Supervision and Investigation of Internet Fraud Crimes

Lei Zhang
(Corresponding author: Lei Zhang)
Railway Police College, Zhengzhou, Henan 450053, China
(Email: zhangleirpc@126.com)
(Received May 31, 2018; revised and accepted Aug. 10, 2018)

Abstract

The popularity of computers and the Internet in recent years facilitates people's life, but the Internet also provides a convenient crime platform for illegal elements. In order to detect and prevent Internet fraud accurately, Association Rules Mining was performed on the information in the samples of some Internet fraud cases using Apriori algorithm in this study. The attribute fields used for searching association rules are location of crime, scope of crime, time of crime, number of cases and degree of loss. Finally, it was found that the inherent rule is that Internet fraud in the suburban community will cause a higher loss. Moreover, several measures were put forward to prevent crimes.

Keywords: Apriori Algorithm; Data Mining; Internet Fraud; Supervision and Investigation

1 Introduction

With the maturity of Internet technology and the popularity of Internet terminals such as computers and mobile phones, the network has rapidly been integrated into our daily life [4]. The Internet has brought us a lot of convenience. Moreover, all aspects of the modern society cannot be separated from the Internet, for example, banking, traffic management and criminal investigation. However, in addition to bringing convenience to people's lives, it also provides a new space for criminals to commit crimes. As a virtual society, the network has many social characteristics; however, the information of users involved in network is unreal because of the virtual property of the network [16]. Criminals commit crimes, such as Internet fraud, by taking advantage of the virtual property of the network. Untrue information makes it very difficult to solve cases.

Many scholars have studied this aspect. Vlasselaer [15] proposed a method for detecting fraudulent credit card transactions on Internet stores. The method combined the characteristics of RFM based on customer transaction and the derived characteristics with the network characteristics of credit card holder and merchants to figure out the time-dependent suspicious score of each network object. The model combining two kinds of characteristics has good effect. Wu *et al.* [17] detected network identity fraud events on Facebook by analyzing the browsing behaviors of SNS users and found that the method could achieve a reasonable detection accuracy within a given observation time. Shen [10] carried out a deception detection experiment with three-stimulus guilty knowledge test paradigm under the simulated network fraud condition and put forward a multi-domain electroencephalography (EEG) signal processing method. The optimal subset was obtained using a genetic algorithm, and the linear discriminant analysis was used for classification. The results showed that the method was effective in detecting fraud under the simulated network fraud condition. In order to detect and prevent Internet fraud more accurately, this study performed association rule mining on the information in some Internet fraud case samples using the Apriori algorithm, and put forward the corresponding suggestions according to the rules.

2 Internet Fraud

2.1 Definition of Internet Fraud

With the rapid development of Internet technology, people's life has become more convenient, and has depended on the Internet more and more. Moreover, because of the non-face-to-face communication and anonymity on the Internet, it provides a hotbed for the new network crimes. Internet crime has emerged with the birth of the Internet. Although it emerged not too long ago, its harm to society is no less than that of traditional crimes. Internet fraud [7, 13] is a kind of Internet crimes. The crime of fraud is constantly changing with the development of the times and society, and the emergence of the network is a convenient tool for the illegal elements.

In general, Internet fraud is defined as the act of making up or concealing the facts through the network or

communication tools which rely on the Internet to achieve the illegal possession of a large amount of others' property, the word "Internet" has two connotations, Internet in broad sense and narrow sense. At present, there is no specialized classification of network fraud in China's criminal law [12]. The main reason is the wide range of network fraud. The content of fraud includes the categories of frauds described in the current criminal law, but it is difficult to be classified as a single category because of the use of the Internet [5].

2.2 Characteristics of Internet Fraud

The popularity of the Internet and computers has a profound impact on people's lives and also makes the network economy develop rapidly. The criminals also take the opportunity to implement fraudulent activities taking the advantage of network vulnerability. Based on the analysis of Internet fraud cases, the characteristics of Internet fraud are concluded as follows:

1) A wide range of crime objects [9]:
Traditional fraud is usually face to face; therefore, the victim of traditional fraud is often fixed, people in an age group or people who engage in some kind of job. However, after the emergence of network, fraudsters release false information regardless of cost and object via the Internet or the group chat function of communication tools such as QQ. It shows that the object of Internet fraud has changed from the single group to all netizens, i.e., the target of crime becomes more extensive.

2) Diversity of ways of crime:
As the Internet has been applied in different kinds of professional fields, people on the Internet engage in different kinds of works. Considering the wide range of crime object mentioned above, fraudsters may try to improve success rate by adopting different fraud means for different objects. Therefore, network fraud also has the feature of diversity. Moreover, the progress of network communication technology provides more network fraud tools for Internet fraud.

3) Professional crimes [2]:
In the early stage of the emergence of Internet fault, netizens were easy to be deceived by simple false information as they have not been familiar with it. However, with the increase of knowledge and the improvement of security awareness, the ability of netizens in identifying false information has also been improved. Fraudsters are more specialized in fabricating false information. Modern defrauders often have two skills: familiarity with the process of network business transactions and proficiency at computer programs.

4) Lowering trend of criminal age [8]:
The amount of information on the network is huge, but there is a lot of negative information. Young people who are shallow and immature and vulnerable to the temptation of bad information are more likely to commit crimes in pursuit of enjoyment. In recent years, the cases of Internet frauds in different regions show that the proportion of youngsters is increasing.

5) The concealment and continuity of fraud behaviors:
In Internet fraud, fraudsters only needs to send false information to victim, and the information sent through the network is not easy to be found. Therefore, the characteristics of fraudsters are difficult to know, and the backwardness of the network identity verification system makes virtual identity a protective umbrella of criminals. As for continuity, because the low implementation cost of Internet frau, it makes criminals feel less guilty.

6) Difficult evidence collection in detection process [14]:
Evidence collection is easy for traditional fraud because of the fixed objects and locations. However, for the network fraud, its concealment makes the crime process not easy to be found. Secondly, the dissemination and storage of its information is in the form of digital, which is very easy to be modified, hidden and deleted, and traces can be eliminated timely before exposure. In addition, because of the right to privacy of citizens, identity authentication on the network in the initial stage is very ambiguous, and most of the network identity information is not true. Once the criminals commit fraud with such an identity, the public security organs are difficult to collect evidences.

3 Apriori Algorithm

Apriori algorithm [3] put forward by Agrawal is used for searching the project relationship in database. It searches project relationship step by step and constantly repeats procedures of connection and pruning. Connection [6] means obtaining candidate set B_k by connecting frequent set M_k and M_{k-1}. Suppose $M_{k-1} = (m_1, m_2, \cdots, m_n)$, m_i, m_j, $(1 \le i \le n, 1 \le j \le n)$ are two elements of M_{k-1}, and $m_t[x]$ as the X-th item of m_t. Two elements in M_k and M_{k-1} can be connected, and moreover if m_i and m_j in M_{k-1} can be connected and $(m_i[m] = m_j[1]) \cap (m_i[2] = m_j[2]) \cap \cdots \cap (m_i[k-2] = m_j[k-2]) \cap (m_i[k-1] = m_j[k-1])$ when and only when the first $k-2$ element of the two elements are the same, then the result set generated after connection is $m_i[m]m_i[2] \cdots m_i[k-1]m_j[k-1]$. Pruning [11] means determining the count of each element in candidate set B_k through scanning database and eliminating elements with infrequent count. The remaining elements constitute a new frequent set M_k. The flow of Apriori algorithm [1] is shown in Figure 1.

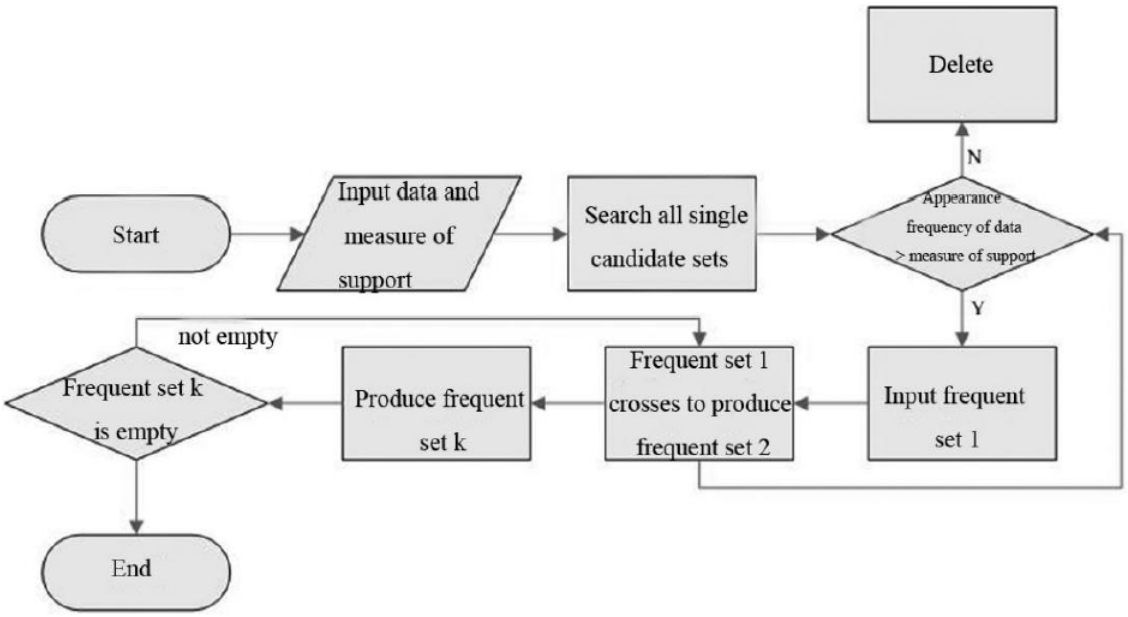

Figure 1: The flow of Apriori algorithm

4 Instance Analysis

With the continuous progress of information technology, the information processing of cases has been advanced. In the public security system, the functions of data processing have been relatively perfect, including recording, storage, classification-type query and static statistics of data, which greatly promote office efficiency, but it is difficult to make linking analysis of information between cases and explore the potential laws of cases. In order to deal with this problem and improve the detection rate of cases, Apriori algorithm was applied to search for information rules in cases.

4.1 Selection of Field

The information of network fraud cases in the database of the public security system was statistically analyzed, and the fields which appeared the most frequent were selected as the attribute values. After statistical analysis, crime scene, scope of crime, time of crime and number of criminals and degree of loss were selected as fields.

4.2 Data Preparation

The essential connection between different crimes and the law of committing crimes can be obtained through analyzing the information of cases. The information of some network fraud cases was selected from the case database of the public security department as the sample data according to the selected fields above. The content of the fields is shown in Table 1.

4.3 Establishment of Database

According to Table 1, the network fraud information was converted to the transaction database D. Crime scene was set as A; scope of crime was set as B; time of crime was set as C; scope of crime was set as D, and degree of loss was set as E. The subclassification was distinguished by number as shown in Table 2.

4.4 Data Mining

Association analysis was performed on the samples using Apriori algorithm, and the flow of the algorithm has been shown in the last section. In this study, the minimum support was set as 30%, the minimum confidence was set as 75%, and the measure of support was 3. Then the detailed data mining is as follows.

1) Candidate set B1 was obtained after performing scanning and statistics on Table 2 as shown in Figure 2.

2) The statistics of different item sets in Figure 2 were compared with measure of support; item sets whose statistics was smaller than 3 were deleted, and finally frequent set M1 was obtained as shown in Table 3.

   It could be found from Table 3 that C3, C4 and D1 were deleted, indicating that time of crime was at 16:00 22:00 and 22:00 4:00, and one criminal had a low association with the other information of the cases.

3) Item sets in frequent set M1 were combined in pairs, and then candidate set B2 was obtained as shown in Figure 3.

Table 1: The information of some network fraud cases

No.	Crime scene	Scope of crime	Time of crime	Number of criminals	Degree of loss
1	Residents community	Suburb	4:00-10:00	3	High
2	Bank	Suburb	22:00-4:00	2	Low
3	Bank	Urban area	16:00-22:00	1	High
4	Residents community	Suburb	10:00-16:00	2	High
5	Bank	Urban area	10:00-16:00	3	Low
6	Residents community	Urban area	10:00-16:00	3	Low
7	Residents community	Suburb	16:00-22:00	2	High
8	Residents community	Suburb	22:00-4:00	1	Low
9	Bank	Urban area	4:00-10:00	2	High
10	Residents community	Urban area	4:00-10:00	3	High

Table 2: Transaction database

No.	Crime scene	Scope of crime	Time of crime	Number of criminals	Degree of loss
1	A1	B1	C1	D3	E1
2	A2	B1	C4	D2	E2
3	A2	B2	C3	D1	E1
4	A1	B1	C2	D2	E1
5	A2	B2	C2	D3	E2
6	A1	B2	C2	D3	E2
7	A1	B1	C3	D2	E1
8	A1	B1	C4	D1	E2
9	A2	B2	C1	D2	E1
10	A1	B2	C1	D3	E1

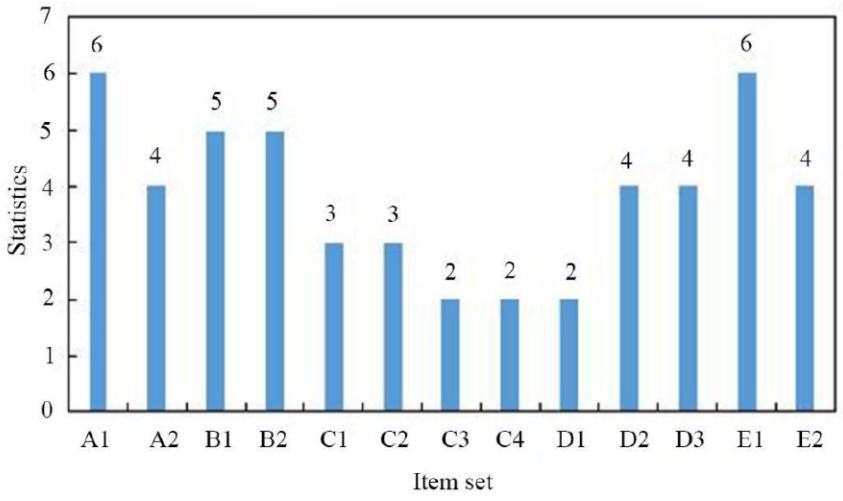

Figure 2: Candidate set B1

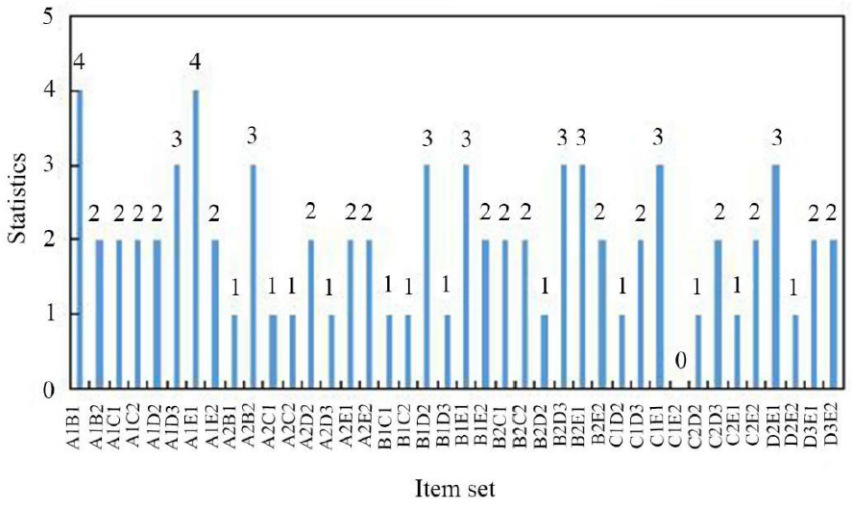

Figure 3: Candidate set B2

4) The statistics of the item sets in Figure 3 were compared with the measure of support; item sets whose statistics was smaller than 3 were deleted, and finally frequent set M2 was obtained, as shown in Table 4.

5) The item sets in frequent sets, M2 and M1, were combined in pairs, and then candidate set B3 was obtained as shown in Figure 4.

6) The statistics of the item sets in Figure 4 were compared with the measure of support; item sets whose statistics was smaller than 3 were deleted, and finally frequent set M3 was obtained, as shown in Table 5.

Frequent set M3 was the final output result, i.e., the potential law between the network fraud cases mined from the sample database, which was numbered A1B1E1; after the conversion according to Table 1, the law was that the loss of victims was high when network fraud crimes happened in residence community and suburb. According to this law, the public security department can pay more attention to the related information in the surrounding of residences communities in suburb, and take corresponding measures to prevent crimes, for example, through popularizing fraud prevention knowledge, reminding residents to use more complex, strong passwords in key settings to protect personal property, warning them not to disclose the password of bank cards, especially to the stranger, reminding residents who like online shopping not to do online transactions in public places, and not clicking on unidentified links sent by communication tools such as QQ or WeChat.

Table 3: Frequent set M1

Item set	Statistics	Item set	Statistics
A1	6	C2	3
A2	4	D2	4
B1	5	D3	4
B2	5	E1	6
C1	3	E2	4

Table 4: Frequent set M2

Item set	Statistics	Item set	Statistics
A1B1	4	B1E1	3
A1D3	3	B2D3	3
A1E1	4	B2E1	3
A2B2	3	C1E1	3
B1D2	3	D2E1	3

Table 5: Frequent set M3

Item set	Statistics
A1B1E1	3

5 Conclusion

This paper first introduced the definition of network fraud, summarized several characteristics of network

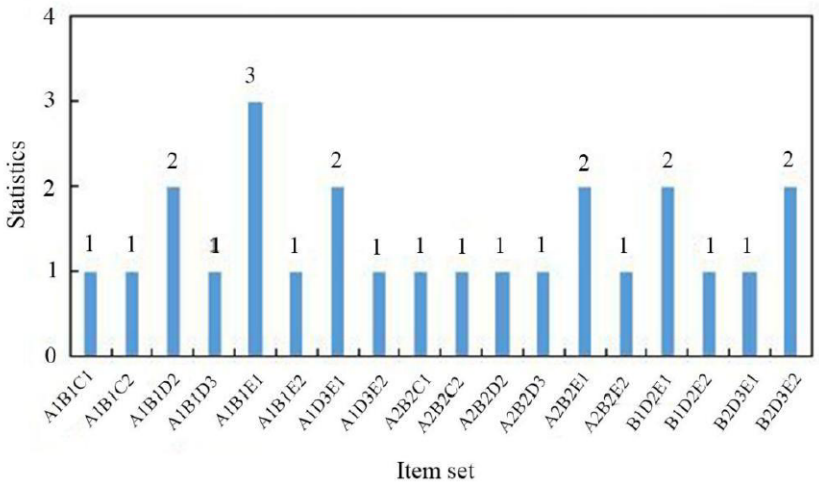

Figure 4: Candidate set B3

fraud, and performed association rules mining on the information of some network fraud case samples. The data attribute fields used for searching association rules are crime scene, scope of crime, time of crime, number of criminals and degree of loss. The inherent law found was that network fraud in residential communities in suburb will cause a higher loss. This rule provides a reference for preventing crimes and finding a way to solve cases, which can help the police arrange the police strength scientifically and rationally. Finally, several measures were put forward for preventing fraud.

References

[1] O. S. Adebayo, N. Abdulaziz, "Android malware classification using static code analysis and Apriori algorithm improved with particle swarm optimization," in *4th World Congress on Information and Communication Technologies (WICT'14)*, pp. 123–128, 2014.

[2] A. S. Bekirev, M. V. Kuzin, M. V. Kuzin, *et al.*, "Payment card fraud detection using neural network committee and clustering," *Optical Memory & Neural Networks*, vol. 24, no. 3, pp. 193–200, 2015.

[3] A. Bhandari, A. Gupta, D. Das, "Improvised apriori algorithm using frequent pattern tree for real time applications in data mining," *Procedia Computer Science*, vol. 46, pp. 644–651, 2015.

[4] K. M. Fanning, K. O. Cogger, "Neural network detection of management fraud using published financial data," *Intelligent Systems in Accounting Finance & Management*, vol. 7, no. 1, pp. 21–41, 2015.

[5] B. Hannibal, H. Ono, B. Hannibal, *et al.*, "Relationships of collapse: Network brokerage, opportunism and fraud in financial markets," *International Journal of Social Economics*, vol. 44, no. 12, pp. 2097–2111, 2017.

[6] A. Khalili, A. Sami, "SysDetect: A systematic approach to critical state determination for industrial intrusion detection systems using apriori algorithm," *Journal of Process Control*, vol. 32, no. 11, pp. 154–160, 2015.

[7] M. Nandhini, B. B. Das., "An assessment and methodology for fraud detection in online social network," in *IEEE International Conference on Science Technology Engineering and Management*, pp. 104–108, 2016.

[8] R. Puri, M. P. Hammer, D. Grottwassink, *Method and System for Defending a Mobile Network from a Fraud*, Patent WO/2016/148685, Sept. 22, 2016.

[9] T. Razooqi, P. Khurana, K. Raahemifar, A. Abhari, "Credit card fraud detection using fuzzy logic and neural network," in *Proceedings of the 19th Communications & Networking Symposium*, 2016.

[10] J. Shen, J. Liang, X. Liu, "P300-based deception detection in simulated network fraud condition," *Electronics Letters*, vol. 52, no. 13, pp. 1136–1138, 2016.

[11] A. K. Singh, A. Kumar, A. K. Maurya, "An empirical analysis and comparison of apriori and FP - growth algorithm for frequent pattern mining," in *International Conference on Advanced Communication Control and Computing Technologies*, pp. 1599–1602, 2015.

[12] J. R. Sun, M. L. Shih, M. S. Hwang, "Cases study and analysis of the court judgement of cybercrimes in Taiwan," *International Journal of Law, Crime and Justice*, vol. 43, no. 4, pp. 412–423, 2015.

[13] J. R. Sun, M. L. Shih, and M. S. Hwang, "A survey of digital evidences forensic and cybercrime investigation procedure," *International Journal of Network Security*, vol. 17, no. 5, pp. 497–509, 2015.

[14] M. Uğurlu, S. Sevim, "Artificial neural network methodology in fraud risk prediction on financial statements; An emprical study in banking sector," *Journal of Business Research Turk*, vol. 7, no. 1, pp. 60–89, 2015.

[15] V. V. Vlasselaer, C. Bravo, O. Caelen, *et al.*, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decision Support Systems*, vol. 75, pp. 38–48, 2015.

[16] V. V. Vlasselaer, T. Eliassi-Rad, L. Akoglu, *et al.*, "GOTCHA! Network-based fraud detection for social security fraud," *Management Science*, vol. 63, no. 9, pp. 2773–3145, 2017.

[17] S. H. Wu, M. J. Chou, C. H. Tseng, *et al.*, "Detecting in situ identity fraud on social network services: A case study on facebook," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2432–2443, 2017.

Biography

Lei Zhang, born in November 1980, is now working in the department of investigation of Railway Police College, Henan, China. She has gained a master's degre from People's Public Security University of China, Beijing, China. She is interested in science of investigation and psychology.