A Hybrid Approach for Credit Card Fraud Detection using Rough Set and Decision Tree Technique

Rajni Jain
Dept. Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, INDIA

Bhupesh Gour, PhD
Prof & Head of the CS Department
Dept. Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, INDIA

Surendra Dubey
Asst. Professor
Dept. of Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, INDIA

ABSTRACT
To make the business accessible to a large number of customers worldwide, many companies small and big have put up their presence on the internet. Online businesses gave birth to e-commerce platforms which in turn use digital modes of transaction such as credit-card, debit card etc. This kind of digital transaction attracted millions of users to transact on the internet. Along came the risk of online credit card frauds.
Hence the need to have secure payment transactions arose and many techniques based on Neural Network, Decision Tree, Artificial Intelligence, Artificial Immune System, Fuzzy based systems, Nearest neighbor algorithm, Support Vector Machines, Genetic Algorithm were developed to detect the fraudulent online credit card transactions.

This paper presents hybrid Approach for Credit Card Fraud detection using Rough Set and Decision Tree Technique which can be used in credit card fraud detection mechanisms.

General Terms
Your general terms must be any term which can be used for general classification of the submitted material such as Pattern Recognition, Security, Algorithms et. al.

Keywords
Credit card fraud detection, J48 classification, rough set, Support Vector Machines etc.

1. INTRODUCTION
In the credit card transaction credit card credentials for online transaction and physical card for offline transaction is used. When physical card is given for the transaction, the credit card details or the card itself can be stolen and if the owner of the card is not aware about the loss of card it may result in the fraudulent transactions by unauthorized users and the financial company is at the loss.

In the online transaction by a credit card, the printed information on the card and phone number suffice to do online transaction, many cardholders are not vigilant about the transaction notices and do not immediately react to any fraudulent transaction thus giving rise to big financial loss. This kind of fraud detection is possible if every card holders spending habits are learnt and analyzed and any irregularity in the established pattern marks the transaction as a 'suspicious' one, which can be further investigated to sop any more frauds.

With the increasing number of e-commerce websites credit card is widely used thus increasing the chances of online frauds. Credit Card fraud is termed as a theft and fraud committed by the use of physical credit card or just the credit card credentials for the financial gain.

Many fraud prevention models are being incorporated for dealing with this problem but the fraudster also evolve technologically and adapt to newer technologies for committing the fraudulent transactions and try to break the preventive models. Inspite of all the preventive measures employed by the financial institutions and the efforts done by the government the online credit card fraudulent cases continue to rise.

In most of the fraud cases the legal cardholder does not know that the card details have been stolen. In real life the transaction are not detected just by simply doing pattern matching. There is a data mining technique involving outlier detection commonly used for fraud detection. In this technique the data which are odds and do not match consistently with the whole lot are segregated as they seem to be coming from a different mechanism.

The outlier detection can be achieved through techniques like neural network, Self Organizing Map, Hidden Markov Model etc.

Online credit card fraudulent transaction is an unauthorized activity where a fraudster makes an electronic payment for his own financial gain. In this kind of fraud the deception done without the knowing of owner permission and it is illegal. With the increasing use of technology the fraudster hide their location and can therefore commit fraud over the internet.

2. BACKGROUND

2.1 Credit Card
Credit Card is plastic money and is widely used as a mode of payment. Credit card owners are increasing at a high rate and all users transact with a sense of security and confidence, hence credit card security is a prime concern for online transactions. It is important to safeguard the process so that the e-commerce platform providers as well as the credit card providers along with the user using the credit card are not at a loss. Secure services can be provided only with a reliable and safe model to protect the transactions. Secure Fraud detection models primarily based on the analysis of a user's spending pattern is a promising method giving quite good results.

2.2 Frauds
The domains where the fraud commonly occurs are financial systems, e-commerce web applications, computer systems, banks, and telecommunications, health and insurance systems.

Fraud is prevalent in many areas of the human society but internet being a global platform, it is most vulnerable medium.

2.2.1 Telecommunication Fraud:

The use of telecommunication services to commit other forms of fraud. Consumers, businesses and communication service provider are the victims.

2.2.2 Computer Intrusion:

Intrusion Is Defined As The Act Of Entering Without Warrant Or Invitation; That Means "Potential Possibility Of Unauthorized Attempt To Access Information, Manipulate Information Purposefully. Intruders May Be From Any Environment, An Outsider (Or Hacker) And An Insider Who Knows The Layout Of The System [7].

2.2.3 Bankruptcy Fraud:

Bankruptcy fraud is a polished way of crime that has four general types. Firstly, debtors hide assets to avoid having to forfeit them. Secondly, individuals deliberately file wrong or incomplete forms. Thirdly, people file multiple times using false information or real information in several states. The fourth type of bankruptcy fraud involves giving money to a court-appointed trustee. Commonly, the criminal will mix one of these forms of fraud with another crime, such as identity theft, fraud of mortgage, money laundering etc.

2.2.4 Theft Fraud / Counterfeit Fraud:

In this section, now focus on theft and counterfeit fraud, which are related to one other. Theft fraud means using a card which is not yours. As soon as the owner give some feedback and contact the bank, the bank will take all the steps necessary to catch the thief as quickly as possible. Likewise, counterfeit fraud occurs when the credit card is used remotely; and only the credit card details are required [20].

2.2.5 Credit Card Fraud:

The behaviour of credit card users are studied and some sort of sequential pattern is obtained and for all the subsequent transactions, they are tested against the behaviour already established and the probability of deviation from that behaviour gives rise to a 'suspicious transaction' which should be further investigated to mark the transaction as 'fraud'. This observation is repeated for a number of transactions and then classifying them into 'genuine' or 'fraudulent' ones.

2.3 Credit Card Frauds

Credit card fraud is defined as "Unauthorized account activity by a person for whom the account was not intended. Operationally, this is an event for which action can be taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future".

Credit Card frauds are of many types, few of them are described below:

Credit Card Fraud: Credit card fraud has been divided into two types: Offline fraud and On-line fraud.

2.3.1 Offline fraud

is committed by using a stolen physical card at call center or any other place.

(i) ***Physical credit-card stealers***: these fraudsters physically steal credit cards at the malls, shops or on the streets by pick pocketing or just by picking any dropped card

(ii) and use the information on it for making e-payment on internet for shopping.

2.3.2 On-line fraud

Is committed on the internet or in absence of card holder.

(i) ***Credit-card information buyers*****:** these fraudsters who either have less professional computer skills so they adapt a simple method of buying stolen credit card credentials from people who float an illegal credit card sales website, with the intension of buying goods and products online.

(ii) **Site cloning**: In Site cloning the fraudsters clones(copies) full website or only the pages on which the customer makes a purchase. Customers do not know that they are not dealing with the same company that they thought to purchase goods or services from because they are viewing the pages that are identical to the real site. The cloned (copied) site will get these card numbers and will send them a receipt of the transaction through the email to the customer exactly in the same way as the real company would have done. The consumer does not suspect anything, and the fraudsters get all the details they needed to commit fraud.

(iii) ***Black hat Intruders****:* "Black hat intruders" are those who attack computer security with intentions of damaging for personal gains. Their process involves targeting, research and information gathering and finally completing the attack. These people are highly capable in Programming and Networking skills and they can get in any network of computers. The main intention is to steal personal or private information of credit-card, bank-account, etc. for their own benefits.

(iv) **False merchant sites**: Many websites often show that they want to offer cheap services for the customers. That site requests the customer to fill his complete details such as name and address to access the page in the website where the customer buys required products. Almost all of these sites say that they are free, but they need a valid credit card number to confirm an individual s age. These websites collects many credit card details. The sites themselves never charge customers for the services they provide. The sites are typically owned by large network of criminals who use the details to generate money or they sell valid credit card information to other fraudsters.

(v) **Credit card generators**: Computer programs are there that produces credit card numbers and expiry dates which are valid. These programs generate lists of credit card numbers from one account number. The software uses the Luhn algorithm that card issuers use to produce valid card number combinations. This allows to unlawfully producing as many numbers as are desired, in any credit card formats.

2.4 Credit Card Fraud Detection

In fraud detection systems various techniques have been used like Neural Network, Genteic Algorithm, Firefly Algorithm, Support Vector Machines, Decision Tree, HMM (Hidden Markov Model) etc.

To prevent credit card fraud educating the customer can serve as a powerful tool and to protect business against the liability.

Advances in research techniques safe protective models are available which help card companies notice irregularities first.

3. TECHNIQUES OF CREDIT CARD FRAUD DETECTION

3.1 Neural network

Neural network Fraud detection methods are the most popular ones. An artificial neural network [15][16][12][14] consists of an interconnected artificial neurons. neural network is inspired by the functions of the brain especially the pattern recognition and associative memory [13] functions. The neural network recognizes patterns which are related, predicts values or events based on the associative memory of the patterns it has learnt. It is widely used in clustering and classification.neural networks' advantages are that these models are able to learn from by experience and thus, results improve as time passes. NNs can also extract rules and predict future activity on the current situation. banks can detect fraudulent use of a card, faster and more efficiently by employing neural networks.

3.2 Blast-Ssaha Hybridization

Blast-Ssaha Hybidization [2] is a combination of BLAST and SSAHA algorithms which is called as BLAH-FDS algorithm. This algorithm is in two-stage sequence alignment and is used for analyzing spending behavior of customers. The performance and accuracy of this algorithm is high. Due to its high processing speed, it is useful in telecommunication and banking fraud detection Its disadvantage is that it is not capable of detecting cloning of credit cards [9].

3.3 Genetic Algorithm

Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. The basic techniques of the GAs are designed to simulate processes in natural systems necessary for evolution.

A credit card fraud detection system that used genetic algorithm minimizes the number of false alerts of the protective model. GA is used in credit card fraud detection for minimizing the wrongly classified number of transactions [11]. Its implementation in programming language is also easy, thus, making it strong in credit card fraud detection. Though this method has high performance it is quite expensive.

3.4 Support Vector Machine

SVM is used in data mining mainly for supervised training where a dataset is analysed and classified linearly. If the dataset is not linearly classifiable the data is projected in higher dimensionality by making use of various kernel functions and are thus classified

3.5 Decision Tree Classification

A decision tree has structure of a tree which tries to separate the given records into mutually exclusive subgroups.

Decision Tree algorithm is a data mining technique that recursively partitions a dataset using breadth-first approach (Shafer et al, 1996) or depth-first greedy approach (Hunts et al, 1966) until all the items of data go in a particular class. A decision tree structure consists of a root, leaf and internal nodes. The tree Structure is used in classifying unknown data records. A decision of best split is made at each internal node of the tree, using impurity measures (Quinlan, 1993). The class labels which the data items have been grouped, the tree leaves are made up of that class.

4. LITERATURE REVIEW

As the electronic commerce technology advanced, credit cards use has accelerated and became quite a popular payment mode for online and offline purchases. Cards are comfortable to use but cards are not risk free. Many Fraud Detection Systems exist but they detect the fraud after the suspicious transaction is done. It is important that banks and commercial organizations develop an efficient Fraud Detection System.

[5] This paper primarily talks about the classification, many types of fraud in the credit card and methods to find the fraud in economical manner.

It presents the challenges confronted by the cardholder as well as the issuer of the card, analysis of persons who commit fraud, some latest information regarding credit card fraudster and gave some techniques of prevention to be followed by the cardholder against the fraudulent activity. The technology for preventing credit card frauds is developing and reduced computing cost helps in introducing complex systems, which can detect a fraud in a fraction of a second.

[6] In this paper fraud detection is done using Hidden Markov Model (HMM) when transaction is in happening. It is shown that HMM is used with minimum false positive number of transactions. Based on expenditure profile set of probability for transaction amount is assigned to each cardholder HMM categorizes customers profile as low, medium and high. Amount of new ongoing transaction is verified against the card holder profile; if it verifies a predefined value of threshold then transaction is marked as 'right' else it is marked as 'fraudulent'. As we have seen that initially HMM will build the profile of the card holders, so during training it is not secure for fraud detection for initial transactions. Hence, HOTP is used as second measure to detect the fraud alongwith HMM to minimize the fraud and increase the security. HOTP is one time password which will be used only once and it is sent to the registered mobile of client. When HMM checks that transaction amount is more than threshold value but the user entered valid HOTP then transaction is allowed to happen, else it is marked as fraud and barred the transaction. HOTP uses 8 digit unique security codes. They have use HMM with HOTP in proposed model to give more security and fraud probability reduction.

[17]In this paper, a credit card fraud detection model is presented which considers present and the past spending behavior. The detection model consists of card validation through Luhn's algorithm. Luhn's algorithm based on address mismatch and spending pattern has two starting probability assignments, a combination of spending-pattern database, heuristic, and Bayes' theorem. Conflict of existing Dempster-Shafer theory is eliminated by the advanced combination of heuristic.

Credit card fraud detection attracted lot of interest in research and a number of techniques, with special emphasis on decision tree classification, neural networks, data mining and distributed data mining. Credit card fraud detection using a neural network was proposed by Ghosh and Reilly [1]. They built a detection system, which is trained on a big sample of credit card account transactions which were labeled. These transactions consisted of fraud cases due to loss of cards, robbed cards,counterfeit fraud , application fraud Recently, Syeda et al. [2] used (PGNNs) parallel granular neural networks for improving the speed of data mining and knowledge discovery process in credit card fraud detection.

T.Abdul Razak , G.Najeeb Ahmed devised A Comparative Analysis on Credit Card Fraud Techniques Using Data Mining.

To improve the risk of merchants' management level in an effective way, creating a precise and easy credit card risk monitoring system is one of the important tasks. Aim of this paper is to discover the user model that identifies fraud cases to a large extent. It is very necessary to reduce unauthorized access to resources and data. To build a completely secure system, along with with the authentication process, behavior analysis is also needed, before a transaction is completed. Modern techniques based on decision tree, Artificial Neural Network, genetic algorithm, hidden markov model etc., has been introduced for credit card fraudulent transactions detection.

Fraud detection based on customer behavior variables is used in financial institutions. Prodromidis and Stolfo [10] make use of an agent-based approach with distributed learning for fraud detection in credit card transactions. It is based on artificial intelligence and mixes inductive learning algorithms and metal earning methods to get higher accuracy. Phua et al. [11] propose the use of Meta classifier similar in fraud detection problems. They consider naïve Bayesian, and Back Propagation neural networks as the base classifiers.

Xu Wei et al, developed an optimized SVM Model for detection of Fraudulent Online Credit Card Transactions.

Due to the risk faced by the credit card online transaction security issues this paper focus on credit card fraud detection and prevention model. Principles of support vector machine algorithm and model selection are discussed; and finally model of credit card fraud detection based on support vector machine is built. The model is applied to anti-fraud system for credit card online payment. Data is preprocessed; best SVM model is found, and compared withID3+BP hybrid model. The performance of the SVM model is found to be higher than the performance of the ID3 + BP hybrid model.

5. PROPOSED WORK

This section deals with the proposed work. This work is meant for improving the credit card fraud detection. This proposed work is made up of various important facts. These are mentioned as follows:

- Preprocessing using Rough Set
  - Rough sets theory was introduced as a mathematical tool for data analysis by Z. Pawlak (1982). Data preprocessing reduces the data complexity and offers better chances for subsequent analysis.
  - The best features can be found by determining the dependency between any conditional feature and the decision feature. Features with higher dependency values are taken in the final subset of best features.
- Classification using J48 classifier
  - In classification items are classified according to the item features with respect to the set of classes which are predefined.
  - J48 classifier is a simple C4.5 decision tree for classification. A binary tree is created in this. The approach of decision tree is most useful in classification. With this method, a tree is built as the model of classification process. Once the tree is made, it is applied to every tuple of the database and classification for that tuple is obtained.

Proposed algorithm is mentioned in figure 1 which is as follows:

Input:

D is a data set of credit card [3].

Output:

Credit Card Fraud Detection is done.

Procedure:

Step 1: Read the dataset 'D'.
Step 2: Apply Rough Set Theory and reduce dataset by following steps:

- Mark the attributes as 'dominant' in the dataset D by selecting dominant features of the dataset.
- Prepare reduced dataset 'Ds' having dominant attributes only

Step 3: Divide this dataset 'Ds' into two parts called training and testing dataset.

Step 4: Train J48 classification by using training dataset.

Step 5: Test J48 classification on testing dataset.

Fig 1: Proposed Algorithm

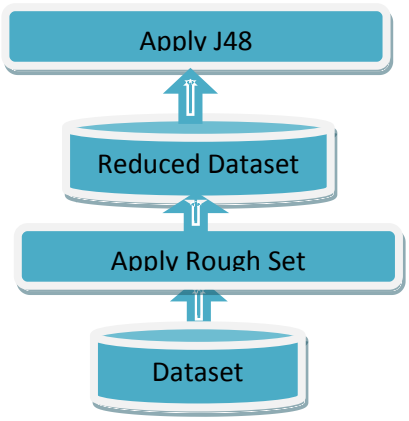

Fig 2: Proposed Architecture

where, fig 2 shows the proposed architecture of the proposed work.

6. EXPERIMENTAL SET UP

In the experimental setup, we have:

1. Dataset to be used
2. Tool to be used

Dataset:

Data set is taken from UCI website and various details are as follows:

- Title: German Credit data
- There are 20 features in the dataset.
- There is one class attribute, which could be
  - Good

- Bad
- It has 1000 tuples.
- Train dataset has 600 tuples
- Test dataset has the remaining 400 tuples

Tool: This work is implemented using two tools, as follows:

- MATLAB
- WEKA

- MATLAB ("MATrix LABoratory") is a tool for numerical computation and visualization. The basic data element is a matrix.
- WEKA is an innovator tool in the history of the data mining
- and machine learning research communities. The proposed algorithm uses WEKA as API in MATLAB.

7. RESULT ANALYSIS

This section is all about the comparison of the proposed work with the existing work based on classification values. Existing and proposed works are executed 10 times and the findings are as follows in table I.

Table 1: 10 execution of existing and proposed work

Base	Proposed
71.5	84.25
71	84.25
70.3	84.25
69.9	84.25
69.2	84.25
69.4	84.25
69.9	84.25
69.4	84.25
68.9	84.25
71	84.25
70.05	84.25

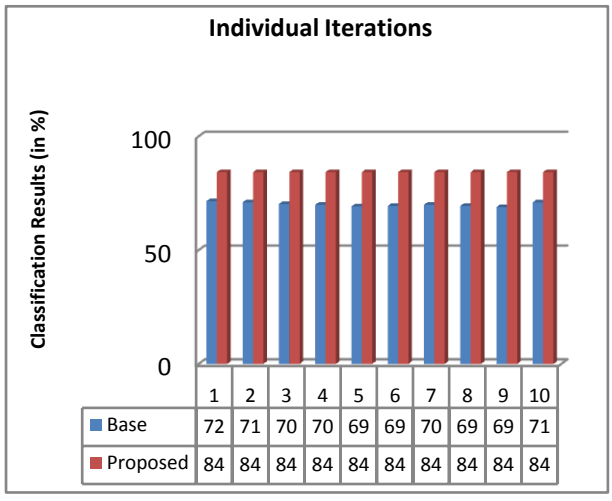

Fig 3: Comparison of existing and proposed work Classification Results

From Table I and Figure 3, it is clearly shown that performance of the proposed work over the existing work is better.

8. CONCLUSION

Credit card use is prevalent and used by everyone. Owing to the internet platform payment system is vulnerable to fraud attacks. The fraud detection systems in credit card transactions need to be very robust and precise, giving minimum false alerts and exactly classifying the fraud and non-fraud transactions. The system proposed in this article is very much effective and accurate. Last section showed that the performance of the proposed system is much better than existing work. Using this framework, financial institutions can compare the transaction information with the historical profile patterns to predict the probability of being fraudulent for a new transaction, and provide a scientific basis for the authorization mechanisms. Furthermore, resources of the institutions can be focused on more suspicious transactions to decrease the fraud levels.

9. REFERENCES

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

[2] M. Syeda, Y.Q. Zhang, and Y. Pan (2002) "Parallel granular networks for fast credit card fraud detection," Proc. IEEE int'l Conf. Fuzzy Systems, pp.572-557.

[3] uciwebsitehttps://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)

[4] Xu Wei, Liu Yuan, "An optimized SVM model for detection of fraudulent online credit card transaction," published in 2012 International Conference on Management of e-Commerce.

[5] N.Sivakumar and Dr.R.Balasubramanian," Fraud Detection in Credit Card Transactions: Classification, Risks and Prevention Techniques," International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015.

[6] Twinkle Patel and Ms. Ompriya Kale," A Secured Approach to Credit Card Fraud Detection Using Hidden Markov Model," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)Volume 3 Issue 5, May 2014.

[7] Anderson. J. P. "Computer Security Threat Monitoring and Surveillance." Technical Report, James P Anderson Co., Fort Washington, Pennsylvania, 1980.

[8] Masoumeh Zareapoor, Seeja.K.R, and M.Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, August 2012

[9] W. Fan, A.L. Prodromidis, and S.J. Stolfo (1999) "Distributed data mining in credit card fraud detection," Proc. IEEE Intelligent Systems, vol. 14, no. 6,pp. 67-74.

[10] Ekrem Duman, M. Hamdi Ozcelik "Detecting credit card fraud by genetic algorithm and scatter search". Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).

[11] R. Brause, T. Langsdorf, M. Hepp "Neural Data Mining for Credit Card Fraud Detection, "International Conference on Tools with Artificial Intelligence; (1999). (103-106).

[12] Raghavendra Patidar, Lokesh Sharma "Credit Card Fraud Detection Using Neural Network". International Journal of Soft Computing and Engineering (IJSCE), (2011). Volume-1, Issue; (32-38).

[13] Ray-I Chang, Liang-Bin Lai, Wen-De Su, Jen-Chieh Wang, Jen-Shiang Kouh "Intrusion Detection by Backpropagation Neural Networks with Sample-Query and Attribute-Query". Research India Publications; (2006). (6-10).

[14] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods". IEEEInternational Conference on Computer, Communication and Electrical Technology; (2011). (152-156).

[15] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick "Credit card fraud detection using Bayesian and neural networks". Proceedings of the First International NAISO Congress on Neuro Fuzzy Technologies; (1993). (261-270).

[16] Sahil Hak, Suraj Singh and Varun Purohit," Credit Card Fraud Detection Using Advanced Combination Heuristic and Bayes' Theorem," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 4, April 2015.