

Modelling Time Series Customer Preference Based on E-commerce Website

Huimin Jiang¹, Chunsheng Li¹, Farzad Sabetzadeh^{2,*}

¹*Department of Decision Sciences, School of Business, Macau University of Science and Technology, Macau, China*

²*Faculty of Business, City University of Macau, Macau, China*

* *Corresponding author. Email: farzad@cityu.mo*

ABSTRACT

In the research of customer preference for products, we need to collect enough customer data first, and then we can analyze it to realize the effective establishment of customer preference model. Although a reasonable model can be obtained, the customer's own preference presents a dynamic state, which changes disorderly with the change of time. This requires a systematic analysis of the time series data of customer preference based on different time periods, so as to complete the modeling of customer preference. Through research, this paper proposes a new time series customer preference modeling method, which can effectively process and analyze online customer comment data of e-commerce websites, including opinion mining and chaos optimization algorithm (COA) based dynamic evolutionary neuro fuzzy inference system (DENFIS). Finally, a case verification method is adopted, and a hair dryer is selected to verify the effectiveness and rationality of the method. The comparison results show that the proposed COA based DENFIS approach performs better than K-means-based ANFIS, fuzzy c-means-based ANFIS, subtractive cluster-based ANFIS and DENFIS in terms of mean absolute percentage error.

Keywords: *time series customer preference, e-commerce website, opinion mining, COA based DENFIS*

1. INTRODUCTION

The research on customers' perception of products is of great significance, especially in product development, which can bring strategic support to develop products to meet customers' needs. Due to volatile market and diversified technological innovation, the preferences, needs and desires of customers always vary with respect to time. It is likely that what customers said in the past or currently will not be relevant by the time that new products are launched in the future [1]. However, previous research mostly focuses on studying the static customer preferences with the assumption that past customers' survey data can reflect customer's behaviours at a certain period of time in the future.

In terms of customer survey, interview or questionnaire survey is the most basic way, so as to obtain certain customer preference data, and then analyze and study it. However, with the continuous change and improvement of customer preferences, it is more difficult to collect the time series data of customer preferences. Moreover, the investigation process needs to consume high economic cost and time cost. With the development

of the Internet, online customers' comments on products have become an important evaluation method, and they are relatively easy to obtain on the Internet. Such comments are very rich in content and have high research value. Therefore, they can be used as data to dynamically analyze customers' product preferences. In fact, in the past, dynamic evolutionary neuro fuzzy inference system (DENFIS) was proposed in the research of customer preference [2]. However, the determination of parameters in DENFIS is difficult which affects the accuracy of modelling time series customer preference. This article is based on this, opinion mining and chaos optimization algorithm (COA) based DENFIS approaches are proposed to model time series customer preference based on e-commerce websites. It is necessary to effectively collect online comment data based on different time periods. Then the opinion mining method is used to analyze the data of each time period, complete the process of emotion analysis, and get the emotion score of customer preference. Based on the setting of design attributes and results of opinion mining, COA based DENFIS is adopted to model time series customer preferences.

2. RELATED WORKS

Through the opinion mining method to process the online comment data, and then get the product attributes and customer preferences, researchers have done a lot of research. In order to realize the effective mining and extraction of sentence level adaptive text, Lee [3] studied and designed the supervised machine learning method, which can accurately identify the needs of customers. Wang et al. [4] also conducted a lot of research. In order to obtain product attributes, they developed customer driven product design. At the same time, in the modelling stage of customer preference, they innovatively adopted Bayesian linear regression method, which has been verified to be highly effective. Chen et al. [5] designed an ontology learning customer requirement representation system on the extraction of customer requirements. The ontology of the system has higher comprehensiveness and coverage in semantics. Zimmermann et al. [6] focused on the extraction of important hidden features in online comment data, developed and designed a framework for effective monitoring and recognition. Opinion mining and case-based reasoning was combined to achieve accurate identification of potential customer needs [7]. Zhang et al. [8] developed and designed a new opinion mining and extraction algorithm, which can be used for various analysis, such as feature opinion relationship and so on. Zhou et al. [9] developed a new effective model for customer preference, which can also integrate customers' negative opinions and positive opinions, so as to optimize and supplement the feature model. Through research, Kang and Zhou [10] developed a method that can extract the objective and subjective features of customers. Tuarob and tucker [11] tried to mine and extract data in social media networks, which can automatically mine potential product features and users' data, which has high research significance and value.

With the change of time, customers' preference for products will also change greatly, showing a dynamic change trend. The research on predicting the future customer preference is very important, and there are many previous research results. Shen et al. [12] obtained the fuzzy trend analysis method through research, which can effectively analyze the importance trend of customer perception over time in quality function deployment. Xie et al. [13] studied and realized the effective prediction of the future importance perceived by customers, and adopted the method of double exponential smoothing technology. Wu et al. [14] proposed a new model theory, which can effectively analyze the future importance and dynamics of customer preference based on the past, present and forecasted data, that is, grey theory model. Huang et al. [15] predicted the future importance based on the artificial immune system. Jiang et al. [16] adopted another method, that is, fuzzy time series method for predicting the future importance of customer preference

based on online customer reviews. DENFIS was proposed to model time series customer preferences based on online customer reviews [2]. However, the optimal parameters of DENFIS cannot be determined which affects the accuracy of prediction.

3. PROPOSED METHODOLOGY

The proposed methods include opinion mining and COA based DENFIS method to model time series customer preferences.

3.1. Opinion Mining From E-commerce Website

The sample products are first identified. Then, customer reviews of sample products on the e-commerce website are obtained by the web crawlers, which are divided into different periods and put into the separate excel files. In this study, Semantria, a well-known text analysis software tool, was used for opinion mining of online reviews. It provides text analysis using Excel plug-in, extracts opinions based on positive, neutral, and negative dimensions and calculates the corresponding sentiment scores.

3.2. Chaos Optimization Algorithm (COA)

The COA adopted in this paper can effectively determine the optimal parameter setting of DENFIS. In fact, COA algorithm has many advantages, especially adding chaos to the optimization strategy, which can accelerate the analysis of the global optimal solution. After a series of iterative calculations, the logistic model generates chaotic variables, as shown in (1).

$$c_{n+1} = f(c_n) = \mu c_n(1 - c_n) \quad (1)$$

In the formula, $c_n \in [0, 1]$ and $\mu = 4$ represent the n th iteration value of the chaos variable c .

The chaotic variable is transformed into optimization variable by using the linear mapping formula:

$$q_n = a + (b - a) \cdot c_n \quad (2)$$

In the formula, a and b represent the lower and upper limits of the optimization variable q , respectively. q_n represents the optimization variable; During the iteration, the chaotic variable will traverse $[0, 1]$ and the optimization variable will traverse $[a, b]$. Based on this, the optimal solution is obtained.

3.3DENFIS Approach

In fact, DENFIS method adopts evolving clustering method (ECM) to realize the clustering division of input,

and then relies on the clustering centers to establish the antecedent of new fuzzy rules. Based on the threshold D_{thr} , the number of clusters can be determined. D_{thr} can effectively control the maximum distance between the cluster center and data points, and then become a constraint to update the cluster radius. Cluster C_1 needs to be initialized. Its radius Ru_1 and center Cc_1 are set to zero and the first data set, respectively. When the new data $Z_i, i = 2, \dots, n$, is presented, the distances, D_{ij} , from Z_i to the existing clusters, $C_j, j = 1, 2, \dots, m$, are calculated by (3).

$$D_{ij} = \|Z_i - Cc_j\|, j = 1, 2, \dots, m \quad (3)$$

In the formula, Cc_j represents the cluster center of C_j ; $\| \cdot \|$ is the Euclidean distance from Z_i to C_j ; m represents the quantity of existing clusters, n represents the number of data sets. In addition, the minimum distance is calculated:

$$D_{i\min} = \min(D_{ij}) = \|Z_i - Cc_{\min}\| \quad (4)$$

C_{\min} represents cluster with $D_{i\min}$ in which The Ru_{\min} represents radius and Cc_{\min} represents corresponding center, If $D_{i\min}$ is no more than Ru_{\min} , $D_{i\min} \leq Ru_{\min}$, Z_i belongs to the cluster C_{\min} . On the contrary, the existing cluster will be updated or a new cluster will be established. Through (5), calculate V_{ij} ; through (6), calculate V_{ia} , which represent the minimum value of V_{ij} .

$$V_{ij} = D_{ij} + Ru_j, j = 1, 2, \dots, m \quad (5)$$

$$V_{ia} = \min(V_{ij}) = D_{ia} + Ru_a \quad (6)$$

C_a represents cluster with V_{ia} in which Ru_a represents radius, Cc_a represents corresponding center. From Z_i to C_a , D_{ia} represents the distance. When $V_{ia} > 2 \times D_{thr}$, a new cluster will be established. Z_i and 0 are set as the center and radius of the new cluster, respectively. When $V_{ia} \leq 2 \times D_{thr}$, the cluster C_a will be updated. The radius Ru_a' will be updated as $V_{ia} / 2$. Cc_a' is new center, will be located at the point on the line connecting Z_i and Cc_a and the distance from the Cc_a' to Z_i is equal to Ru_a' . Then all data sets are processed to complete the ECM algorithm process.

Based on the input clustering, the corresponding fuzzy rules can be obtained:

If x_1 is MF_{11} , x_2 is MF_{12}, \dots , and x_q is MF_{1q} , then y is $f_1(x_1, x_2, \dots, x_q)$

If x_1 is MF_{21} , x_2 is MF_{22}, \dots , and x_q is MF_{2q} , then y is $f_2(x_1, x_2, \dots, x_q)$

⋮

If x_1 is MF_{m1} , x_2 is MF_{m2}, \dots , and x_q is MF_{mq} , then y is $f_m(x_1, x_2, \dots, x_q)$ (7)

In the formula, $x_i, i = 1, 2, \dots, q$ represent the i th input variable of DENFIS and the number of inputs is q . In fact, the input includes the emotional score of customer preference in the previous periods and the setting of design attributes; $MF_{ji}, j = 1, 2, \dots, m, i = 1, 2, \dots, q$ represents the j th membership function of x_i ; m is the number of membership function and is equal to the number of clusters based on ECM; " x_i is MF_{ji} " are $m \times q$ fuzzy propositions as antecedents of m fuzzy rules; y is the output of DENFIS which is the future sentiment scores of the customer preference; $f_j(x_1, x_2, \dots, x_q), j = 1, 2, \dots, m$ represent the first-order Sugeno fuzzy models in the subsequent parts of fuzzy rules. The triangular membership function is calculated as follows.

$$\mu_j(x_i) = \begin{cases} 0, & x_i < a_j \\ \frac{x_i - a_j}{b_j - a_j}, & a_j \leq x_i \leq b_j \\ \frac{c_j - x_i}{c_j - b_j}, & b_j \leq x_i \leq c_j \\ 0, & x_i > c_j \end{cases} \quad (8)$$

In the formula, b_j represent the center value of the j th cluster; the left and right values of the membership function are $a_j = b_j - d \times D_{thr}$ and $c_j = b_j + d \times D_{thr}$, $1.2 \leq d \leq 2$;

Based on the weighted recursive least squares estimation method, the establishment of the first-order linear model of the subsequent part of each fuzzy rule is realized, which is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q \quad (9)$$

The data sets $\{[x_1^l, x_2^l, \dots, x_q^l], y_l\}, l = 1, 2, \dots, n$ is to get the regression coefficient, $\beta = [\beta_0, \beta_1, \beta_2, \dots, \beta_q]^T$, where n is the number of data pairs. Using (10) and (11), the calculation and derivation of the initial inverse matrix P and the coefficients β can be realized based on the weighted least squares estimation.

$$P = (A^T W A)^{-1} \quad (10)$$

$$\beta = P A^T W Y \quad (11)$$

where

$$A = \begin{pmatrix} 1 & x_1^1 & x_2^1 & \dots & x_q^1 \\ 1 & x_1^2 & x_2^2 & \dots & x_q^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^n & x_2^n & \dots & x_q^n \end{pmatrix}, W = \begin{pmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & W_n \end{pmatrix}$$

$$Y = [y_1, y_2, \dots, y_n]^T$$

of which $W_l = 1 - Dis_l$, $l = 1, 2, \dots, n$ and Dis_l represent the length between the current cluster center and the l th data set ; $[]^T$ and $()^{-1}$ represent the transpose and inverse of a matrix.

P and β are initialized. When the new data set is entered, the coefficients β_{l+1} and inverse matrix P_{l+1} at the $(l+1)$ th iteration will be updated.

$$\beta_{l+1} = \beta_l + W_{l+1} P_{l+1} \alpha_{l+1} (y_{l+1} - \alpha_{l+1}^T \beta_l) \tag{12}$$

$$P_{l+1} = \frac{1}{\lambda} (P_l - \frac{W_{l+1} P_l \alpha_{l+1} \alpha_{l+1}^T P_l}{\lambda + \alpha_{l+1}^T P_l \alpha_{l+1}}) \tag{13}$$

where $\alpha_{l+1}^T = [1 \ x_1^{l+1} \ x_2^{l+1} \ \dots \ x_q^{l+1}]$ and it is the $(l+1)$ th row vector of matrix A ; y_{l+1} represent $(l+1)$ th element of Y ; and λ is a forgetting factor and $0 < \lambda \leq 1$.

Through the above learning, the calculation of DENFIS' l th prediction output can be realized, with the weighted average of each rule output.

$$y_l = \frac{\sum_{j=1}^m w_j f_j(x_1^l, x_2^l, \dots, x_q^l)}{\sum_{j=1}^m w_j} \tag{14}$$

$$w_j = \sum_{i=1}^q \mu_j(x_i^l) \tag{15}$$

where $\mu_j(x_i^l)$ is determined by (8).

4. CASE STUDY

The method proposed in this paper needs case analysis and verification. It adopts the method of time series customer preference modeling in the line with the customer comments of hair dryer products, which can systematically verify and compare the effectiveness and rationality of the method proposed in this paper. During the research, a variety of hair dryer products were compared, and ten typical ones were selected and numbered, expressed as A ~ J. In terms of online comment data acquisition, it is based on the method of the fixed time period, and there are four specified time periods. The online platform selects Amazon. com, which has authoritative and comprehensive online customer comment data on hair dryer products. The above data are collected and uniformly summarized into an excel file. Then the collected data are analyzed and processed, and the corresponding opinion mining can be

carried out by using the semantria excel add-on, which is more convenient and reliable. In the research process, the customer preference "performance" which is expressed as y , is used to illustrate the proposed COA based DENFIS method. The obtained emotional score of "performance" is shown in Table 1.

Table 1 The sentiment scores of "performance"

Product	Performance			
	Stage I	Stage II	Stage III	Stage IV
A	0.31	0.31	0.32	0.29
B	0.28	0.23	0.3	0.25
C	0.38	0.35	0.34	0.34
D	0.22	0.26	0.31	0.25
E	0.35	0.31	0.34	0.44
F	0.34	0.33	0.34	0.28
G	0.34	0.48	0.23	0.19
H	0.35	0.31	0.36	0.32
I	0.18	0.27	0.36	0.27
J	0.39	0.33	0.42	0.4

This paper summarizes the design attributes that are closely related to customer preference "performance", as shown in Table 2. The four design attribute include weight x_1 , power x_2 , heat setting x_3 and speed setting x_4 .

Table 2 Design attributes settings

Product category	Design property			
	Weight (x_1)	Power (x_2)	Heat setting (x_3)	Speed setting (x_4)
A	1.75	1875	2	2
B	1.50	1875	3	2
C	2.30	1875	3	2
D	2.55	1875	3	2
E	0.50	1000	2	2
F	1.60	1875	3	2
G	0.30	1875	3	3
H	1.00	2000	4	3
I	1.65	1875	3	3
J	1.80	2000	3	3

In this study, stage 4 was treated as the future period and the sentiment scores of stage 4 were denoted as $y(t)$, where t is the time period. Stage 1 ~ 3 are considered as the historical periods. The emotion scores of stages 1~3 presented as $y(t-3)$, $y(t-2)$, and $y(t-1)$, respectively, as well as the setting of four design attributes, x_1 , x_2 , x_3 , and x_4 , were used to predict the future sentiment scores of "performance" of stage 4, $y(t)$. Based on the COA based DENFIS, the time series customer preference models can be developed. The following shows some examples of the fuzzy rules generated.

Fuzzy rule I is as follows:

If x_1 is MF_{11} , x_2 is MF_{12} , x_3 is MF_{13} , x_4 is MF_{14} , $y(t-3)$ is MF_{15} , $y(t-2)$ is MF_{16} , and $y(t-1)$ is MF_{17} , then $y(t) = 0.0165 + 0.0141x_1 + 0.0139x_2 + 0.0496x_3 + 0.0331x_4 + 0.0063y(t-3) + 0.0058y(t-2) + 0.0056y(t-1)$.

Fuzzy rule II is as follows:

If x_1 is MF_{21} , x_2 is MF_{22} , x_3 is MF_{23} , x_4 is MF_{24} , $y(t-3)$ is MF_{25} , $y(t-2)$ is MF_{26} , and $y(t-1)$ is MF_{27} , then $y(t)=0.0204+0.0129x_1+0.0172x_2+0.0409x_3+0.0409x_4+0.0063y(t-3)+0.0063y(t-2)+0.0065y(t-1)$.

Fuzzy rule III is as follows:

If x_1 is MF_{31} , x_2 is MF_{32} , x_3 is MF_{33} , x_4 is MF_{34} , $y(t-3)$ is MF_{35} , $y(t-2)$ is MF_{36} , and $y(t-1)$ is MF_{37} , then $y(t)=0.0198+0.0125x_1+0.0167x_2+0.0397x_3+0.0397x_4+0.0061y(t-3)+0.0061y(t-2)+0.0063y(t-1)$.

Fuzzy rule IV is as follows:

If x_1 is MF_{41} , x_2 is MF_{42} , x_3 is MF_{43} , x_4 is MF_{44} , $y(t-3)$ is MF_{45} , $y(t-2)$ is MF_{46} , and $y(t-1)$ is MF_{47} , then $y(t)=0.0219+0.0138x_1+0.0184x_2+0.0437x_3+0.0437x_4+0.0068y(t-3)+0.0068y(t-2)+0.0070y(t-1)$.

Fuzzy rule V is as follows:

If x_1 is MF_{51} , x_2 is MF_{52} , x_3 is MF_{53} , x_4 is MF_{54} , $y(t-3)$ is MF_{55} , $y(t-2)$ is MF_{56} , and $y(t-1)$ is MF_{57} , then $y(t)=0.0184+0.0116x_1+0.0155x_2+0.0369x_3+0.0369x_4+0.0057y(t-3)+0.0057y(t-2)+0.0059y(t-1)$.

Through comparative analysis with other methods, the reliability and accuracy of the proposed method can be obtained. The methods of comparative analysis

Table 3 The validation results for the “performance” based the five approaches

Validation Test	Validation datasets	K-means-ANFIS	FCM-ANFIS	SC-ANFIS	DENFIS	COA -DENFIS
1	A, B	0.1671	0.1678	0.1670	0.1654	0.1566
2	C, D	0.1489	0.1480	0.1481	0.1458	0.1379
3	E, F	0.3044	0.2996	0.3095	0.2764	0.2578
4	G, H	0.5562	0.5549	0.5491	0.4905	0.4870
5	I, J	0.1915	0.1921	0.1922	0.1842	0.1565

5. CONCLUSION

Through research, this paper proposes a new time series customer preference modelling method based on e-commerce website, including opinion mining of online comments and COA based DENFIS modelling method. Then, the proposed method is verified by the case analysis, and the hair dryer products based on online comments are selected as the analysis object. By comparing the mean absolute percentage error of the proposed method and the other four methods, it is obtained that the COA based DENFIS method has more prominent advantages than the K-means-based ANFIS, FCM-ANFIS, SC-ANFIS and DENFIS. The method proposed in this paper has higher modelling accuracy.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China [grant number 71901149].

REFERENCES

[1] M.E. Adams, G.S. Day, D. Dougherty, Enhancing new product development performance: an

include K-means-based ANFIS, fuzzy C-means-based ANFIS (FCM-ANFIS), subtractive cluster-based ANFIS (SC-ANFIS) and DENFIS. The verification results of the five methods are compared, and then the mean absolute percentage error (MAPE) method is used to get the comparison results.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y'_i(t) - y_i(t)|}{y_i(t)} \tag{16}$$

The results obtained from the verification are shown in Table 3. The MAPE of the other four methods are higher than that of the method proposed in this paper, so the method proposed in this paper has higher prediction accuracy.

organizational learning perspective, Journal of Product Innovation Management, 15 (1998) 403-422.

[2] H. Jiang, C.K. Kwong, G.E. Okudan Kremerec, W.Y. Park, Dynamic modelling of customer preferences for product design using DENFIS and opinion mining, Advanced Engineering Informatics, 42 (2019) 100969.

[3] T. Lee, Automatically Learning User Needs from Online Reviews for New Product Design, In: Proceedings of America Conference on Information Systems, 2009, pp. 22.

[4] L. Wang, B. D. Youn, S. Azarm, P.K., Kannar, Customer-driven product design selection using web based user-generated content, In: Proceedings of ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2011, pp. 405-419,

[5] X. Chen, C. H. Chen, K. F. Leong, X. Jiang, An ontology learning system for customer needs representation in product development, The International Journal of Advanced Manufacturing Technology, 67 (1-4) (2013) 441-453.

- [6] M. Zimmermann, E. Ntoutsis, M. Spiliopoulou, Discovering and monitoring product features and the opinions on them with OPINSTREAM, *Neurocomputing*, 150 (Part A) (2015) 318–330.
- [7] F. Zhou, R. J. Jiao, J. S. Linsey, Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews, *Journal of Mechanical Design*, 137 (7) (2015) 071401-1-11.
- [8] H. Zhang, A. Sekhari, Y. Ouzrout, A. Bouras, Jointly identifying opinion mining elements and fuzzy measurement of opinion intensity to analyze product features, *Engineering Applications of Artificial Intelligence*, 47 (2016) 122–139.
- [9] F. Zhou, R. J. Jiao, X. J. Yang, B. Lei, Augmenting feature model through customer preference mining by hybrid sentiment analysis, *Expert Systems with Applications*, 89 (2017) 306-317.
- [10] Y. Kang, L. Zhou, RubE: Rule-based methods for extracting product features from online consumer reviews, *Information & Management*, 54 (2) (2017)166-176.
- [11] S. Tuarob, C.S. Tucker, Automated discovery of lead users and latent product features by mining large scale social media networks, *Journal of Mechanical Design*, 137 (7) (2015) 071402-1-11.
- [12] X.X. Shen, M. Xie, K.C. Tan, Listening to the future voice of the customer using fuzzy trend analysis in QFD, *Quality Engineering*, 13 (3) (2001) 419-425.
- [13] M. Xie, K. C. Tan, T.N. Goh, *Advanced QFD applications*, ASQ Quality Press, Wisconsin, 2003.
- [14] H.H. Wu, A.Y.H. Liao, P.C. Wang, Using grey theory in quality function deployment to analyse dynamic customer requirements, *International Journal of Advanced Manufacturing Technology*, 25 (11) (2005) 1241-1247.
- [15] A.H. Huang, H.B. Pu, W.G. Li, G.Q. Ye, Forecast of importance weights of customer requirements based on artificial immune system and least square support vector machine, In: *Proceedings of 2012 International Conference on Management Science and Engineering*, 2012, pp. 83-88.
- [16] H. Jiang, C.K. Kwong, K.L. Yung, Predicting Future Importance of Product Features Based on Online Customer Reviews, *Journal of Mechanical Design*, 139 (11) (2017) 111413-1-10, 2017.