

PROPOSED DISTRIBUTED DATA MINING IN CREDIT CARD FRAUD DETECTION

Priyanka Yadav¹, Pavan Wangade², Manish Thakur³, Mohammed Fakhri⁴, Gayatri Hegde⁵

¹²³⁴Department of Information Technology, PIIT, Maharashtra, India

⁵Professor, Department of Information Technology, PIIT, Maharashtra, India

Abstract-Data Mining is popularly used to combat frauds because of its effectiveness. Using data mining techniques or model the Credit card fraud can be detected. The Credit card provides cashless shopping all over the world. Hence the risk of fraud using Credit card is been increasing. The clustering model used to classify the legal and fraudulent transaction using data cauterisation of region of parameter. Hidden Markov Model can detect whether an incoming transaction is fraudulent or not with low false alarm. HMM does not require fraud signatures and still it is capable to detect the frauds by bearing in mind a cardholder's spending habit. An HMM is initially trained with the normal behaviour of a cardholder. If an incoming credit card transaction is not accepted by the trend HMM with sufficiently high probability, it is consider to be fraudulent. At the same time, try to ensure that genuine transactions are not rejected. The financial losses due to Credit card fraud effect not only the individuals but also the merchants. Therefore the security measures need to be taken to detect the Credit card fraud.

1. INTRODUCTION

The usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. Many E-commerce website use the Credit card payment method for transaction or buying purpose. Credit card transactions take place over an Internet and hence the fraud and risks related to it is also increasing day by day. The organizations has to build systems which can detect the fraud before it takes place, as now the fraud is done and then the user understand that fraudulent has happened. In the credit card fraud, only some important information about a card (card number, expiration date, secure code) 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 way to detect this kind of fraud is to analyse the spending patterns 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 cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviourist 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, etc. Deviation from such patterns is a potential threat to the system. This system is mainly use by bank organization and also by credit card providers to detect the fraudulent transaction. Hidden Markov Model will be helpful to find out the fraudulent transaction by using spending profiles of user. It works on the user spending profiles.

2. LITERATURESURVEY

Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on neural networks, data mining and distributed data mining have been suggested.

Metalearning [1][3] is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A metaclassifier is thus 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. They use Java agents for Meta learning (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. Alekerov et al. 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.

Ghosh and Reilly [2] have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labelled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and non-received issue (NRI) fraud. Recently, Syeda et al. 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. Stolfo et al. 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.

Sam and Karl [4] suggest a credit card fraud detection system using Bayesian and neural network techniques to learn models of fraudulent credit card transactions. Kim and Kim have identified skewed distribution of data and mix of Legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection. This paper investigates the usefulness of applying different learning approaches. Clustering helps in grouping the data into similar clusters that helps in uncomplicated retrieval of data. Cluster analysis is a technique for breaking data down into related components in such a way that patterns and order becomes visible. This model is based on the use of the parameters' data cauterisation regions.

Vatsa et al. [5] have recently proposed a game-theoretic approach to credit card fraud detection. They model the interaction between an attacker and an FDS as a multi stage game between two players, each trying to maximize his payoff. HMM-based applications are common in various areas such as speech recognition, bioinformatics, and genomics. In recent years, Joshi and Phoba have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. An HMM-based intrusion detection system that improves the modelling time and performance by considering only the privilege transition flows based on the domain knowledge of attacks.

Phua et al. [6] suggest the use of metaclassifier similar to in fraud detection problems. They consider naïve Bayesian, 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.

3. EXISTING SYSTEM

Credit Card Fraud detection works on the concept of trying to find out a fraud in the online transaction done by a user. Various Systems is generated for this purpose:

3.1 Dempster-Shafer Theory and Bayesian Learning:

Dempster-Shafer Theory basically proposes Fraud Detection System using information fusion and Bayesian learning in which evidences from current as well as past behaviour are

combined together and depending on certain type shopping behaviour establishes an activity profile for every cardholder. It requires large amount of probability data and the independence of evidences assumptions is often not valid. Relationship between hypothesis and evidence is reduced to a number and also the explanations for the user difficult.

3.2 Bayesian and Neural Networks:

Bayesian and Neural network approach is automatic credit card fraud detection system and type of artificial intelligence programming which is based on variety of methods including machine learning approach, supervised and data mining for reasoning under uncertainty. This approach has Difficulty to confirm the structure. It requires Excessive training and efficiency of training and so on.

3.3 Evolutionary Fuzzy System:

The Evolutionary-Fuzzy system which uses genetic programming for evolving fuzzy logic rules. It classifies the transactions into suspicious and non-suspicious. It comprises of Genetic Programming (GP) search algorithm and a fuzzy expert system. This approach has very high accuracy and produces a low false alarm. It is not applicable in online transactions also it is highly expensive and processing speed is low.

4. PROPOSED SYSTEM

The propose system is the Hidden Markov Model (HMM) which is used to detect the credit card fraud by maintaining the database for spending behaviours of cardholder's. The details of items purchased in Individual transactions are usually not known to any Fraud Detection System (FDS) running at the bank that issues credit cards to the cardholders. Hence, the use of HMM is an ideal choice for addressing this problem. The system has two security levels i.e. the security questions and the OTP. The questions are asked while registration based on some personal details.

HMM-based approach is a drastic reduction in the number of False Positives transactions identified as malicious by an FDS although they are actually genuine. An FDS runs at a credit card issuing bank. 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. The types of goods that are bought in that transaction are not known to the FDS. It tries to find any anomaly in the transaction based on the spending profile of the cardholder, transaction amount, and no. Of transactions, etc. If the FDS confirms the transaction to be of fraud, it raises an alarm, and the issuing bank declines the transaction.

Advantage:

- The detection of the fraud use of the card is found much faster than the existing system.
- In case of the existing system even the original card holder is also checked for fraud detection. But in this system no need to check the original user as there is a log.
- The log which is maintained will also be a proof for the bank for the transaction made.
- This system can find the most accurate detection using this model.
- This reduce the tedious work of an employee in the bank

5. HMM MODEL

A Hidden Markov Model is a finite set of states; each state is linked with a probability distribution. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. To map the credit card transaction processing operation in terms of an HMM, it start with first deciding the observation symbols in our model. The actual price range for each symbol is configurable based on the spending habit of individual cardholders. These price ranges can be determined dynamically by applying a clustering algorithm on the values of each cardholder's transactions. It is only the outcome, not the state that is visible to an external observer and therefore states are "hidden" to the outside; hence the name Hidden Markov Model.

A credit cardholder makes different types of purchases of different amounts over a period of time. One possibility is to consider the sequence of transaction amounts and look for deviations in them. However, the sequence of types of purchase is more stable compared to the sequence of transaction amounts. The reason is that, a cardholder makes purchases depending on his need for procuring different types of items over a period of time. This, in turn, generates a sequence of transaction amounts. Each individual transaction amount usually depends on the corresponding type of purchase. The HMM model keeps the tracks of all transactions of cardholders like the amount of transaction, type of purchase, no. of transaction over a period of time, etc. The first 3 transactions are made with security questions and then based on the last transactions the HMM creates threshold value, every after transaction the threshold value is updated and new value is set. During transaction the threshold value is checked with the transaction amount if it is greater than the threshold value then the user has to pass through the security levels. The cardholder has to give OTP

for completing the process if he/she fails to provide OTP then the transaction is blocked.

Advantages:-

- Simple and easy to understand.
- Does not require high speed processing system.
- Produces maximum true positive alarms.

6. SYSTEM ARCHITECTURE

In system architecture represents actual flow of credit card fraud detection system. User can carry out online transactions safely with the help of HMM model. If the user/cardholder has already registered details then he/she can directly login and if user has not registered details then he has to register. In which our system will take users personal details, bank accounts and security questions. Note that while doing so user must keep in mind the security answers that he/she has answered while registration. After registration user can carry out transactions.

During first three transactions system will ask for OTP which is generated and send to user mobile no which he registered during registration process. Once the OTP is verified correctly then transaction processed and if not then card will get blocked and cardholder will get alert message. Once user has carried out three transactions then system will save details of each transaction in its database. While doing next transaction system will check the pattern in which user is doing transactions by using HMM model. If system catches any unusual transaction then our system will authenticate the user by security questions and one time password (OTP). If the security questions are answered correctly then user can process with transaction otherwise transaction is stopped immediately, card holder gets alert message and credit card is blocked.

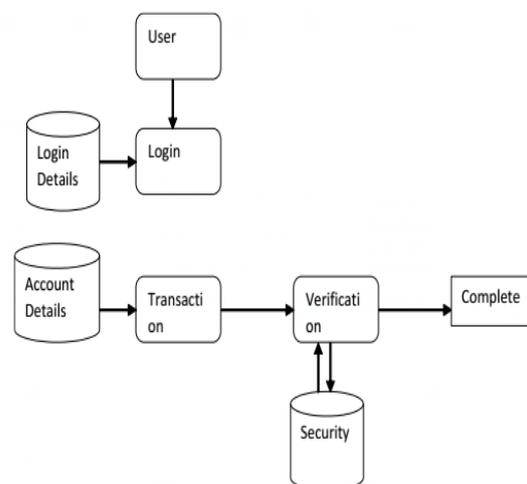


Fig1. System Architecture

7. CONCLUSION

HMM model is used to detect fraud activities on credit card. The model maintains the database in which users transaction behaviours and pattern are saved and if any unusual transaction is carried out which is different from passed behaviour of that user then system raise alarm and the transaction is blocked. The Hidden Markov Model makes the processing of detection very easy and tries to remove the complexity. Hence the security is maintained and transactions are secure from fraud.

9. REFERENCES

- [1] Aashlesha Bhingarde, Avnish Banger, "Credit Card Fraud Detection using Hidden Markov Model", *IJRCCCE*, ISSN: 2278-1021, vol.4, March 2015.
- [2] Bilonikar Priya, Deokar Malvika, "Survey on Credit Card Fraud Detection Using Hidden Markov Model", *IJRCCCE*, Vol. 3,

Issue 5, May 2014.

- [3] Shailesh S. Dhok, "Credit Card Fraud Detection using Hidden Markov Model", *IJSCE*, ISSN:2231-2307, Vol. 2.March 2012.
- [4] V. Dheepa, Dr. R. Dhanapal "Analysis of Credit Card Fraud Detection Methods" *IJRTE*, Vol 2, No. 3, November 2009.
- [5] V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-theoretic Approach to Credit Card Fraud Detection," *Proc. First Int'l Conf. Information Systems Security*, pp. 263-276, 2005
- [6] C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 50-59, 2004.
- [7] bsys [Online] .Available:
www.bsys.monash.edu.au/people/cphua