Advances in Intelligent Systems Research, volume 147

International Conference on Network, Communication, Computer Engineering (NCCE 2018)

Application of Adaboost Algorithm and Immune Algorithm in Telecommunication Fraud Detection

Bing Wu [a)], Mengtao Li [b)] and Chunlai Zhou [c)]

School of Communication University of China, Beijing 1000000, China.

[a)] 280101747@qq.com,
[b)] 1449047566@qq.com,
[c)]clzhou@cuc.edu.cn.

Abstract. Fraud detection is one of the biggest challenges facing the telecommunication industry now and, in the future, the fight against fraud and anti-fraud has also reached a new stage. Finding a time-sensitive fraud detection method is an important way for operators to solve this problem. In this article, to solve the low accuracy of the general algorithm, an adaptive improvement algorithm is proposed. The common algorithms are combined and enhanced, which greatly improves the accuracy of the detection results. Here we take the artificial immune algorithm as an example. The main application is combining machine learning and immune algorithm to apply to telecommunication fraud detection. The combination of the two is more conducive to pushing research on telecommunication fraud to a new stage for the future telecommunication industry.

Key words: Adaboost algorithm, Artificial immune algorithm, Fraud detection.

INTRODUCTION

With the advent of the information age, mobile phones have become an important part of people's lives. It has become an important medium for people to obtain social information and provides people with efficient services to ensure the high efficiency of mobile phone calls is the responsibility of mobile operators. Along with the popularization of mobile phones and the development of communication technologies in China, the demand for mobile amateurs is getting higher and higher, and the demand for services in the communications industry is also increasing. At the same time, fraud in the telecommunications industry also occurs frequently. According to Juniper Research survey by market research institution [9] in the communications industry in 2016, the value of operating revenue was nearly 200 billion US dollars. Global estimates of $300 billion due to billing issues and telecommunications fraud losses. Losses in mobile communications in Africa and the Middle East account for as much as 15% of total revenue in revenue each year, and Europe and North America account for approximately 1% and 2.8% [9]. Some analysts and research institutes predict that the losses from fraud in the future mobile network may increase further.

With the advent of the information age, mobile phones have become an important part of people's lives. It has become an important medium for people to obtain social information and provides people with efficient services to ensure the high efficiency of mobile phone calls is the responsibility of mobile operators. Along with the popularization of mobile phones and the development of communication technologies in China, the demand for mobile amateurs is getting higher and higher, and the demand for services in the communications industry is also increasing. At the same time, fraud in the telecommunications industry also occurs frequently. According to Juniper Research survey by market research institution [9] in the communications industry in 2016, the value of operating revenue was nearly 200 billion US dollars. Global estimates of $300 billion due to billing issues and telecommunications fraud losses. Losses in mobile communications in Africa and the Middle East account for as

much as 15% of total revenue in revenue each year, and Europe and North America account for approximately 1% and 2.8% [9]. Some analysts and research institutes predict that the losses from fraud in the future mobile network may increase further.

The current fraudulent behavior is diversified [6-12]. Operators prevent or prevent fraud by many means. Telecom operators in Europe and the United States and other developed countries have specially installed anti-fraud systems at network control centers and set up specialized Anti-fraud teams and departments. This can reduce the loss by 30% [8]. Domestic anti-telecom fraud measures are still not in place. Only a few provinces and cities follow the model of foreign countries to conduct research and lack systematic solutions, In the future development, China's mobile network is indispensable. Under the background of the substantial improvement of people's living standards in China, people's demand for services on the mobile side will certainly increase and will not decrease. Under the temptation of huge benefits, fraudsters will Complete fraud through multiple means [8-14].

ANALYSIS OF FRAUD CHARACTERISTICS

There are various types of fraud. Depending on the target, the fraud can be divided into many types. The most common ones are credit card fraud and telecommunication fraud. The way of credit card fraud is mainly through handling fake cards, cashing out, stealing brushes and so on. The occurrence of fraud caused customers to feel that their property was threatened and mistrust the bank, which may lead to the loss of many customers. Nowadays, the development of the communications industry [11, 13], telecom operators are gradually driven by technology. Market-driven, customer-driven conversions, the market competition of various operators attracts users by lowering the threshold of entry into the network, making the customer market expand rapidly. The benefits of doing so are obvious, but at the same time the drawbacks brought by it are also very prominent, although most customers are not malicious. Arrears, but it is inevitable that lawbreakers have received high profits to engage in fraud through technology or in conjunction with local mobile operators [11].

Understanding the characteristics of fraud and the types of fraud can better provide us with the direction and focus of the research. From the existing data combined with domestic and foreign research, we summarize the characteristics of fraudulent customers as follows [6]:

(1) The customer's talk time is unusually long in several months and the length of time is significantly extended.
(2) The customer's history is stable, but the amount of spending in one month has increased dramatically.
(3) The customer has a short card-running period and has a high arrears.
(4) The customer's history is not good and his credibility is poor.
(5) The customer is not a real name, the long-distance charge is very high, and the call charge is very low.
(6) International phone disguised as local phone fee fraud, such as making friends hotlines, foreign consulting hotlines, etc. refused to pay telephone charges.
(7) Phone hijackers use technical means to record services on others' bills.
(8) The above is the most common fraud model at this stage.

Research Status

There is still some gap in the research on fraud at the current stage. There are many fraud detection systems currently used by operators in China [14], anti-fraud based on rules, anti-fraud based on neural networks, anti-fraud based on rules and neural networks, anti-fraud system based on data mining, trend analysis anti-fraud, Law, the availability of these methods is higher than the anti-fraud system based on data mining.

The current fraud products mainly include Fraud Office of Ericsson and Fradview of ECtel, Teradata of NCR and other products. NCR Teradata's data warehouse is more complete, it can be close to real-time detection, feature analysis of customers, multi-threaded multi-process load system, OLAP reporting capabilities. In China, GoldenEye, a real-time monitoring and analysis system for anti-fraud and anti-fraud, has certain shortcomings in these products. In the above products, the accuracy of detection is relatively low, and there are many elements that cause this phenomenon. The most important one is the huge amount of communication data and the endless number of frauds. Existing detection technologies have a low accuracy for fraud detection, it is often easy to generate false positives or false negatives. It is urgent to propose a new detection method [2]. The main reason for this situation is the low recognition rate of the algorithm. A new detection method combining machine learning and artificial immune algorithm is proposed to be applied to telecommunication fraud to not only improve the detection accuracy but also make the system structure have good learning ability, making the system more flexible, this approach is mainly

framed by adaboost.

Basic Definition of Adaboost

The Boosting method is a method to improve the accuracy of weak classification algorithms by constructing a series of predictive functions and then combining them into a predictive function in a certain way [14, 8]. He is a framework algorithm, which mainly obtains sample subsets through the operation of sample sets, and then uses a weak classification algorithm to train a sample subset to generate a series of base classifiers. He can be used to improve the recognition rate of other weak classification algorithms, that is, to put other weak classification algorithms as the base classification algorithm in the Boosting framework, and to use the Boosting framework to operate the training sample set to obtain different training sample subsets. The sample subset is used to train the generated base classifier; each base classifier is used to generate a base classifier on the sample set. In this way, n base classifiers can be generated after a given number n of training rounds, and then the Boosting framework algorithm then weights the n base classifiers to produce a final result classifier. This improves the recognition rate of the weak classification algorithm

SYSTEM IMPLEMENTATION

Principle of Algorithm

The main working principle of AdaBoost in English is “Adaptive Boosting” [14]:

(1) First, the weight distribution D1 of the training data is initialized. Assuming that there are N training sample data, each training sample will be given the same weight when it starts: W1=1/N.

(2) Training weak classifier HI, If the training sample point is somewhere, it is classified by the weak classifier accurately. In the next training set, its corresponding weight needs to be reduced accordingly. On the contrary, if it is misclassified, its weight should increase. The weighted updated sample set is used to train the next classifier and the entire training process is iteratively repeated.

(3) Finally, each weak classifier is reorganized to form a strong classifier. After the training process of each weak classifier is over, the weight of the classifier has a smaller error rate. Make it play relative role in the final classification function and reduce the classifier with large classification error rate and make it play a relatively small role in the final classification function.

(4) Classifier---Immune Algorithm [14]:

(5) The introduction of the artificial immune system originated in Japan in December 1996, it was the first time that an international symposium based on the immune system was held in Japan. Inspired by the biological immune system, the concept of the "Artificial Immune System" (AIS) was first proposed. The biological immune system is an adaptive, self-organizing, distributed system that can withstand invading self-defense systems of foreign pathogens. What we call the artificial immune system is inspired by this self-defense mechanism. At present, the main application area of artificial immunology is computer security. As an intrusion detection mechanism, it prevents the invasion of foreign viruses. At home, Harbin Institute of Technology professor applied the artificial immune system to anomaly detection. Use graphs to represent our correspondence [2].

The fraud detection method is basically divided into two kinds of misuse detection and anomaly detection. The misuse detection method is mainly to model the known fraud characteristics, and then the user uses these established models to detect the user's communication behaviour. If a matching model is found, this user will be judged to be fraudulent. The main advantage of this detection mode is that it is simple and convenient, but its disadvantage is that the false alarm rate is high. The anomaly detection method is based on the user's daily behaviour as a standard and associates fraudulent behaviour with daily behaviour. When the user's consumption is unconventional, this will be recorded as a record. When this behaviour occurs many times, the user is tracked and sent out. Warning, the main advantage of anomaly detection is that it can detect fraud patterns that have not previously appeared [2].

Telecommunications users are the object of telecommunication fraud detection. The characteristics of different users' communication behaviours are also different. The normal communication consumption behaviour of users is easy to obtain. Given the diversity of fraud, we choose artificial immune algorithms to deal with the normal communication behaviour of known users. However, we divide the user's data in this way, the accuracy rate is very low, and can be federated through the framework of adaboost.

TABLE 1. Comparison of immune algorithm and telecommunication data [2]

Comparison project	Artificial immunology	Telecommunication
Normal set	self	Normal user and communication data
Abnormal set	Non-self	Abnormal user and communication data
Modeling	Lymphocyte B	Fraud recognition model

Algorithm Architecture

Combining immune algorithm with adaboost algorithm to form a new detection model with high accuracy, Combination method as shown:

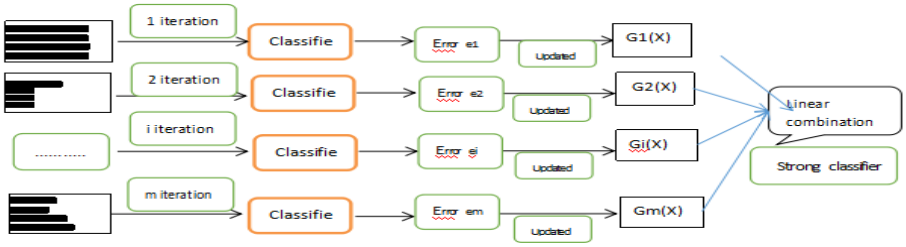

FIGURE 1. Combination method [14]

The following is the process of combining the adaboost algorithm with the immune algorithm [8, 9]:

The first Weight distribution of initial training data $D_1(i)=(w_1,w_2.....w_N)=(\frac{1}{N}.....\frac{1}{N})$,second, iterate t=1...T, Select the classifier h with the lowest error rate as the tth basic classifier Ht, and calculate the weak classifier Ht:X→ {-1,1}. The error of the weak classifier on the distribution Dt is $e_t=P(H_t(x_i)\neq y_i)=\sum_{i=1}^{N} w_{ti}I(H_t(x_i)\neq y_i)$,Calculate the weight of the weak classifier in the final classifier (if the classifier weight is represented by α) , $\alpha_t=\frac{1}{2}\ln(\frac{1-e_t}{e_t})$ and Update weight distribution of training samples D_{i+1} , $D_{t+1}=\frac{D_t(i)\exp(-\alpha_i\ y_i H_t(x_t))}{z_t}$ (Where Zt is a normalization constant), $Z_t=2\sqrt{e_t(1-e_t)}$,Finally, according to the role of the weak classifier sign, the strong classifier is: $H_{final}=sign(f(x))=sign[\sum_{t=1}^{T}\alpha_t H_t(x)]$.

Because telecommunication customer data is relatively large, we need to focus on the customer's classification and processing and focus on detection and supervision of the customer groups that have a high probability of fraud. The specific method is based on the customer's [2] history call bills and other information to establish, we calculate the weight of customer proportions from the data provided by the existing data and communication operators and based on the possibility of fraudulent conduct at the corresponding point in the data. This is formula: Loyalty = (15%* Month Call Duration + 5% Monthly SMS Total +15% Monthly User Flow + 30% Monthly Total Consumption + 35% Total Card Opening Time) * 100.

The setting of this pre-processing is mainly to separate the customer's rating, so that we can ease the burden on the server in the detection of customers. When the score reaches 80 points, you can default to this user basically does not happen fraud! Based on the above test data, we can focus on testing the score below 60.

TABLE 2. Fraud and loyalty

Loyalty score	Customer level	Possible fraud
>=80	A	1%
>=60	B	10%
<=60	C	50%

SUMMARY

The highlight of this study is to propose combining the immune algorithm with the adaboost algorithm and propose a new detection method that includes but is not limited to the telecommunications industry. Specific good flexibility and convenience of the system, our weak classifier can be upgraded or replaced depending on the situation. In terms of real-time detection, we can obtain the user's communication behavior in real time, rely on our detection system to mine fraudulent users, and provide them to telecommunications companies the system has a dynamic update function that can prevent the occurrence of new fraud mechanisms in real time. Different users can be classified in preprocessing and set different thresholds.

REFERENCES

1. Yongdeng, husun “Fault dictionary technology measurement point selection based on artificial immune algorithm[J]”. Control and decision.2017,32(05): pp.925-929.
2. Feizhang, yiwenliang, hongbindong, “Multi-granularity immune method for telecommunication fraud alert[j]”. Journal of Harbin Engineering University.2006,7: p p.223-227.
3. Cuiqiongyang, hongjiang, xiaoleiyu and yanzhou. “Research on clustering of telecommunication fraud[j] Computer and Digital Engineering”.2010,6(248): pp.99-103.
4. Peter Burge and John Shaw-Talor, an unsupervised neural network approach to profiling the behavior of mobile phone users for use in fraud detection. Journal of Parallel and Distributed Computing, vol.61, July.2001, PP.915-925, doi10.1006/jpdc.2000.1720.
5. .Murynets, M. Zabarankin, R. P. Jover and A. Panagia, "Analysis and detection of SIMbox fraud in mobility networks," IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, 2014, pp. 1519-1526, doi: 10.1109/INFOCOM.2014.6848087.
6. .Lingfenglin, fangyuhu and peikangwang. “Anti-fraud technology in mobile communication billing [J]”. Telecommunication technology.2000 (01): pp.79-83.
7. Strengthening Information Protection and Payment Security Preventing Telecommunications Network Fraud Basic Knowledge of 13 Questions [J]. Wuhan Finance.2017 (02):89.
8. Xiangli. “Application and Research of Boosting Classification Algorithm [D]”. Lanzhou Jiao tong University, 2012.
9. nizhang, yangfanchen, zhijunwang, zhengli and yetao. “Effective and scalable telecommunications fraud management system design and implementation [J].” Information and Communication Technology.2015,9(06): pp.50-56+72
10. Jinminggan, liwan and jiemingwu. “Spam filtering model based on hierarchical collaboration [J]”. Science Bulletin.2012, 28 (10): pp.116-118.
11. Xiangmingliu. “Telecomer fraud prediction system research and application[D]”. Chongqing University.2005.
12. Xueqinmou, mingjiexu and qinghuali. “The new development of mobile communications fraud [J].” World Telecom.2001 (06): pp.20-23.
13. Informationon:https://www.cnblogs.com/pinard/p/6133937.html?utm_source=tuicool&utm_medium=referral.
14. Information on: https://blog.csdn.net/guyuealian/article/details/70995333.