RESEARCH ARTICLE OPEN ACCESS

CO DETECT: FINANCIAL FRAUD DETECTION SYSTEM BASED ON BEHAVIOR CERTIFICATE (BC)

V.Santhi [1], Mr. A.Senthil Kumar [2]
1(Mphil (Research Scholar), Tamil University, Thanjavur, Tamilnadu, India)
2 (Assistant Professor, Department of Computer Science, Tamil University, Thanjavur, Tamilnadu, India)

Abstract:

In today's world frauds are increasing in both regular purchasing and online shopping. More increases in online shopping. For every transaction bank should check whether the transaction is fraudulent or not. Fraudster uses various techniques to generate the fraudulent transactions. Nowadays online transactions are more fraudulent than offline transactions. Due to increase in frauds more advance fraud detection system needed the most. Banking mostly use data mining techniques for credit card fraud detection. In this paper we propose a new credit card fraud detection system based on Behavior Certificate (BC) which reflects the cardholders' transaction habits using Random Tree data mining technique.

Keywords — **fraudulent, data mining technique, online shopping, Behavior Certificate, Random Tree.**

I. INTRODUCTION

Now a day the modes of payment method are changed into online communication. Banking scheme provides different type of payments like e-cash, card payments, internet banking, and eservices for improving online transaction. Credit card is one of the most custom ways of online transaction. Credit card is medium of selling goods or services without having cash in hand. With increased number of such cashless transaction, number of fraudulent transactions also increasing. During the online transaction we do not need any physical card, we need only card number, cvv, expiry date so there is more chances of fraud should be happen. In this method of fraud detection we generate behavior certificate on the basis of cardholder's transaction habits. Most of the credit card fraud detection methods based on anomaly detection try to extract the historical behavior patterns as rules and compute the similarity between an incoming transaction and these behavior patterns. The main idea of this kind of approach is that people may have personalized transaction habits that depend on their different accounts, different income sources, and different motivations and so on.

II. LITERATURE REVIEW

Chawla et al proposed Synthetic Minority Over-Sampling Technique (SMOTE). It is an over-sampling approach in which the minority sample is over sampled by creating synthetic samples rather than by oversampling with replacement. The minority class is over-sampled by taking each minority class sample and introducing synthetic samples along the line segments joining any/all of the k minority class' nearest neighbours. Depending upon the amount of over-sampling required, neighbours from the k nearest neighbours are randomly chosen. This approach effectively forces the decision region of the minority class to become more general.

Wael et al proposes a fraud engine based on an Artificial Neural Network (ANN). It was implemented on a Smart card in order to assess the performance and the general feasibility of this approach. The motivation was that, the intelligent behaviour-based security mechanisms can provide added protection for critical systems. Of particular interest is the real-time detection and reaction to

fraudulent behaviours. Any suspicious or unusual activities are captured and prevented instantly. With real artificial intelligence implemented using Neural Networks, behaviour based security mechanisms promise to be at the same step as the attacker and not a step behind. Using this kind of approach to security brings it down to the personal level, meaning fraud should be detectable for every single user or customer depending on his usage characteristics.

D.Nali et al.proposes a universal infrastructure and protocol for IDF detection, which is called CROO (Capture Resilient Online One-time password scheme). According to this proposal, each user must carry a personal device used to generate One-Time Passwords (OTPs) verified by online trusted parties. These OTP generation and verification procedures are universal, in the sense that they can be associated with any user transaction, regardless of the transaction's purpose (e.g. user identification, user authentication, or financial payment), associated credentials (e.g. driver's license or credit card), and online or on-site (e.g. point-of-sale) nature. For increased scalability, multiple OTP verification parties may be used. OTPs are not sent in clear text; they are used as keys to compute MACs of hashed unique transaction information (e.g. list of bought items). This allows OTP-verifying parties to confirm that given user credentials (i.e. OTP-based MACs) correspond to claimed hashed transaction details. Hashing transaction information increases user privacy. Online OTP-verifying parties detect IDF when OTPs of received user credentials or the associated transaction information do not have expected values.

Vapnik describe a training algorithm for optimal margin classifiers in which they showed that maximizing the margin between training examples and class boundary amounts to minimizing the maximum loss with regards to the generalization performance of the classifier. This idea was initially explored because binary class optimal margin classifiers achieve errorless separation of the training data, given that separation is possible, and outliers are easily identified in such a classifier.

Cortes and Vapnik extended the algorithm to account for linearly inseparable data; attempts soon followed to also extend the results to multiclass classification problems. This type of learning machine was later dubbed the Support Vector Machine (SVM). The SVM is a machine learning technique with a strong and sound theoretical basis. It is interesting to note that, in most cases, researchers claim that SVMs match or outperform neural networks in classification problems. Methods like Time-Delay Neural Networks, Plate's method and NARX networks are examples of learning algorithms that are applicable to short time gap problems only.

Chen et al. proposed a novel approach to solve the fraud problem. They proposed to develop a personalized model to detect fraud. One can first gather personal transaction data of users by an online, self-completion questionnaire system. The gathered Questionnaire-Responded Transaction (QRT) data from the online system are considered as the transaction records and are utilized to build up a personalized model, which is sequentially in use to predict whether a new, actual transaction is a fraud or not. Since the illegal user's consumer behavior is usually dissimilar to the cardholder, the fraud can be avoided from initial use of a credit card, even without any transaction data. The QRT approach is promising. However, there are still some problems needed investigating. One of the most important issues regarding the QRT approach is how to predict accurately with only few data, say 100 to 200, since the users are usually not willing to answer too many questions. In this paper, the influences of data number and data distribution on the prediction accuracy are examined. Several typical ways for improving the prediction accuracy are also employed to study their influences. Figure 2.2 represents the QRT approach to fraud detection.

Ren proposes an efficient outlier detection method with clusters as by product, which works efficiently for large datasets. The contributions include: a) introducing a Local Connective Factor (LCF); b) Based on LCF, an outlier detection method which can efficiently detect outliers and group data into clusters in a one-time process is proposed. This method does not require the beforehand clustering process, which is the first step in other state-of-the-art clustering-based outlier detection methods; c)

The performance of this method is further improved by means of a vertical data representation, Ptrees.

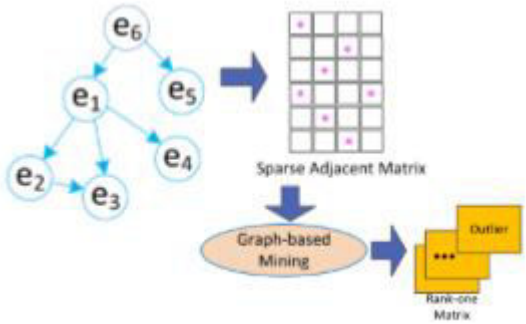
Manzoor et al.introduces the cluster-based local outlier factor to identify and find the outliers in dataset. It proposes an outlier detection algorithm based on spectral clustering. Experimental results show that it outperforms the traditional K-means based algorithm.

Syeda et al have used parallel granular neural network for improving the speed of data mining and knowledge discovery process for credit card fraud detection. But it could achieve reasonable speed up to 10 processors only, more number of processors introduces load imbalance problem.

Chiu et al have proposed web services based collaborative scheme for fraud detection in the banking industry. The proposed scheme supports the sharing of knowledge about fraud pattern with the participant banks in a heterogeneous and distributed environment. Phua et al have done an extensive survey of existing data mining based FDSs.

III. IMPLEMENTATION

Frauds are increasing in both regular purchasing and online shopping. More increases in online shopping. For every transaction bank should check whether the transaction is fraudulent or not. Fraudster uses various techniques to generate the fraudulent transactions. Nowadays online transactions are more fraudulent than offline transactions. Due to increase in frauds more advance fraud detection system needed the most. Banking mostly use data mining techniques for credit card fraud detection.



A. Random tree

Random Tree is the supervised classifier Random Tree is use to construct the random set of data for constructing decision tree. Random Tree algorithm deals with classification and regression problems. Random Tree is the group of tree predicators that called as forest. In Random Tree, classifier get input feature vector and classify it with every tree in forest and output of class label received majority votes. In regression, the classifier reply average of responses over tree in forest. Random Tree is the combination of two algorithms from machine learning. Single Model Tree combines with Random Tree to improve the functioning of Random Tree. Single Model Tree is the decision tree in which leaf node hold linear model. Random Tree improve the performance of decision tree. Random Tree is reasonably balance tree [6]. In this Random Tree one global setting works across the all leaves and thus simplifying optimization procedure. This feature of Random Tree reduces the time and efforts. Random Tree produce slightly better classification accuracy than Random Forest. Random Tree yielded the highest accuracy value of 94.32% [6]

B. Credit card fraud detection system

We have to perform FDS process in this section to detect fraud in credit card. FDS consist of two components BC construction and fraud detection. These two components are connected with each other using database. The BC construction process carried out offline while fraud detection process carried out online.

In FDS, we have to find Behavior Certificate using BFV (Behavior Feature Vector). BFV which consist of 13 dimensions which describes cardholder transaction behavior.

BFV = (Weekday, Weekend, Festival, Normal Day, Interval1, Interval2, Interval3, Interval4, Location, Range1, Range2, Range3, Range4) is a behavior feature vector, where “Weekday” and “Weekend” are to represent whether a transaction take place in weekday or weekend.

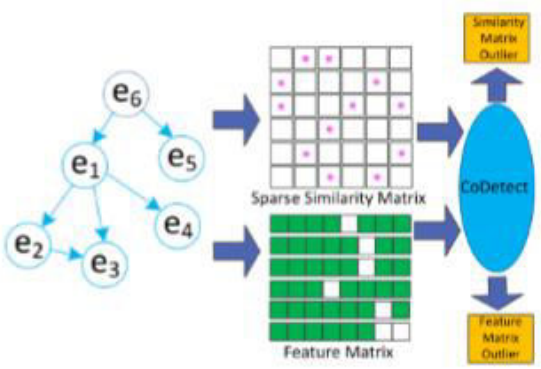
 “Festival” and “Normal Day” are to represent whether a transaction take place in festival or not.

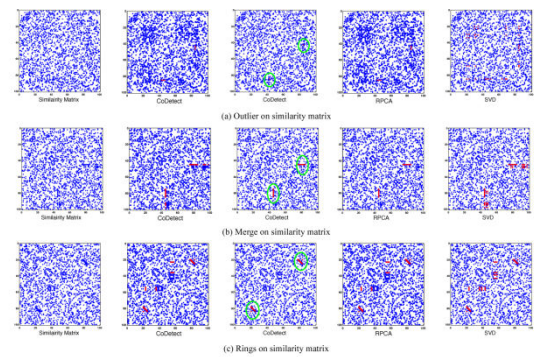
 Interval1- Interval4 are four time-intervals.

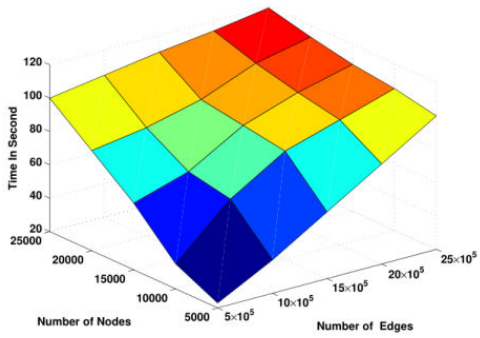
 “Location” is the area code of the place in which the transaction take place.

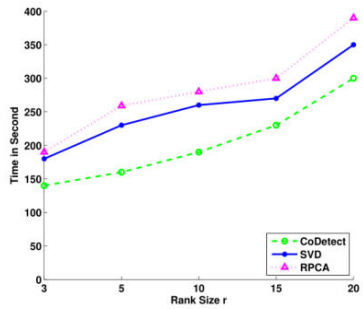
 Range1 – Range 4 are four amount-ranges. Location dimension has only value string as in numbers. All other dimensions return Boolean value. Boolean values are in form of 0 or 1.

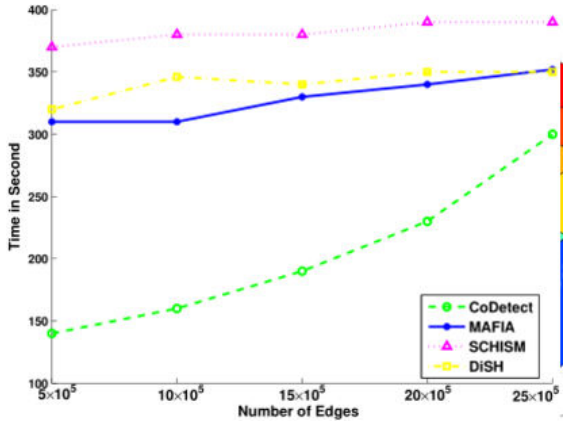
We propose a new credit card fraud detection system based on Behavior Certificate (BC) which reflects the cardholders’ transaction habits using Random Tree data mining technique.

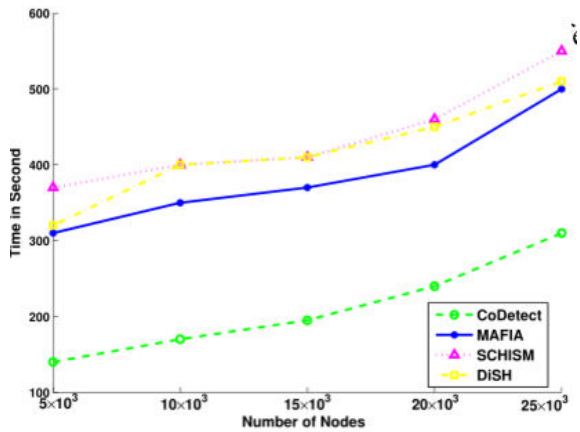












IV. CONCLUSIONS

Fraud detection is an important part of the modern finance industry. In this research, we have investigated the current practices in financial fraud detection using intelligent approaches, both statistical and computational. Though their performance differed, each technique was shown to be reasonably capable at detecting various forms of financial fraud. In particular, the ability of CI methods such as neural networks and support vector

machines to learn and adapt to new situations is highly effective at defeating the evolving tactics of fraudsters. There are still many aspects of intelligent fraud detection that have not yet been the subject of research. Some types of fraud, as well as some data mining methods, have been superficially explored but require future study to be completely understood. There is also the opportunity to examine the performance of existing methods by using customization or tuning, as well as the potential to study cost benefit analysis of computational fraud detection. Finally, further research into the differences between each type of financial fraud could lead to a generic framework which would greatly enhance the scope of intelligent detection methods for this problem domain.

REFERENCES

1. *Bai B, Yen J, and Yang X (2008) False financial statements: characteristics of China's listed companies and CART detecting approach. International Journal of Information Technology & Decision Making 7, 339-59.*
2. *Bermúdez L, Pérez J, Ayuso M, Gómez E, and Vázquez F (2008) A Bayesian dichotomous model with asymmetric link for fraud in insurance. Insurance: Mathematics and Economics 42, 779-86.*
3. *Bhattacharyya S, Jha S, Tharakunnel K, and Westland JC (2011) Data mining for credit card fraud: A comparative study. Decision Support Systems 50, 602-13.*
4. *Bose I and Wang J (2007) Data mining for detection of financial statement fraud in Chinese Companies. Paper presented at the International Conference on Electronic Commerce, Administration, Society and Education, Hong Kong, 15-17 August 2007.*
5. *Cecchini M, Aytug H, Koehler GJ, and Pathak P (2010) Making words work: Using financial text as a predictor of financial events. Decision Support Systems 50, 164-75.*
6. *Duman E and Ozcelik MH (2011) Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications 38, 13057-63.*
7. *Glancy FH and Yadav SB (2011) A computational model for financial reporting fraud detection. Decision Support Systems 50, 595-601.*
8. *Holton C (2009) Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. Decision Support Systems 46, 853-64.*
9. *Hoogs B, Kiehl T, Lacomb C, and Senturk D (2007) A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. Intelligent Systems in Accounting, Finance and Management 15, 41-56.*
10. *Humpherys SL, Moffitt KC, Burns MB, Burgoon JK, and Felix WF (2011) Identification of fraudulent financial statements using linguistic credibility analysis. Decision Support Systems 50, 585-94.*
11. *Jans M, van der Werf JM, Lybaert N, and Vanhoof K (2011) A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications 38, 13351-9.*
12. *Kirkos E, Spathis C, and Manolopoulos Y (2007) Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications 32, 995-1003.*
13. *Koh HC and Low CK (2004) Going concern prediction using data mining techniques. Managerial Auditing Journal 19, 462-76.*
14. *Ngai E, Hu Y, Wong Y, Chen Y, and Sun X (2011) The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems 50, 559-69.*