Developing Effective Fraud Detection Methods for Online Auction

Jau-Shien Chang
Dept. of Information Management
Tamkang University, Taipei, Taiwan
jschang@mail.tku.edu.tw

Yu-Hung Liu
Dept. of Information Management
Tamkang University, Taipei, Taiwan
clfdliu@hotmail.com

Ching-Fen Lee
Dept. of Business Administration
Tamkang University, Taipei, Taiwan
leecf@mail.tku.edu.tw

ABSTRACT

The past decade has witnessed the rapid growth of online auctions. However, the low cost and anonymity in joining online auctions provided an easy path for fraudsters. The simple binary reputation system promoted by the auction site is clearly not enough to protect consumers from fraud. In view of this, many fraud detection methods have been proposed. Nevertheless, there are still many weaknesses needed to be improved. To help secure the online trading environment, this study aims at developing more effective methods to identify the fraudsters in online auctions. First, a novel selection method is proposed for deriving a concise attribute set used to build efficient detection models, which allow a reduction in detection costs while improving detection accuracy. In addition, a two-stage detection procedure is proposed wherein multiple mutual-complement models are combined for promoting overall detection accuracy. To evaluate the proposed methods, actual auction transaction histories were collected for testing. The experimental results show that these methods can outperform those in the previous work.

Keywords: fraud detection, feature selection, classification, online auction, e-commerce

I. INTRODUCTION

In recent years, online auctions have become one of most successful business models in E-commerce [15]. Inevitably, the rapid growth of online auction trading will spur a subsequent increase in trade disputes with fraud being the most serious [26]. There are various types of fraudulent schemes, such as non-delivery, delivered of blemished commodities, and exaggerated item descriptions [17,18]. The low cost and anonymity in joining online auctions has provided an easy path for fraudsters. Therefore, auction houses provide simple binary reputation systems to help traders choose appropriate partners and reduce the chances of being defrauded. However, it has been proven that binary reputation systems are not effective in combatting fraudsters [19,20]. Fraudsters have developed various tricks to accumulate positive feedback scores to attract potential traders [17,18,26]. These schemes degrade the effectiveness of existing reputation systems and contribute to a general unease in trading [21]. Thus, researchers have proposed many detection methods to help traders identify fraudsters [2].

One of the most important steps of fraud detection is to select a set of measured attributes, from which an effective fraud detection model can be constructed. Various attribute sets describing the features of fraudsters have been proposed in previous research. Chau et al. [12-14] observed that the list price and trade volume of auctioned items tend to increase significantly in the last phase of a fraudster's active lifespan. Therefore, a set of price-related attributes was created from which classification trees were constructed to model these characteristics. To raise the accuracy of fraud detection, Chang and Chang [9] proposed a set of rating-related attributes with a density calculation. Their work attempted to describe the difference in feedback accumulation between fraudsters and normal traders. In fact, the trading relations in an auction site can be represented as a network in which complicated organized fraud could occur [28]. For organized fraud detection, researchers have developed various attributes to describe the trading network structure, trading information and personal information [13, 29-31]. To find collaborative frauds, Chau et al. [13] and Tseng et al. [29] used belief propagation to identify the accomplices in a trading network. When the trading network became too complicated, an approximate estimation method was used to reduce calculation time [14]. To fulfill early detection requirement, Chang and Chang developed the concept of multiple-phase modeling, in which the transaction histories of fraudsters were partitioned to illustrate behavioral features present during different periods of a fraudster's lifespan [9-11].

It can be seen from the above discussions that previous research has accomplished various predefined goals; however, there are still several weaknesses which need to be addressed: (1) First, to promote the detection accuracy, multitudinous attributes have been developed in the previous work to describe the characteristics of various types of fraudsters. However, more attributes implied more costs on data retrieval and data analysis. In addition, a large attribute set is not necessarily helpful in obtaining better detection accuracy [22]. Therefore, to develop a cost-effective fraud detection system, discovering a concise but effective attribute set is more important than extending the attribute set at will. (2) Secondly, the previous work inclines to apply a single, but powerful model for fraud detection. Such an approach could constrain the effectiveness of detection systems as fraudsters use complicated tricks to disguise their intentions.

To cope with the above problems, this study aims at developing effective methods to help construct a practical fraud detection system for online auction. First, a novel attribute selection method, RESF, is proposed for choosing a concise but effective attribute set, which can reduce the cost of detection while maintaining overall accuracy. In addition, a two-staged multimodel detection procedure is then developed to promote detection accuracy in which results of mutually-compensated detection models are combined to examine suspicious accounts. To validate the proposed methods, the transaction histories of Yahoo!Taiwan were used for testing. When compared to the results of previous work, the detection model built with RESF-

selected attributes was seen to outperform other methods in detection accuracy. When the proposed balanced detection procedure was used, the detection accuracy was further improved. In conclusion, the proposed methods actually provide a feasible way to build a cost-effective fraud detection system for online auctions.

The remainder of this article is organized as follows: Section 2 introduces the techniques and concepts related to our work. Section 3 proposes an iterative reselection method for devising a concise attribute set. Section 4 explains the proposed balanced detection procedure. Experimental results are given in Section 5. Conclusions are presented in the final section.

II. Preliminaries

To facilitate the subsequent discussions, related work and techniques for online auction fraud detection will be introduced in this section.

A. Fraud Detection

Fraud detection is important for securing the profits of business transactions. Basically, fraud detection is a kind of anomaly detection. Researchers generally use data mining approaches to help identify fraudulent behaviors. For instance, support vector machines and random forests have been applied to discover credit card frauds [5,6,7]. Ravisankar et al. use several data mining techniques, like neural network, genetic programming and support vector machines, to detect financial statement fraud [29]. Similarly, researchers also tried to compose several data mining methods with learning schemes to help identify evolutionary financial fraud [8,34]. For online auction fraud, researchers usually applied classification techniques to identify abnormal behavior in the transaction histories [10,12]. Many different feature sets for fraud detection have been proposed. Due to the length limits, the details of these features please refer to [12,16].

In general, there are two main steps for developing an anomaly detection method. First, a set of effective measured attributes needs to be determined. Next, based on the selected attribute set, an appropriate learning method is used to construct a detection model. A detection model for identifying abnormal instances could correspond to a decision tree [12], a Bayesian network [27], self-organizing maps [3], or an artificial neural network [31]. The accuracy of detection is directly affected by the selected attribute set and learning algorithms used. In practice, measured attributes can be directly extracted from existing data fields in transaction histories.

B. Measured Attribute Selection

Even though a fine-grained attribute set is indispensable for detection model construction, a massive diverse set of attributes is not necessarily helpful in increasing the overall accuracy of fraud detection. In fact, there could exist irrelevant and redundant attributes which are not ineffective for model construction but just increase the detection costs [22,33]. Thus, it is critical to select a compact attribute set to construct a cost-effective detection model. Attribute selection is a procedure to remove irrelevant, redundant or noisy attributes from a given attribute set [24,33]. The performance of the selected subset is expected to be similar to the original one or at least not worse than a predefined threshold. More formally, the problem statement of attribute selection can be described as follow:

Given a data set D, a attribute set S, a learning algorithm L, and a performance evaluation function E, attribute selection is a procedure to find a small subset from $\{S' \mid S' \subset S, E(S',L,D) \geq E(S,L,D)\}$, where $E(S,L,D)$ represents the performance of building a detection model by L with S for classifying the instances in D.

Obviously, the time complexity of finding an optimal solution for the above problem would be $O(2^{|S|})$, which is intractable for a large S. Thus, different heuristic methods have been proposed for discovering a small S' with a reasonable $E(S',L,D)$. There are two general approaches for attribute selection, which are the filter approach [4,23,33] and the wrapper approach [24]. The basic concept of the filter approach is to select useful attributes by ranking these attributes with an evaluation function. In addition, redundant attributes can also be removed by applying more elaborated selection rules. Because no leaning algorithm is involved in the filter approach, it can be performed quite efficiently and can prevent form the over-fitting problem. However, when a learning algorithm is chosen for model construction, the performance of the filter-based attribute set could be not as expected as that indicated by the scoring function values. For the wrapper approach, the effectiveness of a selected attribute subset is evaluated by a predetermined learning algorithm. Thus, the resultant attribute set can match this learning algorithm very well. However, the whole process would be very time-consuming and may result in the problem of over-fitting [24,25].

III. A Novel Filtering Method with Reselection for Attribute Selection

In this section, the proposed attribute selection method will be introduced. A modified filter method with reselection has been developed to generate a concise but effective attribute set. As described in Section 2, to construct an effective detection model, a suitable attribute set should be determined first. And, a compact attribute set would be preferable to a large one. The reason is obvious because less attributes implies less computation effort in data retrieval and data analysis. To this end, this study proposes a modified filter method with reselection to discover a small but effective attribute set.

The conventional filter methods for attribute selection can be performed in an efficient way. However, several disadvantages could be incurred by such a concise procedure. To help explain the details, assume we have an attribute set FS, a filtering method FM, and a scoring function E for evaluating the result of filtering. Let $F = FS \cup FD$, where FS is the

attribute subset selected by *FM*, and $FS \cap FD=\varnothing$. In addition, assume *C* is the target attribute for classification, whose value is either 'Fraud' or 'Non-Fraud' in this study.

(1) In the filter approach, to remove irrelevant or weakly irrelevant attributes, only those attributes in *F* with *E* values larger than a predefined threshold are retained. However, considering the case that the phased partitions [10] are used to shrink the transaction histories in advance, only very few attributes will survive by such a filter procedure. This is due to fewer collected data resulting in small *E* values, and thus causing a lot of attributes dropped before further processing. In practice, a too small attribute set would not be beneficial in characterizing complicated fraudulent behavior. This means that the performance of a detection model constructed using these selected attributes will be inferior.

(2) Not all the attributes passing the threshold filtering is necessary for classification. In fact, some of them are redundant and could be removed from the attribute set. An attribute F_j is declared as redundant if it becomes useless in the case of including another attribute F_i in the attribute set. However, if both $E(F_i)$ and $E(F_j)$ are quite small and contribute less to the classification of *C*, it is too arbitrary to remove F_j simply because F_i is already selected in the attribute set.

(3) Individually irrelevant or redundant attributes are not necessarily useless for classification if they collaborate with other attributes. In fact, as described in [24], presumably redundant attributes may contribute to better class separation and two attributes that are useless by themselves can be useful together.

From the above discussions, it is clear that these removed attributes should be carefully examined for further promoting the effectiveness of attribute selection. In view of this, this study proposes a novel **F**iltering method with **Res**election (RESF), in which those attributes in *FD* are reconsidered by different evaluation processes. And, the Fast Correlation-Based Filter algorithm [33] (FCBF hereafter) is chosen as the underline filter-based attribute selection method.

The proposed RESF method comprises three sub-procedures: *Basic*, *Reselection-by-Cover***,** and *Reselection-by-Information-Gain*. Given a candidate attribute set *G*, a union of classification attributes (*F*), and target class (*C*) ($G=F\cup C$), the details of these steps are as follows:

(1) **Basic**: First, the modified FCBF is performed on *F* (denoted by FCBF(*F*,*C*)), from which a selected attribute set, $FS=\{FS_1, FS_2, \ldots, FS_m\}$, is obtained. For each $F_i \in FS$, the set Cover(F_i) denotes those attributes removed which result from applying the second deletion rule on F_i. Thus, the removed attribute set can be represented as $FD=F\text{-}FS=\cup\text{Cover}(F_i)$. Different from the traditional filtering method (such as FCBF), *FD* is not discarded immediately in RESF. Instead, a chance to reselect members of *FD* into *FS* if they contribute to classification accuracy is given. For this purpose, two reselection procedures are performed subsequently.

(2) **Reselection-by-Cover**: In this step, every Cover(F_i) will be refined for re-selection by performing FCBF(Cover(F_i),F_i). Next, the result of reselection on each Cover(F_i), is added to the reselected set *FR1*. The purpose of this step is to find members of Cover(F_i) which are highly correlated with F_i and possibly helpful in raising the accuracy of classification.

(3) **Reselection-by-Information-Gain**: In this step, attributes in *FD* will be reconsidered by checking the correlations between the attribute with the highest information gain (denoted as *H*) and those deleted attributes. To this end, attribute *H* is used as the target for classification. Next, FCBF(*FD*, *H*) is performed to extract attributes highly correlated to *H*. Then, the obtained attribute set will be added to *FR2*.

(4) Finally, the resultant attribute set *TRS*, the union of *FS*, *FR1* and *FR2*, i.e., $TFS=FS\cup FR1\cup FR2$, is determined.

IV. A Balanced Detection Procedure with Multiple Detection Models

In practice, a single powerful detection model may be ineffective in detecting various types of fraudsters. Thus, a Balanced Detection Procedure (BDP) is proposed in this study to take advantage of multiple detection models for fraud detection.

Based on the selected measured attribute set, the proposed detection procedure consists of two stages in which three different detection models are applied. In the first stage, two reciprocal detection models are used to construct a conceptual model using account transaction histories from the past 90 days. The *M1* model can precisely identify fraudsters, while the *M2* model accurately identifies normal traders and serves as the companion for *M1*. In the second stage, model *M3* is built with balanced capabilities for detecting normal traders and fraudsters.

To facilitate the discussions on the balanced detection procedure, some notations are first introduced below. For a set of account *S* to be tested, F_S and NF_S denote the sets of identified fraudsters (F) and non-fraudsters (NF). Therefore, *S* represents the union of F_S and NF_S. If a model *M* is applied to classify S, $F_S(M)$ and $NF_S(M)$ are used to denote the sets of fraudsters and non-fraudsters classified by *M* respectively. Based on these notations, the balanced detection procedure can be described as follows (the flowchart of BDP is given in Figure 1) :

1. Given a set of account *S* to be tested, collect the transaction data from the last *d* days of each member in *S*.
2. Test *S* using *M1* and *M2* to obtain $F_S(M1)$, $NF_S(M1)$, $F_S(M2)$ and $NF_S(M2)$.
3. let $F1=F_S(M1)-NF_S(M2)$,
   $NF1=NF_S(M2)-F_S(M1)$, and
   $S2=\text{S-}F1\text{-}NF1$.

4. Test *S2* using *M3* to obtain $F_{S2}(M3)$ and $NF_{S2}(M3)$
5. let the final result be $F_S = F1 \cup F_{S2}(M3)$ and $NF_S = NF1 \cup NF_{S2}(M3)$

The design principle of the above procedure is based on the concept of mutual-compensation, a process of applying multiple detection models with specific characteristics used to classify suspicious accounts. Through application of *M1* and *M2*, $F_S(M1)$ and $NF_S(M2)$ can be obtained with high accuracy. That is, members in $F_S(M1)$ are in all likelihood actual fraudsters, while members in $NF_S(M2)$ are truly normal traders. As a result, *M1* and *M2* work in unison to filter out the cases in which an account is deemed either a fraudster or normal trader. However, some contradictory results may occur using *M1* and *M2*. In certain cases, an account may be classified as a fraudster by *M1*, and simultaneously as a non-fraudster by *M2*. To deal with such cases, members in $F_S(M1) \cap NF_S(M2)$ are treated as uncertain cases and fed into Stage 2 which uses the *M3* model for a final decision. Because *M3* is an unbiased detection model, all uncertain cases from Stage 1 will be re-examined by *M3* to decide their final classification (i.e., F or NF).

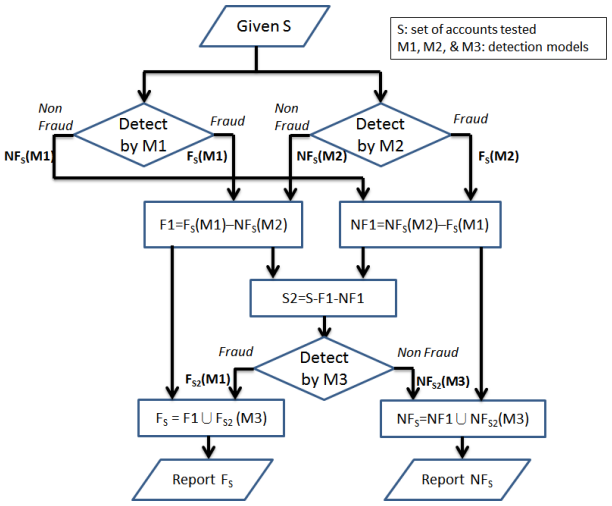

Figure 1：The balanced two-stage multi-model detection procedure

The model M1, M2, and M3 is built as follows: for model M1, the ratio of normal users to fraudsters in the training set was set at m to 1, where m>2. A high imbalanced ratio for NF to F (say10:1) in the training set is beneficial for building a detection model that can provide a better fraudster detection precision. Similarly, the M2 model is also built with an imbalanced ratio of NF:F at 1:n in the training set, where n≥2, which allows for higher precision in identifying normal traders. The M3 model is constructed with a ratio of normal users to fraudsters set at 2:1 and is obviously has a balanced capability of differentiating normal users from fraudsters.

V. Experimental Results

In this section, to validate the effectiveness of these detection methods proposed in this study, real transaction histories from Yahoo!Taiwan were acquired for testing. The methods of both the measured attribute selection and the balanced detection procedure are examined respectively below. The experimental results will be compared with the outcomes of other methods proposed in related work.

A. Experimental settings

The authors collected transaction histories of 1,115 normal traders and 514 proven fraudsters from the Yahoo!Taiwan auction site. All collected data were parsed into values corresponding to measured attributes for each account. And, the Weka's implementation of C4.5 classification algorithm is used to build the detection model. The evaluation metrics of the following experiments are based on the following four indices: True Positive (TP) denotes the number correctly identified as positive ; False Positive (FP) is the number identified as positive but actually is negative; False Negative (FN) is the number identified as negative but actually is positive; True Negative (TN) the number correctly identified as negative.

A. Attribute selection algorithm validation

The performance of RESF is validated in the following section. The candidate set of measured attributes is consisted of 52 attributes from Chau's work[12], Chang's work[10] and those in Cheng's work[16]. And, the C4.5 tree classification algorithm is used to build the detection model. To demonstrate the effectiveness of the attribute set selected by RESF, the evaluation results are compared with those by applying the compound candidate attribute set (including all the 52 attributes) and those selected by the FCBF[33], Relief and CfsSubset. ReliefF is an attribute selection method that can operate on both discrete and continuous class data. It evaluates the value of the given attribute for the nearest instance of the same and different class. The CfsSubset selection method estimates the value of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. In the following experiments, we use the Weka's implementation [32] for ReliefF and CfsSubset.

The ratio of normal traders to fraudsters was set at 2:1 in each training set and test set. Training sets contained 700 normal trader profiles and 350 fraudster profiles, while test sets contained 300 normal trader profiles and 150 fraudster profiles. Instances were chosen randomly from the collected data for both training and test sets. The training set and the test set were mutually exclusive. In total, 10 trials were performed and the experimental results were averaged from the outcomes of the same. In practice, scarcity of available data occurs in real-world applications and could affect the capability of detection models. This scarcity of data is due to the fact that auction houses may not maintain complete transaction records. Therefore, to validate the effectiveness of various attribute sets for limited transaction data, the researchers used 90 days as constraints to partition transaction histories and construct detection models using different measured attribute sets for comparison.

Table 1：Comparisons of different attribute selection methods for fraud detection assumed that complete transaction histories are available for analysis (d=∞)

Training set: Normal traders 700; Fraudster 350, Testing set: Normal traders 300; Fraudster 150							
Attribute Set	No. of Attribute	Detection Accuracy	Non-Fraud/Fraud (NF/F), where d=∞				
			TP Rate	FP Rate	Precision	Recall	F-Measure
Compound[1]	52	0.849	0.905/0.737	0.263/0.095	0.873/0.796	0.905/0.737	0.888/0.764
FCBF	21	0.857	0.914/0.743	0.257/0.086	0.877/0.812	0.914/0.743	0.895/0.775
ReliefF[2]	18	**0.862**	0.914/0.757	0.243/0.086	0.883/0.815	0.914/0.757	0.898/0.784
CfsSubset[2]	21	0.856	0.902/0.763	0.237/0.098	0.884/0.796	0.902/0.763	0.893/0.779
RESF[3]	**13**	0.858	0.914/0.747	0.253/0.086	0.879/0.817	0.914/0.747	0.896/0.778

[1] The Compound attribute set is consisted of 52 attributes collected from three attribute sets proposed in different work.
[2] Weka's implementation of the two algorithms are used for experiments.
[3] the proposed attribute selection method of this study.

Table 2：Comparisons of different attribute selection methods for fraud detection assumed that only transaction histories in the last 90 days are available (d=90)

Attribute Set	No. of Attribute	Detection Accuracy	Non-Fraud/Fraud (NF/F), where d=90				
			TP Rate	FP Rate	Precision	Recall	F-Measure
Compound	52	0.805	0.845/0.725	0.275/0.155	0.861/0.705	0.845/0.725	0.852/0.712
FCBF	14	0.802	0.844/0.719	0.281/0.156	0.858/0.698	0.844/0.719	0.851/0.708
ReliefF	16	0.813	0.855/0.729	0.271/0.145	0.863/0.716	0.855/0.729	0.859/0.722
CfsSubset	18	0.822	0.860/0.746	0.254/0.140	0.873/0.734	0.860/0.746	0.865/0.736
RESF	**12**	**0.822**	0.863/0.741	0.259/0.137	0.871/0.735	0.863/0.741	0.866/0.734

Table 3：Comparisons of the proposed BPD procedure with single detection model

BPD procedure proposed by this study (referring to Section 4)						
Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.835	0.902/0.702	0.298 /0.098	0.858 /0.783	0.902/0.702	0.879/0.739	NF/F
RESF selected attributes (single detection model, d=90)						
Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.822	0.863/0.741	0.259/0.137	0.871/0.735	0.863/0.741	0.866/0.734	NF/F

For the case of complete transaction histories available for analysis (denoted as d=∞), the experimental results are shown in Table 1. It can be seen that the accuracy of applying the Compound attribute set (0.849) is lower than those of all other selection methods. This demonstrates that a large attribute set is not necessarily more effective for detection than a smaller one. In other word, it is possible to devise a concise attribute set which contains less attributes and incurs less effort for detection. In this case, the detection success rate of RESF selected attributes is 0.858, which is better than other methods except for RelifF (0.862). However, only 13 attributes are selected by RESF but 18 are needed for RelifF.

For the case of only the transaction histories for the last 90 days available for download, referring to Table 2, the performance of all attribute sets are all degraded. It is reasonable because less transaction data give less information for fraud detection. In this case, RESF obtains the highest success rate (0.822) and a smallest attribute set (only 12 attributes are needed for detection). Although CfsSubsetEval also achieves the same detection accuracy (0.822) as RESF, however, it needs 18 attributes to build the detection model. The above results demonstrate that, in comparison with the previous researches, RESF can consistently select a more compact attribute set and achieves better detection accuracy.

B. Validation of Balanced Detection Procedure

In this section, a validation of the performance of the proposed Balanced Detection Procedure (BDP) by applying multiple detection models in two stages is expressed below. Each of those detection models in BDP is constructed by the C4.5 classification algorithm. Referring to Section 4.2, the ratios of normal users to fraudsters for detection model M1, M2 and M3 are set at 10:1 (700 normal users to 70 fraudsters), 2:1(350 normal users to 175 fraudsters). All training sets and test sets contained only transaction histories occurring within

the last 90 days. Table 3 presents the averaged results of applying BDP in the 10 trials. The success rate of BPD was 0.835, superior to the best single model detection methods (0.822 by RESF attribute set) shown in Table 2. It can be seen that compared to RESF with *d*=90, this improvement mainly comes from the promotion of detection precision for fraudsters, which is 0.783 versus 0.735. This also results in an improvement in the recall rate for normal traders (from 0.863 to 0.902). In conclusion, the result shown in Table 3 demonstrates that the multimodel-based BDP method does have the potential to further improve fraud detection accuracy and achieve satisfactory performance.

VI. Conclusions

Online auctions have become one of the most popular platforms in e-commerce. To further extend the reach of online auction, more elaborate designs are needed to facilitate trading procedures. In particular, safety of users in completing a transaction should be carefully considered by auction house authorities. In addition to applying cryptography to secure the trading process, uncovering fraudsters before they strike is undoubtedly a priority. To this end, this study proposed a novel attribute selection method, RESF, to generate a concise attribute set to build effective detection models. It was observed that a small but effective attribute set can protect consumers from fraud at lower costs. Subsequently, a new detection procedure, BDP, has been developed to further improve the effectiveness of detection. BDP applies three detection models in two stages to carefully examine each account. The experimental results show the effectiveness of these proposed methods. With the help of an effective fraud detection system, the traders can avoid money loss and relieve to trade in the auction site.

References

[1] Abbasi, A., Lau, R. Y.K., Brown, D. E., "Predicting Behavior," IEEE Intelligent Systems, May/June 2015, pp. 35-43.

[2] Abdallah, A., Maarof, M. A., Zainal A., "Fraud detection system: A survey," Journal of Network and Computer Applications, vol. 68 (2016) 90-113.

[3] Almendra, V., et al., "Using Self-Organizing Maps for fraud prediction at online auctions sites", in Proceedings of 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 2014, pp. 281-288.

[4] Almualim H. and Dietterich T.G., Learning with many irrelevant features, In Proceedings of AAAI-91, Anaheim, CA, 1991, 547-552.

[5] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C., Data mining for credit card fraud: A comparative study, Decision Support Systems 50 (2011) 602–613.

[6] Carneiro, N., Gigurira, G., and Costa, M., "A Data Mining Based System for credit-card fraud detection in e-tail," Decision Support Systems 95 (2017) 91-101.

[7] Kumar, M.-S. et al., "Credit Card Fraud Detection Using Random Forest Algorithm," IEEE 3rd International Conference on Computing and Communication Technologies (ICCCT), 2019, p. 149-155.

[8] Xuan, S., et al., "Random Forest for Credit Card Fraud Detection," IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018.

[9] Chang, W.H., & Chang, J. S., Using Clustering Techniques to Analyze Fraudulent Behavior Changes in Online Auctions, In Proceedings of 2010 International Conference on Networking and Information Technology, Manila, Philippine, June 11-13, 2010.

[10] Chang, W. H., & Chang, J. S., A Novel Two-Stage Phased Modeling Framework for Early Fraud Detection in Online Auctions, Expert System with Applications, vol. 38, no. 9, (2011) 11244-11260.

[11] Chang, W. H., & Chang, J.S., An Effective Early Fraud Detection Method for Online Auctions, Electronic Commerce Research and Applications, Jul.-Aug, (2012) 346–360.

[12] Chau, D. H., & Faloutsos, C., Fraud Detection in Electronic Auction, In Proceedings of European Web Mining Forum at ECML/PKDD, October 3-7, 2005.

[13] Chau, D. H., Pandit, S., & Faloutsos, C., Detecting Fraudulent Personalities in Networks of Online Auctioneers, In Proceedings of PKDD. Sep. 18-22, 2006, pp. 103-114.

[14] Chau, D. H., Pandit, S., Faloutsos, C., & Wang, S., NetProbe: A fast and scalable system for fraud detection in online auction networks, In Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 201-210.

[15] Chen, J., et al., "Big Data based fraud risk management at Alibaba," The Journal of Finance and Data Science 1 (2015) 1-10.

[16] Cheng, H.-R., Effective Detection Methods for Latent Fraudsters in Online Auctions, Master Thesis, Dept. of Information Management, Tamkang Univ., Taiwan, Jun, 2011.

[17] Chua, C. & Wareham, J., Fighting Internet Auction Fraud: An Assessment and Proposal. IEEE Computer. Vol. 37(10), 2004, pp. 31-37.

[18] Criminal Investigation Bureau (CIB), Taiwan. The latest criminal trick awareness guidance, https://www.cib.gov.tw/crime, (accessed 2020.08.25).

[19] Dellarocas, C., Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems, International conference on Information Systems, Dec, 2000.

[20] Dellarocas, C., Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms, ACM EC'01, Oct. 14-17, 2001, pp. 171-179.

[21] Dellarocas, C. & Wood, C. A., The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias, Management Science, 54(3), March, (2008) 460-476.

[22] Guyon, I. & Elisseeff, A., An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, vol. 3, (2003) 1157-82.

[23] Kira, K. and Rendell, L. A., A practical approach to feature selection, in Proceedings of 9th International Conference on Machine Learning, Aberdeen, Scotland, 1992.

[24] Kohavi, R., & George, H. J., Wrappers for feature subset selection, Artificial Intelligence (1997) 273-324.

[25] Kononenko, I., "Estimating attributes: analysis and extension of relief," in ECML'94: Proceedings of the 7th European Conference in Machine Learning. Springer-Verlag, 1994, pp. 171–182.

[26] National White Collar Crime Center (NW3C). 2019 Internet Crime Report, https://pdf.ic3.gov/2019_IC3Report.pdf

[27] Panda, M., & M. R. Patra, Network Intrusion Detection Using Naïve Bayes. International Journal of Computer Science and Network Security, VOL.7 No.12, Dec, (2007).

[28] Shen, Z. & Sundaresan, N., eBay: an E-commerce marketplace as a complex network, In Proceedings of the fourth ACM international conference on Web search and data mining, 2011.

[29] Tsang, S., et al., "SPAN: Finding collaborative frauds in online auctions," Knowledge-Based Systems Volume 71, November 2014, Pages 389-408.

[30] Wang, J., & Chiu, C. Q., Detecting online auction inflated-reputation behaviors using social network analysis, In Proceedings of the NAACSOS conference, June 26-28, 2005.

[31] Wang, J.-C., Chiu, C.-C., & Ker, H.-Y. Ker, Detecting Online Auction Fraud of Reputation Inflation through Social Network Structures Embedded in Transaction Records (in Chinese). Journal of Information Management (ISSN:1608-5752), vol 12(4), (2005) 143-184.

[32] Witten, I. H., Frank, E., & Hall, M. A., Data mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2011.

[33] Yu, L., & Liu, H., Efficient feature selection via analysis of relevance and redundancy. The Journal of Machine Learning Research, Vol. 5, (2004) 1205-1224.

[34] Zhou, W., & Kapoor, G., "Detecting evolutionary financial statement fraud", Decision Support Systems 50 (2011) 570–575.