A Survey on Credit Card Fraud Detection System

Manoj S Ishi
RCPIT, Shirpur, India

Abstract— In today's world of technology peoples are making many online activities like online shopping, net banking for that purpose they take help of E-Commerce mostly called as Internet Banking. Due to increasing in online transaction fraud activities are also increasing. Hidden Markov model is designed to provide genuine transaction by providing one time password generated by server and it is sent to mobile of user. Banking sector is also victim of fraud activities in online transaction. It is necessary for bank to save money from such fraud transaction. The objective of study of this paper is that detecting credit card fraud using various techniques. Credit card fraud is found and then it gives solution for money fraud. Two types of credit card frauds are as fraudulent and legitimate transactions. And this is based on supervised and unsupervised learning. In this paper credit card fraud detection technique like decision tree, HMM, SVM, GASS etc. algorithms of credit card fraud detection are discussed.
Key words: E-Commerce, HMM, SVM, GASS

I. INTRODUCTION

In today's world of technology credit card users are increased to handle online transaction. Online as well as regular purchasing is done with the help of credit card. Virtual card is provided for online transaction and physical card is provided for offline transaction. While handling transaction 30% of fraud is increases in 2008 due to ambiguity in issuing and managing credit cards. Credit card fraud is classified as inception of mail for new card, copying and cloning of card through cloned website, Phishing where credit card password and number detected through mail and transaction where fraudsters make authentic look website and provide sell of things at lower price where unaware user attracts towards site and make online transaction. They submit their card information and fraudsters make valid transaction and users found them in Fraud Activity. In site cloning of fraud transaction entire site or payment page is cloned where customer make payment. Customer thinks that he deals with valid site and handles credit card details to those fraudsters and valid receipt is generated to that user and now with that details user can make invalid transaction. If credit card is lost or stolen then no need of technology for fraudsters for making fraud. In Skimming actual data of credit card is copy with electronic way. Valid credit card is generated with number and available with free download off the internet. Lots of false emails are sent in phishing activity. This email is asked for credit card number and use credit card fraud is detected. Customer personal information is stolen for credit card fraud. In this way numbers of activities are performed by fraudsters to create credit card fraud.

II. LITERATURE SURVEY

A. *Decision Tree [1]*

Decision tree has been developed to deal with continuous data. It is tree shaped data to connect data with available nodes. In this tree each node is branch node and one leaf node is there for classification. It divides the complex problem into simpler one and provides solution to this problem through data mining method to discover training various kind of classifying knowledge by designing decision tree. It provides advantage like high flexibility, Good haleness, and explainable for varied utilization. The main objective of this algorithm to design website to restrict and block transaction from attackers if they are using valid credit details of user.
Advantage: Decision system provides more transaction limit for transaction.
Disadvantage: Need to check each transaction one by one.

B. *Hidden Markov Model [2]*

Hidden Markov Model is used with double embedded stochastic process as compared to previous model of Markov. If with sufficient high probability incoming request for transactions is not arrived in trained HMM Model then it is treat as fraud transaction. Hidden Markov Model is based on the behavior of customer. User profile is classified as Lower, Middle and Higher profile. For finding fraudulent transaction inconsistency of user profile is considered. It tries to find fraud using profile of cardholder, shipping address and billing address.FDS is used to verify transaction. FDS determined whether that transaction is valid or not. If FDS declares that transaction as malicious then alarm is generated for bank to decline that transaction. Then concerned cardholder is notified and alerted for misused of the cards. Log is maintained for proof for the transaction made.
Advantage: HMM Provides good false alarm rate and high false positive.
Disadvantage: Overhead of Maintaining log file.

C. *Supervised Machine Learning [3]*

For the generation of synthetic data transaction probability is checked for input value in this algorithm. This algorithm is consisting of following Steps:

1) Read input data
2) Five groups are defined for transaction like month, date, day, amount of transaction and difference of amount between previous two transactions.
3) Vector of five fields are designed for each transaction
4) Transactions are classified into two groups True or False.
5) Select from Linear, Quadratic and RBF Kernel.
6) Train SVM and classifier is saved.
7) Read transaction and place saved classifier and generated vector and generate decision from SVM classifier.

SVM toolbox is used to train machine by passing machine parameter.
Advantage: This algorithm with RBF algorithm provides better result.
Disadvantage: Dynamic improvements are needed in this algorithm for training of classifier using different SVM Model.

D. Detection of Fraud in Outline Credit-Card Transactions [4]

The objective of this system is identify and detects fraud in online credit card transaction. In this algorithm system follows multilayered approach for security on transaction. This algorithm considers previous transaction to calculate threshold value. It classifies threshold values into low, medium and high. Each current transaction compared with threshold value and then classified whether it is fraud or not. Certain authentication mechanism needs to detect fraud in real time and block that transaction if user is not valid. In this system while doing registration author take required information to detect fraudulent user activity. HMM algorithm is used for fraud detection and after certain number of transactions threshold value is found and then current transaction is compared with this value to detect fraud or legitimate transaction. To check transaction valid or not OTP and security questions are used. This algorithm provides more protection for first 10 transactions due to high risk. Encryption is done at registration by using Secure Hash Algorithm for protecting from hackers. Counter is maintain for successful transaction and incremented for each iteration. If transaction is having number less than 10 then particular details need to check while doing transaction. If value is more than 10 then threshold value used to check valid transaction.
Advantage: Threshold value provides more accurate result.
Disadvantage: Less accuracy for first ten transactions.

E. GASS Algorithm [5][6]

This technique is fusion of genetic algorithm and scatter search. The basic idea of genetic algorithm is survival change for strong number of population is more as compared to weaker member and if generation changes the average fitness of population gets better. New generation is production of crossover of two parent members. Diversity of population is produced if random mutation is occurred. Less fit members are eliminated and fitter member are selected for next generation. The procedure is repeated for finding best generation. The scatter search operates on set of solution and combines this solution to produce new solution. The combine solution is obtained with the help of linear combination. Diversity is very important in SS Algorithm to find best solution from first best and with number of diverse solution to create new set of solution. GASS combines both techniques to provide credit card fraud detection. Using GA technique population is kept small and minimum diversity is maintained for each generation. Mutation Operator is maintained for GA and SS Algorithm. The step of algorithm is as follow:

1) Number of parent solution defined according to size of problem. It will generate maximum number of alerts and Minimum number of alerts.
2) Number of Children are obtained by recombining possible pair of parent solution
3) Reproduction or recombination of solution is obtained with weighted average of two parent solution to obtained child solution.
4) Mutant operator is applied for random range
5) Recombination and Mutant Probabilities are applied to select one of the children and then mutant operator is applied
6) Fitness Function is obtained from saving of fraud losses.
7) Selection of best three members are done on the basis of MAX, Min and PRD parameters
8) If no improvement then that generation is terminated.

Advantage: Provides more accurate solution by finding fittest population
Disadvantage: Processing speed is too slow.

F. A Hybrid Approach Using Dempster-Shaper Theory and Bayesian Learning [7]

The approaches are used to detect credit card fraud i.e. rule-based filtering, Dempster shaper theory and Bayesian Learning. Multiple evidence from rule based component are collected and combined for every incoming transaction and then Dempster rule is applied. The suspect score is updated on the basis of Bayesian Learning using history database of fraudsters as well as genuine cardholder. Transaction repository component of fraud detection system is denotes by THD. Record of fraudulent transaction and genuine transaction are maintained to extract characteristic of two groups from given data. GTH i.e. good transaction history of customer is maintained from past behavior and FTH i.e. Fraud Transaction History for fraud data is build. History transaction is also having attribute like card number, transaction amount and time since last purchasing done. Current and past both behaviors are considered for transaction to analyzed and accumulated. Transaction amount is also need for detecting the outlier. Bayesian Learning is used to provide dynamic approach to adapt the behavior of genuine customer and fraudsters. The FDS architecture is also dynamic for rule-based component.
Advantage: Dempster- Shafer theory provides good performance and Bayesian Learning improves accuracy.
Disadvantage: This technique is expensive and processing speed is low.

G. Stream Outlier Detection Based on Reverse K- Nearest Neighbors (SODRNN) [8]

Two procedures are used in this algorithm Stream Manager and Query Manager. Incoming Stream of object is received by stream manger and updates the memory window. It only update the list whenever new object arrives to maintain current window perfect. The list of that object in K- Nearest algorithm is called as knnlist. And for reverse K- Nearest algorithm rknnlist is maintained. If the object is expired then rknnlist is updated. Whenever Query is demanded by user for fraud detection with given list Query manager makes scan of Current Window and returns object form rknn list whose false rate is small as possible. Credit card errors are detected and validity is check in sequence of number to detect valid and invalid number.

Advantage: Number of Scan is reduced to one.
Disadvantage: Two List need to Maintain.

H. Credit Card Fraud Detection using Clustering Data Mining Techniques [9]

In this paper unique pattern for each customer is designed not only for representing normal behavior but also finding fraud transaction. LINGO clustering data mining algorithm is designed to replace Apriori algorithm for detecting legal or Fraud transaction and summarize the behavior of customer. More chance is created by this algorithm as fraudsters always try to behave like customer behavior; instead of study fraudsters' behavior the customer behavior is study to detect legal or fraud transaction. Simulated test is used to found meaningful summarized patter using LINGO Algorithm. It is fast fraud detection process to detect fraud or legal pattern and used to verify transaction in real time.
Advantage: False alarm rate is decreased causes for improvement in performance of algorithm.
Disadvantage: It is slow for huge record and need to work on speed of algorithm.

I. Fraud Detection in Credit Card using Data Mining Techniques [10]

Stochastic process of Markov Based Model is used to detect credit card fraud in this technique. It does not require fraud signature to detect fraud instead it observes the behavior of customer. FDS is unaware of details of purchase item for individual transaction while issuing credit card. These transactions are represented by Markov chain and stochastic process observes that behavior, amount of money spent. When online transaction is performed it is submitted to FDS for verification, if FDS declares it as fraud then alarm is generated and bank declines that transaction. Here transaction is divided into three categories high, medium and low on different range of transaction. The stochastic process is consisting of different step performed in credit card transaction. Profile of cardholder is studies and then model parameters are designed to detect whether given transaction is fraud or not. There is drastic reduction of number of false rates. It achieves accuracy of 92% for wide variation of input data.
Advantage: It is scalable for large volume of transaction.
Disadvantage: It is complex algorithm.

J. Artificial Neural Network [11]

Credit card fraud detection techniques consist of expert system, data mining, knowledge detection, but they are not good enough to detect fraud at a time when fraudulent transaction is in progress. Hidden Markov model, Neural Network able to detect fraudulent process while it is in progress. Customer uses a particular pattern of credit card during transaction. So using this data neural network is trained for customer. Neural network is trained on the basis of income, location, occupation, number of large purchase, etc. By using this neural network decide whether given transaction is genuine or fraudulent. It provides the output in real value between 0 to 1. If the output value is less than 0.7 then transaction is ok and if output grater that 0.7 then illegal processes are increases.

Advantage: The detection of fraud when fraudulent transaction in progress.
Disadvantage: Number of parameters need to be set for training without any clear rule and neural network topology optimal for the performance of algorithm.

III. Conclusion

Due to changes in technology the use of credit card is increased and its causes for increasing in fault. This study provides techniques to detect fraud in efficient manner. If this techniques are implemented in field of E-Commerce then credit card fraud will be minimize. Comparative studies of credit card fraud detection techniques are discussed. All this techniques are compared on the basis of three parameters such as accuracy, speed and cost. The techniques discussed in this paper are having own strength and weakness. HMM is having fast speed for detecting fraud but less in accuracy. Decision tree is also having good processing speed to detect credit card with good speed. Like HMM, Decision tree other techniques are also having its own strength. By considering the strength of all algorithms hybrid approach can be designed to develop some effective algorithm which performs well with high speed, more accuracy and low cost for credit card fraud detection.

References

[1] Sahin Y., Duman E., "Detecting Credit Card Fraud by Decision Treeand Support Vector Machine", Proceeding of the International multi conference of engineers and computer scientist.
[2] Abhinav Srivastava, AmlanKundu, ShamikSural, Arun K. Majumdar. "Credit Card Fraud Detection using Hidden Markov Model". IEEE Transactions on dependable and secure computing,Volume 5; (2008) (37-48).
[3] Sitaram Patel, SunitaGond," Supervised Machine (SVM) Learning for Credit Card Fraud Detection", International Journal of Engineering Trends and Technology (IJETT) – Volume 8 Number 3- Feb 2014.
[4] Deepak Pawar, SwapnilRabse, Sameer Paradkar, NainaKaushik, "Detection of fraud in online credit – card transactions", "International Journal of Technical Research and Applications",2016.
[5] S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011
[6] Hung, W. N. N., Song, X., Aboulhamid, E. M., & Driscoll, "BDD minimization by scatter search. IEEE Transactions on Computer- Aided Design on Integrated Circuits and Systems", 21(8), 974–979, 2002.
[7] SuvasiniPanigrahi, AmlanKundu, ShamikSural, A. K. Majumdar "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning", Information Fusion 10 (2009) 354–363.
[8] VenkataRatnamGanji, Siva Naga Prasad Mannem, "Credit card fraud detection using anti-k nearest neighbor algorithm", International Journal on Computer

Science and Engineering (IJCSE), Vol. 4 No. 06 June 2012.

[9] Mohamed Hegazy, Ahmed Madian, Mohamed Ragaie, "Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques", Egyptian Computer Science Journal, Volume 40 – Issue 03, September 2016

[10] Mr.P.Matheswaran,Mrs.E.SivaSankari ,Mr.R.Rajesh, "Fraud Detection in Credit Card Using DataMining Techniques", IJRSET-February-2015

[11] RaghavendraPatidar and Lokesh Sharma, et al. "Credit Card Fraud Detection Using Neural Network" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011

