (CP-ABE) with revocable attributes is proposed. The algorithm first uses the SM4 encryption algorithm to encrypt the medical data, and then the SM4 encryption key is used as the encrypted data for encryption by CP-ABE. At the same time, aiming at the requirement of CP-ABE algorithm to revoke attributes, this paper proposes to embed an revocation table in the ciphertext to realize the direct revocation and restore of specific attributes of specific users. Through theoretical analysis and experimental verification, the revocable attribute hybrid encryption algorithm proposed has great advantages for the security and efficiency of personal sensitive information in cloud medical data.

Keywords-SM4 encryption algorithm, Data sharing, CP-ABE, Attribute revocable.

1. Introduction

With the rapid development of science and technology, people pay more and more attention to health. Medical data is a valuable resource in the medical and health field. Medical data has guiding significance for scientific research in the medical field. However, in recent years, medical data breaches have occurred frequently, and the disclosure of medical privacy data has brought many problems to individuals. Unexpected data leakage of the medical data may cause significant psychological harm to the patients or even threaten their lives[1].

There are many different roles in the medical system: patients, doctors, researchers, insurers, etc. How to protect the privacy of patient data when these different roles query patient data through the network. CP-ABE[2] encryption algorithm relates the user's private key to a series of user attributes. Only users who meet the specified access structure can decrypt the information. On the contrary, users who do not meet the access structure can not decrypt the encrypted information, which conforms to the multi-role encryption scenario.

The earliest prototype of CP-ABE comes from Sahai[3]. In the fault tolerance problem of biometric encryption, using bilinear pair knowledge, this article proposes a concept of fuzzy identity based encryption, which leads to the concept of ABE, and gives the ABE scheme of small-scale attribute set and large-scale attribute set. However, this article is limited to threshold access structure and can not be widely used. on the basis of Sahai, Goyal et al.[4] first divided ABE into KP-ABE and CP-ABE schemes and introduced the access structure tree. Attrapadung et al.[5] proposed a hybrid attribute revocation CP-ABE scheme, which supports the data owner to choose direct and indirect ways, but can not be changed later. Yang et al. [6]’s CP-ABE scheme that supports fine-grained attribute revocation in cloud storage environment does not require the data owner to be online in real time, but the security is proved based on random oracle model, which is not ideal. Zhang et al. [7] proposed a CP-ABE scheme supporting AND gate, which supports direct attribute revocation, but only supports AND access policy. Xue et al. [8] proposed a fine-grained attribute revocation CP-ABE scheme based on non-monotonic access structure, but the defect is that it is constructed based on composite order group, which is inefficient.

Compared with the symmetric encryption algorithm, the encryption efficiency of CP-ABE is slower. Therefore, the hybrid encryption combining symmetric encryption algorithm and CP-ABE encryption can not only meet the requirements of high encryption efficiency, but also achieve the requirements of access control of CP-ABE encryption.
algorithm. Chong Bao et al. [9] implemented a hybrid encryption algorithm combining DES[10] and CP-ABE to solve the problem of data access control in cloud storage. Fugkeaw[11] implemented the combination of AES and CP-ABE to solve the access control problem in the mobile cloud environment. SM4 encryption algorithm is a symmetric encryption algorithm issued by National Cryptography Administration of China, which is used to replace AES[12] and DES. At the same time, there are a large number of users in the medical system, and the user identity often changes, so there is a need to revoke the user attributes. To solve the above problems, this paper proposes a revocable algorithm based on SM4[13] and CP-ABE hybrid encryption attributes, which can effectively protect data security and realize the access control of medical data sharing. At the same time, the revocation and recovery of attributes are realized by revocation table in the ciphertext.

2. Basic Knowledge

2.1. SM4 Encryption Algorithm
SM4 encryption algorithm has officially become the ISO/IEC international standard in June 2021[14]. The algorithm is a block cipher, the plaintext block is 128 bits, and the key length is also 128 bits. It can run in the microprocessor, and its encryption and decryption adopt an initial key.

2.2. DBDH Problem Hypothesis
Two multiplicative groups $G_1$ and $G_2$ with prime order $p$ are known, and exists $G_1 \times G_1 = G_2$ bilinear mapping, random selection of generators $g \in G_1$, random number $c_1, c_2, c_3 \in Z_p$ and $n \in G_2$. If there is no method to judge $e(g, g)^{c_1c_2c_3}$ and $e(g, g)^n$ with non-negligible advantages in polynomial time, the Decisional Bilinear Diffie Hellman (DBDH) hypothesis is true.

2.3. Lagrange Polynomial
Given $d$ points $(x_0, y_0), (x_1, y_1), \ldots, (x_{d-1}, y_{d-1})$. You can determine $d - 1$ this polynomial, $d - 1$ polynomial can be determined. The polynomial can be expressed as:

$$q(x) = \sum_{i=0}^{d-1} y_i \delta_i(x)$$

Lagrange factor $\delta_i(x) = \prod_{j=0, j \neq i}^{d-1} \frac{x-x_j}{x_i-x_j}$

3. Design of Revocable Hybrid Encryption Scheme for Medical Data

3.1. Overall structure of the scheme
This paper proposes a revocable medical data sharing architecture based on hybrid encryption attributes. According to the principles of medical data generation, processing and application, the medical data sharing system is divided into three layer architecture, namely device terminal layer, data processing layer and user management layer, as shown in Fig.1.
(1) Device terminal layer: mainly Internet of things devices (IOTD), including medical data generation device. For example, CT machine, Mr machine, etc. Data generated by the medical data generation device is encrypted with SM4 encryption algorithm to generate the medical data ciphertext. According to the access policy specified by the system, CP-ABE is used to encrypt the SM4 encryption key to obtain the key ciphertext, and the two ciphertexts are uploaded to the cloud database (CD) of the data processing layer for storage.

(2) Data processing layer: the main function is to store the medical data of patients and provide services to the user management layer. The data processing layer not only stores the medical data uploaded by the Internet of things devices in the device terminal layer, but also stores the patient diagnosis reports, patient past medical records and other data uploaded by patients and doctors in the user management layer. These medical data are stored after encryption using the hybrid encryption scheme in this paper. The entities providing services to the user management layer include cloud server (CS), revocation mechanism (RM) and authorization center (AC).

(3) User management layer: mainly users involved in different roles of medical data: patients, doctors, researchers, insurers, etc. Users in these different roles have requests to request data sharing from the data processing layer.

Different entities of the medical system have different tasks:

(1) Data owner (DO) is the owner of data, the encryptor of data, and the maker of access policy of hybrid encryption algorithm. It is credible. (2) The authorization center is the entity of CP-ABE public-private key generation and data requester registration, which is completely trusted. (3) The revocation mechanism is the entity that revokes and restores the user attribute of CP-ABE in the scheme of this paper, which is trusted. (4) Cloud server and cloud database are outsourced by medical institutions to service providers to provide stable and reliable services for the system. The outsourced service providers are honest but curious and not completely credible. Therefore, the data handed over to the cloud server and stored in cloud database are encrypted. (5) The data requester (DR) has a request to query medical data. At the same time, they may also be the leaker of medical data, which is untrusted. Therefore, they need to get medical data that conforms to their identity attributes.

Among them, patients and doctors belong to both data owners and data requesters, and the Internet of things devices in the device layer also belong to DO.

3.2. Implementation of the Scheme

Based on the work of Brent waters et al. [15], this paper proposes a revocable medical sharing scheme based on hybrid encryption attributes. The proposed scheme supports the revocation and recovery of user attributes. The access policy of CP-ABE algorithm adopts the way of user customized access tree. The scheme consists of seven stages: system initialization, revocation mechanism initialization, key generation, attribute revocation, attribute recovery, hybrid encryption and hybrid decryption.

3.2.1 System initialization

Authorization center AC generates $G_1,G_T$ and $g$, where $e:G_1 \times G_1 \rightarrow G_T$ is a bilinear mapping, $G_1$ and $G_T$ is a multiplicative cyclic group of prime order $p$, and $g$ is Generator of $G_1$. Random selection $\alpha, \beta \in Z_p$. Calculate $g^\alpha,g^\beta,Y = e(g,g)^\alpha$. Finally, AC sets the master key $\text{MSK}=g^\alpha$, Public key $\text{pk}=<Y, g^\beta>$ as output.

3.2.2 Revocation mechanism initialization

Revocation mechanism RM randomly generates prime set $P_1, P_2$ and $P_1 \cap P_2 = \emptyset$, and generate two tables $\text{list1} <\text{userID, sn1}>, \text{list2} <\text{attr, sn2}>$. list1 is the user table, which stores the user's identity information userID and prime sn1; list2 is an undo table that stores the user attribute attr and the prime number sn2.

3.2.3 Key generation

The authorization center AC selects the random number $t \in Z_T$. Calculate $D = g^\alpha g^\beta t, D_0 = g^t$. Assign a prime number $sn1 \in P_1$ to each registered data requester, and the userID and sn1 of the data requester are stored in list1 in the form of binary $<\text{userID, sn1}>$. Select the hash function $H(x)$, and calculate $D_i = H(i)^t$. The authorization center outputs revocation table list2 and user private key $sk = <D,D_0,U_i=\text{attr, sn1}>$. 

---

Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie
3.2.4 Attribute revocation

Suppose the attribute \( \text{att}_1 \) of user \( U \) is revoked, input the user \( U \)’s \( \text{userID} \), attribute \( \text{att}_1 \), user table \( \text{list1} \) and revocation table \( \text{list2} \). The revocation mechanism queries the \( \text{sn1} \) corresponding to the \( \text{userID} \) from the user table \( \text{list1} < \text{userID}, \text{sn1} > \) and gives the revocation attribute \( \text{att}_1 \) assigns a prime number \( \text{sn2} \in P_2 \), and calculate \( \text{sn2} = \text{sn1} \times \text{sn2} \). Finally, the attribute \( \text{att}_1 \) and the prime number \( \text{sn2} \) are stored in the revocation table \( \text{list2} \) as a binary \( < \text{attr, sn2} > \).

3.2.5 Attribute recovery

When restoring attribute \( \text{att}_1 \) of user \( U \), input user \( U \)’s \( \text{userID} \), attribute \( \text{att}_1 \), user table \( \text{list1} \) and revocation table \( \text{list2} \). The revocation mechanism queries the user table \( \text{list1} \) and revocation table \( \text{list2} \) to obtain \( \text{sn1} \) and \( \text{sn2} \) corresponding to the user’s \( \text{userID} \) and attribute \( \text{att}_1 \). The processing method of the two prime numbers is \( \text{sn2} = \text{sn2}/\text{sn1} \), and the value of \( \text{sn2} \) in the revocation table \( \text{list2} \) is updated.

The attribute has been restored, but there is still the restored \( \text{sn2} \) value in the revocation table \( \text{list2} \), which will cause the length of the revocation table embedded in the ciphertext to be too long. Therefore, a mechanism for regularly updating the revocation table should be established to delete the \( \text{sn2} \) of the restored attribute. The update mechanism of the revocation table is as follows: within a fixed time, such as one week, the revocation mechanism calculates \( \text{gcd} (\text{sn1}, \text{sn2}) \) for \( \text{sn2} \) in the revocation table. If \( \text{gcd} (\text{sn1}, \text{sn2}) \neq 1 \), it means that no user has revoked the attribute corresponding to \( \text{sn2} \), then the binary \( < \text{attr, sn2} > \) is deleted from the revocation table \( \text{list2} \).

3.2.6 Hybrid encryption

The data owner runs the hybrid encryption algorithm. The encryption process is mainly divided into two steps:

Step 1: the data owner randomly generates the 128bit encryption key \( k \) required by the SM4 encryption algorithm, executes the SM4 encryption algorithm to encrypt the plaintext \( m \), and outputs the ciphertext \( dc \).

Step 2: execute CP-ABE encryption algorithm, randomly generate secret value \( v \) and use SM4 algorithm encryption key \( k \) as encrypted data. Calculation \( C = e(g, g)^{avk}, C_0 = g^v \).

Generate access policy tree \( T_W \). For the access policy tree \( T_W \) treatment is as follows:

According to index of tree breadth first search algorithm, number \( x \in \{0, 1, 2, \ldots, n \} \) nodes for each tree. Make root node \( q(x) | x = 0 = v \), then, according to the index value \( x \) order, randomly select one for other non leaf nodes \( cf \in Z_r \).

Construct \( d - 1 \) degree polynomial \( q(x) = cf[0] + cf[1]x^1 + \cdots + cf[d - 1]x^{d - 1} \) with index value \( x \) and its threshold value \( d \), and let \( cf[0] \) be the polynomial \( q(x) \) value of its parent node. Then the generated \( q(x) \) value is shared as a secret partition to its child nodes; For attribute \( i \) of leaf node, random selection \( r \in Z_r \), calculation \( C_i = g^{\beta q(x)}H(i)^{-r_i}, C_i^2 = g^{r_i} \), eventually, the ciphertext is output \( kc = < C, list2, C_0, (C_i^1, C_i^2)_{i \in T_W} > \).

Finally, the medical data ciphertext \( dc \) and key ciphertext \( k \) are output through the hybrid encryption algorithm.

3.2.7 Hybrid decryption

The data requester runs the hybrid decryption algorithm, which is divided into three steps:

Step 1: the data requestor's private key attribute set \( C \), start with leaf node \( i \) of access policy tree \( T_W \), if \( i \in C \), the algorithm obtains \( sn2 \) corresponding to attribute \( i \) from the revocation table \( list2 \), obtains \( sn1 \) from the private key \( sk \), and determines \( \text{gcd} (\text{sn1}, \text{sn2}) \). If \( \text{gcd} (\text{sn1}, \text{sn2}) = 1 \), it indicates that the attribute \( i \) of the data requestor has not been revoked. If \( \text{gcd} (\text{sn1}, \text{sn2}) \neq 1 \), indicates that the attribute \( i \) of the data requestor has been revoked, remove attribute \( i \) from set \( C \) to obtain attribute set \( C' \).

Finally, the attribute set \( C' \), which satisfies the access policy tree \( T_W \) can successfully decrypt the secret value \( v \), otherwise the algorithm outputs *, and the decryption algorithm is terminated.

Step 2: if the attribute set \( C' \) is consistent with the leaf node \( i \) attribute of the access policy tree \( T_W \), calculate

\[
S_i = e(C_i^1, D_i)e(C_i^2, D_i) \\
= e(g^{\beta q(x)}H(i)^{-r_i}, g^v) e(g^{r_i}, H(i)^v) \\
= e(g, g)^{\beta q(x)v}  \quad (2)
\]

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 01 Nov 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use
$S_1$ is used as a secret partition to recover the secret by Lagrange Polynomial, and the recovered secret value is recovered in the form of $e(g,g)^{\text{rev}}$. Then calculate:

$$\frac{e(C_0D)}{e(g,g)^{\text{rev}}} = \frac{e(g,g)^{\text{rev}e(g,g)\text{av}}}{e(g,g)^{\text{rev}}} = e(g,g)^{\text{av}} \tag{3}$$

Finally, after decryption, the SM4 key $k = \frac{c}{e(g,g)^{\text{av}}}$ is obtained.

Step 3: the data owner runs the SM4 decryption algorithm, inputs the SM4 key and the medical data ciphertext $dc$, decrypts the data ciphertext to obtain the medical data plaintext $m$.

4. Security Analysis of the Scheme

The CP-ABE encryption security model is proved by the game between efficient attacker $A$ and Challenger $C$. Under the assumption of DBDH, the indistinguishability under selective plaintext attack is achieved.

1. Parameter setting phase: Challenger $C$ runs the system initialization algorithm to obtain $MSK$, $PK$, and sends $PK$ to efficient attacker $A$. $A$ simulates data owner output access policy tree $T$.

2. Key query phase 1: Efficient attacker $A$ queries a set of keys and user identities $user_1$ related to attribute set $S_1$. $A$ sends $S_1$ and $user_1$ to $C$, and then $C$ returns the private key $SK$ corresponding to $A$. However, $SK$ cannot meet the access policy $T$.

3. Challenge phase: Efficient attacker $A$ randomly submits two messages $M_0, M_1$ with the same length to Challenger $C$. $C$ randomly selects $b \in \{0,1\}$, simulates the CP-ABE encryption algorithm of the data owner to encrypt $M_b$ to obtain the challenge ciphertext $CT$, and then sends the $CT$ to $A$.

4. Key query phase 2: Similar to key query phase 1, efficient attacker $A$ queries another set of key user identities $user_2$ related to attribute set $S_2$, and $C$ returns $A$ private key $sk_2$. Similarly, $sk_2$ cannot meet the access policy $T$.

5. Guess stage: Efficient attacker $A$ outputs the guess result $b_1 \in \{0,1\}$. If $b_1 = b$, the game succeeds.

The following conclusions describe the security of the hybrid encryption scheme proposed in this paper.

Conclusion 1: Suppose there is an attacker $A$ who succeeds in the game with a non-negligible advantage $\gamma$, the advantage of Challenger $C$ in solving the DBDH problem is $\frac{\gamma}{2}$.

Certification:

(1) Parameter setting phase: Challenger $C$ selects multiplication groups $G_1$ and $G_T$ of order $p$, and there is $e: G_1 \times G_1 \rightarrow G_T$ is a bilinear map, and random selection generator $g \in G_1$. random number $c_1, c_2, c_3 \in Z_p$. $C$ conduct an experiment, order $Z = c_1c_2c_3$, record as experiment 1, let $Z = n$ and $n \in Z_r$, is recorded as experiment 2, $C$ randomly selects one for the above experiment, then $C$ gets $(g, g^{c_1}, g^{c_2}, g^{c_3}, e(g, g)^{c_1})$. $C$ simulation authorization center selection $\alpha, \beta \in Z_r$, calculation $g^\alpha, g^\beta$, get $MSK$ and public key $pk = < e(g, g)^\alpha, g^\beta >$, the public key $pk$ is returned to attacker $A$, which simulates the data owner and outputs the access policy tree $T$.

(2) Key query phase 1: Attack $A$ generates a set of and attributes $S_1 = \{i_1, i_2 \ldots i_n\}$ and their user identities $user_1$, and $S_1$ does not satisfy the access policy tree $T$. $A$ sends $S_1$ to $C$, $C$ calculation $D_t = H(i_m)^t (m = 1,2 \ldots n)$, pick random number $t \in Z_r$, calculation $D = g^\alpha g^\beta$, $D_0 = g^t$, private key $sk_1 = < D, D_0, \{D_t\}_{i \in S_1}, s1 >$, then $C$ returns $sk_1$ to $A$.

(3) Challenge phase: Efficient attacker $A$ randomly submits two messages of equal length $M_0, M_1$ to Challenger $C$. $C$ random selection $b_1 \in \{0,1\}$, calculate: $C_1 = e(g, g)^{\alpha b}M_0, C_2 = e(g, g)^{\beta b}M_1, C_0 = g^\nu$. Then $C$ sends the calculation result $CT$ to $A$.

(4) Key query phase 2: Similar to key query phase 1, repeat key query is performed.

(5) Guess stage: Attacker $A$ outputs the guess result $b_1 \in \{0,1\}$. If $b_1 = b$, then the game of $A$ is successful, According to the guess $b_1 = C$, $C$ will also guess, If $C$ guesses to conduct experiment 1, that is $Z = c_1c_2c_3$, form a quintuple $(g, g^{c_1}, g^{c_2}, g^{c_3}, e(g, g)^{c_1c_2c_3})$, this quintuple constitutes a valid DBDH group. If $C$ guesses to conduct experiment 2, a quintuple $(g, g^{c_1}, g^{c_2}, g^{c_3}, e(g, g)^{c_1c_2c_3})$ will be formed. The advantages of attacker $A$:
Pr [b₁ = b] = 1 + γ  \hspace{1cm} (4)

So the advantage of Challenger C:

\[ Adv_C = \frac{1}{2} \text{Pr} [b_1 = b] - \frac{1}{2} = \frac{\gamma}{2} \hspace{1cm} (5) \]

Therefore, conclusion 1 is established.

5. Performance Analysis of the Scheme

The encryption scheme proposed in this paper realizes the implements revocation of CP-ABE attributes by embedding tables in the ciphertext. The following is a comparison between the hybrid encryption scheme in this paper and similar literatures in terms of public key size, private key size, key length, revocation mechanism and access policy. The comparison results are shown in Table 1.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>pk size</th>
<th>sk size</th>
<th>Ciphertext size</th>
<th>Revocation mechanism</th>
<th>Access policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>literature [6]</td>
<td>2BL₁</td>
<td>(B + 3)L₁</td>
<td>(2D + 1)L₁ + Lₚ</td>
<td>Attribute revocation</td>
<td>Access tree</td>
</tr>
<tr>
<td>literature [8]</td>
<td>(B + 2)L₁ + 2Lₚ</td>
<td>(B + 2)L₁</td>
<td>(2D + 1)L₁ + Lₚ</td>
<td>Attribute revocation</td>
<td>Access tree</td>
</tr>
<tr>
<td>literature [16]</td>
<td>BL₁ + Lₚ</td>
<td>(B + 1)L₁</td>
<td>5L₁</td>
<td>Attribute revocation</td>
<td>AND</td>
</tr>
<tr>
<td>literature [17]</td>
<td>nL₁</td>
<td>(B + 3)L₁</td>
<td>(2D + 2)L₁</td>
<td>Attribute revocation</td>
<td>Access tree</td>
</tr>
<tr>
<td>literature [18]</td>
<td>(B + 3)L₁ + BLₚ</td>
<td>(B + 1)L₁</td>
<td>2L₁ + Lₚ</td>
<td>User revocation</td>
<td>LSSS</td>
</tr>
<tr>
<td>our</td>
<td>L₁ + Lₚ</td>
<td>(B + 2)L₁</td>
<td>(D + 1)L₁ + Lₚ + R</td>
<td>Attribute revocation</td>
<td>Access tree</td>
</tr>
</tbody>
</table>

In Table 1, B represents the maximum number of attributes in the system, R represents the length of the revocation table in the ciphertext, D represents the number of attributes in the data owner’s access policy, n represents the number of attribute authorization agencies in the system, and L₁ represents the bit length of G₁ cyclic group, and Lₚ represents the bit length of Gₚ cyclic group.

According to Table 1, in terms of the public key size of CP-ABE encryption, compared with other comparative literatures, the public key length of the hybrid encryption scheme in this paper is the smallest, because the public key of the hybrid encryption scheme in this paper is not generated based on user attributes, but is encrypted after the public key and user attributes are processed in the process of data encryption, and the public key in document [17] is not generated based on user attributes. However, the length of its public key is related to attribute authorization agencies in multiple systems, and its advantage in public key length is not as good as the hybrid encryption scheme in this paper; In the comparison of private key length and ciphertext length, the hybrid encryption scheme in this paper has disadvantages compared with literature [16] and literature [18]. Although literature [16] and literature [18] achieve a fixed ciphertext length, the access policy in literature [16] adopts an and structure, which is not as flexible as the access tree structure of the hybrid encryption scheme in this paper. Literature [18] only implements user revocation in terms of revocation mechanism, but does not realize the revocation of user attributes, The fine granularity of revocation is coarser than the hybrid encryption scheme in this paper.

To sum up, the hybrid encryption scheme in this paper realizes the access control of user attributes through the access structure of the access tree, and realizes the direct attribute revocation for a specific user by embedding the revocation list in the ciphertext, achieving better fine granularity.

6. Experimental Analysis

In view of the low efficiency of CP-ABE encryption algorithm, this paper proposes a hybrid encryption algorithm based on CP-ABE and SM4 to solve this problem. This section uses the above experimental machines and environments for comparative testing. The test environment is as follows:

Experimental machine: Inter(R) Core(TM) i5-7200U 2.50GHz Memory: 8GB Operating system: Windows 10 testing
The CP-ABE access policy tree used in the test has five attributes. The hybrid encryption algorithm and other encryption algorithms in this paper are used to test and compare the encryption/decryption of file data of different sizes. From the comparison of the hybrid encryption algorithm described in Fig.2 with SM4 encryption algorithm and CP-ABE encryption algorithm, it can be seen that when encrypting data files of the same size, the hybrid encryption algorithm in this paper has greater advantages in encryption efficiency than CP-ABE encryption algorithm, and is close to SM4 encryption algorithm. As can be seen from Fig.3, when decrypting data files of the same size, the hybrid encryption algorithm designed in this paper is close to SM4 encryption algorithm in terms of decryption efficiency, and CP-ABE encryption algorithm is far behind the hybrid encryption algorithm in this paper.

In order to further test the efficiency of the hybrid encryption algorithm in this paper, the hybrid encryption algorithm in this paper encrypts files of the same size while increasing the number of attributes of the CP-ABE access tree. As shown in Fig.4, when using the hybrid encryption algorithm in this paper to process 1MB file data, the encryption time of CP-
ABE increases with the increase of the number of attributes in the access tree. However, when the number of attributes of the access tree is large, CP-ABE encryption algorithm still maintains good encryption efficiency. Similar to CP-ABE encryption process, the decryption time of CP-ABE algorithm also increases with the increase of the attributes of the access tree, but its decryption time increases slowly. Fig.4 shows that when the number of attributes of the access tree increases from 3 to 11, the decryption time of CP-ABE algorithm is only increased by 40ms.

From the above experimental results, we can see that the attribute revocable encryption algorithm based on SM4 and ciphertext strategy proposed in this paper has better encryption/decryption efficiency and can meet the needs of user.

7. Conclusion

This paper proposes an attribute revocable encryption algorithm based on SM4 and ciphertext policy, which can effectively solve the problems of privacy protection and low efficiency of CP-ABE encryption and decryption caused by different users' access in medical data sharing. At the same time, the revocation table is embedded in the ciphertext to realize the direct revocation and recovery of user attributes. Through experimental verification and comparison, the hybrid encryption scheme in this paper has certain advantages in security and efficiency.

References

Data collection and talent demand analysis of recruitment website based on principal component analysis

Jihui Fan*\textsuperscript{a,b}, Fengshan Yuan\textsuperscript{b}
\textsuperscript{a}Department of graduate school, St.Paul University Philippines, Tuguegarao, Cagayan, Philippines
\textsuperscript{b}School of computer science and engineering, Guangzhou Institute of Science and Technology, Guangzhou, China
*Corresponding author: (2519639989@qq.com)

ABSTRACT

The talent structure has been flattened, from the pyramid talent structure to the olive talent structure. Jobs are more complicated, industries are upgraded from low-end to high-end, and the situation faced by jobs is more complex. Talent specification compounding, industry understanding, data thinking, programming ability, innovative thinking and comprehensive problem-solving ability. With the rapid development of science and technology, the demand for social talents is constantly changing, and the big data wave is coming. How to match the talent training of the school with the talent demand of the society. This paper uses the crawler technology to obtain data from the recruitment website, through data collection, cleaning, and data preprocessing, and finally carries out data analysis and visual display.

Keywords: Big Data, Talent Structure, Visual Display, principal component analysis

1. INTRODUCTION

1.1 Background

New technology: Based on the new generation of information technology (Internet of things, cloud computing, big data, artificial intelligence, etc.).

New energy: use environmentally friendly electricity as energy to replace gasoline

New business type: the car is just a piece of hardware. The hardware itself is not worth money. What is worth is software. It is the data and intelligence generated by the car. Based on online campus recruitment, some scholars use Python and SPSS tools to analyze the employment status \cite{1}. If you can calculate, there will be an intelligent ecosystem based on auto drive system in the future. In fact, its logic is the same as that of Apple mobile phones, Don't simply think it's just hardware.

New industry: unlike traditional vehicles, intelligent electric vehicles will bring about changes in the entire industrial chain, involving complete vehicles and automobile cores. Film, automatic driving, Internet of vehicles, 5g+ Beidou, big data, data security and other industrial links will be like traditional Toyota and Honda in the future. It will not become the mainstream, but a new era of intelligent vehicles, In view of the coexistence of "difficult employment" of job seekers and "difficult recruitment" of enterprises in the current human resources market \cite{2}.

1.2 Target

Crawl recruitment websites nationwide for recruitment information of big data, data analysis, data mining, machine learning, artificial intelligence and other related posts. Analyze and compare the salary and education requirements of different positions, and visually present them. Analyze and compare the demand of different regions and industries for relevant talents, and visually present them. Analyze and compare the knowledge and skill requirements of different positions. Give relevant suggestions on big data talent training.

2. INFORMATION CRAWLING

The URL is the web address of the web page. The seed URL is the web address of the web page that the crawler should crawl first. Determine which web pages your crawler should crawl from first. A set of seed URLs refers to one or more web page addresses. After the crawler starts to work, the seed URL will first be added to the queue of the web page to be
crawled. The crawler obtains the web page URL from the queue according to the first in first out principle. The crawler starts to crawl the web page. The crawler will download the entire web page content, then extract the web page content, analyze the URL contained in the web page content, and add the new URL to the queue. It accurately obtains a large amount of talent recruitment information from major recruitment websites, and uses numpy, pandas, Matplotlib and word segmentation technology to visually display the characteristics of enterprises' demand for talents from multiple dimensions in the way of big data visualization[3].

Crawling posts: big data, data analysis, machine learning, artificial intelligence and other related posts

Crawling fields: company name, position name, work address, salary, publishing time, job description, company type, number of employees, and Industry

Tools used: python+requests+lxml+pandas+time

Parsing method: XPath

Directory page information XPath path

Position name: //div[@class= "El"]//p//span//@target= "\u blank"/text ()
Company name: //div[@class= "El"]/span[@class= "T2"]//@target= "\u blank"/text ()
Working address: //div[@class= "El"]/span[@class= "T3"]//text ()
Salary: //div[@class= "El"]/span[@class= "T4"]
Published on: //div[@class= "El"]/span[@class= "T5"]//text ()
Job link: //div[@class= "El"]/p//span//@target= "\u blank"//@href

Job description page information XPath path

Job description: //div[@class= "tcompany\u main"]//div[@class= "bmsg job\u MSG inbox"]//p//text()
Company type: //div[@class= "tcompany\u sidebar"]//div[@class= "com\u tag"]//p[1]//@title
Number of employees: //div[@class= "tcompany\u sidebar"]//div[@class= "com\u tag"]//p[2]//@title
Company industry: //div[@class= "tcompany\u sidebar"]//div[@class= "com\u tag"]//p[3]//@title

Some recruitment information has no salary information and needs to be processed separately

d = dom. XPath ('//div[@class= "El"]//span[@class= "T4"]') ,
salary_ = [i.text for i in d] « some salaries are blank, so they are written in the derivation record. If it is blank, it is recorded as none.
Store the obtained data by page.
d.to_csv('job_info.csv', mode='a+', encoding='gbk', header=None)

3. DATA PREPROCESSING

Data preprocessing includes: Duplicate value processing; Post name processing; Payroll data processing; Workplace handling; Company type processing; Industry data processing; Job description data processing; Company headcount processing. In fact, data preprocessing is strictly a link before data analysis. When we get data from the database or other channels, the next link is not to analyze the data immediately. Because there are always some data that are not complete or are not needed by our business, the unnecessary data is often referred to as "dirty data". At this time, we need to preprocess the data before data analysis, that is, do some processing on these "dirty data". The four steps of data preprocessing are data cleaning, data integration, data transformation and data reduction; Data preprocessing refers to the necessary processing such as review, screening and sorting before classifying or grouping the collected data; Data preprocessing, on the one hand, is to improve the quality of data, on the other hand, it is also to adapt to the software or methods of data analysis. The python crawler is used to capture the relevant recruitment information data in the website, and then the crawled recruitment information data is processed and stored in the database in time[4].

Data transformation is to transform into appropriate forms to meet the needs of software or analysis theory.
(1) Simple function transformation

Simple function transformation is used to transform data without normal distribution into data with normal distribution. Commonly used are square, square, logarithm, difference, etc. For example, in the time series, the logarithm or difference operation is often performed on the data to convert the non-stationary series into a stationary series.

Use the pandas database module to clean and analyze the recruitment data file csv crawled from the storage.

(2) Normalization

Normalization is to eliminate the influence of variable dimensions. For example, the difference between height and weight can be directly compared. Different units and value ranges make it impossible to compare directly. Min max normalization: also known as standard deviation normalization, it performs linear transformation on data to change its range to [0,1]. Zero mean normalization: also known as standard deviation normalization. After processing, the mean value of the data is equal to 0 and the standard deviation is 1. Decimal scaling normalization: move the decimal digits of attribute values and map attribute values to [-1,1].

Use WordCloud library module and matplotlib library module to realize data visualization. Word cloud visualization

The code is as follows:

```python
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import jieba
path_txt = 'data.txt'
f = open(path_txt,'r',encoding='gbk').read()
cut_text = " ".join(jieba.cut(f))
wordcloud = WordCloud(font_path="C:/Windows/Fonts/simfang.ttf",background_color="white",width=1000,height=880).generate(cut_text)
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

(3) Principal component analysis

When conducting empirical research on a thing, in order to reflect the characteristics of the thing more comprehensively and accurately, we often need to consider multiple indicators related to it, which are also called variables. In practical problems, on the one hand, consider as many indicators as possible to avoid missing important information, on the other hand, the complexity of the problem increases with the increase of indicators. Because these indicators reflect the same thing, it is inevitable to cause information overlap, which will interfere with the research and affect the results. Principal component analysis can deal with the above problems. In this way, more information can be obtained through fewer variables. The specific implementation steps of principal component analysis are as follows:

Assume that the study involves P indicators, respectively using $X_1, X_2, ..., X_P$. P means that the P-dimension random vector formed by these P indicators is $X = (X_1, X_2, ..., X_P)$. First, standardize the original data:

$$X_{ik} = \frac{(X_{ik} - x_k)}{s_k}, i = 1, 2, ..., n, k = 1, 2, ..., p$$

(1)

After standardization, X is linearly transformed to form a new comprehensive variable, which is represented by Y, that is, the new comprehensive variable can be linearly represented by the original variable, which satisfies equation (2):
After linear transformation, we call that $Y_i = u_i^*X_i$, is the $i$th principal component of $X$. Generally, only the first few principal components with the largest variance are selected, and the cumulative variance contribution rate usually reaches 70% or more than 80%, thus simplifying the problem.

4. RECRUITMENT DATA ANALYSIS AND VISUALIZATION

Use the pandas library module to clean up and analyze the data of stored and crawled recruitment data pieces of CSV incoming data:

1) Recruitment cycle data analysis:

By analyzing the average recruitment cycle of different positions and job levels, we can strive for more reasonable recruitment time for recruitment activities in the future.

2) Recruitment cost data analysis:

Analyze the input and output of each recruitment channel. Data analysis can be performed in a certain type of recruitment channel, or horizontal comparison can be performed in multiple recruitment channels. For example, you can analyze and compare the network, school recruitment, internal recommendation, job fairs and other channels, or analyze multiple recruitment websites that are used at the same time. You can see the ratio between the recruitment expenses invested by different recruitment websites and the output (the number of people on the job, the number of effective resumes, and even the number of people on the job). You can see which recruitment channel is more effective. Analyze the recruitment cost per person: provide basis for formulating recruitment budget and reducing recruitment cost.

3) Recruitment channel data analysis:

Recruitment channel distribution is the ratio of employees employed through different recruitment channels to the total employees employed. The distribution of recruitment channels directly affects the selection of recruitment channels and the expenditure of recruitment expenses. It can be analyzed from the dimensions of job level, position, time, cost and overall. The selection of recruitment channels determines the change of recruitment plan completion rate, recruitment cycle and recruitment cost data to a certain extent.

4) Analysis of recruitment conversion rate data:

That is, the ratio between the number of invitations received from the resume, the number of people interviewed, and the number of employees: These data are transformed step by step like a funnel, and finally suitable employees are generated. When the recruitment goal is not achieved, we can pay more attention to which link is insufficient. In addition, continuously reducing the proportion can improve the employment efficiency and reduce the recruitment cost.

In today's society, on the one hand, a large number of graduates often shout "the most difficult employment season". On the other hand, it is still difficult for enterprises to recruit suitable talents for heavy recruitment. The traditional professional training requirements make it difficult to cultivate new talents suitable for today's industrial upgrading. Therefore, the national "six excellences and one top-notch" and "new engineering, new medical, new agricultural and new liberal arts", and the talent training in Colleges and universities should "keep pace with the times".

The data of the recruitment website is obtained by using Python crawler technology and automatically saved locally. The file access type is SCV file.

Determine the tool used for data analysis. Because two workbooks are used to facilitate the association of two tables and the display of data charts, I decided to use power bi, Open power Bi and click to obtain data, which can be either excle or CSV. At the same time, you can also select a database. Depending on the storage type of your files, click load directly.
Preprocess the data and obtain the data form used for drawing through counting and other operations. Next, draw the histogram. The vertical column diagram is obtained as follows, and the horizontal column diagram can also be drawn.
Figure 3 Salary Distribution Based On Education Background

Figure 4 Age Average
Note 1:
(1) Beijing, Shanghai and Shenzhen rank among the top three in terms of demand and treatment. In particular, Beijing has a very large demand, high treatment and many opportunities.
(2) There is little demand for data analysts in Suzhou, but the relative salary is relatively high.
(3) There are many demands for data analysis posts in Nanjing, but the salary is low.

Note 2:
(1) The undergraduate course basically meets the requirements of data analysis
(2) Basically, 1-5 years of work experience is required, indicating that this profession has only been emerging for a few years.
(3) The salary in 1-3 years is about 2k, and that in 3-5 years is 20K. The overall increase is OK, indicating that the development space is very good.

Preprocess the collected recruitment data, use the word segmentation tool to segment the data, remove the stop words, and then extract the core words that belong to the characteristics of job requirements, and analyze the recruitment data from multiple dimensions[5].

5. CONCLUSION

According to the recruitment data of the data industry, it can be observed that the salary of big data jobs is generally higher than that of the previous IT industry. Many operation and maintenance personnel and developers have begun to turn to the big data industry. More are in Beijing, Shanghai, Shenzhen, Guangzhou and Hangzhou, especially in Beijing. Large and medium-sized companies and big data start-ups are strongly attracting help. Relatively speaking, jobs are easier to find. The big data industry does not require high academic qualifications. The undergraduate course basically meets the requirements. High level positions such as algorithm, precise recommendation and machine learning have a high degree. Big data positions in skills are basically developed using the Java language, while data mining is mostly using the Python language. The platform system is Linux, and the architecture is hadoop, history, storm, hbase, spark, kafka, zookeeper, etc. The use of big data helps us better analyze the current employment market and employment trend, and provides better guidance for employees and recruitment companies.
REFERENCES

[1] Hupeng, lichangyun, furongxin Demand analysis of College Students' employment market based on data mining

[2] Fu Mingliang, Ligang Research on matching between job seekers' needs and enterprises' needs based on big data
2021.08.016.

[3] Jinhua Crawling method and implementation of IT talent demand information based on big data [j] Information
and computer (theoretical Edition), 2021,33 (04): 238-241

[4] Liuxiaozhi Information crawling and data analysis of recruitment websites based on Python [j] Electronic testing,

[5] Wanghuiling Design and implementation of data visualization analysis platform for recruitment website [d] Qufu
Research on Remote sensing image classification based on RA-UNet

Qihang Zhao, Bin Zhou*, Ben Wang, Jin Lu, Luxiao Zhu
School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China
azqh@stu.hznu.edu.cn, bzhoubin@hznu.edu.cn, c20170056@hznu.edu.cn, d188125865@qq.com, e1654603110@qq.com

Abstract

With the development of satellite remote sensing technology, the quality and quantity of remote sensing images are constantly improved. Remote sensing feature classification is also playing an increasingly important role in urban planning, resource exploration and other fields. In the early stage of remote sensing feature classification, machine learning algorithms such as SVM and K-means are mainly used. Nowadays, with the expansion of deep learning, various kinds of research in the computer vision field emerge in an endless manner. Remote sensing images are also mostly classified by different neural networks. According to the characteristics and advantages of U-NET, channel attention mechanism, ResNet, large convolution kernel and structural reparameterization, this paper proposes a network structure called RA-UNET. This paper uses the remote sensing ground object classification dataset LoveDA to conduct experiments. The results show that the network classification effect of this paper is better, with mIoU reaching 59.4% and mPA reaching 72.6%. And use the network in this paper and the four mainstream neural networks of FCN, SegNet, PSPNet and UNet to conduct comparative experiments. The comparative experimental results show that the classification effect of the network in this paper is better than the above four mainstream neural networks.

Keywords-Remote sensing; ground objects classification; semantic segmentation

1. Introduction

Remote sensing images are images captured by satellites or aerial machines, which are used to record the electromagnetic waves of various objects. Remote sensing ground objects classification is the study of obtaining information such as the distribution of various ground objects types by analyzing remote sensing images. As a frontier research field of remote sensing information processing, remote sensing ground object classification plays an extremely important role in many fields. For instance, it is involved in urban evolution and planning, natural resource research, agricultural research, natural disaster detection, military reconnaissance, emergency planning after natural disasters, land cover and other fields [1].

Initially, the classification results of ground objects in remote sensing images mainly rely on researchers with professional knowledge to make manual judgments based on image characteristics and actual conditions. However, with the continuous change of satellite remote sensing technology, remote sensing images gradually show the characteristics of high resolution and massive data. At the same time, the requirements of various industries and fields for timeliness and accuracy of remote sensing ground object classification results are also constantly increasing. In order to adapt to the application needs of the big data era. Improving the update speed and classification accuracy of remote sensing ground object classification can reduce the input of human and material resources. The automation of remote sensing image classification has gradually become an important topic in the field of remote sensing [2].

In the early days, with the development of machine learning, traditional algorithms such as SVM[3], K-Means[4], decision tree and random forest[5-6] were used to classify remote sensing ground objects. Nowadays, deep learning has been continuously developed, and neural networks have been widely used in many fields such as computer vision. Pixel-by-pixel classification tasks such as ground object classification are also called semantic segmentation [7]. The classic neural networks include: CNN (Convolutional Neural Network) [8], FCN (fully convolutional network) [9], ResNet [10], U-Net [11], PSPNet (Pyramid Scene Parsing Network) [12], SegNet [13] et al. The U-Net is especially suitable for detail-focused semantic segmentation tasks due to its feature of fusing low-level and high-level semantic information. The ResNet solves the problems of gradient explosion and gradient disappearance caused by too deep convolutional neural network, which greatly increases the depth of neural network and improves the classification effect significantly.
Therefore, in recent years, many new neural networks have been researched based on this, such as ResUNet [14], RefiNet [15], ResUNet++ [16], DeepLabV3+ [17].

With the in-depth development of deep learning, Bahdanau proposed an attention mechanism in the field of natural language processing [18]. The attentional mechanism mimics the internal processes of biological observation behavior, with reference to the human visual mechanism. In cognitive science, humans selectively focus on parts of all information and ignore others. Similarly, when the neural network processes a large amount of information input, the attention mechanism is able to quickly focus on some of the key information and temporarily, ignoring the unimportant parts. Especially, the continuous research of attention mechanism has been widely implemented in computer vision. The types of attention mechanisms generated by different needs are also increasing. For example, the spatial attention mechanism (Spatial attention) pays more attention to the feature regions of the desired target in the image to improve the detection accuracy [19]. The channel attention mechanism (Se attention) allows the network to distinguish the most interesting channels by focusing on the weight of each channel in the input feature map [20]. CBAM (Convolutional Block Attention Module) combines the respective advantages of spatial attention mechanism and channel attention mechanism to improve the accuracy of network prediction. [21].

Based on above background, this paper proposes a multi-scale feature fusion U-shaped neural network (RA-UNet) for remote sensing ground object classification. The network adopts the U-Net framework globally and is divided into two parts: decoder and encoder. The encoder refers to the ResNet50 structure. In view of the characteristics of large objects and complex edges in remote sensing ground objects, a large convolution kernel (9*9) is used to improve the receptive field. The identity shortcut module is used to compensate for details neglected by large convolution kernels and prevent network degradation. At the same time, the structural re-parameterization method is adopted, and 3*3 small convolution kernels are added in parallel with the 9*9 convolutional layers in the encoder for training. And then these two results are combined. After training, the parameters of the small convolution kernel are also fused into the large convolution kernel, so that the network in this paper has a large receptive field and attention to details. A channel attention module is added to the backbone network to improve the training accuracy and classification effect. The channel attention module is also added to the skip connection between low and high layer to improve the accuracy of network training. In this paper, the LoveDA remote sensing feature classification data set recently proposed by Wuhan University is taken as the research object. The network structure is compared with the current mainstream neural network, and the mIoU and mPA indicators are used for evaluation. The experimental results show that the mIoU and mPA indices of the newly suggested network are better than the comparison networks in remote sensing ground object classification. The classification accuracy is also good, which proves the effectiveness of the proposed network for remote sensing ground object classification.

2. Fundamental

2.1. General Structure

The overall framework of the network in this paper adopts the U-shaped structure proposed by U-Net. The U-shaped structure connects the low-level information with the high-level information, which is conducive to the multi-scale feature fusion. The classification processing at the pixel level is much deeper and more sensitive to the details of the target image. The processing results are quite fine, and they are especially suitable for recent high-resolution remote sensing images with rich details and complex feature structures. This helps to improve the problems of surface feature missing detection, false detection and surface feature edge roughness in high-resolution remote sensing images.

The new general structure of the network is shown in Fig. 1(a). It consists of five stages, each of which is used for the encoder and decoder, and jump connections for adding “Se” attention mechanism. After the remote sensing image is input to the encoder, five stages of processing are performed. In each stage, the semantic information acquired by itself is fed into the “Se” attention block. The block filters the information in it. Sufficiently important image features are selected by the weight of each channel and less important feature information is suppressed. The decoder receives jump connection information and up-sampling information at each stage and transposes the convolutional amplification. Finally, the output of the feature classification result map is performed by Softmax function at the end of the network.
2.2. Detailed structure of each stage

In this paper, there are five stages in the encoder of the network. Each stage consists of a large convolutional layer of 9*9, a small convolutional layer of 3*3, a convolutional layer of 1*1, an identity shortcut module, BN, and ReLU. The components of each stage are all different. We take stage 4, which has the largest number of layers, as an example, as shown in Fig. 1(b). The stage 1 consists of a 7*7 convolutional layer, BN layer and ReLU layer in order. Stages 2-5 all contain the Bottleneck A module. The stage 2 has two Bottleneck B modules. The stage 3 has three Bottleneck B modules. And the stage 4 has five Bottleneck B modules. Finally, there are two Bottleneck B modules in the stage 5.

As shown in Fig 1(b), the stage 4 contains six modules, one Bottleneck A and five Bottleneck B modules, respectively. In order to improve the perceptual field size of feature extraction, it confirms to the characteristics of large proportion and blockiness of all types of features in remote sensing feature images. In both Bottleneck A and Bottleneck B, a large convolutional layer of 9*9 is set. At the same time, a small convolutional layer of 3*3 is set simultaneously to assist feature extraction. During training, the feature extraction results of the two convolution kernels are superimposed. After training, the parameters of the 3*3 small convolution kernel are merged into the parameters of the 9*9 large convolution kernel. That is, the receptive field is improved without ignoring the detailed features of the complex edges between some features, so it improves the classification accuracy as well.

The difference between Bottleneck A and Bottleneck B is that Bottleneck A is generally the first module in each stage. As the image undergoes convolution in either stage, the size changes. Therefore, in the identity shortcut module of Bottleneck A, convolution is performed based on size matching to meet the subsequent processing conditions.
Firstly, the image is processed by Stage3 and then input to Stage4. After entering Stage4, the Bottleneck A module is started first, and the image size is changed from Channel*Image size to (2*Channel)*(Image size/2)*(Image size/2). Then five Bottleneck B module operations are performed. Finally, it is input to Stage 5.

2.3. Se block

The Se Attention is a lightweight channel attention mechanism module. The loss is calculated during the training of the network model, and the channel weights of the feature map are learned iteratively. SE Attention assigns larger weights to effective feature map channels and smaller weights to less effective ones. Such operations allow image segmentation models to selectively learn important semantic information.

In our network, the decoder receives shallow image features from the encoder. While receiving useful image feature information, some useless image feature information will be received. Therefore, introducing an attention mechanism into skip connections can allow the network model to learn more important features, thereby enhancing the segmentation accuracy of the network model.

3. Dataset

The dataset used in this paper is LoveDA, a remote sensing feature classification dataset released by Wuhan University in 2021 [22]. The LoveDA dataset contains areas of Nanjing, Changzhou, and Wuhan with some urban areas and some suburban areas. There are 4, 191 images, and each image is a RGB true color satellite image. The image size is 1024*1024*3, and the image type is PNG format. A total of 8 data types (7 feature types) are included, and each image is labeled with the following 8 categories. That is, no data other, buildings, roads, water bodies, wasteland, forest grassland, and cultivated land. Here, no data means that some pictures of certain areas are not collected due to satellite photography and other reasons, so they are replaced by no data and the type is ignored during training. The pixel resolution reaches 0.3 m and the images are extremely clear. However, the 1024 size image is too large for the training effect of the neural network. Therefore, the images are randomly cut to the size of 256*256*3, and generates 67056 images after cutting. According to the ratio of 9:1, the training set and the validation set are divided, and the training set includes 60, 352 images and the validation set has 6704 images.

4. Experimental results

This experiment will use mIoU (Mean Intersection over Union) and mPA (Mean Pixel Accuracy) to compare the classification accuracy of each neural network. The compared mainstream neural networks are FCN, SegNet, PSPNet, and UNet. The comparison results are shown in Table 1 and Table 2. In this case, the validation set images are randomly selected, and the classification results of each neural network are visually compared in Fig 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>mIoU(%)</th>
<th>Other</th>
<th>Building</th>
<th>Road</th>
<th>Water</th>
<th>Barren</th>
<th>Forest</th>
<th>Agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN-32</td>
<td>50.73</td>
<td>49.68</td>
<td>50.15</td>
<td>35.60</td>
<td>65.09</td>
<td>33.04</td>
<td>58.98</td>
<td>62.55</td>
</tr>
<tr>
<td>SegNet</td>
<td>55.14</td>
<td>52.60</td>
<td>54.50</td>
<td>48.90</td>
<td>68.40</td>
<td>62.68</td>
<td>64.14</td>
<td>64.79</td>
</tr>
<tr>
<td>PSPNet</td>
<td>55.37</td>
<td>51.98</td>
<td>50.96</td>
<td>48.90</td>
<td>68.67</td>
<td>38.36</td>
<td>62.22</td>
<td>66.48</td>
</tr>
<tr>
<td>UNet</td>
<td>55.79</td>
<td>54.61</td>
<td>56.26</td>
<td>53.07</td>
<td>71.48</td>
<td>25.99</td>
<td>62.42</td>
<td>66.69</td>
</tr>
<tr>
<td>RA-UNet</td>
<td>59.40</td>
<td>56.01</td>
<td>57.51</td>
<td>57.55</td>
<td>73.79</td>
<td>37.25</td>
<td>64.05</td>
<td>69.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>mIoU(%)</th>
<th>Other</th>
<th>Building</th>
<th>Road</th>
<th>Water</th>
<th>Barren</th>
<th>Forest</th>
<th>Agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCN-32</td>
<td>65.16</td>
<td>71.94</td>
<td>69.13</td>
<td>38.98</td>
<td>74.75</td>
<td>48.44</td>
<td>71.78</td>
<td>80.12</td>
</tr>
<tr>
<td>SegNet</td>
<td>67.99</td>
<td>71.23</td>
<td>68.36</td>
<td>54.18</td>
<td>81.25</td>
<td>42.49</td>
<td>76.18</td>
<td>82.23</td>
</tr>
<tr>
<td>PSPNet</td>
<td>68.21</td>
<td>67.61</td>
<td>60.00</td>
<td>59.58</td>
<td>79.72</td>
<td>53.47</td>
<td>74.36</td>
<td>82.73</td>
</tr>
<tr>
<td>UNet</td>
<td>67.66</td>
<td>75.61</td>
<td>72.00</td>
<td>58.56</td>
<td>81.23</td>
<td>29.36</td>
<td>76.81</td>
<td>80.03</td>
</tr>
<tr>
<td>RA-UNet</td>
<td>72.60</td>
<td>74.33</td>
<td>68.62</td>
<td>70.65</td>
<td>81.20</td>
<td>51.74</td>
<td>80.94</td>
<td>80.74</td>
</tr>
</tbody>
</table>
As shown in Tables 1 and 2. The final results of this network are better than the four mainstream networks in terms of mIoU and mPA evaluation indexes, and are higher than 4% or higher. Fig. 2 shows the visualization results of six kinds of ground object classification. The original is the original manual labeling, and all others are the prediction results of neural network. The RA-UNet is the prediction result of this paper’s network. As shown in Fig. 2, in other categories, only a few ground object are marked manually. However, other networks have a large number of wrong labels, and the network error label in this paper is obviously the least. In the water and agriculture classes, the classification shape of our network is also most similar to the manual labeling. In the barren and forest classes, the other four networks have a large number of incorrect labels, and the shape prediction of our network is obviously better than the other four networks. Compared with the other four networks, the misclassification and omission of our network are significantly improved, and the classification of edges is clearer and closer to the manual labeling of the original images. It proves that our network has certain advantages in the remote sensing image classification.

5. Conclusion

According to the characteristics and advantages of U-Net, channel attention mechanism, ResNet, large convolution kernel and structural reparameterization, this paper proposes a neural network structure RA-UNet for remote sensing ground object classification. This network structure takes into account the large size and block shape of remote sensing image ground object. That is, it has a larger receptive field for ground object extraction and is used to classify large-scale objects. It also has the control of details for edge extraction between different complex ground objects. In experiments of the latest remote sensing ground object classification dataset LoveDA, the mIoU is 59.4%, and the mPA is 72.6%. In comparative experiments, the results of the network in this paper are better than the four mainstream neural networks FCN, SegNet, PSPNet, and UNet. It can be seen that the proposed network has certain advantages in the field of remote sensing land object classification.

However, there are still some limitations in the network. Because of the combination of large convolution kernel and small convolution kernel synchronous training, the number of parameters is increased, and the training cost and time are
relatively increased. It is necessary to optimize the text network to reduce the amount of parameters and the training costs. In the future, we will continue to conduct research and analysis on this issue.

Acknowledgments

This paper is funded by projects: Zhejiang Province (LS21F020003).

References

Analysis of Heterogeneous Data Model Based on Federated Learning

Yating Gao*, Xingjie Huang, Jinmeng Zhao, Jing Zhang, Xinyu Liu
State Grid Information and Telecommunication Branch, Beijing, 100053, China
*yating-gao@sgcc.com.cn

Abstract

The rapid development of edge network devices has led to the explosive growth of their data, and the difficulty of dealing with heterogeneous data in edge devices has further increased. To solve the problem of heterogeneous data fusion without interaction, this paper proposes a data heterogeneous model analysis based on federated learning. Preprocess the multi-source heterogeneous data to obtain the main features of the condensed data. Then, the multi-source heterogeneous data nodes are positioned to avoid multi-fusion results, and Spatio-temporal correlation degree of the multi-source heterogeneous data is calculated to improve the accuracy of fusion. Finally, a multi-source heterogeneous data fusion model is established based on federated learning to ensure the security of data fusion. Compared with the traditional model, the data fusion of the proposed model is more stable and the error is smaller. The effectiveness of the proposed model is verified by the stability and accuracy of the fusion of the heterogeneous data. The multi-source heterogeneous data fusion model studied in this paper can improve the quality of Internet of Things data and promote the development of edge devices in China.

Keywords-Federated learning, Multi-source heterogeneous data, Data fusion, Modeling.

1. Introduction

In order to deal with the challenge that there are data islands among groups and it is difficult to jointly build a group intelligence model, federated learning (FL) is proposed. The concept is proposed, which aims to make the original data of the group jointly construct the machine learning model without leaving the local premise. Federated learning can be generally divided into horizontal federated learning, vertical federated learning and federated transfer learning. Horizontal federated learning refers to the fact that participants have user data with nearly the same characteristics, but almost completely different samples. Vertical federated learning refers to the fact that participants have almost the same user groups, but almost completely different characteristics. Federated transfer learning refers to scenarios where both user groups and user characteristics are almost completely different.

With the rapid development of science and technology, China has entered the era of national information, and edge devices have been continuously developed. The diversity of edge devices generates massive multi-modal data, and more valuable new data can be obtained by fusing different modal data sources. Multi-source heterogeneous data refers to data with different data sources and different data structures. With the rapid development of information and communication technology, the growth rate of data is very fast, not only the scale is large, but also the structure is diversified. The data sources are rich, and the multi-source heterogeneous big data is composed. Multi-source heterogeneous big data has the following characteristics: Multi-source. The dispatching center constructs databases independently in different fields and receives the model parameters transmitted in the same field at the same time, so that the source of the parameters is diversified. Different specialties in the dispatching centers are modeled independently, and the selected coding rules are very different, showing isomerization and poor consistency. Different fields of dispatching center maintain their own database independently, and there is no collaborative processing in the process of updating, which will lead to the lack of consistency of the database to a large extent. Multi-source heterogeneous data is typical unstructured data, and there is a risk of data leakage in the process of fusion. Therefore, the fusion of multi-source heterogeneous data collected by edge devices is a hot research topic. Wang Gang et al. [1] applied the fusion of multi-source heterogeneous data to the power cabin of the pipe gallery, which effectively improved the level of operation and maintenance. Liu Chenxi et al. [2] applied the fusion of multisource heterogeneous data to reduce the prediction error in the prediction of traffic flow. Wang Yan et al. [3] used multi-source heterogeneous data fusion to realize lightning protection data analysis of substation, which has good generalization ability. This paper refers to the relevant literature, and in-depth study of heterogeneous data models, hoping to help give full play to the value of heterogeneous data.
2. Preprocessing multi-source heterogeneous data

In the construction of the multi-source heterogeneous data fusion model [4], because the original data has the characteristics of different types and repeated alarms, it cannot be quantified. Thus, to make the multi-source heterogeneous data fusion results more accurate, this paper must first preprocess the data. The method comprises the following steps: firstly, divide data into a numerical type and a non-numerical type; uniformly standardize all the non-numerical type data to change the non-numerical type data into numerical type data; and then, normalize all the numerical type data. At the same time, the interference data in the multi-source heterogeneous data is removed for feature selection, so that the dimension of the heterogeneous data is reduced [5], which is convenient for subsequent fusion. The preprocessing process of multi-source heterogeneous data is shown in Figure 1.

![Flow chart of multi-source heterogeneous data preprocessing](image)

Carry out statistical analysis on the collected multi-source heterogeneous data, and then convert the characters in the non-numerical data into numerical values to ensure that all the fields of the multi-source heterogeneous data are numerical values [6]. Since there is a certain error in measurement when non-numeric data is converted into numeric data, it is necessary to normalize all multi-source heterogeneous data to the range of [0,1] through the Min-Max normalization method:

$$g = \frac{A - A_{\min}}{A_{\max} - A_{\min}}$$

(1)

Where $A$ is the original data; $A_g$ is the normalized data; $A_{\min}$ is the minimum and $A_{\min} = \min \{ A \}$; $A_{\max}$ is the maximum and $A_{\max} = \max \{ A \}$. There are some noise data in the multi-source heterogeneous data after normalization, so it is necessary to select the best denoising method to remove the noise data. Although the multi-source heterogeneous data has been preprocessed, there are still some redundant or repetitive features in the data, which cannot highlight the key features of the data [7]. Thus, in order to reduce the complexity of the multi-source heterogeneous data, this paper uses the principal component analysis to screen the data features, and to obtain more concise data main features, so that the subsequent data fusion is simpler and easier.

3. Locating multi-source heterogeneous data nodes

Due to the diversity of the multi-source heterogeneous data [8], there will be a variety of results during fusion. To obtain a single and accurate fusion result, this paper locates the nodes in the multi-source heterogeneous data before data fusion. First, the node coordinates of the heterogeneous data anchor points need to be obtained. If the node coordinates of the
three anchor points of the multi-source heterogeneous data are $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, and $M_3(x_3, y_3)$, then the coordinate of the heterogeneous data point is calculated as follows:

$$
\begin{bmatrix}
  x_g \\
  y_g
\end{bmatrix} = \begin{bmatrix}
  2(x_1 - x_3), 2(y_1 - y_3) \\
  2(x_2 - x_3), 22(y_2 - y_3)
\end{bmatrix}^{-1}
$$

It can be seen from this formula that the three anchor nodes can form a triangle, so the formula for obtaining the hop counts of nodes and anchor nodes according to the triangle is:

$$
x_{t+1} = MZ_{t+1}M^T + J
$$

In the formula, $x_{t+1}$ represents the prior estimation data when the coordinate of the heterogeneous data point is $t+1$; $M$ represents the anchor point node; $Z_{t+1}$ represents the posterior estimation data matrix when the coordinate of the heterogeneous data point is $t$; $J$ represents the state estimation data of the coordinate node of the heterogeneous data point; and $T$ represents the time data. The coordinate data of all mobile nodes in the multi-source heterogeneous data can be obtained through the formula (2) and the formula (3). The distance data between the anchor node and the hop count is calculated through the MCL algorithm. The space position of the multi-source heterogeneous data is determined, and the position coordinate of the mobile node to be calculated is speculated according to all sample data in the space:

$$
J(d_t | d_{t-1}) = \begin{cases}
\frac{1}{\pi v^2} \cdot (d_t - d_{t-1}) \leq v_{\text{max}} \\
0, (d_t - d_{t-1}) > v_{\text{max}}
\end{cases}
$$

In the formula, $v$ represents the moving speed data of the mobile node and is within the range of $[0, v_{\text{max}}]$. $J(d_t | d_{t-1})$ represents the state estimation data of the multi-source heterogeneous data mobile node moving from time $t-1$ to time $t$. According to the communication radius data between hops, all nodes in the multi-source heterogeneous data can be screened, and the nodes that do not meet the requirements can be eliminated [9]. Then, all nodes that meet the requirements can be integrated to obtain the fusion mapping relationship to realize the single fusion of the multi-source heterogeneous data.

3.1. Calculate the Spatio-temporal correlation degree of the multi-source heterogeneous data

Before establishing the multi-source heterogeneous data fusion model, it is necessary to calculate the Spatio-temporal correlation of data [10], and realize the data fusion according to the correlation. If the Spatio-temporal correlation between multi-source heterogeneous data is not strong, the effect of data fusion will not be ideal. Thus, this paper calculates the Spatio-temporal correlation of data through dynamic time planning distance, and establishes a multi-source heterogeneous data fusion model with reference to the strength of data correlation. When the DTW algorithm is used to calculate the sequence similarity of the multi-source heterogeneous data, it needs to calculate the sequence of all data for complete matching, which greatly increases the amount of calculation and prolongs the calculation time [11]. Therefore, in order to reduce the resource consumption in the calculation and improve the calculation efficiency, this paper introduces a sliding window to analyze the data sequence in the DTW algorithm. In the sliding window time, the distance between the multi-source heterogeneous data is calculated as follows:

$$
d_{u,v} = \begin{cases}
s(E_u(t), F_v(t)), u \neq v \\
0, u = v
\end{cases}
$$
In the formula, \( u \) and \( v \) respectively represent the multi-source heterogeneous data of the terminal; \( E_u(t) \) and \( F_v(t) \) respectively represent the time series of \( u \) and \( v \); \( s(E_u(t), F_v(t)) \) represents the dynamic time warping distance data of \( u \) and \( v \). When dividing the multi-source heterogeneous data, it is necessary to refer to the correlation between time and space to remove the multi-source heterogeneous data with poor correlation. In the multi-source heterogeneous data fusion model established in this paper, it is necessary to improve the fusion accuracy according to the distance between the multi-source heterogeneous data of the terminal. When the distance between the heterogeneous data is 0, it is impossible to realize the data weight allocation. At this time, the spatial correlation degree between the multi-resource heterogeneous data can be quantified through the exponential function:

\[
f(d_{u,v}) = e^{-\frac{d_{u,v}^2}{2}}
\]

The distance data between the processed heterogeneous data can better describe the Spatio-temporal correlation between the multi-source heterogeneous data to improve the accuracy of the subsequent fusion calculation.

4. The multi-source heterogeneous data fusion model based on federated learning

The basic paradigm of federated learning can be stated as follows:

When the multiple data owners \( i \) (1), \( Fi N = "want to federate their data sets Di jointly while training the machine learning model M, the traditional scheme usually integrates the data first to obtain the total data set \( 1 N D = D D \cup " \cup " \), and then trains on D Practice to get the model Msum. This practice has the hidden danger of privacy disclosure, and may even violate relevant laws and regulations. The federal learning is in data ownership. Under the condition that the party Fi does not need to directly provide the data Di, the federated model Mfed is trained only by transmitting the intermediate result after the security processing.

Assuming that there is \( n \) kind of the multi-source heterogeneous data [12] in the fusion model to participate in the training, then in the initialization phase of the training, the multi-source heterogeneous data fusion model is:

\[
O = \langle Q \cdot H \cdot C \rangle
\]

In the formula, \( Q \) represents the feature extraction of the multi-source heterogeneous data; \( H \) represents the feature fusion of the multi-source heterogeneous data; \( C \) represents the feature decision of the multi-source heterogeneous data; \( \langle \cdot \rangle \) represents the splicing operation. In the feature fusion of the multi-source heterogeneous data, a high-order tensor data can be established, and the spatial dimension characteristics of the multi-source heterogeneous data can be displayed through the training of the data. In the feature decision of the multi-source heterogeneous data, the feature training results of multi-source heterogeneous data after fusion are used to deeply mine the potential relationship between data to enhance the feature expression power of multi-resource heterogeneous data. In the training stage of the fusion model, the adaptive problem in the multi-source heterogeneous data fusion model is solved by setting the feature graph of the feature extraction sub-network and the constraint tensor value. In the aggregation phase of the fusion model, the training features of each data node are merged. An average aggregation algorithm and a federated learning algorithm are used to obtain a multi-source heterogeneous data fusion model with global features:

\[
O' = \langle Q' \cdot H' \cdot C' \rangle = \left\langle Q' \cdot \sum_{m=1}^{b} h_{m} \times H' \cdot \sum_{m=1}^{b} h_{m} \times C' \right\rangle
\]

In the formula, \( Q' \), \( H' \) and \( C' \) respectively represent the updated multi-source heterogeneous data feature extraction, fusion and decision modules; \( m \) represents the number of edge nodes; and \( b \) represents the multi-source heterogeneous data samples participating in the fusion model training. As the federated learning [13] is a distributed learning framework, it enables the multi-source heterogeneous data to be trained in different edge nodes, so the federated learning is applied to the fusion model to improve communication efficiency and ensure data security.
5. Experimental verification

5.1. Preparation for the experiment

The multi-source heterogeneous data fusion model established in this paper is mainly to ensure the quality of the heterogeneous data after fusion, so through the training experiments of the fusion model on different heterogeneous data sets, the performance of the fusion model established in this paper is judged according to the accuracy and error of the data after training. This test environment simulates the real network environment, using three hosts with 4GB memory and dual-core processors to build the experimental environment, collecting three groups of representative multi-source heterogeneous data as training data sets, and inputting them into the proposed model and other traditional models for fusion training.

5.2. Experimental results

The two fusion models are trained for 100 s, and the estimation error covariance data is obtained every 10 s. The error change curve is drawn by the MATLAB as follows:

![Error covariance curve of estimation error of the model in this paper](image1)

**Fig.2 Covariance curve of estimation error of the model in this paper**

![Estimation error covariance curve of traditional model](image2)

**Fig.3 Estimation error covariance curve of traditional model**
It can be seen from Figure 2 that the data after the fusion of the model in this paper is relatively stable, and the error fluctuation of each group of training data is small and within the range of 10 ~ 20; it can be seen from Figure 3 that the data stability after the fusion of the traditional model is insufficient, the error fluctuation of training data set 1 is large, and the error of three groups of training data sets is within the range of 5 ~ 25. Therefore, compared with the traditional model, the stability and accuracy of the model established in this paper have been improved, which verifies that the heterogeneous data fusion model established in this paper based on federated learning is effective.

6. Conclusion

To sum up, data quality issues need to be considered when building a heterogeneous data fusion model, so this paper introduces federated learning to make full use of the characteristics of heterogeneous data to achieve accurate data fusion on the basis of considering the relevance of multi-source heterogeneous data and the complexity of data fusion. At present, there are some deficiencies in this study, which does not take into account the limited heterogeneous data resources of the Internet of Things terminals.

In the future, we still need to continue to improve and build a better heterogeneous data fusion model. Future federated learning research is full of challenges and opportunities. Overcoming these difficulties will help to establish a new Internet-based swarm intelligence theory and method system, thus promoting the implementation of the artificial intelligence dividends and injecting new impetus into the prosperity and development of modern society.

References

Stock Market Trend Prediction Using CBAM and CNN

Yong Wang\textsuperscript{1a*}, Zhiyu Xu\textsuperscript{1b*}, Yisheng Li\textsuperscript{1c}

\textsuperscript{1}School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401135, China
\textsuperscript{a*}ywang@cqut.edu.cn, \textsuperscript{b*Xyzaer_Xu@2020.cqut.edu.cn, \textsuperscript{c}yamisora@2020.cqut.edu.cn

Abstract

In recent years, deep learning has been increasingly used to analyze financial data. For deep learning to predict the Buy, Sell, and Hold points of stocks are prone to over-fitting, unreasonable feature extraction, and other issues. This paper builds a CBAM-CNN model based on Convolutional Neural Network (CNN) and Convolutional Block Attention Module (CBAM) to predict the Buy, Sell and Hold points. In order to verify the applicability and superiority of the proposed method, the shares of Dao 30 and SHH 50 from stock listing to August 11, 2021 are selected, and the accuracy of the deep learning algorithm is evaluated using confusion matrix, weighted F1 score, and Kappa coefficient. The analysis results show that this algorithm has a high classification prediction accuracy because it can identify most of the buy and sell instances and therefore has a better effect. In addition, compared with CNN that do not use the CBAM attention mechanism, classification performance is significantly improved. The results from this analysis can help investors determine their better investment strategies.

Keywords- CNN; CBAM; CBAM-CNN; stock market; deep learning

1. Introduction

Stock market forecasting is one of the most essential topics in the financial markets, and many researchers have conducted extensive research in this area. Due to its inherent characteristics of large data volumes, non-linearity, and non-smoothness, and thus the effectiveness of using only traditional statistical methods for stock market forecasting is insufficient\textsuperscript{1}(2). As model prediction based on deep learning becomes the best performance in various fields of application, it outperforms classical computational intelligence methods such as BP. And computer vision-based image processing problems exemplify the types of applications where these deep learning models exceed other techniques\textsuperscript{3}.

In recent years, many deep learning techniques, such as recurrent neural networks (RNN) \textsuperscript{4}, convolutional neural networks (CNN) \textsuperscript{5} and long short-term memory (LSTM) \textsuperscript{6} have been applied in stock market research to improve prediction accuracy. Chen proposed a deep learning method based on convolutional neural networks to extract features and demonstrated the reliability of using a deep learning method based on convolutional neural networks to forecast stock price movements \textsuperscript{7}. Kim T proposed a model called feature fusion LSTM-CNN and proved that the feature fusion LSTM-CNN model outperformed a single model in predicting stock prices. The model consists of an LSTM and a CNN that combines features learned from different representations of the same data for prediction, which can effectively reduce the prediction error\textsuperscript{8}. Zhou proposes using deep convolutional neural networks (DCNN) for stock price trend prediction based on stock images, focusing on how to generate stock images and label the images to train the DCNN model. It has been proved that the Convolutional Block Attention Module (CBAM) can improve the performance of DCNN \textsuperscript{9}.

To address the problems of overfitting and unreasonable feature extraction in predicting stock buy, sell and hold points by deep learning, this paper constructs a model based on a combination of CNN and CBAM to make predictions and achieve the purpose of finding more buy and sell points and hold points. Firstly, a one-dimensional time series is transformed into a high-dimensional image dataset as the input of a deep convolutional neural network; secondly, firstly, a BN layer is introduced between the convolutional and pooling layers to reduce overfitting and accelerate convergence; secondly, a CBAM module is submitted after the BN layer to extract more important features; finally, SHH 50 and WMT stocks are used as examples to verify the generalization performance of the CBAM-CNN model.
2. Methodology

2.1. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning model, similar to the multi-layer perceptron of artificial neural networks, characterized by local connectivity and weight sharing. CNN has been widely used in computer vision, speech recognition, natural language processing, and other fields as a result (10).

The convolutional layer of a convolutional neural network is mainly used for feature extraction by convolving the feature vectors of the previous layer using an activation function with a convolutional kernel. Assume that \( Y_i \) is the output of the neuron and \( X_j \) is the input vector. The correspondence between the input vector and the output vector can be expressed as:

\[
Y_i = G \left( \sum_{j=1}^{n} X_j \ast W_{ij} + B_j \right)
\]  

(1)

In Formula (1), \( n \) is the total number of vectors input to neuron \( j \), \( \ast \) is denoted as the convolution, \( W_{ij} \) is the weight parameter of the input vector \( X_j \) connected to neuron \( j \), \( B_j \) is the bias parameter and \( G() \) is the activation function.

A typical CNN network is mainly used to process image data with two-dimensional features of length and width as input samples (11). In this paper, 225 high quality 1D features selected using the feature selection technique are transformed into 2D images as input for time series data such as opening and closing prices of stocks, etc. The parameter sharing mechanism of CNN with fixed weights of neurons within the convolutional kernel effectively reduces the complexity of the network and improves the training efficiency of the model.

2.2. Convolutional Block Attention Module (CBAM)

Accurate extraction of useful feature factor information is a prerequisite for improving the accuracy of predicting buy, sell and hold points. The CNN convolutional layer and other layers consider each channel processing the feature map to have the same weight. Still, since the importance of each feature factor is different, it is unreasonable to assign the same weight to each channel. The Convolutional Block Attention Module CBAM is designed to improve the flow of information within the network by understanding what information to emphasize or suppress (12), finding the more important feature factors, and focusing all attention on these critical regions to improve classification accuracy (13)(14).

The CBAM structure is shown in Figure 1 and contains two separate sub-modules, the channel attention module (CAM)(15) and the spatial attention module (SAM)(16). It compares with SENet (17), which uses global average pooling (GVP) to migrate spatial information into the channel process, and spatial information is changed from multiple to one so that information loss occurs (18). The CBAM uses both global average pooling (GVP) and global max pooling (GMP), as well as the spatial attention SAM module, to compensate for the lack of spatial information in the migration process, which helps to improve the prediction accuracy.

![CBAM structure diagram](image)

Figure 2 shows the CAM structure of the channel attention module. The input feature graph \( F \) undergoes both global maximum pooled GMP and global average pooled GAP operations in the spatial dimension. Then two channels' attention feature vectors are joined together by element addition through two layers of shared multi-layer perception (MLP)(19), and then the channel domain attention feature map is obtained by sigmoid activation function, which is recorded as \( F' \).
Figure 3 shows the SAM structure of the spatial attention module. The channel domain Attention Feature Map (F') above is combined with the global maximum pooling GMP and the global average pooling GAP along the channel dimension(20). Finally, two different channel feature descriptors are obtained. These two feature vectors are stitched together, and convolution with a convolution kernel of 7*7 is passed through the sigmoid activation function. Finally, the spatial domain Attention Feature Map is obtained and recorded as F''.

2.3. Batch Normalisation Layer (BN)

An essential disadvantage of the deep learning algorithm is the difficulty of network training(21). The gradient descent method will converge to the optimal local value, the sigmoid activation function will disappear from the gradient, and as the number of network layers increases, the hyperparameters in training will have a significant impact on the training process and results(22). For this reason, the batch normalization algorithm is added to the network designed in this paper to batch normalize the image input from the network.

The BN algorithm is formulated as follows:

$$\hat{X}^i = \frac{X^i - E[X^i]}{\sqrt{Var(X^i)}}$$

(2)

In Formula (2), $\hat{X}^i$ denotes the input of a neuron in a layer, $\hat{X}^i = Wh + b$, where h denotes the output of the previous layer and W is the weight of this layer; $E[X^i]$ denotes the mean of all the input data of a batch of this neuron in stochastic gradient descent, and $\sqrt{Var(X^i)}$ denotes the standard deviation of all the inputs of a batch of this neuron. Using the above procedure, the input to a layer of neurons can be normalized to a standard normal distribution with mean 0 and variance 1.

2.4. CBAM-CNN

In this paper, CBAM-CNN is a practical improvement to CNN networks. CBAM-CNN consists of an input layer, a convolutional layer, a BN layer, a CBAM, a pooling layer, and a fully connected layer, as shown in Figure 4. The BN layer is added between the convolutional layer and the pooling layer, and the CBAM(14) module is added after the BN layer. The convolutional layer extracts the different features of the input, the BN layer increases the training speed and reduces the risk of overfitting, the CBAM module extracts the crucial features, the pooling layer reduces the number of model parameters to optimize the workload, the Dropout increases the sparsity of the network and alleviates overfitting. The fully
connected layer connects all the features and feeds the output values into the classifier for classification. The fully connected layer combines all the features and provides the output values to the classifier for classification.

![CBAM-CNN structure diagram](image)

Fig.4 CBAM-CNN structure diagram

3. Experiments and Analysis

3.1. Dataset Preprocessing

Our raw numerical data was obtained by calling the Alpha Vantage API for training and testing purposes. We use a sliding window tagging method, where all daily closing lows are labelled as Buy and highs as Sell by identifying the top and bottom points in the sliding window, and the remaining points are labelled as Hold points. The time period is from the Stock listing to 11 August 2021. The raw data consisted of six fields: timestamp, open, high, low, close, adjusted-close, volume. There are different evaluation indexes for stock data, their dimensions or units are different, and they are in different orders of magnitude. In order to resolve the comparability between the characteristic indexes, after normalization, each index is in the same order of magnitude so that it is easy to compare them comprehensively. Therefore, we used MinMaxScaler to normalize the data in the range of [0,1], as shown in the Formula (3).

\[
X_{\text{scaled}} = \frac{X - X_{\text{min}}(\text{axis}=0)}{X_{\text{max}}(\text{axis}=0) - X_{\text{min}}(\text{axis}=0)} \cdot (\text{max} - \text{min}) + \text{min}
\]  

(3)

Max, min are the maximum and minimum values of a given zoom range.

To improve the training efficiency, this paper refers to the traditional technical surface analysis means and calculates many feature factors as the original input of the model. The Technical indicator Factors are constructed in the manner shown in Table 1. The OverBoughtOverSold Type abbreviation is OBOS Type.

<table>
<thead>
<tr>
<th>Primary indicators</th>
<th>Secondary indicators</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBOS Type</td>
<td>Relative Strength Index (RSI)</td>
<td>trade signals</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Williams %R (W%R)</td>
<td>trade signals</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Money Flow Index (MFI)</td>
<td>short-swim trade signals</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Receiver Operating Characteristic (ROC)</td>
<td>trade definition signals</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Chaikin Money Flow (CMF)</td>
<td>trade signals</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Chaikin Money Flow (CMF)</td>
<td>extremely OBOS conditions</td>
</tr>
<tr>
<td>Moving Average Type</td>
<td>Simple Moving Average (SMA)</td>
<td>N-day moving average</td>
</tr>
<tr>
<td>Moving Average Type</td>
<td>Exponential Moving Average (EMA)</td>
<td>future trend of price</td>
</tr>
<tr>
<td>Moving Average Type</td>
<td>Weighted Moving Average (WMA)</td>
<td>calculate weighted average</td>
</tr>
<tr>
<td>Moving Average Type</td>
<td>Hull’s Moving Average (HMA)</td>
<td>moving average price</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Triple Exponentially Smoothed Average (TRIX)</td>
<td>long-swin trade signals</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Commodity Channel Index (CCI)</td>
<td>bull/bear market trade signals</td>
</tr>
<tr>
<td>Trend Type</td>
<td>Detrended price Oscillator (DPO)</td>
<td>bull/bear market trade signals</td>
</tr>
<tr>
<td>Trend Type</td>
<td>Know Sure Thing (KST)</td>
<td>trade momentum oscillator</td>
</tr>
<tr>
<td>Trend Type</td>
<td>Directional Movement Index (DMI)</td>
<td>bull/bear market trade signals</td>
</tr>
<tr>
<td>Moving Average Type</td>
<td>%Bollinger Band (%BB)</td>
<td>bull/bear market trade signals</td>
</tr>
<tr>
<td>Trade Volume Type</td>
<td>Force Index (FI)</td>
<td>Strength of upward or downward</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Raw Stochastic Value (RSV)</td>
<td>real fluctuation range of a market</td>
</tr>
<tr>
<td>OBOS Type</td>
<td>Stochastic Indicator (KDJ)</td>
<td>trade signals</td>
</tr>
<tr>
<td>Trade Volume Type</td>
<td>Ease Of Movement (EOM)</td>
<td>measure change degree</td>
</tr>
</tbody>
</table>

After calculating these indicators, mapping them to image sets according to their type (OBOS Type, Moving Average Type, Trend Type, etc.), the model does not learn enough at this point, perhaps the features are not good enough. We used a large number of indicators rather than strictly following the rules of calculation for different periods. We then used two feature selection methods, f-classif and mutual-info-lassif, and selected common features from the two results. The feature selection technique resulted in the selection of 225 high quality features.

### 3.2. Training Process

The CBAM-CNN model constructed in this paper is an effective improvement of the CNN network model, and the training process is as follows. The first step is to obtain the stock data dataset through an API to perform pre-processing operations on it, so that the low-dimensional numerical dataset is transformed into a higher-dimensional image dataset. The image dataset is fed into the CBAM-CNN we constructed to train the model architecture and analyse the model learning. The hyperparameters involved in the model mainly include the number and size of convolutional kernels in the convolutional layer, the proportion of neurons retained in the Dropout layer, standard deep learning training parameters such as learning rate, batch size, weight initialization method, etc. The values of each hyperparameter will be adjusted and optimized in the training set. By constantly training and tuning the parameters, the final trained CBAM-CNN model is obtained. The training process of the model is shown in Figure 5.

![Fig.5 Flowchart of CBAM-CNN](image_url)
3.3. Parameter and Evaluation Indicators

In this paper, the optimal hyperparameters of the model are obtained by cross-validating the training set with a sliding window training method, a convolutional kernel size of 3\times3, a padding of same, and a stride of 1. The retention ratio of the Dropout layer is 0.5, the learning rate is set to 0.001, the batch size is 60, and the ADAM optimizer is used for training.

To measure the prediction effectiveness of the model, the most straightforward and most crude measure for the classification problem in machine learning is accuracy (accuracy). Still, for data with an extremely uneven distribution of sample labels, accuracy can be misleading. In this paper, confusion matrix, weighted F1 scores, and Kappa coefficients are used as classification accuracy evaluation metrics, which are calculated as shown in Formulas (4) to (6).

\[
F1 \text{ Score} = \frac{2PR}{P + R} \tag{4}
\]

In Formula(4), The weighted F1 Score indicator combines the results of the output of Precision and Recall.

\[
k = \frac{p_o - p_e}{1 - p_e} \tag{5}
\]

In Formula(5), \( p_o \) is the sum of the number of correctly classified samples in each category divided by the total number of samples, the overall classification accuracy.

Assume that the true number of samples in each category is \( a_1,a_2,...,a_C \), and the predicted number of samples in each category is \( b_1,b_2,...,b_C \). The total number of samples is \( n \). Then the \( p_e \) is calculated as in Formula(6):

\[
p_e = \frac{a_1b_1 + a_2b_2 + \cdots + aCb_C}{n \times n} \tag{6}
\]

3.4. Ablation experimental

The ablation experimental results under WMT dataset are shown in Figure 6. The comparison results of CNN, CNN+CBAM and CNN+BN+CBAM in ACC, precision, F1 score and kappa coefficient are verified respectively. It can be seen from the figure that the increase of BN layer and CBAM module improves the personality ability index to varying degrees, which verifies the effectiveness of CBAM-CNN model in this paper.

![Fig.6 Performance comparison results of Ablation Experiment](image-url)
stage in Figure 7 and Figure 8. Val-F1 becomes larger and larger and finally tends to stabilize, Val-Loss is getting smaller and smaller, finally tends to be stable. CBAM-CNN adopts a more efficient network feature extraction module with fast network convergence and high accuracy. It has particular practical significance for the research of stock trading strategy in the future.

Fig. 7 Model loss diagram of Dao 30 Experiment

![Model loss diagram of Dao 30 Experiment](image1)

Fig. 8 Performance comparison results of Ablation Experiment

![Performance comparison results of Ablation Experiment](image2)

4. Results

Using the established CBAM-CNN network model, in order to verify the applicability and superiority of the proposed method, Dow 30 and SSH 50 are selected to predict the trading point and hold point. The prediction performance of the model is analyzed. Although the performance of each stock is considered separately, due to space constraints, the summary results will be displayed. It is worth noting that in the actual prediction process of the whole dataset, the iterative prediction is carried out in the form of window movement. The confusion matrix, weighted F1 score, kappa coefficient, and other indicators of the statistical model are shown in Table 2 to Table 5.
Table 2 Confusion Matrix of Dow 30 Test Data

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Buy</td>
</tr>
<tr>
<td>Predicted</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>34</td>
</tr>
</tbody>
</table>

Table 3 Evaluation of Dow 30 Test Data

<table>
<thead>
<tr>
<th></th>
<th>Buy</th>
<th>Sell</th>
<th>Hold</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Score</td>
<td>0.90</td>
<td>0.74</td>
<td>0.90</td>
</tr>
<tr>
<td>Precision</td>
<td>0.64</td>
<td>0.76</td>
<td>0.93</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc</td>
<td>87.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 lists the confusion matrix of Dao 30 test data. Table 3 illustrates the performance evaluation of the results obtained by the Dow 30 confusion matrix. The accuracy of buy and sell points is worse than hold points. For a stock trading system, accurate prediction of buy and sell is very important for the overall success of the trading algorithm. In our experiment, most "buy and sell" points were correctly captured by the CBAM-CNN model. Because the frequency of buy and sell points is much lower than that of hold points, this leads to many false buy and sell points. Because when the hold point is close to the buy and sell point, the deep learning neural network is likely to be unable to distinguish the buy and sell points from the hold points. Therefore, it is challenging to find a small number of buy and sell points with the deep learning algorithm. Kappa coefficients are basically within the range of the model, which proves the rationality of this experimental dataset. Table 4 provides the confusion matrix of SHH 50 test data, and Table 5 illustrates the evaluation of the confusion matrix of SHH 50 test data. Comparing the experimental results of SHH 50 and Dao 30 shows that this CBAM-CNN model can well solve the problem of determining the optimal network scale so as to reduce the network scale and improve the generalization performance of the network.

Table 4 Confusion Matrix of SHH 50 Test Data

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Buy</td>
</tr>
<tr>
<td>Predicted</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
</tbody>
</table>

Table 5 Evaluation of SHH 50 Test Data

<table>
<thead>
<tr>
<th></th>
<th>Buy</th>
<th>Sell</th>
<th>Hold</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Score</td>
<td>0.91</td>
<td>0.75</td>
<td>0.90</td>
</tr>
<tr>
<td>Precision</td>
<td>0.64</td>
<td>0.76</td>
<td>0.91</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acc</td>
<td>87.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition, this paper compares the accuracy and model size of CNN and CBAM-CNN in the test set. The results show that after adding the CBAM attention mechanism, although the amount of model parameters increases, the calculation speed decreases by nearly 1%, and the accuracy increases by 2.5%, which proves the effectiveness of CBAM-CNN in the Dow 30 and SSH 50 datasets.

5. Conclusion

Aiming at the prediction of stock buy, sell and hold point, combined with the respective characteristics of the revolutionary block attention module and convolutional neural network, this paper establishes the CBAM-CNN model, which is verified by dao30 and SHH50 stock indexes. To solve the problem of insufficient feature extraction, this paper adds a CBAM
attention mechanism to the network, which not only enhances the weight of the important feature layer in the previous training network but also suppresses the influence of the invalid feature layer. Compared with CNN, the experimental results show that CBAM-CNN has higher accuracy than CNN, and the calculation speed is not significantly reduced. The test results show that the model has high accuracy in F1 score, accuracy, confusion matrix, and other evaluation indexes and has good timeliness and stability. The relevant results verify the modeling and prediction ability of deep learning in financial time series and provide a reference for investors and quantitative researchers.

6. Abbreviations

Table 6 Technical indicator Factors

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBAM</td>
<td>Convolutional Block Attention Module</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>RNN</td>
<td>recurrent neural networks</td>
</tr>
<tr>
<td>LSTM</td>
<td>long short-term memory</td>
</tr>
<tr>
<td>DCNN</td>
<td>deep convolutional neural networks</td>
</tr>
<tr>
<td>CAM</td>
<td>channel attention module</td>
</tr>
<tr>
<td>GVP</td>
<td>global average pooling</td>
</tr>
<tr>
<td>GMP</td>
<td>global max pooling</td>
</tr>
<tr>
<td>SAM</td>
<td>spatial attention module</td>
</tr>
<tr>
<td>BN</td>
<td>Batch Normalisation</td>
</tr>
<tr>
<td>OBOS</td>
<td>OverBoughtOverSold</td>
</tr>
<tr>
<td>RSI</td>
<td>Relative Strength Index</td>
</tr>
<tr>
<td>W%R</td>
<td>Williams %R</td>
</tr>
<tr>
<td>MFI</td>
<td>Money Flow Index</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>CMF</td>
<td>Chaikin Money Flow</td>
</tr>
<tr>
<td>CMO</td>
<td>Chande Momentum Oscillator</td>
</tr>
<tr>
<td>SMA</td>
<td>Simple Moving Average</td>
</tr>
<tr>
<td>EMA</td>
<td>Exponential Moving Average</td>
</tr>
<tr>
<td>WMA</td>
<td>Weighted Moving Average</td>
</tr>
<tr>
<td>HMA</td>
<td>Hull's Moving Average</td>
</tr>
<tr>
<td>TRIX</td>
<td>Triple Exponentially Smoothed Average</td>
</tr>
<tr>
<td>CCI</td>
<td>Commodity Channel Index</td>
</tr>
<tr>
<td>DPO</td>
<td>Detrended price Oscillator</td>
</tr>
<tr>
<td>KST</td>
<td>Know Sure Thing</td>
</tr>
<tr>
<td>DMI</td>
<td>Directional Movement Index</td>
</tr>
<tr>
<td>%BB</td>
<td>%Bollinger Band</td>
</tr>
<tr>
<td>FI</td>
<td>Force Index</td>
</tr>
<tr>
<td>RSV</td>
<td>Raw Stochastic Value</td>
</tr>
<tr>
<td>KDJ</td>
<td>Stochastic Indicator</td>
</tr>
<tr>
<td>EOM</td>
<td>Ease Of Movement</td>
</tr>
</tbody>
</table>

Acknowledgments

This work was financially supported by the Science and Technology Research Project of Chongqing Municipal Education Commission (KJQN202001118)
References


Research on AOD Furnace End-point Temperature Prediction Based on Sparrow Search algorithm optimization ELM

Lei Yang1a*, Lin Chai2b*, Huikang Liu2c, Zhen Yi1d
1School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081
2Engineering Research Center of Metallurgical Automation & Testing Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
a*1321243661@qq.com, b*chailin@wust.edu.cn, cliuhuikang@wust.edu.cn, d486842759@qq.com

Abstract

In order to make the endpoint temperature prediction hit rate and velocity meet the requirements, in this paper, the extreme learning machine(ELM) is used to build a model which predicts the endpoint temperature of the AOD refining furnace, and the model is trained with the real production data to search the nonlinear variation law between the input quantity and the endpoint temperature. The sparrow search algorithm(SSA) is used to adjust the parameters of model to improve the hit rate of ELM. By comparing with other prediction models, found that the model is better than the pre-optimized temperature model in prediction hit rate and better than support vector regression(SVR) in prediction velocity, which is more suitable for guiding actual production.

Key words: AOD refining furnace, Endpoint temperature, ELM, SSA, Temperature prediction model

1. Introduction

Steel is an essential metal material in the building, machinery and Daily life, while stainless steel plays an significant role in steel materials because of its remarkable performance of corrosion resistance and recycling. AOD(argon oxygen decarburization) method decreases the CO partial pressure in furnace by adjusting the proportion of argon oxygen mixture pumped into furnace many times, thus reducing the content of carbon in molten steel. It is mainly used for making stainless steel. Accurate control of terminal parameters is the main goal of refining. There is nonlinear and coupling connection between endpoint temperature and many variables, and some value cannot be measured in real time. Therefore, building an accurate endpoint temperature prediction model is of great significance for improving the hit rate, velocity and production efficiency[1].

S-Asai etc.[2-3]conducted mathematical simulation on the stainless steel refining procedure, assuming that the heat accumulated in the molten pool during the refining procedure is equal to the heat produced by the system, and they built a heat equilibrium computation model. Finally, the model is applied to 40T electric furnace with high accuracy. But this model is ideal, the heat loss of the system is ignored. Wang Haijiang etc.[4]studied the law between carbon, chromium and temperature in the decarburization and chromium preservation reaction of AOD furnaces, and then developed a full molten iron smelting model. It is also concluded that when AOD full molten iron is smelted, the heat of chemical reaction and the sensible heat of the furnace lining occupy the main position, and the heat of the molten iron is secondary. Wang Yunan etc.[5]established the RH-TOP refining terminal temperature prediction model. The prediction error within ±5°C is improved to about 80%. Li qiang etc.[6]based on BP neural network to build the LF furnace temperature prediction model, but the model training tempo is to moderate, and usual to remain local optimum. Wei Fuhao etc.[7]setup a forecast model of endpoint temperature based on NARX neural network, which is accurate and a hit rate of 90% within the error ±5°C in RH furnace endpoint temperature prediction. Wu Yang[8]established an RH-MFB refining temperature prediction model based on the NARX neural network, and the error within 5°C was 89.5%. However, this method is not stable enough. Zhang Xianzhe[9]combined the energy conservation principle and partial least squares method, and then established the LF furnace temperature prediction model with Ada-Boosting algorithm and GA to improve ELM. The prediction effect is good, but it is intricate and has a large amount of calculation. Wang Miaoy etc.[10]established the TSVR model for 80t converter end point prediction, which applies twin support vector regression to temperature prediction. The model has high accuracy, and the carbon temperature double hit rate is 90%. Wang
Zhenyang etc.\cite{11} based on SVR and ELM respectively established molten iron temperature prediction models, the SVR hit rate is higher.

All the above mentioned approaches pay too much attention to the hit rate of the model and think little of the response speed of the model. Slow model responsiveness is not suitable for production practices. This paper proposes an endpoint temperature prediction model based on ELM, and optimizes the model with SSA. ELM has fast response speed, but relatively low accuracy. Improve model accuracy through SSA. After the simulation experiment, the results exhibit that the rate of convergence and exactness of the optimized model are more in line with the actual production requirements, and the endpoint temperature can be predicted more quickly, which provides a basis for the post-sequence steelmaking process.

2. ELM

Extreme Learning Machine (ELM)\cite{12-13} is different from the general neural network, only iteratively train the weights and thresholds between the implicit layer and the output layer, thereby eliminating the time of drilling the weights and thresholds between the input layer and the implicit layer, which steps up the velocity of the network.

Figure 1 depicts the ELM structure, which is similar to BP neural network.

$$\sum_{j=1}^{n} \beta_{j} g(W_{i} * X_{j} + b_{i}) = y_{j}, \quad j = 1, 2, \cdots, n \quad (1)$$

$\beta_{i} = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{iL}]^{T}$ is the output weights between the implicit layer and the output layer; $g(x)$ is activate function, Sigmoid is selected in this paper; $W_{i} = [\omega_{i1}, \omega_{i2}, \cdots, \omega_{im}]^{T}$ is the input weight between the input layer and the implicit layer; $b_{i}$ is the bias value of the i-th neuron in the hidden layer; $W_{i} * X_{j}$ is the inner product of the $W_{i}$ and $X_{j}$\cite{14}.

![ELM structure diagram](image-url)
3. SSA optimized ELM temperature model

Sparrow Search Algorithm[15](SSA) is an intelligent optimization algorithm proposed by scholars Jiankai Xue and Bo Shen in 2020 inspired by the foraging behavior and anti-predation behavior of sparrows. Sparrows are divided into two categories, one is the seeker who seeks food, and the other is the follower who follows the seeker to get food. At the same time, a warning value was introduced, and too high it would alarm the sparrow, imitating the anti-predatory behavior of the sparrow to escape natural enemies.

The Seeker's location update is described below:

\[
X_{i,j}^{t+1} = \begin{cases} \alpha \exp \left( -\frac{i}{\alpha \cdot \text{iter}_{\text{max}}} \right), & R_2 < ST \\ X_{i,j}^{t} + Q \cdot L, & R_2 \geq ST \end{cases}
\]

(2)

where \( i \) is the current iteration number, \( j=1,2,3,...,d \), \( \text{iter}_{\text{max}} \) is the maximum number of iterations. \( X_{i,j} \) indicates the position information of the \( i \)-th sparrow in the \( j \) dimension. \( \alpha \in (0,1) \) is a random number. \( R_2 \in [0,1] \) and \( ST \in [0.5,1] \) represent warning values and safety values, respectively. \( Q \) is a random number that follows a normal distribution. \( L \) represents a matrix of \( 1 \times d \), and all elements within the matrix are 1.

The location update for followers is described as follows:

\[
X_{i,j}^{t+1} = \begin{cases} Q \exp \left( \frac{X_{\text{worst}} - X_{i,j}^{t}}{2} \right), & i > \frac{n}{2} \\ X_{i,j}^{t+1} + Q \cdot L, & i \leq \frac{n}{2} \end{cases}
\]

(3)

where \( X_p \) is the optimal position occupied by the current discoverer, and \( X_{\text{worst}} \) indicates the current global worst position. \( n \) represents a matrix of \( 1 \times d \), where each element is randomly assigned a value of 1 or -1, and \( A^+ = A^T(\text{AA}^T)^{-1} \).

The Aware sparrow location update describes the following:

\[
X_{i,j}^{t+1} = \begin{cases} X_{\text{best}} + \delta \left| X_{i,j}^{t} - X_{\text{best}} \right|, & f_i > f_g \\ X_{i,j}^{t} + \gamma \left( \frac{X_{\text{worst}} - X_{i,j}^{t}}{f_i - f_w} \right) + \varepsilon, & f_i = f_g \end{cases}
\]

(4)

where \( X_{\text{best}} \) is the current global optimal position. \( \delta \) controls step and is a random value that follows the standard normal distribution. \( \varepsilon \in [-1,1] \), and \( f_i \) is the fitness value of the current sparrow individual. \( f_g \) and \( f_w \) are the current global best and worst fitness values, respectively. \( \varepsilon \) is a number slightly greater than zero.

The predicted results minus actual temperature, the absolute value of difference as a fitness function. The SSA for finding the optimal weights \( W \) and thresholds \( b \), and the optimal parameters are passed to ELM. To eliminate the error, operating for many times, average error metrics. Replace the network only with pre-optimization ELM and SVR, and use the above data to train and test model and record related data.

4. Simulation Experiment

4.1 Data selection, pre-processing and error measurement

More than 400 sets of production data were collected from a steelmaking plant, and after deleting missing data and erroneous data, 362 groups of production data were sorted out. Take 240 of these groups as training data and the remaining 122 groups as test data. Each set of data contains several influencing factors, picking out the factors that are significantly correlated with the endpoint temperature, and the factors with significant correlation are used as the inputs of the model.

The end point temperature was used as the dependent variable, and the correlation was analyzed with influencing factors...
such as molten steel quality, start temperature, treatment time, oxygen blowing, and platform temperature. Calculate the correlation coefficient between each influencing factor and endpoint temperature, discard the factors with no significant correlation (less than 0.05) \[16\].

<table>
<thead>
<tr>
<th>Input</th>
<th>Pearson's correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank condition</td>
<td>-0.2777</td>
</tr>
<tr>
<td>Molten steel quality</td>
<td>-0.1702</td>
</tr>
<tr>
<td>Slag thickness</td>
<td>-0.4141</td>
</tr>
<tr>
<td>Start temperature</td>
<td>0.6372</td>
</tr>
<tr>
<td>Platform temperature</td>
<td>0.9206</td>
</tr>
<tr>
<td>Vacuum</td>
<td>-0.3190</td>
</tr>
<tr>
<td>Arrival oxygen content</td>
<td>0.2386</td>
</tr>
<tr>
<td>Processing time</td>
<td>0.7139</td>
</tr>
<tr>
<td>Nitrogen argon flow rate</td>
<td>0.7066</td>
</tr>
<tr>
<td>Blow oxygen</td>
<td>0.0749</td>
</tr>
</tbody>
</table>

Through the above correlation analysis, the input variables were determined as: tank condition, molten steel quality, slag thickness, start temperature, platform temperature, vacuum, arrival oxygen content, processing time, nitrogen argon flow rate, and blow oxygen. The data are normalized to eliminate the differences between various influencing factors and avoid orders of magnitude influencing factors guiding the model training direction. Use Equation (5) to normalize the data to [-1,1].

\[ x_j = 2 \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} - 1 \]  

where: \(x_i\) is the ith original input or output; \(x_{\max}\) and \(x_{\min}\) are the maximum and minimum values of the original input or output, respectively; \(x_j\) is the input or output after the ith original input or output normalized mapping.

The quality of the training results is measured by the mean absolute error (MAE), correlation coefficient \(R^2\), model response rate (Duration) and the proportion of the data with error within 5°C to the total data (Hit rate).

\[ MAE = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(X_{\text{test}} - X_{\text{predict}})^2} \]  

\[ R^2 = \frac{(N \sum_{i=1}^{N} X_{\text{test}} - X_{\text{predict}} - \sum_{i=1}^{N} X_{\text{test}} \sum_{i=1}^{N} X_{\text{predict}})^2}{[N \sum_{i=1}^{N} X_{\text{test}}^2 - (\sum_{i=1}^{N} X_{\text{test}})^2][N \sum_{i=1}^{N} X_{\text{predict}}^2 - (\sum_{i=1}^{N} X_{\text{predict}})^2]} \]  

In equations (6) and (7), \(N\) is the total number of sample data for the test set; \(X_{\text{test}}\) is the test sample data; \(X_{\text{predict}}\) is forecast data.

### 4.2 Establishment of end-point temperature prediction model

To avoid blindly testing the weights and thresholds of ELM, SSA is cited to find the best weights and thresholds. The flow chart of SSA-ELM is shown in Figure 2. Steps as follows:

1. Normalize 362 groups of original data.
2. Initialize \(\text{iter}_{\text{max}}, \text{ST}\) and other parameters. Calculate the fitness value and sort it.
3. Use Equations (2), (3) and (4) to update the position and obtain the parameter \(X_{\text{best}}\).
4. The prediction model was obtained by substituting the optimal parameters into ELM.
5. Judge the prediction effect of the model. If the deviation is large, return to step 2 until the requirements are met.
Sets the parameters of the algorithm
Randomly select the training and testing sets
Train the training set and build an SSA-ELM model
Calculate the fitness of the individual
Sort and classify individuals according to fitness
Grouping and categorization
Meet the termination criteria?
Update the location
Outputs optimal weights and thresholds

Fig. 2 Flow chart of SSA-ELM prediction model

4.3 Simulation experiment and analysis
To compare the performance of the above model, 122 groups were randomly picked from 362 groups of data as the test set, and record the relevant performance index data. The average of 50 times was recorded. The final prediction results are shown in Table 2 and Figure 3. It can be seen intuitively from Figure 3 that most of the error points of the SSA-ELM model fall within the of \( \pm 5 \)\(^\circ\)C error bars.

Fig. 3 Comparison of prediction errors between SSA-ELM model and ELM model
Table 2  Comparison of end point temperature predicted by each model

<table>
<thead>
<tr>
<th>Indicators</th>
<th>ELM</th>
<th>SVR</th>
<th>SSA-ELM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>4.2376</td>
<td>1.8546</td>
<td>2.4591</td>
</tr>
<tr>
<td>R²</td>
<td>0.5064</td>
<td>0.8768</td>
<td>0.8388</td>
</tr>
<tr>
<td>Hit rate/%</td>
<td>66.3934</td>
<td>90.1639</td>
<td>86.1789</td>
</tr>
<tr>
<td>Duration/s</td>
<td>12.8239</td>
<td>23.4468</td>
<td>19.4286</td>
</tr>
</tbody>
</table>

5. CONCLUSION

1) Temperature is critical to the unfolding of the steelmaking process. Limited by the process environment can only predict the temperature, the existing model accuracy and convergence rate can't meet the process requirements at the same time. In this paper, SSA is combined with ELM to improve the hit rate. The model has certain reference value for endpoint temperature prediction.

2) To solve the cumbersome problem of selecting ELM parameters by manual testing and other methods, the SSA is introduced to optimize the ELM. The best ELM parameters are found by the excellent global optimization performance of SSA. The optimized SSA-ELM model is applied to the AOD furnace.

3) According to the metallurgical process flow and the correlation analysis, the factors with significant correlation with the endpoint temperature were taken as the independent variables. To compare model performance, three models based on SVR, ELM and SSA-ELM were established. The SSA-ELM was trained and tested through 362 sets of actual production data. According to the statistics, the hit rate can reach 86.1789% under the condition of 5℃error, and the average response time is 19.43s. The SSA-ELM is better than ELM in accuracy, better than SVR in response speed, and more suitable for guiding actual production.

References

Transformer in Galaxy: A study of galaxy classification based on the Swin Transformer

Zhenyu Huang
School of Physics and Astronomy, Sun Yat-Sen University
Huangzhy77@mail2.sysu.edu.cn

Abstract

By replicating the CNN part of Galnet network and inspired by Swin Transformer, the article constructs a Transformer-based model called SwinT Galnet by adding the Swin Transformer block into Galnet thus introduces the attention mechanism into that traditional model, trying to investigate the improvement of Transformer over CNN in a single Sérsic galaxy model data. This article compares Sérsic parameters of SwinT with those of CNN. The paper compares the parameters trained by Galnet and SwinT Galnet and shows that the Transformer has an improvement by calculating global attention in accuracy over the pure CNN-based network in galaxy parameters regression tasks, indicates that Transformer has a large potential in the field of galaxy morphology.

keywords: Galaxy; Convolution Neural Network; Transformer

1. Introduction

With the progress of observational techniques improvements, we are able to observe millions of extragalactic galaxies. The study of galaxy morphology now accounts for a significant part of galaxy cosmology, for galaxies of all shapes and sizes. The study of galaxy morphology can not only disclose the evolution of galaxies, but also reproduce the evolution scenario of the Universe and validate cosmological models. Most existing studies of galaxy morphology are based on the study of photometric measurements. With the widespread use of computer vision, which has flourished with deep learning, the use of deep learning methods can help researchers to quickly complete the work of galaxy identification and classification in massive amounts of data.

The traditional methods analyze the observational data with empirical scaling functions fitting. It takes lots of time to fit parameters comparing with machine learning. Tuccillo [1] for the first time applied CNNs to two-dimensional light profile galaxy fitting on HST/CANDELS data. Roy Li adding PSF information in Galnet [2] shows that CNN helps researchers gain high accuracy in a short time. The simple idea behind the Galnet is that we could put the galaxies images into that model and predict parameters required through the networks. Since the neural network could be regarded as a kind of regression method, having CNN to learn how labelled profiles would be in images with the background structure and seeing of real observed galaxies. Comparing with fitting tools used as benchmarks in galaxies fitting, the advantage of the deep-learning method is that we could design a general model for various empirical model which help us better eliminate galaxy models and discover more reliable models.

Currently, although the CNN-based methods have achieved excellent performance in the field of galaxy identification, it could hardly improved accuracy in several parameters especially Sérsic index regression tasks with different CNN-based backbones. Image identification is still a challenge task in astronomy image analysis. Since the intrinsic locality of convolution operation, it is hard for CNN-based approaches to learn explicit global and long-range semantic information interaction[3]. Transformer was first proposed for the machine translation task in [4]. In the NLP domain, the Transformer-based methods have achieved the state-of-the-art performance in various tasks [5]. Driven by Transformer’s success, the researchers introduced a pioneering vision transformer (ViT) in [6], which achieved the impressive speed-accuracy trade-off on image recognition task. Compared with CNN-based methods, the drawback of ViT is that it requires pre-training on its own large dataset. To alleviate the difficulty in training ViT, Deit [7] describes several training strategies that allow ViT to train well on ImageNet. It is worth mentioning that an efficient and effective hierarchical vision Transformer, called Swin Transformer, is proposed as a vision backbone in [8]. Based on the shifted windows mechanism, Swin Transformer achieved the state-of-the-art performance on various vision tasks including image classification, object detection and semantic segmentation.
Since Transformer could require millions of data to train, transformer would be a excellent galaxy morphological classification candidate due to the increasing amount of galaxy images for future sky surveys. J.Yao in [9] purposed an approach to classified galaxies from galaxyzoo[10]. This work is the first attempt of applying Swin Transformer on Sérsic galaxy morphological classification tasks. We redeveloped CNN-based model Galnet, as the standard to compare. Hereafter, we will build the SwinT block as the Conv2d. With the new SwinT block, the new model ,SwinT Galnet, could have the ability to calculate global attention. We could demonstrate in that the performance of SwinT Galnet is competitive with the Galnet, the pure CNN model with the number of training data being only around a hundred thousand. Finally, we find that transformer models are specifically good at classifying galaxies in that one Sérsic profile parameters regression takes comparing with the pure CNN-based Galnet. With this promising preliminary result, It can be seen that in the field of astronomy, Transformer still has great potential for processing complex theoretical models as well as observational data.

2. Models

2.1. Sérsic Galaxies Simulating

Majority of the analyses of galaxy structural parameters have been mainly based on one-component Sérsic profile[11]. Galaxy structural parameters, like the total magnitude mag, the effective radius $R_{\text{eff}}$ and Sérsic index could be derived by some empirical scaling relations. With three free structural parameters (the total magnitude $I$, the effective radius $R_{\text{eff}}$, and the Sérsic index $n$), together with other shape/position parameters (the axis ratio $q$, the position angle $\theta$ and the galaxy center coordinates $x_{\text{cen}}; y_{\text{cen}}$), this model is able to match most of galaxies’ light distribution of variety morphology.

The light distribution of galaxies can be defined as

$$I_R = L_e \exp\left(-b_n \left[ \left( \frac{\sqrt{q(x-x_{\text{cen}})^2 + (y-y_{\text{cen}})^2}}{R_{\text{eff}}} \right)^{1/n} - 1 \right] \right)$$

(1)

Here, $R_{\text{eff}}$ is the effective radius, $I_e$ is the surface brightness at the effective radius, $x_{\text{cen}}$ and $y_{\text{cen}}$ is the position of the galaxy center. The position angle $\theta$ gives the angle between the center plane and the sky coordinate frame, which is defined by assuming to be positive from North to East. And $q$ is the axis ratio, while $n$ is the Sérsic index which refers to early-type galaxy when $n \leq 2.5$. Considering galaxies in low redshift[12] $z < 0.5$, we would have following relations.

For $b_n$ provided by [13] can be written as the following

$$b_n = \begin{cases} 
2n - 1.3 + 4/(405n), & n \geq 0.36 \\
0.01945 - 0.8902n + 10.95n^2, & n < 0.36 
\end{cases}$$

(2)

$F_{\text{tot}}$ is the total flux of the galaxies, defined as

$$F_{\text{tot}} = \pi R_{\text{eff}}^2 I_e e^{b_n} n^{-2n} \Gamma(2n+1) q$$

(3)

The total magnitude $\text{mag}$ is defined by

$$\text{mag} = -2.5 \log(F_{\text{tot}}) + zpt$$

(4)

The observed galaxies images is the result from the convolution of the intrinsic (real) galaxy fluxes and the PSFs. The PSFs largely change surface brightness distribution and result in degency in some parameters which could account for bad result of the traditional fitting tools. Using prepared PSFs, our data could be written as the following

$$M(BG, \{p_R\}) = BG + I(\{p_R\}) \circ S$$

(5)

where $I$ is the galaxy surface brightness distribution, $BG$ refers to the value of local background, $S$ is the PSF while $\circ$ denotes convolution.

\(^1\) point-spread-function
2.2. The CNN architecture

Inspired by the network structure of VGGnet[14], a convolutional neural network[2] is built to identify one Sérsic galaxies with predicting 5 Sérsic parameters, and the overall network framework is shown in the following figure.

![The architecture of Galnet](image)

There are two parts in the model: 1) convolution blocks 2) fully connected layers. The CNNs structure, the backbone of the model is used as encoder to extract features from those input images. The main path in backbone process the figures while the other path gives the weights learned from provided PSF information. These two paths are concentrated together after flattening. The fully connected layers are set as decoder to derive the relations between features extracted by CNNs and labels. In short, the fully connected layers are located at the end of convolution blocks of networks to perform a non-linear combination of the existing features from high-level to low level.

2.3. The SwinT Galnet architecture

Each swin transformer block is composed of LayerNorm (LN) layer, multi-head self attention module, residual connection and 2-layer MLP with GELU non-linearity. The window based multi-head self attention (W-MSA) module and the shifted window-based multi-head self attention (SW-MSA) module are applied in the two successive transformer blocks, respectively.

![The architecture of SwinT block](image)

From this two SwinT block, the attention is computed as the follows:

\[
\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V
\]

where \( Q, K, V \in \mathbb{R}^{m \times d} \) denote the query, key and value matrices. \( M^2 \) and \( d \) are the number of patches in a window and a dimension of the query or key. The values of \( B \) are taken from the bias matrix \( \bar{B} \in \mathbb{R}^{(2M-1) \times (2M+1)} \).
We have rewritten a plug-and-play SwinT module which is similar in usage to Conv2d and can replace the convolution module in Galnet more easily. Since SwinT block provides a way to calculate attention in the whole image, we replace one of the VGGnet block with SwinT. The SwinT Galnet would be like in the following picture.

![Image of the architecture of SwinT Galnet]

Fig. 3: The architecture of SwinT Galnet

### 3. Experiments

#### 3.1. Training and Testing Data

Using KIDS[15] background as BG and PSFs provided by KIDS[16], we have simulated galaxies from ([5]). Setting simulated galaxies in the center of images, it is reasonable to set \( x_{cen} \) and \( y_{cen} \) to zero which can be ignored. The remaining distribution of 5 parameters are in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Units</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{cen} )</td>
<td>0</td>
<td>arcsec</td>
<td>-</td>
</tr>
<tr>
<td>( y_{cen} )</td>
<td>0</td>
<td>arcsec</td>
<td>-</td>
</tr>
<tr>
<td>( \text{mag} )</td>
<td>17 – 22</td>
<td>–</td>
<td>exponential (( s=1 ))</td>
</tr>
<tr>
<td>( q )</td>
<td>0.2-1.0</td>
<td>–</td>
<td>uniform</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0-3.14</td>
<td>rand</td>
<td>uniform</td>
</tr>
<tr>
<td>( \text{Ref} )</td>
<td>0.2 – 4</td>
<td>arcsec</td>
<td>normal (( \log(\text{Ref}) )) (( \mu = -0.1, \sigma = 0.4 ))</td>
</tr>
<tr>
<td>( n )</td>
<td>0.2 - 8.0</td>
<td>–</td>
<td>( F ) (( n=30, d=5 ))</td>
</tr>
</tbody>
</table>

Adopting an SNR > 50 for the training sample is made for uniformity with typical choices made for real galaxies to ensure reasonable accuracy on the fitting parameters.

#### 3.2. Training the Network

Using galaxies images in the size of 64 × 64 pixels and the PSFs in the size of 25 × 25 trained model. The outputs are 5 parameters describing the light distribution of Sérsic galaxies. We train model by minimizing the SmoothL1 loss function which defined as the following with an Adam[17] optimizer.

\[
L(a) = \begin{cases} 
\frac{1}{2} a^2, & a < 1 \\
\|a\| - 0.5, & a \geq 1 
\end{cases}
\]  

(7)
where $a = y_{\text{true}} - y_{\text{predict}}$. This loss function penalizes outlier points less, is more robust and achieves the required accuracy faster and more stable.

When building SwinT Galnet, we replace the activation function following the SwinT block with SELU[18], which could avoid the dead RELU problem due to excessive learning rate in the initial stage.

### 3.3. Testing on simulated data

After training Galnet and SwinT Galnet with the same galaxies images, we use test samples to make comparison on their performances. We used R-squared to show the linear relations between these parameters. The $R^2$ is defined as

$$R^2 = 1 - \frac{\sum (p_i - t_i)^2}{\sum (t_i - \bar{t})^2}\quad(8)$$

In linear least squares multiple regression with an estimated intercept term, $R^2$ equals the square of the Pearson correlation coefficient between the observed and predicted data values of the labels value.

### 3.4. Assessing the statistical errors

$$e_s = \frac{1}{N-1} \sum_{i=1}^{N} (p_i - t_i)^2\quad(9)$$

$$\Delta e = \frac{e_s}{\bar{p}}$$

where $\bar{p}$ is the mean value of labeled data and $\Delta e$ means the error between prediction and label.

### 4. Conclusion

#### 4.1. Testing result

After training 200 epochs, we could draw the testing results in the following figures. The black base line refers to the perfect situation that the prediction is totally equal to the real data. The blue dots are the prediction of Galnet while the orange dots are the predictions of SwinT Galnet.

Fig. 4: The top left figure is the photometry, the top middle panel is the value corresponding to Sérsic index, the top right figure is the effective radius $R_{\text{eff}}$ distribution, the bottom left is the position angle distribution, and the bottom middle panel the axis ratio $q$. 

---

**Downloaded From:** https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 01 Nov 2023
**Terms of Use:** https://www.spiedigitallibrary.org/terms-of-use
Table 2: Analysis of Prediction Results Table

<table>
<thead>
<tr>
<th>Model</th>
<th>Magnitude ma</th>
<th>Effective radius</th>
<th>Sérsic index n</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galnet</td>
<td>0.02232</td>
<td>0.04854</td>
<td>0.7806</td>
</tr>
<tr>
<td>SwinT Galnet</td>
<td>0.01724</td>
<td>0.02966</td>
<td>0.6734</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galnet</td>
<td>0.974513</td>
<td>0.877638</td>
<td>0.6874</td>
</tr>
<tr>
<td>SwinT Galnet</td>
<td>0.980314</td>
<td>0.915300</td>
<td>0.7289</td>
</tr>
</tbody>
</table>

From the table, we can see that after 200 training sessions, the performance of Galnet with the SwinT module shows an improvement over the traditional method. However, it is worth noting that for the data convolved with the PSF, both the original Galnet and the model with the SwinT module, the predictions of Sérsic index n still do not reach the accuracy of the first two sets of parameters, especially for universes with n > 5, which we believe is a possible reason for the small training sample of the model, and secondly for galaxies with n larger than 5 the difference in luminosity is not significant. There is a relatively large jump in position angle from 0 to a predicted value of π in radian, which is already a cycle in the actual image.

4.2. Analyzing the training process

We recorded the predictions of the model at 1, 10, 30 and 50 training epochs, which are shown in the following figures. While the black base line y = x refer to the most accuracy prediction, the blue dots are predictions of Galnet and the orange dots are SwinT Galnet.

Fig. 5 The distribution of photometry at 1 (top left), 10 (top right), 30 (bottom left) and 50 (bottom right) training epochs, where the black dashed line is the baseline y = x, the blue dots are the original galnet, and the orange dots are the predictions of SwinT Galnet.
Fig. 6 The distribution of Sérsic index $n$ at 1 (top left), 10 (top right), 30 (bottom left) and 50 (bottom right) training epochs.

Fig. 7 The distribution of effective radius $R_{\text{eff}}$ training at 1 (top left), 10 (top right), 30 (bottom left) and 50 (bottom right) training epochs.

From the experimental result in table 2, SwinT Galnet has higher accuracy on parameters than the Galnet. It shows that deep learning model with attention has ability to extract more features in shallow layers with learning global and long-
range information which enables model to reach the same accuracy with less epochs comparing with the pure CNN structure Galnet.

5. Conclusion

In this experiment, Swin Transformer block performs well in a regression task with 5 specific parameters, and we are also able to test the limits of its feature extracting ability in more complex galaxy models, and thus extract or exclude specific galaxy models in the field of astronomy. Training and testing with one Sérsic objects, the experiment compared Galnet with SwinT Galnet. The small leading prediction of SwinT Galnet with the introduction of the transformer block design can be still improved for some parameters, which may be due to the insuficient design of our de-PSF convolutional network structure. Exploring the new network structure to reach the model to master the operation of de-PSF is the next direction that can be experimented, while specific feature engineering can make our model better to extract the target model. It can be seen that in the field of astronomy, Transformer still has great potential for processing complex theoretical models as well as observational data.

6. Acknowledgements

Author acknowledged Dr. Rui Li for sharing the code and data. Author acknowledged Mr. Chen Qiu for sharing the experience with training the Galnet.

References


iDHS-DPPE: a method based on dual-path parallel ensemble decision for DNase I hypersensitive sites prediction

Xueping Lv*, Yufeng Wang, Hongwen Liu

*aSchool of Electrical Engineering and Automation, Anhui University, Hefei, China
* Corresponding author: lvxp0616@163.com

ABSTRACT

The DNase I hypersensitive site (DHS) is the chromatin region that exhibits a hypersensitive response to cleavage by the DNase I enzyme. It is a universal marker for regulatory DNA and associated with genetic variation in a wide range of diseases and phenotypic traits. However, traditional experimental methods have limited the rapid detection of DHS as well as its development. Therefore, effective and accurate methods to explore potential DHSs need to be developed urgently. In this task, a deep learning approach called iDHS-DPPE to predict DHSs in different cell types and developmental stages of the mouse. iDHS-DPPE uses a dual-path parallel integrated neural network to identify DHSs accurately. First, the DNA sequence is segmented into 2-mers to extract information. Then, the BiGRU-Attention model captures remote dependencies and the MSFRN model enables hierarchical information fusion. The dual models are trained separately to enhance the feature information. Finally, the ensemble decision of two models yield the prediction results, enabling the integration of information from multiple views. The average AUC across all datasets was 93.1% and 93.3% in the 5-fold cross-validation and independent testing experiments, respectively. The experimental results demonstrate that iDHS-DPPE outperforms the state-of-the-art method on all datasets, proving that iDHS-DPPE is effective and reliable for identifying DHSs.

Keywords: mouse, DNase I hypersensitive site, deep learning, dual-path parallel, ensemble decision

1. INTRODUCTION

In genetics, specific chromatin regions that are insensitive to cleavage by DNase I nuclease are known as DNase I hypersensitive sites (DHSs). DHSs can help localize transcriptional regulators in mammals. Also, they are associated with chromatin state status. Additionally, genetic research have identified a number of non-coding variants related to diseases and traits, such as cancer and cardiovascular disease. And these variants are clustered in regulatory DNA tagged by DHSs. Therefore, the accurate identification of DHSs is beneficial in further exploring the regulation of genes and the pathogenesis of common diseases.

Currently, methods such as Southern blots and DNase-seq have been frequently used in biology laboratories to identify DHSs and a variety of other components of gene regulation. Given the inefficient and expensive drawbacks of traditional biological methods, an increasing number of computational methods are being applied. Noble et al. predicted DHSs using a support vector machine (SVM) classifier based on the RevKmer feature. SVM-PseKNC, iDHS-DMCAC, iDHSs-MFF, iDHS-SVM and iDHS-DT were similarly chosen SVM as classifiers. The iDHS-DXG and CEPZ had each developed a model with XGboost based on different dinucleotide property matrices. In addition, iDHS-EL, iDHSs-PseTNC and iDHS-DSAMS machine learning models were also developed. However, these predictors were developed for humans. Dao F Y et al explored models based on CNNs and LSTMs to predict DHSs in mouse.

DHSs on genes at different developmental stages in mouse are dynamically variable and sequences in different tissues are highly variable. Accurate identification of multiple types of data is always difficult to achieve with a single model. However, for predicting DHSs in multiple developmental stages and tissues in mouse, there is a need to construct predictors with powerful generalization capabilities. Integrating models to extract information from different views can improve generalisation and get better performance than a single predictive model. This offers us a solution idea. At the same time, natural language processing techniques have demonstrated their full value in bioinformatics text mining. Sequences might be better represented by treating them as sentences, with nucleotides forming words.
All things considered, we construct a deep learning model (iDHS-DPPE) with dual-path parallel ensemble decision making to accurately predict DHSs for mouse. The multifaceted experimental results show that iDHS-DPPE offers better performance and generalisation than existing methods as a powerful method for predicting DHSs.

## 2. MATERIALS AND METHODS

### 2.1 Datasets

Since there is currently no published database of mouse DHSs, we use the benchmark datasets from iDHS-Deep\(^1\). The DHS sequences were obtained from the comprehensive map of mouse created by Breeze et al\(^2\). DHS sequences of 50-301 bp length were selected for positive samples. To construct suitable negative samples, the centroids were first set between suitable adjacent DHS sequence fragments. Sequence fragments on either side of these centroids were selected as negative samples. The length of these negative samples is also between 50 and 301 bp. CD-HIT\(^2\) was used to eliminate sequences with a similarity threshold of 0.8 to reduce redundancy and homology bias. Samples of different cell types and developmental stages were divided into different datasets, and 70% of the samples were selected as training datasets and 30% as independent test datasets. Details of the datasets used in this paper can be found in Table 1.

<table>
<thead>
<tr>
<th>Types of Data</th>
<th>Training Datasets</th>
<th>Independent Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Posesitive Nega</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forebrain</td>
<td>7114</td>
<td>7114</td>
</tr>
<tr>
<td>Midbrain</td>
<td>10299</td>
<td>10299</td>
</tr>
<tr>
<td>Hindbrain</td>
<td>5766</td>
<td>5766</td>
</tr>
<tr>
<td>Liver</td>
<td>6519</td>
<td>6519</td>
</tr>
<tr>
<td>Lung</td>
<td>7424</td>
<td>7424</td>
</tr>
<tr>
<td>Heart</td>
<td>30929</td>
<td>30929</td>
</tr>
<tr>
<td>Kidney</td>
<td>6316</td>
<td>6316</td>
</tr>
<tr>
<td>Limb</td>
<td>4978</td>
<td>4978</td>
</tr>
<tr>
<td>Stomach</td>
<td>1062</td>
<td>2125</td>
</tr>
<tr>
<td>Thymus</td>
<td>1612</td>
<td>1612</td>
</tr>
<tr>
<td>Craniofacial</td>
<td>2515</td>
<td>2515</td>
</tr>
<tr>
<td>Retina</td>
<td>3511</td>
<td>3511</td>
</tr>
<tr>
<td>Muller retina</td>
<td>2877</td>
<td>2877</td>
</tr>
<tr>
<td>Neural tube</td>
<td>1224</td>
<td>1224</td>
</tr>
<tr>
<td>Stages</td>
<td>ESC</td>
<td>7521</td>
</tr>
<tr>
<td>Early-Fetal</td>
<td>52418</td>
<td>52418</td>
</tr>
<tr>
<td>Late-Fetal</td>
<td>16172</td>
<td>16172</td>
</tr>
<tr>
<td>Adult</td>
<td>21247</td>
<td>21247</td>
</tr>
</tbody>
</table>

### 2.2 Feature Extraction

The mouse DNA sequence is composed of four types of nucleotides, including 'A' (Adenine), 'C' (Cytidine), 'G' (Guanine) and 'T' (Thymine). We encode sequences by forming nucleotides into "words".

Convolutional neural networks often require a constant size of input. And the sequence length in the datasets varies from 50 to 301bp. When the sequence is less than 301bp long, add "N" to the end of the sequence. Padding the sequence to its maximum length, which is uniformly 301bp. There are five types of bases in the populated sequence, making up a total of 21 position-specific dinucleotide pairs. Splitting DNA sequences into 2-mers by sliding window, e.g. GACGT ... CANN → 'GA ', 'AC ', 'CG ', 'GT ', ......, 'CA ', 'AN ', 'NN '. We define a dictionary that maps the different ‘words’ to the corresponding values in the range 0–20. In this way, the sequence is encoded as a numeric vector of length 300.

Convolutional layers are often used in CNNs to further extract more complex features, and pooling layers to reduce feature dimensionality. The generated sequence feature embedding matrix is further feature extracted by the convolutional and maxpooling layer.
2.3 Ensemble Models

Ensemble models predict DHS and non-DHS from different views. The extracted features are fed into the BiGRU-Attention model and the MSFRN model respectively. The architecture of the iDHS-DPPE to be shown in Figure 1.

![Architecture of iDHS-DPPE](image)

Figure 1. The architecture of iDHS-DPPE. (A) Flow chart of iDHS-DPPE. The DNA sequence is mapped from the C2 encoding scheme to the embedding layer, and feature information is further extracted by the convolutional and maxpooling layers. After that, local information and remote dependencies are captured by BiGRU-Attention and MSFRN basis models to get a probabilistic representation. Ultimately, the results of the two models are evaluated together to get a final output to determine whether DHSs. (B) (C) The details of BiGRU+Attention layers and Multi-scale Residual Network.

Recurrent neural networks (RNNs) have been intensively used in research related to natural language processing to address long-term dependence in sequences. BiGRU splices the generated forward and backward sequences and the resultant output is passed to the downstream network, able to fully learn the contextual information in the sequences\textsuperscript{22}. The attention mechanism allows the model to be trained with more attention to important characteristics and avoid the interference of useless information\textsuperscript{23}. The MultiHead attention layer assigns different weights to the output of BiGRU, obtaining an intermediate representation within each head. The results of each head are then concatenated and mapped back to the original dimension.

The formula for the MultiHead attention layer is presented below:

$$\text{head}_i = \text{Attention} \left(qW^q_i, kW^k_i, vW^v_i\right)$$ \hspace{1cm} (1)

$$\text{MultiHead}(q, k, v) = \text{Concat} \left( \text{head}_1, \ldots, \text{head}_n \right)W^O$$ \hspace{1cm} (2)

where $W^q_i \in \mathbb{R}^{d_{\text{model}} \times d_i}$, $W^k_i \in \mathbb{R}^{d_{\text{model}} \times d_i}$, $W^v_i \in \mathbb{R}^{d_{\text{model}} \times d_i}$, $W^O \in \mathbb{R}^{d_{\text{model}} \times d_i}$. In the BiGRU-Attention model, BiGRU captures remote dependencies and the MultiHead attention layer selects key feature information.

Residual networks can avoid gradient disappearance and overfitting compared to traditional CNNs\textsuperscript{24}. Multi-scale convolution can obtain information at different scales of the sequence and reduce the loss of information. In the MSFRN...
Model, six convolutional layers of different scales (all with 64 convolutional kernels) are cascaded to achieve fusion of layered features. In addition to this, a convolutional layer (with 192 convolutional kernels) is added to fuse the features with residual connections and organize the information across channels. The MSFRN model extracts local relationships between adjacent features by fusing features at different scales to enhance the model representation.

2.4 Dual-path Parallel Ensemble Strategy

Ensemble methods can effectively improve generalisation and robustness by combining multiple base models. For the BiGRU-Attention and MSFRN base models trained in parallel for dual paths, we choose to use the greedy fuzzy hybrid strategy in PmliPEMG25 to integrate at the decision level according to certain rules. In identifying whether to be DHSs, selecting the model with a higher confidence level can help us to make better predictions. Here, the confidence level of the two base models is measured using $G$. The prediction probabilities obtained from the base models are evaluated together to get better classification results. $G$ is calculated as:

$$G = ab(2cp - 1), cp \in [0,1]$$

As the above equation shows, when the confidence probability $cp$ is higher, then the value of $G$ is larger, which can be considered that this base model is more reliable to determine whether it is DHS. When $cp \geq 0.5$, the sequence is believed to be a DHS, or not.

3. RESULTS AND DISCUSSION

3.1 Evaluation criteria

We use several common metrics to evaluate the performance of the model, including sensitivity (Sn), specificity (Sp), accuracy (Acc), Matthew correlation coefficient (Mcc). These metrics can be formulated as follows:

$$Sn = \frac{TP}{TP + FN}$$

$$Sp = \frac{TN}{TN + FP}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}}$$

where TP, TN, FP and FN indicate the number of true positive, true negative, false positive and false negative samples, as respectively. Furthermore, area under the receiver-operating characteristic curve (AUC) is also used for assessment.

3.2 Comparison with different feature representations

Initial features significantly affect the performance of the model. To determine the best length of the 'word', we have experimented. Due to limitations on the length of the sequence, we only considered the cases $k = 1, 2, 3$ for $k$ in k-mers. In the meantime, the onehot coding scheme, common in biology, is also added. The AUC values from 5 fold cross validation of all datasets are shown in Figure 2.

Both 1-mers and onehot extract information from individual bases, but may ignore base-to-base connections. Compared to 1-mers and onehot, 2-mers takes advantage of the long-term dependency between adjacent nucleotides by treating two bases as a whole. 3-mers can also capture the sequential information of multiple base pairs, so good prediction performance is obtained. Nevertheless, higher order coding often implies higher computational costs and is prone to overfitting. In some of the smaller data sets, 2-mers would be more suitable than 3-mers. The experimental results demonstrate that the 2-mers encoding scheme is capable of extracting sufficient sequence information and is effective for identifying DHSs.
3.3 Comparison with different neural network structures

The feature vectors are fed into several different neural networks to verify the effectiveness of the dual-path parallel integration strategy. We will compare this with 3 structures: the CNN-LSTM, the BiGRU-Attention model and the MSFRN model. The network structure of CNN-LSTM is the same as in iDHS-Deep\textsuperscript{19}.

The average AUC values of CNN-LSTM, BiGRU-Attention, MSFRN and iDHS-DPPE are 0.910, 0.926, 0.928 and 0.931 respectively. As shown in Figure 3. We can find that iDHS-DPPE achieves the best results among all methods. As a comparison with CNN-LSTM illustrates that the network structure of iDHS-DPPE is able to represent the features more accurately. Compared with BiGRU-Attention and MSFRN model verifies the effectiveness of the dual-path parallel ensemble strategy, which enhances the prediction performance of the model.

3.4 Comparison with existing methods

To be able to visually present the superiority of our model, we compare to the following published methods on a neural tube benchmark dataset, including SVM-RevcKmer\textsuperscript{8}, iDHS-EL\textsuperscript{16}, iDHSs-PseTNC\textsuperscript{17}, iDHS-DSAMS\textsuperscript{18}, iDHS-DXG\textsuperscript{14}, iDHS-Deep\textsuperscript{19}. Of these, iDHS-Deep\textsuperscript{19} is a prediction tool specifically for DHSs in the mouse genome, while the other methods are used to predict humans. The comparison results of each method are shown in Figure 4. It is clear that iDHS-DPPE is optimal in all the evaluated metrics. Clearly, iDHS-DPPE is more robust and reliable in predicting the differentiation of mouse DHSs.
4. CONCLUSION

We present a new deep learning model iDHS-DPPE to identify DHS in various tissues and developmental stages of the mouse genome. The 2-mers feature is utilized to obtain sequence information in the mouse genome. The base models MSFRN and BiGRU-Attention explore local and global contextual dependencies. For the dual-path training results, the integration strategy serves to enhance generalisation and prediction performance. Various experiments have proven that iDHS-DPPE performs optimally.

Next, we will explore more effective feature extraction methods to mine more potential information. In addition, because of the large genetic differences among multiple species such as mice, humans and plants, the majority of methods are targeted at identifying DHSs of a specific species, with poor cross-species prediction capabilities. Therefore, a general prediction framework capable of identifying multi-species DHSs is also urgently needed to be developed.

REFERENCES


Exploration of Profit Model of Beach Sports Based on Data Drive

GuoQiang  Zenga, ZiYi Rena , Huan Zhoua, Hong Wangb, JianRong Tanga, Hua Tangb

a Hainan University, School of Applied Science and Technology, Haikou, 570228, China
b Hainan University, School of Physical Education, Haikou, 570228, China
* Corresponding author: z452730354@163.com

ABSTRACT

The innovative integration of beach sports and the tourism industry, as a breakthrough in sports tourism, is the current development direction of the tourism industry. With the development of the economy and the improvement of living standards, artificial intelligence technology is gradually applied in various fields. Using the data accumulated by the tourism industry, the application of big data technology in the tourism industry can be realized. First, the beach sports and the tourism industry are integrated; second, the risk assessment model is used to determine the economic profit trend brought by beach sports in the development of tourism; then, the validity of the model is verified by experiments.

Keywords: Beach sports; Tourism industry; Innovation and integration; Artificial intelligence technology

1. INTRODUCTION

At present, the leisure style of Chinese residents is becoming more and more popular, but most of them are passive "leisure", such as hanging out, chatting, watching TV, etc. These leisure methods lead to the disharmony between my country's social and cultural environment and the pace of the times to a certain extent. While sports as an active way of life has been widely accepted in developed countries, the World Health Organization (WHO) regards sports as a way of "active life"[1-2], and held in Geneva in 1997 with "Active Life": Sports and health", and launched the “Global Active Living Movement”, making sports a way of leisure that can stimulate people’s high spirits and positive progress, and create a vibrant social and cultural atmosphere[3]. To this end, the development of sports work in our country should change the way of thinking, overcome the tendency of the biological sports concept that overemphasizes will training and physical enhancement, vigorously develop leisure sports, and make the development of sports have a "happy" aspect. The taste of life, "leisure" life style and "leisure" life enrichment meet people's pursuit of a high-quality and harmonious life after the achievement of the economic indicators of a well-off society, and promote the full realization of a well-off society and a harmonious society[4-5].

Beach sports in sports can be well integrated into the tourism industry, forming a new situation of sports tourism and bringing a new tourism experience to tourists[6]. Marine ecology brings pressure. Most cities have many high-quality tropical and subtropical natural beach resources, which are natural beach sports holy places. Human movement is between the beach, the ocean and the blue sky, forming a realm of harmony and unity between man and nature, which is also the original intention of developing beach sports[7-8]. Cities that are in the stage of optimizing and upgrading their economic structure can abandon the previous model of relying on sacrificing the ecological environment to develop the real estate industry to drive economic development, and transform to vigorously develop tourism and modern service industries and other pillar industries to ensure a balanced ecological environment[9-11].

Therefore, the combination of beach sports and tourism can effectively promote economic growth, form an effective economic growth model, and drive the development of regional economy. This paper uses data analysis technology and relevant risk assessment models to effectively combine beach sports and tourism to realize the judgment of economic trends.

2. MODEL DESIGN

Structural entropy weight method can effectively evaluate the hierarchical relationship in the indicator system of economic development trend, and decompose the indicators into independent hierarchical structures. First, the Delphi survey method is used to collect the evaluation opinions of experts, and the importance of the indicators is typically ranked; secondly, the uncertainty theory is used to analyze the ranking structure, and the existing deviation data are
statistically processed to obtain the importance of different indicators at the same level. Finally, the index weight is determined according to the importance ranking value, so as to analyze the effective integration of beach sports and economic development from expert opinions.

2.1 Collection of expert opinions

First, industry experts related to the evaluation field are invited to rank the importance of indicators at different levels. If there are n indicators, the importance of the indicators is ranked from 1 to n, where 1 represents the most important indicator, and n represents This indicator has the least importance, and some indicators can be considered equally important. The indicator importance questionnaire filled out by experts is shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
<th>Y5</th>
<th>Yi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scoring1</td>
<td>Scoring2</td>
<td>Scoring3</td>
<td>Scoring4</td>
<td>Scoring5</td>
<td>Scoring(i)</td>
</tr>
<tr>
<td>2</td>
<td>Scoring1</td>
<td>Scoring2</td>
<td>Scoring3</td>
<td>Scoring4</td>
<td>Scoring5</td>
<td>Scoring(i)</td>
</tr>
<tr>
<td>3</td>
<td>Scoring1</td>
<td>Scoring2</td>
<td>Scoring3</td>
<td>Scoring4</td>
<td>Scoring5</td>
<td>Scoring(i)</td>
</tr>
<tr>
<td>i</td>
<td>Scoring1</td>
<td>Scoring2</td>
<td>Scoring3</td>
<td>Scoring4</td>
<td>Scoring5</td>
<td>Scoring(i)</td>
</tr>
</tbody>
</table>

2.2 Indicator blindness analysis

In the process of experts' evaluation of the actual situation, due to different experts' inconsistent understanding of the scoring rules and data entry errors, it will lead to abnormal values in the initial data evaluated, which will hinder the subsequent analysis process and lead to the existence of sorting. Bias and Uncertainty. The entropy theory is one of the better methods to solve the uncertainty problem, and has a wide range of applications. Therefore, the bias caused by sorting is reduced by calculating the entropy value.

First, set the number of experts participating in the questionnaire survey as k, then you need to return k questionnaire data, then the questionnaire survey table data is Y1, Y2, Y3, ..., Yk, the expert evaluation data set is Y, and Yi represents the arrangement of experts The array set is a1, a2, ..., an, where i=1, 2, 3, ..., k, and a1, a2, ..., an represent any natural number between [1, n], the smaller the value, the more important it is higher. The index ordering matrix A can be obtained from the k tables,

\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn}
\end{bmatrix}
\]

In the formula, ai represents the evaluation of the j index by the i expert.

The sorted results can be converted into quantitative results through membership, which can be expressed as:

\[
\chi(I) = -\lambda p_n(I) \ln p_n(I)
\]

In the formula, \( p_n(I) = \frac{m-I}{m-1} \), \( \lambda = \frac{1}{\ln(m-1)} \). Substitute them into the formula to get:

\[
\chi(I) = -\frac{1}{\ln(m-1)} \left( \frac{m-I}{m-1} \right) \ln \left( \frac{m-I}{m-1} \right)
\]

Assuming 1-\( \chi(I)/(m-I)/(m-1) = \mu(I) \), and dividing by \( (m-I)/(m-1) \) on both sides of the formula, we can get:

\[
\mu(I) = \frac{\ln(m-I)}{\ln(m-1)}
\]
In the formula, I refers to the qualitative ranking numerical value of an indicator evaluated by an expert. If experts rank indicators \( Y_1, Y_2, \) and \( Y_3 \) as 3, 1, and 2, it means that the \( Y_1 \) indicator is more important than other indicators. At this time, \( I=1, \) set \( m \) as the conversion parameter, and \( j \) represents the indicator number, let \( m=j+2. \)

The quantitative conversion value of \( b_{ij} \) can be obtained by bringing the qualitative ranking value of \( I \) into the formula, that is, \( b_{ij}=\mu(a_{ij}) \), the formula represents the membership value of \( I, \) and the matrix \( B=(b_{ij})_{k*n} \) is regarded as the membership matrix. Set the average awareness to \( b_j \). It represents the consistency effect of \( k \) experts on the evaluation of the indicator \( Y_j, \) which can be expressed as:

\[
b_j = \frac{b_{1j} + b_{2j} + \cdots + b_{kj}}{k}
\]

After normalization, the weight value of the indicator can be obtained. The calculation formula is as follows:

\[
\omega_j = \frac{\chi_j}{\sum_{j=1}^{k} \chi_j}
\]

In the formula, \( \omega=(\omega_1, \omega_2, ..., \omega_j) \) represents the weight vector of the index set \( Y=(Y_1, Y_2, ..., Y_j). \)

In the evaluation process, it should be noted that: in the process of economic development, the growth of high-tech industries and leading industries is in a leading position. The industries that are given priority to develop will radiate and drive the development of other industries in the city or region. Because of the obvious differences in the economic foundation and development environment of each region, the industrial development plan can refer to the industrial development of Jimo to formulate specific development strategies. Different positioning and planning of each region should follow the concept of regional development differences, and specific differential positioning may not be suitable for the differential positioning of the region.

3. MODEL APPLICABILITY ANALYSIS

The data of 100 people's evaluation results were selected, and the satisfaction analysis was carried out on the effect of the risk assessment model. The data were divided into 10 groups, and the average satisfaction of each group was obtained by calculating the mean value.

<table>
<thead>
<tr>
<th>Group</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.82</td>
</tr>
<tr>
<td>2</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>0.86</td>
</tr>
<tr>
<td>4</td>
<td>0.83</td>
</tr>
<tr>
<td>5</td>
<td>0.88</td>
</tr>
<tr>
<td>6</td>
<td>0.82</td>
</tr>
<tr>
<td>7</td>
<td>0.84</td>
</tr>
<tr>
<td>8</td>
<td>0.85</td>
</tr>
<tr>
<td>9</td>
<td>0.83</td>
</tr>
<tr>
<td>10</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Through the analysis of data satisfaction, we can find that the average satisfaction of the evaluation models proposed in this study has reached more than 80%.

Through the study of the evaluation model, we can think deeply: take sports as the core, develop "sports +", create beach sports event performance industry, beach sports fitness and leisure industry, beach sports tourism industry, beach sports training industry, beach sports culture industry, Beach sports equipment manufacturing and other industrial chains centered on beach sports. Taking the sports industry service as an effective extension, integrating high-tech elements to strengthen services, promoting the supply of sporting goods, and finally organically combining sports with manufacturing, technology, culture, tourism, etc., training, medical, outdoor product sales and design, manufacturing, marketing, logistics and other productive industrial chain of the entire sports service industry chain. With the theme of beach sports, it promotes the development of the sports tourism industry, gathers high-end professional talents, high-end
training schools, physical fitness monitoring and other high-end sports elements, and creates special service projects such as national physical fitness monitoring, sports medical treatment, and sports training. It can also undertake high-level sports preparation tasks and become a new base for the training of competitive sports talents.

At the same time, the emergence of a new economic form of "Internet + sports" can further promote the enhancement of people's constitution, people can use a series of products under the Internet to keep healthy. Body APP, fitness watch, smart bracelet and other products can more intuitively obtain the data of physical exercise and provide data basis for your fitness. "Internet + Sports". The economic model of China will become a new sports industry model, and its development will further drive sports. The development of sports will also strengthen people's interest in physical exercise. Analysis through research can be concluded that "Internet + sporting goods", "Internet + sports venues", "Internet + sports training", etc. are new sports formats and industries that are developing rapidly under the new perspective of the sports industry.

4. CONCLUSION

When the traditional operation mode and the new industrialized operation mode are gradually mature, the development of sports will add strength to the vigorous development of China's sports industry, which is a new attempt for economic development. The development and operation of beach sports is still in the exploratory stage, and a new and relatively mature operation model has not yet been formed. With the gradual maturity of policies and markets, economic development and operation are just around the corner. There are still many problems and deficiencies in the positioning of the characteristic industry of beach sports. The completion process is a process of finding problems and solving them. In the future development process, how to find a suitable path within a limited time and how to stand out among many sports towns and become a characteristic model benchmark will be the next development direction of the integration of sports and economy.

REFERENCES


Logistics data sharing method based on Federated learning

ZhiHui Wang\textsuperscript{a}, DeQian Fu\textsuperscript{a*}, Z hang jiawei \textsuperscript{b}

\textsuperscript{a}School of Information Science & Engineering, Linyi University, Linyi, China
\textsuperscript{b}Linyi Institute of Commerce and Logistics Technology Industry, Linyi, China
\* Corresponding author: fudeqian@lyu.edu.cn

ABSTRACT

In today's era of big data, the logistics supply chain generates massive amounts of data at all stages, and the privacy issues of logistics data are increasingly prominent. To efficiently utilize the logistics data of each enterprise to meet the needs of the enterprise and achieve secure data sharing, a federated learning-based logistics data sharing scheme is proposed. Using federated learning to federate multiple sources of data for modeling, the reputation value of each enterprise is stored on the blockchain and the enterprises that provide high-quality data sharing are rewarded. Finally, the effectiveness of the scheme and the impact of data quality and algorithm selection on model training is verified through simulation experiments.

Keywords: logistics data sharing, federal learning, blockchain

1. INTRODUCTION

With the widespread use of big data, the Internet, and other new generation information technology, making logistics enterprises to information and digital transformation, the logistics field, and supply chain data is explosive growth, data multi-source complex and diversified development \[1\]. According to the National Bureau of Statistics, in 2021, the national logistics volume reached 108.3 billion pieces and business revenue was 1,033.2 billion yuan, an increase of 29.9\% and 17.5\% respectively year-on-year \[2\]. In the year-on-year growth of logistics scale environment, logistics big data also gradually become an important driver of logistics wisdom, if logistics enterprises from the overall situation, use good big data analysis technology to optimize resource allocation, it can effectively achieve enterprise cost reduction and efficiency. However, as logistics data involves the privacy of users and is stored independently in the database of each enterprise's logistics information system, data security issues can arise in the process of data sharing \[3\]. In recent years, data protection issues have been taken seriously at home and abroad, and several relevant regulations have been issued, such as the "Information Security Technology - Personal Information Security Code" promulgated by China and the EU’s General Data Protection Regulation, all of which address issues related to data protection, while national awareness of personal data privacy is also gradually increasing. Therefore, there is an urgent need to address the issue of data privacy protection in the logistics sharing process, as well as the issue of computing efficiency to better support the application of intelligent technologies. Federated Learning (FL) is a machine learning framework proposed by Google Labs to address privacy issues in data sharing \[4\]. Federated learning allows for efficient machine learning by federating data from multiple parties while maintaining data security. The paper \[5\] uses federated learning combined with blockchain technology in an IoT scenario to solve the problem of secure sharing of industrial data. The literature \[6\] combines federated learning with blockchain technology for securing the privacy of healthcare data sharing.

In summary, federal learning techniques are a natural fit for the privacy issues in logistics data sharing. Therefore, this paper introduces federal learning techniques to the logistics domain to address the privacy issues in logistics data sharing.
In this paper, a logistics data sharing scheme is constructed to federate the logistics data scattered in various enterprises for modeling and providing intelligent decisions for logistics enterprises. Finally, the effectiveness of the scheme is verified by conducting simulation experiments with logistics data.

## 2. SYSTEM MODEL

To achieve logistics data sharing, a three-layer system model architecture is set out in this paper, namely the demand layer, the federal learning layer, and the blockchain layer, as shown in Figure 1.

### 2.1 Demand layer

The requirements layer is made up of multiple model requirements issued by the data requestor and metadata for data that can be shared by the enterprise. A model requirement contains the data provider's selection requirements (e.g., enterprise reputation values), model objectives (e.g., return of optimal routes), model convergence values, and metadata for locally available data in the requirement model. A data requester is an enterprise party that needs to integrate a large amount of data for analysis to meet its requirements.

### 2.2 Federal Learning Layer

In the federation learning layer, the demand layer-based enterprise needs to feed data from multiple parties for model training to protect data privacy. Each data provider needs to satisfy the data requester's rules, using reputation values as an evaluation metric for reliable data providers. Each enterprise then iteratively trains the shared global model generated by the server based on its local logistics big data. All participants then encrypt and upload their local model parameters to the server to update the global model. Finally, the training process is repeated until the global model accuracy reaches the pre-defined model convergence value in the requirements. The model is passed into the blockchain layer, and the data requester performs a quality assessment of the updates to each participant's local model, generating reputation values and contribution value ratings for each participant.
2.3 Blockchain layer

The blockchain layer is responsible for recording and updating information such as the client's reputation value to provide a more secure and trusted data-sharing environment. At the end of a training mission, the reputation value of companies is updated via the blockchain, and companies are rewarded for their active participation in data sharing using an accurate payment mechanism.

3. SPECIFIC PROCESSES

The model sharing solution is divided into the requirement initiation process and the cryptographic model training process.

3.1 Requirements release process

The requirement initiation phase is when a company that has been authenticated in the system posts a modeling requirement. The federal learning layer receives the modeling requirement and responds with a signal of whether it was posted correctly or incorrectly. The requirements include metadata for locally available training data $md_L$, partner reputation values $rv_p$, model goals $M_{obj}$, model convergence values $M_{con_v}$, specified partner flags $sign_p$ and other parameters. The process is shown in Figure 3.

![Figure 2. The requirements release process.](image-url)
the modeling request and sends the confirmation back to the Federated Learning Layer, which confirms that all enterprises have confirmed their cooperation in the modeling and returns a successful confirmation message to the data requestor.

3.2 Enterprise data sharing processes

After the requirements are published, the federated learning layer is used to train the encrypted model, ensuring data security and data sharing. The federated learning training process is described in Algorithm 1. Firstly, the system is initialized and the initial global model weight parameters are generated at $W^i$ and distributed to all participants in the modeling. The local model accuracy or training rounds are trained before the preset values are reached, and the local parameters are aggregated to the federated learning model when the training is completed. If the model does not reach the preset model target, the aggregated parameters are sent to a subset of logistics companies for the next round of training until the preset model target is reached.

Algorithm 1: Federal training algorithm

**Inputs** Subset of logistics firms $m = 1, 2, ..., M$, Total number of modeling partners $M$, Learning rate $\eta$, Initialized global weight parameters $W^\circ$, Training rounds $epochsE$.

**Output** Global model parameters

1. Initialize $W^\circ$; //system initialization

2. $W^m \leftarrow W^\circ$; // weight parameters are issued to a subset of logistics companies

3. WHILE(epoch or accuracy<preset value)DO

4. FOR $k \in \{0,...,local\_train\_data/bath\_size\}$ DO

5. $w^{f}_{m} \leftarrow w^{f-1}_{m} - \eta \nabla f(w^{f-1}_{m})$; // updated weights for companies participating in training

6. END FOR

7. FOR each enterprise $j \in m$ DO

8. $w \leftarrow \sum_{i=1}^{M} \frac{n_{i}}{n} w^{i}_{r+1}$; // Aggregation parameters

9. END FOR

10. END WHILE
4. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, we set the logistics scenario to achieve the logistics delivery time prediction requirement, and evaluate the effectiveness of the solution by building the SecureBoost model and linear regression model respectively through the e-commerce and logistics dataset situation shown in Table 1. Each enterprise sample will be aligned data samples locally for feature engineering processing, the experimental training process, and the overall data set randomly divided 70% as the training set, and 30% as the test set. The experimental environment was configured as follows: system using CentOS 7.6.1810; CPUs were two 4-core Intel(R) Core (TM) i5-9300HF CPUs @ 2.40GHz; 32G RAM; Python version 3.8.

Table 1. Information on the validity assessment dataset.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>E-commerce data</th>
<th>Logistics data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data details</td>
<td>Data set on customer orders on the e-commerce side</td>
<td>Data sets relating to the logistics side of the delivery note</td>
</tr>
<tr>
<td>Sample size</td>
<td>24000</td>
<td>24000</td>
</tr>
<tr>
<td>Number of features</td>
<td>16</td>
<td>37</td>
</tr>
<tr>
<td>Availability of y-tags</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Aligned sample size</td>
<td>23689</td>
<td>23689</td>
</tr>
</tbody>
</table>

Figures 3-4 show a comparison of the loss values of the SecureBoost model and the linear regression model before and after adding the feature engineering process respectively, and it is evident that model training with the feature engineered data can reduce the model loss better. Figure 5 shows a comparison of the federal learning loss values for the two models, showing that the model chosen for different needs affects the model loss values. A comparison of the model training time for two different encryption algorithms on the SecureBoost model is shown in Figure 6, indicating that the choice of encryption algorithm can improve the training efficiency of the model.
During the training process, the data does not leave the local area and the training is completed by encrypting the model parameters for transmission. For the parameter transfer, the model parameters of each logistics company $\mathbf{W}^M$ need to be encrypted locally using an encryption algorithm before being uploaded and subsequently sent to the federal learning layer, ensuring the security of the data.

5. SUMMARY

To address the data privacy issues in the logistics data sharing process, this paper constructs a system model architecture based on federated learning. The architecture allows the data sharing process to be realised through federated learning techniques, using blockchain to store enterprise reputation and rewarding organisations that provide high quality data sharing, with joint modelling under the premise of ensuring data security. Finally, simulation experiments are conducted with logistics data to verify the effectiveness of the solution, and comparative experiments are done for application scenarios, and the results show the impact of model and algorithm selection on the performance of the training model.

REFERENCES

Application and Prospect of Big Data Technology in Medical Image Enhancement

Yucheng Huang1a, Jiyun Chen1b, Qiang Wang1c, Chun Chen1d, Haifeng Liu1e, Zijun Mo1f, Yimin Lin1g*, Tianyuan Zhu1h*

1Zhuhai College of Science and Technology, Zhuhai 519040, China

Abstract

With the continuous development of the information age, big data technology has emerged in the context of a large number of information-derived clusters, and the advantages of its application in image processing have become increasingly prominent. This paper takes medical image enhancement technology as the landing point, based on the current situation and development trend of big data technology in image processing applications, analyzes its application principles, compares the effects of different medical image enhancement algorithms, discusses the aspects of big data technology that can be improved such as image transformation, restoration, compression, etc., focuses on the application of medical image enhancement and proposes countermeasures for the application of big data technology in the future, so that researchers can fully exploit the value of big data and promote the application of big data technology especially for image processing. So that researchers can fully exploit the value of big data and promote new developments in image processing technology especially for medical applications.

Keywords-big data technology; image processing; medical image enhancement; application

1. INTRODUCTION

Big data technology is an advanced form of technology that can scientifically analyze and effectively process massive amounts of data [1]. The core technologies include the collection, management, preprocessing, storage, and analysis of information. The advantage of big data technology is that it can break through the boundaries of data type and style diversity, and has a very fast computing speed to analyze and dig out valuable information from a large variety and quantity of data, and apply it precisely to the related fields. At present, as the level of socio-economic development and the level of scientific and technological development continue to improve, the demand for big data technology in various industries has increased, and the application and development of big data technology have shown a trend of diversification, showing significant advantages in the development and innovation of information management, finance, medical, manufacturing, scientific research, education and other fields [1].

Among them, the application of big data technology in image processing usually combines the use of hardware facilities and related software to process images and to develop deeper in the direction of image recognition to obtain the data information reflected in images more quickly and accurately. Big data technology can refine the image processing, ensure the clarity of the image in the process of efficient image transmission, and simulate the image by using two-dimensional data sets [2]. This can be used to increase the image pixels, and also to compress and segment the image to meet the actual needs of the work.

The research concept of this paper on the application of big data in medical image enhancement mainly involves (1) closely combining the development trend of big data technology and analyzing the role it can play in image processing one by one, (2) because medical images have the characteristics of ambiguity and inhomogeneity compared with ordinary images [3], it is necessary to consider not only the denoising effect but also the ambiguity of images when analyzing the image enhancement effect, (3) the mammogram X-ray images with high quality and more features are selected as the experimental data, and the histogram equalization algorithm is evaluated to have better performance in processing the effect of texture and details of this type of images by comparing several perspectives such as image features and image classification accuracy, (4) the research of the thesis is closely related to the actual needs in the medical field at present, and driven by these actual needs, the research and development related to artificial intelligence activities, the production and application
of AI embedded devices, and the construction of intelligent platforms and infrastructures, there is a huge driving force [2]. In the future, the allocation of limited medical resources will remain the main issue. In this paper, we will propose countermeasures for the application of big data technologies in the field of medical image processing to promote new developments in artificial intelligence technologies and smart medical projects.

2. Medical Image Enhancement Technology And Current Status

Image enhancement is a basic and important technique in the field of image processing to highlight some difficult information in the image according to the actual needs while minimizing the interference information [3]. It is a fundamental and important technique in the field of image processing. The main purpose of image enhancement is to make the image easy to understand and analyze. Image enhancement seeks to extract the information needed in the image according to the actual application, improve the clarity of the image, and enhance the readability and intelligibility of the image, rather than requiring the image to be the same as the real one [4].

The purpose of medical image enhancement is to improve the recognition of the required information in the image. Although medical image enhancement techniques cannot improve the quality of the information in the image, they can make the information in the image clear. Medical image enhancement mainly includes contrast enhancement, noise suppression, edge sharpening, pseudo-color processing and so on. Medical image data enhancement is processed in two ways. One is single-sample data enhancement and the other is multi-sample data enhancement. Single-sample data enhancement refers to augmentation processing based on the existing data with a single image sample as the center, which can be simply described as overall image level processing and image pixel level processing. It can be simply described as the overall image level processing and image pixel-level processing. Multisample data augmentation is distinguished from single-sample data augmentation in that multisample data augmentation is obtained by processing multiple samples to produce new samples [5]. The problem of unbalanced number of samples from different medical data and insufficient number of samples can be solved by multisample data augmentation.

Some traditional medical image enhancement techniques have been used for a long time, although the traditional methods are simpler and faster, the application effect has not been very good, because it can not take into account the before and after messages in the image. In recent years, with the deep development of Internet technology, the technical means of image processing and recognition based on the neural network have made a great breakthrough. At this stage, with the continuous application of convolutional neural networks in many low-level image processing tasks, medical image enhancement techniques have achieved important technical breakthroughs in image super recognition, deblurring, defogging, noise removal, and image enhancement. But this supervised learning has certain drawbacks, requiring human identification of each image, and the data set is often quite small. To solve these problems, self-supervised or weakly supervised methods need to be researched and applied continuously. At the same time, due to the increasing complexity of image enhancement computation, there is an increasing demand for the development of hardware to implement enhancement computation.

3. The Function and Characteristics of Big Data Technology

Along with the continuous development of information technology, big data technology is increasingly used in various industries. Comprehensive and efficient big data technology plays a key role in image processing. Big data technology has a strong image reproduction function, which can be used to maintain the original appearance of the image and achieve efficient transmission without degrading the image quality in the process of transmission. Big data has diverse characteristics and its data sources are very rich, so it can reflect the true size of objective things, which is also very helpful for the development of various technologies [6][7].

3.1. Application of big data technology in image processing

3.1.1. Image transformation function

A very critical step in image processing is image exchange. In the process of image exchange, as the image display increases, the rate of change of the image will also increase, thus affecting the image exchange. The image exchange process can be controlled by big data technology to reduce the amount of computation and improve the quality and speed of image processing [8].

3.1.2. Image restoration and enhancement

In the process of image processing, big data can be used to restore and enhance the image.
In the process of generating and transmitting images, images are easily disturbed and affected by noise, such as weather, equipment, environment and so on, which will affect the quality of images. Big data technology can comprehensively analyze and process a large amount of image data, so as to improve the clarity of the image. It can also highlight the key parts of the image by removing noise, and truly restore the image [16].

3.1.3. Encoding compression

Big data technology helps to save the time of image transmission and the storage of images because the coding and compression technology can reduce the amount of image data in the process of image processing. At the same time, the coding and compression technology can save the image processing results truly, clearly and efficiently, which is more convenient for subsequent practical application [11].

3.1.4. Image description

At present, the main way of image processing is the two-dimensional shape description method including image area description and image boundary description. During the development of big data technology, image processing has changed from purely two-dimensional to three-dimensional, and new methods such as image surface description, image volume description, and generalized cylinder description have gradually emerged [11].

3.2. Application of big data technology in medical image enhancement

In clinical medicine, it is a very common technique for digital X-ray to assist doctors in clinical diagnosis, but in the process of imaging, digital X-ray can not focus with a lens, which leads to some small blur in the medical image, and the structure of the lesion area is similar to that of human organs and tissues [11], so the contrast between the normal area and the lesion area is relatively low. Therefore, there will be some missed diagnosis and misdiagnosis in clinical image analysis.

Digital X-ray medical images generally use histogram equalization enhancement technology, but the traditional histogram equalization enhancement technology has some problems, which will cause some details of the image to disappear after enhancement transformation; in addition, there will be a phenomenon that the enhancement intensity is not well controlled and excessive enhancement leads to unnatural images. By combining big data with the traditional histogram equalization image enhancement technology, and making use of the advantage of big data's flexible processing, it can improve the flexibility of image processing and reduce the occurrence of image detail disappearance and excessive enhancement. It makes the lesion area of X-ray image more accurate and clear.

Mammograms are one of the most common techniques used to assist physicians in screening patients for breast cancer, generally enhanced using grayscale histogram equalization, but the grayscale of this technique is basically uniform, which makes some grayscale in the high grayscale area and low grayscale area which is difficult to be observed by the naked eye. After mammogram enhancement, the high gray area mainly falls in the middle, while the low gray area converges on the edge. In this case, if the histogram operation is used to enhance the mamogram, the contrast is still not obvious. However, there is some important information such as skin on the edge of the mamogram, which is also essential for doctors to diagnose. Therefore, in this case, some scholars have proposed a method of mammogram enhancement based on background homogenization: after background homogenization, improve the brightness of the edge region on the premise of ensuring the overall brightness of the breast image. In this case, using the traditional histogram equalization enhancement technology, we can see that the contrast is obviously enhanced after imaging, and most of the gray values are distributed in the low and medium grayscale. Then the local contrast enhancement algorithm is explored and simulated, and the adaptive enhancement algorithm based on local mean and local variance is applied, which avoids the excessive brightness of the image and makes the details of the mammogram more obvious and the edge information of the image more clear.

The comparison before and after adopting the adaptive enhancement algorithm based on local mean and local variance is shown in Fig. 1 [13].
Through the image, it is obvious that the edge of the breast image has been enhanced, but the medical image is huge and comes from different sources, so it is obviously inefficient to use this kind of mammogram enhancement based on background homogenization. Therefore, the application of big data technology in this technology can process image data systematically and effectively. In the process of image reproduction or transmission, the image clarity may be reduced. Big data technology can reproduce the image, and its application in this technology can highlight the clarity of the breast area, make the image details more detailed, help health care workers to diagnose to a greater extent, and improve the diagnostic accuracy.

3.3. Problems of big data technology

Under the background of big data, the amount of data is increasing, and the computing power of hardware systems and devices that store data need to be upgraded. Compared with before the introduction of big data technology, the amount of data today is very large, and the corresponding way of information storage will also change greatly. Increasing the requirement of hardware may increase the cost of image processing, and combined with cloud computing technology, this problem can be solved effectively. Combined with cloud computing technology, users can choose to rent cloud servers to reduce costs, but also promote the development of graphics processing technology. Therefore, improving the storage mode of data to make it more reasonable and efficient is one of the ways to improve the image processing technology under the background of big data.

In the process of image data collection, the data flow will be very large, so it is easy to have data security problems, and thus their security needs to be improved both in terms of hardware and software aspects. As for the image processing, a perfect system to ensure data security should be built, and the collected image data should be deeply detected to ensure the privacy and security of image data.

4. Commonly Used Medical Image Enhancement Algorithms

4.1. Contrast stretching

Medical images may be acquired with a lack of dynamic range in the image sensor due to insufficient light. The idea of this method is to enhance the dynamic range of the image grayscale during image processing.

The formula is as follows.

\[
I(x, y) = \frac{I(x, y) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} (\max - \min) + I_{\text{min}}
\]

(1)

The formula \(I_{\text{max}}\) is the maximum gray level in the image, \(I_{\text{min}}\) is the minimum gray level in the image, \(\max\) is the maximum value of the gray range to be stretched, \(\min\) is the minimum value of the gray range to be stretched.

This method can be applied to most medical images with good applicability. As in Fig. 3(a) and (b), the images before and after contrast stretching are performed.
4.2. Grayscale slicing

In medical images, the grayscale part of the image is a relatively high proportion, so it is possible to emphasize specific grayscale values and suppress other unnecessary grayscale values to achieve the effect of image enhancement. This technique is used more in MRI images, CT images, etc. Fig.3(c) show that the area with the gray value of [100,180] in the figure is emphasized and the other areas are suppressed.

4.3. Gaussian filtering

Gaussian filter is a kind of smoothing filter, which is mainly used to suppress the noise that obeys normal distribution. In medical images, the acquired images have noise, and most of these noises obey normal distribution. Therefore, Gaussian filtering is also commonly used for medical image processing.

The Gaussian filtering is specified as follows:

1. Scan using a template for each pixel in the image
2. Calculate the weighted average gray value of the pixels in the field by the template
3. Use the obtained field pixel-weighted average grayscale value to replace the value of the center pixel point of the template

Here the template obeys a two-dimensional Gaussian distribution: the

\[
f(x, y) = \frac{1}{(2\pi \sigma^2)^{1/2}} e^{-(x-\mu_x)^2+(y-\mu_y)^2)/(2\sigma^2)}
\]  

(2)

where \( \frac{1}{(2\pi \sigma^2)^{1/2}} \) is a constant, what is actually calculated is \( e^{-(x-\mu_x)^2+(y-\mu_y)^2)/(2\sigma^2)} \), where \( \sigma \) is the standard deviation, \((x,y)\) is the coordinate of any point inside the mask, and \((\mu_x,\mu_y)\) is the coordinate of the center point inside the mask, which can be an integer during processing.

Since Gaussian filtering takes into account the pixels in the field and the fact that template used obeys a two-dimensional Gaussian distribution. So there is less blurring of the image, thus better maintaining the overall detail of the image.

4.4. Image inversion

Image inversion is suitable for images where the background is dark and the image subject is off-white. This is especially true when the black area is predominant in size. Some medical images such as (CT, MRI) have a mostly black background and the image itself is rendered in a predominantly grayish-white color.

The inversion formula is as follows.

\[
S = L - 1 - r
\]  

(3)

Where \( S \) is the gray value of the inverted target pixel point of the image, \( r \) is the gray value of the corresponding point of the original image, and \( L \) is the number of gray levels. As in Fig.3(d), the images after image inversion are performed.

4.5. Histogram equalization

Histogram equalization is an image enhancement algorithm that changes the gray level of each pixel in an image by changing the gray histogram of the original image. When medical images are acquired, there may be underexposure and overexposure, making the grayscale distribution in more concentrated intervals, causing problems such as unclear images and features not being highlighted. The steps of the rectangular equalization are.

1. Get the grayscale of the image to be processed.
2. Calculate the cumulative distribution function of the grayscale histogram.
3. Obtain the mapping relationship between input and output based on the cumulative distribution function and the equalization principle of the histogram.
4. Finally, the output result of the image is obtained according to the mapping relationship.

Where the mapping method is:
\[ S_k = \sum_{j=0}^{n_j} \frac{n_j}{n} \quad (4) \]

Where \( k = 0, 1, 2, \ldots, n-1 \).

Where \( S_k \) is the value of the current gray level mapped by the cumulative distribution function, and \( n_j \) is the number of pixels of the current gray level, \( n \) is the number of pixels of the current gray level, and \( l \) is the total number of gray levels in the image.

Fig.2 show the gray histogram of the image before and after histogram equalization and Fig.3 (e) show the image after histogram equalization.

At the same time, histogram equalization also has disadvantages. This method is suitable for images that are brighter overall, or darker overall. If the image is not filtered, then it may add disturbing information and reduce the gray level of a feature or detail, making the processed image features even less obvious than the original image or even disappear.

4.6. Laplace Image Sharpening

In medical images, some errors in the acquisition process may cause the image edges to be inconspicuous and lead to the problem of inconspicuous image features. Image sharpening is to compensate for the edge details of the image, which outlines the image more obvious and makes the image clear. Image sharpening can be broadly divided into two methods: spatial domain processing and frequency domain processing. The following is a simple 3*3 filter for sharpening.

\[
\begin{bmatrix}
0 & -1 & 0 \\
-1 & 5 & -1 \\
0 & -1 & 0
\end{bmatrix}
\quad (5)
\]

At the same time, image sharpening is an operation on the whole image, it is also thought that if the image exists earlier, then the noise will also be magnified by the image sharpening, so the original image needs to be smoothed and other operations to remove the noise before sharpening. Fig.3 (f) show the image after Laplace image sharpening.

5. Experiments and Analysis of Results

5.1. Experimental data and pretreatment

In order to accurately indicate the location of human lesions and pathological types and to make an objective evaluation of study results, standard digital X-ray images are essential. Currently, only a few developed countries have established a standard database of disease X-ray images, thus some images from the more commonly used high-resolution mammographic image analysis society (MIAS) data set were selected as experimental data in this paper. MIAS includes a variety of sign images, with 161 cases The 322 digitized MLO images were reduced, cropped and rotated to obtain 1024×1024 pixel images, which contained 208 normal images and 114 abnormal images (63 with benign masses and 51 with malignant masses), and the abnormal areas of the images had the annotated radius length and center coordinates of the smallest outer circle containing the lesion area.

In this experiment, 90 normal, 45 benign and 45 malignant lateral mammograms were selected according to the annotation information, the abnormal area of 128×128 pixels was segmented and the texture features were extracted from the samples processed by different image enhancement algorithms, the image quality was evaluated and the feature values were standardized by Z-Score method. The processed data conform to the standard normal distribution.
5.2. Subjective evaluation

Figure 3 shows the experimental results of X-ray image enhancement. It can be seen that the X-ray image processed by the image enhancement algorithm studied in this paper has a more uniform grayscale distribution than the original image, and the contrast is sharper and retains the details of the original image, and no artifacts are generated.

![Image showing effect comparison of breast X-ray image enhancement methods](image)

5.3. Objective evaluation

The evaluation of image quality mainly focuses on image fidelity and image intelligibility. Image quality directly depends on the influence of various factors such as optical performance of imaging equipment, image contrast, and instrument noise, etc. Quality evaluation can provide monitoring means for various aspects of image acquisition and processing. At present, people still do not fully understand the characteristics of human vision, especially the psychological characteristics of human eye vision is still difficult to find quantitative description methods, so the image quality evaluation still needs to be studied in depth [14].

In order to verify the objectivity of the algorithm mentioned in this paper, as shown in Table 1, three evaluation metrics of information entropy, contrast and variance for reference-free images are used here to evaluate the image quality [15].

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information entropy</td>
<td>$f_1 = \sum P(i, j) \ast [-\ln P(i, j)]$</td>
<td>It is used to measure the randomness of the grayscale distribution. If the grayscale is randomly distributed in the image, the theoretical expected entropy will be high. If pixels have similar grayscale values, the contrast will be lower.</td>
</tr>
<tr>
<td>Contrast</td>
<td>$f_2 = \sum P(i, j) \ast (i - j)^2$</td>
<td>It is used to display the average value of grayscale in an image. If the sum of the image grayscale values is higher, the sum of the average values is larger. It is used to measure the unfolding scale of the gray level distribution. If the image gray distribution is large, the variance will be larger.</td>
</tr>
<tr>
<td>Grayscale mean</td>
<td>$f_3 = \sum i \ast P(i, j) = \mu_i$</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>$f_4 = \sum P(i, j) \ast (i - \mu_i)$</td>
<td></td>
</tr>
</tbody>
</table>

Table.1 Image quality evaluation indicators

Information entropy indicates the degree of information contained in the image, and the larger the value, the more information the image contains and the more detailed information the image has. The larger the contrast value, the more detailed
information the image can show and the better the quality of the image. The variance measures the scale of the unfolding of the gray level distribution, and the larger its value, the more significant the image variation.

Table 2 Comparison of objective criteria for different algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Objective evaluation metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information entropy</td>
</tr>
<tr>
<td>Original image</td>
<td>2.86</td>
</tr>
<tr>
<td>Contrast stretching</td>
<td>4.38</td>
</tr>
<tr>
<td>Grayscale slicing</td>
<td>5.36</td>
</tr>
<tr>
<td>Gaussian filtering</td>
<td>4.56</td>
</tr>
<tr>
<td>Image inversion</td>
<td>5.26</td>
</tr>
<tr>
<td>Histogram equalization</td>
<td>5.58</td>
</tr>
<tr>
<td>Laplace image sharpening</td>
<td>4.96</td>
</tr>
</tbody>
</table>

As can be seen from Table 2, the contrast after processing by the grayscale slicing algorithm is relatively large, indicating that the contrast of the image can be effectively improved. The information entropy and variance after processing are the largest, indicating that the algorithm can see more detailed information. In summary, the histogram equalization algorithm for X-ray image enhancement, in terms of subjective vision, can see more detailed information and improve the contrast of the image compared with other algorithms, and in objective evaluation, its evaluation indexes are generally better than other algorithms. In addition, some researchers have proposed several improved algorithms for coronary angiography image enhancement [16] as shown in Table 3, which can further remove the image background noise and make the imaging finer, it should be noted that the effect of image enhancement algorithms is also affected by the actual application scenario.

Table 3 Summary of the advantages and limitations of several improvement options

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hessian matrix based image enhancement improvement algorithm</td>
<td>Removing many complex backgrounds.</td>
<td>Inevitably some noise will be mixed in.</td>
</tr>
<tr>
<td>Improved algorithm for image enhancement based on Gaussian first-order derivative matched filter</td>
<td>Removal of noise is effective, and tissue structures disturbed by light and other factors can be found.</td>
<td>High error rate of detection.</td>
</tr>
<tr>
<td>Improved bilateral filter image enhancement algorithm</td>
<td>Higher level of refinement for micro lesion enhancement processing.</td>
<td>High time complexity and high hardware requirements such as image printing equipment.</td>
</tr>
</tbody>
</table>

In order to further evaluate the performance of several image enhancement algorithms with better results in Table 2 in practical applications, we composed vectors of extracted features as the input of the support vector machine classifier for benign and malignant classification. The SVM algorithm model has excellent characteristics in processing small sample data of breast cancer medical images in this experiment, and in this paper, under the premise of keeping the training set and parameters are selected consistently, we used comparative experiments with 5-fold cross-validation, respectively, and the experimental data are shown in Table 4.

Table 4 Cross-validation results

<table>
<thead>
<tr>
<th>Crossover Round</th>
<th>Grayscale slicing Accuracy(%)</th>
<th>Image inversion Accuracy(%)</th>
<th>Histogram equalization Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72.22</td>
<td>72.22</td>
<td>83.33</td>
</tr>
<tr>
<td>2</td>
<td>83.33</td>
<td>72.22</td>
<td>72.22</td>
</tr>
<tr>
<td>3</td>
<td>66.67</td>
<td>61.11</td>
<td>77.78</td>
</tr>
<tr>
<td>4</td>
<td>72.22</td>
<td>66.67</td>
<td>83.33</td>
</tr>
<tr>
<td>5</td>
<td>83.33</td>
<td>72.22</td>
<td>77.78</td>
</tr>
<tr>
<td>Mean</td>
<td>75.56</td>
<td>68.89</td>
<td>78.89</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>6.66</td>
<td>4.44</td>
<td>4.16</td>
</tr>
</tbody>
</table>

From the results in the above table, it can be seen that the support vector machine has different classification effects on the breast cancer image data set processed by different image enhancement algorithms. Compared with the other two algorithms, using histogram equalization to process the image data makes the model have higher prediction accuracy, but the robustness of the model for this experimental data set has not yet reached the strict medical diagnosis requirements, and further validation and optimization are still needed.
6. Outlook

6.1. Strengthen technology research and development

In the application of big data technology to the field of image processing, there are still problems such as data security, high cost, and functional limitations of storage devices that need to be solved. Relevant technical departments should set up research projects related to the field of big data image processing and organize technical teams to continuously improve the image processing technology, such as building a perfect data security system to ensure that the privacy and security of image data will not be leaked and stolen in the process of image processing; upgrading the function of the facilities and systems for storing data to achieve a larger storage capacity and higher data processing speed. The research will be carried out in the following areas Through a series of studies, we will promote the maturity of big data-based image processing methods and achieve universal application and cost reduction in the future.

6.2. Pay attention to cultivating professional talents

The application of big data technology in image processing technology needs the support of talents with high professional skills. Therefore, it is necessary to strengthen the cultivation of talents in the field of big data technology, and it is also possible to jointly cultivate talents from other disciplines and big data professionals, so as to provide more comprehensive developmental talents for the development of different fields of application of big data technology. Universities can offer professional courses on big data image processing to make more students understand and master image processing technology. The government invests funds in universities and related research institutions, which on the one hand is helpful to absorb professional talents, and on the other hand can be used for upgrading and R&D of research equipment.

7. Conclusion

Big data technology has greater advantages in image processing, which can provide scientific reference for image processing work, improve image processing efficiency, and ensure scientific and accurate image data analysis. Especially in the processing of medical images, it is helpful to accurately diagnose the actual condition and provide information support for the smooth development of the treatment work. With the development of the information age, the scope of application of big data technology is constantly expanding, and its influence is also expanding. Therefore, we should pay more attention to the problems and hidden dangers in the process of its development, strengthen technological research and strengthen the support of human resources. Make big data technology better serve social development.

This paper evaluates the performance of image enhancement algorithms by extracting image texture features and classifying benign and malignant mammogram images, verifies the superiority of histogram equalization for processing such images and points out the limitations of its relatively average gray scale and weak contrast, summarizes the related improvement methods and presents the prospect of data security and equipment development for future big data applications in the medical field, especially image enhancement.

References


FSC-based Data Fusion Sensitivity Assessment Scheme

Caiwen Wang*, Rui Liu*, Ke An*
(*China Aerospace Academy of Systems Science and Engineering, Beijing 100048)
*Corresponding author: clevenwang@163.com

Abstract

In recent years, with the large amount of data generated, data fusion and sharing have become more and more important. However, the expansion of the scope of data fusion and sharing has led to the increasing problem of data leakage risk. In this paper, we improve the FSC fusion algorithm to evaluate the sensitivity change during data fusion from three aspects: data set, data items and data values, so that the fused data sensitivity is more accurate, and later the fused data is anonymized for sharing. The algorithm reduces information loss and provides better privacy protection.

Key words: Data Fusion, Privacy Data Security, Sensitivity, K-anonymity

1. Introduction

In recent years, with the gradual expansion of data sharing, data based on original applications can no longer meet the needs of data decision analysis. The value of data has received unprecedented attention. Data concentration, integration and sharing are becoming more and more frequent, and data fusion is becoming more and more important. However, the expansion of data fusion and sharing has led to an increasingly prominent problem of data leakage risk. In particular, if sensitive data is leaked, it will have a large negative impact on society and individuals. In recent years, the number of data privacy leaks has increased exponentially, and the losses caused by the incidents have been expanding, and the protection of privacy information has become a huge problem in the regulation of cyberspace security in countries around the world [1].

This paper investigates the protection scheme of sensitive information in the process of data fusion and sharing. Data originates from different enterprise systems and contains different sets of attributes of individuals. Now multiple parties need to fuse data to make decisions together. However, there are competitive relationships between enterprises and they do not want to share the attribute sets among themselves. So how to achieve information sharing while effectively protecting sensitive information from leakage after fusion is particularly important. In this paper, we use the idea of hierarchical privacy protection, starting from three dimensions: data set, data items, and data values. The relationship between the source data, the selected attribute set and the fused attributes is considered comprehensively to finally determine the data sensitivity. While achieving privacy protection, data availability is ensured as much as possible. It accelerates the opening and sharing of big data and promotes the rapid development of data industry.

Data publishing that directly removes or suppresses sensitive attributes and retains only identifier attributes can break the correlation between sensitive and quasi-identifier attributes. It may lead to inaccurate results of subsequent data analysis. Therefore sensitive attributes usually need to be retained. And anonymous privacy-preserving data publishing models are needed to resist attribute disclosure and prevent sensitive information leakage.

The k-anonymity model was first proposed by Sweeney et al. in 2002 at the PODS (ACM on Principles of Database Systems) conference [2]. It can resist the information leakage caused by link attacks. However, it does not constrain the sensitive attributes, so some studies have introduced the concept of sensitivity. It refers to the corresponding value after doing the numericalization of the sensitivity level. Accordingly, the anonymity model based on sensitivity measure is proposed. For example, Machanavajjhala A [3] set different L-diversity parameters for each sensitive attribute value according to its different sensitivity level customization, and optimize the model results accordingly. Later, during the research of privacy anonymity protection technology, specialized research on sensitive information metrics emerged [8,9]. The literature [4] proposed a privacy model oriented to the hierarchy of sensitive values. The personalized privacy protection of sensitive values is achieved by setting the data attributes and the sensitivity of the attribute values through...
the data provider's requirements for the degree of anonymity of the data.

FSC is a sensitivity assessment algorithm for fused data based on data hierarchy. The algorithm assesses the dynamic sensitivity of data changes in the process of big data fusion and sharing, and assigns a reasonable sensitivity to the resultant dataset. However, the FSC algorithm only considers the relationship on structural features when evaluating data sensitivity. It cannot discover the hidden association between data, which leads to low fusion sensitivity. In this paper, we explore the influence of data feature relationships on the sensitivity of fused data based on the FSC algorithm.

3. Scheme

For the problem of sensitivity level change, this paper proposes a sensitivity evaluation model. As shown in the figure, the dynamic sensitivity assessment of data in multiple data sources with different sensitivity levels in the process of data fusion and sharing is carried out from three dimensions: data set, data items and data values. The impact of the sensitivity level of data items before and after fusion is integrated, and a reasonable sensitivity level is assigned to the resultant dataset of fusion and sharing[10]. The sensitivity of the result dataset is quantitatively evaluated to achieve effective access control of the fused and shared dataset and to avoid leakage of important data contents to the maximum extent.

Figure 1. Data sensitivity assessment dimensions

In terms of dataset dimensions, the data comes from different enterprises. Different enterprises have different attributes of user information, and there are some attributes, such as disease information, asset information, which are more sensitive compared with academic information. Therefore, their sensitivities should be different. In terms of data item dimension, this paper analyzes data item in terms of both data features and structural features. Structural features are rated by experts for the sensitivity of the data items. Data features determine the correlation between different data items by spearman correlation coefficients, from which a more accurate sensitivity is obtained. In the data value dimension, this paper ensures the security of data sharing by anonymizing the fused data attribute values.

3.1 Data Set Sensitivity

In order to better describe and reveal the sensitive data situation in the dataset, the sensitive attributes are classified according to the difference of sensitivity level among individuals. Sensitivity F is used to quantitatively reflect the sensitivity level of the data. The data sensitivity level is divided into levels as shown in Table 1. The sensitivity level is expressed as a sequence of numbers 0, 1, 2, and 3. Where 0 indicates the lowest sensitivity level, and the larger the sensitivity value is, the higher the sensitivity level of the data. Since this paper assesses the sensitivity of data by sensitivity factor, the assessment results are kept in two decimal places.

<table>
<thead>
<tr>
<th>Sensitivity level</th>
<th>Value setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely sensitive</td>
<td>3</td>
</tr>
<tr>
<td>more sensitive</td>
<td>2</td>
</tr>
<tr>
<td>moderate</td>
<td>1</td>
</tr>
<tr>
<td>Insensitive</td>
<td>0</td>
</tr>
</tbody>
</table>

When data from multiple data sources are fused and shared, the degree of their influence on the sensitivity of the resultant data set varies due to the different sensitivities of each data source. This algorithm uses the sensitivity adjustment coefficient $\alpha$ to reflect the weight of each data source's sensitivity. The larger the sensitivity adjustment coefficient is, the higher the importance of the data source and the greater the impact on the result sensitivity. From the
dataset dimension, the sensitivity of the datasets from different enterprises is rated.

Definition 1: (Sensitivity factor $\alpha$) Starting from the dataset dimension, there are $m$ data sources, and defining the data source as the collection data object $E_1, E_2, \ldots, E_n$. Suppose the total number of attributes contained in the collection data object $E_k$ is $m$. $E_k$ has $P_1, P_2, \ldots, P_m$ attributes and the corresponding sensitivity factors are $\alpha_1, \alpha_2, \ldots, \alpha_m$. Then the sensitivity factor satisfies $\sum_{i=1}^{m} \alpha_i = 1$, $0 < \alpha_i < 1$. The higher the degree of privacy, the higher the value of $\alpha_i$.

### 3.2 Data Item Sensitivity

The next division is from the data item dimension. The traditional data item sensitivity level classification method is mainly based on the experience of data clerks. The results of this qualitative classification vary from person to person and are highly influenced by individual experience and subjective factors. In contrast, there are many different combinations of quasi-identifier attributes in the data table that can be used to identify individual information. If the judgments are made by data clerks based on their own experience only, there may be omissions, which may lead to privacy leakage risks.

In this paper, the pre-fusion data sensitivity is evaluated in two dimensions: dataset and data items. In evaluating the association between attributes $P_i, P_j$ the evaluation is done in terms of both structural features and data features. Structural features are defined by experts based on experience as well as data context. In terms of data features, different values of data item attributes correspond to different sensitivities. The sensitivity in this paper is defined as a range.

Definition 2: The sensitivity of each collected data object $E_i (1 \leq i \leq n)$ takes the value range $[F_{\text{min}, i}, F_{\text{max}, i}]$, where $F_{\text{max}, i}$ is the highest sensitivity of the $i$th data object $E_i$, and $F_{\text{min}, i}$ is the lowest sensitivity of the $i$th data object $E_i$.

In exploring the data association feature relationships, the Spearman correlation coefficient is used to measure the data feature correlation to obtain the sensitivity factor $\beta$.

Definition 3: The Spearman correlation coefficient is denoted by $\rho$. It evaluates the correlation of two statistical variables using monotonic equations. For two sets of data $X$ and $Y$ of size $n$, they are transformed into rank data $x_1, y_1 \ (i=1,2,\ldots,n)$, $\bar{x}, \bar{y}$ denote the mean of $x_i, y_i$ respectively, when the same rank exists in the rank data, the correlation coefficient $\rho$ can be expressed as:

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

When all rank values are different, the Spearman correlation coefficient can be calculated in a simpler way. The rank difference of two corresponding variables is denoted by $d_i = x_i - y_i$:

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad (2)$$

where $d_i$ is the equivalence difference between $x_i$ and $y_i$. It follows that the Spearman's correlation coefficient indicates the direction of correlation between the variable $X$ and the variable $Y$. If $Y$ tends to increase when $X$ increases, the Spearman correlation coefficient is positive. If $Y$ tends to decrease when $X$ increases, the Spearman correlation coefficient is negative. A zero Spearman correlation coefficient indicates that there is no tendency for $Y$ to increase when $X$ increases.

In summary, the sensitivity of the data items is adjusted by the sensitivity factor $\beta$ of the data structure features to obtain the pre-fusion data sensitivity.

### 3.3 Data Fusion and Sensitivity Correction

Define the result dataset of data fusion and sharing as the shared data object $S$. In the fusion or sharing process, the sensitivity of each acquisition data object $E_i \ (1 \leq i \leq n)$ is $F_i$ and $F_i \ (1 \leq i \leq n)$ is the highest sensitivity of the $i$th acquisition data object $E_i$.

Definition 4: The sensitivity weight $Q_i$ of each acquisition data object is calculated as follows. $Q_i$ reflects the weight of the $i$th data object sensitivity.
\[ Q_I = \frac{F_{\max} - F_{\min}}{\sum_{i=1}^{n} (F_{\max} - F_{\min})} \] (3)

Attribute selection is included in the data fusion process. For example, a part of the structural attribute content is cropped off on the collected data object for sharing. Along with the change of the data object structure, the sensitivity of the shared content will also change.

Definition 5: The selection object sensitivity is corrected according to all data attributes sensitivity distribution. and the weight distribution of the sensitivity on m attributes is \( \mu_k \) (1 \( \leq \) \( m \)) respectively, which should satisfy:

\[ \sum_{i=1}^{m} \mu_k = 1 \] (4)

If l of these attributes are selected for sharing, the sensitivity \( EM_k \) is calculated as follows:

\[ EM_k = (\sum_{i=1}^{n} \mu_k) \times EM_k \] (5)

The sensitivity of an instance \( s \) of S is denoted by SM, which corresponds to the sensitivity of n instances of the collected data object. After being modified by Equation 3, they are denoted by \( EM_1', EM_2', ..., EM_n' \), respectively. Among these values, the sensitivity \( EM_i' \) (1 \( \leq \) \( n \)), which satisfies the following conditions, is selected as the benchmark value for estimating SM.

After obtaining \( EM_i' \), the sensitivity of the shared data object instance can be corrected using the sensitivities of other collected data object instances other than \( EM_i' \).

Definition 6: The sensitivity correction \( \Delta F \) is calculated as:

\[ \Delta F = \sum_{j=1}^{n} \frac{q_j}{EM_j - EM_j + 1} \] (6)

Therefore, the sensitivity of the shared data object instance is:

\[ SM = EM_i' + \Delta F \] (7)

### 3.4 Data Sharing

The fused data needs to protect the user information from being leaked during the sharing process. In this paper, the shared data S data is anonymized. That is, a reasonable desensitization operation is performed on the attributes of the data set. It is required that while mining the value of the data, the corresponding individuals cannot be anti-identified. The individual information is protected from being leaked.

Definition 7: Equivalence class. When multiple data records are indistinguishable on quasi-identifiers, these records are said to constitute an equivalence class R. The total number of tuples contained in an equivalence class is called the size of that equivalence class. If the data table satisfies k anonymity, the number of equal tuples on the quasi-identifier property in any equivalence class is at least k.

Definition 8: k-anonymity. Given a dataset S and an equivalence class R. If the number of records in R are all at least k \( (k \geq 2) \), then T satisfies k-anonymity.

It follows that the probability of an attacker determining the association between a particular individual and tuple data when the dataset satisfies k-anonymity does not exceed \( 1/k \). The linkage attack is effectively prevented.

In this paper, we use Mondriaan algorithm to implement k-anonymity partitioning. The algorithm seeks to maximize the accuracy of data partitioning while satisfying the k-anonymity model. The quasi-identifier space is continuously partitioned until the sub-identifier space is indivisible. In this paper, the accuracy of dimensional partitioning is determined based on the fused data sensitivity values. Part of the algorithm process is as follows.

1. Anonymize(partition)
2. if(no allowable multidimensional cut for partition)
   return partition
3. else
   dim= choose dimension(EMi', SM)
4. fs = frequency set(partition, dim)
5. splitVal = find median(fs)
\[ \text{lhs} = \{ t \in \text{partition} : t.\dim \leq \text{splitVal} \} \]
\[ \text{rhs} = \{ t \in \text{partition} : t.\dim > \text{splitVal} \} \]
\[ \text{Anonymize(rhs)} \]

3.5 Model Evaluation

In this paper, the model is evaluated using Normalized Certainty Penalty (NCP) [6], which measures the information loss of a single equivalence class. For numerical data items, the NCP of the equivalence class R is defined as [7]:

\[ NCP_{P_{\text{Num}}}(R) = \frac{\max_{P_{\text{Num}}}-min_{P_{\text{Num}}}}{\max_{P_{\text{Num}}}-\min_{P_{\text{Num}}}} \]

where \( R_{\text{Num}} \) denotes the value of \( P \). \( \max_{P_{\text{Num}}} \) denotes the maximum value on attribute \( P \). \( \min_{P_{\text{Num}}} \) denotes the minimum value on attribute \( P \). \( \max_{P_{\text{Num}}} \) denotes the maximum value on attribute \( P \) in the equivalence class \( R \). \( \min_{P_{\text{Num}}} \) denotes the minimum value on attribute \( P \) in the equivalence class \( R \). So NCP measures information loss by the proportion of the generalized range of attributes in the equivalence class to the range of values taken by the attributes. For categorical data items, NCP is defined relative to the categorical tree of the data items:

\[ NCP_{P_{\text{cat}}}(R) = \begin{cases} 0, & \card(u) = 1 \\ \card(u)/|P_{\text{cat}}|, & \text{otherwise} \end{cases} \]

where \( P_{\text{cat}} \) denotes the range of \( P \). \( u \) is the lowest common ancestor of all \( P_{\text{cat}} \) values in \( R \). \( \card(u) \) refers to the number of leaves. \( |P_{\text{cat}}| \) is the total number of distinct \( P_{\text{cat}} \) values. The NCP of the class \( R \) over all quasi-identifier attributes is:

\[ NCP(R) = \sum_{i=1}^{d} w_i NCP_{P_i}(R) \]

d is the number of data items in \( S \). \( P_i \) is a numerical or categorical attribute with weight \( w_i \), where \( \sum w_i = 1 \).

4. Experiment & Results

Experiment with fusion and sharing of data from multiple data sources. Each data table includes different attributes of a common individual, connected by a common id. It is guaranteed that the fused and shared data satisfy two conditions. One is that the data is available and the other is that the data is not compromised.

All algorithms are implemented in python3 and the experiments were run on an Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz 2.80 GHz and Windows 10. The compiler tool is PyCharm 2020.1.2 x64.

4.1 Experimental data

This paper uses the Adult dataset from the UCI Machine Learning Repository, which has been widely used in related research. The data is cleaned and 30,000 randomly selected data are used for the experiment. Adding the id column to the dataset results in a total of 16 attributes in the dataset. The experiment uses 8 attributes from the dataset as \( E_1 \), including \{ id, age, work-class, fnlwge, marital-status, relationship, race, occupation \}. Other attributes were used as \( E_2 \), including \{ id, education-num, education, sex, capital-gain, capital-loss, hours-peer-work, Native-Country, income \}. In the dataset, nine of these data items are categorical and six are of numeric type. The anonymization algorithm used in this paper requires the input to be numeric, so the classified data are converted to numeric. For example, incomes of >50k and <=50k are converted to 0 and 1. From the dataset perspective, the sensitivity factor \( a \) is set to (0.5,0.5). Then define the sensitivity of structural features in data items. Due to the specificity of some occupations and confidentiality requirements, the workclass sensitivity is set to (1,3). Set the sensitivity of income to 3. So the initial sensitivities are set as follows, \( F_1 \): \{0,1,(1,3),0,1,1,1,2\}, \( F_2 \): \{0,1,1,1,(2,3),(2,3),2,1,3\}.

4.2 Experimental results analysis

We use spearman to calculate the attribute association degree. The correlation between age and marital-status in \( E_1 \) is 0.51 with p-value of 0. However, their original sensitivities are the same, so the sensitivities are unchanged. income and education in \( E_2 \) have a correlation of 0.33. Since the sensitivity of income is 3 and the sensitivity of education is 1, the adjusted sensitivity is 1.66. The correlation between education and education-num is 1. This is because the two columns denote education level and the corresponding grade level, respectively. The sensitivity of education-num is also 1.66.
because the sensitivity of both columns is the same. The final adjusted sensitivity is $F_1:\{0,1,1,3)\,0,1,1,2\}$, and $F_2:\{0,1,1.66,1.66,(2.27,3),(2,3),2.26,1,3\}$

Figure 2 reflects the comparison of the execution time variation between the model in this paper and the FSC algorithm. As the value of $k$ increases, the algorithm takes less time to execute. This is mainly because this paper takes out the $k$ records with the highest similarity on the quasi-identifier attributes until the dataset is empty. As the value of $k$ increases, the number of equivalence classes decreases and therefore the execution time decreases. The sensitivity of the data is determined and the data of different sensitivity levels are generalized to different degrees. So the model in this paper takes less time compared to the FSC algorithm.

Figure 2 k-value and execution time

Figure 3 K-value and information loss

The degree of information loss is measured by Eqs. (8), (9) and (10). NCP measures information loss by the proportion of the generalized range of attributes in the equivalence class to the range of values taken by the attributes. As shown in Figure 3, the information loss degree of this algorithm and the FSC algorithm are compared when the value of $k$ is changed. The degree of information loss increases as the value of $k$ increases. The larger the value of $k$, the higher the number of tuples in each equivalence class, the higher the generalization of tuples in the equivalence class, which leads to more information loss. The reason for less information loss in this algorithm is that the classification results of this algorithm are significantly related to the quasi-identifier attributes, sensitive attributes and sensitive values of the attributes. In this paper, the sensitivity values of the fused data are calculated by considering the data set, data items, and data values. Thus, we can better apply different degrees of generalization to different attributes to achieve privacy protection. When the degree of generalization of the attributes becomes lower, the NCP also becomes lower and the information loss also becomes lower.

5. Conclusion

In this paper, an FSC-based fusion sensitivity assessment method is proposed. The sensitivity of multidimensional sensitive attributes is evaluated from three aspects: dataset, data items, and data values. By conducting simulations on the dataset, it is demonstrated that the proposed model can make the fused data sensitivity more accurate. The proposed model can provide better privacy protection and has strong practical significance.

References


[5] Shulin LI, Quan HU, Xiaojie XIN, Rui LIU. Dataset Sensitivity Assessment Method and System for Big Data Fusion and Shared Results, G06F21/62, G06F21/62; G06F16/2458.


Theoretical analysis of vibration conditions in high voltage power equipment and on-line processing of field data

Shiling Zhang, Song Wei
1State Grid Chongqing Electric Power Company Chongqing Electric Power Research Institute, Chongqing, 401123
zhangshiling@cq.sgcc.com.cn

Abstract

Through detailed formula derivation process, the 6-order modal vibration spectrum is obtained. The axial dynamic characteristics of the winding are analyzed. During the analysis, the "mass-spring-damping" dynamic analysis model of the winding is established. Combined with the modal analysis method, the dynamic displacement response of winding cake under different loads is calculated, and the frequency spectrum is analyzed. The influence of the preload on the dynamic force of the cake is obtained, and the appropriate axial preload is finally determined. This paper attempts to deduce the formula from the theoretical level to obtain the relevant physical concepts and process parameter visualization in the vibration process of high-voltage power equipment. The relationship between the displacement data of key components and time is obtained by physical model inversion. In this paper, at present, the sinusoidal excitation function is used to approximate the field complex excitation waveform, and the closed analytical solution has been expected to be obtained. For the field complex excitation waveform, numerical simulation results are obtained through numerical simulation such as boundary element method. The research content of this paper has good guiding significance for the deformation of power equipment under the action of vibration signals and operation and maintenance measures.

Keywords: 6-order modal vibration spectrum, modal analysis method, frequency spectrum, closed analytical solution, numerical simulation

1. Introduction

The oil paper insulation structure inside the transformer may cause permanent mechanical damage to paperboard, insulating paper and other solid insulation under action of mechanical stress, affecting the normal operation of the transformer. When the common AC power transformer is in short circuit, its short-circuit electromagnetic force can reach hundreds of times of the rated operation. Although the short-circuit impedance of converter transformer is larger than that of ordinary transformer, and the short-circuit current will be smaller, its short-circuit electromagnetic force can still reach tens or hundreds of times of the rated operation[1-3]. Therefore, it is of great significance to calculate and analyze the electromagnetic force of UHV converter transformer during operation, and accurately analyze the short-circuit strength of UHV converter transformer winding, so as to ensure long-term and reliable operation of converter transformer[4-7].

The most serious short circuit condition of the transformer or wall bushing is the sudden short circuit when the terminal voltage passes through zero. Considering most serious situation, using the "field circuit coupling" finite element model established above, set the initial phase angle of the applied voltage at the grid side as, and calculate that the short-circuit current of the winding at the grid side and valve side changes non sinusoidally with time[8,9]. The maximum short-circuit current occurs at time \( t = 0.01 \) s. The maximum short-circuit current at the grid side is about 6000A, and the maximum short-circuit current at the valve side is 22127A, which is more than 6 times the rated current[10]. With the passage of time, the short-circuit current decreases gradually. The calculation and analysis of short-circuit electromagnetic force are based on the assumption that the winding is fixed. In fact, the body, winding and insulation structure will vibrate under the action of short-circuit electromagnetic force during the short-circuit process. Therefore, there may be a large deviation when calculating and analyzing the winding as fixed. In dynamic analysis, in addition to the short-circuit electromagnetic force borne by the conductor, the inertia force of the conductor, the elastic force of the insulating material, the internal friction force of the insulating material, and the friction force between the conductor and the insulating material should also be considered. Since the radial natural frequency of the winding is always much greater than 100Hz. Therefore, the radial short-circuit force of winding can be calculated by static method without large deviation. However, the axial natural frequency of large capacity transformer winding is often between 50~100Hz, so its dynamic characteristics must be calculated accurately.
analyzed[11]. In this paper, from the point of view of theoretical derivation, the characteristic control equations are established for UHV valve hall with super long wall bushing and transformer winding cake. Through detailed formula derivation process, the 6-order modal vibration spectrum is obtained. The axial dynamic characteristics of the winding are analyzed. During the analysis, the "mass-spring-damping" dynamic analysis model of the winding is established. Combined with the modal analysis method, the dynamic displacement response of winding cake under different loads is calculated, and the frequency spectrum is analyzed. The influence of the preload on the dynamic force of the cake is obtained, and the appropriate axial preload is finally determined. This paper attempts to deduce the formula from the theoretical level to obtain the relevant physical concepts and process parameter visualization in the vibration process of high-voltage power equipment. The relationship between the displacement data of key components and time is obtained by physical model inversion. In this paper, at present, the sinusoidal excitation function is used to approximate the field complex excitation waveform, and the closed analytical solution has been expected to be obtained. For the field complex excitation waveform, numerical simulation results are obtained through numerical simulation such as boundary element method. The research content of this paper has good guiding significance for the deformation of power equipment under the action of vibration signals and operation and maintenance measures.

2. Theoretical analysis of vibration mode of super long through wall casing structure

For the transformer winding, the super long wall bushing, structural system between the GIS/GIL internal conductor and metal jacket and other applications, there are vibration phenomena under the action of current or the weight of metal conductor itself. The above vibration signals exist objectively under the normal operation state of the power equipment, including current action (such as central conductor) and mechanical action (such as switching process). The normal vibration signal has good regularity and traceability, and can be used as a good representation of the normal operation of power equipment[12]. Abnormal vibration signal has random and high frequency characteristics, which can be used as effective indicator of abnormal operation of power equipment. This section first introduces the beam structure modal analysis, which mainly corresponds to the transformer bushing structure with one end fixed and one end similar to the free state[5]. The partial dimension equation of the response of the beam under transverse vibration is:

\[
\frac{d^2}{dx^2} (EI \frac{d^2 \omega(x,t)}{dx^2}) + m \frac{d^2 \omega(x,t)}{dt^2} = 0 \quad (1)
\]

Figure 1 shows a super long casing structure with a length of more than 10 meters. The indoor and outdoor ends will have beam vibration under the action of seismic waves. The vibration waveform is the calculation result of partial differential equation of position and time. The partial differential control program is shown in formula (1). Where \(E\) is the young's modulus, \(I\) is the moment of inertia of the beam cross section, \(I = bh^3 / 12\). \(EI\) is the bending stiffness, \(b\) is the width, \(h\) is the thickness of the beam, and \(m\) is the mass per unit length. If the beam is uniform, its free vibration can be expressed as:
For the harmonic vibration, according to the modal analysis method, the displacement of the structure is divided into two parts, time and space: \( \omega(x, t) = \phi(x)\eta(t) \). Where \( \phi(x) \) and \( \eta(t) \) respectively represent the structural mode shapes and modal coordinates. By separating variables from time \( t \) and space \( X \), two constant dimensional fractional equations can be obtained:

\[
\frac{d^4\omega(x, t)}{dx^4} + \frac{m}{EI} \frac{d^2\omega(x, t)}{dt^2} = 0
\]

(2)

For the harmonic vibration, according to the modal analysis method, the displacement of the structural mode shapes and modal coordinates. By separating variables from time \( t \) and space \( X \), two constant dimensional fractional equations can be obtained:

\[
\frac{d^4\phi(x)}{dx^4} - \frac{m\omega^2}{EI} \phi(x) = 0
\]

\[
\frac{d^2\eta(t)}{dt^2} + \omega^2 \eta(t) = 0
\]

(3)

Where \( \omega \) is the natural frequency of the beam. Let the general solution of the differential equation be:

\[
\phi(x) = A\sin(kx) + B\cos(kx) + C\sinh(kx) + D\cosh(kx)
\]

(4)

For simply supported the boundary conditions:

\[
k = \frac{n\pi}{L_x}, \quad x = 0 \text{ or } L_x.
\]

By substituting the simply supported boundary conditions, the \( B = D = 0 \), and \( C = 0 \), For a simply supported beam, the \( N_{th} \) mode shape function is:

\[
\phi_n(x) = \sin\left(\frac{n\pi}{L_x}x\right)
\]

(5)

Corresponding natural frequency can be obtained:

\[
\omega_n = \sqrt{\frac{EI}{m}} \left(\frac{n\pi}{L_x}\right)^2
\]

(6)

For ultra long UHV wall bushing, the governing equation of isotropic, undamped plate free vibration can be expressed as:

\[
D\nabla^4\omega(x, y, t) + m \frac{\partial^2\omega(x, y, t)}{\partial t^2} = 0
\]

\[
\nabla^2 \omega = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, \quad D = \frac{(h)^3 E}{12(1 - v^2)}
\]

(7)

For harmonic free vibration, it can be expressed as the superposition of infinite vibration mode functions, namely:

\[
\omega(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \phi_{mn}(x, y) \eta_{mn} e^{i\omega t}
\]

(8)

It has the following properties:

\[
\int_0^{L_x} \int_0^{L_y} m \phi_{mn}(x, y) \phi_{jk}(x, y) dy dx = \begin{cases} M_{mn}, & m = j, n = k \\ 0 & \text{otherwise} \end{cases}
\]

(9)

Where \( \eta_{mn} \) is the modal amplitude of the \( (m, n) \)th mode in the plane and \( M_{mn} \) is the modal mass. By transforming the above formula, we can get:

\[
D\left(\frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2\partial y^2} + \frac{\partial^4}{\partial y^4}\right)\phi_{mn}(x, y) - \omega_{mn}^2 M_{mn}\phi_{mn}(x, y) = 0
\]

(10)

Where \( \omega_{mn} \) is \( (m, n)_n \)th natural frequency. The vibration mode of the structure can be chosen arbitrarily as long as it meets the quasi orthogonal and boundary conditions. The mode shape function can be expressed as the product of two independent beam functions:
\[ \phi_{mn}(x, y) = X_m(x) \cdot Y_n(y) \]  

(11)

If and only if the mode shape function \( X_m(x) \) and \( Y_n(y) \) satisfy the quasi orthogonal and boundary conditions at the same time, it can be arbitrarily selected:

\[
\int_0^L X_j(x)X_k(x)dx = \int_0^L \frac{\partial^2 X_j(x)}{\partial x^2} \frac{\partial^2 X_k(x)}{\partial x^2} dx = 0
\]

(12)

\[
\int_0^L Y_j(y)Y_k(y)dy = \int_0^L \frac{\partial^2 Y_j(y)}{\partial y^2} \frac{\partial^2 Y_k(y)}{\partial y^2} dy = 0
\]

(13)

By using the orthogonal relation, the natural frequency can be obtained as:

\[
\omega_{mn} = \sqrt{\frac{D}{m_r}} \sqrt{\frac{I_1I_2 + 2I_1I_4 + I_4I_6}{I_2I_6}}
\]

(14)

\[
I_1 = \int_0^L \frac{\partial^4 X_m(x)}{\partial x^4} X_m(x)dx, \quad I_2 = \int_0^L (Y_n(y))^2 dy
\]

(15)

\[
I_3 = \int_0^L \frac{\partial^2 X_m(x)}{\partial x^2} X_m(x)dx, \quad I_4 = \int_0^L \frac{\partial^2 Y_n(y)}{\partial y^2} Y_n(y)dy
\]

(16)

\[
I_5 = \int_0^L \frac{\partial^2 Y_m(y)}{\partial y^2} Y_m(y)dy, \quad I_6 = \int_0^L (X_m(x))^2 dx
\]

(17)

For simply supported boundary, the mode shape function is selected as follows:

\[
X_m(x) = \sin\left(\frac{m\pi x}{L_x}\right), \quad Y_n(y) = \sin\left(\frac{m\pi y}{L_y}\right)
\]

(18)
For the fixed supported plate, the vibration mode function is selected as follows:

\[
X_m(x) = \cosh\left(\frac{\lambda_m x}{L_x}\right) - \cos\left(\frac{\lambda_m x}{L_x}\right) - \beta_m[\sinh\left(\frac{\lambda_m x}{L_x}\right) - \sin\left(\frac{\lambda_m x}{L_x}\right)]
\]

\[
X_n(y) = \cosh\left(\frac{\lambda_n y}{L_y}\right) - \cos\left(\frac{\lambda_n y}{L_y}\right) - \beta_n[\sinh\left(\frac{\lambda_n y}{L_y}\right) - \sin\left(\frac{\lambda_n y}{L_y}\right)]
\]

\[
\beta_m = \frac{\cosh(\lambda_m) - \cos(\lambda_m)}{\sinh(\lambda_m) - \sin(\lambda_m)}
\]

Where \(\lambda_m\) and \(\lambda_n\) are equations root of \(\cosh(\lambda)\cos(\lambda) = 1\).

\[\text{Fig. 3 First 6 modes of fixed support plate}\]

\[\text{Fig. 4 (2,3) mode amplify vibration waveform}\]

3. Theoretical analysis of winding vibration mode of UHV converter transformer

The short-circuit electromagnetic force of converter transformer winding is related to the distribution of the leakage magnetic field inside the transformer. Therefore, before calculating the short-circuit electromagnetic force of the winding, it is necessary to calculate leakage magnetic field distribution when the converter transformer is suddenly short circuited. In this section, the "field circuit coupling" method is used to calculate and analyze the short-circuit impedance, short-circuit current waveform and the leakage magnetic field distribution inside the UHV converter transformer during short circuit. The equivalent circuit of "field circuit coupling" model of converter transformer is shown in Figure 5.
Figure 5 shows that for UHV converter transformer winding, the column 1 and column 2 are the core winding area, which can be treated as the finite element part. The input and output terminals of the winding voltage and current can be treated as circuits. The equations governing the finite element domain and the circuit domain must be solved simultaneously [13]. According to the magnetic field calculation principle, the magnetic field of the converter transformer is analyzed, and the finite element matrix equation of "field circuit coupling" can be obtained, as shown in equation (20). It is assumed materials and structural parts in winding area have constant permeability and conductivity:

\[
\begin{bmatrix}
0 & 0 & 0 \\
C^{ii} & 0 & 0 \\
C^{ii} & 0 & 0 \\
\end{bmatrix} \frac{\partial \mathbf{A}}{\partial t} + \begin{bmatrix}
K^{AA} & K^{Ai} & 0 \\
0 & K^{ii} & 0 \\
0 & K^{ii} & K^{ie} \\
\end{bmatrix} \begin{bmatrix}
\mathbf{A} \\
\mathbf{I} \\
\mathbf{E} \\
\end{bmatrix} = \begin{bmatrix}
\mathbf{0} \\
\mathbf{0} \\
\mathbf{0} \\
\end{bmatrix}
\]

(20)

Establish the "field circuit coupling" model in case of short circuit at the valve side winding outlet of the transformer. See Table 1 for the structural parameters of the converter transformer.

Table 1. Winding parameters of converter transformer

<table>
<thead>
<tr>
<th>Winding</th>
<th>Grid side winding</th>
<th>Valve side winding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rated current/A</td>
<td>1176.5</td>
<td>3674.4</td>
</tr>
<tr>
<td>Internal diameter/mm</td>
<td>866</td>
<td>1195</td>
</tr>
<tr>
<td>External diameter/mm</td>
<td>910.8</td>
<td>1294.4</td>
</tr>
<tr>
<td>Height/mm</td>
<td>2461.6</td>
<td>2401.76</td>
</tr>
<tr>
<td>Number of cakes</td>
<td>140</td>
<td>179</td>
</tr>
<tr>
<td>Winding mode</td>
<td>Tangled continuous</td>
<td>Single screw type</td>
</tr>
</tbody>
</table>

For the calculation of initial conditions in the transient process, it can be assumed that the system starts to calculate under the steady-state conditions. Taking the transient process of sudden short circuit of transformer as an example, whether it is sudden short circuit of load or sudden short circuit of no-load to ground, the transformer is in steady-state operation before the short circuit occurs. By solving the magnetic field process from steady state to transient state, the short-circuit impedance, short-circuit current and leakage magnetic field distribution of the valve side winding of the converter transformer can be obtained [14]. Table 2 shows the current data of the converter transformer short-circuit simulation test.

Table 2. Short circuit simulation test of converter transformer

| Applied voltage at grid side/kV | 308.88 | 154.44 | 77.22 | 38.61 | 51.48 | 50.60 |
| Valve side current /A           | 22427  | 11213.6 | 5580.5 | 2079.3 | 3720.4 | 3673.9 |
| Network side current /A         | 6239.2 | 3104.8 | 1552.5 | 776.25 | 1035.0 | 1022.7 |
The relationship between the transient displacement of the converter transformer winding and time is obtained through the fractional differential equation\[11\]. According to the above winding structure characteristics of the converter transformer, the fractional nonlinear equation is selected as shown in equation (21):

\[
\frac{d^\alpha x}{dt^\alpha} = x - x^3 + R\sin(\omega t), \alpha \in (0, 2]
\]  

(21)

Wherein, \( F(t) = R\sin(\omega t) \) is characteristic high-frequency sinusoidal component of analog converter transformer, \( \omega = 0.5 \), \( f(x) = x - x^3 \) represents the structural characteristic quantity of the converter transformer, and the amplitude of \( R \) is related to the amplitude of short-circuit current in Table 2. The predictive correction method is used to simulate the equation (21), and the relationship between the transient displacement of converter transformer and time is shown in Figure 6.

![Fig. 6 Relationship between transient displacement of converter transformer and time](image)

Fig. 6 shows that for the specific expression of formula (21), the analytical formula and field circuit analysis method are in good agreement, and the control factor \( \alpha \) can effectively control the winding degeneration of converter transformer, and its value corresponds to the number of gaskets between winding cakes and the stress tension. In fact, for actual converter transformer, fractional differential equation of winding vibration can be described by the "mass spring damping" model of the vibration system\[15\], as shown in Figure 7.
According to the vibration theory, the system response is divided into two parts: free vibration and forced vibration. The free vibration part is caused by the initial conditions of the equation, that is, the initial tension of the spring and the initial state of the capacitive damper. The forced vibration is partly caused by the external excitation. Due to the existence of damping, part of the energy of free vibration is quickly consumed. As shown in Figure 6, the vibration is quickly established through the transition process in the first 50s, and the system enters the steady-state response stage containing only forced vibration from 50s to 100s [16, 17]. The vibration system in Figure 6 is described by fractional derivative formula of the simple harmonic function, as shown in equation (22).

\[
\frac{d^2 x}{dt^2} + \delta \frac{d^\alpha x(t)}{dt^\alpha} + \omega_0^2 x(t) = f \cos(\omega t) + F \cos(\Omega t)
\]  

(22)

First, find the approximate solution in the following equations under actual conditions:

\[
\frac{d^2 \psi}{d\tau^2} + \delta \frac{d^\alpha \psi}{d\tau^\alpha} + a_1 \psi = F \cos(\Omega_1 \tau) + F \cos(\Omega_2 \tau)
\]  

(23)

Let the solution of the above equation be:

\[
\psi = \frac{F}{\mu} \cos(\Omega_1 \tau - \theta)
\]  

(24)

Obtained by undetermined coefficient method:

\[
\mu = \sqrt{(a_1 + \partial \Omega_1^\alpha \cos(\frac{\alpha \pi}{2} - \Omega_1^2))^2 + (\partial \Omega_1^\alpha \sin(\frac{\alpha \pi}{2}))^2}
\]

\[
\theta = \arctan \left( \frac{\partial \Omega_1^\alpha \sin(\frac{\alpha \pi}{2})}{a_1 + \partial \Omega_1^\alpha \cos(\frac{\alpha \pi}{2} - \Omega_1^2)} \right)
\]  

(25)

Similarly, find the approximate solution of \( Y \) in the following linear equations:

\[
\frac{d^2 Y}{d\tau^2} + \delta \frac{d^\alpha Y}{d\tau^\alpha} + \omega_1^2 Y = f \cos(\omega_1 t)
\]  

(26)

Let the steady state solution of \( Y \):
\[ Y = A_L \cos (\omega_L \tau - \phi) \] (27)

Obtained by undetermined coefficient method:

\[
\begin{align*}
A_L &= \frac{f}{\sqrt{[\omega_r^2 - (\omega_i^2 - \omega_{\alpha^L}^2 \cos \frac{\alpha \pi}{2})]^2 + (\omega_{\alpha^L}^2 \sin \frac{\alpha \pi}{2})^2}} \\
\phi &= \arctan \frac{\omega_{\alpha^L}^2 \sin \frac{\alpha \pi}{2}}{\omega_r^2 - (\omega_i^2 - \omega_{\alpha^L}^2 \cos \frac{\alpha \pi}{2})}
\end{align*}
\] (28)

The amplitude gain of the system response is:

\[
Q = \frac{1}{\sqrt{[\omega_r^2 - (\omega_i^2 - \omega_{\alpha^L}^2 \cos \frac{\alpha \pi}{2})]^2 + (\omega_{\alpha^L}^2 \sin \frac{\alpha \pi}{2})^2}}
\] (29)

Firstly, the test is carried out in combination with actual vibration of the actual transformer winding. The relationship between the displacement and time is shown in Figure 8.

Figure 8 shows the dynamic displacement response of the winding cake is non-sinusoidal, and the dynamic displacement trends of the cake at different positions are quite different. The displacement of the wire cake (No. 1 and 179) at the end of the winding is very small, and the displacement of the wire cake (No. 90) at the middle of the winding is obviously larger, but it is not the position with the maximum vibration amplitude. The position with the largest vibration amplitude occurs near the winding 1/4 and 3/4 (No. 45 and No. 135). The change period of dynamic displacement of each thread cake is no longer consistent, and the first peak value of dynamic displacement is not at time \( s \), with obvious lag. It shows that the dynamic displacement response lags behind its excitation force (axial short-circuit electromagnetic force), and the closer it is to the middle of the winding, the more obvious this lag phenomenon is.

4. Conclusion

1) The analytical formula and field circuit analysis method are in good agreement, and the control factor \( \alpha \) can effectively control the winding degeneration of the converter transformer, and its value corresponds to the number of gaskets between
winding cakes and the stress tension. In fact, for actual converter transformer, fractional differential equation of winding vibration can be described by the "mass damping" model of the vibration system.

2) The relationship between the displacement data of key components and time is obtained by physical model inversion. In this paper, at present, the sinusoidal excitation function is used to approximate the field complex excitation waveform, and the closed analytical solution has been expected to be obtained. For the field complex excitation waveform, numerical simulation results are obtained through numerical simulation such as boundary element method.

3) Dynamic displacement response of winding cake is non sinusoidal, and the dynamic displacement trends of the cake at different positions are quite different. The position with the largest vibration amplitude occurs near the winding 1/4 and 3/4.

References

Application of Attention Mechanism in Image Search

Bingchen Gou a, Chengfeng Cai *a

a School of Mechanical Engineering, Northwestern Polytechnical University, Youyi West Road, Beilin District, Xian, China.

* Corresponding author e-mail: lhlccf@163.com

ABSTRACT

Under the background of high precision requirements of image search technology, new image search technologies emerge in an endless stream. Image search is one of the important research directions in deep learning. Among them, chromosome image search is a key step in medical imaging and modern genetic methods. In this paper, the principle of image search technology is analyzed, and a VGG16 neural network model using attention mechanism CBAM is presented to search for chromosome images. Finally, the optimization model is experimented and analyzed. The conclusion is that the addition of attention mechanism enhances the ability of the model to extract chromosome image features and efficiently completes the image search task. This method provides a new idea for the development of image search technology in the future.

Keywords: Chromosome image, Attention mechanism, Deep learning, CBAM

1. INTRODUCTION

The related technology of image search is not only widely concerned in the international academia, but also develops rapidly in the domestic industry, metallurgical forging and medicine. In the related ore smelting, image search is also needed to quickly identify the related materials, so as to complete the related operations such as batching smelting with high efficiency. In the industrial industry, the identification and classification of workpieces or products through image recognition and search is also one of the important directions of research. In medicine, molecular biology and cell biology have many applications: the collection and collation of pathological cell tissue images plays an important role in modern medicine, and chromosome image search technology is also one of the important research directions in the field of medical images.

The deep learning related algorithm has a wide application prospects in graphics and is the core of chromosome image search technology [1]. However, the operation of traditional technology is so complex that the efficiency of image search is not high that it can not really meet the enormous needs of medical image diagnosis and analysis applications in hospitals [2]. How to search and match chromosome pictures with similar morphological characteristics in the database is the focus and main problem of map search technology.

Traditional image search techniques, such as perceptual hash algorithm, compare fingerprint information features displayed on different images to achieve similar image search. However, the traditional method has low accuracy, can only input low-precision images, and the operation is cumbersome, which makes it difficult to update the data set quickly. Therefore, using deep learning to search images is an effective way to solve this problem.

This paper designs and implements the chromosome image search of chromosome image and its dataset, and applies it to the image retrieval of human chromosome database. This technology not only has a high accuracy rate, but also greatly saves manpower and material resources, and provides new strength for the development of new technologies in the future.

2. METHODS AND PRINCIPLES

2.1 Image search technology

Image search technology uses existing images to find the most matching pictures from masses of picture databases. Image search uses computer software to perform feature classification operations to obtain high-dimensional feature vectors that can be understood by the computer in place of image content, and then builds a feature information base by saving the feature signals for further operations such as feature matching and retrieval [3].
2.2 VGG16

Training network models for image search has been one of the important research applications and work focuses in the field of deep machine learning and visual graphics applications. VGGNet is a network developed by Oxford University in collaboration with Google [4]. After systematic analysis, it explores the depth distribution of topological structure between convolution and neural network layers, stacks a 3*3 convolution calculation core and 2*2 maximum pooling layer through repeated calculation, and constructs a 16-19 layer complex neural network. The chromosome image search technique designed in this paper is also implemented using VGG16 convolution neural network in deep learning.

2.3 Attention mechanism

The attention mechanism (CBAM), which is well known in the academic community by virtue of the attention of neural networks, can simulate the operation of the human brain [5]. For example, when people observe things, the eyeball uploads the acquired information to the brain through the nervous system, and the information transmitted by the brain always allocates attention to the relatively important area of the observed object. The signal processing function of human brain needs certain weight division. In the corresponding research papers, the model has conducted ablation experiments on the calculation methods of the two spatial attention mechanisms mask. The results show that the average pooling method and the maximum pooling method can extract more accurate signals. At the same time, the methods of combining the two mechanisms are compared. It is found that the efficiency of the method of combining the channel attention mechanism with the spatial attention mechanism is better. The basic architecture of CBAM module is shown in Figure 2.

![Figure 1. Block diagram of total body frame](image)

2.4 Similarity measurement and evaluation index

This design needs to analyze and evaluate the effect of the model used in the training, so as to judge the performance of the real effect of the model application. Error matrix is one of the commonly used evaluation indexes [6]. Generally speaking, the situation that will occur when the predicted results of the task are combined with the actual situation is shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSTIVE</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>TN</td>
<td>FN</td>
</tr>
</tbody>
</table>

According to the error matrix in the above table, the formulas of accuracy and recall are shown in Formula (1).

\[
\text{Accuracy} = \frac{TP}{TP+FP}
\]

(1)

The loss function for this design is shown in Formula (2).

\[
L(y|f(x)) = \max (0, -f(x))
\]

(2)

3. DATA SET CONSTRUCTION AND MODEL TRAINING

3.1 Image definition and data set

The data set used in this design is composed of chromosome pictures in different tissues of the human body. The chromosome pictures in this data set are taken by professional medical institutions through electron microscopy.
images are clear, similar in size, complete in labeling, and the resolution is $1017 \times 896$, after manual sorting, a total of about 600 chromosome images were obtained, including 250 chromosome images from peripheral blood, 250 chromosome images from amniotic fluid, and 100 abnormal chromosome images. After preliminary observation, the chromosome color of peripheral blood is darker and longer, and the chromosome color in amniotic fluid is lighter and shorter. The abnormal chromosome images often have chromosome deletion. Examples of chromosome images in some data sets are shown in Figure 3 below.

![Chromosome image example](image)

Through preliminary screening, it is found that some images are not clear enough and there is noise, so image preprocessing is required. The picture preprocessing technology is the most important key technology step in the picture post-processing technology. The picture processed by the preprocessing technology has a higher dynamic definition range and static contrast range. Smooth filtering and equalization are adopted for chromosome images. The comparison before and after processing is shown in Figure 3.

![Comparison before and after pretreatment](image)

### 3.2 Model pretraining and parameter setting

For the improved model, CBAM module is added after convolution operation. After the convolution operation, the channel score of the feature map can get the most important points in the model, that is, the most meaningful spatial features can be obtained. Secondly, the full connection layer of the model is replaced by the convolution layer. The size limitation of the received image is solved, so that the model can receive images of any resolution and obtain fuller feature information. At the same time, the function of information exchange is realized by combining the signals of each channel to ensure the integrity of the spatial structure of the image. In addition, compared with the modeling parameters of the full connection layer, the convolution method of the new model can effectively reduce the modeling parameters and shorten the time limit of the forward calculation of the model. The parameters are shown in Table 2 below.
This paper uses perceptron loss as the loss function, which is a variant of the Hinge loss function. Hinge loss has a high degree of penalty for determining points near the boundary. Perceptron loss is satisfied as long as the category of the sample is correct, regardless of the distance between the boundary, so it is simpler than Hinge loss and more suitable for the application of this experiment. The selection of training cycle mainly depends on the change of the loss function. Based on the analysis of the loss function curve during the training process, according to Figure 4, the design finally takes 45 as the size of the training cycle.

In addition, this paper uses some additional open data sets for analysis, pre-training and testing the improved model and the original model. Due to limited space, the results are briefly described. After pre-training, the improved model can obtain about 96% accuracy. Compared with the original model, the number of parameters is reduced and the accuracy is improved by 4.6%, which shows that the improved model has better feature extraction performance.

4. EXPERIMENTAL RESULTS AND EVALUATION

4.1 Image search results display

During the experiment, several experiments were performed to compare the searched images with the known chromosome images in the database, to evaluate the matching chromosome images of the corresponding categories in the database, and to compare the similarities between the images according to Euclidean distance, and to output the image results according to the similarity from high to low. The following is a screenshot of the chromosomes selected in the experiment using the map-searching system. The three maps are the chromosome image of peripheral blood, the chromosome image of amniotic fluid and the abnormal chromosome search results from top to bottom. The results show that the search accuracy is high, and the three different types of chromosome images can be distinguished and output in the order of similarity.
To evaluate the improved model, the accuracy of the chromosome image search system implemented by the original model and the improved model was analyzed. Excluding contingency, a hundred random search experiments were performed to calculate the accuracy of the first five, ten, twenty, fifty, and one hundred times, respectively. The accuracy of the two experiments is shown in Figure 6.

The evaluation results show that with the increase of the number of experiments, the accuracy of the evaluation results approximates to 90%, and the accuracy increases by about 10% compared with the original model. The evaluation results show that chromosome image searching with increased attention mechanism is more accurate and more suitable for medical detection of human chromosome images.

5. CONCLUSION

In order to overcome the low accuracy of chromosome image search technology, this paper presents a VGG16 deep learning model based on deep learning theory. The role of attention mechanism in deep learning is analyzed. By investigating traditional image search methods, an attention mechanism module is added to the original model, which improves the performance of some details in the model. By building datasets, image preprocessing and model validation, the effectiveness of the attention mechanism is verified in many ways. The experimental results show that by adding attention mechanism, the accuracy of chromosome image search based on depth features is improved compared with the original model, which provides a feasible solution for image search in the future. In the future, a new deep learning model can be combined to improve the accuracy of image search more effectively.
REFERENCES

Correlation Analysis and Prediction Between Bitcoin Price and Its Influencing Factors
Yinhao Liu
Department of Computer Science, Wenzhou Kean University, Wenzhou, Zhejiang, China
* Corresponding author: liuyinh@kean.edu

ABSTRACT

As the oldest and most famous cryptocurrency, the price of Bitcoin has increased nearly 2 million times in the last decade. As a result, predicting the price of Bitcoin through machine learning has become a big hit in recent years. This paper analyzes the correlation between the price of Bitcoin and market or social factors that may affect the price of Bitcoin. Then the author uses these factors with higher correlation to predict the price of bitcoin by LSTM. The experiments show that the average absolute percentage error of the LSTM prediction of bitcoin price decreases from 11.52% to 10.16%, 9.79%, 9.73%, 9.59%, 8.82%, and 8.50%, respectively, after the introduction of external correlation factors.

Keywords: Bitcoin, Gold, ETH, Baidu index, Google Trends, Tweet, Correlation, LSTM.

1. INTRODUCTION

Bitcoin is the longest-running and best-known cryptocurrency, first released as open source in 2009 by the anonymous Satoshi Nakamoto. Bitcoin acts as a decentralized digital medium of exchange where transactions are verified and recorded in a public distributed ledger(blockchain) without needing a trusted record-keeper or central intermediary. As with any currency/commodity in the market, bitcoin transactions and financial instruments soon continued to grow following the public adoption of bitcoin. In 2010, the price of per bitcoin was only $0.03, while in 2017, it reached $20000. Such an anomalous price fluctuation has made Bitcoin increasingly attractive to investors. As a result, predicting the price of Bitcoin has become a big hit in recent years for such an attractive virtual currency[1]. Among the many methods to predict the price of Bitcoin, one of the most effective machine learning methods is through long and short-term memory networks. However, traditional prediction methods only predict by the characteristics of Bitcoin itself and do not consider other factors that may be relevant, which also leads to the low accuracy of its prediction. This paper attempts to filter out five social and market factors that may be correlated with bitcoin price based on previous literature in order to improve the accuracy of bitcoin prediction models and avoid the problem of simplified bitcoin price prediction. The factors with higher correlation are screened out through correlation analysis, thus improving the LSTM prediction model. This improvement will help bitcoin investors comprehensively analyze the factors that affect the rise and fall of bitcoin prices when investing in bitcoin and provide investors with a more reliable solution for allocating their money wisely.

2. RELATED WORK

When it comes to the subject of Bitcoin price formation, the literature is primarily made up of empirical studies that evaluate the factors. Combined with previous research results, the factors that influence the price of bitcoin can usually be divided into social attention and market factors.

2.1 Social attention

Ladislav Kristoufek believes that the virtual currency market consists only of speculative traders. He relates digital currencies to Google trends and examines the relationship between them. In his article, he points out that search queries about digital currencies are linked to their prices, but there is also an apparent asymmetry in the effect of increased trader interest in the currency when it is above or below its trend value. It suggests that speculation and trend-chasing drive the price dynamics of Bitcoin in the cryptocurrency market[2]. Meanwhile, market confidence and perceived utility can influence the price of bitcoin[3]. Chen et al. also use Google trends and the Baidu index in their paper to measure the attention and media hype surrounding bitcoin investors. They argue that in speculative markets, investor sentiment reflects market confidence, and therefore investor sentiment plays an essential role in the price formation of digital...
currencies, while investor attention and media hype precisely reflect investors’ sentiment towards bitcoin[4]. However, some researchers consider Google Trends or the Baidu index to measure how much attention is paid to information about bitcoin[5]–[7]. Nevertheless, investors who genuinely understand cryptocurrencies will not search for bitcoin in these search engines; they will be more likely to talk about bitcoin on Twitter and even tweet about future price trends for bitcoin. Therefore, Shen et al. believe that the number of tweets with bitcoin as a keyword is a better indicator of the attention span of bitcoin investors[7].

2.2 Market factors
At its inception, Bitcoin was capped at 21 million, making most of its value come from its mining cost and scarcity, and the characteristics of Bitcoin’s store of value are similar to gold[8]. In addition to being compared to physical currencies such as gold, Bitcoin is often compared to other virtual currencies. In their experiments on Bitcoin’s connectivity with conventional assets, Zeng et al. demonstrate that the impact between Bitcoin and conventional assets is limited. However, when analyzing the connectivity between cryptocurrencies, they found a 34.7% correlation between Bitcoin and ETH[9]. Therefore, in this paper, ETH, Gold, Google Trend, Baidu Index and Tweet are chosen as the external factors to predict Bitcoin’s price (as shown in Table. 1)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Definition</th>
<th>Including</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>Historical data of Bitcoin exchanges</td>
<td>Open, High, Low, Close, Volume</td>
<td>Marketing</td>
</tr>
<tr>
<td>Ethereum</td>
<td>Historical data of Ethereum exchanges</td>
<td>Close, Volume</td>
<td>Marketing</td>
</tr>
<tr>
<td>Gold</td>
<td>Historical data of Gold exchanges</td>
<td>Close, Volume</td>
<td>Marketing</td>
</tr>
<tr>
<td>Tweet</td>
<td>Historical data on the number of tweets with the keyword &quot;bitcoin&quot;</td>
<td>Number of tweets</td>
<td>Attention</td>
</tr>
<tr>
<td>Google trend</td>
<td>Historical popularity of Google searches for the keyword &quot;bitcoin&quot;</td>
<td>Popularity</td>
<td>Attention</td>
</tr>
<tr>
<td>Baidu index</td>
<td>Historical popularity of Baidu searches for the keyword &quot;bitcoin&quot;</td>
<td>Popularity</td>
<td>Attention</td>
</tr>
</tbody>
</table>

3. DATA COLLECTION & CORRELATION ANALYSIS

3.1. Data preprocessing and normalization
The data on Bitcoin, ETH, and gold used in this article is from Kaggle, and a website called Bitinfochart provided historical data on the number of tweets for the keyword “bitcoin”. Also, Google trends and the Baidu index for the keyword “bitcoin” can be found on their respective websites. As shown in Fig. 1, for experiments and analysis, this paper uses daily data for each feature from March 10, 2016, to July 10, 2022[10][11][12][13].
Figure 1. Bitcoin features charts
3.2. Correlation analysis

To verify the correlation between several external factors mentioned in the related work and the bitcoin price and obtain the exact magnitude of the correlation between these factors and the bitcoin price, the Pearson correlation coefficient is introduced in this section to describe the correlation between the characteristics. The correlation coefficient is usually expressed by \( r \). The magnitude of \( r \) measures the strength of correlation between two features\([14]\). The correspondence between the magnitude of \( r \) and the correlation is shown in the Guidelines for Interpreting the Correlation Coefficient.

\[
\begin{align*}
    r &= \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \\
\end{align*}
\]

(1)

Table 2. Guidelines for interpreting the Correlation coefficients

<table>
<thead>
<tr>
<th>Correlation Coefficient (r)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r = -1 )</td>
<td>A perfect negative linear relationship</td>
</tr>
<tr>
<td>-1 &lt; ( r &lt; -0.7 )</td>
<td>A strong negative linear relationship</td>
</tr>
<tr>
<td>-0.7 &lt; ( r &lt; -0.3 )</td>
<td>A moderate negative linear relationship</td>
</tr>
<tr>
<td>-0.3 &lt; ( r &lt; 0 )</td>
<td>A weak negative linear relationship</td>
</tr>
<tr>
<td>( r = 0 )</td>
<td>No linear relationship</td>
</tr>
<tr>
<td>0 &lt; ( r &lt; 0.3 )</td>
<td>A weak positive linear relationship</td>
</tr>
<tr>
<td>0.3 &lt; ( r &lt; 0.7 )</td>
<td>A moderate positive linear relationship</td>
</tr>
<tr>
<td>0.7 &lt; ( r &lt; 1 )</td>
<td>A strong positive linear relationship</td>
</tr>
<tr>
<td>( r = 1 )</td>
<td>A perfect positive linear relationship</td>
</tr>
</tbody>
</table>

Based on the daily data collected for each feature, a heat map of the correlation coefficients was plotted as shown in Figure 2.
Based on the Heapmap about Correlation Coefficients, the following inferences can be made:

- The closing price of bitcoin (CLOSE) is strongly correlated with the closing price of gold (GOLD CLOSE), the number of tweets (with bitcoin as a keyword)(TWEET), and the closing price of ETH (CLOSE ETH).
- The closing price of bitcoin has a moderately positively correlated with the trading volume of bitcoin (VOLUME), Google trend (with bitcoin as a keyword) (GOOGLE TRENDS), and Baidu index (with bitcoin as a keyword) (BAIDU TRENDS).
- The closing price of Bitcoin and the trading volume of ETH (VOLUME ETH) have a weak positive correlation.
- The closing price of bitcoin and the trading volume of gold (GOLD VOLUME) have a modest negative correlation.

4. PREDICTION

Based on the above, this section will construct a prediction model for the bitcoin price and evaluate the model through a machine learning approach.

4.1. Long Short Term Memory

Long Short Term Memory, or LSTM, is a neural network that can remember long and short term information. It was first proposed by Hochreiter and Schmidhuber in 1997 and then improved by Felix Gers, Fred Cummins, and other computer science professors to form the present systematic and complete LSTM framework[15][16]. The motivation behind LSTM is to overcome the previously described long-term dependency problem. LSTM introduces the gate mechanism 2 to control the flow and loss of features. The structure of LSTM is shown in Figure 3.

\[
\begin{align*}
  f_i &= \sigma(W_f \cdot [h_{i-1}, x_i] + b_f) \\
  i &= \sigma(W_i \cdot [h_{i-1}, x_i] + b_i) \\
  \tilde{C}_i &= \tanh(W_c \cdot [h_{i-1}, x_i] + b_c) \\
  C_i &= f_i \cdot C_{i-1} + i \cdot \tilde{C}_i \\
  o_i &= \sigma(W_o \cdot [h_{i-1}, x_i] + b_o) \\
  h_i &= o_i \cdot \tanh(C_i)
\end{align*}
\] (2)
4.2. Parameters

To exclude the gap in the accuracy of the model due to different parameters of the neural network model, this experiment will use the original features of Bitcoin (open, low, high, close, volume) as the input values for training the LSTM neural network model, and then find the optimal number of LSTM layers, Dense layers, and neural unit nodes by grid search. Also, this experiment chooses to use Root Mean Squared Error 3 (RMSE) as the loss function of the neural network model [18][19]. Meanwhile, the Mean Average Percentage Error 4 (MAPE) is used as a metric to evaluate the LSTM networks built with different parameters, and the parameters in model with the smallest MAPE are selected in this experiment[20].

\[
RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{y}_j)^2}
\]

\[
MAPE = \frac{100}{n} \sum_{j=1}^{n} \left| \frac{y_j - \hat{y}_j}{y_j} \right|
\]

Table 3. Parameters of LSTM model

<table>
<thead>
<tr>
<th>LSTM layers</th>
<th>Dense layers</th>
<th>Units</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>32</td>
<td>11.52%</td>
</tr>
</tbody>
</table>

4.3. Build LSTM models

Based on the results of the correlation analysis between the daily closing price of Bitcoin and other external features in Section III, the strongly and moderately correlated features are selected as the input values for the training model. Then, train the LSTM model with the same parameters as in Table 3. Different combinations of features and their MAPE for the LSTM model are shown in Table 4.

Table 4. Combinations of features & MAPE for the LSTM models

<table>
<thead>
<tr>
<th>Features</th>
<th>Correlation Coefficient</th>
<th>Correlation</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>None</td>
<td>None</td>
<td>11.52%</td>
</tr>
<tr>
<td>Bitcoin+Baidu index</td>
<td>0.45</td>
<td>Moderate</td>
<td>10.16%</td>
</tr>
<tr>
<td>Bitcoin+Google trends</td>
<td>0.53</td>
<td>Moderate</td>
<td>9.79%</td>
</tr>
<tr>
<td>Bitcoin+Gold Close</td>
<td>0.73</td>
<td>Strong</td>
<td>9.73%</td>
</tr>
<tr>
<td>Bitcoin+Tweet</td>
<td>0.8</td>
<td>Strong</td>
<td>9.59%</td>
</tr>
<tr>
<td>Bitcoin+ETH Close</td>
<td>0.94</td>
<td>Strong</td>
<td>8.82%</td>
</tr>
<tr>
<td>Bitcoin+ETH+Tweet+Gold</td>
<td>0.82</td>
<td>Strong</td>
<td>8.50%</td>
</tr>
</tbody>
</table>
4.4. Prediction result

To test the LSTM model constructed in Section IV-C, the test set data was fed into the LSTM model as input values, and the closing price of Bitcoin was predicted from November 23, 2021, to July 1, 2022. The prediction results for the seven LSTM models are shown in Figure 4 and Figure 5. As shown in these figures, Figure 4 is the result predicted by the LSTM only from Bitcoin’s features, while Figure 5 is the result predicted by the LSTM Bitcoin’s features combined with other external features. The solid red lines in Fig. 4 and Fig. 5 represent the daily closing prices of Bitcoin from November 23, 2021, to July 1, 2022, while the solid black lines are the predicted daily closing prices of Bitcoin obtained from the LSTM model. The LSTM model does an excellent job of predicting the daily price trends of Bitcoin. However, in general, the predicted prices are lower than the actual ones.

Figure 4. Prediction chart (original)

Figure 5. Prediction chart
5. CONCLUSION

Based on past research, this paper has compiled and filtered out some of the two market factors (GOLD and ETH) and three social concerns (Google Trend, Baidu index, and Tweet) that may affect the price of Bitcoin. In addition, this paper analyzes and explains the correlation between the five factors mentioned above and the bitcoin price, then uses the factors with high and moderate correlation as essential features for training LSTM models. The LSTM is shown to have an excellent ability to process and analyze time series data as it overcomes the 'long term dependency problem' that recurrent neural networks have long suffered from. Comparing the features’ correlation magnitude and the model’s MAPE, the LSTM models trained with more correlated features have a smaller MAPE. It proves that LSTM models trained with more correlated features are more accurate in predicting the price of Bitcoin. Also, the MAPE of the LSTM model is better than the MAPE of the model trained with the three features that have a strong correlation with the bitcoin price at the same time as the input values for the model training. Although the MAPE of the LSTM model improved after the introduction of the outside correlation feature, according to Figure 4, although the LSTM model was able to predict the trend of the bitcoin price accurately, the predicted values given were, in general, smaller than the actual values. Therefore, the data and experiments in this paper still have some limitations, which suggests that the study is extensible. To improve the accuracy of the model needs to look for more factors related to bitcoin price to increase the dimensionality of the training data. Further research will fully explore the relationship between the bitcoin price and other external factors to provide better data support for investors.

ACKNOWLEDGMENT

Firstly, I would like to express my deep gratitude to my professor, who used to teach me data structures and algorithms in college and who generously provided all the expertise related to machine learning. Without his guidance, I would not have been able to venture into the field related to machine learning. In addition, I am also grateful to my classmates and teachers for their enthusiastic feedback and editorial guidance. Finally, I would be negligent if I did not mention my family, particularly my parents. Their confidence in me has maintained my spirits and motivation strong throughout the process.

REFERENCES


Multi-source data fusion evaluation method based on Uncertainty Quantification and Bayes method for success-failure type test

Wang Jiajia\textsuperscript{1,2*}, He Jingyuan \textsuperscript{1b*}, Ma Jing \textsuperscript{1,2c}

\textsuperscript{1} Beijing Jinghang Research Institute of Computing and Communication Beijing, China
\textsuperscript{2} Applied Mathematics Research Center of China Aerospace Science and Industry Group Beijing, China
\textsuperscript{a*}gaga214@126.com, \textsuperscript{b*}244172596@qq.com, \textsuperscript{c}jingzi0197@126.com

Abstract

In this paper, we use Uncertainty Quantification method to analyse the data from three types of success-failure type test sources: physical test, semi-physical test and computer simulation test. Bayes method is used to evaluate the data fusion of multiple source test, and the validity and applicability of the method are verified by experiments.

**Keywords** - Uncertainty Quantification; Multi-source data fusion evaluation; Bayes; success-failure type test

1. Introduction

The Multi-source test data fusion technology makes full use of the prior information to participate in the comprehensive evaluation. There are many methods of data fusion, and the fusion objects involved are quite wide. The methods of processing and fusion information for different objects have different emphasis. The basic idea of multi-source data fusion is to use appropriate methods to fuse the prior information of large sub samples and the confirmatory test information of small samples for comprehensive analysis and evaluation. The theoretical core is Bayes method. Since Bayes method has the minimum error rate when the prior probability and conditional probability are known, it has obvious advantages in pattern classification. However, the priori information obtained under different experimental conditions is often of different types, subject to different overall distribution, and must be processed before it can be used. This method comprehensively considers the characteristics of the prior information, adopts the uncertainty quantification method that can handle the uncertainty information, and combines the Bayes method to transform the prior information under different conditions into a whole for statistical analysis, so as to realize the multi parent joint inference, and combines the small sample field test data for inspection. It has a significant effect in increasing the amount of information and improving the evaluation reliability.

2. Multi-source Data Fusion Evaluation Process Based on Uncertainty Quantification and Bayes Method

The with H-type main steel beams and steel channels, lightweight precast panels set upon the steel skeleton, shear keys connected to the main steel beams and post-pouring concrete layer.
2.1. Calculation of reliability and success probability of computer simulation test

The reliability of a computer simulation test is determined by the similarity between the simulation system and the prototype system. The value of the reliability is a decimal between 0 and 1. The reliability of 1 indicates that the test is completely credible, that is, all the results of the simulation test can be completely realized in reality; A reliability of 0 indicates that the simulation system is completely unavailable.

Assuming that all parts of the computer simulation system are independent of each other, the simulation test reliability $C_{sim}$ is as follows: $C_{sim} = \sum_{i} \eta_i B_i$. Where $\eta_i$ is the importance coefficient of the $i$th part of the simulation object, and its value depends on the purpose of simulation (its value depends on the importance of each part during simulation, which is generally determined by experience), and meets the following requirements: $\sum_{i} \eta_i = 1$. For the success probability of the simulation test, the maximum likelihood estimation method is adopted in the prior data, namely: $P_{sim} = \frac{s_{sim}}{n_{sim}}$. Where, $n_{sim}$ is the number of computer simulation tests, $s_{sim}$ is the number of successes.

2.2. Calculation of reliability and success probability of semi physical test

Reliability of alternative physical tests based on normalization conditions $C_{phy}$ is weighted and distributed, then $C_{phy} = 1 - C_{sim}$. The structural model of the system is divided by the discriminant method of the reliability structure of the engineering system, and the probability of equivalent results is estimated by the least squares method. The reliability model of a system shows the logical and quantitative relationship of reliability between the system and its units. When the reliability structure model of the system is a series model, the estimated probability of success from the equivalent information source is $P_{phy}$.

$$P_{phy} = \Pi_i^n \hat{P}_i = \Pi_i^n \frac{s_i}{n_i} \quad (1)$$

When the reliability structure model of the system is a parallel model, the estimated probability of success from the equivalent information source is $P_{phy}$.

$$P_{phy} = 1 - \Pi_i^n (1 - \hat{P}_i) = 1 - \Pi_i^n \frac{L_i}{n_i} \quad (2)$$

For more complex systems, the reliability model of hybrid systems is needed and calculated according to the actual situation.

2.3. Prior probability calculation based on Uncertainty Quantification method

A priori probability refers to the probability of success of an equivalent test after it is fully analyzed, calculated and evaluated from the substitution test data and converted into a real field test. The results of a priori probability calculation.
can include the success probability of the field test, the total number of equivalent tests of the substitution test, the number of equivalent successes, and the probability density function of the success of the field test.

2.3.1. Estimate success probability

In the previous calculation, the reliability and probability of success estimates of computed simulation tests and physical substitution tests are obtained respectively, and then the probability of success estimates of all test information sources are calculated.

\[
\hat{P} = C_{sim} \cdot P_{sim} + (1 - C_{sim}) \cdot P_{phy}
\]  

(3)

2.3.2. Multi-source information fusion based on uncertainty quantization and information entropy

The basic function of information is to eliminate people's uncertainty about things. In 1948, Claude Elwood Shannon, the father of information theory, borrowed the concept of thermodynamics to propose "information entropy", which was used to represent the relationship between probability and information redundancy, and to solve the problem of quantifying information. According to Shannon's theory, things contain accurate information expressed as:

\[
H(x) = -C \sum_{x \in X} p(x) \ln p(x)
\]  

(4)

Where \(x\) is the possible event in the system, \(X\). For the event set, \(p(x)\) is the probability function of each event. In this invention, uncertainty is introduced as the weight, and \(C\) is selected as the relative standard deviation of prior data in a single trial as the uncertainty index, indicating the randomness of the data. The smaller the standard deviation, the smaller the uncertainty, and the more reliable the test results. The data uncertainty \(C\) for a single trial is:

\[
C = \frac{s}{d} \times 100\% \quad S = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (d_i - \bar{d})^2}
\]  

(5)

Where the test data \(d = (d_1, d_2, \ldots, d_n)\), \(S\) is the variance of data \(d\), \(\bar{d}\) is the mean of test data \(d\).

There are \(N\) alternative trials as \(N\) independent sources of information for success or failure trials. In the \(i\)-th information source, test \(n_i\) times, successes \(s_i\), failures \(f_i = n_i - s_i\). Based on the principle that the total amount of test information is constant and the probability of success of the test, the total number of tests \(n_0\), the number of Success \(s_0\), and the number of failures \(f_0\) is calculated. Success \(s_0\).

\[
\begin{aligned}
n_0 &= \sum_{i=1}^{N} \frac{C_i s_i f_i (n_i - s_i) \ln(n_i)}{C \left(\ln(P) + \ln(1 - P) + \ln(1 - P)\right)} \\
 &= \frac{s_0}{n_0 \hat{P}} \\
f_0 &= n_0 (1 - \hat{P})
\end{aligned}
\]  

(6)

2.3.3. Posterior probability calculation based on Bayes method

Previous calculations converted computer simulation and alternative physical test data into equivalent success or failure tests and yielded some valid results. However, if you want to apply the equivalent test results to an integrated evaluation method for multisource data, you also need to use the Bootstrap method to calculate the prior distribution of the above data according to the Bayes School of thought.

According to the probability theory, the true probability density of the success or failure test conforms to the Beta function, \(\pi(Y) = Be(a, b)\). Based on the properties of the Beta distribution function, it can be concluded that:

\[
\begin{aligned}
\frac{a}{a + b} &= E \\
\left(\frac{ab}{(a+b)(a+b+1)}\right) &= S^2
\end{aligned}
\]  

(7)

From the above formula, the hyperparameters \(a\) and \(b\) can be calculated, and the prior probability density function \(\pi(Y)\) of \(X\) can be obtained.

A priori probability density function is the probability of success of a success or failure type of test calculated by a variety of simulations and substitution tests. A posterior probability density function is a modification to this function after the results of a real environmental test are obtained.
Assume that the field samples obtained under the real test environment are \( X = \{x_1, x_2, \ldots, x_n\} \), \( n \) times in total, of which \( s \) times are successful. Based on Bayes theory, a posteriori distribution of \( \pi(y) \) in a success or failure test can be obtained by combining \( \pi(y) \) with a field sample as \( X \).

\[
\pi(y|X) = \frac{\pi(y|f(y|X))}{\int_y \pi(y|f(y|X))dy} = \text{Be}(a + s, b + (n - s))
\]  

(8)

The point estimate of \( \pi(y|X) \) is:

\[
\hat{y} = \frac{a + s}{a + b + n}
\]  

(9)

the confidence level \( \alpha \) is:

\[
\alpha = \int_{R_L}^1 \pi(y|X)dy
\]  

(10)

Usually we set \( \alpha = 0.95 \), that is, the probability that the true value appears within the confidence interval is 95%. In engineering, there are often requirements of 0.99 and 0.999. When \( \alpha \) Given, there are:

\[
\sum_{i=0}^{b+(n-s)} C_i^{d+b+n} R_L^{d+b+n-i} (1 - R_L)^i = 1 - \alpha.
\]  

(11)

The lower confidence limit of \( R_L \) can be calculated.

### 3. Test Results and Discussions

#### 3.1. Experimental data

we conduct a success or failure multi-source evaluation test for a certain combat equipment. The test contents include field test, computer simulation test and two physical substitution tests for different independent subsystems. The number of each test is different and independent of each other, without any impact. The reliability structure model of the test system is a series model. The test results are shown in the Table 1.

<table>
<thead>
<tr>
<th>Test category</th>
<th>Test times</th>
<th>Success times</th>
<th>Probability of success</th>
<th>Data uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation test A</td>
<td>1000</td>
<td>932</td>
<td>93.2%</td>
<td>0.14</td>
</tr>
<tr>
<td>Physical substitution test</td>
<td>Test B1</td>
<td>30</td>
<td>28</td>
<td>93.3%</td>
</tr>
<tr>
<td></td>
<td>Test B2</td>
<td>5</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>Field test C</td>
<td>2</td>
<td>2</td>
<td>100%</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 1 The data of field test, computer simulation test and two physical substitution tests

In the simulation test, there are three independent parts. Their importance and simulation accuracy are determined by experts, as shown in the following Table 2.

<table>
<thead>
<tr>
<th>Importance coefficient ( \eta_i )</th>
<th>Accuracy ( B_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation part①</td>
<td>0.5</td>
</tr>
<tr>
<td>Simulation part②</td>
<td>0.2</td>
</tr>
<tr>
<td>Simulation part③</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 2 The data of three independent parts.

#### 3.2. Calculation process

##### 3.2.1. Test data processing

For simulation test A, its simulation reliability \( C_{\text{sim}} = \sum_i \eta_i B_i = 0.938 \) according to Formula 1, the success probability of simulation test \( P_{\text{sim}} = 0.932 \). For physical substitution test, its simulation reliability \( C_{\text{phy}} = 0.093 \) according to Formula 1, and the uncertainty of data is \( C_1 = 0.14, C_2 = 0.16, C_3 = 0.20, C = 0.18 \) according to Formula 5. Since the reliability structure model of the system is a series model, the success probability \( P_{\text{phy}} = 0.9586 \) according to Formula 3.
3.2.2. Prior probability calculation

Estimated probability of success \( P_{\text{in}} = 0.9586 \), total number of equivalent tests \( n_{\text{in}} = 1037 \), success times \( s_{\text{in}} = 967 \), failure times \( f_{\text{in}} = 70 \). The information of equivalent test is: total number of equivalent tests \( n_0 = 1159 \) according to Formula 6, success times \( s_0 = 1111 \), failure times \( f_0 = 48 \).

When the information entropy theory is applied in practice, the success / failure may be 0 in the success / failure type test. At this time, the value of 0 is added \( \delta = 1.0 \times 10^{-15} \) to facilitate calculation.

According to the above information, a regenerated sub sample with a success probability of \( Y = 0.9586 \) is generated to form a new Bernoulli distribution, and bootstrap is used for the test. The expected value of the success probability of the sample is \( E_{\text{in}} = 0.959 \), and the variance is \( S^2 = 7.5160 \times 10^{-5} \) according to Formula 7. According to this result, the hyperparameters \( a = 501 \) and \( b = 21 \) of the beta distribution are obtained according to Formula 8, that is, the distribution function of the prior probability density is \( \pi(y) = \text{Be}(501,21) \).

3.2.3. Posterior probability calculation

Take confidence level \( \alpha = 0.95 \). This case has been tested twice in the field and succeeded twice, i.e. \( n = 2 \), \( s = 2 \). The posterior distribution probability function \( \pi(y|X) = \text{Be}(501,21) \) is obtained. At this time, the point estimate \( y \) of the posterior distribution \( \hat{y} = 95.99\% \), lower confidence limit \( R_L = 94.98\% \) according to Formula 11.

3.3. Analysis and comparison of results

In order to verify the effect of multi-source data fusion method based on information entropy theory, three classical methods are used to analysis and calculate the experimental data, and the results are compared with the research results of this paper. The classical probability theory method is used for analysis. When all the data are considered comprehensively, the comprehensive success probability \( p = 0.95 \) is obtained, which is close to the data studied in this paper. However, after a little consideration of its principle, it can be seen that this practice compares the importance of all kinds of tests equally, which is not only inconsistent with reality, but also inconsistent with the idea of data fusion of alternative tests, and has great loopholes.

We analysis the field data, the success probability \( p = 1 \) is obtained. However, the Bayes theory is used to test the value and it is found that the method is taking the confidence level \( \alpha = 0.95 \), the lower confidence limit is \( R_L = 22.4\% \). A lower confidence limit indicates that the data source and sample size used in this method are too small, resulting in the evaluation results being too conservative and not suitable for use in engineering practice.

4. Conclusion

This paper explores the evaluation method of multi-source data fusion of success or failure type test based on uncertainty quantification and Bayes method, calculates the uncertainty and information entropy of data uncertainty and randomness respectively, fully considers the reliability of data, and then integrates Bayes method to give the evaluation method of multi-source data fusion based on physical test data, semi physical test data and physical test data, The validity of the method is verified by experiments, and the analysis conclusion is given.

References


Database Management and Data Mining

Zhiying Qiao
College of Arts, Media and Technology, Chiang Mai University, Chiang Mai Thailand, 50200
qiao_zhiying@elearning.cmu.ac.th

ABSTRACT

Today, driven by new technologies such as cloud computing, big data, the Internet of Things and artificial intelligence, the era of artificial intelligence has arrived. "Digital transformation" has become the only way for enterprises to seek breakthroughs and innovations. As the carrier of information, the applications of data are fundamental. This paper discusses the database construction, data mining, analysis methods and the application of optimizing the database. Each of these applications will be discussed in details.

Keywords-Big Data, Database, Data mining, Mining methods

1. Introduction

Data play an essential role in modern life, whether people works in a prosperous innovative office or simply enjoy the benefit of the internet. Data are physical symbols or combinations of these physical symbols that record the nature, state, and mutual relationship of objective things. It can take many forms, including words, numbers, images, etc. We can now refer to all information recorded in electronic form as "data." [1].

Vast amounts of data are generated constantly, they can be stored and organized efficiently in database. Database is an organized, shareable collection of data stored in a computer for a long time [2]. Databases are everywhere. We use database when we look up someone’s phone number in the contact list. When searching on the internet, all the information is inside the database. When we log in to any privacy accounts, we also need to rely on the database to verify our name and password. Even when using an ATM (Automated Teller Machine), the bank uses their database for PIN (personal identification number) verification and balance checking. Database is a collection of files, warehouse for storing data, and file system. The database stores data in specific format and users can add, delete, modify, and query the stored data. In everyday life, people can directly describe things in natural language. The computer needs to abstract out the features of these things to form a descriptive record. These records are called electronic file cabinets and users can add, intercept, update, and delete data in the files.

Database has obvious advantages in managing, organizing and exchanging information resources. It has strong technical capabilities in data security and management efficiency [3]. A structure that contains multiple units of information can be built by connecting units such as file entities, file managers. The database can comprehensively realize the collection, editing and modification of archive data due to these structures links and ensure the effective execution of the archives management process [4]. Taobao needs to collect massive data, such as online product conversion, clicks number, current influence and other information to create user portraits. Merchants can increase revenue by carefully examining issues faced with consumers and products through product ranking, product selection, and competitive product analysis [5]. There are still many problems with databases which face of large customer base, such as the inability to build databases efficiently, or the inappropriate methods of data collection and analysis. This paper discusses database construction, data mining, analysis methods, and the application of database optimization. While understanding such technologies, in-depth discussions, descriptions of specific applications and future development will be conducted. Suggestions for practical applications will be provided to make them more valuable in the development of modern society.

This paper is mainly divided into three parts, 1) Database Construction, including Demand Analysis, Data Collection, and Data Conversion. 2) Database data mining and analysis methods include six methods: Association Analysis, Cluster Analysis, Classification Mode, Time Series, Variance Analysis, and Abnormal Group. 3) Suggestions on the Application of Databases in Management include the confidentiality and security of files, strengthening network supervision, and improving the efficiency of information storage.
2. Database Construction

With the development of the IT industry, many service providers' business models and operation methods have changed. For example, the traditional business model of the company is single dimension of profit while the Internet model of the company is a multi-dimensional profit, multi-dimensional can better cope with risks; Traditional corporation are logic of "goods", while internet corporation are logic of "people", with private domain traffic will be able to sell things. The "ideal data processing core organization" is development by the researchers to meet the dramatic increasing demand of data processing. It is a software system that provides contents by data which achieves storage, maintenance and search. The storage problem has been well solved with the help of software system or server [6,7]. At the same time, the high-speed processing capability and large-capacity memory of computer provide the condition for the realizing of data management automation. To standardize the content of the database, the first need to analyze the structure of the database. The construction of archival database can be divided into three categories: demand analysis, data collection and data conversion.

2.1 Demand Analysis

Everyone is unique, with personal views, interests, backgrounds, cultures and experiences, so dissimilar customers have various needs and values. Enterprises can better identify customer groups through user description and market different research. According to the demand of customer groups, different sales methods are adopted to better guide customer relationship management, attract customers, maintain customers and build customer loyalty [8]. Application requirement analysis mainly solves how to obtain a large number of key data information from archives. For more accurate synthesis processing, database technology can be used. Enterprises need to analyze users, grasp user preferences and divide customer into dissimilar groups, such as users' age, occupation, gender, viewing time and so on. Refer to a large amount of information and data to draw corresponding conclusions and provide more intuitive decision-making [9]. This information support will assist its users to make correct decisions.

2.2 Data Collection

Data are the database's primary content, and data collection is significant in constructing database archives. The collected data is the basis of data analysis and data mining [10]. With data support, companies can better convert data into cash flow and quickly take the lead. Today is the era of big data, and the sources of data collection are diversified. There are six primary sources of data collection: terminal data, open data, data from other platforms, physical data, subjective data and data from databases. Data collection methods can be mainly divided into four types: namely manual method, barcode method, RFID (Radio Frequency Identification) method, and data analysis tool input. These data sources are different in the data structure. However, they can also be aggregated through data collection and then summarized into a unified archive database to form an archive information source.

2.3 Data Conversion

After the case data are collected, the existing archive needs to be transformed. This conversion must be carried out according to the information source [11]. For example, some users have different information registration methods of taking the analysis of archives users. Some users prefer to use numbers, others prefer letters. In the face of these various formats of data, archival managers need to convert these data into the same mode. Facing these various types of data, a same format can be realized by the standardized management of the archive database.

3. Database data mining and analysis methods

3.1 Association Analysis

The primary technique used in data mining is association analysis, which identifies relationships between various fields in light of the connections between data sets. It can be utilized in various marketing contexts, including correlation analyses of consumer purchase patterns [12]. Therefore, Businesses can use the proper marketing strategies to popularize their products and encourage the sales of new products by better understanding consumer behavior patterns when they buy things and make queries. Additionally, it can aid in the consumer's discovery of contemporary goods, assisting in their marketing and sales. For example, suggesting products to the customers based on their recent shopping patterns.

There are two steps in the association rule recommendation algorithm. According to the association regulation, the product recommendation system firstly determines the recommendation degree of the items the current customer has not explored. Secondly, it suggests the products to the current customer based on the magnitude of the recommendation degree.
3.2 Cluster Analysis

Cluster analysis makes the similarity of data within a category as great as possible and the similarity between data within different categories as small as possible. One of the most widely used methods in the field of data mining is clustering analysis [13]. This is often referred to as categorization based on similarity and dissimilarity. We can achieve the desired results by repeatedly eliminating or adding variables to the classification process [14].

Cluster analysis is frequently used in market segmentation in e-commerce. Customers with similar consumption patterns can be grouped based on available customer information data. Marketing strategies can be adjusted more specifically to offer customers a more convenient and satisfying individualized service [15]. For instance, it can offer high-quality service content during the product marketing process depending on the current customer situation, the customers’ inconstant consumption levels, and the customers’ inconstant situations.

3.3 Classification Mode

According to the classification mode, data with resemblant properties are grouped into one class, and data with comparatively distinct features are grouped into various classes. Therefore, classification analysis aims to identify the shared characteristics of a collection of data objects. The goal is to use a classification model or function to associate each data item in the database with a particular category; Decision Tree Algorithm [16], Support Vector Machine, and Naive Bayes Classifier [17].

The easiest way to utilize is the classification model. It breaks occurrences into objects also uses this technology to track and foretell data. The classification mode sort the data items into categories to create a thorough and accurate prediction model. For instance, e-commerce enterprises predict what customers will do next according to their main situation and carry out diversified business marketing activities for various users to provide tailored service content.

3.4 Time Series

Databases related to e-commerce transactions should contain information about items type, items name, prices, purchase times, and user id. Businesses use databases to determine the relationship between user behavior and purchase timing to implement tailored marketing strategies. Analyzing the frequency of events over a while is referred to as time series pattern analysis [18].

This type of time series is used to discover the relationship between content of data sets and timing. It also analyzes projects on a project-by-project basis, making scientific predictions and observations about future data. However, the main goal of the sequence pattern is to discover the situation of the products purchased by customers in the database during a specific period. For example, the sequence pattern can find out that the purchasing of computer always comes along with the purchasing of other computer’s accessories.

3.5 Variance Analysis

Variance analysis is a small proportion of objects in a data set; biased objects are frequently referred to as outliers which includes abnormal occurrences, observational deviations, and random changes [19]. It is used to reveal abnormal phenomena, detect and analyze the deviation data in the database also some abnormal records summarized by the database, whose characteristics differ significantly from the majority of the data in the database.

Outliers in data reflect anomalous patterns in daily activity, and their flexible use can be advantageous to businesses. It has been used for customer churn warning, discovery, analysis, identification, and evaluation of aberrant customer information. Rare events simplify identification problems in some specialized data mining applications. Therefore outliers cannot be ignored. Outlier analysis has also emerged as a helpful security monitoring technique in network penetration and credit card theft. For instance, if a customer's statement suddenly shows a sizable transaction, it was most likely made fraudulently with a credit card.

3.6 Abnormal Group

Abnormal group is composed of a few similar data objects in a given large data set, which is different from most data objects and forms anomalies [20-22]. It is a kind of high-value and low-density data. Morphology is also a new type of big data mining task in a big data environment.
Abnormal group, Cluster analysis, and Variance analysis are all data mining tasks that are divided according to the similarity of data objects [20-22]. However, abnormal group mining differs from Cluster analysis and Variance analysis in terms of problem definition, algorithm design, and application effects [23].

Data on human performances show how people act proactively. An abnormal group is generated when these deeds objects show a significant percentage of anomalies and a specific number of the same or similar behavior patterns. Finding those few anomalies among numerous behaviors is the goal of abnormal group mining.

A higher credit rating will bring in more buyers. As a result, some sellers employ the “credit brush” method to achieve high credit scores. For example, most online trading platforms, such as eBay and Taobao, have credit scoring systems for both parties in a transaction. If a group of accounts consistently provides a high credit score for the same sellers, this group of accounts is suspicious. Identifying these accounts will aid the trading platform's reputation fraud detection.

4. Suggestions on the Application of Database in Management

The first is to ensure the confidentiality and security the documents. The database storage personal privacy information [24], Security and confidentiality of data are particularly important, those disclosure personal information will directly affect individuals and even endanger society. At the same time, network security will be connected to the security of file information, the computer database management system can encrypt the data, or combine with the special data encryption system to improve the security of data. However, the technical weakness of computer networks in file management needs to be solved by technical improvement.

Second, Strengthen network supervision, timely database backup and recovery work. Data information loss, system crashes and other accidents often occur in computer network applications, resulting in the data recorded in the database can not be called normally. The computer database management system is mainly a collection of data storage [25,26]. The management system will store data according to a particular pattern in this storage process and establish a more complex structural relationship [27]. In this way, the data can not only be processed separately but also the existing data can be merged, used and processed, in order to achieve high-efficiency data recovery ability. Enterprises should pay more attention to network supervision and carry out network security strategies for local terminal security configuration. For the Internet, effective management and control of network security access, in order to better play the role of network security and gateway protection. The maintenance of network resources can also prevent illegal use and access. On the other hand, the database backup should be done in time to prevent the loss and destruction of the database in order to prevent unnecessary losses to file management.

Finally, to improve the efficiency of information storage, the most significant advantage of the computer database management system is that it can store manage data and information for a long time [28]. This feature has a wide range of applications in many fields. Always adhere to the scientific development of the database, using new file management software with higher security. The information sharing of data between different software must adopt one or more database transformation methods and find the appropriate method according to various standards. Prevent the invasion of illegal users, ultimately reduce malicious access, hacker attacks and other problems, to ensure data and information security. In addition, it is necessary to consider the increasing speed of database information and the development trend of database technology in time to upgrade or update the database in time to improve the effectiveness of file management.

5. Conclusion

The database is core information management technology. To the rapid development of modern science and technology. Databases can help businesses become more efficient and effective. We must construct the database by following the requirements, including customer demand analysis, data collection, and data conversion. Also, we must apply accurate database data mining and analysis methods. In addition, the effectiveness of information storage is improved while solving the problems of network supervision, confidentiality and security, database application and so on. Strengthen the use of cryptography to ensure the security of database information. In order to update the database promptly, increase the accuracy and quality of the information as well as the efficiency and level of service, We should not only pay closer attention to improving and updating the database, but also fully utilize cutting-edge science, technology, and scientific applications. With the continuous progress and development of information technology, the function of database systems at the beginning of data storage then converts to the current online transaction processing and data mining.
This paper mainly summarizes the composition of the database, data mining and analysis methods. It introduces the database more comprehensively so that people can more systematically understand it and its mining methods. At the same time, some suggestions can remind people to use the database more safely to prevent improper use, resulting in loss of database information and leakage of customer information. From the beginning of simple development to the current integration of models and knowledge bases, the data security market and databases are evolving very rapidly. There is reason to believe that in the coming years, the use of databases will become more common, data security and privacy will receive more attention. The accumulation of data is wealth, and the protection of data is the protection of wealth [29]. In the future, the author will continue to focus on data information security and do more in-depth research.

Reference

Design and Implementation of art grading Line Dance Scoring System based on Android and SQLite

Ting Meng¹,a, Han Li¹,b*
¹Physical Education Department, Guangdong University of Foreign Studies, Guangzhou, China
³Mito_hi@163.com
*b Corresponding author: 123203711@qq.com

Abstract

Combined with the actual needs of scoring in line dancing competition, an Android-based competition scoring system is designed. The system is mainly divided into Android terminal and PC terminal. The administrator can enter the main interface of the PC terminal by entering the correct user name and password, and perform user management, parameter setting, data import, printing score table, and data export operations. On the Android side, the judges can perform user management, parameter setting, scoring and other operations after entering the correct user name and password. The implementation of the system effectively improves the efficiency of scoring in the line dance competitions and ensures the correctness of the final score.

Keywords: Android system; SQLite database management system; The scoring system; Line dance.

1. INTRODUCTION

In major sports competitions nowadays, sports timing and scoring system has become an indispensable basic technical facility, it plays an increasingly important role in sports competitions, and most importantly, it provides a guarantee for well-running progress of the competitions smooth progress of the games and helping the judges to make objective and fair judgments [1-2]. Sports timing and scoring system in addition to being able to assist the judges in function of timing, scoring, statistics data, implement ranking decisions, and also for such as broadcasting the sports events, provides a convenient and possible way for the audience to understand the status of the sports event in time, and the result of the match and other related information. It is also a specific equipment to achieve digital sports [3].

The scoring system is mainly used by the personnel of the competition organizer to complete the management of the information and scores of the competitors, as well as the operation of exporting data and printing the ranking table. The system can effectively manage the scoring situation, make the data management more automatic, intelligent and humanized, improve the efficiency and accuracy of the scoring of the competition ranking, and make it more organized and scientific.

At present, most line dancing competitions use EXCEL for data input, score processing and score distribution manually, which is not only slow, inefficient and error-prone [4], but also inconvenient for data sharing and various related information release of line dancing competitions. With the increase of the number of participants, the amount of the project team increases, especially in the individual events, the competition process is short, there are many people present, and the judges need to quickly and accurately give the correct score, the workload of the judges will be particularly complicated. Therefore, it is extremely urgent and imperative to develop a general information system of line dancing through the application of information technology according to the characteristics of line dancing.

The scoring system is a software and hardware specialized system which is responsible for all kinds of sports competition technology and supports data collection and distribution at the forefront of the system (competition site). It is responsible for the collection, processing, transmission and distribution of the results and results information of various sports competitions. The data of the results will be transmitted to judges, coaches, computer information systems, TV transmission and comment systems, and on-site large screen display systems through special technical interfaces (interfaces, protocols) [2]. The scoring system is a management system for all kinds of stadiums and venues, including: team information management, game scoring, multimedia display, real-time video display, post-match statistics and other functions. Because of the unrepeatability of sports competition, the timing and scoring system is an electronic service system with the core of strong real-time and high reliability. Therefore, the scoring system itself consists of an
independent collection, distribution, evaluation, display and release system, to achieve all information in real time, accurate, fast and authoritative [7]. According to different events held by the venue, their timing and scoring equipment is not the same, the specific scheme can be designed according to the competition rules of each event and the operation requirements of the venue. The main function of the system is to obtain the results of the athletes reliably and accurately, display the relevant information by the large screen equipment in time, so that the present guest of honors, judges and audience can timely understand the latest information of the competition in time. Nowadays, electronic timing and scoring system and related equipment have become indispensable electronic equipment in all kinds of sports competitions. Whether the design of timing and scoring system is reasonable is related to the stability and reliability of the operation of the whole sports competition system, and directly affects the smooth progress of the whole sports competition.

Taking the scoring work of the line dance competition in China as the research object, this paper design the scoring system of line dance, and develop and implement it on Android platform with SQLite database technology.

2. SYSTEM FUNCTION

The scoring system is to collect information, process data, monitor and quantify the information in the competition according to the competition rules of different events. According to the working progress, the scoring system should meet the following basic functions:

- (1) It can timely and accurately complete the information collection, processing and transmission of competition items, retain the original score data and have backup safety measures, so as to improve the quality of competition and work efficiency.
- (2) The system should have the ability to adapt to various competition environment conditions, and provide the ability to distribute and transmit result information to other systems through the timing and scoring system.
- (3) The system structure must be complete in function, reasonable in scheme, rigorous in structure, and minimize human intervention while meeting the requirements of intelligent and large-scale stadiums.

The function module of line dance scoring system based on Android and SQLite is mainly divided into two parts, Android terminal and PC terminal. The PC side is written by C++ programming language, the development environment is Microsoft Visual Studio 2010 running in Windows7 operating system, QT Creator 2.7.1 or above. Android: Java +SQLite, Eclipse 4.3.0 + Android SDK Tools.

2.1 Android Port

The software Android is mainly used for the scoring work of the competition in the line dance project, which plays the role of scoring device. During the competition, the judge holds a tablet to score the numbered teams. The working process is shown in Figure 1.

![Figure 1. The working process of the Android Port](image)

The administrator input the correct user name and password can enter the main interface of the system, to control such as user management, team information setting, judge information input, etc. The judge log in system can check the competition order, competition groups, review the scoring rules, scoring operations.
The administrator can modify the judges' name, password, and referee level, delete referee information, add new referee information, or import referee information in batches from the file.

In the parameter setting function, the administrator can set the information of the participating teams, such as the team name, participating groups, provinces, the number of participants, the number of judges, the highest score, the lowest score and other parameters.

In terms of scoring function, if the user type is set to "Judge" and the verification password is correct, the user can enter the judge interface to modify personal information, score and verify. When the score given by the referee is too high or too low, the referee will reject the score input by the referee and ask the referee to change his score as prompted by the system. After the final adjustment, the total score will be sent directly to the chief judge for confirmation and publicity.

2.2 PC Port

The administrator enters the correct user name and password to enter the PC main interface to manage users, set parameters, import data, print rating tables, and export data. The working process chart of the operation is shown in Figure 2.

![The working process of the PC Port](image)

Figure 2. The working process of the PC Port

The user management function is used for user management. You can add and delete users and modify the passwords of corresponding users.

Parameter setting function, which is mainly used to set the scoring information of the competition, competition groups, the number of participants and so on, to ensure the integrity of the scoring data. Group setting is very important, because there are many groups of line dancing events, many teams will compete in many events at one time, so they must be consistent with the actual competition group and the competition group set by Android before the competition. Otherwise, in the later data import, the score will be inconsistent and the data cannot be imported because of the failure of verification.

Data import function, which is used to import judge score data from Android terminal. Before use, connect android tablet with PC terminal through data cable, and select the corresponding judge number or name information after connection to import.

The function of printing the score sheet is used to import the data of the corresponding referee into the PC and print it out in the built-in Excel template format for the chief referee to sign and confirm for archival use. After clicking the "print" button, the dialog box will pop up to select the corresponding score for printing out.

The function of printing the score sheet is used to import the data of the corresponding referee into the PC and print it out in the built-in Excel template format for the chief referee to sign and confirm for archival use. After clicking the "print" button, the dialog box will pop up to select the corresponding score for printing out.
3. SYSTEM FUNCTIONS AND IMPLEMENTATION

3.1 Database Design

SQLite database is used in the system. SQLite is a C language library that implements a small, fast, self-contained, highly reliable, full-featured SQL database engine. SQLite is the most used database engine in the world. SQLite is built into all mobile phones and most computers, bundled with countless other applications that people use every day. SQLite is stable, cross-platform, and backward-compatible, and SQLite database files are often used as containers for transferring rich content between systems and as a long-term archiving format for data. SQLite source code is in the public domain, known as open source, and everyone is free to use it for any purpose without restriction.

3.2 Android Development Environment

Android, the operating system for smart phones and other devices, was developed by The company and later bought by Google. The Android platform is based on the Linux operating system and features many applications developed by Google and third-party developers.

3.3 Function Module Implementation

According to the requirements of the function analysis, specific implementation of each module function. The following takes the judging and scoring function as an example for detailed introduction.

In order to facilitate the judging and scoring, the main scoring interface sets the information of the upcoming competition, competition categories and groups, participating teams, etc., and displays the judges' scores in real time. At the same time, the main interface will also have the order of appearance, which is convenient for the judges to record their own scores and rankings in the process of competition, and also for the reference of the following scoring work. The interface after each judge input score is shown in Figure 3.

![Figure 3. The interface of the scoring system](image)

Considering that there will be many teams in each group, the judges are likely to lose the impression of the teams in front after a long time of judging. The scoring function also sets up the function of checking the order of the matches in the group. This function is to ensure the accuracy of ranking results, and assist the judges to give accurate scores. Its interface is shown in Figure 4.

![Figure 4. The interface of the scoring system](image)
Because of dozens of classification projects being subjective in nature, namely different judges have different requirements for athletes and sports teams, even so, the team scoring judgment should be according to the rules, because the line dance project at the same level of other teams in the competition, levels are close to, in general, between the team and the team score does not differ too much. According to this feature, the scoring function sets the scoring prompt function. After the judges in this mode score the corresponding team, click OK, and the score will be sent to the panel of the judge. If the score is within the reasonable scoring range, the judge can return to the main scoring interface, the page will pop up the information of the next team, and the judge can continue scoring; If the judge score is beyond the range, the referee will give a hint, the arrow up is just to give too low points, need extra points; The arrow down is overrated and needs to be subtracted. The interface is shown in Figure 5.

### Table 1: Sample Scores

<table>
<thead>
<tr>
<th>Entrance no.</th>
<th>Team</th>
<th>Participants</th>
<th>Group</th>
<th>Item</th>
<th>No. of Participants</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jining College</td>
<td>Lu Ting</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>8.75</td>
</tr>
<tr>
<td>2</td>
<td>Guangzhou University of Chinese medicine</td>
<td>Lou Kejie</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>9.02</td>
</tr>
<tr>
<td>3</td>
<td>Jining College</td>
<td>Gao Zilin</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>8.31</td>
</tr>
<tr>
<td>4</td>
<td>Jining College</td>
<td>Wu Siao</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>8.43</td>
</tr>
<tr>
<td>5</td>
<td>Guangzhou University of Chinese medicine</td>
<td>Chen Jun</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>8.10</td>
</tr>
<tr>
<td>6</td>
<td>Zhang College</td>
<td>Zhao Wenzhi</td>
<td>College Group</td>
<td>Individual</td>
<td>1</td>
<td>9.11</td>
</tr>
</tbody>
</table>

4. **Conclusions**

As people pay more and more attention to sports, they put forward higher and higher requirements for the holding of sports events. Fair, open and fair competition is the first condition. In order to make the results of sports more fair, the use of electronic timing and scoring systems can greatly reduce the error, helping the judges to make objective and fair judgments.

The rapid development of information technology brings a new era for the development of sports. Through the research and development of the scoring system for line dancing, the scoring efficiency has been greatly improved, which not only avoids mistakes and omissions in the process of scoring, but also prevents artificial modification of scores, so as to ensure fairness and fairness in scoring and accurate ranking. This study is an effective exploration of the application of mobile terminals in the field of process scoring of sports and art competitions, which can provide a certain reference for the development of sports and art projects. The software developed has practical significance for provincial and municipal organizations to carry out unified scoring of sports and art competitions.
ACKNOWLEDGMENT

This work was not supported by any funds. We want to show sincere thanks to those who have helped this research and all the authors of the references.

REFERENCES

A High Performance Bitcoin Trading Strategy Prediction Model

YunJing Zhang1, JiaXu Sun2, HengBin Liu3, ShaoQing Shi4, and QiuFeng Wang5

1 School of Computer and Information Engineering, Henan University of Economics and Laws, Henan Province, China
2 Kansas International College, Zhengzhou Sias College, Henan Province, China
3 Mathematical Finance, International College, Zhengzhou University, Henan Province, China
4 Communication Engineering, International College, Zhengzhou University, Henan Province, China
5 Academy of Cyberspace Security, International College, Zhengzhou University, Henan Province, China

Email: zyjdeyouxiang163@163.com (author: Yunjing Zhang)

Abstract

First released as open source in 2009 by the pseudonym Satoshi Nakamoto, Bitcoin is the longest running and best known cryptocurrency. Bitcoin’s transaction history is characterised by openness and transparency, and Bitcoin has become an important part of financial transactions. Therefore, it is increasingly important to be able to make accurate predictions about the development of the Bitcoin market. In this study, we construct a prediction model for bitcoin trading strategies based on the LightGBM algorithm, and show that our model has an accuracy of over 95.1%. The results show that our model achieves an accuracy of over 95.1% and has a higher performance compared to popular machine learning models.

Keywords: bitcoin, LightGBM, trading, Machine learning, feature engineering

1. Introduction

Bitcoin, as a cryptocurrency, is limited in quantity but can be exchanged for most countries’ currencies, which is arguably the biggest value of Bitcoin at the moment. For cryptocurrency trading, the root of its biggest problem comes from the instability of the market. Since the cryptocurrency market is traded around the clock, the market reflects rapid changes, making it difficult to dynamically track cryptocurrency positions. But at the same time, this is something that automated trading algorithms and trading robots are good at handling.

Yang Xiaochen and Zhang Ming[1] elaborated on the operation principle of Bitcoin and dissected the typical characteristics of Bitcoin. Li Jing [2] used BP neural networks to build a bitcoin market forecasting model and concluded that the more recent the data, the better it is for accurately predicting the price of bitcoin, and the more short-term the forecast, the better it fits the bitcoin development trend. Xu Bo [3] found that price risk is the most significant risk in the Bitcoin market. Zhao Lei and Liu Qing [4] tested and predicted the bitcoin price bubble based on the established asset bubble theory. Hanyi Zhong [5] focused on the analysis of Bitcoin risk and its legal regulation. Ai Qing [6] focused on the predictability of bitcoin trading trends, the implementation of traditional models applied to bitcoin trading trend prediction.

Various machine learning algorithms can be used to generate trading signals to try to predict market movements. By using machine learning algorithms to make predictions, we can classify the next day’s movements into three categories: market up (long positions), market down (short positions) or market sideways (no positions). Since we get the predictions through the algorithm, we gain a huge advantage in making decisions.

This study applies the machine learning technique LightGBM for bitcoin trading strategy prediction. The original bitcoin dataset is subjected to data cleaning and feature engineering to extract features, and a bitcoin trading strategy prediction model based on the LightGBM algorithm is constructed. By comparing the current common classification algorithms, LR (logistic regression), Xgboost and Random Forest algorithm, etc., we compare the prediction results of Bitcoin trading strategies and make a reasonable analysis of the results in order to draw certain conclusions.
2. Method

2.1. General framework

The Bitcoin trend prediction model uses the LightGBM algorithm. The overall framework is shown in Figure 1.

![Figure 1. Overall structure](image)

1. Loading the files of the Bitcoin model dataset. 2. Perform data cleaning. 3. Feature engineering of dataset files. 4. Slicing the data set into a training set and a test set. 5. Training on the training set using the LightGBM algorithm to obtain model data. 6. Using model data to make predictions on a test set. 7. Validate the prediction effect of LightGBM algorithm.

2.2. LightGBM

LightGBM (Light Gradient Boosting Machine) [7] is a framework for implementing the GBDT algorithm, which supports efficient parallel training and has the advantages of faster training, better accuracy, and distributed support for fast processing of large amounts of data. To solve the time consuming problem in the large sample high latitude data environment, LightGBM adopts two solutions, GOSS (Gradient-based One-Side Sampling) and EFB (Exclusive Feature Bundling). We call the gradient boosting tree (GBDT) using GOSS algorithm and EFB algorithm as LightGBM, and the principle is as follows.

2.2.1. GOSS algorithm

Instead of using the used sample points to calculate the gradient, the GOSS (Gradient Based One-Side Sampling) method samples the samples to calculate the gradient. Therefore, in order to maintain the accuracy of the information gain evaluation, when we downsample the samples, we keep these sample points with large gradients, while the sample points with small gradients can be randomly sampled proportionally. The specific algorithm description is shown in Table 1.
Table 1. GOSS algorithm

| Inputs: training data, number of iterations d, sampling rate a for large gradient data, sampling rate b for small gradient data, loss function and several weak learners |
| Output: Strong learner for training |

1. Sort the samples in descending order according to their absolute value gradients
2. The first a x 100% of the sorted sequence is selected as a subset of the large gradient samples
3. For the remaining sample set (1 - a) x 100% of samples, b (1 - a) x 100% of samples are randomly selected as a subset of the small gradient samples
4. Merge large gradient samples and sampled small gradient samples
5. Multiplying a small gradient sample by a weighting factor \( \frac{1-c}{b} \)
6. Learn a new weak learner using the sampled samples above
7. Repeat steps 1 to 6 until the specified number of iterations is reached or until convergence is achieved

2.2.2. EFB algorithm

The EFB (Exclusive Feature Bundling) method bundles some features together to reduce the dimensionality of the features and reduce the consumption of finding the best cut point. Data sampling and feature sampling are performed in the LightGBM implementation, making the training speed of the model further reduced. The feature sampling is also different from the general feature sampling in that it binds mutually exclusive features together thereby reducing the feature dimensionality. The main idea is that data at high latitudes in practical applications are often sparse (e.g., one-hot encoding), and many features in the sparse feature space are mutually exclusive (e.g., rarely have non-zero values at the same time). This means that we can safely bind mutually exclusive features together to form a single feature, thus reducing feature dimensionality. LightGBM binds mutually exclusive features together based on a histogram (histograms) approach. The specific algorithm description is shown in Table 2.

Table 2. EFB algorithm

| Input: feature F, maximum number of conflicts K, graph G |
| Output: feature bundle set bundles |

1. construct a graph with weights on the edges whose weights correspond to the total conflict between features
2. Sort features in descending order by their degree in the graph
3. Check each feature in the ordered list and assign it to an existing bundling with a minor conflict (controlled by \( \gamma \)), or create a new bundling.

3. Experiments

3.1. Data set

In terms of average daily trading volume, we use data from Bitstamp [4], which is one of the largest and most authoritative bitcoin exchanges in terms of time. The data covers a wide range, including all bitcoin trading prices from January 2012 to May 2017. We extract important momentum indicators and trends from the obtained data and use the momentum indicators and trends as features to be added to improve the accuracy of our predictions. The original dataset of Bitcoin is shown in Figure 2.

![Figure 2. Original dataset](https://example.com/fig2.png)

3.2. Feature engineering

The signaling problem of the Bitcoin Prediction Trading Strategy is defined in the classification framework, where a value of 1 for the predictor variable means we are going to buy and a value of 0 for the predictor variable means we are...
going to sell. This predictor variable is determined by comparing the short-term and long-term price trends.

First and foremost, creating target variables determines to some extent the merit of our model, and we will create target variables for our model. We will divide the prices into short-term and long-term prices, where the 10-day rolling average is the short-term price and the long-term price is defined as the 60-day rolling average.

The current data for Bitcoin includes date, open, high, low, close and volume. Using this data, we calculate momentum indicators such as Moving average, Stochastic oscillator %K, Relative strength index (RSI), Rate of change (ROC), and Momentum (MOM). The following is a detailed description of several of them.

**Moving Averages**

The principle of moving averages is simple: they indicate price trends by reducing the amount of noise in the data series.

**Momentum (MOM)**

Momentum is, in short, the rate of change or acceleration of the trend in trading volume or security prices, i.e. the rate at which they change in price.

### 3.3. Evaluation indicators

Among machine learning problems, it is usually necessary to build models to solve specific problems. For the goodness of the model, that is, the generalization ability of the model, some evaluation metrics are usually defined to measure the strengths and weaknesses of the model, such as accuracy, precision, recall, and F1 value. Definitions denote true examples, denote true negative examples, denote false positive examples, and denote false negative examples for accuracy, precision, recall, and F1 value.

Accuracy is the percentage of the total sample that predicts the correct outcome.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

(1)

Precision is the probability that a sample is actually positive among all the samples predicted to be positive.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

(2)

Recall is the probability that a positive sample will be predicted from an actual positive sample.

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

(3)

F1 considers both precision and recall, and is the summed average of precision and recall, which better reflects the combined effect of model performance.

\[
F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]

(4)

### 3.4. Analysis of results

The original dataset was input into the LightGBM algorithm model after data cleaning and feature engineering, and Table 3 shows the confusion matrix obtained after training.

<table>
<thead>
<tr>
<th>Type name</th>
<th>0 (real)</th>
<th>1 (real)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (prediction)</td>
<td>8943</td>
<td>470</td>
</tr>
<tr>
<td>1 (prediction)</td>
<td>504</td>
<td>10083</td>
</tr>
</tbody>
</table>

Also, to better demonstrate the superiority of our model, we compared our model with LR (logistic regression), the XGBoost and Randomforest models, and the results are shown in Figure 3.
The comparison chart shows that the accuracy, precision, recall and F1 values are better than the classical LR, XGBoost and Randomforest algorithms when using the LightGBM algorithm for classification.

4. Conclusion

The article introduced the background and significance of building a LightGBM-based Bitcoin trading strategy prediction model, and proposed a new research method based on the classical research model. The original bitcoin dataset collected was firstly cleaned and feature engineered, and then the LightGBM algorithm was applied to predict the bitcoin trading strategy. The experimental results show that our proposed LightGBM model outperforms the current classical model.

REFERENCES

[1] Yang, X. Chen and M. Zhang, Bitcoin: Operating principles, typical features and prospects [j]. Economic and Management Science.F49;F821
[6] Ai Qing, A study of bitcoin trading trend forecasting [j]. Economic and Management Science.F822; F832.5; TP18
[8] https://www.bitstamp.net/
Abstract

The power grid business has the characteristics of wide coverage, diverse terminal equipment types, large numbers, and rich business scenarios. The blockchain technology itself has the characteristics of non-tampering and traceability, which is highly compatible with the opening of the power grid business and can be powerful supporting the digital transformation strategy of the power grid company. In order to ensure the authenticity and reliability of electric power data, data security and reliable storage technology was studied based on blockchain technology. On the data source side of the power system, a set of oracles mechanism for the grid data on-chain process was designed to achieve data reliability and consistent transmission. In addition, to prevent data damage or malicious tampering after power grid data on the blockchain, a hybrid data storage scheme based on shadow chain was proposed to realize data safe storage and improve value-added services such as data storage and sharing.

Keywords-power grid; power data; storage; blockchain;

1. Introduction

In September 2020, at the 75th Session of the United Nations General Assembly, General Secretary Xi Jinping made a solemn commitment of carbon peak and carbon neutral. Later, at the ninth Meeting of the Central Financial and Economic Commission, General Secretary Xi Jinping again made important deployment on carbon peak and carbon neutrality, stressing the need to build a new power system with new energy as the main body and defining the direction of China’s energy and electric power transformation and development under the background of “dual carbon”. The new power system has the characteristics of “extensive interconnection”, “intelligent interaction”, “flexibility and flexibility”, and “safety and controllability”[1].

Taking the extensive interconnection of new power systems as an example, at the beginning of 2019, tate Grid Corporation of China (SGCC) has been connected to more than 540 million terminals (sets) such as smart meters, and the daily increase of data collected exceeds 60TB level. According to the planning of SGCC, it is expected that the access terminal devices will exceed 1 billion in 2025 and 2 billion in 2030 [2].

The new power system faces key problems such as the trusted access of massive terminals and the trusted storage and sharing of massive data. Research on terminal access will provide key technical support for the construction of IOT (Internet of Things) in power systems, ensure the comprehensive ability to perceive and interact with the terminal status of the power grid, ensure the safe operation of the power grid, and improve the quality and efficiency of power grid operation and social comprehensive energy efficiency.

The essence of blockchain technology is a shared database, and the data or information stored in it has the characteristics of “unforgeable”, “full traces”, “traceable”, “open and transparent” and “collective maintenance”. On October 24, 2019, General Secretary Xi Jinping emphasized during the 18th collective study of the Political Bureau of the CPC Central Committee: “The integrated application of blockchain technology plays an important role in new technological innovation and industrial transformation. We will take blockchain as an important breakthrough for independent innovation of core technologies, clarify the main direction of attack, increase investment, focus on conquering a number of key core technologies, and accelerate the development of blockchain technology and industrial innovation”. General Secretary Xi clearly stated that it is necessary to promote the combination of the underlying technology services of the
blockchain and the construction of new smart cities, and explore the promotion and application in the fields of information infrastructure, energy and power, etc[3]. On April 20, 2020, the National Development and Reform Commission clarified for the first time three aspects of new infrastructure, among which information infrastructure includes new technology infrastructure represented by artificial intelligence, cloud computing, and blockchain.

At present, the application of blockchain in the field of energy and power has gradually landed, and it has formed ten scenarios such as new energy cloud, power transaction, high-quality service, comprehensive energy, material procurement, smart finance, smart law, data sharing, safe production, and financial technology[4-8].

In this paper, a set of oracles mechanism for the grid data on-chain process was proposed to achieve data reliability and consistent transmission. In addition, a hybrid data storage scheme based on shadow chain was proposed to realize data safe storage and improve data added services such as data sharing and transaction.

2. Power data collaborative governance technology based on blockchain

In the power blockchain system, the application of smart contracts will greatly improve the production efficiency, but the smart contracts themselves do not have the ability to automatically trigger and require external information to be provided by the outside world. The oracle machine writes external information into the blockchain to complete the data exchange between the blockchain and the real world. It allows certain smart contracts to react to the uncertain external world, and is the only way for smart contracts to interact with the outside world, and it is also the interface between blockchain and the real world[9].

2.1. Oracle Technology

Based on the existing oracle solutions, the general technical architecture and workflow of the oracle mechanism are shown in Figure 1.

![Fig.1 The general technical architecture of the oracle mechanism](image_url)

First, a user or software creates and issues a smart contract. The universal oracle mechanism starts with the requester creating a smart contract specifying the data needed to trigger the execution and deployment of the contract on the blockchain. The requester can be a user or a component of a software system. In some cases, the requester can directly trigger the oracle to include a value into the blockchain for future use.

Second, the oracle receives a request from the smart contract. Among them, the oracle can be centralized or decentralized, and the centralized oracle can automatically identify the requirements specified by the smart contract. Distributed oracles contain multiple redundant oracles that provide the same functionality to check external state. Some oracles are people with blockchain accounts who can manually enter oracle data and sign transactions.

Then, the oracle interacts with the external data source to get the data. If an automated oracle is deployed, it will communicate with its external data sources (such as physical sensors or web services) to collect the required data.

Finally, the oracle injects the data into the blockchain system, and the requester can get the data when executing the smart contract.

The types of oracles can be divided into the following dimensions. According to the number of oracle nodes, it can be divided into three categories: centralized oracles, trusted alliance oracles, and decentralized oracles. Among them, centralized oracles rely on authoritative central institutions as nodes to provide data, and there is a single point of failure problem. Decentralized oracles are provided by multiple nodes to obtain data and provide services, and create a data
provider network through a distributed consensus mechanism to ensure data reliability; Alliance oracles are provided by designated distributed trusted individuals or institutions, and are a special form of decentralized oracles.

According to the different sources of data obtained by oracles, they are divided into software oracles, hardware oracles and artificial oracles. Among them, software oracles are mainly obtained from third-party service provider applications or websites. The hardware oracle mainly obtains data such as ambient temperature, humidity and location information from hardware facilities such as IoT sensors and RFID data collectors; artificial oracles mainly refer to data such as human input.

According to the interaction direction between smart contracts and external data, oracles are divided into input oracles and output oracles[10]. Among them, the input oracle machine provides the external data of the smart contract to the smart contract, and starts to execute after receiving the information. Output oracles transmit data on smart contracts to external resources.

2.2. Power Data Collaborative Governance Design

The basis for the collaborative management and application of power data is the standardization of power data, and different data formats corresponding to different power terminals[11]. The standardization of power data refers to the unification of data names, data definitions, data types, assignment rules, scientific processes and other aspects of the data scattered in various systems to form a universal on-chain data. Research data collaborative governance, carry out data standard co-construction, ensure the high quality of smart terminal data on the chain, maintain the consistency of the data model, improve the efficiency of data collaborative governance on and off the chain of the power grid alliance chain, and ensure the openness, sharing and security of data thereby improving the level of on-chain application of smart terminal data[12-13].

![Fig.2 The on-chain and off-chain collaboration framework for power data and blockchain systems](https://example.com)

The realization of the goal of on-chain and off-chain data collaborative governance is mainly based on the characteristics of blockchain distributed ledger, encryption algorithm and consensus mechanism, through the structure and method flow of collaborative governance model, data sharing and access control, data update and smart contract. In view of the authenticity of off-chain power data, it mainly consists of data processing module, identity authentication module, authority management module and supervision management module, which is mainly responsible for collecting, processing and transmitting panoramic power grid data, verifying user status, rewarding users and supervising user behaviors.

For the data on the power block chain system, responsible for storing and reading the data information of power block chain. Power block chain mainly stores various intelligent power terminal data of the upper chain and identity authentication information of each user entity and other related power information. Each block is composed of block head and block body. Block head mainly stores block identification information, including identification number, time stamp, Merkle root, etc. Block main body mainly stores power information data, including real-time information, key instruction, address source, equipment operation, data lock time, identity authentication information of each user entity, etc. The new block is verified by consensus algorithm and linked to the power block chain after passing, making the chain block height gradually increase.
3. Power data security storage technology based on blockchain technology

Currently, power data typically has the following risks. First, power data is usually stored in a data center or cloud data center. Attackers may tamper with the power data in the data center or cloud server, which seriously threatens data privacy. Secondly, relying on centralized storage, once the centralized platform fails or completely collapses, a large amount of data will be lost, threatening the security of power data and the reliability of storage[14-16].

In the process of processing and storing power data, there are risks such as unauthorized reading and writing, natural disasters in the data dispatch center, and server failures. Therefore, protecting power data from malicious tampering is an urgent problem that needs to be solved in the current power industry[17-18].

3.1. Blockchain System shadow chain Technology

Based on the blockchain system, shadow chain expansion technology is designed. The blockchain system for data storage is called the main chain, and the blockchain system for operation records is called shadow chain. A log storage system that satisfies the reliability and validity constraints on the data source, the data immutability in the storage method, and the additional constraints, as shown in Figure 3. Among them, the data source refers to the log data generated by monitoring the data reading behavior on the database (main chain). The reliability of the data source means that all log data generated can be stored in shadow chain under the premise of monitoring all data reading behaviors. The validity of the data source means that all log data generated are real and valid, rather than fake data created out of nothing or tampered with. The immutability of data in the storage method means that the data cannot be tampered with during the storage process. The additional constraints include two points. On the one hand, the impact of the introduction of shadow chain on the latency of reading and writing data to the original data storage system cannot exceed the threshold. On the other hand, it means that the storage overhead brought by the introduction of shadow chain cannot exceed the critical value. When the above constraints are met, it is considered that shadow chain can achieve the purpose of giving data traceability.

![Fig.3 Schematic diagram of shadow chain technology](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

3.2. Design of secure storage of power data based on shadow chain technology

In the actual scenario of the power grid, there is a need for data sharing between multiple regions or multiple institutions. In order to solve the disputes and liability issues after data leakage caused by data sharing, shadow chain technology is introduced to solve related difficulties.

The detailed design of constructing shadow chain is as follows. First of all, in the power grid, some servers need to be provided as nodes to join shadow chain to form a blockchain network. Only these regional institutions have permission to access and write in this blockchain network. Secondly, deploy a listener program on the original data system to provide a data source for shadow chain, and all read operations of shared data are based on the listener program. Finally, when the shared data is leaked, the data in shadow chain will serve as a credible basis to help resolve disputes and liability issues. In this process, shadow chain, as a blockchain, in order to ensure that its fairness is not destroyed, on the one hand, regulators can participate in supervision to ensure that there will be no oligarchs in the network, on the other hand, all parties in the network Institutions can also monitor each other to avoid an oligarchic situation. Specifically, first, when regional institution A initiates a request to read the data shared by regional institution B, the read behavior will be intercepted by the listener and an operation log will be generated. Then, the operation log will be submitted to shadow chain for storage. Finally, when the submission is successful, the listener will execute the data read request and return the result to regional agency A.
Ceph is an open source, scalable, high-performance and single point of failure distributed storage system [17-19]. This paper proposes the combination of Ceph distributed storage system and block chain extended shadow chain, and designs a distributed trusted storage and management method for multi-system data on power grid chain. The technical architecture is shown in Figure 4. Through monitoring the data on the database (main chain) read the behavior of log data, can solve the problem of data are difficult to trace, and then through specific data set to create a master database and multiple copies of the database, by synchronizing on different master copy of log data in the database, for each system terminal of data exchange and collaboration, to improve the consistency of the data; Finally, Ceph is used to build a distributed storage system, combining with the high concurrency mechanism of distributed block chain to improve data storage computing resources, which solves the problem of limited storage computing resources of block chain technology, so as to adapt to large-scale user access in power grid.

Theoretical analysis shows that the research on the hybrid storage technology of shadow chain and Ceph has significantly reduced the data storage overhead, which can prove that this method effectively solves the problem of limited storage and computing resources in blockchain technology[20]. Through the analysis of the distributed data storage technology and the efficient data redundant storage mechanism of multiple management and maintenance subjects, the unification and integration of different versions of power grid can be realized;The system log is processed on the chain through the blockchain extension of shadow chain, which can solve the problem of using the data tampering mechanism to attack in the process of power grid transaction, and realize the anti-tampering and traceability functions of data.

4. Conclusion

With the development of blockchain technology and the increasing demand of blockchain in power business scenarios, the integration of power and blockchain has become the main factor limiting the application of blockchain in power scenarios. Based on the results and discussions presented above, the conclusions are obtained as below:

(1) There are still technical shortcomings in the application of energy and power, and it is urgent to overcome key core technologies such as on-chain and off-chain collaboration, secure data storage and sharing, efficient and reliable smart contracts, and cross-chain interaction.

(2) This paper focuses on the analysis and design of the power data on-chain based on oracle technology, and the reliable storage and sharing based on shadow chain after the power data is on-chain.

(3) With the proposed method, the security, reliability and practicability of off-chain data on-chain, and the security and reliability of on-chain data were improved.

Acknowledgments

This work is supported by the Science and Technology Project of State Grid Corporation of China: “research on digital
mutual trust and cross domain sharing technology for “One Network Service” of electric power. (No.: 5700-202153179A-0-0-00).

References


Research on adaptive model of English translation based on data fusion

Ruying Huang
Nanjing University of Finance & Economics Hongshan College Nanjing, 210000 China
* Corresponding author: carolineh42@126.com

ABSTRACT

This research is based on the attention mechanism English translation adaptive model. After analyzing the key factors that affect English language translation, the attention mechanism is used to extract the detailed features of such factors in each region to form a feature sample set, and the feature sample set is fused and normalized, so as to obtain a brand-new feature sample set. Input to build an English language translation model and output the translation results. According to the results, the overall translation effect of the model is predicted. The results show that the prediction model of this method has high prediction accuracy in training and testing.

Keywords: Attention mechanism; English; Translation; Adaptive model

1. INTRODUCTION

As the most popular machine translation method at present, neural machine translation method loads specific algorithms into the neural model framework, and optimizes the model in a node to node manner. The current mainstream neural machine translation algorithm is NMT model, and the neural framework uses coding-decoding structure. The framework converts the language to be translated into input data[1-2]. The input data is first encoded into a special vector structure by an encoder, and then the vector structure is decoded into the translated language by a decoder. The algorithm is connected between the encoder and the decoder through the attention mechanism. The traditional decoding decoding algorithm framework focuses on the translation of single sentences, that is, the translation of specific sentences. If the constructed algorithm model is used to translate word by word, it will cause a sense of separation between sentences. This is not conducive to the translation of the text and will result in the text the semantic discontinuity between sentences and the whole translation is not smooth. In addition, the translation of a single sentence may cause ambiguity[3-4].

Although traditional human translation methods have the advantages of stable, reliable and high-quality translation, they also have the disadvantages of high human cost and unable to process large quantities of data quickly. In today's big data era, in the environment of massive Internet information, the machine translation system can use computers to carry out efficient and large-scale automatic translation, and it is rapidly favored by people[5-6]. Similar to the process of human translation, the process of machine translation is divided into two stages: the first stage is to understand the meaning of the source language text (also known as the source text); In the second stage, based on the source language text, the target language text (also called translation) is generated according to the grammar and syntax of the target language. Since the pioneer of information theory, American scientist W. weaver and British engineer A.D[7-8], both proposed the idea of automatic language translation by computer in 1947, machine translation has experienced rule-based machine translation model and case-based machine translation model in the development history of more than 70 years, machine translation model based on statistics and machine translation model based on Neural Network type. To build a practical and rule-based machine translation system, it is often necessary to establish various knowledge bases to describe the lexical, syntactic and semantic knowledge of the source language and the target language, and even to describe more abundant and complex common sense of life and world knowledge. This knowledge base often needs many trained experts to create and maintain. With the arrival of the big data era, how to ensure that the newly introduced knowledge does not contradict the old knowledge has become a new problem. Therefore, the acquisition and maintenance of knowledge has become the bottleneck of the development of rule-based machine translation systems[9-10].

Compared with the statistical based machine translation model, the adaptive language translation model has the following significant advantages:

(1) Vector representation of automatic learning data
The statistical based machine translation model requires human experts to design the implicit structure, design the corresponding features and model the corresponding translation process. However, the existing neural machine translation model can automatically learn to compress the semantic, grammatical and other information contained in the input sentence into the generated vector representation through the neural network, without manual intervention. At the same time, this vector representation makes sentences with the same semantics but different syntactic structures aggregate, while sentences with different semantics but the same syntactic structure are separated in the vector space.

(2) Long distance context information modeling

One of the challenges faced by the statistical based machine translation model is how to adjust the order of the translated text to make it a smooth and reasonable natural language sentence. However, statistical machine translation cannot deal with this problem well. The reasons are: on the one hand, it uses local features and dynamic programming to perform approximate search in the exponential structure space, and its modeling ability is limited; On the other hand, it adopts the discrete data representation method constructed by hand, which leads to serious data sparsity problem, and it is difficult to capture and model long-distance context dependence, which is easy to lead to accurate word by word translation and non smooth and unreasonable translation sentences. In contrast, the neural machine translation model solves the problem of data sparsity through the dense sentence vector representation learned by the neural network. At the same time, the long-distance context information is modeled by using the cyclic neural network models such as the long-term memory network and the gating unit and the attention mechanism, which greatly improves the fluency of the translation.

2. ESTABLISHMENT OF AN ADAPTIVE ENGLISH LANGUAGE TRANSLATION MODEL

The attention mechanism is a process of feature extraction of more detailed features[11-12]. The main principle is to quickly scan the whole scene of its visual interval through the human visual system, select the key target interval from it through the signal processing mechanism of the brain, and put more attention resources into the interval. In the process of English language translation, many key value pairs can be searched through the attention mechanism, and the coincidence degree of such key value pairs can be obtained[13-15]. The coincidence degree is proportional to the amount of attention resources allocated.

$L_i - W_i$ represents the weight of influencing factors of energy consumption; The question is represented by $Q$, which represents the amount of English words. Using the feature information of the key influencing factors extracted by the attention mechanism to build a prediction model can improve the prediction accuracy and efficiency of the model. Wherein the calculation process for attention is:

Step 1: calculate the coincidence between the weights of each key influencing factor to obtain the weight. Here, the splicing method is selected to calculate the coincidence degree. The calculation equation is:

$$g(Q, L_i) = v_a [Q; L_i]$$

In the above formula, the splicing coefficient is expressed by $v_a$.

Step 2: weight normalization. The calculation formula of the weight obtained in the function normalization step (1) is:

$$\text{Soft max}[g(Q, L_i)] = \frac{\exp[g(Q, L_i)]}{\sum_j \exp[g(Q, L_j)]}$$

Step3: attention calculation. Attention is obtained by weighted summation and normalized weight and influence factor weight. The operation equation is:

$$A(Q, L, W) = \sum_i \alpha_i W_i$$

In the above formula, the normalized weight is expressed by $\alpha_i$. 
Particle swarm optimization (PSO) proposed by Dr. Kennedy in 1995 is a global evolutionary algorithm. Due to its small parameter adjustment range and simple calculation process, this method is suitable for solving complex problems. The concept principle of PSO algorithm comes from the predator-prey behavior of birds. When solving complex problems, each bird is compared to a particle, which has two parameters of "speed" and "position". On the premise that the optimal objective function of the evaluation system is known, the particles are substituted and the iterative calculation is carried out continuously (this process can be regarded as the process of the particles adapting to the objective function). Finally, the optimal solution of the evaluation objective function of the system is obtained. The judgment matrix is constructed by the tomographic analysis method. The weight value of the indexes in the judgment matrix can be used as the coordinate value of the particles in different dimensions. After the iterative process, the optimal weight value of each index in the judgment matrix of each layer is obtained.

The particle iterative calculation steps are as follows:

1. Determination of particle dimension. Combined with the characteristics of particle swarm optimization, the unknown weight of the influencing factor is regarded as the dimension coordinate of each particle. After iteration, the coordinate value of each dimension of the optimal particle is the weight value of the influencing factor of the judgment matrix.

2. Set the initial position and speed of particles. The initial position of the particle is denoted as \(X_i(0)\), and the initial velocity is denoted as \(v_i(0)\). The size (including dimension, quantity, etc.) is set according to the actual situation. The larger the size, the more accurate the calculation results and efficiency are. The position and speed of particles need to set the feasible range according to the situation. The particle calculation results in this paper represent the weight value of the influencing factors. If the weight value range is between 0 and 1, the feasible range is \([0,1]\). Therefore, the initial position \(X_i(0)\) and the velocity \(v_i(0)\) are any number of \([0,1]\).

3. The individual optimal position \(S_i(T)\) and the global optimal position \(S_g(T)\). When the optimal objective function value of a single particle is obtained in the t-round iteration, the coordinate of particle I is the individual optimal position \(S_i(T)\); The corresponding coordinates of the particles with the best objective function values of all particles in this round are the global optimal position \(S_g(T)\).

4. According to the PSO evolution equation, the coordinates and velocities of the iterative particles are updated. The calculation formula is as follows:

\[
x_i(t+1) = x_i(t) + 0.1v_i(t+1)
\]

\[
v_i(t+1) = \omega v_i(t) + c_1r_1(t)[s_i(t) - x_i(t)] + c_2r_2(t)[s_g(t) - x_i(t)]
\]

In equations (4) - (5), \(V_i(T + 1)\) and \(V_i(T)\) respectively represent the spatial velocity of particle I at time \(t + 1\) and \(T\); \(X_i(T + 1)\) and \(X_i(T)\) represent the spatial positions of the particles I at times \(T + 1\) and \(T\), respectively; \(\omega\) It indicates that the current particle inherits the state of the previous particle. The smaller the value, the better the local optimization ability is. On the contrary, the better the global optimization ability is. In this paper \(\omega\) Of 0.1; \(C_1\) and \(C_2\) represent the group learning rate and the individual learning rate. \(R_1\) and \(R_2\) are random numbers from 0 to 1, representing the randomness of the particle change process. According to the previous literature and experience, the value of \(C_1\) and \(C_2\) in this paper is 0.2.

5. Repeat steps (2) - (4) to obtain the optimal coordinate position \(sg_{best}(T)\) of the evaluation system and output the optimal weight \(\omega_i\).
3. TEST RESULTS AND DISCUSSIONS

Select 20 documents at random and output the translated language results by using the adaptive English language translation model. Compare with the traditional model to verify the availability of the model proposed in this study. The output results are shown in Table 1.

<table>
<thead>
<tr>
<th>Data</th>
<th>Traditional model</th>
<th>The model proposed in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>4</td>
<td>0.28</td>
<td>0.30</td>
</tr>
<tr>
<td>5</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>6</td>
<td>0.47</td>
<td>0.48</td>
</tr>
<tr>
<td>7</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>8</td>
<td>0.50</td>
<td>0.52</td>
</tr>
<tr>
<td>9</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>10</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>11</td>
<td>0.11</td>
<td>0.29</td>
</tr>
<tr>
<td>12</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td>13</td>
<td>0.27</td>
<td>0.29</td>
</tr>
<tr>
<td>14</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td>15</td>
<td>0.35</td>
<td>0.39</td>
</tr>
<tr>
<td>16</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>17</td>
<td>0.41</td>
<td>0.44</td>
</tr>
<tr>
<td>18</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>19</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td>20</td>
<td>0.37</td>
<td>0.43</td>
</tr>
</tbody>
</table>

According to the experimental results, we can find that the model proposed in this study has better advantages than the traditional model. It runs faster and can be effectively used in English language translation.

4. CONCLUSION

In this paper, an English language translation model based on data fusion algorithm is established. The conclusions are as follows:

In this study, the attention mechanism is used to extract many detailed features of the key influencing factors. After fusion and normalization, the feature sample set is obtained and input into the machine learning algorithm to build the translation model and output the translation results;

ACKNOWLEDGMENTS

Research on Ideological-political Course in College English Classrooms in the Digital and Intelligent Era——A Case of ‘Xi’s Moments’ of Chinadaily (JGYB2225)

REFERENCES


FL-Lightgbm Prediction Method of Unbalanced Small Sample Anti-breast Cancer Drugs

Chenxiao Zhou\textsuperscript{1a*}, Lianying Zou\textsuperscript{1b*}, Chuang Liu\textsuperscript{1c}, Ziwei Song\textsuperscript{2d}
\textsuperscript{1}School of Electrical and Information, Wuhan Institute of Technology, Wuhan 430205, China
\textsuperscript{2}School of Mechanical and Electrical Engineering, Wuhan Institute of Technology, Wuhan 430205, China
\textsuperscript{a*}2653464846@qq.com, \textsuperscript{b*}zou_ly@126.com, \textsuperscript{c}1464684153@qq.com, \textsuperscript{d}912225007@qq.com

Abstract

The problem of small amount data and sample imbalance exists in the machine learning prediction of the molecular properties of anti breast cancer candidate drugs. Proposing a FL-Lightgbm prediction model based on WGAN-GP data enchanee model in order to solve this problem. Firstly, WGAN-GP model is used for data enhancement to increase the sample size of the training data set. Considering the small difference between positive and negative samples, the enhanced data of positive and negative samples are generated respectively, and then combined them according to the original order to ensure that the generated data and the original data maintain the same distribution; Then the Focal Loss function is introduced into the Lightgbm model to increase learning ability for unbalanced samples, the model constructed is called FL-Lightgbm prediction model. After the training of the enhanced data set, the proposed model shows excellent prediction accuracy for 178 randomly selected validation samples in the experiment, and its highest accuracy, AUC and F1 values reach 0.882, 0.851 and 0.7272 respectively. In these three indexes, the proposed model has better prediction ability than the original Lightgbm model with over sampling algorithms such as BorderlineSMOTE and ADASYN.

Keywords-anti breast cancer drugs; small sample unbalanced dataset; WGAN-GP model; Focal Loss algorithm; Lightgbm model

1. Introduction

Current machine learning models often need a large number of data for training to obtain good prediction performance. However, in the medical field, for the prediction of molecular data of anti breast cancer candidate drugs [1], researchers often have very limited data and small sample size, which is not enough to effectively train the model and obtain high accurate prediction results. In addition, there is a problem of sample imbalance in small sample data sets:1) due to the imbalance of training data, the classification results of machine learning models tend to favor the samples of most classes; 2) Due to the scarcity of minority samples, machine learning models are often unable to effectively train minority samples. Therefore, the solution of the problem of sample size scarcity and sample imbalance is of great significance for the application of machine learning in medical, mechanical and other fields.

To solve the problem of small sample size, the main research method in the field of machine learning is data enhancement. Data enhancement method means to expand the original dataset to obtain more sufficient training samples. Xiaoru song improved the deep convolution generation countermeasure network (DC-GAN) to simulate the data distribution in order to solve the problem of small and unbalanced fire image samples. In his work, he relied on the improved network to generate highly diversified balanced fire image data sets [2], and realized the training of unbalanced small samples by using exponential decay learning rate, L2 regularization method and Adam optimization algorithm. It is proved in the experiment that 99% of the fire images can be recognized. Wang Yu extracted the upper and lower bounds of the airline passenger flow time series to build a ternary interval number data structure, and then converted it into three independent time series. Then he established the airline passenger flow prediction model by using the grey system theory [3], and used the ternary interval number structure to make up for the defect that the accurate real number contains less information.

To solve the problem of sample imbalance, the current research methods are mainly divided into two categories, one is to balance the dataset, the other is to improve the algorithm and adjust the algorithm’s weight value of positive and
negative samples [4]. Typical sample balance algorithms include SMOTE algorithm and improved SMOTE algorithms [5]. Jinglin Huang uses BSMOTE algorithm to oversample minority samples in the unbalanced fault sample set [6], which can reduce the imbalance of various samples. Based on Borderline-SMOTE algorithm, Liu Yang proposed SMB-SMOTE algorithm [7], which can balance samples and distinguish individual importance of boundary samples on the basis of strengthening boundary samples. Oversampling method can effectively solve the problem of absolute sample imbalance, and can also play the effect of data enhancement. However, there are often problems of confusion with the boundaries of minority samples and majority samples, and the distribution of generated samples does not conform to the original samples [5], and the labeling of new samples requires a lot of time and manpower. Therefore, in recent years, researchers often focus on the second methods. A typical algorithm is the focal loss algorithm [8] launched by He in 2017. This algorithm replaces the traditional overall cross entropy loss function with the cross entropy loss containing the overall sample weight, most class sample weight and easily divided sample weight, so that the model can focus more on the classification of difficult samples during training.

The data set selected in this paper is the 1974 compounds obtained from the drugbank drug molecular database of the University of Alberta, Canada for the treatment of breast cancer target a Biological activity and ADMET property data. At present, the research on this data set mainly focuses on the prediction of biological activity. Gu Chengliang used the genetic algorithm combined with BP neural network optimized by genetic algorithm and multi-objective optimization idea to construct the optimization model of anti breast cancer candidate drugs based on genetic algorithm neural network [9]. Xu Pu used BP neural network combined with particle swarm optimization algorithm to build a biological activity prediction model for anti breast cancer candidate drugs [10]. Xumeixian used BP neural network with particle swarm optimization algorithm and SVM model with particle swarm optimization algorithm to predict the bioactivity and ADMET properties of candidate drugs respectively [11]. However, the above research did not consider the problems of too small sample size and sample imbalance in the prediction process of the model.

Inspired by the above research, if the data enhancement method is combined with the Focal Loss algorithm, not only a small number of original samples can be used to obtain sufficient training samples that conform to the data distribution of the original samples, so that the model can be fully trained, but also the Focal Loss algorithm can be used to change the weights of positive and negative samples and difficult samples in the model, which makes the model obtain better training accuracy.

In this paper, the Wasserstein generative adversarial networks with gradient penalty (WGAN-GP) is used to expand the training dataset, and then the Lightgbm model with focal loss algorithm is used to predict the the samples to be predicted.

2. FL-Lightgbm prediction model based on WGAN-GP data enhancement

2.1 Data enhancement based on WGAN-GP network

A main method of data enhancement is to use the general adversarial networks (GAN) to generate analog data. GAN is a neural network framework proposed by Google researcher goodfellow in 2014 [12],this neural network framework mainly composed of generator G and discriminator D. generator G is used to generate simulated samples close to real samples, and discriminator D is used to identify real data and simulated data in samples. In the game between generator G and discriminator D, researchers can get a set of simulated samples that are closest to the real samples. This neural network framework is now widely used in data enhancement of small sample data and has achieved remarkable results.

In recent years, an excellent improvement of GAN is the Wasserstein distance countermeasure generation network with gradient penalty (WGAN-GP) proposed by Dr. Martin Arjovsky of New York University [13]. The objective function of the original GAN model generator adopts the -logD loss function, and the gradient can be transformed into a combination of KL divergence and JS divergence when the discriminator reaches the optimal value. In the generation process, while the KL divergence is maximized, the JS divergence tends to be minimized, resulting in the disappearance of the gradient and the mode collapse [14]. In order to solve above problem, Martin Arjovsky introduced Wasserstein distance into the loss function of Gan's discriminator [15].

Wasserstein distance is defined as:

\[
W(P_r, P_g) = \inf_{\gamma \sim \Pi(P_r, P_g)} E_{(x, y) \sim \gamma} [||x - y||]
\]
In the formula, $\prod(P_r, P_g)$ is the set of all possible joint distributions that combine the probability distribution of the real sample $P_r$ and the probability distribution of the generated sample $P_g$. For each possible joint distribution $\gamma$, a real sample $x$ and a generated sample $y$ can be obtained from sampling. The distance between the two samples is $\|x - y\|$, and the expected value of the sample for the distance under the joint distribution is:

$$E_{(x, y) \sim \gamma}[\|x - y\|]$$

(2)

The lower bound of the expected value can be obtained in all joint distributions, which is called Wasserstein distance. Compared with the -$\log D$ loss function, Wasserstein distance is more sensitive to gradient changes and can provide more meaningful gradients, thus avoiding the problems of gradient disappearance and model collapse.

When Martin Arjovsky introduced Wasserstein distance into the original Gan model, he initially adopted weight clipping to ensure that the value of the loss function is limited to the 1-lipschitz limit, but weight clipping also has the risk of weakening the modeling ability of the model and destabilizing the gradient. Therefore, Martin Arjovsky added the loss value of the discriminator to the penalty gradient [13]. If and only if the gradient norm of a differentiable function does not exceed 1 at any place, the function satisfies the 1-lipschitz condition. Thus, WGAN-GP model is constructed. The loss function of the model is:

$$L = \E_{x \sim P_g} [D(\hat{x})] - \E_{x \sim P_r} [D(x)] + \lambda \E_{x \sim P_g} [\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1]^2$$

(3)

In the formula, the first term is the loss value added with Wasserstein distance, and the second term is the penalty gradient. By adding the guarantee of penalty gradient, WGAN-GP avoids the risk of gradient instability, thus improving the speed of training and the quality of generated samples.

This paper uses WGAN-GP for data enhancement. However, in the data set of this study, there are certain differences between positive and negative samples. As shown in Figure 1, among all properties in the data set, ETA ranked the highest in contribution after ranking by random forest dimensionality reduction and distance correlation coefficient. Alphap (relative molecular size of all non hydrogen vertices of the molecule $\alpha$ On the distribution map of positive and negative samples with the properties of sum of values) and bcutp-1l (lowest polarizability weighted bcut), it can be seen that there are small differences between positive and negative samples.

![Fig.1 the distribution of ETA_AlphaP and BCUTp-1l of positive and negative samples](image)

Although the difference between positive and negative samples is small, if the positive and negative samples are input into WGAN-GP model as a whole for data enhancement, the generated data will be distributed between positive and negative samples. Such generated data will interfere the model in the training process, and the model will be difficult to judge the true label of the sample.

In order to solve this problem, this paper generates enhanced data of positive samples and negative samples respectively. After the data is generated, they are combined according to the original order to ensure that the original data set and the
enhanced data set maintain the same distribution, so as to reduce the interference of the difference between positive and negative samples on the enhanced data generation.

2.2 Lightgbm prediction model based on Focal Loss

Traditional machine learning algorithms often have shortcomings of insufficient robustness and convergence in the face of small sample data [16]. Integrated learning has a better performance in the field of small sample learning because it adopts the method of combining multiple learners according to certain strategies to achieve better learning results. In this paper, the cutting-edge integrated learning model lightgbm (light gradient boosting machine) proposed in recent years is selected for prediction [17].

Lightgbm model was proposed by Ke et al. In 2017, is an integrated learning model based on decision tree. The previous decision tree models (such as xgboost, etc.) need to save the feature values of data and the results of feature sorting, and need to calculate the splitting gain when traversing each partition point, resulting in huge consumption of space and time. To solve this problem, the lightgbm model uses Goss (gradient based one side sampling) algorithm and EFB (exclusive feature bundling) algorithm to optimize the traditional gradient lifting tree. Goss algorithm uses the split variance gain to measure the information gain, and only retains the samples with large gain for calculation, so that the algorithm has a balance between reducing the amount of data and maintaining the accuracy. The variance gain formula of Goss algorithm is:

$$\hat{V}_j = \frac{1}{n} \left( \frac{\left( \sum_{x_i \in A_i} g_i + \frac{1-a}{b} \sum_{x_i \in B_i} g_i \right)^2}{n_j^l(d)} + \frac{\left( \sum_{x_i \in A_i} g_i + \frac{1-a}{b} \sum_{x_i \in B_i} g_i \right)^2}{n_j^r(d)} \right)$$

(4)

In the formula, j is the split feature used, d is the split point of the sample feature, n is the number of samples, A and B are the split large and small gradient samples respectively, l and r are the left subtree and right subtree respectively, and g is the sample gradient.

EFB algorithm bundles the completely mutually exclusive or comparative mutually exclusive features to reduce the number of features and reduce the time complexity. Therefore, the problem to be solved by EFB algorithm is divided into two steps: the first step is how to select the features to be bound, that is, the build bundled problem; The second step is how to bind features together, that is, the merge feature problem. For the build bundled problem, EFB algorithm solves it by constructing a weighted undirected graph. By sorting the degree of conflict between features on the graph, the features to be bound are obtained. For the merge feature problem, the key is that the original feature can be separated from the merged feature. In the previous problem, different bound feature values are divided into the same bundle. EFB algorithm uses the histogram algorithm to save the continuous values as discrete bin, and divides the bundle feature values into different bin.

However, despite the above advantages, Lightgbm model still lacks the ability to predict unbalanced samples. The reason is that the loss function commonly used in lightgbm model is the cross entropy loss function. This loss function will be dominated by most samples when processing unbalanced data [8], resulting in poor prediction ability of the model for a few samples. The formula is:

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise} \end{cases}$$

(5)

In the formula, P is the probability that the predicted sample belongs to 1, and Y is the sample label. If Pt is used to represent the correctly divided sample, the formula can also be abbreviated as:

$$CE(p, y) = CE(p_t) = -\log(p_t)$$

(6)

From formulas (5) and (6), it can be seen that cross entropy can measure the difference between two different probability distributions in the same random variable, and is expressed as the difference between the real probability distribution and the predicted probability distribution in machine learning. However, the weight of cross entropy loss is the same for all kinds of samples, which leads to poor learning effect for a small number of samples. At the same time, because The proportion of majority samples in the dataset is too large, and most of them are easy to classify, the model will not be optimized in the direction envisaged by the researchers. Therefore, this paper considers introducing the Focal Loss loss function, which is more suitable for unbalanced sample classification, into the Lightgbm model.
The Focal Loss function adds the modulating factor [8] on the basis of the cross entropy function to solve the problem that the weight of each sample is consistent. The formula is:

\[ FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \]  

(7)

In the formula, \( p_t \) represents the correctly divided sample. Its modulating factor is:

\[ -(1 - p_t)^\gamma \]  

(8)

In the formula, \( \gamma \) is a modulating factor greater than 0. In formula (8), if \( p_t \) is smaller, it means that the sample is more difficult to be successfully classified; while if \( p_t \) is larger, it means that the sample is less difficult to be successfully classified. As shown in Figure 2, if \( \gamma \) equals to 0, it can be regarded as the cross entropy function, that is, the blue line in the graph; Other lines in the figure represent Focal loss function when \( \gamma \) takes other values. As can be seen from Figure 2, the larger value \( \gamma \) takes, the closer the loss value of the difficult samples is to the loss value of the easy samples. This shows that through the mediation of the modulating factor, the model can effectively reduce the weight of the easy samples and make the model pay more attention to the classification of the difficult samples, thus solving the problem of classification difficulty caused by the small sample size.

Fig.2 Curve of loss value-true samples’ probability while focus parameter \( \gamma \) of Focal Loss function takes different value

2.3 Algorithm flow

First, the WGAN-GP model is used to generate the positive sample data with label 1 and the negative sample data with label 0 respectively, and then combined together according to the original order to obtain the enhanced data set.

After the enhanced data set is obtained, its input loss function is replaced by the Lightgbm model of Focal Loss algorithm (hereinafter referred to as FL-Lightgbm model) for prediction. By adjusting the modulating factor of focal loss algorithm, the model achieves the best prediction effect on the validation dataset.

The algorithm flow chart is shown in Figure 3:
3. Experiment And Result Analysis

3.1 Dataset

The dataset studied in this paper is the molecular descriptor data set of anti breast cancer drugs obtained from the drugbank drug molecular database of the University of Alberta, Canada. The original data set gives 729 molecular descriptor information (i.e. independent variables) of 1974 compounds. Molecular descriptors of compounds are a series of parameters used to describe the structure and property characteristics of compounds, including physical and chemical properties (such as molecular weight, logP, etc.), topological structure characteristics (such as the number of hydrogen bond donors, the number of hydrogen bond receptors, etc.), and so on. However, the sample label of the original data set adopts HOB (human oral bioavailability), and HOB label has only two values of 1 and 0, of which 1 represents high oral bioavailability and 0 represents low oral bioavailability.

Because there are a large number of abnormal data with vacancies, null values and abnormal values in the original dataset, it is necessary to preprocess the data. Deleting outliers out of range according to the pauta criterion, and all 0 columns and null value rows are eliminated. After removing the abnormal data, 88 data with 80% contribution are extracted by using the random forest model, and then the distance correlation coefficient method is used to remove the variables with high correlation between variables, and only the variables with high contribution are retained to achieve the effect of secondary feature dimensionality reduction. After data preprocessing, this dataset was reduced to 13
molecular descriptors containing 1278 compounds. The data with label 1 accounted for 19.40%, and the positive and negative sample ratio was about 1:4.

We randomly extracted 178 data as validation data set to verify the effect of the model. The remaining data sets are randomly divided into training set and test set according to the ratio of 4:1.

3.2 Model evaluation indicators
In order to accurately evaluate the prediction effect of the model, this paper adopts three indicators: accuracy, AUC (area under curve) and F1 score [17].

Accuracy represents "the number of correctly predicted data in the data predicted as positive examples". The closer the value is to 1, the higher the accuracy is. AUC is the area under the ROC curve. The closer the value is to 1, the better the prediction effect of the model is; F1 value is the harmonic average of accuracy rate and recall rate. The closer the value is to 1, the better the prediction effect of the model will be.

3.3 Comparison experiment of sample quality generated by Gan and WGAN-GP
In order to verify the advantages of WGAN-GP in generating samples, we use GAN model and WGAN-GP model to generate positive and negative samples respectively. The quality of the generated samples is evaluated according to the average deviation between the generated samples and the original samples. The smaller the deviation value, the closer the generated sample is to the original data, and the higher the quality of the generated sample [13]. The calculation formula of deviation value is:

\[ p = \frac{V_g - V_r}{V_r} \]  

(9)

In the formula, \( p \) is the deviation value. \( V_g \) is the value of the generated sample, and \( V_r \) is the value of the original sample.

We use the data of 13 molecular descriptors in the preprocessed dataset to generate samples. These molecular descriptors (abbreviated as MD in Figure 4) are ETA_AlphaP、BCUTw-1h、BCUTp-11、nBase、MAXDP2、ATSml、ETA_dEpsi1on_D、VCH-6、LipinskiFailures、ETA_dBetaP、ATSc2、ETA_EtaP_B_RC and maxhbint2 are abbreviated as properties 1 to 13 in the comparison diagram below.

The average deviation value of sample data generated by GAN model and WGAN-GP model on 13 molecular descriptor information is shown in Figure 4:

![Fig.4 the comparison of the deviation value of chemical properties’ generated data](image-url)
As can be seen from Figure 4, the average deviation values of the generated samples of WGAN-GP model in the data of various properties of positive and negative samples are lower than those of the generated samples of GAN model. This shows that the quality of the samples generated by WGAN-GP model is better than that of the original GAN model.

3.4 Generated sample quality evaluation experiment

In order to evaluate the quality of generated samples and verify the impact of the number of generated samples on the model, and compare the performance improvement of FL-Lightgbm model with that of the original Lightgbm model, this paper first uses WGAN-GP model to expand, and generates 220, 440, 660, 880 and 1100 generated sample datasets respectively, and then mixes these datasets with the original training datasets. As shown in Table 1, the original data set represents the original training data set without input of generated samples, and 220, 440, 660 and 1100 of augmentation represent input of 220, 440, 660, 880 and 1100 generated samples into the original training data set. Finally, we input these datasets into the original lightgbm model and FL lightgbm model for comparative experiments.

In order to facilitate the comparison of the model prediction effects under different conditions such as whether the dataset is enhanced or not and whether the Focal Loss algorithm is adopted, the parameter values of Lightgbm are uniformly set as the default parameter values.

The experimental results of the original lightgbm model are shown in Table 1:

Tab.1 the index of original LightGBM model trained by different augment data

<table>
<thead>
<tr>
<th></th>
<th>Original dataset</th>
<th>Expand 220 case</th>
<th>Expand 440 case</th>
<th>Expand 660 case</th>
<th>Expand 1100 case</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>0.8314</td>
<td>0.8595</td>
<td>0.8707</td>
<td>0.8707</td>
<td>0.8707</td>
</tr>
<tr>
<td>AUC</td>
<td>0.7224</td>
<td>0.7723</td>
<td>0.7793</td>
<td>0.7793</td>
<td>0.8008</td>
</tr>
<tr>
<td>F1</td>
<td>0.5588</td>
<td>0.6376</td>
<td>0.6567</td>
<td>0.6567</td>
<td>0.676</td>
</tr>
</tbody>
</table>

Table 1 shows that the prediction performance of the original Lightgbm model increases with the increase of the expanded sample data, which shows that the simulation samples generated by WGAN-GP can effectively help the model improve the prediction performance.

The experimental results of FL-Lightgbm model are shown in Table 2:

Tab.2 the index of FL-LightGBM model trained by different augment data

<table>
<thead>
<tr>
<th></th>
<th>Original dataset</th>
<th>Expand 220 case</th>
<th>Expand 440 case</th>
<th>Expand 660 case</th>
<th>Expand 1100 case</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>0.8483</td>
<td>0.8707</td>
<td>0.8764</td>
<td>0.8764</td>
<td>0.882</td>
</tr>
<tr>
<td>AUC</td>
<td>0.8084</td>
<td>0.8116</td>
<td>0.8151</td>
<td>0.8259</td>
<td>0.851</td>
</tr>
<tr>
<td>F1</td>
<td>0.6582</td>
<td>0.6849</td>
<td>0.6944</td>
<td>0.7027</td>
<td>0.7272</td>
</tr>
</tbody>
</table>

Table 2 shows that the prediction performance of FL-Lightgbm model increases with the increase of expanded sample data. This shows that the simulation samples generated by WGAN-GP can effectively help the model improve the prediction performance on the whole, and the performance improvement reaches the best when the expanded sample size is the same as the original sample size. It is worth mentioning that on the FL-Lightgbm model, although the accuracy of the model with 1100 expanded samples is only improved by 0.56% compared with the model with 660 expanded samples, the improvement in AUC and F1 are 2.51% and 2.45% respectively, which shows that the generation of big data under this model can effectively improve the generalization ability of the model.

Combining table 1 and table 2, comparing the performance of the original Lightgbm model and the FL-Lightgbm model under the expansion of the same augmented data, we can see that the performance indicators of the FL-Lightgbm model on each augmented data are better than the original Lightgbm model in most cases. Moreover, on the original Lightgbm model and FL-Lightgbm model, the prediction performance of the models has been improved with the increase of augmented data.

3.5 Comparison experiment with other models

In order to compare the prediction effect of the model proposed in this paper with other models, the SMOTE model [19] mentioned in literature [20], the BorderlineSMOTE model [20], the ADASYN model [21] mentioned in literature [21] and the SMOTETomek model [22] mentioned in literature [22] are introduced. Since the new data generated by these
four oversampling models can balance the positive and negative samples, Lightgbm model without Focal Loss algorithm is used for prediction. The introduction of these four data balance models will be more conducive to comparing the advantages and disadvantages of various models.

After the four oversampling models are used to expand the dataset, the sample size of the original data set has been expanded by about 600 cases. In order to better compare the performance of this model with the above four oversampling models, we also use WGAN-GP model to expand the original data set by 600 cases.

The comparison results of each model are shown in Table 3:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>0.8314</td>
<td>0.8426</td>
<td>0.8595</td>
<td>0.837</td>
<td>0.8258</td>
<td>0.8764</td>
</tr>
<tr>
<td>AUC</td>
<td>0.7224</td>
<td>0.7726</td>
<td>0.8154</td>
<td>0.7691</td>
<td>0.7513</td>
<td>0.8151</td>
</tr>
<tr>
<td>F1</td>
<td>0.5588</td>
<td>0.6216</td>
<td>0.6753</td>
<td>0.6133</td>
<td>0.5866</td>
<td>0.6944</td>
</tr>
</tbody>
</table>

It can be seen from table 3 that although SMOTE, BorderlineSMOTE and other oversampling models have oversampling and expanded the training data, so that the model can fully learn minority samples and increase the generalization ability of the model, there is also the risk of oversampling. In the verification data set with unbalanced samples, they still lack the ability to identify minority samples. Compared with the performance of the original model, the performance shown in the experiment has only achieved limited improvement. The model with Focus Loss function can effectively distinguish the minority samples from the majority samples by changing the weight values of difficult samples and easy samples. Through the enhancement of the simulation data of WGAN-GP model, the model has achieved good prediction results on the validation data set. In the experiment, the FL-Lightgbm model with 1100 cases enhanced by WGAN-GP improved the accuracy, AUC and F1 by 5.06%, 12.86% and 16.84% respectively compared with the original model, and was significantly better than the other oversampling models selected in the paper under the same amount of expanded data.

4. Conclusion

This paper proposes a method to solve the problem of small sample size and unbalanced samples in the prediction of anti-breast cancer candidate drug by combining WGAN-GP model and Focal Loss algorithm on Lightgbm model. Based on the above discuss and experience, the conclusion shows that:

(1) The samples generated by WGAN-GP model can effectively expand the original dataset, and WGAN-GP model are more suitable than several traditional SMOTE algorithms to help improve the prediction effect of Lightgbm model. Compared with SMOTE model, BorderlineSMOTE model, ADASYN model and SMOTETomek model, the accuracy of WGAN-GP model is improved by 3.52% on average, the AUC of WGAN-GP model is improved by 3.8% on average, and the F1 of WGAN-GP model is improved by 7.02% on average.

(2) Focal Loss algorithm is used to adjust the weight value of difficult samples, so that Lightgbm model pays more attention to difficult samples. In the unbalanced sample prediction experiment, compared with the traditional original Lightgbm model using cross entropy as the loss function, the lightgbm model using Focal Loss as the loss function increases accuracy, AUC and F1 by 1.69%, 8.6% and 9.94% respectively, which means FL-Lightgbm model can achieve better prediction results on unbalanced samples.

(3) The experience conclusion on breast cancer molecular descriptor dataset shows that combining WGAN-GP model and Focal Loss algorithm on Lightgbm model can effectively improve the prediction effect of Lightgbm model on small sample size and unbalanced samples. The Lightgbm model with double data enhancement can improve the accuracy, AUC and F1 by 3.93%, 7.84% and 11.72% compared with the original Lightgbm model. After the Focal Loss algorithm is introduced to the Lightgbm model with double data enhancement, the accuracy, AUC and F1 can be improved by 1.13%, 5.02% and 5.12%.
Through the analysis of the experimental results, it can be seen that the model proposed in this paper can be effectively applied to the prediction of unbalanced small sample data, so this model can greatly improve the problems of small size and unbalanced data in the prediction of anti-breast cancer drug candidate.

References


Cointegration Identification with Metric Learning

Zeyu Xia1,a, Changle Lin2,3b*
1Department of Economics, Shenzhen University, Shenzhen, China
2China Aerospace Academy of Systems Science and Engineering, Beijing, China
3Institute for Interdisciplinary Information Core Technology, Xi’an China
*aixazeyu@outlook.com, bchanglelin@tsinghua.edu.cn

Abstract

Cointegration is an important topic for time series analysis, especially in finance pair trading and hedging area. Cointegration is a kind of structure in which a linear combination of two (or more) time series is stationary. Traditional way to identify cointegration is to use the OLS estimator, firstly run a regression and secondly run a unit root test on residuals. But such method is easy to lead to ambiguous and unstable result. Therefore, we developed a dimensionality reduction model based on automatically calculated common factors, and adopted the Metric Learning method to find a method that can quickly reduce the dimensionality and test the cointegration relationship of stock pairs.

Keywords-Metric Learning, pair trading, cointegration

1. Introduction

Time series analysis try to find stable relationships to help people better understand and predict time series data. While in a real world, various phenomena are usually complexly combined, the data are usually not stationary, making it impossible for us to simply observe and predict. While cointegration is an important concept that by calculating the differences between two or more time series, we may get a more predictable series than any individual time series itself.

In finance, traders want to make huge profit by having a better understand of price time series. Simply if a signal is stationary, people can expect it to have characteristics of mean-reverting, so we could bet it would decrease if already too big and bet it would increase if already too small. Pairs trading is an important type of quantitative trading strategies thinking in this way. It is very important for quantitative traders to identify stable cointegration relations between a pair of, or a tuple of assets. By suitably assigning the holding weights, traders can offset the systematic risk and get a high return-risk ratio. Like the example stated in [1], Pepsi and Coca Cola are the two cartels of cola industry, who has similar prices except for the fluctuation of them. When the price margin of them rises to a set threshold, one can buy the plunged one and sell the other. When both prices return to mean or exceed another threshold, do the inverse trade.

The first step of pairs trading is to find cointegrated pairs before constructing a portfolio. Given that there can be tens of thousands of possible assets groups to be traded, it can cost much to compute each pair, or 3-tuple and even larger tuples of assets. In this study, we focus on pair trading. And to calculate the p-value of two time series, we used the traditional regression and unit root testing method as [2] states. The p-value calculated is used to train our model and evaluate its performance. We propose a method to compute the possible pairs using neural network (metric learning network), which uses much less compute power with satisfying performance.

2. Preliminary Knowledge

2.1 Cointegration

Two time series are said to be cointegrated if some linear combination of them is a stationary series. That is, for time series xt and yt, they are cointegrated if and only if zt = x t − βyt is a stationary process, in which β is not equal to 0. By suitably assigning the parameters, and then calculate unit root for the residuals of the combination, the marginal significance level (p-value), can be computed. The lower the p-value, the more possible the two series are cointegrated. The critical value we select here for p-value being is lower than 1%. So, the problem is to find as many significant pairs as we can and the best pairs.
Given a finite set of time series, how to find a pair with smallest p-value? We can search all O(n^2) pairs, and choose the best one. This is the theoretical optimal pair, and is used as the best possible result we can get in this study. While given the possibly longer and more detailed data in the real world, this approach is not always feasible. After testing, we found Metric Learning can help finding such significant pairs.

2.2 Metric Learning

Metric Learning is a supervised learning method that can automatically construct task-specific distance metrics. Traditional machine learning method requires researchers to specify the distance between the data. Common distances such as Euclidean, City-Block, Cosine distance, while such distance may not be well-suited to the particular data and task of interest. The goal is to learn a distance metric that can better be classified.

3. Algorithm Design

Quantitative trading, like many applications with the input of large vectors with temporal or spatial correlation between different dimensions. (e.g., For image classification, the input image has high spatial correlation between close pixels. For quantitative trading, the input stock value curve has high temporal correlation between nearby time.) It is not efficient to machine learning approaches. In order to deal with this problem, metric learning offers a good way to turn such input into a low dimension high level (highly abstract) feature. This method through neural network is firstly used in 1991 by Friedman [3], that successfully turned 63 × 61 pixels images into 80 features each, and then perform classification on the 80 features.

For this Quantitative trading, we propose a neighborhood method by learning the p-value through deep neural networks. Our method is mainly based on Neighborhood Component Analysis (NCA) [4]. Since p-value is a metric of similarity for two stocks, we set up a dimension reducing neural network to turn a stock value curve into a 12-dimension vector, and hope the distance of the vector estimating p-value as precise as possible.

3.1 Dimension Reducing Neural Network

Since stock value has high noise, it is not suitable to use a huge model. So, we designed a well-generalization light-weight model from the idea of LeNet [5]. LeNet is an image classification network with following pipeline:

\[
\text{Conv} \to \text{Pool} \to \text{Conv} \to \text{Pool} \to \text{Linear} \to \text{Linear} \to \text{Linear}
\]

Such architecture inspires our design. We turn the 2d-convolution to 1d-convolution. After a little tuning, our model is as Figure 1 shows.

![Figure 1 The architecture of the model.](image-url)
• The input is a stock value curve of a year, norm the vector to zero-mean and one-variance. Regard as a 1 channel 243 length tensor.

• Layer C1 is a 1d-convolution layer with kernel size 3 and stride 2, turn the shape of tensor to 12 channel 121 length. (ReLU after this.)

• Layer P1 is a 1d-pooling layer with size 3 and stride 2, turn the shape of tensor to 12 channel 60 length.

• Layer C2 is a 1d-convolution layer with kernel size 3 and stride 2, turn the shape of tensor to 24 channel 29 length. (ReLU after this.)

• Layer P2 is a 1d-pooling layer with size 3 and stride 2, turn the shape of tensor to 24 channel 14 length.

• Layer C3 is a 1d-convolution layer with kernel size 3 and stride 2 with padding 1, turn the shape of tensor to 40 channel 7 length. (ReLU after this.)

• View (reshape) the tensor into one dimension. (i.e., no more channels.) • Layer L1 is a linear layer turn the length to 90. (ReLU after this.)

• Layer L2 is a linear layer turn the length to 12 (the output vector).

3.2 Skewed Square Log Loss
Due to the real distribution of log p-value over all stock’s pairs (Figure 2).
We regard the significant pairs to be true pairs, then it is suitable to use common NCA method. The usual loss function of NCA is,

\[ L_{NCA}(x, y, Z) = \log \left( \frac{\exp(-d(x, y))}{\sum_{z \in Z} \exp(-d(x, y))} \right) \]  \hspace{1cm} (1)

But this loss function is not very suitable in this task for two reasons. One, there are lots of wrong samples and wrong samples also have information and should also be considered. Two, we also want the extreme small p-value pair to be extremely close. (e.g., If p-value is 10^-6, we want their 12d-vector almost the same.) Hence, we propose the SSL (skewed square log loss). (In which, M is the model.) The idea of SSL is intuitive. Note that the factor 1/100 if label(x, y)=0, this is skewed because false samples are much more than true samples. We want the model to make more effort on fitting the log level of p-values.

\[ \text{pred}(x, y) = I(||M(x) - M(y)||_2 < 1) \]  \hspace{1cm} (2)

\[ \text{label}(x, y) = I(p\text{value}(x, y) < 1\%) \]  \hspace{1cm} (3)

\[ L_{SSL}(x, y) = \begin{cases} 
\frac{1}{30} \left( \log \frac{||M(x) - M(y)||_2}{1000 \ p\text{value}(x, y)} \right)^2 & \text{if \ pred}(x, y) > \text{label}(x, y) \\
\frac{1}{30} \left( \log \frac{||M(x) - M(y)||_2}{1000 \ p\text{value}(x, y)} \right)^2 & \text{if \ pred}(x, y) < \text{label}(x, y) \\
0 & \text{if \ pred}(x, y) = \text{label}(x, y)
\end{cases} \]  \hspace{1cm} (4)
3.3 Connection to Multi-Factor Model

Consider a multi factor model,

$$R_t = \alpha_t + \sum_k \beta_{tk} F_k + \epsilon_t, \forall t$$  \hspace{1cm} (5)

A perfect zero-beta portfolio w has a goal being,

$$\sum_i w_i \beta_{ik} = 0, \forall k$$  \hspace{1cm} (6)

Compared with dimension reduction model, that the output is fitting p-value. Then a pair trading portfolio w will have a goal that

$$\sum_x w(x) M(x) = 0$$  \hspace{1cm} (7)

If we believe that there is an underlying multi-factor model characterizing the market, then our neural network will learn the coefficient of the factors (i.e., the $\beta$ terms) without explicit representation of the factors F. The factors are given in the final 12-dimensional output of the network, each dimension of which is a $\beta$ value of a stock.

Learning of the factors (or in fact coefficients of implicit factors) is driven by the attempt to minimize p-value evaluation loss, which can also be a starting point for machine learning based multi-factor model.

4. Experiment Results

4.1 Data Preprocessing

We use a dataset of all A-shares stocks on all trading days from 2018 to 2022 with more than 700 trading days. By omitting stocks with too many missing data, we selected 2012 stocks to carry out metric learning. All data are downloaded from open data source Tushare.

4.2 Z-score Policy

We use the approach of z-score mentioned in [6]. First calculate the ratio of prices of a pair to represent the relative trend of the two stocks. On the ratio, there are two sliding averages, one spreads over 5 days, and the other spreads over 30 days. The former one is used to indicate the fluctuation of the stock, and the second one is a measure of “average” ratio. The standard deviation of the ratio is also calculated, which is based on the larger window. From the ratio, define z-score to be the difference between the average of smaller window and the other average, divided by the standard deviation. This is the parameter used to capture the divergence of the stocks. When on any day, the z-score reaches a pre-set threshold, for
example 0.75, it will emit a buying or selling signal. Then some stocks will be traded, one bought and the other is sold in this situation.

4.3 Model Accuracy

With model architecture and loss function set previously, we keep resampling data to train the model. In detail, to improve efficiency, we don’t actually perform 1/30 when label(x, y) = 0. Instead, we use importance sampling. That is, the positive pairs with sample importance 30 while the negative pairs with sample importance 1. Table 1 and Table 2 shows the result of our model. Note that we have data of 2012 stocks, hence TP + FP + FN + TN = 2012×2011/2.

Recall_train = 16027/(16027+8029) = 66.62%
Recall_test = 10952/(10952+12240) = 47.22%
Precision_train = 16027/(16027+132714) = 10.78%
Precision_test = 10952/(10952+114435) = 8.74%

Although here Precision is quite small, it is mainly because negative samples are much more than positive samples (about 100 times.). Which means, it is reasonable with some “lucky” negative samples to have wrong classification. The probability given a negative sample to reject it is still high.

Table. 1 RESULT ON THE DATA IN 2018 (TRAIN SET)

<table>
<thead>
<tr>
<th>Train Set</th>
<th>Actual P</th>
<th>Actual N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict P</td>
<td>TP = 16027</td>
<td>FN = 132714</td>
</tr>
<tr>
<td>Predict N</td>
<td>FP = 8029</td>
<td>TN = 1866296</td>
</tr>
</tbody>
</table>

Table. 2 RESULT ON THE DATA IN 2019 (TEST SET)

<table>
<thead>
<tr>
<th>Train Set</th>
<th>Actual P</th>
<th>Actual N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict P</td>
<td>TP = 16027</td>
<td>FN = 132714</td>
</tr>
<tr>
<td>Predict N</td>
<td>FP = 8029</td>
<td>TN = 1866296</td>
</tr>
</tbody>
</table>

4.4 Trading Based on Metric Learning Classification

We use the p-value in the data of year 2019, pick the pairs if their p-value less than $10^{-6}$ (which identified 980 pairs in total). For each pair, we run z-score policy trading in the data of year 2020 with some hyper parameters. Buy if $Z$-score < −0.5, Sell if $Z$-score > 0.5, Clear if $|Z$-score| < 0.2.

Pick an example to show its efficiency. Figure 3 shows its $Z$-score curve and signals on it. We can find out it oscillates quite often and such hyper parameters will catch the chances. Figure 4 shows how it works. When back-testing, we set no service.
For each pair trading, we mainly focus on two benchmarks, Return and Max-Hold. Return parameter is the money we gained at last if we have zero initially, and Max-Hold parameter is the maximum value of stock we hold over the whole trading time. Figure 5 shows the histogram of Return of all these 980 pairs, and Figure 6 shows the histogram of Max-Hold.

Moreover, we can get

\[
\begin{align*}
\text{mean}(\text{Return}|1e^{-6}) &= 29501.38, \\
\text{std}(\text{Return}|1e^{-6}) &= 101859.79, \\
\text{mean}(\text{MaxHold}|1e^{-6}) &= 484737.89, \\
\text{std}(\text{MaxHold}|1e^{-6}) &= 193320.19.
\end{align*}
\]
These results are over the pairs with p-value less than $1 \times 10^{-6}$. While if we set the p-value critical value to be $1 \times 10^{-3}$, the results are still quite good.

5. Conclusions

In order to conduct pair trading based on cointegration relationship, the traditional method needs to find out pairs that may have cointegration relationship from a large number of stocks. But this method will lead to huge amount of computation. However, based on the multi-factor model, the co-integration relationship between assets can be attributed to the corresponding factors. Therefore, the method based on Metric Learning can quickly achieve dimensionality reduction, find the distance based on the common factor measurement, and then initially detect such pairs. We also tested the application of related algorithms in pair trading.

References


Tracking Pedestrians from a Moving Camera Based on Kalman Filter

Yingxu Wang1, *

1School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan, 430070, China

*Corresponding author’s e-mail: 299513@whut.edu.cn

Abstract

The target tracking and object tracking are defined in this paper and the difference between multi-target tracking and multi-object tracking is also be illustrated. The Bayes filter, Kalman filter, EKF, JPDA and Hungarian Algorithm are introduced with formulars and an example of moving camera to track the pedestrians used by Kalman filter are shown. In this example, the method which is based on Kalman filter that track pedestrians from a moving car which is installed with camera in the field of the multi-object tracking is analysed with steps. The algorithm initializes boundary boxes to track the pedestrians and predict the pedestrians based on the previous position. Then, update the tracks and delete the useless tracks. The final step is creating the tracks. After displaying the result, the algorithm based on Kalman filter can successfully track the pedestrians with boundary boxes. However, when the camera is moving fast, some of the pedestrians cannot be recognised.

Keywords: Tracking Pedestrians, Kalman Filter

1. Introduction

Target tracking is an important part of computer vision. The key to successful target tracking is extracting the features of targets’ observations [1]. After holding the feature, another significant step is to get the features in every frame of the observation. Then, in tracking the object, the movement of the target is random so there are some algorithms to tracing the target. In random movement, Bayes filter can be used to handle the signal. Under the prior of Bayes filter, Kalman filter [2] has six hypothesises. Kalman filter is usually to handle problem in the circumstance of Gaussian mode. First, the Kalman filter assumes two random variables which agrees with the Gaussian distribution. Every variable has an average, which is the centre of the random distribution and a variance to evaluate the uncertainty. Owing to no connection between the variables, it is impossible to get other variable from one. However, sometimes these two variables have association, which maybe important on target tracking. The purpose of Kalman filter is obtaining as much as information from the random variables. The covariance matrix can get the relation of the two variables. The covariance matrix and the external uncertain factors can be connected to get the predictive formular. The new predictive formular is based on the original prediction and the external influence and the new uncertainty is founded by original uncertainty and the external uncertainty. The original time of the position is used to predict the next time position. The system collects the whole data and get a new matrix, which also obey Gaussian distribution. The new matrix can be calculated to get the covariance. The original matrix plus the final matrix and get the overlap which is the most possible place for the object to emerge. Particle filter based on Monte Carlo method is used in nonlinear Bayesian filtering situation [3]. One of the particle filter algorithms is sequential importance sampling (SIS) [4]. First, the system should get the sample with the Monte Carlo method. Second, the sample can be calculated to get the importance of particle weight. Finally, the system can repeat the last step and get the sequential weight of the sample.

Practically, it is widely to use multi-target tracking (MTT). MTT is made up by target tracking and data association which is to select a hypothesis that can divided the detection into disjoint subsets [5]. When the targets are very close in space, the joint probabilistic data association (JPDA) [6] performs great. JPDA judges all points of observation and analyses the connection between the points and the traces of target [7] and probability of the reflection from the target is different. JPDA needs no prior information so it can have a great performance in clutter environment. However, when the data is too large, the amount of the calculation is too much. In MTT, a common algorithm to deal with the association of data is Hungarian Algorithm [8]. Hungarian Algorithm divides the whole group into two parts as Bipartite graph and uses a matrix to depict the relationship. Hungarian Algorithm is used to settle the maximum matching problem of Bipartite graph. Firstly, in the matrix, every row and column will be misused by the minimum value of the row or the column. After that, some lines will connect the value of zero and if the number of lines is less than the order of the matrix, the process will be circulated again until the number of the lines is same as the order of the matrix. Finally, according to the ultimate matrix, every individual should have one relationship with other individual in different group and the correspondence is the result.
Multi-object tracking (MOT) is a combination of MTT and recognition to the object. Thus, in MOT, a significant step is matching the object to every track and a method to solve it is calculating the areas’ boundary overlap ratio. The union rate is that the intersection of the areas divided by the union of the areas. The minimum is that the intersection divided by the minimum of the areas and get the overlap ratio.

2. Problem Formulation

The problem is about recognizing the pedestrian from the picture and tracking. In this part, Bayes filter and Kalman filter are introduced and Kalman filter is used in the method of tracking.

2.1 The model of tracking and recognizing

This paper is based on MOT to solve the problem that moving robots or vehicles can recognize the objects including pedestrians and other objects. This algorithm can be divided into two parts. The first part is MTT to track targets. In this part, the Kalman filter achieves prediction and update. The second part is recognition which is the difference between MTT and MOT. In recognition step, the overlap ratio can be calculated from the cost of targets and Munkres’ version of the Hungarian Algorithm matches the cost of targets and the objects.

2.2 Bayes Filter

Filter is a method to estimate the hidden value based on the observed value with error in time-varying system and Bayes Filter is a kind of filter that adopts Bayes formular to get from $x_k, y_k, p(x_k|x_{k-1})$. and $p(y_k|x_k)$.

The assumption is that the model fits Markov property which formulars are

$$p(x_k|x_{1:k-1}, y_{1:k-1}) = p(x_k|x_{k-1}) \quad (1)$$
$$p(x_{k-1}|x_{k:T}, y_{k:T}) = p(x_{k-1}|x_k) \quad (2)$$

and condition independence which formular is

$$p(y_k|x_{1:k}, y_{1:k-1}) = p(y_k|x_k). \quad (3)$$

The first step is initializing the probability of hidden value $p(x_0)$. Second, based on the previous observed values $y_{1:k-1}$, the hidden condition value $x_k$ can be predicted by total probability formular to get

$$p(x_k|y_{1:k-1}) = \frac{1}{z_k} p(y_k|x_{k-1}) p(x_k|y_{1:k-1}) dx_{k-1}. \quad (4)$$

Finally, the previous prediction and the prediction of $k$ can be combined with total probability formular so that the final formulars of the update step are

$$p(x_k|y_{1:k}) = \frac{1}{z_k} p(y_k|x_k)p(x_k|y_{1:k-1}) \quad (5)$$

and

$$Z_k = \int p(y_k|x_k)p(x_k|y_{1:k-1}) dx_k. \quad (6)$$

2.3 Kalman Filter

There are two steps in Kalman filter. The first one is prediction and the second step is update. Prediction is to predict the condition this time with the foundation of the last time’s condition. Update is combining the observed value with this time prediction value to get the best condition. In prediction step, the formulars are

$$x_k = Ax_{k-1} + Bu_{k-1} \quad (7)$$

and

$$P_k = AP_{k-1}A^T + Q. \quad (8)$$

In update step, the formulars are

$$K_k = P_kH^T(HP_kH^T + R)^{-1}, \quad (9)$$
$$x_k = x_k + K_k(z_k - Hx_k) \quad (10)$$
and

\[ P_k = (I - K_k H) P_k \tag{11} \]

In these formulas, the \( x_k \) is the condition in the time of \( k \).

Although Kalman filter is suitable in linear system, it cannot be utilized in non-linear system. Thus, an extended Kalman filter (EKF) is used in non-linear system. In EKF, first order Taylor series which formular is

\[ T(x) = f(a) + f'(a)(x - a) \tag{12} \]

can construct a linear formular approximately. After the approximation, the formular can be combined with Kalman Algorithm to get an approximate Gaussian distribution. The approximation has error, but considering the speed of calculation, the approximation is necessary. The prediction formulars of EKF are

\[ x_t = g(u_t, x_{t-1}) \tag{13} \]

and

\[ z_t = h(x_t) + \delta_t. \tag{14} \]

The formular \( g \) and formular \( h \) are non-linear formulars and the former is the location posteriori of movement model and the latter is the posteriori of measuring data. After approximate by Taylor series, the formulars turn to

\[ g(u_t, x_{t-1}) \approx g(u_t, \mu_{t-1}) + g'(u_t, \mu_{t-1})(x_{t-1} - \mu_{t-1}) = g(u_t, \mu_{t-1}) + G_t(x_{t-1} - \mu_{t-1}) \tag{15} \]

and

\[ G_t := g(u_t, \mu_{t-1}). \tag{16} \]

The posteriori possibility of \( G \) at the time of \( t \) is

\[ p(x_t | u_t, x_{t-1}) \approx \det(2\pi R_t)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ x_t - g(u_t, \mu_{t-1}) - G_t(x_{t-1} - \mu_{t-1}) \right]^T R_t^{-1} \left[ x_t - g(u_t, \mu_{t-1}) - G_t(x_{t-1} - \mu_{t-1}) \right] \right\} \tag{17} \]

Similarly, the formulars of probability of posteriori of measuring data are

\[ h(x_t) \approx h(\mu_t) + h'(\mu_t)(x_t - \mu_t) = h(\mu_t) + H_t(x_t - \mu_t) \tag{19} \]

and

\[ H_t := h'(\mu_t). \tag{20} \]

After getting the possibility, the measuring probability can be calculated and the final formular is

\[ p(z_t | x_t) \approx \det(2\pi Q_t)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ z_t - h(\mu_t) - H_t(x_t - \mu_t) \right]^T Q_t^{-1} \left[ z_t - h(\mu_t) - H_t(x_t - \mu_t) \right] \right\} \tag{21} \]

Combining with Kalman filter, the final state prediction, measurement update, calculation of Karman gain and calculation of posteriori state and covariance are clear. The state prediction has two formulars containing

\[ \dot{x} = f(x) \tag{22} \]

and
\[
P\hat{x} = FP \hat{P}_T + Q. \tag{23}
\]
The measurement update formulas are
\[
y = z - h(\hat{x}) \tag{24}
\]
and
\[
S = H \hat{P} H_T + R. \tag{25}
\]
The calculation Kalman gain formula is
\[
K = \hat{P} H_T S^{-1}. \tag{26}
\]
The calculation of posteriori state and covariance formulars are
\[
x = \hat{x} + Ky \tag{27}
\]
and
\[
P = (I - KH)\hat{P}. \tag{28}
\]

3. The method of recognizing and tracking

3.1 Tracking

3.1.1 Auxiliary Input and Global Parameters of the Tracking System

The tracking system requires a data file which includes the pixel position and the size of the bounding box of the pedestrian's location. The prior knowledge lays in the vector. In this vector, the \( n \)-th entry means the estimated the height of the pedestrians. The reference \( n \) refers to the estimated \( y \)-coordinate of the pedestrians' feet.

In order to obtain the vector, a collection of the training images is gotten from similar view and scene in test environment. The training images include pedestrians which have different distance to the camera and the app-Image Labeler can note the bounding boxes of the pedestrians. The height of the bound boxes and the location of the pedestrians are used to generate the scale data file through regression.

3.1.2 Create System Objects for the Tracking System Initialization

Firstly, the creating system object is used to create the system to read and display the video frames and load the scale data file.

The vector stores in the scale data file, encoding the prior knowledge of the target and the scene. After obtaining the regressor training from the samples, the height of every \( y \)-position in the image can be calculated. The values are stored in the vector.

3.1.3 Initialize Tracks

Creating an array of tracks by introducing some arrays containing id, color, boundary boxes, scores, age, pedestrians' position, total visible count and confidence, where each tracks express the construction of moving object. The purpose of the construction is to maintain the condition of tracking target. The condition includes the information of detecting track assignment, track termination and display.

3.1.4 Predict New Location of Existing Tracks

In this step, the Kalman filter can predict every centroid of the track in current frame. In the previous frame, the height and the length of the bounding boxes is the prediction to the current frame.

3.1.5 Detect Lost Tracks

Delete a sequence of tracks which are invisible in frames because the invisible tracks are constructed by mistake. In this step, the no-use tracks will be deleted.

3.1.6 Create New Tracks

The new tracks are created in undistributed detections. Suppose any undistributed detections are started with new tracks,
practically, other cues can be used to eliminate noisy detections.

### 3.2 Detect People

In this step, the pretrained upright people detector using aggregate channel features will be returned. The detector can recognize the specific objects in images, based on training [9] and using aggregate channel features object detector to detect objects [10] [11].

The aggregate channel can make a difference in returning the centroids, the bounding boxes and the classification scores of the people who are detected.

The database is split into training and testing, which makes it easy to compare the methodology. The database can be divided into two parts. The former is people with different gesture and the latter is pedestrian [9]. In recognizing pedestrians, the images will be resampled to reserve important features. The gradient histogram method can handle the data. The $I(x, y)$ is discrete signal and $\frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial x}$ are the discrete derivatives of $I$. The gradient magnitude and orientation are defined as

$$M(i,j)^2 = \frac{\partial I}{\partial x}(i,j)^2 + \frac{\partial I}{\partial y}(i,j)^2$$

and

$$O(i,j) = \arctan\frac{\frac{\partial I}{\partial y}(i,j)}{\frac{\partial I}{\partial x}(i,j)}$$

and the pixel of the images has a vote and make a sequence by the weight. The orientation is quantized into $Q$ bins so

$$O(i,j) \in \{ 1, Q \}$$

and the number $q$ bin’s histogram is defined by

$$h_q = \sum_{i,j} M(i,j) \textbf{1} [O(i,j) = q]$$

and 1 is indicator function.

### 3.3 Data Association

#### 3.3.1 Assign Detections to Tracks

The minimizing cost assigns the object in current frame to the existing tracks. The cost can be calculated to get the overlap rate of the predicted bounding boxes and the detected bounding boxes. The overlap ratio can be calculated by two formulas which are

$$\text{Union} = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)}$$

and

$$\text{Min} = \frac{\text{area}(A \cap B)}{\min[\text{area}(A),\text{area}(B)]]}$$

In this paper, the pedestrians are moving gradually with low velocity.

#### 3.3.2 Update Assigned Tracks

Kalman filter is to correct the location estimate. Next, it gets the nearest bounding boxes which less than 5 bounding boxes and calculates the average to store the new bounding boxes and adds one to the age and the total visible number. Finally, the formula adjusts the confidence score based on previous detected score.

#### 3.3.3 Update Unassigned Tracks

Sign every undistributed track to invisible and adding one to the age and attaches the predicted bounding boxes to the track. Then, the new boundary boxes will be stored by calculating the average of the size which up to 4 boxes and the age of the track and the total visible count are added by one. Finally, based on the previous detection scores, the confidence of the score will be changed. Note that the Probability Hypothesis Density filter can also be applied if unique label is not required for each target [12][13][14].
4. Result

The result is conducted by MATLAB demo “Tracking Pedestrians from a Moving Car” in Computer Vision Toolbox and some constructed functions are used.

4.1 Pedestrians Tracking

In the figure 1, the pedestrian who are riding a bike can be recognized and some information and a boundary frame are also presented. The moving cars are not be recognized as pedestrians and no misrecognition are emerged.

The people riding bike is detected, but the pedestrian is not detected. In the detection method, the people who is comparably vague is not considered as a meaningful message to the moving camera. Thus, the pedestrian may be deleted.

![Figure 1](image1.png)

Figure 1. Figure shows a recognition to the pedestrian even he is riding a bike and some message and a boundary box is shown.

4.2 Pedestrians’ recognition

In pedestrian’s recognition, based on the result of figure 2 (a), the algorithm can only recognize the people who are close to the moving camera. Plus, when the camera moving fast, the success possibility is decreased from the figure 2 (b).

In the detect people step of the method, the people who are not moving or the reflection on the glasses are not be regarded as useful data. Thus, the undetected people are moving slow comparing with the moving camera.

![Figure 2](image2.png)

(a)

(b)

Figure 2. Figure shows two pictures of failed recognition.

5. Conclusion

In this paper, the Bayes filter and Kalman filter are introduced and the JPDA, EKF and Particle filter are also be mentioned. Mainly, this paper focuses on the application of the Kalman filter in MOT and the result is that the moving camera installed
in a car to track the pedestrians. The result shows that the Kalman filter can track the pedestrians successfully but when the camera moving in a high velocity, the accuracy of the tracking will be down.

Reference

[12] Chen J and Dames P 2020 Distributed multi-target search and tracking using a coordinated team of ground and aerial robots in Robotics science and systems
A Prediction of Human Body Key-points Based on Kalman Filter in Transient Visual Masking

Rui Li\textsuperscript{1}, Xiaoxiao Zhu\textsuperscript{2}, Zhaohui Liang\textsuperscript{2}, Yanhong Ma\textsuperscript{3}, Huipeng Shi\textsuperscript{3} and Qixin Cao\textsuperscript{2}

\textsuperscript{1}SJTU Paris Elite Institute of Technology, Shanghai Jiaotong University, Dongchuan Road, No.800, Shanghai, CN
\textsuperscript{2}School of Mechanical Engineering, Shanghai Jiaotong University, Dongchuan Road, No.800, Shanghai, CN
\textsuperscript{3}Shanghai Sixth People Hospital, Yishan Road, No.600, Shanghai, CN
qxcao@sjtu.edu.cn

Abstract

While extracting dynamic 3D human key-points data for human pose estimation using only one RGBD camera, some key-points may be covered by other body parts during users’ movement. In this case, some key-points’ depth data can’t successfully be fused with aligned RGB data and that will influence final estimation of movement in this period. In this article, we present a method to predict users’ movement when camera can’t read exact 3D information of human key-points. This method proposes a Kalman filter model with constraints for predicting movement while there is a short sensor failure. The experiments show that these constraints can enhance robustness of system and help the filter to have a smooth performance. During sensor failure, 60% location data is well predicted.

Keywords: Pose estimation, Kalman filter, Visual masking

1. Introduction

Human Pose Estimation is a popular subtopic of computer vision in recent years. The application of this technology can be expanded into sports analysis [1], action recognition [2], rehabilitation [3], robotics [4]. The basic idea of human pose estimation is to firstly find key-points’ location and then group them in a certain way, bottom-up or top-down. “Bottom-up” predicts firstly all key-points in the screen then assembles them into different skeletons, like Openpose [5]. “Top-down” means that the system outputs firstly proposal detectors and pose estimators then connects key-points for each people [6].

However, when some key-points are covered, it is difficult for the system to collect exact 3D data for pose estimation due to the lack of exact depth data. An example is shown in Figure 1 that depth data of two ankles are both default value 0. Meanwhile, neural network may connect key-points incorrectly in visual masking, especially when there are intersections of membrum in 2D plane shown in Figure 2.

To combine observation and prediction, one of commonly used method is Kalman filter (KF). It is wildly used in computer vision field. Tao Hu, Chunxia Xiao use Kalman filter to track peoples’ movement with visual information [8]. In this paper, a portable system is proposed to collect 3D key-points locations of users. Then model error will be analysed. To continuously estimate user’s pose when there is a visual masking, a constrained KF model is proposed. Finally, the performance of filter with/without constraints will be shown and compared.
2. Methods

In this paper, data collection system consists of Realsense D435i and an Intel Core i5-8250U CPU.

2.1 3D pose data collection

In this part, a light weight network based on Openpose is used to collect 2D location information of human body key-points. Daniil changes network structure then reduces complexity [7], so that this network can be applied in a CPU only computer. With this network, the system uses video data as input, and outputs 18 human key-points, including two ears, two eyes, nose, chest, two shoulders, two elbows, two wrists, two hips, two knees and two ankles. Since this system has no GPU, OpenVino toolkit is used for accelerating programs. Figure 4 presents the global process of data collection. The network will output 2D location of key-points, then depth data is merged.

Figure 3. 18 human body key-points model

Figure 4. Global process of data collection

2.2 Design for Kalman Filter

While collecting 3D key-points locations, the sampling frequency is about 16 Hz. A hypothesis is proposed that the sampling frequency is big enough so that during every sample the movement is linear and rotation can be ignored. Therefore, the movement can be described by matrix:

\[
\begin{bmatrix}
    x_k \\
    y_k \\
    z_k \\
    VX_k \\
    VY_k \\
    VZ_k
\end{bmatrix} = \begin{bmatrix}
    1 & 0 & 0 & dT_{k-1} & 0 & 0 \\
    0 & 1 & 0 & 0 & dT_{k-1} & 0 \\
    0 & 0 & 1 & 0 & 0 & dT_{k-1} \\
    0 & 0 & 0 & 1 & 0 & 0 \\
    0 & 0 & 0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
    x_{k-1} \\
    y_{k-1} \\
    z_{k-1} \\
    VX_{k-1} \\
    VY_{k-1} \\
    VZ_{k-1}
\end{bmatrix}
\]  

(1)

Here, \(x_k, y_k, z_k\) are coordinates of human joints in geodetic coordinate at time k, and \(VX_k, VY_k, VZ_k\) represent speed of joints in three directions at time k. \(dT_{k-1}\) shows the time different value between two visible frames. The main process of Kalman filter is shown in (2)(3)(4)(5)(6):
\[ X_k = F_kX'_{k-1} + B_kU_k \quad (2) \]
\[ P_k = F_kP'_{k-1}F_k^T + Q_k \quad (3) \]
\[ X'_k = X_k + K'(Z_k - H_kX_k) \quad (4) \]
\[ P'_k = P_k - K'H_kP_k \quad (5) \]
\[ K' = P_kH_k^T(H_kP_kH_k^T + R_k)^{-1} \quad (6) \]

\( X_k \) is the state vector of measured system at time \( k \), which includes three coordinates and speed in three directions represented in (1). \( F_k \) is status transfer matrix. \( B_kU_k \) can be considered as an extra parameter which is controllable and originates from outside of system. In this model, there is no external stimulation, so \( B_kU_k \) is zero. \( Q_k \) is process excitation noise covariance which is assigned by experience. \( P_k \) is a priori estimated covariance. And \( P'_{k-1} \) is posterior estimated covariance. \( R_k \) is the covariance of estimation noise. \( Z_k \) is observation value and \( H_k \) represents transformation matrix, connecting status and observation. \( K' \) is coefficient of Kalman, renewing iteratively.

While testing different movements, some coordinates are less relative. For example, \( x \) plays an insignificant role during squats while \( y \) is important. In this case, a big filter can be divided into three small filter:

\[
\begin{bmatrix}
  y_k \\
  VY_k
\end{bmatrix} = \begin{bmatrix}
  1 \\
  0
\end{bmatrix} \begin{bmatrix}
  dT_{k-1} \\
  1
\end{bmatrix} \begin{bmatrix}
  y_{k-1} \\
  VY_{k-1}
\end{bmatrix} \quad (7)
\]

Formula (7) shows a filter involving only information on \( y \) axis of camera coordinates. And it is similar on \( x \) and \( z \) axis. In order to enhance the performance of filter, certain constraints are put into the algorithm. Figure 5 shows the process of Kalman filter with constraints and formula (8)-(9)-(10) represent constraints.

\[
d(P_i, P_j) < \varepsilon_{ij}(Z_i, Z_j) + \gamma \quad (8)
\]
\[
d(R_i, R_j) < \varepsilon_{ij}(Z_i, Z_j) + \gamma \quad (9)
\]
\[
|d(P_i, P_j) - d(R_i, R_j)| < \gamma \quad (10)
\]

Here, \( P_i \) and \( P_j \) are two adjacent key-points predicted by filter. \( R_i \) and \( R_j \) are two adjacent key-points captured by camera. Figure 3 is the reference to check whether two key-points are adjacent. \( d(,\) \) is Euclidean distance in camera coordinates. \( \varepsilon_{ij}(z) \) is corresponding body’s length, it can be calculated by depth data and \( d(R_i, R_j) \) of first frame, the depth data of \( R_i \) and \( R_j \) of first frame. And \( \gamma \) is a coefficient relative to modelling error which will be measured in 2.3. It should be 4 times bigger than modelling error.

After key-points’ location is predicted, each point will be checked if they satisfy the constraints. If no, filter will trace back to nearest frame. Kalman filter can only predict location when there is a short sensor failure. A threshold is set to judge whether the filter will be restarted. Here, if key-point is continuously lost more than 5 frames, filter will record faults and restart.

**Figure 5. Process of Kalman filter with constraints**

### 2.3 Modelling error detection

Since the recognition of key-points is related to heatmaps and heatmaps are generated based on Gaussian model [7], there is a fluctuation of predicted key-points 3D coordinates. In order to measure this part of error, participants are...
required to stand face to camera and keep still, then the system records 3D coordinates of key-points. Figure 6 shows the fluctuation of 3D coordinates of participant’s right shoulder joint. The horizontal axis is time (frames). It is chosen as an example because shoulder joint is one of the most stable joints of human body.

This fluctuation, noted as $N$, can be divided into two parts. The first part originates from slight sway of body, noted as $N_{\text{sway}}$. The second part may originate from model errors which relate to neuro-network, noted as $N_{\text{model}}$. $N_{\text{model}}$ on X axis should approach to $N_{\text{model}}$ on Y because they should follow Gaussian distribution. Here $N_{\text{sway}}$ should not be considered as noise because it can be correctly observed, and it is a part of the real movement. The real noise should be $N_{\text{model}}$, like a kind of measurement error.

![Fluctuation of 3D coordinates](image)

Figure 6. The fluctuation of 3D coordinates of participant’s right shoulder joint

A hypothesis is proposed that $N$ is close to $N_{\text{model}}$ on Y axis. To examine this hypothesis, another test for the measurement of $N_{\text{model}}$ is done using 10 video streams of 100 frames and each video is consisted of only one picture. For each video, the amplitudes of 3D coordinates fluctuation of every key-point are recorded and then RMSE (Root Mean Square Error) is calculated. The average values of RMSE of 10 videos are shown in Table 1.

<table>
<thead>
<tr>
<th>Average of RMSE</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_x$ 0.06</td>
<td>Estimation error on x axis</td>
</tr>
<tr>
<td>$R_y$ 0.07</td>
<td>Estimation error on y axis</td>
</tr>
<tr>
<td>$R_z$ 0.12</td>
<td>Estimation error on z axis</td>
</tr>
</tbody>
</table>

$R_x$, $R_y$, $R_z$ are estimation error on x, y and z axis of camera coordinate. They are also components of $R_e$ which represents covariance matrix of estimation error in Kalman Filter. Other 3 components are $R_{vx}$, $R_{vy}$, $R_{vz}$. They are estimation error of $VX$, $VY$, $VZ$. The value of $R_{vx}$ should be $\frac{4R_x}{16}$, because $R_{vx}$ is calculated by the difference of coordinates between two frames and the sampling rate is 16Hz. So are $R_{vy}$ and $R_{vz}$.

### 3. Experiments and Results

During experiments, participants are told to directly face camera, do squats periodically.

#### 3.1 Result of experiments

During squats, hip joint is the most important key-point because it is probably covered by other parts of body. Meanwhile, the value of x is less relative to the movement. Figure 7 shows hip joint’s y and z coordinates before and after Kalman filter (without constraints). Figure 8 shows the performance of filter with constraints and filter without constraints while participants doing squats periodically. The judgement of two filter will be discussed in 3.2.
3.2 Discussion

From raw data shown in figure 7, hip joint is covered irregularly by other parts of body. Usually, they are thighs and arms. When they are covered, system can successfully predict 2D coordinates but the depth data is 0. Kalman filter without constraints can deal with this problem. The wave form of Y is well retained by filter, but with some delay. One probable reason is that the movement is not strictly linear and sampling rate is not big enough.

It is shown in Figure 8 that filter with constraints has a smooth performance on z axis. It is probably because the backtrck mechanism give filter chances to find reliable data to modify filter coefficients. This mechanism makes system robust to unusual variation while recording normal movement data. But when participants do squats slowly, there might be continuous sensor failure which is hard to modified by filter. The only solution of this article is to shut down the filter and restart when system can successfully observe key-point’s position. In order to judge these two filters, precision and smoothness are calculated and shown in table 2 and table 3.

Table 2. Precision of two filters (judged by formula 8, 9, 10).

<table>
<thead>
<tr>
<th>Participant number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data</td>
<td>90%</td>
<td>88%</td>
<td>83%</td>
<td>75%</td>
</tr>
<tr>
<td>Filter without constraints</td>
<td>95%</td>
<td>93%</td>
<td>90%</td>
<td>84%</td>
</tr>
<tr>
<td>Filter with constraints</td>
<td>99%</td>
<td>99%</td>
<td>96%</td>
<td>90%</td>
</tr>
</tbody>
</table>

Table 3. Smoothness of two filters.

<table>
<thead>
<tr>
<th>Participant number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter without constraints</td>
<td>0.039</td>
<td>0.041</td>
<td>0.058</td>
<td>0.081</td>
</tr>
<tr>
<td>Filter with constraints</td>
<td>0.022</td>
<td>0.023</td>
<td>0.029</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Table 4. Root mean squared error (visual masking frames excluded).

<table>
<thead>
<tr>
<th>Participant number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter without constraints</td>
<td>0.098</td>
<td>0.086</td>
<td>0.054</td>
<td>0.042</td>
</tr>
<tr>
<td>Filter with constraints</td>
<td>0.087</td>
<td>0.079</td>
<td>0.044</td>
<td>0.038</td>
</tr>
</tbody>
</table>

During experiments, participant 3 and 4 are required to move slowly while 1 and 2 move quickly. Table 2 and 3 shows that filter with constraints has a better performance on smoothness and on precision than ordinary Kalman Filter. And it can predict and replace at least 60% of sensor failure data in the cases of all four participants. For slow movement, filters have less precision. It is probably because there is more long-time visual masking, which offers little information to fusion and causes filter restart. But from table 4, filters have better track effects during valid observation during slow movement. In this article, from point of view of prediction in visual masking, precision should be more important in final evaluation of filters. Therefore, this constrained filter is more adequate for quick movement.

4. Conclusion

This article aims to predict human body’s location during movement in transient visual masking. First it applies a light weight network on a CPU-only computer and reaches 3D data collection of 16 Hz. Then a constrained Kalman filter is proposed and modelling error is measured. After that, squats experiments are recorded and performance of constraints is discussed. The results show that these constraints enhance robustness of filter when there is abnormal variation of location data, and retain normal movement data. During visual masking, enhanced filter can predict at least 60% location data. And the filter has a better performance when it deals with slow movement.

Acknowledgments

This work is supported by Science and Technology Supporting Projects in the Biomedical Field, Shanghai Municipal Science and Technology Innovation Action Plan 2019 (Grant No. 19441908300).

References

Research on vehicle tracking technology in expressway cross-monitoring area based on radar and video fusion

Jianzhen Liu *,a, Baoguo Fengb

a JIAOKE Transport consultants Ltd., Beijing, China 100191;
b Hebei Deguroon Electronic Technology Co., Ltd., Shijiazhuang, China 050030;
* Corresponding author: 1911623157@qq.com

ABSTRACT

It is an important way to improve expressway monitoring management level to realize continuous real-time tracking and monitoring of abnormal driving vehicles such as overspeed, long time and low speed occupation of overtaking lane, continuous lane change, and dangerous goods transportation vehicles under different cameras. This article through studies the expressway ray regard convergence across monitoring area joint tracking technology, build the camera comprehensive control, camera control balance compensation and the monitoring area based on target motion continuity principle joint track model, such as through the accurate control of radar to detect vehicle more monitor cameras, solves the continuous tracking target vehicle monitoring technical problems. It has been successfully applied in Yanchong expressway, providing more convenient monitoring services for the expressway management department, and providing strong support for the decision-making of the expressway management department, which has high application value.

Keywords: Expressway, radar and video fusion, monitor systems, joint tracking

1. INTRODUCTION

By the end of 2021, China had 169,100 kilometers of expressways, ranking first in the world in mileage. Expressway operation monitoring technology mainly adopts vehicle detector, video image recognition, Internet data sharing and other ways to realize[1][2]. How to realize the real-time continuous tracking monitoring of vehicles with abnormal driving and dangerous goods transportation under different cameras, that is, target matching, coordination and continuous tracking across the monitoring domain, and improve the level of expressway monitoring management, is an urgent problem to be solved in the industry[3].

This paper focuses on the joint tracking technology of expressway cross-monitor area based on radar and video fusion. Radar tracking and video tracking fusion technology is a comprehensive application technology[4][5]. It not only needs to study a core model algorithm to dynamically adjust the horizontal and vertical angles of monitor cameras according to the position of radar detection target[6], but also needs to study the scheduling and tracking model algorithm of multiple monitor cameras along the line[7][8]. In addition, monitor cameras of different brands and models must be fully supported. In addition, the installation positions and layout methods must meet the requirements of actual scenarios, and the Angle error of onsite camera installation must be considered.

2. RESEARCH ON CAMERA OMNIDIRECTIONAL CONTROL MODEL

2.1 Camera control model algorithm

The precondition for the algorithm to work well is to ensure that accurate radar detection target position data is transmitted in real time. The important condition of this model is to ensure the radar real-time stable tracking and monitoring target. This model mainly analyzes the position relationship of multi-node latitude and longitude, and adopts the multi-point position analysis algorithm of two-dimensional plane coordinate system. Combined with the preset reference point position, the relative Angle and absolute Angle of the horizontal rotation of the camera are analyzed in real time. At the same time, the straight-line distance between the location of the monitoring target and the monitoring camera is calculated in real time, and the relative and absolute angles of the camera are adjusted vertically according to the mounting height of the camera.

The camera zoom adjustment refers to the vertical Angle analysis algorithm, and the standard zoom is calculated according to the straight-line distance between the target and the camera. The specific control model calibration formula is as follows:
\[
\begin{align*}
\mathbf{s} \cdot \mathbf{m}' &= A \cdot [R|t] \cdot M' \\
\mathbf{s} \cdot \mathbf{v} &= \begin{bmatrix} f_x & y & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}
\end{align*}
\]

(1)

2.2 Scheduling and tracking model algorithm of multiple monitor cameras

In the actual road, the relatively uniform distribution of monitor cameras makes it possible to continuously track and monitor the target\(^9\). However, it is necessary to develop perfect scheduling algorithm and control algorithm to realize continuous tracking of multiple cameras\(^10\). In actual scenarios, roads are often two-way roads, even urban roads, and the uncertainty of vehicle driving direction may cause a large waste of computing resources and low efficiency. Therefore, the main difficulty of this algorithm is how to locate the next scheduling resource quickly and reduce the unnecessary waste of computing resources.

At the beginning of the system operation, the topology of road monitor cameras is analyzed, and the camera scheduling relationship network is constructed, and the scheduling combination of neighboring monitor cameras is collected for each monitor camera. During the monitoring target tracking, the location of the monitoring target is analyzed in real time, and the appropriate adjacent camera in the scheduling relationship network of the current main monitor camera is calculated timely. When the monitoring range of the moving adjacent camera of the target is monitored, the monitoring camera is adjusted. And so on, gradually dispatching monitor cameras. The scheduling tracking model algorithm is as follows:

\[
A(\theta) = -\min \left[ 12 \left( \frac{\theta}{\theta_{3dB}} \right), A_m \right], \quad -180 \leq \theta \leq 180
\]

(2)

2.3 Multiple camera control methods

Through the above algorithm, the basic tracking data calculation of real-time tracking and monitoring of the camera is basically solved. The horizontal Angle, vertical Angle and scaling times are converted into horizontal parameters, vertical parameters and scaling parameters of the camera, and the control command is sent to the camera through the unified standard control interface to realize camera Angle adjustment.

But consider the monitor camera manufacturers numerous, numerous models. And the manufacturer of various types of monitor cameras may have different levels of parameters or Angle and the different characteristics of the hardware level, which may lead to the actual monitoring Angle slightly deviation, in the actual practice step by step according to the manufacturer all kinds of camera parameter adaptive adjustment, can be a good solution to the brand characteristic deviation of the problems of many models.

2.4 The influence of the placement position and mode of monitor cameras

The standard tracking model algorithm was born in the experimental environment\(^11\). In the actual road scene, there may be different camera placement positions, various placement methods and installation heights. The tracking algorithm based on position relationship is a position and Angle sensitive algorithm. Due to the special situation of the placement position and mode in the actual scene, large errors may occur in the target monitoring. In order to solve such problems, it is necessary to carry out research on the impact caused by actual installation and layout, such as layout location and layout method.

Through the road test installation, central isolation belt installation, road test installation, gantry installation, independent rod or common rod installation and other installation methods of the test. A variety of fault-tolerant mechanisms and analysis algorithms are gradually applied to ensure that the monitoring target can still be tracked and monitored with high precision under different installation positions and installation methods.

2.5 Camera control balance compensation technology

During the actual installation of the camera or due to natural conditions such as wind, the vertical Angle deviation of the camera often occurs\(^12\). The tracking algorithm based on position relation belongs to position and Angle sensitive algorithm. The slight deviation of Angle may cause a large deviation in the tracking and monitoring process when the camera is tracking the distant target with high multiple scaling. Therefore, it is necessary to carry out research on vertical Angle balance compensation technology.
In actual projects, the system configuration personnel lock the relative reference angles of the left and right sides according to the monitoring deviation of the two sides of the camera in the horizontal direction\[^{[13]}\]. The system automatically calculates the balance compensation Angle according to the internal model according to the left and right reference angles. To ensure that the camera vertical Angle deviation problem is well solved. The vertical Angle balance compensation algorithm is as follows:

\[
\begin{align*}
X_i &= x \cdot \cos(P) + Y \cdot \sin(R) \cdot \sin(P) - Z \cdot \cos(R) \cdot \sin(P) \\
Y_i &= Y \cdot \cos(R) + Z \cdot \sin(R)
\end{align*}
\]

(3)

3. JOINT TRACKING MODEL ACROSS MONITORING AREA BASED ON RADAR AND VIDEO FUSION

3.1 Object detection model based on location correlation

Two thresholds are used in the cross-monitor joint tracking model based on radar and video fusion. The first threshold is directly selected by using the relative amplitude value of the sample sequence. Suppose that the echo amplitude sequence of N range units in a working cycle obtained at the IF output end is \(X_1, X_2, \ldots, X_N\), from which we choose the maximum value: \(X_{\text{max}} = \max_{i=1}^{N} \{X_i\}\). Then, the first threshold can be selected as: \(Th_1 = Z \cdot X_{\text{max}}\), where \(Z\) is the confidence coefficient, which can be determined according to the signal-to-noise ratio. The research shows that its value can be selected between 0.3 and 0.7. Set the echo amplitude sequence \(X_1, X_2, \ldots, X_N\). Where \(\arg\{\}\) denotes the operation of taking variables.

\[
I_i = \begin{cases} 
(i, x_i \geq Th_1) \\
0, x_i < Th_1 
\end{cases}
\]

(4)

Then the output after the first threshold detection is the location information that exceeds the threshold. If the position sequence number \(i=0\) is removed, the remaining set is: \(I = \{i_1, i_2, \ldots, i_K\} \{1, 2, \ldots, N\}\), that is, there are \(K\) possible "point targets" exceeding the first threshold, and they are respectively distributed in \(i_1, i_2, \ldots, i_K\) distance cells. Based on the first scatterer of the \(K\) possible "point targets", the second, the third... , the \(K\)TH scatterer is successively compared with the previous scatterer without repetition to obtain the distance information, then the scatterer position information set \(I\) can be converted into the distance information set.

\[
\begin{align*}
IR_1 &= \{r(1, 2), r(1, 3), \cdots r(1, K)\} \\
IR_2 &= \{r(2, 3), r(2, 4), \cdots r(2, K)\} \\
IR_{K-1} &= \{r(K-1, K)\}
\end{align*}
\]

(5)

Where, \(r(j, k) = (k-j) \times \delta R; K > j\); \(J = 1, 2, \ldots, K - 1\); \(k=j+1, j+2, \ldots, K\); \(\delta R\) is radar resolution.

The basic principle of the second threshold detection method is: continuously compare the maximum radial distance between the elements in the set \(\{IR_1, IR_2, IR_{K-1}\}\) and the target, and set the counter \(I_{\text{UC}}(c \text{ is the target batch})\) to count the echo sampling points belonging to the target batch \(c\) within the length of \(LR\). In order to avoid false detection or missing detection, assume that within the \(LR\) distance, there is a scatterer \(i_k\), the distance between the former \(i_{k-1}\) and the latter \(i_{k+1}\) is within \(LR\), but the distance between \(i_{k-1}\) and \(i_{k+1}\) is not within \(LR\). In this case, we judge that \(i_k\) belongs to the following target group.

\[
\begin{align*}
\text{If } r(k, k + 1) < r(k, k + 1), \text{Then } i_k \text{ and } i_{k-1} \text{ belong to the same target group} \\
\text{If } r(k, k + 1) > r(k, k + 1), \text{Then } i_k \text{ and } i_{k+1} \text{ belong to the same target group}
\end{align*}
\]

(6)

According to the above principle, \(C\) possible target groups are obtained by comparing, judging and counting to the KTH sampling point, and then the second threshold \(Th_2\) is set as the sampling points distributed along the radial direction of the target (generally 5-8). In other words, the number of sampling points exceeding \(Th_2\) is considered as the target. However, in the detection process, it is found that it is sensitive to the selection of threshold and coefficient. If the threshold is set slightly higher, the target cannot be detected, while if the second threshold is set slightly lower, it is easy to judge one target as two or even several targets. We adopt the following methods to improve the performance of target detection: that
is, the first threshold and the second threshold are set low, so as not to miss the target. After the target is detected, according to the preset threshold, the targets with very close locations are combined into one target.

### 3.2 Joint tracking model across monitoring area based on the principle of target motion continuity

According to the principle of continuity of target motion, the trajectory of each target is continuous\(^{[14]}\), and the derivatives are also continuous\(^{[15]}\). According to the sampling theorem, as long as the frequency of radar sampling the target motion is high enough, the spatial position of the target can be predicted and the trajectory of the target can be fitted by the target position information collected from each frame. Based on this idea, the specific principle is as follows: Let the position of the target in the data of the first frame be \(A(X_{11}, Y_{11})\), and the possible position in the second frame should be in the circle with \(A(X_{11}, Y_{11})\) as the center and the maximum motion speed \(R_2\) as the radius. All the targets in the circle (suppose there are two) \(B_1(X_{21}, Y_{21}), B_2(X_{22}, Y_{22})\). It is possible that the target \(A\) moves to the position of the second frame. According to point \(A\) and \(B_1\), the target moving speed is:

\[
\begin{align*}
V_{1x} &= (X_{21} - X_{11})/\Delta t \\
V_{1y} &= (Y_{21} - Y_{11})/\Delta t
\end{align*}
\]  

(7)

But considering \(\Delta t\) as a constant, the unit can be 1, which can be written as:

\[
\begin{align*}
V_{1x} &= X_{21} - X_{11} \\
V_{1y} &= Y_{21} - Y_{11}
\end{align*}
\]  

(8)

Extrapolation from \(V_{1x}\) and \(V_{1y}\) can be obtained as follows: the extrapolation position \(O_1(a_1, Z_1)\) of \(A\) via \(B_1\) in frame 3 is expressed as:

\[
\begin{align*}
a_1 &= X_{21} + V_{1x} \\
Z_1 &= (Y_{21} + V_{1y})
\end{align*}
\]  

(9)

The extrapolation error is set as a circle with radius \(R_3\). In the circle with center \(O_1(a_1, Z_1)\) and radius \(R_3\), all the targets (suppose there are two) \(C_1(X_{31}, Y_{31})\) and \(C_2(X_{32}, Y_{32})\) may be the images of target \(A\) moving through \(B_1\) to the third frame. Let’s say the velocity of \(C_1(X_{31}, Y_{31})\) is \(V_{2x}, V_{2y}\).

\[
\begin{align*}
V_{2x} &= X_{31} - X_{21} \\
V_{2y} &= Y_{31} - Y_{21}
\end{align*}
\]  

(10)

The acceleration \(a_x, a_y\) can be expressed as:

\[
\begin{align*}
a_x &= V_{2x} - V_{1x} \\
a_y &= (V_{2y} - V_{1y})
\end{align*}
\]  

(11)

Further extrapolated by \(V_{2x}, V_{2y}\), the extrapolated position \(O_2(a_2, Z_2)\) of \(A\) via \(B_1, C_1\) in frame 4 is obtained.

\[
\begin{align*}
a_2 &= X_{31} + V_{2x} + \frac{1}{2} a_x \\
Z_2 &= Y_{31} + V_{2y} + \frac{1}{2} a_y
\end{align*}
\]  

(12)

Similarly, if the extrapolation error range is set as \(R_4\), all the targets (suppose there are two) \(D_1(X_{41}, Y_{41})\) and \(D_2(X_{42}, Y_{42})\) in the circle with \(O_2(a_2, Z_2)\) as the center and \(R_4\) as the radius may be the result of extrapolation of target \(A\) through \(B_1\) and \(C_1\). They directly extrapolated the distance deviation of position \(O_2(a_2, Z_2)\) as \(\Delta r_1, \Delta r_2\). Angle deviation \(\Delta h_1, \Delta h_2\), so that the synthesis error is

\[
\Delta I_j = (\Delta r/C_jD_j)^2 + \Delta h^2, \quad i = 1, 2, 3, 4, \quad j = 1, 2, 3, 4, 5, 6, 7, 7
\]  

(13)
It is assumed that there may be two targets in the prediction circle after each target is extrapolated, $C_i$ represents the target that may appear in the third frame, $D_j$ represents the target that may appear in the fourth frame, so there are four possibilities for $i$ and eight possibilities for $j$. Each branch is tracked to obtain $\Delta l_1, \Delta l_2, \ldots, \Delta l_8$. The branch with the smallest $\Delta l_i$ is selected as the most likely trajectory. If the minimum $\Delta l$ is greater than a set value (Accept_R), that is, the most likely trajectory error is too large, then all the tracking is invalid.

4. EXAMPLES OF APPLICATION

The research results were demonstrated and applied on the smart expressway in Yanzhong (Hebei section), where omnidirectional millimeter-wave radar high-reliability traffic edge computing equipment, license plate feature recognition camera and other auxiliary equipment were deployed along the expressway, and a visual management platform was also deployed.

Through practical application, the problem of continuous tracking and monitoring technology of target vehicle is successfully solved, which provides more convenient monitoring service for the expressway management department and provides strong support for the decision-making of the expressway management department.

![The camera is continuously regulated according to the vehicle position in real time to realize vehicle tracking](Image)

Figure 1. Continuous vehicle tracking and monitoring function across the monitoring area

Through multiple groups of test data of multiple expressways, the function of continuous tracking and monitoring of vehicles across the monitoring area is realized, and the accuracy of continuous tracking and monitoring of vehicles reaches more than 95%, which provides solutions for the management of special vehicles and has broad application prospects.

5. CONCLUSION

This paper studies the realization of continuous real-time monitoring under different cameras for abnormal driving vehicles such as overspeed, long time low speed occupation of overtaking lane, continuous lane change, and special vehicles such as dangerous goods transport. This paper provides more convenient monitoring services for the highway management department, and provides strong support for the decision-making of the highway management department. In this paper, the joint tracking technology of thundersight fusion across monitoring area is studied. By constructing the model of camera omnidirectional control, camera control balance compensation and cross-monitoring joint tracking based on the principle
of target motion continuity, the problem of continuous tracking and monitoring technology of target vehicle is solved. It has been successfully applied to the smart expressway in Yanchong (Hebei section) and has high application value.

REFERENCES

An image analysis tool for cell migration assay with a large sample size

Di Yin\textsuperscript{1}a, Hongbo Zhang\textsuperscript{1}b, Shih-Mo Yang\textsuperscript{2}c, Ruixue Yin\textsuperscript{1}d, Wenjun Zhang\textsuperscript{1*3}e

\textsuperscript{1}School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai, P.R. China
\textsuperscript{2}Biomedical Science and Technology Research Center, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, P.R. China
\textsuperscript{3}Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, Canada

\textsuperscript{a}zhizhuxia.oscar@gmail.com, \textsuperscript{b}hbzhang@ecust.edu.cn, \textsuperscript{c}smyang@shu.edu.cn,
\textsuperscript{d}yinruixue@ecust.edu.cn, \textsuperscript{e}chris.zhang@usask.ca

Abstract

Cell migration assay is the most common research approach for cell migration. Quantitative research and analysis are carried out by measuring the migration of cells into the region that is artificially created among confluent monolayer cells. To improve the efficiency and accuracy of the analysis, the software/tools were developed to assist the image analysis process. However, these software/tools are still at the stage of measuring a single sample, which can not satisfy the requirement of large sample size for cell migration assay device. In this paper, an image analysis tool based on Fiji is developed, which can segment multiple samples from a scanned image and then analyze a single sample in batch. In addition, the screening function should be added for the application scenario of large sample size. The samples can be filtered according to different conditions to improve the consistency of experimental conditions. The results show that the developed analysis tool ATCA has high accuracy in identifying cell-free zones, with a difference of 2.3\% from the tool WHST and 2.9\% from manual operation. The analysis efficiency of this tool is 15 times that of manual operation.

Keywords—cell migration; cell motility; image analysis tool; image segmentation

1. Introduction

Cell migration refers to the movement of cells from one location to another. As a ubiquitous form of movement of living cells, cell migration is a crucial part of many physiological and pathological processes, such as wound healing, tissue homeostasis and morphogenesis\textsuperscript{[6-7]}. On the other hand, cell migration can also lead to serious consequences, including tumor formation and metastasis, rheumatoid arthritis\textsuperscript{[4, 5]}. To build the in vitro model of cell migration, the widely used approach is cell migration assay that is to make the confluent monolayer cells migrate into artificially created cell-free zone (CFZ). By measuring the changes of the cell-free areas during the process, the migration ability of cells, the response of cell populations to extracellular stimuli, and the interaction between cells were investigated\textsuperscript{[6, 7]}. To measure the changes, the most commonly used method is to capture the images in the bright field or fluorescence field or phase field during cell migration, and extract the information of cell-free areas from images. However, this extraction process is mostly done by hand. This complicated process consumes a lot of time and energy of the operator. In addition, the process was influenced by the subjective factors of the researchers, which leads to the high error rate, low accuracy and low comparability of the data. Therefore, many software tools based on the technology of image segmentation or machine learning were developed, such as Wound-healing-size-tool (WHST), and Microscope image analysis toolbox (MiToBo)\textsuperscript{[8, 9]}. These tools can accurately identify CFZ under unified standards, and automatically calculate the width, area, and migration rate. With proper adjustment, these tools can be applied to the images of different cell types by different imaging technologies. However, due to the current limitations of algorithms, these tools are still limited to identify a single CFZ, and can not recognize multiple CFZs in a scanned image because of the data noise. This obviously can not meet the current demand for analysis of experimental data in cell migration assay\textsuperscript{[10, 11]}. To this end, an image analysis tool based on Fiji is developed, which can segment multiple CFZs from a scanned image and then analyze a single CFZ in batch. In addition, the tool integrates a screening function to select samples. The function in this paper is applied to screen CFZ samples under similar fluid shear environment to improve the consistency of experimental conditions. The scanned images are from a cell migration assay device made in previous work. The quantitative analysis results of fluid shear environment are derived from the shear stress distribution simulated by finite element method in our previous work\textsuperscript{[12]}.

2. Method

The tool consists of three sub-modules. Each sub-module is coded using the built-in language of Fiji, an open source image processing software. The data flow diagram of the tool is shown in Figure 1. The function of the first sub-module (M1) is to segment all CFZ in the scanned image and save them by number. The function of the second sub-module (M2) is to find all CFZ within the region of interest and record the CFZ number in the region. The third sub-module (M3) is controlled by the output of the first two sub-modules. It processes and analyzes the segmented CFZ in batches according to the number of interest, and outputs CFZ’s number and corresponding area data.

![Data Flow Diagram](image)

Figure 1. The overall analysis pipeline illustrated by data flow diagram.

The detail of M1 is shown in Figure 2. After inputting the scanned image, call Fiji’s built-in rectangle tool to manually select the boundaries of all CFZs (Figure 2(a)), so as to eliminate the interference of other regions. Then a rectangular array automatically expanded within the boundary according to the pre-programmed pattern. The CFZs were segmented one by one and numbered from left to right, top to bottom (Figure 2(b)). The segmented images are output to a new folder and named with their number.

![Segmentation and Numbering Process](image)

Figure 2. segmentation and numbering process, input operation (a) and output result (b) of scanned image. The scale bar is 1 mm.

M2 use of pixel value differences in different channels to find the number of CFZ within the region of interest. The specific analysis process is shown in Figure 3(a). The input image is a simulation result of the shear stress distribution, where the region of interest is the region of shear stress error less than 10% (Figure 3(b)). After the input, call the built-in straight tool to draw a straight line across the circle of CFZ, and then calculate the approximate CFZ’s radius and area. On the other hand, the program splits the image into red and green channels. The red channel, shown in Figure 3(c), is analyzed to find the coordinates of all the CFZs in the image (Figure 3(d)) and the CFZs are numbered from left to right, top to bottom. The image of the green channel is shown in Figure 3(e) and can be analyzed to find the coordinates of the CFZs within the shadow (Figure 3(f)). Finally, by matching the coordinates of CFZs (within a certain error) obtained in figure 3(c) and figure 3(e), the number of interested CFZs can be obtained and output.
The third sub-module (M3) processes a single CFZ image and obtains its area. This module is based on variance filter that distinguish CFZ by comparing the pixel value of given point to that of the peripheral. The steps are refined from the methods of other literatures [8, 9, 13]. After input of the original image (Figure 4(a)), the image is processed by powering, variance filter, top-hat filter, Otsu segmentation and hole filling, and the output is shown in Figure 4(b). Then, the particle analysis function is called to identify the white blocks in the image and calculate their number of pixels. According to the size characteristics of CFZ, the largest white block was chosen as CFZ. With the collection of CFZs’ in different time period, the cell migration process can be accurately quantified by ATCA (Figure 4(c)).

By integrating the above three sub-modules, an image analysis tool with screening function for cell migration assay with a large sample size can be realized. The code of the tool has been uploaded to the github (https://github.com/Bayoscar/Analysis_tool_for_CFZs_array.git). The tool only needs to input scanned image and region of interest to output segmented images, image masks and corresponding area data.
3. Results and discussion

Sub-module M3 is the core module of the tool ATCA. To validate its accuracy, the analysis results are compared with those from manual measurement as well as other analysis tools. The analysis tools are open source and freely available such as MRI_Wound_Healing_Tool (MRI), Mitobo, WHST\(^8,9,13\). During processing, no parameter should be changed to ensure the consistency of the processing standard. The results are shown in Figure 5, and it can be seen that the delineation of CFZ boundaries by other tools is very similar or even coincides, except MiToBo. Because MiToBo's algorithm modifies the boundary determination and is more suitable for elongate CFZ. The difference of other tools can be more clearly observed from Figure 5(b), and these tools have different treatment of gaps between cells. MRI and ATCA were used to correct these gaps, which was more similar to manual processing. To further quantify the differences among tools, a comparison is made by relative errors, as shown in Figure 5(c). The minimal difference is ATCA and WHST (2.37% ± 5.8), which share the same processing principle i.e. variance filter. The maximal difference is ATCA and MRI (-8.93% ± 4.0). The main reason is that MRI is highly dependent on parameter adjustment (5 parameters), and large differences will occur when analyzing large quantities of samples with the same parameters. The difference between ATCA and manual measurement is also small (-2.9% ± 4.7), which indicates that the tool is highly accurate and can replace an experienced operator. In addition, the processing speed of ATCA is 15 times that of manual operation, which greatly improves the efficiency of image analysis.

ATCA was used to analyze the scanned images of HUVEC under different shear stress in cell migration assay. The simulation results were introduced to screen the samples with shear stress error less than 10%\(^12\). Figure 6 shows the cell migration area under shear stress of 0, 1 and 4 dyne/cm\(^2\) in 24 hours. The fastest migration occurs in the group without shear stress and the slowest migration is the group under shear stress of 4 dyne/cm\(^2\).
4. Conclusion

An cell-free area analysis tool, which consists of three functional sub-modules is introduced in this paper. The tool provides a solution to the CFZs’ analysis with large sample size and integrates the screening function to select the samples of interest for analysis. By comparing with other tools and manual measurement, it is proved that the tool has high accuracy and high efficiency. And the practicability of the tool is proved by being used to analyze the images of HUVEC in cell migration assay.

References

Continuous sign language recognition based on 3DCNN and BLSTM
Hengbo Zhang, Daming Li *, Nana Fu
College of Physics, Electronics and Electrical Engineering, Ningxia University,
Yinchuan, Ningxia 750021, China
*Corresponding author: nxldm@126.com

ABSTRACT

Sign language recognition can make the communication between deaf mutes and healthy people more convenient and fast. In recent years, with the continuous development of deep learning, the research method of deep learning has also been introduced into the field of sign language recognition. Compared with the recognition of isolated words, the recognition of continuous sign language is more time-dependent. The current research still has shortcomings in recognition accuracy. Therefore, we proposed a continuous sign language recognition method based on 3DCNN and BLSTM. Based on the spatial feature information extracted by 3DCNN and the short-term temporal relationship established, the global temporal modeling of the video information of continuous sign language is carried out by using the bidirectional semantic mining ability of BLSTM. The CTC loss function is used to solve the problem of time series label misalignment. At the same time, we add the calculation of auxiliary loss function and auxiliary classifier. Experiments show that the auxiliary loss function and classifier can effectively reduce the error rate of the network. The word error rate of the continuous sign language recognition algorithm proposed in this paper on the large continuous sign language dataset RWTH-PHONEIX-Weather 2014 is as low as 23.5%, which is lower than the previous algorithm.

Key words: deep learning; Continuous sign language recognition; 3D CNN; BLSTM; CTC;

1. INTRODUCTION

As the spoken language in the daily conversation of the deaf, sign language plays a vital role in the communication of the deaf. Millions of people around the world are using sign language, which includes human subjective factors (such as gestures) and non subjective factors (such as facial expressions) [1]. And sign language has its own vocabulary information and grammar, the task of video based continuous sign language recognition is more complex, which is different from isolated word sign language recognition. Its goal is to translate a sign language video into a corresponding sign language context sequence, so it is not only necessary to extract the corresponding features of each frame image in the video, but also need to model the temporal context relationship between each frame and extract corresponding temporal-features. With the rapid development of deep learning technology in the field of computer vision in recent years, the recognition of isolated words and static single pictures has made good progress [2~4], but there are still many deficiencies in the accuracy and anti over fitting ability of continuous sign language recognition tasks.

This paper proposes a continuous sign language recognition algorithm based on 3D CNN [5] and BLSTM [6], which uses the information fusion extraction ability of 3D CNN for spatial and temporal dimensions to extract the corresponding features of sign language recognition for each frame of pictures in the video, and extracts the features in the short-term time dimension. We use BLSTM network to establish bidirectional time series relations, conduct global time series modeling for short-term spatiotemporal features extracted from 3D CNN network, solve the sequence alignment problem in the translation process by using CTC algorithm [7], and add auxiliary loss function calculation to the loss function to avoid over fitting [8].

2. ALGORITHM DESIGN

The overall architecture of this algorithm is shown in Figure 1. In order to enhance the anti-noise ability of the network and reduce over fitting, video frames are first enhanced by data such as cut and flip, color gamut transformation, and then continuous video frames are extracted through 3DCNN network for temporal and spatial features, and the extracted features are sent to BLSTM for context global bidirectional temporal modeling. The CTC loss function is used to calculate the loss of label alignment. Before BLSTM, a classifier and CTC auxiliary loss calculation are also set up to
enhance the ability of 3D CNN network to extract sign language information and suppress the occurrence of over fitting. In the process of translation, the results of short-term spatio-temporal information extracted from 3D CNN network are also introduced.

![Figure 1. Overview of algorithm structure in this paper](image)

### 2.1 3D CNN

3D CNN is developed from 2D CNN and is mainly used in the field of video classification and action recognition. Its advantage is that it can extract information in both temporal and spatiotemporal dimensions. In the process of 3D convolution, multiple consecutive video frames are stacked into a cube. Its convolution core can move in the height, width, and channel of the image. Compared with 2D convolution, 3D convolution increases the time dimension. For different positions in the convolution kernel window, weight sharing is carried out, so each feature map in the convolution layer will be connected with multiple adjacent continuous frames in the previous layer, so motion information can be captured. The definition of 3D CNN is shown as follow:

\[
y_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_{m} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} w_{ijm}^{pqr} y^{(x+p)(y+q)(z+r)}\right)
\]

(1)

Where, \(y_{ij}^{xyz}\) is the output value on the \(j\)th feature map of layer \(i\), \(b_{ij}\) is the offset term, \(m\) is the coordinate of the set of layer \(i-1\) feature maps connected to the current feature map, and \(w_{ijm}^{pqr}\) is the weight coefficient. \(P\), \(Q\), and \(R\) are the length, width, and time dimensions of the convolution kernel.

Because of the excellent effect of 3D CNN in video understanding, we use the network to extract sign language features from consecutive video frames in sign language video and establish short-term temporal relationships. The 3D CNN network used in this paper is ResNet18-3D network transformed from ResNet network by using I3D method. This paper makes some modifications to the network structure for sign language recognition. See Table 1 for its structure. The parameters of 3D CNN network are large, so a large amount of sample data is needed for training. This paper uses the method of transfer learning to solve the problem of small number of samples. The data set of Kinetics includes actions between people and objects. The dataset has 400 human action classes, each class has 400 or more clips, each clip comes from a unique video, and there are a total of 240k training videos. These clips lasted about 10 seconds, and there was no untrimmed video. First, the 3D CNN network is trained with the Kinematics dataset, and the weight obtained from the pre training is used as the initialization parameter of the sign language recognition network model, which is migrated to the sign language dataset to fine tune the entire network.
2.2 BLSTM

The information segmentation of the time dimension of the continuous sign language recognition task is a key point. At present, the common scheme is to convert it to the segmentation and recognition of isolated words. However, a lot of calculations are introduced, and the segmentation of isolated words is not accurate. RNN [9] has the ability to establish context temporal relationship, but its memory ability for time is limited. With the increase of sequence length, the distance between relevant information also increased. When the distance is too large, RNN will lose the ability to learn the relationship between long-distance information. LSTM (long short term memory) network is a special RNN network added with a gating structure, which can overcome the defect that RNN cannot remember for a long distance, and solve the problem of small gradient disappearance in the process of back propagation through a special gating structure. The video of continuous sign language contains rich contextual information. For the sentence information expressed in sign language, it is not only required to contact the information before the words, but also the information in the following text. BLSTM has two two-way information flow routes, forward and backward, which can capture the relationship between the preceding text and the following text at the same time, and can more easily extract the contextual semantic information of sign language. As shown in Figure 2. BLSTM is composed of a forward LSTM and a backward LSTM network, which ensures that each time step in the network includes forward and backward time relationship calculation, and mining the semantic information in the continuous sign language video stream more fully by combining the two opposite time stream information from the past to the future and from the future to the past. The structure of each hidden cell is shown in Figure 3, and the calculation method of each hidden cell is shown as follows.

![Figure 2. Structure of BLSTM](image)

$$
i_t = \sigma(U_i v_i + W_i h_{t-1} + b_i)$$
$$f_t = \sigma(U_f v_i + W_f h_{t-1} + b_f)$$
$$o_t = \sigma(U_o v_i + W_o h_{t-1} + b_o)$$
$$c_t = \sigma(U_c v_i + W_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = \tanh(c_t) \odot o_t$$

(2)

Where $i_t$, $f_t$, $o_t$ correspond to input gate, forgetting gate and output gate, and vector $h_t$ and $c_t$ are implicit and cellular. Where $U$ and $W$ are the weight matrix input to hide and the weight matrix hidden to hide, and $b$ is the deviation vector. The unit sigmoid is represent by $\sigma$. $\odot$ represents Hadamard product. In this paper, BLSTM temporal network is used to build a global bidirectional temporal information model by combining the tensor information extracted by 3D CNN with

### Table 1. Structure of 3D CNN

<table>
<thead>
<tr>
<th>Layer-name</th>
<th>Input-channels</th>
<th>Kernel-size</th>
<th>Output-channels</th>
<th>Input-feature-size</th>
<th>Output-feature-size</th>
<th>Stride</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1-t</td>
<td>3</td>
<td>7×7×7</td>
<td>64</td>
<td>3×180×224×224</td>
<td>64×180×112×112</td>
<td>1×2×2</td>
</tr>
<tr>
<td>Maxpool</td>
<td>64</td>
<td>3×3×3</td>
<td>64</td>
<td>64×180×224×224</td>
<td>64×90×56×56</td>
<td>2×2×2</td>
</tr>
<tr>
<td>Layer1</td>
<td>64</td>
<td>3×3×3</td>
<td>128</td>
<td>64×90×56×56</td>
<td>128×90×56×56</td>
<td>1×2×2</td>
</tr>
<tr>
<td>Layer2</td>
<td>128</td>
<td>3×3×3</td>
<td>256</td>
<td>64×90×56×56</td>
<td>256×90×28×28</td>
<td>1×2×2</td>
</tr>
<tr>
<td>Maxpool</td>
<td>256</td>
<td>3×3×3</td>
<td>256</td>
<td>128×45×28×28</td>
<td>256×45×14×14</td>
<td>2×2×2</td>
</tr>
<tr>
<td>Layer3</td>
<td>256</td>
<td>3×3×3</td>
<td>256</td>
<td>128×45×28×28</td>
<td>256×45×14×14</td>
<td>1×2×2</td>
</tr>
<tr>
<td>Layer4</td>
<td>256</td>
<td>3×3×3</td>
<td>512</td>
<td>256×45×14×14</td>
<td>512×45×7×7</td>
<td>1×2×2</td>
</tr>
<tr>
<td>Ave pool</td>
<td>512</td>
<td>——</td>
<td>512</td>
<td>512×45×1×1</td>
<td>512×45</td>
<td>——</td>
</tr>
<tr>
<td>Flatten</td>
<td>512</td>
<td>——</td>
<td>512</td>
<td>512×45×1×1</td>
<td>512×45</td>
<td>——</td>
</tr>
</tbody>
</table>

**Ave pool** and **Flatten** are added with a gating structure, which can overcome the defect that RNN cannot remember for a long distance, and solve the problem of small gradient disappearance in the process of back propagation through a special gating structure.
short-term temporal features and image spatial features, to deeply mine the semantic information in continuous sign language videos, and to prevent over fitting, the algorithm in this paper adds the Dropout layer after BSLTM.

2.3 Loss Function

CTC (Connectionist Temporary Classification) algorithm is designed for end-to-end time classification tasks with non segmented data. To provide more effective supervision, CTC introduces a "blank" to represent unlabeled data (such as motion insertion or non gesture fragments in continuous sign language recognition), and aligns them through dynamic programming. This not only solves the problem of misalignment between neural network label and output, but also avoids the problem of manual alignment of complex input and output. CTC defines a many to one function \( F: G^T \rightarrow G^{\leq T} \), aligns the tag sequence of the path \( \pi \in G^T \) and marks \( l \in G^{\leq T} \) by removing the duplicate tags and blank tags in the path in turn. For example: \( F(-aaa--aabbb-) = F(-a-ab-) = aab \). CTC uses this function to calculate the parameters of feature extractor and alignment module through the probability sum of all feasible paths \( \theta \) The formula for supervision is as follows:

\[
L_{CTC} = \log p(l|X; \theta) = \log \left( \sum_{\pi \in F^{-1}(l)} p(\pi|X; \theta) \right) \tag{3}
\]

The formula of \( p(\pi|X) \) can be obtained according to the conditional probability calculation, shown as follows:

\[
p(l|X; \theta) = \prod_{t=1}^{T'} p(\pi_t|X; \theta) \tag{4}
\]

In this paper, the BSLTM network is used to model the global bidirectional time series of the local attitude motion information extracted from the 3D CNN network, and the output information is sent to the classifier. After the classifier, a CTC loss function \( L_{CTC} \) is added to supervise the training of the network. In order to enhance the feature extraction ability of the network, this paper introduces an auxiliary classifier after the 3D CNN network to obtain the auxiliary classification weight, and also adds a CTC loss function \( L_a \) after the auxiliary classifier to assist the network training. The auxiliary loss function directly provides supervision for the feature extraction network. This auxiliary loss forced feature extractor can only predict based on local visual information, which is compatible with the main CTC loss function and can flexibly adapt to the network design. The Loss function as follows:

\[
Loss = L_{CTC} + \alpha L_a \tag{5}
\]

3. EXPERIMENTS

3.1 Dataset

We use the large German continuous sign language data set PHONEIX-Wather2014 \(^{[10]} \). PHOENIX14 dataset is a widely used continuous sign language recognition data set, recorded from the German TV weather forecast, and executed by 9 sign language interpreters with normal hearing. It contains 6841 sentences and 1295 different annotations. The dataset is divided into 5672 Train sentences, 540 Dev sentences and 629 Test sentences for multi signer settings.
3.2 WER
Word Error Rate (WER) is a commonly used indicator to evaluate the performance of recognition algorithms in continuous sign language recognition tasks \cite{11}. It is also called the length normalized editing distance. First, it aligns the identified sequence with the reference statement, and then calculates the number of operations transferred from the reference to the identified sequence, including the number of times of substitution (sub), deletion (del), and insertion (ins). N is the total number of words. The calculation method is shown as follows:

\[
WER = \frac{\text{sub} + \text{del} + \text{ins}}{N} \times 100\% \tag{6}
\]

3.3 Experimental Setup
The environment platform used in this experiment is Ubuntu 22.04 system+intel i7 10700k+Nvidia 3090 24G+16G memory. In the training process, set 80 epoches of training, the batchsize is set to 2, the initial learning rate is set to 0.0001, and divide by 10 in 40th epoch and 60th epoch respectively, using the Adam optimizer. All video frames are unified to 256 × 256, training set is randomly trimmed to size 224 × 224, and perform random horizontal flipping and random time scale scaling.

3.4 Experimental results and analysis

![WER training curve(left) and training loss curve(right)](image)

The algorithm proposed in this paper based on 3D CNN and BLSTM can reduce the word error rate on the PHOENIX2014 dataset to 23.5% on the development set (Dev) and 23.6% on the test set (Test), which is lower than the algorithm proposed in the past, and performs better on the PHOENIX2014 dataset, as shown in Table 2. This is due to 3D CNN's fine grain feature extraction of video frames and the fusion of local attitude and motion information, as well as BLSTM's modeling of two-way information in time dimension, which enables the network to more deeply mine the semantic information in sign language.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Backbone</th>
<th>Dev WER(%)</th>
<th>Test WER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SubUNet\cite{12}</td>
<td>CaffeNet</td>
<td>40.8</td>
<td>40.7</td>
</tr>
<tr>
<td>Stage-Opt\cite{13}</td>
<td>GoogLeNet</td>
<td>39.4</td>
<td>38.7</td>
</tr>
<tr>
<td>Re-Sign\cite{14}</td>
<td>GoogLeNet</td>
<td>27.1</td>
<td>26.8</td>
</tr>
<tr>
<td>SFL\cite{15}</td>
<td>ResNet18</td>
<td>26.2</td>
<td>26.8</td>
</tr>
<tr>
<td>STMC\cite{16}</td>
<td>VGG11</td>
<td>25.0</td>
<td>25.4</td>
</tr>
<tr>
<td>CrossModel\cite{17}</td>
<td>CNN</td>
<td>23.9</td>
<td>24.0</td>
</tr>
<tr>
<td>FCN\cite{18}</td>
<td>Custom</td>
<td>23.7</td>
<td>23.9</td>
</tr>
<tr>
<td>Ours</td>
<td>3D-ResNet</td>
<td><strong>23.5</strong></td>
<td><strong>23.6</strong></td>
</tr>
</tbody>
</table>

The impact of the structure in 3D ResNet on the extraction of video stream information is also explored as follows: 1. The impact of MaxPooling layers on feature extraction. 2. The influence of sampling rate on feature extraction in time dimension. 3. The influence of Conv1-t layer convolution kernel size and step size on feature extraction.
Table 3. Influence of different maxpooling layers on the network

<table>
<thead>
<tr>
<th>MaxPool layers</th>
<th>Dev WER(%)</th>
<th>Test WER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>26.8</td>
<td>27.0</td>
</tr>
<tr>
<td>1</td>
<td>24.1</td>
<td>24.0</td>
</tr>
<tr>
<td>2</td>
<td>23.5</td>
<td>23.6</td>
</tr>
<tr>
<td>3</td>
<td>24.5</td>
<td>24.7</td>
</tr>
</tbody>
</table>

Table 3 shows that it is necessary to add the maximum pooling layer in 3D CNN. However, adding too many maximum pooling layers may cause the loss of fine-grained information in the features, thus increasing the error rate. In addition, using the maximum pooling layer can reduce the space occupied by GPU memory and improve the computing efficiency.

Table 4. Influence of sampling ratio on feature extraction in time dimension of network

<table>
<thead>
<tr>
<th>Downsampling ratio</th>
<th>Dev WER(%)</th>
<th>Test WER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>32.6</td>
<td>33.5</td>
</tr>
<tr>
<td>2</td>
<td>27.1</td>
<td>27.7</td>
</tr>
<tr>
<td>4</td>
<td>23.5</td>
<td>23.6</td>
</tr>
<tr>
<td>8</td>
<td>26.3</td>
<td>26.9</td>
</tr>
</tbody>
</table>

Table 4 shows that downsampling the time dimension can increase the receptive field of the network to the time dimension, thus reducing the word error rate. In addition, the motion information extracted from the sampling rate is too short for BLSTM network to perceive the amplitude change of the motion information under low time. Too large down sampling rate makes the receptive field too large, so the perception of motion fine grain information is not sufficient, which may lead to misjudgment for similar but different motion changes.

Table 5. Influence of conv1-t parameters on feature extraction

<table>
<thead>
<tr>
<th>Kernel Size/Stride</th>
<th>Dev WER(%)</th>
<th>Test WER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3,7,7)/(1,2,2)</td>
<td>24.8</td>
<td>24.7</td>
</tr>
<tr>
<td>(7,7,7)/(1,2,2)</td>
<td>24.3</td>
<td>25.0</td>
</tr>
<tr>
<td>(11,7,7)/(1,2,2)</td>
<td>25.4</td>
<td>25.6</td>
</tr>
<tr>
<td>(7,7,7)/(2,2,2)</td>
<td>27.5</td>
<td>28.0</td>
</tr>
<tr>
<td>(7,7,7)/(1,2,2)</td>
<td>23.5</td>
<td>23.6</td>
</tr>
</tbody>
</table>

Conv1-t is the first layer of 3D CNN network, which is the key layer for converting RGB video image information into neural network recognizable tensor. Resnet can extract excellent features from a single image in a video frame. However, for continuous sign language recognition tasks, feature extraction and information fusion of time dimension are different from video classification tasks, so network layer structure is very important for processing time dimension. Table 5 shows that when the step size of the time dimension is 2, the WER of Conv1-t increases significantly, indicating that the features extracted by Conv1-t are not rich enough when the step size is 2, and the use of too large and too small convolution kernels in the time dimension will affect the quality of feature extraction.

For the design of the loss function, the main loss function $L_{CTC}$ supervises the results of classifier $C$ after passing through the BLSTM network, while the auxiliary loss $L_a$ directly supervises the results of classifier $C_a$ after the 3D CNN network to force the 3D network to extract features that are helpful for sign language recognition. See Table 6. for the role of auxiliary loss and classifier.

Table 6. Influence of different loss function proportion on recognition effect

<table>
<thead>
<tr>
<th>Loss Function</th>
<th>Dev WER(%)</th>
<th>Test WER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{CTC}$</td>
<td>26.6</td>
<td>27.5</td>
</tr>
<tr>
<td>$L_{CTC}+0.5L_a$</td>
<td>24.3</td>
<td>25.0</td>
</tr>
<tr>
<td>$L_{CTC}+L_a$</td>
<td>23.9</td>
<td>24.0</td>
</tr>
<tr>
<td>$L_{CTC}+2L_a$</td>
<td>24.5</td>
<td>24.7</td>
</tr>
<tr>
<td>$1.5L_{CTC}+L_a$</td>
<td>23.5</td>
<td>23.6</td>
</tr>
</tbody>
</table>

The addition of auxiliary loss function improves the recognition effect of the network. Experiments show that $1.5L_{CTC}+L_a$ has the best loss ratio recognition effect. When translating network feature vectors into text information, the weight of the network classifier is explored. Table 7. shows that the error rate can still be reduced to 25.8% when only auxiliary classifiers are used for decoding and translation, which proves the importance of introducing auxiliary classifiers and auxiliary loss functions in this paper and the powerful ability of 3D convolution network for spatio-temporal feature extraction. The decoding effect obtained by adding the weights of main classifier and auxiliary classifier 1:1 is the best.
Table 7. Influence of different classifier weights on recognition effect

<table>
<thead>
<tr>
<th>Classifier ratio</th>
<th>Dev WER (%)</th>
<th>Test WER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cₐ</td>
<td>25.4</td>
<td>25.8</td>
</tr>
<tr>
<td>C</td>
<td>24.8</td>
<td>24.9</td>
</tr>
<tr>
<td>C₊₀.₅Cₐ</td>
<td>24.3</td>
<td>24.5</td>
</tr>
<tr>
<td>0.₅C₊Cₐ</td>
<td>24.5</td>
<td>24.7</td>
</tr>
<tr>
<td>C₊Cₐ</td>
<td><strong>23.5</strong></td>
<td><strong>23.6</strong></td>
</tr>
</tbody>
</table>

4. CONCLUSION

The network algorithm of continuous sign language recognition based on 3D CNN and BLSTM proposed in this paper can recognize and translate the sign language information in the video stream with high accuracy. The network introduces an auxiliary classifier and an auxiliary loss function, which can effectively reduce the word error rate in the process of sign language recognition. Compared with the previous continuous sign language recognition work, the word error rate is lower. However, there are still some shortcomings in this method, which is also the direction that needs to be worked on in the future: (1) introducing multimodal mechanism to consider the change information and depth information of facial expressions. (2) Use a network with stronger time modeling ability, and consider the relationship between redundant features and useful features in feature information.

REFERENCES


Blind Image Super-Resolution Reconstruction Based on Dual Regression Network
Hongpeng Tian\textsuperscript{a}, Shengzhou Jiang\textsuperscript{b}

\textsuperscript{a}College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, CHINA;
\textsuperscript{b}College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, CHINA;

* Corresponding author: tianhp@xust.edu.cn

ABSTRACT

Existing deep learning-based Super Resolution (SR) reconstruction algorithms achieve remarkable performance on images with known degradation. Most of the degradation models exists problems in self-adaptations when facing with the deviation of the degradation model of the image of the real scene, and the effect is not good. Therefore, this paper proposes a blind image super-resolution reconstruction algorithm based on dual regression, which aims to solve the problem of poor performance of super-resolution networks in real scenes. Firstly, the closed-loop network is used to constrain the mapping space, and the optimal reconstruction function is found to improve the network reconstruction performance. Secondly, the attention mechanism is adopted into the residual block of feature extraction to expand the receptive field of the feature map, improve the reuse of features, and strengthen the reconstruction of high-frequency information. Finally, the frequency-domain blur kernel map estimates the downsampling kernel and reconstructs the low-resolution image, adaptively extracts the feature expression, enhances the ability to restore texture details, and reconstructs the real-world image better.

Keywords: blind image super resolution reconstruction; dual regression network; attention mechanism;

1. INTRODUCTION

Image super-resolution is a long-term research topic, which aims to convert the input Low Resolution (LR) images into High Resolution (HR) images. Due to the powerful feature representation and model fitting ability of deep Neural Network, the SR reconstruction method of Convolutional Neural Network (CNN) has achieved significant performance improvement compared with traditional methods. Recently, in real-world applications, image super-resolution technology has been widely used in security monitoring, medical images, satellite images and other fields. The popularity of smart edge devices, such as smartphones and VR glasses, has created a higher demand for super-resolution images. Given the limited resources of edge devices, the application of efficient SR on these devices is critical.

Although the SR method based on CNN has achieved good results, it is not effective in real scenes. Most SR algorithms assume that the fuzzy kernel of the input LR image is known and fixed in the process of overscore reconstruction, but this is difficult to apply to images with complex degradation types. In addition, it is a typical "ill-conditioned problem" that the super-resolution reconstructed low-resolution image is nonlinearly mapped to high-resolution image, that is, the solution space of the mapping function is too large, which easily leads to the problem of model adaptation. In view of the appeal problem, the main contributions of this paper is as follows:

- We design a blind Attention Dual Regression Network (CADRN) based on Dual Regression strategy, which improves the performance of Network reconstruction by nonlinear mapping and constraining function mapping space.

- We introduce coordinate attention into the residual block of feature extraction, and capture feature dependencies with accurate location information by modeling the channel relationship.

- We use frequency-domain fuzzy kernel to estimate the downsampling kernel and reconstruct the low-resolution image, and adaptively extract the feature expression to enhance the texture detail recovery ability. Experimental results show that compared with SRCNN, DBPN, EDSR and RDN, the proposed method has better reconstruction effect.
2. RELATED WORK

2.1 Blind super-resolution reconstruction

As a special category of image super-resolution, blind super-resolution is an active research topic in real world image restoration applications and has attracted more and more attention. Most existing methods assume a predefined degradation process from HR image to LR image (for example, bicubic downsampling), which is difficult to apply to real images with complex degradation types. To solve this problem, Wei et al.[2] captured real LR/HR images by building paired LR/HR cameras, but it took a lot of work to collect a large amount of well-aligned data, and the learning model was easily limited by the degradation kernels defined by the captured LR images. There is a growing interest in learning super-resolution models of unbiased data in unsupervised environments. Based on CycleGAN[3], Yuan et al. proposed a CinCGAN[4] to generate HR images without pairwise data. Lugmayr et al.[5] proposed to learn deep degradation maps through cyclic consistency loss functions, and then generate paired LR/HR for supervised training. Ji et al.[6] proposed a cyclic framework to extract different noise images from LR images using a variety of fuzzy kernels, and then applied traditional degradation models to synthesize different LR images. This accurate degradation modeling method achieves certain results, but for complex and severely degraded data, the accuracy of fuzzy kernel and noise estimation will be reduced, resulting in unreliable LR image synthesis. The above blind super-resolution degradation method does not utilize paired synthetic data, although these data can be easily obtained and used to enhance training.

2.2 Attention mechanism

By assigning weights to different features, the attention mechanism makes the neural network pay attention to "good" features and ignore "bad" features, so as to improve the performance of the network. At present, attention mechanism has been introduced into some hyperpartition networks, and good results have been achieved. According to RCAN[7] (super-resolution Using Very Deep Residual Channel Attention Networks), the structure is shown in Figure 1. Where H, W, C represent the high, wide and channel number of the features, and \( r \) represents the scaling factor; \( f \) represents the Sigmoid activation function; \( H_{GP} \) represents the average pooling; \( W_D \) represents the number of reduced channels; \( W_U \) represents the number of expanded channels; \( \otimes \) represents the multiplication of feature map elements. By learning the importance of different channels, the model obtains the weight value, adaptively adjusts the features of each channel, strengthens the high-frequency channel information, and enhances the representation ability of convolutional network.

In addition, Liu et al. proposed enhanced attention block ESA[8], as shown in Figure 2, which uses residual features to focus on key spatial information. In this mechanism, \( 1 \times 1 \) convolution is used to reduce the number of channels, and then \( 2 \times 2 \) step-convolution and \( 2 \times 2 \) maximum pooling layer are used to expand the receptive field for feature maps. Secondly, the channel dimension is raised and the activation layer is used to output the feature to extract the block. Finally, the high-resolution features are output to the end of the residual block using a step join. By integrating the salient information, the model obtains more representative features.
3. COORDINATE ATTENTION DUAL REGRESSION NETWORK

3.1 Reconstruction of the network

CADRN is built based on the dual regression network, and the network structure is shown in Figure 3, including the original regression network (black line in the figure) and the dual regression network (red line in the figure). The original regression network adopts U-NET network structure, which is composed of down-sampling and up-sampling modules. The down-sampling module adopts a convolution layer with step size of 2, and Relu activation function is connected with the convolution layer in series. This module uses the pixel-level modeling ability to extract more complex details. Then, a progressive upsampling network is constructed based on CA (Coordinate Attention) module, residual jump connection and subpixel convolution, which makes the network have more powerful feature expression and related feature learning ability, and extract more abundant feature vector related information. Finally, the regression network is used to reduce the function mapping space and find the optimal solution.

3.2 Coordinate attention module

The coordinate attention module decomposes the feature map into feature maps in different directions and channels to enhance the ability of feature expression in the network. It consists of two steps: coordinate information embedding and coordinate attention generation. The structure of coordinate attention module is shown in Figure 4.
Coordinate attention as a new efficient attention mechanism, the first step is to embed the bit information into the channel attention, so that the lightweight network can obtain more information, reduce the number of attention module parameters and avoid too much calculation overhead. In order to avoid the loss of location information caused by the two-dimensional global pooling, the two-dimensional global pooling is decomposed into two parallel one-dimensional feature codes, which are respectively aggregated along two spatial directions, and the spatial coordinate information is efficiently integrated into the attention map. In the process of coordinate information embedded in the $X$ direction input by size $(H,1)$ and $(1,W)$ the pooling of nuclear along the horizontal coordinate direction and the vertical coordinate direction to encode each channel, $H$, $W$, respectively, for the current focus mode figure the height and width of the corresponding input characteristics, value will change with mining under the network of the sample. The output of the $c$-th channel with height $h$ is shown in Equation 1:

$$Z_h = \frac{1}{W} \sum_{i=0}^{W} x_c(h,i)$$

Similarly, the output of the $c$-th channel with width $w$ is shown in Equation 2:

$$Z_w = \frac{1}{H} \sum_{j=0}^{H} x_c(j,w)$$

In the second step, in order to make better use of the coordinate information, the information with global receptive field and precise position is embedded. In the process of channel information encoding of feature map, the unidirectional coded feature map $Z_h$ and $Z_w$ are cascaded first, and the dimension of transformation channel is reduced by using a common $1 \times 1$ convolution of $F_1$ to obtain $f$, as shown in Equation (3). $Cat$ represents the cascade along the spatial dimension, and $\delta$ is the nonlinear activation function.

$$f = \delta(F_1(cat(Z_h,Z_w)))$$

Secondly, $f$ is partitioned into two independent tensors $f_h$ and $f_w$ along the spatial dimension. Then, two $1 \times 1$ convolution $F_h$ and $F_w$ are used to change the feature maps $f^h$ and $f^w$ to the same number of channels as the input $X$, and the results are shown in Equations (4) and (5).

$$g^h = \sigma(F_h(f^h))$$

$$g^w = \sigma(F_w(f^w))$$
\[ g^h = \sigma(F_h(f^h)) \]  

In order to reduce the complexity of the model, the number of channels of \( f \) is reduced by decreasing the ratio \( r \) appropriately. \( g^h \) and \( g^w \) are respectively obtained along the X and Y axes to obtain the attention weight characteristics. The final output of the coordinate attention module is shown in Equation (6).

\[ y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \]  

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Datasets

The CADRN network used the DIV2K and Flickr2K synthetic datasets to train the neural network, including 800 and 2650 training images, respectively. By randomly intercepts the 48 \( \times \) 48 size image of RGB input low-resolution image and the corresponding high-resolution image region as paired training data, and rotates the image by 90º, 180º and 270º respectively, corresponding to the flip, to obtain the variant of each image, so as to achieve the purpose of increasing the training data. In this paper, four common single image super-resolution test sets are used for testing and comparison at 4x and 8x scaling factors, including Set5, Set14, BSD100 and Urban100. In addition, the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are used to quantitatively evaluate the reconstructed images.

4.2 Experimental parameter setting

This article uses Adam optimizer for parameter optimization, set \( \beta_1 = 0.9, \beta_2 = 0.999 \), set minibatch to 32. The initial learning rate is 10^{-4} and then decreases by half for every 100 epochs.

4.3 Analysis of experimental results

The algorithm in this paper is compared with advanced superscore algorithms such as DRCN[9], SRDenseNet[10], SRRAM[11], EDSR[12] and RDN[13]. The superscore results of 4-fold and 8-fold are shown in Table 1. It can be seen that the proposed method is only slightly lower than the comparison model in part of the data, while the PSNR and SSIM indexes on almost all benchmark test sets are optimal. Compared with the algorithm with the second-best index, the proposed algorithm has smaller number of network parameters and less computational burden. Taking 4-fold superfraction reconstruction as an example, the parameter number of the proposed algorithm is 4.87m, which is 88%, 77% and 53% less than EDSR, RDN and DBPN, respectively. In this paper, the subjective visual effects of different algorithms are compared and tested. The images are from Set5, Set14, BSD100 and Urban100 test sets, and the results are shown in Figure 5.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Scale</th>
<th>Set5 PSNR/SSIM</th>
<th>Set14 PSNR/SSIM</th>
<th>BSD100 PSNR/SSIM</th>
<th>Urban100 PSNR/SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicubic</td>
<td>4</td>
<td>28.43/0.811</td>
<td>26.00/0.702</td>
<td>25.96/0.668</td>
<td>23.14/0.657</td>
</tr>
<tr>
<td>EDSR</td>
<td></td>
<td>32.46/0.896</td>
<td>28.80/0.788</td>
<td>27.71/0.742</td>
<td>26.62/0.803</td>
</tr>
<tr>
<td>RDN</td>
<td></td>
<td>32.47/0.899</td>
<td>28.81/0.787</td>
<td>27.72/0.742</td>
<td>26.61/0.800</td>
</tr>
<tr>
<td>CADRN</td>
<td></td>
<td>32.48/0.900</td>
<td>28.86/0.792</td>
<td>27.72/0.743</td>
<td>26.60/0.806</td>
</tr>
<tr>
<td>DRCN</td>
<td>8</td>
<td>26.58/0.755</td>
<td>24.30/0.615</td>
<td>24.49/0.580</td>
<td>21.89/0.587</td>
</tr>
<tr>
<td>SRDenseNet</td>
<td></td>
<td>25.99/0.704</td>
<td>24.23/0.581</td>
<td>24.45/0.530</td>
<td>21.67/0.562</td>
</tr>
<tr>
<td>EDSR</td>
<td></td>
<td>27.03/0.774</td>
<td>25.05/0.641</td>
<td>24.80/0.595</td>
<td>22.55/0.618</td>
</tr>
<tr>
<td>CADRN</td>
<td></td>
<td>27.20/0.777</td>
<td>25.06/0.645</td>
<td>24.90/0.602</td>
<td>22.62/0.624</td>
</tr>
</tbody>
</table>

As can be seen in the reconstruction figure, the Bicubic algorithm has poor effect, such as ambiguity, ringing effect and sawtooth. DBPN, EDSR and RDN algorithms have been improved to a certain extent, and ringing effect has been...
basically eliminated, but there are still some problems such as too smooth edge information and fuzzy details. The proposed algorithm CADRN has better reconstruction effect, significantly improved the subjective visual effect, richer texture details, and clearer edge details of the image.

![Figure 5. Visual comparison of different methods for (a) 4× image super-resolution.](image)

### 5. CONCLUSION

This paper designs a blind image super-resolution network based on dual regression strategy. In order to solve the problems of insufficient detail feature extraction and unclear texture region reconstruction, coordinate attention is injected into the residual block of feature extraction, and the accurate position information is used to capture the feature dependence relationship through modeling the channel relationship, so as to obtain richer feature information and improve the reconstruction ability of detail information. In order to solve the problem that the existing hyperpartition network is not good in the real world image, the frequency domain fuzzy kernel is used to estimate the downsampling kernel and reconstruct the low-resolution image. The feature expression is extracted adaptively, the texture detail recovery ability is enhanced, the details of the reconstructed image are clearer, and the overall robustness and comprehensive performance are better. The experimental results show that the proposed method is superior to the comparison method in both objective indexes and subjective vision.

### REFERENCES

Remote sensing image segmentation network based on attention mechanism feature fusion

Ji Gao, Huan Zhao, Wufei Liu and Zipeng Zhang*
School of Computer science, Hubei University of Technology, Wuhan, China
*Corresponding author’s e-mail: 782477656@qq.com

Abstract

For remote sensing image segmentation projects, it is usually necessary to segment the boundary of ground objects finely, but the fitting of the boundary is a difficult problem in deep learning image segmentation. At the same time, for remote sensing images taken from high altitude, usually many large-scale targets in natural scenes will become small target objects in remote sensing images. To solve this problem, in order to improve the recognition effect of remote sensing image scene targets with huge scale differences, this paper combines UNet and FPN. This paper uses the framework of pytorch1.7.1 to establish the UNet + FPN model. Through the experimental results and visual analysis of building data sets and road data sets, the model in this paper has obvious advantages.

Keywords - Remote sensing image segmentation; Deep learning; UNet; FPN

1. Introduction

In the current stage of development, many remote sensing image semantic segmentation algorithms are based on the improvement of the decoding part of U-Net. At the same time, Deeplab, ccnet, DANet and so on use to enhance the image context information or establish the relationship between pixels to improve the effect of image segmentation. At the same time, there are also patch attention module (PAM), attention embedding module (AEM), and linear attention mechanism (LAM) to solve practical problems, but the above structures ignore the importance of multi-level features. Multi-level features have different kinds of information of deep learning features. Remote sensing image segmentation is different from natural scene image segmentation. There are many natural scene categories, and there is often a strong interdependence between different object categories, which is different from natural scene. The target categories that need to be recognized in remote sensing images are often few, or sometimes only specific categories in the remote sensing image scene, such as buildings or roads, need to be recognized. Usually, these categories are reflected as small targets or targets with large scale differences in remote sensing images, which requires the algorithm to effectively use the feature maps with different kinds of information in the feature extraction stage. The above-mentioned frameworks and attention mechanisms do not improve the recognition effect of scene targets in remote sensing images with huge scale differences.

Based on the UNet model, this paper designs a new end-to-end semantic segmentation network structure. By using the UNet basic model to combine the FPN structure, and combine the detailed features such as shallow color and texture with the semantic features, the features in the convolution network structure can be reused, reducing the adverse impact of feature information loss caused by down sampling. At the same time, by combining with the shallow neural network, the application of low-resolution texture information is strengthened, which makes the network have better detail recovery information to a certain extent. In this paper, the AG module is designed for feature fusion. By using the AG module to filter and screen the fused feature map, the representation ability of the feature map is optimized.

2. UNet + FPN ALGORITHM

This section starts with the basic FPN structure, improves the FPN structure, adopts the block non local attention mechanism to optimize the characteristic diagram, and combines UNET with FPN. The main improvement strategies are as follows.

2.1. Characteristic pyramid network structure

Feature Pyramid Networks (FPN) is a feature extractor, which solves the problem of scale difference by predicting feature mapping on multiple scales. In recent years, FPN [4] has been widely used in image segmentation, inputting feature information to a certain extent. In this paper, the AG module is designed for feature fusion. By using the AG module to filter and screen the fused feature map, the representation ability of the feature map is optimized.
information into FPN. Using the features of multiple scales, FPN further fuses the deep and shallow feature information, so that the feature information is fully studied. As shown in Figure 1, a complete feature pyramid network structure includes two feature map branches, one of which is bottom-up. The network branch directly uses the feature map extracted by the backbone network. The backbone network extracts the feature map group from bottom to top through several stages of convolution neural network, and gradually reduces its size.

Pyramid characteristics are calculated by the following formula:

\[ P_i = \zeta(C_i) + \Gamma(P_{i+1}) \]  

(1)

Where, \( P_i \) represents pyramid feature, \( C_i \) represents backbone feature, \( \zeta \) represents horizontal connection, and \( \Gamma \) represents nearest neighbor upsampling with scale factor of 2. Horizontal connection is 1 × 1. Realized by convolution layer.

2.2. Fusion of UNet and FPN

UNet is a general segmentation framework in the field of medical image and remote sensing image segmentation. For remote sensing images, there are usually many small targets in remote sensing images, and the target size differences are large. Therefore, the segmentation network framework that needs to be designed can make full use of the information of different levels of image feature map, so as to recognize targets of different scales. The network effectively combines UNet and FPN, as shown in the FPN network part in Figure 2.

First, the network performs deep convolution on the input image, convolutes and reduces the dimension of the original image, and obtains the information layer1 of the first stage. Then, three times of convolution down sampling is calculated, and the information of layer2, layer3 and layer4 is obtained respectively. Finally, the feature maps of different levels output by FPN have different kinds of information of the original image, so it is convenient for the feature map to calculate the semantic segmentation and decoding part.

2.3. Block non local computing module

Non-local Neural Networks is somewhat similar to Non-local Means (non-local means denoising filtering). Ordinary filtering is 3 × Convolution kernel of 3, and then move on the whole picture, processing 3 × 3 partial information. The
Non-local Means operation combines a relatively large search range and is weighted. This paper uses the Non-local attention mechanism module proposed by Wang Xiaolong et al. in 2018 to capture long-distance dependencies, which is very important to grasp the global information in remote sensing images, so as to improve the segmentation accuracy. The calculation formula of Non-local module is as follows:

\[ y_i = \frac{1}{C(x)} \sum_{vj} f(x_i, x_j)g(x_j) \]  
\[ g(x_j) = W_g x_j \]  
\[ f(x_i, x_j) = e^{\phi(x_i)T \phi(x_j)} \]  
\[ C(x) = \sum_{vj} f(x_i, x_j) \]

Where \( x \) and \( y \) are input signals and output signals respectively, \( i \) and \( j \) represent the response of eigenvalues at the current position and global position respectively \((x_i, x_j)\) is a function of calculating the similarity relationship between any two points, and calculating the similarity between \( i \) and \( j \). \( g \) function is a mapping function, usually expressed as a linear transformation, and \( W_g \) represents the weight of the network. \( W_f \) refers to weighting the features corresponding to the \( j \)th pixel, \( T \) refers to the transpose matrix of the original matrix, and formula (4) is the core calculation formula of Non-local module \( f(x_i, x_j) \). The mathematical method of matrix multiplication is adopted, so that the global information and local information are fused to obtain the dependency of the whole region. \( C(x) \) is the factor normalized by softmax.

\[ \theta(x_i) = W_{\theta} x_i \]  
\[ \varphi(x_i) = W_{\varphi} x_i \]

\( W_{\theta} \) and \( W_{\varphi} \) represents the weight matrix to be learned, which can be realized by 1×1 convolution in space. The proposed non-local operations directly capture remote dependencies by calculating the interaction between any two locations, rather than being limited to adjacent points. It is equivalent to constructing a convolution kernel as large as the size of the feature map, so that more information can be maintained. At the same time, non-local can be combined with other network structures as a component.

2.4. Feature fusion module based on attention computing

Attention mechanism can also be divided into two aspects: space and channel. The Ag module proposed in this paper is a kind of soft attention focusing on spatial domain. It can be easily embedded into the FCN framework to improve the performance of the model without adding a large number of network computing parameters.

This paper combines the FPN structure by using the UNet basic model, and combines the detailed features such as shallow color and texture with semantic features, so that the features in the convolution network structure can be reused, reducing the adverse impact caused by the loss of feature information caused by down sampling. At the same time, by combining the shallow neural network, the application of low-resolution texture information is strengthened, which makes the network have better detail recovery information to a certain extent. At the same time, by using block processing to optimize the non-local module calculation method, the dependency relationship of the high-level feature map is constructed, and the network can learn the region of interest of the feature map more. For feature fusion, the AG module performs feature fusion. By using the AG module to filter and screen the fused feature map, the representation ability of the feature map is optimized.

3. Experimental process and analysis

3.1. Data set description

This paper uses the WHU building segmentation data set and the CHN-CUG road segmentation data set. Buildings and roads are closely related to people's livelihood in the remote sensing targets. Therefore, we set the experimental targets as
buildings and roads. The WHU building data set consists of more than 220000 independent buildings, with a total of 4736 training set images, 1036 verification set images and 2416 test set images, each of which is 512 in size × 512 size. CHN-CUG data set is published by China University of Geosciences. This data set is selected from different cities in different regions of China and is highly representative in domestic actual engineering projects. At the same time, the effect on the data set also shows that the algorithm in this paper can achieve significant results in practical applications.

3.2. Experimental platform in this experiment

Linux server system is selected, NVIDIA GeForce RTX 2080 graphics card is selected as computing resource, and CPU is 6×Xeon E5-2678 v3, 32G memory, 8G video memory. The development framework of deep learning algorithm is pytorch1.7.1.

3.3. Training process in this experiment

In order to enhance the diversity of training data, during the training process, the samples of the training set were augmented by geometric transformation. Data augmentation is the most commonly used skill in deep learning. Its main function is to make the data set as diverse as possible, so that the trained model has stronger generalization ability. Figure 3 is the schematic diagram of data augmentation operation.

![Figure 3. rotation and flip transformation](image)

3.4. Training process in this experiment

For remote sensing image feature classification algorithm, it usually needs to carry out visual analysis according to the semantic segmentation evaluation index and the results of the segmentation algorithm. Through visual analysis, the algorithm can be further optimized and the effect can be improved. As shown in Table 1, according to the accuracy calculation results of confusion matrix theory, the prediction results of some remote sensing image slices are selected for visual display, so as to compare the different results of different algorithms.

Table 1. Target classification accuracy of WHU building segmentation dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Val IoU</th>
<th>Val Recall</th>
<th>Val Precision</th>
<th>Test IoU</th>
<th>Test Recall</th>
<th>Test Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSPNet</td>
<td>0.8943</td>
<td>0.958</td>
<td>0.930</td>
<td>0.854</td>
<td>0.892</td>
<td>0.953</td>
</tr>
<tr>
<td>UNet</td>
<td>0.900</td>
<td>0.956</td>
<td>0.939</td>
<td>0.868</td>
<td>0.945</td>
<td>0.914</td>
</tr>
<tr>
<td>EncNet</td>
<td>0.905</td>
<td><strong>0.962</strong></td>
<td>0.938</td>
<td>0.871</td>
<td>0.917</td>
<td><strong>0.946</strong></td>
</tr>
<tr>
<td>Deeplabv3+</td>
<td>0.908</td>
<td>0.959</td>
<td>0.943</td>
<td>0.884</td>
<td>0.939</td>
<td>0.938</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>0.911</strong></td>
<td><strong>0.962</strong></td>
<td><strong>0.946</strong></td>
<td><strong>0.896</strong></td>
<td><strong>0.957</strong></td>
<td><strong>0.935</strong></td>
</tr>
</tbody>
</table>

For the road segmentation experiment, a detailed comparative experiment is carried out on the algorithm network proposed in this paper. The following table 2 shows the comparison results of the segmentation accuracy between the network proposed in this chapter and the classical network model in the CHN6-CUG road data set.

Table 2. CHN6-CUG road dataset segmentation accuracy results

<table>
<thead>
<tr>
<th>Model</th>
<th>RoadIoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSPNet</td>
<td>47.98</td>
<td>66.24</td>
</tr>
<tr>
<td>UNet</td>
<td>48.63</td>
<td>67.12</td>
</tr>
<tr>
<td>RefineNet</td>
<td>54.12</td>
<td>66.75</td>
</tr>
<tr>
<td>Deeplabv3+</td>
<td>55.96</td>
<td>68.78</td>
</tr>
<tr>
<td>DlinkNet</td>
<td>56.64</td>
<td>69.32</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>61.09</strong></td>
<td><strong>73.22</strong></td>
</tr>
</tbody>
</table>
As shown in table 2 above, the algorithm in this chapter also produced excellent results on the CHN6-CUG dataset. Overall, it is better than most classical road segmentation algorithms on chn6-cug. It can be seen from Figure 4 that the feature fusion algorithm based on attention mechanism proposed in this paper is more suitable for real tags, and can reflect more superior segmentation results in building targets. At the same time, this chapter visualizes the prediction results of chn6-cug dataset and compares the visualization results of different algorithms, as shown in Figure 5 below.

![Figure 4. Visual comparison results of feature fusion network in WHU dataset](image1)

![Figure 5. Visual comparison results of feature fusion network on CHN6-CUG dataset](image2)

In order to reflect the advantages of the algorithm proposed in this paper, a detailed experiment is designed. The accuracy results of the proposed algorithm and the classical semantic segmentation algorithm are compared through the semantic segmentation evaluation indicators F1 score, IOU and other indicators. At the same time, the visualization effects of the segmented image results obtained by this algorithm and different networks are also compared. Through the experimental results and visual analysis on building data sets and road data sets, the results confirm the superiority of the feature fusion algorithm based on attention mechanism in the task of remote sensing image target extraction.

### 4. Conclusion

This paper takes the WHU building segmentation data set and CHN-CUG road segmentation data set as the experimental objectives, and the effect on this data set also shows that the algorithm in this chapter can achieve significant results in practical applications. At the same time, for remote sensing images, this paper also describes the scheme of pre-processing remote sensing image data. Finally, according to different data sets and network structures, this paper makes a detailed comparison between the proposed algorithm and the classical image segmentation algorithm, including the comparison of quantitative indicators and visual analysis results. The comparison results show that the algorithm can get superior recognition results in remote sensing image segmentation projects.
References


Agricultural Text Named Entity Recognition Based on the BiLSTM-CRF Model

Yongqiang Qian\textsuperscript{1a}, Xiaojin Chen\textsuperscript{1b}, Yaojun Wang\textsuperscript{1c*}, Jingbo Zhao\textsuperscript{1d*}, Di Ouyang\textsuperscript{2c}, Shihao Dong\textsuperscript{1f}, Lan Huang\textsuperscript{1g}

\textsuperscript{1}College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
\textsuperscript{2}College of Land Science and Technology, China Agricultural University, Beijing 100193, China
\textsuperscript{a}qianyongqiang@cau.edu.cn, \textsuperscript{b}chenxiaojin@cau.edu.cn, \textsuperscript{c*}wangyaojun@cau.edu.cn, \textsuperscript{d*}zhao.jingbo@cau.edu.cn, \textsuperscript{e}ouyangdi@cau.edu.cn, \textsuperscript{f}dongshihao@cau.edu.cn, \textsuperscript{g}hlan@cau.edu.cn

Abstract

Named entity recognition (NER) is an essential part of many natural language processing applications and serves as the basis of natural language processing. NER tasks aim to identify the entity boundaries and types of specific meanings in texts, which include person names, place names, organizations, and dates. Due to the complexity of the composition of agricultural entities and limited quantity of agricultural corpora, often with low quality, agricultural text named entity recognition (ANER) is a more difficult and challenging problem compared to general NER tasks. Existing models often have poor performance in ANER tasks. To solve the existing problems of ANER, we constructed a named entity dataset of agricultural context through manual annotation using collected massive agricultural texts. The dataset contains 8873 entities composed of six entity categories: crop type, disease type, pest type, pesticide type, fertilizer type, and crop part type. We propose an agricultural text named entity recognition model based on BiLSTM-CRF. Experimental results show that the F1 score of the model on the test set achieved 91.20\%, and our model also achieved the highest precision of 93.23\% and the highest recall of 89.39\% compared to other existing models.

Keywords- Named Entity Recognition; Agricultural Text; Natural Language Processing; BiLSTM; CRF

1. Introduction

A fundamental task in information extraction is Named entity recognition (NER). It is an essential component of natural language processing (NLP) applications, which include information retrieval, automatic text summarization, question answering system, machine translation and knowledge graphs, etc. NER tasks aim to identify entity boundaries and types of specific meanings in texts, which typically include person names, place names, organizations, and dates. Compared with English named entity recognition (ENER), Chinese named entity recognition (CNER) has more challenges\textsuperscript{(1)}. Chinese entity names have no special indications, such as capitalized English entity name, in the text. The meaning of the Chinese entity name is highly dependent on context. Same words have different meanings in different contexts. In Chinese language, there is no precise word segmentation. This characteristic causes boundary ambiguity and further increases the difficulty of CNER task.

An ANER task is a named entity recognition task for agricultural text. It is one of the research hotspots in NLP. Unlike entity recognition and extraction of common entities in general cases, such tasks on agricultural text are difficult.

Based on the current research progress, this paper aims to solve the problems of ANER, including the small number and low quality of public datasets, the limited types of entities identified by existing research, and the poor performance of existing models for complex agricultural entity recognition in agricultural texts. The primary contributions of this paper can be summarized as follows:

1. We collected and constructed a large self-annotated corpus AgriNER for plant protection technology that includes six categories: crop type, disease type, pest type, pesticide type, fertilizer type, and crop part type. We labeled 8873 entities, including 7421 entities in the training set, 679 entities in the verification set, and 773 entities in the test set.
We propose a new agricultural text named entity recognition model based on the BiLSTM-CRF framework. BiLSTM is used to obtain contextual text features, and CRF is used to calculate sequence labels, which improves the overall precision and recall of agricultural text named entity recognition. Experimental results show that the F1 score of the BiLSTM-CRF agricultural text named entity recognition model is 91.20%, higher than the other three existing models.

The structure of this paper is organized as follows: Section 2 introduces the relevant research work of NER based on rules and the machine learning approach; Section 3 introduces the dataset and proposes an agricultural text named entity recognition model based on BiLSTM-CRF; Section 4 presents the experimental results and analysis; and Section 5 provides the summary and discusses future work.

2. Related Work

The machine learning approach to accomplish a NER task is to consider it as a sequence labeling problem. Manually labeled corpus is used for training and the recognition results are given in the form of serialized and marked characters of the text. Statistical machine learning models, including Hidden Markov Models (HMMs), Maximum Entropy (ME), Conditional Random Fields (CRFs) and Support Vector Machines (SVMs), have been successfully used for the serialization and annotation of named entities (NE) and achieved satisfactory results. Fang et al. (2011) proposed a method for named entity recognition in the agricultural field based on the Cascaded Conditional Random Field Model (C-CRF). This method uses words as the segmentation granularity at the bottom layer to identify simple entities. The experimental results showed that the F1 score of the model reached 92.69%. Li et al. (2017) proposed a CRF-based recognition model of the crop, pest, and pesticide named entities, which identifies agricultural named entities in its self-annotated corpus that includes crops, drugs, and diseases. The experimental results showed that the F1 scores of the model for recognizing named entities of crops, diseases, pests, and pesticides reached 97.72%, 87.63% and 98.05%, respectively. Zhang et al. (2018) proposed a CRF-based ANER model, based on agricultural texts, to classify agricultural named entities into four types: pests, crops, fertilizers, and pesticides. By adding a variety of different features to improve the recognition rate, the experimental results showed that the average accuracy rate of the model reached 96.7%, and the recall rate reached 88.0%, which demonstrated the feasibility of the method. Guo et al. (2020) proposed a named entity recognition model based on Chinese agricultural diseases and pests with multi-scale local context features and a self-attention mechanism. By adding a CNN layer for feature extraction to improve the BiLSTM-CRF model, the F1 scores reached 94.15%, 94.56% and 90.55% on three corpora: AgCNER, Resume, and MSRA, respectively. However, it is difficult for the method to identify entities, such as fertilizers, crop varieties, and weeds. Zhang et al. (2021) proposed a Chinese named entity recognition model based on character enhancement for Chinese named entity recognition in apple diseases and pests. Dictionaries and similar words were incorporated into the character-based BiLSTM-CRF model to enhance character representation. Experiments were carried out on the manually constructed Chinese apple diseases and pests corpus ApdCNER with 21 entity categories, and the F1 score reached 92.14%.

Named entity recognition models based on BiLSTM-CRF have been widely used, but named entity recognition of agricultural text based on BiLSTM-CRF still has the following challenges:

1. Agricultural-based public corpora are small in quantity and low in quality, making entity recognition more difficult.
2. The length of some entities in agricultural texts is quite long, and some agricultural entity structures contain a mixture of numbers, letters, and special characters, which increase the difficulty of entity recognition.
3. Existing ANER methods only recognize limited entity types and are not effective in complex agricultural entity recognition in agricultural texts.

3. Methods

In this section, we describe the AgriNER dataset that can be applied to agricultural text named entity recognition, built through manual annotation of text entities. We also describe the agricultural text entity recognition model based on BiLSTM-CRF. Deep learning models were constructed for agricultural text named entity recognition with the AgriNER dataset.

3.1. AgriNER dataset

The most common datasets in the existing literature and open-source databases are standard datasets about people, places, and organizations. Currently, there are no publicly available agricultural named entity recognition datasets that can be used...
for research. Therefore, we first constructed the agricultural named entity recognition dataset AgriNER to fulfill our needs. The AgriNER text corpus data came from the agricultural knowledge websites of Chinese government departments. We mainly collected unstructured texts on themes of plant protection technology. We then adopted the BMOES notation method to annotate text entities. As shown in Figure 1, the "B label" represents the beginning of an entity, the "M label" represents the interior of an entity, the "E label" represents the end of an entity, the "O label" represents a non-entity, and the "S label" represents an individual word as an entity. The AgriNER dataset contains 8,873 entities. The dataset is divided into training, validation, and testing sets. The training set contains 7,421 entities, the validation set contains 679 entities, and the testing set contains 773 entities. The types and quantities of each entity are shown in Table 1.

![Figure 1  Example of agricultural text annotation](image)

### 3.2. BiLSTM model

LSTM network is a variant of recurrent neural networks. A recurrent neural network is based on time series expansion, which is suited for feature extraction of time series data and enables the neural network to learn contextual feature information.

Since LSTM can only encode the information of forwarding time series, the weight of the word at the current moment in the LSTM network is greater than the weight above. The BiLSTM network concatenates the sequence obtained from the forward input of the sentence and the sequence obtained from the reverse input. The forward sequence traverses backward from the first word on the left side of the sentence, and the reverse sequence traverses forward from the last word. The two sequences are spliced together to obtain the hidden layer vector BiLSTM. Through forward and reverse feature learning, useful context features can be learned at each moment, which solves the problem that the weight of words after the current moment is greater than the weight of previous words in the LSTM network. The BiLSTM model structure is shown in Figure 2.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crop</td>
<td>1690</td>
<td>141</td>
<td>165</td>
</tr>
<tr>
<td>Disease</td>
<td>940</td>
<td>119</td>
<td>149</td>
</tr>
<tr>
<td>Pest</td>
<td>383</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>Crop Part</td>
<td>2282</td>
<td>130</td>
<td>141</td>
</tr>
<tr>
<td>Fertilizer</td>
<td>571</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Pesticide</td>
<td>1555</td>
<td>196</td>
<td>213</td>
</tr>
</tbody>
</table>
3.3. CRF model

CRF (Conditional Random Field) is a sequence modeling framework that has all the advantages of MEMM (Maximum Entropy Markov Model) but solves the label bias problem of MEMM models. A major difference between CRF and MEMM is that MEMM uses an exponential model of the current state to determine the conditional probability of the following state given the current state. In contrast, CRF has a separate exponential model for the joint probability of the entire sequence of markers for the observation sequence. Therefore, the weights of different features of CRF in different states can be balanced with each other.

CRF attempts to model the conditional probability of multiple variables given observations. Let \( x = \{x_1, x_2, \ldots, x_n\} \) be the observation sequence and \( y = \{y_1, y_2, \ldots, y_n\} \) be the corresponding marker sequence, then the goal of CRF is to build a conditional probability model \( P(y \mid x) \). If the random variable \( y \) form a Markov random field represented by an undirected graph \( G = (V, E) \), for any node \( v \in V \):

\[
P(y_v \mid x, y_w, w \neq v) = P(x_v \mid x, y_w, w \sim v)
\]

where \( P(y \mid x) \) is the conditional distribution, and \( y_v \) is the label variable corresponding to node \( v \), \( w \neq v \) means \( w \) is all nodes except \( v \), \( w \sim v \) means \( w \) is the adjacent node of node \( v \).

Theoretically, graph \( G \) can have any structure as long as it can represent the conditional independence relationship between the labeled variables. However, practical applications usually refer to a linear-chain conditional random field, especially when modeling labeled sequences and this discriminative model predicts the output sequence from the input sequence.

3.4 BiLSTM-CRF model

The BiLSTM model can predict the probabilities of the correspondence of text characters to different labels and uses the label with the highest probability as the label prediction value for a character. The BiLSTM model ignores the constraint relationship between labels. For example, the characters corresponding to the B-CRO label are predicted to be marked as the M-CRO label. To allow the model to learn the constraint relationship between labels, we add a CRF layer after BiLSTM to form a BiLSTM-CRF model.

The BiLSTM-CRF model is a structure that combines a bidirectional LSTM network and a CRF network, as shown in Figure 3. One LSTM processes the sequence from start to end in the forward time direction, and the other LSTM processes the sequence in reverse from start to end in the reverse time direction. Therefore, the BiLSTM-CRF model can learn future input features. Learning these future input features can enhance labeling accuracy.
To verify the recognition performance of the BiLSTM-CRF model on complex agricultural corpus, we also used the HMM, CRF, and BiLSTM models for comparative experiments in addition to the BiLSTM-CRF model. We compared the precision, recall and F1 score of the four models on AgriNER. As shown in Table 2, we calculated the average values of various entity recognition results on the test set to present results. As can be seen from Table 2, the accuracy rate of BiLSTM-CRF model recognition reached 93.23%, the recall rate reached 89.39%, and the F1 score reached 91.20%. BiLSTM-CRF outperformed HMM, CRF and BiLSTM models in accuracy, recall and the F1 score. Compared with the BiLSTM model, the BiLSTM-CRF model used the adjacent labels between entities to dynamically plan the optimal sequence annotation by adding a CRF layer, and the model accuracy was improved by 2.44%.

Figure 4 shows the F1 scores of the four models for six entities of crop type, disease type, pest type, pesticide type, fertilizer type and crop part type. The four models have higher F1 scores for crop and disease, and for fertilizer and pesticide F1 score is low. Compared with BiLSTM model, BiLSTM-CRF has improved F1 score on five types of entities except crop part.

5. Conclusions

The task of text named entity recognition in agriculture is challenging, mainly because agricultural text entities contain numerous categories and complex character compositions of entities. This comprehensive study showed that the fusion of traditional machine learning and deep learning models performed better than using these two models individually for agricultural text named entity recognition. Compared with the other three mainstream machine learning models, the BiLSTM-CRF model improved entity recognition accuracy and recall rate.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>0.8774</td>
<td>0.8680</td>
<td>0.8726</td>
</tr>
<tr>
<td>CRF</td>
<td>0.9289</td>
<td>0.8823</td>
<td>0.9043</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>0.9079</td>
<td>0.8486</td>
<td>0.8764</td>
</tr>
<tr>
<td>BiLSTM-CRF</td>
<td>0.9323</td>
<td>0.8939</td>
<td>0.9120</td>
</tr>
</tbody>
</table>
Acknowledgments

This work was supported by the National Natural Science Foundation of China (72171005) and the 2115 Talent Development Program of China Agricultural University.

References

Multi-Task and Feature Attention for Semisupervised Medical Images Segmentation

Congcong Li\textsuperscript{1a}, Zhaoyang Liu\textsuperscript{1b}, Jizhe Li\textsuperscript{1c}, Jinshuo Zhang\textsuperscript{1d}, Xiuyang Zhao\textsuperscript{1e}\textsuperscript{*}

\textsuperscript{1}The School of Information Science and Engineering, University of Jinan, Jinan, 250024, China
\textsuperscript{*}zhaoxy@ujn.edu.cn

Abstract
Semisupervised learning (SSL) algorithms have received extensive attention in medical images segmentation. However, most SSL methods ignore the significance of boundary regions, and they cannot fully extract the significant image features, resulting in unsatisfactory boundaries and nonsmooth objects. In this study, we propose an SSL network, named Multi-Task and Feature Attention network (MTFA-Net). Our network can produce segmentation map and signed distance map (SDM) to fully mine the geometric information of object boundaries. To extract much more meaningful features, we devise a feature attention module and embed it into V-Net to discover the significance of each neuron, strengthen the important features, and suppress secondary features. Our experiments demonstrate that the method can achieve more accurate segmentation results than current methods on 3D left atrial image database.

Keywords-semisupervised segmentation; signed distance map; feature attention

1. Introduction
Medical images segmentation is a significant direction of medical images analysis and a significant part of computer-aided diagnosis and treatment. The field's key task is to obtain objects of interest in medical images. Recently, the segmentation of medical images has achieved satisfactory results. However, these methods require a large number of annotations. For 3D medical images, sufficient annotated data for use as manual annotation are difficult to obtain because they are costly and the task is time consuming. Aiming to reduce the labeling cost, semisupervised learning (SSL) is used for medical images segmentation.

In this article, we propose an SSL method, named Multi-Task and Feature Attention Network (MTFA-Net), for 3D left atrium (LA) segmentation. We use limited annotated data to achieve high-precision segmentation of 3D medical images, assist doctors in designing surgical plans for patients, and popularize computer-aided diagnosis. Our network is a dual-task model that outputs segmentation map and Signed Distance Map (SDM) [1]. The network can mine different features of objects from different perspectives to make full use of the intrinsic geometric information. Following Ref. [2], we encourage networks to learn from each other and explore useful knowledge during training. As many redundant features are expected to be generated in the feature extraction stage, we design a feature attention module embedded into V-Net [3]. This module can strengthen the important features and suppress the secondary features, enabling the neural network to extract more meaningful features.

2. Related Work
SSL can be trained with both annotated and unannotated data. Medical images segmentation has made tremendous progress. For instance, Zhang et al. [4] designed a adversarial network that enables consistent segmentation of all data. Li et al. [5] designed the SASSNet, and introduced an adversarial loss to encourage the segmentation of unannotated data, allowing them to be similar to annotated data. Luo et al. [1] used the SDM to design a Dual-Task Consistency framework for guiding the consistency of annotated data and unannotated data after performing different tasks. Although the abovementioned methods have achieved excellent performance, the importance of target geometric features is ignored, and generate many redundant features in the feature extraction stage. The above constraints have led to unsatisfactory boundaries, and nonsmooth objects.
3. Method

Fig.1(a) illustrates our MTFA-Net for 3D LA segmentation. First, we embed a head into the original V-Net to form a Dual-Task framework, which outputs the segmentation map and SDM. Second, we utilize the mean-teacher model to achieve multi-task output, and use this to compute multi-task consistency. Inspired by the Simam in Ref. [6], we propose a feature attention module (Fig.1(b)) and embed it into V-Net to optimize feature extraction.

![MTFA-Net Diagram](https://example.com/mtfa-net-diagram.png)

**Fig.1** Our MTFA-Net. Here, (a) is our main framework, which takes 3D data as input and outputs segmentation maps and SDM. The loss function for this model includes the supervised losses $L_{seg}$ and $L_{sdm}$ on the annotated data and the consistency loss $L_{unsup}$ for all data. (b) is the encoding part of the V-Net, in which the feature attention module is embedded.

3.1. Segmentation Network

$N$ is the number of annotated data, and $M$ is the number of unannotated data. $X \in \mathbb{R}^{H \times W \times D}$ is the input data, and $Y \in \{0,1\}^{H \times W \times D}$ is the ground-truth. $Y_{s,seg} = f_S(X, \theta_{seg}), Y_{s,sdm} = f_S(X, \theta_{sdm}), Y_{T,seg} = f_T(X, \theta_{seg}), Y_{T,sdm} = f_T(X, \theta_{sdm})$.

3.2. Feature Attention

Ref. [6] demonstrates that information neurons often have distinct values from surrounding neurons and proposes Simam. Inspired by this, we transform it into a 3-dimensional operation and propose feature attention. This method can help the neural network extract much more meaningful features to discover the importance of each neuron as follows:

$$E = \frac{(F - \mu)^2}{\alpha(\varphi^2 + \beta)} + \gamma,$$

where

$$\mu = \frac{1}{HWD} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{D} F_{ijk}, \quad \varphi^2 = \frac{1}{HWD - 1} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{D} (F - u)^2_{ijk},$$

$F \in \mathbb{R}^{H \times W \times D}$ is the input feature in a single channel. $\alpha, \beta,$ and $\gamma$ are all hyperparameters. $F_{ijk}$ is the value at the position of the feature vector $(i,j,k)$. Eq. (2) utilizes the mean squared error (MSE) to calculate the mean dispersion of all neurons in a single channel. An active neuron corresponds to a larger $E$, and a larger weight is assigned to this neuron. Moreover, an active neuron suppresses the activities of surrounding neurons.

After calculating Eq. (1), the input feature $F$ is given a new weight.

$$F = \alpha(E) \odot F,$$

where $\odot$ represents the Hadamard product. All computations in our module take the form of element-wise operations. According to Eq. (3), the shape of the input feature $F$ is the same as the original input after processing. As shown in Fig.1(b), we add this implementation after each convolutional layer in the V-Net encoding stage. In adopting this method, we can achieve the goal of enhancing the important features and suppressing the redundant features.
3.3. Loss Function

For annotated data, the student network is mainly guided by a supervised loss to learn reliable representations for segmentation tasks. Eq. (4) is the supervised loss:

\[ L_{sup} = \sum_{i=1}^{N} \left( L_{seg}(Y_{S,seg}^{i}, Y^{i}) + L_{sdm}(Y_{S,sdm}^{i}, G(Y^{i})) \right), \]  

where \( Y_{S,seg}^{i} \) and \( Y_{S,sdm}^{i} \) represent the i-th segmentation map and SDM from the student network, respectively. \( Y^{i} \) is the i-th ground-truth. We adopt Dice loss and MSE loss to calculate \( L_{seg} \) and \( L_{sdm} \), respectively.

For all data (both unannotated and annotated data), the consistency loss is as follows:

\[ L_{con1} = \sum_{i=1}^{N+M} MSE(Y_{T,seg}^{i}, Y_{T,seg}^{i}), \quad L_{con2} = \sum_{i=1}^{N+M} MSE(Y_{T,sdm}^{i}, Y_{T,sdm}^{i}), \]  
\[ L_{con3} = \sum_{i=1}^{N+M} MSE(Y_{T,seg}^{i}, G^{-1}(Y_{T,sdm}^{i})), L_{con4} = \sum_{i=1}^{N+M} MSE(Y_{S,seg}^{i}, G^{-1}(Y_{T,sdm}^{i})), \]  

where \( G^{-1}(\cdot) \) is the transformation of SDM to segmentation maps as described in Ref. [1]. We use the MSE loss to compute all consistency losses, and finally, the unsupervised loss is defined as:

\[ L_{unsup} = L_{con1} + L_{con2} + L_{con3} + L_{con4}, \]  

the total loss is shown in Eq. (8):

\[ L_{total} = L_{sup} + \lambda \cdot L_{unsup}, \]  

where \( \lambda \) is a weight coefficient that assigns the weight between supervised loss and unsupervised loss.

4. Experiments and Results

4.1. Dataset

We evaluated our method on the LA dataset\(^1\). The dataset contained 100 3D imaging scans and corresponding masks. We select 80 of them for training and the remaining 20 for testing.

4.2. Implementation Details and Metrics

The segmentation network was implemented PyTorch, using an RTX 2080Ti GPU. We adopted the SGD optimizer, with the iteration set as 6000 times. The batch size was 4 (2 annotated images and 2 unannotated images). We randomly cropped 112 × 112 × 80 sub-volumes as the input. Following Ref. [1], we set \( k = -1500 \). Following Ref. [6], we set \( \alpha = 4, \beta = 0.0001 \), and \( \gamma = 0.5 \).

4.3. Comparison with Other Semisupervised Methods

We evaluated our method under two different settings and compared the network with eight semisupervised segmentation methods, including DAN [4], CPS [7], MT [8], Entropy Mini [9], UA-MT [10], ICT [11], SASSNet [5], and DTC [1]. Table 1 shows the results, in which we first present the result of V-Net under the fully-supervised settings, followed by the two individual settings.

The first setting involved 20% scans for the annotated data, and 80% scans for the unannotated data. Our MTFA-Net produced excellent performance, and our method obtained comparable results, e.g., 90.05% vs. 91.41% for Dice. Among the semisupervised methods, DTC achieved the top result among all of the previous methods. However, our method was better in all metrics compared with DTC. In particular, Dice was improved by 0.63%, Jaccard was improved by 1.03%, ASD was improved by 0.33, and 95HD was improved by 0.68. We also implemented a more challenging setting by training with 10% annotated data. Experimental results show that our method outperforms other methods on all metrics.

\(^1\) https://atriaseg2018.cardiacatlas.org/.
Table.1 Comparison between our method and other methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scans used</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Labeled Unlabeled</td>
</tr>
<tr>
<td>V-Net [3]</td>
<td>80 0</td>
<td>91.14</td>
</tr>
<tr>
<td>V-Net [3]</td>
<td>16 0</td>
<td>86.03</td>
</tr>
<tr>
<td>DAN [4]</td>
<td>16 64</td>
<td>87.52</td>
</tr>
<tr>
<td>CPS [7]</td>
<td>16 64</td>
<td>87.87</td>
</tr>
<tr>
<td>MT [8]</td>
<td>16 64</td>
<td>88.42</td>
</tr>
<tr>
<td>UA-MT [10]</td>
<td>16 64</td>
<td>88.88</td>
</tr>
<tr>
<td>ICT [11]</td>
<td>16 64</td>
<td>89.02</td>
</tr>
<tr>
<td>SASSNet [5]</td>
<td>16 64</td>
<td>89.27</td>
</tr>
<tr>
<td>DTC [1]</td>
<td>16 64</td>
<td>89.42</td>
</tr>
<tr>
<td>MTFA-Net (Ours)</td>
<td>16 64</td>
<td>90.05</td>
</tr>
<tr>
<td>V-Net [3]</td>
<td>8 0</td>
<td>79.99</td>
</tr>
<tr>
<td>DAN [4]</td>
<td>8 72</td>
<td>75.11</td>
</tr>
<tr>
<td>CPS [7]</td>
<td>8 72</td>
<td>84.09</td>
</tr>
<tr>
<td>MT [8]</td>
<td>8 72</td>
<td>84.24</td>
</tr>
<tr>
<td>Entropy Mini [9]</td>
<td>8 72</td>
<td>85.90</td>
</tr>
<tr>
<td>UA-MT [10]</td>
<td>8 72</td>
<td>84.25</td>
</tr>
<tr>
<td>ICT [11]</td>
<td>8 72</td>
<td>85.39</td>
</tr>
<tr>
<td>SASSNet [5]</td>
<td>8 72</td>
<td>86.81</td>
</tr>
<tr>
<td>DTC [1]</td>
<td>8 72</td>
<td>87.49</td>
</tr>
<tr>
<td>MTFA-Net (Ours)</td>
<td>8 72</td>
<td><strong>87.83</strong></td>
</tr>
</tbody>
</table>

Fig.2. shows a visual comparison of our result evaluation, which includes UA-MT [10], SASSNet [5], and DTC [1], with 20% annotated training data. The corresponding ground-truth is shown on the right part. Our MTFA-Net has achieved more accurate results, as shown by the yellow circles in the figure, in both 2D and 3D plots compared with the other SOTA methods.

4.4. Ablation Study

We presented the ablation of our MTFA-Net with 20% annotated data. To investigate the individual impacts of different tasks, we conducted a number of corresponding experimental studies to validate the effectiveness of our framework. As shown in Table.2, in the first row we only use the annotated data to calculate $L_{seg}$, and the performance in Dice is 86.03%. Then in the second line, we added $L_{sdm}$ to calculate the loss on the annotated data together on the basis of $L_{seg}$, and the performance is 88.25%, which is an improvement of 2.22%. Subsequently, we successively added a series of
consistency losses, and the performance finally increased to 89.69%. Finally, we added the feature attention module, and the approach achieved the highest performance.

<table>
<thead>
<tr>
<th>#</th>
<th>seg</th>
<th>sdm</th>
<th>con1</th>
<th>con2</th>
<th>con3</th>
<th>con4</th>
<th>FA</th>
<th>Dice(%)↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86.03</td>
</tr>
<tr>
<td>2</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88.25</td>
</tr>
<tr>
<td>3</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88.29</td>
</tr>
<tr>
<td>4</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>89.33</td>
</tr>
<tr>
<td>5</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>89.14</td>
</tr>
<tr>
<td>6</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>89.69</td>
</tr>
<tr>
<td>7</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>90.05</td>
</tr>
</tbody>
</table>

5. Conclusion

In this study, we proposed a semisupervised medical images segmentation network called MTFA-Net. With the MTFA-Net, the SDM can interact with segmentation map between tasks and subsequently mine potential boundary information. The two networks optimize each other and jointly explore useful information. The method can also enhance the generalization ability of the model by extracting much more meaningful features. Extensive experiments demonstrate the effectiveness of our method, as it outperforms all state-of-the-art methods.

We believe that our semisupervised training framework provides a new segmentation method for unannotated 3D medical data and will contribute to medical diagnosis in the future.

Acknowledgments

This work is supported by the Research on key technology research and development of medical functional model platform for bone repair under Grant (2018GXRC023), Research on Image Example Retrieval Method Based on Geometric Features to Enhance Deep Learning Features under Grant (ZR2019MF013)

References


The Method of Grounding Fault Location of Distribution Line Based on Transient Signal

Lanjiang Wu1*, Liyuan He1, Chuanwei Zhou1
1 Dalian University of Science and Technology, Dalian, China
*E-mail:1850508590@qq.com

Abstract

With the rapid development of distribution lines, which is also accompanied by many problems, the most common is the problem of the ground fault. When a ground fault occurs, it will greatly affect the normal operation of the distribution line. At present, the existing grounding fault location methods for distribution lines have a large amount of calculation, which leads to a large relative error of the location results in practical applications and can not meet the actual needs. Based on this, this paper proposes a method of grounding fault location for distribution lines based on transient signals. The Kaernbauer transform matrix is used to decouple the three-phase lines of distribution lines, and the real wavelet transform is used to extract the amplitude information of transient signals. The fault location is estimated by using the extracted amplitude information. The experiment proves that the relative error of the design method for grounding fault location is less than that of the traditional method, which provides strong theoretical support for grounding fault location of distribution lines.

Keywords: Transient signal, Distribution line, Grounding fault, Location

1. Introduction

When the single-phase grounding fault occurs in the distribution network while the phase voltage exceeds the peak value, the transient zero-sequence current is mainly composed of high-frequency capacitance components. The transient characteristics are very obvious. When a single-phase grounding fault occurs in the distribution network at the phase voltage zero-crossing point, the transient zero-sequence current is mainly an inductance component. The transient characteristics are not obvious, but the attenuation DC component of the fault feeder caused by the arc suppression coil inductance is larger. The attenuation DC component of the sound feeder is smaller, and only flows through the faulted feeder and does not flow through the sound feeder. When the high resistance grounding fault occurs, the transient high-frequency component is small and the attenuation rate is fast. At present, many scholars have proposed a variety of transient fault line selection methods, mainly using the transient zero-sequence current mutation, energy, amplitude and polarity, making the effect of line selection improved to varying degrees. However, after the occurrence of grounding fault, the function of arc suppression coil makes the zero-sequence current weak, especially when intermittent arc grounding occurs. The burning and extinguishing of arc is uncertain, which makes the detection method based on a single one no longer reliable.

The main reason for the formation of ground fault is that the structure of a single line is chaotic and the conductor is complex, which is easy to connect with the ground accidentally, resulting in ground fault. According to relevant statistical data, in 2018, the number of grounding faults of domestic distribution lines accounted for 86.16% of the total number of faults, which is an increase of 2.26% compared with 2017, and the average daily number of grounding faults of a single distribution line was 5 [2]. With the continuous development of the power industry, the specifications, capacity and power distribution of the distribution network are gradually improved. In order to meet the power supply needs of users, it is necessary to continuously increase the number of branch lines and electrical equipment, which to a certain extent increases the probability and probability of ground fault. At present, the grounding fault has become one of the most important problems in the stable operation of the distribution system.

At present, the difficulties in determining the location of the line fault are as follows: ① The grounding resistance is relatively large. In this case, the fault signal of the 10 kV distribution line is relatively weak, which makes it difficult to locate the fault of the 10 kV distribution line under the strong noise interference. Workers have used many ways to locate the fault of the 10kV distribution line, but they can not succeed. ② The line has many branches. The branch point of the
line has an important impact on the transient signal and has a certain degree of weakening effect on the transient signal. In this case, the transient signal received by the signal receiver is very weak, and the location value is greatly reduced, so it can not be used as the basis for fault location.

In order to quickly determine the fault location and reduce the losses caused by line faults, the power sector has higher and higher requirements for the automation of fault location in distribution network, and the intelligent location method using fault transient signals has become one of the emerging technologies of the distribution automation in power system [3]. At present, the research on transient signals mainly includes: Sun Bo et al. [4] proposed a fault location method based on transient line voltage and transient zero-mode current, which uses the product of Hilbert transform of transient line voltage and transient zero-mode current to calculate the fault direction parameters, and determines the fault point according to the amplitude and phase of the magnetic field. Because magnetoresistive sensors are used to measure the magnetic field under the tower, once the magnetic field around the sensors is disturbed, bad data will be generated, which will greatly reduce the reliability of the algorithm. Weng Lantian et al. [6] mainly used the linked list method to locate the fault and improve the fault-tolerant capability. However, if the fault occurs at the end node, underreport false positives may occur.

To ensure the safety and reliability of distribution lines, it is necessary to take effective means to accurately locate the location of grounding faults, timely repair and deal with grounding faults, and avoid causing large-scale blackouts. As that research on the position of the grounding fault of the distribution line in China starts relatively late, the existing technology and theory are not mature enough, the existing position method needs to be based on massive actual operation data of the distribution line. The calculation amount is large, so that a large relative error exists in the positioning result in practical application. The traditional methods have been unable to meet the actual needs. Therefore, this paper proposes the research on the grounding fault location method of distribution lines based on transient signals.

2. Distribution line grounding fault locating method

2.1 Distribution line three-phase decoupling

Under normal circumstances, the distribution lines are not laid in parallel. Some lines are intersected, so there will be an alternating magnetic field between the distribution lines, which will have a certain impact on the distribution line, resulting in the skin effect, that is, when the alternating current is transmitted forward through the line, it will concentrate on the surface of the distribution line. At this time, the impedance parameters of the distribution line will be affected by the alternating frequency, and the transient signal transmission characteristics of the line will change at different frequencies. The frequency domain mathematical expression of the distribution line is:

\[
\begin{align*}
\frac{\partial U}{\partial x} &= ZI \\
\frac{\partial I}{\partial x} &= YU
\end{align*}
\]  

Where \( U \) is the voltage frequency domain of the distribution line; \( x \) is the alternating frequency; \( Z \) is the impedance per unit length of the distribution line; \( I \) is the current frequency domain of the distribution lines; and \( Y \) is the admittance per unit length of distribution lines. The above is a single-phase distribution line, but in practice, most of the distribution lines are three-phase. The lines have interphase coupling, and the transient signal propagation on the phase line of the distribution line will be affected by the coupling, so the two equations in formula (1) are not independent equations. In order to accurately represent the transient process of the distribution line, Kaernbauer transformation matrix is used to decouple the three-phase line of distribution line, and three independent mode components are obtained, which are expressed by the formula:
The transient signal is extracted by using the time wavelet transform method. The smaller the inductance of the circulation circuit is, the real wavelet transform scale is:

\[
\begin{align*}
-\frac{\partial U}{\partial x} &= T Z W I \\
-\frac{\partial I}{\partial x} &= W Y T U
\end{align*}
\]  

(2)

Where, \( T \) is the Kaernbauer transformation matrix used for the mode transformation of the distribution line voltage, and \( W \) is the transformation matrix of the distribution line current column vector. The module components of the voltage and current of the distribution line are transformed by using the above formula to eliminate the coupling between the phase lines of the distribution line.

### 2.2 Transient signal extraction

On this basis, the transient signal of distribution line is extracted. Transient signals are mainly charging transient signals and discharging transient signals, which have obvious attenuation characteristics and are mainly affected by the inductance of the circulation circuit of the distribution line. The smaller the inductance of the circulation circuit is, the faster the transient signal attenuates and the higher the oscillation frequency is. Therefore, the transient signal is mainly manifested as oscillation attenuation. The transient signal is mathematically expressed as:

\[
\alpha = E \sin (2f + \theta) \quad (3)
\]

Where, \( \alpha \) represents the transient signal of the single distribution line; \( E \) represents the amplitude of the transient signal of the single distribution line; \( f \) represents the frequency of the transient signal; and \( \theta \) represents the initial phase angle of the distribution line. The transient signal is gradually attenuated in the process of propagation on the distribution line, and the attenuation coefficient is:

\[
\mu = \frac{K}{2L} \quad (4)
\]

Where \( \mu \) is the attenuation coefficient of the transient signal propagating on the distribution line; \( K \) is the equivalent resistance of the distribution line; and \( L \) is the equivalent inductance of the distribution line. According to the above transient signal performance characteristics, the transient signal is extracted by using the time wavelet transform method, assuming that the original data of the distribution line is \( Q \), the amplitude information of the transient signal is extracted from the original data \( Q \), and the real wavelet transform scale is:

\[
s = e_a e_s / e_v \quad (5)
\]

In the formula, \( s \) represents the real wavelet transform scale of original data; \( e_a \) represents the center frequency of mother wavelet; \( e_s \) represents the sampling frequency of distribution line operation signal; and \( e_v \) represents the transmission frequency of distribution line transient signal. The original signal is transformed according to the transformation scale calculated by the above formula, and the transformation formula is:

\[
\varphi = \text{angle} \rho(t) \psi(t) \phi(t) / s \quad (6)
\]

In the formula, \( \varphi \) represents the amplitude of the extracted transient signal of the distribution line; \( \rho(t) \) represents the phase difference between the charging and discharging frequency voltage and current of the distribution line at time \( t \); \( \psi(t) \) represents the charging and discharging voltage of distribution line at time \( t \); and \( \phi(t) \) represents the charging current and discharging current of distribution line. The transient signal of distribution line is extracted by real wavelet transform.

### 2.3 Earth fault location

When a grounding fault occurs, the amplitude of the transient signal propagating on the distribution line increases to 2-3 times that of the normal signal. The transient signal waveform at the fault point will change abruptly. According to the extracted transient signal amplitude information, whether a grounding fault occurs in the power distribution line is...
determined, and a transient signal amplitude threshold is set here. If the transient signal amplitude extracted by formula (6) exceeds the threshold, it indicates that a grounding fault occurs in the power distribution line. In case of a fault, the transient signal propagated on the distribution line will pass through the fault point and form a circulation loop with the earth. Without considering the line load current and system imbalance, the high-frequency component of the transient signal is used to locate the grounding fault. The following Figure 1 is a schematic diagram of the grounding fault location based on the transient signal.

![Distribution line](image)

Figure 1 Schematic diagram of ground fault location based on transient signal

As shown in Figure 1, it is assumed that a ground fault occurs at the fault point A in the 1, 2 and 3 phase lines of the distribution line. N is the measuring point at the outgoing line side of the distribution line. The fault resistance of the power distribution line is P. At this time, the current under the transient signal propagation frequency is M. The impedance is B, and each sequence component of the power distribution line can be expressed as:

\[
\begin{align*}
I_1 &= \left(\frac{U_{M1} - U_{F1}}{Z_{L1}}\right) \\
I_2 &= \left(\frac{U_{M2} - U_{F2}}{Z_{L2}}\right) \\
I_3 &= \left(\frac{U_{M3} - U_{F3}}{Z_{L3}}\right)
\end{align*}
\]  (7)

In the formula, \(I_1\), \(I_2\) and \(I_3\) respectively represent the positive-sequence, negative-sequence and zero-sequence current of the line under the transient signal frequency; \(U_{M1}\), \(U_{M2}\) and \(U_{M3}\) respectively represent the positive-sequence, negative-sequence and zero-sequence voltage of the line under the transient signal frequency at the fault point; \(U_{F1}\), \(U_{F2}\) and \(U_{F3}\) respectively represent the positive, negative and zero-sequence voltage of the line at the measuring point; \(Z_{L1}\), \(Z_{L2}\), \(Z_{L3}\) respectively represent the positive sequence, negative sequence and zero sequence impedances of the faulty line at the transient signal frequency. When an interpretation fault occurs, the positive-sequence, negative-sequence and zero-sequence currents of the line at the fault point are zero, and the positive-sequence, negative-sequence and zero-sequence currents of the line at the non-fault point rise to 1/2 times of the normal current. The extracted transient signals are sorted in time domain according to the characteristics. The transient signals are propagated to the fault point after a period of time by taking the sudden change time of the waveform of the transient signals as the starting and ending time. The positive sequence current, the negative sequence current and the zero sequence current of the line are zero under the transient signal frequency. The time is used as the end time to record the propagation time of the transient signal of the fault line from the measuring point to the fault point. If the propagation speed of transient signal is known, the distance from the fault point to the measurement point can be calculated, and the
calculation formula is:
\[ l = (t_1 - t_2) \cdot v \quad (8) \]

In the formula, \( l \) represents the distance from the grounding fault point to the measuring point; \( t_1 \) represents the time domain corresponding to zero positive sequence, negative sequence and zero sequence current of the line under the transient signal frequency; \( t_2 \) represents the time domain corresponding to the sudden change of transient signal waveform; \( v \) represents the transient signal propagation speed. And calculate that distance from the grounding fault point to the measure point by using the formula so as to position the grounding fault of the line.

3. Experimental Demonstration

To verify the applicability and reliability of the proposed location method for grounding faults in distribution lines, a distribution line is selected as the test object. The length of the distribution line is 12643. The design method and the traditional method are used to analyze the data and locate the grounding fault position. Through three-phase decoupling of distribution line and transient signal extraction, the transient signal waveform of distribution line is extracted as follows.

![Transient signal waveform of distribution line](image)

The line grounding fault is located according to Figure 2. In this experiment, the principal component value of the natural frequency of the fault traveling wave is taken as the variable, and the principal component value of the natural frequency of the fault traveling wave is selected between 1200Hz and 2400Hz. The grounding fault location results are compared with the actual fault location to calculate the absolute error of the two methods, and the data are recorded with a spreadsheet. The specific data is shown in the following table.

<table>
<thead>
<tr>
<th>Principal component value of natural frequency of fault traveling wave/Hz</th>
<th>Maximum error limit/m</th>
<th>Designed method/m</th>
<th>Traditional method/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200</td>
<td>1.55</td>
<td>0.05</td>
<td>7.16</td>
</tr>
<tr>
<td>1400</td>
<td>1.55</td>
<td>0.01</td>
<td>5.35</td>
</tr>
<tr>
<td>1600</td>
<td>1.55</td>
<td>0.01</td>
<td>8.14</td>
</tr>
<tr>
<td>1800</td>
<td>1.55</td>
<td>0.02</td>
<td>9.14</td>
</tr>
<tr>
<td>2000</td>
<td>1.55</td>
<td>0.01</td>
<td>8.42</td>
</tr>
<tr>
<td>2200</td>
<td>1.55</td>
<td>0.01</td>
<td>6.23</td>
</tr>
<tr>
<td>2400</td>
<td>1.55</td>
<td>0.03</td>
<td>8.46</td>
</tr>
</tbody>
</table>

Through the analysis of the data in the above table, the following conclusions can be drawn:

By using the design method to locate the grounding fault of the distribution line, the location error can be effectively controlled within 0.1m, and less than the specified maximum error limit, indicating that the design method has a high accuracy of location results and can accurately locate the grounding fault location of the distribution line. Using the traditional method to locate the grounding fault of the distribution line, the relative error is much higher than that of the design method, and it also exceeds the prescribed range. This is because the transient signal theory is used to analyze and locate the fault characteristics. The fault location is accurately located according to the traveling wave velocity by
calculating the time domain of the fault transient traveling wave. Therefore, the experimental results show that the design method is superior to the traditional method in terms of location accuracy, and is more suitable for distribution line grounding fault location than the traditional method.

4. Conclusion

In this study, combined with the demand of distribution line grounding fault location and the shortcomings of traditional methods, a new idea of fault location is proposed, which effectively improves the accuracy of grounding fault location. This study has a good practical significance to solve the problem of distribution line grounding fault and improve the technical level of grounding fault location.

Reference

Embedding diverse features in latent space for face attribute editing
Rui Yuan, Xiping He*, Dan He, Yue Li

*Chongqing Key Laboratory of Intelligent Sensing and Blockchain Technology, School of Artificial Intelligence, Chongqing Technology and Business University, Chongqing 400067, China;
Corresponding author: jsjhxp@ctbu.edu.cn

ABSTRACT

Face attribute editing, one of the important research directions in face image synthesis and processing techniques, aims to photorealistic editing single or multiple attributes of face images on demand using editing and generation models. Most existing methods are based on generative adversarial networks, using target attribute vectors to control the editing region or Gaussian noise as conditional input to capture texture details. However, these cannot better control the consistency of attributes in irrelevant regions, while the generation of fidelity is also limited. In this paper, we propose a method that uses an optimized latent space to fuse the attribute feature maps into the latent space. At the same time, make full use of the conditional information for additional constraints. Then, in the image generation phase, we use a progressive architecture for controlled editing of face attributes at different granularities. At last, we also conducted an ablation study on the selected training scheme further to demonstrate the stability and accuracy of our chosen method. The experiments show that our proposed approach, using an end-to-end progressive image translation network architecture, obtained qualitative (FID) as well as quantitative (LPIPS) face image editing results.

Keywords: Face attribute editing, generative adversarial networks, latent space, image translation.

1. INTRODUCTION

In recent years, with the continuous advancement of deep learning and generation capabilities supported by hardware computational, face attribute editing techniques are developing rapidly, gaining more attention, and becoming a popular research topic. The flourishing face attribute editing method is also applied in a wide range of scenarios, including makeup in social media, dataset augmentation, editing-face attacks for intelligence security, character re-building for digital people and film production, etc. [1, 2, 3, 4, 5, 26]. The task of face attribute editing is usually to edit individual attributes of a face (e.g., eyes, nose, mouth, eyebrows) into the desired form or to achieve two or more multi-attribute manipulations simultaneously. However, the editing process is prone to failure cases, as shown in Figure 1, such as the edited result is not what we want (i.e., the wrong part of the attribute was edited, the edited result suffers from poor realism and the generated image is imposed with artifacts). The edited result loses identity information, which changes the original purpose of face attribute editing (some failures in Figure 1).

When editing is carried out, we want to do at least i) Accuracy. During face attribute editing, we want to change where we want, edit the specified attribute part, and strive to avoid the entanglement of attributes. ii) Stability. Not only should we ensure stable editing performance of the model, but also guarantee that the face image's identity and other basic attribute structures should not be changed before and after.

Generative adversarial networks (GAN) [6]. Its emergence and dramatic development allow the better transformation of low-dimensional data into high-dimensional, contributing significantly to the progress of target generation tasks, including face attribute editing. For example, image synthesis [7, 8], image super-resolution [9, 10], age editing [11, 12], expression editing [13, 14, 15], makeup creation and migration [16, 17] and many other impressive generation tasks.

Conditional information. By assisting the model in editing the attributes of faces in this way, the synthetic performance of generators is limited, and the intensity of attribute manipulation gets controllable. CGAN [18], as an extended model of GAN, additional conditions are introduced in the generator and discriminator, respectively. It can assist the discriminator in distinguishing whether the images satisfying the requirements are synthesized in the generation phase, and improve the stability of the model to some extent. IcGAN [19] represents and manipulates the latent and conditional vectors of attributes by using an encoder, which enables multi-attribute editing of face faces. Both conditional options are weak in controlling detailed attribute features and poor in realism. StarGAN [20] takes a cross-domain and hybrid dataset training approach to reconstruct the original image by ground truth (GT) labels and generated images and further uses label constraints or correspondences to optimize the model. StarGANv2 [21] adopts a mapping network to replace
GT labels with a style-encoded format to improve editing flexibility. All Stars are overly focused on GT, and those structured guides are not well protected against irrelevant attributes. Similar programs are ModularGAN [22] and ComboGAN [23]. SaGAN [24] introduces an attention mechanism by replacing the generator with an attribute manipulation network (AMN) and a spatial attention network (SAN). HiSD [25] was able to control the facial attribute structure stably by introducing a hierarchical tree structure label style. Nevertheless, some accurate results are generated, but there are fluctuations in diversity and stability, and there is still entanglement in manipulating multiple attributes. Its reliance on whether the original GT annotation is disentanglement.

Latent space. StyleGAN [26] passes the latent vector z through a mapping network to obtain the latent vector w and inputs the transformed w+ into a generative network structure similar to ProGAN [27] to achieve attribute disentanglement. Although photorealistic images were harvested, fine-grained problems (artifacts and substandard textures) at high resolution were exposed. Around StyleGAN, different mapping methods [28, 29, 30, 31] all yield high-fidelity face images full of details and structural integrity; [32, 33, 34, 35] try to make the editing of GAN controllable by dissecting the latent variables.

Image-to-Image translation. The supervised learning pix2pix borrows from CGAN [18] and uses the source domain image as a new constraint. pix2pixHD [36] proposes a multi-level coarse2fine generator with a multi-scale discriminator; the unsupervised cycleGAN [37] and MUNIT [38] adopt a new model of potential spatial information sharing, which is the subsequent new models [39, 40, 20] are the main focus of development. However, irrelevant attributes are vulnerable to tampering by methods that share attribute information.

In general, these methods have struggled to balance i) Accuracy and ii) Stability. To some extent, the face attribute editing models only perform well on a few face images because the boundaries between these attributes are inherently relatively distinct. For example, the performance is disappointing in scenes such as face shape and hairstyle, skin tone, and background, presumably because these models have a limited ability to separate or control the potential space of attribute information is restricted.

Compared to optimizing the latent space or performing interpretive editing, embedding with an encoder not only preserves the properties of GAN better but also is a solution that works well so far. Therefore, we first applied the embedded solution to the end-to-end network structure. Then we obtained competitive results using the face attribute editing model after a progressive generation network. Our work is summarized as follows:

1. We propose a method using an optimized latent space to embed diverse attribute feature maps into the latent space to fuse the attribute features while improving editing Accuracy.

2. We make full use of the existing labels and masks for additional constraints, utilize a progressive generation architecture, and further demonstrate through experiments that we can increase the Stability of face attribute editing.

3. We conducted an ablation study in different translation paths and proved that our setup was necessary, and the competitive results showed that we improved the face attribute editing.

2. MATERIALS AND METHOD

2.1. Experiments Setup

Dataset. We chose the CelebA-HQ [27] dataset for the evaluation and analysis of our experiments, which contains 30,000 images with labels that include glasses, beards, different hair colors, skin color, age, etc. It also contains masks for key attributes (five senses, face contours). We divided CelebA-HQ into a training set and a test set in a ratio of 9:1, and they have 27,000 and 3,000 images, respectively. In order to make a fair comparison with baselines, we adjusted the images to 256 uniformly for training and testing. No data or other constraints are added throughout the process.

Baselines. We use StyleGANv2 (with its Style-mixing), which is the method related to the latent space; using StarGANv2 (especially the style encoding processing that inspired us) not only makes comparisons in terms of style but also evaluates terms of performance of image translation. All methods are performed using only the implementation and pre-training weights publicly provided by the authors. Figure 4 shows our competitive results.

Evaluation metrics. We use FID [41] and LPIPS [42] to qualitatively evaluate the realism and diversity of generated results. As in Baseline, we use the same means to calculate the average of FID and LPIPS (picking reasonable results for comparison in an equivalent editing domain). Table 1 shows the quality of our reconstruction, where no competitive...
solution is taken, but strong and convincing FID values demonstrate that we obtain excellent results in terms of reconstruction.

2.2. Embedding Diverse Features

In StyleGAN [26], noise \( z \in \mathbb{N}(0, 1) \) is used as the input to the model. It passes through the mapping network as the input to the generative models of different layers, which gives diverse results but easily introduces unnecessary image information and imposes artifacts on the image, as shown in Figure 1(a)(b), which presents a piece of utterly incompatible attribute information when performing the synthesis of the face. Also, it has hair and background certain effects. However, StyleGAN has excellent properties, such as its progressive generation scheme, which inputs attribute information at different resolutions into the generation model. This model can control the change of face image from the shape (coarse style) to the texture (fine style), as shown on the left side of Figure 1(c), which shows the synthesis with rich facial. On the right side of (c), the result of noise in the coarse layers is shown, which has a greater impact on the background of the image and the overall structure of the face.

![Figure 1](image1.png)

Figure 1. Some failures. Except in the middle (c) is a picture of the generation explanation of the progressive architecture (Reproduced with permission of Ref. 28 and Ref. 21, Copyright of © 2020 IEEE.).

As same as StarGANv2, although it replaces the labels in the first version with style encoding, it inevitably manipulates irrelevant regions due to the fuzzy descriptions, as shown in Figure 1(d), which produces more severe artifacts when editing hairstyles of face images and the generated images are not of high quality and behave somewhat fuzzy. Such an image translation task is equally prone to the above problems when editing multiple attributes (gender and hairstyle), as shown in Figure 1(e).

![Figure 2](image2.png)

Figure 2. Optimized latent space, which reduces the effect of random noise input, controls artifacts and cares more about quality. Our method (right side).

**Optimized latent space.** Our research focuses on exploring i) Embedding diverse feature information to optimize the latent space by replacing operation of introducing noise in StyleGAN generation phase (as shown \( z \) at the bottom). We choose the information such as GT and masks. ii) Adapting the traditional latent space style so that GT and attribute
details can be fully embedded into the model. As shown in Figure 2 right side, we optimized the latent space and replaced the vector encoding pattern in StyleGAN (left side of Figure 2) by transforming $1 \times 512$ into latent space style $64 \times 63 \times 3$. The optimized latent space contains not only random noise but also rich texture and shape information, which makes the attribute editing more accurate, and the robustness will be improved. By embedding different sizes in the generation model stage, the optimized latent space is more preserved in the original GT and attribute information, and thus the generated images are more realistic and natural. We demonstrate the effectiveness of our proposed method in subsequent experiments.

**Goals.** Our goal is to have a better blend of semantic and label constraint editing advantages, with their respective advantages being i) greater influence of potential space manipulation on attribute globalization, as in the case of the effect generated by StyleGAN, where the image generation quality has been impressively improved under the flexible use of $W^+$ space, and ii) greater manipulation of regional attributes by label constraints. As in StarGANv2, the region style is encoded to improve the accuracy of editing and manipulating the part of the face attributes we specify. Now, we propose a solution that mixes their respective advantages, using the optimized latent space $W$, using the reference image as the new style constraint label, which includes the mask of the reference image. We also use the progressive generation architecture, which can generate more desirable editing results, and our network model as a whole, shown in Figure 3.

![Figure 3. The overall structure of our model.](image)

The structure of $E$ as shown on the right-top of Figure 3. It shows how we encode the reference image $I_{ref}$ and the original GT *Ground Truth Labels*. The module of down-sampling in $E$ consists of instance normalization (IN), and uses the average pooling method. MLP stands for Mapping network, which includes a sequence of 8 fully connected layers that can share weight information, and uses ReLU as its activation function. We discard the pixel normalization, which fuses the transformed noise $S'^i = M(z')$, $i \in \{1,2,...,N\}$ with the coding of reference image, and the fusion is used as a branch input of the generative model $G$. In particular, we adopt AdaIN[26], make generator has several input layers $S^{'input} = AdaIN(W^{i+}, E'(I_{ref}))$ with different resolutions $(8 \times 8, 16 \times 16)$, and we input different transformed images in each layer to control the style with varying degrees of fineness. The image to be edited is also encoded as another branch of input to the generative model, which replaces the input method of using constant vectors in StyleGAN, while appreciating the design pattern of ResNet [43]. We discard the introduction of noise in additional branches, which is also demonstrated in the subsequent image reconstruction experiments. $D$ represents the discriminator, which consists of eight pre-activation residual blocks that determine $D(G(S^{'input}))$ whether the generated image was obtained by the generator or was originally the real image.
2.3. Training Phases

In order to better optimize the latent space, we design additional image translation sessions which train itself to map and encode attribute information, and our Training Phases can be summarized as follows:

Pure-Translation path. We use only the structure of the encoder and the structure of the Decoder to train the encoder's ability to represent attributes $D(I, G(E(I'))), which is similar to the structure of U-Net [44], and we try to keep the information consistent.

Additional-Translation path. We then use the introduced noise and reference $D(I', G(E(I', W)))$ to improve the encoder's quality further and train and enhance the quality of the generator.

Whole-Translation path. In this one, we want to evaluate the performance of the model as a whole, with the complete translation step $D(I', G(AdaIN(W^m, E(I', W))))$, to get the image we want after the training is done.

Our optimizer is Adam [45], the parameter initialization is Kaiming [46], set the rest of the bias to 0 except for the bias of AdaIN, which is 1, and set the learning rate to 0.0001 uniformly.

2.4. Training Objectives

In this section, we present the loss functions required for each training phase, and we also conclude with a summary of the Objectives.

Reconstruction loss. In all training terms for image translation, we want to obtain images with as good a structural reconstruction as possible.

$$L_{\text{rec}} = E_i[||PT(E'(I') - I')||_1] + E_i[||AT(G(I') - I')||_1] + E_i[||WT((S_{input}^i) - S_{input}^i)||_1].$$ (1)

Where the PT, AT, WT indicate Training Phases as above. $I'$ refers to the input information, and $S_{input}^i$ refers the complete input, which is what we mentioned above.

Style loss (StarGAN). We encourage the styles to be well reconstructed in the first two training terms for image translation.

$$L_{\text{sty1}} = E_i[||E^i(S_{input}^i) - E^i(G(S_{input}^i))||_1].$$ (2)

In the last one, we force the style differences to be somewhat discriminative:

$$L_{\text{sty2}} = E_i[||E^i(S_{input}^i) - E^i(S_{output}^i)||_1].$$ (3)

Where $S_{output}^i$ means the results after editing.

Adversarial loss. The discriminator always tries to distinguish those images that have been edited by our model, which may come from different translation paths. At the same time, our generator tries to generate images that can spoof the discriminator. This method allows us to accomplish the editing task that satisfies realism and diversity as much as possible, and we employ non-saturating loss [6] as our antagonistic loss. In particular, we will demonstrate the validity of our choice in the quantitative results evaluation section of the ablation study.

$$L_{\text{adv}} = 2E_i[\log(D(S_{output1}^i))] + E_i[\log(1 - D(S_{output2}^i))] + E_i[\log(1 - D(S_{output3}^i))].$$ (4)

Full objective.

$$L_{\text{full}} = \lambda_1L_{\text{rec}} + \lambda_2L_{\text{sty1}} + \lambda_3L_{\text{sty2}} + \lambda_4L_{\text{adv}}.$$ (5)

Where $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\lambda_4 = 2$. 
3. RESULTS AND DISCUSSION

In this section, we show our experimental results. Notably, throughout, we do not make beneficial additions in the following fixed sections.

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Reconstruction</th>
<th>StyleGANv2</th>
<th>StarGANv2</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangs</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Nose</td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td>Hair Color</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td>Pale Skin</td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>Eyeglasses</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 4. The results of some competing experiments around baselines (Reproduced with permission of Ref. 28 and Ref. 21, Copyright of © 2020 IEEE.) are shown in the last three columns (Red-box: what we want to change).

As shown in Figure 4, our method achieved more realistic results in image reconstruction, with an average FID value of 4.68. For the overall comparison image, we have taken the best image reconstruction method of the two methods provided by Open Access in order to be able to make a sharp comparison in terms of editing capabilities. The labels on the left, showing the type of attribute operation we are currently performing, give a slight edge to our method of bangs editing. For glasses and hair color, StyleGAN's editing is not strong enough under settings that avoid destroying the shape of the face, and StarGAN suffers from similar problems. For the nose, StyleGAN carried out the wrong operation, giving serious color bias to the image, while StarGAN also suffers from tampering with irrelevant regions, which is also seen in Pale Skin, such as the effect of nose. Our method effectively enhanced editing accuracy and improved editing quality by using style encoding and improving the latent space. The corresponding results are also explained below.
Table 1. Comparison with the baselines, the result of our method. (Bold: best.)

<table>
<thead>
<tr>
<th>Method</th>
<th>FID</th>
<th>LPIPS</th>
<th>Disentanglement</th>
<th>( \text{FID}_{\text{avg}} )</th>
<th>( \text{LPIPS}_{\text{avg}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleGANv2</td>
<td>34.18</td>
<td>0.288</td>
<td></td>
<td>38.78</td>
<td>0.275</td>
</tr>
<tr>
<td>StarGANv2</td>
<td>25.49</td>
<td>0.244</td>
<td></td>
<td>29.09</td>
<td>0.299</td>
</tr>
<tr>
<td>Ours</td>
<td>22.26</td>
<td>0.311</td>
<td></td>
<td>23.28</td>
<td>0.281</td>
</tr>
</tbody>
</table>

On the left side of Table 1, it shows our comparative advantage in face attribute editing. In terms of realism, our model was able to adapt more to visual perception and achieved lower scores (FID) than the other two models. However, in StyleGAN, unnecessary modifications occurred, and irrelevant attributes were edited, so it also competes slightly less in terms of visual effects. StarGAN, although with some enhancements or improvements, does not achieve the preservation of details, and the global visual effects are vulnerable to editing due to style coding, which is obviously not what we want. Also, in terms of generated diversity, our model is better competitive, and the higher score (LPIPS) shows our advantage.

**Disentanglement.** The right side of Table 1 presents our generative ability within the same domain, and the different results reflect not only the diversity but also the average FID scores of different domains, which also reflect that the disentanglement ability of our method is quite competitive.

Table 2. Dependence of selection loss function. (Bold: best.)

<table>
<thead>
<tr>
<th>Removed Loss</th>
<th>Usage Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adversarial Loss</td>
<td>√</td>
</tr>
<tr>
<td>Style Loss</td>
<td>√</td>
</tr>
<tr>
<td>Reconstruction Loss</td>
<td>√</td>
</tr>
<tr>
<td>( F_{\text{score}} )</td>
<td>199.68</td>
</tr>
</tbody>
</table>

Ablation study. As illustrated in Table 2, we mainly evaluated the editing results of removing any of the losses, again quantitatively and qualitatively. In particular, we took the fusion score rule for the evaluating of disentanglement:

\[
F_{\text{score}} = \text{FID} - \text{LPIPS}
\]  

### 4. CONCLUSION

In this paper, to alleviate the problem of face attribute editing suffering from accuracy and stability, we propose a method using an optimized latent space for editing face attributes. The key point of our idea is the full use of the latent space dimension, which has rich fine information and effectively improves the quality of the generated images when fused to different resolution layers using AdalN. We focus on i) how to optimize the latent space and embed diverse feature information, and ii) the impact of the embedded GT information on the accuracy of face attribute editing. Our method avoids global manipulation of face attributes, and the competitive results of local editing appear reliable from both qualitative (FID) and quantitative (LPIPS). Experiments demonstrate that our plan also has stable evaluation scores for maintaining attribute structure and disentanglement, which competes impressively with other methods. However, our approach is unsatisfactory in partially unaligned image editing operations, so we will focus on addressing such issues in the future.

**ACKNOWLEDGEMENT**

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202200803) and “Innovative scientific research project” for graduate students of Chongqing Technology and Business University (Grant No. yjscxx2022-112-68; Grant No. yjscxx2022-112-180; Grant No. yjscxx2022-112-181).
REFERENCES


Image Harmonization with Spatial Feature Interaction and Back-projection Upsample

Tianyanshi Liu\textsuperscript{1a}, Yuhang Li\textsuperscript{1b}, Youdong Ding\textsuperscript{1c*}

\textsuperscript{1} Shanghai Film Academy Shanghai University, Shanghai, China
\textsuperscript{a}liutian@shu.edu.cn, \textsuperscript{b}yuhangli@shu.edu.cn, \textsuperscript{c}\textsuperscript{*}1312652782@qq.com

Abstract

Without any processing, the synthetic image visually unrealistic due to the established differences in the appearance of the foreground and background. In view of this situation, the task of image harmonization arises at the historic moment, and its purpose is to adjust the foreground appearance of a synthesized image to be closer to the background, thereby eliminating local visual differences. However, due to the limitation of the spatial feature interaction range in the feature extraction process, the global appearance transfer effect is not good. Therefore, to solve this problem, we propose an enhanced spatial feature interaction module. Meanwhile, we propose a back-projection upsampling module, which refines the reconstruction error during the reconstruction upsampling process and better restores the details of the reconstruction foreground. Our experiments on a public dataset, iHarmony4, show that the method effectively generates synthetic images with consistent overall appearance and enhanced detail.

Keywords-image harmonization; feature interaction; involution; back-projection upsampling

1. Introduction

In the image synthesis operation, the foreground and background are often very different in appearance (such as lighting, color, etc.) due to different sources or shooting conditions. The simple operation of superimposing the foreground on the background will make the overall image uncoordinated. Image harmonization aims to adjust the foreground appearance of the composite image, making it closer to the background, so that the picture achieves the effect of overall consistency and enhances the realism of the composite image. Image harmonization has practical requirements in image editing and data enhancement, and image local color correction.

The existing deep learning-based image harmonization \cite{1,2,3,4,5,6} algorithms have achieved good results, especially in the process of generative antagonistic interactive game, the generative adversarial network \cite{2} is introduced to maintain the reality of the image. Although, existing methods focus on adjusting the global consistency of the foreground and background, the global transfer of appearance is not effective because of the limited interactive scope of spatial features. In addition, the loss of reconstruction details in the foreground area will also affect the effect of image harmonization.

Based on the above analysis, we adopt a generative adversarial network structure and redesign the generative network part to solve the above problems. The backbone network of the generator is a U-Net \cite{7} based on skip-connections. In the encoder part, we design an enhanced spatial feature interaction module, which can aggregate the context in a wider space in the feature extraction stage, thereby enhancing the long-range interaction modeling ability and making the background appearance better transfer to the foreground region. At the same time, in the decoder part, a back-projection upsampling module is designed to continuously refine the foreground detail information in the process of image upsampling and reconstruction to improve the quality of image reconstruction.

2. Image Harmonization Network

As shown in Fig. 1, basis of our generative network is a skip-connected encoder-decoder network. On this basis, we propose an enhanced spatial feature interaction and back-projection upsampling module to optimize the generation network to obtain better image harmonization effects. More details of the modules are as follows:
2.1. Enhanced Spatial Feature Interaction Module (ESFIM)

The spatial independence of ordinary convolution requires the convolution kernel to share parameters in all spatial positions, and the ability to adapt to different visual modes in different spatial positions is weak. In addition, the locality of convolution also limits the receptive field and cannot efficiently capture long-distance relationships in space. The involution [8] innovatively changes the spatial sharing of convolution into spatial specificity, and the channel specificity becomes channel sharing, thus overcoming the difficulty of modeling long-range interactions with ordinary convolutions.

Based on this, as shown in Fig. 2, we propose an enhanced spatial feature interaction module, which uses the characteristics of involution to solve the problem of the limitation of the spatial feature interaction range in the feature extraction process, pay more attention to long-distance visual elements, and increase the ability to interact with foreground and background features.

First, we use 1×1 convolution to complete data dimensionality reduction and cross-channel fusion of features and add nonlinear excitation to the representation of the previous layer to improve the generalization ability of the network. After that, we use average pooling for down sampling. Finally, through a layer of involution, since the involution kernel is dynamically generated as shown in Fig. 2, it can adaptively assign different weights to different spatial positions, which is beneficial to find out visual elements that contribute more to the foreground in the spatial domain. The convolution kernel is generated as follows:

\[ \omega_{ij} = c_i \sigma(c_0 X_{ij}) \]  

(1)

Among them, \( X_{ij} \) represents the feature vector of position \((i,j)\). \( c_0, c_i \) represent 1x1 convolution. \( \sigma \) represents BN (Batch Normalization) and activation function LeakyReLU. \( \omega_{ij} \) represents the parameter value of the kernel at the \((i,j)\) position.

Fig. 2. Enhanced Spatial Feature Interaction Module (ESFIM) and Back-projection Upsampling Module (BPUM)
2.2. Back-projection Upsampling Module (BPUM)

In order to increase the foreground detail information in the synthesized image and improve the quality of image reconstruction, due to the influence of image super-resolution [9], the back projection method is introduced. Back-projection is an effective iterative process that provides an error feedback mechanism for the projection error of each stage, which can minimize the reconstruction error and better capture the interdependence between the input image and the reconstructed image pair. This provides higher quality reconstruction results.

The back-projection upsampling structure is shown in Fig. 2. First, the input feature map $L$ is mapped to the $\times2$ space to obtain the feature map $L_u$, and then the process of downsampling is simulated, and feature map $L_u$ is mapped to the $\times0.5$ space to obtain the feature map $L_d$, and back-projection is performed. The error $e$ is calculated during the mapping and de-mapping process to guide the image reconstruction mapping mechanism. Then perform the $\times2$ upsampling operation on $e$ to obtain $e_u$, increase the high-frequency information of the image, and output the up-sampling feature map $L_h$ to obtain a reconstructed image with enhanced details. The back-projection upsampling module is defined as follows:

\[
L_u = (L \ast c_u) \uparrow_{\times2} \tag{2}
\]
\[
L_d = (L_u \ast c_d) \downarrow_{\times0.5} \tag{3}
\]
\[
e_u = ((L - L_d) \ast c_e) \uparrow_{\times2} \tag{4}
\]
\[
L_h = L_u + e_u \tag{5}
\]

Among them, $c_u$, $c_e$ represents deconvolution layers, $c_d$ represents convolution layer. $\ast$ represents a convolution operation.

3. Experiments and Analysis

3.1. Implementation details

We use the iHarmony4 dataset collected by Cong et al. [2] to train and test our model. Among them, the iHarmony4 dataset contains HCOCO, HFlickr, HAdobe5k and Hday2night 4 sub-datasets. Among them, there are 7,3146 pairs of images in the dataset, and each pair of images contains synthetic images and corresponding ground truth images and foreground masks. Among all image pairs in the iHarmony4 dataset, 6,5742 pairs are used for training and 7404 pairs are used for testing. We input the iHarmony4 training dataset to train the model. To generate training images, the original images in the iHarmony4 dataset were resized to 256$\times$256 resolution. We initiate the network parameters using the method in [1] and update them by the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The initial learning rate is set to 0.0002. We trained 100 epochs and set the batchsize to 12. All the experiments have been conducted on NVIDIA TITAN Xp GPUs using PyTorch 1.7.

Table 1. Quantitative performance on iHarmony4 test set

<table>
<thead>
<tr>
<th>Method</th>
<th>HAdobe5k</th>
<th>HCOCO</th>
<th>Hday2night</th>
<th>HFlickr</th>
<th>iHarmony4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>MSE</td>
<td>PSNR</td>
<td>MSE</td>
<td>PSNR</td>
</tr>
<tr>
<td>DIH</td>
<td>32.28</td>
<td>92.65</td>
<td>34.69</td>
<td>51.85</td>
<td>34.62</td>
</tr>
<tr>
<td>DoveNet</td>
<td>34.34</td>
<td>52.32</td>
<td>35.83</td>
<td>36.72</td>
<td>35.18</td>
</tr>
<tr>
<td>RainNet</td>
<td>35.84</td>
<td>44.01</td>
<td>36.58</td>
<td>28.07</td>
<td>35.67</td>
</tr>
<tr>
<td>Ours</td>
<td>37.16</td>
<td>29.37</td>
<td>36.80</td>
<td>25.23</td>
<td>37.00</td>
</tr>
</tbody>
</table>

Terms of Use: https://www.spiedigitallibrary.org/terms-of-use
3.2. Quantitative results

As shown in Table 1, we select the existing image harmonization methods DIH[3], DoveNet[2], RainNet[1] and our method for testing on the dataset iHarmony4 and all sub-datasets HCOCO, HFlickr, HAdobe5k and Hday2night. We use mean square error (MSE) and peak signal-to-noise ratio (PSNR) quantitative indicators to evaluate the effect of image harmonization. It shows better performance on all test datasets of iHarmony4. Our method is able to obtain synthetic images with a more consistent overall appearance and enhanced detail, enabling better image harmonization.

We test the prospects of different proportions, and use the mean square error (MSE) and the foreground mean square error (fMSE) to evaluate. fMSE [2] is calculated as follows:

\[ f_{MSE} = \frac{1}{\sum_{i,j} M_{i,j}} \sum_{i,j} M_{i,j} \cdot (H_{i,j} - R_{i,j})^2 \]  

We divide the proportion of prospects into 0~5%, 5%~15% and 15%~100% for testing. The smaller index, the better performance. As shown in Table 2, both indices of our method become smaller under different proportions. In particular, the expansion of the spatial feature interaction range and the foreground details enhance the synthetic images with a large proportion of the foreground, and the effect of image coordination is significantly improved, which reflects the effectiveness of our method.

3.3. Qualitative results

As shown in Fig.3, we test on the iHarmony4 dataset, as shown, there are composite images from left to right, DIH[3], DoveNet[2], RainNet[1], our method, real images. Comparing all methods with ground truth, our method can better capture the global lighting and color style, especially capture more effective reference information in the background and apply it to the foreground to obtain the overall consistency of the synthetic image and more adequate foreground details. It is further observed that because the scope of the hard masks in the image harmonization method is limited, our method still has the problem of over-refining the boundary.
3.4. Ablation study

To verify the effectiveness of the two modules of enhanced spatial feature interaction and back-projection upsampling in our method, the corresponding ablation experimental results are shown in Table 3. The interaction of enhanced spatial features improves the performance measurement of the model and makes the overall appearance of the image more consistent. After that, we analyze the loss of details in the reconstruction process and the back-projection upsampling module is added, which doubles the performance index of the model and enhances the foreground details.

<table>
<thead>
<tr>
<th>Combination</th>
<th>iHarmony4</th>
<th>PSNR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>base</td>
<td>35.75</td>
<td>42.28</td>
<td></td>
</tr>
<tr>
<td>+ESFIM</td>
<td>36.02</td>
<td>39.18</td>
<td></td>
</tr>
<tr>
<td>+ESFIM+BPUM (Ours)</td>
<td>36.33</td>
<td>34.90</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion

This paper proposes enhanced spatial feature interaction and back-projection upsampling modules, based on which the image harmonization generation network is redesigned to perform image coordination on synthetic images to expand the spatial feature interaction range and increase foreground details. Compared to other methods, our model performs better and results in images with more consistent overall appearance and enhanced detail.
Acknowledgments

This work was supported by the National Natural Science Foundation of China Nos. 61303093, 61402278.

References


Research and Implementation of Dynamic Gesture Recognition System based on ZYNQ

Jie Li\textsuperscript{1,2}\textsuperscript{*}, Qingying Liu\textsuperscript{1}, Zengzhen Li\textsuperscript{1}, Wei Chen\textsuperscript{1}
\textsuperscript{1}Chongqing Engineering Research Center of New Energy Storage Devices and Applications, Chongqing University of Arts and Sciences, Chongqing 402160, China
\textsuperscript{2}College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
\textsuperscript{*}rodgersli@163.com

Abstract

At present, gesture has become an important channel of human-computer interaction, and gesture recognition has been widely used in various fields. In this paper, the dynamic gesture recognition technology is studied from algorithm and system implementation for portable devices which require high real-time performance. The algorithm mainly uses the region of interest extraction based on face recognition, skin color detection based on HCrCg color space and gesture motion track marking based on scanline seed filling algorithm. The system is implemented by Xilinx ZYNQ, and a SOPC system architecture based on ARM Cortex-A9 hard core and ARM Cortex-M3 soft core and FPGA is proposed. The scanline seed filling algorithm with long running time is designed as a hardware accelerator to improve the running speed. Through the test of the prototype, the recognition accuracy can reach 95.75% in a simple background and 90.83% in a complex background. The average running time of the system is only 0.68 seconds, which is more than 30% faster than using pure software method. The system has good performance in recognition accuracy and running speed.

Keywords- dynamic gesture recognition; skin area detection; FPGA; SOPC

1. Introduction

Gestures are indispensable in people's daily communication, and language combined with gestures can express richer semantic information\cite{1}. With the rapid development of computer and artificial intelligence technology, gesture has become a new and mainstream human-computer interaction channel after keyboard, mouse and touch screen\cite{2}. It has penetrated into every corner of social life, such as smart home, telecommuting, health care, education and entertainment, and other fields. Therefore, the demand for lightweight and real-time portable devices is increasing, which puts forward more requirements for gesture recognition algorithms, software and hardware.

Pang\cite{3} used YCrCb color space for skin color segmentation and gesture classification based on HMM-NBC model. Lan\cite{4} effectively improved the segmentation accuracy by combining color information, motion information and shape positioning for gesture segmentation. In paper \cite{5}, dual-channel convolutional neural network (DC-CNN) is used for dynamic gesture recognition. Each channel has a separate weight and softmax classifier for classification of output results. In paper\cite{6}, Kinect sensor is used to collect depth images, color model combined with convolutional neural network is used to adjust the threshold and weight of neural network. The recognition accuracy reached 98.52%.

It can be seen that compared with traditional image processing methods, deep learning has a series of advantages such as high accuracy, good fault tolerance and strong robustness. However, the computational load is large, and the real-time performance of recognition needs to be improved, so it is not suitable for implementation in portable devices.

Zhao\cite{7} designed a real-time gesture information extraction algorithm based on FPGA through color gamut separation and Gaussian threshold optimization, which improved the running speed. Raheja\cite{8} used FPGA to implement a static gesture recognition method based on principal component analysis algorithm. Paper \cite{9} proposed a retractable real-time visual gesture recognition method based on FPGA, which can detect both static hand shapes and dynamic gestures. In paper \cite{10}, dual cameras were used to construct 3D images for gesture recognition, and VLSI was used to implement the system. The whole system was also verified on FPGA.
In this paper, the algorithm scheme of dynamic gesture recognition is firstly proposed for portable devices with high real-time requirements, and then a SOPC system architecture based on ARM Cortex-A9 hard core and ARM Cortex-M3 soft core and FPGA is designed. The algorithm module with long running time is designed as hardware accelerator to improve the running speed effectively.

2. Algorithm Design

The main algorithm flow of the system is shown in Fig.1. When the gesture starts, get a frame from the video captured by the camera. Firstly, the region of interest (ROI) of the image is extracted, and then the image is segmented according to the detected skin area. Finally, the geometric center point of the segmented skin area is marked with white dots on the corresponding position of the black image. The motion trajectory of the gesture can be obtained by accumulating the marker points of multiple frames. After the gesture is completed, template matching is performed on the motion trajectory image to obtain the gesture result.

2.1. ROI extraction based on face detection

The system adopts the ROI extraction method based on face detection. First, the haar classifier is used for face detection. If there are multiple faces in the image, only the face with the largest area is retained and marked with a rectangular frame. Then, the width L of the rectangular frame is calculated, and the area is expanded by 1.5L to the left and right, 0.5L upward, and 0.8L downward as the ROI area and intercepted. If the expansion exceeds the boundary of a certain side of the image, the boundary is used as the boundary of the ROI for interception. Due to different cameras and different positions of people, the size of the ROI area after interception is inconsistent. In order to use the same size for subsequent image processing and template matching, it is necessary to normalize the size of the ROI area. In this system, the ROI area is normalized to 640×480 using bilinear interpolation algorithm.

2.2. Skin area detection based on HCrCg color space

Commonly used color spaces are RGB, HSV, YCrCb, etc. Histogram analysis of each color channel in each color space shows that H, Cr and Cg channels have better clustering for skin color region and can be used to build skin area detection model. According to the simulation analysis and Chebyshev's law of Large numbers, the HCrCg model can be used to extract the skin area more accurately within the threshold range: 5≤H≤33, 135≤Cr≤176, 111≤Cg≤126.

2.3. Image segmentation based on marker-based watershed algorithm

In order to obtain more accurate skin color region, the system uses watershed algorithm to further segment the image. Firstly, the image extracted by the HCrCg model for the skin color region is binarized and median filtered, and then 5 times of corrosion operations are performed to obtain the foreground image. Dilate the median filtered image for 5 times.
and invert the pixels to obtain the background image. The labeled image is obtained by adding the corresponding pixel values of foreground image and background image. Finally, based on the labeled image, the watershed algorithm is performed on the normalized ROI image, and then the image is binarized.

### 2.4. Gesture trajectory marking based on scanline seed filling algorithm

Based on the segmentation result of watershed algorithm, the system uses scanline seed filling algorithm to fill the white area in the image, and judge the boundary information and location information of the skin area according to the filling result.

According to the processing results of scanline seed filling algorithm, the position coordinates of upper left and lower right corner of each skin color segmentation region can be obtained. Since the region of interest of the image has been extracted before seed filling, the main object in the image is only a person. In the image after image segmentation based on the watershed algorithm, the three areas with the largest white area will be the head of the person and the two hands, and each area will be larger than 6000 pixels. Therefore, after scanline seeds are filled, only the three largest regions with an area of more than 6000 pixels and white areas accounting for more than 50% are reserved, and the areas are framed by rectangle and there geometric center points are marked.

In gesture recognition, when the two hands do not perform gesture movement, both hands are naturally down, then there will be only one rectangular area in the image. When it is detected that the number of rectangular areas in the image is greater than 1, it means that there is gesture movement. At this time, the coordinate of geometric center point of the rectangular areas in each frame are marked as a white point with a radius of 7 pixels on an black image. The accumulated marks of multiple frames of images describe the trajectory of the gesture movement. When a gesture recognition is completed, all the markers are cleared for the next gesture recognition.

### 2.5. Gesture trajectory template matching

This system mainly realizes four kinds of gestures: zoom in, zoom out, slide and rotate. The normalized correlation coefficient template matching method is used to match the binary image of the gesture trajectory obtained in the previous section with the four templates stored in the system, and four matching scores are obtained. If the four scores are all less than 0.1, the gesture is considered a failure. Otherwise, the match with the highest score will be output as the corresponding gesture result. The four template images are shown in Fig.1 respectively.

### 3. Design of SOPC System Based on ZYNQ

The main chip is Xilinx ZYNQ 7020, and combined with the peripheral OV5640 camera, LCD display module, power module, LEDs, buttons, buzzer, serial port and other circuits, which constitute the whole gesture recognition system.

#### 3.1. Hardware and software division

The ZYNQ 7000 series chips are divided into PS (Processing System) side and PL (Programmable Logic) side. The PS side is a dual-core ARM Cortex-A9 processor, which can be developed through the C language or PYNQ platform. The PL side uses the Xilinx 7 series FPGA architecture, which can be developed using hardware description language. Therefore, the ZYNQ 7000 series has not only the advantages of ASIC in terms of power consumption, performance and compatibility, but also the advantages of FPGA hardware programmability.

In order to give full play to the advantages of the software and hardware of the system, it is necessary to carry out a reasonable division of software and hardware. In order to improve the processing speed, the algorithm step which is easy to be designed by finite state machine with hardware description language is designed as hardware accelerator by FPGA. The rest of the algorithms are processed by ARM Cortex-A9 on the PS side by means of software design to reduce the design difficulty.

<table>
<thead>
<tr>
<th>Algorithm steps</th>
<th>ROI extraction and normalization</th>
<th>HCrCg skin color detection</th>
<th>Image segmentation</th>
<th>Scanline seed filling</th>
<th>Template matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running time</td>
<td>89</td>
<td>3</td>
<td>22</td>
<td>765</td>
<td>31</td>
</tr>
</tbody>
</table>
The algorithms of the whole process of dynamic gesture recognition described in the above section were simulated on PC, and the average running time required by each algorithm step was obtained, as shown in Table 1. It can be seen from the table that the running time of scanline seed filling algorithm is much longer than other algorithms. Therefore, the scanline seed filling algorithm is designed as hardware accelerator in order to improve the running speed.

In order to maximize the advantages of ZYNQ logic resources and improve the parallel processing capability of the system, the ARM Cortex-M3 soft core is added to the PL side, which is specially used to control serial port, buzzer, buttons and LEDs.

The overall software and hardware division and control flow of this system is shown in Fig. 2. When cortex-M3 detects the start button, it writes a shared register to tell Cortex-A9 to start the OV5640 camera to capture gesture images and store them into the DDR. Cortex-A9 reads image data from DDR through PYNQ for ROI extraction and normalization, skin area detection and image segmentation, and then writes the shared register to start the hardware accelerator for scanline seed filling. After filling, the shared register is written to notify Cortex-A9 for subsequent template matching algorithm to obtain gesture recognition result and store the result in the shared register. The Cortex-M3 reads the results from the shared register and sends them to the PC via the serial port, so as to control the zoom-in, zoom-out and flip pages of PPT, PDF, word and other documents.

3.2. SOPC Architecture

The SOPC architecture based on ZYNQ 7020 is shown in Fig. 3. The PS side uses the AXI bus to connect with the hardware logic of the PL side. The PS side is mainly ARM Cortex-A9 hard core with Linux operating system, using PYNQ and OPENCV for image processing. The PL side mainly includes camera control module, LCD display module and SYSTCTRL module. The SYSTCTRL module contains an ARM Cortex-M3 soft core Cortex_M3_top, a shared register module Share_REG and a hardware accelerator module HW_ACC for implementing the scanline seed filling algorithm.
By configuring the corresponding registers in the shared registers, the Cortex-A9 and Cortex-M3 cores can also start the camera, control the buzzer and LEDs, set the ROI normalization size and configure the HCrCg three-channel threshold.

4. Test Results and Analysis

ZYNQ 7020 core board, OV5640 camera, 4.3-inch LCD, USBto232 and other modules were used to complete the prototype of the dynamic gesture recognition system, as shown in Fig.4. A large number of experiments and analysis were conducted based on the prototype.

VIVADO 2019.2 was used for ZYNQ development. In this system, the main frequency of the ARM Cortex-A9 is 766MHz, the main frequency of the Cortex-M3 is 40MHz, and the clock frequency of the hardware accelerator is 50MHz.
4.1. Recognition accuracy analysis

In order to verify the accuracy of the ZYNQ-based dynamic gesture recognition system, the recognition accuracy was tested on 20 individuals under different backgrounds. Firstly, the four gestures were tested 100 times under simple background, as shown in Fig.5(a). The test results are shown in Table 2. As can be seen from the table, the average recognition accuracy is 95.75%.

![Fig. 5 Simple and complex test background](image)

Table 2. Recognition accuracy test under simple background

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Test number</th>
<th>Correct number</th>
<th>Accuracy</th>
<th>Average accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom in</td>
<td>100</td>
<td>95</td>
<td>95%</td>
<td>95.75%</td>
</tr>
<tr>
<td>Zoom out</td>
<td>100</td>
<td>96</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Slide</td>
<td>100</td>
<td>95</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>Rotate</td>
<td>100</td>
<td>97</td>
<td>97%</td>
<td></td>
</tr>
</tbody>
</table>

Four gestures were tested 100 times under three complex backgrounds, as shown in Fig.5(b,c,d). The test results are shown in Table 3. It can be seen from the table that the recognition accuracy is lower than that of simple background, and the average recognition accuracy is 90.83%, which is due to the interference of other colors close to the skin color in the complex background.

<table>
<thead>
<tr>
<th>Background</th>
<th>Gesture</th>
<th>Test number</th>
<th>Correct number</th>
<th>Accuracy</th>
<th>Average accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bookcase</td>
<td>Zoom in</td>
<td>100</td>
<td>87</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zoom out</td>
<td>100</td>
<td>92</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slide</td>
<td>100</td>
<td>88</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rotate</td>
<td>100</td>
<td>91</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>Culture wall</td>
<td>Zoom in</td>
<td>100</td>
<td>85</td>
<td>85%</td>
<td>90.83%</td>
</tr>
<tr>
<td></td>
<td>Zoom out</td>
<td>100</td>
<td>90</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slide</td>
<td>100</td>
<td>88</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rotate</td>
<td>100</td>
<td>88</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>Outdoor</td>
<td>Zoom in</td>
<td>100</td>
<td>93</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zoom out</td>
<td>100</td>
<td>94</td>
<td>94%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slide</td>
<td>100</td>
<td>91</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rotate</td>
<td>100</td>
<td>93</td>
<td>93%</td>
<td></td>
</tr>
</tbody>
</table>

By comparison, the recognition accuracy of the system is improved compared with other gesture recognition systems based on FPGA[9] and VLSI[10]. Compared with the gesture recognition using deep convolutional neural network in paper[11], although the recognition accuracy is 1.5% lower in the simple background, it is 5% higher in the complex background.
4.2. Recognition speed analysis

Table 4. System running time (ms)

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Running time</th>
<th>Average running time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No hardware accelerator</td>
<td>Hardware accelerator</td>
</tr>
<tr>
<td>Zoom in</td>
<td>992</td>
<td>677</td>
</tr>
<tr>
<td>Zoom out</td>
<td>983</td>
<td>665</td>
</tr>
<tr>
<td>Slide</td>
<td>1015</td>
<td>684</td>
</tr>
<tr>
<td>Rotate</td>
<td>998</td>
<td>687</td>
</tr>
</tbody>
</table>

By adding timing function in PYNQ’s Python code to test the processing speed of a frame image in dynamic gesture recognition, the running time is printed out after the system runs. Each gesture was tested with and without hardware acceleration in order to test how fast the hardware accelerator improved the system. Without hardware accelerator, scanline seed filling was designed in Python code and executed by ARM Cortex-A9. The test results are shown in Table 4. As can be seen from the table, the average running time without hardware acceleration is 997 milliseconds. When the scanline seed filling algorithm is implemented by FPGA hardware logic circuit, the average running time is 678 milliseconds, and the speed is increased by 32%. Therefore, the SOPC system architecture is very effective to improve the running speed.

5. Conclusion

In this paper, The main algorithms of dynamic gesture recognition are studied and the ARM Cortex-A9+ Cortex-M3+FPGA SOPC architecture based on ZYNQ is adopted. Extensive experiments were performed on the experimental prototype for four gestures of zoom in, zoom out, slide and rotate. It can be concluded that the average recognition accuracy of the system in the simple background is above 95%, and the average recognition accuracy in the complex background can reach 90%. Through the test of recognition speed, the average recognition time of the system is only 678 milliseconds, which is 30% faster than that without hardware accelerator. The system has good performance in recognition accuracy and running speed.

Acknowledgments

This work was financially supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN2021101321), General Project of Chongqing Natural Science Foundation (cstc2021jcyj-msxmX0271), College Students’s Innovation and Entrepreneurship Training Program (S202110642009).

References


VisMole: A Molecular Representation based on Voxel for Molecular Property Prediction

Qiang Tong12a*, Jiahao Shen2b*, Xiulei Liu12c*

1Beijing Advanced Innovation Center for Materials Genome Engineering, Beijing Information Science and Technology University, Beijing, China
2Laboratory of Data Science and Information Studies Computer School, Beijing Info. Science & Technology University, Beijing, China
a*tongq85@bistu.edu.cn, b*sy5622_5@126.com, c*xiuleiliu@hotmail.com

Abstract

To make computers understand the molecules, the first and important thing is to represent molecules in a proper way, which will affect the efficiency of chemistry tasks like property prediction and molecular design. In this work, we introduce a molecular representation for noncrystalline small molecules based on the theory of quantum physics. This representation captures the microscopic spatial structure of the molecule, which ensures it reflects more visual perception information about the molecule. We use Drug3DNet as our baseline and test the efficiency of our representation. By comparing with several other representations, we prove that our representation performs better on most of the properties.

Keywords-molecular representation; quantum chemistry; property prediction

1. Introduction

Since deep learning develops in chemistry, molecular representation always plays an important role in it. A proper representation should carry as much information as it can to make sure that the computers can understand the molecules better. At present, research on molecular representation is becoming more popular. Although everything around us is made up of molecules, the size of the molecules is too small to be directly observed by human eyes. By establishing a visual molecular model to understand and analyze the microstructure of molecules from visual perception while retaining the microscopic characteristics, it can raise the cognition of these molecules, so as to promote the progress in many related fields, e.g. drug creation [1] and materials discovery [2].

Over the past few years, the developed deep learning has a significant improvement in the field of molecular simulation [3]. The molecular simulation method based on deep learning firstly needs to encode the molecular structure, and then use the deep learning network to learn the molecular properties contained in it. A comprehensive reference [4] lists several thousand different descriptors to encode molecular structures. The most popular general classes of such encodings include string, graph, and 3D representations. However, string and graph representations just indicate the abstraction of molecules, which is a simplification of the microscopic view from a macroscopic perspective. It is difficult to accurately describe the conformation of molecules by using these representations.

Some works try to build 3D molecular models for specific tasks, such as protein-ligand prediction [5] or crystal prediction [6]. Protein-ligand is usually a biomacromolecule containing thousands of atoms. Because of the large number of atoms, it’s unnecessary to explore the microscopic structure of a single atom but using atom-level representation for protein modeling. Meanwhile, the property of a crystal depends on its regular grid structure rather than the single molecule. Thus the 3D crystal models mostly focus on the regular structure but not molecular structure.

Currently, the conventional 3D representation for non-crystalline small molecules often lacks an accurate description, but a rough approximation of microscopic structure. Although the macroscopic view of molecules can help us understand molecular properties in some ways, it is difficult to have a more accurate understanding limited by the gap between the macro and the micro world. The key contributions of this paper are as follows:

- We analyze the weakness of the existing representation.
- We design a new diffusion method based on the theory of quantum physics.
We introduce a representation for non-crystalline small molecules that conforms to the microscopic spatial structure.

The organization of this paper is as follows. In Section 2, we survey various representations of molecules and compare the pros and cons of them. Section 3 introduces the existing voxel-based representation and Section 4 introduces the molecular representation we proposed. Section 5 analyzes the experimental results of using different settings through multiple comparative experiments. We can find the improvement from the results of these experiments. The conclusion of our work is in the last section.

2. Related Works

2.1. 2D Molecular Representation

As mentioned in the previous section, there are several approaches to encoding molecular structure. Fig. 1 illustrates the different views of these representations on a molecule.

Simplified molecular input line entry system (SMILES) [7] is the most widely used string representation for encoding molecular structure, which uses a linear sequence of letters and numbers to extract the structural features of compounds. International Chemical Identifier (InChI) [8] is a more verbose string representation that explicitly specifies a chemical formula, atoms’ charges, hydrogens, and isotopes. Extended connectivity fingerprint (ECFP) [9] is a widely used molecular fingerprint to construct a quantitative structure-activity relationship (QSAR) model of compounds.

Since string representation focuses on the basic consist of a molecule, it lacks the spatial information of the atoms. Researchers turn to use graph representation to keep the combination of each pairs of atoms in molecules. Graph representation has long been used to extract features from molecular structure [11]. The MolGAN model [12] represents the molecular graph as adjacency matrix and node feature matrix. Junction Tree VAE [13] process molecules in terms of their subgraphs. However, all these methods neglect the 3D molecular structure, which means that they lose information about the abundant molecular conformation.

2.2. 3D Molecular Representation

3D representation using 3D descriptors has gradually attracted attention in recent years. The extended 3D fingerprint (E3FP) [14] extends ECFP to 3D space for molecular conformation. Some researchers [15] created a probabilistic graphical model that encoded semantic and geometric relationships among shape components. Due to the recent successes in point cloud processing [16,17], some works try to use point cloud to represent protein molecules [18]. Vote3Deep [19] proposed "voting" in convolutions for the sparse point cloud input. These methods mainly aim at biomacromolecules with a large number of atoms, and each atom is represented as a point in 3D space.

Some works introduce the voxel-based model for various downstream tasks [20]. 3D ShapeNets [21] proposed a volumetric representation in the form of binary voxels with a convolutional Deep Belief Network. VoxNet [22] introduced the volumetric occupancy grid representation for 3D classification tasks. OctNet [23] used unbalanced octrees...
for hierarchical partitioning of the space. However, these methods are limited to the intensive computational cost caused by the exponential growth of the number of parameters, the voxel-based models have not yet been widely applied.

3. Voxel-based Molecular Representation

In this section, we begin by introducing the baseline voxel-based representation of molecular conformation. Firstly, we need to select a unique coordinate system in 3D space for every molecule to avoid ambiguity. We compute the geometric center of each molecule as the origin and apply principal components analysis (PCA) to extract three axes of this coordinate system. Then we translate the molecule into the origin and rotate it for the extracted orientations. Finally, we discretize the 3D space into a regular grid with a fixed cell size to make sure that every voxel (3D cell) contains no more than one atom. We represent each voxel of the grid as a one-hot vector that has at most one non-zero entry, where each component of this vector corresponds to the presence of specified atomic type in the exact voxel.

Fig. 2 Several voxel-based representations of CNC₃OH₉ while H atoms are not shown. Atoms are colored using the CPK notation: C = gray, O = red, N = blue, and so on.

To make sure that a voxel contains up to one atom, we use 0.5Å as the cell size because it’s small than the minimal distance between any pairs of atoms. The whole voxel space would be 16Å × 16Å × 16Å so that each molecule can be set into the space. In conventional discrete voxel-based molecular representation, each atom occupies just one voxel where the center of the atom is located. Fig. 2(a) shows a sample molecule CNC₃OH₉ by using discrete representation, where different colors show different atomic types based on CPK notation.

Although the discrete voxel-based 3D representation of molecule captures structural properties as well as geometrical properties of the molecule, it contains a lot of empty space around atoms, which means the severe sparsity in 3D space. Moreover, this representation cannot reveal the difference between atomic types. Since we use QM9 dataset [24] as our evaluation dataset in this paper, which contains more than 130, 000 molecules, we make statistics about the volume of molecules and numbers of atoms each molecule contains in this dataset. The detailed statistical result is shown in Table 1.

Table 1 Statistics of molecules in QM9 dataset. vol. means volume of molecule, i.e. the number of voxels occupied by molecule, while num. means number of atoms in the molecule.

<table>
<thead>
<tr>
<th>vol. num.</th>
<th>(0, 2090]</th>
<th>(2090, 4180]</th>
<th>(4180, 6270]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 10]</td>
<td>834</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(10, 20]</td>
<td>93,097</td>
<td>12,803</td>
<td>31</td>
</tr>
<tr>
<td>(20, 30]</td>
<td>12,301</td>
<td>14,751</td>
<td>68</td>
</tr>
</tbody>
</table>

It can be found that relative to the number of voxels occupied by each molecule, the number of atoms contained is less than 1%, nearly 0.5% while using the discrete version of representation. Since each atom occupies only one voxel, less than 1% of voxels contain non-zero vectors, and the remaining voxels are zero vectors. High data sparsity may cause under-fitting during the training process since the propagated gradients turn out to be sparse. A conventional strategy on this problem is to apply Gaussian smoothing [25] by convolving the original 3D model with a Gaussian kernel defined as

$$k_{Gauss}(x, y, z) = \exp\left(-\frac{x^2 + y^2 + z^2}{2\sigma^2}\right)$$  \hspace{1cm} (1)$$

The convolution operation here is defined as
\[
\hat{X}_{\text{Gaus}}(p) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} X(p + \delta_p)k_{\text{Gaus}}(\delta_p)
\]  
(2)

where \( p(x, y, z) \) is a point in 3D space, and \( \delta_p = (\delta_x, \delta_y, \delta_z) \) is the offset of \( p \). Parameter \( \sigma \) is used to control the diffusion range of atom. With larger \( \sigma \), the atom spreads to a larger area of voxel space, which means that more voxel vectors have non-zero values. Fig. 2(b) and (c) shows different illustrations with different \( \sigma \). This Gaussian smoothing leads to an exponentially decaying sphere for every atom that fills multiple voxels. The vectors corresponding to these voxels no longer have the one-hot property. The same voxel may be affected by multiple different atoms. Although using Gaussian smoothing in voxel-based representation can reduce the sparsity problem, there is still a large gap between the actual spatial structure, and it still cannot effectively distinguish atoms of different atomic types.

4. VisMole: Visual Molecular Representation

We first discuss how to generate the visual representation for an atom following the microscopic spatial structure.

According to the theory of quantum physics, as the component of a molecule, an atom consists of an atomic nucleus \( p_0 \) at the center and several electrons moving around the atomic nucleus. Compared with the atom, the volume of the nucleus is so small that can be ignored in the 3D model. Each electron is distributed in different orbits outside the nucleus according to its energy level, thus forms a spherical area around the nucleus, which is called the electron cloud. The density of the electron cloud indicates the probability of electrons appearing in different regions. When a point coordinate is used to indicate the position of an atom, it often refers to the position of the nucleus, i.e. the position of the center of the atom. When several atoms are combined into molecules, the motion states of electrons in the atom are affected by other atoms to form chemical bonds, thereby affecting the properties of molecules. In other words, by expressing the distribution of electrons in atoms, the 3D representation can capture the relationship between atoms in a molecule.

Table 2 Percentage of atomic types in QM9 dataset. The first row is the percentage of each atomic type in all atoms in the dataset, and the second row is the percentage of molecules contain every atomic types.

<table>
<thead>
<tr>
<th>type</th>
<th>H</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>atoms pct. (%)</td>
<td>51.09</td>
<td>35.16</td>
<td>5.80</td>
<td>7.81</td>
<td>0.14</td>
</tr>
<tr>
<td>molecules pct. (%)</td>
<td>99.96</td>
<td>100.00</td>
<td>61.89</td>
<td>85.10</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Following the previous voxel-based representation, we represent the atom as a spherical region, which is the diffusion range of the electron cloud. The radius of this region is mainly affected by two factors: the number of electron layers and the number of nuclear charges. Generally speaking, the more electron layers, the smaller the nuclear charge, and the larger the range of electron diffusion, that is, the larger the atomic radius. Since the QM9 dataset used in this paper contains only five atomic types: H, C, N, O, and F, and the electron orbitals of these atomic types do not exceed two layers, we only consider two layers of electron orbitals in this paper. Table 2 gives more detailed information about these atomic types in QM9 dataset.

According to quantum physics description of molecular microscopic physical structure, We use the following function [26] to describe two layers of electron orbits outside the nucleus. For more layers, it can be expressed by extending this function.

\[
\Psi_1(p) = \frac{1}{\sqrt{\pi}} \left( \frac{Z}{a_0} \right)^\frac{3}{2} e^{-\frac{Zr}{a_0}} 
\]  
(3)

\[
\Psi_2(p) = \frac{1}{4\sqrt{2\pi}} \left( \frac{Z}{a_0} \right)^\frac{1}{2} (2 - \frac{Z}{a_0}r) e^{-\frac{Zr}{2a_0}} 
\]  
(4)

where \( r = \|p, p_0\| \) is the distance between \( p \) and the nucleus \( p_0 \). \( Z \) is the number of nuclear charge of atom, and \( a_0 \) is a constant number. \( \Psi_i(p) \) describes the distribution law of electron orbitals in the i-th layer. This function reflects the characteristics of the electron orbital, that is, the inner orbital is narrower and the outer is wider. In addition,
the function also reflects the shrinking effect caused by the increase in the number of nuclear charges on the electron orbitals of each layer.

The kernel of wave transform is the combination of every electron orbitals of the atom, as shown in the following equation.

\[ k_{\text{VisMole}}(p) = 4\pi r^2 \sum_{i}^{l} \Psi_i^2 \]  

(5)

where \( l \) is the number of layers of electron orbitals. In this paper we apply \( l = 2 \) since we only consider two layers of electron orbitals. Furthermore, to avoid the electron cloud spread to infinity, we clip points that were further than \( 2 \cdot a_0 \) away from the nucleus. Different from the simple exponential decay brought by the Gaussian kernel, this wave kernel based on the electron orbital reflects the physical structure of the atom, which is achieved by taking into account the influence of the electron and nuclear charge at the same time.

Next, we apply convolution operation through the entire voxel space to generate molecular representation. We use the wave transform kernel to perform convolution operation channel by channel with 3D window size \( h \times h \times h \), where \( h = 2 \cdot a_0 \), as shown in the following equation.

\[ \hat{X}_{\text{VisMole}}(p) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} X(p + \delta_p) k_{\text{wave}}(\delta_p) \]  

(6)

Since this convolution operation only operates one channel at a time, it does not mix information between different atomic types which lie in different channels. The interference between different channels will be further combined in the subsequent learning network.

Fig. 2(d) shows an illustration of the resulting visual molecular representation, which is named as VisMole. It can be seen that the proposed VisMole significantly improves the effect of molecular representation by simulating the atomic microscopic structure. By modeling the influence of the number of nuclear charges and electron layers on the atomic level, this VisMole not only expresses the difference between different atomic types but also reflects the relationship between atoms, to achieve a more accurate representation of molecules.

5. Experimental Results

5.1. Data and Settings

In this paper, we uses QM9 dataset [24] for our experiments. QM9 dataset contains the geometric information and physical properties of 133,885 molecules. For each molecule in the dataset, we predict its 8 properties, e.g. dipole moment(\( \mu \)), isotropic polarizability(\( \alpha \)), electronic spatial extent(\( R^2 \)), etc. Because the range of these properties is varied, the arctan operation is applied to normalize all property values. The dataset is divided into the training set, validation set, and testing set according to the ratio of 8 : 1 : 1. We use Drug3DNet [27] as our baseline. It combines the spatial feature and the channel feature from different atomic types to improve the performance of predicting molecular properties.

During the training process, the batch size is 8, the learning rate is 1e-2, and the optimizer is AdaDelta. We also apply batch normalization and dropout (ratio=0.5) to avoid overfitting. Use MAE to evaluate the regression performance of the network. All experiments are carried on CentOS 7 with Intel® Xeon® CPU E5-2620 v4 @ 2.10GHz and 2 NVIDIA GeForce GTX 1080 Ti GPUs (2GB memory in total).
5.2. Normalization of molecular size

Table 3 Comparison of different normalization strategies, using MAE as the evaluation. The properties shown in the table are dipole moment ($\mu$), electronic spatial extent ($R^2$), energy of HOMO ($\epsilon_{HOMO}$), energy of LUMO ($\epsilon_{LUMO}$), gap between $\epsilon_{HOMO}$ and $\epsilon_{LUMO}$ ($\Delta \epsilon$), heat capacity at 298.15K ($C_v$), zero point vibrational energy (ZPVE), and isotropic polarizability ($\alpha$).

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$R^2$</th>
<th>$\epsilon_{HOMO}$</th>
<th>$\epsilon_{LUMO}$</th>
<th>$\Delta \epsilon$</th>
<th>$C_v$</th>
<th>ZPVE</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>naïve</td>
<td>0.0440</td>
<td>0.7076</td>
<td>0.4104</td>
<td>0.0427</td>
<td>0.0317</td>
<td>0.0758</td>
<td>0.0233</td>
<td>0.2570</td>
</tr>
<tr>
<td>exp</td>
<td>0.1659</td>
<td>20.6434</td>
<td>0.4107</td>
<td>0.0430</td>
<td>0.0398</td>
<td>1.9927</td>
<td>0.0210</td>
<td>0.0101</td>
</tr>
<tr>
<td>ratio</td>
<td>0.0445</td>
<td>20.6346</td>
<td>0.4107</td>
<td>0.0672</td>
<td>0.0400</td>
<td>0.0580</td>
<td>0.0201</td>
<td>0.0091</td>
</tr>
</tbody>
</table>

As mentioned above, the input size of each voxel is restricted. However, the size of some molecules is much smaller than others along one or more coordinate axes directions. It is necessary to explore the necessity of performing a normalization process and find out which strategy to choose.

In this part, we follow three different normalization strategies and evaluate their performances in property prediction tasks. The naïve strategy directly fills the empty voxels to the desired input size with zero vectors. It does not apply any scaling to original molecules. Both the other two strategies which are named "exp" and "ratio" rescale the original molecule according to the actual size of small molecules. The former uses an exponent of 2 to enlarge the size of a small molecule along every coordinate axis direction, and then fills the remaining empty voxels with zero vectors. The latter first calculates the minimum enlargement ratio of the three coordinate axis directions, rescales the entire molecule with this ratio, and then fills in space with zero vectors.

The comparison result is shown in Table 3. It seems that the effect of every normalization strategy is similar, while the performance of the naïve version is a little better than the others. The result shows that although there is a lot of empty space around the naïve representation of molecules, it’s better than rescaling the size of molecules. As Fig. 3 shows, normalization would enlarge the distance between atoms so that the connection and the relationship would be reduced. In other words, normalization may lose the information embedded in the bonds. We can also see that the ratio version is better at predicting some properties than the exp version. We believe that it’s better to normalize all the molecules than to enlarge small molecules.

![Fig. 3](image_url)

(a) naïve  (b) exp  (c) ratio

Fig. 3 Different normalization effects for molecule C4NOH9. Each color corresponds to one atomic type as the same notation as in Fig. 2.

5.3. Comparison of diffusion method

In the second experiment, we compare the different voxel-based 3D molecular representations. The first representation is the naïve discrete voxel-based 3D representation, each atom only occupies the voxel at its center. The following cases use Gaussian smoothing to diffuse the atomic range. We use different $\sigma$ as we have mentioned in the previous section to explore the effect of this diffusion parameter. We also compare another wave-based atomic diffusion method [28] for molecular representation. This method also replaces the Gaussian convolution kernel, however, its new kernel neither effectively reflects the atomic spatial structure. We randomly selected several molecules from the dataset to generate
their voxel-based representations in 3D space by using different diffusion methods. Fig. 4 illustrates the effects of these molecules. The experimental results are shown in the Table 4.

Table 4 Comparison of different atomic diffusion methods with the same baseline, using MAE as the evaluation. The properties shown here are the same as those in Table 3.

<table>
<thead>
<tr>
<th>network</th>
<th>diffusion</th>
<th>$\sigma$</th>
<th>$\mu$</th>
<th>$R^2$</th>
<th>$\epsilon_{\text{HOMO}}$</th>
<th>$\epsilon_{\text{LUMO}}$</th>
<th>$\Delta\epsilon$</th>
<th>$C_v$</th>
<th>$\text{ZPVE}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug3DN et</td>
<td>-</td>
<td>-</td>
<td>0.1660</td>
<td>0.7509</td>
<td>0.4108</td>
<td>0.0670</td>
<td>0.0407</td>
<td>0.0734</td>
<td>0.0132</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>gauss</td>
<td>1</td>
<td>0.1653</td>
<td>1.2312</td>
<td>0.4110</td>
<td>0.0478</td>
<td>0.0436</td>
<td>0.0797</td>
<td>0.0243</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.0585</td>
<td>1.1970</td>
<td>0.4107</td>
<td>0.0462</td>
<td>0.0417</td>
<td>0.0719</td>
<td>0.0237</td>
<td>0.0093</td>
</tr>
<tr>
<td></td>
<td>wave[28]</td>
<td>1</td>
<td>0.1657</td>
<td>1.2456</td>
<td>0.4107</td>
<td>0.0477</td>
<td>0.0444</td>
<td>0.0774</td>
<td>0.0242</td>
<td>0.0103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.0600</td>
<td>1.2283</td>
<td>0.4106</td>
<td>0.0481</td>
<td>0.0433</td>
<td>0.0783</td>
<td>0.7022</td>
<td>0.0108</td>
</tr>
<tr>
<td>VisMole</td>
<td>-</td>
<td>-</td>
<td>0.0440</td>
<td>0.7076</td>
<td>0.4104</td>
<td>0.0427</td>
<td>0.0317</td>
<td>0.0638</td>
<td>0.0233</td>
<td>0.2570</td>
</tr>
</tbody>
</table>

Compared with the original network, the results show that we improve the performance on some of the properties after using the diffusion strategies and get better by changing the $\sigma$. The diffusion ‘gauss’ and ‘wave’ get similar improvements on most of the properties. The reason could be that they have similar kernels and focus only on diffusion while neglecting the microscopic information in the molecules. The molecules have their own physical rules and they cannot be calculated by simply applying a Gaussian kernel. Since our method is based on the microscopic physical structure of molecules, the functions we use can represent the electron distribution around the nucleus, so that we can describe the molecules as they are in the real world. The result in the table shows that our diffusion strategy performs better than other diffusion strategies.
Fig. 4 Different diffusion effects for several molecules. The four columns correspond to the discrete version, the gaussian version with 
1 as $\sigma$, the gaussian version with 4 as $\sigma$, and the VisMole version respectively. Each color corresponds to one atomic type as the same 
notation as in Fig. 2. Notice these examples and example in Fig. 2 are isomers.

6. Conclusion

Since deep learning technology has been widely used in molecule-related tasks these days, molecular representation has 
been a popular topic in deep learning in chemistry. However, the representation for the non-crystalline small molecule is 
still an open problem since its arbitrary microscopic spatial structure. In this paper, we propose a voxel-based 3D 
representation that conforms to the microscopic spatial structure for non-crystalline small molecules, which helps to 
understand and analyze molecules by visual perception. Existing representations for non-crystalline small molecules 
generally lacked precise descriptions of the microscopic spatial structure, which limited the performance of molecule-
related tasks. Although the Gaussian smoothing can reduce the sparsity problem, it’s just a rough approximation of the 
distribution of electrons in 3D space. Our proposed kernel based on quantum physics describes the distribution of 
electrons in a better way. The evaluations support that it can carry out more information, which is important for chemistry 
tasks, such as properties prediction, drug creation and so on.

Acknowledgments

This work was financially supported by National Key R&D Program of China (2021YFB2600600), Beijing Education 
Commission for General Project of Science and Technology Plan (KM202111232003), Promoting the classified 
development of universities - key research and cultivation projects (No.2121YJPY225) and Innovation capacity building 
of scientific research institutions-Institute of Data Science.

References

Reviews 116(10), 6107–6132.


Weld defect recognition method based on improved DenseNet

LI Huadu*a, LUO Renzea, TANG Xiangb, WU Yongc, LI Yalongc

*aFaculty of Electrical Information, Southwest Petroleum University, Chengdu, Sichuan 610500, China;
bFaculty of Science and Technology, Southwest Petroleum University, Chengdu, Sichuan 610500, China;
cSichuan Jiacheng Oil and Gas Pipeline Quality Inspection Ltd, Chengdu, Sichuan 610000, China

* Corresponding author: 2595796081@qq.com

ABSTRACT

There are many subjective influencing factors, poor recognition effect and low efficiency in manual evaluation of pipeline weld defects. An intelligent identification method of pipeline weld defects based on improved DenseNet network is proposed. This method firstly uses the form of multi-channel convolution of different scales to improve the DenseNet network, thereby improving the generalization ability of the network; Then, the feature extraction ability of the network is improved by stacking two convolutions of the same scale; Finally, an attention mechanism module is introduced into the dense connection block of the network to achieve the effect of improving beneficial features and suppressing useless features. The experimental results show that the method can achieve 92% accuracy in the identification of pipeline weld defects, which is about 13% higher than the original method, and has high efficiency, which can fully achieve the purpose of industrial application.

Key words: pipe weld defects; DenseNet networks; attention mechanism; defect identification: recognition accuracy

1. INTRODUCTION

With the wide application of pipelines in petroleum, natural gas and other industrial fields, it is necessary to further improve the existing pipeline weld defect identification methods. Most of the oil and gas pipelines are welded by straight pipes and elbows, and the generated welds are the high-incidence part of pipeline defects. Therefore, the defect detection of welds is an important guarantee for the safe operation of pipelines [1]. The traditional manual evaluation method of pipeline welds is time-consuming, laborious, inefficient, and greatly influenced by human subjective factors, which can no longer meet the increasing needs of pipeline weld defect detection.

The automatic identification method of pipeline weld defects based on deep learning [2-12] can effectively solve the problems of time-consuming labor and low efficiency of manual film evaluation. Liu Han et al. [13] used convolutional neural network to automatically extract deep-level features of defects, which not only realized automatic classification, but also improved the time-consuming and laborious problem of manual feature extraction; Jiang Hongquan et al. [14] used random forest algorithm to improve the feature selection method of CNN model, and improved the parameter learning ability of traditional CNN method; Daniel Bacioiu et al. [15] used the fully connected network architecture combined with convolutional neural network to identify 6 types of defects, 4 types of defects, and 2 types of defects, with accuracy rates of 70%, 85%, and 95%, respectively, it can be seen that the classification accuracy of the model decreases significantly in the face of multiple defects.

Aiming at the problems of poor generalization ability and single extraction feature parameters of existing deep learning methods, this paper proposes a pipeline weld defect identification method based on multi-channel and multi-scale convolution DenseNet network. This method firstly reduces the loss of network feature parameters with the characteristics of DenseNet structural feature reuse; Secondly, the network generalization ability is improved through multi-channel; Thirdly, the diversity of model feature parameters is improved by using different scales of convolution; Finally, stacked convolution is used to combine The attention mechanism method improves the network feature extraction ability. The results show that the method effectively solves the shortcomings of the original deep learning defect identification methods, and improves the generalization ability and parameter diversity of the model.
2. DENSENET NETWORK

DenseNet is a convolutional neural network that optimizes information transfer by reusing features [16]. Its model structure is shown in Figure 1. The input defect image are transferred to Denseblock after Conv(n×n)+BN+ReLU+Avgpooling operation. In Denseblock, each layer and the output of all previous layers will be merged into the input of the next layer in the channel dimension [17]. If the Denseblock has $L$ layers and the input image is $X_0$, it contains $\frac{1}{2}L(L+1)$ connections. The input to the layer is as follows:

$$X_l = H_l([X_0, X_1, \cdots, X_{l-1}])$$  \hspace{1cm} (1)

In the formula, $X_l$ is the input of layer $L$; $H_l$ is the nonlinear function of the BN+ReLU+Conv(n×n) structure; $[X_0, X_1, \cdots, X_{l-1}]$ is the output from layer 0 to layer $l$ of the network.

After the dense connection, the feature map is transferred to the next dense block through the Transition layer, as follows:

$$X_{l+1} = T_l([X_0, X_1, \cdots, X_l])$$  \hspace{1cm} (2)

In the formula, $X_l$ is the input of the next Denseblock, $H_l$ represents the Transition layer of the BN+ReLU+Conv(1×1)+Avgpooling structure, and $[X_0, X_1, \cdots, X_{l-1}]$ is the output of the previous Denseblock.

The final output feature map is then normalized. After the activation function and the average pooling operation, the recognition result is output through the fully connected layer.

3. IMPROVE DENSENET NETWORK

The existing pipeline weld defect recognition algorithm has a weak global perception effect of the convolution kernel, so the extracted feature map information is relatively simple, which leads to a poor recognition effect of the model and a weak generalization ability.

In view of the shortcomings of existing methods, the model proposed in this paper adopts a multi-channel multi-scale convolution network structure based on DenseNet. Using the characteristics of multi-channel and multi-scale convolution, the model can obtain receptive fields of multiple scales and features of different scales, which effectively improves the recognition accuracy and generalization ability of the model, and changes the dilemma of its single parameter quantity. Its network structure is shown in Figure 2.
The model firstly uses the original DenseNet network structure in the first channel. The dense connection in the structure can effectively increase the amount of feature parameters extracted by the network and improve the recognition rate.

Secondly, in the branch channel, convolution stacking [18] is used to increase the feature extraction capability of the model. The principle is to use Conv(n×n)+Conv(n×n) to replace Denseblock in the original network, and use the increase of the convolution kernel to reduce the loss of parameters. The increased parameters are beneficial to the model’s generalization ability and anti-interference ability, to improve the stability of the model. Its structural principle is shown in Figure 3.

In the figure, two 3×3 convolution blocks are superimposed, and the size of the receptive field is the same as 5×5 convolution, which not only improves the feature extraction ability of the convolution layer but also enhances the nonlinear effect of the model. Its output is as follows:

\[ X_n = T_1(X_{n-1}) \]  

(3)

\( X_n \) represents the nth convolution stack; \( T_1 \) is the output of nonlinear variation.

Finally, the introduction of the convolution stack layer SE (squeeze-and-excitation) module is an attention mechanism that learns the spatial correlation and weight values between channels of the feature map [19]. After the network is subjected to the squeeze excitation operation [20], the weights of each layer of the network are automatically learned, so that the beneficial features are strengthened. Its structure is shown in Figure 4.
In the figure, $X_L$ is the input nonlinear feature matrix; $M, H, C, M', H', C'$ are the dimension information of the feature map; $X_{L+1}$ is the output feature matrix; $U$ is the feature information obtained by the $L$-th layer network after the convolution operation, which can be expressed as:

$$U = W^L \otimes f(BN(W^{(L-1)} \otimes f(BN([X_0X_1\cdots X_{L-1}]))))$$

(4)

In the formula, $W^{(\cdot)}$ is the convolution operation matrix; $\otimes$ is the convolution operation; $f(\cdot)$ is the activation function; $BN(\cdot)$ is the normalization operation; $[X_0X_1\cdots X_{L-1}]$ is the previously input feature map.

It can be seen that the SE module has a simple structure and is easy to implement. When it is applied to the dense connection blocks of the network, it can learn the feature weights of each channel and improve the model recognition effect.

### 4. EXPERIMENT AND RESULT ANALYSIS

#### 4.1 Experiment process

The experimental process is divided into data preprocessing, model selection and experimental comparison. In the experiment, firstly, the weld defect image was preprocessed and the experimental data set was made. Then, the improved model with the best experimental effect was selected by comparing the improved model with the existing weld defect identification method. The overall experimental process is shown in Figure 5.
4.2 data preprocessing

In this paper, the oil and gas pipeline weld defect images provided by Sichuan Jiacheng Oil and Gas Pipeline Quality Inspection Ltd are selected as the experimental data set. The data of two kinds of pipeline weld defects including slag inclusion and round defect are mainly analyzed. Among the types of pipeline weld defects, slag inclusion is a defect formed by residual foreign solid substances in the weld metal; round defect refers to a circular defect, with pores accounting for the vast majority, and pores are voids caused by gas dissolved in the weld metal. The image of slag inclusion before the experimental treatment is shown in Figure 6, and the image of the round defect is shown in Figure 7.

Figure 6 and Figure 7 show that the proportion of defect areas in the entire weld image is small, and there are some other influencing factors and artificial damage in the image, so it is more difficult to use the entire image to train the model. Therefore, the preprocessing of the weld image data is divided into the following steps: Firstly, the original weld image (above4000×1500pixel) is segmented to obtain image data with the size of 500×500pixels, and the specific method is to use the slide frame of 500×500pixel to segment every 500 pixels in the horizontal direction. After the horizontal segmentation is completed, move down 500 pixels to continue the segmentation operation. For the area less than 500 pixels, the next segmentation operation is directly abandoned until the segmentation of the entire weld image is completed. Secondly, the images containing defects and peripheral weld areas were selected from the segmented images. Finally, the selected weld images were classified into defects, and the obtained image data were divided into three types: slag inclusion, round defect and non-defect, as shown in Figures 8, 9, and 10.
There are a total of 2030 images obtained by segmentation and selection, using 85% of the images as the training set and 15% as the test set. The sample data of each defect images are shown in Table 1.

<table>
<thead>
<tr>
<th>Defect type</th>
<th>total number of samples</th>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>round cut</td>
<td>550</td>
<td>450</td>
<td>100</td>
</tr>
<tr>
<td>Slag</td>
<td>625</td>
<td>525</td>
<td>100</td>
</tr>
<tr>
<td>flawless</td>
<td>855</td>
<td>725</td>
<td>130</td>
</tr>
</tbody>
</table>

### 4.3 Evaluation indicators

In order to verify the recognition effect of the algorithm in this paper. In the paper, Confusion matrix, Precision, Accuracy, Specificity, Recall, F1-Score, loss function, training time, model parameters and model calculation are used as the evaluation index of model performance. The main formula of its evaluation index is as follows:

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{Accuracy} = \frac{TP}{TP + FP} \tag{6}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{7}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{8}
\]

\[
F1 \cdot \text{Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}
\]

In the formula, TP (True Positive) represents the number of samples that are actually defects and predicted to be defects, FN (False Negative) is the number of samples that are actually defects and predicted to be non-defects, and FP (False Negative) stands for the number of samples that are actually predicted as defects. TN (True Negative) is the number of samples that are actually non-defective and predicted to be non-defective.
4.4 Model selection

For the algorithm model proposed in this paper, experimental comparison is needed to obtain the optimal performance of the model. Firstly, the training effect of the model is compared with different channel number and network layer number, and the optimal network channel number and network layer number are selected. For the number of network channels, the experiment was conducted by 1-channel, 2-channel, 3-channel improved network experimental comparison. The number of layers is set to 12, the number of iterations is 50, the learning rate is set to 0.003, and the batch size is set to 20. Using the Adam optimizer, the Cross Entropy loss function is used. The comparison results are shown in Table 2.

<table>
<thead>
<tr>
<th>Number of channels</th>
<th>Accuracy</th>
<th>Loss</th>
<th>Time(s)</th>
<th>Calculated (10^8)</th>
<th>Parameter (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85%</td>
<td>0.389</td>
<td>1451.53</td>
<td>37.43</td>
<td>8.00</td>
</tr>
<tr>
<td>2</td>
<td>92%</td>
<td>0.006</td>
<td>9798.51</td>
<td>3436.89</td>
<td>32.83</td>
</tr>
<tr>
<td>3</td>
<td>88%</td>
<td>0.169</td>
<td>137058.77</td>
<td>11472.43</td>
<td>74.36</td>
</tr>
</tbody>
</table>

According to Table 2, compared with 1-channel network, the recognition accuracy of 2-channel network is improved by about 7%, the convergence degree of loss function is increased, and the model is more stable. However, the 3-channel network not only has a significant increase in the number of parameters, the amount of computation and training time, but also the recognition accuracy is decreased compared with the 2-channel network, and the model stability is weak. Therefore, considering the comparison of the results, it is more appropriate to adopt the 2-channel network as the recognition network.

According to the results of the previous experiment, the same parameters are used for experimental comparison of the 8-layer, 12-layer, and 16-layer network models in the 2-channel state. The comparison results are shown in Table 3.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Accuracy</th>
<th>Loss</th>
<th>Time(s)</th>
<th>Calculated (10^8)</th>
<th>Parameter (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>75%</td>
<td>0.139</td>
<td>5543.89</td>
<td>1178.52</td>
<td>22.43</td>
</tr>
<tr>
<td>12</td>
<td>92%</td>
<td>0.006</td>
<td>9798.51</td>
<td>3436.89</td>
<td>32.83</td>
</tr>
<tr>
<td>16</td>
<td>86%</td>
<td>0.003</td>
<td>134561.65</td>
<td>30621.08</td>
<td>157.87</td>
</tr>
</tbody>
</table>

As can be seen from Table 3, when the training time of the 12-layer network is not increased greatly, the number of parameters is 76.04m, and the recognition accuracy is about 92%. Compared with the 8-layer network, the effect is improved by 17%, and the loss function is more convergent, and the improvement of the training time, the number of parameters and the computation amount is not obvious. While the training time, the number of parameters and the computation amount of the 16-layer network are greatly increased, the recognition accuracy is reduced and the convergence of the loss function is not ideal. Therefore, the 2-channel 12-layer network is adopted as the network model in this paper. The convergence of accuracy and loss function are shown in Figure 11.

![Accuracy rate curve and loss function curve of 2-channel 12-layer network model](image)

Fig.11  Accuracy rate curve and loss function curve of 2-channel 12-layer network model

As can be seen from Figure 11, the model tends to be stable after 10 iterations, and the accuracy rate remains at about 90%, indicating that the model has strong generalization ability and can effectively avoid overfitting; the loss value is about 0.02, indicating that the model has good convergence and strong stability.
4.5 Experimental comparison

In order to better verify the advantages and disadvantages of the improved network, this experiment compares and analyzes the algorithm model in this paper with the VGG model [21], the DenseNet model [22], and the AlexNet model [23]. The learning rate is set to 0.003, the batch size is 20, and the changes of each evaluation index after the training times are 50 times are compared. The comparison results are shown in Figure 12 and Table 4.

![Comparison of experimental confusion matrices of each model](image)

**Fig.12** Comparison of experimental confusion matrices of each model

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Specificity</th>
<th>F1-measure</th>
<th>Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG</td>
<td>79.00%</td>
<td>79.97%</td>
<td>79.80%</td>
<td>89.43%</td>
<td>79.88%</td>
<td>9348.10</td>
</tr>
<tr>
<td>DenseNet</td>
<td>79.39%</td>
<td>79.83%</td>
<td>79.57%</td>
<td>89.43%</td>
<td>79.70%</td>
<td>9451.53</td>
</tr>
<tr>
<td>AlexNet</td>
<td>80.00%</td>
<td>81.27%</td>
<td>82%</td>
<td>90.33%</td>
<td>81.63%</td>
<td>9202.88</td>
</tr>
<tr>
<td>Improve DenseNet</td>
<td>92.00%</td>
<td>86.87%</td>
<td>86.50%</td>
<td>92.87%</td>
<td>86.68%</td>
<td>9798.51</td>
</tr>
</tbody>
</table>

In Figure 12, the abscissa of the confusion matrix represents the real value of the image marker, and the ordinate represents the predicted value of the model. Therefore, the larger value on the positive diagonal of the matrix, the better performance of the model and the higher accuracy. It can be seen that the improved DenseNet model has the highest recognition accuracy. Among the 330 test set images, 284 images are correctly recognized, and the round defect recognition error rate is the highest, 20 images are recognized as slag inclusions, and 4 images are recognized as non-defects. The VGG model and the DenseNet model identified 262 correct images, and the AlexNet model identified 267 correct images. It can be seen that the improved DenseNet model is the best, the overall recognition accuracy is higher, and the wrongly classified images are the reason for the higher shape similarity between defects.

It can be seen from Table 4 that the precision rate, recall rate, specificity, and F1-measure of the improved DenseNet model are 86.87%, 86.50%, 92.87%, and 86.68%, which has obvious excellent effect compared with other models. The recognition accuracy rate is as high as 92%, which is the best among all models. Compared with the VGG model, the DenseNet model, and the AlexNet model, the recognition accuracy is improved by 13%, 13.61%, and 12% respectively.
Therefore, the algorithm proposed in this paper has better recognition effect than other models.

5. CONCLUSION

Aiming at the problems of weak generalization ability and low defect identification accuracy of existing deep learning methods, a pipeline weld defect identification method based on improved DenseNet network is proposed. Firstly, the multi-channel and multi-scale convolution structure is used to improve the parameter diversity of the model, and then the stacked convolution is used to improve the feature extraction ability of the model, and finally attention mechanism is introduced to realize feature weight distribution and improve its generalization ability. The experimental results show that the defect recognition accuracy of the model is 92%, which is about 13% higher than that of deep learning methods in recent years, and can meet the standard of industrial application. However, due to the lack of defect types in the current welding seam data, more types of defect data will be collected in the later research to fully demonstrate the multi-defect identification effect of the model.

REFERENCES


Design of vulnerability scanning method for video surveillance system

Yan Yu1a*, Mingyang Li1b*, Yixin Zou1c, Yuanyuan Luo1d
1 China Telecom Research Institute, Shanghai 2100127, China

a* yuy37@chinatelecom.cn, b* limy11@chinatelecom.cn, c zouyixin@chinatelecom.cn, d luoyy9@chinatelecom.cn

Abstract

Recently, video surveillance system network security incidents have occurred frequently, all communication levels of video surveillance systems face severe information security risks and have many vulnerabilities. So a vulnerability scanning system of video surveillance device is proposed. This paper presents the architecture of vulnerability scanning system, describes the process of detection model, researches on the verification of vulnerability, and finally carries out an experiment of the vulnerability scanning system.

Keywords-Video surveillance system; Vulnerability scanning; PoC database; device security

1. Introduction

With the huge-scale deployment of video surveillance systems in public, diverse intelligent surveillance devices are deployed in more and more scenarios including smart homes, smart stores, smart farming, intelligent transportation [1], etc. These surveillance devices can not only capture real-time video and audio streaming, but also support a variety of intelligent analysis functions including face recognition, license plate recognition, pet recognition, behavior recognition, etc., generating massive sensitive data, and bringing convenience to people’s lives.

While most of these surveillance devices are deployed in public environments and telecommunication networks, involving a large amount of video and image data information including personal privacy information, are vulnerable to the physical hijacking (like private connect to physical interface, lens occlusion, lens shift, etc.) and cyber-attacks [2]. These factors have made video surveillance devices a prime target for hackers in recent years. More and more hackers try many kinds of attack methods, such as the brute force of weak password, Cross Site Script Attack (XSS) [3], sensitive file tampering, process privilege escalation, and even use hijacked video surveillance devices to launch Distributed denial of Service Attack (DDoS) resulting in huge data leakage [4],[5], in the meantime, with the recent frequent discovery of the third-party application software vulnerability in the video surveillance devices, more and more attack methods are exposed, which will also lead to large-scale security incidents and data leakage.

Video surveillance systems have great security risks, such as vulnerabilities in the system itself, vulnerabilities in device configuration and negligence in management. There are four main types of vulnerabilities, as shown in Table 1:

Table 1 Vulnerability classification and types of video surveillance system

<table>
<thead>
<tr>
<th>Classification</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction Vulnerability</td>
<td>Weak password</td>
</tr>
<tr>
<td></td>
<td>Cross Site Request Forge</td>
</tr>
<tr>
<td></td>
<td>Remote command injection</td>
</tr>
<tr>
<td></td>
<td>Password information disclosure</td>
</tr>
<tr>
<td>Firmware Vulnerability</td>
<td>User privilege promotion</td>
</tr>
<tr>
<td></td>
<td>Buffer Overflow Vulnerability</td>
</tr>
<tr>
<td></td>
<td>Anonymous user authentication bypass</td>
</tr>
<tr>
<td></td>
<td>Hard coded credential</td>
</tr>
<tr>
<td></td>
<td>Execute arbitrary code</td>
</tr>
<tr>
<td></td>
<td>Path traversal</td>
</tr>
<tr>
<td>Protocol Vulnerability</td>
<td>Hyper Text Transfer Protocol (HTTP) authentication bypass</td>
</tr>
<tr>
<td></td>
<td>Access via Real Time Streaming Protocol (RTSP)video stream</td>
</tr>
<tr>
<td></td>
<td>Use hard coding to bypass RTSP</td>
</tr>
</tbody>
</table>
For the problems above, a vulnerability scanning system based on video surveillance device is proposed. This system is mainly using dynamic analysis, homology analysis and optimum detection to discriminate the vulnerabilities of video surveillance devices.

2. Architecture of Vulnerability Scanning System

The Architecture of vulnerability scanning system consists three main parts: device acquisition, detection model, vulnerability scanning and analysis, as shown in Fig. 1.

Device acquisition is responsible for collecting identification and feature information from video surveillance devices, due to firmware version, software version, SDK and other information of the devices produced by different manufacturers, it is necessary to identify and pre-process the device information to form a standardize feature database.

Detection model is the processing of detection tasks. The vulnerability scanning system does not need to invoke the entire detection database and detection methods. According to the digital modelling results of video surveillance device, necessary detection items can be formed and useless detection items can be discarded. The priority of detection subtasks can be dynamically adjusted to improve the efficiency of vulnerability analysis. The detection model includes homology analysis and homology vulnerability mining. The homology analysis includes model extraction of source code, feature extraction of historical vulnerabilities and device feature, and finally analyzing the vulnerabilities according to the consistency of vulnerability identification and device identification.

Vulnerability scanning mainly judges whether target host has vulnerabilities through the following two methods [6]. After port scanning, we can know the ports opened by the target host and the network services on these ports, then match these relevant information with the vulnerability library provided by the vulnerability scanning system to check whether there are vulnerabilities that meet the matching conditions; Or scan the target host system for offensive vulnerabilities by simulating the attack tactics of hackers. If the simulated attack is successful, it indicates that there are vulnerabilities in the target host system.
Finally, the upper layer serves as human-computer interaction interface, display vulnerability scanning results and form scanning reports, including the port scan results, vulnerability name, vulnerability level, description of vulnerability, impact of vulnerability, and repair suggestions.

The key parts of vulnerability scanning system are vulnerability database and PoC database. According to the case analysis and actual test experience of security experts on vulnerabilities of video surveillance system, and rule-based verification technology, a set of standard detection queues is automatically formed. By invoking the PoC module and the fuzz testing module, the vulnerability scanning system can automatically scan device vulnerabilities.

3. Experimental of vulnerability scanning system

The experimental environment of vulnerability scanning system is divided into hardware environment and software environment. The hardware environment includes processors, memory, adapters, and other standard PC components. The software environment is CentOS operating system and can be deployed in the cloud. This vulnerability scanning system has more than 200 test items of PoC database, taking into account multiple environments or platforms and multiple common versions of vulnerable applications.

Input the IP address of the video surveillance device in the detection parameter configuration interface, and automatically perform the scanning task. The report produced after scanning is shown in Fig.2.

<table>
<thead>
<tr>
<th>IPC 192.168.1.108</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>5000</td>
</tr>
<tr>
<td>554</td>
</tr>
<tr>
<td>8086</td>
</tr>
</tbody>
</table>

(a) Port scanning

(b) Vulnerability distribution

<table>
<thead>
<tr>
<th>Vulnerability Name</th>
<th>http-slowloris denial of service vulnerability exists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vulnerability Level</td>
<td>High</td>
</tr>
<tr>
<td>Vulnerability Description</td>
<td>This slow attack is very similar to http flood, which consumes the resources of the server, including the database server. However, there are differences between them. Slow attack is to maintain the connection with the server, so that the connection of the normal client cannot be established, so as to achieve the purpose of denial of service.</td>
</tr>
<tr>
<td>Vulnerability Impact</td>
<td>Since the web server has a certain upper limit on the number of concurrent connections, if these connections are maliciously occupied and not released, all connections of the web server will be occupied by maliciously connected connections, so that new requests cannot be accepted, resulting in a denial of service.</td>
</tr>
<tr>
<td>Repair Suggestion</td>
<td>For the defense against these slow attack, you can simply modify some parameters according to its characteristics, including: limiting the number of connections of a single machine IP, limiting the timeout time of http request header/body and so on.</td>
</tr>
</tbody>
</table>
This scan report shows that there are 4 ports opened and the service type of each port, such as HTTP, RTSP, Universal Plug and Play (UPnP), Distributed Service Network (DSN). This scan result also found that there are 4 high risk vulnerabilities, 10 medium risk vulnerabilities and 21 low risk vulnerabilities. Therefore, this video surveillance system condition is dangerous. This scanning report contains detailed information about the vulnerability. The user can decide how to deal with the scanned vulnerability according to whether the vulnerability level is high enough. Through the vulnerability description information, the impact of the vulnerability can be understood. The vulnerability impact describes the specific security problems that may be caused by the vulnerability in detail. The vulnerability repair suggestion provides a positive repair method for this vulnerability, users can timely handle and repair the vulnerabilities in the system according to the repair suggestion to ensure the system security.

The experiment results show that the vulnerability scanning system designed in this paper can scan the vulnerability of the video surveillance devices efficiently and quickly reflect the scanning result to the user.

### 4. Conclusion

We have proposed and experimentally demonstrated a vulnerability scanning method for video surveillance system. By forming a set of vulnerability database and PoC database for video surveillance system, combining our self-developed vulnerability detection module, vulnerabilities can be effectively recognized. The scanning report will be generated after the vulnerability scanning system is completed, and the security status will be listed in details. Our research results have been applied in the safety detection of video surveillance devices, which greatly improve the detection efficiency and ensure the security of video surveillance system. Further study will be in-depth detection methods to ensure more accuracy of vulnerability scanning, expand the PoC verification method, and collect more comprehensive vulnerability data.

<table>
<thead>
<tr>
<th>Vulnerability Name</th>
<th>hidden dangers in ONVIF configuration exists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vulnerability Level</td>
<td>Medium</td>
</tr>
<tr>
<td>Vulnerability Description</td>
<td>Open network video interface forum (ONVIF) is a global open industry forum. Its goal is to promote the development and use of global open standards for secure product interfaces based on physical IP. ONVIF has created a standard for how IP products in video surveillance and other physical security fields communicate with each other.</td>
</tr>
<tr>
<td>Vulnerability Impact</td>
<td>This vulnerability allows attackers to freely call relevant interfaces to obtain device information and device entry points.</td>
</tr>
<tr>
<td>Repair Suggestion</td>
<td>Add protocol authentication for ONVIF.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vulnerability Name</th>
<th>MiniUPnP MiniUPnPd up to 2.1 pcp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vulnerability Level</td>
<td>Low</td>
</tr>
<tr>
<td>Vulnerability Description</td>
<td>Miniupnp is a set of UPnP tools of miniupnp project that can be used for embedded systems. This tool can connect the devices in the home network and the company network. Miniupnpd is the daemon of miniupnp. Miniupnp pcserver in miniupnpd version 2.1 There is a code problem vulnerability in copyIPv6IdDifferentof pcserver.c file. This vulnerability originates from improper design or implementation in the code development process of the network system or product.</td>
</tr>
<tr>
<td>Vulnerability Impact</td>
<td>CVE-2019-12111</td>
</tr>
<tr>
<td>Repair Suggestion</td>
<td><a href="https://github.com/miniupnp/miniupnp/commit/cd506a67e174a45c6a202eff182a712955ed6df0f">https://github.com/miniupnp/miniupnp/commit/cd506a67e174a45c6a202eff182a712955ed6df0f</a></td>
</tr>
</tbody>
</table>
References


Entity extraction based on the parts of speech attention mechanism

Jie Xu*a, Lijun Wangb, Jing Xub, Huan Heb, Jiaying Lia, Jingru Liaoa

*aUniversity of Electronic Science and Technology of China, 2006 Xiyuan Avenue, Chengdu, China; bThe 29th research institute of CETC, Chengdu, China.

* Corresponding author: xuj@uestc.edu.cn

ABSTRACT

Entity extraction is an information extraction technique that aims to locate and classify named entities (e.g., organizations, locations, persons...), which is a very important and fundamental problem in natural language processing. On the research of entity extraction, numerous models ignore the learning of grammatical structure. Considering the shortcomings of previous models, this paper first proposes the PALC (POStag-Attention-LSTM-CRF) model, which adds POS (part of speech) features to entity extraction. Specially, PALC fuses POS features with other features through a multi-layer bidirectional LSTM network and attention mechanism to improve the effect of entity extraction. The experimental results show that the accuracy of the PALC model in this paper on the CONLL03 dataset can be 90.65%, on the CONLL03 dataset can be 84.86%, and on OntoNote 5.0 English dataset can be 86.99%.

Keywords: entity extraction, part of speech, multi-layer bidirectional LSTM, attention mechanism

1. INTRODUCTION

Artificial intelligence technology has carried out a lot of research and application in NLP (Natural Language Processing)\(^1\), image\(^2\), video\(^3\) and so on. Entity extraction is an important technique in the field of NLP to process unstructured data, which plays a significant role in semantic text understanding and information extraction. Due to the flexibility and complexity of unstructured data and the rapid growth of data volume, traditional NLP technologies are not qualified apparently for their inherent limitations. So in this paper, we propose a PALC (POStag-Attention-LSTM-CRF) model based on the POS attention mechanism that aims to lift the entity extraction model’s ability to learn the semantic expression of words. Furthermore, we propose a feature fusion method based on the attention mechanism and multi-layer bidirectional LSTM network to solve the problem of feature fusion in PALC.

2. ENTITY EXTRACTION MODEL

Because the importance of learning grammatical structure in entity extraction task is always ignored, we propose the PALC model, an entity extraction model composed of attention mechanism, LSTM, and CRF. In detail, the attention mechanism learns POS features, LSTM learns semantic features, and CRF learns close dependence among words. In this study, the input of the model is unstructured text and the output is the entity type of words.

2.1 Overall structure of the PALC model

In this section, we propose the PALC model based on the POS attention mechanism to finish the entity extraction task. Figure 1 shows the overall structure of the PALC entity extraction model. First, the original unstructured data need to be pre-processed. Specifically, the pre-training model performs sentence segmentation and word segmentation processing on these data, and then further obtains the word, POS, and character composition of each sentence. After that, the pre-training model is utilized to encode the obtained three parts and then get respective representation vectors.

After pretraining, POS vectors of different words in a sentence are sent to the BiRNN (Bidirectional RNN) network to learn the attributes and categories in the specific context of the sentence. Take the vector set in hidden layer of BiRNN network as the POS feature set \(P = \{p_1, p_2, p_3, ..., p_n\}\), where \(p_i\) denotes POS feature vector of the \(i\)-th word in the sentence and \(n\) refers to sequence length. Meanwhile, the characters of a word are sent to the BiLSTM (Bidirectional LSTM) network to learn the internal structure so as to obtain different morphological features of the word, such as learning the effect of word prefix on word interpretation.
Then, the morphological features of the word are spliced with the encoded word vector to generate the initial word vectors set $S = \{s_1, s_2, s_3, \ldots, s_n\}$, a character-level initialization word vector set, where $s_i$ denotes the initialization word vector of the $i$-th word. $s_i$ usually contains the morphological characteristics and the coding expression of a word.

After that, the vectors in $S$ are sent to the multi-layer BiLSTM in order of position and the corresponding hidden layer vector $h_i$ is generated. The output of LSTM is a set of hidden layers, denoted as $H = \{h_1, h_2, h_3, \ldots, h_n\}$. At this point, $h_i$ contains the morphological features and the contextual semantic features of words in the context of sentence. Then $h_i$ is sent into the attention mechanism together with POS set $P$ to learn the importance of different POS $p_i$ to $h_i$. The output vector of the attention mechanism is spliced with $s_i+1$ to continue learning in the LSTM network; then, we can get set $O = \{o_1, o_2, o_3, \ldots, o_n\}$, which is composed of the hidden layer outputs of multilayer LSTM. Finally, we use two fully connected layers and a CRF layer as the classifier model to obtain the final result of label classification.

### 2.2 POS feature learning module based on the attention mechanism

This paper proposes to introduce POS features to help obtain accurate semantic features. The use of words in different contexts reflects the POS of the word. Similarly, the POS can also express the contextual information of the word in the sentence to a certain extent. Therefore, the learning of POS features can assist in understanding the exact information of entities in sentences and better express entity information.

We use the Stanford Parser POS tagging tool to perform fine-grained POS tagging on words. The Stanford Parser provides nearly 40 kinds of POS tags. For example, POS tag “NN” means “noun, singular or mass”, POS tag “VB” means “verb, base form” and “JJ” means “adjective”.

As mentioned in Section 2.1, the POS feature learning model sequentially inputs the POS vector of each word in the sentence into the BiRNN network to get the POS feature set $P$. In BiRNN network, each neuron in the hidden layer contains context information of the POS in the sentence. Then, we use the attention mechanism to calculate the weights of the relationship between POS feature set $P$ and the hidden layer vector $h_i$ extracted at some point from LSTM. The weights indicate the importance of the POS to the semantic expression and the weight matrix is denoted as $Score$. Finally, we can obtain the POS context feature of the word by adding the vectors obtained by multiplying $Score$ and $P$. The specific module is shown in Figure 2.
Figure 2. POS feature learning module based on the attention mechanism

Figure 3 shows the calculation method of Score mentioned above. As shown in Figure 3, before splicing $h_i$ with $P$, the dimensions of $h_i$ need to be transformed to be the same as the dimensions of $P$ to get the matrix $M$ with the same dimensions of $P$. Then the weight matrix $Score$ can be calculated as follows:

$$ T = v^T \tanh(W^T M) $$

$$ Score = \text{soft} \max(T) $$

As the formula (1) describes, $M$ is mapped by a fully connected layer to become the vectors with the same dimensions of $P$ and then input into $\tanh$ activation function to get new vectors. Next, we initialize a learnable parameter matrix $v^T$ and multiply the new vectors to get the matrix $T$. To some extent, matrix $T$ can express the influence of different POS on semantics. After softmax function, the weight matrix $Score$ is obtained from $T$.

![Figure 3. Calculation method of weight matrix](image)

2.3 Improved multilayer BiLSTM

When fusing POS and other features, it is necessary to explore the relationship between $h_i$ and $P$. At this time, $h_i$ should be able to fully express the context information of the sentence and we need to obtain the semantic expression of words with POS features in the learning process. However, the traditional BiLSTM is not applicable. Instead, a backward LSTM layer needs to be added. Given that, this paper proposes an improved multilayer BiLSTM as shown in Figure 4.

First, we input the initialization vector set $S$ into a backward LSTM ($b_1$) to obtain its hidden layer vector set $H_s = \{ h_{s1}, h_{s2}, h_{s3}, ..., h_{sn} \}$. $h_i$ indicates the context information of the $i$-th word in the sentence. The input of forward LSTM ($f_1$) is the word vector $q_i$ integrated with the POS feature as well as backward LSTM ($b_2$). $q_i$ is obtained by POS fusion, which needs to fuse POS features with the initial word vector $s_{i+1}$. We can get a new input $q_{i+1}$ as follows:

$$ q_{i+1} = \text{concat}(\sum_{j=1}^{n} Score \times P, s_{i+1}) $$

(3)
where \textit{concat} represents the splicing of vectors and \( q_1 \) is formed by splicing \( s_1 \) and an initialized zero vector and the weight matrix \textit{Score} is obtained in the attention mechanism.

After completing the forward LSTM\( (f_1) \), backward LSTM \( (b_2) \) is needed to learn the following information after integrating POS features. Take out the hidden layer vectors of forward LSTM\( (f_1) \) and backward LSTM\( (b_2) \) and splice them to form the set \( O \), which is used as the input of the subsequent fully connected layer.

![Multilayer BiLSTM model](image)

Notably, in forward LSTM \( (f_1) \), the last hidden layer of BiRNN (depicted in Figure 1) in the POS feature extraction is spliced as the hidden vector of the LSTM \( (f_1) \) input. The last hidden layer vector contains the initial understanding of the whole sentence in terms of POS, which can be used as external information to help subsequent learning.

### 3. EXPERIMENT AND ANALYSIS

#### 3.1 Dataset and Experimental Settings

We validate our PALC model on three public datasets including the CONLL03\([4]\), CONLL02\([5]\), and OntoNote 5.0 English dataset\([1]\). Since POS feature labeling in these datasets is incomplete, we use the open source tool Stanford Parser to annotate the POS.

About the parameter settings, the experiment uses glove as the encoding dictionary; hence, the dimension is 300 after glove encoding. The vector dimension of character encoding is 25. The vector dimension of POS coding is 25. The dimension of the hidden layer in multilayer BiLSTM is 100. The dimension of the hidden layer in morphological feature extraction network LSTM is 50 and the dimension of hidden layer in the POS feature extraction network is 50. The learning rate of the model is 0.002. The rate of neurons that stop working is 0.2.

#### 3.2 Entity Extraction Results

Table 1 summarizes the experimental results of the existing models proposed by other scholars in recent years and the PALC model on the three datasets and shows that our model performs better than the other state-of-the-art models.

Theoretically, each model has its limitations. The CNN-CNN-LSTM model ignores the order of input to some extent. The IE model assumes that partial data label information is incomplete to learn and then obtains the linear chain CRF of missing labels. Similar to the IE model, AdaPU constructs some data without labels, which reduces the learned information. Compared with the other methods, PALC adopts LSTM and RNN to extract features and pays more attention to the sequence of input, so the PALC model has more complete and reasonable theoretical advantages.
Table 1. F1-scores of Different Models on Entity Extraction.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CONLL03</th>
<th>CONLL02</th>
<th>OntoNote 5.0 English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D-CNN</td>
<td>90.54</td>
<td>-</td>
<td>84.24</td>
</tr>
<tr>
<td>CNN-CNN-LSTM</td>
<td>90.69</td>
<td>-</td>
<td>86.15</td>
</tr>
<tr>
<td>IE</td>
<td>89.50</td>
<td>82.00</td>
<td>-</td>
</tr>
<tr>
<td>AdaPU</td>
<td>82.94</td>
<td>75.85</td>
<td>-</td>
</tr>
<tr>
<td>EDG</td>
<td>88.60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PALC</td>
<td><strong>90.84</strong></td>
<td><strong>84.86</strong></td>
<td><strong>86.99</strong></td>
</tr>
</tbody>
</table>

From the perspective of the dataset, the PALC model is based on the CONLL03 training set, on which it has the best performance. For the CONLL02 dataset, the language is Spanish and the PALC model considers the learning of word transformation in grammar structure which is ignored by the IE nor AdaPU model. Conceivably, both of them are not good at cross-language datasets. The F1 score of the PALC model in CONLL03 and CONLL02 are higher than that of IE model and AdaPU model. Although the 1D-CNN model achieves comparable performance with PALC on CONLL03 dataset, PALC focuses on the effect of lexical transformation on the results in syntax. Consequently, the PALC model has better performance on the OntoNote 5.0 English dataset. The EDG model is based on the attenuation error to sample uncertainty, which can accelerate the convergence of the results to some extent, but there is still a gap compared with the PALC model. According to the experimental results, we have verified the effectiveness and validity of PALC in the task of entity extraction.

3.3 Analysis of PALC Model Ablation Experiment

To further study the influence of POS on extraction results, this experiment starts from the basic model, named LSTM-CRF, to explore the effects of the introduction of different modules on extraction results. We study the results of different models on the CONLL03 test set to verify the effectiveness of the introduction and improvements of POS features. In detail, different models use different types of features and different networks to conduct experiments.

Table 2. Results of Different Methods to Improve the Basic Model on the CONLL03 Test Set

<table>
<thead>
<tr>
<th>Model</th>
<th>Semantic Features</th>
<th>POS Features</th>
<th>Morphological Features</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM-CRF(W)</td>
<td>√</td>
<td>-</td>
<td>-</td>
<td>86.03</td>
<td>86.22</td>
<td>85.99</td>
</tr>
<tr>
<td>LSTM-CRF(WC)</td>
<td>√</td>
<td>-</td>
<td>√</td>
<td>89.42</td>
<td>89.23</td>
<td>89.32</td>
</tr>
<tr>
<td>PALC(POS LSTM)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>90.22</td>
<td>89.78</td>
<td>89.99</td>
</tr>
<tr>
<td>PALC(POS RNN)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>90.65</td>
<td>91.06</td>
<td>90.84</td>
</tr>
<tr>
<td>PALC</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>90.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the results of different models on the CONLL03 test set and all the experimental results are reproduction results. The most basic model, LSTM-CRF(W), only uses the semantic features of word learning in text. Based on the LSTM-CRF(W) model, the LSTM-CRF(WC) model introduces the BiLSTM network to learn the morphological features of words in advance and then extracts semantic features, which reduces the granularity of learning. Therefore, compared with the LSTM-CRF(W), the LSTM-CRF(WC) model shows significant improvement in all indicators.

The POS feature can represent the categories and attributes of words in context more accurately. On the basis of LSTM-CRF(WC) model, PALC (POS LSTM) and PALC (POS RNN) respectively adopt BiLSTM and BiRNN network to learn the POS feature. According to the experimental results, RNN has more advantages than LSTM in learning the POS feature in this task. The possible reason is that the conversion of POS features often depends on close dependence and RNN has advantages in short distance relationship learning. However, the LSTM network is good at capturing long-distance dependence and often learns some long-distance information. When the long-distance information becomes noise, it can interfere with the result judgment.

Considering the effectiveness of RNN in POS feature learning, we advance our PALC model based on the PALC (POS RNN). When fusing the POS feature with other features, both PALC (POS LSTM) and PALC (POS RNN) models use a simple splicing method where each word only stitches its own POS, ignoring the influence of the POS of other words. However, our PALC model adds the attention mechanism and multilayer BiLSTM to consider the global influence of the POS of all words. Experiments show that PALC outperforms the other models on the CONLL03 test set.
4. CONCLUSION

In this paper, we propose a novel PALC entity extraction model combining the attention mechanism. The PALC model introduces the word label and makes judgment on the type and boundary of an entity from the aspects of word, morphology, and semantics so as to improve the precision of entity extraction. The experimental results show that the PALC model achieves good performance in entity extraction task.

ACKNOWLEDGEMENTS

This work was supported by CETC Key Laboratory of Electromagnetic Domain Operation; Natural Science Foundation of China (Grant No.62071098); Sichuan Science and Technology Program (Grant Nos. 2021YFG0307, 2022YFG0319).

REFERENCES

A Fast Point Clouds Registration Method with ISS-Shot Feature Descriptors

Shuo Xu¹a, Jun Liu¹b, Hao Cai¹c*
¹The School of Computer Science and Artificial Intelligence, Wuhan Textile University, Hubei 430200, China
ashuoxushuo@yeah.com, bjunliuwtu@yeah.net, c*2019048@wtu.edu.cn

Abstract

With the rapid development of high-precision sensors such as LiDAR and Kinect, point cloud has become the main data format to represent the three-dimensional world. And point cloud registration is an important step in the processing of these data. However, there are problems that wrong corresponding points, low registration accuracy and slow speed in the process of point cloud registration. We propose a fast point cloud registration method based on ISS-Shot feature descriptors, SVD algorithm, and an improved ICP algorithm. The proposed method firstly uses a voxel grid filter for down-sampling. Next, the feature points are extracted by the ISS algorithm and described by the SHOT algorithm. Afterwards, the ISS-Shot features are used for rough registration with SVD algorithm. Finally, the ICP algorithm with normal vector constraints and KD tree acceleration is used for accurate registration. The experimental results show that the proposed algorithm has higher registration accuracy and faster registration speed than the compared algorithms.

Keywords- Point clouds registration; intrinsic shape signatures (ISS); signature of histograms of orientations (Shot); singular value decomposition (SVD); iterative closest point (ICP)

1. Introduction

Point cloud registration is a key problem for computer vision applied to SLAM [1], medical imaging [2], and other applications [3]. This problem involves finding a rigid transformation from one point cloud into another so that they align. Traditional point cloud registration is to establish correspondence between point clouds by using local similarity features, then transform estimation is performed based on the matching point pairs, and ICP [4] algorithm is used to further refine the registration results. However, due to the lack of representativeness or the number of feature points, these methods often lead to low registration accuracy. In recent years, with the discrimination and representation ability of deep learning, the deep point cloud registration method has gained more and more research interest. But these methods rely on a large number of high-performance GPU training models, which take a long time. To achieve high efficiency along with high accuracy, we propose a fast point cloud registration method with ISS-Shot feature descriptors. And our main contributions are:

- We introduce a new point cloud registration method with corresponding point pairs matched by ISS-Shot feature descriptor and then SVD coarse registration. Use ISS-Shot feature to better describe points. It can get enough, and representative feature points for efficient matching. Coarse registration with SVD which can get a useful transformation matrix and fine registration with ICP algorithm which used normal vector constraint and KD tree acceleration. This structure further improves the registration accuracy.
- Our method is insensitive to the relative initial position of the point cloud. Through experiments, it is found that our method is more effective for point cloud registration with large initial positions.
2. Related work

The "coarse to fine" strategy [5] has been used in point cloud registration, and the work is completed in two steps. First, establish the correspondence between the two point clouds to estimate a coarse registration; second, carry out fine-tuning through the fine registration algorithm.

2.1. Coarse registration

Coarse registration algorithms can be divided into [6] based on global features and [7] based on local features. Global descriptors encode object geometry. Global feature descriptors such as VFH, CVFH, OUR-CVFH, ESF, GFPFH, GRSD, SSC, and PCAN, are calculated for the entire cluster rather than a single point, and represent an object. And the method based on local features is more robust to interference such as partial overlap. Therefore, the point cloud registration method based on local features is mainly considered in this paper. Local descriptors are used for object recognition and registration to describe the local geometry around a point. There are many local descriptors in the open-source point cloud library PCL, like PFH, FPFH, RSD, 3DSC, USC, Shot, the Spin Image, RIFT, NARF and so on.

2.2. Fine registration

Fine registration algorithms can be divided into iterative nearest point algorithms and normal distribution transformation algorithms. The iterative nearest point (ICP) algorithm does not need to estimate position or extract features. Compared with ICP, 3D-NDT [18] does not need to calculate the nearest neighbour matching point for less computational complexity.

3. Proposed method

The flow chart of the proposed method is shown in Figure 1 and mainly consists of five steps. Firstly, the voxel grid filter is used to down-sample the point clouds, which can reduce registration time. Secondly, feature points are extracted from the down-sampled point cloud with the ISS algorithm. Next, the extracted feature points are described using the Shot algorithm to form ISS-Shot features. Thirdly, the ISS-Shot feature is used to coarse registration with the SVD algorithm. Finally, the ICP algorithm is used for fine registration. Furthermore, to improve the registration accuracy and efficiency, the normal vector angle is used to effectively eliminate noise point pairs, and the KD tree algorithm is used to speed up the process.
3.1. Downsampling

The voxel grid filter divides the point cloud into small cubes, each of which is a voxel. The centroid of each voxel is calculated as a point. Its neighbours instead of the voxel for simplifying the entire point cloud. It not only realizes downsampling but also maintains the original feature information.

3.2. ISS-Shot point feature

Feature points are stable and unique point sets in the image point cloud that can be extracted by defining detection criteria, and the number of feature points is far less than the original points. Compared with Harris, NARF and SIFT feature point extraction algorithm, ISS algorithm is faster and the extracted feature points are distributed more evenly. ISS feature point is a method to express the feature degree of a point by connecting with neighbourhood information and using the relationship between feature values. SHOT stands for the signature of histograms of orientations. It encodes information about the topology with a spherical support structure as same as 3DSC. The basic idea of the SHOT feature descriptor is that the local coordinate system is established based on the neighbourhood points, and divides the neighbourhood space of the points into several subspaces. Then, the histogram by counting on the normal feature for each point in the subspace is encoded. And get the 3D descriptor by combining the histograms of each subspace.

4. Experimental evaluation

The experiment dataset contains “Bunny”, “Dragon”, and “Buddha” from Stanford University Graphics Lab and “Chair”, “Car”, “Piano” from the basic geometry library.

$RMSE$ (Root Mean Square Error) is used as one of the evaluation indexes of point cloud registration. It is a common similarity measure to evaluate the matching degree between two groups of point clouds. By using the sum of squares of distances corresponding ($f$-score) to the nearest points of the registered point cloud as the evaluation of the registration effect, the situation where the $RMSE$ value is small but the registration effect is not ideal can be avoided.

<table>
<thead>
<tr>
<th>name</th>
<th>SVD time/s</th>
<th>SVD RMSE/E-03m</th>
<th>number(P)</th>
<th>number(Q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris</td>
<td>0.4700</td>
<td>6.3075</td>
<td>277</td>
<td>243</td>
</tr>
<tr>
<td>NARF</td>
<td>0.4210</td>
<td>5.854</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>SIFT</td>
<td>1.7580</td>
<td>3.1905</td>
<td>1547</td>
<td>1431</td>
</tr>
<tr>
<td>ISS</td>
<td>0.4487</td>
<td>4.5692</td>
<td>415</td>
<td>366</td>
</tr>
</tbody>
</table>

4.1. Contrast experiment

4.1.1. Comparison test of feature point selection. Table 1 shows the experimental results of coarse registration of the down-sampled “Bunny” point cloud by four feature point extraction algorithms. We find that although the NARF algorithm takes the least coarse registration time, it extracts few feature points. And although the SIFT algorithm has the smallest $RMSE$, it extracts a large number of feature points and takes a long time for coarse registration. Likewise, although the number of feature points extracted by Harris algorithm is moderate, its coarse registration $RMSE$ is the largest. In addition, Figure 4 visualizes the 4 feature points of the “Bunny” point cloud after the down-sample. As shown in Figure 2(a), the Harris algorithm extracts the wrong feature points. Therefore, we use the ISS algorithm to extract feature points.

![Figure 2. Different feature points about Harris, NARF, SIFT, and ISS of the “Bunny” P_ds and Q_ds.](https://example.com/image.png)
Table 2 presents the registration results of the four feature points under the proposed method. Experiments show that using ISS feature points in the proposed method has a better registration effect than the other three types of feature points.

Table 2. The registration results of different feature points about Harris, NARF, SIFT, and ISS under the proposed method with FPFH.

<table>
<thead>
<tr>
<th>name</th>
<th>Total time/s</th>
<th>SVD RMSE/E-05m</th>
<th>F-score/E-05m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris</td>
<td>0.570</td>
<td>7.71</td>
<td>2.27</td>
</tr>
<tr>
<td>NARF</td>
<td>0.626</td>
<td>8.35</td>
<td>22.01</td>
</tr>
<tr>
<td>SIFT</td>
<td>1.812</td>
<td>7.79</td>
<td>2.25</td>
</tr>
<tr>
<td>ISS</td>
<td>0.552</td>
<td>7.62</td>
<td>2.22</td>
</tr>
</tbody>
</table>

4.1.2. Comparison experiment of feature selection of feature points. Table 3 presents the registration results of the six features of feature points under the proposed method. Among them, the registration time using the Shot feature is only 0.028s apart from the registration time using the FPFH feature, and the RMSE and F-score values are both the smallest. Experiments show that using the Shot feature of feature points in the proposed method has a better registration effect.

Table 3. The registration results of different features about PFH, FPFH, RSD, 3DSC, USC, and Shot under the proposed method with ISS.

<table>
<thead>
<tr>
<th>name</th>
<th>PFH</th>
<th>FPFH</th>
<th>RSD</th>
<th>3DSC</th>
<th>USC</th>
<th>Shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time/s</td>
<td>0.713</td>
<td><strong>0.611</strong></td>
<td>0.729</td>
<td>1.239</td>
<td>1.374</td>
<td>0.639</td>
</tr>
<tr>
<td>RMSE/E-05m</td>
<td>8.10</td>
<td>7.84</td>
<td>7.87</td>
<td>7.75</td>
<td>7.75</td>
<td><strong>7.64</strong></td>
</tr>
<tr>
<td>F-score/E-05m²</td>
<td>20.81</td>
<td>2.26</td>
<td>2.29</td>
<td>2.26</td>
<td>2.29</td>
<td><strong>1.06</strong></td>
</tr>
</tbody>
</table>

4.2. Comparative Experiment

Table 4 shows the final F-score and Time values of ICP [4], GICP [9], NDT+ICP [10], SAC-IA+ICP [11], SAC-IA+NDT [12], 3DSC+ICP [13], and our method for the six models, and the best results are highlighted in bold font. The registration results of each algorithm are shown in Figure 3. The source point cloud is green, the target point cloud is red, and the registered point cloud is blue. It can be seen from the experimental results that our method is the most effective. It can be seen from the experimental results that the ICP algorithm, the SAC-IA+NDT algorithm and the 3DSC+ICP algorithm are suitable for models with small initial pose deviations. And the F-score of three models in Table 4 (left) is much smaller than the F-score of three models in Table 4 (right) using these three algorithms and is closer to zero. And since sufficient correspondence cannot be found, the F-score value of the Car model with ICP algorithm is greater than 1, so the registration cannot be performed. Furthermore, the GICP algorithm, NDT+ICP algorithm and SAC-IA+ICP algorithm are not sensitive to the initial pose deviations of the point cloud. Throughout Table 4, the F-score obtained by registration using our method is smaller than other algorithms. In addition, the F-score obtained by our method is roughly at the same level as other algorithms for models with small initial pose deviations, and it is at least 3 levels higher than other algorithms for models with large initial pose deviations. What is more, we can see that the running time of our method is far less than that of other algorithms.

Table 4. F-score and time values of ICP, GICP, NDT+ICP, SAC-IA+ICP, SAC-IA+NDT, 3DSC+ICP, and our method for the three models with small(left) and large(right) initial displacement.

<table>
<thead>
<tr>
<th>Point</th>
<th>method</th>
<th>F-score/E-05m²</th>
<th>time/s</th>
<th>Point</th>
<th>method</th>
<th>F-score/m²</th>
<th>time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny</td>
<td>ICP</td>
<td>87.08</td>
<td>28.526</td>
<td>Chair</td>
<td>ICP</td>
<td>0.249415</td>
<td>6.853</td>
</tr>
<tr>
<td></td>
<td>GICP</td>
<td>1.17</td>
<td>4.953</td>
<td></td>
<td>GICP</td>
<td>0.0129826</td>
<td>3.534</td>
</tr>
<tr>
<td></td>
<td>NDT+ICP</td>
<td>14.85</td>
<td>2.648</td>
<td></td>
<td>NDT+ICP</td>
<td>9.99E-13</td>
<td>1.944</td>
</tr>
<tr>
<td></td>
<td>SVD-IA+ICP</td>
<td>6.66</td>
<td>3.427</td>
<td></td>
<td>SVD-IA+ICP</td>
<td>8.81E-06</td>
<td>26.569</td>
</tr>
<tr>
<td></td>
<td>SVD-IA+NDT</td>
<td>2.99</td>
<td>3.503</td>
<td></td>
<td>SVD-IA+NDT</td>
<td>0.00112474</td>
<td>79.636</td>
</tr>
<tr>
<td></td>
<td>3DSC+ICP</td>
<td>2.30</td>
<td>7.185</td>
<td></td>
<td>3DSC+ICP</td>
<td>0.013732</td>
<td>4.069</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td><strong>1.06</strong></td>
<td><strong>0.776</strong></td>
<td></td>
<td>Ours</td>
<td><strong>6.08e-14</strong></td>
<td><strong>1.243</strong></td>
</tr>
<tr>
<td>Dragon</td>
<td>ICP</td>
<td>4.39</td>
<td>4.909</td>
<td>Car</td>
<td>ICP</td>
<td>1.39343</td>
<td>1.308</td>
</tr>
<tr>
<td></td>
<td>GICP</td>
<td>2.96</td>
<td>1.299</td>
<td></td>
<td>GICP</td>
<td>0.0653315</td>
<td>16.649</td>
</tr>
<tr>
<td></td>
<td>NDT+ICP</td>
<td>7.19</td>
<td>1.636</td>
<td></td>
<td>NDT+ICP</td>
<td>0.000326436</td>
<td>5.922</td>
</tr>
<tr>
<td></td>
<td>SVD-IA+ICP</td>
<td>2.88</td>
<td>4.158</td>
<td></td>
<td>SVD-IA+ICP</td>
<td>0.0114426</td>
<td>85.027</td>
</tr>
<tr>
<td></td>
<td>SVD-IA+NDT</td>
<td>4.34</td>
<td>3.857</td>
<td></td>
<td>SVD-IA+NDT</td>
<td>0.00226523</td>
<td>3.355</td>
</tr>
<tr>
<td></td>
<td>3DSC+ICP</td>
<td>1.12</td>
<td>6.765</td>
<td></td>
<td>3DSC+ICP</td>
<td>0.00849583</td>
<td>3.442</td>
</tr>
<tr>
<td>Ours</td>
<td>ICP</td>
<td>GICP</td>
<td>NDT+ICP</td>
<td>SVD-IA+ICP</td>
<td>SVD-IA+NDT</td>
<td>3DSC+ICP</td>
<td>Ours</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>1.12</td>
<td>0.370</td>
<td>0.00014123</td>
<td>0.540</td>
<td></td>
<td>0.33</td>
<td>0.271</td>
</tr>
<tr>
<td>Happy Buddha</td>
<td>4.10</td>
<td>4.911</td>
<td>6.16</td>
<td>0.65</td>
<td>0.79</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Piano</td>
<td>ICP</td>
<td>GICP</td>
<td>NDT+ICP</td>
<td>SVD-IA+ICP</td>
<td>SVD-IA+NDT</td>
<td>3DSC+ICP</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td>0.503377</td>
<td>1.76E-10</td>
<td>5.35E-05</td>
<td>0.000522504</td>
<td>0.000786614</td>
<td>0.0173907</td>
<td>2.39E-13</td>
</tr>
</tbody>
</table>

Figure 3. Registration results of each algorithm.

5. Discussion and Conclusions

In this article, we proposed a fast point cloud registration based on ISS-Shot feature descriptors, SVD algorithm and an improved ICP algorithm. It is specifically described that extracting feature points using ISS algorithm, describing feature points by SHOT algorithm, then getting a good initial pose by SVD algorithm, and fine registration with normal vector constraints and KD tree acceleration in order to achieve better registration results and reduce time cost. More importantly, this method is insensitive to the relative initial position of the point cloud. Experiments show that our method has high registration accuracy, small error and high efficiency. However, the adaptive threshold of some parameters and the construction of the covariance matrix in this method still needs to be further optimized.

References


Accuracy analysis and development of multi-channel acquisition device for domestic Intelligent Programmable Logic Controller

Qiang Liu*, Yining Wang
(NARI Group (State Grid Electric Power Research Institute) Co. Ltd., Nanjing 211106, Jiangsu Province)
(*Corresponding author: Qiang Liu E-mail: liuqiang1@sgepri.sgcc.com.cn)

Abstract

In this paper, an multi-channel acquisition device suitable for domestic Intelligent Programmable Logic Controller (IPLC) is developed. The GD32F407RE of GigaDevice is used as MCU control system and the AD chip SGM5208 of SGMICRO is used. The upper computer based on Loongson CPU realizes data interaction with the acquisition device through CAN communication. In the experiment, the precision current and voltage source is used to verify and test the developed multi-channel acquisition device. The device can realize the acquisition and transmission of DC analog 0-5V / 0-20mA , achieving a measurement accuracy of 0.1%. The developed multi-channel acquisition device achieves reliable operation in engineering applications. The influencing factors of measurement accuracy and real-time are analyzed, and the software is optimized. The device is designed based on domestic electronic components, which ensures the safety of the chip supply chain.

Key words: Localization; Multi-channel acquisition; CAN communication; Accuracy analysis

1. Introduction

The Intelligent Programmable Logic Controller (IPLC) is gradually applied in intelligent manufacturing, robotics and other fields[1-3]. Recently the wave of chip application’s localization, the development of domestic PLC is of great importance to solve the key problem of chip supply chain, and is of great significance to the intelligent manufacturing industry. It can ensure the national chip supply chain’s safety and self-control. The domestic semiconductor chip industry has developed vigorously in recent years, and many excellent domestic chip companies have emerged. The chip performance of MCU chips designed by GigaDevice, analog digital chips designed by SGMICRO and other domestic companies has reached or exceeded the same type of chips abroad, which provides a great possibility for domestic alternative implementation. In this paper, a domestic multi-channel acquisition device for the Intelligent Programmable Logic Controller (IPLC), which is based on GigaDevice’s MCU chip GD32F407 and SGMICRO’s AD chip SGM5208 is designed.

2. Basic Principle

The Intelligent Programmable Logic Controller (IPLC) can be called industrial computer, including a CPU module and many I/O modules. Through the series of IO modules, the switching value, analog value and temperature value are collected and transmitted to the CPU module through the backplane bus for data exchange. The Intelligent Programmable Logic Controller (IPLC) is widely used in many industrial fields, including intelligent manufacturing, robots, etc. The multi-channel acquisition device used for current and voltage acquisition in this paper can be used to acquire 0-5V voltage signal or 0-20mA current signal converted by the sensor.
As shown in Figure 1 above, the multi-channel acquisition device has two working modes: current sampling and voltage sampling. IV conversion is required for current sampling, and then it is sent to AD acquisition after channel selection. The data is transmitted to MCU through SPI. The MCU realizes data interaction with the upper computer CPU through the CAN bus of the backplane.

### 2.1 I/V conversion

I/V conversion uses a 250 Ω resistor to convert 0-20mA current signal into 0-5V voltage signal. In order to realize the difference between current and voltage acquisition modes, mode selection is carried out in channel input mode and software design.

### 2.2 Channel selection

The multi-channel acquisition device adopts the time slice polling mechanism to realize the data acquisition of the eight channel, and realizes the real-time acquisition of channel n (n = 1, 2, 3, 4, 5, 6, 7, 8) by MCU controlling the charge and discharge of the fit capacitor.

### 2.3 MCU part

As shown in Figure 2 below, the MCU adopts GigaDevice company’s GD32F407RE. The chip adopts arm cortex-m4 core, external clock frequency: 4MHz ~ 32MHZ, working voltage range: 2.6V ~ 3.6V, program flash capacity: 512KB, RAM capacity: 192KB, number of CAN channels: 2, number of GPIO ports: 51, working temperature range: -40 °C ~ +85 °C. As the main control unit of the multi-channel acquisition device, the MCU micro control module uses the 3.3V power interface to supply power. The MCU controls the AD acquisition timing through the SPI interface, communicates with the upper computer CPU through the backplane CAN bus, and sends the measured values of the collected analog signal.

### 2.4 AD sampling

The multi-channel acquisition device adopts the SGMICRO company’s AD chip SGM5208, with a sampling rate of
500kSPS, 14 bit chip and SPI interface, which can sample analog signals at high speed and facilitate the implementation of filtering algorithm.

The circuit design principle of direct sampling of AD chip SGM5208 is as follows. The capacitor is charged by the input analog signal (the capacitor is on the field side), the voltage held on the capacitor is completely disconnected from the field side through the switching of the solid-state relay. Then the signal is input to SGM5208 through the voltage follower composed of SGM8965A-2 operational amplifier, which is converted into a digital signal for MCU to read, so as to obtain the magnitude of the input analog. SGM8965A-2 operational amplifier is a domestic operational amplifier of SGMICRO company. Its gain bandwidth product reaches 50MHz, and can realize rail to rail input and output with low distortion. It can be seen from the above that the 0-20mA current signal is finally converted into the 0-5V voltage signal. Therefore, a 5V reference source chip is selected and its output voltage is 5V.

2.5 CAN communication mechanism

CAN bus adopts two-wire serial communication mode, with strong error detection ability, and can work in high noise interference environment. With priority and arbitration functions, multiple control modules are connected to the CAN bus through the CAN controller to form a multi host local network. It has the advantages of strong real-time, long transmission distance, strong anti electromagnetic interference ability and low cost [5-6].

As an important part of domestic intelligent programmable logic controller (IPLC), unified communication protocol is needed in software design. The communication flow between the CPU module and the multi-channel acquisition module is as follow. The communication is always initiated by the CPU and the multi-channel acquisition module responds. No response is required for the time setting message.

1) Power up the CPU module, initialize the program, and read the configuration of the multi-channel acquisition module.

2) Load the IO module, the CPU module sends the loading message (CODE_IOLOAD), and the multi-channel acquisition module responds (CODE_ACK).

3) For calling measurement, the CPU module sends a calling measurement message (CODE_IOPOLL), and the multi-channel acquisition module responds with a response message (CODE_AIVAL).

4) IO module time setting: the CPU module broadcasts time setting to the multi-channel acquisition module. The time setting command is sent once a minute, and the time setting command has no response message.

Once the multi-channel acquisition module is successfully loaded, the CPU module repeats step 3). In step 3), if the CPU module does not receive the measured value or response of the multi-channel acquisition module, it is considered that the communication with the multi-channel acquisition module has failed, and the loading message for the module should be sent again. The communication described in step 3) of the module is not resumed until the multi-channel acquisition module has correctly responded.

2.6 Upper computer software design

It has designed the upper computer software for hardware configuration and data monitoring for the Intelligent Programmable Logic Controller (IPLC) using Visual C++ 6.0, as shown in Figure 3 below.
3. Experimental results and analysis

3.1 Experimental configuration

As shown in Figure 3 above, a power module, a CPU module and a developed multi-channel acquisition module are configured in the experiment. AI0009 ~ AI0016 can be configured with current type or voltage type acquisition mode. In order to facilitate the upper computer to process the collected analog signals, 0~5V voltage signals or 0~20mA current signals are mapped into 0~20000 code values in the software. Several groups of experimental measurements were carried out on the current and voltage acquisition functions and accuracy of the eight channels respectively. The experimental results of the group with larger error were recorded as shown in Table 1 and 2 below.

Table 1: Experimental results of 0-5V voltage measurement

<table>
<thead>
<tr>
<th>Vin</th>
<th>MeVal</th>
<th>Ch1</th>
<th>Ch2</th>
<th>Ch3</th>
<th>Ch4</th>
<th>Ch5</th>
<th>Ch6</th>
<th>Ch7</th>
<th>Ch8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1V</td>
<td>4000</td>
<td>3993</td>
<td>3997</td>
<td>3998</td>
<td>4000</td>
<td>3994</td>
<td>3996</td>
<td>3997</td>
<td>3999</td>
</tr>
<tr>
<td>2V</td>
<td>8000</td>
<td>7995</td>
<td>7998</td>
<td>7990</td>
<td>8000</td>
<td>8001</td>
<td>7993</td>
<td>7992</td>
<td>7996</td>
</tr>
<tr>
<td>3V</td>
<td>12000</td>
<td>11993</td>
<td>11992</td>
<td>11994</td>
<td>12001</td>
<td>12000</td>
<td>11994</td>
<td>11993</td>
<td>11995</td>
</tr>
<tr>
<td>4V</td>
<td>16000</td>
<td>15992</td>
<td>15990</td>
<td>15995</td>
<td>15994</td>
<td>16000</td>
<td>15997</td>
<td>15995</td>
<td>15993</td>
</tr>
<tr>
<td>5V</td>
<td>20000</td>
<td>19991</td>
<td>19997</td>
<td>19994</td>
<td>19995</td>
<td>19992</td>
<td>19996</td>
<td>19995</td>
<td>20000</td>
</tr>
<tr>
<td>Maximum error value</td>
<td>-0.45</td>
<td>-0.50</td>
<td>-0.50</td>
<td>-0.25</td>
<td>-0.40</td>
<td>-0.35</td>
<td>-0.40</td>
<td>-0.35</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vin</th>
<th>MeVal</th>
<th>Ch1</th>
<th>Ch2</th>
<th>Ch3</th>
<th>Ch4</th>
<th>Ch5</th>
<th>Ch6</th>
<th>Ch7</th>
<th>Ch8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1V</td>
<td>4000</td>
<td>3993</td>
<td>3997</td>
<td>3998</td>
<td>4000</td>
<td>3994</td>
<td>3996</td>
<td>3997</td>
<td>3999</td>
</tr>
<tr>
<td>2V</td>
<td>8000</td>
<td>7995</td>
<td>7998</td>
<td>7990</td>
<td>8000</td>
<td>8001</td>
<td>7993</td>
<td>7992</td>
<td>7996</td>
</tr>
<tr>
<td>3V</td>
<td>12000</td>
<td>11993</td>
<td>11992</td>
<td>11994</td>
<td>12001</td>
<td>12000</td>
<td>11994</td>
<td>11993</td>
<td>11995</td>
</tr>
<tr>
<td>4V</td>
<td>16000</td>
<td>15992</td>
<td>15990</td>
<td>15995</td>
<td>15994</td>
<td>16000</td>
<td>15997</td>
<td>15995</td>
<td>15993</td>
</tr>
<tr>
<td>5V</td>
<td>20000</td>
<td>19991</td>
<td>19997</td>
<td>19994</td>
<td>19995</td>
<td>19992</td>
<td>19996</td>
<td>19995</td>
<td>20000</td>
</tr>
<tr>
<td>Maximum error value</td>
<td>-0.45</td>
<td>-0.50</td>
<td>-0.50</td>
<td>-0.25</td>
<td>-0.40</td>
<td>-0.35</td>
<td>-0.40</td>
<td>-0.35</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Experimental results of 0-20mA current measurement

<table>
<thead>
<tr>
<th>Cin</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
<th>MeVal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0mA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4mA</td>
<td>4000</td>
<td>3992</td>
<td>3993</td>
<td>3994</td>
<td>3996</td>
<td>3994</td>
<td>3998</td>
<td>3993</td>
</tr>
<tr>
<td>8mA</td>
<td>8000</td>
<td>7993</td>
<td>7994</td>
<td>7999</td>
<td>8000</td>
<td>7992</td>
<td>7994</td>
<td>7993</td>
</tr>
<tr>
<td>12mA</td>
<td>12000</td>
<td>11994</td>
<td>11993</td>
<td>11996</td>
<td>12001</td>
<td>11993</td>
<td>11996</td>
<td>11994</td>
</tr>
<tr>
<td>16mA</td>
<td>16000</td>
<td>15993</td>
<td>15992</td>
<td>15995</td>
<td>15991</td>
<td>16000</td>
<td>15993</td>
<td>15995</td>
</tr>
<tr>
<td>20mA</td>
<td>20000</td>
<td>19990</td>
<td>19993</td>
<td>19995</td>
<td>19991</td>
<td>19994</td>
<td>19992</td>
<td>20000</td>
</tr>
</tbody>
</table>

Maximum error value

<table>
<thead>
<tr>
<th>Ch1</th>
<th>Ch2</th>
<th>Ch3</th>
<th>Ch4</th>
<th>Ch5</th>
<th>Ch6</th>
<th>Ch7</th>
<th>Ch8</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.50‰</td>
<td>-0.40‰</td>
<td>-0.25‰</td>
<td>-0.45‰</td>
<td>-0.40‰</td>
<td>-0.35‰</td>
<td>-0.40‰</td>
<td>-0.35‰</td>
</tr>
</tbody>
</table>

3.2 Analysis of experimental results

Each channel of the multi-channel acquisition device is tested by the precision current voltage signal source in different select modes. According to the data analysis in Table 1 and Table 2, each channel’s measurement accuracy error of the multi-channel acquisition device is less than 1 ‰, which meets the requirements for high-precision analog signal acquisition in intelligent manufacturing, robotics and other industrial fields. The calculation algorithm in the Table 1 and 2 is dividing the maximum error value by the full scale value to obtain the accuracy error. The accuracy error of the multi-channel acquisition device depends on the sampling accuracy of the AD chip, the output accuracy of the 5V voltage reference source and the capacitor’s charging and discharging timing control based on the time slice rotation. In addition, in order to eliminate the pulse error, the signal is collected in both the forward and reverse directions’ acquisition timing control in the software, and then the error is eliminated by the average algorithm of 7 times. From the experimental results, the error has reached a high accuracy.

4. Conclusion

In this paper, a multi-channel acquisition device for the Intelligent Programmable Logic Controller (IPLC) is designed, which realizes the domestic substitution from basic electronic components such as capacitors and resistors to the MCU and AD chip. The designed multi-channel acquisition module has realized the acquisition of eight channel’s analog current and voltage signal. The module has passed through four level static electricity, four level surge, four level fast transient and other electromagnetic compatibility (EMC) experiments. The measurement accuracy is less than 1 ‰. The developed multi-channel acquisition device can be used for intelligent manufacturing and precision control of robots based on analog control. The domestic hardware design ensures the supply safety of multi-channel acquisition device’s chips, and also provides ideas for the domestic replacement of other IO modules such as switching value and temperature value that are applicable to the Intelligent Programmable Logic Controller.

Acknowledgement

This project has been supported by the science and technology project of State Grid Electric Power Research Institute (Project No.: 5246C5220002). I hereby express my thanks.

Reference

Research on new recognition method of web component based on machine learning

Kaiming Yang, Tianyang Zhou, Junyi Wang

ABSTRACT

The security of web components is increasingly concerned by the industry, and identifying web components is of great significance to both network attack and defense. Aiming at this problem, we propose a novel identification method of web components based on machine learning. We transform the Web component identification problem into a classification problem. On this basis, we further construct the framework for realizing this method, and implement the framework. The experimental results show that the application of this method can have a good recognition effect on Web components.

Keywords: web components, machine learning, feature selection, ensemble learning, classification

1. INTRODUCTION

With the maturity of website construction technology, a large number of web components have entered the cyberspace. Broadly speaking, web servers, databases, web containers, web frameworks, and web plugins all belong to web components. Web components are used by many website builders due to their convenience and functionality. But Web components also bring security risks, and more and more network attacks are carried out against the components. As long as you master the vulnerability of a certain component and find out the websites that use this component in the cyberspace, you can make these websites fall completely. Therefore, accurate identification of web components is increasingly important for network security.

At present, the mainstream Web component identification method is feature matching. Based on the component fingerprint database, the method performs feature matching on website data, and further analyzes the matching results to obtain the Web component identification results. Based on this mainstream identification method, there are mainly two solutions.

(1) Method 1: Only strong fingerprints with high recognition accuracy are considered. When such fingerprints appear, it is determined that the website has applied the components corresponding to the strong fingerprints. Typical representatives include tools such as BlindElephant and Httppint. The advantages and disadvantages of this method are obvious. The advantage is that the Web component identification process is simple and efficient. The first disadvantage is that it cannot cope with the situation that the component features are modified, resulting in a low recognition accuracy. The second disadvantage is that the role of weak fingerprints is ignored, thus reducing the completeness of component identification.

(2) Method 2: A linear calculation model is introduced. This method sets a confidence level for each fingerprint of each component. When judging whether the target website applies a certain component, match the analysis result of the target website with the fingerprint of the component, and linearly add the confidences of the successfully matched fingerprints as the probability P of judging whether to apply the component. If P is greater than or equal to 1, it is determined that the component is applied; otherwise it is determined that the component is not applied. Compared with the first method, the second method has obvious advantages. First of all, the second method considers the situation that the component features may be modified, so that the strong fingerprint is also given confidence, which can improve the recognition accuracy. Secondly, method 2 utilizes weak fingerprints with low recognition accuracy, which can improve the completeness and accuracy of component recognition. However, method 2 still has some limitations. The first
disadvantage is that in practical applications, the setting of confidence is often based on expert experience or statistics on a small number of samples, which is not accurate. The second disadvantage is that the fitting effect of the simple linear calculation model on the component identification problem is not very satisfactory, and the accuracy of the identification still needs to be improved.

To address the deficiencies of the two existing methods, we propose a new method that introduces a machine learning approach to solve the web component identification problem. Our method does not need to set a quantitative confidence level for each fingerprint, and can make full use of the recognition function of weak fingerprints. At the same time, the fitting effect of the machine learning model is more ideal than that of the linear calculation method.

The rest of the paper is organized as follows. In Section 2, we introduce the basic theories and key technologies on which to realize the new identification method of Web components. In Section 3, we implement a novel identification method for web components. In Section 4, we conduct experiments on the previous method and analyze the experimental results. In Section 5, we summarize the full text and give a brief overview of the next steps.

2. BASIC THEORY AND KEY TECHNOLOGY

2.1 Web Component Fingerprint Design Method

At present, the mainstream method for Web component identification is to rely on the Web component fingerprint database to perform feature matching. Whatweb [6] and Wappalyzer [7] are two well-known web component recognition tools, which mainly use the html keyword features of web pages for recognition. Tools such as Test404 and Gwhatweb utilize static file hashes for identification. However, the above-mentioned Web component identification tools all face the defect of insufficient selection of feature types. Fabian Marquardt [8] performed component identification by analyzing the html document structure, but this method can only identify a small number of CMS components and is not universal. Kozina [9] used the link pattern of url to identify components, but this method has limited application scope and low identification accuracy. Therefore, we comprehensively analyze the existing Web component feature selection methods, and select 6 categories of features to build Web component fingerprints, namely: Html source keywords, static file Hash values, response header information, special file paths, static file content, other information, etc.

2.2 Feature selection methods

Feature selection can effectively reduce the complexity of the problem and improve the performance of machine learning models. Recursive Feature Elimination (RFE) algorithm [10] [11] is one of the more popular feature selection methods. By recursively reducing the size of the feature set, the computational complexity can be greatly reduced.

First, the RFE algorithm selects a classifier for feature selection and trains the classification model on the initial feature set. Second, the RFE algorithm calculates the importance measure of each feature, deletes the feature with the lowest importance measure, and forms a new feature subset. Then, the RFE algorithm iteratively performs the “training-computing-deleting” process until the set number of features or iterations is reached. Finally, the optimal feature combination is output.

Using the RFE algorithm may cause over-fitting problems, so the Recursive Feature Elimination Cross Validation (RFECV) algorithm based on cross-validation is introduced, and the resampling process is added on the basis of the RFE algorithm. The RFECV algorithm trains the classification model on the initial feature set and calculates the importance measure of n features. The algorithm sorts the features in descending order of importance, forms n feature subsets with different numbers of features respectively, and calculates the model scores that apply different numbers of feature subsets. Average the scores of the models with the same number of features generated by k resampling to obtain the number of features with the highest score. Using the number of features as the termination condition, the RFE algorithm is performed on the original data set to obtain the optimal feature set.

2.3 Ensemble learning method

Ensemble learning is an important machine learning method. Different from other common machine learning methods, ensemble learning is not a separate algorithm, its main principle is to combine multiple simple models in a certain way, and finally produce a better model. There are mainly three types of ensemble learning methods with greater influence, namely Boosting [12], Bagging [13], and Stacking [14] algorithms. The Bagging algorithm generates several training subsets in a certain random manner, and then inputs different training subsets into each base classifier, and finally, the trained base classifiers are combined into a whole. The Boosting algorithm iteratively trains the base classifier, and each
round of misclassified samples will be given a larger weight in the next round of training, and the base classifier generated in each round will be weighted and voted to combine into a strong classification device. The stacking algorithm is divided into two layers. The first layer is to generate several independent base classifiers, and the second layer is a meta classifier. The results of all the base classifiers in the first layer are used as input to train the meta classifier, thereby obtaining a combined classification model.

3. SYSTEM DESIGN

We design a novel identification method for web components based on machine learning. We transform the problem of identifying which components are applied to a target website into \( n \) problems of whether a certain component is applied to a target website. Among them, \( n \) is the number of possible application components screened out in the early fingerprint matching process. Compared with the traditional method, the robustness of the model and the accuracy of recognition have been improved.

3.1 Building a New Recognition Framework for Web Components

There are two main methods for applying machine learning models to identify web components. The first method is to establish an overall machine learning classification model for all components. The advantage of this method is that only one model needs to be trained to identify all web components. The disadvantages of this method are also obvious. Since web components are constantly updated, whenever new web components are added or the original web components are updated, the entire model must be retrained, which consumes a lot of time and computing power, and also seriously affects the practicability of this method. At the same time, the way of training an overall model to recognize Web components is effective when faced with a small number of components. However, in the face of a large number of components, the robustness of the model is reduced due to the extremely large number of features.

In view of the shortcomings of the first method, we propose another solution. We train an independent machine learning model for each web component. In view of the shortcomings of the first method, we propose another solution. We train an independent machine learning model for each web component, where \( n \) is the number of web components that may be applied. Then, the \( n \) feature vectors are respectively input into the classification model of the corresponding component, and the classification results are aggregated to obtain the Web component identification result. We transform the Web component identification problem into a set of \( n \) sub-questions whether a component is applied, as shown in Figure 1.

![Figure 1. Web Component Recognition Model](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

In the fingerprint matching stage of the two model construction methods, each fingerprint of each component in the Web component fingerprint library needs to be traversed once, and the time complexity is the same. In the model training and classification stage, the second method is equivalent to dismantling an overall model into several sub-classification models, and each sub-model can run in parallel, which greatly improves the efficiency. At the same time, when a new web component or web component update is found, only the sub-category model related to the component needs to be updated, which has better flexibility and practicability.
3.2 Construct eigenvectors

When identifying the web components of the target website, first, crawl the website content and analyze the website data. Then, the parsed data is matched with the Web component fingerprint library, and the component corresponding to the matching fingerprint is marked as a possible application component. Finally, a feature phasor is constructed for each possible applied component based on the fingerprint matching results. The dimension of the feature phasor is the number of fingerprints of the component in the fingerprint database, and each feature value of the feature phasor is the number of matches between the corresponding fingerprint and the website data. Taking the WordPress [15] component as an example, when the parsed website data is matched with the component fingerprint database, a set of feature phasors are generated according to the matching results between the website data and all relevant fingerprints of WordPress. The feature phasors are combined with labels as input data for the next step of model training or component identification, as shown in Figure 2.

3.3 Train the combined model

Combined model training is performed for each web component. The training process of our Web component classification model is implemented in two steps. The first step is to use the RFECV algorithm for feature selection, and the second step is to train the classification model based on the optimal feature subset, as shown in Figure 3.
In the feature selection stage, the performance of the RFECV algorithm depends on which estimator is used in the iterative process. With the change of the estimator, the optimal feature combination will also change. We use 6 algorithms as the base evaluator of RFECV for feature selection, including: Support Vector Machine (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Classification Regression Tree (CART), Stochastic Gradient Descent (SGD), Gradient Boosting Decision Tree (GBDT). For the training process of the Web component recognition model, in the feature selection stage, different optimal feature subsets will be obtained by using different algorithms.

In the model training stage, we selected six classifiers to train the classification model, including Decision Tree (DT), Multi-layer Perceptron (MLP), and Bayesian Regression (BR), Random Forest (RF), AdaBoost, Stacking algorithm. Among them, Random Forest, AdaBoost, and Stacking are the three representative algorithms in the ensemble learning method.

In the model training process, different feature subsets and classification methods are selected to obtain different classification models. There are also differences in performance indicators between different models. Therefore, by combining the feature selection process and the classification model training process, a combined model can be formed by combining different RFECV algorithms and classification algorithms. Through experimental verification, the combination model that is most suitable for the Web component recognition scene can be screened out.

4. EXPERIMENTAL VERIFICATION

4.1 Experimental data

In the experiment, we obtained websites with different components applied by viewing the official websites of different web components respectively. We crawl the corresponding website data and construct the "component-website data" database. Among them, from the four types of components: Web Content Management System (CMS), JavaScript Libraries, Web Frameworks, and Ecommerce Components, two components were selected as the experimental objects respectively. We crawled a total of 1755 website data for the next step of feature extraction and training the classification model. The number of websites corresponding to each component crawled is shown in Table 1.

<table>
<thead>
<tr>
<th>Component Type</th>
<th>CMS</th>
<th>JavaScript libraries</th>
<th>Web frameworks</th>
<th>Ecommerce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component Name</td>
<td>Wordpress</td>
<td>Drupal</td>
<td>JQuery</td>
<td>Lodash</td>
</tr>
<tr>
<td>Number Of Sites</td>
<td>352</td>
<td>178</td>
<td>326</td>
<td>147</td>
</tr>
</tbody>
</table>

The experiment is based on the machine learning module Scikit-Learn, and the parameters of the model are all searched by GridSearchCV. In the experiment, the K-fold cross-validation method is used to train and evaluate the model. We divide the data samples into k parts, of which k-1 is the training set and 1 is the test set. Repeat k times in sequence, where k is 5.

4.2 Evaluation indicators

We adopted the accuracy and AUC values as evaluation metrics. Accuracy (Acc) refers to the proportion of correctly predicted samples to the total number of samples. The ROC curve is often used to judge the classification effect of the model. The ordinate is the true positive rate (TPR), which represents the probability that positive samples can be correctly classified, and the abscissa is the false positive rate (FPR), which represents the false positive rate of negative samples. The probability of being classified as a positive class. The area under the curve is called the AUC value, and the larger the value, the better the classification performance of the model.

4.3 Comparison of the effects of RFECV methods based on multiple evaluators

In the feature selection stage using the RFECV algorithm, we used different estimators, which will generate different optimal feature subsets. The experiment in this section takes the WordPress component as an example. In the early stage, a total of 25 fingerprints of the component were collected as the 25-dimensional features of the component. When using the SVM algorithm as an evaluator to perform 5-fold cross-validation on the dataset, the variation of the obtained cross-validation score with the number of features is shown in Figure 4. At this time, the optimal number of features is 16.
When different estimators are used, multiple sets of optimal feature subsets with different numbers are obtained. The feature quantity scores of the RFECV algorithm based on different evaluators are shown in Figure 5.

As can be seen from Figure 5, the optimal number of features may be different when using different estimators. For example, when using the SVM and LDA algorithms as estimators, the corresponding optimal number of features are 16 and 23, respectively. However, when using different estimators, it may also happen that the optimal number of features is the same. For example, when the GBDT and SGD algorithms are used as estimators respectively, the corresponding optimal number of features is 12.

4.4 Comparison of the effects of different models based on multiple feature subsets

Based on the different optimal feature subsets obtained in the previous section, we used Decision Tree, Multi-layer Perceptron, Bayesian Regression, Random Forest, AdaBoost and Stacking algorithms to training model, so that a combined model that combines different RFECV algorithms with classification algorithms can be constructed. By comparing the AUC values of different combination models, the best combination can be screened out. The experimental
results are shown in Table 2 below, where SVM-RFECV represents the RFECV algorithm with the SVM algorithm as the estimator, and the rest are the same.

<table>
<thead>
<tr>
<th>Feature Selection Algorithm</th>
<th>Number Of Features</th>
<th>DT</th>
<th>BR</th>
<th>RF</th>
<th>MLP</th>
<th>AdaBoost</th>
<th>Stacking</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feature Selection</td>
<td>25</td>
<td>0.838</td>
<td>0.852</td>
<td>0.886</td>
<td>0.889</td>
<td>0.875</td>
<td>0.922</td>
</tr>
<tr>
<td>SGD-RFECV</td>
<td>12</td>
<td>0.848</td>
<td>0.864</td>
<td>0.857</td>
<td>0.887</td>
<td>0.886</td>
<td>0.913</td>
</tr>
<tr>
<td>SVM-RFECV</td>
<td>16</td>
<td>0.836</td>
<td>0.858</td>
<td>0.878</td>
<td>0.912</td>
<td>0.892</td>
<td>0.897</td>
</tr>
<tr>
<td>LR-RFECV</td>
<td>13</td>
<td>0.869</td>
<td>0.862</td>
<td>0.879</td>
<td>0.869</td>
<td>0.908</td>
<td>0.919</td>
</tr>
<tr>
<td>CART-RFECV</td>
<td>11</td>
<td>0.897</td>
<td>0.897</td>
<td>0.906</td>
<td>0.912</td>
<td>0.952</td>
<td>0.956</td>
</tr>
<tr>
<td>LDA-RFECV</td>
<td>23</td>
<td>0.897</td>
<td>0.897</td>
<td>0.906</td>
<td>0.912</td>
<td>0.952</td>
<td>0.956</td>
</tr>
<tr>
<td>GBDT-RFECV</td>
<td>12</td>
<td>0.898</td>
<td>0.902</td>
<td>0.904</td>
<td>0.932</td>
<td>0.957</td>
<td>0.969</td>
</tr>
</tbody>
</table>

It can be seen from the above results that from the perspective of the RFECV feature selection algorithm, the combined model using GBDT and LDA as the RFECV algorithm estimator has a better effect, and is better than the performance without feature selection. From a classification model perspective, Stacking and AdaBoost work better when used as classifiers. From the perspective of the combined model, the combined model when using GBDT and LDA as the estimator and Stacking and AdaBoost as the classifier works better.

### 4.5 Comparison of classification effects of various components based on different models

Based on the conclusions obtained in the previous section, we used four different combined models to perform feature selection and model training on eight different web components, and use cross-validation for evaluation. We compared the ACC value and AUC value to judge the pros and cons of different combination models. The experimental results are shown in Table 3. The “GBDT + Stacking” combination model means that the GBDT algorithm is used as the estimator and the Stacking algorithm is used as the classifier.

<table>
<thead>
<tr>
<th>Component Name</th>
<th>LDA + AdaBoost</th>
<th>LDA + Stacking</th>
<th>GBDT + AdaBoost</th>
<th>GBDT + Stacking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wordpress</td>
<td>0.964</td>
<td>0.952</td>
<td>0.936</td>
<td>0.956</td>
</tr>
<tr>
<td>Drupal</td>
<td>0.883</td>
<td>0.903</td>
<td>0.877</td>
<td>0.927</td>
</tr>
<tr>
<td>JQuery</td>
<td>0.925</td>
<td>0.936</td>
<td>0.910</td>
<td>0.967</td>
</tr>
<tr>
<td>Lodash</td>
<td>0.926</td>
<td>0.922</td>
<td>0.916</td>
<td>0.968</td>
</tr>
<tr>
<td>Django</td>
<td>0.947</td>
<td>0.953</td>
<td>0.931</td>
<td>0.921</td>
</tr>
<tr>
<td>Laravel</td>
<td>0.913</td>
<td>0.919</td>
<td>0.919</td>
<td>0.926</td>
</tr>
<tr>
<td>Magento</td>
<td>0.921</td>
<td>0.947</td>
<td>0.916</td>
<td>0.932</td>
</tr>
<tr>
<td>Shopify</td>
<td>0.931</td>
<td>0.939</td>
<td>0.933</td>
<td>0.938</td>
</tr>
<tr>
<td>Average Value</td>
<td>0.926</td>
<td>0.934</td>
<td>0.922</td>
<td>0.929</td>
</tr>
</tbody>
</table>

It can be seen from the above results that when the “GBDT + Stacking” combined model is used, the average ACC of different Web components is 0.951, and the average AUC is 0.957, which are higher than other combined models. It shows that the Web component recognition effect of the combined model is the best.

### 4.6 Comparing the recognition performance of our method with existing tools

We applied different web component identification tools respectively to identify 8 kinds of web components and compare with our method. The experimental results are shown in Table 4. It can be found that the application of our method has higher accuracy than other tools and websites in the identification of web components.

<table>
<thead>
<tr>
<th>Identification Tools</th>
<th>Wappalyze</th>
<th>Whatweb</th>
<th>FOFA</th>
<th>Our Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.912</td>
<td>0.923</td>
<td>0.826</td>
<td>0.951</td>
</tr>
</tbody>
</table>

### 5. CONCLUSION

We studied how to identify the Web components applied in the website, and proposed a new identification method of Web components based on machine learning, and implemented the method. The experimental results show that our
method has better recognition effect in the process of Web component recognition. In the next step, we will further improve the feature selection method and classification model to improve the recognition effect as much as possible while taking into account the efficiency.

REFERENCES

Commodity Knowledge Graph-based TransD-KGAT Method for Recommendation
Chunfu Xie
Institute of Yongyou Company
Yonyou Network Technology Co. LTD
100000, Beijing, China
xiechfu@yonyou.com

ABSTRACT
In the field of item recommendation, traditional collaborative filtering-based approaches face the problems of data sparsity and cold start. As a heterogeneous network with multiple entities and relationships, the knowledge graph has rich semantic association information to help explore reasonable associations between users-items and items-items and assist recommendation systems in alleviating the problems of sparsity and cold start. Graph neural network-based approaches can explicitly encode key collaboration signals for user-item interactions and have been proven to be effective recommendation methods. However, previous work has not explored how the entity superordinate words and synonyms of the knowledge graph affect the metrics. In this paper, we discuss and experiment from two perspectives of knowledge graph construction and graph neural network embedding to explore how to improve the recommendation metrics comprehensively. In this paper, we construct commodity knowledge graphs and propose the model called TransD-KGAT, which improves three indexes compared with the KGAT model in terms of precision, recall, and ndcg by 2.64%, 4.96%, and 3.9%. It is also concluded that adding knowledge graphs of semantically similar and superordinate words improves the recommendation metrics.

Keywords: recommendation, graph neural network, TransD, knowledge graph

1. INTRODUCTION
With the advent of the significant data era, users are also facing the problem of information overload, and a recommendation system [1] is an effective information filtering tool. It recommends information of interest to users by analyzing their basic information and historical behaviors. The recommendation algorithm based on collaborative filtering is widely used [2], but there are problems with data sparsity and cold start [3]. Using a knowledge graph [4-7] to assist the recommendation system, the recommendation model-assisted using a knowledge graph can make the recommendation results have three characteristics: accuracy, diversity, and interpretability [8-11]. There are three main approaches: the path-based approach, the embedding-based approach, and the graph neural network-based approach.

The path-based approach [12, 13] mainly mines multiple connections between graph-based users and items and provides additional suggestions. However, manually designing many meta-paths and the need to reconstruct them when the recommendation scenario or the graph changes is difficult to optimize in practice. The embedding-based approach [14-16] uses the knowledge graph embedding [17] algorithm to preprocess the knowledge graph and merge the learned entity embeddings into the recommendation framework. However, this approach is focused on modeling strict semantic relevance with transformation constraints and is more suitable for graph-related tasks, such as link prediction.

Graph neural network-based approaches [18-21] can explicitly encode key collaboration signals for user-item interactions to enhance user-item representations through the propagation process. It shows that graph neural network-based approaches have significant advantages over path-based and embedding-based approaches. One of the KGAT models [21] is the classical model of this family of models. We use this model idea to explore how knowledge graphs and graph neural networks affect recommendation metrics.

We summarize the contributions of this work.

Automating the construction of commodity knowledge graphs, including semantic similarity word knowledge graphs, superordinate word knowledge graphs, and extracting sub-commodity knowledge graphs based on CN-DBpedia, and proposing a new model called TransD-KGAT, considering the high accuracy as well as efficiency of TransD [22]...
compared to TransR [23] model in calculating the similarity information degree of commodities. Compared to the baseline model, we obtained soft results on the company's dataset.

2. PROBLEM FORMULATION

2.1 Knowledge Graph

The knowledge graph describes the knowledge base composed of entities and the relationships between them in the form of a graph consisting of nodes, edges, and data to form a semantic network graph. Where nodes are sets of entities or concepts and edges are sets of attributes or relationships. The knowledge graph is stored in structured triples. There are "entity-relationship-entity" to express the relationship between entity and entity and "entity-attribute-attribute value" to express the value of an attribute of an entity, which we denote by KG.

2.2 Hypernym

Hypernymy describes the subordinate relationship between concepts, which helps to uncover the intrinsic hierarchical structure between concepts.

2.3 Task description

Considering the user purchase id, item id, and item name, the likelihood of a user purchasing an item can be predicted.

3. METHODOLOGY

3.1 Commodity Knowledge Graph Construction

We collect user transaction data based on the data from our internal marketing cloud department, including user id, user purchase product id, and product name. We build a product knowledge graph based on product names, mainly divided into the following parts.

The product name is the information freely marked by merchants on the platform, and the product name contains much information, such as product name, brand name, and model name, which are valid for the product knowledge graph. However, some redundant information will also lead to the failure to link the knowledge graph. We believe that only nouns can represent the subject entity of the transaction in this list, and we have adopted the lexical nature for nouns using the pyltp1 open-source tool for word separation and lexical annotation and selected the lexical nature for nouns to form the commodity labels.

We consider the product label information too thin and adopt semantic similarity word expansion and hypernym extraction. To expand the label information, we assume the Tencent word vector2 for semantic similarity word expansion, five semantic similarity words expand each label, denoted as KGsmi. Commodity hypernym extraction labels the superordinate commodity category. We adopt an open-source dictionary3, and two hypernyms are extracted for each original label, i.e., the primary superordinate word and the secondary superordinate word denoted as KGhpy.

For the original tags, the tags formed by the superordinate word extraction and those formed by the commodity semantic similarity words are jointly used as a list of tags corresponding to commodity ids. The sub-knowledge graphs are extracted from the Chinese open knowledge graph CN-DBpedia4, denoted as KGcn. All sub-knowledge graphs are linked to form the product knowledge graph called Commodity Knowledge Graph. We use an actual commodity to represent how to construct a commodity knowledge graph, as shown in figure 1 below.

---

1 https://github.com/HIT-SCIR/pyltp
2 https://ai.tencent.com/ailab/nlp/embedding.html
3 http://www.bigcilin.com/
4 http://kw.fudan.edu.cn/apis/cndbpedia/
3.2 TransD-based node similarity calculation

According to the user id, product id, and product id corresponding to the product knowledge graph as input, the KGAT model is selected as the reference framework benchmark. The KGAT model uses the TransR model for knowledge graph embedding and takes the in-computer approach to calculate the attention score based on the distance between the head entity and the tail entity in the embedded relationship space as the attention score mechanism. In this regard, We believe that the TranR model selected by the KGAT model in the knowledge graph embedding layer has specific problems. Firstly, the head and tail entities project themselves onto the hyperplane using the same transformation matrix. However, the head and tail entities are usually different types of entities. For example, they are represented in the commodity knowledge graph as one is the brand name, and the other is the merchant name. This will have errors in calculating the item similarity information (attention score). Second, the number of parameters of TransR is too much, and it isn't easy to apply to large-scale knowledge graphs due to its complexity. Therefore, we make the following improvements: we use the TransD method, which considers the diversity between entities and relationships, is less computationally intensive, and can be applied to large-scale knowledge graphs. The head entity and tail entity are denoted as \( h \), \( t \), and the relation is denoted by \( r \). \( h_p \), \( r_p \), and \( t_p \) is denoted as their vectors, respectively. \( M_{hr} \) And \( M_{rt} \) denote the matrix mapping from entity space to relation space, respectively, as shown in the following two formulas:

\[
M_{hr} = r_p h_p^p + I^{m,n} \quad (1) \\
M_{rt} = r_p t_p^p + I^{m,n} \quad (2)
\]

Entities and relations determine the mapping matrix. This approach makes the projection matrix sufficiently interactive. \( i \) is the unit matrix, \( r \) is the projection vector for the head entity and the tail entity projection vectors, and \( f_r \) is the loss function. The structure of TransD is shown in figure 2.

\[
h_\perp = M_{hr} h_t = M_{rt} t \quad (3) \\
f_r(h,t) = \|h_\perp + r - t_\perp\| \quad (4)
\]
3.3 Why apply graph attention networks as the basic embedding framework

This part uses the KGAT model as the primary reference and is divided into the attention embedding propagation and prediction layers. The attention embedding propagation layer consists of four parts: information propagation layer, commodity information attention layer, information aggregation layer, and higher-order propagation layer. The information dissemination layer is based on the similarity of goods, the purchase of goods as a starting point, and the collection of information about neighbors.

Consider the entity \( h \), denoted by \( N_h = \{(h, r, t) | (h, r, t) \in KG\} \) for a set of three, where \( h \) is the head entity, \( r \) is the relation, and \( t \) is the tail entity.

\[
e_{N_h} = \sum_{(h,r,t)\in N_h} \pi(h, r, t)e_i
\]  

(5)

The product information attention layer is to calculate the similarity between the product purchased by the user and the surrounding products, which is achieved by the relational attention mechanism \( \pi(h, r, t) \) with the following equation:

\[
\pi(h, r, t) = (t_i^T \tanh(h_i + r))
\]  

(6)

Why do we adopt graph attention networks as the basic encoding model?

(1) It can explicitly encode key collaborative signals of user-item interactions to enhance user/item representations through the propagation process.

(2) It is more efficient and convenient because it avoids the labor-intensive process of specifying paths compared to path-based approaches.

(3) It considers the effect of different relationships on recommendation effectiveness.

4. EXPERIMENTS

Evaluation metrics: for each user in the test set, we considered all items not purchased as negative items. Three widely used evaluation protocols were used: precision@K, recall@K, and ndcg@K. We set \( K = 20, 40, 80, \) and 100.

Baseline: To demonstrate its validity, we compared our proposed TransD-KGAT with CKE, NFM, and KGAT.

NFM: This method is a state-of-the-art factorization model, which includes FM under neural networks.

CKE: This is a knowledge graph embedding-based method that uses semantic embeddings from TransR to enhance matrix decomposition.

KGAT: KGAT is a state-of-the-art gnn-based recommender. It applies the attention neighborhood aggregation mechanism to the overall graph and combines KG with the user-item graph to generate user and item representations.

<table>
<thead>
<tr>
<th></th>
<th>precision@20</th>
<th>recall@20</th>
<th>ndcg@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKE</td>
<td>0.0459</td>
<td>0.3450</td>
<td>0.2061</td>
</tr>
<tr>
<td>NFM</td>
<td>0.0538</td>
<td>0.4377</td>
<td>0.2376</td>
</tr>
<tr>
<td>KGAT</td>
<td>0.0607</td>
<td>0.5139</td>
<td>0.2774</td>
</tr>
<tr>
<td>TransD-KGAT</td>
<td>0.0623</td>
<td>0.5394</td>
<td>0.2822</td>
</tr>
</tbody>
</table>

A table comparing the overall performance of different models is shown in Table 1. TransD-KGAT consistently produces the best performance on all datasets. It improves 2.64%, 4.96%, and 3.9% in three metrics compared to the KGAT model in precision, recall, and ndcg. It proves that the TransD model is more effective compared to the TransR model in the embedding aspect of calculating the similarity of goods. Figures 3, 4, and 5 compare recall, accuracy, and ndcg metrics on different @K, which also proves the above conclusion.
To explore the effect of knowledge graph on graph neural networks, we remove semantic similarity words and superordinate words part and both superordinate words and semantic similarity words, respectively. w/o hyp stands for removing superordinate words part, w/o smi stands for removing semantic similarity words part, and w/o hyp&smi stands for removing semantic similarity words and superordinate words. Table 2 shows the overall performance comparison table of the effect of the knowledge graph on graph neural networks. On the three indicators of precision, recall, and ndcg, the indicators of removing the superlative part decreased by 2.09%, 0.41%, and 4%, respectively, and the indicators of removing the semantic similarity part decreased by 3.05%, 7.27%, and 6.17%, respectively, and the indicators of removing semantic similarity words and superlatives decreased by 4.98%, 7.12%, and 10.24%, respectively. This experiment proves that both semantic similarity words and hypernyms have a significant impact on the recommendation index, where semantic similarity words have a more significant impact than superordinate words because semantic similarity words contain superordinate words to a large extent. Figures 6, 7, and 8 show the comparison of recall, accuracy, and ndcg metrics on different @K, which also prove the above conclusion.

<table>
<thead>
<tr>
<th></th>
<th>precision@20</th>
<th>recall@20</th>
<th>ndcg@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>TransD-KGAT</td>
<td>0.0623</td>
<td>0.5394</td>
<td>0.2822</td>
</tr>
<tr>
<td>w/o hyp</td>
<td>0.0610</td>
<td>0.5372</td>
<td>0.2709</td>
</tr>
<tr>
<td>w/o smi</td>
<td>0.0604</td>
<td>0.5002</td>
<td>0.2648</td>
</tr>
<tr>
<td>w/o hyp&amp;smi</td>
<td>0.0592</td>
<td>0.5010</td>
<td>0.2533</td>
</tr>
</tbody>
</table>
best metrics on our dataset.
This work demonstrates that the graph neural network approach based on knowledge graphs is naturally applicable to recommender systems. Knowledge graphs have rich semantic association information, and the rich semantic relationships between items in knowledge graphs can help explore the relationships and improve the representation of items. At the same time, using knowledge graphs to organize auxiliary information can also extend user interests, reasonably disperse recommendation results, and more easily help us build interpretable recommendation systems.

REFERENCE

[12] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. 2014. Personalized entity recommendation: A heterogeneous information network approach. In Proceedings of the 7th ACM International Conference on Web Search and Data Mining. 283–292.


Abstract

The domain of computer vision's most popular study area has always been face recognition which aims to identify different face images and predict the corresponding identity information through feature analysis and modeling. In practical applications, a number of variables, including lighting, posture, and clarity, have an impact on facial recognition accuracy. Among them, the most challenging scene is occlusion face recognition, which will cause feature loss, local coherence and alignment errors to greatly inhibit the accuracy and generalization ability of the face model. Based on detailed literature research and analysis, this paper provides a comprehensive evaluation of the research progress of occluded face recognition. Specifically, based on the introduction of classical face recognition technology, we further discuss the design ideas, basic framework, advantages and disadvantages of representative occlusion face recognition methods from two aspects: robust feature extraction and robust classifier. Finally, we summarize the main challenges and give an outlook on the future research development of object detection.

Keywords—Face recognition; Occlusion; Deep Learning; Robust features and classifier

1. Introduction

Face recognition is a biometric method, which realizes human identity recognition by extracting and comparatively analyzing the visual feature information of faces. Due to the advantages of easy acquisition of face image data (compared with iris, fingerprint, retina, DNA, etc.), low hardware cost, non-contact acquisition, and large data set scale, face recognition technology has become a cutting-edge technology in the field of computer vision. In recent years, thanks to the powerful feature learning ability and good model generalization ability of convolutional neural networks, face recognition researches based on deep learning have achieved breakthroughs in both accuracy and speed, which are widely used in face access control attendance, community access control management, station face gates and other fields. Although face recognition technology has made great progress, the actual face recognition environment is affected by factors such as illumination, expression, posture and especially occlusion, which brings great challenges to the further application of face recognition. Nowadays, under the background of the epidemic of new coronavirus pneumonia, masks have become a necessity for people to travel. Face occlusion caused by masks greatly reduces the recognition accuracy of face recognition technology. Therefore, it is crucial to figure out how to reduce how occlusion affects face recognition.

According to the different causes of occlusion, the occlusion in reality mainly includes three categories: illumination occlusion, object occlusion and self-occlusion caused by attitude change. Illumination occlusion is caused by uneven external lighting, while object occlusion is generally caused by the covering of external objects (such as masks, sunglasses, etc.). In addition, the change of the human body's own posture will also lead to self-occlusion. For example, the face will be occluded when the arm is raised. The above-mentioned occlusions directly lead to the lack of local information in the face image and the introduction of irrelevant noise, which will greatly affect the final recognition accuracy. Occlusion face recognition technology has been the subject of extensive study and experimentation in recent years, and numerous practical solutions have been put forth. Representative models and frameworks of face recognition include mtcnn, facenet, Dlib and their subsequent improvements. algorithm etc. For occluded face recognition methods, they still follow the basic framework of face feature representation and classification, but focusing on how to achieve high-quality feature representation or build robust classifiers. According to the different improvement links, existing research on occluded face recognition can be classified into two categories: robust feature extraction and robust classifier construction.
1.1 Robust feature extraction

The extraction and comparison of the key features of the face is the key to the face recognition algorithm. When using computer vision algorithms to build a face recognition model, the completeness of the facial features is critical. Occlusion can cause problems such as feature loss, noise, and local aliasing, preventing face recognition algorithms from making accurate decisions. The main methods of robust feature extraction include traditional feature extraction methods, deep learning-based feature learning methods, and hybrid methods that combine the traditional and deep learning framework. To describe, restrict, or remove occlusion areas and increase the accuracy and efficiency of partially occluded face recognition algorithms, traditional methods primarily leverage the global and occlusion structure aspects of face images. Convolutional neural networks' strong feature extraction capabilities are used by feature learning techniques based on deep learning to extract deep features that better represent the data's (such as face data) abstract semantic information. In order to propose a more complete face recognition model, a lot of work have been done by researchers on the improvement of the network structure, the construction and adaptability of the loss function. For example, DeepID2 facilitates traditional face recognition algorithms. In addition, there are some works that utilize the 3D geometric information of faces to detect, locate and classify occlusions in 3D space, providing new ideas for face recognition in 2D images.

1.2 Robust classifier construction

For occlusion images, no feature is more informative than the original image itself, and feature extraction methods diffuse occlusion noise in local or global regions. Therefore, some studies consider that feature extraction is not important for building robust classifiers, and focus on building robust classifiers. The methods of robust classifiers include subspace regression and robust error coding. The basic idea of the subspace regression method is to divide the face images of different categories into different subspaces, and to establish an independent occlusion subspace for occlusion, and to convert the occlusion face image recognition problem into the face image and occlusion regression respectively. Corresponding subspace problem. The main difficulty of such methods is how to construct high-quality occlusion subspaces. The basic idea of the robust error coding method is to regard the occluded image as a combination of the original unoccluded image and the occlusion error, which mainly include two types of "additive model" and "multiplication model", which synthesize the original image and occlusion error through addition and multiplication respectively.

Focusing on the main technical framework of occluded face recognition, this paper describes traditional face recognition algorithms and reviews existing occluded face recognition methods from robust feature extraction and robust classifier design. Finally, we look forward to future trends in this research area.

2. Classic face recognition method

2.1 Face recognition based on geometric features

Face recognition based on geometric features refers to the prior knowledge of geometric relations of face topological structure, using structure-based methods in the knowledge level to extract facial features of the main organs, the face with a set of geometric feature vectors, through the matching of feature vectors, to complete face recognition. Its basic idea is to use the relative position and relative distance of some feature points of the face (such as eyes, nose, mouth, etc.), supplemented by the shape information of the contour of the face. When matching feature vectors, distance measurement is a good measure of the similarity between two feature vectors, and a decision based on Euclidean distance is the most commonly used method. The measurement of features is a crucial step in this approach. Therefore, before measuring, the characteristics of the measurement must be standardized, the purpose of which is to make the measurement result independent of the azimuth, scale and rotation of the face, and the measurement results are stored in the form of vectors.

Due to the limitations of the shooting conditions of the original image, the extraction of feature points will cause interference. Therefore, before the extraction and classification, it is necessary to do the standardization of the image, remove the redundant information that is useless or interfere with face recognition in the image, retain the main information of the face, and obtain a standardized image with less interference, which is conducive to subsequent feature extraction and face recognition. The normalization of face image is divided into geometric normalization and gray normalization. Geometric normalization refers to the face changes to the same position and size in the face image from the input image. Gray-scale normalization is the processing of illumination compensation for the positioned face image to overcome the influence of illumination change. In this paper, geometric normalization is used to calibrate the boundary of the face. Here, two functions are introduced first.
Let $F(x, y)$ be the input image with size of $m \times n$, $I(x, y)$ represents the gray value of the image $f$ at point $(x, y)$, we can make the vertical integral projection function of the area $V(x)$ and $H(y)$ by formula (1) and (2).

$$V(x) = \frac{1}{y_2-y_1} \sum_{y=1}^{y_2-1} I(x, y), \quad x \in [1, m] \quad (1)$$

$$H(y) = \frac{1}{x_2-x_1} \sum_{x=1}^{x_2-1} I(x, y), \quad y \in [1, n] \quad (2)$$

Here, $V(x)$ represents the vertical integral projection function of the area $[y_1, y_2]$, and $H(y)$ represents the horizontal integral projection function to the area $[x_1, x_2]$. A graph of the projection function is shown as Figure 1.

![Fig. 1. Visualization of the projection curves](image)

Observing the face image and projection curve, it can be seen that since the average grayscale of the face is different from other objects in the background, the grayscale value of the area where the face is located is relatively average, and there will be obvious grayscale changes at the junction with the background, reflecting the integral projection curve, a point with a relatively large gradient value will appear. Therefore, by calculating the gradient value of the vertical projection curve and setting the threshold, the upper boundary of the face can be obtained. Similarly, the left and right boundaries of the face can be obtained by performing horizontal integral projection on the face image.

### 2.2 Face recognition based on eigenface

EigenFace's goal is to convert the face from space in pixels to another space and do a similarity computation there. In preprocessing, Principal Component Analysis (PCA), widely employed to eliminate the relationship between sample feature dimensions, is the spatial transformation technique selected by EigenFace. The primary face distribution components are obtained by the EigenFace approach using PCA. The primary face distribution components are obtained by the EigenFace approach using PCA [1]. EigenFace conducts an eigenvalue decomposition on the correlation matrix of every facial image in the test examples in order to achieve the associated eigenvectors. "Eigenfaces" are these eigenvectors. Each eigenvector or eigenface is equivalent to describing the changes or features between faces.

An image can either be expanded into a vector or thought of as a matrix made up of pixel values. For instance, a $N \times N$ pixel picture can be thought of as a vector of length $N^2$, making the image a point in the $N^2$-dimensional space. This image's initial image space is its vector representation, but there are numerous more spaces in which an image might be represented or recognized. The implementation process of face recognition based on eigenface can be described as following: (1) Each face image in the training set should be extended by one column, then combined to create a sizable matrix $A$. Assuming that each face picture has a size of $M \times M$, each sample of a face has a dimension of $d = M \times M$ from being arranged into a column. The dimension of the sample matrix $A$ is $d \times N$ if there are $N$ face images. (2) The average of $N$ faces is calculated to get the "average face." (3) To get the data matrix $\Phi$ of the difference image, subtract the ordinary face image from all $N$ photographs. (4) Make a covariance matrix $C = \Phi \Phi^T$ calculation. You may then apply eigenvalue decomposition on it to obtain the necessary eigenvector (eigenface). (5) The classification is realized by...
projecting the images of training set and test set onto these feature vectors to find the nearest neighbor or k nearest neighbors in the training group for each image in the evaluation set.

### 3. Robust feature extraction

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>Basic idea</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td>2015</td>
<td>PCA</td>
<td>Filter learning does not involve regularization parameters or require a numerical optimization solver, making training simple</td>
</tr>
<tr>
<td>[3]</td>
<td>2014</td>
<td></td>
<td>The proposed method is robust to the interference of illumination change, occlusion and camouflage</td>
</tr>
<tr>
<td>[4]</td>
<td>2021</td>
<td></td>
<td>It has certain robustness to light conditions</td>
</tr>
<tr>
<td>[5]</td>
<td>2021</td>
<td>Attention mechanism</td>
<td>The Convolutional Block Attention Module (CBAM) Attention mechanism is integrated into ResNet50 and MobileFaceNet networks, which has been proved to improve the performance of face recognition compared with other methods</td>
</tr>
<tr>
<td>[6]</td>
<td>2022</td>
<td></td>
<td>Low cost, high precision masked face synthesis method - masked transfer method for data enhancement. It also improves the performance of masked face recognition</td>
</tr>
</tbody>
</table>

The human eye can predict the higher-order properties of occlusion according to the existing unmasked parts, which also tests the inference ability of machine vision, such as color, brightness, texture, and direction, which are called lower-order features. Posture, age, facial expression, race, and so on are called high-order features. The robust feature extraction method decomposes high-order features and low-order features.

#### 3.1 Principal Component Analysis (PCA)

PCA is a dimensionality reduction approach. It is mainly used to simplify large dimensional data sets in data mining. With the development of deep learning in recent years, more and more deep learning methods are used in face recognition technology, and remarkable results have been achieved. Chan et al. proposed Principal Component Analysis Net (PCANet) in 2015, which combined cascaded PCANet, binary hashes and block histograms [2]. PCANet uses PCA to learn multi-stage filter banks, combining feature extraction and deep learning for the first time. The experiment results show that the dimension of features would increase exponentially with the number of PCANet advanced stages. At the same time, in the face of more challenging image databases, PCANet may not be enough to solve the problem of image variability.

At present, the treatment of various conditions under occlusion conditions is still the mainstream research. The possible influencing factors can be divided into three types: complex environment such as illumination, Hats, glasses, and other artificial shielding, identify the changes in posture and expression. Motivated by the most recent findings on robust principal component analysis for low-rank matrix recovery, a approach using the sparse error components obtained by robust principal component analysis was proposed by Luan et al. [3]. However, it has some limitations because it does not take into account explicit occluded areas and environment. Based on low-rank matrix decomposition, the method can solve the problem of PCA sparse error component is too large, which has certain robustness to occlusion environment.
Zhou et al. took the features extracted by local binary mode as the input of deep network according to that local binary mode was not sensitive to illumination changes [4]. Combining the local binary mode with the deep learning method, the experiment proves the effectiveness of the method. Locally Linear Embedding (LLE) and Correlation Nearest Neighbor (CNN) (LLE-CNN) method for occlusion face detection, which can recover many lost facial feature information to a large extent and greatly reduce the influence caused by various occlusions.

3.2 Attention mechanism

When face information occlusion occurs, the features of other unoccluded parts of the face can be used to assist in the extraction of occluded face features, that is, the image content of the missing region can be supplemented, restored and predicted according to the neighborhood information of the occluded region. Then perform feature extraction. To this end, introducing an attention mechanism to selectly focus on non-occluded regions to improve the quality of learned features has become another major framework for robust feature extraction.

The convolutional neural network (ACNN) with attentional mechanism for face emotion detection under partial occlusion attempts to focus on multiple regions of the face image, mainly the non-occlusion region, and automatically calculates the adaptive weight according to the degree of occlusion. Three modules are mainly used: proposal module, embedded module and validation module. A novel mask face recognition method based on an attention mechanism was proposed by Liu et al. named faceX-Zoo [5]. The author incorporate two networks with the Convolutional Block Attention Module (CBAM) attention mechanism and program the modules to focus on the region around the eyes. By employing the FaceX-zoo approach to create masked face images, the module's speed is improved. The outcomes of the experiments demonstrate how well this technology can achieve facial recognition to enhance the effectiveness of masked face recognition. Zhang et al. proposed an Attention-aware Masked face recognition technology (AMaskNet), which realized accurate feature representation through simple matrix multiplication and optimized the whole model by using an end-to-end training strategy [6]. This technique focuses on the non-occluded region and can eliminate the interference of sample background to some extent, thus obtaining the discriminant region of face recognition. The suggested AMaskNet module can develop an efficient contribution matrix and provide more weight to the activated feature graph of the unmasked part, which improves the face recognition performance of masked people, according to experimental data.

4. Robust classifier

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>Basic idea</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>[7]</td>
<td>2012</td>
<td>Subspace regression method</td>
<td>The system can recognize faces efficiently and effectively under various real conditions, and has a high degree of robustness and stability to illumination changes, image misalignment and partial occlusion</td>
</tr>
<tr>
<td>[8]</td>
<td>2012</td>
<td>Subspace regression method</td>
<td>The kernel dictionary is used to reconstruct the sample and the face image is classified according to residuals, which has good effect</td>
</tr>
<tr>
<td>[9]</td>
<td>2011</td>
<td>Subspace regression method</td>
<td>It has certain robustness under the influence of illumination and Angle</td>
</tr>
<tr>
<td>[10]</td>
<td>2014</td>
<td>Robust Error Coding</td>
<td>It is robust to severe illumination variation, occlusion and random pixel noise destruction</td>
</tr>
<tr>
<td>[11]</td>
<td>2013</td>
<td>Subspace regression method</td>
<td>Solve the problem of surface occlusion and data loss caused by subspace analysis technology</td>
</tr>
<tr>
<td>[12]</td>
<td>2014</td>
<td>Subspace regression method</td>
<td>It has good illumination robustness</td>
</tr>
<tr>
<td>[13]</td>
<td>2014</td>
<td>Subspace regression method</td>
<td>It has good illumination robustness</td>
</tr>
<tr>
<td>[14]</td>
<td>2020</td>
<td>Subspace regression method</td>
<td>It has good illumination robustness</td>
</tr>
</tbody>
</table>

The final step in a complete face recognition system is to classify the extracted features. The quality of the classifier affects the recognition performance of the entire algorithm. For occluded face images, learning a robust classifier is a feasible
solution to improve the recognition performance. Methods for robust classifiers include subspace regression and robust error coding.

4.1 Subspace regression method

The Sparse Representation Classification method (SRC) based on Sparse Representation can help the algorithm predict the size of occlusion that can be processed and select training images to reduce the error caused by occlusion. This algorithm has brought widespread attention to the problem of partial occlusion face recognition. In 2012, Wagner et al. further proposed Robust Sparse Representation Classification (RSRC) model [7]. Experiments show that feature extraction and feature selection should be considered when the RSRC model is used for recognition. Face recognition using the Kernel Metaface learning method developed by Zhu et al. in 2012 was successful, although the issue of parameter selection needs further discussion [8]. A sparse representation face recognition method based on deep learning features solves the problem that traditional sparse representation face recognition methods are not robust to in-class changes in the case of small samples, and has certain advantages in the case of small samples.

Based on the consideration of computational performance, Zhang et al. analyzed the mechanism of SRC algorithm in 2011, which believed that SRC and related work emphasized the importance of sparsity while ignoring the use of Collaborative Representation (CR) [9]. They pointed out that the most powerful role of SRC is CR and proposed a simple and efficient face classification scheme named Collaborative Representation based Classification (CRC). Experiments show that the classification performance of CRC is almost the same as RSRC, but the computational complexity is greatly reduced. Zhang et al. 2014 found that the computational complexity of the sparse representation classification algorithm is too high, which can be improved by the collaborative classification algorithms [10]. In view of the advantages of combining the collaborative classification and the Gabor feature extraction, Zhang et al proposed a Gabor and collaborative classification of a traditional algorithm combining algorithm, which can make the recognition rate of face recognition relatively stable under the influence of affection, lighting and angles.

Face recognition also can be realized based on a 3D model. This kind of recognition and authentication through more camera sensors to obtain a 3D model of the face, for the face posture, expression, lighting, occlusion, and other changes have very robust performance. Alyuz proposed a fully automatic 3D face recognition system with robustness to occlusion [11]. It is mainly aimed at face recognition with occluded surfaces and data loss caused by subspace analysis technology. The unmasked part is trained and the occluded part is removed. The masking strategy adopted in the classification stage is a subspace analysis technique using incomplete data, and further, it is to improve the global quality by using the local processing method. Ali proposed a 3D face recognition model with unchanged posture in a remote frame that could automatically acquire graphics to complete face recognition [12].

4.2 Robust Error Coding

Unlike the common errors caused by Gaussian noise, the errors caused by occlusion generally have a certain spatial structure. Therefore, the occlusion structure can be considered to encode the occlusion error, so as to improve the recognition accuracy of the model for occluded faces. The existing structured coding ideas for occlusion or occlusion errors mainly include three types: (1) The occlusion dictionary learning method, which describes occlusion by constructing an occlusion dictionary that matches the occlusion structure. (2) describe the inherent structure of the face image itself to separate the occlusion; (3) directly encode the structural error by embedding the inherent structure of the occlusion itself into the reconstruction error.

The structure of robust error coding is usually used to describe the face image, so this method can separate the occlusion in the face. A novel face technique was proposed by Wei et al. in 2014. It can partition the initial training data into a collection of representative bases for improved modeling and can also incorporate the structural incoherence constraint [13]. For the problem that robust statistics are not robust enough to dense coarse errors, Zheng et al. proposed a coding residual modeling method based on sparse representation, which can co-process error detection and error correction [14]. String matching is a very powerful partial matching technique, but it is not suitable for positive face recognition, because faces include continuous and discontinuous features, which need to be represented by global sequences. Chen and Gao proposed a face recognition approach utilizing global string matching, which uses very compact syntax description string to represent faces and can complete matching between two discontinuous string faces [15]. Because the sequential order and direction of the face strings remain the same, this potential allows the algorithm to automatically exploit every unshaded area in the recognition process, regardless of the shape. This method is also effective when the training sample size is small, which not only has remarkable performance in partially occluded face recognition but also has the ability to match the sketched face with the face in the photo. Chen et al. proposed a discriminant descriptor based on Weber's Law.
(WLD), which uses differential excitation and position in a specific direction to calculate histograms [16]. This method has good illumination robustness, but a poor recognition effect in the case of occlusion. This method will be applied to face recognition in the future.

5. Experiment

5.1 Mask face recognition commonly used data set

Real-world Masked Face Recognition Dataset (RMFD) Recently, the new coronavirus is rampant around the world, and almost everyone wears masks in areas with serious epidemics, which has a massive sample base. The samples were collected to build the world's largest mask face dataset RMFD and opened to the society to accumulate data resources for intelligent control of current and possible future similar public safety events. The samples are crawled from the network, after finishing, cleaning and labeling, containing 5,000 mask faces of 525 people and 90,000 normal faces.

FaceMaskDetection contains 853 images belonging to the 3 classes, as well as their bounding boxes in the PASCAL VOC format. The classes are: With mask; Without mask; Mask worn incorrectly.

FaceMaskDataset All pictures are translated into jpg format, contain only properly wearing medical masks; All annotation files are converted from XML to txt YOLO compatible format. FaceMaskDataset contains pictures of people wearing medical masks along with txt files containing their descriptions. There are 632 pairs overall.

MAFA (MAasked FAces) is a masked face detection benchmark dataset, of which images are collected from Internet images. MAFA contains 30,811 images and 35,806 masked faces. Faces in the dataset have various orientations and occlusion degrees, while at least one part of each face is occluded by mask. In the annotation process, each image contains at least one face occluded by various types of masks, while the six main attributes of each masked face, including locations of faces, eyes and masks, face orientation, occlusion degree, and mask type.

WIDER FACE dataset is a face detection benchmark dataset, of which images are selected from the publicly available WIDER dataset. Including 32,203 images and label 393,703 faces with a high degree of variability in scale, pose and occlusion as depicted in the sample images. WIDER FACE dataset is organized based on 61 event classes. For each event class, we randomly select 40%/10%/50% data as training, validation and testing sets. We adopt the same evaluation metric employed in the PASCAL VOC dataset. Similar to MALF and Caltech datasets, we do not release bounding box ground truth for the test images. Users are required to submit final prediction files, which we shall proceed to evaluate.

Table 3. Commonly used face recognition occlusion data set description

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMFD (2020)</td>
<td><a href="https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset">https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset</a></td>
<td>The samples are crawled from the network, after finishing, cleaning and labeling, containing 5,000 mask faces of 525 people and 90,000 normal faces.</td>
</tr>
<tr>
<td>Face Mask Detection (2020)</td>
<td><a href="https://www.kaggle.com/datasets/andrewmvd/facemask-detection">https://www.kaggle.com/datasets/andrewmvd/facemask-detection</a></td>
<td>This dataset contains 853 images belonging to the 3 classes, as well as their bounding boxes in the PASCAL VOC format.</td>
</tr>
<tr>
<td>MAFA (2017)</td>
<td><a href="http://www.escience.cn/people/geshiming/mafa.html">http://www.escience.cn/people/geshiming/mafa.html</a></td>
<td>each image contains at least one face occluded by various types of masks, while the six main attributes of each masked face, including locations of faces, eyes and masks, face orientation, occlusion degree, and mask type.</td>
</tr>
<tr>
<td>WIDER FACE (2015)</td>
<td><a href="http://shuoyang1213.me/WIDERFACE/">http://shuoyang1213.me/WIDERFACE/</a></td>
<td>For each event class, we randomly select 40%/10%/50% data as training, validation, and testing sets.</td>
</tr>
</tbody>
</table>
5.2 Commonly used evaluation indicators

Evaluation indexes are mainly used to evaluate the quality of face recognition models, that is, to evaluate the accuracy of trained models in test sets. The more common evaluation indicators are accuracy, which is the matching degree of face recognition under different occlusion conditions, and so on.

5.3 The experimental results

Table 4. Relevant experimental data statistics

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Testset</th>
<th>Percent occluded</th>
<th>Recognition accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCANet (Chan et al.,2015)</td>
<td>RandNet</td>
<td>0%</td>
<td>97.77%</td>
</tr>
<tr>
<td>LDANet</td>
<td>Extended YALE B</td>
<td>80%</td>
<td>54.38%</td>
</tr>
<tr>
<td>sparse error component obtained RPCA (Luna et al.,2014)</td>
<td>Extended YALE B</td>
<td>0%</td>
<td>99.58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80%</td>
<td>51.73%</td>
</tr>
<tr>
<td>SRC (Wagner et al.,2012)</td>
<td>Extended YALE B</td>
<td>10%</td>
<td>99.78%</td>
</tr>
<tr>
<td>CR based classification with regularized least square (CRC_RLS) (Zhang et al.,2011)</td>
<td>AR</td>
<td>10%</td>
<td>100%</td>
</tr>
<tr>
<td>face recognition system (Alyuz et al.,2013)</td>
<td>Extended YALE B</td>
<td>70%</td>
<td>60.53%</td>
</tr>
<tr>
<td>fully automatic 3D</td>
<td>Scarf</td>
<td>70%</td>
<td>11.40%</td>
</tr>
<tr>
<td>low-rank matrix decomposition (Wei et al.,2014)</td>
<td>Bosphorus</td>
<td>80%</td>
<td>45.10%</td>
</tr>
<tr>
<td>CMU MULTI-PIE</td>
<td>Masked</td>
<td></td>
<td>93.18%</td>
</tr>
<tr>
<td>CMU MULTI-PIE</td>
<td></td>
<td></td>
<td>93.40%</td>
</tr>
</tbody>
</table>

Table 5. The correct recognition rate with different algorithms in YALE-B

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Correct recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>95.41%</td>
</tr>
<tr>
<td>SVM</td>
<td>96.83%</td>
</tr>
<tr>
<td>LBP</td>
<td>96.12%</td>
</tr>
<tr>
<td>DBN</td>
<td>95.74%</td>
</tr>
<tr>
<td>LBP+DBN</td>
<td>97.33%</td>
</tr>
</tbody>
</table>

In the past, Extended YALE B data set was often used. In this case, low blocking rate had little effect on the model, such as PCANet and Sparse Error Component obtained RPCA, which generally achieved nearly 100% accuracy. However, with the increase of shielding degree, Face recognition rates have dropped significantly. For example, the low-rank matrix decomposition method can still achieve 93% face recognition accuracy in the case of occlusion of eyes, that is, the occlusion rate is about 20%. However, with the increase of occlusion rate, the influence on the model is further enhanced. The occlusion rate of scarf is about 80%, while the recognition rate of common methods is reduced to about 50%. The CRC_RLS method, which performed best, covered most of the faces, but still achieved 95% accuracy.

6. Discussion

Face recognition with occlusion is a major challenge in computer vision in recent years. Although occluded face recognition has undergone a comprehensive and rapid development process in the past research, there are still many changes in the existing algorithms. By analyzing the existing occluded face recognition algorithms, we find that there are still many problems in this field.

6.1 Computer-perceived occlusion is still a problem

When the human eye perceives an occluded image, the vision will automatically judge whether there is occlusion and filter out the occluded part. However, the computer still cannot judge the exact location of occlusion through the training of a large number of data sets. The limited training samples also restrict the learning process of the computer. How to make the computer accurately perceive the occlusion position is still a research direction.
6.2 Missing dataset
At present, there are few masked datasets. For example, LF W, CASIA webface and celeba datasets are all face datasets without occlusion. Wearing masks on faces has also become a problem to be solved. The problems of data set and perceptual occlusion have been solved to varying degrees, but the location of occlusion is still difficult to detect accurately and has a great impact on the accuracy of recognition. When using 0%-99% random occlusion of baboon face, it can be proved that the NS classifier with known occlusion support can have 100% recognition accuracy even when combined with "shallow" Weber and LBP features. NN classifier with unknown occlusion support, even when combined with "deep" PCA net feature, the recognition accuracy will decline sharply with the increase of occlusion ratio. In real identity recognition systems, the location of occlusion is random and the area of occlusion may be large, so the research on the location of occlusion is still the focus of future research.

6.3 Optimize the loss function
A well-designed loss function can maximize the distance of inter-class features and the compactness of intra-class features, which is particularly critical to improve the network's ability to discriminate specific feature vectors.

6.4 3D face recognition research
Strengthen the construction of 3D face datasets and the research on 3D face recognition algorithms. Make full use of its stable spatial geometric relationship to reduce the recognition error of the face due to changes in illumination, viewing angle, etc.

7. Conclusion
Following the development of technical design ideas, this paper summarizes the current representative algorithms in occluded face recognition from the aspects of robust feature extraction and robust classifier. The main methods of robust feature extraction include traditional feature extraction methods, deep learning-based feature learning methods and hybrid methods that combine the traditional and deep learning framework. While for the robust classifier, existing methods includes subspace regression and robust error coding. By introducing and analyzing these representative occluded face recognition algorithms, we finally compare their performance and look forward to future trends in this research area.

REFERENCES
Face-based Gender Recognition with Small Samples Generated by DCGAN Using CNN

Haoyu Feng1a*
1Department of Electrical Engineering, Chongqing University, Chongqing 400044, China
a*20186093@cqu.edu.cn

Abstract

Gender recognition of face images is an important research field in image classification. Convolutional neural network (CNN) has a good performance in this field. To obtain successful results, large and high-quality data sets are essential. However, data scarcity has always been a common challenge in deep learning. The deep convolutional generative adversarial network (DCGAN) has a powerful application in data enhancement. The face image generated by DCGAN can be used to enhance face data. In this study, the fake face images generated by DCGAN and the real face images of CelebA datasets are used to train CNN classifier for gender classification of human faces. In the experiment, the training set and test set both are divided into three types, including real face images, fake face images, and real-fake mixed face images. The type of test set is the same as that of the training set, and the number of face images in both the training set and test set is small, including 1000 face images. By comparing the training accuracy and test accuracy of datasets under various conditions in CNN classifiers, it can judge whether the fake face images generated by DCGAN have a good data enhancement effect. The experimental results show that the data enhancement ability of DCGAN is also well applied to CNN classifier, and the face images generated by DCGAN can effectively improve the performance of the CNN classifier to a certain extent.

Keywords-component: DCGAN; CNN; Gender classification; Small samples; CelebA dataset

1. Introduction

Human face recognition is a branch of biometric technology. It is often used as the only feature to be identified by machines in life. Human gender recognition from face images is an important research field because it is widely used in life. In the criminal investigation, when the public security department obtains the photos of criminals, it can use gender identification technology to reduce the search index, which greatly saves the time and energy to solve the case. In addition, for some places that are allowed to be visited based on gender, such as toilets, special passages, special train carriages, and changing rooms, any passenger or visitor who violates this access regulation can be monitored according to gender identification technology. For commercial applications, advertising and promotional activities of a product for a specific sex group can be recommended, such as cosmetics and skirts for women, and electronic products and razors for men. This kind of targeted product marketing will greatly increase the publicity effect. Gender identification technology is also used to collect customer and visitor population data in specific areas, such as amusement parks, shopping malls, airports, and railway stations. According to the gender data of business areas or public areas, effective business plans can be better formulated for the area.

In recent years, gender classification based on face images has attracted more and more attention in machine learning. Many studies were carried out on different technologies of human gender recognition. Including gender classification based on geometric features, gender classification based on subspace, and gender classification based on neural networks. The recognition method based on the neural network can avoid the influence of a complex environment to a certain extent and improve the recognition rate of the algorithm. In addition, in the field of image classification, a convolutional neural network (CNN) has been proved to be an excellent algorithm compared with other machine learning algorithms [1]. All layers of a CNN have multiple convolution filters to work, scan the complete characteristic matrix, and reduce the dimension. This makes CNN become a very suitable network for image classification and processing.

However, to obtain successful results from deep learning, large and high-quality data sets are essential. For most people, a large number of data sets are difficult to obtain, and data scarcity is always a common challenge in deep learning, especially in medical image processing [2-4]. The data enhancement method can obtain a large number of data sets and
build powerful classifiers from insufficient data effectively. Data augmentation methods are widely used in the practical application of deep learning, such as medical images [5], text classification [6], autonomous vehicles [7] and speech recognition [8]. Image-based Data augmentation is also called image augmentation. Image augmentation technology makes a series of random changes to the training image to produce similar but different training samples, to expand the scale of the training set [9]. However, in the field of face-based gender recognition, the application effect of this technique has not been deeply studied. Training CNN classifier for gender recognition also requires a large number of face image samples. DCGAN has a powerful application in image augmentation. The expected result of this experiment is to improve the performance of the CNN classifier.

In this study, the fake face images generated by DCGAN and the real face images of data sets are used to train CNN classifier for gender classification of human faces. In the experiment, the training set is divided into three types, including real face images, fake face images, and true-fake mixed face images. The type of test set is the same as that of the training set, and the number of face images in both the training set and test set is small. By comparing the accuracy of three sets of CNN classifiers, it can judge whether the fake face images generated by DCGAN have a good data enhancement effect. In the experiment, TensorFlow is used to classify the images into male and female. To verify the effectiveness of the classifier, the accuracy of the model is calculated. And by comparing the accuracy of the three data sets, the conclusion of the study can be attained.

2. Method

2.1. Dataset preparation and preprocessing

The dataset chosen by the model is CelebFaces Attributes Dataset (CelebA) [10]. It contains 202, 599 celebrity images in RGB format, each with 40 attribute annotations. The background of the image set is disordered, the brightness is different, and the characters show a variety of expressions.

DCGAN generates 2, 000 fake images using CelebA as a test set and training set. In the DCGAN, the number of training sets is 202, 599 pictures. CNN employs 2, 000 fake images generated by DCGAN and 2, 000 true images in CelebA as data sets. In CNN, the datasets are divided into three types. Both training sets and test sets are including these three types: 1000 fake images generated by DCGAN, 1000 true images in CelebA, or 500 fake images generated by DCGAN and 500 true images in CelebA. Each type contains 500 females and 500 males. Because three training sets and three test sets are combined in pairs, the CNN model can be divided into 9cases. In the DCGAN, the pictures are transformed to 3x64x64 size and cropped based on the central zone. Fig. 1 presents some sample real and fake images.

Fig. 1 Sample real and fake images

2.2. Employed DCGAN Model

Setting parameters and image set routes is the first step. The model then preprocesses the pictures in the data set and divides them. In Pytorch, data is loaded in the following order: construct a dataset and a data loader object, loop the data loader object and then pass the data and label to the model for training. All model weights in the experiment should be
randomly initialized from a normal distribution with a mean of 0 and a standard deviation of 0.02. The weight-initial function takes the initialized model as the input and reinitializes all convolution layers, deconvolution layers, and batch standardization layers to meet this standard. This function is instantly applied to the model when it has been initialized. The epoch of the model is set to 5.

The discriminator is a classifier that takes a set of pictures as input and outputs a scalar once it has passed through the network. The higher the value of the scalar, the greater the probability that the input picture is real. The smaller the scalar value, the more likely it is that the image is a fake. The model uses 1 to represent the real image, and 0 to represent the image generated by the generator. The architecture of the designed discriminator network can be found in Fig. 2.

It's not the same as the discriminator's convolution process for extracting the image's features. The generator is a reverse procedure that converts a latent space vector into data space. Fig.3 presents the structure of the generator network.

SGD optimizer and Adam optimizer were chosen in this experiment to generate fake images. And it is estimated by two typical metrics, namely IS and SSIM. Finally, define the loss function and optimizer, start training, and save the trained model in a file.

2.3. Proposed CNN model

The experiment involves the creation of a convolution neural network that can detect a person's gender based on their facial appearance. The catalogue classification and several pictures in the test and training sets are identical. The images of the two sets are contained in a directory called CNN images. Each set contains three directories: real-real, real-fake, and fake-fake. Each directory includes two subdirectories: man and woman. All 12 subdirectories contain jpg images that are trying to input the neural network. CNN is a process of extracting information from these images and marking or
classifying them. The binary classification was carried out in the experiment. In this sort of classification, the experiment’s outcome is a binary value of 0 or 1, with the female output being 0, and the male output being 1.

Import the necessary libraries at the beginning of the experiment. Import datasets, load data, and basic EDA. The number of images of men and women will be visualized in the model. Then prepare training and test data. Set that rotation_range = 40, width_shift_range = 0.2, height_shift_range = 0.2, shear_range = 0.2, zoom_range = 0.2. Complete architecture implemented in TensorFlow. Starting with a convolution block with 64 filters, the kernel size is (3×3), the step size is 2, and then is the relu activation function. There are three convolution layers and three pooling layers in total. The number of convolution cores in the convolution layer is 64, 128, and 256 respectively. The size of the input data is 64×64×3. The pooled layers are all maximum pooled. The number of images of men and women will be visualized in the model. Then prepare training and test data. Set that rotation_range = 40, width_shift_range = 0.2, height_shift_range = 0.2, shear_range = 0.2, zoom_range = 0.2. Complete architecture implemented in TensorFlow. Starting with a convolution block with 64 filters, the kernel size is (3×3), the step size is 2, and then is the relu activation function. There are three convolution layers and three pooling layers in total. The number of convolution cores in the convolution layer is 64, 128, and 256 respectively. The size of the input data is 64×64×3. The pooled layers are all maximum pooled. The number of convolution cores in the convolution layer is 64, 128, and 256 respectively. The size of the input data is 64×64×3. The pooled layers are all maximum pooled.

3. Test Results and Discussions

3.1. The performance of DCGAN

![Generator and discriminator loss during training](image)

Fig.4 Generator and discriminator loss during training

According to Fig.4, it can be seen that the loss of discriminator is low at first, demonstrating that the network can readily distinguish between real images and fake images. However, the loss of the generator is very large, and the maximum value reaches 35.6299, this indicates that the images generated by the generator are not real enough. As training iterations rise, the loss of the generator and discriminator steadily approaches zero, indicating that the network is producing increasingly realistic images. However, overall, the loss of the generator is always more than that of the discriminator, indicating that although being more real, the created image is inferior to the true image.

Fig.1 is some samples of real images and fake images. Although the real images and the fake images can still be recognized by the human eye, the fake images can see the human face. It is very similar to the real image in color, background, and brightness. Fake images can be tried to use as training samples to train the CNN model.

3.2. The performance of CNN

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>600:0</td>
<td>600:0</td>
<td>300:300</td>
<td>0.6667</td>
<td>0.7050</td>
</tr>
<tr>
<td>2</td>
<td>1000:0</td>
<td>1000:0</td>
<td>500:500</td>
<td>0.7490</td>
<td>0.8240</td>
</tr>
<tr>
<td>3</td>
<td>1000:0</td>
<td>500:500</td>
<td>500:500</td>
<td>0.7450</td>
<td>0.7510</td>
</tr>
<tr>
<td>4</td>
<td>1000:0</td>
<td>0:1000</td>
<td>500:500</td>
<td>0.7350</td>
<td>0.6950</td>
</tr>
<tr>
<td>5</td>
<td>500:500</td>
<td>1000:0</td>
<td>500:500</td>
<td>0.6840</td>
<td>0.7360</td>
</tr>
<tr>
<td>6</td>
<td>500:500</td>
<td>500:500</td>
<td>500:500</td>
<td>0.7000</td>
<td>0.7890</td>
</tr>
</tbody>
</table>
The accuracy of various training sets and test sets are shown in Table 1. The first row contains 300 real images as the training set and 300 real images as the test set. The accuracy is shown in Table 1. From row 2 to row 10, the number of images in both the training set and the test set is 500. The training set and test set are both divided into three types: 1000 real images, 500 real images and 500 fake images, or 1000 fake images. These three datasets constitute 9 training set and test set combinations.

From Table 1, test accuracy is bigger than training accuracy in every experiment, and it can be observed that the more the real images in the test set, the higher the test accuracy and training accuracy when training sets are all real images in the case of comparing row 2, row 3 and row 4. Comparing row 8 to row 9 and row 10, when training sets are all fake images, the more the fake images in the test set, the higher the test accuracy and the training accuracy.

Comparing row 2 to row 5 and row 8, when test sets are all real, the more the real images in the training set, the higher the training accuracy and the test accuracy. Comparing row 4 to row 7 and row 10, when test sets are all fake, the more the fake images in the training set, the higher the training accuracy and the test accuracy.

In summary, the training accuracy and test accuracy are higher when the type of training set and test set are the same. Therefore, using the same kind of training set and test set to train the model is better. Furthermore, the two accuracies are the highest when both the training set and test set are all real images. Thus, using real images train model is more effective than using fake images. However, compared to row 1, increasing fake images in the training set and test set indeed improves accuracy.

<table>
<thead>
<tr>
<th></th>
<th>500:500</th>
<th>0:1000</th>
<th>1000:0</th>
<th>500:500</th>
<th>0.7170</th>
<th>0.6980</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0:1000</td>
<td>1000:0</td>
<td>500:500</td>
<td>0.6510</td>
<td>0.6970</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0:1000</td>
<td>500:500</td>
<td>500:500</td>
<td>0.6830</td>
<td>0.7010</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0:1000</td>
<td>1000:0</td>
<td>500:500</td>
<td>0.6920</td>
<td>0.7110</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5 The curve for training and test accuracy
As shown in Fig. 5, with the increase of epoch, the training accuracy and test accuracy also increase. When epoch=15, the training set is 1000 real images, and the test set is 1000 real images. The training accuracy reaches the maximum value of 0.7490, and the test accuracy also reaches the maximum value of 0.8240.

As shown in Fig.6, with the increase of epoch, the training loss and test loss decrease. When epoch=15, the training set is 1000 real images and the test set is 1000 real images. The training loss reaches the minimum value of 0.5157, and the test loss also reaches the minimum value of 0.4198.

4. Conclusion

DCGAN has a powerful application in data enhancement. Face images generated by DCGAN can be used to enhance face data. The face images generated by DCGAN are added to the training set and test set to form three types of training sets and test sets (full real data set, full fake data set, real-fake mixed data set) and nine combined models of training sets and test sets. By comparing the accuracy of nine data set combination models, it can be concluded that DCGAN can indeed expand the data set, achieve the purpose of training CNN classifier, and realize the application of data enhancement. However, the effect of training the CNN model with images generated by DCGAN is not as good as that with real images, because there is still a gap between the images generated by DCGAN and the real images, and they cannot completely replace the real images. In the future, further studies can consider increasing the authenticity of the generated image to make the generated image closer to the real image and better replace the real image.
References


Wild Animal Recognition Based on Effective-Class-Balanced Softmax Loss

Wen Chen\textsuperscript{a}, Qianzhou Cai\textsuperscript{a}, Jin Hou\textsuperscript{b}, Jindong Zhang\textsuperscript{b}, Bochuan Zheng\textsuperscript{*c}

\textsuperscript{a}School of Mathematics and Information, China West Normal University, Nanchong, Sichuan, China;
\textsuperscript{b}School of Life Science, China West Normal University, Nanchong, Sichuan, China;
\textsuperscript{c}School of Computer Science, China West Normal University, Nanchong, Sichuan, China

* Corresponding author: zhengbc@vip.163.com

ABSTRACT

Wild animal recognition is important for wild animal protection. Because the number of different wild animals is different in the wild. The wild animal image dataset collected in filed by using camera trap is a typical long tail dataset. This paper proposes an Effective-Class-Balanced Softmax Loss (ECBSL) to solve the long tail problem of self-built wild animal dataset. Firstly, a new cross entropy loss function is obtained by using pointwise mutual information instead of conditional probability for modeling. Then the improved effective number of samples calculation method is used to approximately calculate the prior probability distribution of different animal species. Finally, the effectiveness of ECBSL is proved by experiments. Experiments on the self-built wild animal dataset show that the proposed method improves the recognition accuracy of the tail classes and the whole dataset. The comparison experiments with other methods show that the proposed method is superior to other methods.

Keywords: wild animal recognition, pointwise mutual information, cross entropy loss, effective number of samples

1. INTRODUCTION

Wild animal protection is of great significance to achieve ecological sustainable development. In recent years, camera trap technology has been widely used in wild animal image acquisition. Biologists can collect a lot of wild animal images in field by installing a large number of infrared cameras in wild animal habitats. They can obtain the information of wild animal from these collected images. One of the main work that biologists have to do is to recognize the wild animal species from the collected images. For a long time, this work is done manually. It requires experienced experts and consumes a lot of time. In recent years, researchers have used computer vision technology to recognize wild animal species from the collected images. This technology greatly improves the efficiency of wild animal recognition.

Most collected wild animal datasets suffer from imbalanced samples due to the differences in the number of different wild animal species. In other words, the sample numbers of some wild animal species are larger than others, and these wild animal species are called head classes. While the sample numbers of some rare wild animal species are smaller than others, and these wild animal species are called tail classes. This problem is usually called long tail problem. The long tail problem has the following characteristic: the sample numbers of head classes are much larger than that of tail classes. Taking the wild animal dataset constructed by ourselves in this paper as an example, the sample number of the first head wild animal species is hundreds of times that of the last tail species. The long tail problem results in neural networks biasing towards head classes, and the recognition accuracy of the tail classes is low. In the wild animal dataset, rare animals only have a small amount of image data, but we hope to recognize rare animals as much as possible. Therefore, it is necessary to find a method to improve the recognition accuracy of rare wild animal without reducing the overall recognition accuracy.

Researchers studied the long tail problem and have proposed many methods to solve it. Most researchers solve the problem by balancing data, such as re-sampling \cite{1, 2} or re-weighting \cite{3}. However, these methods underestimate the influence of huge head class data on the neural network in training process. Cao et al. \cite{4} proposed Label-Distribution-Aware Margin (LDAM) from the perspective of generalization error. They argued that a larger classification distance should be given to the tail class data. However, their algorithm is not suitable for multi-classification problem because it

Keywords: wild animal recognition, pointwise mutual information, cross entropy loss, effective number of samples

1. INTRODUCTION

Wild animal protection is of great significance to achieve ecological sustainable development. In recent years, camera trap technology has been widely used in wild animal image acquisition. Biologists can collect a lot of wild animal images in field by installing a large number of infrared cameras in wild animal habitats. They can obtain the information of wild animal from these collected images. One of the main work that biologists have to do is to recognize the wild animal species from the collected images. For a long time, this work is done manually. It requires experienced experts and consumes a lot of time. In recent years, researchers have used computer vision technology to recognize wild animal species from the collected images. This technology greatly improves the efficiency of wild animal recognition.

Most collected wild animal datasets suffer from imbalanced samples due to the differences in the number of different wild animal species. In other words, the sample numbers of some wild animal species are larger than others, and these wild animal species are called head classes. While the sample numbers of some rare wild animal species are smaller than others, and these wild animal species are called tail classes. This problem is usually called long tail problem. The long tail problem has the following characteristic: the sample numbers of head classes are much larger than that of tail classes. Taking the wild animal dataset constructed by ourselves in this paper as an example, the sample number of the first head wild animal species is hundreds of times that of the last tail species. The long tail problem results in neural networks biasing towards head classes, and the recognition accuracy of the tail classes is low. In the wild animal dataset, rare animals only have a small amount of image data, but we hope to recognize rare animals as much as possible. Therefore, it is necessary to find a method to improve the recognition accuracy of rare wild animal without reducing the overall recognition accuracy.

Researchers studied the long tail problem and have proposed many methods to solve it. Most researchers solve the problem by balancing data, such as re-sampling \cite{1, 2} or re-weighting \cite{3}. However, these methods underestimate the influence of huge head class data on the neural network in training process. Cao et al. \cite{4} proposed Label-Distribution-Aware Margin (LDAM) from the perspective of generalization error. They argued that a larger classification distance should be given to the tail class data. However, their algorithm is not suitable for multi-classification problem because it
is derived from Cross Entropy (CE) loss under binary classification. Ren et al. [5] proposed Balanced-Meta Softmax (BALMS) method which combine balanced softmax and meta sampler to improve the recognition performance for tail classes. But this method sacrifices too much recognition performance of head classes. Cui et al. [6] realized the reweighted strategy by using the effective number of samples, but the parameters with the effective number of samples need to be set manually. This way results in difficulty of finding the optimal parameters.

Because the classification boundary between the two classes in the long tail problem biases towards the class with larger samples number. And there exists feature overlap among samples in same class. This paper tries to optimize the classification boundary between two classes by using effective number of samples. Therefore, based on literature [6], we propose the Effective-Class-Balanced Softmax Loss (ECBSL) function. Our method combines balanced softmax with effective number of samples to solve the long tail problem.

The main contributions of this paper are as follows: 1) The proposed method improves the recognition accuracy of rare wild animal. Biologists can quickly and accurately find rare animals from wild animal image dataset. 2) The proposed ECBSL can more accurately locate classification boundaries between different classes than before. 3) A equation of auto counting effective number of samples is proposed, avoiding manually set.

2. RELATE WORK

2.1 Wild Animal Recognition

In order to dynamically monitor wild animal, researchers recognize wild animal species through computer vision technology from images collected by using camera traps. Villa et al. [7] can recognize 26 species of wild animal from Snapshot Serengeti database [8]. Verma et al. [9] first used Deep Convolutional Neural Network (DCNN) to detect wild animal in camera trap database. Then they used the machine learning algorithms to recognize wild animal species. The accuracy reach 91.4 %. Islam et al. [10] proposed a Convolutional Neural Network (CNN) to classify wild animal images. These images are collected from the Bastrop Country, Texas, and achieved SOTA results. Vatsavayi et al. [11] classified wild animal species in Serengeti dataset by using Deep Residual Convolutional Neural Network (DRCNN). Norouzzadeh et al. [12] recognized 48 species of animals in Snapshot Serengeti database by using multiple DCNN models. And the recognition accuracy reached 94.9 %.

2.2 Long Tail Problem

Researchers have proposed a variety of solutions to solve the long tail problem. These solutions can be divided into four categories: data re-balancing, improving the loss function, meta-learning strategy and transfer learning.

Data Re-balancing. Data re-balancing mainly includes two strategies: data re-sampling and data re-weighting. The re-sampling strategy [1, 2, 13–16] achieves quantitative balance among classes by over-sampling the tail classes or under sampling the head classes. However, over-sampling easily causes overfitting because that the tail-class samples are repetitive. While under-sampling will result in poor generalization ability of the model. The re-weighting strategy [3, 6, 17, 18] is mainly to assign higher weight to the tail classes than the head classes in the loss function. It removes the influence of unbalance samples. In general, data re-balancing method may decrease the overall recognition accuracy of all samples.

Improved Loss Function. Cao et al. [4] showed that the generalization error boundary can be minimized by increasing the margin of the tail classes. Ren et al. [5] derived the generalized boundary of multiple softmax regressions. They minimized the error boundary by improving the loss function. Hayat et al. [19] designed a mixed loss function based on the “affinity measure” in Euclidean space. Ye et al. [20] made the tail classes have a larger decision value to prevent feature deviation by adding Class Dependent Temperature (CDT) in the training process.

Meta-Learning. Jamal et al. [21] explicitly estimated the conditional distribution difference among classes by using meta-learning methods. This enhanced the classical class balance learning methods. Ren et al. [22] proposed a new meta-learning algorithm that can dynamically learn the weight of each sample according to gradient change. Shu et al. [23] proposed a method that can adaptively learn explicit weighting functions. However, these meta-learning methods need a clean and unbiased data set as the meta set.

Transfer Learning. Ouyang et al. [24] designed a hierarchical learning scheme. In the first stage, all data are clustered into different clusters by clustering algorithm. In the second stage, deep features of different clusters are learned respectively to obtain a long tail classification model. Zhong et al. [25] proposed a phased training strategy to process
head class and tail class data in different ways. The first step is to learn face features from the head class data. And the second step is to learn more stable feature information to achieve auxiliary discrimination by using the tail class data. Zhou et al. [26] proposed a unified Bilateral-Branch Network (BBN), whose two branches share parameters. One branch learns from the original data and the other branch learns from the data after re-sampling. The two branches are dynamically weighted.

3. THE METHOD

3.1 Dataset

In this paper, wild animal images are collected from Wolong Nature Reserve, Sichuan Province, China. Firstly, we set a large number of infrared cameras in the wild animal habitat to get wild animal images. Then we manually screened out clear and complete wild animal images. The collected wild animal images are shown in Figure 1 (a). Finally, the wild animal dataset is obtained after manually labeling and cropping. The cropped wild animal images are shown in Figure 1 (b). The dataset contains 16,707 images of 26 wild animal species, including giant pandas, golden monkeys, Tibetan sheep and Tibetan yaks. The distribution of 26 wild animal species are shown in Figure 2. It can be seen that the wild animal dataset collected in this paper is a typical long tail dataset. In the experiment, according to the ratio of 8:1:1, the dataset is randomly divided into training set, validation set and test set. The training set contains 13,365 images, the validation set contains 1,671 images and the test set contains 1,671 images.

(a). The collected wild animal images                              (b). The cropped wild animal images

Figure 1. The wild animal images.

Figure 2. The image number distribution of wild animal species.
3.2 Improvement Measures

Cross Entropy Loss based on the Pointwise Mutual Information. Suppose there are $N$ training samples in the dataset $D$, and the $D$ has $C$ classes. The input feature of the sample $i$ is $x_i$, and the corresponding label is $y_i \in \{1, 2, \ldots, C\}$, so the training data $(x_i, y_i) \in D$. Assume the conditional distribution of $y_i$ is $p(y_i | x_i)$, the CE loss function of all samples is:

$$L_{CE} = - \frac{1}{N} \sum_{i=1}^{N} y_i^T \log p(y_i | x_i)$$

(1)

Where,

$$y_i^T = [y_{i1}, y_{i2}, \ldots, y_{iC}]$$

(2)

$$p(y_i | x_i) = \frac{1}{\sum_{j=1}^{C} e^{f_j(x_i; \theta)}} [e^{f_1(x_i; \theta)}, e^{f_2(x_i; \theta)}, \ldots, e^{f_C(x_i; \theta)}]^T$$

(3)

It can be seen that the network output $f(x; \theta) = [f_1(x; \theta), f_2(x; \theta), \ldots, f_C(x; \theta)]^T$ actually fits the probability of each sample belonging to each class. Since the proportion of each class can be obtained according to the statistics of sample dataset, we can obtain the distribution information of each class if we replace the probability with the proportion of each class. Therefore, in order to use the class distribution information of each class, we make the network output $f(x; \theta)$ fit the Pointwise Mutual Information (PMI). The calculation equation of PMI is as follows,

$$PMI(x, y) = \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(y | x)}{p(y)}$$

(4)

Because $f(x; \theta)$ fits PMI, so

$$\log \frac{p(y | x)}{p(y)} \approx f(x; \theta) \Rightarrow \log p(y | x) \approx f(x; \theta) + \log p(y) \Rightarrow p(y | x) \approx e^{f(x; \theta) + \log p(y)}$$

(5)

Thus, equation (2) can be transformed into:

$$p(y | x_i) = \frac{1}{\sum_{j=1}^{C} e^{f_j(x_i; \theta) + \log p_j(y)}} [e^{f_1(x_i; \theta) + \log p_1(y)}, e^{f_2(x_i; \theta) + \log p_2(y)}, \ldots, e^{f_C(x_i; \theta) + \log p_C(y)}]^T$$

(6)

Where, $p_j(y)$ represents the probability of each class.

Calculating the Probability of Each Class. Literature [6] proposed the concept of effective number of sample size. Suppose that $n_i$ is the number of samples collected in the $i$ class, and $N_i$ is the number of all possible samples in the feature space of this class. The effective number of sample size of the class $i$ is defined as:

$$E_i = \frac{1 - \beta^n}{1 - \beta}$$

$$\beta = \frac{N - 1}{N}$$

(7)

Since it is impossible to know the number of all possible samples $N_i$ in the feature space of the class $i$. Literature [6]
uniformly set the possible sample size of all classes as $N$, and then $\beta$ was set as 0.9, 0.99, 0.999, 0.999 respectively. They respectively calculated the loss function to take the optimal value.

For the class $i$, although it is impossible to know the number of all possible samples, the $n_i$ can be calculated statistically. Therefore, we use $n_i$ instead of $N_i$, and then calculate effective sample size is as follows.

$$E_i = \frac{1 - \beta^n}{1 - \beta}, \quad \beta = \frac{n_i - 1}{n_i}$$

(8)

Assume $S = \sum_{i=1}^{C} E_i$, the calculation of $p_i(y)$ is:

$$p_i(y) = \frac{E_i}{S}$$

(9)

Finally, the following is our ECBSL function:

$$L_{ECBSL} = -\frac{1}{N} \sum_{i=1}^{N} y_i^T \log \frac{e^{f_i(x_i;\theta) + \log p_i(y)}}{\sum_{j=1}^{C} e^{f_j(x_i;\theta) + \log p_j(y)}}$$

(10)

Where $y_i^T = [y_{i1}, y_{i2}, \cdots, y_{ic}]$, $p(y) = [p_1(y), p_2(y), \cdots, p_c(y)]^T$, $f(x_i;\theta) = [f_1(x_i;\theta), f_2(x_i;\theta), \cdots, f_c(x_i;\theta)]^T$

4. EXPERIMENT AND RESULT ANALYSIS

4.1 Experimental Environment and Parameter Setting

Experimental computer hardware configuration: dual-core Intel Xeon CPU E5-2650v4 @ 2.20GHz, 252GB memory, 4 Tesla P40 graphics cards, each having 24GB display memory. The software system configuration: Ubuntu 16.04 LTS, CUDA 9.0, CUDNN 7.3, Python 3.6 programming language, PyTorch deep learning framework.

The batch size in the training is 512. The Adam optimizer is adopted the momentum parameter is set to 0.9, and the total number of training epochs is 2000. The initial learning rate is 0.001. When the loss function does not decrease for 10 consecutive epochs during training, the learning rate is attenuated, the attenuation weight is 0.0005.

4.2 Experimental Results

For the evaluation criteria, we divide the dataset into two subsets: head class $D_h$, tail class $D_t$. In the head class, each specie has more than 100 images. In the tail class, each specie has 20-100 images. The mean classification accuracy of the head class subset is shown in equation (11), the mean classification accuracy of tail class subset is shown in equation (12). The mean classification accuracy of the whole dataset is shown in equation (13).

$$mAP_h = \frac{1}{C_h} \sum_{i=\text{headclass}} mAP_i$$

(11)

$$mAP_t = \frac{1}{C_t} \sum_{i=\text{tailclass}} mAP_i$$

(12)

$$mAP_d = \frac{1}{C} \sum_{i=1}^{C} mAP_i$$

(13)
Where $mAP_i$ is the classification accuracy of class $i$, $C_t$ is the number of wild animal species in tail class subset, $C_h$ is the number of wild animal species in head class subset, $C$ is the number of wild animal species in dataset.

The experimental classification network includes Resnet34, Resnet50, Resnet101 and Resnet134. For each network, softmax, LDAM, balanced softmax and ECBSL loss functions are used for comparative experiments. The experimental results of the test dataset are shown in Table 1. The results show that:

1. From the classification results of tail class, no matter what network is used, the classification accuracy of softmax loss function is obviously worse than the other three loss functions. The main reason is that the other three loss functions consider the distribution of samples, but softmax loss function does not. At the same time, the classification accuracy of ECBSL is better than the other three loss functions. It can be seen that our improvements are effective.

2. From the classification results of head class, the classification accuracy of ECBSL is better than the other three loss functions when using Resnet34 and Resnet134 networks. When Resnet50 and Resnet101 networks are used, the classification accuracy of ECBSL is not optimal. But it is not significantly different from the best classification accuracy. Therefore, ECBSL loss function does not reduce and improve the classification accuracy of head class, because it focuses on the tail samples.

3. From the classification results of the whole dataset, no matter what network is used, the classification accuracy of softmax loss is obviously worse than that of the other three loss functions. At the same time, the classification accuracy of ECBSL is better than the other three loss functions. Therefore, ECBSL loss function improves the classification accuracy of whole dataset.

In summary, the ECBSL loss function proposed by us improves the classification accuracy of the tail class and the whole dataset. However, it does not reduce and improve the head class.

Table 1. Comparison of experimental results.

<table>
<thead>
<tr>
<th>Method</th>
<th>Network Model</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Head class</td>
</tr>
<tr>
<td>Softmax Loss</td>
<td>Resnet34</td>
<td>85.2</td>
</tr>
<tr>
<td>LDAM Loss</td>
<td></td>
<td>84.6</td>
</tr>
<tr>
<td>Balanced Softmax Loss</td>
<td></td>
<td>84.4</td>
</tr>
<tr>
<td>ECBSL</td>
<td></td>
<td>85.3</td>
</tr>
<tr>
<td>Softmax Loss</td>
<td>Resnet50</td>
<td>86.8</td>
</tr>
<tr>
<td>LDAM Loss</td>
<td></td>
<td>85.5</td>
</tr>
<tr>
<td>Balanced Softmax Loss</td>
<td></td>
<td>85.9</td>
</tr>
<tr>
<td>ECBSL</td>
<td></td>
<td>85.7</td>
</tr>
<tr>
<td>Softmax Loss</td>
<td>Resnet101</td>
<td>87.0</td>
</tr>
<tr>
<td>LDAM Loss</td>
<td></td>
<td>87.9</td>
</tr>
<tr>
<td>Balanced Softmax Loss</td>
<td></td>
<td>87.0</td>
</tr>
<tr>
<td>ECBSL</td>
<td></td>
<td>87.0</td>
</tr>
<tr>
<td>Softmax Loss</td>
<td>Resnet134</td>
<td>85.6</td>
</tr>
<tr>
<td>LDAM Loss</td>
<td></td>
<td>86.3</td>
</tr>
<tr>
<td>Balanced Softmax Loss</td>
<td></td>
<td>86.4</td>
</tr>
<tr>
<td>ECBSL</td>
<td></td>
<td>86.6</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The rare wild animal species are focus of wild animal protection. It is possible to use camera traps to collect a lot of wild animal images in the wild animal habitat [27]. Installing infrared camera from different angle can capture complete wild animal images and biologists can distinguish and label different wild animal images. This provides a data basis for more efficient tracking and monitoring of wild animal. The wild animal recognition is the key technology to realize tracking and monitoring wild animal. It can count the number of wild animals and grasp their tracks. However, the collected wild animal image dataset has the long tail problem, because different wild animal species number exist huge difference in same wild environment. The long tail problem affects the recognition of rare wild animal species.
The ECBSL function proposed in this paper can improve the classification accuracy of tail classes by fitting PMI. It also calculates the distribution of wild animal species with effective number of samples. At the same time it improves the classification accuracy of the whole dataset. Therefore, our method provides a technical guarantee for effective tracking and monitoring of rare wild animal. But our method does not improve the head classes, and our dataset is small. In future work, we will test our method on more and larger datasets.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China under Grant 62176217, the Innovation Team Funds of China West Normal University under Grant KCXTD2022-3 and the Fundamental Research Funds of China West Normal University under Grant 19B045.

REFERENCES


Adversarial attacks on cross-resolution Person Re-Identification

Zhikang Song\textsuperscript{1a}\textsuperscript{*}

\textsuperscript{1}Department of International School of Information Science and Engineering, Dalian University of Technology, Dalian 116081, China
\textsuperscript{a}\textsuperscript{*}songzhikang@mail.dlut.edu.cn

Abstract

Person Re-Identification (re-ID) is a task that involves matching individuals captured by various cameras and in various poses. And because of the uncertainty of input image resolution in practice, traditional Person Re-Identification tends to have poor performance. Cross-Resolution Person re-ID is a hot branch of Person re-ID that aims to address the image with a different resolution. Besides, Person re-ID has been applied in lots of fields because of its practicality, including public security systems. Therefore, this poses a challenge to the security of the re-ID model. This paper discusses the results of attacks on PS-HRNet, VGG16, and ResNet networks using the Fast Gradient Sign Method. The experiment result of this paper proves the effectiveness of the Fast Gradient Sign Method on Cross-Resolution Person Re-Identification. It tests the transferable of the adversarial samples generated by the Fast Gradient Sign Method.

Keywords-Person Re-Identification; Cross-Resolution; Fast Gradient Sign Method; Adversarial attacks.

1. Introduction

The Person Re-Identification (re-ID) task is designed to compare the identity of the person from various perspectives captured by other cameras. Person re-ID has become a hot field in machine learning due to its practicality as a branch of recognition tasks. Many works have been done to design various network structures to improve the models' recognition accuracy, generalizability, and robustness. Although these network structures can alleviate the effects of changes in human posture and the influence of contextual content, they are all predicated on the idea that training and test sets will have the same resolution in particle. However, the resolution of the input image in the real world varies due to the camera parameters, the different positions, and the fact that people often make up only a certain percentage of the picture.

The inputs of low-resolution images tend to degrade the model's performance because of the difference in their distribution with high resolution images and the missing information caused by the low-resolution. Therefore, a Cross-resolution person's Re-ID has surged as a hot branch of Person re-ID to address the images with cross-resolutions. The earliest work mainly used metric and dictionary learning methods [1]. However, those methods only calculate the distances in the feature space and cannot recover the lost information due to low-resolution. Therefore, some works applied the super-resolution technique to cross-resolution person re-ID. One of the first applications of super-resolution modules to the re-ID task was SING, which formed a new network by combining SRCNN with the backbone of the network of re-ID [2]. Later, Zhang et al. improved the model's training process by incorporating the super-resolution module by training the low-resolution network with the super-resolution module and the high-resolution network jointly, improving the recognition performance substantially [3]. Besides, Chen et al. presented a new network structure to address cross-resolution Re-ID, replacing the original super-resolution model with an adversarial learning strategy [4]. The adversarial learning strategy is used to extract resolution-invariant re-recognition representation, like the auto-encoder structure, preserving the identification capability of the images.

In adversarial machine learning (AML), an attacker tries to confuse the machine learning model into making wrong decisions, usually by modifying the input data fed to the machine learning model during the training or inference phase. Due to the widespread use of AI in various fields in recent years, research on AML is on the rise to maintain the robustness of the model. Person Re-ID has been a hotspot in security systems for private and public to match mugshots due to its efficiency and accuracy. However, some criminals with strong anti-surveillance awareness may target the Person Re-Identification model to make a targeted attack to hide. A well-known adversarial attack was the Fast Gradient Sign Attack (FGSM), which was proposed by Ian J. Goodfellow et al. [5]. The method makes use of the loss function's gradient which is as part of input data to increase loss. However, since the training data and model structure are often unavailable to the attacker in most cases, later work has been done to use more general methods to attack the model without knowing the
training data and model structure. Tom B. Brown et al. presented a method for launching the attack by completely replacing a part of the image with a patch [6].

Three main networks are used in this paper to study the effect of adversarial samples on Cross-Resolution Person re-ID: PS-HRNet, ResNet50, and VGG16. And because of the efficiency and simplicity of the FGSM method, FGSM is used to generate adversarial samples to attack the Cross-Resolution Person re-ID model. This paper tests the robustness of PS-HRNet, ResNet50, and VGG16 under different attack intensities by FGSM. In addition, we also tested the transferable of the adversarial samples generated by FGSM in a specific model to attack other models.

2. Related work

This section introduces relevant research on classic re-ID, cross-resolution person re-ID, and adversarial machine learning.

2.1 Person re-ID

Due to its potential applications across numerous industries, person re-ID has lately become a hot topic in computer vision. Many network structures have been designed to alleviate the effects of background and human pose on model recognition performance. Zhao et al. proposed Spindle Net. Spindle Net could extract the 14 human vital points by skeleton key point extraction [7]. Then, these features are combined in a specific way at various scales to create a human re-identification feature that combines global and local features at various scales. Longhui We et al. proposed a GLAD to address the problem of human pose variation [8]. Currently, single-frame re-ID research is still mainstream because the data set is relatively small, and even a single GPU does not take too long to do one experiment. However, the information in a single frame image is usually limited, so many works focus on person re-ID methods using video sequences. Some results are based on video sequences for re-ID [9].

2.2 Cross-Resolution Person re-ID

Due to the variable resolution of the photos captured by cameras input to the model. Many methods have recently been put forth to address the cross-resolution re-ID to improve the robustness of the model. The earliest study mainly used metric and dictionary learning methods [1]. Those methods only compute the distances in the feature space, and do not recover the lost information caused by the low resolution. Later, because convolutional neural networks (CNN) have great performance in computer vision tasks, especially in super-resolution field, some approaches advocate combining super-resolution modules with CNN recognition networks to improve the overall recognition accuracy [10, 11]. Jiao et al. firstly propose an approach to connect SRCNN with recognition networks and corresponding training strategy [2]. To enhance the SR model's compatibility with re-ID in HR photo feature space, Cheng et al. provided INTACT, which could allow model to find potential association information between picture recovered by super-resolution model and person re-ID task and use it as an additional learning constraint [12]. And based on INTACT, Zhang et al. presented PS-HRNet, which uses VDSR and HRNet to improve model performance [3]. Other works have also attempted to address this problem without the super-resolution module. Chen et al. proposed new network structures to solve cross-resolution person reidentification, replacing the original applied image super-resolution model [4]. An adversarial learning strategy is proposed to extract resolution-invariant re-recognition representation, similar to the auto-encoder structure, and preserve the identification capability of the images.

2.3 Adversarial machine learning

Due to the widespread use of Artificial intelligence (AI) in the real world, there is a surge of work on the security of AI models. One of the branches is Adversarial machine learning (AML). Previously, researchers mainly focused on attacks based on accessible models or training datasets. The most primitive adversarial attacks mislead the model through slight additive noise. The FGSM [5] method was proposed by Ian J. Goodfellow et al. It makes advantage of the gradient calculated from the loss function and modifies the input data to increase the loss of model. But in real life, the model structure and training dataset are challenging to access. As a result, Patch attacks—a type of limited and obvious continuous disturbance of image pixels—emerged. Tom B. Brown et al. [6] presented a method for launching the attack by a patch which fully replaces a portion of the image. It could mask the patches to give them any shape and then train various images, applying random panning, scaling, and rotation to the patches in each image and optimizing using gradient descent.
3. Methods

This section introduces how to train PS-HRNet, ResNet50, and VGG16 in cross-resolution Person-ID problems. And it also introduces how to attack those models by FGSM.

3.1 PS-HRNet

Since PS-HRNet requires both low-resolution images and high-resolution images for training, we first need to down-sample the images. We define the set of N high resolution images as \( D_h = \{x_h^i, y^i\}_{i=1}^N \), where \( x_h^i \in \mathbb{R}^{H \times W \times 3} \), and the down sample rate \( r = 4 \). The low-resolution images obtained by down sample are denoted as \( D_l = \{x_l^i, y^i\}_{i=1}^N \). And PS-HRNet is composed of two modified modules which should be introduced firstly.

VDSR-CA is a super-resolution module. Adding a channel attention mechanism based on the VDSR [13] model structure, VDSR could have different weights in the channel, enabling the module to recover the high-frequency features better. Therefore, this mechanism can further help to improve the performance of PS-HRNet at low resolution. The \( c - th \) weight of channel \( w_c \) can be defined as following formula:

\[
w_c = H_{GP}(p_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W p_c(i, j)
\]

Where \( H_{GP} \) can be calculated by a global pooling function, and \( H, W \) denotes the hight and width of channel \( c \), respectively. \( p_c(i, j) \) is the pixel value at \( (i, j) \) of channel \( c \). For training the super-resolution model, the loss \( L_{SR} \) presenting the distance between low-resolution image and images recovered by the super-resolution module is calculated as:

\[
L_{SR} = \sum_{i=1}^N \| \mathcal{H}(x_l^i) - x_h^i \|_1
\]

Where \( \mathcal{H} \) denotes the super-resolution module, \( x_l \) denotes the LR image from dataset \( D_l \) and \( x_h \) denotes the HR image from dataset \( D_h \). And \( N \) is the mini-batch size.

HRNet is a human posture estimation network that is able to keep a high-resolution representation throughout the forward process [14]. For HRNet, several subnetworks which can propagate different resolution information among subnetworks are jointed in parallel after a high-resolution network. By continuously sharing data in parallel multi-resolution sub-networks, repeated multi-scale fusion is accomplished. Therefore, it can better retain the information of the corresponding scale in the process when receiving the low-resolution images that are already missing some information after the super-resolution module. As a result, PS-HRNet decides to use HRNet as the backbone to enhance the feature extraction capability to the LR image in Cross-Resolution Person Re-ID task.

For metric learning, PS-HRNet introduces a triplet loss to make the distances of the same objects as close as possible and to make the distances of different things as far as possible in the feature space, which can be calculated as follows:

\[
L_{BH} = \sum_{i=1}^P \sum_{a=1}^K \sum_{l=1}^5 \left[m + \max_{p=1...K} \| Se_{a,l}^{(i)} - Se_{p,l}^{(i)} \|_2 - \min_{j=1...p} \min_{n=1...K} \min_{j\neq a} \| Se_{a,l}^{(i)} - Se_{n,l}^{(i)} \|_2 \right] +
\]

where \( a, p \) and \( n \), respectively, stand for an anchor picture, a positive sample image, and a negative sample image. \( m \) is the margin parameter of triplet loss, which controls the boundary between intra- and inter-distances. Besides, for representation learning, PS-HRNet uses cross-entropy label smooth loss, which is defined as:

\[
L_{CE} = \sum_{n=1}^{P \times K} \left[ - \sum_{y=1}^M \log (p(y))q(y) \right]
\]

where \( M \) is the number of labels in the training set, and \( p(y) \) stands for the likelihood that label \( y \) will be present. Moreover, the \( q(y) \) in Equation (5) is defined as following:

\[
q(y) = \begin{cases} 
1 - \frac{M-1}{M} \varepsilon, & \text{if } y = y_{\text{truth}} \\
\frac{M}{M}, & \text{others}
\end{cases}
\]
To summarize all the above loss functions, the loss $L_{ID}$ of the HRNet section is defined as:

$$L_{ID} = \lambda_{CE}L_{CE} + \lambda_{BH}L_{BH}$$  \hspace{1cm} (6)

There are two primary steps in the training process. In the first step, HRNet-w32 pretrained in ImageNet is firstly trained on high-resolution dataset. This pre-trained HRNet-w32 with the addition of a classification module is called HRNet-ReID-H, which mainly helps the low-resolution network to train efficiently by distillation and provides guidance information corresponding to high-resolution images as the teacher network. In the second step, VDSR-CA, HRNet-w32 pretrained in ImageNet, and the classification block are combined to form HRNet-ReID-L. Then HRNet-ReID-H and HRNet-ReID-L are joined together in parallel. To enable the two networks to learn similar features, the distance between features is evaluated using Manhattan distance, denoted as $L_{ID}$.

$$L_{PS} = \sum_{i=1}^{m} \| C_h^{(i)} - C_t^{(i)} \|_1$$  \hspace{1cm} (7)

PS-HRNet define the overall loss $L_{TOTAL}$ as follows by weighing all the aforementioned loss functions:

$$L_{TOTAL} = L_{ID} + \lambda_{SR}L_{SR} + \lambda_{PS}L_{PS}$$  \hspace{1cm} (8)

To keep the teacher network from being affected by the results of HRNet-ReID-L, the teacher network is needed to freeze when training HRNet-ReID-L. In other words, we do not update the parameters of the teacher model during back-propagation. For testing, we only need to input low-resolution images into HRNet-ReID-L and use HRNet-ReID-L to extract features and output.

### 3.2 VGG16 and Resnet50

VGG16 is possible to gather more picture features while keeping the number of parameters under control, preventing overly complex structure and excessive computation caused by the network structure with more depth, smaller convolutional kernel, and pooling domain [11]. Therefore, it has been used as the backbone of many computer vision networks in the early years. Still, its depth is limited by the problem of gradient disappearance and explosion. This paper uses the original VGG16 pre-trained in ImageNet adding with a classification block to form the classification network as VGG16-ReID.

Through residual connections, Resnet50 [10] alleviates the gradient disappearance and gradient explosion caused by too deep a network, thus allowing the construction of deeper networks. Therefore, Resnet50 is widely used as a backbone for many computer vision tasks. This paper uses the original resnet50 pretrain in ImageNet plus a classification block to form the classification network as resnet50-ReID.

While training the two networks, they accept low-resolution datasets as inputs. To be consistent with the gradient calculation as PS-HRNet, the two models use cross-entropy label smooth loss for representation learning, consistent with what is mentioned in Equation (8) above. For metric learning, this paper modified the triplet loss in Equation (7) because of different outputs with PS-HRNet. And the new triplet loss $L_{tripletloss}$ is calculated as:

$$L_{tripletloss} = \sum_{i=1}^{p} \sum_{a=1}^{K} [m + \max_{p=1..K} \| Seq_{a,i} - Seq_{p,i} \|_2 - \min_{n=1..K} \| Seq_{a,i} - Seq_{n,j} \|_2]_+$$  \hspace{1cm} (9)

And the total loss of VGG16-ReID and Resnet50-ReID can be defined:

$$L_{TOTAL} = \lambda_{CE}L_{CE} + \lambda_{tripletloss}L_{tripletloss}$$  \hspace{1cm} (10)

### 3.3 Fast Gradient Sign Attack

A classic adversarial attacks in AML is the Fast Gradient Sign Attack (FGSM), which employs a straightforward but efficient technique to generate adversarial samples. It seeks to undermine neural networks by taking advantage of how the model is trained and optimized. The concept is straightforward: rather than decrease the loss by optimizing the weights by the back-propagation gradient, FGSM modifies the input data to increase the loss by the same back-propagation gradient. In other words, the approach makes use of the loss function’s gradient to increase the loss of image. The train dataset is defined as $D_{test}$, and the adversarial sample is calculated as below:
\[ x_p = x_t + \epsilon \text{sign} \left( \nabla_x J(\theta, x_t, y) \right) \]  
\hspace{1cm} (11)

Where \( x_p \) denotes the adversarial sample, \( x_t \) denotes the test image of low resolution, \( \epsilon \) control the intensity of attack, \( J(\theta, x_t, y) \) denotes the loss function of model with parameter \( \theta \). Fig. 1 shows an example about FGSM. This paper uses the loss gradient of the three models to generate adversarial samples, respectively. The generated adversarial samples are used to attack themselves and the other two models, respectively. The rank-k and mean average precision (mAP) of each model are calculated to evaluate the performance of the models.

![Fig. 1 An image in gallery after normalization using FGSM attack and how the image changes visually.](image)

4. Experimental results

This section first introduces dataset used for training and evaluation, and then specific settings for the experiments. Finally, the results of our adversarial attack experiments are presented.

4.1 Datasets

The Market-1501 dataset [15] is a classic dataset for person re-ID. It includes 1501 people and 32,668 rectangular person frames that were spotted by six cameras (5 HD cameras and one low-definition camera). At least two cameras are used to record each person, and each camera record more than one image of one person. The test set comprises 750 persons and has 19,732 photos in total, with an average of 26.3 test data per person. The training set has 751, with 12,936 images in total with 17.2 training data per person. Additionally, 3368 person detection rectangles are manually constructed for the query set. In contrast, the DPM detector is used to capture the person detection rectangles in the gallery set. Since this paper focuses on the classification task, this dataset's person detection rectangle boxes are not used. And the labels which consists of 751 classes’ label in the training dataset and 750 in the test dataset are extracted from the image files’ name. In this paper, the low-resolution images are obtained by down sampling the original training and test sets with a sampling rate of \( r = 4 \) from the high-resolution images.

4.2 Settings

When training for PS-HRNet, both low-resolution and high-resolution datasets are used to train the model. Two steps make up the model training procedure.: In the first step, HRNet-ReID-H is trained on high-resolution dataset, improving his feature extraction ability. In the second step, HRNet-ReID-H is a teacher network whose gradient does not backpropagate and provides HRNet-ReID-L with guidance information to help HRNet-ReID-L training. Therefore, HRNet-ReID-L can recover and retain high-resolution information efficiently.

In the testing phase, since the test set is divided into two parts, the query section and gallery section, the images of the query are downsampled, and the high resolution of the gallery is maintained. When producing confrontation samples, the gradients are only added to the features generated by the query while extracting features directly from the gallery. The model rank-k and mean average precision (mAP) are calculated to evaluate the model's performance.

**PS-HRNet.** Before training, images from training set are resized to same size with the length of 256 pixels, the width of 128 pixels, and the channel count of 3. A mini-batch for training consists of 24 pairs of images from \( P = 4 \) separate
identifications, with \( K = 6 \) pairs of LR and HR images as input for each identification. The specific setting is listed in Table 1.

**VGG16 and Resnet50.** Before training, all images are resized like PS-HRNet. A mini-batch for training consists of 128 pairs of images from \( P = 4 \) separate identifications, with \( K = 32 \) pairs of LR and HR images as input for each identification. VGG16 was trained for 40 epochs. ResNet50 was trained for 55 epochs. The specific setting of VGG16 and ResNet50 are listed in Table 2.

<table>
<thead>
<tr>
<th>hyper-parameters</th>
<th>( \lambda_{CE} )</th>
<th>( \lambda_{SH} )</th>
<th>( \lambda_{SR} )</th>
<th>( \lambda_{PS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Optimizer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>weight decay</td>
<td>Learning rate for HRNet-ReID</td>
<td>Learning rate for VDSR-CA</td>
<td></td>
</tr>
<tr>
<td>SGD</td>
<td>( 5 \times 10^{-4} )</td>
<td>( 8.5 \times 10^{-3} )</td>
<td>( 8.5 \times 10^{-4} )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hyper-parameters</th>
<th>( \lambda_{CE} )</th>
<th>( \lambda_{ripletioss} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Optimizer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>type</td>
<td>weight decay</td>
<td>Learning rate for VGG16 and ResNet50</td>
</tr>
<tr>
<td>SGD</td>
<td>( 5 \times 10^{-4} )</td>
<td>( 8.5 \times 10^{-3} )</td>
</tr>
</tbody>
</table>

Besides, when training, to improve the generalizability of three networks, some data augmentation techniques are utilized, including random flipping, padding, and cropping. When conducting adversarial experiments, \( \varepsilon \) is set to 0, 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3 to test the model's robustness under different attacks' intensity, respectively. Experiments are performed with PyTorch of version 1.8.1 and Cuda of version 11.1 on a 24GB NVIDIA RTX A5000 GPU.

### 4.3 Results

In the adversarial experiments, gradients from the model itself and other models are used to perform an adversarial attack with different intensities. And the impact of the attack on the model is evaluated by calculating the model's rank-k, and mAP, as shown in Fig. 2. PS-HRNet performs much better than ResNet50 and VGG16 in the cross-resolution person re-ID task when not under attack because of the addition of the super-resolution module and parallel training. FGSM is remarkable in generating adversarial samples for adversarial attacks based on each network's own generated gradients. And the adversarial samples generated using the gradients of other networks cause less performance degradation. Moreover, Adversarial samples generated using the model's gradients can enormously impact the model's performance even when the attack intensity is weak. Because differing little from the original images at low attack intensity, FGSM can be effective without being observed visually. The gradients of a high-performing model produce adversarial samples with better transferable, especially when targeting performing samples. In high-intensity attacks, the adversarial samples generated by the excellent performance model gradient can make the rank-k of the poor performance model down to the same level as those generated by the present model's gradient. The effect of the adversarial samples generated by the gradient of the low-performance model on the high-performance model is weak at low-intensity attacks. And only after the changes in the picture can be detected visually can they cause some attack effect. Therefore, the FGSM attack using the gradient of the low-performance model is less effective.
5. Conclusion

This paper investigates the effectiveness of the FGSM on cross-resolution person re-ID network. PS-HRNet, ResNet50, and VGG16 are taken as the study subjects. This paper firstly describes how each of these networks is trained. Then FGSM attacks were performed on these three networks, which were tried separately using the adversarial samples generated by the network's gradients and those generated by the gradients of other networks. The performance of PS-HRNet is the best among the three networks when no attack is performed. And the FGSM attack, when performed with the network's own generated gradients, could rapidly degrade the model's performance without visually detecting the attack. Finally, the result verifies that the adversarial samples generated by the model gradients with high performance can be better transferable, especially when attacking a model with poor performance.

References


A Method for Wireless Communication Interference Signal Identification Based on Extreme Learning Machine

XiaoZheng Liu\textsuperscript{a*}, Yue Wang\textsuperscript{b*}, XiaoFei Wang\textsuperscript{c*}, Jian Geng\textsuperscript{d*}
\textsuperscript{1}Beijing Institute of Space Long March Vehicle, China Academy of Launch Vehicle Technology, Beijing 100076, China
\textsuperscript{a*}liuxiaozheng12@163.com, \textsuperscript{b*}151933327@qq.com, \textsuperscript{c*}wangxiaofei8202@126.com, \textsuperscript{d*}gengjian_garry@sina.com

ABSTRACT

Intelligent anti-jam communication is a new generation of anti-interference technology combined with artificial intelligence, and the identification of interference signals is the basis of the technology. It is required to achieve better identification results with lower computational complexity in engineering applications. However, previous research has shown that they cannot balance these two sides. Here, we report an interference signal identification algorithm based on Extreme Learning Machine (ELM). Five typical oppressive interference signals were recognized based on ELM which is based on feature extraction. The overall correct identification rate is more than 96% under the condition of 40 neurons in a single hidden layer, and it has certain generalization ability. This study objectively promotes the engineering application of this technology.

Keywords-neural network; interference identification; ELM

1. Introduction

Artificial intelligence has been gradually applied to many fields with its powerful nonlinear mapping ability. As a new generation of anti-interference technology, intelligent anti-jam communication is deeply integrated with artificial intelligence, and adopts the method of learning and intelligent decision-making to achieve efficient and reliable information transmission.\textsuperscript{[1]} The effective identification of interference signals is the basis of intelligent anti-interference communication. The current research shows that the identification algorithm based on the back-propagation neural network has a low calculation amount, but its identification rate is not high,\textsuperscript{[2]} while the identification algorithm based on the convolutional neural network has a large amount of calculation due to its large number of network layers,\textsuperscript{[3]} which both couldn’t meet the real-time requirements in the engineering applications. Therefore, it is necessary to study the algorithm with low calculation amount and high identification rate. Extreme Learning Machine (ELM), as a new feedforward neural network, has attracted wide attention due to its advantages of simple implementation and fast operation speed.\textsuperscript{[4]} Therefore, in this paper, based on ELM, the typical oppressive interference signals are identified, and it is expected that a neural network with low computational complexity and high identification rate can be obtained through training, so as to promote the engineering application of this technology.

2. The System Model

2.1 Extreme Learning Machine\textsuperscript{[5]}

Given \( N \) arbitrarily different training dates \( \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^{N} \), where \( \mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \mathbb{R}^n \) is the input vector and \( \mathbf{t}_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in \mathbb{R}^m \) is the matching predicted output vector. The mathematical model of the ELM network, which has \( n \) input neurons, \( L \) hidden layer neurons, and \( m \) output neurons with activation function \( g(x) \) is as follows:

\[
H\beta = T
\quad (1)
\]

Where
\[ H = \begin{bmatrix} h(x_1^T), \ldots, h(x_N^T) \end{bmatrix}^T \]
\[ = \begin{bmatrix} g(w_1 \cdot x_i + b_i) & \cdots & g(w_L \cdot x_i + b_i) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_i) & \cdots & g(w_L \cdot x_N + b_i) \end{bmatrix}_{N \times L} \]  

(2)

\( H \) is also called as random feature map matrix in ELM, \( w_j = [w_{j1}, w_{j2}, \cdots, w_{jm}]^T \) represents the input weight connecting the \( j \)th hidden neuron and the input neuron, \( b_j \) represents the bias of the \( j \)th hidden neuron, \( \beta = [\beta_1, \beta_2, \cdots, \beta_L]^T \) represents the weight matrix between the output layer and the hidden layer, \( T = [t_1, t_2, \cdots, t_L]^T \) represents the expected output matrix of the training samples. After the hidden layer neuron parameters \((w_j, b_j)\) are randomly generated according to any continuous sampling distribution probability and given training samples, the hidden layer output matrix \( H \) is actually known and remains unchanged. In this way, equation (1) is transformed into the least norm least squares solution \( \hat{\beta} \) for solving the linear system \( H\beta = T \).

\[ \hat{\beta} = H^+ T \]  

(3)

Where \( H^+ \) represents the Moore-Penrose generalized inverse of the hidden layer output matrix \( H \).

### 2.2 The Interference Signals

In this paper, five typical oppressive interference signals are used as the signals to be identified, mainly including Single-Tone Jamming (STJ), Multiple-Tone Jamming (MTJ), Linear Frequency Modulation Jamming (LFM), Partial Band Noise Jamming (PBNJ) and Noise Frequency Modulation Jamming (NFM). The interference models are shown in the following table.

<table>
<thead>
<tr>
<th>Type of Interference</th>
<th>Interference Equation</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Tone Jamming (STJ)</td>
<td>( J(t) = A e^{j(2\pi f_c t + \varphi)} )</td>
<td>( A ) is interference amplitude, ( f_c ) is interference frequency, ( \varphi ) is interference initial phase</td>
</tr>
<tr>
<td>Multiple-Tone Jamming (MTJ)</td>
<td>( J(t) = \sum_{m=1}^{M} A_m e^{j(2\pi f_m t + \varphi_m)} )</td>
<td>( A_m ) is interference amplitude, ( f_m ) is interference frequency, ( \varphi_m ) is interference initial phase</td>
</tr>
<tr>
<td>Linear Frequency Modulation Jamming (LFM)</td>
<td>( J(t) = A e^{j(2\pi f_c t + \pi k^2 t^2 + \varphi)} ) 0 ( \leq t \leq T )</td>
<td>( A ) is interference amplitude, ( f_0 ) is interference frequency, ( k ) is frequency modulation rate, ( \varphi ) is interference initial phase, ( T ) is the single cycle</td>
</tr>
<tr>
<td>Partial Band Noise Jamming (PBNJ)</td>
<td>( J(t) = U_n(t) e^{j(2\pi f_c t + \varphi)} )</td>
<td>( U_n ) is Gaussian noise with mean 0 and variance ( \sigma_n^2 ), ( f_c ) is the center frequency of the interference, ( \varphi ) is interference initial phase</td>
</tr>
<tr>
<td>Noise Frequency Modulation Jamming (NFM)</td>
<td>( J(t) = A e^{j(2\pi f_0 t + 2\pi k_{mf} \int \xi(t) dt')} )</td>
<td>( A ) is interference amplitude, ( f_0 ) is the carrier frequency of the interference, ( k_{mf} ) is frequency modulation rate, ( \xi(t) ) is narrowband Gaussian noise with mean 0 and variance ( \sigma_n^2 )</td>
</tr>
</tbody>
</table>

The parameters of the above interference signals are set as follows:
Table 2 The interference parameters

<table>
<thead>
<tr>
<th>Type of Interference</th>
<th>Sampling Parameters</th>
<th>Interference signal parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Tone Jamming (STJ)</td>
<td>Bandwidth: 10MHz</td>
<td>( A ) is 1</td>
</tr>
<tr>
<td></td>
<td>Sampling period: ( 10^{-7} ) s</td>
<td>( f_c ) is a random value in -5~5MHz</td>
</tr>
<tr>
<td></td>
<td>Number of sampling points: 1024</td>
<td>( \varphi ) is a random value</td>
</tr>
<tr>
<td></td>
<td>Number of FFT points: 1024</td>
<td>( m ) is a random value from 2 to 10</td>
</tr>
<tr>
<td>Multiple-Tone Jamming (MTJ)</td>
<td></td>
<td>( A_m ) is 1</td>
</tr>
<tr>
<td>Linear Frequency Modulation Jamming (LFM)</td>
<td></td>
<td>( f_m ) is a random value in -5~5MHz</td>
</tr>
<tr>
<td>Partial Band Noise Jamming (PBNJ)</td>
<td></td>
<td>( \varphi_m ) is a random value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( m ) is a random value from 2 to 10</td>
</tr>
<tr>
<td>Noise Frequency Modulation Jamming (NFM)</td>
<td></td>
<td>( A ) is 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( f_0 ) is a random value in -5~5MHz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( k ) is a random value between 1~8GHz/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \varphi ) is a random value</td>
</tr>
</tbody>
</table>

Table 3 Feature parameters

<table>
<thead>
<tr>
<th>Feature parameters</th>
<th>Calculation Equation</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Metric of Peak Energy</td>
<td>( C = \sum_{n=m-k}^{n-k} F^2(n) / \sum_{n=1}^{N} F^2(n) )</td>
<td>( F(n) ) is the spectral amplitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( m ) is the index of spectral peak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( k ) is set to 1</td>
</tr>
<tr>
<td>Average Spectrum Flat Coefficient</td>
<td>( F_c = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (P(n) - P_0(n))^2} )</td>
<td>( P_s(n) = \frac{P(n)}{P(0)} ), where ( P(n) ) is the power spectrum of the signal</td>
</tr>
</tbody>
</table>

3. Interference Signal Preprocessing and Feature Extraction

In order to avoid the influence of signal strength on identification, it is necessary to perform power normalization preprocessing on the sampled time domain signal as shown in equation (4), where \( J(n) \) represents the complex baseband signal with \( N \) sampling points.\(^{[5]}\)

\[
\overline{J}(n) = \frac{J(n)}{\sqrt{\frac{1}{N} \sum_{n=1}^{N} |J(n)|^2}} \tag{4}
\]

At this time, if the preprocessed time domain signal is directly input into the interference identification network, the identification effect will be poor due to the excessively high sample dimension. The feature parameter extraction of the interference signal is the key to the identification algorithm based on feature extraction. The feature parameters extracted for the interference signal in this paper are: Aggregate Metric of Peak Energy \( C \), Average Spectrum Flat Coefficient \( F \), Frequency Domain Moment Skewness Coefficient \( b_3 \), Frequency Domain Moment Kurtosis Coefficient \( b_4 \), Time Domain Moment Kurtosis Coefficient \( a_4 \), and Envelope Variation Parameters \( R_f \). Their calculation method is shown in the table below.
Set the simulation Jamming to Noise Ratio (JNR) to -20~20dB, the interval is 2dB, and count the average value of 100 samples under each JNR. The following feature extraction results after time domain signal preprocessing can be obtained:

- **Aggregate Metric of Peak Energy**
  
  ![Graph of Aggregate Metric of Peak Energy](image1)

- **Average Spectrum Flat Coefficient**
  
  ![Graph of Average Spectrum Flat Coefficient](image2)

- **Frequency Domain Moment Skewness Coefficient**
  
  ![Graph of Frequency Domain Moment Skewness Coefficient](image3)

- **Frequency Domain Moment Kurtosis Coefficient**
  
  ![Graph of Frequency Domain Moment Kurtosis Coefficient](image4)

- **Time Domain Moment Kurtosis Coefficient**
  
  ![Graph of Time Domain Moment Kurtosis Coefficient](image5)

- **Envelope Variation Parameters**
  
  ![Graph of Envelope Variation Parameters](image6)
4. Simulation Results and Discussion

4.1 The Influence of the Number of Hidden Neurons

In the ELM algorithm, the number of input neurons and output neurons is determined by the number of features of the input data. In the third section, we extract 6 kinds of features for the interference signal, so the number of inputs is 6. The number of output neurons correspond to the number of the interference signal types to be identified equal to 6.

The identification efficiency is determined by the number of hidden neurons. We first examine the effect of the number of hidden neurons on the identification results.

The JNR of the training data is -5 to 30dB, the interval is 0.5dB, and the number of samples of each interference signal under each JNR is 100; the JNR of the test data is -10 to 20dB, the interval is 0.5dB, the number of samples of each interference signal under each JNR is 30, and the activation function is set as

\[ \text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (5) \]

The following training results can be obtained as Figure 2.

As shown in the Figure 2, when the number of hidden neurons is between 10 and 40, as the number of neurons increases, the overall correct recognition rate rises slowly, but to a limited extent. The overall identification rate is stable at 94.6% when the number of neurons is greater than 40. The simulation results show that the interference recognition algorithm based on feature extraction can obtain a better identification effect with a small number of training samples after reducing the dimension of the input signal, but its recognition performance is also limited by the parameters of feature extraction. In order to reduce calculated amount, shorten the calculation time and ensure the recognition rate, the number of hidden layer neurons is set to 40 in the future training.
4.2 The Influence of the Number of Training Samples

After determining the number of input neurons, output neurons and hidden neurons, the number of training samples will also affect the training effect of the network. The effect of the number of training samples on the identification effect is studied as follows.

The JNR of the training data is -5 to 30dB, and the interval is 0.5dB; the JNR of the test data is -10 to 20dB, and the interval is 0.5dB, and the number of samples of each interference signal under each JNR is 30. The simulation results for different numbers of training samples are as follows:

It is shown that with the increase of the number of training samples, the overall correct identification rate finally stabilizes at 94.44%. The maximum overall correct identification rate is 96.12% at the number of 60 training samples per interference signal under each JNR. The network parameters with the largest overall correct identification rate were selected as the future research object.

4.3 The Influence of Different Training Sample Distributions

Different distributions of training samples also affect the identification performance of the network.

The JNR of the training samples are set to be 0 to 30dB, -5 to 30dB and -10 to 30dB respectively, and the intervals are all 0.5dB. The number of samples of each interference signal under each JNR is 60. Set JNR of the test samples to -20 to
20dB with 0.5dB interval. The number of samples of each interference signal under each JNR is 30. The following test results can be obtained:

![Figure 4](image-url) The overall correct identification rate under different training sample distributions

It can be seen from the Figure 4 that when the training data distribution is 0 to 30dB, the overall correct identification rate reaches 98.68% when the JNR is -2.5dB; when the training data distribution is -5 to 30dB, the overall correct recognition rate reaches 99.58% at the JNR of 5dB; when the training data distribution is -10 to 30dB, the overall correct recognition rate reaches 98.17% at the JNR of -6.5dB. The results suggest that with the expansion of the distribution range of training samples, the overall correct identification rate gradually increases under the same JNR.

### 4.4 Correct Identification Rate on Different Interference Signals

Take the optimal network parameters obtained in 4.2 as the network input. Set the JNR of the test data to -20 to 20dB, the interval is 0.5dB, and the number of samples of each interference signal under each JNR is 30. The following test results can be obtained:

![Figure 5](image-url) The correct identification rate under different interference signals

From the above figure, it can be seen that when the JNR is greater than -6dB, the correct identification rate of all interference signals reaches more than 98%. STJ reaches 100% identification rate at -10dB. MTJ reaches 100% identification rate at -8dB. LFM reaches 99.5% identification rate at -6dB, and 100% at -4dB. PBNJ reaches 98.6% identification rate at -5dB, and 100% at -4dB. NFM reaches 99.1% identification rate at -6dB, and 100% at -4dB. The correct identification rate of non-interference (pure noise) signals remained at 98.1%.
From -20dB to 20dB, the overall correct identification rate is 77.67%, when it is -10dB to 20dB, the overall correct identification rate is 96.02%, and when it is -5dB to 20dB, the overall correct identification rate is 99.65%. The above results suggest that the identification algorithm based on ELM can still maintain high recognition efficiency with low computational complexity, which verifies our previous conjecture.

5. Conclusion

In this study, the researcher took five typical oppressive interference signals as the identification object and studied an ELM-based neural network with low computational complexity and high identification rate. This study also analyzed the factors that affect its identification performance, including: the number of hidden neurons, the number of training samples, and the distribution of training samples. Based on the results and discussions presented above, the main conclusion can be summarized as follows:

(1) The interference identification algorithm based on ELM can achieve a high level of overall correct identification rate with low computational complexity. Under the condition of 40 neurons in a single hidden layer, the overall correct recognition rate can achieve 96.02% for typical oppressive interference signals with JNR between -10 and 20dB, and 99.65% for interference signals with JNR between -5 and 20dB.

(2) The number of hidden neurons affects the identification effect of the network. In the early stage, with the increase of the number of hidden neurons, the overall correct identification rate increases gradually. When the number of hidden layer neurons reaches 40, the overall correct identification rate tends to be stable at 94.6%.

(3) The number and distribution of training samples also affects the identification effect of the network. Other things being equal, when the number of training samples is 60, there is the largest overall correct identification rate, and the larger the distribution range of training samples, the higher the overall correct identification rate of the identification network under the same JNR.

References

Knowledge Graph Construction of Component Quality Management

HaiMing Zhang\textsuperscript{1a}, XiaoMing Fan\textsuperscript{1b}, JiaQi Zhang\textsuperscript{1c}, ChengZhi Jiang\textsuperscript{1d}, Jiang Li\textsuperscript{1e}, HanTian Gu\textsuperscript{1f}, BoHan Li\textsuperscript{1g}, Hao Hu\textsuperscript{2h}, ChengXi Liu\textsuperscript{1i}  
1 China Academy of Space Technology, Beijing;  
2 Aerospace ShenZhou Smart System Technology Co.,Ltd.  
\textsuperscript{a}dantahaiming@sohu.com, \textsuperscript{b}fanxming@163.com, \textsuperscript{c}zhangjiaqiz77@163.com, \textsuperscript{d}jiangchengzhi@tsinghua.org.cn, \textsuperscript{e}lijiangyio@163.com, \textsuperscript{f}ght19860111@sina.com, \textsuperscript{g}yuanyi56@sina.com, \textsuperscript{h}huhao32993@aliyun.com, \textsuperscript{i}wzb_lcx@163.com  

Abstract

With the development of Industrial Internet of Things, the types and functions of components are increasing, the application environment is becoming more and more complex. Also, the quality management of components is becoming more and more important. In order to understand the knowledge related to component quality management more conveniently and build an intelligent system for component quality management, this paper proposes a method to construct component quality management knowledge graph based on BERT word embedding model and entity relationship joint extraction method based on annotation strategy. Combining entity extraction and relationship extraction parts into one not only reduces the consumption of computing resources, but also reduces the propagation of wrong entities. In this paper, the sequence to sequence model of Bert-BiLSTM-CRF is adopted. Through the BERT word embedding layer, the context information can be better utilized and the accuracy of extraction can be improved. Experimental results show that compared with other classical deep learning term extraction models, this model has a significant improvement in accuracy, recall rate and F1 value.

Keywords-component quality management; BERT; sequence model

1. Introduction

Along with the development of the Industrial Internet of Things, the complexity of electronic equipment used in various household appliances, communication hardware, medical equipment and weapons equipment has become increasingly high. And the environmental conditions used by the more severe, the density of electronic devices is also increasing, so the electronic components put forward higher requirements, not only requires good characteristics, but also requires a highly reliable work\cite{1}. Electronic components constitute the basic unit of equipment, the reliability of components is the basis of all kinds of electronic equipment reliability. Therefore, how to effectively manage the quality of electronic components and improve the reliability of products is an urgent problem to be solved\cite{2}. However, in the field of component quality management, there is a lack of related term base and knowledge graph research.

Aiming at the task of building component quality management knowledge graph, this paper adopts the end-to-end entity relationship joint extraction model based on Bert-BiLSTM-CRF. Using BIOES's field-oriented labeling method for joint extraction of overlapping entity relations, directly model RDF, improve the efficiency of model recognition, reduce entity overlap, error propagation and entity redundancy and other issues. This model solves the problem of entity extraction with multiple meanings and the long distance between entities. It also adopts the entity relationship joint extraction model based on labeling strategy, which reduces the errors caused by the pipeline method and the joint extraction model based on parameter sharing Propagation and entity redundancy, relationship loss and other issues. Use the entities and relationships extracted from this model to build a knowledge graph of component quality management, help relevant staff to better understand and use component quality management related technologies, build an intelligent component quality management system, and improve component quality.
2. Related Work

Knowledge graph is a graphical representation of structured knowledge from the real world, where nodes represent entities of interest and edges represent relationships between those entities. In essence, knowledge graph aims to describe various entities or concepts and their relationships in the real world. It forms a huge semantic network graph, where nodes represent entities or concepts and edges are composed of attributes or relationships. Now knowledge graph has been used to refer to a wide variety of large-scale knowledge bases. Back in 2001, Berners Lee, the father of the Internet, proposed the concept of the Semantic Web. He believed that web pages could be used to store semantic information that computers could understand, thus enabling machines to understand human semantic knowledge. In 2006, it was proposed to use ontology language OWL and RDF triples to represent semantic knowledge. Knowledge is formed in the form of triples, a vast network of interconnected triples. Each node in the network represents an entity, and each edge represents a relationship or intrinsic property of the entity. Since Berners Lee proposed semantic web in 2001, a large number of new generation knowledge bases based on Internet resources have emerged. For example, Google's Freebase, China's Zhisi.me, Wikipedia's Dbpedia, CN-dbpedia.

The architecture of the knowledge graph includes its own logical structure and the technology (system) architecture used to construct the knowledge graph. Logically, the knowledge graph can be divided into a pattern layer and a data layer. The data layer is composed of a large number of established triples stored in graph databases such as Neo4j, Microsoft Trinity, Twitter's FlockDB and so on. The pattern layer is a structure built on top of the data layer and usually consists of many ontology. With the help of ontology, it further manages and organizes the connections among ontology, relationships and attributes, so as to reduce redundancy.

In recent years, various fields have constructed their own corresponding knowledge graphs, among which the development of knowledge graphs in the fields of medicine, education, finance and e-commerce has been very mature. Haussmann proposed a comprehensive food knowledge graph, reflecting knowledge of healthy food, recipes and nutrition. Penghe expresses knowledge in the form of graphs, which improves learners' learning efficiency. Wang uses the financial knowledge graph to detect online fraud. Kim applied knowledge graph technology in the music field and formed a knowledge graph belonging to the music field. Wang uses a comprehensive corpus composed of geological dictionaries and geological and mineral resource terms and classification codes to construct a knowledge map from earth science literature.

In view of the excellent performance of the BERT pre-training language model in English natural language processing tasks in recent years, automatic mining of hidden features can effectively solve the characteristics of discovering new words, while reducing the problems of artificially defined features and excessive dependence on domain knowledge. This research uses the method of joint recognition of Chinese entity relationship based on BERT, extracts a large number of knowledge information related to component quality management from books and literature related to component quality management, and establishes a component knowledge map. These books and documents are unstructured data sets, so this research will use deep learning-based natural language processing technology to extract knowledge content related to component quality management.

3. Term extraction model based on BERT-BiLSTM-CRF model

3.1. Overall framework of component quality management term extraction model

The overall structure of the BERT-BiLSTM-CRF component quality management term extraction model is shown in Figure 1. The first is the BERT pre-training language model layer. The labeled character-level corpus passes through this layer to convert each character into a low-dimensional word vector. The second is the BiLSTM layer, which inputs the word vector sequence output from the previous layer to this layer for semantic coding and automatically extracts sentence features. Finally, the CRF layer uses this layer to decode the predicted label sequence with the highest output probability, obtain the label type of each character, extract and classify the entities in the sequence, and finally realize the extraction of component quality management terms. Compared with other deep learning term extraction models, the main difference between this model and other deep learning term extraction models is that BERT acquired by Google in large-scale Chinese corpus is used to pre-train Chinese vector, because of its stronger ability of context-long distance semantic learning. It can better solve the problem of polysemy of word vector, dig the features of component quality management text, and provide richer semantic information for downstream tasks.
3.2. BIOES labeling strategy

This research adopts the BIOES labeling strategy. Compared with BIO's conventional tagging method, BIOES can better reflect the diversified attributes of words. Among them, B stands for the beginning character of the word, E stands for the end character of the word, I stands for the non-beginning and ending characters in the word, O stands for other characters, and S stands for a single character. Since this study adopts the method of joint identification of entity relations, relationship types should be added to the annotated data. For example, CP refers to country-president relationship. Secondly, the position of the entity in the relationship needs to be added to the label information. The rules are as follows: \{1 (head entity), 2 (tail entity)\}, which represents the position of the entity in a set of relationships. Among them, this rule can also be divided into 1-1 and 1-N annotation methods. 1-N corresponds to the relationship containing overlapping entities in a sentence.

![Fig.1 BERT-BiLSTM-CRF model framework](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)
matrix. Then, through matrix operation, the final result matrix is input into feedforward neural network. In order to form a different semantic environment.

BERT also uses a positional coding technique to solve the problem of interpreting the word order of the input sequence. Here, BERT introduces a position vector, which determines the position of the current word, or the distance between different words in a sentence. Finally, add the position vector to the value of the word vector as input to the first Encoder.

In the BERT pre-training process, it includes two techniques, Masked Language Model and Next Sentence Prediction. In MLM, 15% of the words in the sentence will be removed for contextual prediction. This step is unsupervised learning, and the accuracy of the pre-training will be increased as much as possible through parameter adjustment. In NSP, the model selects two sentence pairs, in which there is 50% possibility that sentence B is the next sentence of sentence A, and the remaining 50% is the sentence pairs randomly selected, which can satisfy the current task of understanding the relationship between sentences in many NLP tasks.

3.4. BiLSTM bidirectional long and short-term memory network

LSTM, which stands for Long short-term Memory, is an upgrade model of RNN, which can effectively utilize the information of Long distance timing. LSTM is ideal for modeling time series data, such as text data, because of its characteristics. BiLSTM is Short for bi-directional Long short-term Memory. It is a combination of forward LSTM and backward LSTM. BiLSTM can better capture the bidirectional context semantic information. Among them, LSTM proposed by Hochreiter et al., leads to the concepts of "forgetting gate", "input gate" and "output gate" \[23\]. Through these gates, the update, deletion, input and output of memory information are completed, and the utilization of long-distance information is realized. The calculation method is as followings:

![Fig.2 BiLSTM structure](image)

Forgotten Gate :
\[ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \]

Input gate:
\[ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \]

Transformation:
\[ \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \]

State Update:
\[ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \]

Output gate:
\[ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \]
\[ h_t = o_t \cdot \tanh(C_t) \]

According to the center, \( \sigma \) is the activation function, \( W_f \) and \( W_i \) are weight matrices, \( b_f \) and \( b_i \) are bias vectors, and \( \tilde{C}_t \) is according to the current state of the incoming units. \( C_t \) is the updated state at time \( T \), and the symbol \( \cdot \) indicates multiplication by elements. \( h_t \) is the output of the entire LSTM unit at time \( T \).
3.5. CRF conditional random field

CRF is a sequence labeling algorithm that receives an input sequence $X=(X_1, X_2... X_n)$ and finally get the target sequence $Y=(y_1,y_2... ,y_n)$, was first proposed by Lafferty in 2001[24]. This is also a model of SEQ2SEQ, which is often used for part-of-speech prediction. $X$ and $Y$ are random variables, $P (Y | X)$ is the conditional probability distribution of a given $X Y$, if random variable $Y$ is a Markov random field that is called conditional probability distribution $P (Y | X)$ is the conditional random field. Where, if the CRF of $X$ and $Y$ have the same structure, the linear chain member random field is formed. CRF learns by means of feature functions and their weights, among which, feature functions are divided into node feature functions and local feature functions. Both node eigen functions and local eigen functions can only be 0 or 1. The feature condition is satisfied or the feature condition is not satisfied. At the same time, we can assign a weight to each eigen function, which expresses how much we trust the eigen function.

Linear-crf can be simplified as:

$$P(Y|X) = \frac{1}{\sum_{y \in Y} w_k f_k (y, x)} \prod \left( y, x \right) \exp \left( w_k f_k (y, x) \right)$$

$$Z(x) = \sum \exp \sum f_k (y, x)$$

4. Experimental verification

4.1. Acquisition of experimental data

The experimental data are derived from the relevant literature abstracts with the keywords "components" and "quality management" in Google Scholar. In this paper, Python libraries such as BeautifulSoup, Selenium, Requests and other Python libraries are used to crawler the abstract parts of the literature with keywords of "components" and "quality management" in Google Scholar through web crawler method. Selenium can be used to simulate browser page clicks, and BeautifulSoup can parse and recognize HTML source code. This method can obtain a large number of relevant literature in a short time and reduce the burden of data collection. At the same time, this paper uses OCR technology to convert PDF files into TXT files for the collected electronic versions of books.

4.2. Data annotation and evaluation indicators

In this paper, a domain-oriented BIOES annotation method based on joint extraction of overlapping entity relations is used, where B represents the beginning character of a word, E represents the end character of a word, I represents the non-beginning and end character of a word, O represents irrelevant character, and S represents single character. Since this paper adopts the method of entity relationship joint identification, relationship types need to be added to the annotated data.

In this experiment, three indexes commonly used in machine learning are used to evaluate the quality of the model, which is accuracy rate $P$, recall rate $E$ and $F1$.

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP + FN}{TP + FN}$$

$$F1 = \frac{2 \times PR \times RE}{PR + RE}$$

4.3. Experimental Environment and Parameter Configuration

4.3.1. Experimental Environment Configuration

The bert-bilstm-crf component quality management term extraction model runs on the 64-bit Ubuntu20.04 operating system, and the hardware parameters are shown in table 1:
The project runs on the Python+Pytorch framework. Anaconda is responsible for managing the Python virtual environment, which is supported by CUDA10.1 and CUDN7.6 and accelerated by GPU.

In the training of deep learning model, the quality of parameter setting often determines the final accuracy of the model. In the bert-BiLSTM-CRF model, the experiment will analyze the influence of learning rate and iteration number on the training effect of the model.

Learning rate is an important parameter to control the convergence rate of deep learning model, which is mainly used in the calculation of the gradient descent method. Large learning rate can make the model converge quickly, but it is more likely to miss the optimal solution. Although a small learning rate slows down the learning speed, it can be more accurate and convergent. Therefore, how to select an appropriate learning rate parameter greatly determines the final training effect of the model. In this experiment, the batch_size of training was fixed as 8, and the learning rate was gradually improved from 5E-6. The final experimental results are shown in the figure. As can be seen from the figure, when the learning rate is 1E-5, it reaches the highest recall rate, but its accuracy is low. When the F1 value and accuracy were observed, it was found that the F1 value and accuracy reached the highest at 2E-5, and then showed a downward trend. Therefore, 2E-5 was selected as the learning rate in the experiment.

The number of iterations is the number of times the model needs to train all the samples. Too small number of iterations will lead to under-fitting, and eventually lead to poor training effect of the model. However, too large number of iterations will lead to over fitting of the model, resulting in weak generalization ability of the model and poor performance on the unseen data sets. In this experiment, the batch_size and learning rate determined above will be fixed, and the number of iterations will be between 3-20 for comparison experiment. The results are shown in the figure below. It can be seen that when the number of iterations reaches 15 rounds, the three indexes reach the highest, and after 15, the indexes gradually tend to be stable, that is, the convergence state. The model before 15 rounds is still in a state of under fitting, and its index keeps rising with the improvement of the number of iterations. Therefore, in the subsequent model comparison, we also determined the number of iterations to be 15 rounds.

Fig.3 Relationship graph of learning index and learning rate
4.4. Experimental results and analysis

In order to demonstrate the superiority of the Bert-BiLStm-CRF model adopted in this paper, three kinds of knowledge extraction models are selected for comparative experiments. They are:

1. HMM model

HMM model is a directed graph probability model\(^25\), which contains hidden state sequence and observable observation sequence. The HMM model is based on two basic assumptions: the state at any time \(T\) depends only on the state at the previous time, and has nothing to do with other observations or states; The observation at any time \(T\) is only related to the state of the Markov chain at that time, and not to other observations or states.

2. BiLSTM model

The bidirectional long term short-term memory network is composed of two reverse long term short-term memory networks\(^26\). The advantage of bidirectional long term short-term memory network is that it can use sequence context information for feature learning at the same time, which is the upgraded version of the RNN model.

3. BiLstm-crf model

Although BiLSTM model has been able to achieve feature extraction and prediction of sequences, its disadvantage is that it cannot learn the relationship between state sequences (output annotations). The advantage of CRF model is that it can model the implicit state and learn the characteristics of a state sequence\(^27\). Therefore, to combine the best of the two models, the NER task typically connects a CRF layer behind the BiLSTM layer to combine the best of the two models.

This experiment is based on Bert_Base developed by Google and the pre-trained Chinese corpus. Bert_base has 12 Transformer layers, 768 hidden units, 12 multi-head-attention, and a total number of 110M. In terms of parameter selection, this experiment adopts the learning rate of 2E-5, 15 iterations, the drop_out rate of 0.5 and the batch_size of 8. The final experimental results are shown in the table 2:

Table 2. Comparison of different knowledge extraction models

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>73.13%</td>
<td>74.06%</td>
<td>75.17%</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>74.71%</td>
<td>77.05%</td>
<td>70.86%</td>
</tr>
<tr>
<td>BiLSTM-CRF</td>
<td>76.60%</td>
<td>78.09%</td>
<td>72.59%</td>
</tr>
<tr>
<td>BERT-BiLSTM-CRF</td>
<td>79.25%</td>
<td>83.92%</td>
<td>78.22%</td>
</tr>
</tbody>
</table>

It can be found from the results that the model using BERT as word embedding layer achieves the best performance in all indicators. Compared with the traditional Word2Vec word embedding model, the accuracy of BERT model is improved.
by 2.35%, the recall rate is improved by 5.87%, and the F1 value is improved by 5.63%. The results show that BERT word embedding model can better combine the context. Thus improving the training effect of the model.

5. Conclusion

This paper introduces the most important link in the process of building component quality management knowledge graph, the establishment and use of automatic knowledge extraction model based on deep learning. Aiming at the problem of entity error propagation in the traditional pipeline model, this paper adopts the joint learning method based on the annotation strategy, so as to reduce the occurrence of errors and redundant information. Aiming at the problem that the traditional Word2Vec model cannot make good use of contextual information, this paper adds the BERT word embedding model developed by Google on the basis of BiLSTM model to make better use of contextual information. After that, the experiment of joint extraction of component quality management is carried out in this paper. After parameter adjustment, the experiment is also compared with the other three models, showing the superiority of the model in this paper. In this paper, the next step of the priority to continue to expand the field of core dictionary, in the existing model based on the results of extracting rules for efficient and feasible filter components quality management terms, automatic tagging and train more large-scale components quality domain-specific term extraction model, further improve the generalization of the model, so as to build a richer components quality domain terminology dictionary.

References

Remote sensing landslide recognition method based on LinkNet and attention mechanism

Jing Yang1a*, Yaohua Luo2b*, Xuben Wang2c, Haoyu Tang1d, Shanshan Rao1e

1 College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China
2 Key Lab of Earth Exploration & Information Techniques of Ministry Education, Chengdu University of Technology, Chengdu 610059, China

a*1914148674@qq.com, b*lyh@cdut.edu.cn, c wangxb@cdut.edu.cn, d thy@stu.cdut.edu.cn, e raoshanshan@stu.cdut.edu.cn

Abstract

Rapid detection and identification of landslide areas are very important for disaster prevention and mitigation. Aiming at the problems of time-consuming and labor-intensive traditional landslide information extraction methods and low recognition efficiency, a remote sensing landslide recognition method based on LinkNet and convolution attention module was proposed. The model adopts the coding-decoding structure to improve the operation efficiency. The Convolutional Block Attention Module (CBAM) is applied to optimize the weight allocation from both channel and spatial dimensions to highlight the landslide feature information. And compared with the traditional U-Net and LinkNet models. The results show that the CBAM-LinkNet model has excellent performance in remote sensing landslide identification, which provides the possibility for rapid and accurate landslide identification.

Keywords: remote sensing, seismic landslide, LinkNet, target recognition, convolutional attention module

1. Introduction

With the aggravation of global extreme climate and the impact of seismic activities, as well as the rapid development of human engineering activities, more and more damage to the natural environment, directly leading to frequent geological disasters. Landslide is a natural disaster phenomenon, its occurrence process usually involves the movement of soil, rock, vegetation and other objects. The types of landslide movement include sliding and flow. Landslides will cause great persecution to residential buildings, roads, service facilities, and personal safety at the foot of the mountain; The landslide flowing to cultivated land damages crops and causes economic harm. Landslide flow to the river caused by river blockage formed a barrier lake, forming greater potential harm. After a landslide, it is the key to rescue and recovery work to acquire the location information of the landslide by using Remote Sensing Image, and it is of great value and significance to detect a landslide. Landslide detection is to detect the geographical location and shape attributes of landslides. Therefore, it is very important to detect the landslide area quickly and accurately.

There are many methods for landslide disaster detection. Traditionally, extraction and analysis of landslide information have been challenging tasks with many problems, including low efficiency, high costs, and unintuitive results. With the rapid development of space technology and electronic information technology, the application of remote sensing technology in large-scale geological disaster investigation is becoming more and more common. At present, remote sensing landslide identification methods mainly include the following four categories.

1.1 Visual interpretation

This method mainly relies on professionals to judge the landslide area according to the shape, tone, texture, and layout of ground objects[1]. Then some non-remote sensing data are used for analysis and inference to achieve higher information accuracy. Although visual interpretation results are generally accurate, it is a time-consuming and costly task[2].

1.2 Pixel-based landslide extraction

Pixel-based landslide recognition classifies each pixel without considering adjacent pixels. The commonly used methods include maximum likelihood, random forest, and support vector machine (SVM)[3,4]. Because only the spectral features
of a single-pixel are considered, enough shape and topographic features cannot be provided for landslide recognition. But because it is fast and saves resources, it is currently the most commonly used method.

1.3 Object-based landslide extraction

The object-oriented landslide recognition method uses objects to reduce spectral variance correlation in each information class, and combines the features related to objects, such as shape and texture, to make the results more reasonable\cite{5,6}. But the disadvantage is that noise information will be generated in the process of segmentation, and these units will be ignored in classification, so there will be some limitations.

1.4 Artificial intelligence algorithm landslide extraction

With the breakthrough of artificial intelligence algorithms, deep learning technology has made substantial progress and become a research hotspot in various fields. Landslide recognition in remote sensing images can be divided into traditional machine learning algorithms and deep learning algorithms\cite{7}. The machine learning algorithm only needs to input feature layers and training samples that can reflect the difference between landslides and surrounding ground objects. The feature threshold is automatically determined by the algorithm without human intervention, which is more intelligent than the traditional method. The deep learning algorithm inputting original image layers does not need to set features and thresholds for landslide extraction. The realization of the deep learning algorithm relies on a deep convolutional neural network and a large number of training samples. These samples can extract deep semantic information and greatly improve the identification accuracy of landslides.

Deep learning techniques for landslide detection primarily involve target detection and semantic segmentation. Target detection mainly deals with the positioning of recognized targets, and the ideal result of recognition is the minimum enclosing rectangular frame of the target. Some commonly used models include R-CNN\cite{8}, Mask R-CNN\cite{9}, Fast R-CNN\cite{10}, Yolo\cite{11}, and other networks. Yuan Zhen Ju et al. used Mask R-CNN for the automatic identification of landslides\cite{12}. The image segmentation process involves not only identifying the image category and target location but also categorizing each pixel in the image at the pixel level. Some commonly used models are FCN\cite{13}, U-Net\cite{14}, SegNet\cite{15}, Deeplab\cite{16} and other networks. For example, Peng Liu et al. proposed an improved U-Net model to identify landslides and improve landslide extraction accuracy\cite{17}.

Aiming at the low contrast between the target and surrounding objects in remote sensing landslide images, the occlusion phenomenon exists, and the requirement of real-time segmentation, in order to further improve the segmentation ability of the model, this paper proposes an improved LinkNet remote sensing image landslide extraction method, which adds a convolution attention mechanism into the residual structure. The convolution attention mechanism module is a lightweight model, combining spatial and channel attention mechanism modules, which can weight the feature information, enhance the feature information of the target, and weaken the interference of useless information, so as to improve the speed and accuracy of the model.

2. Method

2.1. CBAM - LinkNet model

CBAM-LinkNet network is proposed on the basis of LinkNet\cite{18} model and adopts the coding-decoding structure. The structure is shown in Fig.1. On the left side of the model, there is an input layer, including a convolution layer and a pooling layer, then four coding layers, four decoding layers on the right, and finally an output layer, including two full-Conv layers and one convolution layer. The left part of the network is mainly the coding structure, which uses four encoder-block to increase the number of image feature channels by 8 times. The right part is primarily a decoding structure. Four decoder-block are used to reduce the number of image feature channels to the original number. After the code-decoding operation, the number of image feature channels remains unchanged, to avoid the increase in parameters and improve the operation efficiency.
For remote sensing landslide information extraction model training, the set of training data with an image size of 256×256 pixels and three feature channels is first input, and then the first convolution layer is added. In this paper, the size of the convolution kernel is set as 7×7, the number is 64, and the step is 2. After the first convolution layer, the image size becomes 128×128 and the number of feature channels is 64. After the convolution operation, each pixel in the image integrates its original information and the information of each pixel in the 3×3 field. To improve the robustness of the algorithm, the second layer is set as the pooling layer. Set the maximum pooling window to 3×3 and the step is 2. After the pooling layer, the image size becomes 64×64 and the number of channels is 64.

In the coding structure, the CBAM attention module is added to ResNet to form the encoder. In this structure, the number of input image feature channels is m, and the number of output image feature channels is n. After four times of coding, the image’s feature channel number 64,128,256,512 respectively. At this time, the number of image feature channels expands to 8 times the original. In the decoding structure, two convolution layers and one full-Conv layer are used. In this structure, the input image feature channel number is m, the output image feature channel number is n. After four decodings, the image feature channel number is 256,128,64,64 respectively. At this time, the image feature channel number shrinks to the original number.

After the encoding-decoding structure, the image enters the first full-conv layer, which is equivalent to the up-sampling operation. LinkNet uses deconvolution to maintain the spatial location information of image data and reduce the complexity of calculation. After the full-conv layer, the image size becomes 128×128 and the number of feature channels becomes 32. Then, the image enters the second convolution layer, and the image size and the number of feature channels remain unchanged. Finally, the image enters the second deconvolution layer, the image size changes to 256×256, and the number of channels changes to 1. At this time, the prediction data extracted from landslide information is output.

In this paper, convolutional attention module is added to the residual structure of LinkNet model to optimize the weight allocation from two aspects of channel and spatial dimension, strengthen the attention degree of landslide information in shallow features, reduce the interference of background noise information, so as to effectively improve the accuracy of landslide extraction model. Due to the multiple down-sampling operations in the encoder, some spatial information will be lost, and it is difficult to recover the lost information only by using the down-sampling output of the encoding. Therefore, the network adds the input of each encoding layer to the output of the corresponding decoder to recover the lost spatial information.
Precisely because each layer of decoder has input from the encoder, the decoder can use fewer parameters, thus improving the efficiency of the network.

2.2. Convolutional Attention Module

Although convolutional neural network has a strong ability to extract local information, it is relatively weak to obtain the overall feature distribution. The attention mechanism in the field of computer vision obtains the important information of each feature through the form of autonomous learning, so as to ignore the other irrelevant feature information and focus on the distribution of the overall landslide feature, which effectively improves the prediction ability of the model.

In this paper, the CBAM structure is used to further process the landslide information in shallow features. The CBAM highlights landslide feature information in channel dimension and spatial dimension by assigning weights, which not only suppresses background information, but also enhances the classification and prediction effect of the model. The CBAM has two main steps: Firstly, Global Max Pooling (GMP) and Global Average Pooling (GAP) of shallow features are respectively carried out in the channel dimension to create two one-dimensional feature vectors, and the weights of the feature vectors are allocated through the fully connected layer. The landslide feature information on the channel domain is strengthened. Secondly, the features extracted from the channel domain are compressed again by GMP and GAP in the spatial dimension to generate a two-dimensional feature map, and then the feature weights are allocated by convolution to realize the enhancement of landslide feature information in the spatial domain. Through the processing of the above steps, the network model in this paper pays more attention to the landslide information in the shallow characteristics. Its CBAM structure is shown in Fig.2, where the Channel attention module (CAM) mainly focuses on the meaningful content in the input data. Its expression is:

\[
M_c(F) = \sigma \left( \text{MLP} \left( \text{AvgPool}(F) \right) + \text{MLP} \left( \text{MaxPool}(F) \right) \right) \\
= \sigma \left( W_1 \left( W_0(F_{avg}^c) \right) \right) + W_1 \left( W_0(F_{max}^c) \right)
\]

(1)

Where, \( W_0 \in \mathbb{R}^{c/r \times c}, W_1 \in \mathbb{R}^{c/r \times c} \).

Spatial Attention Module (SAM) mainly focuses on which location information is meaningful, which is a supplement to channel Attention. Its expression is:

\[
M_c(F) = \sigma \left( f^{7 \times 7} \left( [\text{AvgPool}(F)] \right) \right) \\
\text{MLP} \left( \text{MaxPool}(F) \right) = \sigma \left( f^{7 \times 7} \left[ F_{avg}^s, F_{max}^s \right] \right)
\]

(2)

2.3. Precision evaluation

In this paper, pixel-based Accuracy evaluation indexes are used to quantitatively evaluate landslide identification Accuracy, including pixel Accuracy (ACC) and Mean Intersection over Union (MIoU), whose calculation formula is as follows.
Accuracy

(1) Accuracy

Used to calculate the percentage of the number of samples correctly divided. The higher the accuracy, the better the classifier. The equation is as follows.

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(4)

TP, FP, FN, and TN are shown in Table 1. TP is truly positive, and the actual landslide pixel is considered a landslide pixel. FP is false positive, and non-landslide pixels are predicted to be landslide pixels. FN is a false negative, and the actual landslide pixel is predicted to be a non-landslide pixel. TN is true negative and non-landslide pixels are predicted to be non-landslide pixels.

Table 1: Confusion matrix between predicted value and true value.

<table>
<thead>
<tr>
<th>Predicted Value</th>
<th>Landslides</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landslides</td>
<td>True Positive (TP)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>Others</td>
<td>False Negative (FN)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

(2) Mean Intersection over Union

Mean Intersection over Union is a standard measure of semantic segmentation [19]. It computes the ratio of the intersection and union of two sets of true and predicted values. The two groups of data presented in this article are explained landslide numbers and predicted landslide numbers. The higher the ratio, the higher the accuracy. Its equation is as follows.

\[
mIoU = \frac{1}{K+1} \sum_{j=0}^{K} \frac{P_{ij}}{\sum_{i=0}^{K} P_{ij} + \sum_{j=0}^{K} P_{ji} - P_{ii}}
\]  

(5)

\[
mIoU = \frac{TP}{TP + FP + FN}
\]  

(6)

Where K +1 is the number of categories (K target classes and 1 background class); i represents the true value, j represents the predicted value; \( P_{ij} \) represents the number of pixels labeled i but predicted to be category j; \( P_{ii} \) represents the number of pixels labeled j but predicted to be category i; \( P_{ii} \) represents the number of correctly classified pixels.

3. Experiment

3.1. Data

The data set adopted in this paper is a remote sensing landslide data set obtained by the Group of Photogrammetry and Computer Vision (GPCV) at Wuhan University Shunping Ji. The data set is composed of satellite optical images and label maps of landslide boundary shapes, including 770 remote sensing landslide images and 770 mask images as label maps. The label maps are artificially binarized labeling of landslide and non-landslide in the training data, in which landslide is labeled 1 and non-landslide is labeled 0. Both datasets were taken from TripleSat satellite imagery in May 2018. The sample data were divided into training data sets, validation data sets, and test data sets. Training data is used to learn target characteristics, validation data is used to select the best-performing model, and test data is used to evaluate model performance. In this paper, 640 images are divided into training sets, part of the training data are shown in Fig. 3, 123 images are divided into verification sets, and the rest are taken as test sets.
3.2. Image preprocessing

Due to the time-consuming and laborious process of data annotation, the number of landslide samples collected and annotated is very limited. To ensure the number and diversity of network training samples, the data set needs to be enhanced. The enhancement method selected in this paper is left-right flipping, which converts the data set into multiple instances in space. The process is shown in Fig. 4.
3.3. Artificial Experimental Environment

The hardware environment of this paper is Intel(R) Core(TM) i5-10210U CPU@ 1.60ghz CPU, MX250 graphics card, memory is 16GB. The software environment is deep learning framework Tensorflow2.3 and Python3.7. We optimize the network using Adam optimizer and relu function. The software environment is Tensorflow2.3 and Python3.7. The batch size of training is 16, the learning rate is 0.0001, and the training cycle is 50.

During the back propagation of the network, the Binary Cross entropy function is used to update the parameters of the network, which is defined as

$$L_{loss} = -\frac{1}{m}\sum_{i=1}^{m}[\log_2 \hat{y} + (1 - y) \log_2 (1 - \hat{y})]$$

(7)

Where: \(y\) is the true value of pixels; \(\hat{y}\) is the predicted pixel value; \(m\) is the number of categories.

3.4. Results

The training data and label data are injected into the improved network, and the prior model of landslide information is obtained by adjusting the parameters. Then landslide recognition is carried out on the remote sensing image in Fig.5 (a), and the recognition results of each model are shown in the figure. It can be seen that, compared with the result of manual interpretation, it is shown in Fig.5. U-net and LinkNet network identification results have some omissions and boundary error classification. The main factor leading to false extraction is the road, and the presence of a slight dirt accumulation on the road, which is similar in color to a landslide, Image1 in Fig.5. For some large landslides, vegetation is covered at the boundary of the landslide, and the mixture of some landslides and vegetation leads to incorrect extraction, as shown in Image2 in Fig.5. However, the landslide information extracted by CBAM-LinkNet network is relatively complete and the extraction result is ideal.

![Image1](image1.png)

![Image2](image2.png)

(a) (b) (c) (d) (e)

Fig.5 (a) Test image; (b) Label image; (c) U-Net segmentation result; (d) LinkNet segmentation result; (e) CBAM-LinkNet segmentation result.

In order to quantitatively evaluate the performance of the model and the extraction results of the test area, the accuracy indexes of each model are calculated. Fig.6 shows the line chart of each evaluation index of CBAM-LinkNet changing with the number of iterations after 50 rounds of iterations. The specific accuracy calculation results of the two methods are shown in Table2. It can be seen that the CBAM-LinkNet proposed in this paper has the highest accuracy in landslide identification accuracy and can accurately identify the location of landslide. In terms of computing time, the average running efficiency of U-Net method is 1026 minutes, and the average running efficiency of the proposed method is 167 minutes, which is 16.2 times higher, and the computational efficiency is greatly improved.
Table 2. Comparison of accuracy and running time of landslide information extraction results

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy/%</th>
<th>MIoU/%</th>
<th>Time/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net</td>
<td>95.376</td>
<td>79.2</td>
<td>1026</td>
</tr>
<tr>
<td>LinkNet</td>
<td>95.071</td>
<td>76.5</td>
<td>62</td>
</tr>
<tr>
<td>CBAM-LinkNet</td>
<td>96.146</td>
<td>81.3</td>
<td>167</td>
</tr>
</tbody>
</table>

4. Conclusions and Prospects

In this paper, a remote sensing landslide recognition algorithm based on LinkNet and convolutional attention mechanism is proposed. The introduced convolutional attention mechanism enhances the utilization of shallow information, obtains rich context information, filters background information, and obtains finer feature maps. The experimental results show that compared with the LinkNet and U-Net models, the proposed algorithm CBAM-LinkNet has improved the accuracy of remote sensing landslide extraction, and can greatly improve the recognition speed, which can be used to quickly respond to the sudden earthquake landslide disaster.

Due to the small amount of published remote sensing landslide data and low resolution, there are still some errors in the process of landslide extraction. Moreover, the effect of remote sensing landslide extraction under more complex scenes is slightly poor, so the next research will focus on improving the accuracy of landslide extraction under complex scenes. At the same time, the training dataset needs to be expanded, including the expansion of remote sensing images of different types and resolutions.

Acknowledgments

This work was supported in part by NSFC under Grant 41930112 and Meteorological Disaster Prediction and Early Warning Research Project of Universities in Sichuan Province under Grant ZHYJ21-ZC01

References


