Hybrid Approach of Credit Card Fraud Detection Based on Outlier Detection

Raman Ghode, Shiv K. Sahu, Amit Mishra

Abstract— financial fraud is constantly rising with the advancement of new technologies and global mediums of the communication, which is resulting in loss of the millions of dollars over the world-wide every year. The organization and the financial firm lose their large volume because of the fraud and the fraudsters that regularly attempting to search for some new rules and strategy to do unauthorized activities. Hence, the fraud detection systems are now became one of the essentials for all the banks that are issuing credit card to decrease their deficit. An efficient use of the techniques of data mining and their algorithms may be applied to find out or to anticipate the fraud by the Knowledge Discovery from the different patterns obtained from collected data-set. In this paper, it is briefly explained regarding the several credit-card deceivers' strategies and with their methods of detection for the cyber credit-card transactions. The unauthorized transactions get never protected from being got cleared, the organization that should be agreed the economical cost of those types of transaction. This decreases the associated cost with the higher rates of interest and their complaints.

Keywords: Accuracy, Credit Card, Clustering, Classification, Fraud Detection.

I. INTRODUCTION

Transactions like online purchase or transfer can be done by the use of credit cards issued by bank in the online transactions. Online purchase may be done by the use of the card like debit or credit which is issued by bank or with the card dependent. Purchase may be divided into two kinds like Physical Card and the Virtual Card. Within both the situations in case if the card or the details of the card have stolen by the fraudster so then the fraudster may commit the fraud transactions very easily that will cause the substantial loss for the holder of card or to the bank. And in situation of the Online Fund Transfer the customer makes the use of information like Username, Security Password and the transaction password. Here also if theses information regarding account may get miss used then also it causes loss that will initiates the fraud transaction. Credit-card fraud is a widely used concept for the theft and the fraud that are committed by the use of a credit card or any other similar types of payment methods as a deceptive source of the funds in the transaction. The reason can be to attain materials without making any payment or to attain money illegally from any account. Credit-card fraud may also be an addition to an identity theft.

The credit card can be of physical card or virtual card [1][2][3][4]. In the physical cards, the holder of card physically shows his card to the merchant for paying the goods.

To commit the fraudulent transactions within these types of purchase, an attacker or fraudster has to steal first the credit-card. And in second type of the purchase, only few Significant details regarding the card like the number of card, its expiration-date, secure code of the card etc, is needed for making the payment. Many times, an authorized holder of the card doesn't know that someone else has stolen or seen the information of his card. In the real life, the fraudulent transactions are spread with the authorized transactions and along with the simple pattern matching. Previous researches of the credit-card fraud protection targeted on the classification and recognition of the methods and their models, consisting of the single pattern recognition approach like the decision trees and the neural networks, distributed data mining or the combination approach. Although, because of the complicacy and scarce of transaction data, these types of approaches are commonly suffered with the problem of the model selection, or model parameter settings, and improper selection in case of working with the huge scale of transaction data, that commonly directs to incur the analysis, over-fitting and issue of the local optimal [5].

Support-vector machine is completely an advance domain in the area of data-mining. This process in the first stage mapped the data collected from input space to the feature space and it then constructed a function called a linear discriminate function in the feature space. However there are various similarities in between the neural network and the support vector machine in their structure, and latter one is comparatively very simple. It is the proof that the performance of the data-mining algorithms and their techniques may be raise considerably by the precise and well managed removing of the deviations. This method of removing and diagnosing of the outliers is termed as the Data-Cleaning, and if it is used in the other type of applications as the pre-processing activity.

Data-mining has a various separate territories or regions that are extremely using the outlier detection mechanism that results in the massive and extremely distant literature for the methods of the outlier detection. Few of the mentioned methods may be used for the more common issues where as the others may be used for some particular type of problems. In this observation it is concerning on the financial types of frauds and it will specifically targeting on the detecting fraudulent credit-card transactions. The observation is required because of the inherent structure of the credit card (CC) transactions.

Revised Version Manuscript Received on May 20, 2016.

Raman Ghode, M.Tech Scholar, Technocrats Institute of Technology, Bhopal (M.P.). India.

Dr. Shiv K. Sahu, Assoc. Professor & Head, Technocrats Institute of Technology, Bhopal (M.P.). India.

Amit Mishra, M.Tech Coordinator, Technocrats Institute of Technology, Bhopal (M.P.). India.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.*

This is regarding the optimizing parametric type of fraud detection solutions. The volume of losses occurred because of the fraud and their awareness regarding the relation in between the loss and their available limits on CC have pressured to create a better performing solution for this type of problems. This provided solution is also got tested on the basis of data-set and the obtained results based on sample databases and the selections of best parameters for solution.

II. CREDIT CARD FRAUD DETECTION

This type of data-mining implementations uses the anomaly detection algorithm to find out the cyber credit card fraud in the online transaction by using credit-card that is implementing the Pattern recognition along with the Neural Networks. An anomaly detection algorithm is a new technique that is used in the applications of Data mining to find out the particular patterns or the relations among the data given for the process of Fraud detection. There is also a fixed type of pattern to find out how the credit-card owners utilize their credit-card on internet.

This type of fixed pattern may be taken from the legitimate frequent activities of the owner of credit-card on the basis of past one or two years transactions details on its credit-card and the websites of continuous merchant of the credit-card owner that are used for regularly making electronic payment for the goods and the services, and their geographical location from where the past all the authorized transactions have been made, also all the geographical location to where the goods have been continuously shipped to the credit-card owner, with the email-address and their phone number which is regularly used by the owner of the credit card for the notification.

2.1 Fraud

One of the major ethical types of issues is Fraud in the industry of credit card. The basic targets are, first, to recognize the various kinds of the credit-card fraud, and then secondly, to analyze their alternative methods which have been used in the fraud detection. The sub-target is to compare, show and analyze the currently published statistics in the credit-card fraud detection. This paper defines some general terms used in the credit-card fraud and their highlighted major statistics and the figures in this domain. Based on the kinds of fraud suffered by the banks or the credit card organizations, several measures may be accepted and applied.

On the basis of their use, credit-card fraud has been categorized into two separate types:

- Offline fraud
- On-line fraud

In the offline fraud, the fraud is committed by the use of a stolen physical card. Whereas in the on-line fraud, the fraud is committed through the phone, internet, web, shopping or in the absence of the card holder.

2.2 Fraud Detection

Detecting a credit-card fraud is a very complex task in case if a normal process is using, so that the development of credit card fraud detection method has become one of the most important method currently in the academic or in the business organizations both. Additionally, the role of the fraud has been modified suddenly while in the last some decades with the modernization of the technologies. The credit card fraud is one of the largest threats to the business and to the commercial organization found today.

2.3 Credit Card Fraud Detection

In this portion, it is explained few conceptual aspects of the credit-card, their issues and few real world issues.

2.3.1 Credit Card

It is a mode of selling or purchasing the goods or the services without having the cash in hand at the same time. A credit-card is a very easy way of providing the credit to the consumer automatically. In the current time, almost each credit card consists of an identifying number which helps in the shopping transactions to be get done quickly.

2.3.2 Fraud

It is a purposely done deception which is made for the personal benefits or to harm other user or individual is a fraudulent. A legal definition comprises by the legal jurisdiction for the fraud. Fraud is one of the civil law violations and it is also a type of crime. Defrauding the people or the entities related with the money is a general purpose of the fraud.

2.3.3 Credit Card Fraud

Credit card fraud detection is simply a private and is not so much exposed in public. Generally this concept is used in the fraud detection methods like the decision trees, rule-induction techniques, LR, Support Vector Machines (SVM), ANNs and Meta heuristics like, genetic algorithms, k-means clustering, or the nearest neighbor algorithms. Admiring the United States, with their huge number of Credit Card transactions that has the low rates of fraud and Ukraine at the top with the list of amazing 19% of fraud rate are closely followed by the Indonesia with 18.3% of the rate of fraud from the countries that are facing the Credit Card Fraud threat now, few other countries are including Malaysia (5.9%), Yugoslavia (17.8%) and Turkey with 9% rate. The authorized users are allowed for the credit-card transactions by the use of the attributes like digital signatures, credit card number, the address of card holders, card’s expiry date etc. The illegal use of those card or the information of card without informing the owner of the card itself and hence it is an act of the criminal fraud that refers to the Credit card fraud.

III. LITERATURE REVIEW

3.1 Credit Card Fraud Detection Using Hidden Markov Model and Its Performance

By the use of E-commerce the people can online make their financial transaction such as the online shopping etc. Very famous medium for the offline and online payment is by making the payment with the credit card, and use of the credit card has severely risen. So as the credit-card is becoming now very popular medium for the online financial types of transactions, simultaneously the fraud related with it are also increasing. In this paper the Hidden Markov Model (HMM) is used to present the series of the operation in the processing of credit card transaction. HMM is trained by using an algorithm called the Baum-Welch along with the normal behavior of the cardholder. If the incoming transaction of the credit card is not adopted or accepted by the trained HMM along with the enough high probability, then it is treated to be as a fraudulent [6].

In this application it consider the three symbol for observation that are spending the ranges of the cardholder which are like low, medium, and the high, in which the kinds of the item have been taken as to be the states of a HMM. An HMM is trained along with the Baum-Welch algorithm for every holder of the card. It has also been described that how an HMM may find out if the incoming transaction is a fraudulent or not valid. At last it is calculated the performance of the system by using the TP and the FP metrics as it is measured the accuracy of the system is at near to the 75%.

3.2 Data Mining Application for Cyber Credit-card Fraud Detection System

Data-mining has the famously obtained identification in combating the cyber credit-card fraud due to its efficient techniques of artificial intelligence (AI) and their algorithms which may be applied to find out or to predict the fraud through using the Knowledge Discovery from the unusual patterns which is originated from the collected data.

In this paper, the model system for the cyber credit card fraud detection method is explained and assigned. This system applies one of the best supervised anomaly detection algorithms for the data-mining to recognize the fraud in various real world transactions over the internet, and also by classifying those transactions as an authorized, doubtful fraud and an illegitimate type of transaction.

To make it understand that how the cyber credit-card fraud are being done, in this paper the various patterns of the cyber fraudsters which have already commit the cyber credit card fraud and their techniques that is used by those cyber fraudsters to make the fraud effective over the internet have been explained here.

3.3 Fraud Detection of Credit Card Payment System by Genetic Algorithm

The advanced technologies that are dependent on the Data-Mining and Genetic Programming etc have been used in finding the fraudulent transactions. The methods of detecting the best solution for these issues and then implicitly produced the results by using the genetic algorithm. The purpose is to introduce a new technique of producing the test data and to find out the fraudulent transaction by the help of this algorithm. This type of algorithm is an optimization method and an evolutionary search dependent on the rules of the genetic and the natural selection, and the heuristic is used to solve the highly complicated computational issues [7].

This paper is presented to find out the detection of the credit card fraud method and analyze the result dependent on the rules of this given algorithm. The advantages of the detecting fraud must be clear for the both, to the credit card companies and also to their clients. The fraudulent-transactions are not protected from being getting cleared and the company should be accepted the financial cost of those transaction [11]. This decreases the cost which is associated with the higher rates of interest, and their charges. If this algorithm is implemented into the bank credit-card fraud detection system, then the chances of the fraud transactions may be anticipated very soon after the credit card transactions by banks. And a sequence of the anti-fraud procedures may be accepted to protect the banks from the huge losses before and decreases the risks.

3.4 Fraud Detection Using Reputation Features, SVMs, and Random Forests

Additionally it uses two one-class Support Vector Machines (SVMs) mechanism to analyze the similarity between the originated reputation feature vectors to the currently measured fraudulent applications and currently measured authorized applications. The combination of both the reputation and their similarity properties are then utilized to train the Random Forest classifier for a new insurance application. The generally present auto insurance fraud data-set is now used to calculate this approach. Cost-savings is difference in the cost for anticipating the all recent insurance applications as a non-fraudulent and anticipating the fraud dependent on the trained data-mining method that are used as the primary evaluation metric [9].

This method presented at 13.6% that is raise in cost-savings than to existing published state of art those results for auto insurance fraud data-set. Although an auto insurance fraud data-set was here used for this representation of the reputation features can be easily implemented to the other fraud detection areas, consisting of the health care insurance-fraud, securities fraud, credit card fraud and an accounting fraud. This method may also be very beneficial for other implementations like the computer network intrusion detection [24] or the credit risk classification [10].

IV. PROPOSED WORK

Here, this section talks about a new Credit Card Fraud Detection technique called ‘Hybrid Approach of Credit Card Fraud Detection Based on Outlier Detection’. The architecture of the proposed work is shown in figure 1.

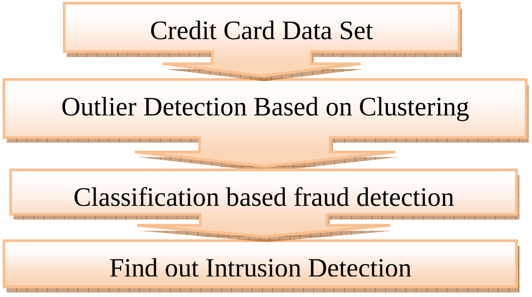
Credit Card Data Set
Outlier Detection Based on Clustering
Classification based fraud detection
Find out Intrusion Detection

Figure 1: Architecture of proposed work.

The algorithm is described in figure 2. As shown bellow. According to the algorithm, the whole process is divided into various sections. First section is dedicated to read the transaction dataset. In second section, natural bounding is found out. Next section is all about finding Intrusion detection process.

Step 1: Start

Step 2: Read Transaction Dataset

Step 3: Cluster dataset based on natural grouping

Step 4: Identify element not fit any group of good or fraud

Step 5: Remove element that not fit to any group of Good or Fraud

Step 6: Recombine rest of data and prepare new dataset.

Step 7: Train classifier with resultant data

Step 8: Apply Intrusion Detection by Applying Classifier

Step 9: Analyze the Intrusion Detection Result

Figure 2: Algorithm of proposed work

V. RESULT

This Section is divided into three parts. First part talks about system on which all experiments are performed. Second section deals with dataset and third part shows the results and its' analysis.

Part One:

Configuration of system is as follows:

Processor	P IV 2 .0 GHz
RAM	2 GB
OS	Win7 with 32 bits

Part Two:

Dataset is taken from UCI. German credit card transaction dataset is taken for consideration. This dataset has 24 various attributes. Various attributes are given bellow:

1. Title: German Credit data
2. Number of Instances: 1000
3. Number of Attributes 24 (24 numerical)

Part three:

It shows the results of proposed work and existing work.

Table I: Comparison between Existing and Proposed work's Accuracy

	Existing	Proposed
Accuracy	0.5	0.62

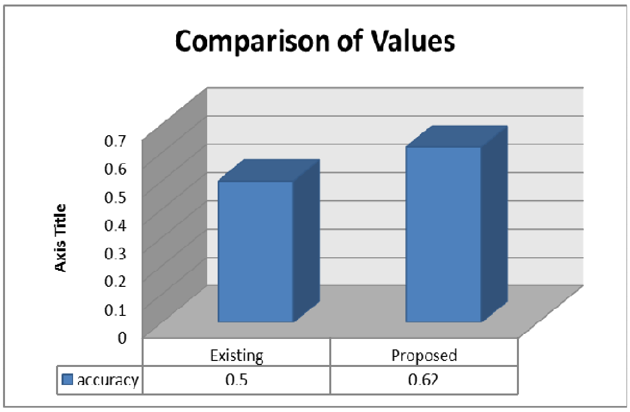

Figure 1: Comparison between Existing and Proposed work's Accuracy

Table I along with the Figure 1 shows the effectively of the proposed work on the parameter Accuracy.

Table II: Comparison between Existing and Proposed work's F-score

	Existing	Proposed
F-score	0.28	0.33

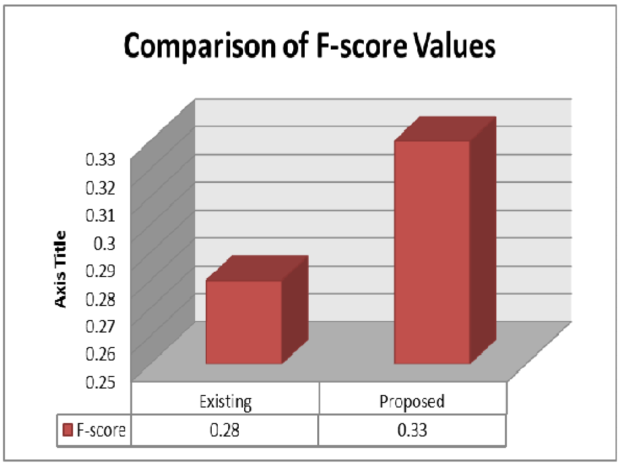

Figure 2: Comparison between Existing and Proposed work's F-score

Table I along with the Figure 1 shows the effectively of the proposed work on the parameter F-score.

VI. CONCLUSION

The utilization of deception for the financial benefit is a generally detected in the form of fraud. The costs for influenced organizations are very high and these incurred costs are then passed to their users. Detection of the fraudulent actions is hence very complex to monitor their costs. Here it is proposed to refer the insurance fraud detection through the use of the reputation and the similarity features which classifies the insurance claims and ensemble learning to compromise for the modifications in the given data distribution. So many ways are there for detection of the credit card fraud. If any one of these ways or the combination of the algorithm is implemented into the bank credit-card fraud detection system and the chances of the fraud transactions that may be anticipated very soon after the credit card transactions done by the banks. A sequence of the anti-fraud methods may be preferred to protect the banks from the huge losses before they get committed and decrease their risks. This paper provides the cooperation in the field of the efficient pattern of the credit-card fraudulent detection. Result shows the performance of the proposed work over existing work is much better.

REFERENCES

1. B.Sanjaya Gandhi , R.Lalu Naik, S.Gopi Krishna, K.lakshminadh, “Markova Scheme for Credit Card Fraud Detection ,” International Conference on Advanced Computing, Communication and Networks,(2011) 144–147.
2. Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar. “Credit Card Fraud Detection using Hidden Markov Model”. IEEE Transactions on dependable and secure computing,Volume 5; (2008) (37-48).
3. Anshul Singh, Devesh Narayan “A Survey on Hidden Markov Model for Credit Card Fraud Detection”, International Journal of Engineering and Advanced Technology (IJEAT), (2012). Volume-1, Issue-3; (49-52).
4. V. Bhusari S. Patil “Study of Hidden Markov Model in Credit Card Fraudulent Detection”. International Journal of Computer Applications, (2011). Volume 20– No.5; (0975 – 8887).
5. V. Bhusari1, S. Patil “Application of Hidden Markov Model in Credit Card Fraud Detection” *International Journal of Distributed and Parallel Systems (IJDPS)* Vol.2, No.6, November 2011.
6. S. Benson Edwin Raj, A. Annie Portia “Analysis on Credit Card Fraud Detection Methods” 2011.
7. Leila Seyedhossein, Mahmoud Reza Hashemi Mining Information from Credit Card Time Series for Timelier Fraud Detection International Symposium on Telecommunications 2010.
8. Genetic algorithms for credit card fraud detection by Daniel Garner, IEEE Transactions May 2011.
9. M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, Improving a credit card fraud detection system using genetic algorithm, International conference on Networking and information technology 2010.
10. Adnan M. Al-Khatib, Electronic payment fraud detection techniques, World of Computer Science and Information Technology Journal (2012), vol. 2, no. 4. pp. 137-141.
11. B.T. Adler, L. de Alfaro, S.M. Mola-Velasco, P. Rosso, and A.G. West. “Wikipedia Vandalism Detection: Combining Natural Language, Metadata, and Reputation Features”, Proceedings of the 12th Intl Conf on Intelligent Text Processing and Computational Linguistics, 277-288, 2011.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.