A Survey on Hidden Markov Model for Credit Card Fraud Detection

Anshul Singh, Devesh Narayan

Abstract*— Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. Fraudsters are so expert that they engender new ways for committing fraudulent transactions each day which demands constant innovation for its detection techniques as well. Many techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, decision tree, neural network, logistic regression, naïve Bayesian, Bayesian network, metalearning, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A steady indulgent on all these approaches will positively lead to an efficient credit card fraud detection system. This paper presents a survey of various techniques used in credit card fraud detection mechanisms and Hidden Markov Model (HMM) in detail. HMM categorizes card holder's profile as low, medium and high spending based on their spending behavior in terms of amount. A set of probabilities for amount of transaction is being assigned to each cardholder. Amount of each incoming transaction is then matched with card owner's category, if it justifies a predefined threshold value then the transaction is decided to be legitimate else declared as fraudulent.

***Index Terms* — Credit card, fraud detection, Hidden Markov Model, online shopping**

I. INTRODUCTION

In today's electronic society, e-commerce has become an essential sales channel for global business. Due to rapid expansion of e-commerce, making use of credit cards for purchases has dramatically amplified. Unfortunately, fraudulent use of credit cards has also become an attractive source of revenue for criminals.

Occurrence of credit card fraud is increasing dramatically due to the security weaknesses in contemporary credit card processing systems resulting in loss of billions of dollars every year Credit cards can be used for doing shopping either offline or online. In offline transaction the card must be physically present and is inserted in the payment machine in the merchant's place for making the payment. But in online transaction only some of the card details like secure code, expiration date and card number etc. is needed to do the transaction as it is mostly done via phone or internet [1]..

Credit card Fraud can be performed in two ways. In first case the card is stolen and misused and in the other case only card details are known which is then entered while doing online shopping for buying the expensive commodities. In this case the owner is not aware that his card details are being used unless notified by some means.

. Some of the challenges that are faced by most detection techniques include [2]:

- Skewed distribution of legitimate and fraudulent data in the database that challenges the detection approaches. Genuine transactions are much higher as compared to fraudulent.
- Count of transaction which is proliferating swiftly. Mining of such immense amount of data calls efficient techniques.
- Availability of labeled data for the purpose of training, as genuine or cheat is not readily available.
- Tracking user's behavior is tough as it changes quite often for all type of users (good users, business and fraudsters). Dealing with old as well as new intellectual is a challenging task.

To deal with credit card fraud, credit card fraud prevention and credit card detection techniques are employed. Prevention approaches include fluorescent fibers, multitone drawings, watermarks, laminated metal strips and holographs on banknotes etc. while detection methods comes into picture when fraud prevention fails. In committing fraud, the range of fraudsters highly varies, some may be masters in doing so and some may be newcomers. For dealing with the masters the detection techniques must be updated constantly as fraudsters are quite prepared enough to penetrate the present detection methods. While for newcomers, the existing methods may work well. So a balanced approach is expected for the purpose of detection of frauds [3].

Manuscript Details: January 2012

Anshul Singh, Computer Science and Engineering Department RCET Bhilai, CSVTU India, Mob.No.09755307738, (e-mail: anshul981@yahoo.com)

Devesh Narayan, Computer Science and Engineering Department. RCET Bhilai, CSVTU, India, MobileNo.09893103576, (e-mail: devesh_nar@yahoo.com).

II. LITERATURE REVIEW ON CREDIT CARD FRAUD DETECTION

Aleskerov et al. [4] brought CARDWATCH, a database mining system used for credit card fraud detection. The system provides an interface to a variety of commercial databases and is based on a neural learning module. Kim and Kim have recognized skewed distribution of data and blend of legitimate and fraudulent transactions as the two main reasons for the complication of credit card fraud detection [5]. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections.

Fan et al. [6] suggest the application of distributed data mining in credit card fraud detection. Chiu and Tsai [7] have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. With this proposal, participating banks share knowledge about the fraud patterns in a varied and dispersed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. D. Sanchez, M.A. Vila L. Cerda and J.M. [8] Serrano adopted association rules for credit card fraud detection

Brause et al. [9] have presented an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage.

Syeda et al. [10] have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection. A complete system has been implemented for this purpose. Leila Seyedhossein and Mahmoud Reza Hashemi [11] suggested Mining Information from Credit Card Time Series for Timelier Fraud Detection. [12] suggest a credit card fraud detection system (FDS) using metalearning techniques to learn models of fraudulent credit card transactions. Metalearning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. Thus a metaclassifier is trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection [13]. They use Java agents for Metalearning (JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them.

Phua et al. [14] suggest the use of metaclassifier similar to in fraud detection problems. They consider naive Bayesian, C4.5, and Back Propagation neural networks as the base classifiers. A metaclassifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use credit card fraud detection as the target application, their approach is quite generic. Phua et al [15] have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report. Prodromidis and Stolfo [16] use an agent-based approach with distributed learning for detecting frauds in credit card transactions. It is based on artificial intelligence and combines inductive learning algorithms and metalearning methods for achieving higher accuracy.

Ekrem and M. Hamdi [17] developed genetic algorithm for fraud detection purpose. Qibei Lu and Chunhua Ju [18] presented Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine. In this method they used principal component analysis (PCA) is been used to reduce the training dimension of SVM (support vector machine) effectively.

Ekrem and M. Hamdi [19] contributed for detection of credit card fraud by the use of genetic algorithm and scatter search.

Sherly K.K. and R Nedunchezhian [20] proposed BOAT adaptive credit card fraud detection system. In this work BOAT supports incremental update of transactional database and it handles maximum fraud coverage with high speed and less cost.

R. Huang, H. Tawfik, and A.K. Nagar [21] worked in AIS (Artificial immune system) model. The AIS based model combines two artificial immune system algorithms with behavior based intrusion detection using Classification and Regression trees (CART).

Tatsuya Minegishi and Ayahiko Niimi [22] developed a decision tree learning called Very Fast Decision Tree learner (VFDT), which regards real data as a data stream and provide solution for imbalanced data stream.

R. Wheeler, S. Aitken [23] used multiple algorithms for fraud detection which is an application of case based reasoning. The approach was towards the problem of reducing the number of final-line fraud investigations in the credit approval process.

Lean Yu, Wuyi Yue, Shouyang Wang a, K.K. Lai [24] proposed Support vector machine based multiagent ensemble learning for credit risk evaluation. Here the impact of the diversity of individual intelligent agents on the generalization performance of the SVM-based multiagent ensemble learning system is examined and analyzed.

Chun-Hua JU and Na Wang [25] did Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum. It finds fraud record by computing similar coefficient sum of every two objects and an example is given to validate the model.

Amlan Kundu, Suvasini Panigrahi, Shamik Sural, and Arun K. Majumdar [26] proposed BLAST-SSAHA Hybridization for Credit Card Fraud Detection where some heuristics improve the performance of sequence alignment algorithm. Basic Local Alignment Search Tool (BLAST) and FAST-All (FASTA) are the two most popular heuristic approaches for local sequence alignment. Sequence Search and Alignment by Hashing Algorithm (SSAHA) is one of the fastest tools for sequence alignment where the alignment process is performed in memory using hash table.

III. APPLICATION OF HIDDEN MARKOV MODEL AS CREDIT CARD FRAUD DETECTION METHOD

A Hidden Markov Model is a finite set of states; each state is associated with a probability distribution. Transitions among these states are administered by a set of probabilities called transition probability. In a specific state a conceivable outcome or observation can be created which is associated symbol of observation of probability distribution. It is only the result, not the state that is evident to an external viewer and therefore states are ``hidden" to the outside; resulting the name Hidden Markov Model [3].

Hence, Hidden Markov Model is a perfect way out for dealing with detection of fraud transaction using credit card. Another significant benefit of the HMM-based approach is large reduction in the number of False Positives transactions acknowledged as malicious by a fraud detection system even though they are categorically genuine [1].
The problem with most of the above mentioned technique is they require labeled data, representing fraudulent or legitimate transaction for training their fraud detection system.

In this prediction process, HMM consider mainly three price value ranges such as
1) Low (l),
2) Medium (m) and,
3) High (h).

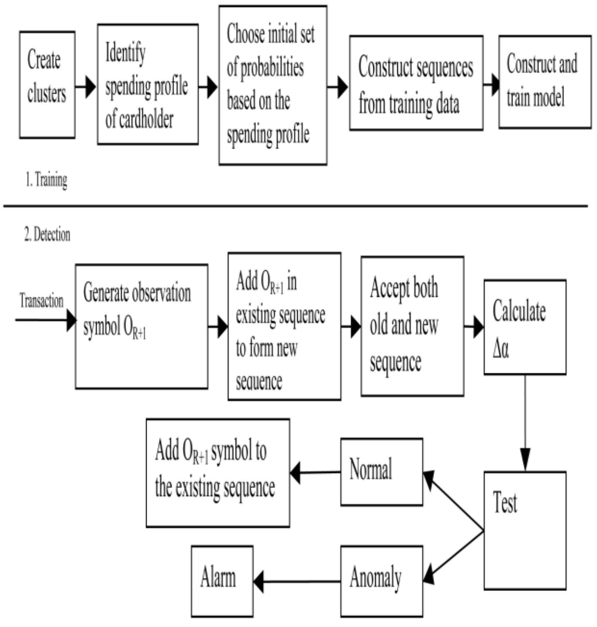
Create clusters
Identify spending profile of cardholder
Choose initial set of probabilities based on the spending profile
Construct sequences from training data
Construct and train model
1. Training
2. Detection
Transaction
Generate observation symbol O_{R+1}
Add O_{R+1} in existing sequence to form new sequence
Accept both old and new sequence
Calculate $\Delta\alpha$
Add O_{R+1} symbol to the existing sequence
Normal
Test
Alarm
Anomaly

Figure 1: process flow of credit card fraud detection system

First, it will be required to find out transaction amount belongs to a particular group either it will be in low, medium, or high ranges.

TABLE 1
Example Transactions with the Dollar Amount
Spent in Each Transaction

Transaction No.	1	2	3	4	5	6	7	8	9	10
Amount (in $)	40	25	15	5	10	25	15	20	10	80

The implementation technique used in HMM is creating clusters of training sets so as to identify spending profile of card holder. The type of items purchased works as states for the model. The transition from one state to another is determined by probability distribution. It requires minimum 10 previous transactions, on the basis of which the fore coming transaction is chosen as fraud or genuine.
The model goes through two stages. In the first stage training of the system is done. Second stage works for the detection of the fraud, based on the expected range of amount the transaction. The expected amount and the actual amount for the next transaction are compared on the basis of probability distribution during training phase.
If the deviation is above a threshold value then it is treated as fraud else legal. In case of fraud alarm is generated and transaction is terminated or else it is routinely accomplished.

The figure 1 below illustrates about the two phases of the detection system used by HMM.

In the training phase clusters are created and based on the initial set of transactions the spending profile of cardholder is identified. This directs for expected transaction amount for each cardholder and the system is trained accordingly.

Then in the detection phase the system looks for the deviation in expected and actual outcome and fraud is recognized.

IV. CONCLUSION

In this survey the various approaches towards credit card fraud detection is been overviewed and a brief discussion of Hidden Markov Model is given which reflects the advantage and simplicity of HMM. The study shows that HMM works on human behavior while doing online shopping which will be a base for further enhancement of the technique, and resulting into a better detection method. The future work on this can to be to make HMM more secure and covering other aspects of human behavior.

V. REFERENCES

[1]*credit card fraud detection using hidden Markov Model.* **Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar.** 2008, Vol. 5.

[2]. *ONLINE CREDIT CARD FRAUD PREVENTION SYSTEM FOR DEVELOPING COUNTRIES.* **Rehab Anwer, Shiraz Baig, Dr. Malik Sikandar Hayat Khiyal, Aihab Khan & Memoona Khanum.** 2009-2010.

[3]. *HMM-based Integration of Multiple Models for Intrusion Detection.* **Chen Xiuqing, Zhang Y ongping, Tang Jiutao.** 2010.

[4]. *"CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection.* **E. Aleskerov, B. Freisleben, and B. Rao.** 1997, pp. 220-226.

[5]. *Minority Report in Fraud Detection: Classification of Skewed Data.* **C. Phua, D. Alahakoon, and V. Lee.** 2004.

[6]. *Distributed Data Mining in Credit Card Fraud Detection.* **W. Fan, A.L. Prodromidis, and S.J. Stolfo.** 1999.

[7]. *A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection.* **Tsai, C. Chiu and C.** 2004.

[8]. *Association rules applied to credit card fraud detection.* **D. Sa´nchez, M.A. Vila, L. Cerda , J.M. Serrano.** 2009.

[9]. *Neural Data Mining for Credit Card Fraud Detection.* **R. Brause, T. Langsdorf, and M. Hepp.** 1999.

[10]. *Parallel Granular Networks for Fast Credit Card Fraud Detection.* **M. Syeda, Y.Q. Zhang, and Y. Pan.** 2002.

[11]. *Mining Information from Credit Card Time Series for Timelier Fraud Detection.* **Hashemi, Leila Seyedhossein and Mahmoud Reza.** 2010.

[12]. *Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results.* **S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan.** 1997.

[13]. *Cost-based Modeling and Evaluation for Data Mining With Application to Fraud and Intrusion Detection: Results from the JAM Project. Salvatore* **J. Stolfo, Wei Fan, Wenke Lee, Andreas Prodromidis and Philip K. Chan, 1999**

[14]. *A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection.* **Kim, M.J. Kim and T.S.** 2002.

[15]. *A Comprehensive Survey of Data Mining-Based Fraud Detection.* **C. Phua, V. Lee, K. Smith, and R. Gayler,.** 2007.

[16]. *Agent-Based Distributed Learning Applied to Fraud Detection.* **Prodromidis, S. Stolfo and A.L.** 1999.

[17]. *Improving a credit card fraud detection system using genetic algorithm.* **M. Hamdi ozcelik, Ekrem Duman, Mine Islk and tugba cevik.** 2010.

[18]. *Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine.* **Ju, Qibei Lu and Chunhua.** 2011.

[19]. *Detecting credit card fraud by genetic algorithm and scatter search.* **Ozcelik, Ekrem Duman and M. Hamdi.** 2011.

[20]. *BOAT ADAPTIVE CREDIT CARD FRAUD DETECTION SYSTEM.* **Sherly K.K, R Nedunchezhian.** 2010.

[21]. *A Novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection.* **R. Huang, H. Tawfik, and A.K. Nagar.** 2010.

[22]. *Detection of Fraud Use of Credit Card by Extended VFDT.* **Tatsuya Minegishi, Ayahiko Niimi.** 2011.

[23]. *Multiple algorithms for fraud detection.* **R. Wheeler, S. Aitken.** 2000.

[24]. *Support vector machine based multiagent ensemble learning.* **Lean Yu a, Wuyi Yue , Shouyang Wang a, K.K. Lai.** 2010.

[25]. *Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum.* **Wang, Chun-Hua JU and Na.** 2009.

[26]. *BLAST-SSAHA Hybridization forCredit Card Fraud Detection.* **Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar.** 2009.

[27]. *Statistical Fraud Detection: A Review.* **Hand, Richard J. Bolton and David J.** 2002.

[28]. *A hybrid model for plastic card fraud detection systems.* **Krivko, M.** 2010.

Ms Anshul Singh BE Information Technology.2010 2-National Papers.

Devesh Narayan MTech Computer Technology 2004, pt RSU Raipur, BE Computer Science and Engg.2000 Amrawati University, 3- national papers.