International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

September 2023 | Issue 56 | Volume 15 | Number 3 | Pages 243-246

# COMBINATION OF DATA MINING TECHNIQUES FOR E-COMMERCE BASED ON CUSTOMER VALUE

**B. Ture Savadkoohi [1] J. Vazifeh Shoae [2]**

*1. Department of Computer and Electrical Engineering, Seraj Higher Education Institute, Tabriz, Iran*
*bita.turesavadkoohi@gmail.com*
*2. Department of Computer and Electrical Engineering, Islamic Azad University, Azarshahr, Iran*
*vazifehshoae@gmail.com*

**Abstract**- Nowadays, customer value and its implications have received increasing attention in new E-commerce paradigm and that is indicated as fundamental to Customer Relationship Management (CRM). On the other hand, data mining in brief is a collecting, processing, analyzing and converting raw data in order to solve E-commerce problems. For this aim, in this paper we apply three data mining techniques namely, fuzzy C-mean clustering, fuzzy analytical hierarchy process and optimized linear regression, to find the customer value properly in order to keep customers and maximize their potential benefit in the world of E-commerce. So that, such an approach causes to help promote customer value by enabling the utilization of data mining. Thus, managers can correctly decide for customers and they can apply specific strategies in order to provide customized services and products for achieving effective CRM.

**Keywords:** Customer Value, Fuzzy C-means, Fuzzy Analytical Hierarchy Process, Optimized Linear Regression.

## 1. INTRODUCTION

The key challenge of E-commerce is to reach customers value as an asset [1-3]. Successful implementation of this concept causes strong relationship in order to customer satisfaction and it leads to contribution of customers during specific periods. Thus, it is foundation of CRM and it causes the available E-commerce can suggest technical, service and social benefits. Data mining technology is an important tool for extracting useful knowledge from existing data to increase the customer value by offering marketing strategy [4-7]. Chen et al. [8] described domain-based data mining for predicting telecommunications customer payment behaviors. Khajvand and Tarokh [9] are segmented the customers based on their lifetime value. Laing [10] is analyzed customer value in automotive industry. Zakrzewska and Murlewski [11] proposed Density-Based Spatial Clustering (DBSCAN) and k-means for bank customer segmentation.

In this paper, we have applied three data mining algorithms such as Fuzzy C-Mean (FCM) clustering, fuzzy Analytical Hierarchy Process (AHP) and optimized linear regression in order to define customer value to develop applications of E-commerce. Thus, we have designed a questionnaire based on customer value and created data base according to the component of the questionnaires. Then, we have used FCM [12], fuzzy AHP [13-15] to determine the customer value. At the end, the predictor model is obtained by applying customer value component with linear regression that is optimized by genetic algorithm [17]. So that, the method of the current paper is intended to provide the customer value for E-commerce in order to interact flexibly with their customers and increase revenues. The article starts with description of the questionnaire that is designed in this paper. Next, in Section 3 an approach to determine customer value and evaluation are presented. At the end, conclusion is given.

## 2. DESIGNE OF QUESTIONNARIES

In the recent years, relational marketing is taken into account specially in the service industrial and manufacturing, thus the concept related to customer value has become more important [1-3]. Beerli, et al. [18] believe that customer value not only increases the value of the business, but also it allows the business to keep it costs low for attracting new customer. Thus, it causes in the competitive world of E-commerce the better product or service repurchases. Customer value creates an obligation in the customer to do business in E-commerce and purchases products and order services continuously. Although, that is a psychological process which includes the evaluation of different options based on different criteria. Moreover, that occurs when the organization provide the customer's need better than the competitors [19-20]. Also, that is the tendency of customers to choose a product or a company among other products to meet a specific need.

According to the definitions that are introduced above, the dimensions of customer value that is considered in questionnaires design in this paper include: functional, emotional, social, situational and cognitive [21-24]. In order to determine the reliability of the questionnaires Cronbach alpha coefficient of the set of questionnaires items that is calculated is more than 70% [25]. This coefficient shows the desired reliability of the questionnaires.

## 3. CUSTOMER VALUE IN E-COMMERCE

In today's highly competitive E-commerce, customer value has been vital among the marketing researcher. Customer value models suggest a good theoretical framework to be used for E-commerce to become customer driven to maintain its competitive advantages [1-3]. On the other hand, in the modern science research, data mining is a terminology for obtaining data for diagnostic or analysis goals that are faced in real applications [5]. The proposed method consists of the following six steps:

- Step 1: Design of questionnaires based on the dimensions of customer value.
- Step 2: Creation of data base according to questionnaire's components.
- Step 3: Clustering the database based on the score of customer value components.
- Step 4: Determining the priority of customers in each cluster.
- Step 5: Specifying customer value position by multiply value of clustering and Fuzzy AHP.
- Step 6: Giving the customer value component for obtaining a predictive model with optimized linear regression.

Fuzzy set theory is defined by Zadeh [26]. This theory has been widely applied in a variety of key areas such as E-commerce domain [27], pattern recognition [28] and medical diagnosis [29]. Assume, $p= \{p_1, p_2, \ldots, p_n\}$ be set of points, $u_k(p)$ be the degree of each point $p$ to be in $k$th cluster and m be the weight index. FCM clustering is suited for classifying collection of data by allowing the same data to be in cluster with different degree of membership [12]. The centroid of each cluster is defined by:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m p_j}{\sum_{j=1}^{n} u_{ij}^m} \quad m \in (1,\infty) \tag{1}$$

An objective function is minimized as follows:

$$J(U, c_1, ..., c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j}^{n} u_{ij}^m \left\| p_i - c_j \right\|^2 \tag{2}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{3}$$

The essential condition for that function is defined by:

$$\bar{J}(U, c_1, ..., c_c, \lambda_1, ..., \lambda_n) = J(U, c_1, ..., c_c) + \\ + \sum_{j=1}^{n} \lambda \left(\sum_{i=1}^{c} u_{ij} - 1\right) = \sum_{i=1}^{c}\sum_{j}^{n} u_{ij}^m d_{ij}^2 + \sum_{j=1}^{n} \lambda i \left(\sum_{i=1}^{c} u_{ij} - 1\right) \tag{4}$$

The following steps are repeated in order to specify the center of cluster and membership matrix:

- Step 1: Initialize matrix as follows:

$$U = \left[u_{ij}\right] \text{matrix}, U^{(0)} \tag{5}$$

- Step 2: Compute the center of vectors at $k$-step by:

$$C^{(k)} = \left[c_j\right] \tag{6}$$

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m p_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{7}$$

- Step 3: Update the $U^{(K)}$ and $U^{(k+1)}$ items as the following:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| p_i - v_j \right\|}{\left\| p_i - v_k \right\|}\right)^{\frac{2}{m-1}}} \tag{8}$$

- Step 4: If the following condition is established then stop the algorithm. Otherwise, go to the step 2.

$$\left\| U^{(k+1)} - U^{(k)} \right\| < \varepsilon \tag{9}$$

After classification of customers, Fuzzy AHP approach is applied in order to specify the performance of each customer within that cluster [13-15]. Suppose, SoI be Scale of Interest, FN be Fuzzy Number, LV be Linguistic Variable and MF be Membership Function, $\tilde{a}$ be the fuzzy comparison matrix's value, $\tilde{r}$ be the fuzzy geometric mean, $\tilde{w}$ be the fuzzy weight's value in each dimension, $BNFP$ be the Best Non Fuzzy Performance, $U_W$, $L_W$ and $M_W$ be Upper, Lower and Middle value of the fuzzy Weight, respectively.

In fuzzy AHP approach, for creating a fuzzy comparison matrix, the scale of linguistics is specified. So that, the scale is the Triangular Fuzzy Number (TFN) scale from one to nine. The Membership Function is shown in Table 1. The Fuzzy comparison matrix by applying TFN is defined as follows:

$$\tilde{A} = \begin{bmatrix} 1 & \cdots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \cdots & 1 \end{bmatrix} \tag{10}$$

Table 1. Membership Function [16]

| SOI | FN | LV | MF |
|---|---|---|---|
| 1 | 1 | Equivalent importance | (1, 1, 3) |
| 3 | 3 | Weakly Important | (1, 3, 5) |
| 5 | 5 | substantially more important | (3, 5, 7) |
| 7 | 7 | Very Strongly Important | (5, 7, 9) |
| 9 | 9 | Completely Important | (7, 9, 9) |

The fuzzy geometric mean is calculated by:

$$\tilde{r}_i = (\tilde{a}_{i1} \otimes \tilde{a}_{i2} \otimes ... \tilde{a}_{in})^{\frac{1}{n}} \tag{11}$$

For each dimension the weight of fuzzy is determined as follows:

$$\tilde{w}_i = \tilde{r}_1 \otimes (\tilde{r}_1 \oplus \tilde{r}_1 \oplus ... \oplus \tilde{r}_1)^{-1} \tag{12}$$

The $BNFP$ is specified by transforming crisp weight and applying Center of Area (COA) method by:

$$BNFP_{w_i} = \frac{\left[(U_{w_i} - L_{w_i}) + (M_{w_i} - L_{w_i})\right]}{3} + L_{w_i} \tag{13}$$

At the end, predictive model will obtain by applying optimized linear regression with genetic algorithm [17]. Let $A$ be estimated variables, $\beta_1$ to $\beta_n$ be the regression ratio, $X_1$ to $X_n$ be absolute variable, $Y_1$ to $Y_n$ be dependent variable, *pre* be the predicted value and *t* be the true value.

Linear regression is applied two or more independent variables in order to approximate the effect of the dependent variable as follow:

$$Y = A + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \quad (16)$$

So that, in the case of $Y \geq 0$, it indicates a positive relationship between variables, otherwise, it indicates the relationship between variables as a negative value.

The main steps of linear regression that is optimized with genetic algorithm are included:

- Step 1: Produce population of $n$ chromosome. So that, each chromosome consist of the gene and each gene or coefficient is equal to $\beta_1$ to $\beta_n$ in Equation (16).
- Step 2: Determine the coefficient of regression.
- Step 3: Define the fitness function $f(x)$ of each chromosome by applying the relation of:

$$pre - t \quad (17)$$

- Step 4: Determine appropriate information by operators such as "selection", "crossover" and "mutation".
- Step 5: Specify the values of the fitness. In the case of local search falls in the local optima, where value indicate proper or close to zero, go to Step 8. Otherwise go to next step.
- Step 6: Looking for fitness value of each chromosome in the population by utilization local search approach.
- Step 7: Update the value of previous step and choose the new particular population for those values for the next generation.
- Step 8: $gen = gen + 1$ (18)
- Step 9: When the following condition is established go to Step 3. Otherwise, the process finishes with the optimal result.

$$gen < \max Gen \quad (19)$$

The evaluation has been done with proposed method of paper with different number of clusters by Root Mean Square Error (*RMSE*) as follows [30]:

$$RMSE = \left[ N_s^{-1} \sum_{i=1}^{N_s} |e_i|^2 \right]^{\frac{1}{2}} \quad (20)$$

where, $N_s$ is the Number of sample and $e_i$ is the difference between the actual value and estimated value for sample *i*.

Comparison of *RMSE* between different numbers of clusters is presented in Table 2.

Table 2. Comparison of *RMSE* between different numbers of clusters

| Number of Clusters | *RMSE* |
|---|---|
| 4 | 2.51 |
| 15 | 1.33 |
| 30 | 0.93 |
| 45 | 0.765 |
| 60 | 0.66 |
| 75 | 0.59 |

## 6. CONCLUSIONS

One of the competitive advantages of the E-commerce is utilization of the customer value. On the other hand, data mining is becoming more and more important tool in E-commerce to convert the data into knowledgeable information. In this paper, the combination of three data mining techniques is applied. The RMSE of the proposed method of this paper with different numbers of clusters has been investigated. Our exploited method and study, has the advantage that it can establish a predictive model for E-commerce in order to generate a simple and efficient approach for providing information related to potential customers and specifying the value of available customers. Thus, from the perspective of Customer Relationship Management (CRM) such an approach can obtain good communication with the customer with good management in competitive world of E-commerce. The proposed method of this paper only considers generation of customer value, whereas in E-commerce era a fast estimation is important as well. Thus, this point should be considered as a limitation of the proposed method of this paper. As future work, parallel processing can be used for fast estimation.

## REFERENCES

[1] I. Lahteenmaki, S. Natti, S. Saraniemi, "Digitalization-Enabled Evolution of Customer Value Creation: An Executive View in Financial Services", Journal of Business Research, Vol. 146, No. 4, pp. 504-517, 2022.

[2] T. Rintamaki, H. Saarijarvi, "An Integrative Framework for Managing Customer Value Propositions", Journal of Business Research, Vol. 134, No. 5, pp. 754-764, 2021.

[3] Y. Shu, "Research on Customer Perceived Value Evaluation of New Chinese-Style Clothing Based on PSO-BP Neural Network", Journal of Scientific Programming, Vol. 2022, Article ID 9273429, 2022.

[4] B. Ture Savadkoohi, "The World of E-Commerce: The Process from Digital Signature up to Market Basket Analysis", Journal of Technical and Physical Problems of Engineering (IJTPE), Issue 52, Vol. 14, No. 3, pp. 202-206, September 2022.

[5] B. Ture Savadkoohi, P. Nik Mohammadi, "Applying Wormhole Approach to Design a Hierarchy in a Relational Database for Quick Data Access", Journal of Technical and Physical Problems of Engineering (IJTPE), Issue 47, Vol. 13, No. 2, pp. 144-148, June 2021.

[6] C.C. Aggarwal, "Data Mining", The Text Book, Springer, Switzerland, 2015.

[7] 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", Journal of Expert Systems with Applications, Vol. 31, No. 1, pp. 101-107, 2006.

[8] C.H. Chen, R.D. Chiang, T.F. Wu, H.C. Chu, "A Combined Mining-Based Framework for Predicting Telecommunications Customer Payment Behaviors", Expert Systems with Application, Vol. 40, No. 16, pp. 6561-6569, 2013.

[9] M. Khajvand, M.J. Tarokh, "Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context", Procedia Computer Science, Vol. 3, No. 3, pp. 1327-1332, 2011.
[10] Y.H. Liang, "Integration of Data Mining Technologies to Analyze Customer Value for the Automotive Maintenance Industry", Journal of Expert Systems with Applications, Vol. 37, No. 12, pp. 7489-7496, 2010.
[11] D. Zakrzewska, J. Murlewski, "Clustering Algorithms for Bank Customer Segmentation", IEEE International Conference on Intelligent Systems Design and Applications, pp. 197-202, Warsaw, Poland, 8-10 September 2005.
[12] B. Zhang, S. Qin, W. Wang, D. Wang, L.Xue, "Data Stream Clustering Based on Fuzzy C-Mean Algorithm and Entropy Theory", Journal of Signal Processing, Vol. 126, pp. 111-116, 2016.
[13] R. Kusumawardani, M. Agintiara, "Application of Fuzzy AHP-TOPSIS Method for Decision Making in Human Resource Manager Selection Process", Information Systems International Conference, pp. 638-646, Surabaya, Indonesia, 2-4 November 2015.
[14] T.L. Saaty, "What is the Analytic Hierarchy Process?", Springer Berlin Heidelberg, pp. 109-121, Germany, 1988.
[15] Z. Gungor, G. Serhadlioglu, S. Erhan Kesen, "A Fuzzy AHP Approach to Personnel Selection Problem", Journal of Applied Soft Computing, Vol. 9, pp. 641-646, 2009.
[16] M. Celik, A. Kandakoglu, I. Deha Er, "Structuring Fuzzy Integrated Multi-Stages Evaluation Model on Academic Personnel Recruitment in MET Institutions", Journal of Expert System with Application, Vol. 36, No. 3, pp. 6918-6927, 2009.
[17] K. Phiwhorm, S. Arc hint, "LDL-Cholesterol Levels Measurement Using Hybrid Genetic Algorithm and Multiple Linear Regression", IEEE International Conference on Information Science and Application, 2013.
[18] A. Beerli, J.D. Martin, A. Quintana, "A Model of Customer Loyalty in the Retail Banking Market", European Journal of Marketing, Vol. 38, No. 1/2, pp. 253-275, 2004.
[19] P. Verhoef, B. Donkers, "Predicting Customer Potential Value an Application in the Insurance Industry", Journal of Decision Support Systems, Vol. 32, pp. 189-199, 2001.
[20] S.A. Taylor, K. Celuch, S. Goodwin, "The Importance of Brand Equity to Customer Loyalty", Journal of Product & Brand Management, Vol. 13, No. 4, pp. 217-227, 2004.
[21] J. Fandos Roig, J. Sanchez Garcia, M. Angel Moliner, J. Llorens Monzonis, "Customer Perceived Value in Banking Services", Journal of Bank Marketing, Vol. 24, No. 5, pp. 266-283, 2006.
[22] T. Chen, P. Chang, H. Chang, "Price, Brand Cues, and Banking Customer Value", Journal of Bank Marketing, Vol. 23, No. 3, pp. 273-291, 2005.
[23] J.C. Sweeney, G.N. Soutar, "Consumer Perceived value: The Development of a Multiple Item Scale", Journal of Retailing, Vol. 77, No. 2, pp. 203-220, 2001.
[24] J.N. Sheth, B.I. Newman, B.L. Gross, "Why We Buy What We Buy: A Theory of Consumption Values", Journal of Business Research, Vol. 22, No. 2, pp. 159-170, 1991.
[25] M. Mohaffyza Mohamad, N. Lisa Sulaiman, L. Chee Sern, K. Salleh, "Measuring the Validity and Reliability of Research Instrument", Procedia - Social and Behavioral Sciences, Vol. 204, pp. 164-171, 2015.
[26] L.A. Zadeh, "Fuzzy Sets", Journal of Information and Control, Vol. 8, No. 3, pp. 338-353, 1965.
[27] Z. Xu, R. Yager, "Dynamic Intuitionistic Fuzzy Multi-Attribute Decision Making", Journal of Approximate Reasoning, Vol. 48, No. 1, pp. 246-262, 2008.
[28] I.K. Vlachos, G.D. Sergiadis, "Intuitionistic Fuzzy Information - Applications to Pattern Recognition", Journal of Pattern Recognition Letters, Vol. 28, No. 2, pp. 197-206, 2007.
[29] S.K. De, R. Biswas, A.R. Roy, "An Application of Intuitionistic Fuzzy Sets in Medical Diagnosis", Journal of Fuzzy Sets and System, Vol. 117, No. 2, pp. 209-213, 2001.
[30] C.J. Willmott, K. Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance", Journal of Climate Research, Vol. 30, No. 1, pp. 79-82, 2005.

## BIOGRAPHIES

Name: **Bita**
Surname: **Ture Savadkoohi**
Birthday: 15.08.1980
Birth Place: Tabriz, Iran
Bachelor: Software Engineering, Department of Computer and Electrical Engineering, Islamic Azad University, Shabestar, Iran, 2003
Doctorate: Computer Science, Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, 2010
The last Scientific Positions: Assist. Prof., Seraj Higher Education Institute, Tabriz, Iran, Since 2012
Research Interests: Computer Graphics, (e.g., Shape Comparison), Analysis of 3D Data, Software Engineering, Data Base and Data Mining
Scientific Publications: 12 papers and international communications

Name: **Jalal**
Surname: **Vazifeh Shoae**
Birthday: 22.02.1978
Birth Place: Tabriz, Iran
Bachelor: Software Engineering from Department of Computer and Electrical Engineering, Islamic Azad University, Shabestar, Iran, 2002
Master: Computer Science from Department of Computer and Electrical Engineering, Islamic Azad University, Azarshahr, Iran, 2018
Research Interests: Data Base System, Data Mining and Software Engineering