Using transaction anomaly analysis to find key account in network pyramid selling

Jianying Xiong

Security Management Department of Jiangxi Police College, Jiangxi Police Institute, China

*special8212@sohu.com

Keywords: Network Pyramid Selling, Database Forensics, Crime Personnel Organization, Key Account, Anomaly Point Mining.

Abstract: Investigation and forensics for network pyramid selling need to analyzed database. The database always has many data types and complex relationships between tables. It is impossible to rely on the experience of technicians to analyze and obtain useful evidence from this massive and changing electronic evidence. It is important to find the key person in the crime which can help to identify network pyramid selling organization structure and to track the flow of funds in network pyramid selling crime. We researched a method of mining transaction anomaly point to find out the key account, and further to find the core members of the organization. Through case experiments, it shows that some suspect key account of network pyramid selling is discovered by transaction abnormal point analysis. Using the correlation analysis of accounts can further confirm user's information, and provide clues for network pyramid selling investigation.

1. Introduction

"The mobile pyramid selling fraud threat situation analysis report in China 2017" shows that there are still a lot of network MLM frauds, which mainly can be divided into the following categories: shopping rebate, rebate and interest game, virtual currency, financial assistance, payment and financial, telecommunications business. pyramid selling fraud always use network marketing promotion, develop Sales levels, buy financial products, shopping platform full back to earn profit. Pyramid selling frauds is an insight into the greed of human nature to money, with high interest as bait, and then a hierarchical system to develop multiple hierarchical system. Once a part of the capital flow is broken, it will easily cause a crash. Network MLM activities is relying on the Internet viral spread, a large number of people and money is involved. And at last it will cause a significant economic loss to the masses, and bring a serious distortion of human values, a serious damage to the social credit system.

The detection of network pyramid selling has some characteristics. Firstly, it is difficult to find the case. The illegal pyramid marketing in the network is mostly in the form of "electronic commerce", "advertising promotion" and so on. Secondly, it is difficult to collect evidence. Network pyramid selling often involves people all over the country, which bring great difficulty to the forensic work. Thirdly, it is difficult to analyze the data. Network pyramid selling operation data is mainly stored in database system, which is also an effective way to manage group data and members effectively. Therefore, the data mining and analysis in the database can improve the effect to combat crime.

It is lack of data forensics model of network pyramid selling. The database as the most important kind of electronic data, has involved profit statistics, the total number of members the amount of each of the major suspects, and many other useful information. It can reveal the distribution of interest and organization etc.. Because there are many database tables, many field names in the table, and the naming way of each table and field is not uniform. Different data in the table (attribute column) may be having duplication. Outlier mining is useful in many fields, such as online stock trading abnormal behavior analysis, network community detection, money laundering in financial banking institutions, suspicious account transaction behavior recognition based on time series

  DOI: 10.25236/systca.18.089

analysis, identification of suspicious foreign exchange transactions. Outlier detection is a basic research direction. Outlier monitoring technology can assist manual analysis to identify suspicious transaction accounts. And it is an important entry point to apply data mining technology to pyramid selling.

2. Model Definition

2.1. The Working Principle of Model

Network pyramid selling investigation and evidence collection requires the number of people involved, personal network structure, core personnel and backbone members of pyramid sales organization, so as to better control the involved persons. However, the virtual identity on the network is difficult to verify, so that the criminal gang members cannot be successfully identified. At the same time, capital flow is also an important evidence for pyramid selling crime. There are many user behaviors such as user transactions, payment, reward distribution and other related data which are directly or indirectly related to capital transfer.

If the user's account trading behavior is significantly different from other users, it may be indicated that this is a key account, and help to investigate the crime and control the involved person. It can also dig out more appropriate account transactions, profits, capital flows of information in a large number of data; help to construct the chain of electronic data. The working principle of the model is as follows:

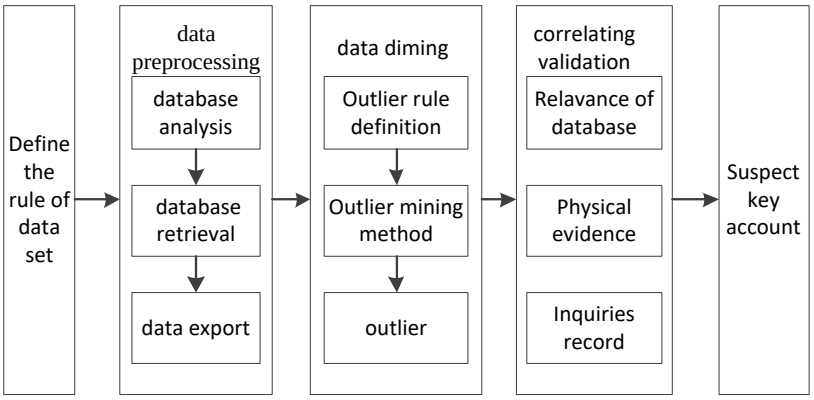

Figure 1 The Process of Model.

Step 1: Define the rule of data set, to determine the content of the information that the target data set needs to be included.

Step 2: data preprocessing, extract useful information from the database of a pyramid marketing site. We need to analyze the database, design the corresponding data retrieval rules to query, and export the results of the query.

Step 3: data mining analysis, define the anomaly rules, use a method to mine the anomaly point in big data.

Step 4: relevance verification, use correlation analysis to feedback the results of evidence, and verify user information through other evidence, and to establish a complete chain of evidence.

2.2. Definition of Mining Model

The outlier is the small pattern data in the data set, which may be caused by the measurement or execution of the error, and may also be the result of the inherent data variability. Hawkins gives its essential definition: the outlier is the unusual data in the data set, which makes people suspect that these data are not random deviations, but arise from completely different mechanisms. There are two basic tasks in the general outlier mining.

(1) The data set rule definition for outliers.

There are many kinds of virtual transactions in network pyramid selling. In this study, the pyramid selling organizer will get money offline, and then it will transfer to the user of the gold

coin through the virtual account in the network marketing platform. But the virtual account transactions of these organizers are different, the transactions frequency is very high, the transaction amount is relatively large, the transaction is relatively dispersed from the common users. In the study, the following is the definition rules of trading frequency, transaction amount, transaction scope.

Transaction Frequency F: the number of transactions for an account in one cycle. If the account has a higher number of transactions, it can indicate that the account is a trading concentration point, and the corresponding virtual account is the core position in the transaction network.

Transaction amount M: the total turnover of the virtual coins by the account, if the amount transferred a larger number of virtual coins, it is indicated that the account is a generation of virtual coins, and the corresponding accounts are more likely to be the beneficiaries of network pyramid schemes.

Transaction scope Q: the number of transaction accounts involved in the transaction activities of the account. The common virtual coin always transfers between ordinary accounts among acquaintances. But the key account bind to the criminal suspect will take transaction with different user, and the core accounts will cover many trading nodes to complete the issuance of virtual assets.

(2)The method of mining the outlier.

Outlier detection is to select K from samples, which is significantly different from other data, it is exceptions or unconformity. Guided learning and unsupervised learning are the two main methods to mine the outlier. In the transaction data of a pyramid marketing website, the core account is hidden in a large amount of transaction data. We can use an undirected learning method based on the basic distribution of users' normal consumption behavior. We can find out the most probable outliers by diagnosing sample data. Here we use the multidimensional clustering based outlier diagnosis method.

From the perspective of comprehensive analysis, we can find the outlier through computing the distance between sample points and data groups, and judging distance. Then achieve the outlier diagnosis and cause analysis.

(3) Definition of outlier mining model.

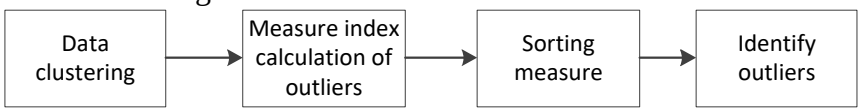

Figure 2 Process of Mining the Outliers.

Data clustering: the transaction frequency, transaction amount and transaction scope are defined as cluster variables Ki. The two step clustering method is used to find various centers, and calculate the mean and variance of each clustering variable Ki, and calculate the sample size of each cluster space. The index of outlier measure is calculated.

GDI: Group Deviation Index. It reflects the sample S adds into category V will caused internal differences increased in category V.

VDI: Variable Deviation Index. It reflects that when the sample S adds into category V will cause internal differences increased in category V, the contribution of each cluster variable.

AI: Anomaly Index, it is a comprehensive visual value of the difference change. If the value is greater, it has a greater possibility of outliers.

VCM: Variable contribution Measures, it is used to reflect the proportion of each cluster variable to the difference contribution.

Ranking the index to determine outliers: arranged the data in descending order by AI, and the former m may be outliers. Meanwhile, for outliers, sorted descending order by VDI, and the variables at the top L bits are the main reasons leading to outliers.

2.3. Correlation Analysis for Accounts

From the electronic data to the real world, forensics is the process of constructing Association of Virtual World, to find the value of virtual coin, the virtual account associated with the real identity. Traditional forensics technology solves the problem of related equipment through IP, MAC and

other information. With the change of the legal environment, we have to solve the problem of the relationship between people and data. Pyramid marketing website database mining forensics also requires multi-level forensics, combined with capital flow, communication flow, data acquisition, analysis and application of network flow, to build a complete chain of evidence.

The correlation analysis of account data forensics, need to use the virtual account through in-depth analysis, including ID card, phone, QQ, WeChat and other personal information, the bank's funds transfer, communications records, and combined with the inquiry transcripts to obtain evidence for authentication of the user information, the establishment of a complete chain of evidence.

3. Experimental Analysis

3.1. Experimental Date

The experimental data originate from a database file of a network Pyramid selling crime. When the criminal receives the money from the user, they used the transaction service between users to allocate the corresponding virtual currency through their associated account. So the object is to mine the key accounts of virtual currency allocation. The data involved 2195801 transactions, involved 48564 of transfer out accounts and 81618 of transfers in accounts. The frequency, amount, and scope of transaction data are extracted.

3.2. Experimental Date

Because there is massive data are recorded, In order to find out the core users, we use two step clustering method, then obtain five categories as in fig.3. The cluster 1 and cluster 2 are eliminated for less prominent. Category 4,5 records are merged to find the outliers, as in Fig.4, there are 9 outliers are detected, derived the data as in fig.5

Figure 3 Clustering results

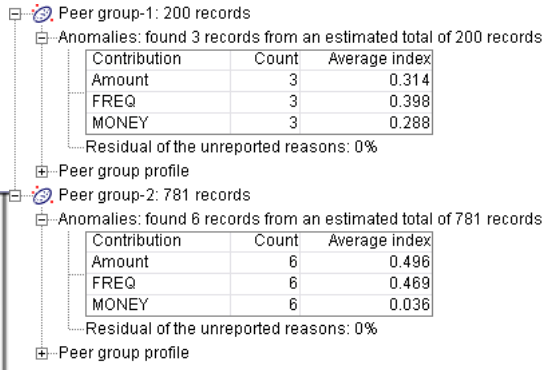

Figure 4 Outliers mining results

File Edit Generate

	USERID	...	Amount	FREQ	MONEY	$O-Anomaly
1	211986	...	620.000	4330....	84524.000	T
2	8407624	...	569.000	3383....	173284622.0...	T
3	877678	...	633.000	3747....	7852096.000	T
4	460394	...	14.000	750.0...	27663.000	T
5	631729	...	59.000	760.0...	362666.000	T
6	458133	...	125.000	216.0...	978816.000	T
7	463775	...	126.000	318.0...	2001015.000	T
8	7018364	...	24.000	759.0...	303810.000	T
9	448797	...	116.000	337.0...	3106183.000	T

Figure 5. Records of outliers

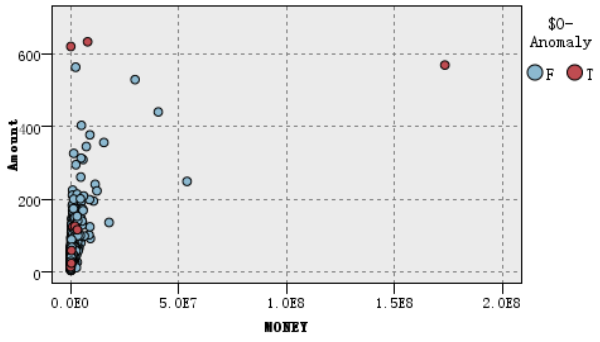

Figure 6. the scatterplot of outliers

Using the data visualization plot, we can see the red label outliers in the figure. and in addition to the system display outliers, according to the investigation to clustering sparse graph is also listed as

doubtful point of investigation, such as the relatively discrete points. It can improve the investigation efficiency of large amounts of data.

4. Conclusion

Network pyramid selling seriously harms the development of network finance and social stability. For network pyramid selling organizations, the core members and backbone criminals and other criminals need to be dealt with severely according to law. The investigation of forensic evidence and clues is in the form of database. For the mass data and complex relations of tables, the paper aims to obtain criminal group related information through the outlier data mining. It can help detector find the suspect account, to find their abnormal tractions. In forensic practice, the extraction and analysis of data can help to detect crime; and data mining can be used in specific case.

Acknowledgements

This research was financially supported by the education department of science and technology plan projects, NO.171056 in Jiangxi province.

References

[1] Singh, J., & Aggarwal, S. (2013). Survey on outlier detection in data mining. International Journal of Computer Applications, 67(19), 29-32.

[2] Zhuang, H., Zhang, J., Brova, G., Tang, J., Cam, H., & Yan, X., et al. (2015). Mining query-based subnetwork outliers in heterogeneous information networks. 1127-1132.

[3] Wei, P., Luo, A., Li, Y., Pan, J., Tu, L., & Jiang, B., et al. (2013). Mining unusual and rare stellar spectra from large spectroscopic survey data sets using the outlier-detection method. Monthly Notices of the Royal Astronomical Society, 431(2), 1800-1811.

[4] Kanhere, P., & Khanuja, H. K. (2015). A Methodology for Outlier Detection in Audit Logs for Financial Transactions. International Conference on Computing Communication Control and Automation(pp.837-840). IEEE.

[5] Gunestas, M., Mehmet, M., & Wijesekera, D. (2010). Detecting Ponzi and Pyramid Business Schemes in Choreographed Web Services. Advances in Digital Forensics VI - Sixth IFIP WG 11.9 International Conference on Digital Forensics, Hong Kong, China, January 4-6, 2010, Revised Selected Papers (Vol.337, pp.133-150). DBLP.

[6] Bhukya, W. N., & Banothu, S. K. (2011). Investigative Behavior Profiling with One Class SVM for Computer Forensics. International Conference on Multi-Disciplinary Trends in Artificial Intelligence (Vol.7080, pp.373-383). Springer-Verlag.

[7] Zhang, G. Z. (2014). Computer forensics based on data mining. Applied Mechanics & Materials, 536-537(536-537), 371-375.

[8] Wu, C. (2017). Study on the computer forensics algorithm in mass data process based on data mining. Boletin Tecnico/technical Bulletin, 55(18), 267-274.

[9] Cheng, P., & Qu, H. (2015). A digital forensic model based on data mining. International Conference on Information Sciences, Machinery, Materials and Energy.

[10] Edem, E. I., Benza?d, C., Al-Nemrat, A., & Watters, P. (2014). Analysis of Malware Behaviour: Using Data Mining Clustering Techniques to Support Forensics Investigation. Cybercrime and Trustworthy Computing Conference (pp.54-63). IEEE.