

Credit Card Fraud Detection Using HMM and K-Means Clustering Algorithm

Suman Kumari
SSGI Bhilai

Dept. of Computer Science and Engineering
Raipur, Chhattisgarh, India
sumankumari516@gmail.com

Dr. Abha Choubey
SSGI Bhilai

Dept. of Computer Science and Engineering
Raipur, Chhattisgarh, India

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 HMM and K-Means Algorithm. The proposed method identifies the fraudulent activities on the credit card.

Keywords— *data mining; credit card; New customer selection; fraud detection; Customer relationship management; repayment prediction.*

I. 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 a 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 [8], [12], and [15].

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 [14] 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 [8], [9], [12], [19] and credit scoring for customer's application [6]. 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 [7]. 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 [18]. The behavioral fraud mean 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 allow 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 cardholder.

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 [17]. The most important reason of CRM activity is that the cost of retention exist customer is lower than a new customer [10]. Data mining was applied clustering [13] and classification.

III. LITERATURE SURVEY

Tanmay Kumar Behera [1], have proposed a novel approach towards credit card fraud detection in which the fraud detection is done in three phases. The first phase does the initial user authentication and verification of card details. If the check is successfully cleared, then the transaction is passed to the next phase where fuzzy cmeans clustering algorithm is applied to find out the normal usage patterns of credit card users based on their past activity.

V.Mareeswari [2], proposed a solving business problems in banking sectors are done mostly by contribution of Data Mining. Better targeting, acquiring new customers, and fraud detection in credit cards, fraudulent transactions can be done by Data mining techniques.

Mohd Avesh Zubair Khan [3], presents a survey of various techniques used in credit card fraud detection mechanisms and Hidden Markov Model (HMM) in detail. HMM categorizes card holder's profile as low, medium and high spending based on their spending behaviour in terms of amount. A set of probabilities for amount of transaction is being assigned to each cardholder. Amount of each incoming transaction is then matched with card owner's category, if it justifies a predefined threshold value then the transaction is decided to be legitimate else declared as fraudulent. Existing fraud detection system may not be so much capable to reduce fraud transaction rate.

Mahesh Singh [4], focused on Luhn algorithm is widely used on the Internet to validate of credit card numbers, but this algorithm suffers from weaknesses, as confirmed by tests.

Jaba Suman Mishra [5], present the necessary theory to detect fraud in credit card transaction processing using a Hidden Markov Model. If an incoming credit card transaction is not accepted by the Hidden Markov Model with sufficiently high probability, it is considered to be fraudulent. At the same time, author try to ensure that genuine transactions are not rejected. Author show how HMM helps to obtain high fraud coverage combined with a low false alarm rate.

E.W.T. Ngai [9], this paper presents a review of — and classification scheme for — the literature on the application of data mining techniques for the detection of financial fraud. Although financial fraud detection (FFD) is an emerging topic of great importance, a comprehensive literature review of the subject has yet to be carried out. This paper thus represents the first systematic, identifiable and comprehensive academic literature review of the data mining techniques that have been applied to FFD. 49 journal articles on the subject published between 1997 and 2008 was analyzed and classified into four categories of financial fraud (bank fraud, insurance fraud, securities and commodities fraud, and other related financial fraud) and six classes of data mining techniques (classification, regression, clustering, prediction, outlier detection, and visualization). The findings of this review clearly show that data mining techniques have been applied most extensively to the detection of insurance fraud, although corporate fraud and credit card fraud have also attracted a great deal of attention in recent years. In contrast, we find a distinct lack of research on mortgage fraud, money laundering, and securities and commodities fraud. The main data mining techniques used for FFD are logistic models, neural networks, the Bayesian belief network, and decision trees, all of which provide primary solutions to the problems inherent in the detection and classification of fraudulent data. This paper also addresses the gaps between FFD and the needs of the industry to encourage additional research on neglected topics, and concludes with several suggestions for further FFD research.

Guoxun Wang [10], Nowadays, with increasingly intense competition in the market, major banks pay more attention on customer relationship

management. A real-time and effective credit card holders' churn analysis is important and helpful for bankers to maintain credit card holders. In this research we apply 12 classification algorithms in a real-life credit card holders' behaviors dataset from a major commercial bank in China to construct a predictive churn model. Furthermore, a comparison is made between the predictive performance of classification algorithms based on Multi-Criteria Decision Making techniques such as PROMETHEE II and TOPSIS. The research results show that banks can choose the most appropriate classification algorithm/s for customer churn prediction for noisy credit card holders' behaviors data using MCDM.

IC. Yeh [11], this research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients. Because the real probability of default is unknown, this study presented the novel "Sorting Smoothing Method" to estimate the real probability of default. With the real probability of default as the response variable (Y), and the predictive probability of default as the independent variable (X), the simple linear regression result ($Y = A + BX$) shows that the forecasting model produced by artificial neural network has the highest coefficient of determination; its regression intercept (A) is close to zero, and regression coefficient (B) to one. Therefore, among the six data mining techniques, artificial neural network is the only one that can accurately estimate the real probability of default. 2007 Elsevier Ltd. All rights reserved.

L. Delamaire [12], Fraud is one of the major ethical issues in the credit card industry. The main aims are, firstly, to identify the different types of credit card fraud, and, secondly, to review alternative techniques that have been used in fraud detection. The sub-aim is to present, compare and analyze recently published findings in credit card fraud detection. This article defines common terms in credit card fraud and highlights key statistics and figures in this field. Depending on the type of fraud faced by banks or credit card companies, various measures can be adopted and implemented. The proposals made in this paper are likely to have beneficial

attributes in terms of cost savings and time efficiency. The significance of the application of the techniques reviewed here is in the minimization of credit card fraud. Yet there are still ethical issues when genuine credit card customers are misclassified as fraudulent.

R. Roselin [13], Consumer financing have become increasingly important in the private sector for the last two decades. With the new reforms in the banking sector, the marketing of financial products has become very competitive, creating a need for strategizing the marketing efforts. This study investigates the shift of consumers towards the use of plastic money, with emphasis on credit cards. A survey of consumers holding one or no credit card was used for data collection. Variables related to demographics such as age, income level and gender have also been taken into consideration. Data mining techniques are useful to analyses the customer behavior. Prediction is done using the relevant data set taken from the database on the basis of the attributes. Classification and Clustering algorithms are used to analyses customer behavior towards credit card. Based on our observations, suggestions have also been made for managers to refine the target market.

IV. METHODOLOGY

In this section we will deal with the process of detecting the credit card fraud using HMM model. The proposed system architecture is presented in fig. 1.

A. Create Cluster

The clustering algorithm is run on the historical data of the bank database. K-means clustering algorithm is used for clustering of transactions. In k-means clustering k is the number of observation. We have clustered it to find the spending behavior of high medium and low transaction customers.

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 HMM model takes an initial probability. The probability are provided to HMM for training purposes.

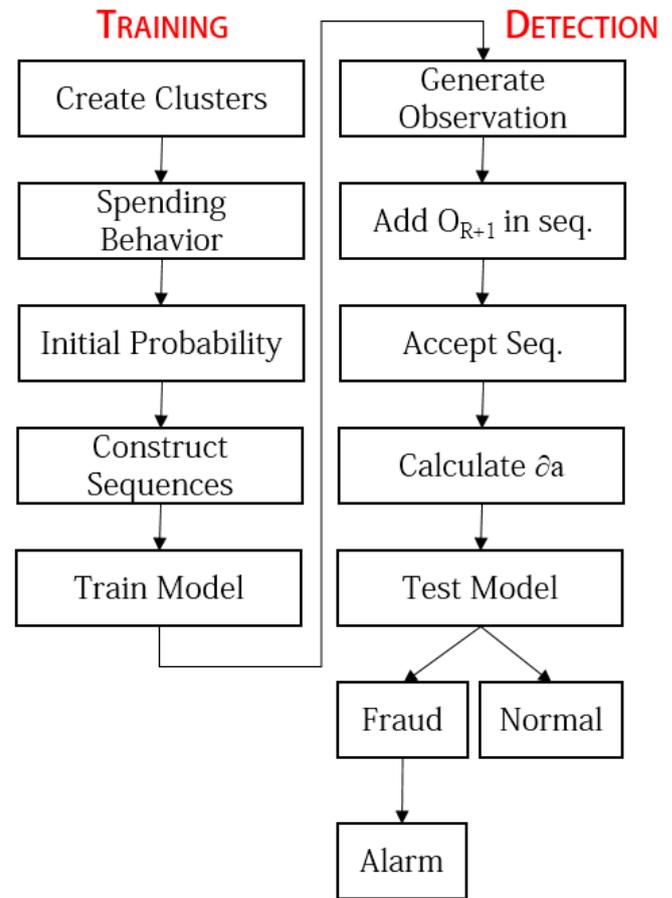


Fig. 1. Proposed System Architecture

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 HMM Model
2. Perform Forward Moves
3. Perform Backward Moves
4. End

E. Fraud Detection

After the learning phase the HMM 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.

V. RESULTS

In this section the output produced by HMM model 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)

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 alarm is raised.

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 customer 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] MohdAvesh Zubair Khan, Jabir Daud Pathan, Ali Haider Ekbal Ahmed, "Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014.

[4] Mahesh Singh, Aashima, Sangeeta Raheja, "Credit Card Fraud Detection by Improving K-Means", International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-2, Issue-5, May 2014.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] E. W. T. NGAI, Y. Hu, Y.H. Wong, Y. Chen and X. Sun, "The application of data mining techniques in

- financial fraud detection: A classification framework and an academic review of literature,” *Decision Support Systems*, 50(3), 2011, pp. 559-569.
- [10] 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.
- [11] IC. Yeh, and C. Lien, “The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card client,” *Expert System with Applications* 36, 2008, pp.2473-2480.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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
- [16] S. H. Ha, and R. Krishnan, “Predicting repayment of the credit card debt,” *Computers & Operations Research* 39.4, 2012, pp.765-773.
- [17] 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.
- [18] V. Jayasree, and R. V. S. Balan, “A review on data mining in banking sector,” *American Journal of Applied Sciences*, 10(10), 2013.
- [19] 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 Taipei, Taiwan*, vol 2., 2004, pp. 749-754.