



Detecting Frauds In Online Auction System

K. Prathyusha*

ECM&K L University
India.

T. Anuradha

ECM&K L University
India.

R. Sai Nikitha

ECM&K L University
India.

K. Meghana

ECM&K L University
India.

Abstract— We know that these days online shopping increased drastically. E-commerce is growing faster than predicted as it is up over 400% in the past 3 years. As customers have the ease to buy things without spending much time there are also criminals who try to fraud and get profit in illegal ways. Many pro-active fraud detection techniques came to reduce the frauds and illegal activities. Machine-learned models especially those are learned online, are able to catch frauds more efficiently and quickly than human tuned rule-based systems. In this paper, we study about online fraud detection.

Keywords— Online auction, Fraud detection, Online equity regression, fraud churn, fraud score, selective labelling.

I. INTRODUCTION

Since the emergence of World Wide Web (WWW) as in [1], electronic commerce, which is commonly called as e-commerce as in [20] become more and more popular. No often do we now think of taking a stroll through the market before buying a mobile handset, but a healthy online research which in some cases is consequently followed by an online purchase. The scenario is not limited to mobiles alone. It covers a wide range of products like home appliances, consumer electronic goods, books, apparels, travelling packages etc. and even the electronic content itself. With e-Commerce as in [20] then, you can buy almost anything you wish for without actually touching the product physically and inquiring the salesman a number of times before placing the final order. In traditional online shopping business model sellers as in [14] sell their products or services at preset price, where buyers can choose what product best suites them which is of good deal. Online auction however is a different business model where the items are sold through price bidding. Usually bidding have starting price and expiration time. Potential buyers in auction bid against each other, and the winner is the one who bids the item for highest price. To provide some assurance against fraud and to give confidence to online auction services as in [7] E-commerce sites provide insurance to victims for those who loss up to a certain amount. Online auction services and e-commerce sites adopt following approaches to control and prevent fraud. To buy a certain product from the online auction website they are to be validated with e-mail, SMS, or phone call verifications as in [4]. In this paper, we study the application of a proactive moderation system as in [18] for fraud detection, where hundreds and thousands of new auction cases are created every day. Due to the limited expert resources only 20%-40% of cases can be reviewed and labelled. Therefore, it is necessary to develop pre-screening moderation system as in [18] that only directs suspicious cases for expert review and passes the rest as clean cases. Human experts are also willing to test and see the results of online feature selection to monitor the effectiveness and stability as in [5] of the current set of features, so that they can understand the pattern of frauds done by fraudulent sellers and further add or remove some features.

Our contribution:

In this paper we study the problem of building online models for the auction fraud detection moderation system as in [10]. We propose a Bayesian online fraud detection model framework for the binary response. We apply the stochastic search variable selection (SSVS) as in [16], a well know technique to handle statistical literature, to handle the dynamic evolution of the feature importance in a principled way. Similar to as in [18], we consider the expert knowledge to bound the rule-based coefficients to be positive. Finally, we consider to combine this online model with multiple instance learning as in [20] that gives even better empirical performance.

II. METHODOLOGY

Our application is to detect online auction where hundreds of thousands of new auction cases are posted every day. Every new case which has been entered are sent to proactive anti-fraudging system as in [15] for pre-screening to assess the risk of being fraud. The current system is featured by:

- **Blacklist:** Human experts with years of experience created many rules to detect whether a user is fraud or not. An example of rule based feature is “blacklist”, as in [21] where the user has been detected or complained as fraud before. If the user has already done fraud he is blacklisted and so prevents this user to access the account in future Each rule can be regarded here as a binary feature which indicates the fraud likeliness.
- **Fraud score:** The existing system only supports linear models. Here frauds are detected by fraud score which is computed as the weighted sum of the feature values on the given set of coefficients (weights).

- **Selective labelling:** By taking a value of bench mark as in [21] on fraud score, if fraud score is above certain threshold, the case will enter the queue which will be handled by human experts for further investigation. Once it is reviewed, the final result will be taken as Boolean i.e., either fraud or clean. Queue is ordered in such a fashion that the case will be given higher priority in queue if it has higher fraud score. The cases whose fraud level is low i.e., which are below the threshold are determined as clean by the system without any human judgment.
- **Fraud churn:** If a case is labeled as fraud by human experts, it means that the seller is not trustable and there is also a chance that he can also sell other fraud products as in [20]. Hence all the items submitted by the seller in that online site is labeled as fraud too. Once that seller is detected to be fraudulent seller his/her cases will be removed from the website immediately once detected. By this the fraudulent sellers can be blocked and removed from the online auction.
- **User rating and feedback:** For every product bought there will be user feedback, by this the buyers who want to buy the product will know whether it is genuine or not by the first buyer who bought it. Buyers can complain to claim loss if they are recently deceived by fraudulent sellers.

III. ONLINE EQUITY REGRESSION

Consider splitting of continuous time into many small equal size intervals as in [16]. For every interval we may have many expert labelled cases which indicates whether they are fraud or not.

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 online fraud detection model as in [8] can be written as

$$P [y_{it} = 1 | x_{it}, \alpha t] = \Phi (x'_{it} \alpha t), \quad (1)$$

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

By data augmentation the online fraud detection model can be expressed 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 $z_{it} \leq 0$, then $y_{it} = 0$. z_{it} can then be modeled by a linear regression as in [19].

$$z_{it} \sim N (x'_{it} \alpha t, 1). \quad (2)$$

In a Bayesian modelling framework it is common practice to put a Gaussian prior on αt , as in [6]

$$\alpha t \sim N (\mu_t, \Sigma_t), \quad (3)$$

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

IV. ONLINE FEATURE SELECTION THROUGH SSVS

For regression problems with many features, proper shrinkage on the regression coefficients as in [12] is usually required to avoid over-fitting. For instance, two common shrinkage methods are L2 penalty (ridge regression) and L1 penalty (Lasso) as in [21]. Experts often want to monitor the importance of rules so that if any adjustments are required they can modify it for effective use. By this experts as in [11] can add new rules or change rules. However, the fraudulent sellers change their behavioural pattern quickly: some rule-based features that does not help today might help a lot tomorrow. For that it is necessary to build an online feature selection framework and intuition as in [16].

At time t , let α_{jt} be the j -th element of the coefficient vector αt . Instead of putting a Gaussian prior on α_{jt} , the prior of α_{jt} now is as in [17]

$$\alpha_{jt} \sim p_{0it} \mathbf{1}(\alpha_{jt} = 0) + (1 - P_{0it}) N (\mu_{jt}, \sigma_{jt}^2), \quad (4)$$

Where p_{0it} is the prior probability of α_{jt} being exactly 0, and with prior probability $1 - P_{0it}$, α_{jt} is drawn from a Gaussian distribution with mean μ_{jt} and variance σ_{jt}^2 . Such prior is called the "spike and slab" as in [18] but how to embed it to online modelling has never been explored before.

V. MULTIPLE INSTANCE LEARNING

In the modern system the procedure of expert labelling is in a bagged fashion as in [13] i.e. when a new labelling process starts, an expert picks the most suspicious seller in the queue and looks through all his/her cases posted in the current batch; the expert determines if any of the cases had been found to be fraud, then all the cases from this seller are labelled as fraud. In these types of scenarios they are to be handled by "multiple instance learning" as in [2].

Suppose for each seller i at time t there are K_{it} 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_{lit} . The multiple instance learning model can be written as

$$y_{it} = \mathbf{0} \text{ iff } z_{lit} < 0, \forall l = 1, \dots, K_{it} \quad (5)$$

Otherwise

$$y_{it} = 1, \text{ and } z_{lit} \sim N (x'_{lit} \alpha t, 1); \quad (6)$$

Where αt can have any types of priors that are described in Section 2.1 (Gaussian), Section 2.2 (spike and slab).

VI. EXPERIMENTAL WORK

Hardware system configuration with minimum requirements are processor of Pentium-III, speed of 1.1 GHz, RAM 256MB, Hard Disk of 20 GB, Floppy Drive of 1.44 Mb. Software system configuration with minimum requirements are

Operating system with Windows 95/98/2000/XP, Application Server of Tomcat 5.0/6.X, Front end of HTML, JAVA, JSP and server side script with JSP and database of MySQL and Database connectivity is of JDBC. Online auction is always recognized as an important issue. Websites extensively uses reputation systems and high end software's although many of websites use native approach.



Fig.1 Home Page

The Home page consists of various products (as shown in Fig.1) which are to be sold by the seller. The Administrator will authorize and allow the products which are to be displayed on the online website.

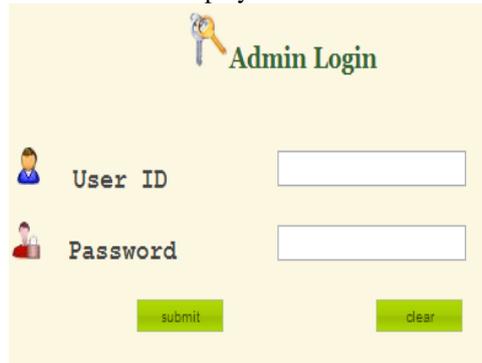


Fig.2 Admin Login

Administrator will login with id and password (as shown in Fig.2) 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.

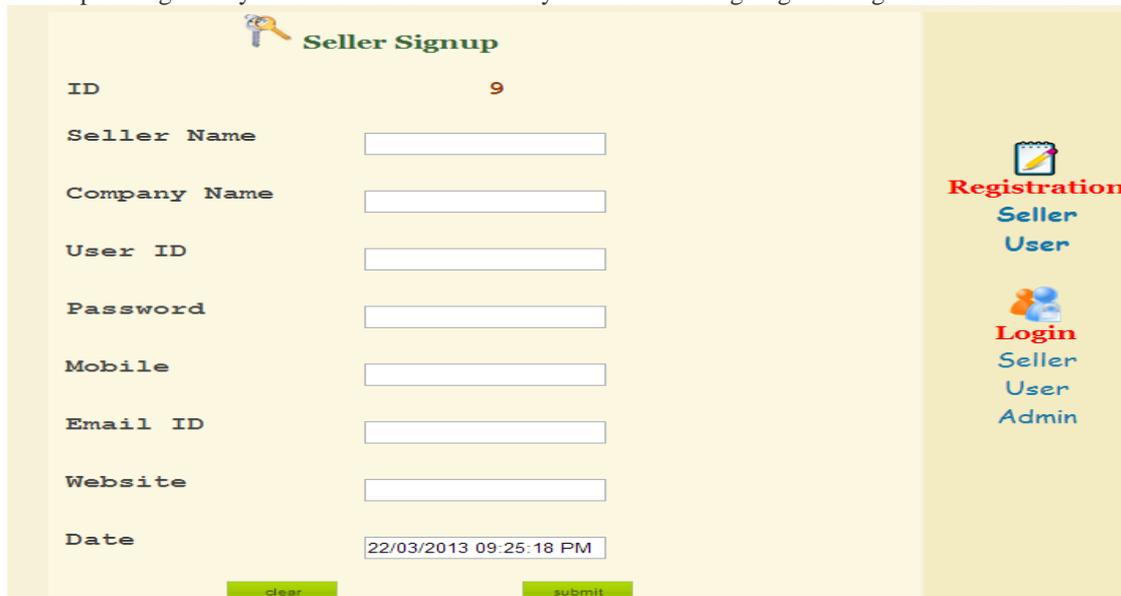


Fig.3 Seller Signup

Seller signup has to be filled up by the seller to sell there products on website (as shown in Fig3). Seller has to choose unique user id and password. These details are stored in Admin's database.

Fig.4 Untrusted Sellers

If the seller is found to be fraudulent their access will be denied by the Administrator. If seller logs in with old id and password, he/she will be set as untrusted (as shown in Fig.4). They can only enter the product details but they are denied to display on website.

Purchase ID	Company Name	Product ID	Product Name	Warrenty date	Product Rate	Description	Complaint
20	soni	8	Pen Drive	06/09/2012 12:05:15	1800	mega offer	Complaint
21	Guptha&co	6	TV	11/09/2013 08:55:04	16000	mega offer	Complaint

Fig.5 Products and Description

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.

Fig.6 Offers and Trustability Percentage

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 (as shown in Fig.6).

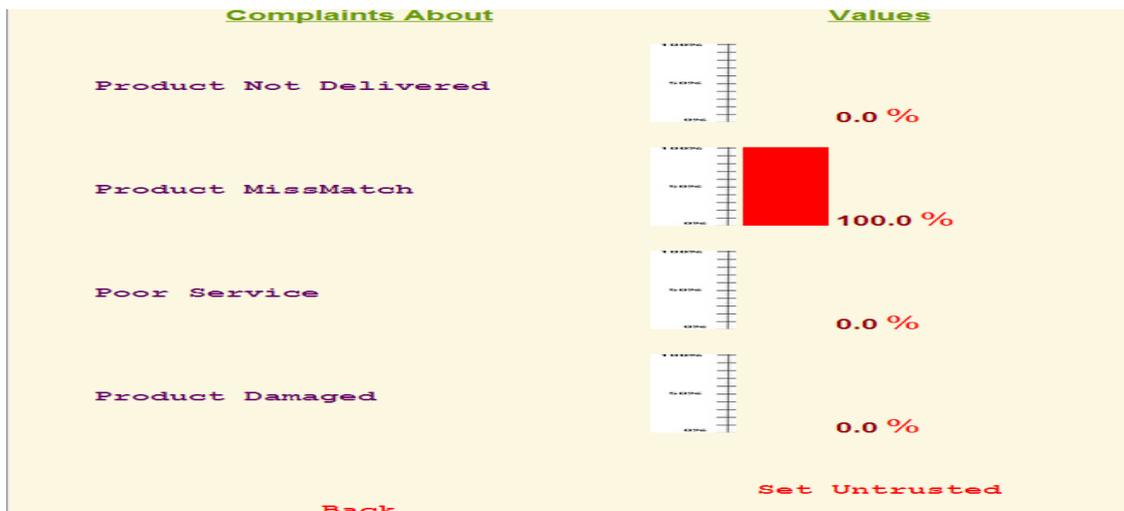


Fig.7 Complaints and Values

Complaints and values gives the details and there values such as product not delivered cases, product mismatches, poor services and product damage cases. (as shown in Fig.7) There values are displayed both in percentages and also in diagrammatic form.

Fig.8 Complaints

Complaints are to be filled by the user who bought the product (as shown in Fig.8). By entering the complaint seller will be able to see the complaint and will take measures to rectify the problem.

VII. CONCLUSION AND FUTURE WORK

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. 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] D. AGARWAL, B. CHEN, AND P. ELANGO. SPATIO-TEMPORAL MODELS FOR ESTIMATING CLICK-THROUGH RATE. IN PROCEEDINGS OF THE 18TH INTERNATIONAL CONFERENCE ON WORLD.
- [2] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. Advances in neural information processing systems, pages 577–584, 2003.
- [3] C. Bliss. The calculation of the dosage-mortality curve. Annals of Applied Biology, 22(1):134–167, 1935.
- [4] A. Borodin and R. El-Yaniv. Online computation and competitive analysis, volume 53. Cambridge University Press New York, 1998.
- [5] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
- [6] R. Brent. Algorithms for minimization without derivatives. Dover Pubns, 2002.
- [7] D. Chau and C. Faloutsos. Fraud detection in electronic auction. In European Web Mining Forum (EWMF 2005), page 87.

- [8] H. Chipman, E. George, and R. McCulloch. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298, 2010.
- [9] W. Chu, M. Zinkevich, L. Li, A. Thomas, and B. Tseng. Unbiased online active learning in data streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 195–203. ACM, 2011.
- [10] C. Chua and J. Wareham. Fighting internet auction fraud: An assessment and proposal. *Computer*, 37(10):31–37, 2004.
- [11] R. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1631–1643, 2005.
- [12] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. Cambridge university press, 2006.
- [13] T. Dietterich, R. Lathrop, and T. Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.
- [14] Federal Trade Commission. Internet auctions: A guide for buyers and sellers. <http://www.ftc.gov/bcp/online/pubs/online/auctions.htm>, 2004.
- [15] S.Pandit, D. Chau, S. Wang and C. Faloutsos. Net Probe: a fast and scalable system for fraud detection in online auction networks. In *proceedings of the 16th international conference on World Wide Web*, pages 201-210. ACM, 2007.
- [16] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* ,58(1): 267-288, 1996.
- [17] J.Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4) : 367-378, 2002.
- [18] H. Ishwaran and J.Rao. Spike and Slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*,33(2): /730-733 ,2005.
- [19] K. Kim. Financial time series forecasting using support vector machines. *Neuro computing*, 55(1-2):307-319, 2003.
- [20] D. Gregg and J. Scott. The role of reputation systems in reducing on-line auction fraud. *International Journal of Electronic Commerce*. 10(3):95-120, 2006.
- [21] USA Today. How to avoid online auction fraud. <http://www.usatoday.com/tech /columnist/2002/05/07/yaukey.htm>, 2002.