2010 International Conference on Intelligent Systems, Modelling and Simulation

A Two Phase Clustering Method for Intelligent Customer Segmentation

Morteza Namvar
Department of Industrial Engineering
Iran University of Science and Technology
Tehran, Iran
info@mortezanamvar.com

Mohammad R. Gholamian
gholamian@iust.ac.ir

Sahand KhakAbi
sahand_khakabi@ind.iust.ac.ir

***Abstract*—Customer Segmentation is an increasingly significant issue in today's competitive commercial area. Many literatures have reviewed the application of data mining technology in customer segmentation, and achieved sound effectives. But in the most cases, it is performed using customer data from a special point of view, rather than from systematical method considering all stages of CRM. This paper, with the aid of data mining tools, constructs a new customer segmentation method based on RFM, demographic and LTV data. The new customer segmentation method consists of two phases. Firstly, with K-means clustering, customers are clustered into different segments regarding their RFM. Secondly, using demographic data, each cluster again is partitioned into new clusters. Finally, using LTV, a profile for customer is created. The method has been applied to a dataset from Iranian bank, which resulted in some useful management measures and suggestions.**

Keywords-customer relationship management; segmentation; data mining; clustering; Iran

I. INTRODUCTION

In the last decade customer relationship management has been concerned by many authors, and it has played critical role in the new business economy. Some potential benefits of CRM are as follows: (1) Increased customer retention and loyalty, (2) Higher customer profitability, (3) Value creation for the customer, (4) Customization of products and services, (5) Lower process, higher quality products and services [1], [2].

Consequently, customer segmentation, as the primary stage of CRM, is an increasingly pressing issue in today's over-competitive commercial area. "More and more literatures have researched the application of data mining technology in customer segmentation, and achieved sound effectives" [2], but most of them segment customer only by single point of view, rather than from systematical framework.

According to [3], [4], and [5], CRM consists of four dimensions: Customer identification, customer attraction, customer retention, and customer development. Based on this categorization, segmentation was placed in the first step of CRM, customer identification. While, for marketing researches, customer segmentation has high contributions to other steps of CRM and it should not be the end in itself, but rather a means to the end [6]. Formation of several segments is possible by using customer profitability. For example, loyalty program could be applied to the most profitable segments consisting of the highest-profit customers in order to retain them. Another possible strategy is reducing the amount of resources dedicated to unprofitable customer groups who generate more costs. "This segment is arguable since unprofitable customers seem to have no worthy of marketing efforts" [7].

Input variables used in the segmentation process determine the step of CRM we are dealing with. Demographic variables, RFM and LTV are the most common input variable used in the literature for clustering customers. However demographic variables deal with all the stages of CRM, their role in customer attraction is more significant. On the other hand, RFM and LTV are mostly used in the customer retention and development. In this study we aim at combining these three input variables in an innovative approach for customer segmentation using the well-known data mining clustering technique, K-means.

The remainder of this paper is organized as follows. Section 2 reviews the previous studies related to customer segmentation. This part presents the limitations of existing studies and explains the background reasons of this study. Section 3 proposes a clustering model for customer segmentation. Consequently, the results of applying the model in the case of banking industry will be presented. Finally, section 4 concludes the paper with some general discussions and an agenda for further research.

II. LITERATURE REVIEW

Although customer segmentation and market segmentation have been considered similarly in the literature, there are some critical differences regarding data availability for their clustering mechanisms. Market segmentation usually aims at acquiring new customers, and deals with the first step of CRM, customer acquisition,

978-0-7695-3973-7/10 $26.00 © 2010 IEEE
DOI 10.1109/ISMS.2010.48

IEEE computer society

Authorized licensed use limited to: Politecnico di Milano. Downloaded on March 07,2022 at 08:35:10 UTC from IEEE Xplore. Restrictions apply.

using socio-demographic data. While customer segmentation works at all steps of CRM using both socio-demographic and transactional data. "We can imagine that customer cultivation and retention are more important than customer acquisition, because lack of information on new customers makes it difficult to select target customers and this will cause inefficient marketing efforts" [8].

Chai & Chan [9] has classified existing customer segmentation methods into methodology-oriented and application-oriented approaches. Most methodology-driven studies modify some data clustering techniques such as SOM, or use a combination of two or more data mining techniques to achieve more accurate clusters or segments (such as [10], [6], [11], and [12]). "On the other hand, application-oriented researches must search for the optimum method for solving segmentation problems in specific applications" [9]. They usually define and create new variable for clustering procedure or use different variables in sequential clustering steps (Such as [8], [7], [13], [14], [9], [15], [16], [2], [17], [18], and [5]).

In the current literature of customer segmentation, LTV has an important role. For example, in [8] an LTV model considering past profit contribution, potential benefit, and defection probability of a customer for wireless telecommunication customers segmentation is suggested. In [7] a framework for analyzing customer value and customer segmentation based on their value is proposed. Then, the offered strategies according to customer segments are illustrated through a case study on a wireless telecommunication company.

Another major input for customer segmentation is RFM. In [13] a self-organizing map neural network is used to identify groups of customers based on repayment behavior and recency, frequency, and monetary behavioral scoring predictors. It also classified bank customers into three major profitable groups of customers. The resulting groups of customers were then profiled by customer's attributes determined by using an Apriori association rule inducer. Lately, in [16] RFM, CHAID, and logistic regression are investigated as analytical methods for direct marketing segmentation, using two different datasets. Finally, in [18] a new procedure, joining quantitative value of RFM attributes and K-means algorithm into rough set theory (RS theory), is proposed to extract meaning rules.

Besides, a combination of above mentioned input variables, also, has been utilized by researchers. For example, in [9] a novel approach that combines customer targeting and customer segmentation for campaign strategies is presented. This investigation identified customer behavior using a RFM model and then uses a LTV model to evaluate proposed segmented customers.

Some authors have used a combination of other different variables and measures to cluster customers. For instance, [17] aimed at providing an easy, efficient, and more practical alternative approach based on the customer satisfaction survey for the profitable customers segmentation. The authors have presented a multiagent-based system, called the survey-based profitable customers segmentation system that executes the customer satisfaction survey and conducts the mining of the customer satisfaction survey, socio-demographic and accounting database through the integrated uses of business intelligence tools such as DEA (Data Envelopment Analysis), Self-Organizing Map (SOM) neural network, and C4.5 for the profitable customers segmentation. In [14] an anticipation model for potential customers in purchasing behavior is proposed. Their model is inferred from past purchasing behavior of loyal customers and the web server log files of loyal and potential customers by means of clustering analysis and association rules analysis. In the same year, in [2] authors have focused on proposing a customer segmentation framework based on data mining and constructs a new customer segmentation method based on survival character. Their new customer segmentation method consists of two steps. Firstly, with K-means clustering arithmetic, customers are clustered into different segments by similar survival characters (i.e. churn trend). Secondly, each cluster's survival/hazard function is predicted by survival analyzing, then, the validity of clustering is tested and customer churn trend is identified.

In [15] integrating data mining and experiential marketing to segment online game customers is investigated. The results can help the firms to predict and understand the new consumer's purchase behavior. According to the authors, online game's manufacturers could draw up the different market strategies to increase the purchase for the new different attributes' consumers.

In [5] groups of retail customers, based on their perception of commitment to the retailer and the degree of use of its technological equipments, are determined and characterized.

In addition, as mentioned before, some authors have focused on the segmentation procedure from technical point of view. For example, in [10] a new methodology for cross-national market segmentation is developed. The authors have proposed a two-phase approach (TPA) integrating statistical and data mining methods. The first phase is conducted by a statistical method (MCFA: multi-group confirmatory factor analysis) to test the difference between national clustering factors. The second phase is conducted by a data mining method (a two level SOM) to develop the actual clusters within each nation. In [11] support vector clustering (SVC) for marketing segmentation is used. In [12], also, a novel clustering algorithm based on genetic algorithms (GAs) to effectively segment the online shopping market is proposed. Simultaneously, in [19] a novel market segmentation approach, namely the hierarchical self-organizing segmentation model (HSOS), for market segmentation of real world multimedia on demand in Taiwan is proposed.

Table I categorizes segmentation models proposed by different authors according to their input variables.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on March 07,2022 at 08:35:10 UTC from IEEE Xplore. Restrictions apply.

TABLE I. SUMMARIZATION OF INPUT VARIABLES USED IN SEGMENTATION MODELS

Input Variables Used	References
Demographic	[19], [1], [17]
RFM	[18]
LTV	[7]
Demographic + RFM	[13], [16]
Demographic + LTV	[8]
LTV + RFM	[9]
Demographic + RFM + LTV	-
Other	[14], [10], [15], [2], [11], [12], [21]

III. RESEARCH FRAMEWORK

We propose a research framework for customer segmentation, considering four stages of CRM. The framework is implemented in a case of banking industry. In order to segment the bank customers and develop marketing strategies, our research approach is categorized into two phases. Firstly, we performed a two-phase method to segment bank customers using RFM and demographic variables. RFM is used in the first stage as the input values of K-means clustering. Then the demographic data, based on age, education and occupation, is used to cluster each segment resulted from stage one. These three demographic parameters were selected after a variable selection procedure using SOM (see Fig. 1).

Finally, LTV is utilized to compare customer value in each cluster. In most previous studies, the quality of a segmentation methodology is measured based on within-segment and inter-segment heterogeneity [20]. Whilst, marketers are concerned with and interested in maximizing the net value of targeted customers, rather than caring about within-segment homogeneity or targeting rate [8, 9]. To solve the core problem that marketers face, LTV model could be applied to assess the fitness between targeted customer groups and marketing strategies, rather than measuring the within segment homogeneity [9]. Therefore, we used neural networks to calculate potential value of each cluster and consequently their LTV and, finally, their profile.

IV. RESULTS

A. Data and measurement

The Iranian bank companies aim at making decisions for customer satisfaction and avoiding their churn. In our case study, in collaboration with them, they provided a part of their data to predict customers up/cross selling opportunities and loyalty programs. Firstly, customer data were selected and filtered and some insignificant records (such as register without transaction records) were discarded. As a result, we reached to 38254 records about 491 customers from data warehouse; each record included 25 attributes.

Besides, after executing a SOM clustering on the entire variables, the importance of demographic variables on the clustering procedure was ranked. As a result, "education level", "occupation level" and "age" were determined as the most influencing variables and were selected for usage in the second stage of clustering.

In the next step, RFM values were defined and calculated for entire customers according to their transactions. The time window for the RFM calculation was set to six menthes earlier than the last existing transaction in the database.

B. Customer clustering using RFM and demographic variables

A two-phase clustering was utilized to segment customers. Firstly customers were clustered according to their RFM scores. As a result, three clusters were detected consisting of 190, 111, and 190 customers.

Secondly, each cluster was internally clustered according to three demographic variables; education level, occupation level and age. Therefore nine clusters were resulted consisting of 104, 52, 34, 118, 23, 49, 32, 42, and 37 customers.

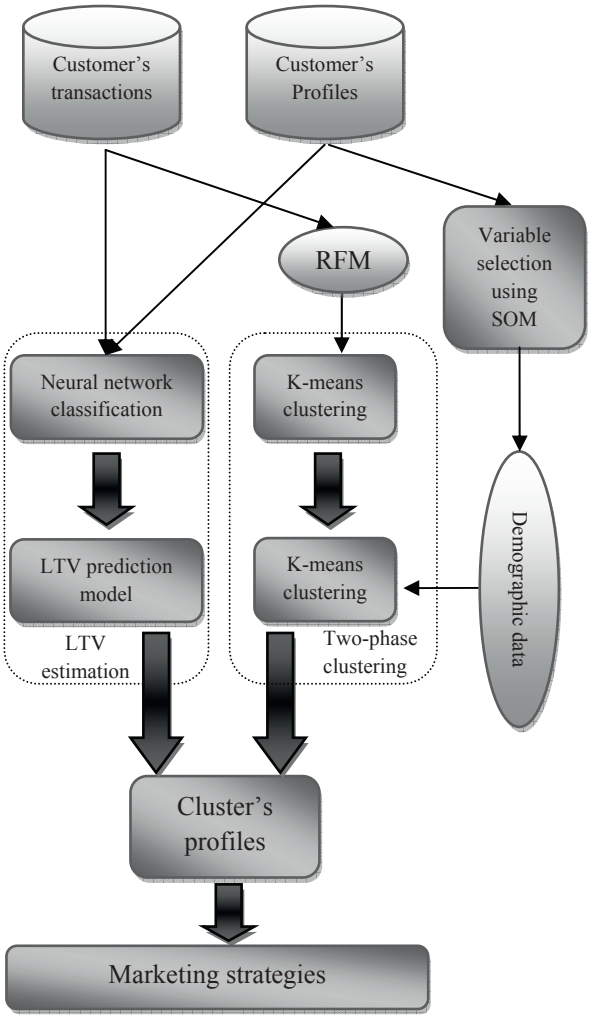

Figure 1. Research framework

C. Analyzing clusters using LTV

Authorized licensed use limited to: Politecnico di Milano. Downloaded on March 07,2022 at 08:35:10 UTC from IEEE Xplore. Restrictions apply.

In [8] a new LTV model considering churn rate of a customer is proposed. This model is shown in Eq. 1.

$$LTV_i = \underbrace{\sum_{t_i=0}^{N_i} \pi_p(t_i)\,(1+d)^{N_i-t_i}}_{\text{Past Profit Contribution}} + \underbrace{\sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i)+B(t_i)}{(1+d)^{t_i-N_i}}}_{\text{Expected future cash flow}} \quad (1)$$

t_i service period index of customer *i*
N_i total service period of customer *i*
d interest rate
$E(i)$ expected service period of customer *i*
$\pi_p(t_i)$ past profit contribution of customer *i* at period t_i
$\pi_f(t_i)$ future profit contribution of customer *i* at period t_i
$B(t_i)$ potential benefit from customer *i* at period t_i

Firstly, according to authors, in order to calculate customer's LTV, current value and potential value of customers should be calculated individually. Therefore, using the balance sheet of each customer in the last six months we calculated their current value. Furthermore, according to the authors, potential value of customers is predicted using Eq. 2.

$$Potential\ value_i = \sum_{j=1}^{n} Prob_{ij} \times Profit_{ij} \quad (2)$$

Where $Prob_{ij}$ is the probability that customer *i* would use the service *j* among *n*-optional services and $Profit_{ij}$ means the profit that a company can get from the customer *i* who uses the optional service *j*.

In order to evaluate potential value, firstly, by interviewing with managers, we found the kind of new services that could be proposed to their customers in the future and the return on each service. For example, in ATM, new services such as transferring money between accounts and bill payment were recommended by managers. Then they predicted an average profit which would be gained from these new services. Secondly the occurrence probability of these new services was predicted using neural network. A prediction model using neural network classification function was utilized to predict new customer's behavior regarding new recommended services. Neural networks are able to illustrate the confidence of model in predicting "yes" or "no" in the response of customer reaction in new service usage. We used this confidence to calculate the $Prob_{ij}$. Therefore, two essential items for customer's potential value prediction, $Prob_{ij}$ and $Profit_{ij}$ were prepared.

Finally, predicted LTV for each customer, were used to calculate mean value of LTV within each cluster. There were nine different clusters resulted from two-phase segmentation, and as a result nine different average LTV were obtained (See table II).

TABLE II. CLUSTER COMPARISON REGARDING DIFFRENT ATTRIBUTES

Cluster Number	1	2	3	4	5	6	7	8	9
RFM rank	6	4	8	7	5	9	3	2	1
Education rank	2	3	8	1	4	9	6	5	7
Occupation rank	1	4	7	2	3	9	5	6	8
Age rank	7	4	8	3	2	5	9	6	1
LTV rank	7	3	9	5	2	8	6	4	1
Largeness rank	1	3	7	2	4	9	5	6	8

D. Management applications

Finally, by analyzing on all kinds of the described features, profile of each cluster could be constructed. This profile is shown in Table II. The rank of each attribute for each cluster is illustrated in comparison with the others. For example, users in cluster 1 are placed in the sixth position regarding RFM. It means that 5 clusters are dominant regarding RFM to this cluster. Occupation level in this cluster is the best; therefore, average occupation level of customers in this cluster would be better than other clusters. Meanwhile, users in this cluster have a good position regarding education level (rank 2). This cluster is also the largest cluster in comparison with the others because of the number of customers. (See table II)

Based on the table which indicates clusters profiles, marketers would be able to make better decisions in order to improve marketing strategies within their organizations. For example, cluster 1 has the youngest users with the highest LTV, so it would be so worthy to investigate on.

VI. CONCLUSION

One of the key purposes of marketing is to identify the target customer portfolios and analyze it by segmentation and then set marketing strategies to each segment in order to reduce the risk of significant customer's defection. To reach this request, this paper proposes a new segmentation method based on data mining and most commonly used CRM models; RFM, LTV and demographic variables.

The method is developed on two-phase clustering model based on k-means technique. In application of the method on our case study (in banking industry), the existing customers were divided into nine groups of customers according to their shared transactional behavior and characteristics. Profiles of customers in each group could be analyzed by marketers to make strategies for each group.

Beyond simply understanding customer value in each cluster, the bank would gain the opportunities to establish better customer relationship management strategies, improve customer loyalty and revenue and find opportunities for up and cross selling. Further researches may aim at using larger data base with more fields to gain more accurate results from the model. Besides, lift charts can be used as a good evaluator on the performance of two-phase clustering model versus other clustering models. Finally, other prediction methods can be used instead of described neural network and their performance can be compared to reach the best selection.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on March 07,2022 at 08:35:10 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Jutla, D., Craig, J., & Bodorik, P., "Enabling and measuring electronic customer relationship management readiness", p*roceedings of the 34th annual hawaii international conference on system sciences organizational systems and technologies track* (pp. 1–10), 2001.

[2] Stone, M., Woodcock, N., & Wilson, M., "Managing the change from marketing planning to customer relationship management", *Long Range Planning*, 29, 675–683, 2006.

[3] Swift, R. S." Accelarating customer relationships: Using CRM and relationship technologies", *Upper saddle river* N.J.: Prentice Hall PTR, 2001.

[4] Parvatiyar, A., & Sheth, J. N., "Customer relationship management: Emerging practice, process, and discipline", *Journal of Economic & Social Research*, Vol.3, pp. 1–34, 2001.

[5] Kim, Yong Seog, Street, W. Nick, Russell, Gary J., & Menczer, Filippo, "Customer Targeting: A Neural Network Approach Guided by Genetic Algorithms", *Management Science*, 51(2), 264–276, 2005.

[6] Jonker, Jedid-Jah, Piersma, Nanda, & Poel, Dirk Van den, "Joint optimization of customer segmentation and marketing policy to maximize long-term profitability", Expert Systems with Applications, 27, 159–168, 2004.

[7] Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh, Hyun-Seok Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study" Expert systems with applications, Vol. 31, pp. 101-107, 2006.

[8] Hyunseok Hwang, Taesoo Jung, Euiho Suh , "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry" Expert systems with applications, Vol. 26, pp. 181-188, 2004.

[9] Chu Chai Henry Chan,"Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", Expert systems with applications, Vol. 34, pp. 2754-2762, 2008.

[10] Sang Chul Lee, Yung Ho Suh, Jae Kyeong Kim, Kyoung Jun Lee, "A cross-national market segmentation of online game industry using SOM", Expert systems with applications, Vol. 27, pp. 599-570, 2004.

[11] Jih-Jeng Huang, Gwo-Hshiung Tzeng, Chorng-Shyong Ong, "Marketing segmentation using support vector clustering", Expert systems with applications, Vol. 32, pp. 313–317, 2007.

[12] Kyoung-jae Kim, Hyunchul Ahn, "A recommender system using GA K-means clustering in an online shopping market", Expert systems with applications, Vol. 34, pp. 1200–1209, 2008.

[13] Nan-Chen Hsieh, "An integrated data mining and behavioral scoring model for analyzing bank customers", Expert systems with applications, Vol. 27, pp. 623-633, 2004.

[14] Horng-Jinh Chang, Lun-Ping Hung,*, Chia-Ling Ho, "An anticipation model of potential customers' purchasing behavior based on clustering analysis and association rules analysis" Expert systems with applications, Vol. 32, pp. 753-764, 2007.

[15] Jyh-Jian Sheu, Yan-Hua Su, Ko-Tsung Chu, "Segmenting online game customers – The perspective of experiential marketing", Expert systems with applications, Vol. 36, pp. 8487–8495, 2009.

[16] John A. McCarty, Manoj Hastak, "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression", Journal of Business Research, Vol. 60, pp. 656–662, 2007.

[17] Jang Hee Lee, Sang Chan Park "Intelligent profitable customers segmentation system based on business intelligence tools", Expert systems with applications, Vol. 29, pp. 1[2]–152, 2005.

[18] Ching-Hsue Cheng, You-Shyang Chen, "Classifying the segmentation of customer value via RFM model and RS theory", Expert systems with applications, Vol. 36, pp. 4176–4184, 2009.

[19] Chihli Hung, Chih-Fong Tsai, "Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand", Expert systems with applications, Vol. 34, pp. 780-787, 2008.

[20] Wedel, M., & Kamakura, W. A., "Market segmentation: Conceptual and methodological foundations" (2nd ed.). Dordrecht: Kluwer, 2000.

[21] Irene Gil-Saura and Maria-Eugenia Ruiz-Molina, "Customer segmentation based on commitment and ICT use", Industrial Management & Data Systems, Vol. 109 No. 2 pp. 206-223, 2009.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on March 07,2022 at 08:35:10 UTC from IEEE Xplore. Restrictions apply.