ISSN 2349-6193
Impact Factor: 2.805

IJESMR

International Journal OF Engineering Sciences &Management Research

COMPARATIVE STUDY OF CREDIT CARD FRAUD DETECTION USING HIDDEN MARKOV MODEL

Divyanshu Kakar*1, Prabhjit Singh Thind2 and Narayanamoorthy M.3
*1,2&3SCOPE, VIT University Vellore, Tamil Nadu, India

ABSTRACT

The use of credit cards has drastically increased in today's world because of the increasing advancement in the e-commerce industry. Credit cards are used for offline as well as online purposes and with its uses come the fraudulent activities associated with it.In this paper, we will model the sequence of operations in credit card transaction processing using a hidden Markov model (HMM) and see how it can be used to detect frauds. HMM is initially trained with the behavior of the user or the cardholder. If credit card transaction is not accepted by model with sufficiently high probability, it is considered to be a fraud transaction. It is ensured that the genuine transactions are not rejected. Detailed experimental results are shown to prove effectiveness of our approach and the model is compared with other techniques already available in the real world.Fraudsters are so expert that they come up with new ways to commit fraud each day which demands constant innovation for its detection techniques as well. Many techniques that already exist in the market are 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

INTRODUCTION

Online Shopping popularity is growing every day.Transactions via credit card are of two types: physical card andvirtual card.In a physical-card purchase, the usergives his card physically to a merchant to make a payment.The attacker must steal the credit card in order to carry out such fraudulent transactions. If the user does not know that his/her card is lost then it can lead to a huge loss to the user as well as the company.

In the case of virtual card, only some of the important information like card number, expiration date, security code is required to make the payment. These kinds of purchases are done on the Internet. To commit fraud a person simply needs to know the card details. The spending patterns on every card are analyzed to detect frauds to check the inconsistency and fraud transaction.

An HMM is a technique that can be represented as simplest dynamic Bayesian network. The states are not directly visible but the output is visible to the user. Each state has a probability distribution over the possible values.

HMM is trained with the purchasing pattern of the user. If the transaction is not accepted by the HMM with sufficient probabilitythen that person is not considered to be genuine and is termed as a fraudulent transaction.

PROPOSED IDEA

The details of items purchased by the users are not usually known to the Fraud Detection System (FDS) branch at the bank. Only the transaction details and the amount of spending is known by the bank. Therefore, HMM is a very good solution to address this problem. HMM has all the details and is very useful in detecting fraud transaction and obtaining a solution. Another advantage of using HMM is reduction in the number of False Positives transactions which were considered to be fraud by FDS where actually they are genuine. FDS runs at a credit card issuing bank. Every transaction is verified by the FDS whether in reality it is fraudulent or not. FDS does not contain any information about the types of goods brought. Anomalies are found in the transactions which arebased on the spending profile of the user, shipping address, and billing address, etc.

Hidden Markov Model finds out fraudulent transaction by checking spending profiles of the user. The spending profiles can be dividedinto three major types:Lower profile, Middle profile and High profile.;

Every user has different spending profile and using that information one can figure out the inconsistency in the profile and find a fraudulent transaction. The records of the spending profile are stored both offline as well as online.Thus, analysis of the goods bought by a user will be usefulin fraud detection. The information of every user contains the details of the user, the money spent by them and the type of product on which the money was spent.Fraud detection is noted when there is deviation from the existing set of patterns. A Hidden Markov Model is a finite set of states where state linked with a probability distribution. There are a set of probabilities for transitioning among the states which are called as transition probabilities. Only the outcome is visible to the user and not the states which are hidden. Therefore, it is called Hidden Markov Model. HMM is veryadvantageous as there are very less number of false positives transactions recognized as fraud transaction while in reality they are actually genuine.

OTHER CREDIT CARD FRAUD DETECTION TECHNIQUES

In this section,we briefly discuss various existing fraud detection techniques that are used in credit card fraud detection. Also, the advantages and the disadvantages of the given techniques are discussed.

Neural Networks

Neural Networks is an AI technique which representslearning system model. It consists of networks of many units that are connected and process some value. It can be imagined as a directed graph with many nodes and interconnections between them. Weighted sum of the inputs are computed and an output is generated. This output then becomes an input to other nodes in the network. The process continues till one or more outputs are obtained. Neural networkscan be usedas a data mining technique to do classification, clustering, generalization, and predictions. For example, it detects Internet fraud on an e-commerce site, predicts which transaction may be a fraudulent transaction, etc.

Neural networks are created for supervised and unsupervised learning. The number of hidden layers along with the number of nodes are specified within a specific hidden layer. Also there exist many disadvantages for the neural networks like difficulty confirming the structure, excessive training and the efficiency of training,etc.

Decision Tree

Decision Tree is a predictive modeling technique that can be used in classification, clustering, and prediction tasks. Decision tree is a tree where the root and each of its internal nodes are labeled with a question about an independent variable. The interconnections from each node represent possible answer to the given question. The leaf node represents a predictionof a solution to the problem.

The decision tree is a table of tree shape with connections to other nodes. Each node is either a branch node followed by more nodes or leaf node. Decision Tree usually divides the problem into sub problems and then solves them. Rather than solving a huge complex problem, decision tree solves multiple easy sub problems and then combines the solution.

Genetic Algorithms

Genetic algorithms are search procedures based on the evolutionary computing methods. Genetic Algorithms can be used for various purposes like clustering, prediction, association rules, etc in the field of data mining. This techniqueis used to find the best model to represent the data.

HMM BACKGROUND

Hidden Markov Modelis a process which consists of two hierarchy levels. It is used to model much more complicated processes as compared to the traditionalexisting models. The model consists ofa finite set of states decided by a set of transition probabilities. The outcome is generated according to the calculated probabilitydistribution. The state is not visible to an external user. HMM-based applications are common in areas such as speech recognition, bioinformatics, and genomics. The behavior of a user is considered to form a model; and deviations are detected. If there exist deviations then an alarm is raised which could mean that the user is not genuine and could be an attacker.

WORKING OF THE SYSTEM

Training Phase

For the training phase in the HMM, the user's transaction details are converted into observation symbols and sequences are formed out of them. At the end of the training phase, we get a modelcorresponding to each user. As all of this work is done offline, it doesn’t affect the transaction processing performance

Detection Phase

After the model parameters are learned, the symbols are taken from the user's training data and an initial sequence of symbols is formed. Let O1, O2…… OR be one such sequence of length R. This sequence is formed from the user's transactions till the time t. This input sequence is input to the HMM and the probability of accepting the transaction is computed.

EXPECTED OUTCOME

After the credit card transaction details are taken and processed, the spending profiles of the user are formed on the basis of which the results are calculated. The observation symbols in the model are the details of a transaction of a user and the items present are considered to be the states of the model.From the spending profiles, the value of the observation symbols and initial estimation of the model parameters are decided. It explains how HMM detects whether a transaction is fraudulent or not.

RESEARCH OBJECTIVE

There do not exist such methods that are implemented by the banks to detect frauds during the transactions to prevent it from taking place.HMM is an appropriate method to stop such frauds. Many research works are done to solve this problem but there exist drawbacks too.Major problem with the existing methods is that they are complicated and require labeled data of the existing data to train the classifiers. Getting real-world fraud data is a big problem associated with credit card fraud detection. Also, these methods do not detect new kinds of frauds for which labeled data is not given. In contrast, a HMM does not require fraud data signatures and is able to detect frauds by taking into account a user's spending patterns. The transaction details of a user are taken as the observation symbols, whereas the types of item are considered to be states of the hidden model. In the end HMM detects whether the transaction is fraudulent or not.

TRANSACTION PROCESS DESCRIPTION USING HIDDEN MARKOV MODEL

Account Creation

This module consists of,

- The bank account is created for the customer and the username and password are provided to ensure security.
- The authentication information is entered by the customer or the user to access the account for various purposes like deposit, withdraw, balance enquiry etc.
- The user details are checked and corresponding response is given.
- The information is stored in the database.

Classification

- This module classifies things into groups according to their characteristics.
- The customer transaction dataset is taken which contains various information like transaction amount, place of purchase,etc.

Fraud Detection

- It detects Anomalies from the data of a real user.
- HMM is used for fraud detection.
- The classification results obtained from HMM help identify the fraud transaction.
- When a fraud transaction occurs, a form appears which asks for security information which the user entered when the account was created.
- The information entered is matched with the data stored in the database. If it matches then the user is verified and the transaction is done successfully otherwise the transaction is terminated.

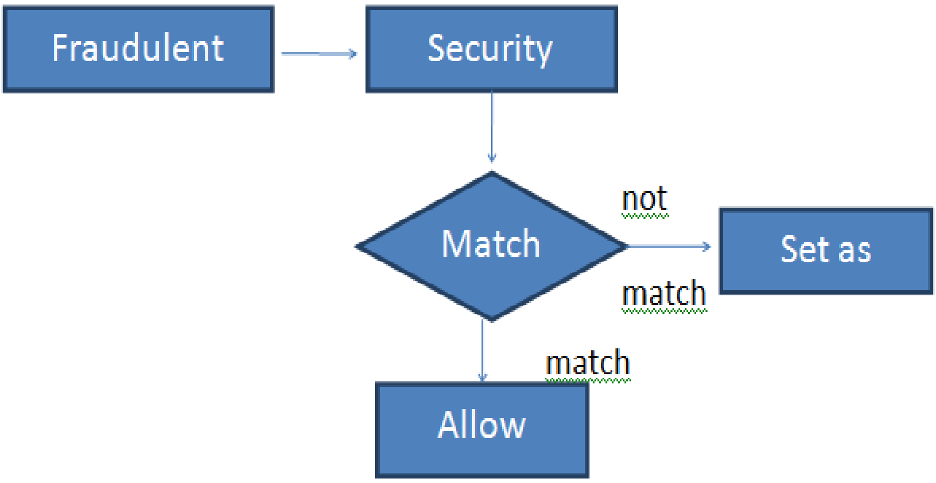

Figure 1 Credit Card Transaction Process with Hidden Markov Model

RESULTS AND DISCUSSION

Comparison

The following is the comparison of the existing methods to detect fraud in credit card transactions. The methods are namely: Decision Tree, Genetic Algorithm and Neural Network and Hidden Markov Model

Table 1 Comparing various Fraud Detection Techniques

Techniques	Advantages	Disadvantages
Rule Induction or Decision Tree	High accuracy, easy usage, easy explanation of results, easy interpretation of rules, easy implementation of applications.	Difficult handling continuous data, difficult handling missing data, over-fitting problem may occur, data size and attributes are usedto split and the order to choose them and number of splits for attribute impacts the performance.
Genetic Algorithm	Easy to handle noisy data, and integrates with other systems.	Requires extensive tool knowledge setting up and operate, and difficult to understand.
Neural Network	Effective in dealing with noisy data, in prediction of patterns, in solving complex problems, and processing new instances. Generates code that are used in real-time systems, it is portable and fast.	Poor explaining capability, less efficient in processing large data sets, difficulty in setting up and operation, sensitive to data format, different data representations produce different results. It works only with numeric data having values between 0 and 1; non-numerical data should be converted and normalized to produce results.
Hidden Markov Model	Gives better compression which helps in finding the important sequences. Flexible in handling inputs of variable length. Easy to read because the ratio of edges to states is low.Separate HMMs can merged together to recognize sequence of structures.	Viterbi algorithm and forward-backward algorithm are expensive as they take a lot of memory and the computing is also more in relativity. It has a lot of unstructured parameters.

CONCLUSION

After evaluation, we observed that there are various methods that can detect fraud in a credit card transaction, each having their own advantages and disadvantages. It was observed that Hidden Markov Modelis more efficient and provides more security as compared to others. Based on strong statistical foundation, it also provides decent compression, is readable and can combine well with separate Hidden Markov Models to recognize sequences of structures.

FUTURE ENHANCEMENT

The application of HMM in credit card fraud detection was proposed. HMM detects whether a transaction is fraudulent or not. In future effective algorithms can be developed which perform well in the field of classificationwith less complexity, faster and more efficient results.

REFERENCES

1. *S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model", 2003*
2. *S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov Models Capabilities in Anomaly Detection", 2005*
3. *C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data", 2004*
4. *C. Chiu and C. Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection", 2004*
5. *M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection", 2002*
6. *Phua, C., Lee, V., Smith, K., and Gayler, R., 2007. A Comprehensive Survey of Data Mining-Based Fraud Detection Research*
7. *Abhinav Srivastava, Amlan Kundu, Shamik Sural," Credit Card Fraud Detection Using Hidden Markov Model", 2008*
8. *Amlan Kundu, Suvasini Panigrahi, Shamik Sural, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection", 2009*
9. *Leila Seyedhossein, Mahmoud Reza Hashemi "Mining Information from Credit Card Time Series for Timelier Fraud Detection", 2010*
10. *M. Hamdi, Mine,"Improving a credit card fraud detection system using genetic algorithm" , 2010*