

Research on the Behavior Analysis and Prediction of Credit Card Fraud

Can Wang

School of Mathematics
Chengdu Normal University
Chengdu, China 611130

Abstract—With the development of economic globalization and the improvement of information technology, the new consumption mode of credit card overdraft is gradually applied in China, and there is a tendency of expanded scope of use. While development and challenges coexist, the problem of credit card fraud is increasingly exposed. This kind of fraud greatly hinders the development of national financial system, brings new problems to the risk control of commercial banks, and seriously affects the normal operation of market economy. Therefore, it is urgent to analyze and solve the problem. Based on this, this paper first explains the current situation of credit card use in China, explains the meaning of credit card fraud, then analyzes the motivation, influencing factors and identification consequences of fraud, and then establishes a monitoring and identification model of credit card fraud based on data mining algorithm and machine learning technology.

Keywords: *credit card fraud, behavior analysis, risk determination, prediction research*

I. INTRODUCTION

Bank card is a convenient financial tool for people to consume in today's society. There are two main payment methods, namely online payment and offline payment. Offline payment is usually realized through self-service machine, and online payment can be divided into credit card payment and debit card payment, among which credit card payment has developed rapidly in China and has been accepted by the general public. This kind of payment means that the card held by the cardholder has a certain limit, within which the card can be consumed first and then repaid on time. With the improvement of Chinese people's living standards and the change of consumption concept, the number of credit card holders has maintained steady growth in recent years.

With the expansion of demand, commercial banks have also gradually invested in the credit card market, constantly optimizing their business methods and contents, and launching a series of credit card activity products to attract relevant customers. According to the "Blue Book on the Development of China's Bank Card Industry (2019)" released by the China Banking Association, as of June 2019, the number of credit cards issued reached 970 million, up

22.8 percent year-on-year in 2018, and the total amount of credit card transactions increased by over 10 times, from 3.5 trillion yuan to 38.2 trillion yuan. Against the background of the huge amount of credit card transactions, the problem of credit card fraud is also increasingly exposed, and the fraud is constantly updated, which has a great impact on the market economy and financial order. There are several ways to commit credit card fraud. In the first circumstance, the cardholder swipes the card for someone else, which increases his bonus points and get cash from the actual consumer. The harm cashing out this way is relatively small, because real consumption has occurred. In the second circumstance, through fake online transactions, cardholders complete fake transactions through e-commerce websites, such as Taobao, etc., and merchants return the money of cardholders in cash, etc., so that merchants get sales volume and cardholders manage to cash out. The third circumstance is cash-out through intermediary agent. In this case, the cardholder makes false card transaction through the illegal merchant who refunds him on the spot, with just an amount of handling fee, generally lower than the withdrawal fee at a bank. In the fourth circumstance, by applying for several credit cards, the cardholder can use the interval of credit card repayment time to pay off the debt of one credit card with the amount of another. In short, it is the behavior of robbing Peter to pay Paul by "repaying one card with another". The main purpose of these frauds is to cash out, but the harm of them should not be underestimated. For the cardholder, having a credit card is a sign of personal credit security. However, fraud is not only a breach of credit, but also an illegal act. Once the fraud is identified and exposed, it will be subject to legal sanctions. Cardholders may think, cashing out by credit card is the equivalent of an interest-free, unsecured loan. But it will cause great loss to commercial banks, and what is relatively serious is the loss of interest. At the same time, the real purpose of the cash taken cannot be grasped, nor can the funds be effectively monitored and controlled. This increases the risk of non-performing loans and bad debts, affects the cash flow and financial statistics of the market economy, which greatly endangers the financial order.

If the holder of a credit card uses the credit line for cash, it essentially turns the credit line of the credit card into an unsecured bank loan. This behavior will affect the normal

cash flow of the country, and the accuracy of financial statistical indicators. And at the same time, it will also have an impact on the normal payment system, increase credit supply in a disguised way, away from the normal credit supply channels, and affect the judgment of the national financial authorities on the direction of money and the implementation of macro policies.

In Europe and the United States and other developed countries, the database information and credit system are very perfect. With this advantage, scholars have done a lot of empirical analysis and research on fraud. Here are some representative studies. Kreps Migrom Roberts and Wilson (1982) emphasized the reputation in credit card market, holding that in balanced cooperation, incomplete information of credit card applicants is a repetitive game; therefore, the basic information of mobile phone customers, including age, education level, assets, etc. should be fully analyzed to the greatest extent, so as to understand the risk level of card holders, which is beneficial for maintaining the stable growth of customers. Rock et al. specifically quantified seven variables such as cardholder's income, occupation and assets, and discussed the relationship between variables and fraudulent behaviors. Siglitz (2000) also emphasized the role of reputation and, on this basis, raises the cost of enforcing contracts. Bhattacharyya S et al. proposed a new idea on the model of credit card fraud prediction by using data mining algorithm. With the change of the consumption pattern of Chinese residents, domestic scholars have also done research on credit card fraud gradually. Representative scholars such as Tong Gou have conducted a hierarchical study on the fraud of cards, and on this basis have given some preventive measures.

II. ANALYSIS OF CREDIT CARD FRAUD

A. Motivation explanation

Motivation was originally a specialized term in psychology. In combination with the fraudulent behavior in financial and economic market, foreign academic circles have conducted a detailed and multi-faceted research on the behavioral motivation of fraud. Among them, Professor Donald Cressey, who made pioneering contributions, is called an anti-fraud expert. He proposed that the emergence of fraud tendency is mainly influenced by the combination of motivation, opportunity and rationality. Based on this, from the analysis of individual factors, the occurrence of credit card fraud is mainly constrained by moral quality, which is not controlled by the regulatory system, but only by the behavioral control of individuals over themselves. From the comprehensive analysis of factors, the occurrence of credit card fraud can be controlled according to the probability of fraud being discovered and the severe punishment after discovery, etc.

The rise of behavioral finance has enriched models and data for the study of fraud, and the results can be summarized in the following three aspects. The first is that individual demographics, such as age, education, and occupation, are important in predicting credit-card fraud.

Studies have found that young individuals without a stable career and lack of professional education have greater probability of credit card fraud; on the contrary, with the increase of age and enhanced stability of career, fraud will reduce due to increased level of life experience and awareness enhancement, accepted the professional education of the individual, especially for the possibility of credit card fraud will be greatly reduced. Individuals with professional education, in particular, are much less likely to engage in credit card fraud. The third is the impact of individual psychological character traits. Studies have found that many credit card frauds begin with an optimistic miscalculation. They believe that other means and methods can cover up the wrong behavior. The fact is often that excessive narcissism will force them to continue with the wrong behavior to cover up the previous fraud. The third is personal experience. Studies have found that individuals who had previously committed a crime, including driving under the influence, using drugs or causing a disturbance, were more likely to commit credit card fraud later, which is primarily a lack of self-control that is consistent across all behaviors. And credit card fraud is less likely to occur in individuals who have served in the military.

B. Influencing factors

The implementation and occurrence of credit card fraud is also restricted and affected by external factors, which are summarized in the following aspects. The first is the development state of the market economy. From the two stages of the depression and the boom, it is believed that credit card fraud will be exposed when the economic state falls; in the boom, especially when the economic value is at the peak, the fraud will decrease with the stable operation of the economic market. The second is the pressure of competition in this industry. From the perspective of the two sides of banks and credit individuals, it is believed that in a less focused competitive economy, the number of institutions encouraging individuals to apply for credit cards increases the likelihood of credit card fraud, and these characteristics are increasingly rampant in the increasingly competitive financial industry. The third is the regulatory environment. Research shows that intensive supervision will restrain credit card fraud to a large extent, especially after the domestic implementation of linking personal credit record with housing, travel, employment and other behaviors, individuals have more concerns before committing credit card fraud, and consider more about the actual interests, so that the occurrence of fraud reduces.

C. Identifying the consequences

Credit card fraud not only damages personal credit reputation, but also restricts the development of bank finance, which disturbs the normal order of economic market, and causes serious impact on the whole society. From a micro perspective, studies found that the impact of credit card fraud on bank value should not be underestimated. The normal operation of an enterprise cannot be separated from the healthy development of banking business. When the banking business is affected by fraud, it will indirectly or directly

affect the orderly credit of the enterprise market. At the same time, individuals once suspected of fraud will face severe punishment, from employment, housing and other aspects of restrictions in life, and even face the consequences of legal sentencing. From a macro point of view of the financial system as a whole, the consequences of credit card fraud on society as a whole are huge, especially the premeditated group fraud. Welfare and social cost analysis from this aspect can be costly, which is also the follow-up focus of scholars.

III. PREDICTION MODEL CONSTRUCTION

A. Algorithm Foundation

Combining with the basic information variables of the credit card holder, the decision tree model is selected for analysis, considering the pruning of the information first. The construction process of decision tree mainly includes two parts: establishment and pruning. The process of pruning is to cut off some subtrees or leaf nodes in the decision tree model. The main purpose is to avoid overfitting by simplifying the decision tree model.

L. Barreiman put forward the CART algorithm based on the theory of information entropy in 1984. In the process of constructing the decision tree, the minimized Gini coefficient is taken as the far side, in which the calculation process of the Gini coefficient $Gini$ is as follows:

First, the sample's index $Gini$ is calculated:

$$Gini(S) = 1 - \sum_{i=1}^n P_i^2$$

In the formula, P_i represents the probability of emergence of category C_j in sample set S

Then, each $Gini$ index divided is calculated

If the sample set S is divided into two subsets S_1 and S_2 , the $Gini$ index divided is :

$$Gini_{split}(S) = \frac{|S_1|}{|S|} Gini(S_1) + \frac{|S_2|}{|S|} Gini(S_2)$$

When the nodes of CART algorithm are split, the smaller $Gini_{split}(S)$, which is the $Gini$ index of each partition is, the more reasonable the partition will be. The CART algorithm is also capable of processing discrete and continuous data and is insensitive to outliers and vacancy values.

The learning of multiple decision trees forms the random forest algorithm, in which each decision tree selects the data set by bagging and randomly classifies the attributes. The

random forest is adopted from bagging based on the small change. In the algorithm, m subsamples are extracted by means of random sampling with reversion on the original data set, and the m subsamples are used to train m base learners, thus reducing the variance of the model and increasing its stability.

B. Modeling and evaluation

The IP address and POS device information in the transaction information can fully reflect the abnormal situation of bank card transaction. On the one hand, since the IP address belongs to the underlying Internet protocol, all devices connected to the Internet will have an IP address. Therefore, the IP address has the features of easy identification for visitors, quick blocking and tracking, and the IP address is managed by the world's five largest Internet agencies and cannot be faked. These characteristics give the IP address an irreplaceable role in the field of network security, risk control, anti-fraud and other fields. The IP address under the fixed network has a fixed geographical location in a certain period of time and through depth analysis, the location of IP address can be more accurate. Through the degree of location dispersion of the user of the IP address, obvious aggression can be identified, such as an IP address appeared around the world in a short period of time. The areas with high risks should be identified. If an IP address used by a region where dark businesses emerge for a long term, the IP address in this region has a degree of risk. Therefore whether an IP is in high risk regions, offsite, abroad and matches the device address or not can be taken as important reference indexes in anti-fraud prediction model. According to the stipulation of the bank that an ip not used at the application address is considered "offsite", as long as a credit card is used offsite, the payment can be refused due to information asymmetry. To some extent, this allows the indicators of equipment in different provinces and cities to predict the risk of machine theft, merchants' illegal transfer, and credit card theft. The merchants that malice cash out usually have the characteristics of block trade, low credit card discount rate, low registered capital, and low popularity, obviously not matching with the characteristics of merchants with low credit card discount rate, low registered capital, and low popularity.

After preprocessing the original data, it is necessary to extract the data features based on business knowledge, that is, to establish the feature engineering. This work is important for building and training models using machine learning because the extracted data features, especially feature dimensions and quality, determine the upper limit of the machine learning model's capabilities. There are many approaches to feature engineering. Combining the business knowledge of credit card anti-fraud and the actual scenario, this paper chooses the dimensionality reduction idea based on principal component analysis method. The specific approach is to extract the basic information of samples of known n first, and obtain advanced distribution characteristics through statistical analysis, and then define these sample information as P indicators, so as to obtain

np data. Then, attribute filtering is carried out from the feature sub-set to make the constructed model more effective, and the input sampling of selected data attributes is processed, and the high-dimensional data is reduced to low-dimensional data through mapping changes. Finally, the

irrelevant redundant variables are taken out to achieve the final goal of the dimension reduction.

In this paper, 14 features are selected after dimension reduction by principal component analysis according to the indicator system of bank card transaction warning mentioned above, as shown in "Table I".

TABLE I. SELECTION OF ANTI-FRAUD FEATURES

Number	Name	Type	Number	Name	Type
1	IP in high risk regions	double	8	The device code is empty	double
2	Grade of transmission	double	9	IP in a different city	double
3	Grade of the card	double	10	Equipment in a different province	double
4	Online transaction	double	11	Equipment in a different city	double
5	IP matches the device address	double	12	High risk trade identification	string
6	IP abroad	double	13	block trade	double
7	IP is empty	double	14	abnormal time	double

The correlation between the selected features is analyzed in a multidimensional way. This process is carried out by Spearman correlation coefficient method of non-parametric statistics, that is, with the rank among variables being considered, and the corresponding correlation analysis is made according to the size relation, instead of the distribution of original variables. The characteristic of this method is that as long as there is a monotonic functional relationship between variables, they can be considered completely correlated. Meanwhile, the requirement for accurate distribution between samples is also reduced, because the overall distribution can be obtained through joint probability density.

After extracting the sample features, the first step is to determine the number of feature variables, $m(m < M)$. M represents the number of characteristic variables, that is, the number of characteristic variables needed to produce a decision tree. The second step is to get k new self-service sample sets by carrying out sampling with replacement through the bootstrap method, according to which K decision-making trees are then set up, and the samples that are not sampled each time make up the k Out-Of-Bag, OOB. The third step is to grow the sample set generated by each self-service sampling into each decision tree. The nodes at each node are selected according to the principle of least impurity, and then the nodes are fully grown without pruning. Finally, the prediction set is predicted according to the generated decision tree, and the average value of the results generated by each tree is the final prediction result.

From the generated classification results, it can be seen that the decision tree branches at such attributes as device exception, card grade and grade of transmission, and IP in a different city. The test set is used to evaluate the model results of the decision tree. The results are shown below. The

evaluation shows that the accuracy of the decision tree model is 85.57% and the ROC 0.917. The training set data is substituted into the random forest model, and the test set data is used to evaluate the model results of the random forest. It shows that the accuracy rate is 86.55%, and the ROC 0.931.

IV. CONCLUSION

With the change of payment method, the number of credit card holders is increasing year by year. The increase in quantity brings hidden dangers to the safety of quality, resulting in an increasing trend of fraudulent behaviors, which brings immeasurable losses to banks and potential hidden dangers to financial order, which seriously disturbs the order of market economy. For this reason, this paper first conducts behavior analysis from the motivation, influence factors, consequences identification and disclosure of credit card fraud. It holds that the motivation of credit card fraud is mainly influenced by the cardholder's psychology, life pressure, occupation, education level, and personal experience and so on. At the same time, outside regulation also plays a role in such fraud. And once the fraud is identified and exposed, the legal punishment also restricts the occurrence of this behavior.

The problem faces the challenge of being solved. The identification and prediction of credit card fraud is a difficult problem in the whole academic field. This paper attempts to use scientific methods to determine and predict the potential bank card fraud in order to improve the efficiency of economic risk management. Combined with the previous research on credit card fraud risk, this paper analyzes how banks manage risk after the occurrence of fraud, uses decision tree to prune data, and then screens all feature engineering of the business. It extracts factors including IP in high risk regions, grade of transmission, grade of card, online transaction, IP matching device address, IP abroad, empty IP,

empty device code, IP in a different country, equipment in a different province, high risk trade identification, block trade and abnormal time, etc., and builds the random forest model to determine whether there is credit card fraud in the process of usage of customers. Compared with the research on the determination of credit card fraud based on multiple linear regression model, the advantage of random forest algorithm over multiple linear regression is that its prediction is more accurate. Through data test set and training set, it is verified that random forest has good applicability and excellent feature selection performance, so that it can provide more targeted suggestions for the development of bank credit, and thus can be popularized and applied to stock investment, medical treatment field, crime field and other aspects in the future, with good applicability.

REFERENCES

- [1] Du Lulu. Research on Defrauding Risk of Chinese Credit Card and Solutions [D]. Jinan University, 2012. (in Chinese)
- [2] Jiang Zihan, Zhou Xinyue, Ye Xinlan, Jin Jiji. Countermeasures Against Malicious Cashing in the Era of Big Data [J]. China Market, 2017(11): 97-98. (in Chinese)
- [3] Huang Miao. The Study for Fraud Detection of Credit Card Based on Imbalanced Data [D]. Nanjing Normal University, 2017. (in Chinese)
- [4] Song Yang. Detection of credit card fraud based on feature selection [J]. Guizhou Science, 2017, 35(06): 83-87. (in Chinese)
- [5] Li Yangzhi. Research on Credit Card Risk Management Based on Decision Tree [J]. Financial and Management Research, 2018(02): 1-21 (in Chinese)
- [6] Chan P K, Fan W, Prodromidis A L , et al. Distributed data mining in credit card fraud detection[J]. IEEE Intelligent Systems, 1999, 14(6): 67-74.
- [7] Xu Shaocheng, Li Dongxi. Weighted Feature Selection Algorithm Based on Random Forest [J]. Statistics & Decision, 2008, 34(18): 25-28. (in Chinese)
- [8] Brause R, Langsdorf T, Hepp M. Neural Data Mining for Credit Card Fraud Detection[C]// Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on. IEEE, 1999.
- [9] Bhattacharyya S, Jha S, Tharakunnel K, et al. Data mining for credit card fraud: A comparative study [J]. Decision Support Systems, 2011, 50(3): 602-613.
- [10] Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism [J]. IEEE Internet of Things Journal, 2018, PP (99): 1-1.
- [11] Xu Qian. Fraud Risk Is Macabre and Anti-fraud Is Urgent — Interview with Derek Wylde, Former Global Head of Fraud Risk of HSBC [J]. The Banker, 2017(4): 30-33. (in Chinese)
- [12] Raj S B E, Portia A A. Analysis on credit card fraud detection methods[C]// 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET). IEEE, 2011.
- [13] Zareapoor M, Seeja. K R, Alam A M. Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria [J]. International Journal of Computer Applications, 2012, 52(3): 35-42.
- [14] Prabowo Y, Hendi. Building our defence against credit card fraud: a strategic view [J]. Journal of Money Laundering Control, 2011, 14(4): 371-386.
- [15] Yan Tianbing, Shen Li. Research on Credit Risk Management of Credit Card in China [J]. Jinan Finance, 2007(08): 47-49. (in Chinese)
- [16] Zhao Gang. Research on Credit Risk Management of Credit Card Business of Commercial Banks [D]. East China Normal University, 2007. (in Chinese)