HMM Based Fault Tolerance in Credit Card Security

P. Prabavathy, S. Priyatharshini and N. Arunachalam

Abstract--- Credit card fraud detection raises unique challenges due to the streaming, imbalanced, and non-stationary nature of transaction data. It additionally includes an active learning step, since the labeling (fraud or genuine) of a subset of transactions is obtained in near-real time by human investigators contacting the cardholders. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In this paper, we model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) algorithm and Fraud detection model show how it can be used for the detection of fraud in card processing. Financial fraud is an ever growing menace with far consequences in the financial industry. HMM, Fraud detection model and image process had played an imperative role in the detection of credit card fraud in online transactions. Credit card fraud detection, which is a data problem, becomes challenging due to two major reasons – first, the profiles of normal and fraudulent behaviors change constantly and secondly, credit card fraud data sets are highly skewed. The using fraud detection algorithm performance of fraud detection in credit card transactions is greatly affected by the sampling approach on dataset, selection of HMM, Fraud detection model. Using fraud detection algorithm and image and image technique(s) used. At the same time, we try to ensure that genuine transactions are not rejected. A reliable augmentation of the target scarce population of frauds is important considering issues such as labeling cost; algorithm HMM, fraud detection; and constantly changing of patterns in the data streaming source. We have approached several scenarios with different legitimate and non-legitimate transaction ratios showing the feasibility of improving detection capabilities evaluated by means of receiver operating characteristic (ROC) curves and several key performance indicators (KPI) commonly used in financial business.

Keywords--- Detector, Signal Processing on Graphs, Credit Card Fraud, Comparative Analysis, Hidden Markov Model and Image Processing, Fraud Detection Model.

I. INTRODUCTION

Credit-card-based purchases can be categorized into two types: 1) physical card and 2) virtual card. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In the second kind of purchase, only some important information about a card (card number, expiration date, secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce

P. Prabavathy, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry. E-mail:Prabavathy110597@gmail.com

S. Priyatharshini, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry.

N. Arunachalam, M.Tech, (Ph.D.), Associate Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry. E-mail:narunachalam85@gmail.com

the rate of successful credit card frauds. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

Now a day the usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In this paper, we model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques available in the literature.

Along with the great increase in credit card transactions, credit card fraud has become increasingly rampant in recent years. In Modern day the fraud is one of the major causes of great financial losses, not only for merchants, individual clients are also affected. Three methods to detect fraud are presented. Firstly, HMM model is used to classify the legal and fraudulent transaction using data image of regions of parameter value. Secondly, Fraud detection model is used to model the probability density of credit card user's past behavior so that the probability of current behavior can be calculated to detect any abnormalities from the past behavior. Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios. The main task is to explore different views of the same problem and see what can be learned from the application of each different technique.

II. RELATED WORK

In present scenario when the term fraud comes into a discussion, credit card fraud clicks to mind so far. With the great increase in credit card transactions, credit card fraud has increasing excessively in recent years. Fraud detection includes monitoring of the spending behavior of users/ customers in order to determination, detection, or avoidance of undesirable behavior. As credit card becomes the mostPrevailing mode of payment for both online as well as regular purchase, fraud relate with it are also accelerating. Fraud detection is concerned with not only capturing the fraudulent events, but also capturing of such activities as quickly asPossible. The use of credit cards is common in modern day society. Fraud is a millions dollar business and it is rising every year. Fraud presents significant cost to our economy worldwide.

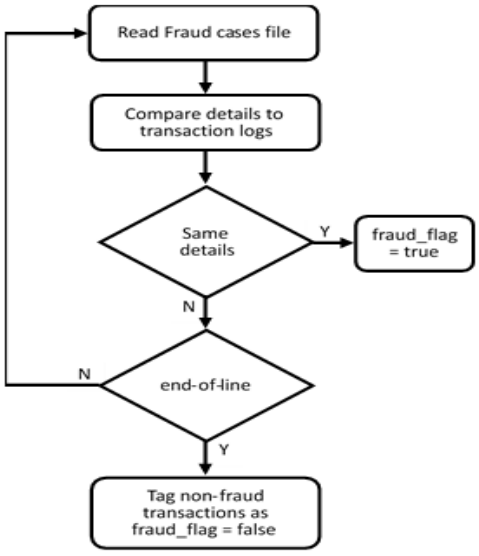
Modern techniques based on fraud detection, Image Processing, Hidden Markov Model Programming, Artificial Intelligence etc., has been introduced for detecting credit card fraudulent transactions. This paper shows how credit card fraud detection techniques can be combined successfully to obtain a high fraud coverage combined with a low or high false alarm rate.This paper presents a survey of current techniques used in credit card fraud detection and telecommunication fraud. The goal of this paper is to provide a comprehensive review of different techniques to detect fraud.

Credit card fraud has become more and more rampant in recent years. To improve merchants' risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. The models are compared in Terms of their performances. To improve the fraud detection system, the combination of the three presented methods could be beneficial. It is possible to use Bayesian Networks based on the input representation method and the developed HMM And Fraud detection model in the real detection system. In the future,

these models can extend to use in health insurance fraud detection.

III. IMPLEMENTATION OF PROPOSED SYSTEM

Fraud is increasing with the extensive use of internet and the increase of online transactions. More HMM and credit card detection algorithm advanced solutions are desired to protect financial service companies and credit card holders from constantly evolving online fraud attacks. The main objective of this paper is to construct an efficient fraud detection system which is adaptive to the behavior changes by combining classification and clustering techniques. This is a two stage fraud detection system which compares the incoming transaction against the transaction history to identify the anomaly using HMM and fraud detection algorithm in the first stage. In second stage to reduce the false alarm rate suspected anomalies are checked with the fraud history database and make sure that the detected anomalies are due to fraudulent transaction or any short term change in spending profile. In this work fraud detection supports incremental update of transactional database and it handles maximum fraud coverage with high speed and less cost. Proposed model is evaluated on both synthetically generated and real life data and shows very good accuracy in detecting fraud transaction.



The internet becomes most popular mode of payment for online transaction. Banking system provides e-cash, ecommerce and e-services by using online transaction. Credit card is one of the best ways for online transaction. In case of risk of fraud transaction using credit card has also been increasing. Credit card fraud detection is one of the ethical issues in the credit card companies, mortgage companies, fraud detection algorithm banks and financial institutes. Many techniques for credit card fraudulent detection but hidden Markova model (HMM) is one of the best engineering practices tool for credit card fraud system. Hidden Markova model generate, observation symbols for online transaction, we has shown the Hidden Markov Model for fraud detection in Credit card Applications. We presented experimental result to show the effectiveness of our approach. Hidden markov model generate, observation symbols for online transaction. Observation probabilistic in an HMM based system is initially studies spending profile of the cardholder and checking an incoming transaction, against spending behavior of the cardholder. We can show clustering model is used to classify the legal and fraudulent transaction using data conglomeration of regions of parameter.

Correlation of Page Fraud Detection Model Classification

Frequent item sets are sets of items that occur simultaneously in as many transactions as the user defined minimum support. The Fraud detection model support () is defined as the fraction of records of database that contains the item set as a subset: For example, if the database contains 1000 records and the item set appears in 800 records, then the support () = 800/1000 = 0.8 = 80%; that is, 80% of transactions support the item set. In credit card transaction data, the legal pattern of a customer is the set of attribute values specific to a customer when he does a legal transaction which shows the customer behavior. It is found that the fraudsters are also behaving almost in the same manner as that of a customer [1].

This means that fraudsters are intruding into customer accounts after learning their genuine behavior only. Therefore, instead of finding a common pattern for fraudster

behavior it is more valid to identify fraud patterns for each customer. Thus, in this research, we have constructed two patterns for each customer—Fraud detection model Or FRAUD pattern and fraud pattern. When frequent pattern mining is applied to credit card transaction data of a particular customer, it returns set of attributes showing same values in a group of transactions specified by the support. Generally the Fraud detection model pattern mining algorithms like that of return many such groups and the longest group containing maximum number of attributes is selected as that particular customer's legal pattern. The training (pattern recognition) algorithm is given below.

Step 1. Separate each customer's transactions from the whole transaction database.

Step 2. From each customer's transactions separate his/her legal and fraud transactions.

Step 3. Apply Fraud detection model algorithm to the set of legal transactions of each customer. The Fraud detection model algorithm returns a set of frequent item sets. Take the largest frequent item set as the legal pattern corresponding to that customer. Store these legal patterns in legal pattern database.

Step 4. Apply Fraud detection model algorithm to the set of fraud transactions of each customer. The Apriority algorithm returns a set of frequent item sets. Take the largest frequent item set as the fraud pattern corresponding to that customer. Store these fraud patterns in fraud pattern database.

Input: Customer Transactions Database,

Support

Output: Fraud detection model Pattern Database FDM, Fraud Pattern Database FPD

Begin

Group the transactions of each customer together.

Let there are "" groups corresponds to "" customers

For to do

Separate each group GI into two different groups FDM

And FDM of Fraud detection model and

fraud transactions.

Let there

are "" legal and "" fraud transactions

FIS = fraud(FDM, ,); //Set of frequent itemset

LP = ; //Large Frequent Itemset

LPD() = LP;

FIS = fraud(FDM, ,); //Set of frequent itemset

FP =; //Large Frequent Itemset

FPD() = FP;

end for

return LPD & FPD;

End

Definition Measures of Hidden Markov Model Pattern and Fraud Detection Algorithm

After finding the HMM and fraud patterns for each customer, the fraud detection system traverses these fraud and HMMand pattern databases in order to detect frauds. These pattern databases are much smaller in size than original customer transaction databases as they contain only one record corresponding to a customer. This research proposes a matching algorithm which traverses the pattern databases for a using fraud Detection algorithm match with the incoming transaction to detect fraud. If a closer match is found with legal pattern of the corresponding customer, then the matching algorithm returns "0" giving a green signal to the bank for allowing the transaction. If a closer match is found with fraud pattern of the corresponding customer, then the matching algorithm returns "1" giving an alarm to the bank for stopping the transaction. The size of pattern databases is where the number of customers is and is the number of attributes. The matching (fraud Detection) algorithm is explained below.

Step 1: Count the number of attributes in the incoming transaction matching with that of the legal pattern of the corresponding customer. Let it be.

Step 2: Count the number of attributes in the incoming transaction matching with that of the fraud pattern of the corresponding customer. Let it be.

Step 3: If and is more than the user defined matching percentage, then the incoming transaction is legal.

Step 4: If and is more than the user defined matching percentage, then the incoming transaction is fraud.

Step 5: If both and are greater than zero and, then the incoming transaction is fraud or else it is legal. The pseudo code of the testing algorithm is given in Algorithm.

Input: Legal Pattern Database LPD, Fraud Pattern Database FPD, Incoming Transaction ,
Number of costumers "", Number of attributes "", matching percentage "mp"
Output: 0 (if legal) or 1 (if fraud)
Assumption:
(1) First attribute of each record in pattern databases and incoming transaction is Customer ID
(2) If an attribute is missing in the frequent itemset (i.e., this attribute has different values in
each transaction and thus it is not contributing to the pattern) then we considered it as invalid

```
Begin
    lc = 0; //FRAUD DETECTION  attribute match count
    fc = 0; //fraud attribute match count
    for  = 1 to  do
        if (LPD(, 1) = (1)) then //First attribute
            for  = 2 to  do
                if (LPD() is valid and LPD() = ()) then
lc = lc + 1;
                endif
            endfor
        endif
    endfor
    for  = 1 to  do
        if (FPD() = (1)) then
            for  = 2 to  do
                if (FPD() is valid and FPD() = ()) then
                    fc = fc + 1;
                endif
            end for
        endive
    endfor
    if (fc = 0) then //no fraud pattern
        if ((lc/no. of valid attributes in legal pattern) ≥
mp) then
            return (0); //legal transaction
End
else return (1); //fraud transaction
endif
    elseif (lc = 0) then //no legal pattern
        if ((fc/no. of valid attributes in fraud pattern) ≥
mp) then
            return (1); //fraud transaction
        else return (0); //legal transaction
        endif
    elseif (lc> 0 && fc > 0) then //both legal and fraud
```

Training and Testing Dataset Creation

The following procedures are used for creating training and testing datasets for evaluating our model.

1. First, we removed the transactions corresponding to those customers who have only one transaction in dataset since it appears either in training or testing dataset only. Now the dataset has been reduced to 40918 transactions.
2. Then we divided these 40918 transactions into two sets—training set with 21000 transactions and testing set with 19918 transactions.

Positives (P): number of fraud transactions;

Negatives (N): number of legal transactions;

True positives (TP): number of fraud transactions predicted as fraud;

True negatives (TN): number of HMM transactions predicted as legal;

False positives (FP): number of FRAUD DETECTION transactions predicted as fraud;

Imbalanced Data

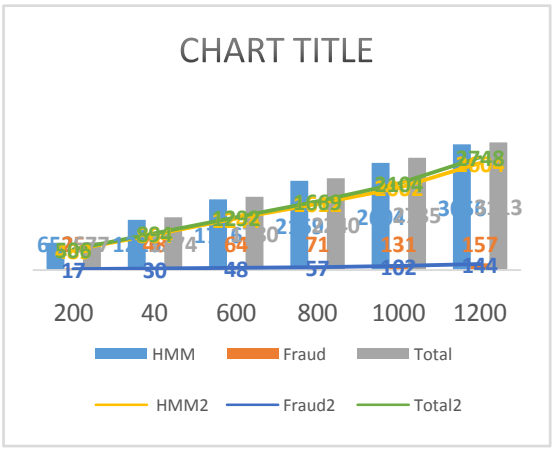
Number of customers	Number of transactions in training set			Number of transactions in testing set		
	FRAUD DETECTION or HMM	Fraud	Total	FRAUD DETECTION or HMM	Fraud	Total
200	652	25	677	489	17	506
400	1226	48	1274	864	30	894
600	1716	64	1780	1244	48	1292
800	2169	71	2240	1612	57	1669
1000	2604	131	2735	2002	102	2104
1200	3056	157	3113	2604	144	2748
1400	3440	158	3598	3083	147	3230

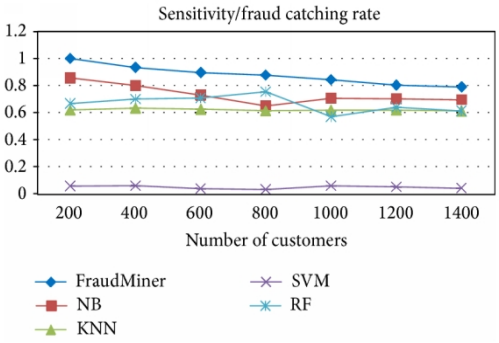
IV. RESULT AND DISCUSSION

In Security information module it will get the information detail and its store's in database. If the card lost then the Security information module form arise. It has a set of question where the user has to answer the correctly to move to the transaction section. Hence, we feel that HMM is an ideal choice for addressing this problem. Another important advantage of the HMM-based approach is a drastic reduction in the number of False Positives transactions identified as malicious by an FDS although they are actually genuine. We have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM.

In fraud detection, the most important measure is sensitivity or fraud detection rate, since the loss due to fraud depends on this metric. From the performance evaluation it is found that Fraud Miner is having the highest fraud detection rate (Figure 2) than other classifiers. The second important measure is the false alarm rate, since it shows the customer dissatisfaction due to false alarm (legal transaction, but suspected as fraud). Fraud Miner shows very less false alarm rate, again from the training dataset we removed the transactions corresponding to those customers who have only one transaction in the training dataset since it is hard to find a pattern from a single transaction. Now the training dataset has been reduced to 19165 transactions.

From this dataset, we have randomly selected different groups of customers and their Corresponding transactions in the training and testing dataset to create different training and testing datasets to evaluate the performance of Fraud Miner with increasing number of transactions. The data distribution is shown in Table 1.The credit card owner initiates a credit card transaction by communicating to a credit card number, and storing therein, a distinguishing piece of information that characterizes a specific transaction to be made by an authorized user of the credit card at a later time. The information is accepted as "network data" in the data base only if a correct personal identification code (PIC) is used with the communication. Now a day the usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising.





Most of the work found in the literature works on customer spending behavior analysis and some of them use some derived attributes also. However, we could not find any research performed on anonymous credit card transaction dataset where the derived attribute concept fails. Thus, the objective of this research was to develop a credit card fraud detection model which can effectively detect frauds from imbalanced and anonymous dataset.

V. CONCLUSION

Credit card fraud detection model for detecting fraud from highly imbalanced and anonymous credit card transaction datasets. The class imbalance problem is handled by finding fraud detection as well as fraud transaction patterns for each customer by using frequent

HMM and Fraud detection model. A matching algorithm is also proposed to find to which pattern (fraud detection or fraud) the incoming transaction of a particular customer is closer and a decision is made accordingly. In order to handle the anonymous nature of the data, no preference is given to any of the attributes and each attribute is considered equally for finding the patterns. The performance evaluation of the proposed model is done on and it is found that the proposed model has very high fraud detection rate, balanced classification rate, image process correlation coefficient, and very less false alarm rate than other state-of-the-art classifiers.

Therefore the fraud detection model should be adaptive to these behavioral changes. These behavioral changes can be incorporated into the proposed model by updating the fraud and pattern databases. This can be done by running the proposed pattern algorithm at fixed time points like once in 3 months or six months or once in every one lakh transaction. Moreover the proposed fraud detection method takes very less time, which is also an important parameter of this real time application, because the fraud detection is done by traversing the smaller pattern databases rather than the large transaction database.

REFERENCES

[1] H.V. Poor, “Information and inference in the wireless physical layer”, IEEE Wireless Communications, Vol. 19, No. 1, Pp. 40-47, 2012.

[2] L. Sankar, S.R. Rajagopalan and H.V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach”, IEEE Transactions on Information Forensics and Security, Vol. 8, No. 6, Pp. 838-852, 2013.

[3] S. Pawar, S. El Rouayheb and K. Ramchandran, “Securing dynamic distributed storage systems against eavesdropping and adversarial attacks”, IEEE Transactions on Information Theory, Vol. 57, No. 10, Pp. 6734-6753, 2011.

[4] L. Lai, S.W. Ho and H.V. Poor, “Privacy–security trade-offs in biometric security systems Part I: Single use case”, IEEE Transactions on Information Forensics and Security, Vol. 6, No. 1, Pp. 122-139, 2011.

[5] S. Bhattacharyya, S. Jha, K. Tharakunnel and J.C. Westland, “Data mining for credit card fraud: A comparative study”, Decision Support Systems, Vol. 50, No. 3, Pp. 602-613, 2011.

[6] V. Hodge and J. Austin, “A survey of outlier detection methodologies”, Artificial intelligence review, Vol. 22, No. 2, Pp. 85-126, 2004.

[7] D.M. Tax and R.P. Duin, “Uniform object generation for optimizing one-class classifiers”, Journal of machine learning research, Pp. 155-173, 2001.

[8] P. Danenas, “Intelligent financial fraud detection and analysis: a survey of recent patents”, Recent Patents on Computer Science, Vol. 8, No. 1, Pp. 13-23, 2015.

[9] A. Salazar, G. Safont, A. Soriano and L. Vergara, “Automatic credit card fraud detection based on non-linear signal processing”, IEEE International Carnahan Conference on Security Technology (ICCST), Pp. 207-212, 2012.

[10] R.H. Girgenti and T.P. Hedley, “Managing the risk of fraud and misconduct: meeting the challenges of a global, regulated and digital environment”, McGraw Hill Professional, 2011.