ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

Design & Analysis of Credit Card Fraud Detection Based On HMM

Ranjit kumar, Sandeep Raj
Asst.Prof, Galgotias College of Engineering &Technology, Gr.Noida

Abstract-The growth of e-commerce increases the money transaction via electronic networks day by day which is designed for hassle free, fast & easy money transaction. But the facility involves greater risk of misuse of facility for fraud; one of them is credit card fraud. It can happen by many ways such as by stolen card, by Internet hackers who can hack your system & get important information about your card or by information leakage during the transaction. Although many people haven proposed their work for credit card fraud detection by characterizing the user spending profile, but in this paper we are proposing the HMM based method with multiple involvements and also including several fields of user profile instead of only spending profile & the simulation result shows improvement in TP (True Positive), TN (True Negative) rates and it also decreases the FP (False Positive) &FN (False Negative) rate.

Keywords: **Fraud detection, HMM (Hidden Markov Model).**

I. INTRODUCTION

Growth in communication network, increased internet speed, easy wireless connectivity & lack of time causes the people to buy through electronic network. Here are some statistics and projections of the Indian credit card industry (http://hubpages.com/hub/Indian-Credit-card-Industry) to show importance of the topic.

1. India is currently the fastest growing Mobile Market in the world and is also among the fastest growing credit card markets in the world.

2. India has a total approx.75 million cards under circulation (25 million credits and 50 million debits) and a 30% year-on-year growth.

3. With 87% of all transactions in plastic money happening through credit cards, debit cards in India continue to be used largely for cash withdrawals.

4. Though Visa, which accounts for 70% of the total card industry is the market leader in India; MasterCard is fast catching up.

5. Every transaction involves payment of an interchange charge to MasterCard or Visa for settlement, which amounted to about $50 million during the year.

6. Internal estimates of Barclaycard have pegged the Indian market with potential to grow to at least 55million credit cards by 2010-11.

The above statistics shows the money involved in transaction through cards & it is required to insure the security of money for both the Bank & for customer.

II. RELATED WORK

As we stated before that many persons have proposed their work on same field some of which we have studied & we think most relevant to our topics are, the work done by AbhinavSrivastava, AmlanKundu, ShamikSural, Arun K. Majumdar[3] has proposed the probabilistic model based on HMM(Hidden Markov Model). They consider the spending history of card holder & characterize the spending pattern by dividing the transaction amount in three categories which shows the TP rate of 0.65 & FP rate of 0.05 and another paper published by Wen-Fang YU & Na Wang [1] proposed the distance based method. This method judges whether it is outlier or not according to the nearest neighbours of data objects. They only showed the highest accuracy of about 89.4 percent but did not talk about FP & FN. A neural network based approach is presented by SushmitoGhosh and Douglas L. Reilly [9]. In their paper they selected large set of 50 fields & after proper relation it is reduced to set of 20 features which is used for training neural network. The neural network used in this fraud detection feasibility study is the P-RCE neural network. The P-RCE is a member of the family of radial-basis function networks that has been developed for application to pattern recognition. The P-RCE is a three-layer, feed-forward network that is distinguished by its use of only two training passes through the data set. Same work is also done by using regression techniques & compared against neural & decision tree methods [4]. This work is done by AihuaShen, Rencheng Tong, Yaochen Deng. Their simulation shows that neural networks model provides higher lift (Lift table and lift chart were used to describe the usefulness of the model to create the scored data set. "Lift" is probably the most commonly used metric to measure the performance of targeting models in classification applications) than a logistic regression and decision tree on the same data, and is slightly better than logistic regression. This provides a key factor in choosing the models. A similar coefficient sum based model analysis was explained by Chun HuaJu& Na wang[2]. They analyzed type I & type II error rate with highest rate of TP up to 89 percent.

III. HMM (HIDDEN MARKOV MODEL)

A Hidden Markov Models is a finite set of states; each state is linked with a probability distribution. Transitions

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

among these states are governed by a set of probabilities called transition probability. In a particular state a possible outcome or observation can be generated which is associated with symbol of observation probability. It is only the outcome, not the state that is visible to an external observer and therefore states are hidden to outside hence the names Hidden Markov Model.

A. Credit Card Fraud Detection Using HMM

In this section, the system of credit card detections based on Hidden Markov Model, which does not require fraud signature and still is capable to detect fraud just by bearing in mind a cardholder's spending habit. The particulars of purchased items in single transaction are generally unknown to any credit card fraud detection system running either at the bank that issues credit cards to the cardholder or at the merchant site where goods is going to be purchased. Each arriving transaction is submitted to the fraud detection for verification purpose. The fraud detection system accepts the card details such as Card number ,CVV number ,card type, expiry date and amount to validate ,whether the transaction is genuine or not, and the implementation techniques of Hidden Markov Model, in order to detect fraud transaction through credit card. It creates cluster of training sets and identifies the spending profile of cardholder. The number of items purchased, types of items that are bought in particular transaction are not known to the fraud detection system, but it only concentrates on the amount of items purchased and use for further processing. It stores data of different amount of transaction in the form of clusters depending on transaction amount. It tries to find out any variance in the transaction based on spending profile of cardholder. The probabilities of initial set chosen are based on spending profile of card holder and construct a sequence for further processing. In various periods of time, purchase of various types with the different amount would be made by credit holder. It uses the deviation in purchasing amount of latest transaction sequence which is one of the possibilities related to the probability calculation.

IV. PROPOSED ALGORITHM

Here we detail the proposed algorithm for classification of Fraud Transactions.

Step 1: Read the given data generated by synthetic method.

Step 2: Re-categorize the data into five groups as transaction month, date, day, amount of transaction & difference between successive transaction amounts.

Step 3: Make each transaction data as vector of five fields.

Step 4: Make two separate groups of data named True & False transaction group (if false transaction data is not available add randomly generated data in this group).

Step 5: Train Hidden Markov Model.

Step 6: Save the classifier.

Step 8: Read the current Transaction.

Step 7: Repeat the process from **step1** to **step3** for current transaction data only.

Step 8: Place the saved classifier & currently generated vector in classifier.

Step 9: Take the generated decision from the classifier.

V. IMPLEMENTATION

Since there is no real data available because of privacy maintained by banks, hence for testing of implementation of our algorithm we generated the data of true & false Transaction using different mean & variance & then mixed them with different probabilities. We used the MATLAB for the implementation of the algorithm because of its rich sets of mathematical functions and also supporting the inbuilt functions for HMM.

VI. RESULTS

The results are simulated for five different Fraud probabilities from 0.3 to 0.5 & changing the training data size from 30 to 100, then according to outputs of program the following tables are drawn which show

TPR = True Positive Rate
TNR = True Negative Rate
FPR = False Positive Rate
FNR = False Negative Rate

Complete details of these parameters are discussed in(http://en.wikipedia.org/wiki/Receiver_operating_characteristic).

Hidden Markov Model for Accuracy Calculation

Total Data	Fraud Prob.	TPR	TNR	FPR	FNR	Accuracy
30	0.30	0.90	0.72	0.15	0.18	0.83
30	0.40	0.61	0.59	0.38	0.41	0.60
30	0.50	0.26	0.77	0.33	0.50	0.56
60	0.30	0.98	0.22	0.38	0.03	0.72
60	0.40	0.77	0.61	0.32	0.26	0.70
60	0.50	0.70	0.75	0.29	0.24	0.73
100	0.30	0.89	0.27	0.39	0.20	0.67
100	0.40	0.65	0.43	0.51	0.38	0.54
100	0.50	0.81	0.48	0.38	0.24	0.67

This shows maximum accuracy of up to 83%, maximum TPR(99%), maximum TNR(98%) & maximum FPR(7%), maximum FNR(6%), it also behaves almost same for all types of data set generated (having very low fraud data & high fraud data).

VII. CONCLUSION

Referring to results we can say that proposed algorithm gives the better results in comparison with the previous papers we have discussed before & hence can be used for automatic Credit Card Fraud detection with excellent accuracy & minimum false alarm.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

The Hidden Markov Model makes the processing of detection very easy and tries to remove the complexity.

REFERENCES

[1] Research on Credit Card Fraud Detection Model Based on Distance Sum Wen-Fang YU, Na Wang 2009 International Joint Conference on Artificial Intelligence.

[2] Research on Credit Card Fraud Detection Model Based on Similar Coefficient Sum Chun-Hua JU, Na Wang 2009 First International Workshop on Database Technology and Applications.

[3] Credit Card Fraud Detection Using Hidden Markov Model by AbhinavSrivastava, AmlanKundu, ShamikSural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 5, NO. 1, JANUARY-MARCH 2008.

[4] Application of Classification Models on Credit Card Fraud Detection AihuaShen, Rencheng Tong, Yaochen Deng School of Management, Graduate University of the Chinese Academy of Sciences, Beijing, 100084, China, 2007 IEEE.

[5] PHUA , C ., Lee , V ., smith , k and gayler , R ., 2007 . A comprehensive survey of data mining based fraud detection research.

[6] CHIU ,C ., and T sai , C 2004 A web services –based collaborative scheme for credit card fraud detection proceedings of IEEE international conference e-technology, e-commerce and e-services.

[7] V. Kecman, Learning and Soft Computing: Support Vector Machines, Neutral Networks and Fuzzy Logic Models. Cambridge, MA: MIT Press, 2001.

[8] Brause, R., Langsdorf, T., and Hepp, M., 1999. Neural data mining for credit card fraud detection, proceeding of IEEE international conference tools with artificial intelligence (1999).

[9] V. N. Vapnik, the Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.