

The Application of Combined Fuzzy Clustering Model and Neural Networks to Measure Valuably of Bank Customers

Raheleh Nasiri Sharifi

Student of Master's Degree in the Field of Computer Software Technology Engineering; Technical & Engineering Faculty, Islamic Azad University, Electronic Branch, Tehran; Iran

Maryam Rastgarpour

Assistant Professor; Technical & Engineering Faculty, Department of Computer, Islamic Azad University; Saveh Branch, Saveh, Iran

ABSTRACT

Currently, acquisition of resources in banks is subject to attraction of the resources of banking customers. Meanwhile, the Bank's valuable customers are one of the best resources to make profit for banks. Several different models are introduced for evaluation of profitability of the customers; but most of them are classical models and they are unable to evaluate the customers in complete and optimized manner. In this way, the conditions are ready to enter artificial intelligence to this scope. In this study, we tried to design an intellectual system for identification of valuable customers of Bank. Such a model is capable of reasoning and explaining the expert system together with adaptation and learning capability of neural network at the same time. Expertise is added to it through data mining. As a result, evaluating the proposed model is done through transaction data of the customer, and accuracy of 95.46 percent is gained.

Keywords

Valuable Customers of Banks, Modern Banking, Data Mining, Fuzzy Clustering, Neural Networks

1. Introduction

Mobilization of financial resources started from the very beginning of human social life with its trend to trade and exchange goods and commodities and it was always the key task of the banking system. There are several components in modern banking affecting on the financial resource mobilization in banks and financial Institutions. Identify and quantify the impact and relationship of these elements to the success of banks in mobilizing financial resources is an important issue. Nowadays, the situations and conditions of the financial institutions and banks are not identical with each other. Influencing factors in the financial resource mobilization may even be different for each branch of a banking group. Obviously, much of the country's banking sources are consist of savings of people. Thus, banks collect surplus deposits available to the

public in order to attract more resources. Meanwhile, a major part of bank customers are not profitable. When banks can be successful in acquiring these resources that in addition to attracting new resources, they can use optimal methods to recognize and maintain their customers. The reason of this condition is due a marketing rule that claims the cost of attracting a new customer is 5 times more than the cost of maintaining old customers, so maintaining the old profitable customers will be more efficient in banking industry. Therefore, it is essential that the country's banking industry make some efforts to design appropriate systems to assess customers. The system will be effective and applicable when it has appropriate measures to classify the customers, so that the valuable banking customers can be identified effectively by the system. From the perspective of bank, the customers are divided in two types i.e. regular customers and valuable or profitable customers. The bank is always trying to work toward retention of profitable customers because this category of customers while are only 20% of the customers of a bank; they bring up to eighty percent of the resources to the Bank. Thus, the most important duty of bank managers in this regard is to identify and classify the customers of banking network so that they can provide them with facilities and services based on a scientific and tested method. Therefore, study of the banking transactions of the customers is very important for their classification based on their value. In this study, we tried to use data mining in the banking field to cluster the customers based on clustering algorithms using the neural network in addition to the simulated expert mind to maintain the confidentiality of banking data. The paper is organized as follows: section 2 includes a review of the literature. Section 3 is describing the proposed model and in section 4 conclusions and recommendations for future research are discussed.

Review of Literature

Many international research papers are presented in the field of customer value evaluation, of which following examples can be provided:

- Fuzzy clustering of customers in a furniture company, due to the high volume and transactions grow, there is no possibility of recognizing the clients manually (Yuliari and et al, 2015)
- Design a customer relationship management system was done using fuzzy clustering for data mining of transaction data of clients in a telecommunications company. The reason for selection of fuzzy clustering method is the large amounts of data that in some cases this volume will also increase up to a terabyte level (Patil and et al, 2015).
- In another article 4459 clients of a bank during the years 2010 to 2012 are studied. In this paper, the customers classification is done using the Kmeansclustering algorithm and then using a three-layer MLP neural network in MATLAB environment, customer analysis was performed (Alborzi, Khanbabaei, 2016).
- Evaluation of bank customers' value has been studied in another article in 2016. Information of 2124 clients is classified in this article using k-means algorithm into four clusters. And to build a prediction model the neural network was chosen after evaluating different models (Ali zadeh, Minaei-Bidgoli, 2016).

Inside the country, different articles and studies are provided in this field, some of which are as follows:

- In the fourth International Conference on Electronic Banking and Payment Systems 2014, an article titled "Applications & Challenges of Data Mining in the Banking Industry" was presented by Dr. Babak Teymorpour. In this article segmentation and identification of bank customers using data mining techniques and important behavioral factors, such as variables of money, frequency and recent transactions of customers so-called RFM analysis model has been studied (Teimourpour, 2014).
- Customer clustering of Refah Bank branches with combining genetic algorithms and thirty means in the fuzzy environment was performed in 2015. Researchers found that the combined algorithm they used is more accurate than Kmeans algorithm in clustering customers (Ghorbanpour et al., 2015).
- 94 In another article examines three approaches of fuzzy clustering, RFM and neural networks in determining customer loyalty in customer-centric environments (Naghimalekshahi, 2015).
- Babaei in his dissertation in 2016 assessed customers of a credit and financial institution. In this thesis, he evaluated customers of the institution with different approaches (Babaei, 1395).

In order to identify valuable customers, similar work has already been done by the author. The models differ in the type of categorizing the customers as training and testing data. 30-70 coefficient and optimal meta-Cost is considered for valuable customers 3.9 versus 1 for ordinary customers. In order to improve former model, 10 fold cross validation is used in the new model to classify the training and testing data. Also by changing the meta-cost to value of 3.5, higher accuracy than the previous model was achieved and in the present article the new method is described.

Given previous research and approaches described in this papers the result can be achieved that using data mining techniques to extract data lies in the large database of the Banks is a good assistance in understanding customers and lower risk customers, and how to identify profitable customers (Yaning and Kai and Fei.G, 2008).

Meanwhile, the value of customer life cycle based on the concept of the discounted sum of net profit for the same customer in its life will be very effective in determining the nature of the customer in terms of profitability (Fang, Ma, 2009). While maintaining data confidentiality in banking is the main challenge in this area.

Method of Research

The overall research model in is shown Figure 1.

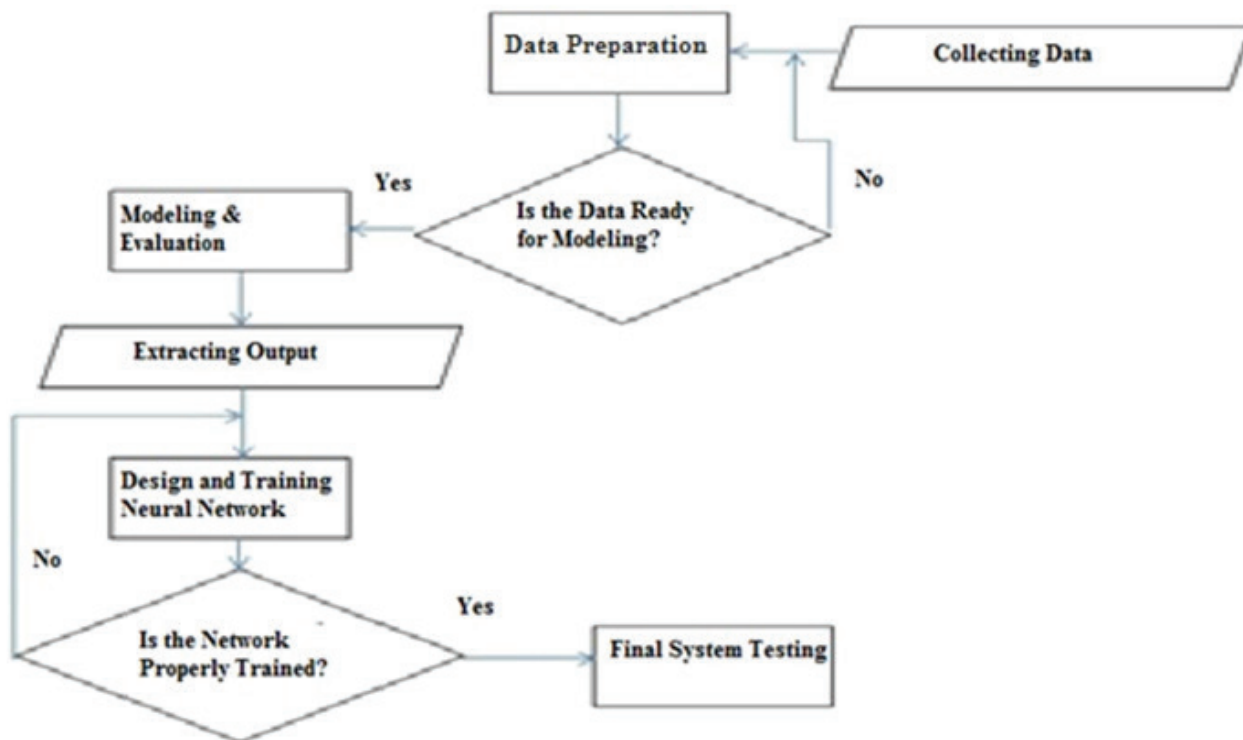


Figure 1- Overview of Proposed Model

Collecting Data

The population used in this study is the transactions of a sample bank with daily capacity of 10 to 12 million transactions that shall be processed. If the amount of unsuccessful transactions for any reason, including lack of balance as well as non-financial transactions such as domestic balancing and bills and so on will be deducted from that amount due to lack of necessity for issuing financial records, we will face an amount of 7 to 10 million documents and registered transaction in this database. In addition, in this research the data of 2015 of the customers in sample bank during high transaction days of Bank including the starting days and ending days of the month (25th of each month to end of month) in which the subsidy (AKA Yaraneh) and salaries are deposited is extracted for all of the customers of bank excluding the customers who have checkbook and offline accounts. The reason for selection of this data set is to extract the actual behavior of bank customers using real data.

Data Preparation

In the first step of this process, the extracted data from the

database of bank is simulated and recorded in real environment. Then the data quality was studied and it was found that no missing value exists in the data. So in this respect there was no need for specific action. In the second step each record extracted from the Bank must be changed in a way that is representative of one customer transaction. For this purpose, with used Java programming language to design and implement the data to correct the records in accordance with the specific circumstances of each transaction. Then, as one of the parameters used for identification of the clients include the number of transactions in the determined period of time, so it is logical that the accounts numbers that are only introduced in bank for the specified purposes and don't represent the client, i.e. internal accounts of the bank, are excluded from the collection of transactions. The reason for such decision is justified from data mining perspective. This type of account numbers are outlier numbers, therefore, they shall be removed. So the third step of the preparation phase was implemented in Java programming language, and the account numbers with nature of certain accounts were removed from the transactions.

The fourth step of the collection of data is its preparation

steps before importing them into RapidMiner software is converting the solar dates to Gregorian dates. Since the RapidMiner software uses the Gregorian dates to calculate the differences of dates, and the software is unable to recognize solar dates, and also due to the fact that in time of RFM calculation for each customer the date interval of the last transaction of customer shall be calculated, therefore, all the dates shall be converted to Gregorian dates. Again, small software that was designed and implemented by the Java programming language had converted the dates and recorded them in a separate field. The fifth step of the data preparation phases was calculation of each customer's RFM. In this line, the data were categorized based on account numbers and for each account number some factors were calculated including Total Number of Customer's Transactions in a Year that is the equivalent of Frequency Variable (F) and the Total Amount of Customer's Transactions in a Year that is the equivalent of Monetary Variable (M) and the Last Date of the Customer's Transaction that is equivalent with the Recent Transaction Variable (R).

Modeling

At this point fully refined data was used for modeling. In the modeling the fuzzy clustering and neural network models have been used simultaneously. Then, the customer's transaction data that was categorized as RFM parameter in the previous steps was segmented using clustering models. Given that in the banking 20-80 law always prevails, by definition that only 20 percent of bank customers provide 80 percent of resources; thus, the output of fuzzy clustering simply proves the accuracy of this law regarding bank's clients. In order to solve the problem of imbalance, sampling solution was used. In this solution instead of using one hundred percent of the members of a particular cluster only a certain percentage can be used. In addition, since the cost of leaving the bank by a valuable customer is more than the cost of regular customers, so the model should be set in such a manner that when predicting customers in neural network, valuable customers will be predicted more accurately. Because, if due to a mistake a valuable customer is recognized as a regular customer, the concentration of bank on this customer will decrease and therefore, the valuable customer may leave the bank and this will be really harmful for the bank. On the other hand, if the model by an error identifies a regular customer as one profitable customer, it will cause the bank less harm than the above mentioned mistake. Because, in this error, the customer is a regular customer but the bank makes an error and considers him as one valuable customer, this error will cause a lot of concentration by bank on this particular customer, while there is a possibility for this person to leave the bank. By the way, erroneous iden-

tification of valuable customer would be more damaging for the bank than ordinary customer and the valuable customers shall be detected with higher accuracy. With regard to the expression of Meta Cost, valuable customers will be predicted more accurately than the ordinary customer because the valuable customers have more costs than normal customers.

In the next step of a series of measures related to the modeling, neural network is used to identify the customers. The main advantage of a neural network is its capability of growth or restructuring (the number of neurons and their connections) that causes better adaptation of neural network with the issue. On issues such as data mining the designer of network is unable to guess the structure of data, therefore, this ability of grow is very important.

Evaluation

Various methods were examined in the evaluation of model; including dividing the data into two parts: training and testing data that is divided into two sets including 70 percent training and 30 percent testing. In this evaluation method, the accuracy of 92.97 percent accuracy on the training data and 88.61 percent in testing data was obtained.

Alternatively Model 10 Fold Cross Validation is used to evaluate the model. In this procedure data is divided into 10 subsets, and each time a set of data is used for validation and the rest is used for training and finally, the mean of results of these ten evaluations is selected as a final estimate. The method uses Shuffled Sampling and the samples are selected without classifying as a random sampling.

The results of the evaluation by 10 Fold Cross Validation was examined by confusion matrix and results are shown in Table 1.

Table 1 – Results of Model

Accuracy	Recall	Precision
95.46	97.14	84.31

4) Conclusion

Managing the relation with customers in banking system can be used as a leverage to gain the customers, gain their loyalty and maintaining them for a long term and also using the customer's advantages. With increasing competition, new customer acquisition costs also increased, so recognizing the bank's current customer and ability to categorize them into profitable and etc. will help a lot in keeping potential sources of Bank. Based on the results gained from the proposed model in this study it can be concluded that using data mining for extracting the knowledge hidden in high level of data of banks is the best tool for recognizing the customers. On the oth-

er hand, according to the fact that the transactional data of the customers is growing data, so fuzzy clustering in this paper showed successful results in the segmentation of customers and its combination with the results of a neural network model provided a proper model for the recognition of banking customers. Therefore, it can be said that using the proposed model of this research in banking area will have a significant effect on grow of banking capital. In the future by expanding the scope of this study to other sectors, in addition to the identification and classification of customers, transactions can be used in other issues such as money laundering and fraud detection.

Research Limitations

The obstacles of this study are as follows:

- Very high volume of customer data: since the data collected had a very high volume, calculation of each customer's RFM consumed a lot of time for refining the data and calculating the RFM of each customer.
- Requirement for powerful hardware: Due to the high volume of transactions in order to process customer data, powerful hardware infrastructure, including CPU, RAM, and storage capacity of the hard disk was needed.
- Lack of accurate customer data record: failure to register complete data about individual customers in the past as well as non-registration of accurate data by users in the branches in time of opening accounts, caused a situation in which the use of data about each customer besides their RFM model to increase the accuracy for determining the valuable customers will be impossible.
- Confidentiality of customer transaction: the banks, including the Bank reviewed in the study are not willing to provide information for data mining researches due to data privacy.

Recommendations for Future Works

In order to develop the results of this research in future, below subjects can be considered:

- In addition to RFM model to examine customer behavior, other parameters such as level of education, age, and amount of granted facilities, etc. can be used.
- Considering a specific weight for each of the RFM parameters can be useful in clustering customers. For example, in a bank the high value of amounts of transactions of customers may have more weight compared to other properties in the calculation of the profitability of customers. Therefore, the weight of M parameter shall be considered to be higher based on the opinion of experts and banking managers in clustering the customers.
- Parameters such as date of opening the account can be used to determine the duration of customer interaction with bank in addition to other parameters. The duration indicates sustainability of customer interaction. So adding

this parameter to RFM parameters may help in identifying profitable customers that are more stable and have more interaction with the bank.

References

- 1) Alborzi, Mahmood; Khanbabaei, Mohammad; (2016) "Using data mining and neural networks techniques to propose a new hybrid customer behavior analysis and credit scoring model in banking services based on a developed RFM analysis method", *Int. J. Business Information Systems*, Vol. 23, No. 1, pp.1-22.
- 2) Alizadeh, Hossein; Minaei-Bidgoli, Behrouz; (2016), "Introducing A Hybrid Data Mining Model to Evaluate Customer Loyalty", *Engineering, Technology & Applied Science Research*, Vol. 6, No. 6, pp.1235-1240.
- 3) Babaei, M., (2016), "Assessment of Bank and Credit Institutions Customers Using Data Mining Techniques (Case Study:Kosar Credit Institution)", a master's thesis in IT Management Field, Islamic Azad University, Electronic Branch, School of Management.
- 4) Fang, Bin; Ma, Shoufeng; (2009), "Data Mining Technology and its Application In CRM of Commercial Banks", *First International Workshop on Database Technology and Applications*, pp.243-246.
- 5) Ghorbanpour, A., Talaei, Gh., Panahi, M. (2015), "Clustering of Customers in Refah Bank Branches by Combining Genetic Algorithms and 30 Means in Fuzzy Environment", *Human Resource Management Research*, Volume 5, Number 3, pp. 153-168.
- 6) Naghmalekshahi, Sh.; Haroonabadi, A., Mirabedini, S.J. (2015), "Studying the Three Approaches for Fuzzy Clustering, RFM and Neural Networks for Determining Customer Loyalty in the Customer-Centric Environments", *National Conference on Computer Engineering and Information Technology Management*, Vol. 2.
- 7) Patil, Abhay; Mane, Utkarsha; Kachare, Arati; Mane, Madhura; (2015), "CRM Application using Fuzzy Clustering Algorithm", *International Journal for Scientific Research & Development*, Vol. 3, Issue 02, PP 571-573.
- 8) Teimourpour, B. (2014), "Applications and Challenges of Data Mining in the Banking Industry", the Fourth Annual Conference on Electronic Banking and Payment Systems.
- 9) Yaning, LI; Kai, LI; Fei, GAO; (2008), "Data mining approach and application in CRM project for internet-focused banking", the International Conference on Wireless Communications, Networking and Mobile Computing.

10) Yuliari, Ni putuputri; Darma Putra, I ketutgede; DwirusjaYanti, Nikadek; (2015), "customer segmentation through fuzzy C-means and fuzzy RFM method", Journal of Theoretical and Applied Information Technology, Vol.78. No.3, PP 380-385.