Credit Card Fraud Analysis And Detection

Rahul V. Thakare, Prof. N. P. Chawande

Abstract— **Due to the rise and rapid growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. As credit card becomes the most popular mode of payment for both online as well as regular purchase, case of the fraud associated with it are also rising. The main objective of this paper is to construct an efficient fraud detection system which is useful for both online and regular purchase. In this proposed fraud detection system, the sequence of operation performed based on rules and probability of input action and compares each transaction with transactional history database.**

***Keywords*— Credit Card Fraud, Rule-based filter, Dempster-Shafer theory, Bayesian learner, Transactional History Database.**

I. INTRODUCTION

Credit card becomes very popular and convenient mode of payment for both online as well as regular purchase. Due to the increase in credit card usage, fraudsters are also finding more opportunities to commit fraud, which affects banks and card holders to great financial losses. To support safe credit card usages, efficient fraud detection system is essential.

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 i.e. card number, pin number, expiration date, address etc. 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 pattern 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 card holder is a promising way to reduce the rate of successful credit card frauds. Since humans tends 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, geographical location, etc. Deviation from such pattern is a potential threat to the system.

The rest of the paper is organized as follows: Section 2 describes the several techniques for the detection of credit card fraud. Section 3 presents details of our proposed credit card fraud detection system. In Section 4, we explain how proposed system is work. We conclude in Section 5 of the paper.

II. RELETED WORK

Credit card fraud detection has drawn lot of research interest and a number of techniques, with special emphasis on data mining and neural networks, have been suggested. Ghosh and Reilly have proposed credit card fraud detection with a neural network. Syeda et al. have used parallel granular neural networks [8] for improving the speed of data mining and knowledge discovery process in credit card fraud detection. Srivastava have proposed credit card fraud detection using hidden markov model [3]. It is initially trained with the normal behavior of a cardholder. HMM reduces the tedious work of an employee in bank since it maintains a log. Fuzzy Darwinian fraud detection system classifies credit card transactions into “suspicious” and “non-suspicious” classes. The complete system is capable of attaining good accuracy and intelligibility levels for real data and produces a low false alarm. Maes and Karl have proposed credit card fraud detection using Bayesian and Neural Network [11]. The credit card fraud detection using Bayesian and Neural network are automatic credit card fraud detection system by means of machine learning approach. These two machine learning approaches are appropriate for reasoning under uncertainty. Amlan have proposed BLAST-SSAHA Hybridization [2] for credit card fraud detection system. The performance of BLAHFDS is good and it results in high accuracy. At the same time, the processing speed is fast enough to enable online detection of credit card fraud. It counters frauds in telecommunication and banking fraud detection. Chan divides a large set of transactions into smaller subsets and then apply distributed data mining for building models of users behavior. The resultant base models are then combined to generate a meta-classifier for improving detection accuracy. Present CARDWATCH, a database mining system used for credit card fraud detection. The system based on neural learning module, provides an interface to variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix legitimate and fraudulent

Manuscript received July 23, 2014.

Rahul V, Thakare, Computer Engg, Mumbai University/ ACPCE,Kharghar, India. Rahul.thakare01@gmail.com

Prof. N. P. Chawande, Computer Engg, Mumbai University/ ACPCE,Kharghar, India. npchawande@acpce.com

transaction and the two main reasons for the complexity of credit card fraud detection. 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. The application of distributed data mining in credit card fraud detection, Brause et al. have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage.

III. EXISTING SYSTEMS

We studied following credit card fraud detection models and at end we propose our work.

A. BLAST-SSAHA Hybridization for Credit Card Fraud Detection

The Hybridization of BLAST and SSAHA algorithm [2] is refereed as BLAH-FDS algorithm. Sequence alignment becomes an efficient technique for analyzing the spending behavior of customers. Basic Local Alignment Search Tool (BLAST) and FAST-All (FASTA) are the two popular heuristic approaches for the local sequence alignment. Sequence Search and Alignment by Hash Algorithm (SSAHA) is one of the fastest tools for sequence alignment where the alignment process is performed in memory using hash table. BLAST and SSAHA are the efficient sequent alignment algorithms used for credit card fraud detection. BLAH-FDS is a two-stage sequence alignment algorithm in which a profile analyzer (PA) determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholders past spending sequences. The unusual transactions traced by the profile analyzer are passed to a deviation analyzer (DA) for possible alignment with past fraudulent behavior. The final decision about the nature of a transaction is taken on the basis of the observations by these two analyzers. BLAST-SSAHA Hybridization When a transaction is carried out, the incoming sequence is merged into two sequences time-amount sequence TA. The TA is aligned with the sequences related to the credit card in CPD. This alignment process is done using BLAST. SSAHA algorithm [9] is used to improve the speed of the alignment process. If TA contains genuine transaction, then it would align well with the sequences in CPD. If there is any fraudulent transactions in TP, mismatches can occur in the alignment process. This mismatch produces a deviated sequence D which is aligned with FHD. A high similarity between deviated sequence D and FHD confirms the presence of fraudulent transactions. PA evaluates a Profile score (PS) according to the similarity between TA and CPD. DA evaluates a deviation score (DS) according to the similarity between D and FHD. The FDM finally raises an alarm if the total score (PS - DS) is below the alarm threshold (AT).

B. Credit Card Fraud Detection using Hidden Markov Model

A Hidden Markov Model [3] is initially trained with the normal behavior of a cardholder. Each incoming transaction is submitted to the FDS for verification. FDS receives the card details and the value of purchase to verify whether the transaction is genuine or not. If the FDS confirms the transaction to be malicious, it raises an alarm and the issuing bank declines the transaction. The concerned cardholder may then be contacted and alerted about the possibility that the card is compromised. HMM never check the original user as it maintains a log. The log which is maintained will also be a proof for the bank for the transaction made. HMM reduces the tedious work of an employee in bank since it maintains a log. HMM produces high false alarm as well as high false positive.

C. Fuzzy Darwinian Detection of Credit Card Fraud

Fuzzy Darwinian Detection system [1] uses genetic programming to evolve fuzzy logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes. It describes the use of an evolutionary-fuzzy system capable of classifying suspicious and non suspicious credit card transactions. The system comprises of a Genetic Programming (GP) search algorithm and a fuzzy expert system. Data is provided to the FDS system. The system first clusters the data into three groups namely low, medium and high. The GP genotypes and phenotypes of the GP System consist of rules which match the incoming sequence with the past sequence. Genetic Programming is used to evolve a series of variable-length fuzzy rules which characterize the differences between classes of data held in a database. The system is being developed with the specific aim of insurance-fraud detection which involves the challenging task of classifying data into the categories:”safe” and ”suspicious”. When the customer’s payment is not overdue or the number of overdue payment is less than three months, the transaction is considered as non suspicious, otherwise it is considered as suspicious. The Fuzzy Darwinian detects suspicious and non -suspicious data and it easily detects stolen credit card Frauds.

D. Credit Card Fraud Detection Using Bayesian and Neural Networks

The credit card fraud detection using Bayesian and Neural Networks are automatic credit card fraud detection system by means of machine learning approach. These two machine learning approaches are appropriate for reasoning under uncertainty. An artificial neural network [7] [11] [12] [13] [14] consists of an interconnected group of artificial neurons and the commonly used neural networks for pattern classification is the feed forward network. It consist of three layers namely input, hidden and output layers. The incoming sequence of transactions passes from input layer through hidden layer to the output layer. This is known as forward propagation. The ANN consists of training data which is compared with the incoming sequence of transactions. The neural network is initially trained with the normal behavior of a cardholder. The suspicious transactions are then propagated backwards through the neural network and classify the suspicious and non suspicious transactions. Bayesian networks are also known as belief networks and it is a type of artificial intelligence programming that uses a variety of methods, including machine learning algorithms and data mining, to create layers of data, or belief. By using supervised learning, Bayesian networks are able to process data as needed, without experimentation. Bayesian belief networks are very effective for modeling situations where some information is already known and incoming data is uncertain or partially unavailable. This information or belief is used for pattern identification and data classification. A neural network learns and does not need to be reprogrammed. Its processing speed is higher than BNN. Neural network needs high processing time for large neural networks. Bayesian networks are supervised algorithms and they provide a good accuracy, but it needs training of data to operate and requires a high processing speed.

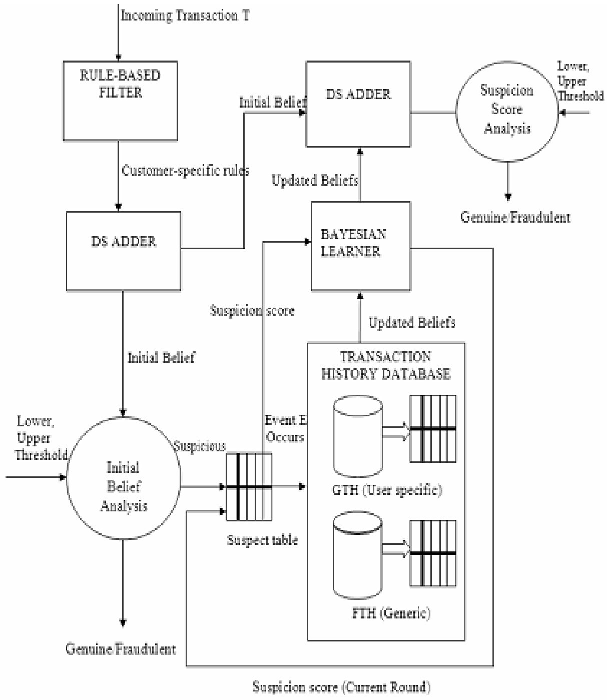
IV. PRAPOSED CREDIT CARD FRAUD DETECTION SYSTEM

We propose a credit card fraud detection system that combines different types of evidences using Dempster-Shafer theory. We propose the four components for our credit card fraud detection system as shown in fig 1. In the proposed FDS, a number of rules are used to analyze the deviation of each incoming transaction from the normal profile of the cardholder by assigning initial beliefs to it. The initial belief values are combined to obtain an overall belief by applying Dempster Shafer theory. The overall belief is further strengthened or weakened according to its similarity with fraudulent or genuine transaction history using Bayesian learning. In order to meet this functionality, the proposed FDS is designed with the following four major components,

- Rule-based filter
- Dempster Shafer adder.
- Transaction history database.
- Bayesian learner.

A. Rule-based filter (RBF)

The RBF consists of generic as well as customer-specific rules which classify an incoming transaction as fraudulent with a certain probability. It measures the extent to which the transactions behavior deviates from the normal profile of the cardholder. This layer can have rules like average daily/monthly spending of a customer shipping address being different from billing address, etc. Address mismatch, the most basic check performed by various credit card companies is billing address and shipping address mismatch. Orders could be shipped to an address different from the billing address. This check does not help us in declaring a transaction as fraudulent with complete certainty since a genuine cardholder could gift some items to his friend. However, a transaction that clears this check can be classified as genuine with very high probability. The transactions that violate this check are labeled as suspect.



B. Dempster Shafer Adder (DSA)

The role of the DSA is to combine evidences and compute an overall belief value for each transaction. For the credit card fraud detection problem, DST is more relevant as compared to other fusion methods since it introduces a third alternative: unknown, along with the measure of confidence in each of the alternatives. It provides a rule for computing the confidence measures of three states of knowledge: fraud, fraud and suspicious (unknown) based on data from new as well as old evidence. Furthermore, in DST, evidence can be associated with multiple possible events unlike traditional probability theory where evidence is associated with only one event. As a result, evidence can be more meaningful at a higher level of abstraction.

C. Trasaction History Database (THD)

THD is the transaction repository component of the proposed FDS. History records of both fraudulent and genuine transactions are used to construct models which allow us to extract characteristic information of the two classes from available data. For accomplishing this, we have built a good transactions history (GTH) for individual customers from their past behavior and a generic fraud transactions history (FTH) from different types of past fraud data. We represent each history transaction by a set of attributes containing information like card number, transaction amount and time since last purchase.

D. Bayesian learner

Bayesian learning is a tool to measure evidences supporting alternative hypotheses and arrive at optimal decisions. The general idea of belief revision is that, whenever new information becomes available, it may require updating of prior beliefs.

We studied the models [1] [2] [3] [7] [11], and compare these architectures with proposed architecture. Some parameters are same; the comparing parameters are listed in following table 1.

TABLE I. COMPARISON BETWEEN PROPOSED AND OTHER MODELS

Parameters	**BLAST-SSAHA**	**HMM**	**Fuzzy**	**ANN-BNN**	**Proposed System**
Method	Sequence Alignment	HMM	Genetic programming and fuzzy logic	Artificial intelligence	Machine learning
Accuracy	High	Medium	Very High	Medium	High
Processing speed	Very High	High	Low	High-Low	High
Cost	Inexpensive	Quite Expensive	High Expensive	Expensive	Quite Expensive
True positive rate	86%	70%	100%	70%-74%	98%

V. CONCLUSION

We proposed a novel credit card fraud detection system based on the integration of three approaches, namely, rule-based filtering, Dempster-Shafer theory and Bayesian learning. Dempsters rule is applied to combine multiple evidences from the rule-based component for computation of initial belief about each incoming transaction. The suspicion score is updated by means of Bayesian learning using history database of both genuine cardholder as well as fraudster. FTH is built from history data about past fraudulent behaviors detected by any credit card company. The FDS architecture is flexible so that new rules using any other effective technique can also be included at a later stage to further augment the rule-based component. In addition, Bayesian learning takes place so that the FDS dynamically adapts to the changing behavior of genuine customers as well as fraudsters over time. While combining rules using Dempster- Shafer theory gives good performance, especially in terms of true positives, Bayesian learning helps to further improve the system accuracy. The fusion of multiple evidences and learning are the appropriate approaches for credit card fraud detection.

REFERENCES

[1] Peter J. Bentley, Jungwon Kim, Gil-Ho Jung and Jong-Uk Choi,"Fuzzy Darwinian Detection of Credit Card Fraud", In the 14th Annual Fall Symposium of the Korean Information Processing Society, 14th October 2000.

[2] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," IEEE Transactions On Dependable And Secure Computing,vol. 6,Issue no.4,pp.309-315,Oct-Dec 2009.

[3] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model," IEEE Transactions On Dependable And Secure Computing, vol. 5, Issue no. 1, pp.37-48, January-March 2008.

[4] Aihua Shen, Rencheng Tong, Yaochen Deng, "Application of Classification Models on Credit Card Fraud Detection" IEEE Transactions 2007

[5] Sherly K.K, R Nedunchezhian, "BOAT ADAPTIVE CREDIT CARD FRAUD DETECTION SYSTEM", IEEE Transactions On Dependable And Secure Computing vol. 4,Issue no. 1, pp . 4244-5967, 2010.

[6] Jon T.S. Quah, M. Sriganesh, "Real-time credit card fraud detection using computational intelligence", 2007 Elsevier

[7] Y. Sahin, E. Duman, "Detecting Credit Card Fraud by ANN and Logistic Regression", IEEE Transactions On Dependable And Secure Computing, 2011

[8] Mubeena Syeda, Yan-Qing Zbang and Yi Pan, "Parallel Granular Neural Networks for Fast Credit Card Fraud Detection", IEEE Transactions On Dependable And Secure Computing, 2002.

[9] M. Hamdi Ozcelik, M. Mine Islk, Ekrem Duman, Tugba Cevik, "Improving a credit card fraud detection system using genetic algorithm", IEEE Transactions On Dependable And Secure Computing, 2010.

[10] R. Huang, H. Tawfik, and A.K. Nagar, "A Novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection", 2012 Published by Elsevier Ltd.

[11] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, " Credit card fraud detection using Bayesian and neural networks", in: Proceedings of the First International NAISO Congress on Neuro Fuzzy Technologies, 2002.

[12] Ghosh S., Reilly D. L., Credit card fraud detection with a neural network, In Proceedings of the 27th Hawaii International Conference on system Sciences, (1994).

[13] J.R. Dorronsoro, F. Ginel, C. Sanchez, C.S. Cruz, "Neural fraud detection in credit card operations", IEEE Transactions on Neural Networks, IEEE, 8 (July) (1997) 827834.

[14] R.C. Chen, S.T. Luo, X. Liang, V.C.S. Lee, "Personalized approach based on SVM and ANN for detecting credit card fraud", in: Proceedings of the IEEE International Conference on Neural Networks and Brain, IEEE, October 2005, pp. 810815.

[15] R. Brause, T. Langsdorf, M. Hepp, "Neural data mining for credit card fraud detection", in: Proceedings of the International Conference on Tools with Artificial Intelligence, 1999, pp 103106.

[16] C. Chiu, C. Tsai, "A web services-based collaborative scheme for credit card fraud detection", in: Proceedings of the IEEE International Conference on e-Technology, e- Commerce and e-Service, IEE, 2004, pp. 177181.

[17] http://yesican.chsoft.biz/dnews-manual/rules.htm

[18] http://en.wikipedia.org/wiki/Bayesian_inference

[19] http://www.blutner.de/uncert/Dempster-Shafer.pdf

Mr. R. V. Thakare received the B.E degree in Computer Engineering from K. K. Wagh College of Ebgineering, Nashik, India, in year 2008. He is doing his M.E in Computer Engineering at Mumbai University, Mumbai,India. He has published various papers in the area of Computer Networking.

Prof. N. P. Chawande received the M.E degree in Computer Engineering from Umiversity of Mumbai, India. He is currently working as Associate Professor in Department of Computer, ACPCE,Kharghar,Mumbai,India. He has published various papers in the area of Computer Networking.