Credit Card Fraud Detection Using Conditional Random Field and Probabilistic Hierarchical Clustering Algorithm

A. Tejhaskar[1], R. Vigneshwar[2]

[1]UG Scholar, Sri Sai Ram Institute of Technology

[2]UG Scholar, Sri Sai Ram Institute of Technology

Abstract- **During the last few years, the credit card system has been widely used as a process to initiate the global economy to develop significantly. A credit card providers has been issued millions of credit cards to their customers. While issuing credit cards to any wrong customers that can be a very crucial factor of a financial crisis. This paper presents an organized analysis and a survey of data mining methods and their applications in the credit card process. Our work focuses on data mining methods applied specifically in the credit card process which helps to emphasize much wider areas. This paper presents a novel mechanism for credit card fraud detection in combination of Conditional Random Field and Probabilistic Hierarchical Clustering Algorithm. The proposed method identifies the fraudulent activities on the credit card.**

Index Terms- **Data mining; credit card; New customer selection; fraud detection; Customer relationship management; repayment prediction.**

INTRODUCTION

In recent years, the rapid increase in using credit cards for making the purchase has caused a huge amount of data. These data can be useful for examine the pattern of the consumption behaviors of the customers. The credit card providers have been interested in planning the default risk of a credit card holder. Negative risk arising from the behaviors of customer which can lead to a big loss of money. Hence, the credit card providers need to use data mining methods for predicting and classifying customers more effectively.

Therefore, data mining is a very important technique for each and every activity of the credit card process. For example, it can be used for classifying good customers or bad customers which is totally based on

their application information and, also, detecting a misuse of a credit card based on purchase information of a customer. The efficiency of predicting the goodness or badness of an applicant can reduce credit risk of credit card providers. While, if the provider make any wrong decision by providing credit cards to bad customers, it will result in big loss of revenue and liquidity. This credit risk issue can leads to the financial crisis of the world economy.

Due to a huge amount of available data, process analysis in the credit card activity need to rely on data mining methods for its effectiveness and efficiency. Basically, data mining is process of extracting the patterns from the data. It helps to combine the method which is used to statistical, machine learning and database in order to extract and identify useful data from a lot of database.

Recently, there have been various works which helps to reviewing the applications of data mining methodology in the banking sector. In [18] the authors have studied data mining used in various activities in the banking sector, i.e., customer relationship management, fraud detection, marketing and risk management. This work did not investigate precisely on the credit card process and, therefore, data mining methods employed in such a manner is not obvious. Another survey of data mining applications in banking has been presented in which the concept of knowledge discovery in database method (KDD) was also discussed. Whereas, they didn't indicate the credit card process. In additional, there are surveying works concerning for particular areas of the credit card method such as fraud detection and credit scoring for customer's application. With respect to be more specific survey which can be for credit card providers and

researchers in this area, we have investigated research on data mining applications in main activities of the credit card methods.

II. ANALYSIS ACTIVITY IN CREDIT CARD PROCESS

In this section, we categorize the credit card fraud detection process into 3 main activities; fraud detection, new customer selection and customer relationship management (CRM).

A. Fraud Detection Activity:

There are various kinds of fraud including credit card fraud, telecommunication fraud, computer intrusion, bankruptcy fraud, Theft fraud or counterfeit fraud, application fraud and behavioral fraud. It has separated fraud into two categories: application based fraud and behavioral fraud. The application fraud refers to a fraudsters situation where a fraud a steals one's information or provides false information when applying for a credit card from a provider. The behavioral fraud means any type of abuses of a credit card by a fraud, for example, using a stolen credit card or counterfeit one.

Usually data mining techniques are used to search patterns of fraud as well as to classify customer behaviors as either fraudulent or non- fraudulent.

Data mining methodology is very important for fraud detection method. This is due to its computation performance on huge data. In additional, those methods need to have high precision results and low cost computation.

B. New Customer Selection Activity:

The aim of the new customer selection activity is to provide a qualified application to hold new credit card that can profit the problem. Credit card providers verify credit's applicant and approve the credit card into good customer group or non-risk group that these groups can score more than threshold that is defined by data mining which using the history of payment data of customers and default rate and other specifics. Credit card providers analyzed the attributes on the applicant. Hence, the credit scoring models use to illustrate the risk and support decision making the selection of new customers. This application allows credit card providers increase profit and reduce credit risk. Data

Mining has been used to manage credit scoring by choosing the attributes that contribute to indicate a low risk for approve the new car holder.

C. CRM:

Customer Relationship Management (CRM) is used to make relationship between card-holder and credit card provider to select a target customer and predicts card-holder churn using data mining which helps to find the pattern of the credit card holder's behavior that expense by their credit card which improve customer loyalty and retention. The most important reason of CRM activity is that the cost of retention exist customer is lower than a new customer. Data mining was applied clustering and classification.

III. PROPOSED METHODOLOGY

A. Create Cluster

The clustering algorithm is run on the historical data of the bank database. Probabilistic hierarchical clustering algorithm is used for clustering of transactions.

B. Spending Behavior

It tells the spending behavior of card holder. How much the card holder spend is his/her spending behavior. Suppose the card holder x spending are less than Rs. 10,000 per month then it come under low transaction customer and so on.

C. Initial Probability

The CRF takes an initial probability. The probability is provided to CRF for training purposes.

D. Training

We have used Baum-Welch algorithm for training transactions of the bank card holders. The algorithm has to provide with initial set of parameters. The initial probability should be uniform and based of N states. Steps involved in training are
1. Initialize Conditional Random Field
2. Perform Forward Moves
3. Perform Backward Moves
4. End

E. Fraud Detection

After the learning phase the CRF produces output in the form of probability. If the probability of false

transaction is more then it will be the fraud otherwise it is a genuine transaction.

3.1 Conditional Random Field

Conditional Random Fields (CRF) are a class of statistical modeling method often applied in pattern recognition and machine learning. CRF fall into the sequence modeling family. Whereas a discrete classifier predicts a label for a single sample without considering neighboring samples. CRF's are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. CRF is being an alternative to the Hidden Markov Model (HMM).

A. Description:

Let $G= (V, E)$ be a graph such that, $Y= (Y_v)$ $v \in V$, so that Y is indexed by the vertices of G. Then (X, Y) is a conditional random field when the random variables Y_v, conditioned on X, obey the Markov property with respect to the graph:

$P(Y_v | X, Y_w, w \neq v) = P(Y_v | X$, where $w \sim v$ means that w and v are neighbors in G.

CRF is an undirected graphical model whose nodes can be divided into exactly two disjoint sets X and Y, the observed and output variables respectively, the conditional distribution $P(Y | X)$ is then modeled.

3.2 Probabilistic Hierarchical Clustering:

A probabilistic hierarchical clustering can adopt the agglomerative clustering framework, but use the probabilistic models to measure the distance between clusters. It provides a probabilistic function by combining both the clusters. The algorithm is as follows:

Algorithm: A probabilistic hierarchical clustering algorithm.
Input: D= {o1... on}: a data set containing n objects;
Output: A hierarchy of clusters.
Method:
•create a cluster for each object $C_i = \{o_i\}$, $1 \leq i \leq n$
•for i = 1 to n
•find pair of clusters C_i and C_j such that C_i, C_j = argmax $i \neq j$ log $P(C_i \cup /C_j) / P(C_i)$. $P(C_j)$
•if log $P(C_i \cup /C_j) / P(C_i)$. $P(C_j) > 0$ then merge C_i and C_j
•else stop;

IV. RESULTS

In this section output produced by Conditional Random Fields is presented. The output are in the form of probability. If there is fraud then SMS is send to the customer. The customer can instantly block the card or roll back the particular transaction.
The input parameters are:

Attributes	Value
Number of HMM States	3
Number of Clusters	3
Threshold Value	30%
Initial State Probabilities	1/3 (all 3 states)

Table1: Output values of fraud detection
The output of above settings is:
Old Customer Profile:
84.4
New Customer Profile
84.4
The customer profile does not deviates from its behavior that means it is genuine transaction.
Old Customer Profile:
74.35
New Customer Profile
102.35
The customer profile deviates from its behavior that means it is fraud transaction and the transaction is rolled back.

VI. CONCLUSION

Basically, classification and prediction assignment are very important in the credit card process. Therefore, data mining has been choose to use in every activity of the credit card process by the credit card provider. Therefore, more importantly there are many credit card providers are interested in finding methods which can be help them to reduce cost as well as increase profit. This survey has found that there has been three important features that helps to make decision models more accurate. They contain an organized analysis, an appropriate selection of data set and a suitable time period of data-set.
Proposed method identifies the fraud and genuine users based on the past transactions. The old and new customers profiles are checked and based on that the clusters are created and fraudulent are identified.

REFERENCES

[1] Tanmay Kumar Behera, Suvasini Panigrahi, "Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network", 2015 IEEE Second International Conference on Advances in Computing and Communication Engineering.

[2] V.Mareeswari, Dr G. Gunasekaran, "Prevention of Credit Card Fraud Detection based on HSVM",International Conference on Information Communication and Embedded System (ICICES 2016).

[3] Jaba Suman Mishra, Soumyashree Panda, Ashis Kumar Mishra, "A Novel Approach for Credit Card Fraud Detection Targeting the Indian Market", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 3, No 2, May 2013.

[4] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," Journal of the operational research society 54.6 (2003):, pp. 627-635.

[5] B. M. RAMAGERI, and B. L. DESAI, "Role of data mining in retail sector," International Journal on Computer Science and Engineering (IJCSE), vol. 5, 2013, pp. 47-50.

[6] D. Sánchez, M. A. Vila, L. Cerda, and J.M. Serrano "Association rules applied to credit card fraud detection," Expert Systems with Applications, vol. 36, 2009, pp. 3630-3640.

[7] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", ISBN: 978-0-12-381479-1

[8] G. Wang, L. Liu, Y. Peng, G. Nie, G. Kou, and Y. Shi, "Predicting credit card holder churn in banks of China using data mining and MCDM," In Web Intelligence and Intelligent Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on vol. 3, 2010, pp. 215-218.

[9] IC. Yeh, and C. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of defualt of credit card client," Expert Syetem with Applications 36, 2008, pp.2473-2480.

[10] L. Delamaire, HAH Abdou, and J. Pointon, "Credit card fraud and detection techniques: a review," Banks and Bank systems, 4(2), 2009, pp. 57-68.

[11] R. Roselin, and C. Hanupriya, "Customer Behaviour Analysis for Credit Card Proposers Based on Data Mining Techniques," International journal of Innovative research in Advanced Engineering (IJIRAE) 2014.

[12] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J.C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, 50(3), 2011, pp. 602-613.

[13] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J.C.Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, 50(3), 2011, pp. 602-613

[14] S. H. Ha, and R. Krishnan, "Predicting repayment of the credit card debt," Computers & Operations Research 39.4, 2012, pp.765-773.

[15] S. Y. Sohn, K. T. Lim, and Y. Ju, "Optimization strategy of credit line management for credit card business," Computers & Operations Research, vol.48, 2014, pp. 81-88.

[16] V. Jayasree, and R. V. S. Balan, "A review on data mining in banking sector," American Journal of Applied Sciences, 10(10), 2013.

[17] Y. Kou, C. T. Lu, S. Sinvongwattana and Y.P. Huang, "Survey of Fraud Detection Techniques," Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control