ISSN:0975-9646

Credit Card Fraud Detection System using Hidden Markov Model and Adaptive Communal Detection

Nabha Kshirsagar[1], Neha Pandey[2],Shraddha Kotkar[3], Suja S. Panicker[4], Amol Pate[5]

[1,2,3,4] *MAEER's Maharashtra Institute of Technology, Pune, India*
[5] *Persistent Systems, Pune, India*

Abstract-Due to enormous growth in e-commerce ,there has been a drastic increase in use of credit cards for various purposes which include online and regular purchases. Associated with this rise, there are an alarming number of fraudulent cases that are reported every year. The proposed fraud detection system detects the fraud before the transaction is completed. Hidden Markov Model (HMM) and adaptive Communal Detection (CD) have been used for increasing the accuracy of the system along with One Time Password (OTP). If any of the two methods detects the incoming transaction as fraud ,OTP is sent to registered card holder.

Keywords-Data mining, Credit card fraud detection, HMM, CD

I. INTRODUCTION

SINCE past few years there is tremendous increase in online shopping. Credit card payment has been widely used for online transaction. The increasing use of credit card results in different kinds of fraud. Credit card based transaction can be of two types- physical and virtual. In physical credit card based transaction, fraudster needs to steal credit card for committing fraud whereas in virtual credit card based transaction fraudster needs to know the details of the registered cardholder. The details like name, credit card number, expiry date, secure code are required for making payment in virtual credit card based transaction. Many a times, genuine cardholder is not aware that the card has been stolen. A Fraud Detection System is required to detect the incoming transaction as genuine or fraud. Since the credit card is not required in virtual credit card based transaction, the only way to detect is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. It is a promising way to detect credit card frauds. The spending profiles like the transaction amount, the way the details are entered can be analysed. Credit Card based Fraud Detection System has already been developed using various data mining techniques like Hidden Markov Model[11], Bayesian Classifier, Neural Network. In credit card based application domain, it is difficult to distinguish between legal behaviour and fraud behaviour. Techniques like Communal Detection and Spike Detection[12] are used for finding real communal relations between the applications. Using such techniques it is possible to detect new kinds of fraud.

We briefly review some Researches for the detection of credit card fraud in Section II.

II. RELATED WORK

In this section we include a detailed survey of existing credit card detection systems.

Following work covers the use of Hidden Markov Model in fraud detection.

Hidden Markov Model (HMM) is a popular statistical tool used by engineers and scientists to solve various problems. Ashphak Khan et al model the sequence of operations in processing of credit card transaction using HMM in which the transition probability related to hidden state and observations is considered [1].

M.Sasirekha et al propose an Intrusion Detection Systems (IDS) combining three approaches -anomaly, misuse and decision making model to produce better detection accuracy and a decreased false positive rate. The integrated IDS can be built to detect attacks in credit card system using HMM approach in the anomaly detection module. The credit card holder's behaviours are taken as attributes and the anomalous transactions are found by the spending profile of the user. The transactions that are considered to be anomalous or abnormal are then sent to the misuse detection system. Here, the transactions are compared with predefined attack types and then sent to the decision making model to classify it as known or unknown type of attack. Finally, the decision-making module is used to integrate the detected results and report the types of attacks in credit card system. As abnormal transactions are analysed carefully in each of the module, the fraud rate is reduced and system is immune to attacks [2].

Hidden Markov Model was used to generate observation symbols for online transactions in [3]. Observation probabilistic in an HMM based system initially studies spending profile of the cardholder and checks an incoming transaction, against spending behaviour of the cardholder. Clustering model was used to classify the transactions into legal and fraudulent using data conglomeration of regions of parameter[3].Khan et al present experimental results to show the effectiveness of this approach.

The necessary theory to detect fraud in credit card transaction processing using a HMM is presented in [4]. HMM was initially trained with the normal behaviour of cardholder. If an incoming credit card transaction was not accepted by the trained HMM with sufficiently high probability, it was considered to be fraudulent. At the same time, it was ensured that genuine transactions are not rejected by using an enhancement of hybrid model to it.

Several existing systems of credit card fraud detection were analysed in [5] and a hybrid model comprising of a

combination of Bayesian classifier and HMM was proposed. It uses the various parameters of many previous proposed systems along with some new parameter to provide a better credit card fraud detection and to limit the fraud. The proposed model can be used at the verification phase of an online transaction [5].

The effectiveness of personalized models as compared to aggregated models in identifying fraud for different individuals was investigated in [6]. Data involved some actual transactions and some collected through online questionnaire. It was observed that aggregated models outperform personalized models. Also, the performances of the random forest and Naive Bayes in creating the models for fraud detection were compared. It was observed that random forest performed better than the Naive Bayes for the aggregated model, while Naive Bayes performed better in the personalized models [6].

A method of interactive construction of global HMMs based on local sequence patterns discovered in data is presented in [7]. Interesting sequences were found based on whose frequency in the database differed from that predicted by the model. The patterns were then presented to the user who updates the model using their intelligence and understanding of the modelled domain. It was observed that this approach leads to more understandable models than automated approaches. Two variants of the problem are considered: mining patterns occurring only at the beginning of sequences and mining patterns occurring at any position; both practically meaningful. For each variant, algorithms were developed allowing for efficient discovery of all sequences with given minimum interestingness. Applications to modelling webpage visitor's behaviour and to modelling protein secondary structure are presented [7].

The focus in [9] is on HMMs applied to the performance evaluation of computer systems and networks. A brief review of background on HMMs, and applications such as channel delay and loss characteristics, traffic modelling and workload generation are surveyed. The power of HMMs as predictors of performance metrics is also highlighted. It also covers few features of the module of the Tangram-II performance evaluation tool that is targeted to HMMs [9].

Following work covers brief survey on the use of Communal Detection (CD).

To address the limitations in existing non-data mining detection system of business rules and scorecards, and known fraud matching, and also to combat identity crime in real time, a new multilayered detection system is proposed in[8]which is complemented with two additional layers: Communal Detection (CD) and Spike Detection (SD).CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. SD finds spikes in duplicates to increase the suspicion score, and is probe-resistant for attributes.

Some of the major developments of an identity crime/fraud stream mining system have been identified in [10].Communal detection is about finding real communities of interest. The algorithm itself is unsupervised, single-pass, differentiates between normal and anomalous links, and mitigates the suspicion of normal links with a dynamic global white list. It is part of the important and novel communal detection framework introduced here for monitoring implicit personal identity streams. For each incoming identity example, it creates one of three types of single link (black, white, or anomalous) against any previous example within a set window. Subsequently, it integrates possible multiple links to produce a smoothed numeric suspicion score. In a principled stream-like fashion and using eighteen different parameter settings replicated over three large window sizes, the authors highlight and discuss significant score results from mining a few million recent credit applications[10].

III. PROPOSED WORK

From the above literature survey, we realized the need for a robust Fraud Detection system that combines the strengths of both the methods – HMM and adaptive CD. Thus the combined benefits will include getting lesser false positives and thereby increasing accuracy of system. Our system proposes to provide combined solutions to both the problems- Identity Theft and Fraud in credit cards. The OTP technique has also been used and hence the system detects fraud before the transaction is completed.

The patterns extracted from credit card data, transaction data and other logs, will be successfully able to characterize individual customers. Thus, we shall be able to profile each customer, based on his unique pattern of spending habits, like – the days when he spends more, the days when he spends less, the days he doesn't shop at all, the typical gap between his transactions, the highest amount he can shop for based on past records, the least amount he has shopped for, the way he enters his details and so on. Once these profiles are made, our proposed system will be able to distinguish between the authentic customer and any other unauthorized access and therefore we can prevent credit card frauds.

Thus one major module of our system is Pattern Recognition [13]. Once we get the usual patterns, we will be able to detect the presence of any outlier in it.

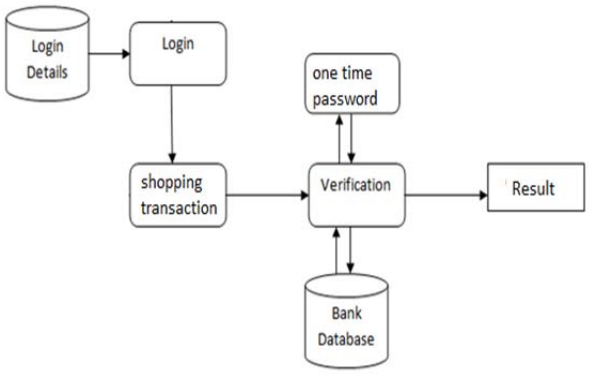

Fig.1 System Architecture of the proposed FDS

IV. METHODS

HMM and adaptive CD are being used along with OTP technique for fraud detection.

An HMM is a doubly stochastic process with an underlying stochastic process that is not observable (it is

hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols. An HMM is initially trained with the normal behaviour 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.[11]

The adaptive Communal Detection is fraud detection method .Basically, it identifies the identity theft where someone tries to steal your identity. For this it uses whitelist approach. The whitelist, a list of communal and self-relationships between applications, is crucial because it reduces the scores of these legal behaviours and False Positives(FPs).The adaptive CD finds relations between applications in the form of link for example if attributes considered are name ,address ,date of birth then link will be of length 3 if particular attribute is matched then 1 else 0 hence link can be from 000 to 111.Whitelist contains all legal relationships hence it contains links that can be acceptable. The adaptive CD then finds a Suspicion score, Delta and Average_Delta.

Delta= (Sum of suspicion scores of all linked applications/Sum of their outlinks)-(Suspicion score of current application/Number of outlinks of current application)

Average_Delta=Sum of the Delta of all linked applications/Number of outlinks of current application

The average delta has been calculated so that we can get an approximate value based on which we can detect the incoming transaction as fraud or genuine. The values have been obtained by testing on various datasets and considering all possible cases.

Under following conditions adaptive CD detects the incoming transaction as fraud:

1) If the link is not present in white list.
2) If Average_Delta is greater than 0.72.
3) If Average_Delta is zero, and Delta is greater than 0.5.

Further, OTP is sent to registered card holder if any of the two techniques detects the incoming transaction as fraud. If OTP shows that the transaction is genuine then status for given application is changed to genuine for further reference

V. RESULT

The dataset for HMM has been simulated using R-Software and for adaptive CD, the dataset is synthetic and is taken from[14].

HMM and adaptive CD sometimes generate random result. Using both together and OTP technique we can reduce the False Positives rate and hence the accuracy of the system is also increased. Since adaptive CD is unsupervised algorithm, labelling of training dataset is not required and it is also able to detect new kinds of fraud. Also the system detects the fraud before the transaction is completed. Hence it is a preventive measure. The system is also scalable for handling large volumes of transaction.

VI. CONCLUSION

In this project we have studied the Hidden Markov Model and adaptive Communal Detection algorithms and have surveyed various applications of HMM and CD. When used separately, both these algorithms suffered important limitations, which include- difficulty in testing because real data is not available in the case of HMM and issues like scalability and time constraints in case of Communal Detection. A Fraud Detection System that combines both these methods will be more robust and efficient in identifying various types of credit card frauds.

REFERENCES

[1] Ashphak Khan, Tejpal Singh, AmitSinhal, “Observation Probability in Hidden Markov Model for Credit Card Fraudulent Detection System”, Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012.

[2] M.Sasirekha,I.SumaiyaThaseen, J.SairaBanu,“An Integrated Intrusion Detection System for Credit Card Fraud Detection” , Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY) July 13-15, 2012, Chennai, India - Volume 1.

[3] Khan, A., Singh, T.,Sinhal, A., “Implement credit card fraudulent detection system using observation probabilistic in Hidden Markov Model”,Engineering (NUiCONE), 2012 Nirma University International Conference on6-8 Dec. 2012.

[4] IyerDivya,MohanpurkarArti ,JanardhanSneha, RathodDhanashree,SardeshmukhAmruta, “Credit card fraud detection using Hidden Markov Model”, Information and Communication Technologies (WICT), 2011 World Congress on11-14 Dec. 2011.

[5] Ashish Gupta, Jagdish Raikwal, “Fraud Detection in credit Card Transaction Using Hybrid Model”, International Journal of Engineering and Computer Science ISSN:2319-7242 , Volume 3 Issue 1 Jan, 2014 Page No. 3730-3735.

[6] Alowais,M.I.,Lay-KiSoon,“Credit Card Fraud Detection: Personalized Aggregated Model”, Mobile, Ubiquitous, and Intelligent Computing (MUSIC), 2012 Third FTRA International Conference on26-28 June 2012.

[7] SzymonJaroszewicz, “Interactively build Hidden Markov Models using interestingsequences, data mining”, Volume 21 Issue 1, July2010 Pages 186 – 220.

[8] K.S.Arthisree, A.Jaganraj, “Crime Detection using Data mining techniques”, International Journal of Advanced Research in Computer Science and Software Engineering 3(8),August - 2013, pp. 977-983.

[9] E. de Souza e Silva, R. M. M. Leão, Richard R. Muntz, “Performance evaluationofHidden Markov Model”,Performance Evaluation of Computer and Communication Systems. Milestones and Future ChallengesLecture Notes in Computer Science Volume 6821, 2011, pp 112-128.

[10] Phua, C. ,MonashUniv, Gayler, R, Smith-Miles, K.,Lee, V.,“Implicit Personal Identity Streams using Communal Detection”, Data Mining Workshops, 2006. ICDMWorkshops 2006. Sixth IEEE International Conference on Dec. 2006.

[11] AbhinavSrivastava, AmlanKundu, ShamikSural, Arun K. Majumdar, “Credit Card Fraud Detection Using Hidden Markov Model”, IEEE Transactions on dependable and secure computing, Volume 5, No. 1, January-March 2008.

[12] Clifton Phua,Kate Smith-Miles, Vincent Lee, and Ross Gayler,“Resilient Identity Crime Detection”,IEEETransactions on knowledge and data engineering Vol.24 No.3 year 2012.

[13] Jiawen Han, MichelineKamber, Jian Pei “Data Mining: Concepts and Techniques”, 3rd edition, The Morgan Kaufmann Series, May 2011,pp 244-254.

[14] Mr.Shakadwipi Amol J.,Prof. P.N.Kalavadekar,”Real-time Credit Application Fraud Detection Syarem Based on Data Mining”,Third Post Graduate Symposium on Computer Engineering cPGCON2014 Organized by Department of Computer Engineering,MCERC Nasik.