Title	A hybrid model of data mining and MCDM methods for estimating customer lifetime value
Authors(s)	Azadnia, Amir Hossein; Ghadimi, Pezhman; Molani-Aghdam, Mohammad
Publication date	2011-12-01
Publication information	41st International Conference on Computers and Industrial Engineering 2011
Conference details	The 41st International Conference on Computers and Industrial Engineering (CIE41), Los Angeles, United States of America, 23-26 2011
Link to online version	https://www.computers-and-ie.org/conferences/2017/1/15/41st-international-conference-1
Item record/more information	http://hdl.handle.net/10197/10167

A Hybrid Model of Data Mining and MCDM Methods for Estimating Customer Lifetime Value

Amir Hossein Azadnia[a,*], Pezhman Ghadimi[b], Mohammad Molani- Aghdam[a]

[a] *Department of Engineering, Ayatollah Amoli branch, Islamic Azad University, Amol, Iran*
[b]*Department of Manufacturing & Industrial Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia*

Abstract: Due to competitive environment, companies want to create a durable relationship with their customers. Building effective customer relationship management, companies should estimate customer lifetime value (CLV). CLV is normally calculated in terms of recency, frequency and monetary (RFM) variables. Allocating resource to customers' segments regarding to communication channel based on CLV can be helpful for managers .In this paper, a hybrid model for estimating CLV based on RFM variables incorporated with data mining and multi criteria decision making (MCDM) methods has been proposed. The proposed methodology contains three phases: data understanding and collection, data preprocessing, modeling and data analyzing in which Fuzzy Analytical Hierarchy Process (FAHP) has been used to determine RFM variables' weights. Then, K-means clustering method as a data mining method was employed in order to customer clustering and segmentation. Customer clusters were then ranked using MCDM method. Based on this ranking the company's marketing and advertising resources have been allocated to different clusters. Finally, the proficiency of the model was shown by conducting a case study of cosmetics industry

Keywords: Customer Lifetime Value, Data mining, Multi Criteria Decision Making

1. Introduction

In today's competitive era, customer relationship management (CRM) can be adopted as a core business strategy in order to help organizations manage customer interactions more effectively [1]. Value added services offered by the company and understanding the customers' needs verify the success or failure of companies [2]. Customer lifetime value (CLV) aims to develop strategies to target customers and to identify profitable customers [3]. Measuring Recency value (R), Frequency value (F) and Monetary value (M) is an important method for assessing CLV. Customer lifetime value evaluation has been discussed in various studies. Customer lifetime value evaluation has been discussed in various studies [4, 5]. Data of CLV could help companies to decide better about strategies development and resource allocation to each customer. So, customers should be clustered in term of their CLV. For example the communicating channel for advertising and product recommendation are should be different for different customers based on their clusters and segments. Because of large amount of customers' data, data mining techniques have been used in order to extract useful data. Using data mining methods integrated with CLV analysis have been practiced by several researchers. One of the applicable methods of data mining in CRM is clustering method which is used to group customers into segments. K-means, kohonen network/self organizing map, two step, and Fuzzy C-means are some of the modeling techniques for clustering [6, 7]. Because of large amount of customers' data, data mining techniques have been used in order to extract useful data. Using data mining methods integrated with CLV analysis have been practiced by several researchers. One of the applicable methods of data mining in CRM is clustering method which is used to group customers into segments. K-means, kohonen network/self organizing map, two step, and are some of the modeling techniques for clustering [6, 7]. In this research, K-means clustering has been used to segment customers based on RFM values because it is very common and simple. Weighing the attributes of RFM can improve the calculation of CLV [8]. For weighting RFM variables, Analytic Hierarchy Process (AHP) as one of the most applicable multi-criteria decision making

(MCDM) methods has been used by various researchers [9, 10]. Fuzzy AHP (FAHP) has been applied in this research in order to solve the problem of vagueness and ambiguity of decision makers' judgments.

In this paper, K-means clustering analysis was applied with the integration of Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and FAHP methods to customer segmentation and ranking. The proficiency of this hybrid methodology was then illustrated by conducting a case study in cosmetics industry.

2. Preliminaries

2.1 K-means Clustering Algorithm

K-means algorithm was developed by MacQueen [11] which aims to find the cluster centers, (c_1, . . . , c_K), in order to minimize the sum of the squared distances (Distortion, D) of each data point (x_i) to its nearest cluster centre (c_k), as shown in Equation below where d is some distance function. Typically, d is chosen as the Euclidean distance. The steps of K-means algorithm are shown as follows:

(1) Initialize K centre locations (c_1, . . . , c_K).
(2) Assign each x_i to its nearest cluster centre c_k.
(3) Update each cluster centre ck as the mean of all x_i that have been assigned as closest to it.
(4) Calculate $D = \sum_{i=1}^{n}[\min_{k=(1....k)} d(xi, ci)]^2$
(5) If the value of D has converged, then return (c_1, . . . , c_K); else go to Step 2.

2.2 TOPSIS

In this work, TOPSIS as a MCDM technique has been utilized to rank the clusters. TOPSIS was proposed by Hwang and Yoon [12]. The steps of TOPSIS are presented as follows:

1) Determining w_j of criteria (RFM), using FAHP.
2) Establishing the data matrix $[x_{ij}]$ which shows the final center of each cluster.
3) Normalization $R = [r_{ij}]_{m*n}$,

$$rij = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}{}^2}} \quad i = 1, \ldots, \text{m}; \; j = 1; \ldots; \text{n}$$

4) Establishing V matrix which is weighted matrix:
$V = [v_{ij}]_{m*n}$, $v_{ij} = [w_j r_{ij}]$

5) positive ideal point v_j^* and negative ideal point v_j^- for $j = 1,2,...,n$

6) Calculating the distance from point v_{ij} to positive ideal point v_j^* and negative ideal point v_j^- for $j = 1,2,...,n$ as follows:

$$d_i^+ = [\sum_{j=1}^{n}(v_{ij} - v_j^*)^2]^{1/2} \qquad d_i^- = [\sum_{j=1}^{n}(v_{ij} - v_j^-)^2]^{1/2}$$

7) calculate the relative closeness to the ideal solution.

$$CLi = \frac{d_i^-}{d_i^+ + d_i^-}$$

3. Methodology

The methodology used in this study combines two approaches to estimate, and evaluate CLV. These approaches are Clustering method and MCDM method. The concept of this methodology is that customers with similar values in RFM variables can be segmented based on their CLV. The methodology is illustrated in Figure 1. The proposed methodology in this study includes three phases as stated below:

Phase1- Data understanding and collection
Phase2- Data preprocessing
Phase3 - Modeling and data analyzing and strategies development

These phases are explained in detail as follows:

- Phase 1- Data were collected from the company data base. RFM scores of customers were considered as basic data of this research.
- Phase 2- Data preparation and data normalization were included in data preprocessing phase. In data preparation step, some noisy and uncompleted data have been omitted. Next, a normalization process is required to put the fields into comparable scales. This process is due to different scales of RFM inputs. In this paper, min-max approach was used which recalled all record values in the range between 0 to1.
- Phase 3- Weights of RFM variables were calculated by FAHP and the normalized data of RFM have been weighted by these weights. Then, based on weighted RFM values customers were clustered by K-means clustering. Finally, the clusters of customers were ranked using TOPSIS method. Then the strategies for resource allocation have been developed based on the clusters ranks.

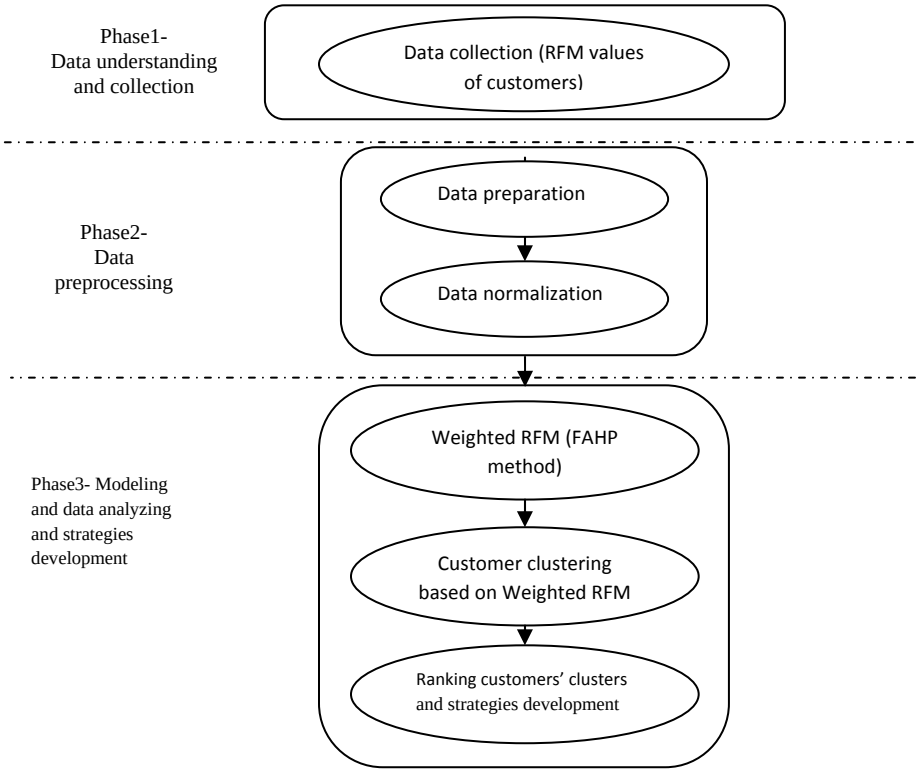

Figure 1. Proposed method

4. Case Study

In order to demonstrate the proposed method, a case study has been conducted in a cosmetics commercial company which is located in Iran. Board of directors of the case company had a plan to segment customers based on their CLV. Then, they wanted to recognize the high value segment in order to allocate their marketing and advertising resources to their customers and cluster their customers. According to the methodology, this study has been conducted as follows and the results of the study have been tabulated.

Phase1 - Data understanding and collection

Based on RFM scores of customers, 1132 customer records have been collected from the historical records of the company. RFM scores of customers are shown in Table 1.

Table 1. RFM scores of customers

Customer No.	Recency	Frequency	Monetary
1	305	18	46965730
2	363	3	2795880
.	.	.	
.	.	.	
.	.	.	
			10034620
1132	308	13	40615500

Phase2 - Data preprocessing

In this phase, a normalization process has been applied in order to convert data in similar scale ranging from zero to one (0 to1). The normalized data are shown in Table 2.

Table 2. RFM normalized data

Customer No.	Recency	Frequency	Monetary
1	0.802111	0.2727273	0.060254
2	0.955145	0.9545455	0.950566
	.	.	.
	.	.	.
	.	.	.
1131	0.300792	0.9545455	0.804658
1132	0.810026	0.5	0.188253

Phase3 - Modeling and data analyzing

In this phase, firstly, RFM variables were weighted by FAHP. Chang's FAHP [13] has been used in order to weight the variables. Three experts were asked to make pairwise comparison. Using pairwise comparison, FAHP has been calculated the weights of RFM variables. Table 3 shows the results of FAHP.

Table 3. RFM variables weights

	Recency	Frequency	Monetary
Weight	0.273056	0.296826	0.430118

Then, as shown in Table 4, normalized RFM scores of customers were multiplied in their corresponded weights. Next, weighted data were clustered by K-means clustering algorithm. SPSS 11.5 software has been utilized to perform K-means clustering. Customers were divided into five from the best to worst clusters and the final centers of clusters were calculated. The clustering results are shown in Table 5. Finally, TOPSIS method has been used to rank clusters based on their RFM final center which is shown in Table 6.

Table 4. RFM weighted normalized data

Customer no	Recency	Frequency	Monetary
1	0.219021	0.080952	0.025916
2	0.260808	0.283333	0.408856
	.	.	.
	.	.	.
	.	.	.
1131	0.082133	0.283333	0.346098
1132	0.221183	0.148413	0.080971

Table 5. Clustering results

Final Cluster Centers				
Cluster no.	Recency	Monetary	Frequency	Number of customers
1	0.109608435	0.151007444	0.064992773	339
2	0.182033947	0.36312595	0.070634954	136
3	0.077944344	0.354582789	0.21364836	188
4	0.20294865	0.18480728	0.20478898	353
5	0.073042616	0.075218273	0.236873816	116

Table 6. Results of TOPSIS

Cluster	*CLi*	Rank
1	0.17891	5
2	0.492909	4
3	0.661831	2
4	0.666229	1
5	0.500225	3

- *3-1Strategies development*

The company has three communication channels which are categorized based on their implementation costs from high to low that are stated as follows, respectively:

1. Verbal communications (Face-to-face, telephone)
2. Written communications (catalogues, proposals, e-mails, letters, and training manuals)
3. External communications (Media advertisement and web pages).

According to ranking of the clusters shown in Table 6, board of directors of the company can decide better about allocating each of this communication channels to the ranked clusters. Consequently, some strategic decisions can be made regarding to attain new customers, maintain current customers, and increase the satisfaction of high value customers.

5. Conclusion and discussion

Maintaining positive relationships with customers, increasing customer loyalty, and expanding customer lifetime value are the basis of Customer Relationship Management (CRM) approach [14-16]. In this research weighted RFM model integrated with K-means clustering and MCDM has been used to measure and rank customer lifetime value and allocate the advertisement and communication resources to clusters. With this proposed method, the company allocated its resource and communication channels such as verbal communications, written communications, external communications to customers' clusters based on their lifetime values ranking. This proposed model helps manager to allocate the resources to segments more precise. For the future works, mathematical model for resource allocation based on company costs and customers' values can be considered.

Reference

[1] Mishra, A. and Mishra, D. (2009). Customer Relationship Management: Implementation Process Perspective. *Acta Polytechnica Hungarica, 6*, 4, 83-99.
[2] King, S.F. and Burgess, T.F. (2008). Understanding success and failure in customer relationship management. *Industrial Marketing Management, 37*, 421-431.
[3] Irvin, S. (1994). Using lifetime value analysis for selecting new customers. Credit World 82, 3, 37-40.
[4] Hwang, H., Jung, Taesoo., Suh, E.(2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, *26*, *2*, 181-188.
[5] Blattberg, R, C., Malthouse,Edward,C., Neslin, S, A. (2009) Customer Lifetime Value: Empirical Generalizations and Some Conceptual Questions. *Journal of Interactive Marketing*, *23*, *2*, *157-168*
[6] Rad, A., Naderi, B. and Soltani, M. (2011). Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran. *Expert Systems with Applications*, *38*, 755-763.
[7] Mostafa, M. M. (2011). A psycho-cognitive segmentation of organ donors in Egypt using Kohonen's self-organizing maps. *Expert Systems with Applications*, *38*, 6906-6915.
[8] Stone, B. (1995). *Successful Direct Marketing Methods*. Lincolnwood: NTC Business Books, IL.
[9] Weiwen, X., Liang, C., Zhiyong, Z. and Zhuqiang, Q. (2008). RFM Value and Grey Relation Based Customer Segmentation Model in the Logistics Market Segmentation. *International Conference on Computer Science and Software Engineering.*
[10] Dan, Z. (2008). Integrating RFM model and Cluster for Students Loan Subsidy Valuation. *International Seminar on Business and Information Management.*
[11] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. 281-297.
[12] Hwang, C.L. and Yoon, K. (1981). *Multiple Attribute Decision Making Methods and Applications.* Berlin: Springer, Heidelberg.
[13] Chang, D-Y. (1996). Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, *95*, 649-655.
[14] Blattberg, R. C. and Deighton, J. (1996). Manage marketing by the customer equity test. *Harvard Business Review*, *74*(4), 136-144.
[15] Brassington, F. and Pettit, S. (2000). *Principles of Marketing (2nd edition)*. London: Prentice Hall.
[16] Ahn, Y.J., Kim, K.S. and Han, S.K. (2003). On the design concepts for CRM systems. *Industrial Management and Data Systems*, *103*(5), 324-331.