RESEARCH ARTICLE OPEN ACCESS

A Survey on Fraud Detection in Internet Banking using HMM and BLAST-SSAHA Hybridization

Ms. Avanti H. Vaidya*, Prof. S. W. Mohod**

*(Department of Computer Science and Engineering, Nagpur University, Wardha (Maharashtra), India
Email: avantivaidya9@gmail.com)
** (Department of Computer Science and Engineering, Nagpur University, Wardha (Maharashtra), India
Email: sudhir_mohod@rediffmail.com)

ABSTRACT
Due to rapid growth of E-commerce the use of credit card has been increased in day to day life. This caused an explosion in credit card fraud as credit card is a most popular mode of payment in internet banking and in regular purchases. In real life, along with genuine transactions the fraudulent transactions also scattered. This paper presents an analysis of different techniques to detect fraud in credit card transaction as well as in internet banking using Hidden Markov Model and BLAST-SSAHA hybridization. The BLAST-SSAHA Hybridization method proposed for the optimization of data in the database to avoid overfitting and underfitting problem of the system and HMM is proposed on optimized data in order to detect fraud.
Keywords – BLAST-SSAHA Hybridization, Hidden Markov Model, Internet Banking.

I. INTRODUCTION

Online banking service is the most popular and provides a fast and easy way to make transaction. Internet banking has their separate account for users. It is managed by banks or retail store. Net banking is a process over the internet to make the banking process effectively. The bank has automatically updates the customer accounts and records. E-commerce applications are now widely used by people. Various companies offer their services through these applications for improving their business. Online banking allows customer to conduct financial transactions on a secure website operated by virtual bank, credit union or building society. Online banking provides many transactional and non transactional features which are application specific. They are as follows.

- Transactional
  1. Funds transfer between two customers
  2. Paying third parties
  3. Investment purchase or sale
  4. Loan applications
- Non transactional
  1. Viewing account balance
  2. Viewing recent transaction
  3. Ordering cheque book
- Supports transaction approval process.
- Wire transfer.
- Financial institution administration.
- Support of multiple users having varying levels of authority.

Internet banking provides a personal financial management support. Some internet banking platform supports account aggregation to allow the customers to monitor all of their accounts in one place whether they are with their main bank or with other institution. As increasing use of internet banking for transaction, the number of fraud transaction is also increased by various thefts. This causes economic loss and makes the bank name unsecured. There are various ways that fraudsters execute an online fraud. By using various technologies they can do fraudulent activities.

1.1 Hidden Markov Model

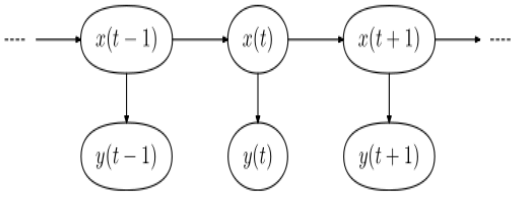

Fig1. Architecture of HMM

Fig 1 shows a general architecture of hidden markov model. Each oval shape represents a random variable that can adopt any number of values. The x(t) and y(t) represents the random variables. The x(t) is hidden state and y(t) is a observation at time 't'. The markov property states that the hidden variables x(t) at all times depends on the values of the hidden variables x(t-1). A hidden markov models are one of the most popular models in machine learning and provides statistics for modeling sequences. A probability distribution over sequences of observations is defined by using HMM. A Hidden Markov model (HMM) is identified as a statistical

model. The system being modeled using HMM is assumed to be a Markov process with unobserved state. The word 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model. Even if the model parameters are known but the model is still 'hidden'.

1.2 BLAST-SSAHA Hybridization.

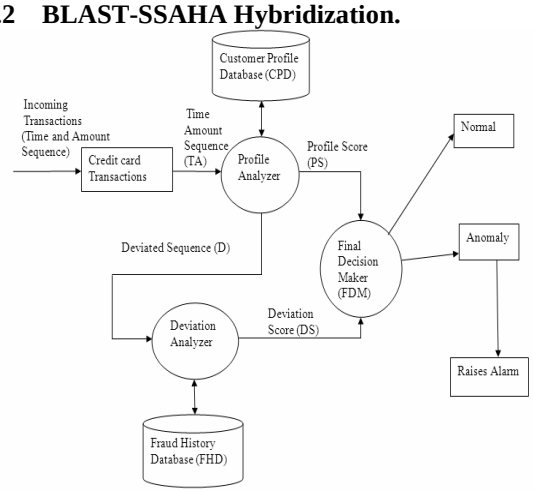

Fig2. Architecture of BLAST-SSAHA fraud detection system

The major components of BLAST-SSAHA fraud detection system are as follows.

1. Profile Analyzer (PA): PA analyzes the similarity of time-amount sequence of the incoming transaction with cardholder's profile database (CPD).

2. Deviation Analyzer (DA): DA analyzes the similarity of the deviated time-amount sequence with company's fraud history database (FHD).

3. Amount Sequence (A): Amount sequence represents a sequence of transaction amounts associated with the last few transactions on that card.

4. Time Sequence (T): Represents a sequence of transaction times associated with the last few transactions on that card.

5. Time-amount Sequence (TA): TA is the merged sequence of T and A.

6. K-tuple Table (KT): KT keeps sequence-index and sequence-offset information of history database. Cardholder's K-tuple information is kept in CKT and fraudster's K-tuple information is kept in FKT.

7. Profile Score (PS) and Deviation Score (DS): The Profile analyzer evaluates a similarity score between TA and CPD which is called PS. Similarly, deviation analyzer evaluates a similarity score between deviated sequence V and FHD which is called DS.

8. Deviated Sequence (V): The sequence of elements which have some deviation from the cardholder's profile form a deviated sequence V.

9. Final Decision Maker (FDM): FDM takes the final decision about the nature of the transaction based on the PS and the DS.

II. LITERATURE SURVEY

Sushimito Ghosh and Douglas L. Reilly in (1994) have proposed a work on neural network based fraud detection system. Using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labeled credit card account transaction and tested on a holdout dataset that consisted of all account activity over a subsequent two months period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail order fraud and non received issue (NRI) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fever false positives (reduced by factor of 20) over rule based fraud detection procedures. They discussed the performance of the network on the data set in terms of detection accuracy and earliness of fraud detection. The system has been installed on an IBM 3090 at Mellon Bank and is currently in use for fraud detection on that bank's credit card portfolio [1].

Emin Aleskrov, Berned Freisleben and Bharat Rao in (1997) proposed a neural network based database mining system for credit card fraud detection which is identified as a CARDWATCH. The system is based on neural network learning module, provides an interface to a variety of commercial databases and has a comfortable graphical user interface. In case of modern corporate databases, copying huge data sets from database is not tolerable. The possibility to directly access different databases types becomes a critical requirement for modern database mining system. Therefore, the CARDWATCH system which is a more sophisticated yet straight forward graphical user interface has created. The CARDWATCH system consists of five main modules. The main purpose of first Global Constant Module (GCM) to bundle all the global variables declared in the system (except the external dynamic link library (DLL) part). The second module that is Core/Graphical User Interface Module not only allows the user to comfortably control the entire system but also serves the "glue" for all other module. It serves as a container for all GUI-related routines, including the call-back code or auxiliary functions for widget control. The Database Interface Module (DBIM) handles the communication between the database and the remaining modules. It contain code for various operations such as initialization, opening and modification of database fields to GUI data control widgets, querying of the database with the help of SQL or assignment of selected record sets to the global variables. The DBIM cooperates with the GUIM, GCM and LAIM module. The fourth Learning Algorithms Library module provides the

neural network learning algorithms. It is limited to only a few neural network architectures with three learning rules, but it is easily extensible to any other adoptive technique used to detect anomalies in customer's credit card usage dynamics. This module is autonomous having its own database access facilities also independent of the core part of the system while retrieving transaction data or making fraudulent records. The fifth Learning Algorithm Interface Module (LAIM) provides an interface between the core and the neural network library. This module contains two functions: train and test with method dependant calls to LAL test/ train functions [2].

Mubeena Syeda, Yan-Qing Zhang and Yi Pan in (2002) proposed a work on Parallel granular neural networks for fast credit card fraud detection. A parallel granular neural network (GNN) is developed to speed up data mining process and knowledge discovery process for credit card fraud detection. The entire system is parallelized on Silicon Graphics Origin 2000, which is shared memory multiprocessor system consisting of 24 CPU, 4G main memory and 200 GB hard drive. In simulations, the parallel fuzzy neural network running on 24-processor system is trained in parallel using training data sets. Then the trained parallel fuzzy neural network discovers fuzzy rules for future prediction. The data are extracted into file from SQL server database containing sample Visa Card transactions. It is then pre-processed for applying in fraud detection. The data are classified into three categories: first for training, second for prediction and third for fraud detection. After learning from training data, the GNN is used to predict on second set of data and latter the third set of data for fraud detection. Around eight scenarios are employed for detecting purpose [3].

Xuan Dau Hoang, Jiankun Hu, Peter Bertok in (2003) presented a new method to process sequences of system calls in order to detect anomaly intrusion. They proposed a Multi-Layer model for anomaly intrusion detection using program sequences of system calls using Hidden Markov Models and enumerating methods. They performed their experiments on Unix Sendmail program have shown that the model is better in detecting anomalous behavior of program in terms of accuracy and response time. Intrusion detection techniques can be broadly classified into two categories: first, misuse detection and second, anomaly detection. Anomaly detection constructs models of subject behavior and any significant deviation from normal behaviors is considered as a part of an attack. They described two main steps to test procedures of sequences in order to find the sequences of system call contained anomaly intrusion. In the first step, the sequences of system calls is compared to those in normal database to find a mismatch or a rare sequences indicated by low occurring frequency. In the second steps the input to the Hidden Markov Model computes the corresponding probability. If the probability is required to produce the sequence is smaller than a predefined probability threshold, it is considered as anomalous sequence [4].

Vasilis Aggelis in (2006) proposed a work on offline internet banking fraud detection. They have demonstrated one successful fraud detection model. The main scope is to present its contribution in fast and reliable detection of strange transaction including fraudulent ones. The Offline internet banking fraud detection system offers many benefits to bank and to customers as well. New data is imported into the database in constant time frames, not in real time [5].

Osama Dandash, Phu Dung Le and Bala shrinivasan in (2007) has worked on internet banking models using a security analysis. They stated that internet banking fraud can be performed internally by genuine staff or externally by customers or suppliers. A security analysis of the proposed internet banking model compared with that of the current existing models used in fraudulent internet payment detection and prevention. Several modern models in preventing and detecting fraud are evolving and being applied to many banking systems. The proposed model facilitates Internet Banking Fraud Detection and Prevention (FDP) by applying two new secure mechanisms, Dynamic Key Generation (DKG) and Group Key (GK). The main function of Dynamic Key Generation allows only legitimate users to access the system and secure user's data to greatly strengthen the authentication. The Group Key mechanism which can identify users, manage them into groups and verify their authorization levels. However, they have no effective mechanism to identify legitimate user and detect unlawful activities. Also they are not secure enough to prevent fraudulent users from performing fraudulent transactions over the Internet. Ones the fraudsters succeeded in obtaining both the secure keys, he can easily perform fraud [6].

Abhinav Shrivastava, Amlan Kundu, Shamik Sural and Arun Majumdar in (2008) proposed an application of Hidden Markov Model (HMM) in credit card transaction processing. HMM model is used for credit card fraud detection. They 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. They have suggested a method for finding the spending profile of cardholders as well as application of the knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has been explained that how the HMM can detect whether an incoming transaction is fraudulent or not [7].

Osama Dandash, Yuiling Wang , Phu Dung Le and Bala shrinivasan in (2008) have proposed an efficient new scheme used to prevent fraud by applying different security algorithms where hacking one secret will not compromise the whole system's security because the system does not rely on fixed values. The higher number the transactions performed there is the less chance the system being compromised. The practical use of this technique has been demonstrated by applying it to Internet banking payment systems. The results show that their technique enhances their security considerably. It has been shown that the proposed technique is secure against key compromise [8].

Qinghua Zhang in (2009) provided a survey on fraud risk prevention of online banks. Their aimed, in the first hand, at giving a discussion on the fraud risks of online banking, introducing the current application situation of information sharing mechanism in respect of internet fraud outside China as well as the development of such concept in China. Then, a system is designed for sharing internet fraud information. Information sharing effectively maintains the security of online banking accounts through which member party could realize the information sharing within the third party frame concerning with internet threat, track internet fraud information and insecurity warnings. But the safety risk presently existed in e-banks provide lack of efficient constraint from point of network itself. So it is necessary to construct a feasible internet system to restrict all activities inside this virtual world. All the online banks should put more joint efforts in perfecting the mechanism for sake of international co-operation [9].

Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arum K. Majumdar (2009) stated that a phenomenal growth in the number of credit card transactions, especially for online purchases, has recently led to a substantial rise in fraudulent activities. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. In real life, fraudulent transactions are interspersed with genuine transactions and simple pattern matching is not often sufficient to detect them accurately. Thus, there is a need for combining both anomaly detection as well as misuse detection techniques. They have proposed to use two-stage sequence alignment in which a profile analyzer (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analyzers. In order to achieve online response time for both PA and DA, they suggested a new approach for combining two sequence alignment algorithms BLAST and SSAHA [10].

Khyati Chaudhary and Bhawna Mallick in (2011) stated that Data mining has been increasing as one of the chief key features of many security initiatives. Data mining involves the use of data analysis tools to discover unknown, valid patterns as well as relationships in large data sets. Decades have seen a massive growth in the use of credit cards as a transactional medium. Data mining become even more common in both the private and public sectors. Data mining has been used widely in industries such as Banking, Insurance, Medicine and Retailing to reduce costs, enhance Research and increase Sales. Credit cards are much safer from theft than is cash and also a promising area for buying and sales. Credit Cards are growing as a popular medium of transaction. Therefore, Fraud Detection involves monitoring the behavior of users/customers in order to estimate, detect or avoid undesirable behavior in future. They investigated the factors and various techniques involved in credit card fraud detection during/after transaction as well. Efficient and well-organized credit card fraud detection system is a greatest requirement for any card issuing bank. Credit card fraud detection has drawn quite a lot of interest from the research community and a number of techniques have been proposed to counter/identify credit card fraud [11].

Francisca Nonyelum Ogwueleka in (2011) has worked on Data mining application in credit card fraud detection system. Data mining is popularly used to combat frauds because of its effectiveness. It is a well-defined procedure that takes data as input and produces models or patterns as output. The design of the neural network (NN) architecture for the credit card detection system was based on unsupervised method, which was applied to the transactions data to generate four clusters of low, high, risky and high-risk clusters. The self-organizing map neural network (SOMNN) technique was used for solving the problem of carrying out optimal classification of each transaction into its associated group, since a prior output is unknown. This network contains two layers of nodes an input layer and a mapping (output) layer in the shape of a two-dimensional grid. SOMNN component learning is a learning procedure that divides a set of input patterns into clusters that are inherent to the input data. In the SOMNN engine, the data set description, number of data points in the dataset and the number of clusters are entered. The data point's entry was created and filled. The entries are sorted in ascending order under the cluster list. When the train/generate clusters are selected, the cluster label becomes ready for filling

depending on the number of clusters entered and then stored on the database. The database was stored in Microsoft Access table and was also used to determine when a card transaction was to be processed, blocked, unblocked, or the alert set off. After each transaction, the data point entry and clusters made are processed by the SOMNN engine and sent into the database. This helps the detection engine to know when any data entry is legitimate or fraudulent, and the reason is given immediately after the alert. The receiver-operating curve (ROC) for credit card fraud (CCF) detection watch detected fraud cases without causing false alarms [12].

Sunil S. Mhamane and L.M.R.J Lobo in (2012) explained about how Fraud is detected and prevented using Hidden Markov Model. At the same time, they have tried to ensure that genuine transactions are not rejected by making use of one time password that was generated by the Bank server and sent to the particular customers through SMS to their mobile number which is registered in the system. Banks are seeking to minimize huge losses through fraud detection and prevention systems. Many different advanced fraud technologies are being applied to fraudulent Internet banking transactions detection and prevention. However, they have no effective detection mechanism to identify legitimate users and trace their unlawful activities. They proposed a model to overcome all these difficulties using Hidden Markov Model [13].

S. Esakkiraj, S. Chidambaram in (2013) designed a model with sequence of operations in online transactions by using Hidden Markov Model (HMM) and decides whether the user act as a normal user or fraud user. The HMM is initially trained with customer's last few transactions. In the trained system, the new transaction is evaluated with transition and observation probability, system finds the acceptance probability and decides the transaction will be declined or not. The model predicts the fraudulent during the transaction time and prevents the money transfer. The main objective was to ensure that the genuine transactions should not be rejected [14].

III. PROPOSED METHODOLOGY

Step1: Create a dummy database for several customers because of privacy of banks.
Step2: Apply BLAST-SSAHA algorithm on database for optimization of data.
Step3: Based on history of banking transaction, database will save customer transaction pattern.
Step4: Trained HMM model to evaluate if the ongoing transaction is fraudulent or not.
Step5: A web service based Tomcat Apache server will be created and deployed to allow users to make use of internet banking.
Step6: Some client applications that will allow the user to make online payments for required service.
Step7: Client application shall communicate with server using Java Networking.
Step8: On finding fraudulent transaction one time password will be sent to client through SMS to their mobile number which is registered in the system.

IV. CONCLUSION

A Hidden Markov Model provides sequence of transactions to different state also at each state to decide whether the transaction is a case of fraud or not. It provides more security to internet banking system. HMM maintains a log of several user transactions which provides a proof for the bank. HMM reduces substantial work of an employee since it maintains a log. The processing speed of BLAST-SSAHA hybridization is fast to enable on-line detection of credit card fraud transaction. The hybridized algorithm named BLAH-FDS identifies as well as detects fraudulent transactions using sequence alignment tool. BLAH-FDS can be effectively used to counter frauds in telecommunication domain also.

REFERENCES

[1] Ghosh and Reilly , Credit card fraud detection with a neural network, IEEE Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994.pp 621-630.

[2] Aleskerov, Emin, Bernd Freisleben, and Bharat Rao, CARDWATCH: A neural network based database mining system for credit card fraud detection, Computational Intelligence for Financial Engineering. Piscataway, NJ: IEEE, 1997, pp.220-226.

[3] Mubeena Syeda, Yan-Qing Zbang and Yi Pan, Parallel Granular Neural Networks for Fast Credit Card Fraud Detection, IEEE Transaction .2002, pp.572-577

[4] X.D. Hoang, J. Hu, and P. Bertok, A Multi-Layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls, Proc. 11th IEEE International Conf. Networks, 2003, Pp.531-536.

[5] Vasili s Aggelis , Offiine Internet Banking Fraud Detection, IEEE Proceedings of the /first International Conference on Availability, Reliability and Security, 0-7695-2567-9/06, 2006.

[6] Osama Dandash,Phu Dung Le and Bala Srinivasan ,Security Analysis for Internet Banking Models, Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, IEEE transaction, 2007, pp. 1141-1146.

[7] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Credit Card Fraud Detection Using Hidden Markov Model, IEEE Transaction, January-March 2008. Pp. 37-47.

[8] Osama Dandash Yiling Wang and Phu Dung Leand Bala Srinivasan, Fraudulent Internet Banking Payments Prevention using Dynamic Key, Journal Of Networks, Vol. 3, No. 1, January 2008".

[9] Qinghua Zhang, Study on Fraud Risk Prevention of Online Banks, International Conference on International Journal of Electronics and Computer Science Engineering 1765Networks Security, Wireless Communications and Trusted Computing IEEE, 978-0-7695-3610-1/09, 2009, pp 181-184.

[10] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arum K. Majumdar, BLAST-SSAHA hybridization for credit card fraud detection, IEEE Transactions On Dependable And Secure Computing, Vol. 6, No. 4, October-December 2009. Pp. 309-315.

[11] Khyati Chaudhary and Bhawna Mallick Exploration of Data Mining Techniques in Fraud Detection System, International Journal of Electronics and Computer Science Engineering 1765, ISSN- 2277-1956.

[12] Francisca Nonyelum Ogwueleka, Data mining application in credit card fraud detection system, Journal of Engineering Science and Technology ,Vol. 6, No. 3 (2011) 311 – 322.

[13] Sunil S Mhamane and L.M.RJ Lobo ,Internet Banking Fraud Detection Using HMM, IEEE-20180, ICCCNT'12 26th_28th July 2012, Coimbatore, India.

[14] S. Esakkiraj and S Chidambaram "A predictive approach for fraud detection using Hidden Markov Model", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 1, January- 2013, ISSN: 2278-0181.