

**The Journal of
Mathematics and Computer Science**

Available online at

<http://www.TJMCS.com>

The Journal of Mathematics and Computer Science Vol. 4 No.2 (2012) 172 - 181

Developing a model for measuring customer's loyalty and value with RFM technique and clustering algorithms

Razieh qiasi^{*1}, Malihe baqeri-Dehnavi², Behrooz Minaei-Bidgoli³, Golriz Amooee⁴

¹ Department of Information Technology, University of Qom, Qom, Iran,

razieghiasi@gmail.com

² Department of Information Technology, University of Qom, Qom, Iran,

Programming_bagheri@yahoo.com

³ Department of Computer Engineering, University of Science and Technology, Tehran, Iran,

minaeibi@cse.mcu.ed

⁴ Department of Information Technology, University of Qom, Qom, Iran,

Golriz.Amooee@gmail.com

Received: January 2012, Revised: April 2012

Online Publication: June 2012

ABSTRACT

In today's competitive world, moving toward customer-oriented markets with increased access to customer's transaction data, identifying loyal customers and estimating their lifetime value makes crucial. Since knowledge of customer value provides targeted data for personalized markets, implementing customer relationship management strategy helps organizations to identify and segment customers and create long-term relationships with them, and as a result, they can maximize customer lifetime value. Data mining techniques are known as a powerful tool for this purpose.

The purpose of this paper is customer segmentation using RFM technique and clustering algorithms based on customer's value, to specify loyal and profitable customers. We also used classification algorithms to obtain useful rules for implementing effective customer relationship management. This paper used a combination of behavioral and demographical characteristics of individuals to estimate loyalty. Finally, the proposed model has been implemented on a grocery store's data,

* Corresponding Author

during 1997 to 1998 in Singapore, to measure customer's loyalty during these two years.

Keywords: customer value, RFM model, K-means algorithm, customer relationship management, loyalty.

INTRODUCTION

Facing with more complexity and competition in today's business, firms need to develop innovation activities to capture customer needs and improve customer satisfaction and retention [1,2]. In this regard, Customer relationship management is a broadly recognized strategy to acquire and retain customers.

On the other hand, the goal of most companies is profitability growth. In order to reach this goal, companies should provide an analysis of how to manage relationship with their customers and offer appropriate corresponding marketing strategies. Bhatnaga and Ghose (2004) provide a new transaction model based on service and customer satisfaction and showed that price is not the only measure that affects customer purchasing decisions, but also it is important that customer and company agree on product value and good customer services [3]. Therefore, organizations should not seek to develop a product to satisfy their customers, but they should track customer purchase behavior and present distinct products for each segment. In other words, customer segmentation based on buying behavior is essential for developing successful marketing strategies, which in turn cause creating and maintaining competitive advantage.

Current methods of customer value analysis based on past customer behavior patterns or demographic variables, are limited for predicting future customer behavior. So today, analysts are trying to use better patterns for analyzing customer value. Nowadays, RFM method is one of the most common methods for segmenting and identifying customer value in the organization. This method use three variables to extract the behavioral characteristics of customers and segment customers based on three variables: Recency, Frequency and Monetary [4]. In this paper we used the results of this method as inputs for clustering algorithms to determine customer loyalty. Providing expected products of customers, improve customer relationship management and finally loyalty promotion are the results of such analyzes.

The rest of the paper is organized as follows: section 2 describes the needed concepts briefly. Section3 reviews related works on customer value. Section 4 expresses stages of model creation and applying data mining algorithms on the data, and last section presents conclusion and suggestions for improving the model in further studies.

BASIC CONCEPTS

Customer Relationship Management

Customer relationship management is a set of processes and technologies for managing current and potential customer relationship and company relationship using marketing, sales and service parts, regardless of the communication channel. The main objective of CRM is to make long-lasting and profitable relationships with customers [2,5]. Therefore, CRM makes three things: providing an integrated and unique view of customers using analytical tools, managing customer relationships in a single way

regardless of communication channel, and improving effectiveness and efficiency of the processes involved in customer relation [6].

Customer Value Analysis

Customer value analysis is a kind of analytic method for discovering customers' characteristics and makes a further analysis of specific customers to extract useful knowledge from large data. On the other hand for organizations, Retention cost is much less costly than acquisition cost. Therefore, they want to use techniques for identifying permanent customers with more lifetime value. Kotler (2000) defines Customer Lifetime Value as the profit net present value that one can obtain in a customer's lifetime. Kim et al. (2006) define customer lifetime value as the net income amount of the business during the entire life cycle of a customer. He emphasizes on long-term continued income and cost, instead of the profits from a specific trading activity [8]. Gledy et al. (2009) have presented an approach to predict customer lifetime value with Pareto/NBD model [9]. Mahboubeh Khajvand et al. are obtained customer lifetime value by weighting behavioral variables using analytic hierarchical process [2].

Loyal Customer

Loyal customer is the one who has a positive attitude towards the organization and is frequently associated with the organization. Creating a loyal customer is not only about maintaining numbers of customer overtime, but it is creating the continuous relationship with customers to encourage purchasing in the future. Different criteria affect customer loyalty in an organization. Some believe that the key factors that influence customer loyalty in organizations are: relationship quality, trust, satisfaction, purchase development, organizational changes and etc. Also, some evaluate customer loyalty with customer transaction behaviors, Including the recency, frequency and monetary which the maximum of these values results in the most loyalty [10].

RFM model

In most of the researches, two methods are common for identifying loyal customers, one of them is in terms of demographic variables (such as age, gender, etc.) and the other is in terms of interactive behaviors of customers that are expressed with the so-called RFM. RFM model is proposed by Hughes in 1994, and has been used in direct marketing for several decades. This model identifies customer behavior and represents customer behavior characteristics by three variables:

- Recency of the last purchase which refers to the interval between latest customer purchase and time analysis of customer data.
- Frequency of the purchases which refers to the number of transactions in a particular period.
- Monetary value of the purchase which refers to consumption money amount in a particular period.

RFM model can be used in different areas by different people; Therefore, RFM can mean different things to different people [11]. Classic RFM ranks each customer based on valuable against other customers and an RFM Score will be assigned to each customer.

Basically, the RFM model applies for dividing each variable to five levels, so that 20% of customers who makes the most purchase from the company are assigned the number of 5. The next 20% will be assigned as 4 and etc. finally, the database divide into 125 equal parts based on recency, frequency and monetary. Customers who have the most score are profitable.

Clustering

Clustering is the process of grouping a set of physical or abstract objects into similar groups. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [4,10,12]. The clustering algorithms which are used in this paper include K-Means and Two-step. K-means, which at first was known as Forgy method [4,11], is one of the well-known algorithms for clustering with top to down approach that desired data objects divide into k groups in terms of its special features and characters. Groups classified based on minimum sum of squares distance between object from center [13].

This method is dependent on the definition of initial centers, so this algorithm must be run with different centers, and a case is acceptable which has the lowest error rate in terms of Euclidean distance. On the other hand, this algorithm is very sensitive to noise in data. In contrast, Two-step algorithm doesn't need the exact number of clusters which was defined by user and it determines the optimum number of clusters in a range which is defined by user.

RELATED WORK

RFM model is identified for discovering customer value of direct sales [14, 15]. Keymak (2001) believes that RFM model is one of the known methods for analyzing customer value [16]. The most important characteristic of RFM is simplicity and speed of its implementation [17,11;16]. Moreover, from the viewpoint of consumer behavior, Schijns et al. believe that RFM Model is a famous method for measuring power of customer relationship [18]. John et al. compare RFM model, CHAID and regression methods for direct marketing and segmented customers with focus on transaction data and concluded that efficiency of RFM model as well as other methods [11].

Godman (1992) applied RFM model to find valuable customers [19]. In another research, Chao and Yang (2003) applied back-propagation neural network on transactions of medical equipment industry based on RFM model [20]. Lin and Tang (2006) also applied RFM model to analyze customers of music product. They classified customers with same value into a cluster. In this study, each variable of RFM model assigned with two levels: High and Low level, and finally eight clusters were established [21].

Lo et al. (2008) adopted RFM model to analyze members of a sports store. The results showed that higher expense customers are male and aged from 26 to 35 [22]. Huang, Chang, and Wu (2009) applied K-means method, Fuzzy C-means clustering method and bagged clustering algorithm to analyze customer value for a hunting store in Taiwan and finally concluded that bagged clustering algorithm outperforms the other two methods [23]. Cheng and Chen (2009) combined quantitative value of RFM attributes and K-means algorithm with rough set theory, to extract meaningful rules [4].

RESEARCH METHODOLOGY

In this section the purposed model to determine customer loyalty is described. This model is based on Data mining and customer value analysis techniques to improve customer relationship management based on RFM variables with Two-step and k-mean algorithms. The purpose of this paper is to determine the degree of customer loyalty for achieving maximum benefit and a win-win situation. Figure 1 shows the required steps for the proposed model which we will discuss further.

Understanding Data

In this paper, we used the data from the Mondrian chain stores, in Singapore in the years 1997 - 1998. Database’s type is SQL server and Contains 10281 records of customers. Customers have three types of credit cards: silver, bronze and gold.

This store does not have the Internet sales channel all these years and people have to transact face to face.

Mondrian stores data warehouse contain a wide variety of products, the demographic profile of their customers and transactional data. According to the goal of the paper, transactions’ and sales’ data have a great role towards the desired purpose.

Data Preprocessing

Preparation and data processing is the most important and time consuming parts of the data mining process. In this step, the data must be converted to a usable format for the RFM model. the next step is to select required fields from database for the RFM model and remove or replace blank values . Finally three fields that are used for our model contains the customer ID, date of purchase, and the Volume of purchase at this time. Using the time of purchase we can obtain frequency and recency of customers and the third field can be used to analyze the monetary.

Using RFM Model

The next step involves determining values and scoring RFM variables and using them as inputs of clustering algorithm. In this paper, three variables have equal weights (1:1:1). Output of the RFM model consists of three fields of each customer: frequency, recency and monetary. Each of these parts are assigned to a number from 1 to 5 (Table 1). Then we normalized the three obtained fields of RFM model using the min-max formula (Eq.1) for the input of the next step.

$$NV = \frac{(V - \max_A)}{(\max_A - \min_A)} \tag{1}$$

In this formula \min_A express minimum value of attribute A and \max_A is the maximum amount of it. After normalization, values of R, F, M are mapped in 0 to 1 interval, and are prepared to enter the next step.

Table 1: quantitative Values of RFM variables for Mondrian data set

<i>Customer ID</i>	<i>Recency(R)</i>	<i>Frequency(F)</i>	<i>Monetary(M)</i>
23	1	1	2
24	4	4	4
25	5	1	2

26	4	3	3
27	2	1	1
28	2	2	4
30	3	2	2
31	1	1	1
32	2	2	3
34	3	1	1
....

Clustering

At this step, the results of RFM model are used for customer segmentation. For clustering with the normalized variables R, F and M we used the K-means and Two-step algorithms and finally results can be compared. Two-step clustering algorithm is one of clustering models that offers the optimal number of clusters. By applying this model the optimum number of 4 is calculated for clusters. But in k-mean algorithm, the appropriate number of clusters is obtained by using the sum of squared errors (SSE), which is shown in the Eq.2.

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist^2(m_i, x) \tag{Eq.2}$$

In this Equation, k is the number of clusters, m_i is the cluster center and x is the record belonging to the cluster. To find the best number of clusters in a range of 2 to 11, Figure 1 is obtained using Eq.2. According to this Figure, the appropriate number of clusters is 5.

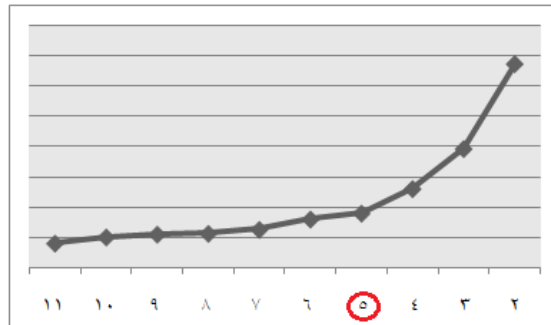


Figure1: Showing the number of appropriate clusters using the SSE equation. After clustering, we use Eq.3 [4] to determine the loyalty of each cluster.

$$C_1=(V_{11},V_{12},V_{13})$$

$$C_2=(V_{21},V_{22},V_{23})$$

$$\dots$$

$$C_5=(V_{51},V_{52},V_{53})$$

$$D_1=\sqrt{(V_{11}-0)^2+(V_{12}-0)^2+(V_{13})^2} \tag{Eq.3}$$

$$D_2 = \sqrt{(V_{21}-0)^2 + (V_{22}-0)^2 + (V_{23})^2}$$

...

$$D_5 = \sqrt{(V_{51}-0)^2 + (V_{52}-0)^2 + (V_{53})^2}$$

In this equation, V_{ij} is the average of RFM variables in each cluster and D_i represents the distance of each cluster's center from the Zero point.

Table 2: Analysis of results and classifying customers based on loyalty by using the Two-step method

Cluster name	NF	NM	NR	D_i	loyalty
Cluster1	0.534	0.507	0.041	0.737486	high
Cluster2	0.191	0.171	0.084	0.269774	low
Cluster3	0.042	0.047	0.171	0.182247	very low
Cluster4	0.012	0.025	0.679	0.679566	medium

Table 3: Analysis of results and classifying customers based on loyalty by using the k-means method

Cluster name	NF	NM	NR	D_i	loyalty
Cluster1	0.05	0.053	0.109	0.131111	Very low
Cluster2	0.55	0.523	0.041	0.760072	high
Cluster3	0.009	0.023	0.778	0.778392	Very high
Cluster4	0.212	0.189	0.071	0.292756	low

In Table (2 and 3) results of implementing both algorithms (k-means and Two-step) is summarized.

Generated Rules Using C5 Algorithm

After implementing the two clustering algorithms and labeling loyalty of each cluster, at this stage, we want to connect the demographic characteristics of customers to the results of clustering. With this aim that customers in the same cluster will have the same characteristics. Therefore in this stage, we apply the C5 model on the obtained clusters from k-means and Two-step algorithms, to produce rules for predicting the amount of loyalty based on demographic variables. An example of these obtained rules is as follows:

Applying C5 model on the results of k-means algorithm:

- 1) City = Sookz, Income = 130-150\$, Number of Car = 2 or less than 2 → very high loyalty
- 2) Member Card = Golden, City = Kamacho, Income = 110-120 \$ → high loyalty.
- 3) Member Card = Burns, City = Lemond, Education = high school, Marriage status = married → medium loyalty.

- 4) Gender = Female, City = Mexico, Income = 50-70 \$, Number of children at home = 0 → low loyalty.
- 5) City = Langly, Income = 10-30 \$, Number of children at home = Less than or equal to 1 →very low loyalty.

Applying C5 model on the results of Two-step algorithm:

- 1) City = Lebanon, Member Card = Burns, Number of children at home = Less than or equal to 1, job = worker, Landlords = No →low loyalty.
- 2) City = Salem, Income = 30-50 \$, Number of children at home = Less than or equal to 1, Education: Academic → low loyalty.
- 3) City = Bellflower, Gender = Male, Member Card = Burns, Number of children at home = Less than or equal to 1, Education = high school, Landlords = No → medium loyalty.
- 4) City = Camacho, Income = 110-130 \$→ high loyalty.
- 5) City = Langly, Income = 10-30 \$, Number of children at home = Less than or equal to 1 →very low loyalty.

The rules show that loyalty mostly depends on the financial situation and also the culture of different regions. Finally, the accuracy of rules generated by the algorithm C5 on results of the two algorithms, k-means and Two-Step, is presented in table4.

Table4: evaluation C5 model on the two cluster algorithms K-means and Two-step.

Cluster algorithm	<i>Two-step</i>	<i>k-mean</i>
Accuracy (%)	72	65

After analyzing and evaluation rules obtained from the C5, we conclude that Applying of C5 model on the clusters of k-means algorithm, is acting really weak in recognizing the medium and very high loyalty. And it's applying on the clusters of Two-step algorithm in recognizing the low loyalty has a very low accuracy.

CONCLUSION

The main purpose of this paper is customer segmentation and measuring their loyalty. First, behavioral variables, recency, frequency and monetary were obtained using RFM model, then customers were segmented by implementing two different algorithms, k-means and two-step, and the results were compared.

Since the acquisition costs of customer is much more than retention costs, determining customer loyalty has an important role for organizations. In this paper, customers of Mondrian store are classified and ranked based on their transactional behaviors. Then relationship between loyalty and demographic characteristic were obtained using the C5 model.

In this paper, our three variables are with the same weight and none of them has superiority over another. Therefore, we recommend future researches to analyze customer value, but weight the variables using experts, depending on the case study, and

compare the results. In addition to the three variables mentioned above, variety of products that customers bought can be also considered as an appropriate parameter for clustering.

Finally, it should be noted that it is much better for the stores within the country, to try to determine and retain profitable and loyal customers using data mining techniques, rather than paying a heavy cost for attracting new customers.

REFERENCE

- [1] H.M. Chuang and C.C. Shen, A study on the application of data mining techniques to enhance customer lifetime value based on the department store industry, *the seventh international conference on machine learning and cybernetics*, PP.168-173.
- [2] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh ."Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study, (2011) *Procedia Computer Science*, Vol.3, pp.57–63.
- [3] A. Bhatnagar, and S. Ghose, A latent class segmentation analysis of e-shoppers, (2004). *Journal of Business Research*, Vol.57(7), pp.758–767.
- [4] C. H.Cheng, & Y. S. Chen, Classifying the segmentation of customer value via RFM model and RS theory, (2009), *Expert Systems with Applications*, Vol.36(3), pp.4176–4184.
- [5] V. Ravi, *Advances in Banking Technology and Management: Impacts of ICT and CRM*, (2008), Information science reference, Hershey, New York, Yurchak Printing Inc.
- [6] D. Injazz, P.Karen, Understanding customer relationship management (CRM), People, process and technology. (2004), Available from: <<http://www.emeraldinsight.com/1463-7154.htm>>.
- [7] P. Kotler, *Marketing management*, (2000), New York: Prentice Hall International, Inc.
- [8] S.Y. Kim, T.S. Jung, E.H. Suh, & H.S. Hwang, Customer segmentation and strategy development based on customer lifetime value: A case study, (2006), *Expert Systems with Applications*, Vol.31(1), pp.101–107.
- [9] N.Glady, B. Baesens, & C. Croux, A modified Pareto/NBD approach for predicting customer lifetime value, (2009), *Expert Systems with Applications*, Vol.36(2 (Part1)), pp.2062–207.
- [10] S.M. Seyed Hosseini, A. Maleki, M.R. Gholamian, Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, (2010), *Expert Systems with Applications*, Vol.37, pp.259–5264.
- [11] J.A. McCarty, M. Hastak. Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression, (2007), *Journal of Business Research*, Vol.60, pp. 656–662.
- [12] J. Han, & M. Kamber, *Data mining: Concepts and techniques*, San Francisco: dwMorgan Kaufmann Publishers. (2001).
- [13] R.S. Wu, P.H. Chou, Customer segmentation of multiple category data in e-commerce using a soft-clustering approach, (2010), *Electronic Commerce Research and Applications*.
- [14] M. L. Roberts, Expanding the role of the direct marketing database, (1992), *Journal of Direct Marketing*, Vol.6(2), pp.51–60.

- [15] W.Y. Chiang, To mine association rules of customer values via a data mining procedure with improved model: An empirical case study, (2011) *Expert Systems with Applications*, Vol.38, pp.1716–1722.
- [16] U. Kaymak, Fuzzy target selection using RFM variables, (2001), *In IFSA World congress and 20th NAFIPS international conference*, Vol. 2, pp. 1038–1043.
- [17] R. Kahan, Using database marketing techniques to enhance your one-to-one marketing initiatives, (1998), *Journal of Consumer Marketing*, Vol.15 , pp.491–493.
- [18] J. M. C. Schijns, & G. J. Schroder, Segment selection by relationship strength, (1996), *Journal of Direct Marketing*, Vol.10, pp.69–79.
- [19] J. Goodman, Leveraging the customer database to your competitive advantage, (1992), *Journal of Direct Marketing*, Vol.55(8), pp.26–27.
- [20] C. M. Chao, C. T. Yang, Applying data mining on prediction of customer value, (2003), *In Conference of an approach to analyze and implement commercial database, Taipei, Taiwan*, pp. 159–183.
- [21] C.S. Lin, & Y.Q. Tang, Application of incremental mining and customer's value analysis to collaborative music recommendations, (2006), *Journal of Information, Technology and Society*, Vol.6(1), pp.1–26.
- [22] C. F. Lo, H. H. Wu, E. C. Chang, & Y. Y. Cheng, Applying data mining to an outfitter's customer loyalty, and value analysis, (2008), *Journal of Quality*, Vol.15(4), pp.293–30.
- [23] S. C. Huang, E. C. Chang, & H. H. Wu, A case study of applying data mining techniques in an outfitter's customer value analysis, (2009), *Expert System with Applications*, Vol.36(6), pp.5909–5915.