

Bank Management under Data Mining Environment towards Fraud Detection

Md.Sarwar Kamal¹, Md. Aktar kamal², Sonia Farhana Nimmy³

^{1,3}Lecturer, Computer Science and Engineering
BGC Trust University Bangladesh.

¹sarwar.saubdcoxbazar@gmail.com, ³nimmy_cu@yahoo.com

²Assistant Professor, Department of Business Administration.
International Islamic University, Chittagong.

²makmgt@gmail.com

Abstract

Banking system is very essential in our daily life and it is very important to manage all the transactions in bank with secure and easy process. In the information superhighway we can incorporate the support of technology to make efficient and secure transactions with increasing data rate. Bank fraud detection is very essential part of the daily transactions to assure better transactions. This work emphasis on the improvement of the start-of -the -art in commercial practice through machine learning data mining. Proper technology to assess huge amounts of transaction data to count efficient fraud detectors in an exact time is a fundamental problem. We analyze the date set by K-Medoids Algorithm (KMA) and trained the system using Artificial Neural Network (ANN). Besides the KMA, we also imposed Fuzzy K-Means and Fuzzy K-modes algorithms. After proper training to the ANN, we noticed that K-Medoids Algorithms is better only for the linear and Quantitative data set. On the contrary, Fuzzy K-Means is better when there are lots of mixed data sets, I mean both for Qualitative and Quantitative. We suggest both of the algorithms for separate environments and policies.

Keywords: Bank fraud detection, machine learning data mining, Artificial Neural Network (ANN), K-Medoids Algorithm (KMA), Fuzzy K-Means and Fuzzy K-modes algorithms.

1. Introduction

The banking industry of Bangladesh has been growing fast during the last decade. Banks are the engines that drive the operations in the financial sector, which is vital for the economy and social change. After Independence, the banks have passed through three stages. They have moved from the character based lending to ideology based lending to today competitiveness based lending in the context of Bangladesh's economic liberalization policies and the process of linking with the global economy. Today, the banking system has entered into competitive markets in areas covering resource mobilization, human resource development, customer services and credit management as well. While the operations of the bank have become increasingly significant banking frauds in banks are also increasing and fraudsters are becoming more and more sophisticated and ingenious. In a bid to keep pace with the changing times, the banking sector has diversified its business manifold. And the old philosophy of class banking has been replaced by mass banking. The challenge in management of social responsibility with economic viability has increased. Bank frauds are the failure of the banker. It does not mean that the external frauds do not defraud banks. But if the banker incorporate the support of technology to make efficient and secure transactions through machine learning data mining it would be possible to reduced bank frauds.

2. The Banking Industry of Bangladesh

Pursuant to Bangladesh Bank Order, 1972 the Government of Bangladesh reorganized the Dhaka branch of the State Bank of Pakistan as the central bank of the country, and named it Bangladesh Bank with retrospective effect from 16 December 1971. Bangladesh Bank has been entrusted with all of the traditional central banking functions including the sole responsibilities of issuing currency, keeping the reserves, formulating and managing the monetary and credit policy, regulating the banking system, stabilizing domestic and external monetary value, preserving the par value of Bangladesh Taka, fostering economic growth and development and the development of the country's

market. After the independence, banking industry in Bangladesh started its journey with 6 nationalized commercialized banks, 2 State owned specialized banks and 3 Foreign Banks. In the 1980s banking industry achieved significant expansion with the entrance of private banks. The reform process was first initiated in 1982 when two of the six nationalized commercial banks were denationalized and a number of licenses were awarded for private sector commercial banks. Now, banks in Bangladesh are primarily of two types: Scheduled Banks: The banks which get license to operate under Bank Company Act, 1991 (Amended in 2003) are termed as Scheduled Banks. Non-Scheduled Banks: The banks which are established for special and definite objective and operate under the acts that are enacted for meeting up those objectives, are termed as Non-Scheduled Banks. These banks cannot perform all functions of scheduled banks. As of 15 December 2012, at present banking sector of Bangladesh comprises four categories of 57 scheduled banks. Among them four are the state owned commercial banks (SCBs), 6 are the state owned development banks (SDBs), thirty seven are the private commercial banks(PCBs) and ten are the foreign owned commercial banks(FCBs) (Bangladesh Bank website).

3. Bank Fraud

The uses of online transaction are growing day by day. From the study of ACNielsen study conducted in 2013, 67% of the world's population is shopping online [1]. Germany and Great Britain are two pivotal countries those are clearly ahead of online shopping, and credit card is the most popular mode of payment (59 percent). Approximately 650 million transactions per year were reportedly carried out by Barclaycard, the famous credit card company in the United Kingdom [2]. Retailers like Wal-Mart, Bata, and E-bay online system are typically handle much larger number of credit card transactions including online and regular purchases. The excessive data sets of credit card user throughout the world are rising and the opportunities for attackers to steal credit card details and, subsequently, commit fraud are also increasing. In a report from the USA the total credit card fraud is reported to be \$5.7 billion in 2012 and estimated to be \$7.0 billion in 2012, out of which \$2.6 billion and \$2.7 billion, respectively, are the estimates of online fraud [3]. Credit-card-based purchases can be categorized into two types: 1) physical card and 2) virtual card. In a physical-card based purchase, direct transactions are occurred at the various outlet or retailers or shopping mall. The persons who are responsible for the fraudulence activities try to steal the pin code or steal cards at various market places. If credit card users do not have much knowledge on online transactions and they are reluctant of the security of the card information, it can lead to a substantial financial loss to the credit card company. In virtual purchase, only digits are inputs to the system and submit to the desired places (card number, expiration date, and secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit virtual transactions fraud, only serial number is enough to make the crime. The only way to identify this kind of crime is to analyze the spending amounts on every card and to compare any inconsistency with respect to the "usual" spending amounts. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds.

4. Literature Review

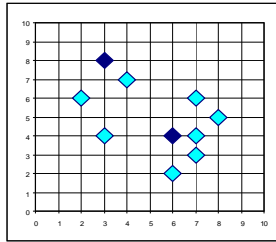
Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on data mining and neural networks, have been suggested. Ghosh and Reilly [4] have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labeled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and non received issue (NRI) fraud. Recently, Syeda et al. [5] have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection. A complete system has been implemented for this purpose. Stolfo et al. [6] suggest a credit card fraud detection system (FDS) using Metalearning techniques to learn models of fraudulent credit card transactions. Metalearning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A Metaclassifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection [7]. They use Java agents for Metalearning (JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them. Aleskerov et al. [8] present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection [9]. Based on

this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections. Fan et al. [10] suggest the application of distributed data mining in credit card fraud detection. Brause et al. [11] have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage. Chiu and Tsai [12] have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. With this scheme, participating banks share knowledge about the fraud patterns in a heterogeneous and distributed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. Phua et al. [13] have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report. Prodromidis and Stolfo [14] use an agent-based approach with distributed learning for detecting frauds in credit card transactions. It is based on artificial intelligence and combines inductive learning algorithms and Metalearning methods for achieving higher accuracy. Phua et al. [15] suggest the use of Metaclassifier similar to [6] in fraud detection problem. They consider naïve Bayesian, C4.5, and Back Propagation neural networks as the base classifiers. A Metaclassifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use credit card fraud detection as the target application, their approach is quite generic. Vatsa et al. [16] have recently proposed a game-theoretic approach to credit card fraud detection. They model the interaction between an attacker and an FDS as a multi stage game between two players, each trying to maximize his payoff. The problem with most of the abovementioned approaches is that they require labeled data for both genuine, as well as fraudulent transactions, to train the classifiers. Getting real-world fraud data is one of the biggest problems associated with credit card fraud detection.

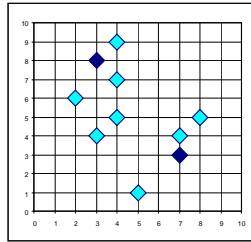
5. The K-Medoids Clustering Method

- i. Determine k data set arbitrarily
- ii. For each pair of non-identified data set h and determined data set i , calculate the total swapping cost TC_{ih}
- iii. For each pair of i and h ,
 1. If $TC_{ih} < 0$, i is changed by h
 2. Then assign each non-determined data set to the most similar data set.
- iv. repeat steps 2-3 until there is no change

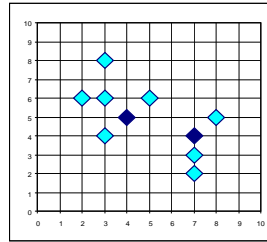
Total swapping cost



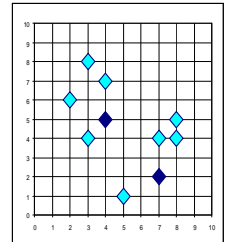
$$C_{jih} = d(j, h) - d(j, i)$$



$$C_{jih} = 0$$



$$C_{jih} = d(j, t) - d(j, i)$$



$$C_{jih} = d(j, h) - d(j, t)$$

Figure 1: The K-medoids clustering for various customer segmented mentioned at table 1.

6. The K-Means Clustering

Given m , the m -Means algorithm is implemented in 4 steps:

- i. Partition objects into m nonempty subsets
- ii. Calculate pivotal points as the centroids of the clusters of the new orientation. The
- iii. Partitioned each object to the cluster with the nearest pivotal point.
- iv. Go back to Step 2, stop when no more new assignment.

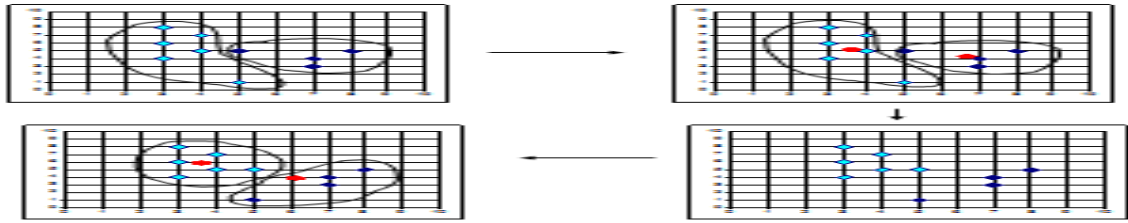


Figure 2: K-Means Clustering Algorithm.

7. Principal Component Analysis

From k original variables: x_1, x_2, \dots, x_k :

Produce k new variables: y_1, y_2, \dots, y_k :

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1k}x_k$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2k}x_k$$

...

$$y_k = a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kk}x_k$$

From k original variables: x_1, x_2, \dots, x_k :

8. Methodology

Here we have segmented the three categories of the customers based on their attitudes, gestures, patients, movements, demands, body languages, communication skills and literacy. According to the algorithms of K-Means and K-Medoids Clustering, we have noticed that K-Medoids perform very well and finally the Principle Component Analysis (PCA) correlates all the classifications results. We define a threshold value of standard behavior after talking with Branch managers of five banks including two governmental Banks. Based on their opinion we then check this parameter with the K-Medoids Algorithm threshold value. Then we finalized the value is $\theta=0.24$. The flow chart of the process is given below.

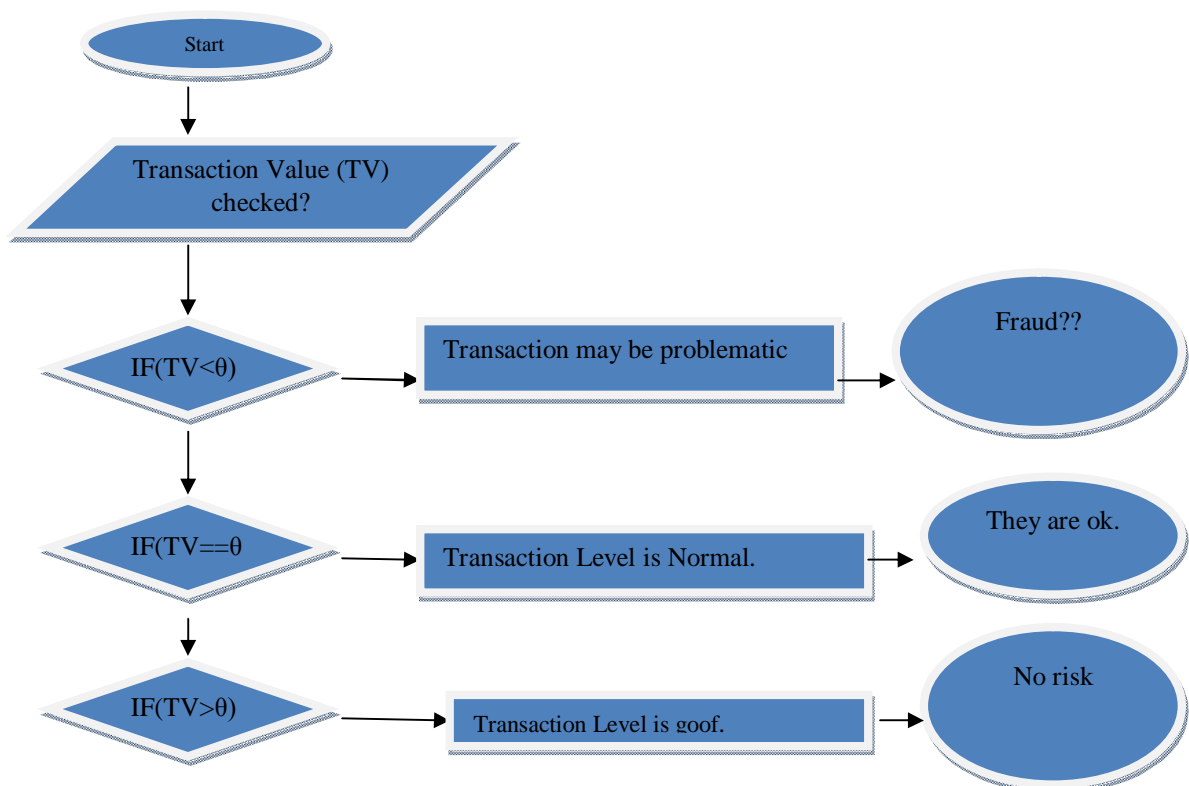


Figure 3: Flowchart for behavior clustering.

9. Result

Bank Name	Customer Name	Loan ID	Loan Account	Location	Fraudness
Brack Bank	Tapos Bormon	4001	124782	Patiya	Normal
Brack Bank	Afjal Hossain	4012	783141	Do	Normal
Brack Bank	Abdul Aziz	4902	321574	Do	Normal
Brack Bank	Hira Das	4092	492345	Do	Normal
Brack Bank	Morjina Begum	4982	438921	Do	Normal
Dhaka Bank	Earsad Ullah	8723	234120	Do	Normal
Dhaka Bank	Emam Uddin	8701	234045	Do	Normal
Dhaka Bank	Abul Khair	8728	230912	Do	Normal
Dhaka Bank	Towhidul Islam	8723	234570	Do	Normal
Dhaka Bank	Karim Uddin	8703	249072	Do	Normal
UCBL Bank	Jahedul Islam	5423	764848	Do	Normal
UCBL Bank	Nishita Shaha	3452	658392	Do	Problematic
UCBL Bank	MNormaltar Hossain	4563	652310	Do	Normal
UCBL Bank	Abutahar Mia	4567	680213	Do	Normal
UCBL Bank	Khursad Fazil	7832	602343	Do	Normal
Janata Bank	Asis Mia	0923	984536	Do	Normal
Janata Bank	Farhan Ali	0876	908765	Do	Normal
Sonali Bank	Tipu Sultan	3424	214567	Do	Normal
Sonali Bank	Korim Monshi	3214	213468	Do	Normal
Agrani Bank	Absar Mia	0123	987566	Do	Normal
Agrani Bank	Josim Uddin	0342	982342	Do	Normal

10. Conclusion

In this work we have found that the unexpected transactions may causes to identify the fraud on credit card transactions. Besides, clustering algorithm used here as K-Medoids algorithm with K-means clustering. But K-medoids performed better classification. After better clustering, the PCA helps us to predict according to the data set. From various bank we have checked more than ten thousand transactions. We have checked all the regular uses of the data set. We only get Nisitha Shaha as the victims who lost her Credit card and consequently get deprived the theft made the illegal transactions. These identification methods will helps to make an automated system for banking system. We have faced problems on extreme data set. To accomplish the total task we need huge data set that have some idea related on fraud transactions.

11. References

- [1] "Global Consumer Attitude Towards On-Line Shopping," http://www2.acnielsen.com/reports/documents/2005_cc_online_shopping.pdf, Mar. 2007.
- [2] D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams, "Data Mining for Fun and Profit," *Statistical Science*, vol. 15, no. 2, pp. 111-131, 2000.
- [3] "Statistics for General and On-Line Card Fraud," <http://www.epaynews.com/statistics/fraud.html>, Mar. 2007.
- [4] S. Ghosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," *Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems: Decision Support and Knowledge-Based Systems*, vol. 3, pp. 621-630, 1994.
- [5] M. Syeda, Y.Q. Zhang, and Y. Pan, "Parallel Granular Networks for Fast Credit Card Fraud Detection," *Proc. IEEE Int'l Conf. Fuzzy Systems*, pp. 572-577, 2002.
- [6] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results," *Proc. AAAI Workshop AI Methods in Fraud and Risk Management*, pp. 83-90, 1997.
- [7] S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," *Proc. DARPA Information Survivability Conf. and Exposition*, vol. 2, pp. 130-144, 2000.

- [8] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection," Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., pp. 220-226, 1997.
- [9] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.
- [10] W. Fan, A.L. Prodromidis, and S.J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection," IEEE Intelligent Systems, vol. 14, no. 6, pp. 67-74, 1999.
- [11] R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. Tools with Artificial Intelligence, pp. 103-106, 1999.
- [12] C. Chiu and C. Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e-Service, pp. 177-181, 2004.
- [13] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research," <http://www.bsys.monash.edu.au/people/cphua/>, Mar. 2007.
- [14] S. Stolfo and A.L. Prodromidis, "Agent-Based Distributed Learning Applied to Fraud Detection," Technical Report CUCS-014-99, Columbia Univ., 1999.
- [15] C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 50-59, 2004.
- [16] V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-theoretic Approach to Credit Card Fraud Detection," Proc. First Int'l Conf. Information Systems Security, pp. 263-276, 2005.
- [17] S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," ACM Trans. Information and System Security, vol. 3, no. 3, pp. 186-205, 2000.