A Novel Idea for Credit Card Fraud Detection using Decision Tree

Prajal Save
St. John College of Engineering and Management
Palghar

Pranali Tiwarekar
St. John College of Engineering and Management
Palghar

Ketan N. Jain
St. John College of Engineering and Management
Palghar

Neha Mahyavanshi
St. John College of Engineering and Management
Palghar

ABSTRACT
Online shopping and banking has increased by the growth of internet and by use of credit card. Along with this number of credit card fraud is also increased. Many modern techniques based on Artificial Intelligence, Data warehousing has evolved in detecting various credit card fraudulent transactions. We proposed a system which detect fraud in credit card transaction processing using a decision tree with combination of Luhn's algorithm and Hunt's algorithm. Luhn's algorithm is used to validate the card number. Address matching rule checks whether the Billing Address and Shipping Address match or not. This check does not guarantee whether a transaction is fraud or genuine. But if the two addresses match, the transaction can be classified as genuine with a high probability. Else, the transaction is labelled as suspect. A customer usually carries out similar types of transactions in terms of amount, which can be visualized as part of a cluster. Since a fraudster is likely to differ from the customer's account, his transactions can be detected as exceptions to the cluster – a process known as outlier detection.

General Terms
Credit card fraud, online Transaction, Electronic Commerce

Keywords
Electronic Commerce, Credit card fraud, address matching, spending pattern, Luhn's Algorithm, Outlier Detection, Heuristic function, Bayes Theorem.

1. INTRODUCTION
With rapid advancement of e-commerce, use of credit cards for purchases has exponentially increased. Unfortunately, fraudulent use of credit cards has also become a source of crime. Credit card fraud is a most popular term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. Credit card fraud is also an appendage to identity theft. According to the United States Federal Trade Commission, while identity theft had been holding steady for the last few years, it saw a 21 percent increase in 2008. However, credit card fraud, that crime which most people's privilege with ID theft, decreased as a percentage of all ID theft complaints for the sixth year in a row.Financial institutions employ various fraud prevention models for tackling this problem. But fraudsters are adaptive, and given time, they devise several ways to intrude such protective models. Despite the best efforts of the financial institutions, law enforcement agencies and the government, credit card fraud continues to rise. Fraudsters nowadays may constitute of a very inventive, intellect and fast moving fraternity. Several techniques for the detection of credit card fraud have been proposed in the last few years.

2. LITERATURE SURVEY
Syeda et al. [1] have suggested the use of parallel granular neural networks for speeding up the data mining and knowledge discovery process. Maes et al. [2] have outlined an automated credit card fraud detection system by Artificial Neural Network - ANN as well as Bayesian belief networks - BBN. They show that BBN gives better results related to fraud detection and the training period is faster whereas the actual detection process is substantially faster with ANN. The neural network based methods are, in general, fast but not so accurate an retraining the neural networks is quite taxing.

Chen et al. [3] propose a method in which an online questionnaire is used to collect questionnaire-responded transaction (QRT) data of users. Further it uses a support vector machine (SVM) trained with this data and the QRT models are used to predict new transactions. Chen et al. [4] have recently presented a personalized approach for credit card fraud detection that employs both SVM and ANN. It tries to prevent fraud for users even without any transaction data. However, these systems are not fully automated and depend on the user"s expertise level.

Chan et al. [5] divide a large set of transactions into smaller subsets and then apply distributed data mining for building models of user behavior. The resultant base models are then combined to generate a meta-classifier for improving detection accuracy. Brause et al. [6] have explored the possibility of combining advanced data mining techniques and neural networks to obtain high fraud coverage along with a low false alarm rate. Use of data mining is also developed by Chiu and Tsai [7] . They consider web services for data exchange among banks. A fraud pattern mining (FPM) algorithm has been elaborated for mining fraud association rules which give information regarding the features that exist in fraud transactions. Banks enhance their original fraud detection systems by using the new fraud patterns to halt attacks. While data mining techniques are relatively accurate, they are inherently slow.

Aleskerov et al. [8] present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and combination of legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection [9] . 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 wrong detections

3. RELATED WORK

Various existing fraud techniques majorly explore decision trees, genetic algorithms, clustering techniques and neural networks. Recently there has been increased in use of parallel granular neural networks for accelerating the data mining and knowledge discovery methods. It seems that BBN (Bayesian Belief networks) gives better results related to fraud detection and the training period is faster whereas the actual detection process is substantially faster with ANN. The neural network based methods are, in general, fast but not so accurate and retraining the neural networks is quite taxing.

Genetic algorithms was first introduced by Holland(1975). Genetic algorithms obtain better solutions as time progresses. Fraud detection problem is classification problem, in which some of statistical methods many data mining algorithms have proposed to solve it. Although decision trees are more popular. Fraud detection has been usually in domain of E-commerce, data mining [10]. GA is used in data mining most probably for variable selection [11] and is mostly coupled with other DM algorithms [12]. And their combination with other techniques has a very good performance. GA has been used in credit card fraud detection for reducing the wrongly classified number of transactions [12]. And is easy accessible for computer programming language implementation, thus, make it powerful in credit card fraud detection. But this method has high performance and is quite expensive.

A Hidden Markov Model is a double embedded stochastic process used to model much more complicated stochastic processes as compared to a traditional Markov model. If an incoming credit card transaction is denied by the trained Hidden Markov Model with sufficiently high probability, it is considered to be fraudulent transactions. HMM[13], Baum Welch algorithm is used for training purpose and K-means algorithm for clustering purpose. HMM stores data in the form of clusters depending on three price value ranges low, medium and high[14]. The probabilities of initial set of transaction have chosen and Fraud Detection System checks whether transaction is genuine or fraudulent. Since HMM maintains a log for transactions it reduces tedious work of employee but produces high false alarm as well as high false positive[15]

Neural networks have been generally used in fraud detection. Neural network is a set of Connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights to guess the correct class labels. Fraud detection methods based on neural network are the most popular ones. An artificial neural network [16][17] consists of an interconnected group of artificial neurons. The principle of neural network is inspired by the functions of the brain especially pattern recognition and associative memory [18]. The neural network diagnose similar patterns, predicts future values or events based upon the associative memory of the patterns it was learned. It is generally applied in classification and clustering. The advantages of neural networks over other techniques are that these models are able to learn from the past and thus, enhance results as time passes. They can also extract rules and predict future activity based on the current situation. By employing neural networks, effectively, banks can detect fraudulent use of a card, faster and more efficiently. Through the reported credit card fraud studies most have focused on using neural networks. In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to design complex relationships between inputs and outputs or to find patterns in data

4. PROPOSED SYSTEM

The proposed model comprised of six steps. Firstly, Luhn's Test is used to validate card numbers. Then, two rules ie. Address Mismatch and Degree of Outlierness are used to analyze the deviation of each incoming transaction from the normal profile of cardholder. These two steps compute initial beliefs. The initial belief values are combined to obtain an overall belief by applying Advanced Combination Heuristic in step four. Step five looks into the spending history to extract characteristic information about genuine and fraud transactions. The overall belief is further strengthened or weakened in the final step using Bayes‟ Theorem, followed by recombination of the calculated probability with initial belief of fraud using advanced combination heuristic

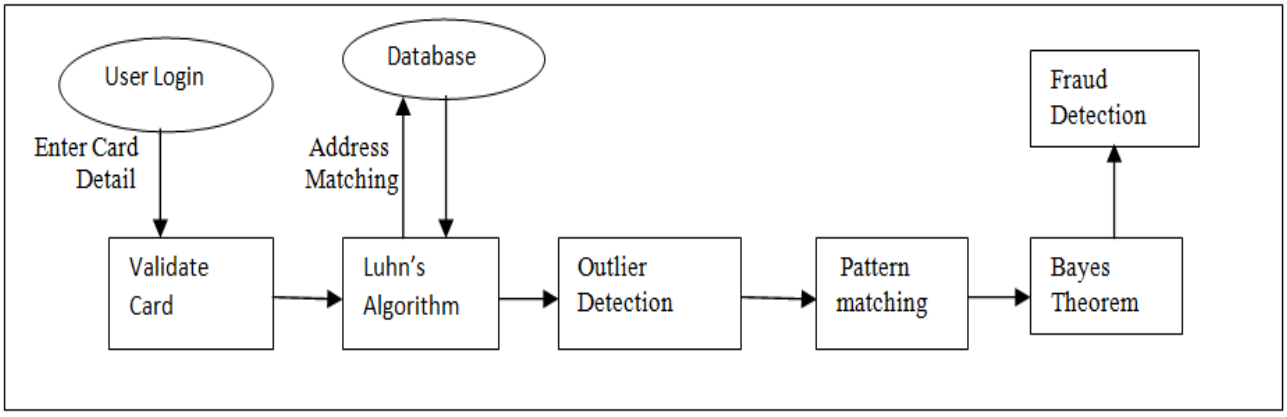

Fig 1: Architecture Diagram

Card Number Validation

Luhn's Algorithm is used to validate card numbers that distinguishing valid numbers from mistyped or otherwise incorrect numbers.

Following standard algorithm is used to validate credit card numbers, [14]

1. Reverse the order of the digits in the number.
2. Take the first, third, ... and every other odd digit in the reversed digits and sum them to form the partial sum S1.
3. Taking the second, fourth and every other even digit in the reversed digits. Multiply each digit by two and sum

the digits if the answer is greater than nine to form partial sums for the even digits.

4. Sum the partial sums of the even digits to form S2.

If S1+S2 ends in zero,then the original number is in the form of a valid credit card number as verified by the

Luhn test.
For example, if the trial number is 49927398716,

1. Reverse the digits:
   61789372994
2. Sum the odd digits:
   6 + 7 + 9 + 7 + 9 + 4 = 42 = s1
3. The even digits:
   1, 8, 3, 2, 9

   Two times each even digit:

   2, 16, 6, 4, 18

   Sum the digits of each multiplication:

   2, 7, 6, 4, 9
4. Sum the last:
   2 + 7 + 6 + 4 + 9 = 28 = s2
5. S1+S2 = 70 which ends in zero which means that 49927398716 passes the Luhn's test.

Address Verification

This step is used for comparing Billing Address with Shipping Address and the check is whether it matches or not. This check does not guarantee whether a transaction is fraud or genuine. But if the two addresses match,t he transaction can be classified as genuine with a high probability. Else, the transaction is labeled as suspect.

Outlier Detection

We have used DBSCAN (Density Based Spatial Clustering of Application with Noise) to generate clusters, using transaction amount as attribute. Any incoming transaction amount, that does not belong to any cluster is detected as fraudulent. These two steps compute initial belief.

Advanced Combination Heuristic Function

The initial belief values are combined to obtain an overall belief.

Spending History Databases

It comprises of genuine Transaction Record(for individual customers from their past behaviour) and Fraud Transaction Record(from different types of past fraud data).We represent each history transaction by set of attributes containing information like card number, transaction amount and time since last purchase. to extract characteristic information about genuine and fraud transactions.

Bayes Theorem

The idea of belief revision is that, whenever new information becomes available, it may require updating of prior beliefs. Bayes Theorem theorem expresses how a subjective degree of belief should rationally change to account for availability of related evidence.

5. RESULT AND ANALYSIS

Credit Card verification is done by using Luhn's algorithm which checks whether the entered card number is valid or not.

The remaining phases of the proposed system will be implemented in future work.

```
Enter Credit Card Number=4577044401759066
Pass Luhn's Test
```

Input to the system is credit card number which performs luhn's algorithm process as explained in credit card validation step .The output for this shows whether it passes the luhn's test or not.

6. CONCLUSION

In this paper we have brief discussion on credit card fraud detection. Here we have shown how the system detect whether an incoming transaction is fraud or genuine.In our proposed model, we have found out validation of card are genuine and very low false alarm.The relative studies and our results sure that the correctness and effectiveness of the proposed system is secure.

7. REFERENCES

[1] M. Syeda, Y.Q. Zhang, Y. Pan, "Parallel granular neural networks for fast credit card fraud detection", Proceedings of the IEEE International Conference on Fuzzy Systems, 2002, pp. 572–577.

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

[3] R.C. Chen, M.L. Chiu, Y.L. Huang, L.T. Chen, "Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines", Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, vol. 3177, October 2004, pp. 800–806.

[4] R.C. Chen, S.T. Luo, X. Liang, V.C.S. Lee, "Personalized approach based on SVM and ANN for detecting credit card fraud", Proceedings of the IEEE International Conference on Neural Networks and Brain, October 2005, pp.810–815.

[5] P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo, "Distributed data mining in credit card fraud detection", Proceedings of the IEEE Intelligent Systems, 1999, pp.67–74.

[6] R. Brause, T. Langsdorf, M. Hepp, "Neural data mining for credit card fraud detection", Proceedings of the International Conference on Tools with Artificial Intelligence, 1999, pp. 103–106.

[7] C. Chiu, C. Tsai, "A web services-based collaborative scheme for credit card fraud detection", Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004, pp. 177–181.

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

[9] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. Int"l Conf.Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

[10] K.RamaKalyani, D.UmaDevi, "Fraud Detection of Credit Card Payment System by Genetic Algorithm",

International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012.

[11] Bidgoli, B. M., Kashy, D., Kortemeyer, G. & Punch, W. F "Predicting student performance: An Application of data mining methods with the educational web-based. (2003). System LON-CAPA". In Proceedings of ASEE/IEEE frontiers in education conference. . (2003).

[12] Ekrem Duman, M. Hamdi Ozcelik "Detecting credit card fraud by genetic algorithm and scatter search". Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).

[13] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, "CreditCard Fraud Detection Using Hidden Markov Model" IEEE, Transactions On Dependable And Secure Computing, Vol. 5, No 1. , January-March 2008

[14] V. Bhusari, and S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011

[15] V.Bhusari ,S.Patil ," Study of Hidden Markov Model in Credit Card Fraudulent Detection ",International Journal of Computer Applications (0975 - 8887) Volume 20-No.5, April 2011

[16] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods". IEEE-International Conference on Computer, Communication and Electrical Technology; (2011). (152-156).

[17] Ray-I Chang, Liang-Bin Lai, Wen-De Su, Jen-Chieh Wang, Jen-Shiang Kouh "Intrusion Detection by Backpropagation Neural Networks with Sample-Query and Attribute-Query". Research India Publications; (2006). (6-10).

[18] Raghavendra Patidar, Lokesh Sharma "Credit Card Fraud Detection Using Neural Network". International Journal of Soft Computing and Engineering (IJSCE), (2011). Volume-1, Issue; (32-38).