A fuzzy clustering approach for determination of ideal points of new products

K.Y. Chan
Department of Electrical and Computer Engineering,
Curtin University, Perth, Australia
kit.chan@curtin.edu.au

Abstract— **Prior to manufacture a new products, consumers with similar purchasing attitudes are grouped into clusters of which their central points are used as ideal points for new product development. However, many clustering methods ignore the fuzziness of consumers in purchasing products or conducing survey. This paper presents a new method which integrates a fuzzy data processing technique for dimension reduction of customer attributes and a fuzzy clustering technique for grouping consumers with similar purchasing attributes. Hence, the central points of each group are treated as the ideal points for new product development. The effectiveness of the proposed method is demonstrated based on a case study of new digital camera design.**

Keywords: Ideal points, fuzzy clustering, fuzzy data, new product development

I. INTRODUCTION

In product planning, designers usually group consumers with similar purchasing attitudes in different segments [1, 2], and the central points of the segments are used as ideal points to reflect the purchasing attitudes of the consumers [12]. Ideal points can be used for development of new products. In addition, they forge the closer relationships between the customers and the product designers. Furthermore, they can be used by the companies to determine the particular competitive strategies (e.g. differentiation, low cost, or focus strategy) [2].

Artificial intelligent methods, including simulated annealing [14], multicriterion clusterwise regression [6], neural networks [5], evolutionary algorithms [15], numerical clustering algorithms [22], etc have been proposed to segment the consumers into clusters. In those approaches, a consumer is classified into a particular cluster if the distance between the consumer to the ideal point of the cluster is the shortest comparing with the others. However, Euclidean distance has commonly been used as the distance measure, and fuzziness of the consumers' data cannot be considered under the distance measure**.** In fact, fuzziness always exists in human feeling or attitudes while purchasing [19]. Therefore, consideration of fuzziness of each consumer is necessary and fuzzy distance measure is necessary to be used for clustering.

This paper presents a methodology in order to perform marketing segmentation based on fuzzy data, which are collected from surveying consumers' purchasing attitudes. In the methodology, the fuzzy data is first compressed into two-dimensions, in order to be presented on a perceptual plot for fuzzy clustering purposes. Principal component analysis (PCA) [16] can be used to underlying structure of data sets into two-dimensions. However, PCA cannot address data fuzziness. A compression method that can handle fuzzy data has been developed by [4, 17]. However, these approaches cannot tackle fuzzy data with very high dimensions. Although an extension of PCA to fuzzy data has also been proposed by generalizing classical principles of PCA [18], it only applies the standard analysis to the centers of the fuzzy data [4]. It cannot take into account the imprecision of the data in the feature extraction process, so it is likely to create local optimal solutions for the fuzzy compression.

Therefore, a fuzzy data processing method [9] is proposed here, which intends to overcome the limitations of those compression methods by finding iteratively linear features from which the original fuzzy data is recovered with minimum error using a linear transformation. The approach based on a three layer neural network which consists of two hidden neurons, and same number of inputs and outputs

After compressing the high dimensional fuzzy data into a two dimensional one, a fuzzy clustering method [13] is used to cluster the fuzzy data into various segments. It clusters fuzzy data into segments which have similar characteristics using the assigned partial memberships of each fuzzy data. This approach is similar to the fuzzy c-number cluster method [21], but an exponential-type distance [22] is introduced to evaluate the fuzzy distance between fuzzy data. Based on the proposed fuzzy distance measure, the method is more robust than the existing fuzzy clustering methods [21], which are strongly influenced by outliers and may give poor clustering results when the fuzzy data includes large numbers of different cluster sample sizes. The central points of clusters, which reflect the interests of the group of consumers for new products, can be identified.

This paper is organized as follows: Section 2 presents a new methodology, which mainly involves fuzzy data processing and fuzzy clustering method, is proposed to identify the ideal points based on multi-dimensional customers' survey data. Section 3 presents a case study of digital camera design in order to demonstrate the outcomes of the proposed methodology. Finally, conclusions are given in section 4.

II. COMMONLY USED METHOD FOR INDENTIFYING NEW PRODUCT OPPORTUNITIES

The identifications of ideal points of products are important, as it allows marketing personnel to understand the marketing positions of various customers and identify possible product opportunities. It involves the following three steps (see pp. 9-11 on [23]). First, it identifies puschasing attitudes from examining competitive products. As the number of puschasing attitudes is usually more than two, the collected puschasing attitude cannot be depicted with a two-dimensional graph. The second step is to transform multi- puschasing attitudes into two major components. Then, the third step is to construct a two-dimensional perceptual map in order to graphically represent competing alternatives in Euclidean space. This allows marketing personnel to identify regions of new product opportunities.

Firms usually use perceptual maps to help them develop a market positioning strategy for their product or service. As the maps are based on the perception of the customer, they are sometimes called perceptual maps which show existing products and services positioned in the market so that the firm can decide where they would like to place their product. Firms have two options they can either position their product so that it fills a gap in the market or if they can compete against their competitors where existing products have placed their product.

However, solely using perceptual maps may produce misleading results, as the data used on generating the perceptual maps is usually from buyers' survey and approaches for producing perceptual maps are usually based on the analysis of Euclidean distances of buyers' customer requirements of which fuzziness of buyers' surveys is not considered. Hence, misleading results is likely to be produced. The following Section III presents a fuzzy approach in order to determine identical points of new products of which buyers' fuzziness can be addressed.

III. IDEAL POINT IDENTIFICATION BASED ON CUSTOMERS' SURVEY DATA

As customer survey data is fuzzy, the methodology involved two main steps namely fuzzy data processing and fuzzy clustering, where fuzzy data processing is used to transfer customer survey data from high dimensions into two dimensions and fuzzy clustering is used to identify ideal points based on processed customer survey data. The mechanisms of the fuzzy data processing and fuzzy clustering are given in Subsection III.A and Subsection III.B.

A. Fuzzy data processing

First n pieces of survey data is collected from n interviewers by asking the interviewers to fill in the designed questionnairs related to the desired d design attributes of the product. Then, a set of survey data can be produced with respect to the d design attributes of the product. As people's purchasing behaviours are human intuition, the n pieces of survey data can be considered as fuzzy data which are denoted as

$$\tilde{X} = \left\{ \tilde{x}(1), \tilde{x}(2),..., \tilde{x}(n) \right\} \tag{1}$$

where the j-th fuzzy data set of the j-th interviewer is denoted as:

$$\tilde{x}(j) = \left(\tilde{x}_1(j), \tilde{x}_2(j),..., \tilde{x}_d(j) \right); \tag{2}$$

$\tilde{x}_i(j) = \left(m_{x_i}(j), \alpha_{x_i}(j), \beta_{x_i}(j) \right)$ is the i-th puschasing attitude of the j-th interviewer in buying the product, with $i=1,2,…,d$ and $j=1,2,…,n$; $m_{x_i}(j)$ 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; $\tilde{x}_i(j)$ is represented by the triangular fuzzy member function:

$$\tilde{x}_i(j)\big|_x = \begin{cases} L\left(\dfrac{m_{x_i}(j) - x}{\alpha_{x_i}(j)} \right) \text{ for } x \le m_{x_i}(j) \\ R\left(\dfrac{x - m_{x_i}(j)}{\beta_{x_i}(j)} \right) \text{ for } x \ge m_{x_i}(j) \end{cases}, \tag{3}$$

of which L and R is a shape function ranged from 0 to 1:

$$\begin{cases} L(x) = 1 \text{ for } x = 0 \\ 0 < L(x) < 1 \text{ for } x > 0 \\ L(x) = 0 \text{ for } x = \infty \end{cases}, \tag{4}$$

and

$$\begin{cases} R(x) = 1 \text{ for } x = 0 \\ 0 < R(x) < 1 \text{ for } x < 0 \\ R(x) = 0 \text{ for } x = -\infty \end{cases}. \tag{5}$$

As this is not easy to analyze fuzzy clustering resulting for multi-dimensions, it is necessary to compress multi-dimensional fuzzy data with high dimensions into two dimensions. Here an neural network based principle component analysis (NPCA) based on [9] is proposed to compress d-dimensional fuzzy data into two-dimensional fuzzy data. In NPCA, a feedforward neural network with three layer is used, where both the input layer and output layer has d nodes and the hiden layer has two nodes. It compresses the fuzzy data with d dimensions into two dimensions through the flow between the input layer to the hidden layer. The detailed mechanism of the NPCA can be refered to [9]. In the hidden layer of the NPCA, the whole population of the compressed fuzzy data sets of the n interviewers is denoted as:

$$\tilde{Y} = \left\{ \tilde{y}(1), \tilde{y}(2),..., \tilde{y}(n) \right\} \tag{6}$$

where $\tilde{y}(j) = \left\{ \tilde{y}_1(j), \tilde{y}_2(j) \right\}$ with $j=1,2,…,n$.

All $\tilde{y}(j)$ in $\tilde{Y}$ are plotted on a two-dimensional perceptual plot with two principle components namely f_1 and f_2 as illustrated in the perceptual plot (Figure 1), where the coordinate of the fuzzy data is represented by the cross points and its fuzziness against f_1 and f_2 is represented by the dashed lines. For the j-th fuzzy data, its coordinates $\left(m_{y_1}(j), m_{y_2}(j) \right)$, and its fuzziness $\left(\alpha_{y_1}(j), \beta_{y_1}(j) \right)$ against f_1 and fuzziness $\left(\alpha_{y_2}(j), \beta_{y_2}(j) \right)$ against f_2 are shown in the perceptual plot, in which the distribution and the fuzziness of the data set can be illustrated. In the perceptual plot, fuzzy clustering can be

performed by grouping the fuzzy data sets with two dimensions, which have small fuzzy distances.
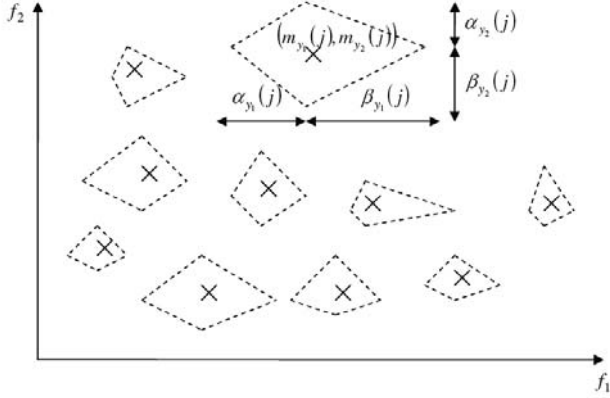

Figure 1 Perceptual plot

B. *Fuzzy clustering*

When the fuzzy data is in two dimensions, the following fuzzy clustering method is proposed to cluster fuzzy data with similar charaterics into fuzzy groups [13], in order to determine marketing strategy. The objective function of the fuzzy clustering method is defined as:

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

(7)

where c is the number of fuzzy clusters;
$\tilde{W}=\left[\tilde{w}(1),\tilde{w}(2),...,\tilde{w}(c)\right]$ in which $\tilde{w}(i)=\left[\tilde{w}_1(i),\tilde{w}_2(i)\right]=$
$\left[\left(m_{w_1}(i),\alpha_{w_1}(i),\beta_{w_1}(i)\right),\ \left(m_{w_2}(i),\alpha_{w_2}(i),\beta_{w_2}(i)\right)\right]$ represents the ideal point of the i-th fuzzy cluster;

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

(8)

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

$$b=\frac{\sum_{j=1}^{n}d_{LR}^2\left(\tilde{y}(j),\bar{\tilde{Y}}\right)}{n}$$

(9)

with $\bar{\tilde{Y}}=\dfrac{\sum_{j=1}^{n}\tilde{y}(j)}{n}$ is a constant;

$$d_{LR}^2\left(\tilde{y}(j),\tilde{w}(i)\right)=\frac{1}{3}\sum_{k=1}^{2}\left\{\left(m_{y_k}(j)-m_{w_k}(i)\right)^2+\right.$$
$$\left[\left(m_{y_k}(j)-l\alpha_{y_k}(i)\right)-\left(m_{w_k}(j)-l\alpha_{w_k}(i)\right)\right]^2$$ (10)
$$\left.\left[\left(m_{y_k}(j)+\gamma\beta_{y_k}(i)\right)-\left(m_{w_k}(j)-\gamma\beta_{w_k}(i)\right)\right]^2\right\}$$

where $l=\int_0^1 L^{-1}(w)dw$ and $\gamma=\int_0^1 R^{-1}(w)dw$ is the fuzzy distance between the two coordinates $\tilde{y}(j)$ and $\tilde{w}(i)$ [21]. J_{FC} intends to perform fuzzy clustering more robust and able to handle tolerate situation [13].

In this approach, partitions of $\tilde{Y}$ into c parts are represented by mutually fuzzy clustering sets $\tilde{\delta}_1$, $\tilde{\delta}_2$, …, $\tilde{\delta}_c$ such that $\tilde{\delta}_1\cup\tilde{\delta}_2...\cup\tilde{\delta}_c=\tilde{Y}$ and $\tilde{\delta}_1\cap\tilde{\delta}_2...\cap\tilde{\delta}_c=0$, where $\tilde{w}(i)$ is the center of the i-th fuzzy cluster $\tilde{\delta}_i$. Equivalently by the indicator functions μ_1, μ_2,...,μ_c such that $\mu_i\left(\tilde{y}(j)\right)$ for i=1,2,…,c and j=1,2,…,n as defined in (8) is a membership functions assuming values in the interval [0,1] such that $\sum_{i=1}^{c}\mu_i\left(\tilde{y}(j)\right)=1$ for all $\tilde{y}(j)$ in $\tilde{Y}$. The fuzzy data $\tilde{y}(j)$ belongs to the i-th cluster with the highest membership value among all the other clusters i.e. $\mu_i\left(\tilde{y}(j)\right)>\mu_k\left(\tilde{y}(j)\right)$ for all $k=1,2,...,i-1,i+1,...,c$. As genetic algorithm performs effectively on solving various fuzzy clustering problems [22], it is used to determine $\tilde{W}$ of the clusters.

As illustrated in Figure 2, the fuzzy central point $\tilde{w}(i)$ in the i-th fuzzy cluster $\tilde{\delta}_i$ with i=1,2,…c, which fuzzy clusters a set of interrelated fuzzy data collected from some people, is used as the fuzzy ideal point on development of a new product, which can satisfy the interests of those people inside the fuzzy cluster. Since the coordinates of the fuzzy ideal points of the fuzzy clusters on the perceptual plot are against the two principle components, it is necessary to map the two-dimensional coordinate of the fuzzy central point of each fuzzy cluster into a d-dimensional coordinate with respect to the d desired purchasing attitudes.

Then, the NPCA [9] developed in is used to map the fuzzy ideal point $\tilde{w}(i)=\left\{\tilde{w}_1(i),\tilde{w}_2(i)\right\}$ from a two-dimensional coordinate into a d-dimensional coordinate.

$$\tilde{w}'(i)=\left\{\tilde{w}_1'(i),\tilde{w}_2'(i),...,\tilde{w}_d'(i)\right\}=\left\{\left(m_{\tilde{w}_1'}(i),\alpha_{\tilde{w}_1'}(i),\beta_{\tilde{w}_1'}(i)\right),\right.$$
$$\left.\left(m_{\tilde{w}_2'}(i),\alpha_{\tilde{w}_2'}(i),\beta_{\tilde{w}_2'}(i)\right),...,\left(m_{\tilde{w}_d'}(i),\alpha_{\tilde{w}_d'}(i),\beta_{\tilde{w}_d'}(i)\right)\right\}$$

where i=1,2,…,c. By referring the c fuzzy ideal points in d-dimensions, the designers can formulate proper design strategies.
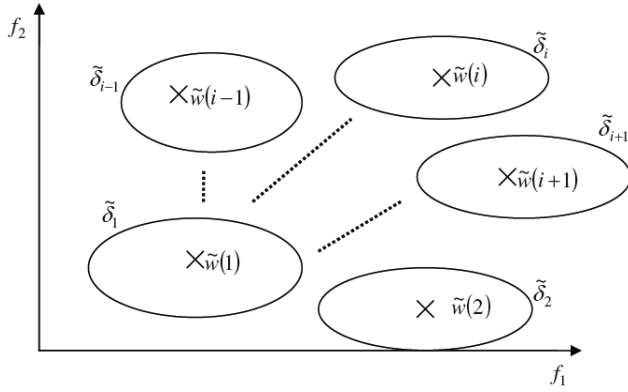
Figure 2 Fuzzy clusters shown on a perceptual plot

IV. CASE STUDY: DIGITAL CAMERA DESIGN

As an illustration of the proposed methodology, determination of ideal points of digital camera design is investigated. First 50 fuzzy data $\tilde{x}(j) = \left(\tilde{x}_1(j), \tilde{x}_2(j), ..., \tilde{x}_{10}(j)\right)$ with $j=1,2,..,50$ were collected, where the purchasing attitudes of the 10 design attributes (D1, D2,…,D10) were collected from surveying 50 interviewers by a questionnaire. The 10 design attributes (D1, D2,…, D10) are described as following:

D1) Fashion such as the size, the design and the colour
D2) Practicability such as the weight, the size and the endurance of the battery
D3) Functionality such as the LCD monitor, internal memory and ISO sensitivity
D4) Easier to use
D5) Modern of the shape of the mobile
D6) Trendy design
D7) Product price
D8) High quality
D9) High product value
D10) Digital camera's brand

Each attitude is graded into five attitude levels, which are described as strongly disagree, disagree, neutral, agree and strongly agree. These five attitude levels defining terms allow for the ambiguity and imprecision inherent to human perception. Since fuzziness exists in human attitudes, those terms are defined using triangular fuzzy numbers:

$$\tilde{A}_{\text{strongly disagree}} = (0,0,2.5), \ \tilde{A}_{\text{disagree}} = (2.5,2.5,2.5),$$

$$\tilde{A}_{\text{neutral}} = (5.0,2.5,2.5), \ \tilde{A}_{\text{agree}} = (7.5,2.5,2.5)$$

$$\text{and } \tilde{A}_{\text{strongly agree}} = (10,2.5,0).$$

To pre-defined the number of clusters, c, companies consider some product design factors such as corporate strategy, amount of investment on new product development, cash flow of companies and competition environments of market segments. After the segmentation process, the product designers estimate the market sizes of individual segments based on the segmentation results. If the market sizes of a particular segment or some segments are not large enough, the pre-defined c is reduced and the segmentation process is performed again. The process is repeated until the product designers feel comfortable about the estimated market sizes of segments.
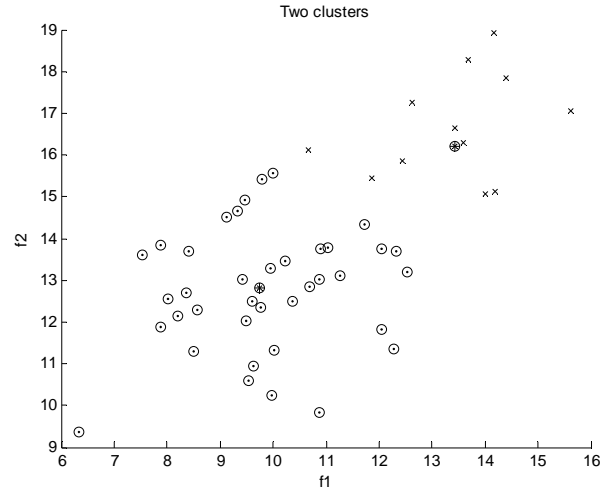

Figure 3 Perceptual plot with three ideal points

As an illustration of the proposed approach, the number of clusters c is set to be 2. The two ideal points of the two clusters can be found in the perceptual plot (as shown in Figure 3) as:

$$\tilde{W}(1) = (7.6471 \ 1.3410 \ 0.1394, \ 7.1531 \ 0.1799 \ 1.3084)$$

and $\tilde{W}(2) = (3.6632 \ 0.1293 \ 0.3769, \ 3.5974 \ 0.2348 \ 0.3021)$

$\tilde{W}(1)$ and $\tilde{W}(2)$ are illustrated as the circle plus star points in the perceptual plot.

Based on the two ideal points of the clusters, the product designers can formulate three design strategies respectively for the two new digital camera designs, which can satisfy the interviewers in the three clusters. The design attributes of the two new digital cameras are summarized in Table I and Table II.

V. CONCLUSION

The paper presented a new fuzzy clustering method for determining market segments and ideal points of consumers' purchasing attitudes toward a product based on fuzzy data. Since consumers' interests are in high dimensions and exist fuzziness, this approach is integrated with a fuzzy dimension compression technique and a fuzzy clustering technique. First the dimension compression extracts the fuzzy data of consumers' interests from high dimensions to two dimensions so that fuzzy data can be analyzed and present on a perceptual plot. Then fuzzy clustering segments the fuzzy data into many clusters and the centrals of the segments are treated as the ideal points of the new products. Based on the proposed approach, fuzziness of customers' purchasing attitudes can be addressed. A case study of digital camera design demonstrated

that the proposed methodology is capable to determine ideal points for new products.

In the future, the proposed approach will be tested by clustering with different number of segments, and the results will be reported shortly.

REFERENCES

1. Anderson, C. and Vincze, J. W., 2000. *Strategic Marketing Management*. New York: Houghton Mifflin.
2. Aaker D.A., 2001. *Strategic Market Management*. New York: John Wiley and Son.
3. Baldi P. and Hornik K., 1995. Learning in linear neural networks: a survey, *IEEE Transactions of Neural Networks*, 6, 837-856.
4. Bock H.H. and Diday E., 2000. *Analysis of symbolic data. Exploratory Methods for Extracting Statistical Information From Complex Data*, Berlin Germany: Springer Verlag.
5. Boone D.S. and Roehm M., 2002a. Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion. *Marketing Letters*, 13(4), 317-333.
6. Bursco M.J., Cradit J.D. and Stahl S., 2002b. Multicriterion clusterwise regression for joint segmentation settings: an application to customer value. *Journal of Marketing Research*, 40(2), 225-234.
7. Butler J.C., Dyer J.S, Jia J. and Tomak K., 2008. Enabling e-transactions with multi-attribute preference models. *European Journal of Operational Research*, 186(2), 748-765.
9. Denoeux T. and Masson M.H., 2004. Principal component analysis of fuzzy data using autoassociative neural networks. *IEEE Transactions on Fuzzy Systems*, 12(3), 336-349.
10. Dhar V. and Chou D., 2001. A comparison of nonlinear methods for predicting earnings surprises and returns. *IEEE Transactions on Neural Networks*, 12(4), 907-921, 2001.
11. Green P.E., Caroll J.D. and Goldberg S.M., 1981. A general approach to product design optimization via conjoint analysis. *Journal of Marketing*, 45, 17-37.
12. Green P.E., Carmone F.J. and Smith S.M., 1989. *Multidimensional Scaling: Concepts and Applications*, Allyn and Bacon.
13. Hung W.L. and Yang M.S., 2005. Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. *Fuzzy Sets and Systems*, 150, 561-577.
14. Leung P.L. and Lau K.N., 2004. Estimating the city-block two-dimensional scaling model with simulated annealing. *European Journal of Operational Research*, 158, 518-524.
15. Liu H.H. and Ong C.S., 2008. Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications*, 34, 502-510.
16. Morrison D.F., 1976. *Multivariate statistical methods*, New York, McGraw-Hill.
17. Pedrycz W., Bezdek J.C., Hathaway R.J. and Rogers G.W., 1998. Two nonparametric models for fusing heterogeneous fuzzy data, *IEEE Transactions on Fuzzy Systems*, 6, 411-425.
18. Rodriguez O., Diday E. and Winsberg S., 2000. Generalization of principal components analysis to histogram data. *4th Europeon Conference on Principles and Practice of Knowledge Discovery in Databases*.
19. Sun C.M. and Wu B., 2007. New statistical approaches for fuzzy data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15, 89-106.
20. Wu K.L. and Yang M.S., 2002. Alternative c-means clustering algorithms. *Pattern Recognition*, 35, 2267-2278.
21. Yang M.S. and Ko C.H., 1996. On a class of fuzzy *c*-number clustering procedures for fuzzy data. *Fuzzy Sets and Systems*, 84, 49-60.
22. Zhou E. and Khotanzad A., 2007. Fuzzy classifier design using genetic algorithms, *Pattern Recognition*, 40, 3401-3414.
23. Chan K.Y., Kwong C.K. and Dillon T.S., *Computational Intelligence Techniques for New Product Design*, Studies in Computational Intelligence, Springer, 2012.

Table 1 1st ideal point for the new digital camera

Design attributes	Customers' purchasing attitudes		
	Mean value	Fuzziness in left spread	Fuzziness in right spread
D1	6.8157	0.0822	0.5449
D2	6.9635	0.3219	0.3680
D3	6.9247	0.3971	0.3050
D4	5.7776	0.3994	0.2005
D5	7.1989	0.2198	0.4700
D6	6.9920	0.2493	0.4282
D7	6.0965	0.3048	0.3039
D8	6.4696	0.2330	0.3943
D9	6.1598	0.4601	0.1865
D10	6.8976	0.4325	0.2744

Table 2 2nd ideal point for the new digital camera

Design attributes	Customers' purchasing attitudes		
	Mean value	Fuzziness in left spread	Fuzziness in right spread
D1	5.3782	0.0983	0.1277
D2	5.3205	0.0874	0.1557
D3	5.2345	0.0827	0.1630
D4	4.3175	0.0653	0.1432
D5	5.5831	0.0966	0.1490
D6	5.3965	0.0918	0.1485
D7	4.6412	0.0753	0.1388
D8	4.9916	0.0848	0.1377
D9	4.5779	0.0677	0.1564
D10	5.1868	0.0804	0.1663