OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Credit Card Fraud Detection System: A Survey

Dinesh L. Talekar [1], K. P. Adhiya [2]
[1,2] *Department of Computer Engineering, SSBT COET, Bambhori, Jalgaon (M.S.), India*

Abstract: *The credit card has become the most popular mode of payment for both online as well as regular purchase, in cases of fraud associated with it are also rising. Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. Fraudsters are so expert that they generate new ways for committing fraudulent transactions each day which demands constant innovation for its detection techniques. Most of the techniques based on Artificial Intelligence, Fuzzy logic, neural network, logistic regression, naïve Bayesian, Machine learning, Sequence Alignment, decision tree, Bayesian network, meta learning, Genetic Programming etc., these are evolved in detecting various credit card fraudulent transactions. This paper presents a survey of various techniques used in credit card fraud detection mechanisms.*

Keywords: *Credit Card Fraud, Hidden Markov Model (HMM), Fraud Detection, Password, Security question.*

I. INTRODUCTION

While performing online transaction using a credit card issued by bank, the transaction may be either Online Purchase or transfer .The online purchase can be done using the credit or debit card issued by the bank or the card based purchase can be categorized into two types Physical Card and Virtual Card. In both the cases if the card or card details are stolen the fraudster can easily carry out fraud transactions which will result in substantial loss to card holder or bank. In the case of Online Fund Transfer a user makes use of details such as Login Id, Password and transaction password. Again here if the details of the account is wrong then, as a result, it will give rise to fraud transaction. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit

Card or any similar Payment mechanism as a fraudulent source of funds in a transaction. The target may be to obtain goods Without paying money, or to obtain unauthorized funds from an account. The fraud begins with either the theft of the physical card or the compromise of data associated with the account, it include the card account number or other information that would routinely and necessarily be available to a merchant during a legal transaction. The compromise can occur by many common routes and can usually be conducted without tipping off the card holder or the merchant at least until the account is ultimately used for fraud. A store clerk

copying sales receipts for later use is a simple example. The speedy growth of credit card use on the Internet has made database security lapses particularly costly; in some cases, millions of accounts have been determined. Stolen cards can be reported emergently by cardholders, but a determined account can be cached by a thief for weeks or months before any miss use, making it difficult to identify the source of the determined.

Popularity of online shopping is growing day to day. Credit card is the easy way to do online shopping. According to an ACNielsen study conducted in 2005 one-tenth of the world's population is shopping online in same study it is also mentioned that credit cards are most popular mode of online payment. In US it is found that total number of credit cards from the four credit card network(Master Card, VISA, Discover, and American Express) is 609 million and 1.28 billion credit cards from above four primary credit card networks plus some other networks (Store, Oil Company and other). If consider the statistics of credit cards in India , it is found that total number of credit cards In India at the end of December-31-2012 is about 18 to 18.9 million [1]. In case of multinational banks, the usage or average balance, per borrower for credit card holder has rise up from Rs. 61,758 in 2011 to Rs. 82,455 in 2012. in the same period, private bank customers' usage rise from Rs 39,368 to Rs. 47,370[1]. As the number of credit card users increases world-wide, the opportunities for fraudster to steal credit card details and, subsequently, commit fraud are also grew up.

II. MOTIVATION

Now a day the customers prefer the most accepted payment mode via credit card for the convenient way of paying bills, online shopping is easiest way. At the same time the fraud transaction risks using credit card is a main problem which should be avoided. So There are many data mining techniques available to avoid

these risks effectively. In existing research they modeled the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and shown how it can be used for the detection of frauds. To avoid computational complexity and to provide better accuracy in fraud detection in proposed work.

III. LITERATURE SURVEY

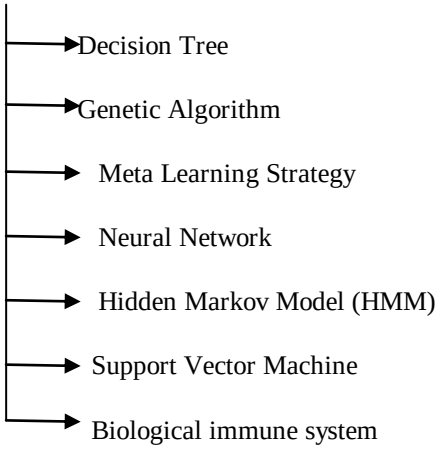
Abhinav Srivastava et al describe the "Credit card fraud detection method by using Hidden Markov Model (HMM)" [2]. In this method, they model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of fraud Transaction. An HMM is initially trained with the normal behavior of a cardholder.

S. Ghosh and Douglas L. Reilly et al describes the "Credit card fraud detection With Neural Network (NN)" [3]. In this method author use data from a credit card issuer, a neural network based credit card fraud detection system was trained on a large sample of labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month of time. The neural network was trained on examples of fraud due to stolen cards, lost cards, application fraud, mail-order fraud, counterfeit fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rule based fraud detection procedures.

IV. VARIOUS TECHNIQUES FOR CREDIT CARD FRAUD DETECTION SYSTEM

In Credit Card Fraud Detection there are many methods, here we present survey of some most powerful method.

Credit Card Fraud Detection Methods



A) Decision Tree

Decision Tree algorithm is a data mining induction Techniques that recursively partitions a data set of records using depth-first greedy approach (Hunts et al, 1966) or breadth-first approach (Shafer et al, 1996) until all the data items belongs to a special class. A decision tree structure is made of root, leaf and internal nodes. The tree Structure is used in classifying unknown data records. So at each internal node of the tree, a decision of best split is made using impureness measures (Quinlan, 1993). The tree leaves are made up of the class labels which the data items have been group [5]. In this method a Credit Card Fraud Detection using algorithm for Decision Tree Learning. Although focus on the Information Gain based Decision Tree Learning in this technique estimating the best split of Purity Measures of Gini, Entropy and Information Gain Ratio to test the best classifier attribute. In this Technique simply find out the Fraudulent Customer/Merchant through Tracing Fake Mail and IP Address. Customer /merchant are suspicious if the mail is fake they are traced all information about the owner/sender through IP Address. It can find out the Location of the customer and Trace all details. Decision Tree is Powerful Technique in Data Mining Decision Tree is vital part of Credit card Fraud Detection [5].

B) Genetic Algorithm

In this Technique fraud detected and fraud transactions are generated with the given sample data set. If this algorithm is applied into bank credit card fraud detection, the chance of fraud transactions can be predicted soon after credit card transactions is in process, and a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks [6].

The Experiment process has four steps:

STEP1: Input group of data credit card transactions, every transaction record with n attributes, and standardize the data, get the sample finally, which includes the confidential information about the card holder.

STEP2: Compute the critical values, Calculate the Credit Card usage frequency count, Credit Card overdraft, current bank balance, Credit Card usage location, average daily spending.

STEP3: Generate critical values found after limited number of generations. Critical Fraud Detected, Monitor able Fraud Detected, Ordinary Fraud Detected etc. using Genetic algorithm.

STEP4: Generate fraud transactions using this algorithm. This is to analyze the feasibility of credit card fraud detection based on technique, then applies detection mining based on critical values into credit card fraud detection and proposes this detection procedures and its process [7].

The initial population is selected randomly from the sample space which has many populations. The fitness value is calculated in each population and is sorted out. In selection process is selected through tournament method. The Crossover is calculated using single point probability. Mutation mutates the new offspring using uniform probability measure. In elitism selection the best solution are passed to the further generation. The new population is generated and undergoes the same process it maximum number of generation is reached.

C) Meta Learning Strategy

The meta-learning aims to filter the legitimate transactions from the fraudulent ones, and by quickly and accurately identifying the fraudulent transactions, fraud losses can be reduced. "Meta-learning" techniques introduced by Chan and Stolfo. There are two methods of combing algorithms that were introduced by Chan and Stolfo, the arbiter and the combiner strategies. Chan and Stolfo found that the combiner strategy performs more effectively than the arbiter strategy. Therefore, the combiner strategy is used. In the combiner strategy the attributes and correct classifications of credit card transaction instances are used to train multiple base classifiers. The predictions of the base classifiers are used as new attributes for the meta-level classifier. By combining the original attributes, the base classifier predictions, and the correct classification for each instance, a new "combined" dataset is created [8] which are used as the training data to generate the meta-level classifier. The predictions from the meta-level classifier are then used as the final predictions in the combiner strategy.

There are four main stages in the meta-learning process:

STAGE 1: Establishes the base classifiers using a training dataset that consists of 50% fraudulent transactions and 50% legitimate transactions [8]. This was done on a month by month basis for the first 8 months where all of the fraudulent transactions for the given month were matched with an equal number of randomly chosen legitimate transactions.

STAGE 2: The base classifiers are applied to a validation dataset to generate base predictions. The validation set consisted of all of the transactions. The predictions from the second stage are then combined with the validation dataset.

STAGE 3: Meta-algorithm is applied to this combined dataset to produce a meta-classifier.

STAGE 4: The forward predicting test stage, the meta- classifier is applied to the testing dataset to produce forward looking predictions [8].

D) Neural Network

Fraud detection using Neural network is totally based on the human brain working principal. Neural network technology has made a computer capable of think. As human brain learn through past experience and use its knowledge or experience in making the decision in daily life problem the same technique is applied with the credit card fraud detection technology. When a particular consumer uses its credit card, There is a fix pattern of credit card use, made by the way consumer uses its credit card. When credit card is being used by unauthorized user the neural network based fraud detection system check for the pattern used by the fraudster and matches with the pattern of the original card holder on which the neural network has been trained, if the pattern matches the neural network declare the authorize transaction. When a transaction arrives for authorization, it is characterized by a stream of authorization data fields that carry information identifying the cardholder (account number) and characteristics of the transaction (e.g., amount, merchant code). There are additional data fields that can be taken in a feed from the authorization system (e.g., time of day) [9]. The neural network is design to produce output in real value between 0 and 1 .If the neural network produce output that is below .6 or .7 then the transaction is ok and if the output is above .7 then the chance of being a transaction illegal increase [9]. In the design of neural network-based pattern recognition Systems, there is always a process of business History descriptors contain features characterizing the use of the card For transactions, the payments made to the account over Some immediately prior time interval. Other some descriptors can Include such factors as the date of issue (or most recent issue) of the credit card. This is important for the detection of NRI (non-receipt of issue) fraud [9].

E) Hidden Markov Model (HMM)

An HMM is a double embedded stochastic process with two hierarchy levels. It can be used to model complicated stochastic processes as compared to a traditional Markov model. An Hidden Markov Model has a finite set of states governed by a set of transition probabilities. In a particular state, observation or an outcome can be generated according to an associated probability distribution. So It is only the outcome and not the state that is visible to an external observer. HMM uses cardholder's spending behavior to detect fraud. In Implementation, three behavior of cardholder are taken into consideration.

1) Low spending behavior
2) Medium spending behavior
3) High spending behavior

Different cardholders has their different spending behavior (low, medium, high).Low spending behavior of any cardholder means cardholder spend low amount (L), medium spending behavior of any cardholder means cardholder spend medium amount (M), high spending behavior of any cardholder means cardholder spend high amount (H). These profiles are observation symbols [10].

Algorithm Steps:

Training Phase: Cluster creation

STEP 1: To Identify the profile of cardholder from their purchasing

STEP 2: The probability calculation depends on the amount of time that has elapsed since entry into the current state.

STEP 3: To construct the training sequence for training model

Detection Phase: Fraud detection

STEP 1: To Generate the observation symbol

STEP 2: To form new sequence by adding in existing sequence

STEP 3: To Calculate the probability difference and test the result with training phase

STEP 4: Finaly, If both are same it will be a normal customer else there will be fraud signal will be provided.

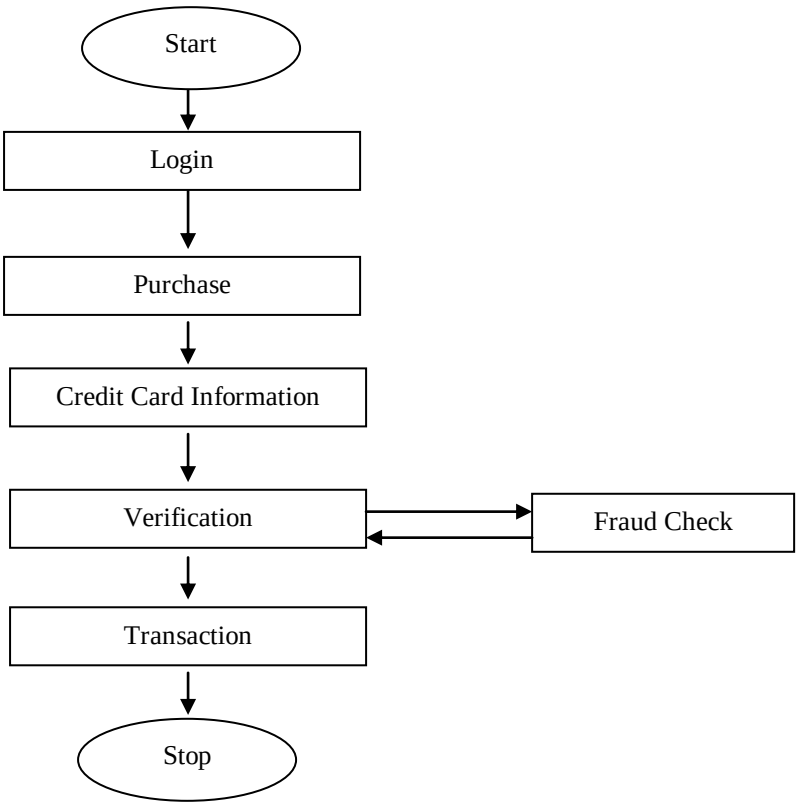


Fig: Flow chart for Credit Card Fraud Detection

In this Technique Clustering algorithm are used for creating three clusters and clusters represent observation symbols. Then calculate clustering probability of each cluster, which is percentage of number of transaction in each cluster to total number of transactions. Then calculate fraudulent Transaction.

But in this Proposed system no need to check the original user as we Maintain a log. We can find the most accurate detection using this technique. This reduces the tedious work of an employee in the bank. A one-time password (OTP) is a password that is valid for only one login session or transaction.

F) Support Vector Machine

Support Vector Machines (SVMs) have developed from Statistical Learning Theory. It have been widely applied to fields such as handwriting digit, character and text recognition, and more recently to satellite image classification. SVMs, like ANN and other nonparametric classifiers have a reputation for being robust. SVMs function by nonlinearly projecting the training data in the input space to a feature space of higher dimension by use of a kernel function. This results in a linearly separable dataset that can be separated by a linear classifier. This process enables the classification of datasets which are usually nonlinearly separable in the input space. The functions used to project the data from input space to feature space are called kernels (or kernel machines) examples of which include polynomial, Gaussian (more commonly referred to as radial basis functions) and quadratic functions. Each function has unique parameters which have to be checked prior to classification and it also usually determined through a cross validation process [11].

The choice of a Kernel depends on the problem at hand because it depends on what we are trying to model. A polynomial kernel, allows us to model feature up to the order of the polynomial. And Radial functions allows to pick out circles (or hyper spheres) in contrast with the Linear kernel it allows only to pick out lines (or hyper planes).

Linear Kernel: The Linear kernel is the simplest kernel function. It is given by the inner product <x,y> plus and constant c as optional. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts, that means. KPCA[11] with linear kernel is the same as standard PCA.

$K(X,Y) = XT\ Y + C$

Polynomial Kernel: The Polynomial kernel means it is a non-stationary kernel. Polynomial kernels are good for problems where as all the training data is normalized.

$K(X,Y) = (\alpha XT\ Y + C)d$

The adjustable parameters are the constant term c, the slope alpha and the polynomial degree d.

Here detail the proposed algorithm for classification of Fraud Transactions.

Step 1: Read the given data.

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

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

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

Step 5: Select one of three kernels (Linear, Quadratic, and RBF)[11].

Step 6: Train SVM.

Step 7: Save the classifier.

Step 8: Read the current Transaction.

Step 9: Restart the process from step1 to step3 for current transaction data only.

Step 10: Replaced the saved classifier & currently generated vector in classifier.

Step 11: Admit the generated decision from the classifier.

Since there is no real data is available because of privacy maintained by respective banks. so for testing of implementation of algorithm author generated the data of true & false Transaction using different mean & variance & then mixed them with different probability. And used the MATLAB for the execution of the algorithm because of its rich sets of mathematical functions and also supporting the inbuilt functions for SVM. Finally author said these technique give near about 90 to 97 % accuracy but future improvement is needed[11].

G) Biological immune system (BIS)

BIS is, a multilayered defense system comprising of cells and molecules which interact in various ways to detect and eliminate infectious agents (pathogens) from our body. BIS differentiates between self, (S), and (ii) nonself (NS) peptides and then assigns the right effectors to eliminate each pathogen. Similarly, detection system which sets apart fraudulent credit card transactions from genuine ones. The input for the system is financial transactions (i.e., source, destination and amount) in the form of a of e-commerce binary string.BIS in turn can be equated to a parallel adaptive information-system (IS) which works on the principle of simple and, localized rules. BIS interacts with pathogens in a localized fashion. Surfaces of BIS cells are covered with receptors, which chemically bind to (i) pathogens, and (ii) other immune system cells or molecules. Also BIS cells circulate around the body via the blood and lymph systems, to form a dynamic system of distributed detection and response. BIS has no centralized control, and hierarchical organization. Similarly, FDSCC detectors can be mobile agents that migrate across networks linking banks, financial institutions, etc.

BIS will comprise of two steps (i) detection and (ii) response. In step 1, detectors will be trained to discriminate between true and fraudulent transactions. In step 2, based on the training the FDSCC will classify a given transaction. It will help to memorize the rule for subsequent detection[12].

Detectors

The mobile detectors of System are analogous to the receptors on lymphocytes (B-cell receptor i.e. antibody or T-cell receptor). The receptors on lymphocytes bind to antigenic determinants (epitopes) on pathogens. Non-self detection results in the activation of the lymphocytes which trigger a series of reactions that can lead to elimination of the pathogens. A lymphocyte only activated when the number of its receptors binding to epitopes exceeds a threshold. Similarly, System detector matches the binary string inputs by using r-contiguous bit algorithm and confirms whether it is genuine or fraud transaction. The specificity of the detector is governed by the length of r-contiguous bits.

Response

BIS has a variety of response mechanism to eliminate different pathogens that attack the human body. One very important response (effector function) is mediated by soluble receptors called antibodies secreted by plasma cell (matured B lymphocytes). Antibody molecule has 2 parts variable region and constant region.Variable region binds to the pathogen and the constant region is responsible for the effector response. This is analogous to the System detector. Selection of effectors in System is determined by mathematical models[12].

BIS based anomaly detection and response system, which augments its performance through self learning. System will be an effective mechanism to detect and eliminate online credit card fraud transactions. This will help promote e-commerce as it will effectively minimize losses and other online credit card frauds.

V. Result

Comparison of Existing Methods

Authors	**Year**	**Techniques / Algorithms**	**Results**
Dr. R. Dhanapal	2012	Decision Tree/ Hunts Algorithm	Fraud detect by using Tracing Email and IP
Rinky D. Patel & Dheeraj Kumar Singh	2013	Genetic Algorithm	Optimizing the parametric fraud detection solution
Joseph Pun, Yuri Lawryshyn	2012	Meta Learning Strategy/ Meta Algorithm	Improvement in catch fraud than Neural Network
Raghavendra Patidar, Lokesh Sharma	2011	Neural Network/ Back Propagation Algorithm	Neural network-based pattern recognition.
Avinash Ingole, Dr. R. C. Thool	2013	HMM/ Clustering Algorithm	Fraud Detect using spending profile
Gajendra Singh, Ravindra Gupta	2012	Support Vector Machine	True Positive rate and false positive rate using MATLAB
Arunabha Mukhopadhyay,Sayali Mukherjee	2011	Artificial Immune System	By Matching Binary string Using detector and response

V. CONCLUSION

Credit card fraud has become more and more rampant in recent years. To improve merchants' risk management level in an automatic and efficient way and building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. There are many ways of detection of credit card fraud. If one of these or combination of algorithm is applied into bank credit card fraud detection system, Then the probability of fraud transactions can be predicted soon after credit card transactions by the banks. This paper gives contribution towards the effective ways of credit card fraudulent detection. In our paper we survey on seven existing Techniques for credit card fraud detection with comparing their results hence we conclude that out of these method HMM model is one of the best model because in HMM model fraud detect using Card holders spending behavior, but we need to improvement HMM in future.

REFERENCES

[1] Avinash Ingole, Dr. R. C. Thool, " Credit Card Fraud Detection Using Hidden Markov Model and Its Performance," International Journal of Advanced Research In Computer Science and Software Engineering (IJARCSSE), vol. 3, 6 June 2013.

[2] Srivastava, Abhinav, Kundu, Amlan, Sural, Shamik and Majumdar, Arun K., (2008) "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions on Dependable and Secure Computing, Vol. 5, No. 1, pp. 37-48.

[3] S. Ghosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems: Decision Support and Knowledge Based Systems, vol. 3, pp. 621-630, 1994.

[4] Pankaj Richhariya et al "A Survey on Financial Fraud Detection Methodologies" BITS, Bhopal," International Journal of Computer Applications (0975 – 8887) Volume 45 No.22, May 2012.

[5] Dr R. Dhanapal, Gayathiri. P, " Credit Card Fraud Detection Using Decision Tree For Tracing Email And Ip," International Journal of Computer Science Issues (IJCSI) Vol. 9, Issue 5, No 2, September 2012.

[6] 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.

[7] Rinky D. Patel, Dheeraj Kumar Singh " Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

[8] Joseph Pun, Yuri Lawryshyn " Improving Credit Card Fraud Detection using a Meta-Classification Strategy", International Journal of Computer Applications (0975 – 8887) Volume 56– No.10, October 2012.

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

[10] Avinash Ingole, Dr. R. C. Thool Credit Card Fraud Detection Using Hidden Markov Model and Its Performance", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) ISSN: 2277 128X, Volume 3, Issue 6, June 2013.

[11] Gajendra Singh, Ravindra Gupta, Ashish Rastogi, Mahiraj D. S. Chandel, A. Riyaz "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581), Volume No.1, Issue No.3, pg : 194-198 01 July 2012.

[12] Arunabha Mukhopadhyay, Sayali Mukherjee and Ambuj Mahanti, " Artificial Immune System for detecting online credit card frauds," Research Front, www.csi-india.org, CSI Communications , December 2011.