Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 3 (2011) 1327–1332

Procedia Computer Science

www.elsevier.com/locate/procedia

WCIT 2010

Estimating customer future value of different customer segments based on adapted RFM model in retail banking context

Mahboubeh Khajvand[a], Mohammad Jafar Tarokh[b] *

[a] Post graduate student, IT Group - Faculty of Industrial Engineering K. N. Toosi University of Technology, Tehran, Iran
[b] Associate professor, IT Group - Faculty of Industrial Engineering K. N. Toosi University of Technology, Tehran, Iran

Abstract

One of the important challenges in customer-based organizations is customer cognition, understanding difference between them, and ranking them. Customer need-based segmentation was common in past years, but recently customer value as a quantifiable parameter could be used for customer segmentation. In this regard, customer segmentation based on customer lifetime value (CLV) and estimating the value of each segment would be useful for making decision in marketing and customer relationship management (CRM) program which can be adapted with the characteristics of each segment. Customer future value as a part of customer lifetime value can be estimated based on customer segmentation. In this regard, this study provide a framework for estimating customer future value based on adapted weighted RFM analysis which is a CLV calculating model, for each segment of customer in retail banking scope.

Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: Customer lifetime value; Customer segmentation; Data mining; RFM analysis; Time series; ARIMA model

1. Introduction

CLV is a concept in customer relationship management scope which was defined more than 30 years ago by Kotler as “the present value of the future profit stream expected over a given time horizon of transacting with the customer” [1]. There are several model for estimating CLV in past researches which were considered by Gupta et al. [2] and Jain and Singh [3], but in retail banking environment, a model to determine CLV should satisfy some conditions that two important of them are [4]: First, in order to be easily implementable, it should focus on the valuation of homogeneous segments of customers instead of individual clients. Second, it needs to be easy to understand and parsimonious in nature to ensure its applicability in many business contexts. This specifically implies limiting data requirements to the information available in an average bank’s information system.

In this regard, in current study customer segmentation methods were exploited to group customers in to homogeneous segments and for satisfying last condition, one of the most powerful and simplest models to implement CRM and estimating customer value which is RFM model – Recency, Frequency, and Monetary value [5, 6,7] was used.

Current study provides a framework for segmenting customers in homogenous segments, calculating CLV value of different segments and finally estimating customer future value of each segment with analyzing trend of customer value in different seasons by time series method. The rest of this study is organized as follows. Section 2 describes

* Mahboubeh Khajvand. Tel.: +98-21-8406 3356; fax: +98-21-8867 4858.
E-mail address: mkhajvand@sina.kntu.ac.ir.

1877-0509
doi:10.1016/j.procs.2011.01.011

the framework, and section 3 explains a case study and empirical results. Finally, section 4 draws conclusions and summarizing the contributions of this work.

2. Framework of estimating future value of different customer segments

Current study provides a framework for segmenting customers in homogenous segments, calculating CLV value of each segment and finally estimating future value of each segment with analyzing trend of customer value in different seasons by time series method. Fig. 1 represents the framework.

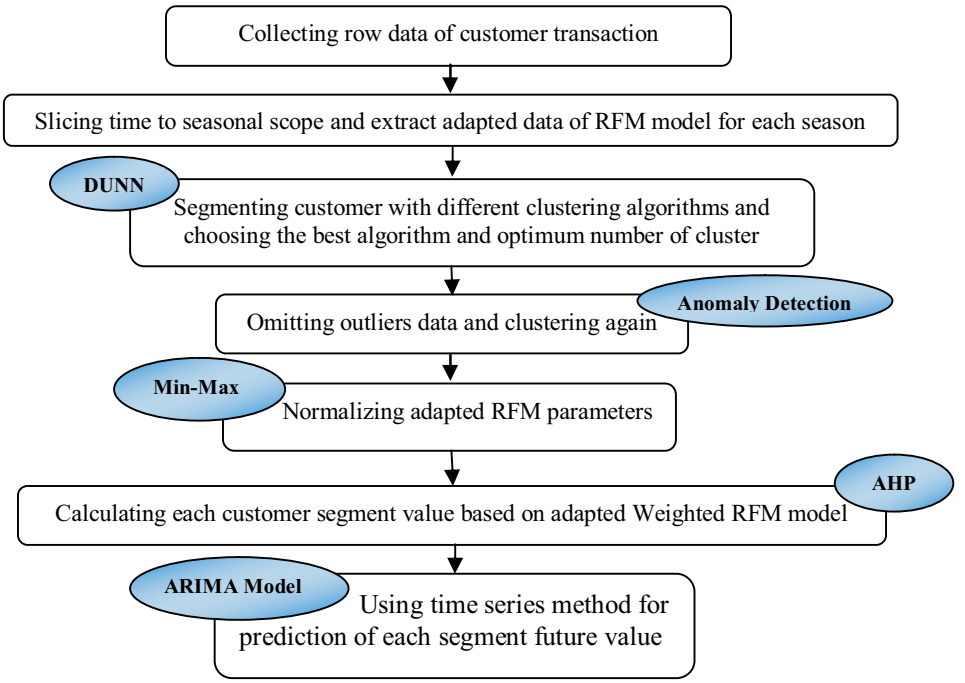


Fig.1. Framework of Customer Segmentation Based on CLV for Estimating Customer Future Value

3. Empirical results

This study as it is shown in Fig.1 is summarized in seven steps:

Step 1: In the first step, data was collected from an Iranian private bank in retail banking scope. The data was row data of fifty hundreds customers' transactions from the first of the spring 2008 to the end of the summer 2009 which was conclude customer ID, type of deposit, serial of customer deposit, branch code, transaction date, the amount of balance before transaction and the amount of transaction.

Step 2: The data which was received in step one, divided to six parts based on seasonal division. Then RFM parameters were extracted for each customer. Table 1 shows definition of basic RFM model parameters and adapted RFM parameters in this study.

Table 1. Definition of RFM and Adapted RFM model parameters

Parameter	Definition	Adapted parameter
Recency (R)	last purchase date in a particular period	Interval between the time of the last transaction and first day of each season
Frequency (F)	Number of purchases in a particular period	Number of days which occur a transaction during each season
Monetary (M)	value of purchases in a particular period	Daily average amount of money in all of the customer's deposits during each season

Step 3: In this step, customer segmentation was done with different clustering algorithms and k-means algorithm was chosen as the best algorithm by Dunn index. This index introduces optimum number of clusters in a clustering algorithm and also the best algorithm in segmenting data. The main goal of this measure is to maximize intercluster distances (distance between different clusters), whilst minimizing intracluster distances (distance between members of a cluster) [8]. For any partition $C = \{C_1, C_2, \ldots, C_k\}$, where C_i represents the i^{th} cluster of such partition, the Dunn index, D, is defined as in equation (1):

$$D(C) = \frac{\min_{i,j=1,\ldots,k;\ i\neq j}\delta(C_i,C_j)}{\max_{i=1,\ldots,k}\Delta(C_i)} \qquad (1)$$

Where $\delta(C_i,C_j)$ defines the distance between clusters C_i, C_j (intercluster distance), and $\Delta(C_i)$ represents the intracluster distance of cluster C_i or the size of the cluster C_i, and k is the total number of clusters. In this study:

$$\delta(C_i,C_j) = d(\bar{C}_i,\bar{C}_j) \qquad (2)$$

$$\Delta(C_i) = \max_{x\in C_i}\{d(x,\bar{C}_i)\} \qquad (3)$$

where $\bar{C}_i$ and $\bar{C}_j$ are the centroid of cluster C_i and C_j. Thus large Value of D corresponds to good clusters. Therefore, the number of cluster that maximizes D is taken as the optimal number of clusters, also the algorithm which has the maximize D is the best clustering algorithm.

The clustering algorithms which are used in this study include k-means [9], Two step algorithm and x-means algorithm [10]. Two step algorithm and x-means algorithm don't need the exact number of clusters which was defined by user while they find the optimum number of cluster in a range which defined by user. The defined range in this study for both of this algorithm was 3 to 10. Both of algorithms choosed 4 clusters as optimum number. But k-means algorithm was executed 8 times and Dunn index was calculated in each time, because this algorithm need . The optimum number of cluster (K) was obtained 4(See Fig. 2).

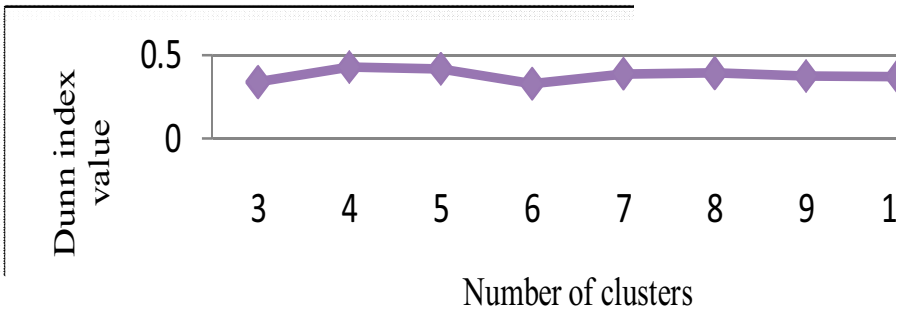


Fig. 2. Optimum number of clusters

Table 2. Result of Dunn index in clustering algorithms

	k-means	Tow step	x-means
Dunn index value	0.427	0.334	0.254

Step 4: K-means algorithm is sensitive about outlier data, so anomaly detection method was used to solve this problem and then the data was clustered again in each season. For example, the result of customer segmentation in summer 2009 by k-means algorithm after anomaly detection is shown in table 3.

Table 3. Clustering customer in summer of 2009

Cluster	Percent of customer	Recency (R)	Frequency (F)	Monetary (M)
c1	44	86.5	9.8	49498711
c2	9	24.1	2.1	3921154
c3	11	91.1	38.3	118829402
c4	36	64.2	4.4	24341737

Step 5: For calculating CLV value of each segment, normalized form of adapted RFM parameters were needed. Thus data was normalized by Min-Max normalization method, as it is shown in equation (4) (Tan et al., 2006):

$$v' = \frac{v - min_A}{max_A - min_A}(newmax_A - newmin_A) + newmin_A \qquad (4)$$

where min_A and max_A are the minimum and maximum values of an attribute, A. Then min-max normalization maps a value, v, of A to v' in the range of $[newmin_A, newmax_A]$.

Step 6: After normalizing RFM parameters, we calculated CLV score of each cluster based on weighted RFM model as follow in equation (5):

$$CLV_{ci} = NR_{ci} \times W_R + NF_{ci} \times W_F + NM_{ci} \times W_M \qquad (5)$$

where NR_{ci} refers to normalized Recency of cluster ci, W_R is the weight of Recency, NF_{ci} is normalized Frequency, W_F is the weight of Frequency, NM_{ci} is normalized Monetary, and W_M is the weight of monetary.

According to the assessments obtained by the AHP method based on expert people idea, the relative weights of the RFM variables are as follow: W_R= 0.081, W_F=0.317 and W_M=0.602. In table 4 the normal R, F and M values for each cluster's centroid and CLV of them is shown for the summer of 2009.

Table 4. Calculating CLV value of each cluster in summer of 2009

Cluster	Percent of customer	Normalized R	Normalized F	Normalized M	CLV Value	CLV Rank
c1	44	0.929	0.103	0.015	0.117	CLV2
c2	9	0.251	0.012	0.001	0.025	CLV4
c3	11	0.979	0.434	0.036	0.238	CLV1
c4	36	0.687	0.039	0.007	0.072	CLV3

In this regard, customers in cluster C3 which are 11% of all customers, with the most CLV value are the best and gold customer. In the other side, customers in cluster C2 with the least CLV value are the most invaluable customers. The rank of each segment of customers also can be seen in table 4.

Step 7: After clustering customers in each season in previous step, in this step we want to predict future CLV value of each segment based on six recent seasons. The trends of CLV value in different seasons in each segment are shown in table 5 and Fig. 3.

Table 5. CLV value of different segments during six seasons

CLV Rank	Spring 2008	Summer 2008	Fall 2008	Winter 2008	Spring 2009	Summer 2009
CLV1	0.163	0.248	0.286	0.225	0.274	0.258
CLV2	0.087	0.120	0.132	0.113	0.131	0.126
CLV3	0.047	0.077	0.083	0.075	0.081	0.078
CLV4	0.015	0.020	0.022	0.025	0.025	0.027

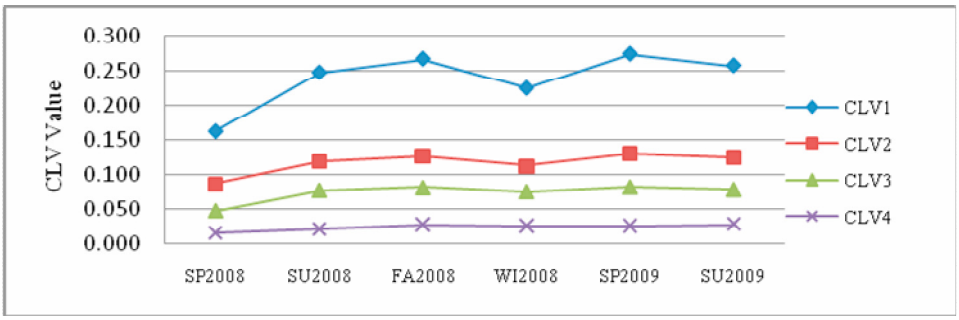


Fig. 3. Trend of CLV value in different segment during six seasons

For estimating customer future value, the multiplicative seasonal ARIMA -Auto Regressive Integrated Moving Average- which is a time series prediction method in nonstationary status and reflect the seasonal behaviour of the process, was exploited. The multiplicative seasonal ARIMA(p,d,q)×(P,D,Q)$_S$ model where p is the order of auto regressive process, d is the order of differencing operator, q is the order of moving average process, also P is the order of seasonal auto regressive process, D is the order of seasonal differencing operator, Q is the order of seasonal moving average process, can be represented by:

$$\phi(B^s)\varphi(B)\nabla_s^D\nabla^d x_t = \Theta(B^s)\theta(B)\varepsilon_t \qquad (6)$$

Where $\varphi(B)$ is auto regressive process, $\theta(B)$ is moving average process and ∇^d is d-fold differencing operator which is used to change a nonstationary time series to a stationary one. $\phi(B^s)$ is seasonal auto regressive process, $\Theta(B^s)$ is the seasonal moving average process and ∇_s^D is D-fold differencing operator. The ARIMA equation which was obtain for segments CLV1, CLV2 andCLV3 is ARIMA(0,0,1)×(1,1,1)×1, thus based on equation (6):

$$(1-\phi_1 B)\nabla_s x_t = (1-\Theta_1 B)(1-\theta_1 B)\varepsilon_t \qquad (7)$$

$$(1-\phi_1 B)(x_t - x_{t-1}) = \varepsilon_t - \theta_1 B\varepsilon_t - \Theta_1 B\varepsilon_t + \Theta_1\theta_1 B^2\varepsilon_t \quad (8)$$

$$x_t = (1+\phi_1)x_{t-1} - \phi_1 x_{t-2} + \varepsilon_t - (\Theta_1+\Theta_1)\varepsilon_{t-1} + \Theta_1\Theta_1\varepsilon_{t-1} \quad (9)$$

The ARIMA equation which was obtain for segments CLV4 is ARIMA(1,0,1)×(0,0,2)×2, thus based on equation (6):

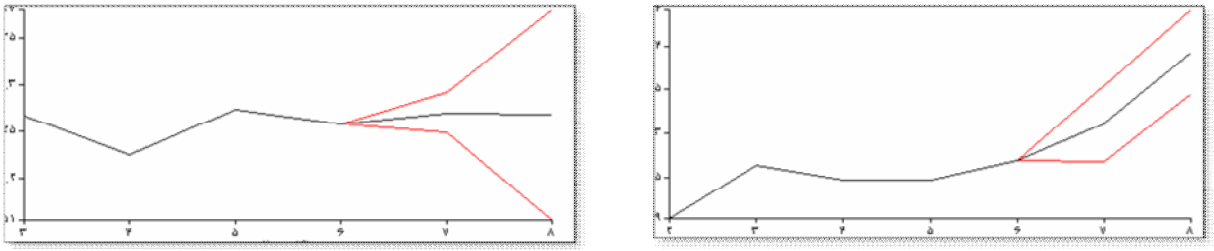
$$\phi(B^s)\varphi(B)\nabla_s^D\nabla^d x_t = \Theta(B^s)\theta(B)\varepsilon_t \qquad (10)$$

$$(1-\varphi_1 B^1)\nabla_s^0\nabla^0 x_t = (1-\Theta_1 B^2 - \Theta_2 B^4)(1-\theta_1 B^1)\varepsilon_t \quad (11)$$

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-2} + \Theta_1\theta_1\varepsilon_{t-2} + \Theta_2\varepsilon_{t-4} + \Theta_2\theta_1\varepsilon_{t-5} \quad (12)$$

The outputs of software for forecasting of CLV value of each segment in next season are represented in Fig. 4. The Fig.4 show that the value of less valuable segment (CLV4) will increment in future. The CLV value of segments CLV1, CLV2 and CLV3 can be predicted just for next season because of lack of data which is the limitation of this study.

Fig4. (a) Forecasting diagram for segments with CLV1, CLV2, CLV3; (b) Forecasting diagram for segment with CLV4



(a) (b)

4. Conclusion

It is so important and vital for firms to know different types of customers to make decision more profitability. Knowing customers one by one in some business is not commodious, so using customer segmentation method could be useful in this situation. Customer segmentation based on their value or their lifetime value represents instead of customer need-based segmentation nowadays. Customer lifetime value includes calculating past and present value of the customers and predicting of future value of the customers. This study prepares a framework for segmenting customers based on their value for estimating future value of different segments of customer in an Iranian private bank in retail banking scope. The result of this study could be as a guideline for marketing strategies, developing and cross selling of new products for each group, and developing of private banking for the most valuable customer group.

Reference

1. Kotler, P. Marketing during periods of shortage. Journal of Marketing 38(3), 20–29, (1974).
2. Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., and Lin, N. Modeling Customer Life-Time Value. *Journal of Service Research*, 9, 2, 2006, 139-155.
3. Jain, D., and Singh, S. S. Customer Lifetime Value Research in Marketing: A Review and Future Directions. *Journal of Interactive Marketing, 16,* 2, 2002, 34-46.
4. Haenlein, M., Kaplan, A. M., and Beeser, A. J. A Model to Determine Customer Lifetime Value in a Retail Banking Context. *European Management Journal*, 3, 5(2007) 221-234.
5. Cheng Yeh, Yang, K. J., and Ting, T. M. Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications,* 36(2009)5866-5871
6. Liu, D. R., and Shih, Y. Y. Hybrid Approaches to Product Recommendation Based on Customer Lifetime Value and Purchase Preferences. *The Journal of Systems and Software, 77*, (2005) 181-191.
7. Hosseini, S.M.S., Maleki, and A. Gholamian, M.R. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(2010) 5259–5264.
8. Brun, M., Sima, C., Huaa, J., Loweya, J., Carroll, B., Suha, E., and Doughertya, E. Model-based evaluation of clustering validation measures. *Pattern Recognition,* 40(2007) 807 – 824.
9. Tan, P.-N., Steinbach, M., and Kumar, V. Introduction to data mining. *pearson education*, (2005).
10. Pelleg, D., Moore, A.W. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *In: Seventeenth International Conference on Machine Learning, (2000)727-734*.