International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Online Auction Fraud Detection Techniques and Security Measures Using Proactive Moderation System

H. D. Gadade[1], Manoj B. Patil[2], Bhushan D. Patil[3]
[1] *Asst. Professor*, Computer Department, Govt. College of Engg. Jalgaon, Maharashtra.
[2] *Student*, Computer Department, Govt. College of Engg. Jalgaon, Maharashtra.
[3] *Student*, Computer Department, Govt. College of Engg. Jalgaon, Maharashtra.

***Abstract*--The traditional online shopping business model allows sellers to sell a product or service at a preset price, where buyers can choose to purchase if they find it to be a good deal. Online auction however is a different business model by which items are sold through price bidding. There is often a starting price and expiration time specified by the sellers. Once the auction starts, potential buyers bid against each other, and the winner gets the item with their highest winning bid. When people are enjoying the benefits from on- line trading, criminals are also taking advantages to conduct fraudulent activities against honest parties to obtain illegal profit. Hence proactive fraud detection moderation systems are commonly applied in practice to detect and prevent such illegal and fraud activities. Machine-learned models, especially those that are learned online, are able to catch frauds more efficiently and quickly than human-tuned rule-based systems.**

***Keywords*--Online Auction, Fraud Detection, Online Modeling, Moderation systems, rule based system etc.**

I. INTRODUCTION

Since the commercialization of the World Wide Web in the mid 1990's, online marketplaces have been widely explored by commercial organizations for brand awareness and revenue sources. Individuals are able to buy and sell a broad variety of goods and services worldwide on online auction and shopping websites e.g. eBay and Amazon. However, criminals have also attempted to conduct fraudulent activities against honest parties for the purpose of illegitimate profit. On Internet auction sites, auction fraud mainly involves fraud attributable to the misrepresentation of a product advertised for sale or the non delivery of products purchased through an Internet auction site.

Malicious sellers may post an (even non existing) item for bidding with false description to deceive the buyer concerning its true value, and request payments to be wired directly to them. By using wire transfer services, the money is virtually unrecoverable for the victim. Similarly malicious buyers may make a purchase via a fraudulent credit card where the address of the card holder does not match the shipping address. Both consumers and merchants can be victims of online auction fraud, as well as the commercial auction websites. Patterns of auction fraud are changing dynamically and rapidly. To maintain the selection or filter accuracy, moderation systems have to be updated periodically to catch the drifting patterns.

It is desirable to design a learning system that is capable of automatically optimizing weights for the rules based on recent observations [3]. Motivated by applications in a commercial online auction website, we develop various advanced machine learning techniques in the proactive moderation system. By noting the imbalanced labels and the limitation of rule based features designed for the fraud catching, we improve the system by constraining the weights to be positive and introducing imbalanced/weighted loss functions. To overcome the selection bias in labeling, we also include the remaining unlabeled cases into training for unbiasedness. Being aware of specific noise patterns in the expert labels, we further enhance the optimization as done in multi instance learning problems. The final model is formulated as optimizing unbounded generalized linear models in multi instance learning problems, with intrinsic bias in selective labeling and massive unlabeled samples [8].

A. *Online Auction*

Auction is Latin word which means augment. Auctions have existed for centuries. Auction is a bid,

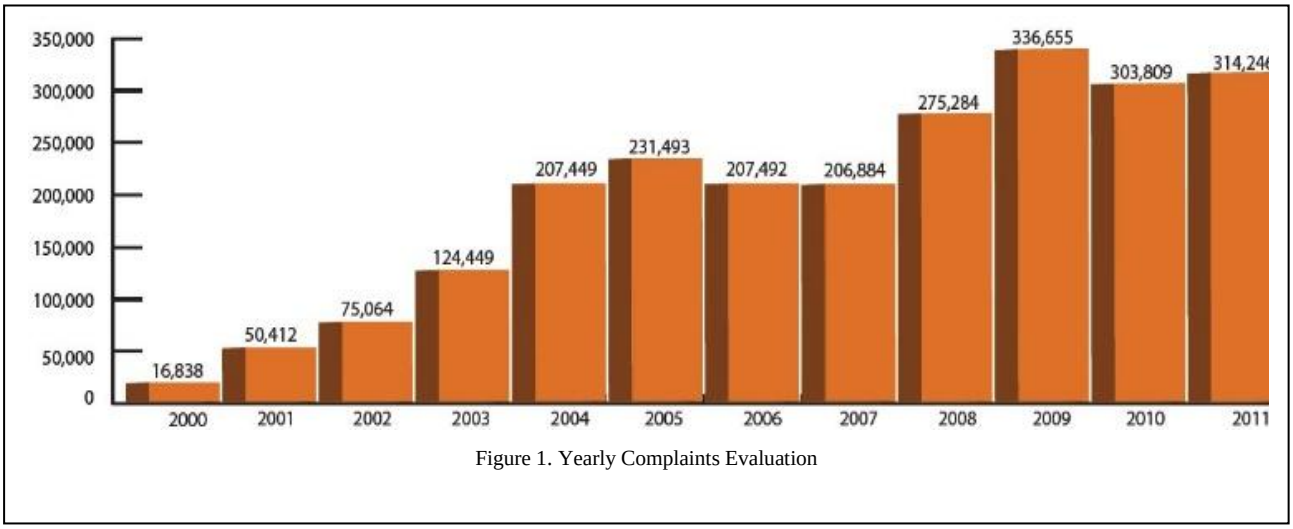

Figure 1. Yearly Complaints Evaluation

a process of selling or buying and services offered take place. There are several different types of auctions and certain rules exist for each auction. There are variations for an auction which may include minimum price limit, maximum price limit and time limitations etc. Depending upon the auction method bidder can participate remotely or in person. Remote auction include participating through telephone, mail, and internet. Many economic transactions are conducted through auctions. Governments sell treasury bills, foreign exchange, publicly owned companies, mineral rights, and more recently airwave spectrum rights via auctions. Art work, antiques, cars, and houses are also sold by auctions. Government contracts are awarded by procurement auctions, which are also used by firm to buy inputs or to subcontract work. There are four commonly used and studied forms of auctions, it's are given below [4].

- Ascending Order Bidding
- Descending Order Bidding
- First-Price Preserved Bidding
- Second-Price Preserved Bidding

These forms are classified according to the price and order of attempt. To pick from the available bidding styles, ascending order variant is the simplest and primitive approach.

B. *Online Auction Fraud*

The vast profits of online auction also attract the attention of criminals who use fraud to cash in on the profitable online trading market. According to annual reports of the Internet Complaint Center, online auction fraud ranks as one of the top serious internet crimes in recent years, the reports shows an increasing of internet crimes respect previous year, in 2001 the centre inward over 300,000 complaints with an increase of 3.4-percent and the total loss was $485.3 million. Yearly comparison of complaints statistics is in "Fig. 1".

Most online auction houses realize that fraud corrodes not only their trustworthiness but also the prosperity of the entire market. For instance, multiple online identities are easy to create and a fraudster could use his many accounts to execute complicated schemes, while camouflaging his malicious target and escaping traditional detection methods that simply examine individual identities [5]. As

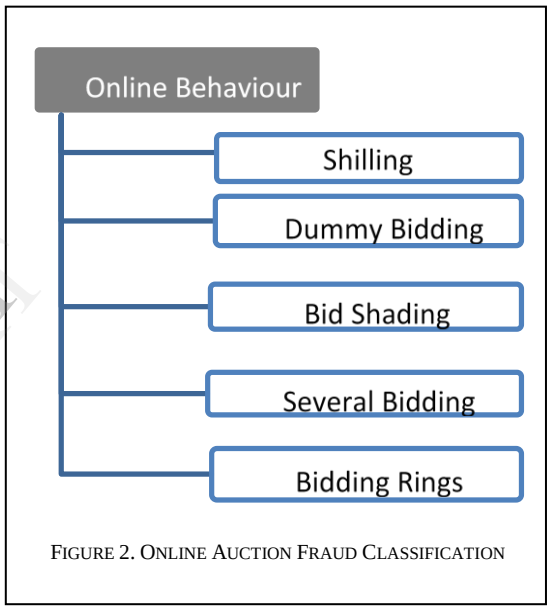

FIGURE 2. ONLINE AUCTION FRAUD CLASSIFICATION

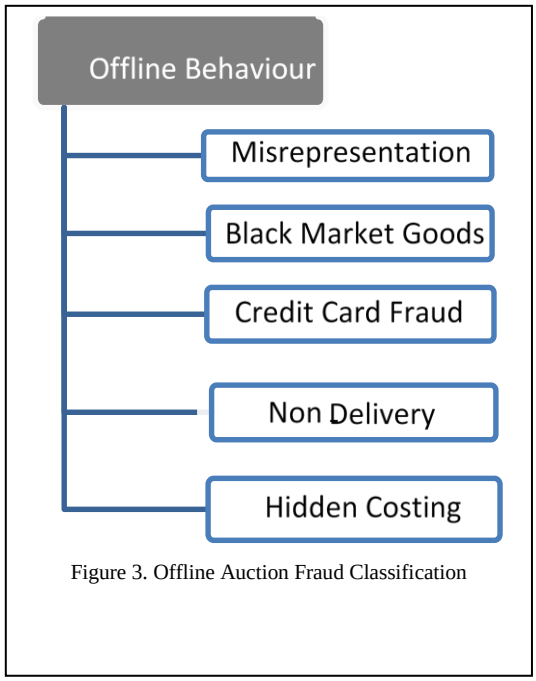

Figure 3. Offline Auction Fraud Classification

more and more inexperienced taders become targeted victims, they begin to distrust the market, resulting in fewer buyers and fewer sellers. To help promote trust in the online market, auction houses developed reputation systems to assist users in evaluating potential trading partners.

II. AUCTION FRAUD CLASSIFICATION

Here we concentrate on the category of online auction frauds. To recognize online auction fraud, it is suitable to first classify the various types of online auction fraud according to the two time periods in which the fraudulent behaviour can take place, it is offline behaviour and online behaviour. In the auction frauds misrepresentation of items, selling of black market goods and triangulation usually happen before the auctions start and non-delivery of goods and fee stacking happen after auctions close, so here classify them as offline such frauds relies more on real-world evidence than on online prevention and detection mechanisms. However, online behaviour frauds happens while transactions are in progress, thus it may occur without leaving direct bodily evidence, and worst of all may not even be noticed by the victims. In "Fig. 2" and "Fig. 3" the auction frauds are classified two ways, one is offline behaviour and next is online behaviour. Here offline auction frauds are classified in five types it happened before the auction start or after the auction close. Online auction frauds are classified five types, it happens while transactions are in progress.

III. FRAUD DETECTION APPROACHES

There are three fraud detection approaches as follows:

A. *Feedback Based Reputation Systems*

Feedback based reputation system is a one of the simplest fraud detection system. In these systems help buyers decide whether to purchase a product based on a feedback score. After the completion of auction, both the seller and the buyer can put down ratings and feedback comments on the other party. These comments and feedback build up in the trader's transaction history, of which the feedback score is one part. So this type of reputation system is simple and easy to understand. It uses positive, neutral, and negative to denote the level of fulfilment for a trade. Analyse the reputation system capacity of finding fraudulent behaviour; it has major drawbacks. Most online auction houses agree to passive approaches to the coordination of reputation systems and management policies that could address fraudulent schemes. Though, if users had more proactive approaches, such as an automatic fraud detector, online auctioning could be safer. The reputation systems development, the most auction houses are used simple method.

The auction houses used the method of marking rates, after a transaction is completed, the traders can only be rated on a positive, neutral and negative scale. The both buyer and seller want to be a good name, so they try to increase the positive feedback, so some of them try to increase their reputation with ambiguous means. Crook could attract buyers by fabricating transaction records that inflate their feedback scores in order to hide their malicious intent. In online auction the buyer using the common fraudulent trick is to first make several small businesses in order earn more positive feedback rating score, but then cheat later on the first expensive product. Another common type fraud is using multiple identities in online auctions. In this type a fraudster first creates multiple identities, dividing them into two groups, fraudsters and legitimate. Then, the fraudsters use the legitimate to artificially boost their reputations by leaving positive ratings. In adding together to the positive rating management, the negative feedback can be prejudiced by fraudulent traders as well. Checking the history of buyer and seller they be worthy of negative feedback ratings intimidate their trading partners into leaving positive ratings, regardless of the actual experience. Here we propose a method to increase the confidentiality of reputation system is using the combination of statistical modelling and automatic anomaly detection based reputation system. It is the best way to improve the quality of reputation system [8].

B. *Data Mining Based Fraud Detection*

Data mining the other name is knowledge discovery is a powerful computer-assisted process designed to analyse and take out useful information from historical data [2]. It allows users to analyse data from different magnitude or perspectives in order to uncover consistent patterns, anomalies and systematic correlations between data elements. The data mining technique used in different areas, the ultimate objective of data mining is to predict future behaviours and trends based on the discovered patterns and association rules. In the area of research most of the researchers have adopted data mining methods to detect shill associations and suspicious patterns [9]. The main steps for implementing this approach for fraud detection within a business organization are [11]. Analyse the fraud objectives and the possible fraudsters, in order to converting them into data mining objectives.

- Data collection and understanding.
- Data cleaning and preparation for the algorithms.
- Experiment design.
- Evaluation results in order to review the process.

Data mining approaches, like reputation approaches, also require analysing huge amounts of historical data, and therefore take a very long time to get results. Using data mining approaches do have the advantage of accuracy compared to reputation systems [4].

C. *Social Network Analysis Based Fraud Detection*

Social network analysis is a set of research procedures for identifying structures in systems based on the relations among actors. Grounded in graph and system theories, this approach has proven to be a powerful tool for studying networks in physical and social worlds, including on the web. In the recent years social network approach has been increasingly applied in fraud detection in internet. With the development of web technologies, there is an increasingly greater amount of interaction by people while on the Internet. Social Networks bring in a multi-disciplinary

approach to solve problems in this domain. It gives new ideas to old problems and gives a new way of looking at them [1]. Social Network Analysis focuses on using quantitative measurement to study the contact among the members to profile the structure of the neighbourhood and its members. It has been developed and practiced in the domain of social studies for decades. Now a day's SNA methods are used among social scientists to investigate the dynamics of a social group.

Focuses have been put on advice network, trust network, and communication network. The basic components of SNA study is the node, as the source of action, and the connection or link, as the relationship developed among nodes. The nodes can be individuals, organizations, or communities [10, 7]. And the links can be one or multiple type of relations or characteristics among the actors for understanding the effects to the social structure to individual members as well as the community as a whole [1]. Network analysis is most effective when combined with other methods in developing theory. Identifying clusters and blocks, or observing patterns in graphic models are technique that enhances descriptions of member's behaviour and the systems within which the behaviour takes place. Finally, testing hypotheses about networks requires comparable data from many networks, or large networks with many subgroups so that the output from network analysis may be imported into standard data files for conventional quantitative analysis.

IV. METHODOLOGY

There is a vast range of business models for online consumer auctions. There are over 200 auction sites on the Web, ranging from the free-standing eBay, which handles 87% of online auction transactions, to the auction sites attached to portals like Yahoo! and MSN, to the auction sites attached to e-commerce sites like Amazon, to specialty auction sites. Some sites charge to list items, others do not (although Yahoo! recently started charging for listings). This variety of business models results in a wide range of practices, which are described in detail below. But despite this variety, some general observations can be made about online auctions. Our application is to detect online auction frauds for a major Asian site where hundreds of thousands of cases posted every day. Every new case is sent to the anticipatory moderation system for pre-screening to assess the risk of being fraud. The current system is featured by:

A. *Rule-based Features*

Human experts with years of experience created many rules to detect whether a user is fraud or not. An example of such rules is "blacklist", i.e. whether the user has been detected or complained as fraud before. Each rule can be regarded as a binary feature that indicates the fraud likeliness.

B. *Linear Scoring Function*

The existing system only supports linear models. Given a set of coefficients (weights) on features, the fraud score is computed as the weighted sum of the feature values.

C. *Selective Labeling*

If the fraud score is above a certain threshold, the case will enter a queue for further investigation by human experts. Once it is reviewed, the final result will be labeled as Boolean, i.e. fraud or clean. Cases with higher scores have higher priorities in the queue to be reviewed. The cases whose fraud score are below the threshold are determined as clean by the system without any human judgment.

D. *Fraud Churn*

Once one case is labeled as fraud by human experts, it is very likely that the seller is not trustable and may be also selling other frauds; hence all the items submitted by the same seller are labeled as fraud too. The fraudulent seller along with his/her cases will be removed from the website immediately once detected.[8]

E. *User Feedback*

Buyers can file complaints to claim loss if they are recently deceived by fraudulent sellers. The following figure shows the registration and login of all the users, sellers and the administrator.

The Administrator will authorize and allow the products which are to be displayed on the online website. Administrator will login with id and password to update database, delete database. Administrator will review the complaints given by users and on the trustability factor he/she is going to recognize the fraudulent seller. Seller signup has to be filled up by the seller to sell their products on website. Seller has to choose unique user id and password. These details are stored in Admin`s database. If seller logins with old id and password, he/she will be set as untrusted. They can only enter the product details but they are denied to display on website. The details of the products of the seller will be stored in the database with details like purchase id, company name, product id, product name, warranty date, product rate, description, complaint etc. It shows all the products of the seller from the day he/she entered into the website. Offers and trustability percentage shows the details of warranty days, product rate, offer rate, offer description, status and trust. Trustability is shown diagrammatically so that it can be easily understood.

Complaints and values gives the details and there values such as product not delivered cases, product mismatches, poor services and product damage cases. The human experts gather all the complaints from the users and they calculates the threshold value. These values are displayed both in percentages and also in diagramatic form. By entering the complaint seller will be able to see the complaint and will take measures to rectify the problems

V. PROPOSED MODEL

Now we describe our Bayesian online modeling framework with details of model fitting via Gibbs sampling. We start from introducing the online probit regression model in we apply stochastic search variable selection (SSVS), a well-known technique in statistics literature, to the online probit regression framework so that the feature importance can dynamically evolve over time. Since it is important to use

the expert knowledge we describe how to bound the coefficients to be positive, and finally combine our model with multiple instance learning.

A. *Online Probit Regression*

Consider splitting the continuous time into many equal size intervals. For each time interval we may observe multiple expert-labeled cases indicating whether they are fraud or non-fraud. At time interval t suppose there are n_t observations. Let us denote the i-th binary observation as y_{it}. If $y_{it} = 1$, the case is fraud; otherwise it is non-fraud. Let the feature set of case i at time t be x_{it}. The probit model can be written as

$P [y_{it} = 1|x_{it}, \beta_t] = \Phi(x_{it}\beta_t)$,

Where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution N (0, 1), and βt is the unknown regression coefficient vector at time t. Through data augmentation the probit model can be ex-pressed in a hierarchical form as follows: For each observation i at time t assume a latent random variable z_{it}. The binary response y_{it} can be viewed as an indicator of whether $z_{it} > 0$, i.e. $y_{it} = 1$ if and only if $z_{it} > 0$. If $z_{it} <= 0$, then $y_{it} = 0$. z_{it} can then be modeled by a linear regression

$z_{it} \sim N (x_{it}\beta_t, 1)$

In a Bayesian modeling framework it is common practice to put a Gaussian prior.

$$\beta_t \sim N(\mu_t, \Sigma_t),$$

Where μ_t and Σ_t are prior mean and prior covariance matrix respectively.

B. *Online Feature Selection Through SSVS*

For regression problems with many features, proper shrinkage on the regression coefficients is usually required to avoid over-fitting. For instance, two common shrinkage methods are L2 penalty (ridge regression) and L1 penalty (Lasso). Also, experts often want to monitor the importance of the rules so that they can make appropriate adjustments (e.g. change rules or add new rules). However, the fraudulent sellers change their behavioral pattern quickly: Some rule-based feature that does not help today might helps a lot tomorrow. Therefore it is necessary to build an online feature selection framework that evolves dynamically to provide both optimal performance and intuition. In this paper we embed the stochastic search variable selection (SSVS) into the online probit regression framework described [8].

At time t, let β_{jt} be the j-th element of the coefficient vector β_t. Instead of putting a Gaussian prior on β_{jt}, the prior of β_{jt} now is

$$\beta_{jt} \sim p0_{jt}1(\beta_{jt} = 0) + (1 - p0_{jt}) N (\mu_{jt} , \sigma^2_{jt}),$$

Where $p0_{jt}$ is the prior probability of β_{jt} being exactly 0, and with prior probability 1-$p0_{jt}$, β_{jt} is drawn from a Gaussian distribution with mean μ_{jt} and variance σ^2_{jt}. Such prior is called the "spike and slab" but how to embed it to online modeling has never been explored before.

C. *Coefficient Bounds*

Incorporating expert domain knowledge into the model is often important and has been proved to boost the model performance. In our moderation system, the feature set x is proposed by experts with years of experience in detecting auction frauds. Most of these features are in fact "rules", i.e., any violation of one rule should ideally increase the probability of the seller being fraud to some extent. A simple example of such rules is the "blacklist", i.e. whether the seller has ever been detected or complained as fraud before. However, for some of such rules simply applying probit regression might give negative coefficients, because given limited training data the sample size might be too small for those coefficients to converge to right values or it can be because of the high correlation among the features. Hence we bound the coefficients of the features that are in fact binary rules, to force them to be either positive or equal to 0. Note that this approach couples very well with the SSVS all the coefficients which were negative are now pushed towards zero.

Suppose feature j is a binary rule and we wish to bound its coefficients to be greater than or equal to 0. At time t, the prior of βjt now becomes

$$\beta_{jt} \sim p0_{jt}\, 1(\beta_{jt} = 0) + (1 - p0_{jt}) N (\mu_{jt} , \sigma_{jt})1(^{\beta}_{jt} > 0),$$

Where $N (\mu_{jt} , \sigma_{jt})1(\beta_{jt} > 0)$ means β_{jt} is sampled from $N(\mu_{jt}, \sigma_{jt})$, truncated by 0 as lower bound

D. *Multiple Instance Learning*

When we look at the procedure of expert labeling in the moderation system, we noticed that experts do the labeling in a "bagged" fashion: i.e. when a new labeling process starts, an expert picks the most "suspicious" seller in the queue and looks through all of his/her cases posted in the current batch (e.g. this day); if the expert determines any of the cases to be fraud, then all of the cases from this seller are labeled as fraud. In literature the models to handle such scenario are called "multiple instance learning". Suppose for each seller i at time t there are K_{it} number number of cases. For all the K_{it} cases the labels should be identical, hence can be denoted as y_{it}. For probit link function, through data augmentation denote the latent variable for the l-th case of seller i as z_{ilt}. The multiple instance learning model can be written as

$$y_{it} = 0 \text{ iff } z_{ilt} < 0, \quad \forall l = 1, \cdots, K_{it};$$

Otherwise $y_{it} = 1$, and $z_{ilt} \sim N (x_{ilt}\beta_t, 1)$, Where β_t can have any types of priors.

VI. EXPERIMENTAL WORK

Our developed application requires implicitly or explicitly collecting visitor purchase information and leveraging that knowledge in your content delivery framework to manipulate what information you present to users.

The steps include:

(1) Collection of data
(2) Analysis of the collected data, and
(3) Determination of the actions that should be performed.

A. *Collection of Data*

Whatever method is eventually used to process the data, information about user's behavior and products must first be collected.

Explicit data collection refers to any method where the user is asked to provide feedback or information about product. Often, this begins after a user purchases a product or used a product. The feedback includes the rating for good, poor delivery, poor manufacturing or usage or general text about the product. All the information will be collected from different users and the status of the product will be updates whether to trust or not.

B. *Analysis of the Collected Data*

The ways that are employed in order to analyze the collected data include are

- Rule-based features:

The trust for particular product(X) can be calculated (in %) by

Trust(X) = 100 - Fraud(X)
Fraud(X) = No of complaints(X) / (No of users(X) * 0.01)

- Selective labeling:

If the fraud score is above a certain level, the case will enter a queue for further investigation by human experts and the cases whose fraud score are below are determined as clean by the human expert.

C. *Decision Making/Final Recommendation*

The decision or the final recommendation after analysis part is to decide whether to ban the product or to trust the product. If the product is banded by the admin then no user can view or buy the product hence providing the user only the trustworthy products.

VII. CONCLUSION

In this paper we build online models for the auction fraud moderation and detection system. This online framework can be easily extended to many other applications like web spam detection, content optimization and so forth. . In this proposed system we provide the responsibility of selling the trustful products by the website itself managed by the admin. So when a customer wishes to buy a product he will get an idea about the product to how much extent he can believe in that product. If he has faced any problem he can make others aware of that product by complaining about the product. This model though it cannot be the ideal way of detecting frauds but it can do the maximum extent in detecting the sellers selling the fraud products.

Regarding to future work, we can include the adjustment of the selection bias in the online model training process and to deploy the online models described in this paper to the real production system, and also other applications.

REFERENCES

[1] Borgatti, P. "What Is Social Network Analysis?" 1998.1998 Social Networks Conference in Barcelona, 5-21

[2] Chau, P. “Catching bad guys with graph mining”. XRDS: Crossroads, The ACM Magazine for Students 17, 3, 2011.

[3] Chua and J. Wareham. Fighting internet auction fraud d: An assessment and proposal. Computer, 37(10):31-37, 2004.

[4] Fei Donga, Sol M. Shatza and Haiping Xub “Combating Online In-Auction Fraud: Clues, Techniques and Challenges”

[5] IC3 annual report about frauds www.ic3.gov/media/annualreport/2011_IC3Report.pdf.

[6] In M. R. Baye, editor, The Economics of the Internet and E-Commerce, volume 11 of Advances in Applied Microeconomics. Elsevier Science, Amsterdam, Netherlands,2002.

[7] Kenneth A. Frank “Mapping interactions within and between cohesive subgroups” Michigan State University, Department of Counselling, Educational Psychology and Special Education, East Lansing, M148824-1034, USA

[8] Online Modeling of Proactive Moderation System for Auction Fraud Detection Liang Zhang Jie Yang Belle Tseng.

[9] Pandit, S., Chau, D. H., Wang, S., and Faloutsos, C. NetProbe: “A fast and scalable system for fraud detection in online auction networks”. In Proceedings of the 16th international Conference on the World Wide Web, Banff, Canada, May 8–12, 2007,201–210.

[10] Roberto Marmo “Data Mining for Fraud Detection System” University of Pavia, Italy.

[11] Shai Rubin, Mihai,Christodorescu,Vinod Ganapathy, Jonathon T. Giffin Louis Kruger Hao Wang:”Auctioning Reputation System Based on Anomaly Detection” Computer Sciences Department University of Wisconsin, Madison