

NOTICE: this is the author's version of a work that was accepted for publication in Applied Soft Computing. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Applied Soft Computing, Vol.12, no.4 (April 2012). DOI: 10.1016/j.asoc.2011.11.026

# **Market segmentation and ideal point identification for new product design using fuzzy data compression and fuzzy clustering methods**

K. Y. Chan<sup>1</sup>, C.K. Kwong<sup>2</sup>, and B.Q. Hu<sup>3</sup>

<sup>1</sup>Digital Escosystems and Business Intelligence Institute, Curtin University, Western Australia, Australia

Email: [kit.chan@curtin.edu.au](mailto:kit.chan@curtin.edu.au)

Tel: 61-8-9266 9269

Fax: 61-8-9266 7548

<sup>2</sup>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Email: [C.K.Kwong@inet.polyu.edu.hk](mailto:C.K.Kwong@inet.polyu.edu.hk)

Tel: (852) 3400 3249

Fax: (852) 2362 5267

<sup>3</sup>School of Mathematics and Statistics

Wuhan University

Wuhan 430072

China

Email : [bqhu@public.wh.hb.cn](mailto:bqhu@public.wh.hb.cn)

***Abstract* - In product design, various methodologies have been proposed for market segmentation, which group consumers with similar customer requirements into clusters. Central points on market segments are always used as ideal points of customer requirements for product design, which reflects particular competitive strategies to effectively reach all consumers' interests. However, existing methodologies ignore the fuzziness on consumers' customer requirements. In this paper, a new methodology is proposed to perform market segmentation based on consumers' customer requirements, which exist fuzziness. The methodology is an**

**integration of a fuzzy compression technique for multi-dimension reduction and a fuzzy clustering technique. It first compresses the fuzzy data regarding customer requirements from high dimensions into two dimensions. After the fuzzy data is clustered into marketing segments, the centre points of market segments are used as ideal points for new product development. The effectiveness of the proposed methodology in market segmentation and identification of the ideal points for new product design is demonstrated using a case study of new digital camera design.**

**Keywords:** Market segmentation, ideal point, product design, fuzzy compression, fuzzy clustering, digital camera design, genetic algorithm, customer requirement

## **1. Introduction**

In product design, market segmentation is essential to partition a large amount of consumers into clusters in which the consumers inside each cluster have similar customer requirements [1-3]. The information of the segments is useful for product designers to manufacture new products which can effectively satisfy all customers' needs [4]. Several benefits can be obtained from the strategy of market segmentation. One of the benefits is the identification of central points of segments, which can be treated as ideal points to reflect the customer requirements of the consumers inside the segment [5]. Ideal points can help companies to plan for manufacturing potential new products, and can allow the decision makers to deploy resources more effectively and efficiently. Also, they can be used to determine the particular competitive strategies, in order to manufacture new products that can bring larger profits [6].

One of the most commonly used methodologies for segmentation on consumer interests and identification of ideal points from segments is conjoint analysis [7]. It assesses the rank-order or overall value for alternatives with different profiles of attribute levels, and then uses the holistic judgment information to estimate discrete levels of single-attribute value functions by regressions, hierarchical Bayes models, or linear programming [8]. Self-organized feature map is widely used for dimension reduction and clustering, concurrently for various applications of which the data is in multi-dimensions [9-11]. However, this approach has only been used on processing real data, while processing data in fuzzy numbers has not been addressed. Other artificial intelligent techniques, including simulated annealing algorithms [12-13], multicriterion clusterwise regression [14], neural network [15-16], genetic algorithms [17-18], clustering algorithms [19], have also been applied to segment consumers into clusters. Consumers are grouped into a cluster of which their buying interests are close. However, these clustering approaches ignore the fuzzy natures of consumers in buying products. In fact, fuzziness always exists in human feeling or attitudes [20-22], thus consideration of consumers' fuzziness for clustering purposes is necessary.

In the paper, a methodology is proposed to perform marketing segmentation based on consumers' survey data, where the fuzzy characteristics regarding consumers' customer requirements are considered. The methodology can be divided into two main tasks: a) it first compresses the consumers' survey data with multi-dimensional customer requirements into two-dimensions, which can be presented on a perceptual plot for analysis purpose; b) it then clusters consumers with similar customer requirements into various marketing segments. The centre points of the segments can be used to identify

potential new products for consumers. The approaches used on the methodology in conducting multi-dimensional compression and clustering of consumers are discussed below:

**a) Multi-dimensional compression approach:** Classical principal component analysis (PCA) [23], which can discover the underlying structure of data sets into a two-dimensional display, can be used. However, this PCA cannot be applied on the cases involving fuzzy data. To compress fuzzy data with multi-dimensions, a fuzzy compression method that can map fuzzy data onto a crisp representation space has been applied by [24-26]. Also, a dimension compression method that can map fuzzy data onto a crisp representation space has been applied by [27-28], and another clustering approach dubbed the vertices method in [26] has also been proposed. However, these compression methods are not oriented toward visualizing high-dimensional fuzzy data, so they can only build suboptimal low-dimensional representations of the fuzzy data.

In this research, a more recent dimension compression method [29] is proposed to compress fuzzy data for consumer perception, which involve multi-dimensions. It attempts to overcome the limitations of those multi-dimension compression methods, where they either cannot address fuzziness of consumer survey data or cannot compress highly dimensional data. The proposed dimension compression method [29] attempts to overcome these two limitations by finding iteratively linear features from which the original fuzzy data is recovered with minimum error using a linear transformation. It applies the three layer neural network with equally number of input and output neurons, and two hidden neurons in the hidden layer. By minimizing an error criterion, the visualization reveals the most salient features of the original fuzzy data [29].

**b) Consumer clustering approach:** In this research, a recent fuzzy clustering method [30] is proposed to cluster consumer survey data into various market segments. It clusters consumer survey data into groups displaying similar characteristics based on the assigned partial memberships of each consumer survey data. This approach is similar to the fuzzy c-number cluster method [31], but an exponential-type distance [19] is introduced to evaluate the fuzzy distance between consumer survey data which is fuzzy. Based on the proposed fuzzy distance measure, the resulting market segments are more robust than the existing fuzzy clustering methods [31-32], which are strongly influenced by outliers and may give poor clustering results when the consumer survey data includes large numbers of different cluster sample sizes. After clustering, the central points of the clusters can be treated as the ideal points of market segments, which reflect consumers' buying interests, can be identified.

The rest of the paper is organized as follows: In Section 2, a new methodology, which mainly involves fuzzy data compression method and fuzzy clustering method, is discussed to determine market segments and to identify the ideal points of segments based on multi-dimensional fuzzy data. In section 3, an example of digital camera design is used to illustrate the effectiveness of the proposed methodology. Finally, conclusions are given in section 4.

## **2. Methodology of identification of market segments and ideal points**

The methodology of identification of market segments and ideal points for new product design consists of five main steps, which can be described briefly as follows:

*Step 1 Data collection:* Customer's data, which concerns on  $d$  customer requirements for product design, is collected by surveying potential customers.

*Step 2 Fuzzy data compression from  $d$ -dimensions into two-dimensions:* The  $d$  customer requirements on the customer's data are extracted into two principle components based on a fuzzy compression method [29].

*Step 3 Perceptual plot:* The customer's data with the two principle components is plotted onto a perceptual plot.

*Step 4 Fuzzy clustering:* Segments on customer's data are identified on the fuzzy perceptual plot by the fuzzy clustering method [30].

*Step 5 Determination of design:* The fuzzy central points of the segments with two dimensions are used as the ideal points of new product design. The  $d$  customer requirements of the new product design can be transformed from the ideal points with two dimensions by the fuzzy compression method [29].

## 2.1. Fuzzy data collection

Data collection is the first step of the methodology. It can be done by surveying  $n$  potential customers regarding  $d$  customer requirements of the product. As customers' opinions are human intuition, these collected customers' data sets are fuzzy data sets.

These  $n$  fuzzy data sets collected by  $n$  customers are denoted as

$$\tilde{X} = \{\tilde{x}(1), \tilde{x}(2), \dots, \tilde{x}(n)\} \quad (1)$$

where the  $j$ -th fuzzy data set with  $d$  dimensions regarding the  $d$  customer requirements is denoted as  $\tilde{x}(j) = (\tilde{x}_1(j), \tilde{x}_2(j), \dots, \tilde{x}_d(j))$ ;  $\tilde{x}_i(j) = (m_{x_i}(j), \alpha_{x_i}(j), \beta_{x_i}(j))$  represents the  $i$ -th customer requirement (with  $i=1,2,\dots,d$ ) regarding the  $j$ -th customer (with  $j=1,2,\dots,n$ );

$m_{x_i}(j)$  is the mean value of  $\tilde{x}_i(j)$ ;  $\alpha_{x_i}(j)$  and  $\beta_{x_i}(j)$  are the left and right spreads of  $\tilde{x}_i(j)$  respectively; and the fuzzy number,  $\tilde{x}_i(j)$ , is defined by Zimmermann's shape function [33 pp. 62-63]:

$$\tilde{x}_i(j)|_x = \begin{cases} L\left(\frac{m_{x_i}(j)-x}{\alpha_{x_i}(j)}\right) & \text{for } x \leq m_{x_i}(j) \\ R\left(\frac{x-m_{x_i}(j)}{\beta_{x_i}(j)}\right) & \text{for } x \geq m_{x_i}(j) \end{cases}, \quad (2)$$

where  $L$  and  $R$  is a decreasing shape function ranged from 0 to 1:

$$\begin{cases} L(x)=1 & \text{for } x=0 \\ 0 < L(x) < 1 & \text{for } x > 0, \\ L(x)=0 & \text{for } x=\infty \end{cases}, \quad (3)$$

and

$$\begin{cases} R(x)=1 & \text{for } x=0 \\ 0 < R(x) < 1 & \text{for } x < 0, \text{ respectively.} \\ R(x)=0 & \text{for } x=-\infty \end{cases} \quad (4)$$

## 2.2 Fuzzy data compression from d-dimensions into two-dimensions

Performing clustering on multi-dimensional fuzzy data is often inefficient, and also it is difficult to analyze the fuzzy data with high dimensions. Therefore, it is essential to compress the fuzzy data from high dimensions into two dimensions, while preserving as much information of the original fuzzy data as possible. After compressed the fuzzy data into two dimensions, the compressed fuzzy data can be illustrated onto a perceptual plot, where fuzzy clustering can be performed by grouping the fuzzy data with correlated customer requirements.

The neural network based principle component analysis namely NPCA [29] (illustrated in Figure 1), is proposed to compress  $d$ -dimensional data sets into two-



dimensional data sets. NPCA is a feedforward three-layer neural network, where there is an input layer of  $d$  units, a hidden layer with two units, and an output layer with  $d$  units. NPCA is trained in an autoassociative mode using the inputs as target outputs so as to learn to approximate the identity mapping. Between the input layer to the hidden layer, NPCA compresses the fuzzy data from  $d$ -dimensions into two-dimensions. Between the hidden layer to the output layer, it works as a feature extractor [34-35] to diverge the fuzzy data from two-dimensions into  $d$ -dimensions. In the hidden layer of the NPCA, the whole population of the compressed fuzzy data sets of the  $n$  customers is denoted as:

$$\tilde{Y} = \{\tilde{y}(1), \tilde{y}(2), \dots, \tilde{y}(n)\} \quad (5)$$

where  $\tilde{y}(j) = \{\tilde{y}_1(j), \tilde{y}_2(j)\}$  with  $j=1, 2, \dots, n$ . The output  $\tilde{y}_k(j) = (m_{y_k}(j), \alpha_{y_k}(j), \beta_{y_k}(j))$  of the  $k$ -th hidden unit with respect to the  $j$ -th interviewer is defined by [16] as:

$$m_{y_k}(j) = \sum_{i=1}^d B_{ik} m_{x_i}(j) \quad (6)$$

$$\alpha_{y_k}(j) = \sum_{\substack{i=1 \\ B_{ik} > 0}}^d B_{ik} \alpha_{x_i}(j) + \sum_{\substack{i=1 \\ B_{ik} < 0}}^d B_{ik} \beta_{x_i}(j) \quad (7)$$

$$\beta_{y_k}(j) = \sum_{\substack{i=1 \\ B_{ik} > 0}}^d B_{ik} \beta_{x_i}(j) + \sum_{\substack{i=1 \\ B_{ik} < 0}}^d B_{ik} \alpha_{x_i}(j) \quad (8)$$

where  $k=1, 2$  and  $B_{ik}$  is the neural network weight of the NPCA. This compressed fuzzy data,  $\tilde{y}(j) = \{\tilde{y}_1(j), \tilde{y}_2(j)\}$ , for the  $j$ -th customer is illustrated in Figure 2, where  $f_1$  and  $f_2$  is the two principle components of the original fuzzy data with  $d$  dimensions.

In the output layer of the NPCA, the fuzzy data sets with two dimensions are decompressed into  $d$  dimensions. The decompressed fuzzy data,  $\tilde{Z}$ , is denoted as:

$$\tilde{Z} = \{\tilde{z}(1), \tilde{z}(2), \dots, \tilde{z}(n)\} \quad (9)$$

where  $\tilde{z}(j) = \{\tilde{z}_1(j), \tilde{z}_2(j), \dots, \tilde{z}_d(j)\}$  with  $j=1, 2, \dots, n$ ; and the  $i$ -th output (or attitude) of the  $j$ -th interviewer  $\tilde{z}_i(j) = (m_{z_i}(j), \alpha_{z_i}(j), \beta_{z_i}(j))$  is defined by:

$$m_{z_i}(j) = \sum_{k=1}^2 B_{ik} m_{y_k}(j) \quad (10)$$

$$\alpha_{z_i}(j) = \sum_{\substack{k=1 \\ B_{ik} > 0}}^2 B_{ik} \alpha_{y_k}(j) + \sum_{\substack{k=1 \\ B_{ik} < 0}}^2 B_{ik} \beta_{y_k}(j) \quad (11)$$

$$\text{and } \beta_{z_i}(j) = \sum_{\substack{k=1 \\ B_{ik} > 0}}^2 B_{ik} \beta_{y_k}(j) + \sum_{\substack{k=1 \\ B_{ik} < 0}}^2 B_{ik} \alpha_{y_k}(j) \quad (12)$$

with  $i=1, 2, \dots, d$ .

**[insert Figure 1: The achitecture of the NPCA]**

**[insert Figure 2: Compressed fuzzy data,  $\tilde{y}(j) = \{\tilde{y}_1(j), \tilde{y}_2(j)\}$ , for the  $j$ -th customer]**

Based on the triangular fuzzy distance in a three-dimensional Euclidean space [35, 41], the reconstruction error between the input and output units of the NPCA is defined as:

$$E(\mathbf{B}) = \sum_{i=1}^d \sum_{j=1}^n \left[ (m_{x_i}(j) - m_{z_i}(j))^2 + (\alpha_{x_i}(j) - \alpha_{z_i}(j))^2 + (\beta_{x_i}(j) - \beta_{z_i}(j))^2 \right] \quad (13)$$

where  $\mathbf{B} = \{B_{1,1}, B_{2,1}, \dots, B_{d,1}, B_{1,2}, B_{2,2}, \dots, B_{d,n}\}$ .

When the reconstruction error,  $E(\mathbf{B})$ , defined in equation (13) is small, the input fuzzy data,  $\tilde{X}$ , matches properly the output fuzzy data,  $\tilde{Z}$ . The compressed fuzzy data,  $\tilde{Y}$ , in the hidden layer also matches properly both  $\tilde{X}$  and  $\tilde{Z}$ . Therefore, the compressed fuzzy data,  $\tilde{Y}$ , are sufficient to represent the input fuzzy data,  $\tilde{X}$ , and the two principal

components of  $\tilde{Y}$  are sufficient to represent the  $d$ -dimensions of the input fuzzy data,  $\tilde{X}$ , when  $E(\mathbf{B})$  is small.

Minimizing  $E(\mathbf{B})$  in equation (13) was performed by the backpropagation method [29]. However, the backpropagation method may only converge to a local minimum, and is sensitive to the values of the initial parameters. Genetic algorithms GAs have been applied on optimization of neural networks satisfactorily in the purposes of modeling [36-37], control engineering [38-39] and prediction [40]. Therefore the GA is proposed to be used here to search the weights  $\mathbf{B}$  of the neural networks. The detailed procedures of the proposed GA namely GA-NPCA is given in the appendix.

### 2.3 Perceptual plot

The  $n$  two-dimensional fuzzy data  $(\tilde{y}_1(j), \tilde{y}_2(j)) \in \tilde{Y}$  with respect to the two principle components,  $f_1$  and  $f_2$ , can be illustrated on a perceptual plot, where  $\tilde{y}_1(j) = (m_{y_1}(j), \alpha_{y_1}(j), \beta_{y_1}(j))$  and  $\tilde{y}_2(j) = (m_{y_2}(j), \alpha_{y_2}(j), \beta_{y_2}(j))$  with  $j = 1, 2, \dots, n$ . In the perceptual plot, fuzzy clustering can be performed by grouping the fuzzy data sets with small fuzzy distances into segments, which aid the determination of the marketing strategy.

### 2.4 Fuzzy clustering

Here a recent approach of fuzzy clustering technique [27], which is robust and able to handle tolerate situation, is proposed to cluster the fuzzy data. In this approach,  $c$  segments of consumers are represented by  $c$  mutually exclusive fuzzy clustering sets,  $\tilde{\delta}_1$ ,

$\tilde{\delta}_2, \dots, \tilde{\delta}_c$ , such that  $\tilde{\delta}_1 \cup \tilde{\delta}_2 \dots \cup \tilde{\delta}_c = \tilde{Y}$  and  $\tilde{\delta}_1 \cap \tilde{\delta}_2 \dots \cap \tilde{\delta}_c = 0$ . For  $i=1,2,\dots,c$ , the ideal point of the  $i$ -th segment,  $\tilde{w}(i)$ , is defined as the center of the  $i$ -th fuzzy clustering set  $\tilde{\delta}_i$ , which can be determined by solving equation (14):

$$J_{FC}(\tilde{W}, \tilde{Y}) = \sum_{j=1}^n \sum_{i=1}^c \mu_i(\tilde{y}(j)) \left\{ 1 - \exp\left(-bd_{LR}^2(\tilde{y}(j), \tilde{w}(i))\right) \right\}, \quad (14)$$

where  $\tilde{W} = [\tilde{w}(1), \tilde{w}(2), \dots, \tilde{w}(c)]$ ;

$\tilde{w}(i) = [\tilde{w}_1(i), \tilde{w}_2(i)] = \left[ (m_{w_1}(i), \alpha_{w_1}(i), \beta_{w_1}(i)), (m_{w_2}(i), \alpha_{w_2}(i), \beta_{w_2}(i)) \right]$  represents the ideal point of the  $i$ -th segment;

$$\mu_i(\tilde{y}(j)) = \frac{\left[ \frac{1}{d_{LR}^2(\tilde{y}(j), \tilde{w}(i))} \right]^{\frac{1}{m-1}}}{\sum_{k=1}^c \left[ \frac{1}{d_{LR}^2(\tilde{y}(k), \tilde{w}(i))} \right]^{\frac{1}{m-1}}} \quad (15)$$

is the fuzzy membership of the fuzzy data,  $\tilde{y}(j)$ , with respect to the  $\tilde{\delta}_i$ , and  $m > 1$  is the index of the fuzziness;

$$b = \frac{\sum_{j=1}^n d_{LR}^2(\tilde{y}(j), \tilde{Y})}{n} \quad \text{with} \quad \tilde{Y} = \frac{\sum_{j=1}^n \tilde{y}(j)}{n}, \quad (16)$$

and

$$d_{LR}^2(\tilde{y}(j), \tilde{w}(i)) = \frac{1}{3} \sum_{k=1}^2 \left\{ (m_{y_k}(j) - m_{w_k}(i))^2 + \left[ (m_{y_k}(j) - l\alpha_{y_k}(i)) + (m_{w_k}(j) - l\alpha_{w_k}(i)) \right]^2 \right. \\ \left. + \left[ (m_{y_k}(j) - \gamma\beta_{y_k}(i)) + (m_{w_k}(j) - \gamma\beta_{w_k}(i)) \right]^2 \right\}; \quad (17)$$

as well as  $l = \int_0^1 L^{-1}(w)dw$  and  $\gamma = \int_0^1 R^{-1}(w)dw$  is the fuzzy distance between the two fuzzy coordinates  $\tilde{y}(j)$  and  $\tilde{w}(i)$  [31]. When  $m$  is smaller, the fuzzy membership

formulated in equation (15) is more sensitive to the fuzzy distance between  $\tilde{y}(j)$  and  $\tilde{w}(i)$ .

The membership functions  $\mu_i(\tilde{y}(j))$  for  $i=1,2,\dots,c$  and  $j=1,2,\dots,n$  as defined in (15) is in the interval  $[0,1]$ , where  $\sum_{i=1}^c \mu_i(\tilde{y}(j))=1$  for all  $\tilde{y}(j)$  in  $\tilde{Y}$ . The fuzzy data  $\tilde{y}(j)$  belongs to  $\tilde{\delta}_i$  with the highest membership value among all other  $\tilde{\delta}_k$  such that  $\mu_i(\tilde{y}(j)) > \mu_k(\tilde{y}(j))$  for all  $k = 1,2,\dots,i-1,i+1,\dots,c$ .

Determining  $c$  involves management subjective judgment, as the number of segments responses to marketing variables such as the nature of competitive activity, management resources, marketing strategies, the number of competitors etc. Also, the number of segments encompasses the complex case of the production line and the manpower for manufacturing. As determining the number of segments is not the main contribution of this research, readers can refer to [49].  $J_{FC}(\tilde{W}, \tilde{Y})$  can be solved by using the gradient based searching method [30], which may only converge to a local minimum, and is sensitive to the values of the initial selected values. It has been shown in many researches that the stochastic method, genetic algorithm (GA), can be performed well on solving various fuzzy clustering problems [41-44], therefore the GA is proposed in this research to search for the ideal points  $\tilde{W}$  of the segments. The detailed procedures of the GA namely GA-JFC is given in the appendix.

## 2.5 Determination of design strategy

The ideal point,  $\tilde{w}(i) = \{\tilde{w}_1(i), \tilde{w}_2(i)\}$ , can be used to develop a new product, which can satisfy the interests of all the potential customers inside the fuzzy cluster  $\tilde{\delta}_i$ . As the dimensions of  $\tilde{w}(i)$  is only against the two principle components,  $f_1$  and  $f_2$ , it is necessary to map the ideal point,  $\tilde{w}(i)$ , from two dimensions into  $d$  dimensions with respect to the  $d$  desired customer requirements. Equations (10) to (12) can be used to map the ideal point,  $\tilde{w}(i)$ , from two dimensions into a  $d$  dimensions:

$$\begin{aligned}\tilde{w}'(i) &= \{\tilde{w}'_1(i), \tilde{w}'_2(i), \dots, \tilde{w}'_d(i)\} \\ &= \left\{ \left( m_{\tilde{w}'_1}(i), \alpha_{\tilde{w}'_1}(i), \beta_{\tilde{w}'_1}(i) \right), \left( m_{\tilde{w}'_2}(i), \alpha_{\tilde{w}'_2}(i), \beta_{\tilde{w}'_2}(i) \right), \dots, \left( m_{\tilde{w}'_d}(i), \alpha_{\tilde{w}'_d}(i), \beta_{\tilde{w}'_d}(i) \right) \right\}\end{aligned}$$

where  $i=1,2,\dots,c$ . Based on the  $c$  ideal points in  $d$  dimensions, the designers can formulate proper design strategies.

The operations of the five steps discribed in Sections 2.1 to 2.5 for the methodology are summarized in Figure 3.

### **[Figure 3: The operations of the five steps for the methodology]**

## 3. Case study: digital camera design

To illustrate the operations of the proposed methodology, a case study of identification of ideal points of digital camera design was conducted. The ten customer requirements regarding digital camera design,  $D_1, D_2, \dots$  and  $D_{10}$ , which are shown in Table 1, were used in this case study. These customer requirements were initiated by experienced product designers and are commonly to design new digital cameras. Defining these customer requirements can be referred to House of quality for digital camera design [48]. 50 pieces of survey data were collected from interviewing 50 potential customers by a

questionnaire, which was designed based on the ten customer requirements. Regarding the survey, each potential customer graded each customer requirement into five attitude levels, which indicate the importance for buying a new digital camera. They are defined as ‘very unimportant’, ‘unimportant’, ‘neutral’, ‘important’ and ‘very important’. These five attitude levels however are all fuzzy, as ambiguity and imprecision inherent to human perception. As fuzziness exists in customer requirements, these five attitude levels are defined using triangular fuzzy numbers as illustrated in Table 2 and Figure 4. Two issues regarding identification of one and three ideal points for digital camera designs are considered in Section 3.1 and 3.2 respectively:

**[insert Table 1: The ten customer requirements for digital camera design]**

**[insert Table 2: Consumers’ five attitude levels in fuzzy numbers]**

**[insert Figure 4: Five consumers’ five attitude levels in triangular fuzzy numbers]**

### **3.1 Identification of one ideal point**

GA-FPCA discussed in Section 2.2 was first used to transform the customer survey data with the dimensions of ten customer requirements into two dimensions. As transforming the customer survey data into two dimensions, it can be plotted on a perceptual plot as shown in Figure 5. If the product designers intend to design a digital camera to satisfy all the potential customers regarding their attitudes, only one segment is enough. Then, the ideal point of the only one segment can be identified by the GA-JFC as discussed in Section 2.4. It is also the ideal point for all the customers for buying the new digital camera. The ideal point,  $\tilde{W}(1)$ , is illustrated as the circle plus star point in the perceptual plot (Figure 5) and is given by a fuzzy two-dimensional coordinate  $(\tilde{W}_1(1), \tilde{W}_2(1))$ ,

where  $\tilde{W}_1(1) = (10.2494, 0.0921, 0.1592)$  and  $\tilde{W}_2(1) = (13.1960, 0.1033, 0.1198)$ .

**[insert Figure 5: Perceptual plot with one ideal point]**

Based on this ideal point of the segment, the customer requirements of the digital camera, which can satisfy the potential customers, can be identified. After the decompression of this ideal point from two dimensions into ten dimensions, the customer requirements of the digital camera with respect to the original ten dimensions which can satisfy all potential customers can be illustrated. The ten customer requirements based on the ideal point for all the customers for buying the new digital camera are summarized in Table 3.

Based on the ideal point, the product designers can formulate a design strategy for the new digital camera design, which can satisfies the potential customers.

**[insert Table 3: Customer requirements of the new digital cameras with one ideal point]**

**3.2 Identification of three ideal points**

When the product designers manage to determine the ideal points of three digital cameras, it is necessary for them to identify three segments. After the three segments are identified, the three ideal points, which can represent the customer requirements of three customer groups regarding the three segments, can be determined. Three segments and their ideal points, point A, point B and point C, can be identified, where both the segments and the ideal points are illustrated in Figure 6. Also, a competitive product of digital camera Z is illustrated in Figure 6. It can be noticed clearly that the digital camera Z is near the ideal point A. Therefore, it is not worth to manufacture a digital camera represented by the ideal point A of which the customer requirements are quite near the competitor's



products, Z. The product designers may only manufacture two digital cameras represented by the ideal points, B and C, which have not been developed in the market.

**[insert Figure 6: Perceptual plot with three ideal points]**

Based on the two ideal points B and C, the product designers can formulate two design strategies respectively for the two new digital camera designs, which can satisfy the potential customers in segments B and C. The customer requirements of the two digital cameras regarding the ideal points B and C are summarized in Table 4. This information can be used by the product designers to determine the appropriate engineering attributes for designing the digital camera, which include ‘battery weight’, ‘storage media support’, ‘resolution support’, ‘aperture exposure control’, ‘LCD size’ etc as discussed in [48]. Based on the functional models which are used to map engineering attributes of digital camera into customer requirements of digital camera, digital cameras with appropriate engineering attributes, which can satisfy customer requirements regarding ideal points B and C, can be designed.

**[insert Table 4: Customer requirements of the new digital cameras with the ideal points, B and C]**

#### **4. Conclusion**

The paper presented a new methodology to determine market segments and ideal points for customer requirements which are essential for new product design. As customers’ data regarding customer requirements are high dimensions and exist fuzziness, this methodology is integrated with a fuzzy dimension compression technique and a fuzzy clustering technique, in order to process and analyze the collected customers’ data. First the customers’ data regarding customer requirements is compressed from high

dimensions to lower dimensions, so that the customers' data can be analyzed and present on a perceptual plot. Then the customers' data is clustered into many segments and the ideal points of the segments are identified as the ideal points of the consumers. In the perceptual plot, the product designers can formulate proper design and marketing strategies for the new products. A case study of digital camera design demonstrated that the proposed methodology is capable to determine market segments and ideal points of individual segments for new product design.

In the future, we will enhance the technique of the compression fuzzy data from high dimensions to low dimensions by using the fuzzy neural network method [45]. It is expected that better data compression with better accuracy can be achieved.

### **Acknowledgement**

The work described in this paper was partly supported by a grant from The Hong Kong Polytechnic University, China (Project No. G-YJ09).

### **References**

- [1] Aaker D.A., Strategic Market Management, New York: John Wiley and Son, 2001.
- [2] Myers J.H., Segmentation and positioning for strategic marketing decisions, Chicago: American Marketing Association, 1996.
- [3] Smith W.R., Product differentiation and market segmentation as an alternative marketing strategy, *Journal of Marketing* 1956; 21(1): 3-8.
- [4] Anderson C., Vincze J.W., Strategic Marketing Management, New York: Houghton Mifflin, 2000.
- [5] Green P.E., Carmone F.J., Smith S.M., *Multidimensional Scaling: Concepts and Applications*, Allyn and Bacon, 1989

- [6] Aaker D.A., Strategic Market Management, New York: John Wiley and Son, 2001.
- [7] Green P.E., Carroll J.D., Goldberg S.M., A general approach to product design optimization via conjoint analysis, *Journal of Marketing* 1981; 45: 17-37.
- [8] Butler J.C., Dyer J.S., Jia J., Tomak K., Enabling e-transactions with multi-attribute preference models, *European Journal of Operational Research* 2008; 186(2): 748-765.
- [9] Aguado D., Montoya T., Borrás L., Seco A. and Ferrer J., Using SOM and PCA for analyzing and interpreting data from a P-removal SBR, *Engineering Applications of Artificial Intelligence* 2008; 21: 919-930.
- [10] Rojanavas P. Dam H.H., Abbass H.A., Lokan C. and Pinngern O., A self-organized, distributed, and adaptive rule-based induction system, *IEEE Transactions on Neural Networks* 2009; 20(3): 446-459.
- [11] Muller D., Self organized mapping of data clusters to neuron groups, *Neural Networks* 2009; 22: 415-424.
- [12] Bursco M.J., Cradit J.D., Stahl S., A simulated annealing heuristic for a bicriterion partitioning problem in marketing segmentation, *Journal of Marketing Research* 2002; 39(1): 99-109.
- [13] Leung P.L., Lau K.N., Estimating the city-block two-dimensional scaling model with simulated annealing, *European Journal of Operational Research* 2004; 158: 518-524.
- [14] Bursco M.J., Cradit J.D., Stahl S., Multicriterion clusterwise regression for joint segmentation settings: an application to customer value, *Journal of Marketing Research* 2002; 40(2): 225-234.

- [15] Boone D.S., Roehm M., Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion, *Marketing Letters* 2002; 13(4): 317-333.
- [16] Boone D.S., Roehm M., Retail segmentation using artificial neural networks, *International Journal of Research in Marketing* 2002; 19(3): 287-301.
- [17] Cowgill M.C., Harvey R.J., Watson L.T., A genetic algorithm approach to cluster analysis, *Computers and Mathematics with Applications* 1999; 37(7): 99-108.
- [18] Liu H.H., Ong C.S., Variable selection in clustering for marketing segmentation using genetic algorithms, *Expert Systems with Applications* 2008; 34: 502-510.
- [19] Wu K.L., Yang M.S., Alternative c-means clustering algorithms, *Pattern Recognition* 2002; 35: 2267-2278.
- [20] Nguyen H.T., Wu B., *Fundamentals of Statistics with Fuzzy Data*, Springer, 2004.
- [21] Sun C.M., Wu B., New statistical approaches for fuzzy data, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2007; 15: 89-106.
- [22] Zadeh L.A., Fuzzy sets, *Information and Control*, 1965; 8: 338-353.
- [23] Morrison D.F., *Multivariate statistical methods*, New York, McGraw-Hill, 1976.
- [24] Hathaway R.J., Bezdek J.C., Pedrycz W., A parametric model for fusing heterogeneous fuzzy data, *IEEE Transactions on Fuzzy Systems* 1996; 4: 270-281.
- [25] Pedrycz W., Bezdek J.C., Hathaway R.J., Rogers G.W., Two nonparametric models for fusing heterogeneous fuzzy data, *IEEE Transactions on Fuzzy Systems* 1998; 6: 411-425.

- [26] Bock H.H., Diday E., Analysis of symbolic data. Exploratory Methods for Extracting Statistical Information From Complex Data, Berlin Germany: Springer Verlag, 2000.
- [27] Lauro C.N., Palumbo F., Principal component analysis of interval data: a symbolic data analysis approach, *Computer Statistics* 2000; 15(1): 73-87.
- [28] Rodriguez O., Diday E., Winsberg S., Generalization of principal components analysis to histogram data, 4<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases, 2000.
- [29] Denoeux T., Masson M.H., Principal component analysis of fuzzy data using autoassociative neural networks, *IEEE Transactions on Fuzzy Systems* 2004; 12(3): 336-349.
- [30] Hung W.L., Yang M.S., Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation, *Fuzzy Sets and Systems* 2005; 150: 561-577.
- [31] Yang M.S., Ko C.H., On a class of fuzzy c-number clustering procedures for fuzzy data, *Fuzzy Sets and Systems* 1996; 84: 49-60.
- [32] Schnatter S.F., On statistical inference for fuzzy data with applications to descriptive statistics, *Fuzzy Sets and Systems* 1992; 50: 143-165.
- [33] Zimmermann H.J., *Fuzzy Set Theory and its Applications*, Kluwer, Dordrecht, 1991.
- [34] Baldi P., Hornik K., Neural networks and principal component analysis: learning from examples without local minima, *Neural Networks* 1989; 2: 53-58.
- [35] Baldi P., Hornik K., Learning in linear neural networks: a survey, *IEEE Transactions of Neural Networks* 1995; 6: 837-856.

- [36] Dhar V., Chou D., A comparison of nonlinear methods for predicting earnings surprises and returns, *IEEE Transactions on Neural Networks* 2001; 12(4): 907-921.
- [37] Farag W.A., Quintana V.H., Lambert-Torres G., A genetic-based neuro-fuzzy approach for modeling and control of dynamical systems, *IEEE Transactions on Neural Networks* 1998; 9(5): 756-767.
- [38] Chen Y.M., O'Connell R.M., Active power line conditioner with a neural network control, *IEEE Transactions on Industry Applications* 1997; 33(4): 1131-1136.
- [39] Lin C.T., Jou C.P., Controlling chaos by GA-based reinforcement learning neural network, *IEEE Transactions on Neural Networks* 1999; 10(4): 846-859.
- [40] Kwon Y.K., Moon B.R. , A hybrid neurogenetic approach for stock forecasting, *IEEE Transactions on Neural Networks* 2007; 18(3): 851-864.
- [41] Gesu V.D., Bosco G.L., A genetic integrated fuzzy classifier, *Pattern Recognition Letters* 2005; 26: 411-420.
- [42] Kaya M., Alhajj R., Genetic algorithm based framework for mining fuzzy association fuzzy, *Fuzzy Sets and Systems* 2005; 152: 587-601.
- [43] Pakbira M.K., Bandyopadhyay S., Maulik U., A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification, *Fuzzy Sets and Systems* 2005; 155: 191-214.
- [44] Zhou E., Khotanzad A., Fuzzy classifier design using genetic algorithms, *Pattern Recognition* 2007; 40: 3401-3414.
- [45] Duniak J., Wunsch D., Fuzzy number neural network, *Fuzzy Sets and Systems* 1999; 108: 49-58.

- [46] Goldberg D.E., Genetic Algorithms in Search, Optimization and Machine Learning, United States of America: Addison Wesley Longman, Inc, 1989.
- [47] Muhlenbein H., Voosen D.S., Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization, *Evolutionary Computation* 1993; 1(1): 25-49.
- [48] Kwong C.K., Chen Y., Bai H., Chan D.S.K., A methodology of determining aggregated importance of engineering characteristics in QFD, *Computer and Industrial Engineering* 2007; 53: 667-679.
- [49] Y. Wind, Issues and advances in segmentation research, *Journal of Marketing Research*, vol. 15, pp. 317-337, 1978.

## Appendix

### The algorithm of GA-NPCA:

**Begin**

$i=1$

*Initialize the population  $P(i)$*

**Repeat**

*Evaluate the fitness of the individuals:  $f(i)=eval(P(i))$*

$P(i+1)=selection(P(i),f(i))$

$P(i+1)=crossover(P(i+1))$

$P(i+1)=mutation(P(i+1))$

$i=i+1$

**Until**  $i=pre\text{-}defined\ number\ of\ generations$

**End.**

GANPCA is similar to the classical GA [46]. First, the first population of solutions  $P(1)$  is randomly initialized, and the generational number  $i$  is set as  $i=1$ . Real encoding is adopted such that each solution is coded by

$$\mathbf{B} = [B_{ij}; i = 1, 2, \dots, d; j = 1, 2, \dots, n]$$

where  $B_{ij} \in [-1, 1]$ ,  $n$  is the number of interviewers, and  $d$  is the number of customer requirements. Therefore the number of variables in each solution is  $n \times d$ .

After that, solutions in the population are assigned fitness values  $f(i)$  according to the qualities of the solutions based on the cost function (13). The objective of cost function (13) is to minimize the input-output error of the NPCA  $E(\mathbf{B})$  by searching  $\mathbf{B}$ . Then a new population  $P(i+1)$  is formed by selecting the better solutions with better fitness values to reproduce more often than the worst ones. The number of generations is updated as  $i=i+1$ . And then the population evolves towards better solutions by crossover and mutation. For selection, the approach of roulette-wheel, which is one of the most common selection methods used for selecting solutions to perform reproduction operations [46], is used for solution selection. For crossover, a commonly used crossover



operation for real encoding representation, intermediate crossover [47], is used in the genetic algorithm. Mutation is carried out by randomly changing one or more variables of the selected solution between its pre-defined upper and lower bound. The population is evolved and improved by the next generation until the pre-defined number generations is reached.

**The algorithm of GA-JFC:**

The procedures of GA-JFC are identical to GA-NPCA except the solution representation and the cost function used on the algorithm.

For the solution representation, real encoding is adopted such that each solution is coded by:

$$\tilde{W} = [\tilde{w}(1), \tilde{w}(2), \dots, \tilde{w}(c)],$$

where  $\tilde{w}(i) = [\tilde{w}_1(i), \tilde{w}_2(i)] = [(m_{w_1}(i), \alpha_{w_1}(i), \beta_{w_1}(i)), (m_{w_2}(i), \alpha_{w_2}(i), \beta_{w_2}(i))]$  with  $i=1,2,\dots,c$ . The number of variables in each solution is depended on the number of segments,  $c$ , defined.

The fitness of the solutions is assigned by evaluating with the cost function (14).