Research and Analysis of Potential Correlation of Telecommunications Fraud Data Based on Big Five Personalities

Jun Gao
Nanan District Branch of Chongqing Public Security Bureau
Chuang Ma (machuang@cqupt.edu.cn)
Chongqing University of Posts and Telecommunications
Jinhao Hu
Chongqing University of Posts and Telecommunications

Research Article

Keywords: Big Five Model, Bert, SVM, Telecommunications Fraud

Posted Date: January 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2392379/v1

Abstract

In recent years, telecommunications fraud cases show a high incidence, more and more scholars have studied how to stop telecommunications fraud cases from the root cause. According to psychological knowledge, we can know that personality traits can affect a person's social behavior. Therefore, this paper proposes a personality trait recognition model based on the Big Five personality theory combining Bert and SVM to study how to block the occurrence of telecommunications fraud cases from the root. Compared with the previous model, the average accuracy of the recognition of personality traits was improved by 0.3%, reaching 61.2%. At the same time, combined with the telecommunications fraud data in a certain place, it is analyzed that in telecommunications fraud, there is a correlation between the probability of being defrauded, the means of being defrauded, the location where the fraud occurs and personality characteristics. Our research results may provide reference for the government and relevant departments to manage telecommunications fraud.

1. Introduction

In China, fraudulent crimes are emerging one after another, causing huge losses to people's economy. In order to comprehensively crack down on fraud crimes, Chinese authorities have issued many policies. For example, on June 13, 2019, the Ministry of Public Security deployed the Operation Cloud Sword [1], emphasizing that the Internet cloud platform and cloud services are used as sharp swords to focus on combating cybercrime. On October 10, 2020, the State Council convened an inter-ministerial joint meeting on cracking down on new types of illegal and criminal telecommunications networks and decided to launch a nationwide “Broken Card” [2] operation starting on October 10. These policies play an important role in combating fraudulent crimes.

With the deepening of governance work, the antagonism of telecommunications fraud crime technology has become increasingly stronger. According to statistics, more than 300 new fraud methods have been detected in 2020[3]. At the same time, after a fraud crime occurs, it is difficult to obtain evidence and solve the case. Therefore, the prevention of fraud cases is more important than the detection of cases. The traditional criminal investigation method analyzes from the level of crime means, which is not conducive to understanding the law of crime, and has a certain lag in the prevention of fraud cases. Each fraud victim has different psychological characteristics in the process of being defrauded. By exploring the criminal behavior from the psychological level of fraud, we can deeply understand the psychological state of the victim when he is defrauded. By dividing the psychological state into different types, it can be concluded that each type of victim prefers what kind of lure information, and then corresponding measures can be taken from the psychological level of ordinary people to prevent the occurrence of fraud cases.

Personality characteristics are the combined characteristics of all of a person's behaviors, motivations, thinking styles, and emotions, which have a huge impact on the choices everyone makes and in life. Many scholars have applied the analysis method based on personality characteristics in various aspects

[4–9]. For example, in sentiment analysis, Lin et al. [4] used an analysis method based on personality characteristics to explore the personality emotions of different users in social texts. Ren et al. [5] studied the emotional properties of users in search texts using a personality trait-based analysis method. In addition, the analysis of personality characteristics is also used in the diagnosis of mental health. Park et al. [6] applied personality trait analysis to mental health diagnosis to explore the impact of personality traits on mental health. Barkauskiene et al. [7] used personality trait analysis to explore the relationship between adolescents' mental health problems and personality traits. In addition, with the increasing number of fraud crimes, many researchers also apply the analysis of personality characteristics in the analysis of fraud crimes. Olukayode et al. [8] conducted research from the level of personal morality and the motivation of fraudsters, and explored the influence of personality characteristics on fraudsters' fraudulent behavior. Xing et al. [9] applied personality trait analysis methods to fraud research, and explored the correlation between the elderly being defrauded and personality traits. These studies based on personality characteristics lay a foundation for the application of personality characteristics in fraud research in this paper. And the main contributions of this paper are as follows:

- The feature detection model based on the Big Five personality is applied to the analysis of fraud case information in a certain place, and the potential correlation between personality characteristics and case information is obtained.
- After comparative analysis of multiple Big Five personality trait detection models, an improved Bert-SVM Big Five personality trait detection model was proposed, which improved the detection accuracy on the existing model.

The rest of this paper is organized as follows: the second chapter presents an overview of the existing results, summarizes the shortcomings of the existing models, selects aspects worthy of reference, and determines the research direction of this paper. The third chapter preprocesses the existing datasets and analyzes the theoretical modeling of the used models. The fourth chapter builds simulation experiments based on data and models, and analyzes the experimental results. The conclusion part summarizes the conclusion and value of the work done in this paper, and puts forward some reference ideas for the next work.

2. Literature Review

2.1. A review of the literature on pre-trained language models

Pretrained language model (PRLM) is beneficial to downstream task execution in natural language processing and can provide better initialization parameters for the model. After years of development, PRLM has become more mature. Xu et al. [10] proposed the application of neural network in pretrained Language Model (PRLM) in 2000. Bengio et al. [11] proposed a classical neural network language model in 2003.The work of Xu Wei and Bengio et al. laid the foundation for the development of the pre-trained

language model (PRLM). Based on this, Wilbur et al. [12] proposed a method to represent a document as a Bag of Words, which takes terms and frequency of occurrence in the document as its core idea. However, this method cannot calculate the similarity between words, and there is sparsity in representation, which is easy to cause dimension explosion. In order to solve this problem and calculate the similarity between words, Mikolov et al. [13] proposed Word2Vec word vector model in 2013. The model can train hundreds of millions of data sets, and the generated word vector can measure the similarity between words well. However, word vector model based on Word2Vec cannot recognize the contextual word vector, so it cannot solve the polysemy problem. In order to make word vectors context-relevant, Google Research [14] proposed Bidirectional Encoder Representation from Transformers (Bert) model in 2018. Transformer Encoder structure is adopted in this model, and the training is divided into pre-training and fine-tuning two stages. This model also considers the context structure of the text, which can mine the text information in a deeper level, promoting the development of NLP field. Later, many scholars improved Bert model and applied it to different fields. For example, Bobur et al. [15] combined Bert model and pierced index model in anomaly detection of judicial documents, which produced better results for searching outliers. Zhou et al. [16] proposed a Bert-based transfer learning method, which established a new network telecom crime monitoring and early warning platform and achieved good results. These Bert-based methods provide ideas for the research of this paper.

2.2. Literature review of the Big Five personality model

There are many personality tests based on psychological research [17], the most accepted of which is the Big Five Personality Model, also known as OCEAN [18]. It contains five different personality traits: openness (OPN), conscientiousness (CON), extraversion (EXT), agreeableness (AGR) and neuroticism (NEU). Open people prefer abstract thinking and have a wide range of interests; Conscientiousness people are usually efficient and organized; Extraversion people display traits such as enthusiasm, sociability, decisiveness, activity, risk taking, and optimism; Agreeableness people are optimistic about resilience and believe that people are inherently good; Neurotic people tend to have unrealistic thoughts and are more prone to negative emotions. The study of personality characteristics is of great significance to the analysis of crowd psychological characteristics. Many researchers applied the Big Five personality model to analyze the psychological characteristics of people in different scenes and achieved good results [5, 19–21].

Mairesse et al. [19] used continuous modeling technology to extract personality characteristics from text information in order to realize automatic recognition of personality characteristics in text information. However, the data set they used was small, and the proposed method did not consider the possible over-fitting of features, resulting in low accuracy of feature recognition. In order to avoid too small data set, Sun et al. [20] proposed a group-level personality detection model based on AdaWalk. This model not only traverses the entire text network, but also relies less on annotations. However, the text network constructed by this model is based on simulated reality rather than real social network, which lacks authenticity. Ren et al. [5] proposed a multi-label personality detection model based on neural network in

order to carry out more accurate personality characteristics detection in real and small data sets. Combining semantic and emotional features, this model can accurately identify different personality characteristics even with a small amount of data. Kazameini et al. [21] proposed an automatic text personality detection model based on deep learning. This model combines support vector machine (SVM) and Bert to study personality characteristics in text without using a large amount of computing resources. And in a recent paper by Kazemeini et al. [22], they used BI-LSTM model with maximum pooling layer, which can provide sentence embedding for mental statements with rich semantics with less computational overhead, and at the same time can better distinguish personality traits. These scholars have made great contributions to the recognition of personality characteristics in text information, and our research is based on their research results.

3. Description Of The Model Used

The model we used includes feature extraction and classification based on big Five personalities. After classification, the potential correlation between case information and personality characteristics is explored by combining real case information. Before input the case text information into the model, we normalized the data. The pre-trained language model is then used to convert the data into word vectors. In feature extraction, multiple SVMS are used to construct the model. The approach we used is detailed below.

3.1. Data preprocessing

Data preprocessing is the first and indispensable step to transform text information into word vector. As the existing data sets contain the text information collected by the public security organs, the expression form is generally close to colloquial, lacking rigorous grammatical structure and fixed format, and also contains some privacy information. Therefore, it is very necessary to perform effective text preprocessing before converting to word vector. Common preprocessing methods include data denoising, part-of-speech tagging, stop word removal and privacy protection. Regular expressions are efficient for replacing text that conforms to certain rules. The data set we have contained partial duplicate information and null value information, so we choose regular expressions for data denoising. Stop words include function words, pronouns, or words without a specific meaning that are present in large numbers in the text. Python's built-in Jieba library is used to stop word processing for text messages. For the privacy protection of victims, pre-defined desensitization rules are used to desensitize the privacy information such as ID number and bank account. The data cleaning process is shown in Fig. 1.

After pretreatment, Bert model was used for the next operation. The basic principle of Bert model is to use the context in each layer of the model to carry out bi-directional pre-training in depth [14]. It is mainly based on Encoder in Transformer, and the deep neural network formed by stacked multi-layer Transformer structure is the main structure of Bert model. Bert's input includes Token Embeddings, Segment Embeddings, and Position Embeddings. The function of tag embedding is to insert the classification token [CLS] at the beginning of each sequence so that the output of the last Transformer

layer corresponding to [CLS] is used to aggregate the entire sequence representation information. Paragraph inserts are inserted after each sentence to separate different sentences. Position embedding represents the order of each word and ensures that the order of fields will not be wrong in the training process.

In order to better pre-train short texts, The Bert Model constructs two pre-training tasks, namely, Masked Language Model and Next Sentence Prediction. MLM makes predictions by extracting 15 percent of the time from a sentence, while randomly erasing some of the data. These data were replaced 80% of the time with a special symbol [MASK], 10% of the time with any word, and the remaining 10% of the time with the original word. Thus, any token representation information can be extracted. The function of NSP is to determine whether sentences can be connected by training them. Through the above operations, Bert model can transform text information into context-related word vectors.

3.2. The Introduction of the Big Five Personality Model

In this section, we will specifically describe how to use Bert-SVM to realize the modeling process based on the Big Five personality model. Firstly, since Bert can only process 512 data at most at one time, but each message in our document is of different length, we processed the text information. Before the information is converted into word vector, some irrelevant data are removed according to the stop word database. The long text is then cut into several short texts, and information less than 512 length is randomly added to the end of irrelevant information, so that all texts have the same length. Finally, we input the processed data into the pre-trained language model Bert. Bert first transformed all case information into context-related word vectors, and then spliced these word vectors to form sentence vectors. The overall dimension of the vectors are 768, and then connect these vectors with the personality trait detection dataset with 84 features developed by Francois Mairesse et al. [19]. The document feature vector that forms the entire document and whose dimension is R^{852}.

Then, in the feature detection and classification part, 20 SVM classifiers are used for parallel computation to improve the classification accuracy. The input data of SVM classification stage is document feature vector composed of context-related word vector. After the document feature vector is input into SVM, it is combined with pre-trained Mairesse feature vector to compare the similarity between the target vector and Mairesse vector, and the scores of personality characteristics of each type are output. After that, the model will take the average of all SVM output results, which is represented as the final prediction results of all SVM classifiers. This result can be represented as the personality characteristics in the big Five personality model represented by the document. The process of personality testing model is shown in Fig. 2.

4. Experimental Verification

In this work, we first perform model validity validation of the personality detection model. Then it is applied to the analysis of the case text information to extract the victim's personality characteristics contained in each case. Firstly, we used the Essays datasets [23], which consists of 2468 essays written

by students. Then, after verifying the effectiveness of our proposed model, the original data of the case information is preprocessed and input into our proposed model. The victim in each case is labeled with the corresponding personality characteristics, and the personality characteristics of the victim in each case are obtained. Finally, we combine personality characteristics and existing case information to analyze the preferences of fraudsters in crime, so as to prevent fraudulent behaviors differently.

4.1. The Big Five Personality Model Experiment

In this part, we mainly focus on the accuracy of the model for the detection of personality traits. Therefore, we improve on the existing personality detection model based on the Big Five personality traits. Through experimental verification, we noticed that the accuracy of the Bert-SVM model in the Big Five personality detection has a certain difference every time. In order to more accurately represent the personality characteristics, we use 20 SVM classifiers to perform personality feature detection in parallel, and finally calculate the average of the output results of all SVMs to obtain the final personality feature. Compared with existing models, our model achieves a 0.3% improvement in the average accuracy of personality detection, reaching 61.2%. The accuracy of other models on our dataset is presented in Table 1.

Table 1
Comparison of our model with other models

	Personality Traits					
Model Name	**EXT**	**NEU**	**AGR**	**CON**	**OPN**	**Average**
Major Baseline	51.7	50.2	53.1	50.8	51.5	51.4
Mairesse	55.1	58.1	55.4	55.3	59.6	56.8
CNN-Mairesse [24]	58.1	57.3	56.7	56.7	61.1	58.0
BB-SVM [21]	59.3	59.4	56.5	57.8	62.1	59.0
BERT-large [25]	63.4	58.9	59.2	58.3	58.9	59.7
BERT-base [25]	**64.6**	59.2	60.0	58.8	60.5	60.6
Previous state of the art [22]	63.9	59.5	**60.2**	**59.5**	61.3	60.9
B-20-SVM(Ours)	63.7	**60.3**	60.1	59.4	**62.5**	**61.2**

4.1.1. Word Embedding Model

We compare and analyze the model with the Word2Vec model, and the results show that the combination of Word2Vec and SVM has lower accuracy in personality feature extraction than the combination of Bert and SVM. Table 2 shows the comparison results. Meanwhile, the study by Devlin et al. [14] showed that concatenating the last four layers of the Bert model can better represent words. Therefore, we use the Bert

model for word embedding in the word embedding stage of the personality detection model to obtain better detection results.

Table 2
Comparison of the average detection accuracy of personality

Model	Word Embedding	Sentence Feature Extraction	Document Feature Extraction	Classifier	Average Accuracy
Word2Vec-SVM	W2V	-	Mean	SVM	57.4
Bert-20-SVM	Bert	-	Mean	SVM	61.2

4.1.2. Personality Detection

In personality feature detection, we also tried a deep learning-based text personality detection method [24]. This method represents the document vector as the addition of multiple sentence vectors, and then inputs the document vector into a convolutional neural network (CNN) to obtain the classification result. However, this method did not improve the accuracy of personality detection on our dataset. Therefore, we finally adopted a combination of Bert and SVM for personality feature detection.

At the same time, according to Professor Peng Danling's book "General Psychology" [26], we know that a person can get higher scores on a variety of personality traits in the Big Five personality assessment. Therefore, we take the two personality characteristics with higher scores as the personality characteristics of the victim in the case. Table 3 shows the number of victims of each type of personality trait in the case.

Table 3
Number of people with various personality traits in cases

Personality Traits	Number
Neuroticism and Openness (NEU&OPE)	6339
Neuroticism and Agreeableness (NEU&AGR)	1272
Neuroticism and Conscientiousness (NEU&CON)	90
Agreeableness and Openness (AGR&OPE)	884
Agreeableness and Conscientiousness (AGR&CON)	271
Conscientiousness and Openness (CON&OPE)	262
Extroversion and Neuroticism (EXT&NEU)	678
Extroversion and Agreeableness (EXT&AGR)	311
Extroversion and Openness (EXT&OPE)	97
Extroversion and Conscientiousness (EXT&CON)	9
Count	10213

4.2. Correlation Analysis of Personality Characteristics and Case Information

This section will combine the personality characteristics obtained in 4.1 with real case information. By applying personality traits to the analysis of fraud cases, potential correlations in cases are explored in multiple dimensions.

4.2.1. Correlation Analysis between Personality Characteristics and the Trend of Fraudulent Means Over Time

Combining the analysis of several types of personality characteristics with a large number of cases and the time trend, the results obtained are shown in Fig. 3. As can be seen from Fig. 3, from the first quarter of 2017, the number of frauds experienced by people with personality traits such as NEU&OPE has fluctuated and increased, from 4 in the first quarter of 2017 to 945 in the second quarter of 2021. pieces. For the personality traits of NEU&AGR, AGR&OPE and AGR&CON, the number of frauds was relatively stable, showing a trend of first increasing and then decreasing. These data illustrate differences in the probability of personality traits being defrauded in fraud cases.

4.2.2. Correlation Analysis between Personality Characteristics and Fraudulent Means

By analyzing the correlation between different types of personality characteristics and fraudulent means and the number of fraudulent, the results shown in Fig. 4 are obtained. From Fig. 4, we can know that from 2018 to April 2021, the number of people with NEU&OPE personality characteristics was very large in various fraud types. Among them, the number of cases of fraudulent provision of false services reached 1,662, and the number of cases of loan fraud and online review credit fraud also exceeded 1,200, the numbers reached 1,254 and 1,240 respectively, which shows that the fraudulent methods often used by criminals for NEU&OPE groups are to provide false services, loan fraud, and online review credit fraud. As for the NEU&AGR type of personality characteristics, the number of cases of loan fraud is the largest, reaching 478, while the number of cases of other types of fraudulent means has not exceeded 200, which shows that the people with NEU&AGR type of personality characteristics suffer more from Loan Fraud. This information suggests that victims with different personality characteristics have large differences in the means of being defrauded.

4.2.3. Correlation Analysis between Personality Characteristics and Fraudulent Means

Our case information comes from a single place with 18 areas. For the combined analysis of personality characteristics and different regions, we selected 10 regions with the top 10 cases, and the analysis results are shown in Fig. 5. From Fig. 5, it can be known that the most fraudulent personality traits in these areas are NEU&OPE, and the NP area has the largest number, reaching 1032 cases. But unlike other regions, the number of fraud cases suffered by people with NEU&AGR personality traits in the HYL region reached 372, which is more than the number of fraud cases suffered by people with NEU&AGR personality characteristics in other regions. This shows that there are regional differences in the proportion of fraud cases experienced by people with various personality types.

Through the experimental analysis in this section, we can know that the number of defrauded personality traits in fraud cases varies with time, the number of defrauded in different fraud methods, and the number of defrauded areas in different regions. Among them, the number of fraudulent NEU&OPE personality traits showed a fluctuating upward trend from the first quarter of 2017 to the second quarter of 2021. The number of cases in a single quarter reached 945 at most. In contrast, the trend of changes in other regions was relatively flat. In terms of the difference between personality characteristics and fraudulent means, Loan Fraud is the most fraudulent among NEU&AGR types, while Offering False Service is the most fraudulent among other types of personality. In terms of personality characteristics and the difference in the number of fraudulent cases in different regions, the number of fraudulent cases of NEU&OPE type and NEU&AGR type in HYL area were 443 and 372, respectively, with a small difference. In other regions, the amount of fraud experienced by these two personality types varied considerably. It can be seen that personality characteristics have many potential laws in fraud cases, which also shows that the work done in this paper is of great significance for exploring potential correlations in fraud cases.

5. Conclusion And Future Work

In this paper, we propose a personality feature detection model for telecommunications fraud cases based on the Big Five. Compared with the models previously proposed by other scholars, the average accuracy of personality detection is improved by 0.3%, reaching 61.2%. At the same time, we also combined the detected personality characteristics with real case information, and found that different personality characteristics have great differences in the probability of being defrauded, the means of being defrauded, and the number of defrauded people in different regions. Our research not only improves the accuracy of personality feature detection, but also innovatively combines personality features with information about fraud cases, and explores the potential correlation between the psychological level of victims and fraud cases. Compared with the existing research on telecommunication network fraud crime in my country, it has the advantages of wider dimension and stronger applicability.

For future work, we will conduct computer modeling of more psychological models and apply them in the prevention of telecommunication fraud cases. In addition, we will also study the law of occurrence of telecommunications fraud cases, and combine the law of occurrence with psychological models to develop a blocking mechanism for telecommunications fraud cases based on personality characteristics. It is expected to prevent the cases from the potential victim level before the occurrence of telecommunications fraud cases, and reduce the occurrence of fraud cases.

Declarations

Acknowledgments

This work was supported by the National Natural Science Foundation (Grant No. 61772099, Grant No. 61772098, Grant No. 61802039); the Science and Technology Innovation Leadership Support Program of Chongqing (Grant No. CSTCCXLJRC201917); the Innovation and Entrepreneurship Demonstration Team Cultivation Plan of Chongqing (Grant No. CSTC2017kjrc-cxcytd0063); Chongqing Research Program of Basic Research and Frontier Technology (Grant No. cstc2018jcyjAX0617); the Science and Technology Planning Project of Guangzhou (Grant No. 202102080382); the Scientific Research Project of the Open University of Guangdong(Grant No. RC2001).

References

1. The Ministry of Public Security deploys the "Yunjian-2020" Campaign to crack down on the loan type telecom network fraud crime cluster [J]. Information network security,2020,20(05):95.
2. Ye Yuqiu, Sang Tao, Shen Panpan. Research on Several Issues of "Broken Card" Action Cases [J]. Chinese Public Prosecutors,2021(14):26–31.
3. Hou Jianbin.(2021). Telecom network fraud crime shows a high incidence situation, experts call for highlighting the source governance. Decision Exploration (PART 1),26–27. doi:CNKI:SUN:JCTS.0.2021-06-013.

4. Junjie Lin, Wenji Mao, Daniel D. Zeng, Personality-based refinement for sentiment classification in microblog, Knowledge-Based Systems, Volume 132, 2017, Pages 204–214, ISSN 0950–7051. https://doi.org/10.1016/j.knosys.2017.06.031.
5. Zhancheng Ren, Qiang Shen, Xiaolei Diao, Hao Xu, A sentiment-aware deep learning approach for personality detection from text, Information Processing & Management, Volume 58, Issue 3, 2021, 102532, ISSN 0306–4573. https://doi.org/10.1016/j.ipm.2021.102532.
6. Park, S., Lee, Y., Seong, S.J. et al. A cross-sectional study about associations between personality characteristics and mental health service utilization in a Korean national community sample of adults with psychiatric disorders. BMC Psychiatry 17, 170 (2017). https://doi.org/10.1186/s12888-017-1322-2.
7. Rasa Barkauskiene, Gabriele Skabeikyte, Lina Gervinskaite-Paulaitiene, Personality pathology in adolescents as a new line of scientific inquiry in Lithuania: mapping a research program development, Current Opinion in Psychology, Volume 37, 2021, Pages 72–76, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2020.08.011.
8. Olukayode Abayomi Sorunke, 2016. "Personal Ethics and Fraudster Motivation: The Missing Link in Fraud Triangle and Fraud Diamond Theories," International Journal of Academic Research in Business and Social Sciences, Human Resource Management Academic Research Society, International Journal of Academic Research in Business and Social Sciences, vol. 6(2), pages 159–165, February.
9. Xing T, Sun F, Wang K, et al. Vulnerability to fraud among Chinese older adults: do personality traits and loneliness matter[J]. Journal of Elder Abuse And Neglect, 2020, 32(1):46–59.
10. AlexRudnicky. Can Artificial Neural Networks Learn Language Models? Sixth International Conference on Spoken Language Processing, ICSLP 2000 INTERSPEECH 2000, Beijing, China, October 16–20, 2000. 2000.
11. Bengio Y, Ducharme R, Vincent P. A neural probabilistic language model[J]. Advances in Neural Information Processing Systems, 2000, 13.
12. Wilbur, W.J., Kim, W. The ineffectiveness of within-document term frequency in text classification. Inf Retrieval 12, 509–525 (2009). https://doi.org/10.1007/s10791-008-9069-5.
13. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space.CoRR, abs/1301.3781, 2013.
14. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019:4171–4186.
15. M. Bobur, K. Aibek, B. Abay and F. Hajiyev, "Anomaly Detection Between Judicial Text-Based Documents," 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT), 2020, pp. 1–5, doi: 10.1109/AICT50176.2020.9368621.
16. Shengli Zhou, XinWang, ZeruiYang. Monitoring and Early Warning of New Cyber-Telecom Crime Platform Based on BERT Migration Learning. CHINA COMMUNICATIONS, 2020, 17(3):9.

17. R. J. Gregory, “The History of Psychological Testing,” in Psychological Testing: History, principles, and applications., 7th ed. Pearson, 2013, ch. 2, pp. 32–58.
18. O. P. John, L. P. Naumann, and C. J. Soto, “Paradigm shift to the integrative Big Five Trait taxonomy,” Handbook of Personality: Theory and Research, pp. 114–158, 2008.
19. Mairesse F, Walker M A, Mehl M R, et al. Using linguistic cues for the automatic recognition of personality in conversation and text[J]. Journal of artificial intelligence research, 2007, 30: 457–500.
20. Sun, X., Liu, B., Meng, Q. et al. Group-level personality detection based on text generated networks. World Wide Web 23, 1887–1906 (2020). https://doi.org/10.1007/s11280-019-00729-2.
21. Kazameini A, Fatehi S, Mehta Y, et al. Personality trait detection using bagged svm over bert word embedding ensembles[J]. arXiv preprint arXiv:2010.01309, 2020.
22. Kazemeini, Amirmohammad, et al. "Interpretable Representation Learning for Personality Detection." 2021 International Conference on Data Mining Workshops (ICDMW). IEEE, 2021.
23. J. W. Pennebaker and L. A. King, “Linguistic styles: language use as an individual difference.” Journal of Personality and Social Psychology, vol. 77, no. 6, p. 1296, 1999.
24. Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. 2017. Deep learning-based document modeling for personality detection from text.IEEE Intelligent Systems, 32(2):74–79.
25. Mehta, Yash, et al. "Bottom-up and top-down: Predicting personality with psycholinguistic and language model features." 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020.
26. Danling Peng. General Psychology: Beijing Normal University Press, 2012.

Figures

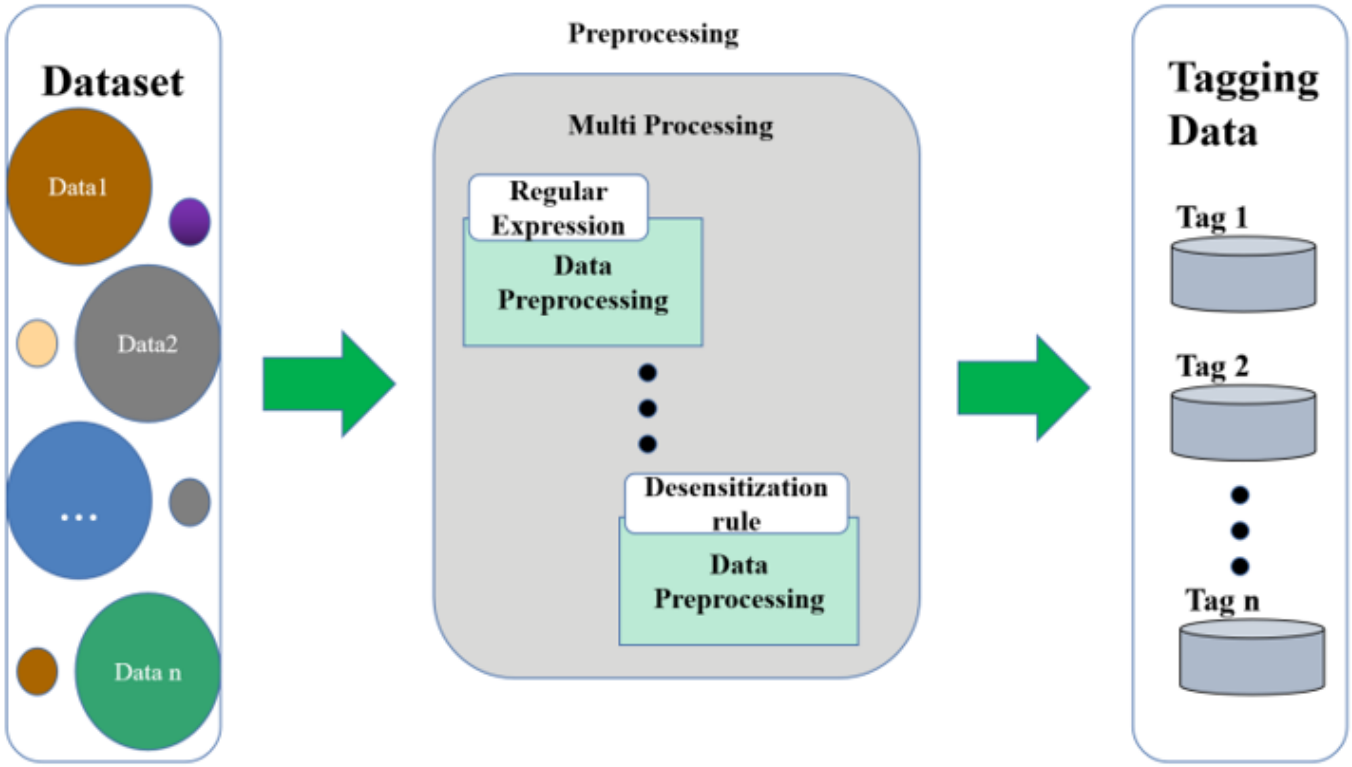

Figure 1

Data cleaning process

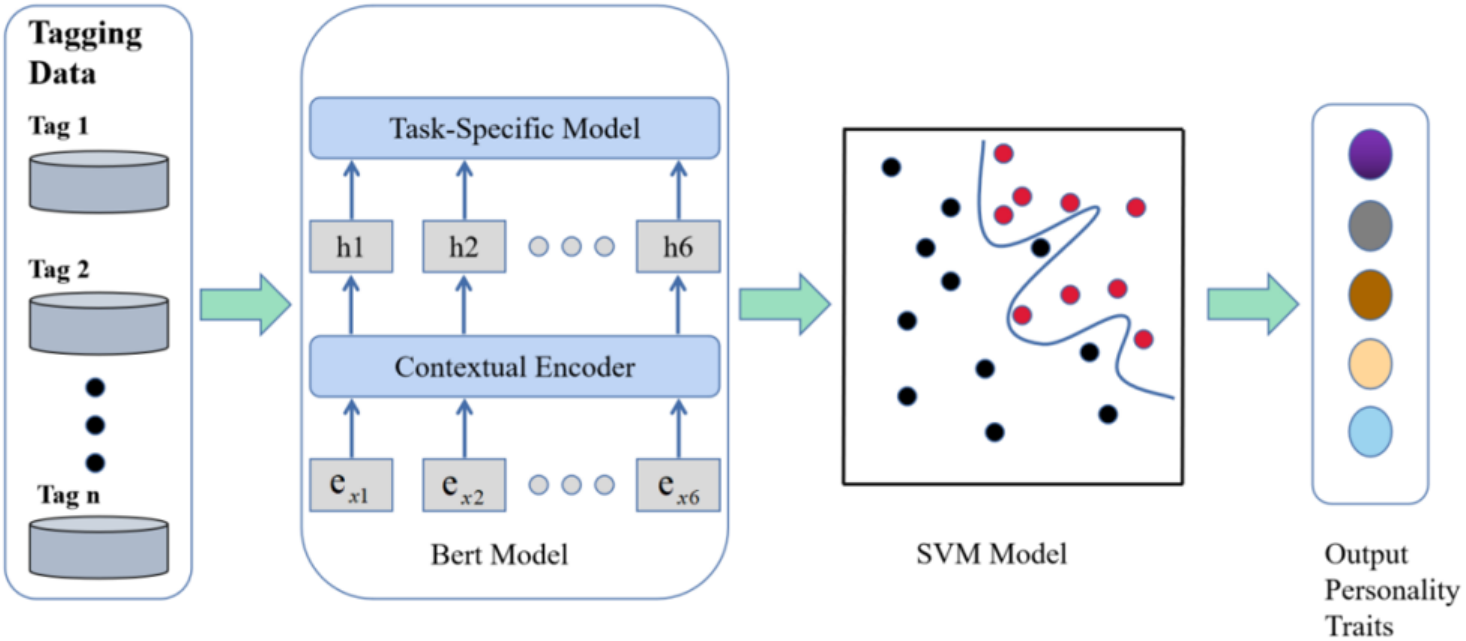

Figure 2

Personality detection model process

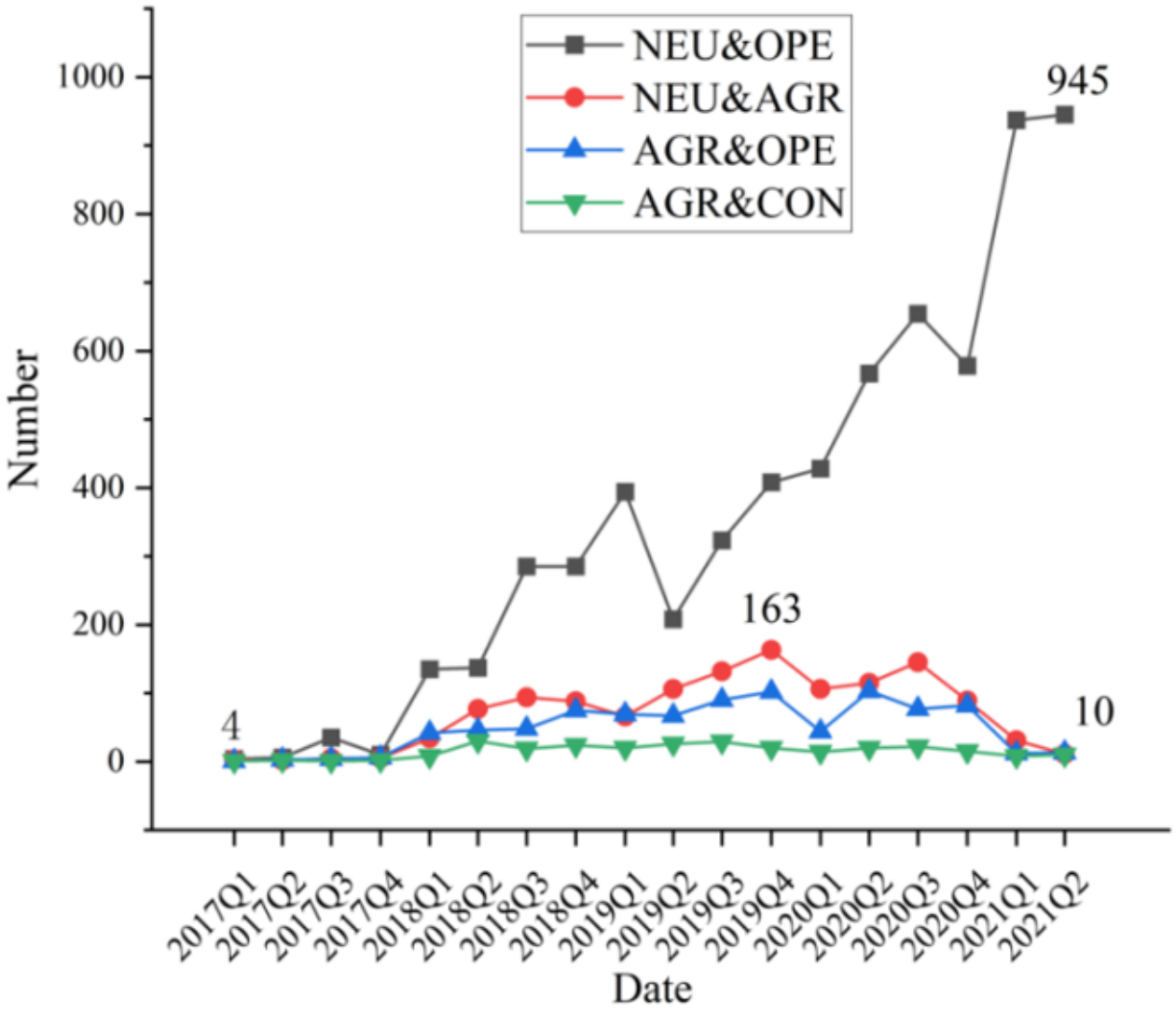

Figure 3

Fraud numbers and time trends of some personality traits

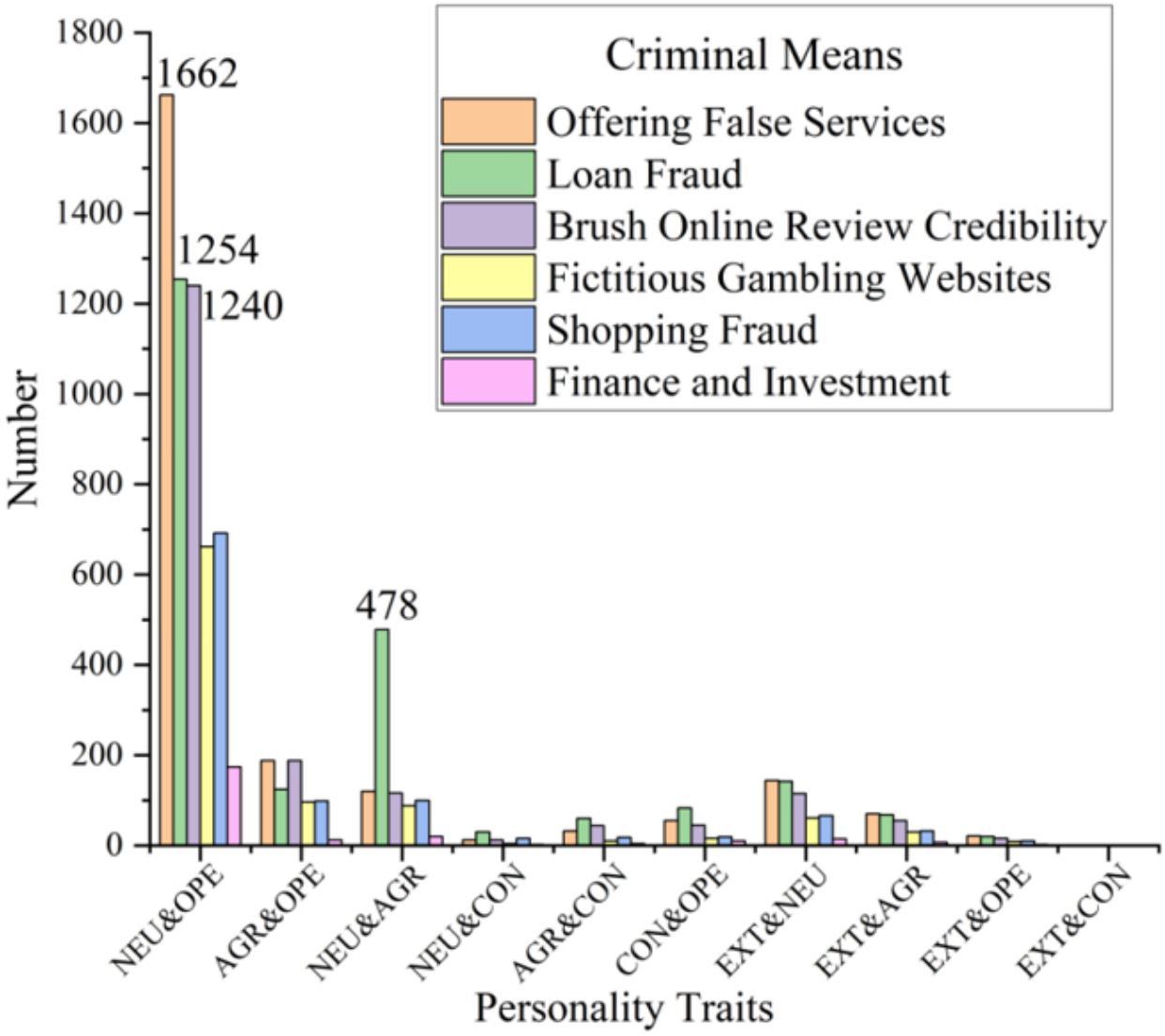

Figure 4

The relationship between various personality characteristics and the means of fraud and the number of frauds

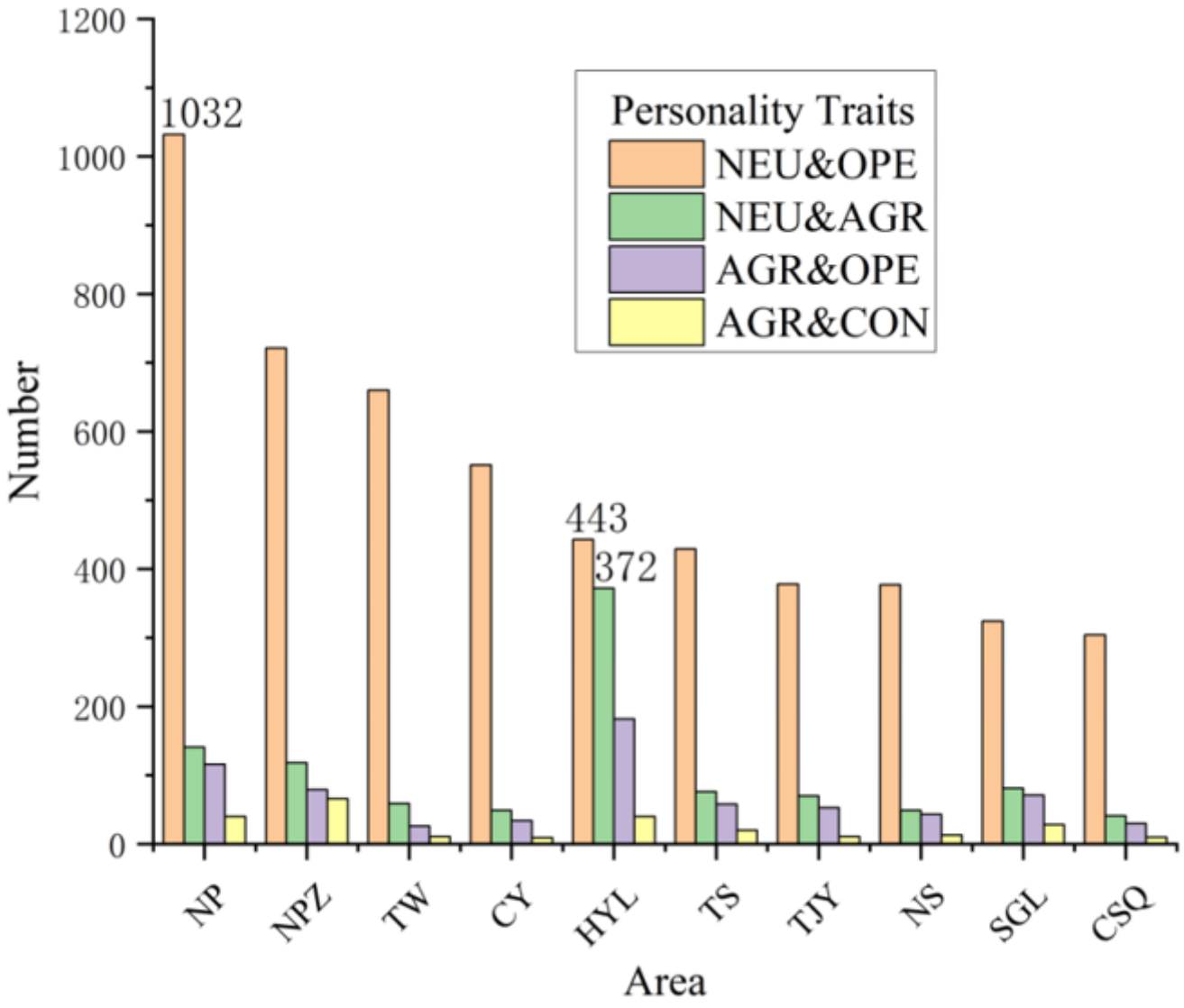

Figure 5

Correlation of personality traits in different regions