ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)

https://doi.org/10.17559/TV-20190110163709
Original scientific paper

A Method of Mining Key Accounts from Internet Pyramid Selling Data

Jianying XIONG

Abstract: Internet pyramid selling causes great harm and has difficulty in obtaining evidence. As most of internet pyramid selling often uses virtual coins to complete cash settlement, criminals can exchange virtual coins by controlling visual accounts to transfer funds illegally. The purpose of the work is to mine the key accounts and build evidence chain of personnel and funds. In order to deal with a large number of transaction data, we characterize the virtual currency trading data, which is divided into seven transaction characteristics. The key accounts are these outliers which have too prominent trading behaviour. So we use positive samples to train One Class Support Vector Machine (OCSVM) classification to find trading behaviours' boundary and detect outliers in classification. Then, the correlation between member's information and pyramid selling hierarchical relationship can also be obtained by further linking the trading behaviour of key accounts. Finally, we can screen out the virtual accounts directly controlled by the pyramid selling organization, analyse the flow of funds, investigate and obtain evidence on amount of money involved in pyramid selling. The experimental results show that the classification model cannot only establish the normal model for account transfer behaviour, but also effectively identify abnormal transfer behaviour, thus improving efficiency of investigation and evidence collection for economic investigation departments.

Keywords: abnormal transaction; electronic forensics; Internet Pyramid Selling; OCSVM classification

1 INTRODUCTION

The State Council has prohibited pyramid selling in China since 1998. In recent years, various forms of Ponzi scheme have emerged, the pyramid selling has increased and further developed into Internet pyramid selling [1]. Internet pyramid selling is more harmful than traditional pyramid marketing, and it is more difficult to govern. According to the reliable data at present, 60% to 70% of pyramid selling cases are operated through the Internet [2]. In China, fraud of Internet pyramid selling mainly includes the following categories: shopping and rebate, compound interest return service games, virtual currency, financial mutual assistance, payment and financial management, telecommunications business. Population recommendation makes business behaviour spread more easily [3]. Aiming at inducing users to join the platform and develop deep levels, they always make use of "promotion division", "developing deep levels ", "purchasing financial product", "platform shopping all return", and so on. The Internet pyramid activities lead to significant economic losses to the masses of people, seriously distort the values of the participants, and seriously damage the system of social integrity [4]. Pyramid selling has a title of "economic cult", and Internet accelerates the spread of pyramid selling crime at all levels of society, which bring difficulties to investigate. A large number of extremely critical data are stored in the database of the internet pyramid marketing platform, including all involved users, the amount of involved money, the profit and so on, which can reveal the way of distribution of interests and the organization structure. Investigators need to find and obtain relevant data, obtain financial clues. However, the information of cash flow is not clear due to the difficulty of network virtual identity verification. It is difficult to accurately identify the funds involved. In this paper, we use outlier analysis and evidence relevance feedback to mine user's fund transfer behaviour, obtain key account information, make an authentication by multiple source of evidence.

2 REALTED WORKS

The means of internet pyramid selling are always changed and complicated, which bring many difficulties to investigation. The forensics has become a key task to govern network pyramid marketing. For these cases, the evidence system is more inclined to electronic evidence. Internet did not change the nature of pyramid selling; it has changed the focus of evidence [5]. Thus, analyzing electronic evidence in internet pyramid selling becomes very important and crucial. Many researchers analyzed the electronic evidence analysis process of internet pyramid selling, but there are few data forensics models [6, 7].

Database is one of the most important electronic data; it involved members' information, the total profit, major suspects' information and some other useful information. The information can reveal the interests distribution and organization [8]. Moreover, there is a large amount of extremely critical data stored in the database; the suspects often disguise the key fields and the data of the website. If the investigators use conventional methods, it is difficult to find and access relevant data case at the first time [9]. The research on database forensics mainly includes the following aspects: 1) Database forensics technology focusing on the database recovery, operation log forensics process, such as Al-Dhaqm defines a model from the identification, collection, storage, analysis process and record database of digital evidence, reconstruction [10]; Fasan & Fruhwirt restored data approach to obtain evidence through a database query log reconstruction [11, 12]. 2) For more complicated and huge data, data mining model is helpful for assisting investigators to find out the law of circulation. For example, Zawoad believes that one of the challenges for forensics is to collect and analyze evidence from a large data set, define the conceptual model of large data forensics, and use examples to show that large data forensics can provide evidence for the determination of criminal events [13]. Edem & Quick believed that data mining can generate insight from a large number of data, including detecting deceptive criminal identity, identifying criminal gangs and various illegal activities and so on, and can also be applied to the field of criminal forensics [14, 15]. 3) The data mining model forensic application research, such as outliners are used to detect abnormal behaviour analysis [16], online securities trading, financial network laundering organizations found, banking institutions customer accounts of suspicious money laundering activity recognition analysis [17, 18], health abnormal findings [19], etc.

Detection and forensics of internet pyramid selling need to collect the people involved, evidence related to the hierarchy, meanwhile find out the network structure, core personnel in the organization. And money flows is also the important evidence. However, it is difficult to verify the virtual identity on the internet. The database records many users' transactions. Pay and reward allocation is directly related to cash flow or indirectly verified the behaviour. The chain of information construction evidence in the database is also complex. In the internet pyramid marketing, the process of money circulation is the core of investigation and forensics. As long as the money flows is grasped, the operation framework of the whole pyramid selling can be grasped. From the existing research, there is a lack of data analysis model based on the business process and electronic data characteristics. In reality, the membership fee is not completely according to the online business process, the amount of money involved cannot be fully recorded in database. So that even if the investigator obtained the database of the platform, the amount of money involved in the case cannot be fully counted. However, whether online charging or offline transfer, when pyramid organizers receive money, they will exchange the corresponding virtual currency to users. If there is no record of the virtual currency allocated from the related table, then the evidence can be found between the members in the virtual currency transfer records.

This research adopts data mining method to discover the abnormal trader information, and solve abnormal sample unbalance classification problem. OCSVM classification method is used to establish the normal model of the user transfer behaviour, and classifying whether the test data belongs to the normal class. It realizes the recognition of abnormal transfer behaviour. The associated personnel level model finds out the virtual account controlled by the organizer, forms the electronic data evidence chain, and combines the actual evidence to carry out the correlation analysis to determine the funds and transferring way of the funds. There are three main contributions of this paper. 1) Discover the crimes evidence from a large number of records from pyramid selling database, and use manual analysis to find out suspicious accounts. 2) According to internet pyramid selling crime mode and evidence collection mode, we build electronic data evidence chain by funds flow. 3) Verify the identity of the virtual account through multidimensional data correlation feedback, and improve the statistics accuracy of amount involved. The content of the paper is arranged as follows: analysis of the development trend and harmfulness of internet pyramid selling in China; the research works of internet pyramid selling forensics and database forensics; the scenario of data mining evidence collection model, and experimental analysis to verify the effectiveness of the model.

3 EVIDENCE ANALYSIS FRAMEWORK

3.1 Scenario

The study uses a scheme of financial-mutual-aid-type internet pyramid selling as a scenario, similar to the "multi-profit splitting disk", which is a kind of transaction based on virtual equity provided to users. User can get sustainable income every day after purchasing a fixed investment. But if users want to get more revenue, they must become the core to get more commissions. It is similar to the superior level concept in pyramid selling, which is to develop deep level.

Mutual-information pyramid schemes essentially use "fund pools" to create interest-broadcasts. However, these fraud organizations often claim that their behaviour is to obtain returns by providing financial assistance to others, through the concept of "charity help". Users who obtain revenue can continue to invest in a "mutual assistance" fund pool to ensure the normal operation of the interactive system. However, this behaviour belongs to the financial fraud with clear pyramid selling nature. According to the security mobile intelligence analysis team, financial mutual aid is generally required to spend a certain amount of money on the purchase of "registration activation code" or "membership "; investment frequency. The investment amount is directly linked to their return. If users want to get more revenue, they must first become investment agents. When members recharge investment, the actual legal currency will be converted into virtual coins. The system will give rebates to the recommender and the agent.

3.2 Forensics Elements

Users pay for the membership fees is not entirely in accordance with the online operation flow in internet pyramid selling. Many members pay the fees offline and the records have not been saved in the database, which makes the money involved is unclear. As shown in Fig. 1, common users need to pay agents for virtual currency (as Processes 1_1 and 1_2). Agents need to pay company for virtual currency (as Processes 2_1 and 2_2). The income of company is mainly consisting of top-up online by agents, and transfers offline (as Processes 3_1-3_4). The company's debits consist of the withdrawal from users and agents (as Processes 4). The characteristics of pyramid marketing also show that the members can get corresponding virtual coins, whether it is the online recharge or offline remittance. Under the offline remittance system, the system cannot form a closed business cycle, and virtual coins cannot be distributed by the platform account. Furthermore, users receive the virtual coins after offline remittances from some key accounts. In this way, as long as mining the virtual coins transferring data between members, the account controlled by the platform can be found, and the amount of money involved can be obtained through the transfer record of the virtual coins.

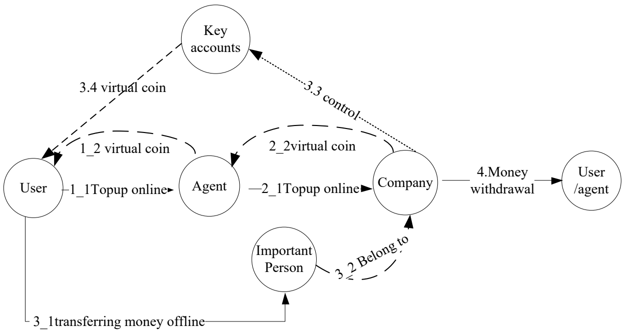

Figure 1 Funds flow process

The electronic evidence analysis of the funds involved includes:

- Obtain all online recharge *STon* received by the company account.
- Obtain the total amount of all offline remittance received by the company *SToff*.
- Obtain the withdraw cash of common member and ordinary agents as *Wuser*, *Wage*.

The amount of money involved is *Amoney*, which can be defined as:

$$Amoney=STon+SToff-Wage-Wuser \quad (1)$$

STon,*Wage*, and *Wuser* is completed online, and they are fully recorded in the website database. The amount of evidence involved in the case is mainly accounting for the *SToff* of the member's offline transfer. In the actual case of pyramid cases, the persons involved in the case are all over the country. We cannot check the details of the transfer money. However, from the process (3_1, 3_2,3_3., 3_4), the amount of the offline transfer can be calculated by mining database to calculate the virtual coins issued by the controlled accounts.

3.2 Framework of Data Mining

It is difficult to verify the virtual identity on the network and it is not possible to confirm the status of criminal gang and other members. Additionally, funds flow is also an important evidence clue for crimes. The database records much behaviour such as user transactions, payments, and award distribution that are directly related or indirectly corroborated with the circulation of funds. If the user's virtual account has an obvious difference in transaction behaviour from other users, it can be proved to a large extent that this is a key user's virtual account, which helps to narrow down the investigation scope of the involved persons, control the involved personnel. Then by mining the transaction, profit and funds flow information to constitute the electronic data evidence chain. The framework works as Fig. 2.

The data analysis process can be mainly divided into the following sections:

Data source analysis: It is necessary to clarify the information content which needs to be included in the target data set to be mined, and extract useful data from the illegal website.

Data pre-processing: It is necessary to analyze the database, design the corresponding data retrieval rules for query, table connection, summary and other operations, and export the query results. The exported data is filtered according to the characteristics of the data for further analysis.

Personnel Information mining: For internet pyramid selling organizations, with the aid of database storage and management of the members and upper and lower levels of the criminal group, the analysis and processing of database data should be carried out to obtain information from a given member and the hierarchical relationship in the organization. The personnel information in the top layer of the pyramid marketing structure needs to be extracted as the important clue of the investigation. Furthermore, it is necessary to combine the main creation and the backbone members of the organization obtained by investigation department to retrieve the personal attribute information in the database. In addition, using two types of the staff attribute information, including telephone, WeChat, QQ, ID card, bank card number and all member information to carry out association analysis, to obtain more related user information.

Virtual currency transfer data mining: Transfers among all members will be stored in the relevant transaction table. There are certain differences in the behaviour of virtual currency transfer between accounts controlled by organizer and ordinary members, such as transfer frequency, transfer amount, etc. The rules of transfer behaviour can be discovered and anomaly points in transaction data can be identified by trained classifier.

Finally, the abnormal account obtained through data mining is compared with all associated user accounts, and then suspicious key accounts information can be found.

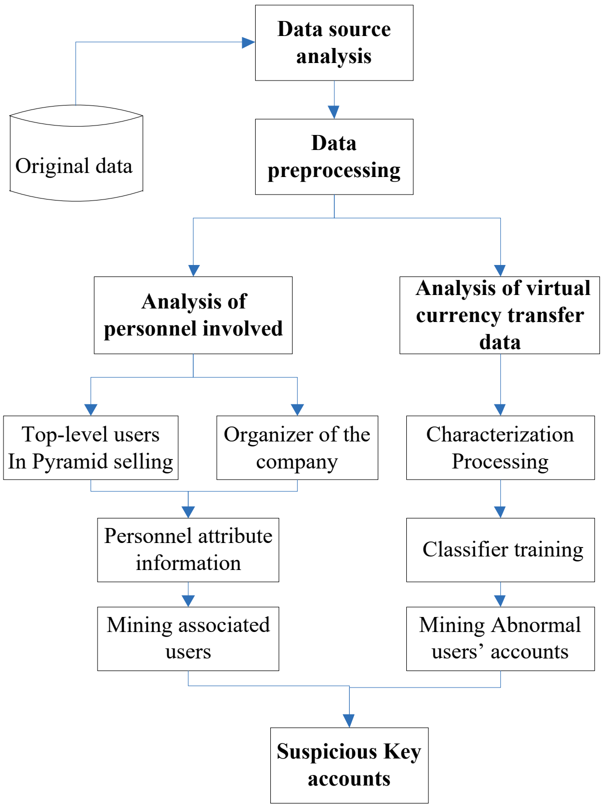
Data source analysis
Original data
Data preprocessing
Analysis of personnel involved
Analysis of virtual currency transfer data
Top-level users In Pyramid selling
Organizer of the company
Characterization Processing
Personnel attribute information
Classifier training
Mining associated users
Mining Abnormal users' accounts
Suspicious Key accounts

Figure 2 Framework of data mining

4 METHOD TO MINING KEY ACCOUNTS

4.1 Mining Associated Users

Investigation of online pyramid marketing needs the reconstruction of its criminal organization structure and the determination of its membership in the entire criminal organization. We need obtain the members and the hierarchical relationship from database files while investigating these cases. The information is also an important link in the evidence chain. Investigation department will also obtain some key members of the organization as suspects for personnel information forensics if the case is registered. The target members' information was screened by data collision comparison, even including the personal identification information of the organization, especially the information of some senior managers of the company.

According to the association collisions of high-level users in pyramid marketing and organizational staff, more suspicious person collections are obtained, as shown in Fig. 3.

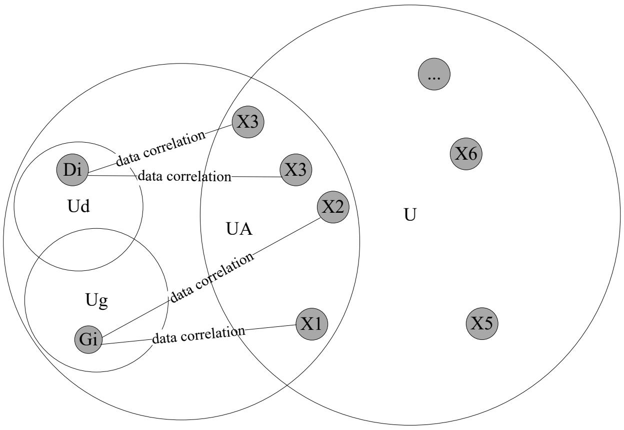

Figure 3 Definition for associated user set

Assume that the top-level users is *Ud*, the company's primary staff is *Ug*, and the associated user set is defined as *UI = UdUUg*.

Get all user-related attributes that can be associated with user informastion in *UI* collections, such as ID number, phone number, bank card number, and collide with all user sets *U* to obtain all associated account set *UA*, which has a correlation with the user set *UI*. So we defined associated account set as *UA =Corr(UI,U).*

4.2 Mining Abnormal Users' Account

Outliers refer to those small-mode data in the data set, which may be due to metrics or execution errors, or they may be the result of intrinsic data variability. According to the size of outlier data, outliers can be divided into outliers and outlier clusters. At present, outlier detection techniques can be roughly classified into: statistical method, neighbor-based method, classification-based method, cluster-based method. There are two basic tasks in outlier mining for account virtual currency transfer transactions.

4.2.1 Rules for Abnormal Transaction Behaviour

There are many kinds of virtual transactions in internet pyramid marketing. In this study, pyramid organizer will transfer virtual coins to the user's account after the user pays money offline, and promises an increment. Users can also exchange the virtual coin with other users. So the accounts controlled by organizer cannot be found directly. But the virtual account transactions of the organizer are different from ordinary users, such as a high frequency to roll out virtual coins, the transaction amount is relatively large, the transaction is relatively dispersed. Some features of the trading behaviors are defined as follows:

Recent frequency RF: The transaction times of the account $user_i$ in the last cycle of *T*, the higher the number of transactions, the more likely to indicate that the account node is a recently active transaction concentration point, and its corresponding virtual account is the core of the transaction network. $Trans_j$ is the transaction record *j*, j_{tim} is the time of record *j*.

$$RF_i=sum(Trans_j \mid j_{tim}\in T \text{ and } Trans_j \in user_i) \tag{2}$$

Recent transfer amount RA: the total amount of virtual coin transferred within a cycle of *T*. A higher value of the *RA* indicated the account is possible centralization point. Mt_j is money of transaction record *j*, j_{tim} is the time of record *j*.

$$RA_i=sum(Mt_j \mid j_{tim}\in T \text{ and } Mt_j \in user_i) \tag{3}$$

Total amount of transferring AT: the total amount of virtual coin transferred generated by the account. A higher value of the *AT* indicated that the account is a virtual coin issuing node, and the corresponding virtual account is more likely a beneficiary of the pyramid marketing. Mt_j is money of transaction record *j*.

$$AT_i=sum(Mt_j \mid Mt_j\in user_i) \tag{4}$$

Account coverage AC: It refers to the number of accounts involved in the trading activities of an account. The virtual coins are often transferred among some acquaintances for common accounts. But abnormal accounts will cover many trading accounts to complete the issuance of virtual coins. So a high value of *AC* may imply the account is abnormal. $Reuer_j$ is receiver of the transaction, repeated receiver should be filtered by dist.

$$AC_i=sum(\text{dist}(Reuer_j) \mid Reuer_j \in user_i) \tag{5}$$

Account repetitiveness AR: In the common accounts' transactions, the transfer behavior is also easier to repeat for it often transferred among a small circle of acquaintances. The abnormal account is mainly for the virtual currency allocation after the membership payment of membership fees and agency membership fees. The payment of the membership fee will not be repeated frequently. So the *AR* value is relatively low. $Reuer_j$ is receiver of the transaction.

$$AR_i=sum(Reuer_j \mid Reuer_j \in user_i)/AC_i \tag{6}$$

Quantity of large transferring NT: Although membership fees are not high, fees of upgrading to agents are relatively high. Virtual coins between transferring common members are generally for withdrawal or reinvestment. The amount is generally not high, and there is very little large transferring behavior. $Trans_j$ is the transaction record *j*, Mt_j is money of transaction record *j*.

$$NT_i=sum(Trans_j \mid Mt_j>TM \text{ and } Trans_j\in user_i) \tag{7}$$

Proportion of large transferring PT: It means the proportion of the quantity of large transferring in the amount of all transactions. Here $Trans_j$ is the transaction record *j*.

$$PT_i= NT_i/ sum(Trans_j \mid Trans_j\in user_i) \tag{8}$$

4.2.2 Design OCSVM Classification Model

The key accounts are hidden in a large amount of transaction data. Artificial empirical judgment is not only inefficient, and the accuracy rate cannot be guaranteed. When the key accounts are relatively small, the positive and negative samples will be extremely unbalanced if using the classification method to detect abnormal accounts. Few abnormal samples may cause the classifier to deviate too much from the number of sample categories, so that the training model has bias. The One-Class SVM (OCSVM) in libsvm is adopted here through using the normal data to train a hyperplane. All testing data is used for classification. If the data is within the hyperplane, it is considered as a normal event. Otherwise, it is considered abnormal. Because the OCSVM is a feature of the unsupervised learning process, training sample precision requirements are greatly reduced, which reduce the human intervention process.

As shown in Fig. 4, the abnormal account identification method can be divided into four steps.

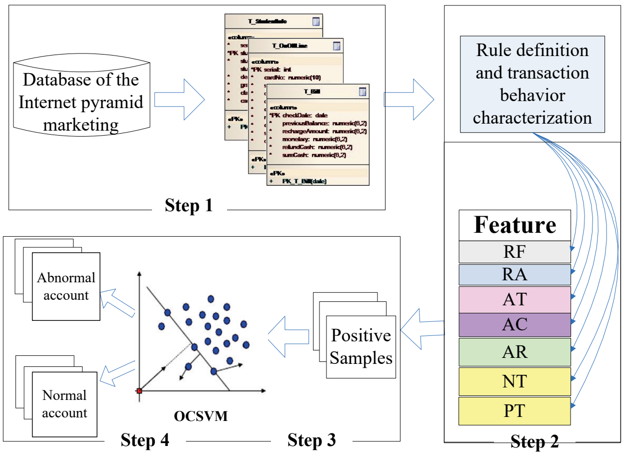

Figure 4 Process of abnormal account identification

Step 1: The study analyzes the table structure of internet pyramid selling website database, extracts and defines 7 dimension trading behavior rules, and takes it as the standard quantification of account.

SD (X_1, X_2,..., X_n) is a given data set, where X_i is a training tuple. Each account is quantized into a 7 dimension vector, and X_i is a feature array. For training tuples, there are 7 dimensional attributes in this case, $X_i=\{x_{i1}, x_{i2},..., x_{i7}\}$. There are 2 kinds of data (abnormal and normal) in data classification. It is classified as {1, −1}. The $n\times p$ data matrix is a normal data set as follows, which has n samples and p dimension characteristics.

$$\begin{Bmatrix} x_{11} & \cdots & x_{1k} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{j1} & \cdots & x_{jk} & \cdots & x_{jp} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{nk} & \cdots & x_{np} \end{Bmatrix}$$

Step 3: Extract a part of account from the massive data, and label the categories manually. Two-thirds is used as a training set of One-class support vector machine, and it only contains the regular samples. One-third is used as test set. All sample sets are quantized as a 7 dimension matrix. The training set is used to train OCSVM classifer.

Step 4: The trained OCSVM classifier can be used to judge the test set, which can be used to identify and predict whether the account is abnormal or not. The kernel function of OCSVM is usually Gaussian kernel, which can map the data sample to high latitude space and make it linear separable. We use LIBSVM package for SVM pattern recognition and regression, which is designed by Professor Lin Chih-Jen of Taiwan University. The SVM type is OC-SVC, and other parameters are the default parameters of SVM in libsvm. The trainer will detect the classification boundary according to the training samples and parameters, and predict according to the signed distance from each test point to the hyperplane. The positive sample is normal, the negative sample is abnormal.

4.3 Association Evidence Chain of User Account

The database forensics of internet pyramid selling website requires multiple levels of evidence collection. It depends on the retrieval, analysis and application of large amounts of data to build a complete evidence chain. The process of forensics is also a behavior of constructing an association between virtual and real world. In order to check the virtual identity, we can use correlation analysis to feedback data forensics results. Some personal identification information reflects the association, such as identity cards, telephones, QQ, Wechat, Alipay, bank cards, bank's funds circulation, communication data, inquiry information, and etc. A complete chain of evidence is used to formulate case facts. The process model of relevance analysis feedback is as follows.

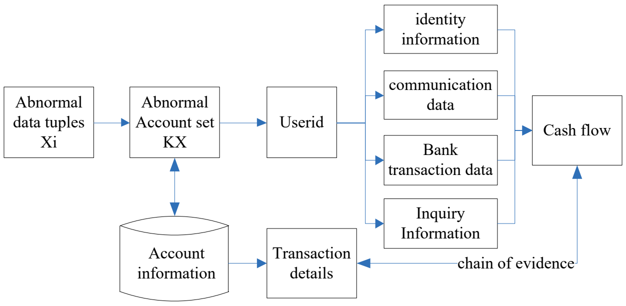

Figure 5 Illegal evidence chain of suspicious accounts

For example, if R_1 is evidence of the circulation of cash offline, we can find abnormal account through the referred "userid", and obtain the associated user's offline circulation of cash records. R_2 is the transfer record of the electronic virtual coin online of the userid. R_1 and R_2 are linked by time and transfer amount, which can form a chain of evidence as follow.

$$R_1 : KX \xrightarrow{Userid} InfUser \xrightarrow{Bankaccount} InfoCash$$

$$R_2 : KX \xrightarrow{Userid} InfTra$$

$$R : R_1 \xleftrightarrow{Ref} R_2$$

5 EXPERIMENT ANALYSIS

5.1 Test Data

The experimental data are derived from database files of the Internet Pyramid Selling platform, which is submitted for inspection and forensics by public security department. It involves all user information, virtual coin trading, cash withdrawal, partial online recharge and other data information. As many users transfer funds to criminals offline or online, the system will control some virtual accounts to pay the corresponding virtual coins to these users. In this study, virtual coin transactions and account entry transactions were extracted for mining. The key users and their corresponding transaction information can be mined, and it will form an evidence chain with real bank cash record.

The transaction data is from October 27, 2016 to June 21, 2017. A total of 2195,801 transaction data were involved. After preprocessing, the missing data was removed. The number of accounts involved in the transfer was 13053 as the training data, and 6400 user data as the test data.The main fields of transaction records of transfer users are as in Tab. 1.

Table 1 Description of each field

No	Field name	Meaning
1	USERID	A virtual user ID number which associated user information in the database
2	QUANTITY	The amount of transfer of virtual coins.
3	RECDATE	Time of transfer for virtual coins
4	RECUSERID	The user ID of received the transfer of virtual coins

The pretreatment is carried out according to the transaction behavior characteristic modeling principle. And each characteristic parameter calculation is as in Tab. 2.

Table 2 Description of each dimension

No.	Field name	Meaning	Calculation method
1	*RF*	Recent frequency	The time period *T* is set to three months, and the recent frequency is the transfer times of $User_i$ within three months.
2	*RA*	Recent transfer amount	Set a period of three months, total amount *of* $User_i$'s virtual coins transferring.
3	*AT*	Total transfer money	All transfer money of the $User_i$
4	*AC*	Account coverage	The number of received users of $User_i$ transfer behavior
5	*AR*	Account repetitiveness	For $User_i$ transfer behavior, the ratio between the total number of received users and the number of transfers
6	*NT*	Large amount transferring	The number of transfer records with an amount greater than 5000 of $User_i$
7	*PT*	proportion of large transferring	The ratio between *NT* and the transferring amount of $User_i$

5.2 Experimental Index

After characterizing the data, the OCSVM model is trained and used to predict. The data results of the detection method are presented through simulation. It is difficult to get effect of correct evaluation algorithm with only one index.Therefore, precision, recall, *F*-measure and accuracy are generally used.

Table 3 Description of experimental index

	Relevant	Non Relevant
Retrieved	*TP* (true positives)	*FP* (false positives)
Not Retrieved	*FN* (false negatives)	*TN* (true negatives)

$$\text{Precision: } P = \frac{TP}{(TP + FP)}$$

$$\text{Recall: } R = \frac{TP}{(TP + FN)}$$

$$F\text{-measure: } F = \frac{2 \cdot P \cdot R}{(P + R)}$$

5.3 Test Results and Analysis

The training sample and the test sample are processed according to the investigation department through other evidence. Some suspicious accounts were obtained, and some accounts with no virtual coins or cash flow will be deleted. 1000 users' information was randomly selected and negative samples were removed through manual verification. The OCSVM classifier is trained and tested. The number of outliers in the process of abnormal points mining is set as (10, 20, 30, 40, 50, and 60). The comparison results are as in Fig. 6. These abnormal points will contain the top three levels users in pyramid. Finally, audited outliers are obtained through data association analysis.

Test result is as in Fig. 6, if the number of abnormal points is small, the OCSVM data mining will have more stringent requirements for abnormal specificity, and easier to get the distinct account nodes with a high accuracy. And with the increase of outlier points, the number of target outlier samples that need to be mined is less, so the number of misjudged outlier points increases, and the detection accuracy index decreases. However, for the recall rate indicator, if the number of abnormal point mining is small, a lot of abnormal account information will be lost.

The number of mining parameters at the outliers is set between 30 and 40, the model can get the highest recall rate. It is indicated that as long as the number of outlier mining is set relatively large, the target key accounts can be found. However, the accuracy rate is relatively low for some active users also have more significant trading characteristics, and easy to confuse with key accounts. But this does not mean that the value of the model is low. Although the accounts are not the key accounts controlled by the platform we are looking for, they are also the key persons involved. Finding out these accounts is also valuable for case analysis. For 30 abnormal accounts in the experiment result, 10 of them are on top three layers of pyramid selling network after analysis. The other 8 accounts have no obvious traces of membership activities, and these accounts maybe are controlled by criminals, which are used to complete the releasing virtual coins after other users transferring money offline. From further data association with account registration information, bank card information, and related feedback with pyramid selling organization, it is confirmed that these accounts are controlled by pyramid selling platform. So in reality, in order to obtain evidence for all abnormal accounts as far as possible, we need to obtain more accounts according to the recall rate index, and then use relevance feedback to filter target key accounts.

However, if the abnormal point reaches a certain scale, the maximum recall rate can be obtained. In reality, in order to get as much evidence as possible from all the abnormal accounts, more accounts need to be obtained according to the recall rate indicator, and then related feedback should be used to filter.

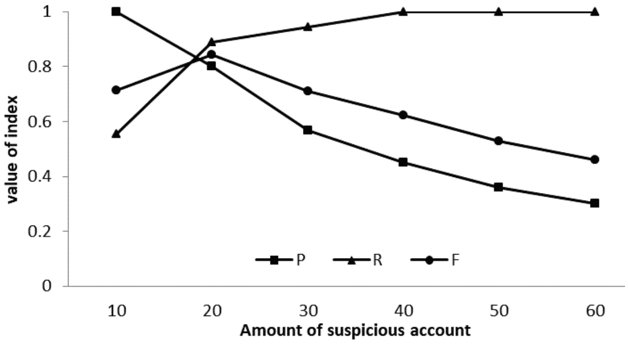

Figure 6 Results of the experiment

There are 8 accounts that are manipulated by the platform from investigation. In traditional way, the identity cards and mobile phone are usually used to associate with the virtual accounts. But only three of them are directly being found in this way. If we set to find 40 outliers in clustering-based mining model in IBM SPSS Modeler application, 13 abnormal accounts are checked, and four of them are controled by the platform. The comparison result is as Tab. 4, the accounts found from association can be directly judged as key accounts, but many accounts have no valid association information, or forged information. So it is difficult to find these accounts. Since the controled accounts are similar to the active accounts in the platform, there is also a large false positive rate when mining these key accounts. However, these misjudged active accounts are also valuable in actual. As the key accounts are easy to aggregate with other active users and may hide anomalous features in a few dimensions. When the positive and negative samples are unbalanced and there are less negative samples, OCSVM is more effective than general clustering method. In our study, OCSVMClassifier has better classification effect in high-dimensional space, and effective for multi-dimensional feature data.

Table 4 Comparison Results

	OCSVM	*Cluster-based*	*Association-based*
Key account	18	13	3
Controled account	8	4	3
P	0.45	0.33	1
R	1	0.73	0.17

6 CONCLUSION

Internet pyramid selling seriously harms the development of network finance and social stability. The core members and key members of network pyramid selling organizations and other illegal criminals need to be investigated and dealt with according to law. A large amount of evidence and clue information are hidden in electronic data. The database analysis is the key problem to investigate the case. The identification and analysis of large data is time-consuming and laborious. This paper researches the transfer behaviour of virtual coins on behalf of pyramid selling organizations. The OCSVM classification model is used to classify the user's transfer behaviour, filter mining exception users, and discover the virtual accounts controlled by core members through data association feedback analysis and personnel hierarchy relationship, forming a chain of electronic data evidence. This model describes a mining method of data analysis for key virtual accounts in the process of database forensics and electronic data reconnaissance in pyramid selling mode. In practice, the pyramid selling model is changing, a universality of the model and data mining objects need to be further explored. It is believed that in the actual electronic forensics, data extraction and analysis in the database involve various aspects. According to the characteristics of pyramid selling model, we should design more data analysis models, find patterns of the data, and help to find evidence.

Acknowledgment

This research was financially supported by the education department of science and technology plan projects, NO. 171056 in Jiangxi province; and Soft science project, and supported by Open Fund Project (NO. JXJZXTCX-018), in Collaborative Innovation Center for Economics Crime Investigation and Prevention Technology, Jiangxi Province.

7 REFERENCES

[1] Cox, J. (2014). Fast money schemes are risky business: gamblers and investors in a papua new guinean ponzi scheme. *Oceania, 84*(3), 289–305.Doi:10.1002/ocea.5062

[2] Tencent. (2017). NPS situation awareness white paper https://slab.qq.com/news/authority/1745.html Published time: 2018-03-30.

[3] Liu, Q., Zhang, X., Zhang, L., & Zhao, Y. (2018). The interaction effects of information cascades, word of mouth and recommendation systems on online reading behavior: An empirical investigation. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-018-9312-0

[4] Moore, T., Han, J., & Clayton, R. (2012). The postmodern ponzi scheme: empirical analysis of high-yield investment programs. *International Conference on Financial Cryptography and Data Security*, Springer, Berlin, Heidelberg, Vol. 7397, 41-56, https://doi.org/10.1007/978-3-642-32946-3_4

[5] Liu Zhijun, Yao Jie, Wu Huadong, & Wang Ning, (2015). Electronic evidence analysis of network pyramid selling. *Computer Science*, 10, 99-101. (In Chinese)

[6] Santra, P., Roy, P., Hazra, D., & Mahata, P. (2018). Fuzzy data mining-based framework for forensic analysis and evidence generation in cloud environment. https://doi.org/10.1007/978-981-10-7386-1_10

[7] Leu, F. Y., Tsai, K. L., Hsiao, Y. T., & Yang, C. T. (2017). An internal intrusion detection and protection system by using data mining and forensic techniques. *IEEE Systems Journal, 11*(2), 427-438. https://doi.org/10.1109/JSYST.2015.2418434.

[8] Olivier, M. S. (2009). On metadata context in database forensics. *Digital Investigation, 5*(3-4), 115-123. https://doi.org/10.1016/j.diin.2008.10.001

[9] Beyers, H., Olivier, M., & Hancke, G. (2011). Assembling Metadata for Database Forensics. *IFIP International Conference on Digital Forensics*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24212-0_7

[10] Al-Dhaqm, A. M. R., Othman, S. H., Razak, S. A., & Ngadi, A. (2015). Towards adapting metamodelling technique for database forensics investigation domain. *International Symposium on Biometrics & Security Technologies*. IEEE. https://doi.org/10.1109/ISBAST.2014.7013142

[11] Fasan, O. M. & Olivier, M. (2012). Reconstruction in Database Forensics. *IFIP International Conference on Digital Forensics*, Springer, Berlin, Heidelberg, Vol. 383, 273-287. https://doi.org/10.1007/978-3-642-33962-2_19

[12] Fruehwirt, P., Kieseberg, P., Schrittwieser, S., Huber, M., & Weippl, E. R. (2012). InnoDB Database Forensics: Reconstructing Data Manipulation Queries from Redo Logs. The Fifth International Workshop on Digital Forensics. *International Conference on Availability*, IEEE, Vol. 17, 625-633. https://doi.org/10.1016/j.istr.2013.02.003

[13] Zawoad, S. & Hasan, R. (2015). Digital Forensics in the Age of Big Data: Challenges, Approaches, and Opportunities. *International Conference on High Performance Computing & Communications*, IEEE, Vol. 5, 1320-1325. https://doi.org/10.1109/HPCC-CSS-ICESS.2015.305.

[14] Edem, E. I., Benzaïd, C., Al-Nemrat, A., & Watters, P. (2015). Analysis of Malware Behaviour: Using Data Mining Clustering Techniques to Support Forensics Investigation. *Cybercrime & Trustworthy Computing Conference*. https://doi.org/10.1109/CTC.2014.10

[15] Quick, D. & Choo, K. K. R. (2014). Data Reduction and Data Mining Framework for Digital Forensic Evidence: Storage, Intelligence, Review and Archive. *Trends & issues in crime and criminal justice* No. 480. Canberra: Australian Institute of Criminology. https://aic.gov.au/publications/tandi/tandi480

[16] Li, Y., Zhang, T., Ma, Y. Y., & Zhou, C. (2016). Anomaly Detection of User Behavior for Database Security Audit Based on OCSVM. *International Conference on Information Science & Control Engineering*, IEEE. https://doi.org/10.1109/ICISCE.2016.55

Le-Khac, N. A. & Kechadi, M. T. (2010). Application of Data Mining for Anti-money Laundering Detection: A Case Study. *2010 IEEE International Conference on Data Mining Workshops*. https://doi.org/10.1109/ICDMW.2010.66

[17] Liu, Q., Huang, S., & Zhang, L. (2016). The influence of information cascades on online purchase behaviors of search and experience products. *Electronic Commerce Research, 16*(4), 553-580. https://doi.org/10.1007/s10660-016-9220-0

[18] Chai, Y. & Dawit, H. (2016). Combined data mining techniques based patient data outlier detection for healthcare safety. *International Journal of Intelligent Computing and Cybernetics, 9*(1), 42-68. https://doi.org/10.1108/IJICC-07-2015-0024

Contact information:

Jianying XIONG, Associate Professor
Jiangxi Police College,
Department of Security Technology,
Nanchang, China 330013
17664376@qq.com