A Survey Paper on Fraud Analysis and Report Visualization in Card System

Shipra Rathore
Computer Science & Engineering
LNCT-E
Bhopal, India

Deepak Kumar Niware
Computer Science &Engineering
LNCT-E
Bhopal, India

ABSTRACT
Due to rapid growth of internet, online shopping for products is an essential part of everyone's daily life most of the time credit card is used to pay online for products. It is an easy way to shopping, people can get your desired product on your computer screen or on smart phone. For online purchase use of credit card increases dramatically but still there is some loop holes in system of online shopping which causes online frauds or credit card frauds. There are many techniques like ANIDS Distributed Data mining WSDL, Parallel granular neural networks, Hidden Markov Model, Clustering and Bayes classification, Cost model are used. These techniques use data mining to detect frauds by the use of datasets like KDD99, KDD cup, NSL KDD. In this paper a study on different techniques and evaluate each technique on the best parameters.

Keywords
E-commerce, Fraudulent transactions, IDS.

1. INTRODUCTION
Data mining is a assembling of data which have keyed out the pattern and made the relationship between data and its multiple attributes. In data mining basically extraction of implicit, previously unknown and potentially useful information from databases is performed. Data mining is sometimes called knowledge discovery. Knowledge discovery is a procedure that extracts implicit, potentially useful or previously unknown information from the data. The knowledge discovery process is identified as follows:

Work to detect frauds need information, but there is a huge quantity of data flooding around companies, organizations even individuals. Because of the quantity of data is so enormous that human cannot process it fast enough to engender the information out of it at the proper time that causes serious threat of frauds, the machine learning technology has been made to work out this problem potentially.

The term data mining is applied for methods and algorithms that allow analyzing data in order to find patterns and figures describing the characteristic properties of the data. Data mining techniques are attractive as they can be used to whateversorts of information in order to determine more about hidden structures and correlations. Nevertheless, this absoluteness also has a shortcoming because the generated knowledge does not have to be meaningful or useful. It is necessary to measure and understand the data mining results with regard to a specific destination or aim of the data analysis.

IDS
Intrusion detection systems (IDS) process large amounts of observing data. As an instance, a server-based IDS examines log files on a

Computer (or horde) in order to detect suspicious actions.Net-based IDS, on the other hand, searches network observing data for risky packets or packet flows. In the late 1990s, progress in data mining research and the necessity to find better methods for network and

host based intrusion detection resulted in research activities attempting to deploy data mining. In our existing paper describe on collective anomaly detection and clustering anomaly which are techniques for anomaly and attack detection generated due to variety of abnormal activities such as credit card fraud detection, mobile phone fraud, banking fraud, cyber-attack etc. an important aspect as the nature of anomaly, existing system introduced the anomaly detection in credit card or any other card system today's life it can be connected through health card or any personal important relevant card.

Need of IDS
Firewall and IDS both are related to network security. An IDS differ from a firewall in that a firewall is primarily utilized as a traffic-filtering device, whereas an IDS is primarily utilized as a traffic–auditing device. An IDS looks for certain traffic patterns Proclivity that has been determined as anomalous or possibly malicious. A firewall is designed to limit the access between networks in order to prevent intrusion. Generally, a firewall is not designed to provide detailed notification of attacks or anomalous network activity. An IDS evaluates a suspected event as it takes place or after it has taken place, creates a detailed audit in one or more secured locations, signals an alarm or notifications. An IDS can also be setup to watch for attacks or events that originate from within an organization's network. This protects the organization from the possible legal punitive actions initiated as a result of attacks from inside the organization.

An IDS cannot replace firewall and vice-versa because firewall does not have the intrusion detection capabilities of an IDS, an IDS does not generally have the firewalling capabilities of a firewall. The two technologies are very complementary and are generally deployed in union to achieve maximum level of security.

Here work is addressing the issues of anomaly detection:

Number of Attributes: since an object may possess many attributes, it may have anomalous values for some properties; an object may be anomalous even if none of its attribute values are individually anomalous.

Global vs. Local Perspective: an object may seem unusual with respect to whole objects, but not with respect to its local neighbours.

Degree of Anomaly: some objects are more extreme anomalies than others; it's desirable to have some assessment of the degree to which an object is anomalous (outlier score).

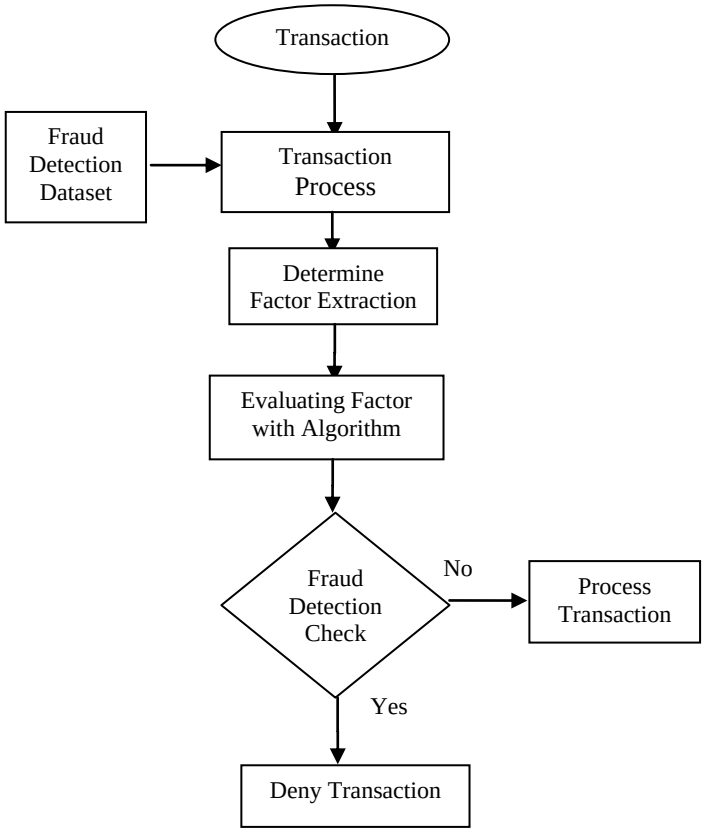

Fig. Fraud Detection Process

One of Time Vs. many at Once: is a better remove anomalous object one at a time or identify a collection of objects together. Two distinct problems: masking, where the presence of an anomaly masks the presence of another; swamping, where normal objects are sorted out as outliers.

Evaluation: find a salutary quantity of valuation for the process of anomaly detection when class labels are available and when class labels are not available (precision, recall, FP-rate, accuracy).

Efficiently: calculate the computational cost of the process of anomaly detection scheme.

The anomaly Detection main approaches statistical approach, Proximity-Based, Density-Based, and Clustering-Based.

1. Statistical Approaches are model-based which are based on establishing a probability distribution model consider how likely objects are under that model an outlier is an object that induces a low probability with regard to a probability distribution model of the information.

2. Proximity-Based Approaches Simplest way to measure proximity: distance to the k-nearest neighbours outlier score is given by the distance to its k-Nearest Neighbours. An object is an anomaly if it is distant from most points (k) this scheme is simple, it's easier to determinate a "good" measure than to determinate the statistical distribution.

3. Density-Based outlier detection is closely related to Proximity-Based outlier detection since. Density is usually defined in terms of Proximity.

4. Clustering: used to find groups of (strongly) related objects Anomaly Detection: used to find objects that are not (strongly) related to other objects.

This paper further organized as follows: - II Literature Review, III comparative analysis, IV Conclusion

2. LITERATURE REVIEW

Monowar Husain Bhuyan, D K Bhattacharyya and J K Kalita introduced a paper with title "Survey on Incremental Approaches for Network Anomaly Detection"[9] in which author have described various related algorithm for the fraud detection, various intrusion detection techniques were involved ,they have mentioned that there are many supervised and unsupervised learning techniques. An anomaly based network intrusion detection system is presented in this paper in which is required to detect frauds.

ANIDS (Anamoly based Network Instruction Detection System)is a technique where the complete mechanism used for solving the intrusion detection problems were solved by introducing above architecture in the field of anomaly detection system. Its performance is based on the arrangement of the features and network in which it uses. Also position with respect to the network is affecting the performance of the system.

Salvatore J. Stool, David W. Fan, Wenke Lee and Andreas L. Prodromidis[2] presented a meta-learning technique in which combine strategy is used to detect fraud for this purpose features and correct events of the fraud detection is used to detect the credit card frauds and train base classifiers. These base classifiers are to generate new features for next meta-learning process and generate new meta-level classifiers and combine these features with original classifiers which enhance the whole process of fraud detection.

Philip K. Chan, Florida Institute of Technology proposed a work in titled "Distributed Data Mining in Credit Card Fraud Detection" [7] In which they have mentioned and proposed methods of cost model in which issues like efficiency, scalability, and other technical problems related to fraud detection are mentioned in this model a costumer independent technique is proposed which based on the cost of the fraud which occurs in recent it calculates average cost and sum of loss caused by the frauds.

It uses to banks chase bank and union banks fraud cases to measure the performance of the model.

V.Dheepa1, Dr. R.Dhanapal Provided a paper titled "Analysis of Credit Card Fraud Detection Methods"[13] in this paper three methods which used to detect fraud are presented. Cluster based method, Gaussian Mixture method and Bayesian Networks. all three methods takes part in the fraud detection process, in this clustering model is use to cluster the data related to previous frauds and use as a measure to detect frauds, in Gaussian mixture model which detects the user's previous behavior to predict about the current behavior to detect the fraud and Bayesian networks are used to detectstatics of a user and fraud events.

Abhinav Shrivastava, Amlan Kundu, Shamik Sural, proposed a paper titled "Credit Card Fraud Detection Using Hidden Markov Model" [6]presented HMM based fraud detection technique which is based on the spending behaviors of the users and form state according their spending habits and calculate the probability of outcome for that state which is used to detect frauds in that state, just like low spending, medium

spending, and high spending. And calculate frauds on that basis.

Ghosh and Reilly proposed credit card fraud detection with a neural network [8]. Authors present a detection system, which is trained on a large data of credit card account transactions. These transactions contain example fraud cases due to the various fraud scenarios like lost cards etc. Recently, a parallel granular neural network (PGNNs) is proposed to improve the speed of fetching data for in credit card fraud detection.

3. COMPARATIVE STUDY

A comparative study for fraud detection techniques is gives in table3.1 which shows accuracy, merits and demerits of existing techniques.

Table 3.1 Comparison Statistics of Different Algorithms

Algorithm/ Methodology	**Dataset**	**Recall %**	**Precision %**	**Comparative analysis**			**Advantage**	**Disadvantage**
				Methods	Recall	Precision		
ANIDS	12000	71	86	WSDL PGNN HMM CBC COST MODEL	+9 -10 +2 +8 +5	+4 -2 +10 +12 -5	Reduced memory utilization, faster and higher detection rate, and improved real time performance.	Limited to specific signature.
Distributed data mining, WSDL.	3000	62	82	ANIDS PGNN HMM CBC COST MODEL	-9 -19 -7 -1 -4	-4 -6 +6 +8 -9	Ease of access, no need to write extra code, user should know the accessing technique.	Proper IP and proper communication links related to the WSDL are required to connect with the web services to access the data associated with the logs.
Parallel granular neural networks (PGNNs).	23000	81	88	ANIDS WSDL HMM CBC COST MODEL	+10 +19 +11 +18 +15	+2 +6 +12 +14 -3	Improving the speed of data mining and knowledge discovery process in credit card fraud detection.	Depending on the existing logs and updated logs are always required.
Hidden Markov Model	12000	69	76	ANIDS WSDL PGNN CBC COST MODEL	-2 +7 -11 +6 +3	-10 +6 -12 +2 -15	It maintains and works with the logs and once the person is detected as fraud it won't be consider for the further transaction.	Person information always kept as fraud if once kept in log as fraud, it may keep sometime in false category to true transaction.
Clustering & Bayes	21000	63	74	ANIDS	-8	-12	Cluster analysis is a technique for breaking	Less recall & precision value.

Classification.				WSDL PGNN HMM COST MODEL	+1 -18 -6 +3	-8 -14 -2 -18	data down into related components in such a way that patterns and order becomes visible, make efficient to process data parallel.	
Cost model	21000	66	91	ANIDS WSDL PGNN HMM CBC	-5 +4 -15 -3 +3	+5 +9 +3 +15 +17	Parallel and data learning classification technique is applied to process it efficiently.	Training distribution based on a defined cost is not automated and can be automated process.

In this table, a comparative analysis among different methods is presented, in which PGNNs performs best in recall and Cost model performs best in precision.

4. CONCLUSION

In this paper, survey about various anomaly detection techniques in credit card fraud detection have presented, like ANIDS, distributed data mining WSDL, PGNNs, HMM, clustering and bayes Classification and cost model are study and a comparative analysis for these methods over

The parameters, precision and recall are shown. Accordingly the algorithm PGNN provide efficient recall among other algorithm and cost model put an efficient precision as compared to different other approaches, thus further work is going define an approach which use advantageous steps from both the algorithm and improve the performance in terms of precision, recall and accuracy in credit card fraud detection technique.

5. REFERENCES

[1] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection," Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., 1997.

[2] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results," Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90, 1997.

[3] Philip K. Chan, Florida Institute of Technology Wei Fan, Andreas L. Prodromidis, and Salvatore J. Stolfo, Columbia University "Distributed Data Mining in Credit Card Fraud Detection", November-December 1999, IEEE.

[4] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, 2002.

[5] C. Chiu and C. Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e-Service.2004.

[6] Abhinav Shrivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar (, Senior Member, IEEE "Credit Card Fraud Detection Using Hidden Markov Model" IEEE Transaction on Dependable and Secure Computing, January-March2008

[7] V.Dheepa1, Dr. R.Dhanapal "Analysis of Credit Card Fraud Detection Methods" International Journal of Recent Trends in Engineering, Vol2, No.3, November 2009.

[8] Ghosh and Reilly "credit card fraud detection with neural network", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI June2011.

[9] Monowar Husain Bhuyan, D K Bhattacharyya and J K Kalita introduced a paper with title "Survey on Incremental Approaches for Network Anomaly Detection", August 2011.

[10] Ramkumar.E &Mrs. Kavitha.P, "Online Credit Card Application and Identity Crime Detection", International Journal of Engineering Research & Technology (IJERT) February- 2013.