Computers, Materials & Continua
DOI: 10.32604/cmc.2023.035016
Article

An Early Warning Model of Telecommunication Network Fraud Based on User Portrait

Wen Deng[1], Guangjun Liang[1,2,3,*], Chenfei Yu[1], Kefan Yao[1], Chengrui Wang[1] and Xuan Zhang[1]

[1]Department of Computer Information and Cyber Security, Jiangsu Police Institute, Nanjing, China
[2]Engineering Research Center of Electronic Data Forensics Analysis, Nanjing, China
[3]Department of Public Security of Jiangsu Province, Key Laboratory of Digital Forensics, Nanjing, China
*Corresponding Author: Guangjun Liang. Email: lianggjun@126.com
Received: 03 August 2022; Accepted: 08 December 2022

Abstract: With the frequent occurrence of telecommunications and network fraud crimes in recent years, new frauds have emerged one after another which has caused huge losses to the people. However, due to the lack of an effective preventive mechanism, the police are often in a passive position. Using technologies such as web crawlers, feature engineering, deep learning, and artificial intelligence, this paper proposes a user portrait fraud warning scheme based on Weibo public data. First, we perform preliminary screening and cleaning based on the keyword "defrauded" to obtain valid fraudulent user Identity Documents (IDs). The basic information and account information of these users is user-labeled to achieve the purpose of distinguishing the types of fraud. Secondly, through feature engineering technologies such as avatar recognition, Artificial Intelligence (AI) sentiment analysis, data screening, and follower blogger type analysis, these pictures and texts will be abstracted into user preferences and personality characteristics which integrate multi-dimensional information to build user portraits. Third, deep neural network training is performed on the cube. 80% percent of the data is predicted based on the N-way K-shot problem and used to train the model, and the remaining 20% is used for model accuracy evaluation. Experiments have shown that Few-short learning has higher accuracy compared with Long Short Term Memory (LSTM), Recurrent Neural Networks (RNN) and Convolutional Neural Network (CNN). On this basis, this paper develops a WeChat small program for early warning of telecommunications network fraud based on user portraits. When the user enters some personal information on the front end, the back-end database can perform correlation analysis by itself, so as to match the most likely fraud types and give relevant early warning information. The fraud warning model is highly scaleable. The data of other Applications (APPs) can be extended to further improve the efficiency of anti-fraud which has extremely high public welfare value.

Keywords: Crawler; user portrait; feature engineering; deep learning; small program development

1 Introduction

As of June 2022, the number of netizens in my country has reached 940 million, an increase of 36.25 million from June 2022. The Internet penetration rate reached 67%, an increase of 2.5 percent points from June 2022 and an increase of about 5 percent points higher than the global average. Among them, my country's mobile phone netizens reached 932 million, an increase of 35.46 million from June 2022. The proportion of netizens using mobile phones reached 99.2%, basically the same as in June 2022. This is due to the large-scale construction of 4G and 5G networks in China, and the fact that mobile phones are easier to carry and operate than computers. The rapid development of the Internet is highly integrated into human society and profoundly changes our work and lifestyle. However, everything has two sides. While the Internet has brought us unparalleled convenience, it has also created some negative problems. One of them is the crime of telecommunication network fraud. According to the "Mobile Phone Security Status Report for the First Half of 2020" jointly released by 360 and the China Academy of Information and Communications Technology, the per capita loss of telecommunications network fraud victims in the first half of 2020 alone has reached 10,037 yuan. Due to the large population base in my country, it is impossible to analyze them one by one. The current anti-fraud early warning mechanism lacks pertinence, and the "mass distribution" effect of anti-fraud publicity is not significant.

At present, the situation of telecommunication and network fraud crimes is severe, and it has become the type of crime with most cases, the fastest rise, the widest coverage, and the strongest response from the people. The five types of fraud in the Public Prosecution Law account for nearly 80% percent of the cases, making them the five most prominent high-incidence cases. Among them, fraud rebate fraud has the highest rate which accounts for about one-third of the total number of cases. Fraudulent investment and wealth management fraud involves the largest amount which accounts for about one-third of all funds involved. According to the survey, 80% percent of netizens have experienced telecom fraud. Compared with the post-70 s generation who have been deceived less, the post-90 s generation is the group with the most victims of telecom and network fraud. An intriguing question is that more than half of the victims were suspicious of a scammer but ended up being scammed anyway. The characteristics of "intelligence, specialization, grouping, and transnationalization" presented by telecommunication network fraud have caused a huge impact on my country's criminal legislation and judicial concept which makes the law enforcement of related cases more difficult. To this end, the Supreme People's Court, together with the Supreme People's Procuratorate and the Ministry of Public Security, successively formulated the "Opinions on Several Issues Concerning the Application of Law in Handling Criminal Cases such as Telecom Fraud" and the "Opinions on Handling Telecom and Internet Fraud and Other Criminal Cases" in 2016 and 2021, respectively. Opinions on Several Issues Concerning the Application of Law (II)", normative legal documents such as the minutes of the "breaking card" action meeting were issued twice last year and this year. In view of the outstanding problems in judicial practice, it is necessary to continuously improve and clarify the applicable standards of law. In order to meet the needs of the current struggle, it is necessary to further regulate and guide law enforcement in handling cases.

Zhuang [1] of the People's Public Security University of China advocated the prevention of crimes by controlling the elements of crime, and proposed a big data early warning model for telecommunication network fraud crimes. Based on the flow of criminal information, big data technology is used to discover false information from the source, identify the propagation process, and intercept it from terminal devices. Based on the flow of criminal funds, a number of monitoring models for suspected victims' financial accounts that adapt to the dynamic changes of fraud methods can be established. From the perspective of crime prevention, Chen et al. [2] of the Xinjiang Police

Department pointed out that the crime of telecommunication and network fraud is a crime involving the vicious fusion of "personnel flow, information flow, and capital flow", and it is the victim's cooperation with the suspect without correct cognition. Investigating and cracking down is no longer the optimal strategy for the governance of such crimes. Prevention is an effective way to deal with such crimes. Pre-prevention based on accurate early warning of crimes can effectively block crimes and prevent harm. Big data analysis, research and judgment, and early warning are very good early warning methods for telecommunication network fraud. References [3,4] conduct in-depth research on this. Wu of Henan Police Academy [3] studied the early warning and countermeasures of cross-border telecommunication network fraud under the background of big data. Through data mining, data profiling, and data modeling of the characteristics of fraud, illegal and criminal behaviors, the automatic identification and discovery of various cross-border electronic fraud violations and crimes, and the intelligent linkage disposal during the event are realized. Fu et al. [4] of the Guizhou Police Academy proposed to conduct research on the early warning and prevention mechanism of telecommunication network fraud under the background of big data. Taking the "pig-killing" fraud crime as an example, according to the characteristics of this type of crime, a big data early warning and prevention and control mechanism for telecommunication network fraud crime is established. By fully integrating social information resources and establishing a data sharing mechanism, an integrated work pattern of "fight, prevention, management and control" has been formed.

This paper proposes a fraud warning scheme based on user portraits. First of all, based on the keyword "defrauded", preliminary screening and cleaning are performed to obtain valid user Identity Documents (IDs) that have been defrauded, and the basic information and account information of these users are used for user tags to achieve the purpose of distinguishing fraud types. Secondly, through feature engineering technologies such as avatar recognition, Artificial Intelligence (AI) sentiment analysis, data screening, and blogger type analysis, the image and text are abstracted into user preferences and personality characteristics, and multi-dimensional information is integrated to build user portraits. On this basis, a WeChat applet for early warning of telecommunications network fraud based on user portraits is developed. When the user enters some personal information on the front end, the back-end database can perform correlation analysis by itself, match the most likely fraud types and give relevant cases, so as to achieve accurate early warning. The previous research results of this paper were published at the ICAIS2021 conference.

The innovations of this paper are summarized as follows:

- By designing a free and interesting prediction applet, the model can provide personalized anti-fraud warnings. Users scan the WeChat Quick Response (QR) code to open the anti-fraud security questionnaire and answer a few simple questions. Then, the anti-fraud warning algorithm accurately matches the cases with the highest similarity to itself. At the same time, the calculation results will also provide the probability of being deceived, real cases and warning messages to enhance the warning effect.
- Through the popularization and application of the public security organs, the model can form a fraud early warning system centered on prevention. Users get a personalized anti-fraud security detection report, which contains the most matching fraud types and the similarity between the two. At the same time, typical cases and persuasive messages of this type of fraud are given, so as to improve users' vigilance against telecom fraud. Relevant reports show that the effect of advance warning is far better than the dissuasion effect of early warning.
- The small program of the proposed early warning model can be used as a detection plug-in in the software background. Through automatic detection directly based on the data in the database, the operation efficiency is greatly improved. At the same time, the Few-shot Learning algorithm

used in the early warning model has higher accuracy than Long Short Term Memory (LSTM), Recurrent Neural Networks (RNN) and Convolutional Neural Network (CNN) algorithms, which can save computing time while ensuring accuracy.

The next chapters of this paper are arranged as follows. The second chapter summarizes the development and application of user portrait technology, paving the way for subsequent chapters. The third chapter, data preprocessing and feature engineering, mainly describes data acquisition and cleaning, as well as feature engineering for user portrait work. The fourth chapter proposes a telecommunication network fraud warning model based on user portraits. The fifth chapter is experimental simulation verification. The last is the summary outlook.

2 Related Work

User portraits are to label and analyze user data to describe the characteristics of real users, so that business personnel can quickly and accurately understand user information, so as to take targeted measures to achieve expected goals. User portrait analysis is one of the key points of human behavior analysis and is widely used in business. Reference [5] uses a user-detailed record dataset from mobile operators and proposes concepts of mobile Internet life roles and potential indexing models. By analyzing the preferences of users of different brands of mobile phones for different Applications (APPs), a user portrait analysis framework based on latent semantic index topic model and association rule mining is constructed to analyze users' mobile Internet life roles. Reference [6] starts from the aspects of image design, data analysis and preprocessing, feature engineering and algorithm model construction, and proposes a grid business travel business customer model based on user portrait analysis to realize the analysis of business travel business customers. Reference [7] proposes a user model with five dimensions based on a large amount of student behavior data accumulated by university business platforms, including students' basic information, learning ability, consumption level, daily habits and interest preferences. Based on the methods of data collection, processing and mining, it realizes the extraction of students' characteristic attributes and the construction of student portraits, and provides data basis for the decision support of college teaching management departments. Reference [8] analyzes user information and social behavior data on the Internet, and proposes a mental modeling method based on computational language features. This method can analyze the Big Five personality characteristics of users on Sina Weibo and their correlation with users' social behaviors. Reference [9] uses some minors' impulse reward data obtained from a certain online live broadcast platform. Combined with the data, the statistical database and association rules of minor users' attribute tags are constructed, and the impulsivity reward behavior model of minors is constructed. The predicted results of the model have a high degree of matching with the actual results, and can predict the trend of impulsive reward behavior of minors.

After completing the modeling and analysis of user portraits and behaviors, many commercial software began to focus on researching recommendation systems. The main task of the recommendation system based on user portrait is to combine users and products, by mapping user portraits and product portraits, combining association rule mining and reordering algorithms, to obtain a personalized product recommendation list. Reference [10] conducts research on the problem of information overload in recommender systems, and points out that helping users find valuable information for themselves and display it in front of users is a win-win for both information consumers and information producers. Reference [11] is based on the data of a Business to Consumer (B2C) e-commerce platform, and the text preprocessing of user data is completed by word segmentation technology and feature selection. Using statistical analysis, clustering and other methods to label

users, a user portrait system based on text mining is designed. Reference [12] quantifies, analyzes and processes data information through related algorithms and models, and proposes an improved music recommendation algorithm, which enriches the application of user portraits in the field of music recommendation. Reference [13] constructed an electric vehicle user portrait index system from three dimensions: electric response potential, time response potential and spatial response potential. By selecting remaining capacity and usage time price as price-sensitive user characteristic parameters, distance, time and electricity fee can be selected as service fee-sensitive user characteristic parameters. An electric vehicle user portrait recommendation system is constructed considering the market response potential. Reference [14] adds four key features to the traditional Key-Value structure, and proposes a multi-domain dialogue-based user portrait model and a method to construct and update user information through historical dialogue mining. Reference [15] extracts user tags from movie information according to movie viewing history, uses traditional machine learning method Naive Bayes, Back Propagation (BP) neural network and deep learning method as classifiers, and proposes a movie recommendation system based on user portrait technology.

References [16–18] are based on the application of user portrait technology in the field of power systems. Reference [16] takes the basic information of the power supplier and the data in the power supply process as the basic data source, and proposes a user portrait model that considers the correlation between different attributes of users and the deviation of analysis perspective. Reference [17] uses the user's operation behavior and the user information of the power equipment account to propose a power system equipment account recommendation system based on user portrait. Reference [18] realizes the risk assessment and defense of electricity bill recovery by analyzing the electricity consumption, payment, and arrears data of the electricity information collection system and marketing system.

3 Data Preprocessing and Feature Engineering

Data preprocessing refers to the necessary processing such as review, screening, sorting, etc., before the collected data is classified or grouped. On the one hand, data preprocessing is to improve the quality of the data, and on the other hand, it is also to adapt to the software or method used for data analysis. Generally speaking, data preprocessing steps are: data cleaning, data integration, data transformation, data reduction, and each large step has some small subdivision points. Of course, these four major steps do not necessarily have to be executed when doing data preprocessing. This paper mainly needs to clean the acquired data.

3.1 Data Acquisition and Cleaning

First, we use web crawler technology to crawl the related Weibo user id through the keyword "deceived". There are roughly three types of user data obtained: ordinary individual users who have experienced Internet fraud, Weibo-authenticated users of anti-Internet fraud propaganda (such as public security, political and legal, anti-fraud, news) and others.

Data cleaning, as the name suggests, "black" becomes "white", "dirty" data becomes "clean", and dirty data is dirty in form and content. Formally dirty, such as: missing values, with special symbols, etc. Dirty content, such as: outliers, etc. This paper uses data cleaning technology to delete the last two types of useless user data, leaving only the users who have been defrauded by the Internet. Then, the data is preprocessed, and special symbols, videos, web page links, etc. are not helpful for subsequent sentiment analysis. Finally, further use user_id will to obtain the user's gender, region, watch list, avatar, and the content of Weibo published by the user. And classify the types of deception of effective users. Fig. 1 shows the flow chart of Weibo data acquisition.

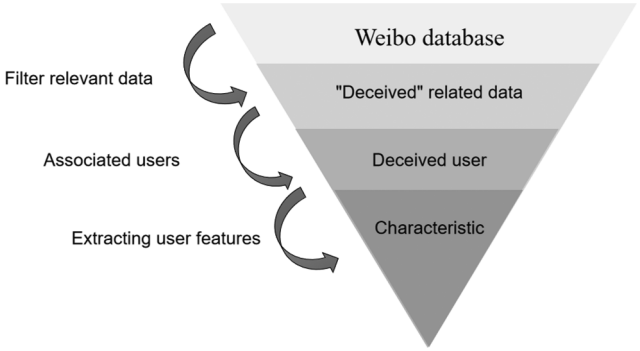

Figure 1: Microblog data crawling and cleaning process

3.2 Feature Engineering

Feature engineering is an integral part of machine learning and occupies a very important position in the field of machine learning. Feature engineering refers to the use of a series of engineering methods to filter out better data features from the original data to improve the training effect of the model. There is a widely circulated saying in the industry that data and features determine the upper limit of machine learning, and models and algorithms are only approaching this upper limit. If better features are used, the model and algorithm can play a greater advantage. Feature engineering usually includes data preprocessing, feature selection, and dimensionality reduction.

Considering the existing research progress of user data sentiment analysis [19] and image classification algorithms, the crawled user data of telecommunication network fraud victims are analyzed. As shown in Fig. 2, the feature engineering of user portraits is divided into two parts, including four types of data. One is Weibo text data, including user registration information, Weibo content, and the following bloggers. This data is mainly analyzed by some algorithms in the field of text sentiment analysis. The other type is Weibo avatar data, which can be analyzed using image-based classification and recognition algorithms. Finally, through the sentiment analysis of Weibo text data and Weibo avatar recognition and classification. Then, the users are labeled, and the user portraits of the deceived groups are shaped, which are used for classification model training.

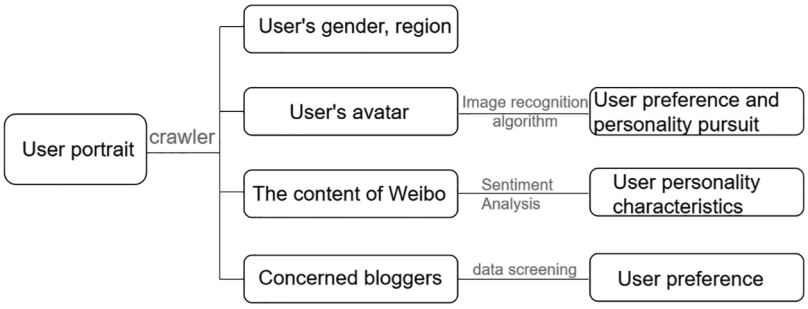

Figure 2: User portrait construction flow chart

4 Telecom Network Fraud Early Warning Model

4.1 User Portrait Based on Weibo Data with Text Feature

The analysis of Weibo text data focuses on users' Weibo content and following bloggers.

Firstly, the sentiment analysis is carried out on the Weibo published by users, and the text content with subjective sentiment is analyzed, summarized, reasoned and judged. Similarly, first of all, on the basis of the establishment of the dictionary, the topics, sentiment words, evaluation words, evaluation objects, opinion classification, emojis, etc. Valuable emotional information is extracted. Second, a classification method based on machine learning [20] is used to classify the extracted Weibo information, including the classification of subjective and objective information and the sentiment classification of subjective information, as well as the summary of core ideas. Finally, the machine-processed sentiment analysis results are presented to the user, categorized by positive, negative, and neutral.

Secondly, for the analysis of the type of bloggers concerned, in Weibo, users are interested in a field or a category of things, and will follow the corresponding bloggers. According to the crawled user attention list, counting the proportion of various bloggers and understanding user preferences can be used as an important feature of the deceived user group. We adopt a tag-based data screening method to accurately and effectively classify the bloggers followed by the deceived by obtaining the Weibo tags of the bloggers followed by the deceived. The main tags are: car, sports, finance, games, photography, shopping, technology, entertainment, animation, beauty, fitness, travel, horoscope, health, real estate, parenting, religion, emotion, lottery.

4.2 User Portrait Based on Weibo Data with Avatar Feature

As shown in Fig. 3, based on the statistical analysis of Weibo user avatars, this paper divides user avatars into five categories: cute pets, stars, landscapes, animations and others.

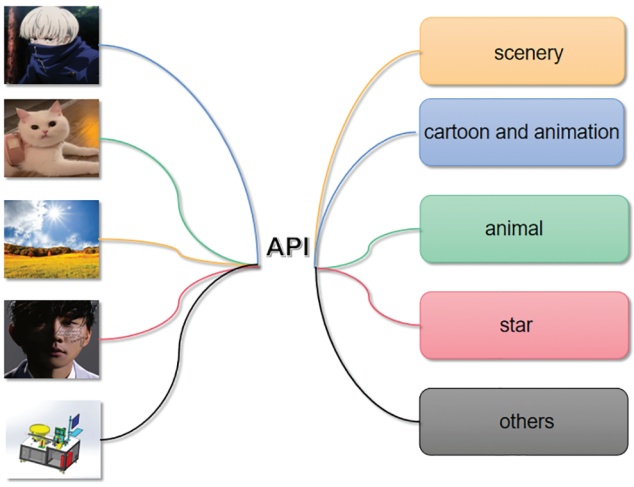

Figure 3: Weibo user avatar classification

The avatar feature analysis is the first impression a user gives to others, which represents the user's preferences and personality pursuits, and is also a very important feature. We use the image classification Application Program Interface (API) publicly available on the Internet to identify images

through a vision-based image recognition algorithm [21], classify the avatars of deceived persons, and explore the relationship between avatar categories and the possibility of being deceived.

The principle of avatar recognition is shown in Fig. 4. The first step is to divide the picture into N∗N pixels. Each pixel represents a neuron, and all the pixels are arranged in a row. As the first layer of the neural network, that is, the input layer, only the data of the input layer is known which is used to generate the first layer of the hidden layer. The second step, the first layer of the hidden layer, is used to identify edges or edges, that is, when the pixels of the first field are input, the edges or edges of the image are generated. That is, only the most insignificant parts of the image can be recognized. In the third step, the first layer of the hidden layer is used as the input of the second layer of the hidden layer to generate corners and contours, and the second layer of the hidden layer produces more detailed image content. Finally, assuming that there are three hidden layers in total, the second layer of the hidden layer is used as the input of the third layer of the hidden layer, which generates more detailed pictures, such as hand and head features, and outputs the classification results, so as to achieve the target picture. Identify and classify.

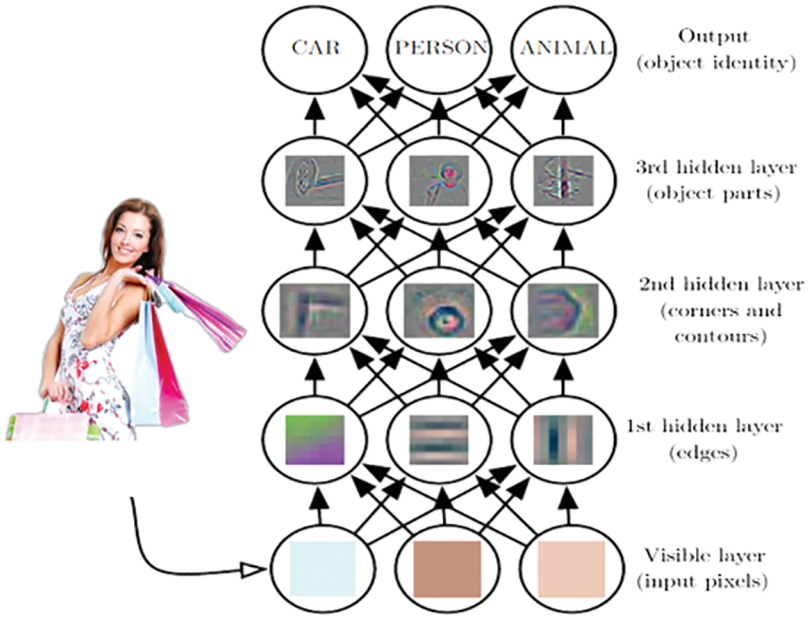

Figure 4: The principle of avatar recognition based on image recognition

This paper uses Baidu artificial intelligence API to classify and label Weibo avatars. First log in to the https://cloud.baidu.com/product/imagerecognition page to register. Baidu AI API is mainly divided into several subcategories of Baidu voice, visual technology, natural language, knowledge graph, and augmented reality. Avatar type recognition belongs to the category of visual technology, the request format is POST, and the content-type of the method call should be application/x-www-form-urlencoded. Then you need to format the request body through urlencode, and the return format is JavaScript Object Notation (JSON) format. Baidu gives two calling methods, we call through the Hyper Text Transfer Protocol over Secure Socket Layer (HTTPS) POST method. We can send the picture to the Universal Resource Locator (URL) https://aip.baidubce.com/rest/2.0/image-classify/v2/dish?access_token=TOKEN. It should be noted here that the token must be obtained in advance. Two keys are required to obtain a token, namely API Key (AK) and Secret Key (SK). After obtaining

the two keys, you can use the following URL to obtain the token, namely https://aip.baidubce.com/oauth/2.0/token?grant_type=client_credentials&client_id=[AK]&client_secret=[SK]. If you have registered a Baidu AI account before, you can get AK and SK for free.

After feature engineering, the user portrait shown in Fig. 5 can be constructed.

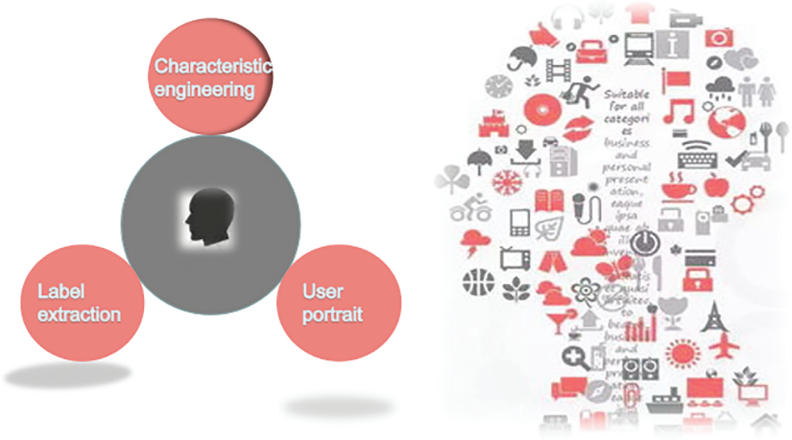

Figure 5: Final form of user portrait

4.3 Early Warning Model

In the previous article, the user data of telecommunication network fraud victims that we crawled is limited, and it cannot be called big data in the true sense. The evaluation of telecom fraud risk model based on "user portrait" based on Weibo data is a small sample learning algorithm [22]. Due to the small number of samples, the optimal hypothesis obtained by minimizing the empirical error is far from the expected hypothesis. In order to solve the problem of a few samples for small sample classes, some information is usually transferred from similar classes for data augmentation. Or you can use some unlabeled or weakly labeled data to help you learn. As shown in Fig. 6, this paper trains and establishes an anti-fraud early warning model through a deep neural network [23].

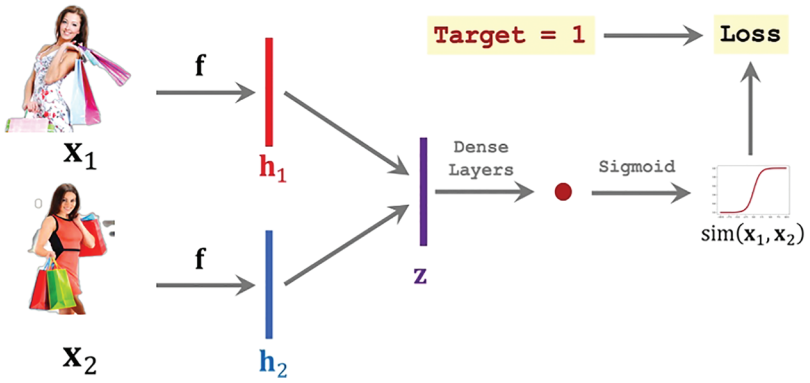

Figure 6: Training anti-fraud early warning model based on user characteristics

In the anti-fraud early warning model shown in Fig. 6, the Sigmoid function is used to predict and evaluate the similarity of two samples. The specific process is as follows. For the two input samples X1

and X2, after subtraction, convolution, pooling and other operations, the value Sigmoid function is calculated. Finally, through the calculation of the sample distance, the similarity of the two samples is judged. During the model training process, the parameters need to be continuously adjusted to ensure the least loss between the predicted value and the true value.

Different features have different degrees of influence on the model. We need to automatically select some features that are important to the problem and remove features that are not very relevant to the problem. This process is called feature selection. The selection of features is very important in feature engineering, and it can often directly determine the quality of the final model training effect. When constructing feature engineering, this model filters almost all valid tags of target Weibo users. In addition to basic age, region, gender, education, etc., feature engineering is also used to analyze user avatars, Weibo sentiments, and watch lists to analyze user personality characteristics. The obtained user characteristics are more objective and comprehensive. Therefore, the construction of user portraits is more observable.

5 Simulation Experiment and Expansion

5.1 Early Warning Model Prediction and Evaluation

This paper constructs an early warning model of telecommunication fraud crime based on Weibo data through the user characteristics induced by feature engineering. The proposed electronic fraud early warning model divides victims into six types, including transaction fraud, free delivery fraud, dating fraud, financial credit fraud, phishing fraud, and part-time fraud. In the data preprocessing stage, the data is labeled with the above-mentioned deceived types.

The proposed anti-fraud model adopts the N-way K-shot prediction model, that is, K samples are randomly selected from N types of samples for prediction each time. Calculate the similarity between the two samples with the sample to be predicted through the previously trained model, and calculate the similarity between the sample to be predicted and each category by calculating the mean value. Fig. 7 shows the user similarity calculation process based on sample distance. Through calculation, if the sample distance is smaller, the similarity between the samples is smaller, and the probability that they belong to the same category of data is larger. Fig. 8 shows the prediction result of a certain piece of test data.

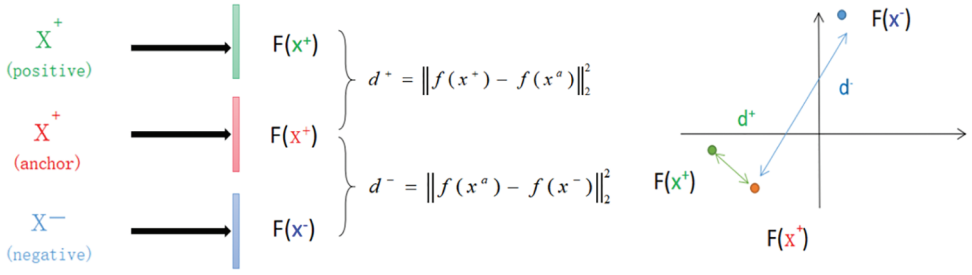

Figure 7: Illustration of user similarity calculation based on sample distance

Next, we compare the accuracy and computation time of the Few-shot Learning algorithm in this model with the deep learning algorithms LSTM [24], RNN, and CNN algorithms. As shown in Fig. 9, the abscissa is the proportion of the training set, and the ordinate is the accuracy of each algorithm (the proportion of correct results in the prediction results). It can be seen from the figure that the accuracy

of the results obtained by the Few-shot Learning algorithm is higher than that of the LSTM, RNN, and CNN algorithms when the training set accounts for 50% to 90%.

```
predict('2950849610')
executed in 11ms,finished 09:19:14 2021-03-16

User 29050849610 and user 2385866161 have the highest similarity
The similarity is 82.3%
It is predicted that this user is most likely to be cheated by part-time job
```

Figure 8: Test results of user deception types

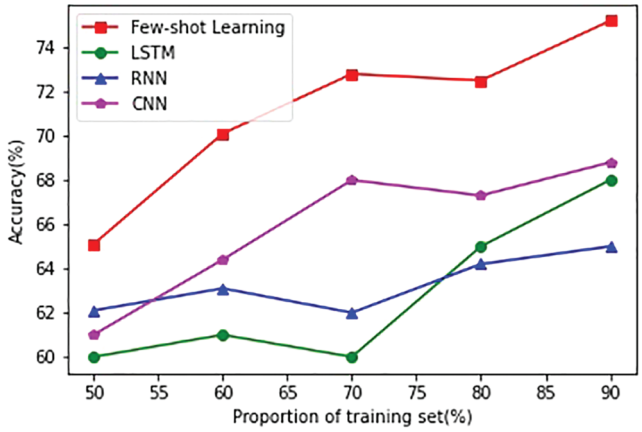

Figure 9: Accuracy comparison between small sample learning and LSTM algorithms

Fig. 10 is a comparison of the computation time between the Few-shot Learning algorithm and the LSTM, RNN, and CNN algorithms. The abscissa is the accuracy of the algorithm result, and the ordinate is the calculation time. It can be observed from the figure that under the same accuracy, the time required for the Few-shot Learning algorithm to obtain the result is shorter than that of the LSTM, RNN, and CNN algorithms. That is, the Few-shot Learning algorithm can save computing time and enhance operating efficiency while ensuring accuracy.

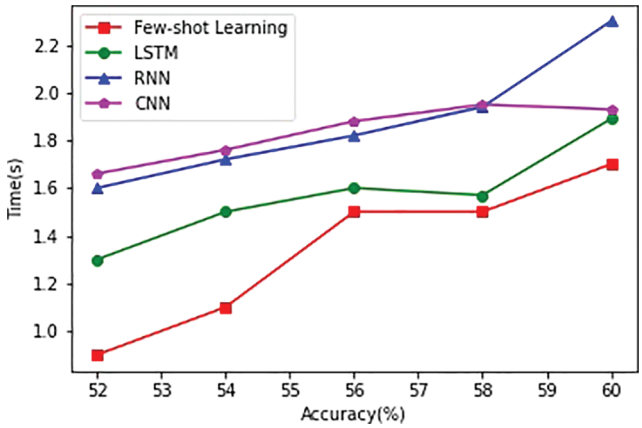

Figure 10: Computing time comparison between small sample learning and LSTM algorithms

5.2 WeChat Mini Program Design and Experiment

This paper constructs a telecommunication fraud crime early warning model. By applying this model to the WeChat applet for better promotion, the social effect of comprehensive anti-fraud can be achieved. As shown in Fig. 11, the design idea of the WeChat applet is simple and easy to use. Test users open the anti-fraud questionnaire by searching for keywords or scanning a QR code. By answering some simple questions, test users can get a self-portrait based on the anti-fraud warning model. Self-portraits can characterize the form in which test users are most likely to be defrauded. Next, the WeChat applet will push some typical cases of electronic fraud with the highest matching degree to test users, reminding them of possible telecom fraud. The anti-fraud early warning model can prevent problems before they happen. Relevant reports show that the effect of advance warning is far better than the dissuasion effect of early warning. The WeChat applet has a flexible design and a friendly interface, which can achieve a relatively good anti-fraud effect. In addition, the embedded core fraud prediction algorithm is highly scalable to adapt to the evolving telecom fraud landscape. After expanding the data set and new forms of telecom fraud, the anti-fraud performance of WeChat mini-programs can be continuously improved which is very suitable for publicity and promotion by public officials.

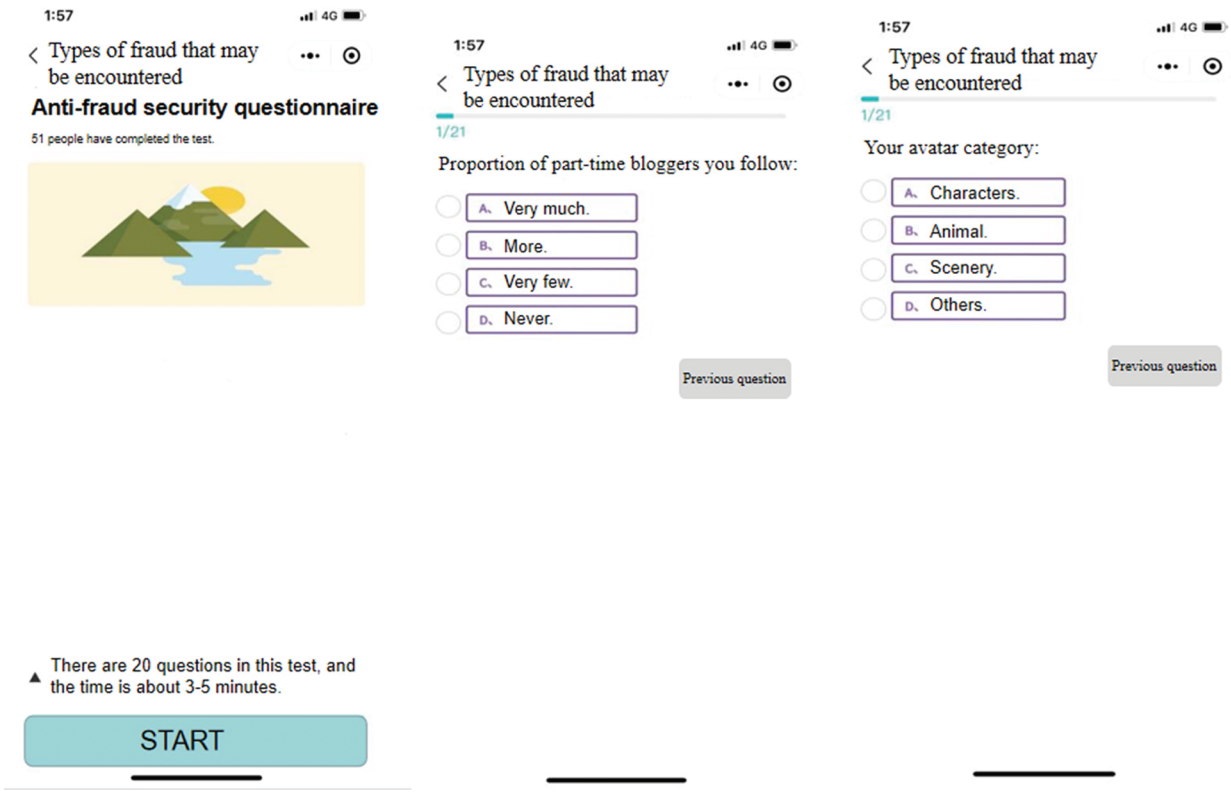

Figure 11: Anti-fraud applet homepage

After the test user enters the WeChat applet, they need to answer a few simple questions, such as "what percentage of you follow up?", "your avatar type" and so on. As shown in Fig. 11, test users choose from a limited number of options according to their own situation. The anti-fraud early warning algorithm matches the back-end database data to obtain a simple detection report as shown in

Fig. 12. The generation of test reports is based on the characteristics of deceived users [25]. The report includes top matching fraud types, similarity, and typical fraud cases. Since WeChat Mini Programs are lightweight codes, they are also popular with young people and can be used as one of the mainstream means of anti-fraud on campus.

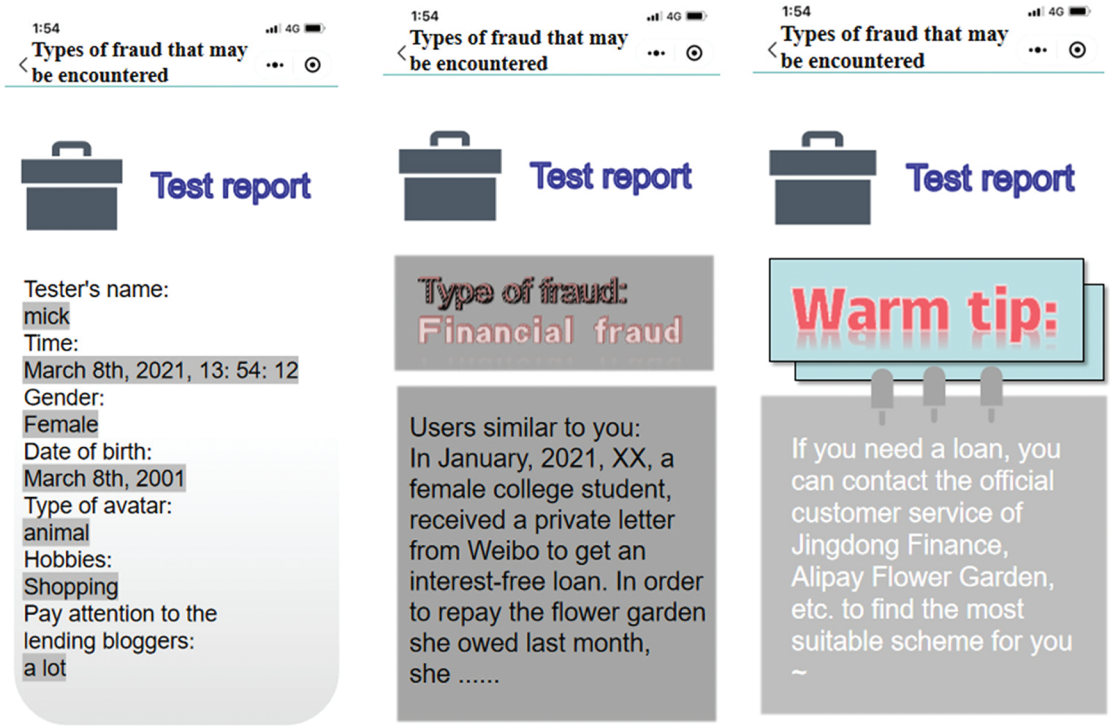

Figure 12: Anti-fraud applet detection report

5.3 Expansion and Application of Early Warning Model

The fraud prevention scheme for Weibo users based on small sample learning proposed in this paper can also be applied to other platforms, such as small video platforms such as Douyin and Bilibili. Now take the Douyin platform as an example to discuss the applicability. Douyin is similar to Weibo. You can interact with any user's private message by just clicking on it. Therefore, the fraud methods and early warning methods based on this platform are also the same as Weibo. Compared with Weibo, the biggest difference in profiling the deceived people on the Douyin platform is the acquisition of user personal data. The information of the Weibo platform is mainly transmitted in text, and it has a web page, which is conducive to the acquisition of data by crawler software such as "Octopus", while the Douyin platform mainly uses short video as the transmission medium, and has no web page. It is characterized by the strategy of "downloading-analyzing-targeting the deceived and crawling account homepage information" [26]. The detailed plan is as follows:

(1) "Download"

Because the videos on the Douyin platform are all manually refreshed, use the Android platform emulator on the Personal Computer (PC), and download Douyin software on the simulator platform to simulate and refresh, obtain batch video sharing links, and use the open source software "Douyin

Video Download Assistant". (URL: https://github.com/kun775/douyinhelper) to download videos in batches and save them to the data disk.

(2) "Split"

Use open source software "Video Framing Magic"

(URL: https://blog.csdn.net/qinlele1994/article/details/98945956?ops_request_misc=%257B%2522request%255Fid%2522%253A%252216159733991678026543947 1%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=161597339916780265439471&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduend~default-1-98945956.first_rank_v2_pc_rank_v29_10&utm_term=%E8%6%A7%886%E9%A26%A7%88%%E5%B8%A7) sub-frame the acquired video to obtain video slice images for better image sentiment analysis.

(3) "Analysis"

Actively use cutting-edge technology to connect server mainland data to Baidu AI interface

(URL: https://ai.baidu.com/tech/imagecensoring), perform image content analysis, look for fraudulent information, and backtrack the fraudulent account.

(4) "Directed Crawl"

Using the account of the deceived obtained in the "analysis" stage, you can crawl the data on the homepage of the deceived and obtain the characteristic values of the deceived. By describing the user profile with the obtained information, the signature database of the deceived person can be built. The methods for characterizing user portraits and building a deceived person's signature database are the same as those described in this paper. Based on this scheme, it can still be done on other apps. At the same time, due to the different user groups applicable to various APPs, more comprehensive and accurate analysis results can often be given. This model can be used as a microblog background detection tool without the need for users to actively call the applet test. By periodically executing the script in the background, the user's recent dynamic and automatic detection of the current deception type and probability of each user. Remind users through private messages, and can directly add anti-fraud measures to cases to achieve the effect of early warning and prevention. At the same time, establish a user information database directly linked to the public security organs, and report to the public security organs on a regular basis. As shown in Fig. 13, major cases should be reported to the public security organs in a timely manner to give advance warnings.

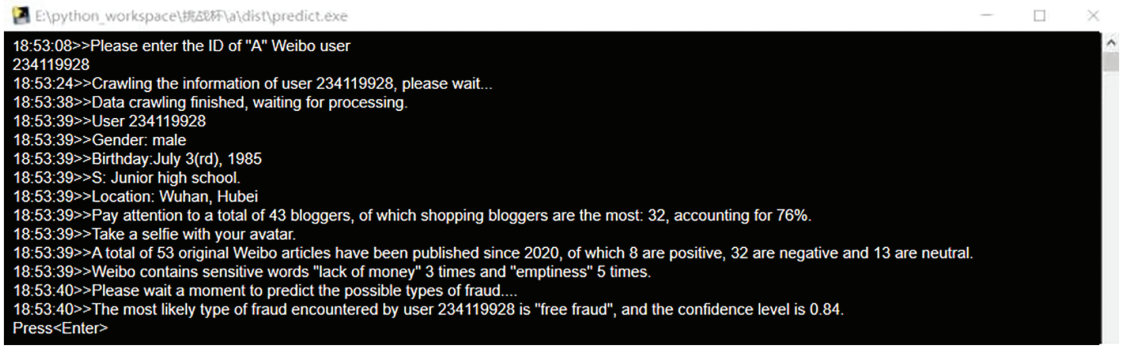

Figure 13: User may be deceived characteristic data feedback

6 Conclusions

The anti-fraud warning based on user portraits established in this paper has broad application prospects. The model constructed in this paper can also effectively improve the anti-fraud awareness of netizens and prevent telecommunication network fraud. Future research could consider adding methods of machine learning or deep learning. By adopting more powerful technical means to reduce the success rate of telecommunication network fraud, so as to give netizens a clear online environment. In addition, WeChat applet anti-fraud warning is also a good exploration.

Acknowledgement: Wen Deng and Guangjun Liang conceived and designed the experiments; Chenfei Yu and Kefan Yao performed the experiments; Chengrui Wang and Xuan Zhang analyzed the data; Guangjun Liang wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding Statement: This research has been supported by the Open Project of the State Key Laboratory of Nanjing University "Research on Intrusion Signal Detection Technology Based on Deep Learning in Complex Electromagnetic Environment", the Open Project of the State Key Laboratory of CAD&CG, Zhejiang University, "Analysis and Visualization of Heterogeneous and Multi-source Cyber Threat Intelligence Data", Jiangsu Province Big Data Management Center Project "Research on Network Security Perception of Jiangsu E-government Extranet".

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. Zhuang, "Big data early warning of telecom network fraud crimes," *Journal of China Criminal Police Academy*, vol. 165, no. 1, pp. 5–13, 2022.

[2] H. Chen, D. Shan and A. Zhao, "An empirical study on early warning of telecom network fraud crimes," *Journal of Xinjiang Police College*, vol. 159, no. 3, pp. 31–40, 2020.

[3] Y. Wu, "Early warning and countermeasures of cross-border telecommunication network fraud crimes under the background of big data——Taking the fraud of impersonating the public security law as an example," *Journal of Hubei Police Academy*, vol. 192, no. 3, pp. 89–96, 2019.

[4] L. Fu, N. Xue and X. Song, "The early warning and prevention mechanism of telecommunication network fraud under the background of big data——Taking the fraud crime of "killing pigs" as an example," *Journal of Guizhou Police College*, vol. 166, no. 1, pp. 93–100, 2021.

[5] H. Jiang, Z. Hu, X. Zhao, L. Yang and Z. Yang, "Exploring the users' preference pattern of application services between different mobile phone brands," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp. 1163–1173, 2018.

[6] Z. Yu, L. Liu, C. Chen, W. Zhang, X. Ju *et al.,* "Research on situational perception of power grid business based on user portrait," in *2019 IEEE Int. Conf. on Smart Internet of Things (SmartIoT)*, Tianjing, China, pp. 350–355, 2019.

[7] X. Li and S. He, "Research and analysis of student portrait based on campus big data," in *2021 IEEE 6th Int. Conf. on Big Data Analytics (ICBDA)*, Xiamen, China, pp. 23–27, 2021.

[8] H. Gu, J. Wang, Z. Wang, B. Zhuang and F. Su, "Modeling of user portrait through social media," in *2018 IEEE Int. Conf. on Multimedia and Expo (ICME)*, San Diego, USA, pp. 1–6, 2018.

[9] N. Wang, "Computer intelligent prediction model of internet impulsive reward based on user portrait algorithm," in *2021 IEEE 4th Int. Conf. on Automation, Electronics and Electrical Engineering (AUTEEE)*, Shenyang, China, pp. 545–548, 2021.

[10] T. Wu, F. Yang, D. Zhang, A. Zhu and F. Wan, "Research on recommendation system based on user portrait," in *2020 IEEE Int. Conf. on Artificial Intelligence and Information Systems (ICAIIS)*, Dalian, China, pp. 462–465, 2020.
[11] Q. Liu, "Business user portrait modeling and clustering analysis under the background of big data," in *2021 13th Int. Conf. on Measuring Technology and Mechatronics Automation (ICMTMA)*, Surabaya, Indonesia, pp. 666–669, 2021.
[12] J. Li, "Music recommendation algorithm based on user portrait," in *2021 IEEE Int. Conf. on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Fuzhou, China, pp. 10–13, 2021.
[13] H. Li, G. Wu, H. Li, F. Gao, X. Long *et al.,* "Electric vehicle user portrait considering market response potential," in *2021 IEEE 5th Conf. on Energy Internet and Energy System Integration (EI2)*, Taiyuan, China, pp. 3941–3946, 2021.
[14] M. Liu, Z. Tu, X. Xu and Z. Wang, "Dialogue-based continuous update of user portraits," in *2021 IEEE Int. Conf. on Services Computing (SCC)*, New York, USA, pp. 193–202, 2021.
[15] Z. Cui, H. Cao and W. Yan, "Research on user portrait based on tag word embedding," in *2021 IEEE Int. Conf. on Information Communication and Software Engineering (ICICSE)*, Chengdu, China, pp. 159–163, 2021.
[16] X. Huang, Z. Qiu, J. Su, Z. Shi and L. Yan, "Knowledge graph-based user portrait construction for electricity enterprise suppliers," in *2021 Int. Conf. on Communications, Information System and Computer Engineering (CISCE)*, Chongqing, China, pp. 81–84, 2021.
[17] D. Li, H. Zhang and Z. Zhao, "Electrical power system equipment account recommendation technology based on the user portrait," in *2020 Int. Conf. on Intelligent Computing, Automation and Systems (ICICAS)*, Chongqing, China, pp. 78–81, 2020.
[18] T. Wang, J. Hu, C. Li and Z. Zhang, "Research on tariff recovery risks assessment method based on electrical user portrait technology," in *2017 6th Int. Conf. on Computer Science and Network Technology (ICCSNT)*, Dalian, China, pp. 218–221, 2017.
[19] S. Zhou, W. Qu, Z. Shi, X. Shi and Y. Sun, "A review of Chinese microblog sentiment analysis research," *Computer Applications and Software*, vol. 30, no. 3, pp. 161–164, 2013.
[20] Y. Zhang and X. Ran, "A step-based deep learning approach for network intrusion detection," *CMES-Computer Modeling in Engineering & Sciences*, vol. 128, no. 3, pp. 1231–1245, 2021.
[21] Y. Xiao, Q. Jiang, X. Wang, Y. Du and Z. Huang, "Image recognition of convolutional neural network model driven by radius interval," *Journal of Xihua University*, vol. 40, no. 2, pp. 71–81, 2021.
[22] K. Zhao, X. Jin and Y. Wang, "A review of small sample learning research," Ph.D. Dissertation, University of Chinese Academy of Sciences, China, 2021.
[23] F. Zhou, L. Jin and J. Dong, "A review of convolutional neural network research," Ph.D. Dissertation, University of Chinese Academy of Sciences, China, 2021.
[24] J. Liang, Y. Chai, H. Yuan, M. Gao and H. Zan, "Sentiment analysis based on polarity shift and LSTM recurrent network," Ph.D. Dissertation, China Institute of Nuclear Science and Technology Information and Economics, China, 2021.
[25] Y. Koi-Akrofi, J. Koi-Akrofi, A. Odai *et al.,* "Global telecommunications fraud trend analysis," *International Journal of Innovation and Applied Studies*, vol. 25, no. 3, pp. 940–947, 2019.
[26] Y. Wang and K. Yang, "Mobile marketing effect influence mechanism and optimization strategies—data analysis based on douyin platform," Ph.D. Dissertation, Beijing Information Science and Technology University, Beijing Information Science and Technology University, 2021.