# Integration Model of Fuzzy C Means Clustering Algorithm and TOPSIS Method for Customer Lifetime Value Assessment

Amir Hossein Azadnia[1], Muhamad Zameri Mat Saman[1], Kuan Yew Wong[1], Abdul Rahman Hemdi[2]
[1]Department of Manufacturing & Industrial Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia
[2]Faculty of Mechanical Engineering, Universiti Teknologi Mara, Pulau Pinang, Malaysia
(zameri@fkm.utm.my)

**Nowadays, companies should establish a long-term relationship with their customers throughout customer relationship management (CRM). In order to be a winner in the market competition, marketing managers want to maximize customer lifetime value (CLV) and customer equity. So, creating a customer value assessment system is obligatory for companies to identify customers' value, develop strategies for customers' segments, and preserve the high value for them. Commonly, customer lifetime value is evaluated by RFM (recency, frequency and monetary) method. In this paper a model for customer value assessment integrated with multi-criteria decision making method and Fuzzy clustering method based on customer purchasing behavior was proposed. Fuzzy Analytical Hierarchy Process was utilized to calculate the weight of RFM variables. Then, based on the weighted RFM values, Fuzzy c-means clustering was used in order to cluster customers. Finally, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) has been employed to rank customer lifetime value. A case study was used to demonstrate the employment of the proposed model.**

*Keywords* – **Customer lifetime value, fuzzy c-means, fuzzy analytic hierarchy process, TOPSIS**

## I. INTRODUCTION

Encountering with complicated situations and contests in today's market environment, organizations need to adjust their structures and processes to capture the customers' needs for acquiring customers, increasing customers' values and retaining valuable customers [1, 2]. Companies want to recognize customers' true value and loyalty with the data gained from customers' transactions and activities by establishing appropriate customer relationship management (CRM). CRM assists companies to make long-term relationship with customers and formulates appropriate strategies for customer segments [1]. CRM is a strategy for establishing, managing and intensifying faithful and valuable customers [3]. This customer-based approach for marketing, tied with the large amount of customer-transaction data has highlighted the importance of estimating and understanding customer lifetime value (CLV) [2]. The concept of CLV presents future cash flow value produced from a customer in CRM [4]. CLV has been used in many researches and has been applied for a lot of purposes such as evaluating customers [5], customer segmentation [6], product recommendation [7], marketing and sales strategies [1].

CLV is normally used to recognize beneficial customers and formulate strategies for customers' segments. One of the most applicable methods for measuring CLV which has been used commonly is RFM (recency, frequency and monetary) method. RFM method has been proposed by Hughes in 1994 [8]. The definitions of RFM model were illustrated as follows [9]: (1) Recency (R) refers to the interval from the time when the previous purchase has been made by a customer; a lower value is a better value (cost criterion); (2) Frequency (F) is the total number of purchases that has been made within a specific period; a higher frequency value shows greater loyalty (benefit criterion); (3) Monetary(M) indicates the total amount of money spent during a certain period; a higher value indicates the customer is more profitable for a company (benefit criterion). RFM has been calculated without considering its variables' weights. In recent researches, researchers suggested weighted RFM. Stone [10] proposed that different weights have to be allocated to RFM variables according to different conditions, but the variables are weighted subjectively without any systematic model.

Some researchers used Analytic Hierarchy Process (AHP) as a systematic method for weighing the variables of RFM [11, 7]. AHP is one of the most applicable multi-criteria decision making (MCDM) methods for weighing criteria and alternatives. However, crisp value is unable to represent the inherent subjectivity and vagueness of the expert perception with an exact number. So, in this paper, Fuzzy AHP (FAHP) was used to solve this problem. Due to the huge amount of customers' information, data mining techniques are widely used to convert data into useful information and knowledge. Data mining techniques are used in CRM for market basket and sequence analysis, customer segmentation, and direct marketing management [3].

One of the most popular techniques of data mining in CRM which is widely used for grouping customers into some segments is clustering [12, 13, 14]. Clustering method can cluster customers with similar lifetime value (LTV). K-means, kohonen network/self organizing map, two step, and Fuzzy C-means are some of the modeling techniques for clustering [15, 16]. K-means clustering is a method frequently used to categorize data into groups [15]. This algorithm performs based on crisp partitioning which means each datum belongs to just one cluster. However, many objects have ambiguous characteristics and belong to more than one cluster with different fuzzy memberships. Thus, a method for soft partitioning is

978-1-4577-0739-1/11/$26.00 ©2011 IEEE

required [13]. For solving this problem, Fuzzy c-means (FCM) algorithm has been utilized which allows objects to belong to more than one cluster. In this research, FCM has been used to cluster customers with similar lifetime value.

The CLV ranking is one of the important issues for companies to develop appropriate strategies for retaining customers, identifying and comparing market segments. In spite of that, in previous researches for ranking CLV, integrated ratings were computed which indicated the sum of multiplication of normalized RFM values of customers in clusters and the corresponded weight for each criterion (Recency, Frequency and Monetary).

In this paper, TOPSIS method as one of the MCDM (Multi Criteria Decision Making) methods was used. The concept of TOPSIS is that the ideal solution in addition to having a maximum distance with the negative ideal solution should have minimum distance with the positive ideal solution. It permits the pursuit of best alternatives for each criterion depicted in a simple mathematical form. It can be used when there are both cost and benefit criteria. Among the RFM variables, recency is a cost criterion, but monetary and frequency are benefit criteria. Because of that, TOPSIS has been used in this research. The rest of this paper is organized as follows: Section II gives the backgrounds, definitions and methods which have been used. Section III presents the research methodology. Section IV illustrates a case study and the approaches which are used to rank CLV. Section V draws conclusions and discussion.

## II. BACKGROUND

### A. (FAHP) for Weighting RFM Variables

In this paper, Chang's [18] FAHP approach for weighting RFM variables, indicated by $W_r$ (weight of recency variable), $W_f$ (weight of frequency variable), and $W_m$ (weight of monetary variable) was applied. It is one of the Linear Weighting Models which works based on pairwise comparisons of human judgments. The biggest weight indicates the highest importance. Initially, Fuzzy AHP (FAHP) was introduced by Buckly in 1985 who generated fuzzy ratios for expressions of decision makers regarding to the pair wise comparisons. Table I shows the triangular fuzzy scale and linguistic variables.

TABLE I
TRIANGULARE FUZZY SCALE

| Triangular fuzzy scale | Linguistic scale |
|---|---|
| (1,1,1) | Just equal |
| (1/2,1,3/2) | Equally important |
| (1,3/2,2) | More important |
| (3/2,2,5/2) | Strongly more important |
| (2,5/2,3) | Very strongly more important |
| (5/2,3,7/2) | Absolutely more important |

### B. FCM for Customer Clustering

Fuzzy c-means (FCM) is a clustering method which proposed by Dunn in 1973 and further developed by Bezdek in 1981. The main aim of this method is minimization of the following objective function:

$$j_m = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^m \left\|x_i - c_j\right\|^2 \qquad 1 < m < \infty$$

Where, $N$ and $c$ are respectively the number of data and clusters. $x_i$ is the $i$th data, $m$ is any real number greater than 1, $c_j$ is the center of the $j^{th}$ cluster, $u_{ij}$ is the membership degree of $x_i$ belonging to the cluster $j$ and $\|*\|$ is the Euclidean vector norm expressing the distance between $j^{th}$ cluster's center and $i^{th}$ data. Fuzzy clustering is done throughout an iterative optimization of the $j_m$, with the update of $u_{ij}$ and $c_j$ by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \quad , \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

If $\|U^{k+1} - U^k\| < \delta$, then the iteration will be discontinued, where $\delta$ is a prescribed accuracy level between 0 and 1, while $k$ is the iteration step. This procedure converges to a local minimum or a saddle point of $j_m$. The algorithm steps are as follows:

*i)* *Initialize* $U = [u_{ij}]\ matrix, U^{(0)}$

*ii)* *At k-step: calculate the center vectors*
$$C^k = [c_j]\ with\ U^{(k)}, c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

*iii)* *Update* $U^{(k+1)}, U^{(k)}, u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$

*iv)* *If* $\left\|U^{(k+1)} - U^{(k)}\right\| < \delta$ *then stop; Otherwise return to step ii*

### C. TOPSIS for Ranking the Clusters

TOPSIS was initially presented by Hwang and Yoon [17]. It is to define the negative ideal solution (NIS) and the positive ideal solution (PIS) for ranking alternatives based on criteria. The PIS is the solution that maximizes the positive criteria (frequency, monetary) and minimizes the negative criteria (recency). The best alternative is the one, which has shortest distance to PIS and the farthest distance to NIS. TOPSIS will be used to rank the CLV based on the RFM scores in clusters.

The steps of TOPSIS are presented as follows:

1) Obtain the weights, $w_j$ of criteria (RFM), using AHP.
2) Establishing the data matrix $[x_{ij}]$ which shows the average score of each group of customers based on criteria (RFM).

3) Normalizing the data matrix $R = [r_{ij}]_{m*n}$ ,

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

4) Establishing weighted matrix; therefore the weighted matrix is configured to be:

$$V = [v_{ij}]_{m*n} \quad , v_{ij} = [w_j r_{ij}]$$

5) Determining the positive ideal point $v_j^*$ and negative ideal point $v_j^-$ for $j = 1,2,\ldots,n$

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

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

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

7) Compute the relative closeness to the ideal solution. (large is better)

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

## III. METHODOLOGY

As in Fig. 1, the proposed method mainly used RFM, Fuzzy AHP, Fuzzy c-means clustering and TOPSIS. The basis of the proposed approach is that the customers with similar values in RFM variables (recency, frequency, monetary) have more likely purchasing behavior and they can be segmented and clustered based on CLV. The attributes of RFM were weighted by FAHP. Based on these weights, the RFM scores have been determined for customers. Then, the Fuzzy C-means clustering method has been applied to cluster the customers in groups with similar RFM. In the next step, the clusters were ranked with TOPSIS which employed the RFM attributes weights and the average RFM scores of each cluster. To illustrate the model, a case study was used. The company produces electrical goods. Managers of the company wanted to establish a system to evaluate customer lifetime value, and develop proper marketing strategies to meet customer needs and acquire more customers. From the company database, the data sets on customers purchasing transactions were collected. Then, because of some noisy and uncompleted data sets, some of them were discarded. Finally, data sets based on RFM values for 806 customers based on their purchase transaction within 18 months have been extracted to calculate CLV. The RFM variables in this research are described as follows:

1. Recency: Period from last purchase (day)
2. Frequency: Count of purchase during 18 months
3. Monetary: Total amount of money spent during 18 months (US Dollar)

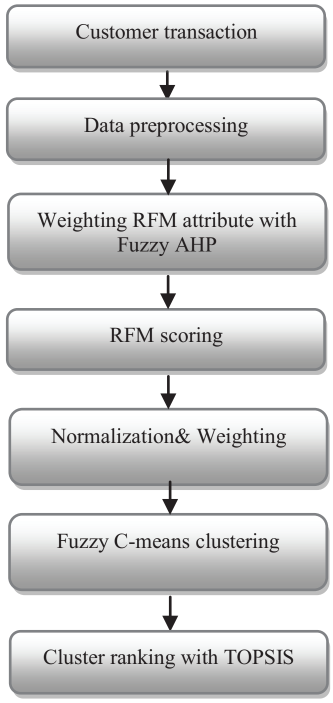


Fig. 1 Proposed method.

## IV. CASE STUDY

### *A. Calculating RFM Variables Weights*

In this phase, the weights of RFM variables were calculated using FAHP.

*Step1*: Pairwise comparisons. In this step, using the fuzzy scale shown in Table I, five experts as a group were asked to make pairwise comparison of the relative importance of RFM variables. The experts group consisted of director manager, sale manager, business manager, administrative manager, and logistic manager of the company. The results are shown in Table II.

*Step 2*: Computing the weights of variables. This calculates the weight of each decision element. This work employed Chang's [18] extent analysis to determine the weights of the RFM. The results are illustrated in Table III.

TABLE II
RFM FUZZY PAIRWISE COMPARISON

| | Recency | Frequency | Monetary |
|---|---|---|---|
| Recency | (1,1,1) | (2/5,1/2,2/3) | (1/2,2/3,1) |
| Frequency | (3/2,2,5/2) | (1,1,1) | (1,3/2,2) |
| Monetary | (1,3/2,2) | (1/2,2/3,1) | (1,1,1) |

TABLE III
RFM WEIGHTS

| Variable | Weight |
|---|---|
| Recency ($W_r$ ,) , | .0970 |
| Frequency ($W_f$) | .3446 |
| Monetary ($W_m$) | .5583 |

*B. Clustering the Customers with Similar RFM Score*

The steps for clustering the customers are described as follows:

*1. Data Normalization and Weighting*

Table IV shows the data matrix which indicates customers purchasing behavior in 18 months.RFM inputs are measured in different scales. Hence, a normalization process is required to put the fields into comparable scales and guarantees that fields with larger values don't determine the solution. In this paper, min-max approach was used which recalled all record values in the range 0-1.

For benefit criteria (monetary and frequency) is equal to (record value- min value of field)/(max value of field- min value of field) .

And for cost criterion (recency) is equal to (max value of field- record value)/ (max value of field- min value of field).

The normalized data have been weighted by $W_r$ , $W_F$ , $W_m$.. The results are shown in Table V.

*2. FCM Clustering*

Subsequently, customers with similarity regarding weighted RFM have been clustered in groups. For this FCM algorithm has been used. In fuzzy c-means algorithm, the number of clusters, (*c*), should be determined first. In this research, 8 clusters were selected because R, F, and M in each cluster can be higher/lower than their overall average. So there are 8 possible combinations (2*2*2).

Matlab 7.10 has been used for fuzzy c-means clustering. Table VI shows the results of customer clustering using Fuzzy c-means. It indicates 8 clusters, each with the related number of customers and their average R, F and M values. The last row in addition shows the total average of R, F, and M for all customers.

TABLE IV
DATA MATRIX

| Customer ID | Recency | Frequency | Monetary |
|---|---|---|---|
| 1 | 139 | 31 | 1979 |
| 2 | 18 | 53 | 770 |
| 3 | 167 | 18 | 4183 |
| 4 | 167 | 41 | 1339 |
| 5 | 68 | 23 | 5100 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 804 | 18 | 42 | 2641 |
| 805 | 96 | 43 | 3880 |
| 806 | 6 | 21 | 3103 |

*C. Estimating and Ranking CLV*

In this paper, TOPSIS was used to rank CLV. TOPSIS method has been selected because it considers costs and benefits criteria and doesn't need to use RFM normalized data, so the average of RFM variable scores can be used in each cluster. The data were obtained from Table VI. The results of TOPSIS ranking are shown in Table VII.

TABLE V
WEIGHTED NORMALIZED DATA MATRIX

| Customer ID | Recency | Frequency | Monetary |
|---|---|---|---|
| 1 | 0.029725 | 0.179503 | 0.090506 |
| 2 | 0.088688 | 0.337466 | 0.019957 |
| 3 | 0.016081 | 0.086162 | 0.219116 |
| 4 | 0.016081 | 0.251305 | 0.05316 |
| 5 | 0.064323 | 0.122062 | 0.272626 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 804 | 0.088688 | 0.258485 | 0.129136 |
| 805 | 0.050679 | 0.265665 | 0.201435 |
| 806 | 0.094535 | 0.107702 | 0.156095 |

TABLE VI
CLUSTERS CREATED BY FUZZY C-MEANS

| | Average in cluster | | | |
|---|---|---|---|---|
| Cluster no | Recency | Frequency | Monetary | Number of customers |
| 1 | 104 | 43 | 1394 | 94 |
| 2 | 108 | 17 | 4443 | 109 |
| 3 | 110 | 42 | 3504 | 115 |
| 4 | 102 | 45 | 6651 | 86 |
| 5 | 88 | 39 | 8490 | 101 |
| 6 | 103 | 16 | 8661 | 98 |
| 7 | 86 | 17 | 1870 | 108 |
| 8 | 107 | 20 | 6654 | 95 |
| Total average | 101 | 30 | 5208 | |

TABLE VII
TOPSIS RESULT

| Cluster number | $R_i$ | Rank |
|---|---|---|
| 1 | 0.293679 | 7 |
| 2 | 0.368973 | 6 |
| 3 | 0.405545 | 5 |
| 4 | 0.7537 | 2 |
| 5 | 0.917718 | 1 |
| 6 | 0.692357 | 3 |
| 7 | 0.06673 | 8 |
| 8 | 0.608084 | 4 |

As shown in Table VII, the customers' cluster ranking in term of their CLV is 5, 4,6,8,3,2,1,7.

## V.CONCLUSION

Customer segmentation is one of the important CLV applications in current studies. Evaluating and ranking customer lifetime value can help companies to make better segmentation based on customer's value. The RFM method as one of the applicable tools for measuring CLV has been used in this research. RFM is a very helpful device for companies in the process of customer segmentation. The supervision and assessment of the recency, frequency and monetary value of the purchases facilitate well-organized segmentation, market targeting and strategies development in customer relationship management. In order to cluster the huge amount of customers' data, FCM method as a data mining method has been utilized. The clustering has been done based on customers' RFM values similarity. TOPSIS method as a systematic approach has been used to rank customers based on their CLV. The main contributions of this study are summarized as follows:

1. A new model for customer clustering has been developed based on customer lifetime value.
2. A new model for CLV assessment has been developed
3. FAHP has been utilized for weighting RFM variables to solve the subjectivity and vagueness associated with the expert opinions.
4. Fuzzy c-means clustering method was used for customer clustering in terms of RFM values.
5. A new approach for ranking CLV has been suggested using TOPSIS.

This research can help managers to make better decision about their policy and strategy in each market segment. By considering the CLV ranks of customers in segmented groups, appropriate marketing strategies can be formulated. There should be a good opportunity for future research to develop strategies based on this segmentation and CLV ranking.

## REFERENCES


[1] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh., "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," *Procedia Computer Science,* vol. 3, pp. 57-63, 2011.

[2] I. Mahdavi, N. Cho, B. Shirazi, N. Sahebjamnia., "Designing evolving user profile in e-CRM with dynamic clustering of Web documents," *Data & Knowledge Engineering,* vol. 65, pp. 355-372, 2008.

[3] K.Tsiptsis and A. Chorianopoulos, *Data mining techniques in CRM inside customer segmentation.* West Sussex, United Kingdom: Wiley, pp. 1-62, 2009.

[4] T. Wiesel, B. Skiera, J. Villanueva., "Customer Lifetime Value and Customer Equity Models Using Company-reported Summary Data," *Journal of Interactive Marketing,* vol. 25, pp. 20-22, 2011.

[5] D. F. Benoit and D. Van den Poel, "Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services," *Expert Systems with Applications,* vol. 36, pp. 10475-10484, 2009.

[6] 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," *Expert Systems with Applications,* vol. 31, pp. 101-107, 2006.

[7] D.-R. Liu and Y.-Y. Shih., "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management,* vol. 42, pp. 387-400, 2005.

[8] A.M. Hughes, Strategic Database Marketing, Probus Publishing, Chicago, 1994.

[9] J.R. Bult, T.J. Wansbeek, Optimal selection for direct mail,Marketing Science 14 (4), 1995, pp. 378–394.

[10] B. Stone, Successful Direct Marketing Methods, Lincolnwood, NTC Business Books, IL, 1995, pp

[11] J. Liu and H. Du, "Study on airline customer value evaluation based on RFM model," vol. 4, ed, 2010, pp. V4-278.

[12] L.-j. Huang and X.-q. Gan, "Customers' Clustering Analysis and Corresponding Marketing Strategies based on Improved SOFM ANN in e-Supply Chain," *Systems Engineering - Theory & Practice,* vol. 27, pp. 49-55, 2007.

[13] Z. Zakaria, K. L. Lo, M. H. Sohod., "Application of Fuzzy Clustering to Determine Electricity Consumers' Load Profiles," ed, 2006, pp. 99-103.

[14] H. Jing-Yuan, M. Li-Qun, W. Tao., "Customer value management based on AHP and K-means clustering method," vol. 5, ed, pp. 2216-2221, 2010.

[15] A. Rad, B. Naderi, M. Soltani., "Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran," *Expert Systems with Applications,* vol. 38, pp. 755-763, 2011.

[16] M. M. Mostafa, "A psycho-cognitive segmentation of organ donors in Egypt using Kohonen's self-organizing maps," *Expert Systems with Applications,* vol. 38, pp. 6906-6915, 2011.

[17] C.L. Hwang and K. Yoon., Multiple Attribute Decision Making Methods and Applications, Springer, Berlin Heidelberg, 1981.

[18] D.-Y. Chang, "Applications of the extent analysis method on fuzzy AHP," *European Journal of Operational Research,* vol. 95, pp. 649-655, 1996.