A Novel Approach to Detect Mischief Activities (Fraud) In On-Line Transaction

Dr. K. Subramanian
Vice-Principal & Head IT, J J College of Arts and Science, Pudukkottai.

Abstract: *Nowadays, Credit cards are used in E-Commerce and for other financial transactions. The Credit card fraud is widespread in the recent years. It leads to the financial loss, for the individuals and also to the merchants. Here, to find these mischief activities clustering and outlier detection techniques can be used. Using the Clustering the data sets are partitioned and outlier detection is used to find the fraudulent data.*

Key words: *Outlier detection, Clustering, Fraud Detection, Credit cards.*

I. Introduction

In Modern world most of the people are using the Credit cards. It is the most popular payment mode. Detecting the frauds in theses on-line transaction is a very difficult task. So, there is need to develop a model to detect these frauds, in the business area and also in the academics. Finding the mischief means identifying suspicious fraud cases. Clustering means grouping the similar data object into a single set or cluster. Outliers can be defined as an observation which appears to be inconsistent with the remainder of the data sets. Outlier is an object whose attribute value is significantly different from the values of its neighbor. Outlier detection is a data mining technique which identifies the abnormal values. In this paper, outlier detection algorithm is employed to find the fraudulent transactions.

II. Related Work

More research work has been carried out with special emphasis to prevent the credit card frauds. In the year 2002 M.J.Kim, Sam *et. al* .identified and suggested some approaches for the outlier data in the credit card transactions. They used the Bayesian and Neural networking concepts. Clustering based outlier detection techniques and grid based techniques are also devised to find the mischief data. Most of the outlier detection algorithms use the distance based approaches. In this paper, a system is developed to find the fraudulent data points and an algorithm is also conceived.

III. Fraud Detection System

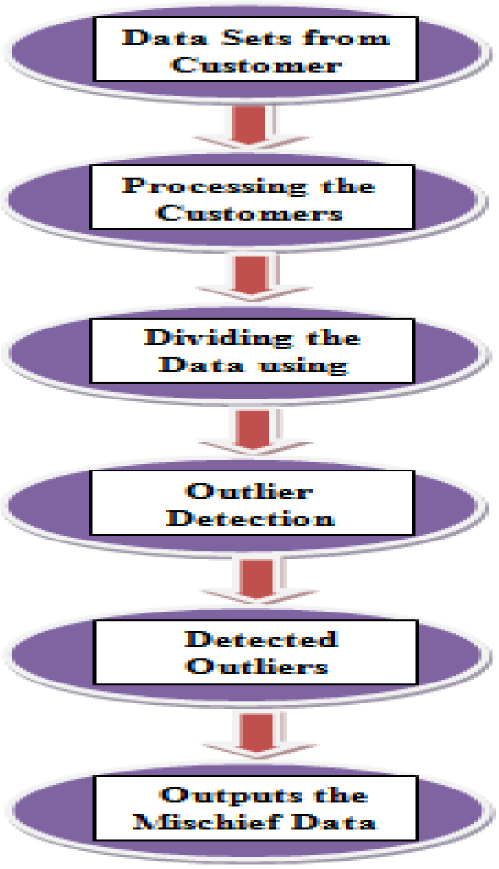

Figure 1: Fraud Detection System

Figure one shows the fraud detection system. Initially the data of the clients has been taken and split into group of similar data called clusters in second and third stage. From fourth stage onwards the standout data is find out using the outlier detection algorithm and if any listed.

IV. Methods for Detecting the Standout data

4.1 Clustering: Clustering helps the users to group the similar data into separate clusters or partitions. In this technique the data is broken down into related components. The goal of clustering is to find common patterns in the data or the similarity among the data. For clustering the customers data point can use the algorithms like k-means algorithm. This algorithm group the data as dense region and sparse region. This algorithm partitions ‘n’ observations into ‘k’ clusters. The cluster mean of ki = {gi1,gi2,…,gim} is defined:

$$Mi = \frac{1}{m}\sum_{j=1}^{m} gij$$

The partitioned data region is shown in the following diagram Density Based Spatial Clustering Application with Noise (DBSCAN). This algorithm was designed to discover the clusters and noise in the data sets. Let us define, D be a data set of points a cluster with respect to EPS and MinPts is a non-empty subset of D satisfying the following conditions: 1. For all p,q: if p belongs to C and q is density-reachable form p with respect to EPS and MinPts then q belongs to C. 2. For all p, q belongs to C: p is density-connected to q with respect to with respect to EPS and MinPts.

4.2 Outlier Detection: Outliers are the data which deviate from the other observations. One has to use different mechanisms to detect outliers from the dense as well as from the sparse region. The data which deviates from the other data indicates the abnormal activity. It is based on the mean and standard deviation of the data observed. Outliers observed from the dense and sparse region is taken and it is ranked. The data with the severity is listed as fraudulent data**.**

V. OUDOLCCT Algorithms

This algorithm is used to find the outlier data points in on on-line credit card transactions.

```
REPEAT WHILE i <= n (in a sorted list)
   CHECK1 r ( i ) = r ( i + 1 )
         CHECK2 r ( i + 1 ) = r ( i + 2 )
         OUTPUT RESULT “Not an Outlier”
         (OR)
         OUTPUT RESULT “The Outlier is” r ( i + 2 )
         END CHECK2
   (OR) OUTPUT RESULT “The Outlier is” r ( i + 1 )
END of REPEAT.
```

(SORTED ON, DATE AND TIME IN ASENDING)

An arrangement of n objects in a given order is called a permutation of objects by taking all the objects at a time. An arrangements of any r<=1 of these objects in a given order is called r-permutation. The permutations of n objects taken r at a time is denoted by p (n,r), observes that the first element in an r-permutation of n objects can be chosen in n different ways, following. The second element in the permutation can be chosen in n-1 ways, then next element in n-2 ways, continues till the last element in n-r+1 ways. Then p(n,r) = n (n-1) (n-2) . . . (n-r)

VI. Discussions

Using this present system the unknown outliers are easily found. Supervised approach can found the anomalies and normal data easily; in contrast, unsupervised approach not uses the labeled records. It finds the mischief or strange data sets. To find the outliers, in this paper Clustering and outlier detection methods are used. Nearest neighbor in the data sets are clubbed using Clustering approaches, this present system extracts mischief transactions, if any, as outlier.

VII. Conclusions and Prospect

This paper presents a system to detect the mischief in online transactions. The clustering approaches are used to group the data sets and outlier detection algorithm presented is for finding the outliers. Finally the standout data is listed as outliers. In future, the algorithm presented in this paper is updated to find the stand out data in the real time.