Winter 12-2-2007

Transactions Behavior Analysis for Internet Auction Fraud

Yungchang Ku

Chaochang Chiu

Proceedings of The Seventh International Conference on Electronic Business, Taipei, Taiwan, December 2-6, 2007, pp. 68-73.

TRANSACTIONS BEHAVIOR ANALYSIS FOR INTERNET AUCTION FRAUD

Yungchang Ku, Yuan Ze University, Taiwan, unicon@mail.cpu.edu.tw
Chaochang Chiu, Yuan Ze University, Taiwan, imchiu@saturn.yzu.edu.tw

ABSTRACT

People often enmesh the Internet auction frauds which damage the benefits of Internet market and threaten transactions security. This research applies social network analysis and data mining to extract characteristic features from two random collected transaction datasets of Yahoo auction site. One dataset is used to construct prediction model and another is treated as validation. The average accuracy ratio of proposed model is at least 90%. The findings are: (1) the abnormal accounts involve circular transaction; (2) fraud accounts can accumulate higher positive reputations in very short time from its circular transaction and rarely play key nodes in transaction network.

Keywords: Internet auction fraud, social network analysis, data mining.

INTRODUCTION

With the progress and rapid development of Internet techniques, there is no doubt that varies business transaction models have formed in Internet. Through Internet, people can easily search for information, order, sell and trade the products or have auctions. Internet auction is one of the few successful business models through Internet [9]. With Internet, "auction" is no longer the activity participated by few people; anyone can sell and buy any products on auction platforms such as Yahoo! Auction and eBay. There are thousands of objects displayed on auction websites at any time. According to the report of MIC [14], American Internet auction market has become one of the important shopping paths for the Americans. By 2010, the market scale will break through $65 billion USD. With the growth of Internet auction scale and the increase of transaction amounts, Internet auction transaction becomes the target of fraud criminals. Because of the lack of effective countermeasures, Internet auction frauds have been the first of top 10 frauds in recent years [16]. The Consumer Sentinel database of Federal Trade Commission (FTC) also showed that the pleadings of Internet auction frauds in 2005 topped among the consumption issues [1]. In economic, the loss in Internet auction frauds reached $ 183 million USD in 2005 and it increased to $ 198 million USD in 2006. The amount of loss was rapidly increasing among all fraud cases which had great impact to the society and became an important issue of Internet development.

There are many researchers do their efforts in counteracting Internet auction frauds in terms of cheating in auctions [11], payment of fake credit cards [4] and quality of auctioned products [3], and so on. But, rare are in discussion the auction transaction behaviors. This research will follow Wang et al [23] treated social network analytical indicators to detect Internet auction anomaly network. Then, we will apply data mining to generate classification rules to discriminate suspicious and fraud accounts from the anomaly network.

In the following sections, section II is the theoretical background, including Internet auction fraud, social network and data mining. Section III is the analytical methods proposed by this research which introduces transaction network construction, discrimination of normal or abnormal transaction networks and the planning of this research. Section IV includes the experimental results, the rules of Internet auction account detection and the characteristics of different types of accounts. Section V is the conclusion and relative future research direction with respect to the control of Internet auction frauds.

THEORETICAL BACKGROUND

Internet Auction Frauds

Internet auction refers to the auction transaction carried out over the Internet [13]. Different from the traditional auctions, Internet auction means that the buyers and sellers had the transactions without facing each other. Besides, Internet auctions allowed anonymous transactions. Thus, the sellers could easily apply for several account. With the mature development of Internet and the continuous increase of Internet users, Internet auctions had transactions with the advantages and characteristics of Internet and gradually boomed everywhere around the world. However, the essence of Internet changed the way of information exchange between the buyers and sellers. The buyers accept unidirectional messages provided by the sellers which resulted in significant information asymmetry in Internet auction transactions [12]. Due to the information asymmetry between the buyers and sellers and the lack of validation on the authenticity of the goods, the buyers are more likely to encounter the frauds [21]. The fraudsters will tend to have many accounts to forge their reputations and do their best to develop varied frauds to enlarge their benefits [22].

Therefore, Chua & Wareham [5] suggested that Internet auction supplier, institutions and governments should know well about different types and characteristics of Internet auction frauds and fraudsters. They suggested these institutions should construct an effective policy to prevent Internet auction frauds to reinforce the related regulations of current Internet auctions. Ba et al. [2] suggested establishing the credibility validation mechanism of the third party and validating individual identities by public key cryptography and digital signatures to increase the authenticity of the accounts as individuals. Other studies indicated

the development of combined evaluation mechanism by different Internet auction suppliers to increase the reliability of auction transactions [10] [18] [19] [25]. These studies mostly discriminated the dishonest or bad participants according to the reputation of participants' past behavior. The Reputations will help the future transaction participants to find the reliable others to avoid the potential Internet auction frauds [7]. However, the methods will increase the management cost of Internet auction. The cost will add the entering barriers to buyers and sellers. It makes Internet auctions activities become more complexity.

Social Network

The application of social network has been adopted by social science for decades. Mitchell [15] defined social network as the individuals' connection in a group. Applying social network can build the connection of relationships and characteristics of these individuals [6]. Social network consisted of actors, relations and ties [8] [20] [24]. Social network analysis is a good method and technique to understand the relationships among individuals, groups or organizations. It provides several structured indicators for analysis [20], such as centrality and subgroup which are related to this research.

1. *Centrality*: In centrality, the measurement is *degree* which applied to observe the transaction frequency of the accounts. When the *degree* of one account is higher, it means there are more accounts having transaction to the account.
2. *Subgroup*: It aims to investigate the subgroup network of certain accounts in the abnormal social network. The measurement of subgroup includes *k-core, k-plex, n-cliques and betweenness*.

Data Mining

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. It combines the techniques of database, machine learning, artificial intelligence, expert system, pattern recognition, statistics, and data visualization. This research will apply decision tree and neutral network to construct the prediction model to discriminate the suspicious and fraud accounts in the auction network.

THE PROPOSED APPROACH

Construction of Auction Transaction Network

The data source of current research was random collected from auction transaction database of Yahoo auction in Taiwan. The research obtained the initial account from the blacklist announced by Yahoo auction and confirmed the account were "invalid". It then constructed the auction transaction network based on initial account according to the following steps and finally acquired 1488 transactions. Blacklist referred to the list of abnormal transaction accounts announced on auction websites. The list was checked by Internet auctions platform supplier, and validated by the police. Transaction relationship of Internet auctions can lead to a complete transaction network and the steps are as below:

1. Selecting the initial accounts (O)
2. Searching for the accounts ($T_i, i = 1 \text{ to } m$) having direct transactions with the initial accounts (O); $O \rightarrow T_i$ denotes O sold products to T_i.
3. Searching for the accounts ($T_j, j = 1 \text{ to } n$) having direct transactions with accounts T_i besides accounts (O); $T_i \rightarrow T_j$ constructing the transaction network based on accounts (O)

Discrimination of Normal /Abnormal Auction Transaction Networks

Resnick [19] pointed out that buyer and seller in normal Internet auction transaction would not remember each other to have repetitive transaction. Thus, when there was "one-to-many" or "many-to-one" transaction in transaction network, it was regarded as normal transaction. When there was transaction circulation of certain accounts in transaction network, it might be abnormal transaction [23].

Research Setting

As mentioned in section II, there are varied indicators of social network analysis for identifying the relationship of accounts. These indicators were applied for the analysis of Internet auction transaction network, as shown in Table 1.

Table 1.The application of social network analytical indicators to the analysis of Internet auction transaction network

Indicator	Classification of features	Descriptions
degree	*degree* *in-degree* *out-degree*	It shows the transaction relationship of the accounts. When we observe the accounts by A, *in-degree* means A buys the products from others; *out-degree* means A sells the products to others. Thus, we construct the transaction network.
betweenness	*nbetweenness* *(normalizing betweenness)*	The moderating capacity of the accounts in abnormal transaction network
k-core	*1-core, 2-core, 3-core,* *4-core, 5-core, 6-core*	When *k* is more than 2, it means there is circular transaction among the nodes
k-plex	*k-plex_k=2_size=5* *k-plex_k=2_size=6* *k-plex_k=2_size=7*	In abnormal transaction network, we find the accounts in the same group. It shows that the there is regular transaction relationship among the accounts in this group. When *size* is 5, it means there are 5 nodes in the subgroup.
n-cliques	*n-clique_n=1_size=3* *n-clique_n=1_size=4* *n-clique_n=1_size=5*	It forms the accounts in the same group and the accounts in the said group have transactions with those out of the group.

Through Table 1, different indicators will generate different subgroup which formed by certain accounts. We can get different types of certain subgroup. When the accounts are in the certain subgroup, these accounts will be marked as 1; if not, it will be marked as 0. The analytical process diagram of transaction network is shown in Figure 1.

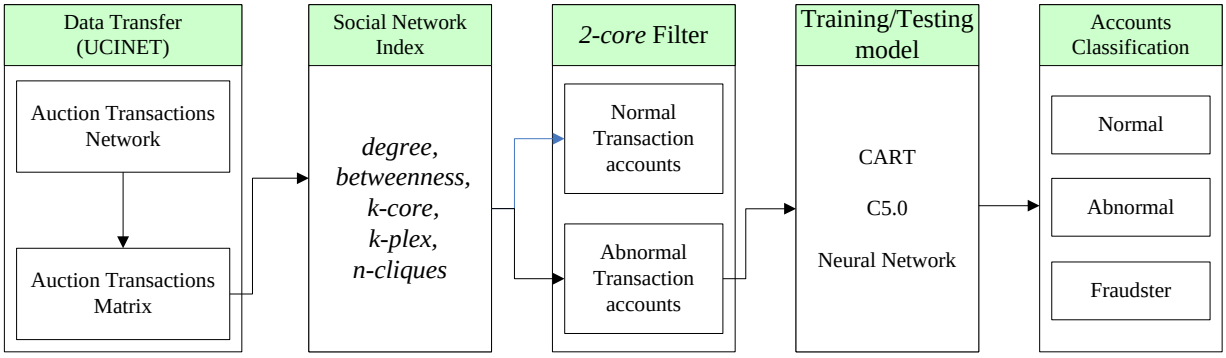

Fig. 1. Analytical process of diagram of Internet auction transaction network

In Figure 1, we first transformed the Internet auction transaction network into transaction matrix and calculated the features of the accounts by different indicators. The step aimed to transform the transaction into indicators. We filtered transaction abnormal accounts by *2-core* and further discriminated suspicious or fraud accounts by data mining. We trained and tested the models by three different methods, including C5.0, CART and Neural Network (NN) to construct and examine different models. In order to validate the accuracy ratio of the models, this research randomly collected 428 transaction accounts (11 of them were validated as fraud accounts) from Yahoo! Auction website to validate the accuracy ratio of the models by training. Table 2 was the numbers of training, test and validation data.

Table 2. Table of training, testing and validation data

	Total accounts	Fraud accounts
Training and testing data	1488	21
Validation data	428	11

THE EXPERIMENT

This research constructed Internet auction transaction network by in-degree and out-degree relationship among the accounts. Then, it generated abnormal transaction network by *2-core* and established the model by CART, C5.0 and Neural Network. Table 3 is the accuracy ratio of 1488 accounts by 5-fold cross-validation training and test.

Table 3. Accuracy ratio of training and test of different models (N=1488; 5-fold cross-validation)

	Training of accuracy ratio			Testing of accuracy ratio		
Model	Average	Worse	Best	Average	Worse	Best
CART	98.03%	97.3%	100 %	92.10%	90.13%	94.15%
C5.0	84.98 %	78.57 %	89.19 %	79.84%	74.36%	84.75%
NN	78.08 %	75.61 %	83.78 %	60.52%	54.68%	67.98%

Since NN is the lowest accuracy ratio for model training and test, we only describe the results of CART and C5.0. First of all, the accuracy ratio of training and test models by CART was the highest. When training the models, the average accuracy ratio could

reach 98.03%; when testing the models, we could have average accuracy ratio up to 92.10%. In C5.0, the average accuracy ratio could reach 84.98 %, and when we tested the modes, the average accuracy ratio could reach 79.84%. The rule description is shown in Table 4 and Table 5.

Table 4. Descriptions of judgment rules constructed by CART

Rule	Rules	Descriptions
1	*If nbetweenness < 0.0135 and 6-core < 0.5 Then fraud =1*	When *nbetweenness* of the accounts is less and 0.0135 and the accounts is in the transaction group upon *6-core*, the accounts are considered as fraud ones.
2	*If nbetweenness >= 0.1605 and k-plex_k=2_n=5 >= 0.5 Then fraud =1*	When *nbetweenness* of the accounts is more than or equal to 0.1605, the accounts are in the transaction group upon 5 nodes(*k-plex_k=2_n=5*) and each account has transaction relationship with at least three other accounts, the accounts are considered fraud ones.

Table 5. Descriptions of judgment rules constructed by C5.0

Rule	Rules	Descriptions
1	*If k-plex_k=2_n=5 > 0 and nbetweenness <= 0.063 Then fraud =1*	When the accounts are in transaction group upon 5 nodes (*k-plex_k=2_n=5*), each account had transaction relationship with at least three other accounts and *nbetweenness* of these accounts are less than or equal to 0.063, they are considered fraud ones
2	*If k-plex_k=2_n=5 > 0 and nbetweenness > 0.15899999 Then fraud =1*	When the accounts are in transaction group upon 5 nodes (*k-plex_k=2_n=5*), each account had the transaction relationship with at least three other accounts and *nbetweenness* of these accounts are more than 0.159, they are considered as fraud ones.

In order to validate the judgment rule effect of the above model, the research validated the other 428 transaction accounts collected. After comparing with the blacklist and website information, we found that there were 38 abnormal transaction accounts and 11 fraud accounts in abnormal transaction accounts. The diagram is shown in Figure 4. The validation results are shown in Table 6 and Table 7

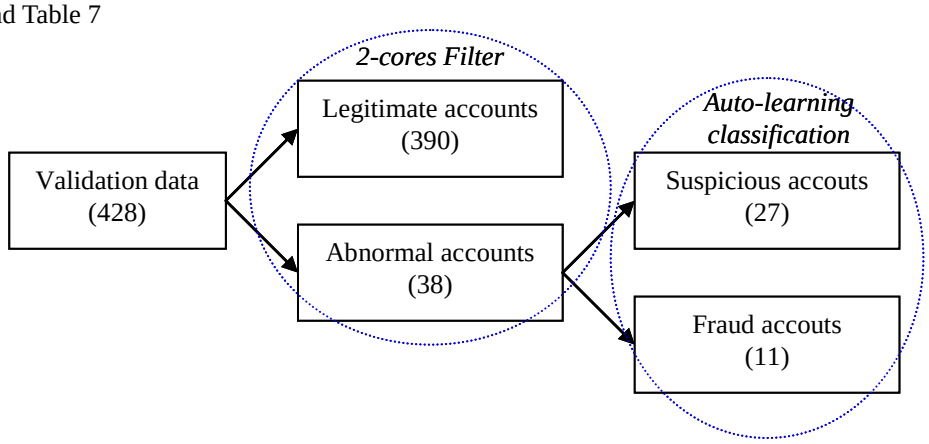

Fig. 4. Classification of validation data (428 pieces of transaction accounts)

Table 6. Prediction result of different training models without the filtering of 2-core (N=428)

Model	CART		C5.0		NN	
	Prediction as Fraud accounts		Prediction as Fraud accounts		Prediction as Fraud accounts	
Fraud accounts	No	Yes	No	Yes	No	Yes
No	414 (99.28%)	3 (0.72%)	407 (97.09%)	10 (2.40%)	401 (96.16%)	16 (3.84%)
Yes	7 (63.64%)	4 (36.63%)	8 (72.73%)	3 (27.27%)	8 (72.73%)	3 (27.27%)

Table 7. Prediction result of different training models with the filtering of 2-core (N=38)

Model	CART		C5.0		NN	
	Prediction as Fraud accounts		Prediction as Fraud accounts		Prediction as Fraud accounts	
Fraud accounts	No	Yes	No	No	Yes	No
No	26 (96.30%)	1 (3.70%)	26 (96.30%)	1 (3.70%)	20 (74.07%)	7 (25.93%)
Yes	2 (18.18%)	9 (81.82%)	2 (18.18%)	9 (81.82%)	8 (72.73%)	3 (27.27%)

After comparing Table 6 with Table 7, we found that when we directly predicted 428 transaction accounts, the accuracy ratio (36.63%) of fraud accounts was much less than that (81.82%) with judgment after filtering the transaction accounts. The accuracy ratio of the model by NN was much less than those of the models upon CART and C5.0. The comparison of accuracy ratios before and after filtering the transaction accounts are shown in Table 8.

Table 8. Comparison of accuracy ratios before and after filtering the transaction accounts by *2-core*

Model	CART		C5.0		NN	
	Before *2-core*	After *2-core*	Before *2-core*	After *2-core*	Before *2-core*	After *2-core*
Accuracy ratio of non-fraud accounts predicts as non-fraud ones	99.28%	96.3%	97.6%	96.3%	96.16%	74.07%
Accuracy ratio of fraud accounts predicts as fraud ones	36.36%	81.81%	27.27%	81.81%	27.27%	27.27%

CONCLUSION AND FUTURE WORK

Although the auction websites provide reputation and blacklist information which might help the buyers and sellers to discriminate the frauds before the transactions, there are still varied frauds. So far, there are no strict laws and complete Internet auction security mechanism in Taiwan or other countries to effectively control Internet auction frauds. This research proposes a simple method to construct the judgment rule on fraud accounts by social network analytical indicators and data mining techniques to reduce the probability of Internet auction frauds. This research can detect the suspicious transaction accounts and fraud accounts in Internet auctions by proposed model. Based on the research results, we realize that social network analysis can transform the Internet auction account transactions into indicators. We can first classify all transaction accounts into normal and abnormal ones by *2-core*. With data mining, we can further divide the abnormal transaction accounts into suspicious and fraud accounts. There are different meanings of these three transaction accounts (normal, suspicious and fraud) in terms of social network indicators and auctions:

1. *k-core* can find the abnormal transaction accounts in the whole transaction network. When there are at least 2 *2-core* in the accounts, it means there are circular and abnormal transactions among the accounts.
2. When *k* value of *k-core* is higher, it means the transaction relationship among the accounts in the group is closer and they tend to be fraud accounts;
3. *k-plex* means there are related groups among the accounts, *nbetweenness* means the possibility of the moderating capacity of the accounts in transaction network. When *k* value is 2 and the accounts consist of 5 nodes (*k-plex_k=2_n=5*), we can find fraud accounts by observing *nbetweenness* of the accounts.

With regard to data mining techniques adopted by this research (CART, C5.0 and Neural Network) except for Neural Network, accuracy ratios of CART and C5.0 are both at least 90%. After comparing the transaction data before and after the filtering by *2-core*, we find that the prediction will be better after filtering transaction data by *2-core* and the accuracy ratio can increase from 36.63% to 81.82%.

Internet auction has become one of the important activities to Internet users in Taiwan. However, the business also increases the frauds. The future studies can be held in the following:

1. The types of Internet frauds will be changed frequently. We can analyze different fraud patterns for future crimes.
2. Through the analysis on actual Internet auction frauds, we can find whether the fraudsters have its organization to commit crime or whether they share experience to other to improve their crime behaviors.
3. This research only analyzes the transaction data collected from current Internet auction transaction websites. If we can cooperate with auction suppliers and obtain the detailed records of validated fraud accounts, we can advance the analysis of

Internet auctions and develop pre-warning detection system to target the potential fraud accounts.

REFERENCES

[1] Anderson, K.B. (2007)"Internet Auction Fraud: What Can We Learn from Consumer Sentinel Data?" , http://www.ftc.gov/be/workshops/internetauction/keithandersonslides.pdf.

[2] Ba, S., Whinston, AB., and Zhang, H.(2003) "Building Trust in Online Auction Markets through an Economic Incentive Mechanism", *Decision Support Systems*, Vol. 35, No. 3, pp.273-286.

[3] Bywell, C.E. and Oppenheim, C. (2001) "Fraud on Internet Auction", *Aslib Proceedings*, Vol. 53, No.7, pp.265-272.

[4] Chae, M., Shim, S., Cho, H., and Lee, B.(2007) "An Empirical Analysis of Fraud Detection in Online Auctions: Credit Card Phantom Transaction", *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pp.155-164.

[5] Chua, C.E.H., and Wareham, J. (2004) "Fighting Internet Auction Fraud: An Assessment and Proposal", *Journal of Computer*, Vol. 37, No.10, pp.31-37.

[6] Freeman, L.C. (1979) "Centrality in Social Networks: Conceptual Clarification", *Social Network*, Vol. 1, pp.215-329.

[7] Gregg, D.G., and Scott, J.E. (2006) "The Role of Reputation Systems in Reducing On-Line Auction Fraud", *International Journal of Electronic Commerce*, Vol. 10, No. 3, pp.95-120.

[8] Haythornthwaite, C., Social Network Analysis: An Approach and Technique for The Study of Information Exchange, *Library and Information Science Research*, 18(4), 323-342, 1996.

[9] Heck, V.E., and Vervest, P. (1998) "How Should CIOs Deal with Web-Based Auction? " , *Communication of the ACM*, Vol. 41, No.7, pp.99-100.

[10] Houser, D., and Wooders, J. (2006) "Reputation in Auctions: Theory, and Evidence from eBay", *Journal of Economics & Management Strategy*, Vol. 15, No. 2, pp.353-370.

[11] Jenamani, M., Zhong, Y., and Bhargave, B.(2007) "Cheating in Online Auction – Towards Explaining the Popularity of English Auction", *Electronic Commerce Research and Applications*, Vol. 6, No. 1, pp.53-62.

[12] Kauffman, R.J., and Wood, C.A. (2007) "Premium Bidding in Online Auctions: An Examination of Opportunism and Seller Preference", http://www.mgmt.purdue.edu/academics/mis/workshop/kw_111805.pdf.

[13] Kumar, M., and Feldman, S.I. (1998) "Internet Auction", *Proceedings of the 3rd USENIX Workshop on Electronic Commerce*, pp.49-60.

[14] Market Intelligence Center (MIC). (2007), "2007 Report of Internet Consumers Behavior and its Shopping Trend in Taiwan", http://mic.iii.org.tw, 2007.

[15] Mitchell, J.C. (1969) *Social Networks in Urban Situation*, Manchester: Manchester University Press.

[16] National Fraud Information Center (NFC). (2007) *Internet Fraud Statistics Reports*, http://www.fraud.org/internet/intstat.htm.

[18] Resnick, P., Zeckhauser, R., Friedman, E., and Kuwabara, K. (2000) "Reputation Systems", *Communication of the ACM*, Vol. 43, No.12, pp.45-48.

[19] Resnick, P., and Zeckhauser, R. (2002) "Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System", *Advances in Applied Microeconomics*, Vol. 11, pp.127-157.

[20] Scott, J. (2000) *Social Network Analysis: A Handbook*, Thousands Oaks, Calif.: SAGE Publications.

[21] Stallaert, B.S., Whinston, J.A.B., and Zhang, H. (2005) "Choice of Transaction Channels: The Effects of Product Characteristics on Market Evolution", *Journal of Management Information Systems*, Vol. 21, No. 4, pp.173-197.

[22] Tokeley, K. (2007) "Vendor Bidding on Online Auctions Protecting Consumers from Vender Bidding on New Zealand Online Auctions", *Journal of Consumer Policy*, Vol. 30, No. 2, pp.137-150.

[23] Wang, J.C., Chiu, C.C., and Ker, H.Y.(2005) "Detecting Online Auction Fraud of Reputation Inflation through Social Network Structures Embedded in Transaction Records", *Journal of Information Management* (Taiwan), Vol. 12, No. 4, pp.144-184.

[24] Wasserman, S., and Faust, K. (1994) *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press .

[25] Weinberg, B.D., and Davis, L. (2005) "Exploring the WOW in Online-Auction Feedback", *Journal of Business Research*, Vol. 58, No. 11, pp.1609-1621.