https://doi.org/10.1016/j.aei.2019.100969

Dynamic modelling of customer preferences for product design using DENFIS and opinion mining

Huimin Jiang[a,b], C.K. Kwong[b,*], G.E. Okudan Kremer[c], W.Y. Park[b]

[a]*College of Management, Shenzhen University, Shenzhen, China*

[b]*Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

[c]*Department of Industrial and Manufacturing Systems Engineering, Iowa State University, IA 50011, USA*

Abstract

Previous studies mainly employed customer surveys to collect survey data for understanding customer preferences on products and developing customer preference models. In reality, customer preferences on products could change over time. Thus, the time series data of customer preferences under different time periods should be collected for the modelling of customer preferences. However, it is difficult to obtain the time series data based on customer surveys because of long survey time and substantial resources involved. In recent years, a large number of online customer reviews of products can be found on various websites, from which the time series data of customer preferences can be extracted easily. Some previous studies have attempted to analyse customer preferences on products based on online customer reviews. However, two issues were not addressed in previous studies which are the fuzziness of the sentiment expressed by customers existing in online reviews and the modelling of customer preferences based on the time series data obtained from online reviews. In this paper, a new methodology for dynamic modelling of customer preferences based on online customer reviews is proposed to address the two issues which mainly involves opinion mining and dynamic evolving neural-fuzzy inference system (DENFIS). Opinion mining is adopted to analyze online reviews and perform sentiment analysis on the reviews under different time periods. With the mined time series data and the product attribute settings of reviewed products, a DENFIS approach is introduced to perform the dynamic modelling of customer preferences. A case study is used to illustrate the proposed methodology. The results of validation tests indicate that the proposed DENFIS approach outperforms various adaptive neuro-fuzzy

* Corresponding author.
E-mail addresses: c.k.kwong@polyu.edu.hk (C.K. Kwong).

inference system (ANFIS) approaches in the dynamic modelling of customer preferences in terms of the mean relative error and variance of errors. In addition, the proposed DENFIS approach can provide both crisp and fuzzy outputs that cannot be realized by using existing ANFIS and conventional DENFIS approaches.

Keywords: Customer preference; Opinion mining; DENFIS; New product development.

1. Introduction

Customer centric product design features with the incorporation of customer needs/preferences into the product design stage such that customer needs/preferences can be translated into a concrete product specification [1]. Various methods and processes have been developed to support customer centric product design. Among them, quality function deployment is the most well-known one for facilitating customer centric product design and has been implemented by numerous companies worldwide [2]. In industries, companies commonly adopt customer surveys to collect and understand customer preferences on products. Since customer preferences could change with time rapidly, companies need to predict customer preferences at the time of new product launch. Suppose a company plans to conduct a customer survey in the early product design stage. Since there exists a time interval, properly six months or more, between the early product design stage and new product launch. The customer preferences obtained from the survey could be quite different from those at the time of new product launch. Therefore, the company has to predict the customer preferences at the time of new product launch which would help her to determine a proper attribute setting of a new products in the product design stage. Apart from the collection and analysis of customer preferences, in the early product design stage, it is also important to understand the relationships between customer preferences on products and product attribute settings that involves the modelling of customer preferences. Developed customer preference models can help companies to determine the proper product attribute settings of their new products quickly. Currently, conduct of customer surveys based on questionnaires and/or interviews is still a common way for companies to obtain survey data, which are then utilized to analyze customer preferences on products and develop customer preference models. However, it is well recognized that customer preferences on products would change over time especially for consumer products. If a customer preference model is developed based on the survey data obtained from a particular time period only, two limitations of the model can be noted. First, the model is incapable of predicting customer preferences of a product for the next time period. Second, the prediction accuracy of the model may not be acceptable. Therefore, time series data of customer preferences should be collected for developing customer preference models. However, the collection of time series data based on customer surveys is difficult as a number of surveys have to be performed under different time periods that would lead to long survey time and substantial resources required especially when interviews are involved. Previous research also commonly adopted customer surveys to obtain survey data for analyzing customer preferences on products and developing customer preference models. Thus, previous studies have the similar problem with the industrial practice where the time series data of customer preferences on products are always unavailable for the modelling of customer preferences.

Nowadays, numerous online customer reviews of products can be found easily from various websites. The reviews are always subjective which are written based on customers' experience of purchasing and/or using products and contain rich information regarding their views and opinions toward products. They could affect potential customers in their purchases of products to a large extent. In fact, customers' views and opinions of products extracted from online reviews can provide a valuable source of information for companies to develop new products. On the other hand, the time series data of customer preferences can be obtained easily from online reviews and nearly without cost as no surveys are required to be conducted. However, a proper modeling approach needs to be introduced in order to develop customer preference models based on the time series data.

In recent years, quite a number of studies have been performed about data and information mining from online reviews for supporting product design. Some previous studies have attempted to extract customer preferences, customer needs and product attributes based on online customer reviews as well as determine their importance and ranking. Some other studies mainly focused on establishing the relationships between customer preferences/satisfaction and product attributes by using rule mining. However, some limitations of previous studies were noted. First, the customers' sentimental expression of online reviews quite often involves fuzziness, but it was not addressed properly in previous studies. Second, although some previous studies have attempted to generate if-then rules for relating customer preferences/satisfaction and product attributes, the rules generated are quite often insufficient to be used in determining the product attribute settings of new products. Finally, the modelling of the relationships between customer preferences and product attributes based on online reviews and time series data were not found in previous studies.

To fill the research gaps, a new methodology for dynamic modelling of customer preferences based on online customer reviews is proposed in this paper, which mainly involves opinion mining from online customer reviews and dynamic modelling of customer preferences using a dynamic evolving neural-fuzzy inference system (DENFIS) approach. DENFIS was proposed by Kasabov and Song [3] based on the Takagi-Sugeno fuzzy inference system for developing a connectionist model in evolving connectionist system architecture. Their study showed that DENFIS could effectively learn complex temporal sequences in an adaptive way and outperformed evolving fuzzy neural network, evolving self-organizing maps, ANFIS and multilayer perceptrons in time series prediction. In the proposed methodology, the online reviews of selected products are first collected under different time periods and their sentiment scores are computed. The sentiment scores are then converted into fuzzy numbers, and DENFIS is introduced to model the relationship between product attributes and customer preferences based on the time series fuzzy data. The models developed based on the proposed methodology can be employed to help determine the optimal product attribute setting of new products.

This paper is organized as follows: Section 2 presents the related works. Section 3 describes the proposed methodology for dynamic modelling of customer preferences based on online customer reviews. Section 4 describes the implementation of the proposed methodology by using a case study. Section 5 presents the validation tests and their results regarding the proposed DENFIS approach for dynamic modeling. Finally, discussion and conclusion are presented in Section 6 and 7, respectively.

2. Related works

Opinion mining is to identify opinion bearing words/phrases from free texts and determine sentiment polarity toward a subject or topic [4], also known as sentiment analysis. Sentiment analysis includes the analysis of the emotions and sentiments of customers and can be used to reveal the customer preferences on various product features [5]. Numerous research works of sentiment analysis have been conducted. Recently, Ma et al. [6] augmented the long short-term memory network with a hierarchical attention mechanism consisting of a target level attention and a sentence-level attention in aspect-based sentiment analysis. Li et al. [7] studied word presentations for sentiment analysis by exploiting prior knowledge. Jasti and Mahalakshmi [8] reviewed various algorithms for sentiment analysis and machine learning algorithms as well as pre-processing techniques that make the data ready for sentiment analysis. In recent years, deep learning approaches have been introduced to sentiment analysis. Do et al. [9] have provided a comparative review of deep learning approaches for aspect-based sentiment analysis.

Various approaches for extracting customer preferences and product attributes based on opinion mining from online reviews were found. Lee [10] proposed a supervised machine learning approach for identifying customer preferences from online reviews. Wang et al. [11] proposed a systematic methodology for extracting product attributes from online reviews and developing customer preferences models using Bayesian linear regression. Chen et al. [12] proposed an ontology-learning customer needs representation system for extracting customer preferences and their study showed that more accurate statements of customer preferences can be generated from the proposed system. A framework for discovering implicit product features and monitoring customers' attitude toward these features from customer reviews on different products was proposed [13]. Zhou et al. [14] proposed a two-layer model which combined opinion mining and use case analogical reasoning approach to extract latent customer needs. An opinion mining extraction algorithm was proposed by using fuzzy logic to jointly identify feature, opinion expressions and feature-opinion [15]. Zhou et al. [16] proposed a domain independent opinion mining method to mine customer preference information from online reviews for augmenting a feature model. Kang and Zhou [17] proposed Rube-unsupervised rule-based extraction methods to extract both subjective and objective features from online reviews. Chiu and Lin [18] developed a case-based method to analyze online customer reviews and extracted customer preferences using an integration of text mining and Kansei engineering. Trappey et al. [19] explored the identification of real-time customer needs based on online customer reviews. Some studies were attempted to mine opinions from social media for product design. Tuarob and Tucker [20] employed information retrieval techniques to mine notable product features and analyzed the sentiment to predict product sales and product longevity from tweets. An automated approach was proposed to identify latent product features from social media networks [21]. Tuarob and Tucker [22] proposed a data mining driven methodology to extract product features and the corresponding customers' opinions from large scale social media data. Lim and Tucker [23] proposed a Bayesian sampling method for the extraction of product features from social media data to minimize the extraction errors. Some approaches were attempted to prioritize product features based on opinion mining in previous studies. Jin et al. [24] employed a supervised learning routine to identify product features from online product reviews and an ordinal classification algorithm to rank product features. Rai [25] extracted customer preferences and their importance levels from online reviews using an unsupervised text mining approach. Yang et al. [26] combined local context information and global context information for the extraction and ranking of features based on feature scores and frequencies.

Customer preference models are developed to relate customer preferences and design/product attributes by using various modelling approaches. Developed customer preference models can be used to predict customer preferences of new products. They can also be employed to formulate an optimization model with the objective of maximizing overall customer preference. By solving the optimization model using proper solving algorithms, the optimal product attribute setting of a new product can be determined. Since the relationships between customer preferences and design/product attributes could be highly complex and nonlinear, no theoetical models were developed thus far which are able to model the complex relationships [36]. Therefore, in academic arena, an empirical approach is always adopted to model the customer preferences. Previous studies applied various statistical techniques to model customer preferences such as partial least squares analysis [27] and statistical linear regression [28]. Yang et al. [29] developed a belief rule-based methodology to model customer preferences by which settings of design attributes can be determined. Artificial neural networks [30] were also employed to model the relationships between design attributes and customer preferences. However, the above approaches cannot address the fuzziness in the modelling, which is mainly caused by the subjective judgments of respondents in customer surveys. To address this issue, quite a few fuzzy approaches have been employed including fuzzy inference methods [31], fuzzy rule-based systems [32], nonlinear programming based fuzzy regression [33] and fuzzy linear regression [34]. Recently, some polynomial modelling approaches based on fuzzy regression have been developed to address both fuzziness and nonlinearity in the generation of customer preference models such as genetic programing-based fuzzy regression [35], chaos-based fuzzy regression [36], a stepwise-based fuzzy regression [37], and a forward selection-based fuzzy regression [38]. However, previous studies of modelling customer preferences were conducted based on survey data and no studies were found thus far about the modelling of customer preferences based on the time series data obtained from online customer reviews under different time periods.

Customer preferences on products would vary over time. Further, a time lag always exists between the analysis of customer preferences for new product development and the launch of a new product. Quite a few previous studies have been conducted for predicting future customer preferences based on survey data. Shen et al. [39] analyzed the trend of importance of customer preferences using a fuzzy trend analysis in quality function deployment. A double exponential smoothing technique was applied to predict importance of customer preferences in the future [40]. Wu et al. [41] presented a grey theory model to analyze the dynamics of customer requirements, while Chong and Chen [42] developed a customer requirements analysis and forecast system based on an artificial immune and neural system approach to predict dynamic customer preferences. Huang et al. [43] adopted an artificial immune system to optimize the parameters of support vector machine approach for predicting the future importance of customer preferences. Regarding the prediction of future customer preferences based on online reviews, limited studies were found. Jiang et al. [44] proposed a fuzzy time series approach to predict the future importance of customer preferences based on online customer reviews. Thus far, only a limited number of studies were found about establishing the relationships between customer preferences and product attributes based on online reviews. Chung and Tseng [45] developed a rule-induction framework by which If-Then rules could be generated to relate customer preferences and product attributes based on online product reviews. Jiang et al. [46] adopted a multi-objective PSO approach to generate association rules for relating customer preferences of product affection and design attributes of products.

3. Proposed methodology

After reviewing the related works, it was noted that quite a number of research works have been attempted in recent years to extract customer preferences and product attributes, determine their rankings, and relate customer preferences with product attributes based on online customer reviews. However, it was also noted that no previous studies were found thus far regarding the development of customer preference models based on time series data obtained from online customer reviews, which can be used to help determine the product attribute settings of new products. To address this research problem and the fuzziness of sentiment expressed by customers in online customer reviews, a methodology for developing dynamic customer preference models based on online reviews is proposed in this paper which mainly involves opinion mining from online customer reviews, and the dynamic modelling of customer preferences using a DENFIS approach. Fig.1 shows a flowchart of the proposed methodology and details of the methodology are described in the following sub-sections.

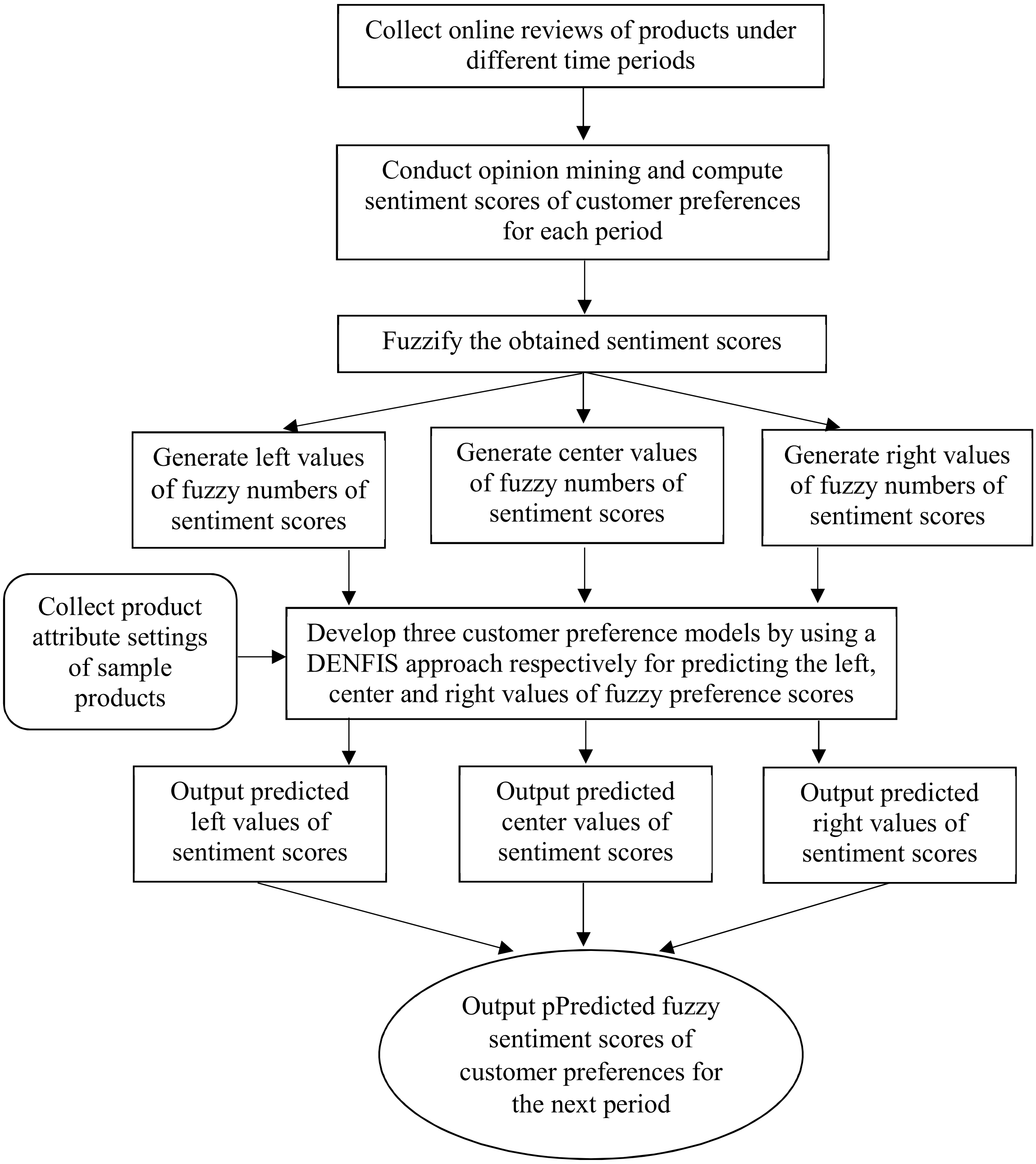


Fig. 1. Proposed methodology.

3.1. Opinion mining from online customer reviews

First, sample products need to be identified. The online reviews of the sample products are found from the website and a web scraping software is used to collect the contents. Based on a fixed-time-period strategy, the collected online reviews were divided according to time period and are stored in different Excel files. Opinion mining is then employed to conduct sentiment analysis to extract customer preferences of the sample products and compute the corresponding sentiment scores.

The process of opinion mining adopted in this study mainly involves the following six steps. First, the pre-processing of the unstructured texts is conducted to remove noise such as html characters, punctuation and stop words. Second, Part-of-Speech (POS) tagging is applied to tag the opinion words in the text as either nouns, adverbs, verbs or adjectives. In this study, nouns are considered as customer preferences, whereas adverbs and adjectives are treated as the corresponding sentiment expressions. Third, relevant sentiment expressions of customer preferences are extracted form online reviews. Fourth, incorrect features and redundant features are eliminated using feature pruning. Fifth, a K-means clustering method is used to cluster the synonymous phrases into the same group. For example, the phrases "fast drying," "dry in short time," and "dry quickly" are grouped under the category "drying time", which is one of the customer preferences on hair dryers. Finally, sentiment analysis is conducted and the semantic polarity and sentiment scores of opinion words for individual customer preferences are determined using SentiWordNet [47]. SentiWordNet is a lexical resource for opinion mining from which a set of three scores representing the notions of positivity, negativity and neutrality are assigned to each synset. The overall sentiment score of customer preference is then computed by using the three scores for each opinion word. In this study, Semantria was adopted to conduct opinion mining from online reviews, which provides an automated sentiment analysis [48]. Semantria is a well-known text analysis software tool which provides Excel add-in to analyse the text in Excel spreadsheets, extract sentiment according to positive, neutral, and negative sentiments as well as compute the corresponding sentiment scores.

3.2. Fuzzification of the obtained sentiment scores

Before the fuzzification of the obtained sentiment scores of customer preferences, outliers are detected and removed from the data sets. For each period, the average value, Ave_i, and standard deviation, σ_i, of the sentiment scores are calculated using equations (1) and (2), respectively.

$$Ave_i = \frac{1}{N}\sum_{j=1}^{N} SS_{ij} \tag{1}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(SS_{ij} - Ave_i)^2} \tag{2}$$

where N is the number of sentiment scores and SS_{ij} is the *jth* sentiment score in the *ith* period. The sentiment score which is beyond the range of [Ave_i -3σ_i, Ave_i +3σ_i] is defined as an outlier.

After the elimination of outliers, the sentiment scores of the *ith* period are transformed into fuzzy numbers and are expressed as (L_i, C_i, R_i), where C_i, L_i and R_i are the central point, left value and right value of the fuzzy number, respectively. They are determined as follows:

$$C_i = Ave_i \tag{3}$$

$$L_i = \underset{j=1,2,\cdots N}{Min} SS_{ij} \tag{4}$$

$$R_i = \underset{j=1,2,\cdots N}{Max} SS_{ij} \tag{5}$$

where Min and Max are the minimum and maximum value of SS_{ij}, $j = 1,2,\cdots N$, respectively. In this research, the obtained fuzzy numbers of the sentiment scores are named as fuzzy preference scores.

3.3. DENFIS approach

Based on the generated fuzzy preference scores and the product attribute settings of products, a DENFIS approach is proposed to dynamic modelling of customer preferences. In the proposed methodology, three DENFIS models are developed respectively for predicting the left, center, and the right values of the fuzzy preference scores. In each DENFIS model, an evolving clustering method (ECM) is applied to partition the input into clusters and the cluster centers are used to form the antecedent of the new fuzzy rule. For the consequent of each fuzzy rule, a first-order linear model is developed based on a weighted recursive least square estimation method. Based on DENFIS, dynamic customer preference models are generated which can be updated with respect to the newly mined data from online reviews. Fig. 2 shows an architecture of a DENFIS where $x_1 \sim x_k$ denote the setting of k product attributes of the products. Using the DENFIS models developed for predicting the left values of fuzzy preference scores as an example, *y(t-1) ~ y(t-T)* represent the left values of fuzzy numbers in the past periods $t\text{-}1\sim t\text{-}T$, which are 1~T periods before the future period t, respectively. $y(t)$ is the corresponding predicted left value of the fuzzy preference score in the future period t.

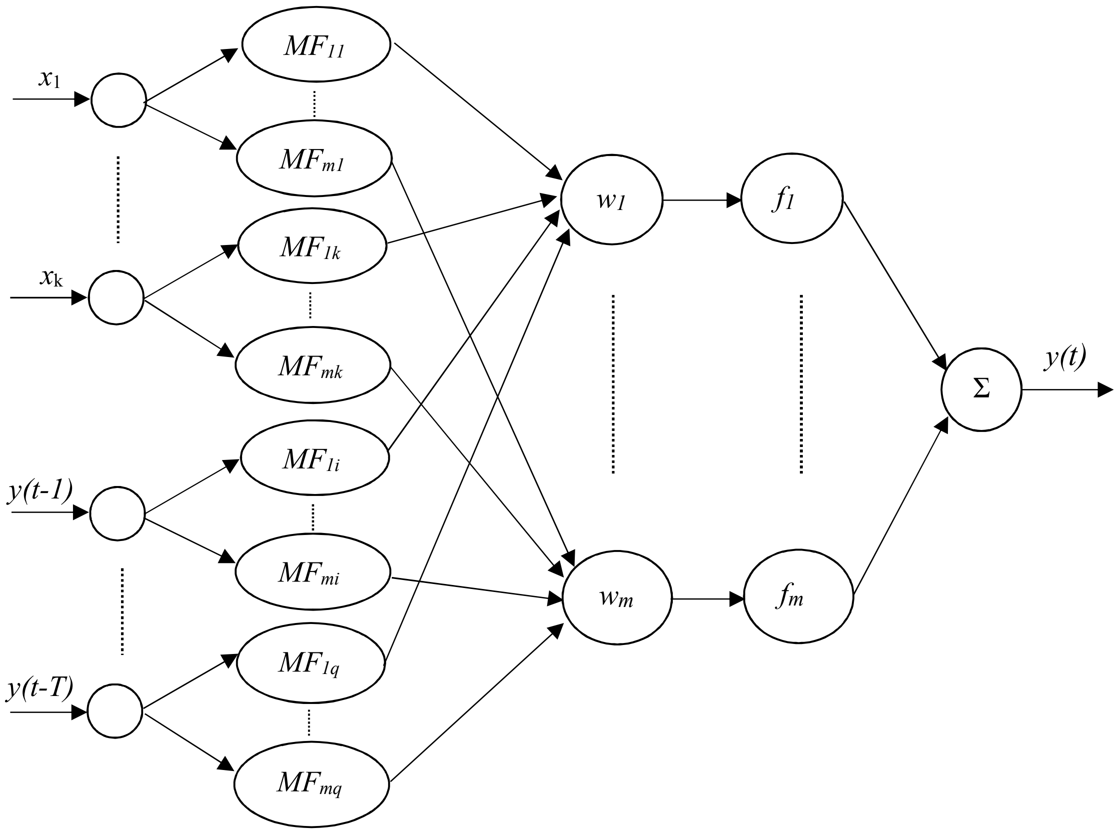


Fig.2 Architecture of DENFIS

3.3.1. Evolving clustering method (ECM)

ECM is a distance-based connectionist clustering method to dynamically estimate the number of clusters and their corresponding centres in the input data space. The input data sets are partitioned into clusters which are determined by their centers and radii. The number of created clusters is self-determined based on a threshold value, D_{thr}, which is a constraint for updating the radii of the clusters and controls the maximum distance between a data point and the cluster center. In the process of clustering, the first cluster, C_1, is first initialized. Its center, Cc_1, and radius, Ru_1, are set as the first data set and zero, respectively. For each input, when the new data Z_i, $i=2,\cdots,n$, is presented, the distances, D_{ij}, from Z_i to the existing clusters, C_j, $j=1,2,\cdots,m$, are computed using (6).

$$D_{ij}=\left\|Z_i-Cc_j\right\|, j=1,2,\cdots,m \tag{6}$$

where $\| \ \|$ denotes the Euclidean distance between Z_i and C_j ; Cc_j is the cluster center of C_j ; n is the number of data sets and m is the number of existing clusters. The minimum distance is defined as follows:

$$D_{i\min} = min(D_{ij}) = \|Z_i - Cc_{\min}\| \tag{7}$$

The cluster with $D_{i\min}$ is denoted as $C_{\min}$ and the corresponding center and radius are defined as $Cc_{\min}$ and $Ru_{\min}$, respectively. If $D_{i\min}$ is equal to, or less than $Ru_{\min}$, $D_{i\min} \le Ru_{\min}$, Z_i belongs to the cluster $C_{\min}$. Otherwise, the existing cluster is updated, or a new cluster is created. The value V_{ij} is calculated by (8) and the minimum value of V_{ij} , V_{ia} , is then determined by (9).

$$V_{ij} = D_{ij} + Ru_j, j = 1, 2, \cdots, m \tag{8}$$

$$V_{ia} = min(V_{ij}) = D_{ia} + Ru_a \tag{9}$$

The cluster with V_{ia} is denoted as C_a and the corresponding cluster center and radius are defined as Cc_a and Ru_a , respectively. D_{ia} is the distance between Z_i and C_a . If $V_{ia} > 2 \times D_{thr}$, Z_i does not belong to any existing cluster and a new cluster is created. Z_i and zero are set as the center and radius of the new cluster, respectively. If $V_{ia} \le 2 \times D_{thr}$, the cluster C_a is updated. Its radius Ru_a is updated as Ru_a' , which is set as $V_{ia}/2$. Its center Cc_a is replaced by the new center, Cc_a' , which is located at a point on the line connecting Z_i and Cc_a and the distance from Cc_a' to Z_i is equal to Ru_a' . If all data sets are processed, the algorithm of ECM is finished.

3.3.2. Learning process in DENFIS

In this study, the inputs include the product attribute settings of sample products, $x_1 \sim x_k$, and the fuzzy preference scores in the past time periods, $y(t-1) \sim y(t-T)$. $x_i, i = 1, 2, \cdots, q$, is used to denote the *ith* input of DENFIS and the number of inputs is q, $q=k+T$. x_i is equal to $x_1 \sim x_k$

when $i=1,2,\cdots,k$ and is equal to *y(t-1) ~ y(t-T)* when $i=k+1,\cdots,q$. Based on the clusters of the input, a set of fuzzy rules can be generated, which are expressed as follows:

$$
\begin{gathered}
If\ x_1\ is\ MF_{11},\ x_2\ is\ MF_{12},\cdots,\ and\ \ x_q\ is\ MF_{1q},then\ y\ is\ f_1(x_1,x_2,\cdots,x_q) \\
If\ x_1\ is\ MF_{21},\ x_2\ is\ MF_{22},\cdots,\ and\ \ x_q\ is\ MF_{2q},then\ y\ is\ f_2(x_1,x_2,\cdots,x_q) \\
\vdots \\
If\ x_1\ is\ MF_{m1},\ x_2\ is\ MF_{m2},\cdots,\ and\ \ x_q\ is\ MF_{mq},then\ y\ is\ f_m(x_1,x_2,\cdots,x_q)
\end{gathered}
\tag{10}
$$

where $MF_{ji}, j=1,2,\cdots,m, i=1,2,\cdots,q$, denotes the *jth* membership function of x_i and the number of membership function is *m*, which 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; $f_j(x_1,x_2,\cdots,x_q),\ j=1,2,\cdots,m$ are the first-order Sugeno fuzzy models in the consequent parts of the fuzzy rules; and *y* is the output of the fuzzy rule. In this study, triangular-shaped membership functions defined below are used.

$$
\mu_j(x_i)=\begin{cases}
0, & x_i<a_j \\
\dfrac{x_i-a_j}{b_j-a_j}, & a_j\le x_i\le b_j \\
\dfrac{c_j-x_i}{c_j-b_j}, & b_j\le x_i\le c_j \\
0, & x_i>c_j
\end{cases}
\tag{11}
$$

where b_j is the center value of the *jth* cluster; $a_j=b_j-d\times D_{thr}$ and $c_j=b_j+d\times D_{thr}$, $1.2\le d\le 2$ are the left and right value of the membership function, respectively;

For the consequent part of each fuzzy rule, a first-order internal model is developed based on a weighted recursive least square estimation method. Each of the internal models is expressed as follows:

$$
y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_qx_q \tag{12}
$$

where *y* is a predicted preference score; x_1 to x_q are product attributes and β_1 to β_q are regression coefficients.

The data sets with *n* data pairs, $\left\{\left[x_1^l, x_2^l, \cdots, x_q^l\right], y_l\right\}, l = 1, 2, \cdots, n$, are used to obtain the regression coefficients, $\beta = \left[\beta_0, \beta_1, \beta_2, \cdots, \beta_q\right]^T$. $x_1^l, x_2^l, \cdots, x_q^l$ and y_l are the inputs and actual output of the *lth* data pair, respectively. The initial inverse matrix P and coefficients β are calculated based on the weighted least square estimator using equations (13) and (14), respectively.

$$P = (A^T W A)^{-1} \tag{13}$$

$$\beta = P A^T W Y \tag{14}$$

where

$$A = \begin{pmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_q^1 \\ 1 & x_1^2 & x_2^2 & \cdots & x_q^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^n & x_2^n & \cdots & x_q^n \end{pmatrix}$$

$$W = \begin{pmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & W_n \end{pmatrix}$$

$$Y = \left[y_1, y_2, \cdots, y_n\right]^T$$

where $()^{-1}$ and $[\ ]^T$ denote the inverse and transpose of the matrix, respectively. W_l can be determined by using (15).

$$W_l = 1 - Dis_l, \ l = 1, 2, \cdots, n \tag{15}$$

where Dis_l is a normalized Euclidean distance between the *lth* data set and the current cluster center that can be defined as $\left(\sum_{i=1}^{q}\left|x_i - b_{ji}\right|^2\right)^{1/2} / q^{1/2}$, where b_{ji} is the *ith* value in the *jth* cluster center.

Based on the initialized P and β, when the new data set is fed, the inverse matrix P_{l+1} and coefficients β_{l+1} at the *(l+1)th* iteration are updated as follows.

$$\beta_{l+1}=\beta_l+W_{l+1}P_{l+1}\alpha_{l+1}(y_{l+1}-\alpha_{l+1}^T\beta_l) \tag{16}$$

$$P_{l+1}=\frac{1}{\lambda}(P_l-\frac{W_{l+1}P_l\alpha_{l+1}\alpha_{l+1}^T P_l}{\lambda+\alpha_{l+1}^T P_l\alpha_{l+1}}) \tag{17}$$

where α_{l+1}^T is the *(l+1)th* row vector of matrix A and $\alpha_{l+1}^T=\begin{bmatrix}1 & x_1^{l+1} & x_2^{l+1} & \cdots & x_q^{l+1}\end{bmatrix}$; y_{l+1} is *(l+1)th* element of Y ; and λ is a forgetting factor with $0<\lambda\leq 1$.

Based on the above learning process, the *lth* predicted output of DENFIS is calculated which is the weighted average of each rule's output.

$$y_l^{'}(t)=\frac{\sum_{j=1}^m w_j f_j(x_1^l,x_2^l,\cdots,x_q^l)}{\sum_{j=1}^m w_j} \tag{18}$$

$$w_j=\sum_{i=1}^q \mu_j(x_i^l) \tag{19}$$

where $\mu_j(x_i^l)$ is determined by (11).

3.4. Computational procedures

The computational procedures of the proposed methodology for dynamic modelling of customer preference based on online customer reviews are described as follows:

Step 1: Online reviews of the sample products are collected under the pre-defined time periods and stored in different Excel files. Semantria is then applied to conduct opinion mining as described in Section 3.1 for each period. Customer preferences are extracted, and the corresponding sentiment scores are obtained.

Step 2: For each period, the outliers are eliminated as described in Section 3.2. The sentiment scores of customer preferences are transformed into fuzzy numbers using equations (3) ~ (5) which are named fuzzy preference scores.

Step 3: Product attributes of the sample products are defined, and their settings are collected. With the opinion mining results and the collected product attribute settings, data sets for dynamic modelling of customer preferences are generated.

Step 4: Based on the generated data sets, a DENFIS approach is introduced to model customer preferences and predict the left, center and right values of future fuzzy scores of the customer preference. In each type of DENFIS model, the parameters used in the DENFIS are first set, including d and λ. The initialization for the first cluster C_1 is conducted as described in Section 3.3.1. The initial inverse matrix P and regression coefficients β are obtained using equations (13) and (14), respectively and the iteration starts.

Step 5: In each iteration, the new input data pair is presented to DENFIS by $l+1 \rightarrow l$. The clusters of the input are updated using ECM according to the process described in Section 3.3.1. Based on the clusters, the membership function $\mu_j(x_i)$ is generated using (11) and the weight of each fuzzy rule is calculated using (19). The regression coefficients β_{l+1} are updated by (16) and the corresponding first-order linear model is developed based on (12). Based on the weight of each fuzzy rule and the linear models, the predicted output $y_l^{'}(t)$ is then computed based on (18).

Step 6: The iteration stops when all the data sets are processed. The DENFIS model is developed and the predicted outputs are the future fuzzy scores of customer preferences.

4. **Implementation**

Nowadays, a large number of online customer reviews about various brands of hair dryers can be found easily from various websites. Those reviews provide valuable information about in what ways customers are happy / unhappy with various brands of hair dryers. However, if the reviews are collected and analyzed by company's staff, tremendous resources and time would be required for the staff the perform the tasks. Therefore, opinion mining needs to be employed to collect and analyze the reviews. However, the results of opinion mining mainly include sentiment scores of brands and product attributes as well as frequencies of the attributes which are not sufficient for companies to predict the customer preferences of new hair dryers and determine their product attribute settings. In the following, a case study of developing a dynamic customer preference model for supporting the design of hair dryers based on the proposed methodology is presented. The customer preference model developed based on the proposed methodology can be used to predict the customer preference score of a new hair dryer by inputting its product attribute setting and determine its optimal product attribute setting by introducing proper optimization techniques. In the case study, a company is planning to develop a new hair dryer and ten competitive hair dryers are identified and denoted as A~J, respectively. The online customer reviews of the competitive hair dryers available on Amazon.com were collected and a total of four periods were defined using a fixed-time-period strategy. Based on the collected online reviews which were stored in Excel files, the Semantria Excel add-in was used to conduct sentiment analysis for all online reviews. Key words and phrases were first mined from online reviews, and high frequency ones which were synonymous or related to the same customer preference were grouped. For example, the extracted phrases "powerful hairdryer", "awesome blow", "good heat", and "positive performance" were grouped as a category "performance", which is one of the customer preferences on hair dryers. Totally, six common customer preferences were summarized which are drying time, weight, price, performance, easy to use and quality. The user category analysis of Semantria was used, and the key words and phrases related to each customer preference were treated as the settings of the "user category". The sentiment analysis was repeated for each time period based on the settings and the sentiment scores of customer preferences for each online review were obtained.

In this paper, the customer preference, "performance", is used to illustrate the proposed DENFIS approach for the dynamic modelling of the customer preference "performance", which is denoted as y. In each period, the outliers were defined and removed. Using product A as an example, after the elimination of the outliers, the numbers of online reviews containing customers' opinions and comments on "performance" were 238, 390, 395 and 395 in the periods 1~4, respectively. Table 1 shows some content examples of the online reviews on "performance" and their corresponding sentiment polarity and scores.

Table 1. Examples of online reviews on "performance" with sentiment polarity and scores.

Performance		
Online reviews	Sentiment polarity	Sentiment score
Absolutely love this hair dryer! It's not loud, dries quickly and my hair shines!	Positive	1.452
…… I've had it for a month or so now and it still works great. The only problem is that after only one month, when the dryer gets hot, it starts to smell, which worries me a little. The performance hasn't changed, though. The attachments fit great and do not fall off at	Neutral	0.2

random. I wish EVERY hair dryer fitted attachments the way this one does! ……		
…… The chord on this dryer is VERY unsafe. It started shorting out on my wife about 4 months ago. Now, it sparks when you start it up. Terrible design.	Negative	-0.8111
Light and it works wonder, it dries my wife's hair really fast like 2-4 minutes …... Worth the price you paid! The best product made by ……	Positive	0.846

The sentiment scores of "performance" in each period were then transformed into fuzzy numbers using equations (3) ~ (5) as shown in Table 2, which are named fuzzy preference scores in the paper.

Table 2 The fuzzy preference scores of "performance".

Product	Performance			
	Period 1	Period 2	Period 3	Period 4
A	[-0.9,0.31,1.49]	[-1.03,0.31,1.57]	[-0.99,0.32,1.6]	[-0.98,0.29,1.45]
B	[-0.54,0.28,1.41]	[-0.64,0.23,0.94]	[-0.45,0.3,1.2]	[-0.98,0.25,1.2]
C	[-0.5,0.38,1.23]	[-0.6,0.35,1.26]	[-0.68,0.34,1.2]	[-0.73,0.34,1.24]
D	[-0.59,0.22,1.17]	[-0.82,0.26,1.16]	[-0.68,0.31,1.35]	[-0.37,0.25,1.2]
E	[-0.67,0.35,1.2]	[-0.68,0.31,1.49]	[-0.75,0.34,1.43]	[-0.9,0.44,1.65]
F	[-0.6,0.34,1.01]	[-0.6,0.33,1.44]	[-0.6,0.34,1.8]	[-0.6,0.28,1.2]
G	[-0.6,0.34,1.74]	[-0.5,0.48,1.16]	[-0.6,0.23,1.32]	[-0.89,0.19,1.2]
H	[-0.59,0.35,1.54]	[-0.9,0.31,1.46]	[-0.87,0.36,1.4]	[-0.74,0.32,1.5]
I	[-0.8,0.18,1.08]	[-0.62,0.27,1.2]	[-0.6,0.36,1.37]	[-0.88,0.27,1.23]
J	[-0.37,0.39,1.2]	[-0.6,0.33,1.32]	[-0.68,0.42,1.44]	[-0.53,0.4,1.3]

Four product attributes related to the customer preference "performance" were defined, which are weight, power, number of heat setting, and number of speed setting, and are denoted as x_1, x_2, x_3 and x_4, respectively. The product attribute settings of the ten sample products were collected and are shown in Table 3.

Table 3 The setting of product attributes of ten products.

Product attributes

Product	Weight (Pounds)	Power (Wattage)	Heat setting	Speed setting
	x_1	x_2	x_3	x_4
A	1.75	1875	2	2
B	1.5	1875	3	2
C	2.3	1875	3	2
D	2.55	1875	3	2
E	0.5	1000	2	2
F	1.6	1875	3	2
G	0.3	1875	3	3
H	1	2000	4	3
I	1.65	1875	3	3
J	1.8	2000	3	3

Based on Tables 2 and 3, the data sets for the dynamic modeling of the customer preference "performance" can be obtained. In this case study, three DENFIS models were developed respectively to predict the left, center and right values of the fuzzy preference scores of "performance" for Period 4. Periods 1~3 are treated as past periods. The fuzzy preference scores of Periods 1~3 denoted as y(t-3), y(t-2), and y(t-1), respectively, as well as the product attribute settings of x_1 to x_4, were used to develop DENFIS models for predicting the fuzzy preference score of "performance" for Period 4, y(t). Taking the prediction of the left value of Product A's fuzzy preference scores for Period 4 as an example, the matrix A and Y in (14) can be obtained as shown below based on the product attribute settings of products C~J shown in the above table and the corresponding left values of the fuzzy preference scores of Periods 1~3 as shown in Table 2.

$$A=\begin{pmatrix} 1 & 0.8535 & 0.8409 & 3 & 2 & -0.5 & -0.6 & -0.68 \\ 1 & 0.9545 & 0.8409 & 3 & 2 & -0.59 & -0.82 & -0.68 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0.6515 & 0.9545 & 3 & 3 & -0.37 & -0.6 & -0.68 \end{pmatrix}$$

$$Y = [-0.73, -0.37, -0.9, -0.6, -0.89, -0.74, -0.88, -0.53]^T$$

In the matrix A, columns 2 and 3 are the normalized values of x_1 and x_2. Since the values of x_1 and x_2 are continuous and have a large scale on the values, previous studies [50, 51] showed that normalizing such kind of values could help to improve training performance of neural systems. Columns 4 and 5 of the matrix are the values of x_3 and x_4. The last three columns of the matrix are the left values of the fuzzy preference scores of Periods 1~3, respectively. Suppose the data set of product C belongs to the k-th cluster and its cluster center is [0.1263, 0.0455, 2, 2, -0.67, -0.68, -0.75]. W_1 of the [W] shown in (13) and (14) can be calculated as follows by using (15).

$$W_1 = 1 - \{[|0.8535-0.1263|^2 + |0.8409-0.0455|^2 + |3-2|^2 + |2-2|^2 + |-0.5+0.67|^2 + |-0.6+0.68|^2 + |-0.68+0.75|^2]^{1/2}\}/(7^{1/2}) = 0.4392.$$

Similarly, W_2 to W_8 can be calculated and thus [W] can be obtained. By applying (13) and (14), the coefficients of the internal models of fuzzy rules can then be determined. The following shows some examples of the fuzzy rules generated by DENFIS, where the first-order linear models shown in the consequent parts of the fuzzy rules are all internal models relating y(t) with x_1, x_2, x_3, x_4, y(t-3), y(t-2), and y(t-1).

Fuzzy rule 1:

$$\begin{aligned}&If\ x_1\ is\ MF_{11},\ x_2\ is\ MF_{12}, x_3\ is\ MF_{13}, x_4\ is\ MF_{14}, y(t-3)\ is\ MF_{15}, y(t-2)\ is\ MF_{16},\\ &and\ y(t-1)\ is\ MF_{17}, then\ y(t) = -0.0417 - 0.0356x_1 - 0.0351x_2 - 0.1251x_3 - 0.0834x_4\\ &+0.0209y(t-3) + 0.0250y(t-2) + 0.0284y(t-1).\end{aligned}$$

Fuzzy rule 2:

$$\begin{aligned}&If\ x_1\ is\ MF_{21},\ x_2\ is\ MF_{22}, x_3\ is\ MF_{23}, x_4\ is\ MF_{24}, y(t-3)\ is\ MF_{25}, y(t-2)\ is\ MF_{26},\\ &and\ y(t-1)\ is\ MF_{27}, then\ y(t) = -0.0347 - 0.0219x_1 - 0.0292x_2 - 0.0695x_3 - 0.0695x_4\\ &+0.0313y(t-3) + 0.0358y(t-2) + 0.0344y(t-1).\end{aligned}$$

Fuzzy rule 3:

$$\begin{aligned}&If\ x_1\ is\ MF_{31},\ x_2\ is\ MF_{32}, x_3\ is\ MF_{33}, x_4\ is\ MF_{34}, y(t-3)\ is\ MF_{35}, y(t-2)\ is\ MF_{36},\\ &and\ y(t-1)\ is\ MF_{37}, then\ y(t) = -0.0420 - 0.0265x_1 - 0.0353x_2 - 0.084x_3 - 0.084x_4\\ &+0.0378y(t-3) + 0.0433y(t-2) + 0.0416y(t-1).\end{aligned}$$

Fuzzy rule 4:

$$\begin{aligned}&If\ x_1\ is\ MF_{41},\ x_2\ is\ MF_{42}, x_3\ is\ MF_{43}, x_4\ is\ MF_{44}, y(t-3)\ is\ MF_{45}, y(t-2)\ is\ MF_{46},\\ &and\ y(t-1)\ is\ MF_{47}, then\ y(t) = -0.0507 - 0.032x_1 - 0.0426x_2 - 0.1014x_3 - 0.1014x_4\\ &+0.0456y(t-3) + 0.0522y(t-2) + 0.0502y(t-1).\end{aligned}$$

Fuzzy rule 5:

$$\begin{aligned}&If\ x_1\ is\ MF_{51},\ x_2\ is\ MF_{52}, x_3\ is\ MF_{53}, x_4\ is\ MF_{54}, y(t-3)\ is\ MF_{55}, y(t-2)\ is\ MF_{56},\\ &and\ y(t-1)\ is\ MF_{57}, then\ y(t) = -0.0394 - 0.0248x_1 - 0.0331x_2 - 0.0787x_3 - 0.0787x_4\\ &+0.0354y(t-3) + 0.0405y(t-2) + 0.039y(t-1).\end{aligned}$$

Fuzzy rule 6:

$$If\ x_1\ is\ MF_{61},\ x_2\ is\ MF_{62}, x_3\ is\ MF_{63}, x_4\ is\ MF_{64}, y(t-3)\ is\ MF_{65}, y(t-2)\ is\ MF_{66},$$
$$and\ y(t-1)\ is\ MF_{67}, then\ y(t)=0.0164+0.014x_1+0.0138x_2+0.0492x_3+0.0328x_4$$
$$+0.0062y(t-3)+0.0057y(t-2)+0.0056y(t-1).$$

Fuzzy rule7:

$$If\ x_1\ is\ MF_{71},\ x_2\ is\ MF_{72}, x_3\ is\ MF_{73}, x_4\ is\ MF_{74}, y(t-3)\ is\ MF_{75}, y(t-2)\ is\ MF_{76},$$
$$and\ y(t-1)\ is\ MF_{77}, then\ y(t)=0.0218+0.0138x_1+0.0183x_2+0.0436x_3+0.0436x_4$$
$$+0.0068y(t-3)+0.0068y(t-2)+0.007y(t-1).$$

Fuzzy rule 8:

$$If\ x_1\ is\ MF_{81},\ x_2\ is\ MF_{82}, x_3\ is\ MF_{83}, x_4\ is\ MF_{84}, y(t-3)\ is\ MF_{85}, y(t-2)\ is\ MF_{86},$$
$$and\ y(t-1)\ is\ MF_{87}, then\ y(t)=0.0266+0.0168x_1+0.0224x_2+0.0533x_3+0.0533x_4$$
$$+0.0083y(t-3)+0.0083y(t-2)+0.0085y(t-1).$$

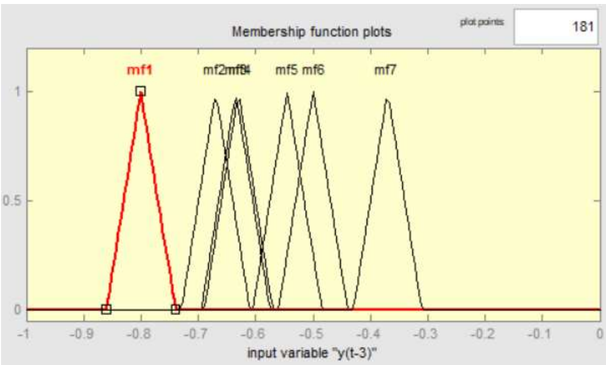
Fuzzy rule 9:

$$If\ x_1\ is\ MF_{91},\ x_2\ is\ MF_{92}, x_3\ is\ MF_{93}, x_4\ is\ MF_{94}, y(t-3)\ is\ MF_{95}, y(t-2)\ is\ MF_{96},$$
$$and\ y(t-1)\ is\ MF_{97}, then\ y(t)=0.0226+0.0143x_1+0.019x_2+0.0452x_3+0.0452x_4$$
$$+0.007y(t-3)+0.007y(t-2)+0.0072y(t-1).$$

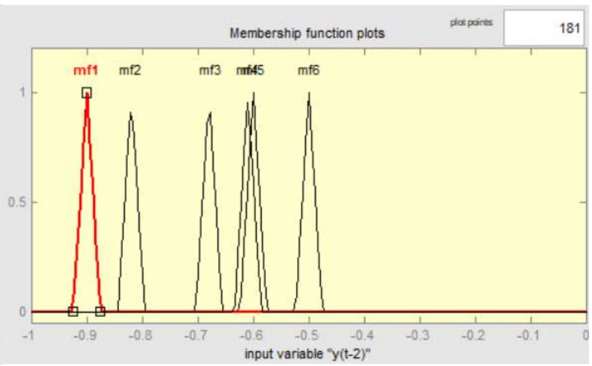
Fuzzy rule 10:

$$If\ x_1\ is\ MF_{101},\ x_2\ is\ MF_{102}, x_3\ is\ MF_{103}, x_4\ is\ MF_{104}, y(t-3)\ is\ MF_{105}, y(t-2)\ is\ MF_{106},$$
$$and\ y(t-1)\ is\ MF_{107}, then\ y(t)=0.022+0.0139x_1+0.0185x_2+0.044x_3+0.044x_4$$
$$+0.0068y(t-3)+0.0068y(t-2)+0.007y(t-1).$$

Figure 3 shows the membership functions of y(t-3) and y(t-2) generated by DENFIS for predicting the left values of fuzzy preference scores during the DENFIS's training process.



(a)



(b)

Fig.3. Generated membership functions: (a) y(t-3) and (b) y(t-2)

After completing the training of DENFIS, the fuzzy inference process of the DENFIS can start by inputting a data set to the DENFIS model. Taking the prediction of the fuzzy preference score of product A for Period 4 (i.e. t=4)as an example, the left values of x_1, x_2, x_3, x_4, y(t-3), y(t-2), and y(t-1) of product A were input to the developed DENFIS and then the fuzzy inference process of the DENFIS operated. Figure 4 shows the fuzzy inference process of the DENFIS by inputting the left values of x_1, x_2, x_3, x_4, y(t-3), y(t-2), and y(t-1) of product A. The predicted left value of product A's fuzzy preference score, y(t), for Period 4 was derived as -0.57 by using (18) and (19).

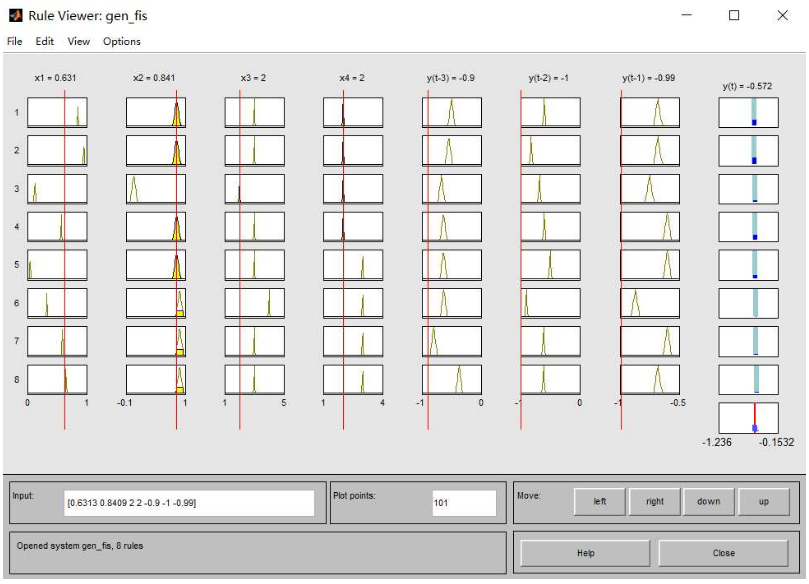


Fig. 4. Fuzzy inference process of predicting the left value of product A's fuzzy preference score for Period 4

5. Validation

DENFIS models were developed for the left, center, and the right values of the fuzzy preference scores, respectively. For each DENFIS model, ten validation tests were conducted. The data sets of products A and B, C and D, E and F, G and H, as well as I and J were selected as validation data sets for validation 1~5, respectively. The remaining data sets were used as training data sets in each validation test. No data set was repeated in the validation tests. To validate the effectiveness of the proposed DENFIS approach in dynamic modelling od customer preferences, the prediction results obtained by the DENFIS approach were compared with those obtained by ANFIS, subtractive cluster-based ANFIS (SC-ANFIS), fuzzy c-means-based ANFIS (FCM-ANFIS), and K-means-based ANFIS. SC assumes each data point is a potential cluster center and determines the cluster centers based on a density measure. In FCM, a given data point can belong to several clusters with a degree of membership. FCM partitions the data sets into fuzzy clusters by minimizing a cost function. K-means method aims to group data sets into K clusters by minimizing the value of an objective function. It alternates between assigning each data point to the cluster with the nearest mean and updating cluster centers until the value of the objective function has no further improvement. In the SC-ANFIS, FCM-ANFIS and K-means-based ANFIS, SC, FCM and K-means methods are incorporated into ANFIS to determine the membership functions of ANFIS, respectively. The mean relative error (*MRE*) and the variance of errors (VoE) defined in (20) and (21), respectively, were adopted to compare the validation results of the five approaches.

$$MRE = \frac{1}{n}\sum_{l=1}^{n}\frac{|y_l'(t)-y_l(t)|}{y_l(t)} \tag{20}$$

$$VoE = \frac{1}{n-1}\sum_{l=1}^{n}\left(\frac{|y_l'(t)-y_l(t)|}{y_l(t)} - MRE\right)^2 \tag{21}$$

where $y_l'(t)$ and $y_l(t)$ are the *lth* predicted and actual score of the customer preference; and n is the number of data sets in a validation test.

In each validation, the same data sets were utilized to model customer preferences based on DENFIS, ANFIS, SC-ANFIS, FCM-ANFIS, and K-means-based ANFIS. However, the ANFIS models could not be developed since a number of inputs were involved in the training which caused the failure of the training process of ANFIS and led to “out of memory” error. For DENFIS, the parameters *d* was set as 1.2 in all the validation tests for developing DENFIS models for the left, center, and right values of the fuzzy preference scores, which is a common setting used in previous studies. To determine the λ value for DENFIS, different λ values within the range of (0, 1) were used to perform the prediction. The prediction errors of the models corresponding to different λ values were then obtained and the λ value leading to the smallest error was selected. The values of λ were set as 1, 0.6, 0.6, 0.9, and 0.6 for the left values, 0.89, 0.88, 0.8, 0.9 and 1 for the center values, as well as 0.6, 1, 0.6, 0.6 and 0.8 for the right values of fuzzy preference scores in the validation tests 1~5, respectively. The number of clusters is usually less than or equal to $\sqrt{n}$, where *n* is the number of data sets [49]. For FCM-ANFIS and K-means-based ANFIS, the number of clusters was set as 3 for all the validation tests. The four approaches for modelling “performance” were implemented using Matlab software. After conducting the five validation tests for the left, center, and right values of sentiment scores, the predicted left, center, and right values of the preference scores in period 4 for products A~J were obtained based on DENFIS as shown in Table 4. Based on Table 4, the range of the future preference scores for each product can be obtained which is between the corresponding left and right value.

Table 4 The predicted future preference scores based on DENFIS.

Product	The predicted sentiment scores of period 4		
	The left value	The center value	The right value
A	-0.57	0.23	1.27
B	-0.66	0.28	1.18
C	-0.48	0.29	1.2
D	-0.5	0.28	1.2
E	-0.73	0.23	1.42
F	-0.61	0.3	1.28
G	-0.68	0.33	1.29
H	-0.82	0.39	1.45
I	-0.66	0.32	1.26
J	-0.65	0.32	1.3

Tables 5~7 show the *MRE* and VoE of the five validation tests for the left, center, right values of the fuzzy preference scores based on the four approaches, respectively. From the tables, it can be seen that the *MRE* and VoE based on the proposed approach are all smaller than those based on the other three approaches.

Table 5 The validation results for the left values of fuzzy preference scores of "performance"

Validation Test		Validation data sets	SC-ANFIS	FCM-ANFIS	K-means-ANFIS	DENFIS (Proposed approach)
1	*MRE*	A, B	0.3751	0.3762	0.3758	0.3727
	VoE		0.004	0.0039	0.0039	0.0037
2	*MRE*	C, D	0.6123	0.5884	0.6079	0.3402
	VoE		0.595	0.5832	0.5936	$1.0155*10^{-5}$

3	*MRE*	E, F	0.2715	0.2632	0.2726	0.1020
	VoE		0.017	0.0190	0.0164	0.0160
4	*MRE*	G, H	0.2019	0.2042	0.2024	0.1739
	VoE		0.0353	0.0393	0.0376	0.0072
5	*MRE*	I, J	0.3193	0.3223	0.3191	0.2378
	VoE		0.2015	0.2060	0.1987	$1.4972*10^{-4}$

Table 6 The validation results for the center values of fuzzy preference scores of "performance"

Validation Test		Validation data sets	SC-ANFIS	FCM-ANFIS	K-means-ANFIS	DENFIS (proposed approach)
1	*MRE*	A, B	0.1670	0.1678	0.1671	0.1654
	VoE		0.0050	0.0036	0.0037	0.0029
2	*MRE*	C, D	0.1481	0.1480	0.1489	0.1458
	VoE		0.0013	0.0015	0.0007	$2.7410*10^{-4}$
3	*MRE*	E, F	0.3095	0.2996	0.3044	0.2764
	VoE		0.1138	0.1235	0.1167	0.0835
4	*MRE*	G, H	0.5491	0.5549	0.5562	0.4905
	VoE		0.1622	0.1640	0.1639	0.1466
5	*MRE*	I, J	0.1922	0.1921	0.1915	0.1842
	VoE		0.0020	0.0021	0.0017	$2.3216*10^{-4}$

Table 7 The validation results for the right values of fuzzy preference scores of "performance"

Validation Test		Validation data sets	SC-ANFIS	FCM-ANFIS	K-means-ANFIS	DENFIS (Proposed approach)
1	*MRE*	A, B	0.1195	0.1198	0.1195	0.0722
	VoE		0.0263	0.0261	0.0264	0.0054
2	*MRE*	C, D	0.0324	0.0317	0.0343	0.0177
	VoE		0.0007	0.0007	0.0007	0.0005
3	*MRE*	E, F	0.2071	0.2083	0.2080	0.1021
	VoE		0.0522	0.0502	0.0500	0.0027
4	*MRE*	G, H	0.1564	0.1544	0.1572	0.0530
	VoE		0.0073	0.0071	0.0071	$9.9037*10^{-4}$

5	*MRE*	I, J	0.1347	0.1371	0.1360	0.0125
	VoE		0.0004	0.0004	0.0004	$1.5335*10^{-4}$

With the predicted customer preferences scores of the competitive products, the company can refer to the scores in the formulation of a positioning strategy for the new product. On the other hand, appropriate product attribute settings of the new hair dryer can also be determined easily using the developed DENFIS.

6. Discussion

In the case study as described in Section 4, the sentiment scores of Periods 1~3 were derived based on opinion mining from online customer reviews and then used as training data sets for developing DENFIS models. However, if a conventional approach were adopted, three times of surveys are required to be conducted respectively for the periods 1~3. The surveys could be paper-based or web-based. Although web-based surveys can take shorter time and less resources to obtain survey data (i.e. customers' views and ratings on products) relatively, it is still common to take months to obtain a sufficient number of time series data sets. For the proposed approach, as customer views, opinions and ratings of products are available all the time on various websites, they can be extracted under different time periods easily and the sentiment scores of customer preferences can be determined shortly by using opinion mining. Then, DENFIS can be adopted to develop customer preference models. If data sets of a new period are available for updating the model, the DENFIS model can be updated easily and needs not to be rebuilt.

ANFIS was commonly employed in previous studies for time series modeling. However, conventional ANFIS is incapable of modelling the problems that involve a number of inputs. In the case study presented in Section 4, the case only involves seven inputs but ANFIS models could not be generated as the number of inputs is too large. In this regard, DENFIS is better than ANFIS in handling high-dimensional problems as it can generate internal models by demand. Nevertheless, DENFIS has its shortcomings. In the training process of the DENFIS model, partitioned regions are formed by using a clustering method for training internal models. The entire input space may not be covered by the partitioned regions of clusters. The DENFIS model can produce accurate prediction if the inputs are inside the regions; otherwise, its prediction is less accurate. To implement DENFIS, we noted that it was quite difficult to select a proper value of λ . In this research, the value of λ for DENFIS was determined based on the smallest errors by using a trial and error method. Therefore, it could take quite a long time to determine a proper setting of λ for DENFIS. To shorten the computational time, DENFIS approach can be modified by adjusting the setting of the λ value adaptively with reference to the lengths of periods. Based on the validation results as presented in Section 4, it can be seen that DENFIS outperforms various ANFIS approaches in time-series modelling in terms of mean relative errors and the variance of errors. In addition, the proposed approach can provide both crisp and fuzzy outputs that cannot be realized by using existing ANFIS and conventional DENFIS approaches. The fuzzy outputs can help product development teams to assess the uncertainty of prediction.

Once the results of opinion mining from online reviews are obtained under various time periods, DENFIS models (i.e. dynamic customer preference models) can be developed based on the time series data obtained from the opinion mining and the specifications of sample products. Product development teams not only can use the developed DENFIS models to predict the customer preference score of a new product by inputting the attribute setting of the new product, but also can employ the models to help determine the optimal product attribute settings of new products with using proper optimization techniques. One possible framework of the determination is shown in Fig. 5 where a developed DENFIS model combining with a genetic algorithm are employed to determine the optimal product attribute setting of a new product with the objective of maximizing customer preference scores.

The proposed methodology basically can be applied on all kinds of products. However, for some kinds of products such as artistic works/products, their product attributes could be highly diverse and are not clear. Without clear identification of product attributes, the proposed methodology may not be able to apply on such kinds of products.

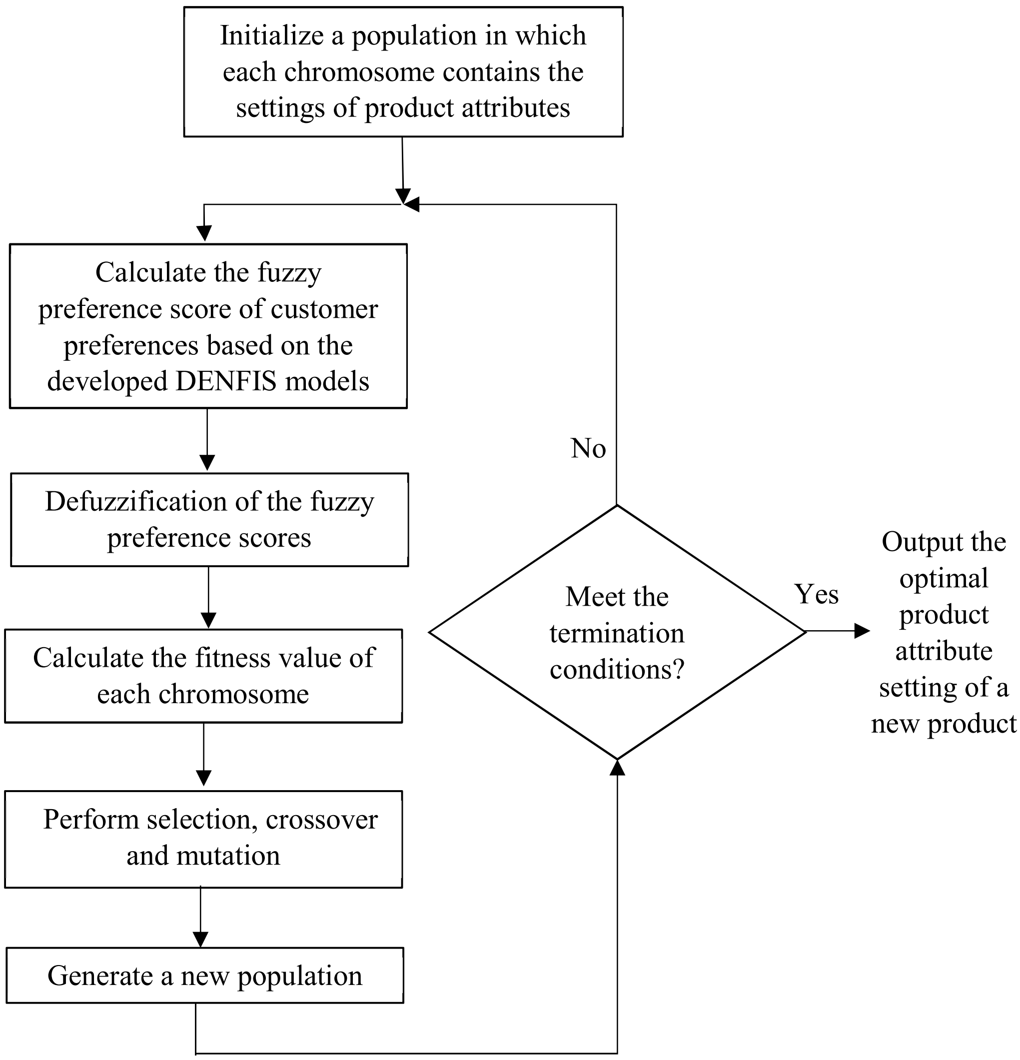


Fig. 5. Flowchart of the determination of product attribute settings.

7. Conclusion

Customer surveys were commonly used in previous studies and industries to collect survey data which was then used to analyze customer preferences on products and develop customer preference models. However, customer preferences on products could vary rapidly with respect to time. Therefore, it is necessary to collect time series data of customer preferences for developing customer preference models such that the developed models are able to make prediction for the future period. In reality, it is difficult to obtain time series survey data of customer preferences under different time periods using customer surveys. Compared with the conventional survey data, online customer reviews contain rich information on customers' opinions and comments toward products, from which time series data of customer preferences can be easily obtained under different time periods. The obtained time series data can then be used to develop customer preference models. However, no publications were found thus far regarding the modelling of customer preferences based on the time series data obtained from online reviews. In addition, previous studies failed to address the fuzziness which exists in the customers' sentiment expression in online reviews. Although some previous studies have generated rules to relate customer preferences/satisfaction and product attributes, the generated rules many times are inadequate to help determine the product attribute settings of new products. To address the research problems, a new methodology for dynamic modelling of customer preferences on products based on online customer reviews is described in the paper, which mainly involves opinion mining from online reviews, and a DENFIS approach for developing dynamic models of customer preferences. In contrast to the conventional DENFIS approach for modeling which only provide crisp outputs, the proposed DENFIS approach is able to provide both crisp and fuzzy outputs. With the predicted fuzzy outputs, companies can consider the worst scenario (i.e. the predicted left values of fuzzy preference scores) and the best scenario (i.e. the predicted right values of fuzzy preference scores) of customer preferences while designing their new products/services. A case study on hair dryer products for dynamic modelling of customer preferences using online reviews was conducted to illustrate the proposed methodology. To evaluate the effectiveness of the DENFIS approach in dynamic modelling of customer preferences, the prediction results obtained based on the DENFIS approach are compared with those obtained by ANFIS, SC-ANFIS, FCM-ANFIS, and K-means-based ANFIS. The training process for ANFIS could not be done completely because of its complex structure that led to "out of memory" error. The results of the comparison showed that the DENFIS approach outperforms the other three approaches in terms of mean absolute percentage errors and variance of errors. Future research work would involve the further evaluation of the effectiveness of DENFIS for dynamic modeling by studying various issues such as the parameter setting of DENFIS, minimum and optimal number of time periods, incorporating backpropagation algorithm into the learning algorithm of DENFIS, and adopting Gaussian membership functions. A study of the improvement of the approach with the adaptive determination of DENFIS parameter settings would also be considered in future work. On the other hand, based on the generated prediction models, a further study of determining the optimal product attribute settings of new products would be conducted.

Acknowledgement

The work described in this paper was supported by a grant from The Hong Kong Polytechnic University (Project No. G-UADL).

References

[1] M.M. Tseng and F.T. Piller, The Customer Centric Enterprise: Advances in Mass Customization and Personalization, (2003) Berlin, Springer.

[2] Y.E. Nahm, A Novel Approach to Prioritize Customer Requirements in QFD based on Customer Satisfaction Function for Customer-Oriented Product Design, Journal of Mechanical Science and Technology, 27(12) (2013) 3765-3777.

[3] N.K. Kasabov, Q. Song, DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction, IEEE Transactions on Fuzzy Systems 10 (2) (2002) 144-154.

[4] C. Clavel, Z. Callejas, Sentiment Analysis: From Opinion Mining to Human-Agent Interaction, IEEE Transactions on Affective Computing 7 (1) (2016) 74-93.

[5] E. Cambria, Affective Computing and Sentiment Analysis, IEEE Intelligent Systems 31 (2) (2016) 102-107.

[6] Y. Ma, H. Peng, E. Cambria, Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM, In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 5876-5883.

[7] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, E. Cambria, Learning Word Representations for Sentiment Analysis, Cognitive Computation 9 (6) (2017) 843-851.

[8] S. Jasti, T.S. Mahalakshmi, A Review on Sentiment Analysis of Opinion Mining, in: P. Mallick, V. Balas, A. Bhoi, A. Zobaa (Eds.), Cognitive Informatics and Soft Computing, Springer, Singapore, 2019, pp. 603-612.

[9] H.H. Do, PWC Prasad, A. Maag, A. Alsadoon, Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review, Expert Systems with Applications 118 (2019) 272-299.

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

[11] 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.

[12] 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.

[13] M. Zimmermann, E. Ntoutsi, M. Spiliopoulou, Discovering and monitoring product features and the opinions on them with OPINSTREAM, Neurocomputing 150 (Part A) (2015) 318–330.

[14] 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.

[15] 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.

[16] 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.

[17] Y. Kang, L. Zhou, RubE: Rule-based methods for extracting product features from online consumer reviews, Information & Management 54 (2) (2017) 166-176.

[18] M.C. Chiu, K.Z. Lin, Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage, Advanced Engineering Informatics 38 (2018) 826-839.

[19] A.J.C. Trappey, C.V. Trappey, C.Y. Fan, I. J.Y. Lee, Consumer driven product technology function deployment using social media and patent mining, Advanced Engineering Informatics 36 (2018) 120–129.

[20] S. Tuarob, C.S. Tucker, Fad or Here to Stay: Predicting Product Market Adoption and Longevity Using Large Scale, Social Media Data, In: Proceedings of 33rd Computers and Information in Engineering Conference, 2013, pp. V02BT02A012.

[21] 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.

[22] S. Tuarob, C.S. Tucker, Quantifying product favorability and extracting notable product features using large scale social media data, Journal of Computing and Information Science in Engineering 15 (3) (2015) 031003.

[23] S. Lim, C.S. Tucker, A Bayesian sampling method for product feature extraction from large-scale textual data, Journal of Mechanical Design 138 (6) (2016) 061403.

[24] J. Jin, P. Ji, Y. Liu, Product characteristic weighting for designer from online reviews: an ordinal classification approach, In: EDBT-ICDT '12 Proceedings of the 2012 Joint EDBT/ICDT Workshops, 2012, pp. 33-40.

[25] R. Rai, Identifying key product attributes and their importance levels from online customer reviews, In: Proceedings of 38th Design Automation Conference, 2012, pp. 533-540.

[26] L. Yang, B. Liu, H. Lin, Y. Lin, Combining local and global Information for product feature extraction in opinion documents, Information Processing Letters 116 (10) (2016) 623–627.

[27] Nagamachi, M., Perspectives and the new trend of Kansei/affective engineering, The TQM Journal 20(4) (2008) 290–298.

[28] You, H., T. Ryu, K. Oh, M. H. Yun, K. J. Kim, Development of customer satisfaction models for automotive interior materials, International Journal of Industrial Ergonomics 36(4) (2006) 323–330.

[29] Yang, J. B., Y. M. Wang, D. L. Xu, K. S. Chin, L. Chatton, Belief rule-based methodology for mapping consumer preferences and setting product targets, Expert Systems with Applications 39 (5) (2012) 4749-4759.

[30] Chen, C. H., L.P. Khoo, W. Yan, An investigation into affective design using sorting technique and Kohonen self-organising map, Advances in Engineering Software 37 (5) (2006) 334-349.

[31] Fung, R., D. Law, W. Ip, Design targets determination for interdependent product attributes in QFD using fuzzy inference, Integrated Manufacturing System 10 (6) (1999) 376-383.

[32] Park, J., S. H. Han, A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design, International Journal of Industrial Ergonomics 34 (1) (2004) 31-47.

[33] Chen, Y., L. Chen, A non-linear possibilistic regression approach to model functional relationships in product planning, The International Journal of Advanced Manufacturing Technology 28 (11) (2006) 1175-1181.

[34] Sekkeli, G., G. Koksal, I. Batmaz, O. T. Bayrak, Classification models based on Tanaka's fuzzy linear regression approach: The case of customer satisfaction modeling, Journal of Intelligent and Fuzzy systems 21 (5) (2010) 341–351.

[35] Chan, K. Y., C. K. Kwong, T. S. Dillon, K. Y. Fung, An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness, Journal of Engineering Design 22 (8) (2011) 523–542.

[36] Jiang, H., C.K. Kwong, W.H. Ip, Z. Chen, Chaos-based fuzzy regression approach to modeling customer satisfaction for product design. IEEE Transactions on Fuzzy Systems 21 (5) (2013) 926–936.

[37] Chan, K.Y., H. K. Lam, T. S. Dillon, S. H. Ling, A stepwise-based fuzzy regression procedure for developing customer preference models in new product development, IEEE Transactions on Fuzzy Systems 23 (5) (2015) 1728-1745.

[38] Chan, K. Y., S. H. Ling, A forward selection based fuzzy regression for new product development that correlates engineering characteristics with consumer preferences, Journal of Intelligent and Fuzzy Systems 30 (3) (2016) 1869-1880.

[39] 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.

[40] M. Xie, K. C. Tan, T.N. Goh, Advanced QFD applications, ASQ Quality Press, Wisconsin, 2003.

[41] 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.

[42] Y.T. Chong, C.H. Chen, Management and forecast of dynamic customer needs: An artificial immune and neural system approach, Advanced Engineering Informatics 24 (1) (2010) 96-106.

[43] 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, 83-88.

[44] 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.

[45] W. Chung, T.L. Tseng, Discovering Business Intelligence from Online Product Reviews: A Rule-Induction Framework, Expert Systems with Applications 39(15) (2012) 11870-11879.

[46] H. Jiang, C.K. Kwong, W.Y. Park, K.M. Yu, A multi-objective PSO approach of mining association rules for affective design based on online customer reviews, Journal of Engineering Design 29(7) (2018) 381-403.

[47] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, In: Proceedings of the 7th Conference on Language Resources and Evaluation (LREC'10), 2010, 2200-2204.

[48] Lexalytics, www.lexalytics.com.

[49] A. Bari, L. Rueda, A New Profile Alignment Method for Clustering Gene Expression Data, in: L. Lamontagne, M. Marchand(Eds.), Advances in Artificial Intelligence, Springer, Germany, 2006, pp. 86-97.

[50] J. Jin, M. Li, L. Jin, Data normalization to accelerate training for neural net to predict tropical cyclone tracks, Mathematical Problems in Engineering (2015) 1-8.

[51] T. Jayalakshmi, A. Santhakumaran, Statistical Normalization and Back Propagation for Classification, International Journal of Computer Theory and Engineering 3(1) (2011) 1793-8201.