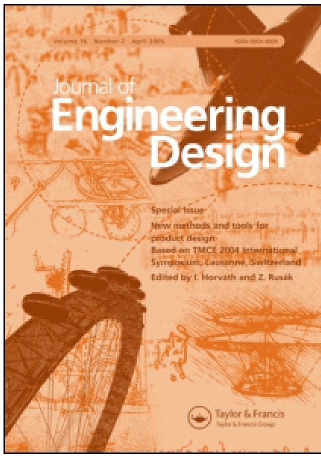


Journal of Engineering Design
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713429619

Market segmentation for product family positioning based on fuzzy clustering

Yiyang Zhang [a]; Jianxin (Roger) Jiao [a]; Yongsheng Ma [a]
[a] Nanyang Technological University, Singapore

First Published: June 2007

To cite this Article: Zhang, Yiyang, Jiao (Roger) , Jianxin and Ma, Yongsheng (2007) 'Market segmentation for product family positioning based on fuzzy clustering', Journal of Engineering Design, 18:3, 227 — 241

To link to this article: DOI: 10.1080/09544820600752781
URL: http://dx.doi.org/10.1080/09544820600752781

Journal of Engineering Design
Vol. 18, No. 3, June 2007, 227–241

Market segmentation for product family positioning based on fuzzy clustering

YIYANG ZHANG, JIANXIN (ROGER) JIAO* and YONGSHENG MA

Nanyang Technological University, Singapore

To compete in the marketplace, manufacturers have been seeking for expansion of their product lines by providing product families. One difficulty for product family positioning is that diverse customer needs can no longer be satisfied by a mass marketing approach. Realizing the importance of customer purchase behaviours for product family positioning, this paper proposes a fuzzy clustering-based market segmentation approach. With a focus on engineering characteristics, the fuzzy clustering-based market segmentation helps plan the right products to target segments effectively and efficiently. An application of the proposed methodology in a consumer electronics company producing vibration motors is reported. The evaluation of the proposed methodology is also discussed.

Keywords: Product family; Mass marketing; Market segmentation; Fuzzy clustering

1. Introduction

To keep the competitive advantage, companies intend to provide product variety by differentiating their product lines. Although high product variety does stimulate sales, companies with expending products face the challenge of controlling inventory costs and providing high quality and good delivery performance for the customers (Hofer and Halman 2005). In addition, high variety results in the proliferation of products and processes, and in turn inefficiencies in manufacturing (Child *et al.* 1991). With platform-based product family design, it turns out that some of the product variants may be more preferred as predicted, while others, although equally sound in technical terms, may not be favoured by the customers (Jiang and Allada 2005; Thevenot and Simpson 2006). As a result, it is imperative for manufacturing companies to position their product families properly while balancing the trade-offs between the diversity of customer needs and manufacturing costs (Olewnik and Lewis 2006).

One difficulty for product family positioning is that the customer needs are too diverse for any single marketing mix to satisfy everyone. In practice, businesses from all industry sectors use market segmentation to facilitate their strategic planning. A range of benefits have been identified in marketing literature for businesses pursuing a segmentation approach.

*Corresponding author. Email: jiao@pmail.ntu.edu.sg

Journal of Engineering Design
ISSN 0954-4828 print/ISSN 1466-1837 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/09544820600752781

The underlying logic is that segmentation can enhance marketing effectiveness and improve an organization's ability to capitalize on marketing opportunities (Beane and Ennis 1987). In theory, the segmentation posits that groups of customers with similar needs and purchasing behaviours are likely to demonstrate a more homogeneous response to products and marketing programmes that target specific customer groups. Thus, businesses adopting a market segmentation approach can enhance their organizational performance (Kotler 1994). Allowing the identification of homogeneous customer requirements and characteristics, enterprises are likely to position their product families to satisfy each target customer group, while keeping economy of scale in product fulfilment (McDonald and Dunbar 1995).

1.1 *Problem description*

A critical issue in successful market segmentation is the selection of segmentation variables. Segmentation variables can be broadly classified into general variables and product-specific variables (Wedel and Kamakura 1997). The general variables include the customer demographic, geographic and psychographic characteristics. Many researchers have devoted themselves to using general variables to partition customers because the general variables are intuitive and easy to operate (Hammond *et al.* 1996). The purpose of grouping the customers by using general variables is to represent the diverse markets according to the customers' characteristics. Similarity in customers' characteristics performs as an indicator to provide product offerings to the market where these customers belong to.

In practice, however, it is doubtful to assume that customers with similar demographic, geographic and psychographic characteristics will exhibit similar purchasing behaviours. Today's customers may gain abundant product information from various medium and channels. Even within a group with similar characteristics, the customers' preferences may vary a lot due to the uniqueness of personality. It is difficult to measure the customers' purchasing patterns using general variables alone. Furthermore, most general variables refer to personal or private information such as the income, occupation and address, which make data collection inhibitive, thus jeopardizing the data credibility. Even though private information may be obtained, the information itself fluctuates and varies over time. For example, occupation, income, and marital status data collected now might not be valid 2 years later if no continuous revision is performed (Drozdenko and Drake 2002). All these issues make market segmentation using general variables questionable.

For the product family positioning problem, product offerings are constructed directly from discrete product attribute values (Jiao and Zhang 2005). Customer preferences are reflected by their choices on different attribute value combinations. In this regard, market segmentation should take into account engineering concerns, such that the enterprises can organize their design, manufacturing and marketing activities to cope with a high level of product variety fulfilment.

1.2 *Strategy for solution*

Clustering analysis is the most popular technique used for marketing segmentation. Clustering refers to a process of grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of objects that are similar to one another within the same cluster, yet dissimilar to the objects in other clusters (Han and Kamber 2001). Considering more engineering concerns, market segmentation for product family positioning inevitably deals with different types of product attributes. The product attributes are usually presented in the form of numerical, binary or nominal types. It is difficult to handle different types of variables at the same time through clustering analysis. In addition, the number of clusters affects the downstream planning of product and process platforms. Too spread out clusters may not be sufficient to utilize the advantage of product family design; while too low-level aggregation may sacrifice the customers' satisfaction.

This research employs a fuzzy clustering approach to implement market segmentation for the product family positioning problem. Compared with the K-means method, one of the most popular techniques applied to segmentation, fuzzy clustering, excels in partitioning different customers according to a hierarchical decomposition of the given set of objects (Deciu *et al.* 2005). Different segments can be derived by adjusting the similarity threshold, thus dealing with the granularity issue inherent in product family positioning.

The remainder of this paper proceeds as follows. In the next section, various existing approaches to market segmentation are reviewed. Section 3 presents the formulation of market segmentation for the product family positioning problem. The implementation of market segmentation is discussed in section 4. Section 5 reports a case study of motor product family positioning. The evaluation of the proposed method is discussed in section 6 and the paper is concluded in section 7.

2. Literature review

Market segmentation has evoked the interests of both practitioners and academicians. To implement market segmentation, it is most important to determine the segmentation variables. According to variables used for segmentation, the related research work can be classified into three categories as follows.

2.1 *Choice-based segmentation*

The choice-based (purchase-based) segmentation is focused on product-specific variables. For years, catalogue companies and other direct marketers have used recency, frequency, and monetary value analysis (RFM) to segment their customers and optimize the purchase response rates of their marketing efforts (Hughes 1994). Although RFM has been challenged by innovative, conceptual approaches, the marketers continue to rely on RFM because of its simplicity and cost-effectiveness. The variables involved in RFM are the purchase frequency and the total monetary value. RFM is suitable for those retail and service businesses with relative high purchase frequency, such as take-out and delivery restaurants. However, for those industrial businesses with relative low purchase frequency, the variables involved in RFM are doubtful. Tsai and Chiu (2004) proposed a purchase-based market segmentation methodology to cluster customers. In their work, the customers are clustered based on the distance between their purchased items. The number of clusters is pre-defined and Genetic Algorithm (GA) is used to produce the best solution. A designed RFM model is used to analyse the profitability of each cluster.

2.2 *Benefit/value-based segmentation*

Benefit-based segmentation is also focused on product-specific variables. Differing from choice-based segmentation, benefit-based segmentation addresses 'why' the customers choose certain products rather than 'what' products are chosen. Heuvel and Devasagayam (2004) proposed a benefit-based segmentation method to help market segmentation based on the intrinsic values that customers derive from the products. It seeks to find out the customers' feeling about the products rather than arbitrarily identifying customers according to their income or address. Benefit-based segmentation intends to identify the reason that the customers purchase certain products, and thus to cluster the customers with similar reason together. Marcus (1998) proposed a market segmentation method by customer value matrix. In his work, the customers are grouped according to the benefits they gain from the products. The customers are then distinguished as high-value and low-value groups and market strategies are determined accordingly to adapt to different groups. Since the customers always choose the products that benefit them most, this method gives a more reasonable way to cluster the customers based on their purchasing behaviours.

2.3 *Demographic segmentation*

Demographic segmentation is focused on the general variables. In the study by Natter (1999), an artificial neural network clustering method is proposed. This methodology incorporates both market segmentation and discriminant analysis of the segments. Customers showing similar characteristics are assigned into the same cluster. Kou *et al.* (2002) introduced a two-stage method encompassing self-organizing feature maps and the K-means algorithm. Their two-stage method outperforms the conventional two-stage method adopting multivariate analysis procedures. Jonker *et al.* (2004) proposed a joint optimization approach to address two issues (i.e. customer clustering and optimal policy). In their work, the customers are clustered using those variables describing customers' characteristics, and the appropriate policy is planned to accommodate each segment.

3. Problem formulation

In the historical database of a manufacturing company, the transaction records contain information about which customers choose what products. Therefore, transaction data can be summarized as C–P pairs in the form of $\langle c_s, p_t\rangle$, where c and p represent the customers and the products; and s and t stand for the customer ID and product ID, respectively. Each pair of such transaction data indicates the relationship between customers and the product offerings.

The customers can be represented by a set, $C \equiv \{c_1, c_2, \ldots, c_S\}$, where S denotes the total number of the customers. For the product offerings that can be produced, each product is deemed as a bundle of attributes and characterized by these attributes, $A \equiv \{a_1, a_2, \ldots, a_N\}$. Each attribute, $a_q|\forall q \in [1, \ldots, N]$, possesses a few possible values, $A_q^* \equiv \{a_{q1}^*, a_{q2}^*, \ldots, a_{qn_q}^*\}$. That is, $a_q =:: a_{qr}^*|\exists a_{qr}^* \in A_q^*$, where $r = 1, \ldots, n_q$, denotes the rth possible value of a_q. Customers are grouped into several clusters, noted as $X = \{\chi_1, \chi_2, \ldots, \chi_L\}$, where $\chi_l \in X|\forall l \in [1, \ldots, L]$, meaning the lth cluster. As a result, all attribute values related to a cluster can be grouped and represented by the characteristics of χ_l – the mean value of these attribute values, $\mu_l \equiv [x_1^l, x_2^l, \ldots, x_N^l]$. Allowing the mean values, all possible product offerings constitute a set, $P' \equiv \{p_1', p_2', \ldots, p_J'\}$, where J refers to the total number of products. A positioned product family, …, is a set consisting of a few selected product profiles (i.e. … $\equiv \{p_j'|j = 1, \ldots, J^\dagger\} \subseteq P'$); $\exists J^\dagger \in \{1, \ldots, J\}$ denotes the number of products contained in the positioned product family. Each product, $\forall p_j' \in \ldots$, is defined as a vector of clustered attribute values (i.e. $p_j' = [x_q^l]_N$).

4. Implementation

4.1 *Distance measure*

In general, each product offering, $p_t = [a^*_{1t}, a^*_{2t}, \ldots, a^*_{qt}, \ldots, a^*_{Nt}] \in A^*$ (where a^*_{1t} means the first attribute takes the tth value, etc.) may involve three types of variables: numerical, binary, and nominal variables.

- *Numerical variables.* For numerical clustering, a lot of methods are proposed, such as the Euclidean distance, Manhattan distance, Minkowski distance and weighted Euclidean distance measure (Han and Kamber 2001). This research employs the weighted Euclidean distance. It is computed as the following:

$$d_{numerical}(p_i, p_j) = \sqrt{\sum_{q=1}^{Q} \left(w_q \left(N\_a^*_{qi} - N\_a^*_{qj}\right)\right)^2}, \tag{1}$$

  where $d_{numerical}(p_i, p_j)$ indicates the numerical distance between two products p_i and p_j, w_q means the relative importance of the qth numerical variable $a_q \in A^{numerical} \subseteq A$, Q represents the total number of numerical variables among the total size N variables ($Q \le N$), and $N\_a^*_{qi}$ and $N\_a^*_{qj}$ denote the normalized values of the original a^*_{qi} and a^*_{qj}.
- *Binary variables.* A binary variable assumes only two states: 0 or 1, where 0 means the variable is absent and 1 means it is present. This research uses a well-accepted coefficient to measure the distance between binary variables, called the simple matching coefficient (Han and Kamber 2001). It is calculated as the following.

$$d_{binary}(p_i, p_j) = \frac{n_2 + n_3}{n_1 + n_2 + n_3 + n_4}, \tag{2}$$

  where $d_{binary}(p_i, p_j)$ indicates the binary distance between two products p_i and p_j, n_1 is the total number of binary variables in A (i.e. $a_q \in A^{binary} \subseteq A$) that equal 1 for both p_i and p_j, n_2 is the total number of binary variables that equal 1 for p_i but 0 for p_j, n_3 is the total number of binary variables that equal 0 for p_i but 1 for p_j, and n_4 is the total number of binary variables that equal to 0 for both p_i and p_j.
- *Nominal variables*. A nominal variable takes on more than two states. This type of variable only can be expressed by qualitative expressions with more than one option. In this regard, this research also adopts the simple matching coefficient to measure the nominal distance between two products containing nominal variables (Han and Kamber 2001):

$$d_{nominal}(p_i, p_j) = \frac{M - K}{M}, \tag{3}$$

  where $d_{nomical}(p_i, p_j)$ indicates the nominal distance between two products p_i and p_j, K means the total number of nominal variables in A (i.e., $a_q \in A^{nomical} \subseteq A$) that assume the same states for p_i and p_j, and M is the total number of nominal variables among the total size N variables ($M \le N$).

The overall distance between p_i and p_j is composed of the numerical, binary and nominal distances. In this regard, a composite distance can be obtained by the weighted sum:

$$d(p_i, p_j) = W_{numerical}\, d_{numerical}(p_i, p_j) + W_{binary}\, d_{binary}(p_i, p_j) + W_{nominal}\, d_{nominal}(p_i, p_j), \tag{4}$$

$$\sum (W_{numerical} + W_{binary} + W_{nominal}) = 1, \tag{5}$$

where $W_{numerical}$, W_{binary} and $W_{nominal}$ refer to the relative importance of numerical, binary and nominal variables, respectively. These weights are determined by applying the analytic hierarchy process.

4.2 *Fuzzy clustering*

Given a collection of objects (i.e. products), $R = A^* = \{a_t^* | \forall t = 1, \ldots, T\}$, a fuzzy set F in R is defined as a set of ordered pairs: $F = \{(r, \varphi_F(r)) | r \in R\}$, where $\varphi_F(r)$ is called the membership function of r in F that maps R to [0, 1]. A certain set of objects that belong to the fuzzy set F at least to the degree λ is called the λ-cut.

Assume R is a finite, non-empty set called the universe. Let F be a fuzzy relation in $R \times R$ (i.e. $F = \{(x, y) | \forall (x, y) \in R \times R)\}$); then (Lin and Lee 1996):

(1) F is reflexive if $\varphi_F(z, z) = 1 | \forall z \in R$;
(2) F is symmetric if $\varphi_F(z, x) = \varphi_F(x, z) | \forall x, z \in R$; and
(3) F is max-min-transitive if $\varphi_F(z, x) \geq \max_{y \in Y} \{\min\{\varphi_F(z, y), \varphi_F(y, x)\}\}$ (i.e. $F \circ F \subseteq F$).

If F satisfies the fist two criteria above, F is said to be a fuzzy compatible relation. If F satisfies all the criteria above, F is said to be a fuzzy equivalence relation.

A fuzzy compatible relation, F is constructed in a matrix form; that is, $F = [\mu(p_i, p_j)]_{T \times T} | \forall (p_i, p_j) \in A^* \times A^*$, where (p_i, p_j) suggests pair-wise relationships among products. A matrix element $\mu(p_i, p_j)$ indicates the similarity grade between any two products p_i and p_j. $\mu(p_i, p_j)$ is a measure of similarity, which is determined by the distance between products. The determination of similarity grade is as follows.

(a) Normalize the distance measure:

$$N\_d(p_i, p_j) = \frac{d(p_i, p_j) - \min\{d(p_i, p_j) | \forall i, j = 1, \ldots, T\}}{\max\{d(p_i, p_j) | \forall i, j = 1, \ldots, T\} - \min\{d(p_i, p_j) | \forall i, j = 1, \ldots, T\}}, \tag{6}$$

where $N\_d(p_i, p_j) \in [0, 1]$ is the normalized value of original distance $d(p_i, p_j)$. Because the variables involve different metrics, expressing a variable in smaller units will lead to a larger range for that variable, thus resulting in larger distance measure. The purpose of normalization is to avoid the dominance of certain variables over others and the dependence on the choice of different metrics.

(b) Derive the similarity grade $\mu(p_i, p_j)$; that is:

$$\mu(p_i, p_j) = 1 - N\_d(p_i, p_j). \tag{7}$$

Hence, we have $0 \leq \mu(p_i, p_j) \leq 1$. It is obvious that $\mu(p_i, p_i) = 1 | \forall i = 1, \ldots, T$; that is, F is reflexive and $\mu(p_i, p_j) = \mu(p_j, p_i) | \forall i, j = 1, \ldots, T$, suggesting that F is symmetrical. As a result, F becomes a fuzzy compatible relation and matrix F is called a fuzzy compatible matrix.

To convert a compatible matrix to an equivalence matrix, the 'continuous multiplication' method is often used:

$$\mu(p_i, p_j) \geq \max\{\min\{\mu(p_i, p_z), \mu(p_z, p_j) | \forall p_i, p_z, p_j \in A^*\}\}. \tag{8}$$

(c) The third step is to determine λ-cut of the equivalence matrix. The λ-cut is a crisp set, F_λ, that contains all the elements of the universe, A^*, such that the similarity grade of F is no less than λ. That is:

$$F_\lambda = [\tau(P_I, P_J)]_{T\times T}, \tag{9}$$

$$\text{where } \tau(P_I, P_J) = \begin{cases} 1 \text{ if } \mu(P_I, P_J) \geq \lambda \\ 0 \text{ if } \mu(P_I, P_J) < \lambda, \end{cases} \quad \mu(P_I, P_J) \in [0, 1]. \tag{10}$$

For each λ-cut, there exists a partition, $\rho(F_\lambda)$, such that each compatible matrix is associated with a set $\rho(F) = \{\rho(F_\lambda)\}$, and the value of $\lambda \in [0, 1]$ indicates the similarity threshold of a λ-cut.

5. Case study

The potential of the fuzzy clustering-based market segmentation for product family positioning has been tested in an electronics company that produces a large variety of vibration motors for major world-leading mobile phone manufacturers. Based on existing product documentation and consultation with design engineers, we know that the functional specification of vibration motors is described by seven attributes. The attributes and their values are presented in table 1. Among these seven attributes, the 'Pbfree' is of binary type and the 'Coating' is of nominal type, while all the rest are numerical variables.

Derived from the historical database, the transaction records are identified indicating which customers choose what product offerings. For illustrative simplicity, only 20 out of hundreds of transaction records are used in the case study here. Corresponding to the 20 customers (end-users of mobile phones), there are 20 vibration motors chosen. These 20 transaction records are represented as pairs of customers and the product offerings, which are presented in table 2.

To prioritize seven attributes, the analytic hierarchy process is applied. A seven-scale rating system is used to provide subjective judgements of preference, as presented in table 3. The result of each weight associated with each attribute is presented in table 4.

Table 1. List of attributes.

Attribute			Attribute value		
$a_q \| \forall q = 1, \ldots, N$	Description	Type	$a^*_{qr} \| \forall r = 1, \ldots, n_q$	Code	Description
a_1	Current	Numerical	a^*_{11}	A11	100 mA
			a^*_{12}	A12	80 mA
			a^*_{13}	A13	60 mA
a_2	Pbfree	Binary	a^*_{21}	A21	1 (Yes)
			a^*_{22}	A22	0 (No)
a_3	Length	Numerical	a^*_{31}	A31	8 mm
			a^*_{32}	A32	12 mm
			a^*_{33}	A33	10 mm
a_4	Coating	Nominal	a^*_{41}	A41	Au
			a^*_{42}	A42	Alloy
			a^*_{43}	A43	None
a_5	Angle	Numerical	a^*_{51}	A51	40°
			a^*_{52}	A52	55°
a_6	Strength	Numerical	a^*_{61}	A61	7 kg
			a^*_{62}	A62	4 kg
a_7	Weight	Numerical	a^*_{71}	A71	2 g
			a^*_{72}	A72	3 g

Table 2. Transaction database.

Sales record	Customer ($c_s \mid \forall s = 1, \ldots, S$)	Products ($p_t \mid \forall t = 1, \ldots, T$)
001	c_1	A11, A22, A33, A43, A52, A62, A71
002	c_2	A11, A21, A32, A41, A51, A62, A71
003	c_3	A12, A22, A33, A43, A52, A61, A72
...	...	...
018	c_{18}	A12, A22, A31, A41, A52, A61, A71
019	c_{19}	A13, A21, A33, A43, A52, A61, A72
020	c_{20}	A13, A21, A33, A42, A52, A61, A72

Among these seven attributes, attribute 2 is a binary variable, attribute 4 is a nominal variable, and others are all numerical variables. For the numerical variables, the SPSS software package (SPSS 12.0 for Windows, http://www.spss.com/) is used to obtain the weighted Euclidean distance measures. The 20 records of product specifications are input into the SPSS software for processing, in which the original data are normalized automatically and then the distances are calculated. Figure 1 shows the raw data for distance measures of numerical variables before the normalization. The normalized distance measure of numerical variables is presented in a 20×20 matrix form, $\lfloor N\_d_{numerical}(p_i, p_j)\rfloor_{20\times 20}$, and the results of distance measures for binary and nominal variables are presented as $\lfloor N\_d_{binary}(p_i, p_j)\rfloor_{20\times 20}$ and $\lfloor N\_d_{nominal}(p_i, p_j)\rfloor_{20\times 20}$, respectively. Based on these three distance components, the composite distances are calculated and presented as a dissimilarity matrix, $\lfloor d(p_i, p_j)\rfloor_{20\times 20}$, for all attribute variables, as shown in figure 2.

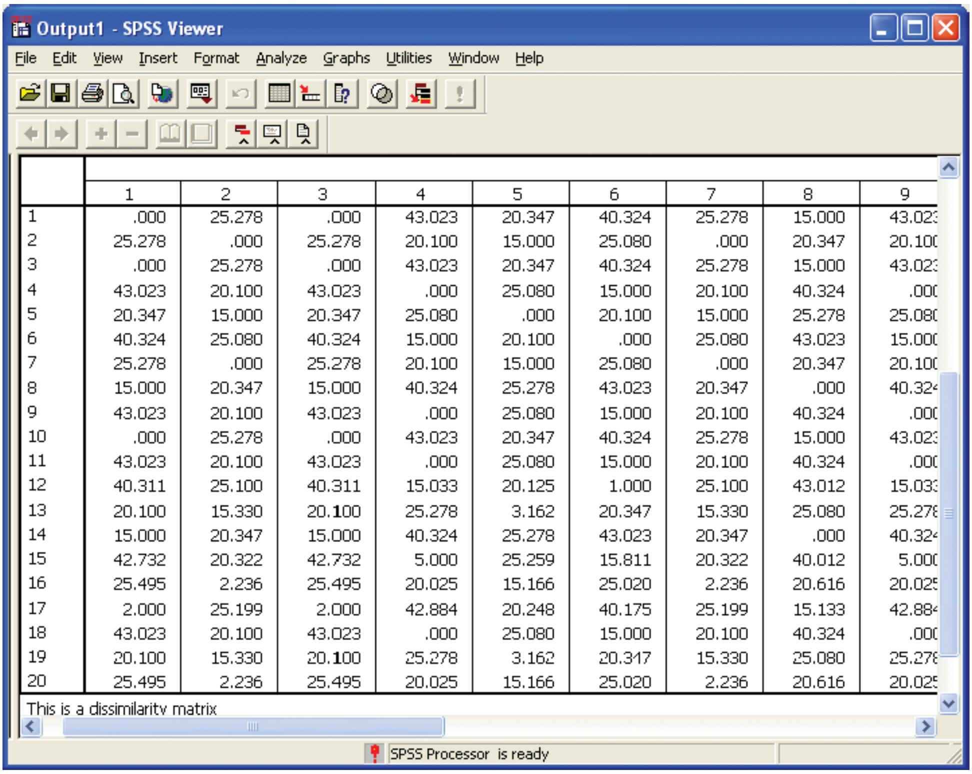
Based on the dissimilarity matrix, a fuzzy compatible matrix, F, is determined, as shown in figure 3. By the maximum–minimum composition method, the fuzzy equivalence matrix F^4 is a result. Based on F^4, the λ-cut is derived with a similarity threshold setting at 0.77. The result of the λ-cut is shown in figure 4.

Table 3. Scale for subjective judgement.

Verbal judgement of preference	Rating
Equally preferred	7
Equally to moderately	6
Moderately preferred	5
Moderately to strongly	4
Strongly preferred	3
Very strongly preferred	2
Extremely preferred	1

Table 4. Relative importance among attribute variables.

Attribute (a_q)	Weight (w_q)
a_1	0.33
a_2	0.273
a_3	0.067
a_4	0.04
a_5	0.035
a_6	0.082
a_7	0.173



	1	2	3	4	5	6	7	8	9
1	.000	25.278	.000	43.023	20.347	40.324	25.278	15.000	43.02
2	25.278	.000	25.278	20.100	15.000	25.080	.000	20.347	20.10
3	.000	25.278	.000	43.023	20.347	40.324	25.278	15.000	43.02
4	43.023	20.100	43.023	.000	25.080	15.000	20.100	40.324	.00
5	20.347	15.000	20.347	25.080	.000	20.100	15.000	25.278	25.08
6	40.324	25.080	40.324	15.000	20.100	.000	25.080	43.023	15.00
7	25.278	.000	25.278	20.100	15.000	25.080	.000	20.347	20.10
8	15.000	20.347	15.000	40.324	25.278	43.023	20.347	.000	40.32
9	43.023	20.100	43.023	.000	25.080	15.000	20.100	40.324	.00
10	.000	25.278	.000	43.023	20.347	40.324	25.278	15.000	43.02
11	43.023	20.100	43.023	.000	25.080	15.000	20.100	40.324	.00
12	40.311	25.100	40.311	15.033	20.125	1.000	25.100	43.012	15.03
13	20.100	15.330	20.100	25.278	3.162	20.347	15.330	25.080	25.27
14	15.000	20.347	15.000	40.324	25.278	43.023	20.347	.000	40.32
15	42.732	20.322	42.732	5.000	25.259	15.811	20.322	40.012	5.00
16	25.495	2.236	25.495	20.025	15.166	25.020	2.236	20.616	20.02
17	2.000	25.199	2.000	42.884	20.248	40.175	25.199	15.133	42.88
18	43.023	20.100	43.023	.000	25.080	15.000	20.100	40.324	.00
19	20.100	15.330	20.100	25.278	3.162	20.317	15.330	25.080	25.27
20	25.495	2.236	25.495	20.025	15.166	25.020	2.236	20.616	20.02

Figure 1. Raw data for distance measure of numerical variables.

$$
\begin{bmatrix}
0 \\
.89 & 0 \\
.22 & .85 & 0 \\
.83 & .7 & .75 & 0 \\
.47 & .21 & .91 & .76 & 0 \\
.87 & .79 & .69 & .16 & .65 & 0 \\
.82 & .12 & .8 & .75 & .08 & .88 & 0 \\
.06 & .7 & .22 & .81 & .81 & .5 & .9 & 0 \\
.94 & .8 & .86 & .08 & .68 & .03 & .74 & .32 & 0 \\
.1 & .25 & .07 & .78 & .89 & .6 & .79 & .15 & .82 & 0 \\
.74 & .7 & .37 & .21 & .77 & .09 & .68 & .74 & .06 & .8 & 0 \\
.58 & .68 & .92 & .16 & .84 & .1 & .82 & .88 & .2 & .86 & .1 & 0 \\
.67 & .11 & .88 & .77 & .14 & .84 & .05 & .68 & .75 & .33 & .72 & .81 & 0 \\
.15 & .6 & .06 & .6 & .73 & .39 & .81 & .1 & .95 & .08 & .87 & .58 & .6 & 0 \\
.38 & .77 & .83 & .17 & .84 & .14 & .7 & .82 & .16 & .82 & .12 & .03 & .8 & .88 & 0 \\
.83 & .06 & .6 & .67 & .1 & .83 & .06 & .86 & .67 & .31 & .77 & .83 & .08 & .86 & .85 & 0 \\
.35 & .88 & .05 & .74 & .81 & .47 & .95 & .03 & .95 & .19 & .95 & .42 & .81 & .11 & .73 & .83 & 0 \\
.83 & .88 & .95 & .15 & .7 & .09 & .7 & .74 & .11 & .81 & .02 & .03 & .85 & .7 & .01 & .88 & .76 & 0 \\
.88 & .03 & .78 & .84 & .02 & .71 & .08 & .55 & .77 & .82 & .49 & .76 & .08 & .84 & .84 & .07 & .87 & .86 & 0 \\
.54 & .01 & .87 & .95 & .08 & .76 & .14 & .77 & .82 & .56 & .88 & .79 & .08 & .81 & .96 & .07 & .79 & .9 & .04 & 0
\end{bmatrix}_{20\times 20}
$$

Figure 2. Dissimilarity matrix for all variables.

$$
\begin{bmatrix}
1 \\
.11 & 1 \\
.78 & .15 & 1 \\
.17 & .3 & .25 & 1 \\
.53 & .79 & .09 & .24 & 1 \\
.13 & .21 & .31 & .84 & .35 & 1 \\
.18 & .88 & .2 & .25 & .92 & .12 & 1 \\
.94 & .3 & .78 & .19 & .19 & .5 & .1 & 1 \\
.06 & .2 & .14 & .92 & .32 & .97 & .26 & .68 & 1 \\
.9 & .75 & .93 & .22 & .11 & .4 & .21 & .85 & .18 & 1 \\
.26 & .3 & .63 & .79 & .23 & .91 & .32 & .26 & .94 & .2 & 1 \\
.42 & .32 & .08 & .84 & .16 & .9 & .18 & .12 & .8 & .14 & .9 & 1 \\
.33 & .89 & .12 & .23 & .86 & .16 & .95 & .32 & .25 & .67 & .28 & .19 & 1 \\
.65 & .4 & .94 & .4 & .27 & .61 & .19 & .9 & .05 & .92 & .13 & .42 & .4 & 1 \\
.62 & .23 & .17 & .83 & .16 & .86 & .3 & .18 & .84 & .18 & .88 & .97 & .2 & .12 & 1 \\
.17 & .94 & .4 & .33 & .9 & .17 & .94 & .14 & .33 & .69 & .23 & .17 & .92 & .14 & .15 & 1 \\
.85 & .12 & .95 & .26 & .19 & .53 & .05 & .97 & .05 & .81 & .05 & .58 & .19 & .89 & .27 & .17 & 1 \\
.17 & .12 & .05 & .85 & .3 & .91 & .3 & .26 & .89 & .19 & .98 & .97 & .15 & .3 & .99 & .12 & .24 & 1 \\
.12 & .97 & .22 & .16 & .98 & .29 & .92 & .45 & .23 & .18 & .51 & .24 & .92 & .16 & .16 & .93 & .13 & .14 & 1 \\
.46 & .99 & .13 & .05 & .92 & .24 & .86 & .23 & .18 & .44 & .12 & .21 & .92 & .19 & .04 & .93 & .21 & .1 & .96 & 1
\end{bmatrix}_{20\times 20}
$$

Figure 3. Result of F.

With the obtained λ-cut, a fuzzy netting graph is constructed as shown in figure 5. Based on the partitions derived from the fuzzy netting graph, three clusters of customers are identified: segment 1 includes customers 1, 3, 8, 10, 14, 17; segment 2 includes customers 2, 5, 7, 13, 16, 19, 20; and segment 3 includes customers 4, 6, 9, 11, 12, 15, 18.

$$
\begin{bmatrix}
1 \\
0 & 1 \\
1 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1
\end{bmatrix}_{20\times 20}
$$

Figure 4. Result of a λ-cut with $\lambda = 0.77$.

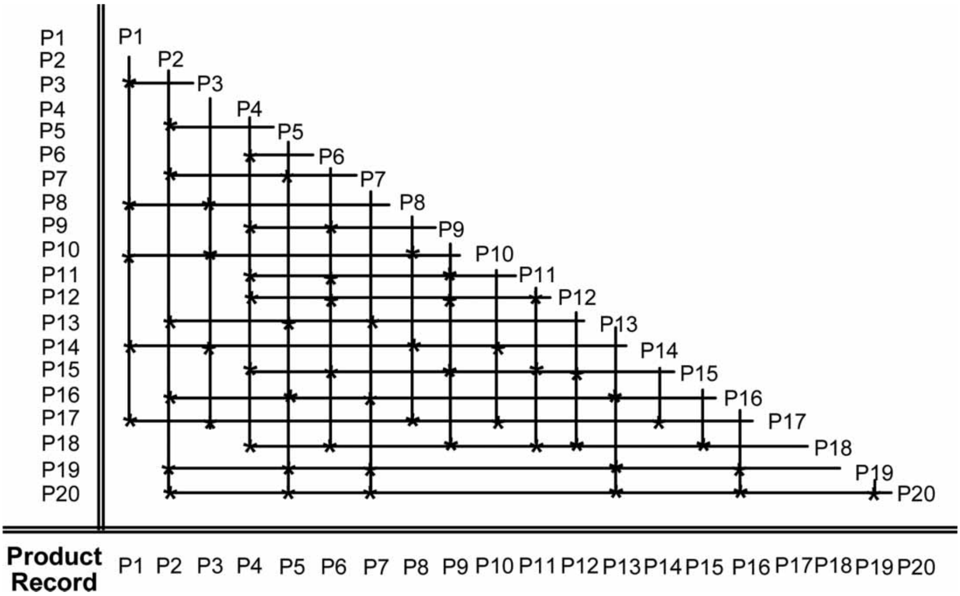


Figure 5. Result of a fuzzy netting graph.

Table 5. Result of clustering.

Cluster		Clustered product instances, $(\{p_t \sim \chi_l \mid \forall t = 1, \ldots, n_l \leq T\})$
χ_l	Mean value (μ_l)	
χ_1	[100,Y, 9.2, Au, 44.5, 6.7, 2.4]	$\{p_1, p_3, p_8, p_{10}, p_{14}, p_{17}\}$
χ_2	[78.3, Y, 11.17, Alloy, 47, 4 .5, 2.42]	$\{p_2, p_5, p_7, p_{13}, p_{16}, p_{19}, p_{20}\}$
χ_3	[67.5, Y, 10.75, None, 42.5, 5.13, 2.38]	$\{p_4, p_6, p_9, p_{11}, p_{12}, p_{15}, p_{18}\}$

The mean value for each cluster is calculated based on those product instances that are grouped into this cluster. The result of clustering is presented in table 5, in which, for example, cluster, χ_1, is associated with its mean, $\mu_l = [100, Y, 9.2, Au, 44.5, 6.7, 2.4]$, and contains six product instances, including p_1, p_3, p_8, p_{10}, p_{14}, and p_{17}.

6. Evaluation

The performance of market segmentation with respect to product family positioning can be assessed in accordance with the performance of the corresponding product and process platform. Jiao and Zhang (2005) apply conjoint analysis and choice models to evaluate customer-perceived utilities of diverse offerings of product families. On the other hand, the construction of the product and process platforms embodies a type of fixed costs (Meyer and Lehnerd 1997, Du *et al.* 2001). Therefore, we introduce a performance measure of product family positioning, Ψ, as the following:

$$\Psi = \sum_{i=1}^{n} \frac{E(U)_i}{C^F} \Big/ n, \tag{11}$$

where $E(U)_i$ denotes the expected product utility as perceived by the ith individual customer (segment), which is determined based on a planning framework (Jiao and Zhang 2005), C^F stands for the fix cost of the corresponding product and process platforms, and n is the total number of individual customers (segments). Furthermore, Jiao and Tseng (2004) posit the rationale of justifying cost implications of the product and process platforms based on process variations. Following Jiao and Tseng (2004), we employ a process capability index to measure the above fixed cost, as the following:

$$C^F = \beta^F e^{(1/PCI)} = \beta^F e^{(6\sigma/(USL-LSL))}, \tag{12}$$

where β^F is a constant indicating the average dollar cost per variation of process capabilities, and USL, LSL and σ are the upper specification limit, lower specification limit and standard deviation of part-worth cost estimates corresponding to individual product offering, respectively. The part-worth cost estimates are determined using a pragmatic approach based on standard time estimation (Jiao and Zhang 2005).

Furthermore, the performance of market segmentation with respect to the positioning of product families entails the specification of an optimal value of similarity threshold of λ-cut. Essentially, it gives rise to a trade-off issue of segment granularity inherent in mass customization (Tseng and Jiao 1996). With a large (small) value of λ-cut, more (less) segments will be identified. These segments affect the downstream positioning of the product and process platforms. At the economic latitude, the cost of introducing more product families (i.e. finer segmentation) and its contribution to customer-perceived values should reach a balance at the right level of aggregation of the product and process platforms. If the differentiation of product families is too spread or at too low a level of aggregation, such as at the nuts and bolts level, then the number of design parameters may be too many and the product fulfilment becomes difficult to leverage the investments. On the contrary, if the family positioning is at a very high level, such as complete subassemblies, then the repetition may not be sufficient to take advantage of mass production efficiency. Therefore, the performance of λ-cut can be assessed with respect to Ψ, where an appropriate value for λ-cut is able to produce a maximum of Ψ.

To analyze the sensitivity of segmentation, a total number of 18 runs of family positioning are generated by changing the λ value from 0.05 to 0.95 with an increment of 0.05. Using process data of vibration motors in Jiao *et al.* (2005) and utility evaluation data of vibration motors in Jiao and Zhang (2005), the result of sensitivity analysis is obtained. As shown in figure 6, the performance measure in equation (11) is presented as a normalized comparison.

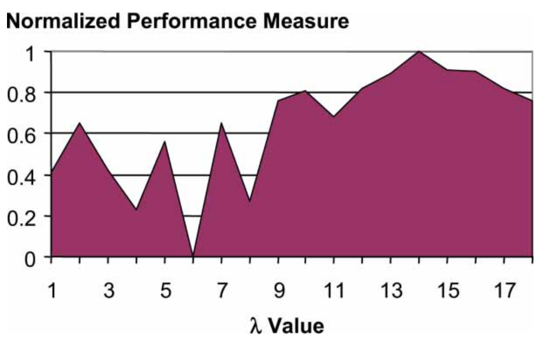


Figure 6. Sensitivity analysis of segmentation with respect to similarity threshold values.

The result clearly shows that a λ value of 0.75 yields the best performance of segmentation for product family positioning.

Without segmentation, as witnessed in the case study, transaction records indicate 20 different motors are provided to satisfy 20 individual customers. Conjoint analysis is implemented to generate the utility evaluation data of vibration motors on individual level. The results are presented in table 6. Using individual level utility evaluation data and process data of vibration motors in Jiao *et al.* (2005), according to equation (11), the performance of product family positioning is calculated as 0.42.

Implementing segmentation, three types of motors are offered to customers belonging to three different segments, as presented in table 5. Conjoint analysis is also adopted to derive the utility evaluation data of vibration motors on a segment level. Table 7 presents the part-worth utilities of three segments with respect to every attribute value. Allowing the utility data on segment level and process data, according to equation (11), the performance of product family positioning is calculated as 0.68.

The performance of product family positioning is improved by implementing market segmentation. Product family positioning is by no means to provide whatever customers may want, as excessive variety results in a dramatic increase of costs and more complexity in management, production process and inventory control (Huffman and Kahn 1998). On the other hand, a single product cannot satisfy the heterogeneous customer needs. A market segmentation study suggests that the customers are willing to choose from those products with attribute values closest to their desired values if they cannot find any product on the market that exactly matches their desired values. As shown in tables 6 and 7, for customer 1, belonging to segment 1, the corresponding expected utility for every attribute value on an individual level may contribute to the improvement of utility to only a modest extent, compared with that on a segment level (e.g. the utility of A1-1 for customer 1 is 0.68 compared with that of 0.65 for segment 1; the utility of A1-1 for customer 2 is 0.44 compared with that of 0.42 for segment 2). This implies that market segmentation helps keep economy of scale

Table 6. Part-worth utilities on an individual level.

Attribute value	Part-worth utility (individual)				
	c_1	c_2	...	c_{19}	c_{20}
A1-1	0.68	0.44	…	0.43	0.43
A1-2	0.87	0.65	…	0.66	0.65
A1-3	1.15	0.89	…	0.88	0.89
A2-1	1.17	1.23	…	1.24	1.23
A2-2	1.14	1.28	…	1.27	1.27
A2-3	…	…	…	…	…
A3-1	1.26	0.79	…	0.79	0.78
A3-2	0.67	1.27	…	1.26	1.26
A3-3	…	…	…	…	…
A4-1	1.64	0.88	…	0.87	0.87
A4-2	1.27	1.19	…	1.18	1.16
A4-3	1.24	0.94	…	0.93	0.93
A5-1	1.37	1.28	…	1.27	1.28
A5-2	0.57	0.40	…	0.38	0.39
A5-3	…	…	…	…	…
A6-1	0.78	1.55	…	1.52	1.52
A6-2	0.88	1.25	…	1.24	1.25
A6-3	…	…	…	…	…
A7-1	0.76	0.37	…	0.35	0.36
A7-2	0.91	0.66	…	0.65	0.65
A7-3	…	…	…	…	…

Table 7. Part-worth utilities on a segment level.

Attribute value	Part-worth utility (segment)		
	s_1	s_2	s_3
A1-1	0.65	0.42	0.71
A1-2	0.86	0.64	1.32
A1-3	1.12	0.87	0.55
A2-1	1.16	1.22	1.25
A2-2	1.13	1.26	0.98
A2-3	…	…	…
A3-1	1.23	0.78	0.36
A3-2	0.66	1.25	1.45
A3-3	…	…	…
A4-1	1.63	0.86	0.33
A4-2	1.25	1.16	0.83
A4-3	1.22	0.93	0.91
A5-1	1.35	1.26	1.31
A5-2	0.56	0.38	0.68
A5-3	…	…	…
A6-1	0.76	1.51	0.93
A6-2	0.87	1.23	0.95
A6-3	…	…	…
A7-1	0.75	0.35	0.62
A7-2	0.89	0.63	0.71
A7-3	…	…	…

in product fulfilment without sacrificing customer satisfaction, and thus to solve the trade-off issues involving in the product family positioning problem.

7. Conclusions

Industries have been seeking to provide sufficient variety to the market while leveraging their manufacturing capabilities. This research allows product families to be offered to target customer segment according to customer purchasing patterns. Thus, diverse customers can be satisfied by offering the 'right' product variety while still taking the advantage of mass production efficiency. Compared with traditional segmentation methods, the proposed fuzzy clustering-based segmentation approach implements the segmentation with more considerations of engineering characteristics. Establishing direct relationships between customer preferences and discrete attribute values, customer preferences can be predicted with respect to various combinations of attribute values. Using engineering characteristics as segmentation variables, fuzzy clustering-based segmentation approach is likely to overcome the preference distortion resulting from other segmentation methods using general variables.

References

Beane, T.P. and Ennis, D.M., Market segmentation: a review. *Eur. J. Market.*, 1987, **21**(5), 20–42.

Child, P., Diederichs, R., Sanders, F.H. and Wisniowski, S., SMR forum: 'The management of complexity'. *Sloan Manage. Rev.*, 1991, **33**(1), 73–80.

Deciu, E.R., Ostrosi, E., Ferney, M. and Gheorghe, M., Configurable product design using multiple fuzzy models. *J. Eng. Design*, 2005, **16**(2), 209–233.

Drozdenko, R.G. and Drake, P.D., *Optimal database marketing: strategy, development, and data mining*, 2002 (Sage: London).

Du, X., Jiao, J. and Tseng, M.M., Architecture of product family: fundamentals and methodology. *Concurrent Eng. Res. Application*, 2001, **9**(4), 309–325.

Hammond, K., Enrenberg, A.S.C. and Goodhardt, G.J., Market segmentation for competitive brand. *Eur. J. Market.*, 1996, **30**(12), 39–49.

Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, 2001 (Morgan Kaufmann Publishers, San Francisco).

Heuvel, D.A. and Devasagayam, P.R., Multivariate cluster analysis model of benefit-based market segmentation: a case study from the recreation and leisure industry, 2004. Available online at: http://www.danavan.net/pdf/cluster.analysis.publication.pdf (accessed November 2005).

Hofer, A.P. and Halman, J.I.M., The potential of layout platforms for modular complex products and systems. *J. Eng. Design*, 2005, **16**(2), 237–255.

Huffman, C. and Kahn, B., Variety for sale: mass customization or mass confusion. *J. Retail.*, 1998, **74**(4), 491–513.

Hughes, A., *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-based Marketing Program*, 1994 (McGraw Hill: NY).

Jiang, L. and Allada, V., Robust modular product family design using a modified Taguchi method. *J. Eng. Design*, 2005, **16**(5), 443–458.

Jiao, J. and Tseng, M.M., Customizability analysis in design for mass customization. *Computer-Aided Design*, 2004, **36**(8), 745–757.

Jiao, J. and Zhang, Y., Product portfolio planning with customer-engineering interaction. *IIE Trans.*, 2005, **37**(9), 801–814.

Jiao, J., Zhang, L. and Pokharel, S., Process platform planning for variety coordination from design to production in mass customization manufacturing. *IEEE Transactions on Engineering Management*, 2006 forthcoming.

Johannesson, H. and Claesson, A., Systematic product platform design: a combined function-means and parametric modeling approach. *Journal of Engineering Design*, 2005, **16**(1), 25–43.

Jonker, J.J., Piersma, N. and Poel, D.V., Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, 2004. Available online at: http://ideas.repec.org/p/dgr/eureir/2002271.html (accessed November 2005).

Kotler, P., *Marketing Management: Analysis, Planning Implementation and Control*, 1994 (Prentice Hall: NJ).

Kou, R.J., Ho, L.M. and Hu, C.M., Integration of self-organizing feature map and K-means algorithm for market segmentation. *Comput. Operat. Res.*, 2002, **29**(11), 1475–1493.

Lin, C.T. and Lee, C.S.G., *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, 1996 (Prentice Hall: NJ).

Marcus, C., A practical yet meaning approach to customer segmentation. *J. Consumer Market.*, 1998, **15**(5), 494.

McDonald, M. and Dunbar, I., *Market Segmentation*, 1995 (Macmillan Press: Basingstoke).

Meyer, M. and Lehnerd, A.P., *The Power of Product Platform-Building Value and Cost Leadership*, 1997 (Free Press: New York).

Natter, M., Conditional market segmentation by neural networks: A Monte-Carlo Study. *J. Retail. Consumer Serv.*, 1999, **6**(4), 237–248.

Olewnik, A. and Lewis, K., A decision support framework for flexible system design. *J. Eng. Design*, 2006, **17**(1), 75–97.

Thevenot, H.J. and Simpson, T.W., Commonality indices for product family design: a detailed comparison. *J. Eng. Design*, 2006, **17**(2), 99–119.

Tsai, C.-Y. and Chiu, C.-C., A purchase-based segmentation methodology. *Expert Systems Applications*, 2004, **27**(1), 265–276.

Tseng, M.M. and Jiao, J., Design for mass customization. *CIRP Annals*, 1996, **45**(1), 153–156.

Wedel, S. and Kamakura, W., *Market Segmentation: Conceptual and methodological foundations*, 1997 (Kluwer: Boston, MA).