Fraud Detection in Online Transactions Using Data Mining Technique

Sayyed Shifanaz[1], Muzaffar Shabad[2], Kshirsagar Vaishnavi[3], Kadlag Pradnya[4], Kadam Nandkishor[5]
[1,3,4,5]*Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India*
[2]*Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India*

***Abstract*: In today's world frauds are increasing in both regular purchasing and online shopping. More increases in online shopping. For every transaction bank should check whether the transaction is fraudulent or not. Fraudster uses various techniques to generate the fraudulent transactions. Nowadays online transactions are more fraudulent than offline transactions. Due to increase in frauds more advance fraud detection system needed the most. Banking mostly use data mining techniques for credit card fraud detection. In this paper we propose a new credit card fraud detection system based on Behavior Certificate (BC) which reflects the cardholders' transaction habits using Random Tree data mining technique.**

Keywords: **Data Mining, Fraud Detection, Behavior Certificate.**

1. Introduction

Now a days the modes of payment methods are changed into online transactions. Banking system provides different type of payments like e-cash, card payments, internet banking, and e-services for improving online transaction. Credit card is one of the most custom ways of online transaction. Credit card is medium of selling goods or services without having cash in hand. With increased number of such cashless transaction, number of fraudulent transactions also increasing. During the online transaction we do not need any physical card, we need only card number, cvv, expiry date so there is more chances of fraud should be happen. In this method of fraud detection we generate behavior certificate on the basis of cardholder's transaction habits [3]. Most of the credit card fraud detection methods based on anomaly detection try to extract the historical behavior patterns as rules and compute the similarity between an incoming transaction and these behavior patterns [6]. The main idea of this kind of approach is that people may have personalized transaction habits that depend on their different accounts, different income sources, and different motivations and so on.

2. Review literature

N. Malini, M. Pushpa [1] the author uses KNN algorithm and outlier detection methods to optimize the best solution for the fraud detection problem. These are most important approaches that are proved to minimize the false alarm rates and increase the fraud detection rate. K-nearest neighbor algorithm is used largely in detection systems. It is also proved that KNN works extremely well in credit card fraud detection systems using supervised learning techniques. In this method the new instance query will be classified depending on the KNN category.

Xu Wei, Liu Yuan [2] in this paper they propose an optimized SVM model for detection of fraudulent credit card model. The model use non-liner SVM and RBF for the sparse transaction data, and use grid algorithm to determine the optional combination of parameters to detect the fraud .The support vector machine algorithm to construct an optimized SVM model for detection of fraudulent online credit card transaction, which helping the merchants make decision on whether to accept the deal. And then analyze the test results of each model. Krishna Modi, Reshma Dayma [3] to detect fraud behavior in this author proposed various methods of data mining such as decision tree, rule based mining, neural network, fuzzy clustering approach, hidden markov model or hybrid approach of these methods. Any of these methods is applied to find out normal usage pattern of customers (users) based on their past activities. M.Kavith, Dr.M.Suriakala, [4] author presents a real-time tree based metaclassifier TBMC that can be used to identify fraudulent transactions in huge imbalanced data. The developed metaclassifier based model operates based on predictions in two levels. The first level of predictions is performed by Random Forest classifier, and the second level predictions are performed by an ensemble created with Decision Trees and Gradient Boosted Trees. The results obtained from first and the second level prediction models are integrated to form the final predictions. Changjun Jiang, Jiahui Song [5] In this paper, they propose a novel fraud detection method that composes of four stages. To enrich a cardholder's behavioral patterns, they utilize the cardholders' historical transaction data to divide all cardholders into different groups such that the transaction behaviors of the members in the same group are similar.They thus propose a window-sliding strategy to aggregate the transactions in each group.Next,we extract a collection of special behavioral patterns for each cardholder based on the aggregated transactions and the cardholder's historical transactions. Then they train a set of classifers for each group on the basis of all behavioral patterns. Finally, they use the classifier set to detect fraud online and if a new transaction is fraudulent, a feedback mechanism is taken in the detection process in order to solve the problem of concept drift. John Richard D. Kho , Larry A. Vea [6] this paper is suggesting that a detection model must be available to capture the possible anomalous transactions – a fallback in case the technology will fail. Several classifiers were evaluated during the model creation however only the Random Tree and J48 yielded the highest accuracy value. Dhiya Al-Jumeily, Abir Hussain [7] states the types of fraud and current techniques which are being used to avoid fraudulent activities.This techniques supports the development of Fraud Detection System. Balasupramanian.N, Imad Salim Al-Barwani [8] suggests the big data analytics techniques to detect and prevent online fraudulent cases.This paper proposes a system in which data is collected cleaned & features are extracted.Using these pattern it can prevent the online fraud before it happens. Kadek Dwi Febriyanti, Riyanarto Sarno, Yutika Amelia Effend [9] belives some fraud occured due to variations in buisness processes.So this can be detected by applying association rule learning approach.This proposes an idea to present solution for detection of fradulent activity by learning the historical data. Dongxu Huang, Dejun Mu, Libin Yang [10] proposes a fraud detection system names as CoDetect which can uses both network and feature information for financial fraud detection. It can detect financial fraud activities and feature patterns associated with it.

3. Methodology

A. Random tree

Random Tree is the supervised classifier Random Tree is use to construct the random set of data for constructing decision tree. Random Tree algorithm deals with classification and regression problems. Random Tree is the group of tree predicators that called as forest. In Random Tree, classifier get input feature vector and classify it with every tree in forest and output of class label received majority votes. In regression, the classifier reply average of responses over tree in forest. Random Tree is the combination of two algorithms from machine learning. Single Model Tree combines with Random Tree to improve the functioning of Random Tree. Single Model Tree is the decision tree in which leaf node hold linear model. Random Tree improve the performance of decision tree. Random Tree is reasonably balance tree [6]. In this Random Tree one global setting works across the all leaves and thus simplifying optimization procedure. This feature of Random Tree reduces the time and efforts. Random Tree produce slightly better classification accuracy than Random Forest. Random Tree yielded the highest accuracy value of 94.32% [6]

B. Credit card fraud detection system

We have to perform FDS process in this section to detect fraud in credit card. FDS consist of two components BC construction and fraud detection. These two components are connected with each other using database. The BC construction process carried out offline while fraud detection process carried out online. In FDS, we have to find Behavior Certificate using BFV (Behavior Feature Vector). BFV which consist of 13 dimensions which describes cardholder transaction behavior.

BFV = (Weekday, Weekend, Festival, Normal Day, Interval1, Interval2, Interval3, Interval4, Location, Range1, Range2, Range3, Range4) is a behavior feature vector, where

- “Weekday” and “Weekend” are to represent whether a transaction take place in weekday or weekend.
- “Festival” and “Normal Day” are to represent whether a transaction take place in festival or not.
- Interval1- Interval4 are four time-intervals.
- “Location” is the area code of the place in which the transaction take place.
- Range1 – Range 4 are four amount-ranges. Location dimension has only value string as in numbers. All other dimensions returns Boolean value. Boolean values are in form of 0 or 1.

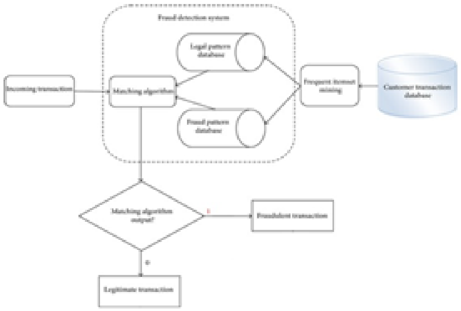

Fig. 1. Fraud Detection System

We have transactions dataset of users and on the basis of that dataset we compute BC. Whenever new transaction is ongoing at the same time from the customer transaction database we find some frequent datasets which are uses by customer for making transactions. In these transaction depending upon the BC factor some transactions are in legal pattern and some others in fraud pattern. These patterns are passed to algorithm. Algorithm compares the incoming transaction with existing datasets and gives output as transaction is fraudulent or not.

4. Conclusion

Through a survey of academic peers who were familiar with other fraud detection systems/tools we were able to gain information required for our prototype FDS. In this paper we proposed a new credit card FDS based on behavior certificates (BC) which reflects cardholder’s transaction habits. The correlation between behavior feature and some special cases such as festival, weekend are considered into BC. By applying this system we can detect fraudulent activities by studying its behavior certificate patterns.

References

[1] N. Malini , M. Pushpa "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection", 2017.

[2] Xu Wei, Liu Yuan "An Optimized SVM Model for Detection of Fraudulent Online Credit Card Transactions", 2012.

[3] Krishna Modi, Reshma Dayma "Review On Fraud Detection Methods in Credit Card Transactions", 2017.

[4] M.Kavith, Dr.M.Suriakala "Real Time Credit Card Fraud Detection on Huge Imbalanced Data using Meta-Classifiers", 2017.

[5] Changjun Jiang, Jiahui Song, Guanjun Liu "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism", March 2018.

[6] John Richard D. Kho, Larry A. Vea "Credit Card Fraud Detection Based on Transaction Behavior", 2017.

[7] Dhiya Al-Jumeily, Abir Hussain "Methods and Techniques to Support the Development of Fraud Detection System", 2015.

[8] Balasupramanian. N, Imad Salim Al-Barwani "User Pattern Based Online Fraud Detection and Prevention using Big Data Analytics and Self Organizing Maps", 2017.

[9] Kadek Dwi Febriyanti, Riyanarto Sarno, Yutika Amelia Effend "Fraud Detection on Event Logs Using Fuzzy Association Rule Learning " -2017

[10] Dongxu Huang, Dejun Mu, Libin Yang "Co Detect: Financial Fraud Detection with Anomaly Feature Detection", 2017.