Credit Card Fraud Detection using Time Series Analysis

R.Devaki
PG Scholar
Department of Computer Science and Engineering
Coimbatore Institute of Engineering and Technology,
Coimbatore, India

V.Kathiresan
Assistant Professor
Department of Computer Science and Engineering
Coimbatore Institute of Engineering and Technology,
Coimbatore, India

S.Gunasekaran, Ph.D
Professor and Head
Department of Computer Science and Engineering
Coimbatore Institute of Engineering and Technology,
Coimbatore, India

ABSTRACT
Credit card usage has been increased tremendously because of the popularity of E-commerce. As the usage of credit card grows the occurrence of fraudulent transactions also increases, thus comes the stipulation of fraud detection. Detection of fraudulent transaction using credit card plays a vital role in financial institutions. In the proposed work, fraud detection is done with data mining approaches. The parameters considered are transaction amount and transaction time. For every cardholder there is always a robust periodic pattern in the spending behaviour, centered on this fact the anomalies in the transaction are detected by analyzing the past history of transactions belonging to an individual cardholder. In this work two levels of detection methods are used. At the first level the fraud is detected by analyzing whether the new incoming transaction is fraud or not by using distance-based method. At the second level the next transaction is predicted by means of label-prediction methodology and compared with the actual transaction, if there is deviation then it is detected to be a fraudulent transaction. If the particular transaction is considered as a fraud then the cardholder is asked to continue the transaction by asking a secret question, if the cardholder does not give correct answer then the transaction will not be allowed to continue further. The approach used in the proposed work has also decreased the false positive situation and hence it is ensured that genuine transaction is not rejected.

General Terms
Data mining, false positive, fraudulent transactions.

Keywords
Fraud detection, distance-based method, label-prediction methodology.

1. INTRODUCTION
In an era of digitalization, credit card fraud detection has become excessive prominence to the financial institutions. Credit card fraud is using a person's credit card by an unauthorized user for transactions while the actual card holder is not attentive to the fraud that has happened. This fraud can be done in on-line via internet and in off-line by stolen credit cards. Fraud detection involves recognizing fraud as quickly as possible once it has been committed. Fraudsters are using refined methods to gain access to credit card data and perpetrate fraud, new technologies are available to detect and prevent fraudulent transactions. Data mining is widely used to combat frauds because of its efficacy. It is a pertinent method that takes data as input and produces models or patterns as output.

In the proposed method, first module is implemented by outlier detection method based on distance-based method as in data mining; outlier detection is mostly done by distance measures. Outlier is detected in the incoming new transaction. In the second module; time-series analysis is done and the prediction of next transaction is done and the deviation from the observed pattern if any is detected based on label-prediction methodology. The fraud is confirmed with the card holder by asking a secret question.

2. RELATED WORK
AbhinavSrivastava et al (2008) stated that Hidden Markov Model (HMM)[1] provides a pragmatic method of detecting the anomaly by analysing the spending pattern of the customer. Basically HMM consists of sequence of states that works on Markov chain property. The input to the model is the observed sequence and the output is the probability for that sequence. For each cardholder HMM is trained and maintained. In HMM based approach there is an extreme decrease in the false positive rate. The objective of the system is to detect the anomaly during the transaction and then the fraud is confirmed with the cardholder by asking some secret questions.

The fraud detection in credit card transaction is one of the important techniques in classification. Dipti Thakur and Shalini Bhatia (2009) provides a technique to perform classification using decision tree methodology[2] in data mining and also the rules are shared among different credit card companies without sharing the data using agent based classification.

Trend offset analysis (TOA) is a local outlier-based supervised learning technique implemented for credit card fraud detection. It focuses on identifying pattern changes at an individual account level. TOA is used credit card fraud detection in such a way that a signature is assigned to each account based on the most recent transaction. Any significant deviation in current behavior from the assigned signature was used for outlier detection.

In genetic approach a authentication mechanism is used while transaction is done, to secure cash card by asking secret question to user for verification in case of credit card & SMS feedback system for ATM transactions. It secures cash card from being cloned via skimmed device& providing more security during the transaction. This work shows AI, image Processing & data mining techniques are used for fraud prevention there by implementing as/which ask secret questions i.e. ATM feedback SMS system with reply and by thumb impressions instead of detecting a fraud, a fraud can also be prevented.

3. PROPOSED WORK

By analysing the spending behaviour of the vendee the fraudulent activity is detected by two modules in the proposed work.

Module 1 - Detecting whether the incoming transaction is an anomaly.

Module 2 - Detecting anomaly by predicting the next transaction.

3.1 Module 1 – Distance Based Method

In data mining outlier detection is mostly based on distance measures. In first module, when the new transaction occurs it is compared with the analyzed spending pattern of the user. If the transaction amount exceeds a threshold value which is obtained during the analysis of previous transactions then the transaction is suspected to be a fraud and the secret question is asked from the user to allow continue the transaction.

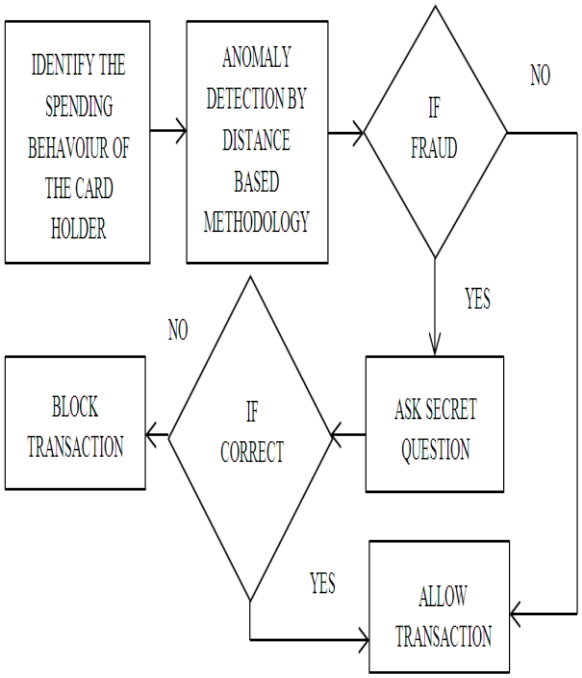

Fig 1: Module 1 – Data flow diagram

3.1.1 Algorithm

Step 1: Identify the vendee's spending behaviour.

Step 2: Find the centroid (x,y) where x is the mean of the transaction number and y is the mean of the transaction amount

Step 3: Find the distance between each point and the centroid using Euclidean distance formula[4].

Euclidean distance = $\sqrt{(x_i - x)^2 + (y_i - y)^2}$,

where i=1, 2, 3,…… Equation (1)

Step 4: Fix the maximum distance as the threshold.

Step 5: Now if a new transaction takes place repeat step 2 and 3.

Step 6: If the distance obtained is less than the threshold then the transaction is accepted else the transaction is rejected.

Step 7: The fraud is confirmed with the cardholder by asking the secret code.

Step 8: Update the threshold for every transaction.

3.2 Module 2 – Label Prediction Methodology

In second module, time-series analysis is used in which the data is recorded at regular intervals. In label-prediction methodology, every transaction is given a label as Low, Medium or High and the amount is clustered into low, medium and high and the transition probability is determined for each cluster. Two levels of anomaly detection can be obtained in this module. Initially the transaction amount is clustered using k-means clustering algorithm [1] and the anomaly is detected when the distance is high and then the incoming new transaction is compared with the predicted range of transaction by means of maximum transition probability, if there is any deviation then it is suspected to be fraud and the secret question is asked from the user to continue the transaction.

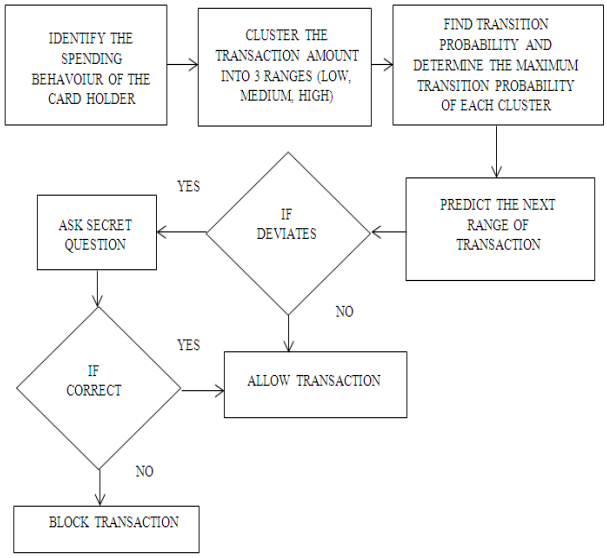

Fig 2: Module 2 – Data flow diagram

3.2.1 Algorithm

Step 1: Identify the vendee's spending behaviour.

Step 2: Cluster the transaction amount of the vendee into three clusters: Low (L), Medium (M) and High (H) by K-Means clustering algorithm. At this level anomaly can be detected based on the distance.

Step 3: Now the range of transaction is (H or M or L) is obtained for every transaction amount.

Step 4: Determine the transition probability of each cluster.

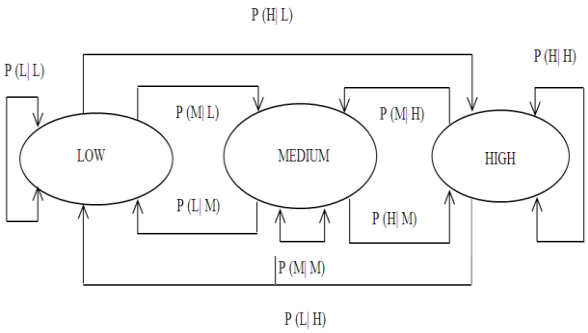

Fig 3: Transition Probability of Each cluster

Step 5: Obtain the maximum transition probability of each clusters. That is maximum transition from the cluster high is, medium or low. Similarly obtain for other clusters.
Step 6: Now the range of the incoming new transaction is found and compared with the obtained maximum transition probability.
Step 7: If the incoming transaction deviates from the predicted range then the transaction is rejected.
Step 8: The fraud is confirmed with the cardholder by asking the secret code
Step 9: Update the transition probability for every transaction.

4. IMPLEMENTATION IDEA

The implementation is, analyzing the spending behaviour of the cardholder and detecting the fraudulent activities if any. It is done by NetBeans and MySQL. The fraud detection system is implemented at the bank server. The system works for the transactions that are done during the online. The secret questions and the respective answers are collected from the cardholder during their registration for online transactions via credit card. The first 10 transactions are recorded in the database and analyzed by distance based method and label prediction method for every customers and from 11th transaction the fraud detection system works for every transaction that is done by the cardholder and if any fraud is detected, the cardholder's transaction is blocked and the further transaction can be done only after answering the secret questions.

5. CONCLUSION AND FUTURE WORK

With an abundant growth in the credit card transactions, credit card fraud has become progressively extensive. In this paper, the credit card fraud detection is done by identifying the anomalies in the transaction with two parameters namely time, amount. The fraud is detected by analyzing the spending behaviour of the cardholder. But the spending behaviour changes over time due to the changes in the life time. Every customer is expected to have his or her own pattern in spending the amount and the pattern is also being observed in the time interval between one transaction and the other. Under such circumstances, when a transaction is made, discrepancies that occurred is compared with the previously analyzed spending behaviour of the cardholder and the transaction will be suspected as a fraud and secret question will be asked. Sometimes the genuine transactions will be blocked leading to false positive situation. In the proposed method such situations are overwhelmed. However in order to overcome this false positive situation the system should be trained by accepting many false positive situations. Further enhancement can be done by making this system secure and covering more aspects of human behaviour like tracking the spending location of the card holder during direct mode purchase.The location of the cardholder can be tracked from the database in which the particular card holder's details are stored. If there is any huge difference in the location where the transaction takes place then it can be suspected as fraud.

6. REFERENCES

[1] AbhinavSrivastava, AmlanKundu, ShamikSural and ArunK. Majumdar(2008) 'Credit Card Fraud Detection UsingHidden Markov Model' IEEE Transactions onDependable and Secure Computing vol. 5 No. 1.

[2] Bhattacharya.S, Jha.S, Tharakunnel. k, and Westland J.C, "Data mining for credit card fraud: A comparative study." Decision Support systems, Vol.50, no. 7, 2011, pp.602-613.

[3] Clifton Phua, Vincent Lee, Kate Smith, Ross Gayler, 'A Comprehensive Survey of Data Mining-based Fraud DetectionResearch', http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf.

[4] Dipti Thakur and Shalini Bhatia (2009) 'Distributed Data Mining Approach to Credit Card Fraud Detection' Proceedings of SPIT-IEEE Colloquium and International Conference, Mumbai, India Vol. 4, 48.

[5] Francisca NonyelumOgwueleka (2011) 'Data Mining Application in Credit Card Fraud Detection System' Journal of Engineering Science and Technology Vol. 6 No. 3.

[6] Jiawei Han and MichelineKamber (2006) 'Data Mining Concepts and Techniques' ElseiverInc, Second edition ISBN:978-81-312-0535-8.

[7] Lawrence R. Rabiner, 'A tutorial on Hidden Markov Models and Selected applications in Speech Recognition', Proceedings of the IEEE, VOL.77, No 2, February 1989.

[8] Leila Seyedhossein and Mahmoud Reza Hashemi (2010) 'A Timelier Credit card Fraud Detection by Mining Transaction Time series' International Journal of Information & Communication Technology vol 2 No 3.

[9] Otto.P.E, Davies.G.B, Chater.N and Stott.H, 'From spending to understanding: Analyzing customers by their spending behaviour,' Journal of Retailing and Consumer Services, vol.16, no. 1, 2009, pp. 10-18.

[10] ParulBhanarkar and Pratiksha L. Meshram (2012) 'Credit and ATM card Detection using Genetic Approach' International Journal of Research & Technology Vol 1 Issue 10 ISSN:2278-0181.

[11] RinkyD.Patel and Dheerajkumar Singh (2013) 'Credit Card Fraud Detection and Prevention of Fraud Using Genetic Algorithm' International Journal of Soft Computing and Engineering' ISSN:2231-2307, Vol-2, Issue-6.

[12] 'A Tutorial on Clustering Algorithms', http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html.

[13] 'Data Clustering Algorithms', https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm.

[14] 'Measures of distance between samples: Euclidean', http://www.econ.upf.edu/~michael/stanford/maeb4.pdf.