ISSN:0975-9646

Data Mining: An Improved Approach for Fraud Detection

Gogu.Sandeep[1],Sachin Malviya[2], Dheeraj Sapkale[2]

[1]*Department of Computer Science,* [2]*Department of Information Technology*
Medicaps Institute of Science & Tech,
Indore, India

***Abstract*---Fraud detection is a technique of identifying prohibited acts that are occurring around the world. It defines the skilled impostor, formalizes the key forms and sub forms of recognized frauds and reveals the gathered data nature. To analyze fraud patterns from data this paper represent preferred data mining techniques. Now a days Data mining is widely used concept, that can be used everywhere for analyzing data patterns. This paper reflects a mannered way by which any number of frauds can be analyzed. This paper also describe different types of frauds and their detection through data mining terminologies and a direction is given by which this detection can be moved in a simplified way. Theme of this paper is to firstly identify the type of fraud patterns using data mining techniques than resolve that criminal aspect in simplified way.**

***Keywords* ---Data mining, fraud pattern detection.**

I. INTRODUCTION

Data mining is about discovering new patterns which are unknown before, statistically reliable and process able from data. Data mining is a field which is concerned to understanding data patterns from huge datasets. We can say that the aim is to find out new patterns in data. A number of data mining techniques are there like classification, clustering, advanced neural networks, prediction and regression models used for different data mining approaches in various areas. Another area we are discussing here is fraud detection.

Fraud detection is the identification of symptoms of fraud where no previous disbelief exists. Firstly we have to learn that given data pattern is fraudulent or not. There are two kinds of learning data set supervised and unsupervised. Supervised learning of data set deals with fraud data that is previously known and unsupervised learning of data set deals with fraud data that is not previously considered as a fraud data but after sometimes they reflect the nature of fraud or crime. Then we treat those data patterns according to their behavior. Different terms are used for doing that task, they are described as techniques and methods for fraud or crime detection.

II. TECHNIQUES USED FOR ANALYSE AND DETECT FRAUD PATTERNS

In Data mining generally four modules of task exist:

1) Classification - Data are arranged into predefined groups with the use of different algorithms. Classification is the grouping of data in predefined classes. We can also say supervised classification, the classification which uses given class to arrange the objects in the data group. Classification techniques generally use a training set for objects those are previously grouped by known class. Algorithms of classification learn from the training set and create a structure for this. This structure is used for groping new objects. Different classification techniques used for fraud data patterns according to their nature.

2) Clustering – It is also like classification but predefined classes do not exist there, so that clustering algorithms attempt to do similar objects together. Like classification technique, clustering is the association of data in classes, but unlike classification, in clustering, classes are not predefined. We can say that Clustering is an unsupervised classification, because the classification is not based on previously known classes. Clustering approaches based on the principle of similarity maximization among intra-class objects and similarity minimization among inter-class objects.

3) Regression – In this process we try to obtain a function which models the data of the minimum error. A general term is to use Genetic Programming.

4) Association rule – It is used to find relationship among data objects. This analysis of association of objects is the discovery commonly called association rules. It observes the frequency sets occurring simultaneously in transactional database.

It is based on two threshold values support and confidence. Support, identifies the frequent item sets and confidence is the conditional probability that an item appears in a transaction when another item appears.

These are the basic terms used in the data analysis using data mining but now we understand techniques used according to fraud patterns.

III. DETECTION OF FRAUD PATTERNS IN CELLULAR NETWORKS

Firstly we inspect call details, billing data of cell users so that we can build up modules and using data mining techniques we detect fraud nature and normal nature of that modules. Firstly we analyze the data set and experience that some elements like Gender, Account type, voucher types, billing and calling comparison reflect a tendency of fraudulent use. After that we define clusters of users according to their behavior of calls. For clustering of users we use k-means clustering technique. We achieved the goal by performing the data mining techniques like Decision tree, Association rule, Neural Networks for training sets and test sets. After that we get results based on performance measures such as exactness, sensitivity.

Steps by which we can analyze and detect communication fraud:

- Firstly understand the required data set(call data, billing data, recharges type, calling zones, duration of calls).
- Establishing relations among data sets to know behavior of mobile user.
- For grouping related data we analyze clusters.
- K-mean clustering and kohenon neural net may be used for clustering but k-mean is faster than others.
- Highly redundant data variables removed after analyzing clustering variables, for that task component extraction technique can be used.
- After that we analyze the relationship among variables to detect the structure of variables. For that we can use different statistical analysis softwares.

IV. Detection Of Swap Card Fraud Patterns:

Swap Cards are increasing as a well-liked transaction medium. Fraud Detection is needed for examining the behavior of users with the intention of detect or avoid undesirable activity in future.

Swap card fraud detection is the method of recognizing those transactions that are bogus and partitioned these databases into classes of real and fraudulent transactions. Swap card frauds can be classified into three classes, conventional card frauds, commercial frauds and Internet frauds. Swap Card Fraud detection is a typical task when using normal data mining techniques, so the model for detecting swap card frauds is different from conventional models.

These models are mostly based on statistics or AI-driven which have the theoretical benefits. Information with time consideration is required to detect fraud activities so that we can specify what action is done by user on which time. Bank and other business fields have large amount of database, so sometimes it is difficult to extract fraudulent data from the database.

Data mining clustering technique is used to make groups of the data with same behaviors. This technique detects the clusters of odd behavior. That behavior is notified according to one moment of time, if someone at once acts roughly .Other technique is to set a breakpoint, that technique notices the use of single card, such as suddenly user withdrawing large amount using card.

Suspicious data detection from large database can be done through Bayesian Classification and also we can use neural networks classification techniques. Bayesian classification uses as an expert system and it draws graphs so as to describe natures of different classes. We can also use fuzzy logics to differentiate suspicious and non suspicious classes from large amount of databases.

V. Fraud Claims Detection

If an organization wants to track fraudulent claims, then it may be possible using data mining techniques. In most cases, the organizations identify fraudulent claims by getting information from related parties then examine their views. In this, we are taking an organization, which identifies patterns that produce fraud claims for that organization.

For finding fraud claim data patterns we analyze following:

1) Data sets those are hidden by the claimer and how to manage that data.
2) Measure the data variables those are provided as inputs
3) How much amount of data variables is fraudulent?
4) Finally we determine factors which are helpful to create prediction rules.

For identifying those fraud patterns we use data mining techniques such as decision trees, regression and neural networks. These methods are very useful for relating input variables with fraud patterns prediction and building rules for new input patterns. But neural networks need to learn neurons, so we use regression and decision trees. Decision tree group all the inputs into smaller groups as predictive of the output field.

According to type and quantity of data available there Decision tree provides best approach for generating rules according to their classification and understanding fraud patterns.

VI. Company Fraud -Patterns Detection By Data Mining Techniques

Data mining techniques are also useful for detecting company related fraud patterns. Many companies use data mining techniques for pattern and trend analysis and also in decision-making processes. Together with fraud detection, data mining on structured data has helped companies in areas like Pricing and Product Analysis, Market Analysis, Claims Trends and so on. The success of techniques of data mining depends mostly on quality and exposure of data.

From available data we can analyze and predict fraud patterns. For building similar behavioral pattern group we can apply Decision Tree Based algorithms and Naïve Bayesian Classification. Then classifier predictions can be interpreted by us using those methods. Now we can say that fraud detection in case of company related data can be done using a model that is supported by Rule-Based Classification, Decision Tree visualization and Bayesian Naïve Visualization.

VII. Conclusion

We studied different fraud detection data mining techniques according to different areas. Data mining is a well known zone of analyzing, predicting and defining rules from the large amount of data and finding true, previously unknown patterns. This paper focuses on data mining techniques as impressive approach for fraud patterns detection in every area.

References

[1]Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth*" From Data Mining to Knowledge Discovery in Databases"* Article.

[2] Zhe Hu, Linlin Ge, Xiaojing Li and Chris Rizos ***"****DESIGNING AN ILLEGAL MINING DETECTION SYSTEM BASED ON DINSAR*"

[3] Ravi S. Sandhu, "*ON SOME RESEARCH ISSUES IN MULTILEVEL DATABASE SECURITY*."

[4] QIANG YANG, XINDONG WU "*10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH*"

[5] *Data Mining Applications for Fraud Detection in Securities Market Koosha Golmohammadi*, Osmar R. Zaiane.

[6]Bentley, P., Kim, J., Jung., G., Choi, J.: Fuzzy Darwinian Detection of Credit Card Fraud.
[7] Sitikantha Kar AND Feroz Khan A L *"Mining Unstructured Text Data for Insurance Analytics"*
[8] Shao, H., Zhao, H., Chang, G.: Applying Data Mining to Detect Fraud Behavior in Customs Declaration
[9] Nikunj shah *"Fraud Detection: Data Mining & Audit Tools"*