mathematics

Article

A Multi-Objective Optimization-Algorithm-Based ANFIS Approach for Modeling Dynamic Customer Preferences with Explicit Nonlinearity

Huimin Jiang [1] and Farzad Sabetzadeh [2,*]

[1] School of Business, Macau University of Science and Technology, Macau 999078, China; hmjiang@must.edu.mo
[2] Faculty of Business, City University of Macau, Macau 999078, China
* Correspondence: farzad@cityu.edu.mo

Abstract: In previous studies, customer preferences were assumed to be static when modeling their preferences based on online reviews. However, in fact, customer preferences for products are dynamic and changing over time. Few research has been conducted to model dynamic customer preferences as the time series data of customer preference are difficult to be obtained. Based on online reviews, an adaptive neuro fuzzy inference system (ANFIS) was introduced to model customer preferences, which can take into account the fuzzy nature of customers' emotions and the nonlinearity of the model. However, ANFIS is plagued with black box problems, and the nonlinearity of the model cannot be directly demonstrated. To address the above research issues, a multi-objective chaos optimization algorithm (MOCOA)-based ANFIS approach is proposed to generate customer preferences models by using online reviews, which has explicit nonlinear inputs. Firstly, a sentiment analysis approach is used to derive information from online reviews by periods, which is used as the time series data sets of the proposed model. A MOCOA is combined into ANFIS to identify the nonlinear inputs, which include single items, interactive items, and terms of second order and/or higher-order terms. Consequently, the fuzzy rules in ANFIS are expressed in polynomial form, which allows for the explicit representation of the nonlinearity between customer preferences and product attributes. A case study of sweeping robots is used to compare the validation results of the proposed approach with those of ANFIS, subtractive cluster-based ANFIS, fuzzy c-means-based ANFIS, and K-means-based ANFIS. Moreover, the proposed approach provides better performance than the other four approaches in terms of mean relative error and variance of error.

Keywords: dynamic customer preferences; explicit nonlinearity; multi-objective chaos optimization algorithm; ANFIS

MSC: 68T07; 68W50

Citation: Jiang, H.; Sabetzadeh, F. A Multi-Objective Optimization-Algorithm-Based ANFIS Approach for Modeling Dynamic Customer Preferences with Explicit Nonlinearity. *Mathematics* **2023**, *11*, 4559. https://doi.org/10.3390/math11214559

Academic Editors: Alexey Filippov, Vilém Novák and Anton Romanov

Received: 8 October 2023
Revised: 2 November 2023
Accepted: 3 November 2023
Published: 6 November 2023

1. Introduction

When designing customer-driven products, establishing the associations between customer preferences and product attributes is crucial to the success of the products [1]. Based on the models of customer preferences, customer preferences can be understood, and new products can be designed to maximize the value of customer preferences. In previous studies, it is assumed that customer preferences are static. It is well known, however, that customers' preferences are dynamic and may change rapidly over time, particularly regarding customer products. When a new product is launched, customer preferences differ significantly from the collected customer preferences at the product design stage. It is likely that the accuracy of customer preference predictions will be adversely affected if static customer preferences continue to be used in the modeling process. Therefore, it is necessary to capture the changing trend of the customer preferences in the research. It is important to consider and incorporate time series data sets regarding customer preferences into the modeling process. However, in previous studies, collecting time series data through interviews or surveys has proven difficult since a number of surveys need to be conducted at different intervals, not only increasing the cost, but also taking up a great deal of time and resources. Currently, online reviews of products can be found easily in online e-commerce websites, which contain rich information about customer preferences for the products. Moreover, these online comments have a significant impact on the purchase decisions of potential customers. Through data mining from online reviews, valuable information can be obtained that is useful for companies to design new products. In addition, by analyzing the online reviews over a period of time, it is possible to obtain time series data regarding customer preferences at a low cost. As a result of collecting the time series data, models can be developed between the preferences of the customer and product attributes. In the research of developing customer preference models using online reviews, two issues need to be addressed: the fuzziness of the customers' sentimental expressions and nonlinear relationships in the models. Most existing approaches can only solve one of the above issues in the modeling. In previous studies, it has been shown that the adaptive neuro fuzzy inference system (ANFIS) is a very effective way to solve both issues, which can establish the nonlinear relationship between customer preferences and product attributes, and the obtained fuzzy rules can address the fuzziness in the data [2]. However, the ANFIS has black box problems and the nonlinearity in the model cannot be demonstrated explicitly, which makes it more challenging to analyze the relationship between customer preference and product attributes. Based on the development of explainable artificial intelligence, it is essential to show the explicit and understandable nonlinearity in the ANFIS for the modeling process. To solve the above research problems, a multi-objective chaos optimization algorithm (MOCOA)-based ANFIS approach is proposed to generate customer preference models based on online reviews that has explicit nonlinear inputs. In this study, we first analyze the online reviews, determine the categories of customer preferences, and compute the sentiment scores of customer preferences for each period to obtain the time series data. Using the data sets, the proposed approach is employed to model dynamic customer preferences. MOCOA is used to determine the optimal nonlinear inputs for ANFIS which include single items, interactive items, and terms of second order and/or higher-order terms. Consequently, the fuzzy rules generated in ANFIS are expressed in polynomial structure, which allows for the explicit representation of the nonlinear relationship between the customer preferences and product attributes. Previous studies have demonstrated that by following the regularity of chaotic motions, the chaos optimization algorithm (COA) has the advantage of avoiding local minimums and producing more accurate solutions. Moreover, it has a fast convergence rate and can provide a variety of solutions [3]. In the proposed approach, the advantages of MOCOA and ANFIS are combined to develop dynamic customer preference models that can cope with the fuzziness and nonlinearity of the model while being understandable to the reader.

The remaining contents of the paper are arranged as follows. In Section 2, we describe the related works. The proposed MOCOA-based ANFIS approach is introduced in Section 3. Section 4 presents a case study of sweeping robots illustrating the implementation of the proposed approach. In Section 5, the proposed approach is compared with the other four approaches, along with their comparison results. Section 6 presents the conclusions of the study.

2. Related Works

Sentiment analysis is the qualitative analysis of sentiments and emotions based on the text contents and the basic tasks include emotion recognition and polarity detection [4]. Based on the sentiment analysis using the product comments written by the customers, the mined features such as the customer preferences and the product attributes can be accurately understood. In addition, the categories and the sentimental scores of the features can be determined. There have been some previous studies conducted on sentiment analysis in the context of product design. In order to extract information about customer needs in the development of products, an ontology learning system is introduced by Chen et al. [5]. In [6], the latent customer preferences were elicited from online product reviews using a two-layer model based on sentiment analysis and case analogical reasoning. Tuarob and Tucker [7] introduced an automated method to identify the lead customer and implicit product features from the large number of social media networks. Using social media data, they also used a data mining methodology to determine the features of the product and customer opinions [8]. Zimmermann et al. [9] proposed a framework called OPINSTREAM in order to analyze product features and determine the polarity of reviews for a variety of products. A Bayesian sampling method has been proposed to mine product features from the large amounts of textual data [10]. Zhang et al. [11] introduced an algorithm to determine the main product features and identify the opinion intensity using the fuzzy measurements. Online comments can be used to identify product features based on a rule-based approach [12]. Zhou et al. [13] proposed a sentiment analysis which is a hybrid combination of various affective lexicons to extract customer preference information using online reviews and incorporated the mined information into the model as an attribute. The researchers have proposed to combine Kansei engineering with text mining to extract customer preferences in the conceptual data-driven design based on online customer reviews [14]. Patent mining and social media were used to identify the real-time customer needs by analyzing the online voice of customers [15]. Zhang et al. [16] described an online-review-based method for identifying to-be-improved product features based on customer reviews. Ali et al. [17] integrated the ontology and the natural language processing system to extract design features using online reviews in the conceptual design phase. Using Kansei engineering and machine learning, Li et al. [18] proposed a method of mining customers' affective responses to products based on online comments. Jin et al. [19] mined online reviews using a Kansei-integrated Kano model to reveal customer affective needs for innovative product design. In addition, few examples of research were conducted on developing the models between customer preferences and product attributes. A rule-induction framework was developed by Chung and Tseng [20] to determine the relationship between customer reviews and customer ratings. In affective design, Jiang et al. [21] applied online comments to mine the association rules between customer preferences and product attributes using a multi-objective particle swarm optimization approach. Nevertheless, the obtained if-then rules are insufficient for determining the optimal product attributes settings.

There have been a number of previous studies conducted in order to construct the models between product attributes and customer preferences. By using the developed models, the optimal settings of product attributes can be better determined for the new products. Quantification I analysis was used to identify important product attributes and generate customer preference models [22]. A Kohonen self-organizing map neural network was applied in constructing the association between the customers' affective requirements and product attributes [23]. To map consumer preferences to product attributes and set optimal target values for improving product quality, a belief rule-based methodology was developed [24]. However, in the process of modeling customer preferences, there is considerable fuzziness in the responses of the customers, and the above approaches cannot account for this fuzziness. In order to accommodate the vagueness in the relationship between the customer requirements and the relevant product attributes, the fuzzy inference technique was employed [25]. To develop models relating product attributes to affective user satisfaction, a fuzzy rule-based approach was proposed [26]. In order to develop functional relationships for product design, a non-linear possibilistic regression approach was used [27]. Using Tanaka's fuzzy linear regression method, classification models were built for data collected through a customer satisfaction survey [28].

In recent studies, in order to account for both fuzziness and nonlinearity in the modeling process of customer preferences, fuzzy regression methods with a polynomial structure were proposed in recent studies. A genetic-programming-based fuzzy regression approach was introduced by [29] in order to address the issues of nonlinearity and fuzziness in affective modeling for product design. As part of the process of developing customer preference models in new product development, they also proposed a stepwise-based fuzzy regression procedure [30] as well as a forward-selection-based fuzzy regression procedure [31]. For the development of nonlinear and fuzzy models of customer preferences, a COA-based fuzzy regression approach was developed [32]. The improved optimization algorithms such as the parallel hybrid genetic algorithm [33] and the multiobjective optimization framework with genetic algorithms [34] can be introduced into the fuzzy regression approach to determine the nonlinear structure of the models. In order to deal with the nonlinearity and fuzzy nature of the process, a rough set and particle-swarm-optimization-algorithm-based ANFIS approach was introduced [2]. However, the ANFIS has black box problems, and the nonlinearity could not be explained explicitly by the customer preference models developed.

3. Proposed Approach

In the proposed MOCOA-based ANFIS approach, the data sets used for the modeling contain the time series sentiment scores of customer preferences, which are obtained from the sentiment analysis on the online reviews by specific time periods, and the settings of product attributes for the sampled products, which are obtained from the description of the products. With the data sets, MOCOA is combined into the ANFIS to identify the nonlinear inputs, which include single items, interactive items, and terms of second order and/or higher-order terms. Based on the nonlinear inputs, the models of customer preferences with the nonlinear fuzzy rules are developed. The process of the proposed approach is shown in Figure 1.

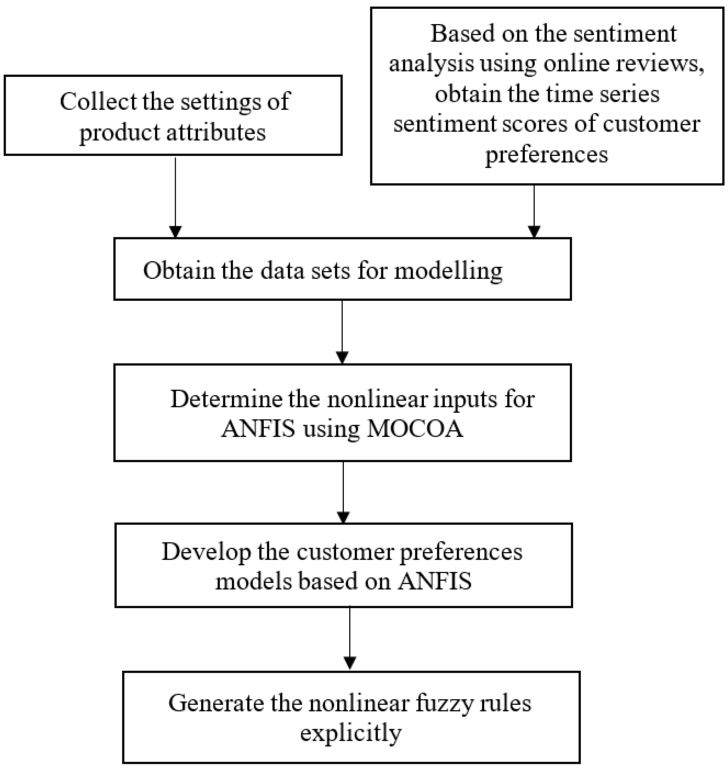

Figure 1. The process of the proposed approach.

3.1. Preparation of the Data Sets for Modeling

The sampled products used for the case study are first determined. By using a web crawler, their online reviews are collected from the websites, which are then stored in Excel files for opinion mining. Using Lexalytics' text and sentiment analysis, called the Semantria

Excel plug-in, the collected online reviews are analyzed, and the dimensions of the customer preferences are extracted. Semantria is a well-known text analysis software tool which contains an industry-specific dictionary and can calibrate the sentiment scores to be more accurate to a specific subject [35]. In Semantria, sentiment analysis consists of six steps: First, the collected reviews are cleaned by removing punctuation, stop words, and HTML characters. Second, the opinion words in the reviews are labelled as either nouns, adjectives, adverbs, or verbs by using Part-of-Speech (POS) tagging. Third, the extracted opinion words or phrases with high frequency are listed and identified as features. Fourth, incorrect and redundant features are eliminated through the process of feature pruning. Fifth, a K-means clustering method is used to cluster the synonymous phrases into the same group. Using the case study of sweeping robot as an example, the words "broken", "durable", "working good", and "good quality" are the synonymous words for the dimension of customer preference "quality", which are used as the settings of the Semantria. Finally, the SentiWordNet sentiment lexicon is used to determine the semantic polarity of opinion words as well as their sentiment scores. Then, the online reviews are divided by the fixed time periods. Based on the setting in the fifth step, the sentiment analysis is again conducted using the online reviews by time periods. As a result of computing the sentiment scores of customer preferences over the consecutive periods, the time series data are obtained, in which the sentiment scores of the latest period are used as the future period and the sentiment scores of the other periods are used as the historical periods. In addition, the product attributes relating to the customer preferences are identified and the corresponding settings are collected from the information of the products. Based on the time series data of the customer preferences and the settings of product attributes, the data sets used for the modeling are prepared.

3.2. Determination of Nonlinear Inputs Using MOCOA

To search for optimal solutions in COA, the carrier wave method is applied, traversing all the status values within the defined range without repeating any of them. There are two phases in the search process in COA, namely, the global search and the local search. The global search is performed first in order to determine the good states based on the ergodic trace of the entire search area. Good states are achieved when the termination requirement is met, which indicates that the optimal solution is approaching. Based on these results, the local search is initiated. Through the use of a small disturbance term, a deeper search is conducted in the local space. If the criterion for final termination is satisfied, then the optimal solution is obtained.

This study utilizes the logic model in COA. Based on (1), the chaos variables can be generated through the iteration of the calculation.

$$c_{m+1} = f(c_m) = \mu c_m(1 - c_m) \tag{1}$$

where m denotes the number of iterations. At the mth iteration, the value of the chaos variable is c_m, which is in the range of [0, 1]; μ is usually set to 4 to achieve all the chaotic states. By using (1), the chaos variables are obtained by the iteration $m + 1 \to m$.

Based on the linear mapping (2), the chaos variable c_m is transferred into the optimization variable q_m.

$$q_m = a + (b - a)c_m \tag{2}$$

The values of the elements in q_m are restricted by the minimum and maximum values, which are denoted as a and b, respectively. Based on (1), c_m traverses in the range of [0, 1] and q_m traverses in the range of [a, b].

The nonlinear inputs of ANFIS are determined by the values of q_m. The elements in q_m contain the inputs of data sets and the connection between every two inputs. The inputs of data sets, $x_1, x_2, \cdots x_k$, $(k = N + n)$, include two parts: the first part is the settings of N product attributes which are denoted as $x_1, x_2, \cdots x_N$, and the second part is the historical sentiment scores of customer preferences in the previous periods which are denoted as

$y(t-n) \ \cdots \ y(t-1)$ where the number of historical periods is n. There are two types of connection which are the separation of inputs and the multiplication, "×". The q_m is explained as follows.

$$q_m = \left[q_m^1, q_m^2, \cdots, q_m^{Ne}\right] \tag{3}$$

where Ne denotes the number of elements in q_m and is an odd number, which is also the length of c_m. The settings of q_m are described in Table 1.

Table 1. The structure of q_m.

Elements	q_m^1	q_m^2	q_m^3	q_m^4	$\cdots$	q_m^{Ne}
Settings	Integer $q_m^1 \in [a,b]$	0 or 1	Integer $q_m^3 \in [a,b]$	0 or 1	$\cdots$	Integer $q_m^{Ne} \in [a,b]$

For the elements at the odd positions, $q_m^1, \ q_m^3, \ \cdots, q_m^{Ne}$, the rounded values indicate the sequence number of inputs. For example, if q_m^3=4, then the input x_4 is used for the 3rd position. For the elements at the even positions, $q_m^2, \ q_m^4, \ \cdots, q_m^{Ne-1}$, the rounded values are either 0 or 1. A value of 0 indicates the separation of inputs, and a value of 1 indicates the multiplication between the two inputs. For example, if the number of elements in q_m is 7 and it is generated as $q_m = [4, \ 0, \ 2, \ 1, \ 3, \ 0, \ 1]$, the nonlinear inputs are determined as x_4, x_2x_3, and x_1. So, there are three input items for ANFIS. Consequently, the fuzzy rules contain these three items. The generated fuzzy rules are in nonlinear structure, and the nonlinearity existing in the modeling is explicit.

In MOCOA, *MRE* (mean relative error) and *VoE* (variance of errors) are the two objective functions used, which are expressed in (4) and (5), respectively.

$$MRE = \frac{1}{nd}\sum_{i=1}^{nd}\frac{|\hat{y}_i - y_i|}{y_i} \tag{4}$$

$$VoE = \frac{1}{nd-1}\sum_{i=1}^{nd}\left(\frac{|\hat{y}_i - y_i|}{y_i} - MRE\right)^2 \tag{5}$$

where the number of the data sets is denoted by nd; the ith actual sentiment score of customer preference in data sets is indicated as y_i; the ith predictive sentiment score based on the generated model is $\hat{y}_i$.

In each iteration of MOCOA, the values of c_m and q_m are updated and the nonlinear inputs of ANFIS are changed. Based on the developed models, the values of *MRE* and the *VoE* are calculated for each iteration. Based on the Pareto dominant theory, the solution with the best values of *MRE* and the *VoE* is selected as the optimal solution. A set for fitness value is expressed as $FV_s = \{FV_1, \ FV_2\}$, where FV_1 and FV_2 represent the *MRE* and *VoE* values, respectively. Solution B dominates solution A in the optimization problem of minimizing if the following two Equations (6) and (7) are satisfied.

$$FV_i(A) \le FV_i(B), \ for \ all \ i \in \{1,2\} \tag{6}$$

$$FV_j(A) < FV_j(B), \ for \ some \ j \in \{1,2\} \tag{7}$$

When the following two conditions are met, solution A dominates solution B in the optimization problem of maximizing.

$$FV_i(A) \ge FV_i(B), \ for \ all \ i \in \{1,2\} \tag{8}$$

$$FV_j(A) > FV_j(B), \ for \ some \ j \in \{1,2\} \tag{9}$$

The Pareto dominant concept is used in MOCOA in order to compare each solution to other solutions according to the above requirements. When there are no other solutions that dominate the optimal solution, this is known as a Pareto optimal solution. The optimal solution denotes the nonlinear inputs of ANFIS.

3.3. Generation of Customer Preference Models Using ANFIS

Essentially, ANFIS is a multilayer feedforward network that combines fuzzy logic with neural networks and converts inputs into an output [36]. A typical ANFIS structure is shown in Figure 2. The system consists of two inputs and one output, where each input has two membership functions. Based on the MOCOA, the optimal solution is obtained as the nonlinear inputs of ANFIS to model customer preference which are denoted as xx_1 and xx_2.

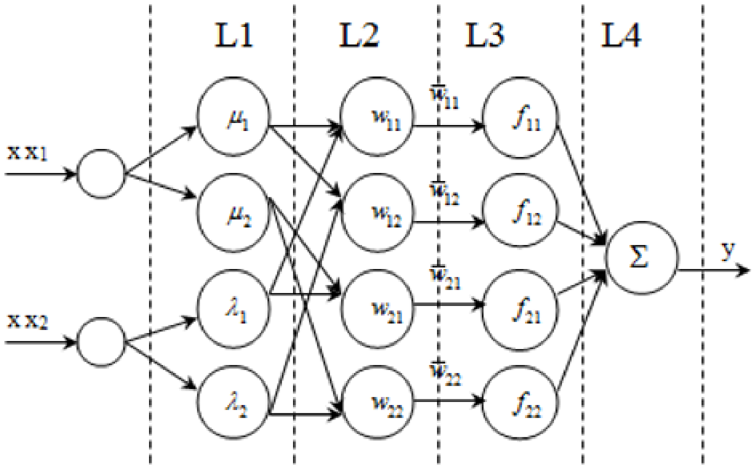

Figure 2. A typical ANFIS architecture.

where μ_i, $i = 1,2$ denote the membership function for the ith linguistic description of xx_1; and λ_j, $j = 1,2$, is the membership function for the jth linguistic description of xx_2. In this paper, the triangular membership function is used as it has the advantage of simplicity by using three points to form a triangle. It also can provide low mean errors and a good fit of data sets in the fuzzy models [37]. The triangular membership function is described as follows.

$$\mu_i(xx_1) = \begin{cases} \frac{xx_1 - a_i}{b_i - a_i} & a_i \le xx_1 \le b_i \\ \frac{c_i - xx_1}{c_i - b_i} & b_i \le xx_1 \le c_i \\ 0 & Otherwise \end{cases} \text{ and } \lambda_j(xx_2) = \begin{cases} \frac{xx_2 - s_j}{t_j - s_j} & s_j \le xx_2 \le t_j \\ \frac{u_j - xx_2}{u_j - t_j} & t_j \le xx_2 \le u_j \\ 0 & Otherwise \end{cases} \quad (10)$$

where (a_i, b_i, c_i) and (s_j, t_j, u_j) are the fuzzy numbers in the triangle shape, which are the antecedent parameters in ANFIS. The subtractive clustering method is applied to determine the clusters' centers and corresponding ranges for the inputs. The determined centers are the values of b_i and t_j. The left-side value of a fuzzy number is calculated as the center value minus the corresponding range, while the right-side value is equal to the center value plus the corresponding range.

As a result of the membership function of the inputs, there are four combinations of them, which are defined as the four nodes in the 2nd layer. The outputs of the 2nd layer are computed as follows, which are called the firing strength:

$$w_{ij} = \mu_i(xx_1)\lambda_j(xx_2) \ (\forall\ i = 1,2, j = 1,2) \quad (11)$$

The weight of the connection between the 2nd and the 3rd layer is the normalized firing strength $\overline{w}_{ij}$ which is also the importance of fuzzy rule R_{ij} in the 3rd layer. It is defined by (12).

$$\overline{w}_{ij} = \frac{w_{ij}}{W} \text{ where } W = \sum_{i}\sum_{j} w_{ij} \ (\forall\, i = 1,2, j = 1,2) \tag{12}$$

If the value of $\overline{w}_{ij}$ is larger, then R_{ij} is more important. Each combination of membership functions of xx_1 with xx_2 is denoted by one fuzzy rule. Therefore, there are a total of four fuzzy rules. The fuzzy rules involve the inputs items and f_{ij} is the internal model in a fuzzy rule which is the nonlinear model.

$$f_{ij} = p_{ij}xx_1 + q_{ij}xx_2 + r_{ij} \ (\forall\, i = 1,2, j = 1,2) \tag{13}$$

By using the least squares method, we can determine the subsequent parameters p_{ij}, q_{ij}, and r_{ij}.

At the 4th layer, a single node denotes the final output which is calculated as follows in (14).

$$\hat{y} = \sum_{i=1}^{2}\sum_{j=1}^{2} O_{ij} = \sum_{i=1}^{2}\sum_{j=1}^{2} \overline{w}_{ij} f_{ij} = \sum_{i=1}^{2}\sum_{j=1}^{2} \overline{w}_{ij}\left(p_{ij}xx_1 + q_{ij}xx_2 + r_{ij}\right) \tag{14}$$

where $\hat{y}$ is the predicted sentiment score of the customer preference based on the models.

3.4. The Computational Process of MOCOA-Based ANFIS

The computational process of the proposed MOCOA-based ANFIS with nonlinear inputs are provided as follows.

Step 1. Consumer products with sufficient online reviews are sampled. The online reviews are collected and divided based on the fixed time periods. Based on the contents of Section 3.1, sentiment analysis is performed in order to compute sentiment scores for customer preferences for the period from which the time series customer preferences can be derived. By using the sampled products, the product attributes settings relating to the customer preferences are obtained. According to the above process, the data sets are ready for modeling, and include the time series sentiment score of customer preferences and the setting of the product attributes.

Step 2. The proposed approach is implemented using MATLAB software. It is first necessary to initialize the parameters of the proposed approach, including the iteration number of MOCOA, the length of the chaos variable c_m, the initial value for c_1, and the values of a and b in (2).

Step 3. The iteration starts from m = 1. By using the initialized chaos variable, the value of c_{m+1} is calculated by using (1) and the corresponding optimization variable q_{m+1} is obtained based on (2). Based on the q_{m+1}, the elements at the odd positions are replaced by the corresponding inputs while the elements at the even positions are substituted by the two types of connection.

Step 4. In order to develop the customer preferences models, ANFIS is used to process the generated nonlinear inputs according to the Formulas (10)~(14). In addition, the nonlinear fuzzy rules are obtained based on (13) and the predicted sentiment scores of customer preferences $\hat{y}_i$ are obtained based on (14). Then, the values of *MRE* and *VoE* are calculated using (4) and (5) for each iteration.

Step 5. By $m + 1 \rightarrow m$, the iteration continues with the same processes repeated from Step 3. In each iteration, the Pareto dominant concept in Section 3.2 is applied to update the optimal solution. Upon reaching the predetermined number of iterations, the iteration is terminated. The optimal solution is found, and the number of iterations associated with it is recorded. Therefore, the optimal solution determines the best nonlinear inputs for ANFIS and the corresponding customer preference models are generated with the explicit nonlinear fuzzy rules.

4. Implementation

The proposed approach is illustrated by an example of a sweeping robot. To serve as our sample products, we selected ten competitive sweeping robots with high sales in the market, denoted by the letters A~J, respectively. The Amazon.com website was used to collect the online comments regarding the ten products. Since all the online comments were collected over a period of approximately two years, a fixed-time-period approach was employed, with a six-month period being applied. In total, four time periods are determined. As a result, the latest time period will be used as the current period, which corresponds to the output of the models y_t. The sentiment scores for the three other time periods, y_{t-3}, y_{t-2}, and y_{t-1}, are applied as inputs for the analysis.

The Semantria Excel plug-in was used to analyze and research online comments stored in Excel files. Based on the process described in Section 3.1, the most frequent words and phrases found in online comments are identified and the synonymous ones are categorized into one group. A name is summarized for each group to indicate the extracted customer opinions for products. For example, the extracted phrases "deep cleaning", "useful", "super-clean", "high performance", and "good cleaner", were categorized as one group, which is named as one customer preference "clean well". Totally, four common consumer preferences can be summarized in this case, namely quality, smart operation, clean well, and sound operation. Afterwards, Semantria is used again to analyze sentiment based on the function of user category analysis for each period. The keywords and phrases relating to consumer preferences are set as the setting of "user category". For the purpose of illustrating the procedure of developing models, the customer preference "clean well" is used in the paper. Table 2 shows the sentiment scores for "clean well" based on sentiment analysis for the four time periods.

Table 2. The sentiment scores of ten products under four time periods.

Products	**Clean Well**			
	Period 1	**Period 2**	**Period 3**	**Period 4**
	y_{t-3}	y_{t-2}	y_{t-1}	y_t
A	0.36	0.43	0.47	0.40
B	0.37	0.29	0.35	0.20
C	0.28	0.32	0.44	0.24
D	0.32	0.33	0.38	0.24
E	0.33	0.33	0.33	0.30
F	0.43	0.40	0.41	0.39
G	0.33	0.31	0.27	0.30
H	0.33	0.32	0.37	0.31
I	0.30	0.31	0.30	0.30
J	0.32	0.32	0.37	0.28

The customer preference "clean well" is associated with four product attributes. They are size, max suction power, dust box capacity, and wet mopping, which are denoted as x_1, x_2, x_3, and x_4, respectively. The units of x_1, x_2, and x_3 are cubic inch, pascal (Pa), and litre (L), respectively. For x_4, a value of 1 indicates that the sweeping robot has the capacity of wet mopping, whereas a value of 0 indicates that it is not capable of doing so. Table 3 presents the settings for the four product attributes of the ten sweeping robots.

By using the data sets in Tables 2 and 3, the MOCOA-based ANFIS approach was applied to develop the associations between y_t and x_1, x_2, x_3, x_4, y_{t-3}, y_{t-2}, y_{t-1}. In the proposed approach, the length of the chaos variable is set as 13; the iteration number of MOCOA is set as 100. For the optimization variable, the ranges of elements at the odd numbers and the even numbers are set as [1, 7] and [0, 1], respectively. The MATLAB software was used to perform the modeling based on the proposed approach. Using validation 1 as an example, the Pareto solutions obtained from the training process are shown in Figure 3.

Table 3. The settings of four product attributes of ten sweep robots.

Products	Attributes of Sweeping Robots			
	Size x_1	Max Suction Power (Pa) x_2	Dust Box Capacity (L) x_3	Wet Mopping x_4
A	438.178	1400	0.5	1
B	490.100	1800	0.6	0
C	417.720	850	0.3	1
D	515.573	2000	0.75	0
E	417.720	1000	0.3	0
F	442.368	1400	0.3	0
G	642.105	1800	0.5	0
H	645.979	1800	0.4	0
I	643.500	2000	0.7	0
J	466.215	1500	0.6	0

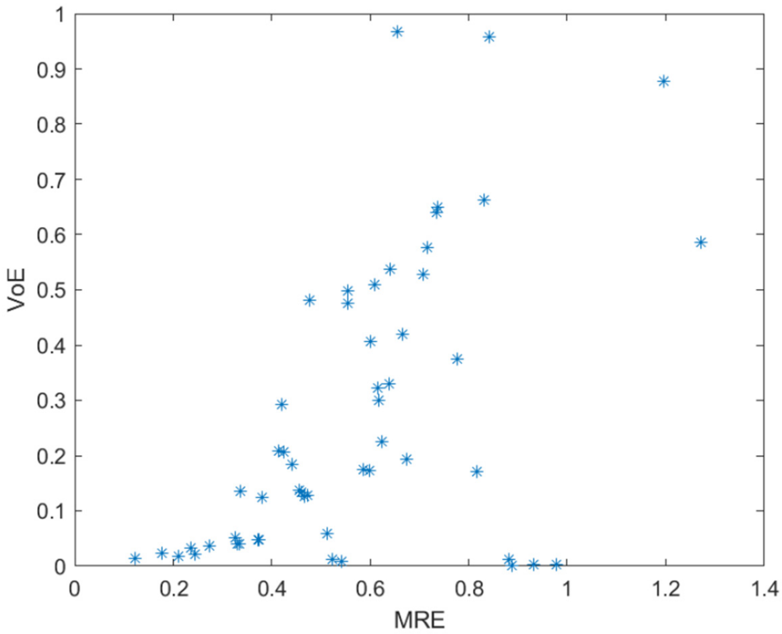

Figure 3. Pareto solutions in validation 1.

Based on the Pareto dominant theory described in Section 3.2, the best solution is selected. In validation 1, the optimal solution for q_m is obtained as follows:

$$q_m = [4,\ 1,\ 7,\ 0,\ 3,\ 1,\ 3,\ 0,\ 5,\ 0,\ 6,\ 1,\ 4]$$

Therefore, the inputs for ANFIS are x_4y_{t-1}, x_3^2, y_{t-3}, and x_4y_{t-2}, which include a single item y_{t-3}, a second-order item x_3^2, and the interactive items x_4y_{t-1} and x_4y_{t-2}. They are denoted as xx_1~xx_4 for the inputs of ANFIS and their membership functions are denoted as μ, λ, σ, and ω, respectively. An example of the generated nonlinear fuzzy rule is expressed as follows.

If xx_1 is μ_1, xx_2 is λ_1, xx_3 is σ_1 and xx_4 is ω_1,

$$THEN\ f_1 = 0.1239x_4y_{t-1} + 0.0253x_3^2 + 0.0789y_{t-3} + 0.0901x_4y_{t-2} + 0.2817$$

5. Validation

The proposed MOCOA-based ANFIS approach was further evaluated by conducting five validation tests using sweeping robots as the case study. In validation tests one to five, the following products were selected as validation data sets: A and B, C and D, E and

F, G and H, as well as I and J, respectively. For the purposes of training, the remaining eight products were used. There was no repetition of data sets during the training and validation process. In each validation, the proposed approach, ANFIS, subtractive-cluster-based ANFIS (SC-ANFIS), fuzzy c-means-based ANFIS (FCM-ANFIS), and K-means-based ANFIS (K-means-ANFIS) are compared. In SC-ANFIS, FCM-ANFIS, and K-means-ANFIS, subtractive cluster, fuzzy c-means, and K-means are introduced as methods for determining the membership function of inputs in ANFIS. Each test used the same data sets for the five approaches.

In ANFIS, the number of membership functions is determined to be 3 for each input, which is a standard setting. If the total number of data sets is n, then the number of clusters is usually $\leq \sqrt{n}$ [38]. Therefore, FCM-ANFIS and K-means-ANFIS were set to have three clusters each. The parameter settings of the proposed approach are provided in Section 4. For the purpose of comparing the validation results resulting from the five approaches, the measurement of *MRE* in (4) and *VoE* in (5) were employed. Table 4 compares the *MRE* and *VoE* values based on the five approaches in the five validations. Using Table 4, we can conclude that the *MRE* and *VoE* values calculated using the proposed approach are significantly lower than those calculated using the other four approaches. Therefore, the proposed approach can better fit the data sets and can model the customer preferences with higher accuracy.

Table 4. The *MRE* and *VoE* values based on the five approaches.

Validation Test		Metrics	ANFIS	SC-ANFIS	FCM-ANFIS	K-Means-ANFIS	MOCOA-Based ANFIS
1	A, B	*MRE*	0.7709	0.4287	0.4275	0.4261	0.0337
		VoE	0.0607	2.9949×10^{-6}	1.3317×10^{-7}	4.3565×10^{-6}	3.1677×10^{-9}
2	C, D	*MRE*	0.9463	0.4194	0.4224	0.4222	0.1597
		VoE	0.0058	0.0270	0.0306	0.0305	4.5578×10^{-5}
3	E, F	*MRE*	0.5888	0.4787	0.4823	0.4829	0.1643
		VoE	0.2218	0.0329	0.0373	0.0377	4.0191×10^{-4}
4	G, H	*MRE*	0.9416	0.0991	0.1101	0.1099	0.0119
		VoE	0.0051	0.0006	0.0006	0.0006	2.3449×10^{-7}
5	I, J	*MRE*	0.8111	0.1245	0.1305	0.1298	0.0455
		VoE	0.0045	0.0290	0.0338	0.0311	7.6151×10^{-5}

The optimal inputs based on the MOCOA-based ANFIS approach are obtained for validation one~five which are shown in Table 5. It can be found that the inputs include single items, interactive items, and terms of second order and/or higher-order terms, which are capable of describing the nonlinearity in the models explicitly.

Table 5. The nonlinear inputs for the five validations.

Validation Test	The Number of Inputs	The Obtained Inputs for the Proposed Approach
1	4	$x_4y_{t-1}, x_3^2, y_{t-3}, x_4y_{t-2}$
2	3	$x_7, x_1y_{t-3}^2y_{t-2}y_{t-1}, x_4$
3	2	$x_2y_{t-3}y_{t-2}y_{t-1}^3, y_{t-1}$
4	4	$x_2y_{t-2}, y_{t-2}, x_1x_4, x_3x_4$
5	3	$y_{t-2}, x_1y_{t-2}y_{t-1}, y_{t-1}^3$

6. Conclusions

Customer preferences are not static and are subject to change over time. However, in previous research on modeling customer preferences, it is assumed that the customer preferences are stable, which affects the modeling accuracy and cannot obtain the changing tendency of the customer preferences. In addition, ANFIS is an effective method to capture the nonlinearity and fuzziness in the modeling. However, ANFIS has back box problems and cannot show the nonlinearity explicitly. Based on the development of explainable artificial intelligence, this paper proposed a MOCOA-based ANFIS approach for modeling customer preferences with explicit nonlinearity based on online reviews. The study presents a case study of sweeping robot products, in which customer preferences are modeled and the example of "clean well" is selected to illustrate the modeling process of the proposed approach. Five validations were conducted, and the proposed approach was compared with ANFIS, SC-ANFIS, FCM-ANFIS, and K-means-ANFIS. According to the comparison, the proposed approach has a smaller *MRE* and *VoE* than the other four approaches, which indicates a higher prediction accuracy. In addition, the nonlinear inputs are determined by using the proposed approach which can display the nonlinearity in the modeling explicitly. Compared with other existing approaches in the literature, the time series values of customer preference are considered in the modeling and the dynamic models are developed based on the proposed approach, which can capture the changes in the customer preferences over time. Based on the developed models, product development companies can predict the future sentiment scores of customer preferences for products by inputting the settings of product attributes and the historical scores of customer preferences, which can help them to investigate the customer's satisfaction with the products in the market and make product adjustments in time. From the obtained nonlinear fuzzy rules, the interaction between the product attributes and their effects on the sentiment scores of the customer preferences can be revealed, which is valuable information for companies to adjust the attributes setting of the existing products or design for new products.

In future research, the optimization of product attributes based on the developed dynamic customer preference model with nonlinear inputs will be performed for product design. On the other hand, the adaptive MOCOA in which the parameters can be set based on the advanced intelligent optimization algorithms will be introduced to enhance the modeling accuracy.

Author Contributions: Conceptualization, H.J.; Data curation, F.S.; Funding acquisition, H.J.; Methodology, H.J.; Project administration, F.S.; Validation, H.J.; Writing—original draft, H.J.; Writing—review & editing, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Faculty Research Grants (FRG) from Macau University of Science and Technology under Grant FRG-23-045-MSB, and in part by the National Natural Science Foundation of China under Grant 71901149.

Data Availability Statement: All datasets generated during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tseng, M.M.; Piller, F.T. *The Customer Centric Enterprise: Advances in Mass Customization and Personalization*; Springer: Berlin/Heidelberg, Germany, 2003.
2. Jiang, H.; Kwong, C.K.; Siu, K.W.M.; Liu, Y. Rough set and PSO-based ANFIS approaches to modeling customer satisfaction for affective product design. *Adv. Eng. Inform.* **2015**, *29*, 727–738. [CrossRef]
3. Safari, A.; Shayeghi, H.; Heidar, A.S. Robust State Feedback Controller Design of STATCOM Using Chaotic Optimization Algorithm. *Serbian J. Electr. Eng.* **2010**, *7*, 253–268. [CrossRef]
4. Cambria, E. Affective Computing and Sentiment Analysis. In *IEEE Intelligent Systems*; IEEE: Piscataway, NJ, USA, 2016; Volume 31, pp. 102–107. [CrossRef]
5. Chen, X.; Chen, C.-H.; Leong, K.F.; Jiang, X. An ontology learning system for customer needs representation in product development. *Int. J. Adv. Manuf. Technol.* **2012**, *67*, 441–453. [CrossRef]
6. Zhou, F.; Jiao, R.J.; Linsey, J.S. Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews. *J. Mech. Des.* **2015**, *137*, 071401. [CrossRef]
7. Tuarob, S.; Tucker, C.S. Automated discovery of lead users and latent product features by mining large scale social media networks. *J. Mech. Des.* **2015**, *137*, 071402. [CrossRef]
8. Tuarob, S.; Tucker, C.S. Quantifying product favorability and extracting notable product features using large scale social media data. *J. Comput. Inf. Sci. Eng.* **2015**, *15*, 031003. [CrossRef]

9. Zimmermann, M.; Ntoutsi, E.; Spiliopoulou, M. Discovering and monitoring product features and the opinions on them with OPINSTREAM. *Neurocomputing* **2015**, *150*, 318–330. [CrossRef]
10. Lim, S.; Tucker, C.S. A Bayesian sampling method for product feature extraction from large-scale textual data. *J. Mech. Des.* **2016**, *138*, 061403. [CrossRef]
11. Zhang, H.; Sekhari, A.; Ouzrout, Y.; Bouras, A. Jointly identifying opinion mining elements and fuzzy measurement of opinion intensity to analyze product features. *Eng. Appl. Artif. Intell.* **2016**, *47*, 122–139. [CrossRef]
12. Kang, Y.; Zhou, L. RubE: Rule-based methods for extracting product features from online consumer reviews. *Inf. Manag.* **2017**, *54*, 166–176. [CrossRef]
13. Zhou, F.; Jiao, R.J.; Yang, X.J.; Lei, B. Augmenting feature model through customer preference mining by hybrid sentiment analysis. *Expert Syst. Appl.* **2017**, *89*, 306–317. [CrossRef]
14. Chiu, M.C.; Lin, K.Z. Utilizing text mining and Kansei Engineering to support data-driven de-sign automation at conceptual design stage. *Adv. Eng. Inform.* **2018**, *38*, 826–839. [CrossRef]
15. Trappey, A.J.; Trappey, C.V.; Fan, C.-Y.; Lee, I.J. Consumer driven product technology function deployment using social media and patent mining. *Adv. Eng. Inform.* **2018**, *36*, 120–129. [CrossRef]
16. Zhang, L.; Chu, X.; Xue, D. Identification of the to-be-improved product features based on online reviews for product redesign. *Int. J. Prod. Res.* **2018**, *57*, 2464–2479. [CrossRef]
17. Ali, M.M.; Doumbouya, M.B.; Louge, T.; Rai, R.; Karray, M.H. Ontology-based approach to extract product's design features from online customers' reviews. *Comput. Ind.* **2020**, *116*, 103175. [CrossRef]
18. Li, Z.; Tian, Z.G.; Wang, J.W.; Wang, W.M. Extraction of affective responses from customer reviews: An opinion mining and machine learning approach. *Int. J. Comput. Integr. Manuf.* **2020**, *33*, 670–685. [CrossRef]
19. Jin, J.; Jia, D.; Chen, K. Mining online reviews with a Kansei-integrated Kano model for innovative product design. *Int. J. Prod. Res.* **2022**, *60*, 6708–6727. [CrossRef]
20. Chung, W.; Tseng, T.-L. Discovering Business Intelligence from Online Product Reviews: A Rule-Induction Framework. *Expert Syst. Appl.* **2012**, *39*, 11870–11879. [CrossRef]
21. Jiang, H.; Kwong, C.K.; Park, W.Y.; Yu, K.M. A multi-objective PSO approach of mining association rules for affective design based on online customer reviews. *J. Eng. Des.* **2018**, *29*, 381–403. [CrossRef]
22. You, H.; Ryu, T.; Oh, K.; Yun, M.-H.; Kim, K.-J. Development of customer satisfaction models for automotive interior materials. *Int. J. Ind. Ergon.* **2006**, *36*, 323–330. [CrossRef]
23. Chen, C.H.; Khoo, L.P.; Yan, W. An investigation into affective design using sorting technique and Kohonen self-organising map. *Adv. Eng. Softw.* **2006**, *37*, 334–349. [CrossRef]
24. Yang, J.-B.; Wang, Y.-M.; Xu, D.-L.; Chin, K.-S.; Chatton, L. Belief rule-based methodology for mapping consumer preferences and setting product targets. *Expert Syst. Appl.* **2012**, *39*, 4749–4759. [CrossRef]
25. Fung, R.Y.; Law, D.S.; Ip, W. Design targets determination for inter-dependent product attributes in QFD using fuzzy inference. *Integr. Manuf. Syst.* **1999**, *10*, 376–384. [CrossRef]
26. Park, J.; Han, S.H. A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design. *Int. J. Ind. Ergon.* **2004**, *34*, 31–47. [CrossRef]
27. Chen, Y.; Chen, L. A non-linear possibilistic regression approach to model functional relationships in product planning. *Int. J. Adv. Manuf. Technol.* **2006**, *28*, 1175–1181. [CrossRef]
28. Şekkeli, G.; Köksal, G.; Batmaz, I.; Bayrak, O.T. Classification models based on Tanaka's fuzzy linear regression approach: The case of customer satisfaction modeling. *J. Intell. Fuzzy Syst.* **2010**, *21*, 341–351. [CrossRef]
29. Chan, K.Y.; Kwong, C.K.; Dillon, T.S.; Fung, K.Y. An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness. *J. Eng. Des.* **2011**, *22*, 523–542. [CrossRef]
30. Chan, K.Y.; Lam, H.K.; Dillon, T.S.; Ling, S.H. A stepwise-based fuzzy regression procedure for developing customer preference models in new product development. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 1728–1745. [CrossRef]
31. Chan, K.Y.; Ling, S.H. A forward selection based fuzzy regression for new product development that correlates engineering characteristics with consumer preferences. *J. Intell. Fuzzy Syst.* **2016**, *30*, 1869–1880. [CrossRef]
32. Jiang, H.; Kwong, C.K.; Ip, W.H.; Chen, Z. Chaos-based fuzzy regression approach to modeling customer satisfaction for product design. *IEEE Trans. Fuzzy Syst.* **2013**, *21*, 926–936. [CrossRef]
33. Akopov, A.S.; Beklaryan, A.L.; Zhukova, A.A. Optimization of Characteristics for a Stochastic Agent-Based Model of Goods Exchange with the Use of Parallel Hybrid Genetic Algorithm. *Cybern. Inf. Technol.* **2023**, *23*, 87–104. [CrossRef]
34. Terano, T.; Naitoh, K. Agent-based modeling for competing firms: From balanced-scorecards to multiobjective strategies. In Proceedings of the 37th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 5–8 January 2004; p. 8.
35. Pereira, F.; Costa, J.M.; Ramos, R.; Raimundo, A. The impact of the COVID-19 pandemic on airlines' passenger satisfaction. *J. Air Transp. Manag.* **2023**, *112*, 102441. [CrossRef] [PubMed]
36. Wang, Y.M.; Elhag, T.M.S. An adaptive neuro-fuzzy inference system for bridge risk assessment. *Expert Syst. Appl.* **2008**, *34*, 3099–3106. [CrossRef]

37. Jaouachi, B. Evaluation of the Residual Bagging Height using the Regression Technique and Fuzzy Theory. *Fibres Text. East. Eur.* **2013**, *21*, 92–98.
38. Bari, A.; Rueda, L. A New Profile Alignment Method for Clustering Gene Expression Data. In Proceedings of the Conference of the Canadian Society for Computational Studies of Intelligence, Québec City, QC, Canada, 7–9 June 2006; pp. 86–97.