International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol. 3, Issue. 4, Jul - Aug. 2013 pp-2507-2509 ISSN: 2249-6645

A Bayesian Probit Online Model Framework for Auction Fraud Detection

Sirisha Vegunta[1], G. Minni[2]
[1]M. Tech, Nimra College of Engineering & Technology, Vijayawada, A.P., India
[2]Asst. Professor, Dept. of CSE, Nimra College of Engineering & Technology, Vijayawada, A.P., India

Abstract: *Individual users are able to buy and sell a broad variety of goods and services worldwide on online auction and shopping websites, e.g. eBay.com and Taobao.com. However, attackers 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 non-delivery of products purchased through an Internet auction site or the misrepresentation of a product advertised for sale. Malicious sellers may post a 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. Similarly, malicious buyers may make a purchase through a fraudulent credit card where the address of the card holder does not match the shipping address. Both consumers as well as merchants can be victims of online auction fraud, as well as the commercial auction websites. In this paper we study the problem of building models for the online auction fraud detection system, which essentially evolves dynamically over time. We propose a Bayesian probit online model framework for auction fraud detection.*

Keywords: *Fraud detection, Probit, Regression, SSVS.*

I. INTRODUCTION

Online auction networks, such as eBay.com and taobao.com, have become popular trading platforms, with a large variety of products available with competitive prices. Today, these networks have hundreds of billions dollars in trading volume, and hundreds of billions dollars in revenue. While online auction networks have many advantages over traditional retail stores, many usres are still reluctant to sell/buy products on these networks with the concern that sellers/buyers on these networks may not be reliable. To help users assess each other's honesty and integrity, online auction networks often use some reputation-based systems. For example, eBay allows the seller and the buyer to leave feedback to each other for each transaction and the feedback may be viewed by other users. A seller or buyer with more positive comments can be regarded as a more reliable user.

Fraudsters, however, can collude with accomplices to accumulate bogus positive feedback to manipulate the reputation systems, which makes it very hard to evaluate a user's reliability according to the reputation (feedback). It has been observed in [1] that the fraudsters and accomplices are likely to form a dense bipartite core as the fraudsters receive most of the feedback from the accomplices, and are interested in receiving a large number of feedback comments as quick as he/she can. In this paper we study the problem of building models for the online auction fraud detection moderation system, which essentially evolves dynamically over time. We propose a Bayesian probit online model framework for the fraud detection. We apply the stochastic search variable selection (SSVS) [2], a well known technique in the statistical literature, to handle the dynamic evolution of the feature importance in a principled way.

II. RELATED WORK

In the past, attempts have been made to help usres identify potential fraudsters. However, most of them are "common sense" approaches, recommended by a variety of authorities such as newspapers articles [3], law enforcement organizations [4], or even from auction sites themselves [5]. These approaches usually suggest that user be cautious at their end and perform background checks of sellers that they wish to transact with. Such suggestions however, require peoples to maintain constant vigilance and spend a considerable amount of time and effort in investigating potential dealers before carrying out a transaction. Reputation systems are used extensively by many auction sites to prevent fraud. But they are usually very simple but can be easily foiled. In [6], the authors summarized that modern reputation systems face many challenges which include the difficulty to elicit honest feedback and to show faithful representations of users' reputation. In [7] and [8], the authors conducted empirical studies which showed that selling prices of goods are positively affected by the seller's reputation, implying people feel more confident to buy from trustworthy sources. In summary, reputation systems might not be an effective mechanism to prevent online fraud because fraudsters can easily trick these systems to manipulating their own reputation.

In [9], the authors have categorized auction fraud into different types, but they did not formulate methods to combat them. They suggest that an effective approach to fight online auction fraud is to allow law enforcement and auction sites to join forces, which unfortunately can be costly from both monetary and managerial perspectives. Authority propagation, an area closely related to online fraud detection, has been studied extensively in the context of Web search. PageRank [10] and HITS [11] treat a Web page as an "important" if other "important" pages point to it. In effect, they propagate the importance of web pages over hyperlinks connecting them. Trust propagation was used by TrustRank [12] to detect Web spam. Here, the goal was to distinguish between the "good" and "bad" sites (e.g, phishers, sites with adult con- tent, etc).

III. PROPOSED WORK

A. Online Probit Regression

Consider splitting the continuous time into many equal- sized intervals. For each time interval we may observe multiple expert labeled cases indicating whether they are considered as fraud or non-fraud. At time interval "t" suppose there are nt observations. Let us denote the i-th binary observation as 'yit". If yit = 1, the case is consider as fraud; otherwise it is consider as non-fraud. Let the feature set of case i at time interval t be xit. The probit model can be written as follows:

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

where "$\Phi(\cdot)$" is the cumulative distribution function of the standard normal distribution N(0, 1), and "β t" is the unknown regression coefficient vector at time t.

Through data augmentation, the probit model can be expressed in the hierarchical form as follows: For each observation i at time interval t assume a latent random variable zit. The binary response yit can be viewed as the indicator of whether zit > 0, i.e. yit = 1 if and only if zit > 0. If zit <= 0, then yit = 0. zit can then be modeled by using 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 on β t as follows:

$$\boldsymbol{\beta_t \sim N(\mu_t, \Sigma_t)}$$

B. Coefficient Bounds for Fraud Detection

It is always important to incorporate domain knowledge into the modeling framework, which can sometimes boost the model performance. In our online fraud detection system, the feature set x was proposed by experts with years of experience. Currently all the features are in fact binary "rules", i.e. any violation of any one of the rule should somehow increase the probability of fraud. However, simply fitting the model might generate a negative coefficient on some of the features, because given limited training data, the sample size might be very small for those coefficients to converge to the right values, or when some features are highly correlated. Hence we bound the coefficients of those binary "rules" to force them to be equal or greater than zero. Specifically, we consider the following optimization problem:

$$\min_{\beta} \quad \sum_i w_i[y_i \log(1+\exp(-x'_i\beta)) +$$

$$(1-y_i)\log(1+\exp(x'_i\beta)) + \rho\|\beta\|_k]$$

C. Online Feature Selection through SSVS

For regression problems with many features, proper shrinkage on the regression coefficients is usually required to avoid the case of over fitting. For instance, two common shrinkage methods are L1 penalty (Lasso) and L2 penalty (ridge regression). Also, experts often want to monitor the importance of the selection rules so that they can make appropriate adjustments (e.g. change rules or add new rules). However, the illegal 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 intuition and optimal performance. In this paper we embed the stochastic search variable selection (SSVS) into our online probit regression framework.

At time interval t, let β jt be the j-th element of the coefficient vector β t. Instead of putting the Gaussian prior on β jt, the prior of β jt now is

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

where p0jt is the prior probability of β jt being exactly zero.

D. Multiple Instance Learning

When we looked into the actual expert reviewing and the labeling process, we noted that the experts actually assign labels in a "bagged" fashion, i.e. for each seller identification number, one expert looks through all of his/her posted items, and if the expert finds any item as fraud, all of this seller id's posted items are labeled as fraud. In literature the models for this scenario is known as "Multiple Instance Learning" . Suppose for each labeled seller i, there are Ki number of cases. For these cases, the labels should be same, thus can be denoted as yi. The multiple instance learning model with the logistic function becomes

$$P[y_i = 1] = 1 - \prod_{j=1}^{K_i} \frac{1}{1+\exp(x'_j\beta)}$$

which is essentially a noisy-or likelihood function. The noisy-or likelihood function only requires a subset of the events in the bag are fraud rather than all are fraud events. The optimization problem can thus be written as:

$$\min_{\beta} \quad \sum_i w_i[-y_i \log(1 - \prod_{j=1}^{K_i} \frac{1}{1+\exp(x_j'\beta)})$$

$$+(1-y_i)\sum_{j=1}^{K_i} \log(1+\exp(x_i'\beta)) + \rho K_i \|\beta\|_k]$$

$$+\sum_i \tilde{w}_j[\log(1+\exp(z_j'\beta)) + \rho\|\beta\|_k]$$

IV. CONCLUSION

Online auction and online shopping have achieved more and more recognition due to the emergence of the world wide open and the problem of building online machine-learned models for identifying the auction deception in e-commerce web sites is considered. As users are enjoying the advantages from online trading, fraudsters are also taking advantages to accomplish deceptive activities against candid parties to obtain dishonest profit. Therefore to detect and prevent such illegal and deceptive activities, proactive fraud detection moderation systems are commonly applied in practice. We show that our proposed online probit model framework is based on the real word online auction fraud detection data, which combines bounding coefficients from proficient knowledge, online feature selection and several instance learning and can extensively develop over baselines and the human-tuned model. This online modeling frame can be simply extended to various other applications. The adjustment of the selection bias in the online model training process is included to one direction and has proven to be very efficient for offline models too.

REFERENCES

[1] D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," in Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases, 2006, pp. 103–114.
[2] E. George and R. McCulloch. Stochastic search variable selection. Markov chain Monte Carlo in practice, 68:203–214, 1995.
[3] Usa today: How to avoid online auction fraud. http://www. usatoday.com/tech/columnist/2002/05/07/yaukey.htm, 2002.
[4] Federal trade commission: Internet auctions: A guide for buyers and sellers. http: //www.ftc.gov/bcp/conline/pubs/online/auctions.htm, 2004.
[5] ebay: Avoiding fraud. http: //pages.ebay.com/securitycenter/avoiding fraud.html, 2006.
[6] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. Communications of the ACM, 43, 2000.
[7] M. Melnik and J. Alm. Does a seller's ecommerce reputation matter? evidence from ebay auctions. Journal of Industrial Economics, 50:337–49, 2002.
[8] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment, 2003.
[9] C. Chua and J. Wareham. Fighting internet auction fraud: An assessment and proposal. In Computer, volume 37 no. 10, pages 31–37, 2004.
[10] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In WWW, 1998.
[11] J. Kleinberg. Authoritative sources in a hyperlinked environment. In Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
[12] Z. Gyongyi, H. G. Molina, and J. Pedersen. Combating web spam with trustrank. In VLDB, 2004.