A Survey on Outlier Detection in Financial Transactions

Pradnya Kanhere
Dept. of Computer Engineering
MMCOE
Karvenagar, Pune, India

H. K. Khanuja
Dept. of Computer Engineering
MMCOE
Karvenagar, Pune, India

ABSTRACT
Outlier detection is a very important concept in the data mining. It is useful in data analysis. Nowadays, a direct mapping can be found between the data outliers and real world anomalies. Hence the outlier detection techniques can be applied to detect the abnormal activities in the real world. Outlier detection has been researched within various application domains and knowledge disciplines. This survey provides an overview of existing outlier detection techniques that can be applied in the financial domain. It mainly focuses on the idea of detecting the suspicious or outlier financial transactions.

General Terms
Data mining, Outlier detection

Keywords
Abnormal transaction, Suspicious transaction, Financial transaction

1. INTRODUCTION
In the Information Technology era, information has become a crucial resource for all organizations. Organizations gather information to aid in everyday activities as well as in decision making. Corporations must have certain processes to detect the abnormal behavior in such activities in order to guarantee the legality, security and quality of their information.

Outlier detection has been found to be useful and directly applicable in a large number of domains. This has resulted in a huge diverse literature of outlier detection techniques. Some of these techniques have been developed in a generic fashion while the others have been developed to solve specific problems in to a particular application domain. This survey aims at providing an overview of the research done in the field of outlier detection for financial domain. We also focus on literature detecting outliers in audit log for financial transactions.

Financial management systems handle high performance databases. Millions of transactions are carried out every day on these databases which result in huge amount of data. In financial system transactions, people hide their actions through a series of steps that make it look like money coming from illegal or unethical sources was earned legitimately. In order to take any action against such cases, financial institutions are required to keep an eye on database transactions to detect the abnormality if any.

1.1 Outliers
An outlier is defined in [1] as a data object that deviates significantly from other objects in the set. Outliers exist in almost every real data set. The important feature of outlier detection is that the outliers can be directly related to the real life behavior. In financial systems, these outliers are often referred to as suspicious activities, anomalies, abnormal behaviors, frauds, etc. In the financial application domain, outlier detection is usually used for credit card fraud detection, tax fraud detection, online transaction fraud detection, ATM card fraud detection and many such areas. Data mining techniques can be used to identify these outliers due to the fact that outliers in data translate to significant information which is useful for analysis and activity reporting. E.g. outliers in ATM card transaction data could indicate ATM card theft or misuse.

This survey discusses techniques which find such outliers in data for financial systems.

1.2 Challenges in Outlier Detection
Identifying and analyzing the unseen area is the major challenge in outlier detection. An outlier is a pattern that does not conform to expected normal behavior. So the simple approach is to define an area indicating normal behavior and declare any data object which does not belong to this normal area as an outlier. But there are several challenges in implementing this approach.

- Defining a normal behavior considering every possible behavior pattern is very difficult.
- The currently defined normal behavior may not correctly represent the normality in future as sometimes the normal behavior keeps evolving.
- Sometimes there is a thin boundary between the outlying and normal behavior. So detecting outliers which lie close to the boundary is difficult.
- The properties of an outlier are different for different applications. Every application has its own set of requirements and constraints.
- In many cases, people behave in a manner such that when observed, their behavior looks normal but in reality they are doing unethical things or suspicious activities.

Considering the above challenges, a detailed specification of the problem is required in order to detect the outliers in specified application domain. Because development of a framework based on a general idea about the problem is a difficult task.

2. LITERATURE SURVEY
The vast amount of literature is available in the area of outlier detection, but very few papers focus on using outlier detection technique for identifying the suspicious or abnormal transactions. Among the several outlier detection techniques, the few which can be easily applied to outlier detection are surveyed in following sections.

2.1 Rule-based Approach
Harmeet Kaur Khanuja et al.[2] 2014, proposed a forensic methodology for private banks, which included the monitoring of transaction audit logs as per Reserve Bank of
India (RBI) guidelines to mark the suspicious transactions if any and the Dempster-Shafer Theory of Evidence to generate reports. In paper by Rajput et al.[3] 2014, they proposed an ontology based system for detecting suspicious transactions based on a set of Semantic Web Rule Language(SWRL) rules and domain knowledge. Nida S. Khan et al.[6] 2013, presented a Bayesian network (BN)-based approach which assigns calculates the customer’s transaction behavior score based on transaction history and later on generates an alert if a significant difference is detected in customers historical transactional pattern and the current behavior. Suvasini Panigrahi et al.[13] 2009, proposed a system for database intrusion detection which made use of rule-based approach along with belief combination, database history and Bayesian learning component to collect the evidences about the transactional behavior, mark the transaction as suspicious and raise alarm.

2.2 Clustering-based Approach
Anu et al.[4] 2014, aimed to identify the illegal customer transactions in banks by differentiating the behavior as non-suspicious and suspicious. Dang Khoa Cao et al.[7] 2012, proposed the money laundering detection approach by using clustering techniques (CLOPE algorithm) for money transferring data of banking system, which considers bank transaction attributes like Transaction account, Sum of sending, Sum of receiving, etc. to define the behavior of accounts. Saleha Raza et al.[11] 2010, presented an approach for reporting suspicious activity using a combination of dynamic Bayesian network and clustering to capture transactional sequence patterns and compare anomaly index against a threshold to mark the transaction as suspicious or non-suspicious. Gao Zengan[12] 2009, proposed a new cluster-based local outlier factor (CLOF) algorithm that is a combination of distance-based unsupervised clustering and local outlier detection techniques, to identify suspicious money laundering transactional behavioral patterns.

2.3 Classification-based Approach
Stefan Axelsson et al.[8] 2012, analyzed the implications of using machine learning techniques for money laundering detection in a data set consisting of synthetic financial transactions and aimed to detect anomalies inside a data set of mobile money financial transactions by using the classification techniques to group transactions as suspicious or non-suspicious. Xingji Wang et al.[14] 2009, proposed a novel algorithm to detect money laundering using an improved minimum spanning tree clustering, an analysis of similarity measure and distance metric.

2.4 Model-based Approach
Mahesh Kharote et al.[5] 2014, proposed the use of transaction flow analysis and customer behavior analysis for profile generation in order to decide whether the customer is suspicious or not. Manuel Meja Lavalle et al.[9] 2011, presented a method to detect outlier or exceptional transaction records by applying an user modeling which works in two stages where the first stage obtains user behavior according to categorical or numerical attributes in the user’s transaction history and the second stage monitors a new transaction, compares it against the corresponding user model and determines if this transaction is unusual. Asma S. Larik et al.[10] 2010, presented an anomaly detection approach that uses clustering to identify customers normal behavior and statistical techniques to determine deviation of a particular transaction from the corresponding group behavior.

3. COMPARATIVE ANALYSIS
From the surveyed literature, we found that each approach has its own requirements and constraints.

- As far as the outlier detection in financial transactions is considered, the rule-based technique needs to refer a standard set of rules. These rules should be obtained from authentic sources like Reserve Bank of India, Financial Intelligence Unit, etc. Other than that this approach is not able to judge the user transaction based on his historic behavior.
- The regular clustering algorithms in data mining like K-means clustering, etc. cannot be directly applied to outlier detection in financial transactions, as the transactions contain not only the numerical attributes but also categorical attributes. Hence a modified approach based on clustering concepts should be developed. The clustering approach can just add the transaction in one of the clusters (suspicious, non-suspicious), but there are chances of confusion when the transaction behavior actually lies on the cluster boundary.
- The classification based approach marks the transaction as a part of one of the category among the several pre-defined categories.
- The model-based technique needs to access user’s transaction history to build the user’s normal behavior profile. During analysis phase every transaction by the user has to be compared against the respective profile to detect the abnormality if any. That means this technique is a two-stage process.

Considering the limitations and challenges for each of the above mentioned approaches, we conclude that the use of more than one approach in the category using a single approach to outlier detection will be more beneficial. This can be framed as a hybrid technique or application of one approach after another in serial fashion to efficiently detect the outliers in financial transactions.

4. PROCESS OF OUTLIER DETECTION
The process of outlier detection involves application of an outlier detection technique on input data and supporting data (if required) to generate the output (a list of outliers) as specified in Figure 1.

4.1 Input Data
In financial system domain, the input usually contains the series of transaction data or it can be the audit log records set on these transactions. Each input can be described using its own set of attributes such as its features, dimensions (numerical, categorical), etc.

4.2 Outlier Detection Technique
Among the several outlier detection techniques, the few which can be easily applied to outlier detection in financial transaction data are as follows:

- Rule-based technique: This technique considers a set of rules. The rules either define the normal behavior or the outlier behavior. The data objects are categorized according to whether their behavior matches with the rules or not.
Model-based techniques: This technique builds the user model based on user's transaction history and later on checks the user's transactions against their own behavior. If any deviation is detected, the data object is detected as an outlier.

Clustering-based technique: This technique partitions the data into groups which contain similar objects. The assumption about behavior of outliers is that they either do not belong to any cluster, or belong to very small clusters, or belong to a cluster which itself indicates the outlier behavior.

Classification-based technique: The outlier detection technique based on classification operates in the two-phases. The first training phase builds a model using the available labeled (normal, outlier) training data and then the testing phase then classifies the unseen objects into one of the learnt classes.

Combination of above techniques

4.3 Output

Different outlier detection techniques provide output in different manner. Different techniques of reporting an outlier can be listed as mentioned by [15]:

Labeling Techniques: These techniques assign a label (normal or outlier) to each transaction which is tested. They work like a classification algorithm. The advantage is that they provide an exact set of outliers for the analysis. The limitation of these techniques is that they cannot differentiate between different outliers, i.e. ranking score is not provided for the outliers.

Scoring Techniques: These techniques assign an outlier score to each pattern by considering the degree of deviation. These techniques provide a ranked list of outliers. Analyst then can choose one threshold value and analyze the outliers above that rank threshold. But choosing the correct threshold is a difficult task.

Supporting data of required e.g. threshold value

Input Data → Outlier Detection Technique → Detected outlier

Fig 1: Process of outlier detection

5. CONCLUSION

This paper provides a survey on outlier detection methodologies for detecting abnormal financial transactions. It also defines the process of outlier detection with selected techniques. We conclude that a framework for suspicious transaction detection can be designed based on the outlier detection approach using one or more of the listed techniques. This framework should be based on the idea to apply the outlier detection techniques in hybrid fashion or one after another in serial fashion, for achieving effective outcome.

In future this outlier detection framework can be extended to detect abnormal transactions from audit logs of financial systems. As audit log is one of the important aspects to be considered in data analysis. The outlier detection in audit logs will greatly help the auditors in data analysis.

6. ACKNOWLEDGMENTS

We would like to thank the Marathwada Mitramandal’s, College Of Engineering, Pune for providing the required facilities, Internet access and important books.

7. REFERENCES

[1] Outlier detection- Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques”