Computer-Aided Design 37 (2005) 149–172
COMPUTER-AIDED DESIGN
www.elsevier.com/locate/cad

Product portfolio identification based on association rule mining

Jianxin Jiao*, Yiyang Zhang

School of Mechanical and Production Engineering, Nanyang Technological University, Nanyang Avenue 50, Singapore 639798

Received 15 December 2003; received in revised form 19 February 2004; accepted 10 May 2004

Abstract

It has been well recognized that product portfolio planning has far-reaching impact on the company's business success in competition. In general, product portfolio planning involves two main stages, namely portfolio identification and portfolio evaluation and selection. The former aims to capture and understand customer needs effectively and accordingly to transform them into specifications of product offerings. The latter concerns how to determine an optimal configuration of these identified offerings with the objective of achieving best profit performance. Current research and industrial practice have mainly focused on the economic justification of a given product portfolio, whereas the portfolio identification issue has been received only limited attention. This article intends to develop explicit decision support to improve product portfolio identification by efficient knowledge discovery from past sales and product records. As one of the important applications of data mining, association rule mining lends itself to the discovery of useful patterns associated with requirement analysis enacted among customers, marketing folks, and designers. An association rule mining system (ARMS) is proposed for effective product portfolio identification. Based on a scrutiny into the product definition process, the article studies the fundamental issues underlying product portfolio identification. The ARMS differentiates the customer needs from functional requirements involved in the respective customer and functional domains. Product portfolio identification entails the identification of functional requirement clusters in conjunction with the mappings from customer needs to these clusters. While clusters of functional requirements are identified based on fuzzy clustering analysis, the mapping mechanism between the customer and functional domains is incarnated in association rules. The ARMS architecture and implementation issues are discussed in detail. An application of the proposed methodology and system in a consumer electronics company to generate a vibration motor portfolio for mobile phones is also presented.

Keywords: Data mining; Mass customization; Product portfolio; Association rules; Variety; Requirement management; Customer satisfaction; Product definition

1. Introduction

Understanding and fulfilling each individual customer need have been recognized as an enormous challenge for companies across industries. Rather than offering market-focused products, which corresponds to an average satisfaction of several customer needs, companies are pursuing a strategy of mass customization [1], which strives to offer customer-focused products with a large degree of individuality. To compete in the marketplace, manufacturers have been seeking for expansion of their product lines and differentiation of their product offerings with the intuitively appealing belief that high product variety may stimulate sales and thus conduce to revenue [2]. While a high variety strategy may offer an effective means for companies to differentiate themselves from their competitors, it unavoidably leads to high complexity and costs in product fulfillment [3]. Moreover, making wide variety of products available and letting customers vote on the shelf may cause customer to be overwhelmed by the huge assortment offered or frustrated by the complexity involved with making a choice [4]. Therefore, it becomes imperative for the manufacturer to determine how to offer 'right' product variety to the target market.

This type of decisions adheres to the general wisdom as suggested in the Boston Consulting Group's notion of product portfolio strategy, one of the most popular contemporary approaches to strategic planning [5]. While representing the spectrum of a company's product offerings, the product portfolio must be carefully set up, planned and managed so as to match those customer needs in the target market [6]. The product portfolio strategy has far-reaching

* Corresponding author. Tel.: +65-6790-4143; fax: +65-6791-1859.
E-mail address: jiao@pmail.ntu.edu.sg (J. Jiao).

0010-4485//$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cad.2004.05.006

impact on the company's business success such as to achieve financial goals in maximizing return and R&D productivity, to maintain the competitive edge of the business by increasing sales and market share, to allocate scare resources properly and efficiently, to forge the link between project selection and business strategies, to better communicate priorities within the organization both vertically and horizontally, and so on [7].

In general, product portfolio planning involves two main stages [8]. The first is called product portfolio identification. The goal is to capture and understand customer needs effectively and accordingly to transform them into specifications of product offerings (e.g. functional features). The second is called product portfolio evaluation and selection. The key issue is to determine an optimal setup or configuration of these planned offerings (e.g. the go/kill decision of an offering) with the objective of achieving best profit performance. Current researchers and industrial practitioners in this field involve themselves mostly in the economic justification of product portfolio (e.g. product line design), viz. the latter stage of product portfolio planning. They usually imply the specification of offerings in a product portfolio is given. However, the first issue—how to identify customer needs and generate product portfolio specifications—has received only limited attention. During this phase, many factors are to be considered including any combination of customer needs, corporate objectives, product ideas and related technological capabilities, etc. Usually, product offerings are represented as a list of functional features and target values. This information is often a mix of quantitative values and qualitative descriptions of product functionality. In most cases, the company may produce a formal document that requires to undergo routinely many amendments along with scrutiny, or to be signed off by many individuals [9]. Even though product portfolio identification is of paramount importance, past research has not addressed it well, nor has actual practice availed to formulate effective means. This may stem from the complications inherent in the product portfolio identification process, as discussed below.

1.1. Fundamentals of product portfolio identification

To leverage the market benefits of customization and the costs of providing variety, it is reasonable to fulfill mass customization within a company's capabilities in design and production. In practice, this is often achieved by developing product and process platforms [10,11]. A product platform performs as a base product from which product families can variegate designs to satisfy individual customer requirements [12]. Corresponding to a product platform, production processes can be organized as a process platform in the form of a bill-of-operations (e.g. standard routings), hence, facilitating build or configure-to-order production for given customer orders [13]. Both product and process platforms originate from, and are thus supposed to conform to, a planned product portfolio.

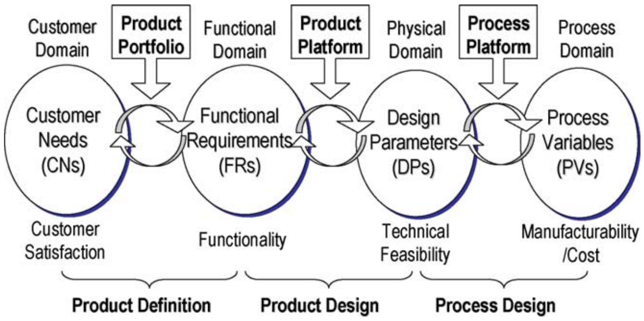


Fig. 1. Product portfolio within the spectrum of product development.

As shown in Fig. 1, a holistic view on product portfolio can be illustrated along the entire spectrum of product development according to the domain framework in axiomatic design [14]. Product development in general encompasses three consecutive stages: (1) product definition—mapping of customer needs (CNs) in the customer domain to functional requirements (FRs) in the functional domain; (2) product design—mapping of FRs in the functional domain to design parameters (DPs) in the physical domain; and (3) process design—mapping of DPs in the physical domain to process variables (PVs) in the process domain. Accordingly, the customer, functional, physical and process domains address the customer satisfaction, functionality, technical feasibility, and manufacturability/cost issues associated with the products, respectively [15]. Within the context of mass customization, product design and process design are embodied in the respective product and process platforms. Product definition is characterized by the product portfolio representing the target of mass customization (i.e. the 'right' product offerings), which in turn becomes the input to the downstream design activities and is propagated to product and process platforms in a coherent fashion. In this sense, a product portfolio represents the functional specification of product families, i.e. the functional view of product and process platforms [16].

Consistent with the product definition process, product portfolio identification involves a tedious elaboration process enacted among customers, marketing folks, and designers, as shown in Fig. 2. Tseng and Jiao [17] point out the difficulties associated with product definition. Their observations are also supported in the study by Tarasewich and Nair [18].

First, the customer requirements are normally qualitative and tend to be imprecise and ambiguous due to their linguistic origins. In most cases, requirements are negotiable and conflict with one another, and thus tradeoffs are often necessary. Frequently, customers, marketing folks and designers employ different sets of context to express the requirements. Differences in semantics and terminology always impair the ability to convey requirement information effectively from customers to designers due to their different

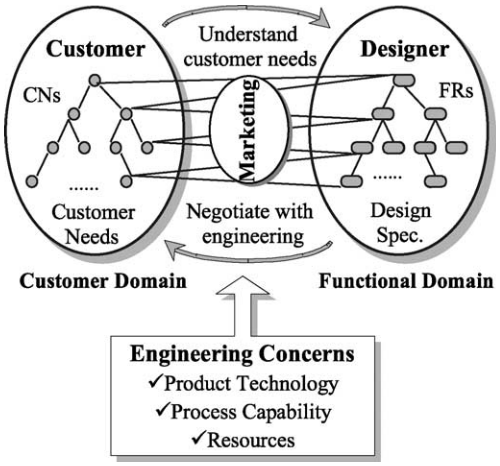


Fig. 2. Product definition process inherent in product portfolio identification.

positions. The differentiation of requirements in terms of CNs and FRs is of practical significance. An organization should put considerable efforts in capturing the genuine or 'real' needs of the customers (CNs), rather than too much focus on the technological issues (FRs) during early stage of product development [19].

Second, there rarely exists any definite structure of requirement information. Variables used to describe requirements are often poorly understood and are usually expressed in abstract, fuzzy, or conceptual terms, leading to work on the basis of vague assumptions and implicit inference. A few researchers have enforced a hierarchical structure or an AND/OR tree structure for the articulation of customer requirements, for example, the requirement taxonomy [20], the customer attribute hierarchy [21], and the FR topology [17]. Nevertheless, the non-structure nature of requirement information itself coincides with those findings in nature language processing [22].

Third, the interrelationships (i.e. mapping) between CNs and FRs are often not clearly available in an early stage of design. Customers are often not aware of the underlying coupling and interrelationships among various requirements with regard to product performance. It is difficult, if not impossible, to estimate the consequences (in particular, in terms of economic, scheduling and quality concerns) of specifying different requirements. Christopher et al. [23] discern customer needs and product specifications and point out the mapping problem between them is the key issue in 'design for customers'.

Fourth, the specification of requirements results from not only the transformation of customer requirements from those end-users, but also considerations of many engineering concerns, involving any internal customer from downstream of the design team along the product realization process [24]. In practice, product development teams must keep track of a myriad of requirement information derived from different perspectives on the product life-cycle, such as product technologies, manufacturability, reliability, maintainability, and environmental safety, to name but a few [25].

Therefore, the process of product portfolio identification can be described as: $\Lambda \leftarrow \Gamma(CNs, Eng.)$, where Λ represents a specification of product offerings, CNs indicate the customer needs of end-users, $Eng.$ means engineering considerations associated with CNs, and Γ denotes the mapping relationship from CNs and $Eng.$ to a particular product portfolio, Λ.

1.2. Strategy for solution

Due to the difficulties inherent in the portfolio identification process, reusing knowledge from historical data suggests itself as a natural technique to facilitate the handling of requirement information and tradeoffs among many customer, marketing and engineering concerns. Tseng and Jiao [17] propose to identify FR patterns from previous product designs for addressing a broad spectrum of domain-specific customer requirements and to organize requirement information during design. In their model, various FRs are grouped according to the similarity among customers (i.e. market segments). The focus is on the functional domain. Du et al. [24] extend the idea to study the patterns of CNs for better customization and personalization. Chen et al. [26] apply neural network techniques to construct a customer attribute hierarchy (CAH) in order to improve customer requirement elicitation. Both ideas emphasize on the customer domain. While these proposed solutions emphasize on the identification of either CN or FR patterns, the mapping relationship between CNs and FRs has not been taken into account. We assert that FR patterns should not be identified in isolation from those patterns of CNs, and vice versa. The patterns of CN–FR mappings play an important role in bringing engineering concerns into product portfolio identification as well as in determining CN and FR patterns within a cohesive context.

To this end, this article proposes to apply data mining techniques to improve the product portfolio identification process. Data mining has been well recognized for decision support by efficient knowledge discovery of previously unknown and potentially useful patterns of information from past data [27]. As one of the important applications of data mining, association rule mining lends itself to the discovery of knowledge associated with mappings from CNs to FRs. Based on association rule mining, this research develops an inference system for effective product portfolio identification.

In Section 2, the background research leading to product portfolio identification is presented. In Section 3, the methodology of product portfolio identification based on association rule mining is described. An association rule mining system (ARMS) architecture and its implementation issues are discussed in Section 4. In Section 5, an application of the proposed methodology and system to generate a vibration motor portfolio for mobile phones is presented. Observations from the case study and managerial implications are derived in Section 6. An outline of future work and conclusions are drawn in Section 7.

2. Related work

Approaches to defining product specifications by capturing, analyzing, understanding, and projecting customer requirements, sometimes called the Voice of the Customer (VoC), have received a significant amount of interests in recent years [28]. A method used for transforming the VoC to product specifications is developed by, Shoji et al. [29], in which semantics methods, such as the Kawakita Jiro (KJ) method (i.e. affinity diagram) and multi-pickup method (MPM), are applied as the basis for discovering underlying facts from affective language. Kano et al. [30] propose a diagram to categorize different types of customer requirements for product definition.

In this regard, market researchers have emphasized customer profiling by applying regression analysis to compare customer characteristics and to determine their overall ranking in contribution towards profitability [31]. Traditionally, market analysis techniques are adopted for investigating customers' responses to design options. For example, conjoint analysis is widely used to measure preferences for different product profiles and to build market simulation models [32]. Louviere et al. [33] use discrete choice experiments to predict customer choices pertaining to design options. Turksen and Willson [34] employ fuzzy systems to interpret the linguistic meaning regarding customer preferences as an alternative to conjoint analysis. Others have taken a qualitative approach and used focus groups to provide a reality check on the usefulness of a new product design [35]. Similar techniques include one-on-one interviews and similarity-dissimilarity attribute rankings [36]. While these types of methods are helpful for discovering the VoC, it is still difficult to obtain design requirement information because marketing folks do not know what engineers need to know. It is difficult to apply the VoC alone to achieve a synergy of marketing and engineering concerns in developing product specifications [37].

A number of complex customer behaviors such as perceptions, motivations, attitudes and personality can be grouped under psychological factors for making rational decisions [38]. These factors influence the way in which customers select, organize and interpret a company and its product offerings. Kansei engineering [39] is a technique for the translation of consumers' psychological feeling about a product into perceptual design elements [40]. As a structured questioning methodology built upon Kelly's repertory grid technique [41], the laddering technique has been widely used to transform customers' psychological factors into useful inputs for design applications [42]. Many methods and tools in the field of knowledge acquisition, such as observation, self-report [43], interview, protocol, ethnographic methods [44], and sorting techniques [45], have some applicability in requirement elicitation for product development [22]. Maiden and Rugg [46] propose a framework called acquisition of requirements (ACRE) to assist practitioners in understanding the strengths and weaknesses of each of the methods for requirement elicitation. Chen and his co-authors propose an integrated approach to the elicitation of customer requirements by combining picture sorts, fuzzy evaluation, laddering, and neural network techniques [26,47,48].

From an engineering design perspective, Hauge and Stauffer [20] develop a taxonomy of product requirements to assist in traditional qualitative market research. To elicit knowledge from customers (ELK), the taxonomy of customer requirements is deployed as an initial concept graph structure in the methodology for question probe—a method used in the development of expert systems. While ELK aims at making customer information more useful to the designer, the taxonomy developed for ELK is too general to be a domain independent framework [17]. A key component of Quality Function Deployment (QFD; [49]) is the customer requirements frame to aid the designer's view in defining product specifications. While QFD excels in converting customer information to design requirements, it is limited as a means of actually discovering the VoC [20]. To empower QFD with market aspects, Fung and Popplewell [50] propose to pre-process the VoC prior to its being entered as customer attributes into the House of Quality (HoQ). In this process, the VoC is categorized using an affinity diagram (KJ method). Fung et al. [51] further adopt the Analytic Hierarchy Process (AHP; [52]) to analyze and prioritize customer requirements. Fung et al. [53] extend their QFD-based customer requirement analysis method to a non-linear fuzzy inference model. Fukuda and Matsuura [54] also propose to prioritize the customer's requirements by AHP for concurrent design. Researchers at IBM have applied structured brainstorming techniques to build customer requirements into the QFD process [55]. McAdams et al. [56] propose a matrix approach to the identification of relationships between product functions and customer needs.

In summary, most approaches assume product development starts from a clean sheet of paper. In practice, most new products evolve from existing products, i.e. so-called variant design. Historical data, product evolution paths, and feedback from customers on current products are often considered only implicitly, if not ignored. As a result, product design seldom has the opportunity to take advantage of the wealth of customer requirement information accumulated in existing products. In addition, these methods do not explicitly differentiate the customer preference from the designer's preference of requirement information [18], nor exists any approach to handling effectively the mapping from the customer domain to the functional domain. Furthermore, new product development in mass customization is facing the challenge of maintaining the continuity of manufacturing and service operations. Therefore, product definition should effectively preserve the strength of product families to obtain significant cost savings in tooling, learning curves, inventory, maintenance, and so on. This demands a structured approach to product definition and to the capturing of gestalt requirement information from previous designs as well as existing product and process platforms.

3. Methodology of product portfolio identification

3.1. Problem formulation

Fig. 3 illustrates the principle of product portfolio identification based on association rule mining. In general, customer needs can be described as a set of features or attributes, $A \equiv \{a_1, a_2, \ldots, a_M\}$. Each feature, $a_i | \forall i \in [1, \ldots, M]$, may take on one out of a finite set of options, $A_i^* \equiv \{a_{i1}^*, a_{i2}^*, \ldots, a_{in_i}^*\}$. That is, $a_i =:: a_{ij}^* | \exists a_{ij}^* \in A_i^*$, where $j = 1, \ldots, n_i$, denotes the jth option of a_i. Suppose all customers comprise a set, $C \equiv \{c_1, c_2, \ldots, c_S\}$, where S denotes the total number of customers. In the customer domain, requirement information of a particular customer, $c_s \in C | \exists s \in [1, \ldots, S]$, can be depicted by a vector of certain options of these features, for example, $\overline{a_s^*} \equiv [a_{13}^*, a_{22}^*, \ldots, a_{M1}^*]$, where a_{13}^* refers to the third option of feature a_1 as desired by customer c_s, a_{22}^* the second option of feature a_2, and a_{M1}^* the first option of feature a_M. All population of customers' needs become a set, $A^* \equiv \{\overline{a_1^*}, \overline{a_2^*}, \ldots, \overline{a_S^*}\}$, which characterizes the customer domain.

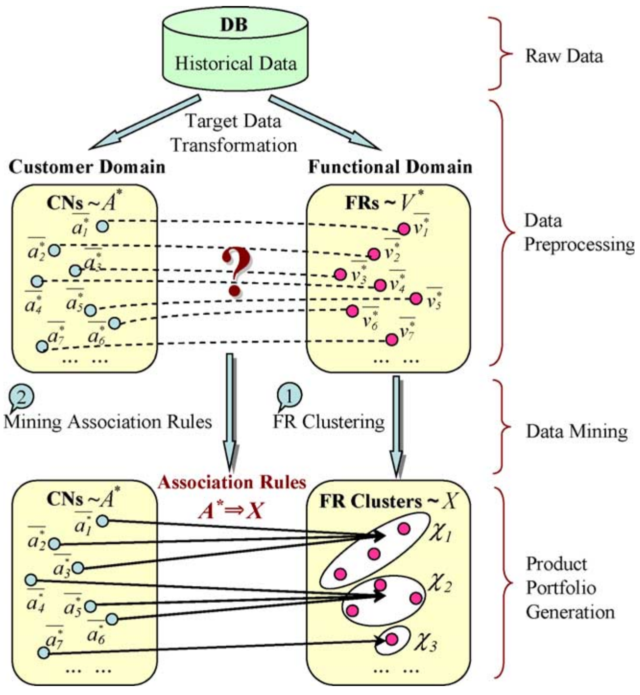


Fig. 3. Product portfolio identification based on association rule mining.

In the functional domain, the functionality of each product is characterized by a set of FRs, $V \equiv \{v_1, v_2, \ldots, v_N\}$. Each FR, $v_q | \forall q \in [1, \ldots, N]$, possesses a few possible values, $V_q^* \equiv \{v_{q1}^*, v_{q2}^*, \ldots, v_{qn_q}^*\}$. That is, $v_q = :: v_{qr}^* | \exists v_{qr}^* \in V_q^*$, where $r = 1, \ldots, n_q$, denotes the rth possible value of v_q. Suppose all existing products comprise a set, $P \equiv \{p_1, p_2, \ldots, p_T\}$, where T refers to the total number of products. The requirement specification of a particular product, $p_t \in P | \exists t \in [1, \ldots, T]$, can be represented as a vector of certain FR values of those FRs, for example, $\overline{v_t^*} \equiv [v_{12}^*, v_{21}^*, \ldots, v_{N5}^*]$, where v_{12}^* means product p_t involves the second value of FR v_1, v_{21}^* the first value of FR v_2, and v_{N5}^* the fifth value of FR v_N. All the instances of FRs (i.e. FR values) in the functional domain constitute a set, $V^* \equiv \{\overline{v_1^*}, \overline{v_2^*}, \ldots, \overline{v_T^*}\}$.

Based on the company's sales records and product documentation, we can extract transaction data related to which customer was met with which product. Therefore, transaction data can be summarized as CN–FR pairs in the form of $\langle \overline{a_s^*}, \overline{v_t^*} \rangle$, where s and t stand for customer ID and product ID, respectively. Each pair of such transaction data not only indicates a specific case of requirement information from both the customer and manufacturer viewpoints, but also implies a particular instance of mapping relationship between the customer and functional domains.

The difference between the customer and functional domains suggests that what a customer de facto perceives is the CNs, rather than FRs. While providing customer-perceived diversity in CNs, the manufacturer must seek for economy of scale in product fulfillment, which is meant by FRs. In addition, mass customization is by no means to provide whatever customers may want, as excessive variety results in a dramatic increase of costs [4]. As postulated in the classic Hotelling–Lancaster model [57], some products close together on the spectrum are better substitutes than those further apart. This implies that customers are willing to choose from those products with functional values closest to their desired values if they cannot find any product on the market that exactly matches their desired values. Consumer behavior study also suggests that the consumers falling into the same cluster usually hold the same purchase trend and thus the customer can be met by providing such a product that the total variations of functionality from what the customer prefers to are the smallest. This implies that individual customers within a cluster can most probably be satisfied with a product whose functional values assume the mean values of different expectations by all customers in the same cluster (namely, the centroid of the cluster).

Therefore, in order to take advantage of commonality in product family design, existing instances of FRs, V^*, should be analyzed and clustered according to the similarity among them [58]. This process is called FR clustering. The result is a few FR 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 FR cluster. As a result, all FR instances related to a FR cluster, i.e. $\chi_l \sim V^{*l} \subset V^*$, can be grouped and represented by the characteristics of χ_l—the mean value of these FR instances, $\mu_l \equiv [x_1^l, x_2^l, \ldots, x_N^l]$, and the variation range of these FR instances within χ_l, $\Delta_l \equiv [\delta_1^l, \delta_2^l, \ldots, \delta_N^l]$. Therefore, each FR cluster can be described as a tuple: $\chi_l = (\mu_l, \Delta_l)$.

Subsequently, these identified FR clusters become the functional specification of product offerings that can be derived from common product platforms and are supposed to be able to accommodate all the customer needs [59]. In other words, the specification of a product portfolio should cover a group of existing and latent CNs by mapping these needs to the identified FR clusters. At this stage, data mining techniques are applied to figure out the mapping relationship between CNs and FR clusters, noted as $A^* \Rightarrow X$, where an association rule, $\Rightarrow$, indicates an inference from the precedent (A^*) to the consequence (X). As a result, a product portfolio specification, Λ, consists of two elements: FR clusters and mappings from CNs to FR clusters, namely, $\Lambda = \langle X, \Rightarrow \rangle$.

3.2. FR clustering

Clustering analysis 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 [60]. The specification of FRs usually presents in the form of numerical, binary or nominal variables. To handle both quantitative and qualitative variables, this research adopts a fuzzy clustering approach to FR clustering. Fuzzy equivalence relations excel in revealing the similarity between any two objects involving subjectiveness and imprecision [61]. Fuzzy clustering is to create a hierarchical decomposition of the given set of objects, in which each object forms a separate group and successively the objects or groups close to one another are merged at different similarity levels. In our case, historical data about FR instances contained in the platform can be used to measure the similarity degree based on the compatibility of FR value ranges. In comparison with the k-means method, fuzzy clustering partitions FR instances based on the similarity degree that is derived from the real data of FR values, rather than based on subjectively pre-defined clusters.

Given a collection of objects (i.e. FR instances), $Z = V^* = \{\overline{v_t^*} | \forall t = 1, \ldots, T\}$, a fuzzy set F in Z is defined as a set of ordered pairs: $F = \{(z, \varphi_F(z)) | z \in Z\}$, where $\varphi_F(z)$ is called the membership function of z in F that maps Z to $[0, 1]$. The membership function is also referred to as the degree of compatibility or degree of truth. A certain set of objects that belong to the fuzzy set F at least to the degree λ is called the λ-cut.

Assume Z is a finite, non-empty set called the universe. Let R be a fuzzy relation in $Z \times Z$, that is, $R = \{(x, y) | \forall (x, y) \in Z \times Z)\}$, then [62]:

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

If R is reflexive and symmetric, R is said to be a fuzzy compatible relation. If R is reflexive, symmetric, and transitive, R is said to be a fuzzy equivalence relation. Fuzzy clustering becomes a set of T objects of Z to be clustered, given a fuzzy compatible relation R defined on Z. Assume R^t denotes the tth power of fuzzy relation R, i.e. $R^t = R^{t-1} \circ R$, where $\circ$ is max–min composition. Then the max–min-transitive closure of R, denoted as R^*, can be defined as $R^* = \bigcup_{i=1}^{T} R^i$. Therefore, R^* is a fuzzy equivalence relation. Assume $0 \leq \lambda \leq 1$ and let $R_\lambda^* = \{(z, x) | \varphi_{R^*}(z, x) \geq \lambda, \forall x, z \in Z\}$. Then we know [63]:

(1) R_λ^* is an equivalence relation on Z; and
(2) Let $G_{R_\lambda^*}$ denote the partition on Z induced according to R_λ^*. Then for each $B \in G_{R_\lambda^*}$, there exists $E \in G_{R_{\lambda'}^*}$, so that $B \subseteq E$, as long as $\lambda' \leq \lambda$.

As a result, the λ-cut of fuzzy equivalence relation R^*, $\times R_\lambda^*$, becomes an equivalence relation. As λ increased, a finer partition can be achieved. With a hierarchy of partitions of objects, k-clusters of objects can be identified. Fig. 4 illustrates the nested partitions corresponding to a fuzzy equivalence relation defined based on the FR instances. Given different values of similarity threshold, λ, different clustering results can be obtained.

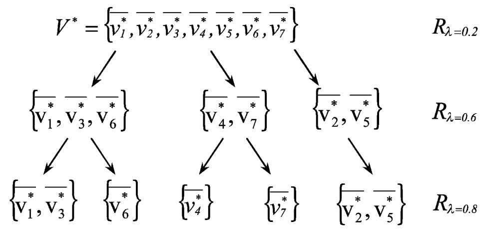
3.3. Association rule mining

FR clustering can separate data items into clusters of items, but cannot explain the clustering results specifically. It needs other methods to figure out the underlying mechanisms of CN–FR mapping between the customer and functional domains. Knowledge is usually represented in the form of rules. Rules are used for deducing the degree of association among variables, mapping data into predefined classes, identifying a finite set of categories or clusters to describe the data, etc. Therefore, this research employs association rules to explain the meaning of each FR cluster as well as the mapping of CNs to each cluster. Association rule mining is one of the major forms of data mining and is perhaps the most common form of knowledge discovery in unsupervised learning systems [27]. Association rules are produced by finding the interesting associations or correlation relationships among a large set of data items. The flexibility of association rule induction lies in its capability to deal with those qualitative data that cannot be treated by traditional operations research methods.

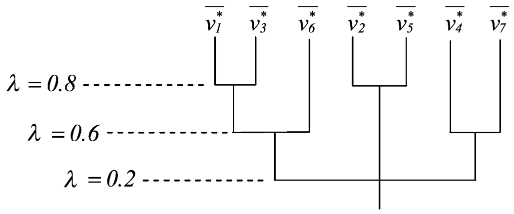
The basic problem of mining association rules is introduced by Agrawal et al. [64]. Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of literals, called items. Let DB be a database of transactions, where each transaction, T, is a set of items such that $T \subseteq I$, and each transaction is associated with an identifier, called TID. Given $Z \subseteq I$, a transaction T

$$R = \begin{array}{c|ccccccc} V^* & \overline{v_1^*} & \overline{v_2^*} & \overline{v_3^*} & \overline{v_4^*} & \overline{v_5^*} & \overline{v_6^*} & \overline{v_7^*} \\ \hline \overline{v_1^*} & 1 & 0.2 & 1 & 0.4 & 0.3 & 0.7 & 0.5 \\ \overline{v_2^*} & 0.2 & 1 & 0.2 & 0.3 & 0.8 & 0.3 & 0.4 \\ \overline{v_3^*} & 1 & 0.2 & 1 & 0.4 & 0.1 & 0.6 & 0.3 \\ \overline{v_4^*} & 0.4 & 0.3 & 0.4 & 1 & 0.1 & 0.5 & 0.7 \\ \overline{v_5^*} & 0.3 & 0.8 & 0.1 & 0.1 & 1 & 0.2 & 0.5 \\ \overline{v_6^*} & 0.7 & 0.3 & 0.6 & 0.5 & 0.2 & 1 & 0.3 \\ \overline{v_7^*} & 0.5 & 0.4 & 0.3 & 0.7 & 0.5 & 0.3 & 1 \end{array}$$

(a) A fuzzy equivalence relation defined on V^*



(b) Nested partitions of V^* induced according to R_λ



(c) Different FR clusters resulted from different values of similarity threshold

Fig. 4. Fuzzy clustering of FR instances.

contains Z if and only if $Z \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I, Y \subseteq I$, and $X \cap Y = \emptyset$. The association rule $X \Rightarrow Y$ holds in DB with confidence c if $c\%$ of the transactions in DB that contain X also contain Y. This is taken to be a conditional probability, $P(y/x|\forall x \in X, \forall y \in Y)$. The association rule $X \Rightarrow Y$ has support s in DB if $s\%$ of the transactions in DB contain X and Y. The support is taken to be a probability, $P(x \wedge y|\forall x \in X, \forall y \in Y)$.

While the confidence denotes the strength of implication, the support indicates the frequencies of the occurring patterns in the rule. Given a minimum confidence threshold, min_conf, and a minimum support threshold, min_sup, the problem of mining association rules becomes a searching for all the association rules whose confidence and support are larger than the respective thresholds. Based on if can meet the thresholds (min_conf and min_sup) or not, association rules are distinguished between strong rules and weak ones. A set of items is referred to as an itemset. An itemset that contains k items is called a k-itemset. Given a minimum support threshold, min_sup, an itemset is called large if its support is no less than min_sup. Association rule mining involves a two-step process [64]:

(1) Discover all large itemsets whose support is larger than the predetermined minimum support threshold. Itemsets with minimum support are called frequent itemsets; and
(2) Generate strong association rules from the large itemsets.

The most crucial factor affecting the performance of mining association rules lies in the first step. After the large itemsets are identified, the corresponding association rules can be derived in a straightforward manner. Efficient counting of large itemsets is hence the focus of most prior studies on algorithms for mining association rules.

4. ARMS architecture and implementation

Knowledge discovery for CN–FR mapping mechanisms is an interactive and iterative process. Based on association rule mining, an inference system can be constructed for effective product portfolio identification. Fig. 5 illustrates

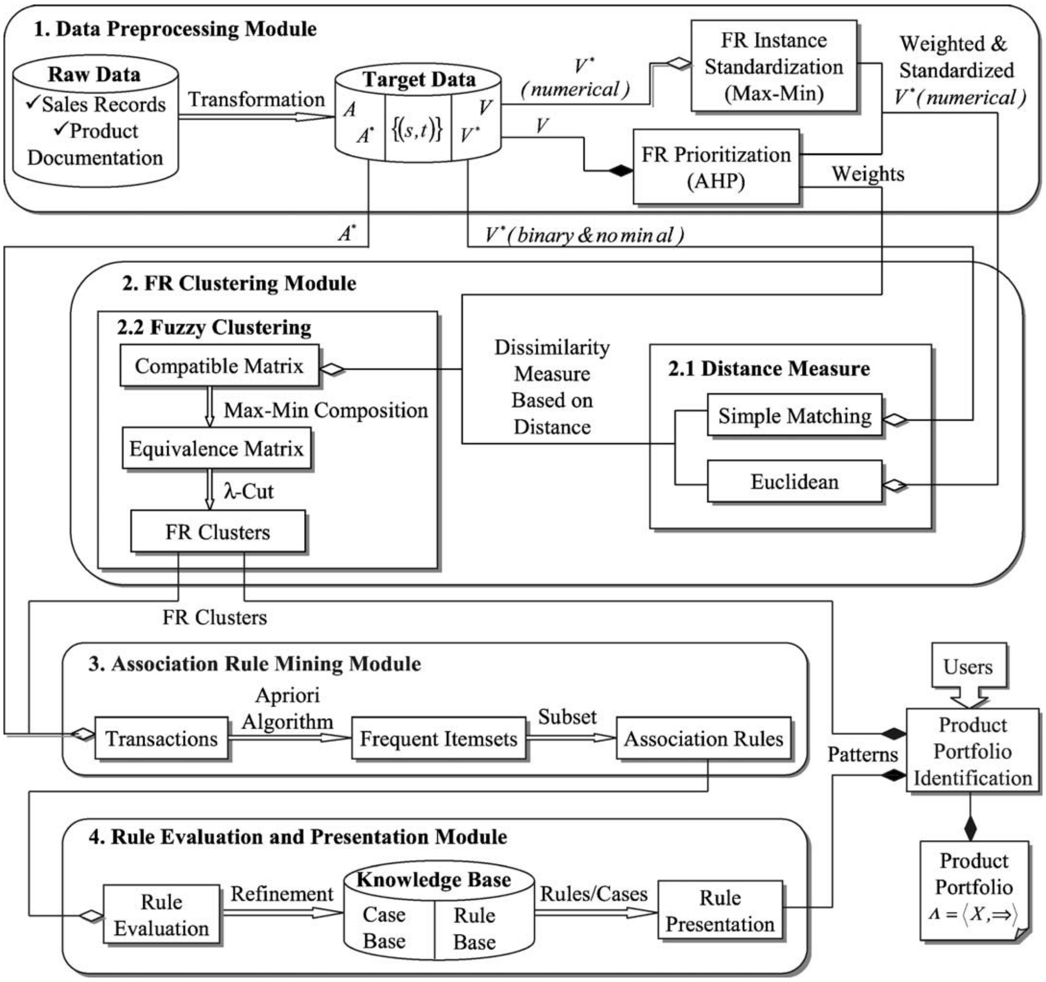


Fig. 5. ARMS architecture.

the architecture of such an association rule mining system (ARMS). The system involves four consecutive stages interacted one and another to achieve the goals, namely the data preprocessing, FR clustering, association rule mining and rule evaluation and presentation modules. First, historical data are selected and transformed to proper target data sets, which are further analyzed and preprocessed for subsequent mining procedures. The data mining procedure then starts to search for interesting patterns by the clustering module and rule-mining module. After mining of association rules, the work of rule evaluation is performed to eliminate any weak rules under the initial criteria predefined by the system. The useful rules are stored with different presentation styles in the knowledge base that may be in the forms of case bases, rule bases, and others. Equipped with such knowledge about the patterns of CNs, FRs and their mappings, the system can provide better recommendations and high-degree predictions to improve portfolio identification.

4.1. Data preprocessing module

Before proceeding to rule mining of data sets, raw data must be preprocessed in order to be useful for knowledge discovery. Three tasks are involved at this stage, as described below.

(1) Target data transformation. Generally, there are lots of data records in a company's databases. Only those records that correlate closely with the mining purpose are taken into account. Based on raw data stored in the company, target data sets should be identified, involving such data cleaning and filtering tasks as integration of multiple databases, removal of noises, handling of missing data files, etc.

All target data should be organized into a proper transaction database. This involves understanding of variables, selection of attributes and metrics, and identification of entity relationships among data. Within the ARMS, sales records and product documentation are transformed into transaction data (TID). Transaction data consists of customer records (C) and their ordered products (P). Each customer is described by his choices of certain options (A^*) for some functional features (A). The product ordered by this customer is described by specific values (V^*) of related FRs (V). The results of CN–FR mappings, i.e. $\langle \overline{a_s^*}, \overline{v_t^*} \rangle$, are embodied in the transaction records ($\langle C, P \rangle$). Fig. 6 shows the entity relationships among these target data sets.

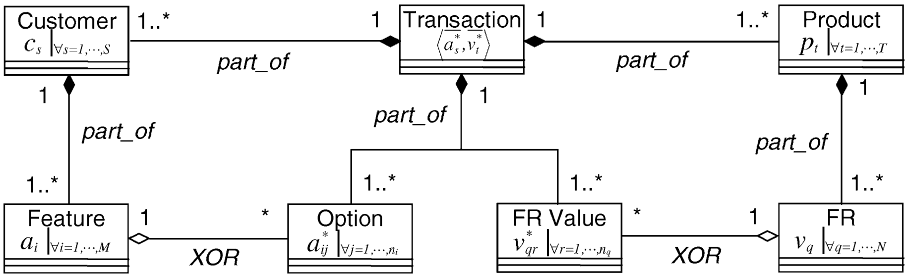


Fig. 6. Entity relationships of target data sets.

(2) Prioritization of FR variables. The specification of FRs involves multiple variables, i.e. $V=\{v_{\hat{q}}|\forall\hat{q}=1,\dots,N\}$. These FR variables contribute to the overall functionality of a product differently—some may play more roles than others. Hence, FR variables should be prioritized to differentiate their different effects, in particular those important ones. The relative importance of FR variables is usually quantified by assigning different weights. That is, each v_q is associated with a weight, w_q, subjective to $\sum_{q=1}^{N} w_q = 1$. For the ARMS, the AHP [52] is adopted for the prioritization of FR variables, owing to its advantages in maintaining consistence among a large number of variables through pair-wise comparisons.

(3) Standardization of FR values. Prior to clustering analysis of FR instances, all V^* data need to be transformed into standard forms because FR variables may involve different metrics and ranges of values. In general, expressing a variable in smaller units will lead to a larger range for that variable, and thus a larger impact on the clustering structure. To avoid dependence on the choice of different metrics or dominance of certain variables over others, those FR instances that are of numerical type should be standardized to become dimensionless. This is achieved by normalization. Many methods are available such as the z-score method, the max–min normalization method [60]. The ARMS adopts the latter method. Assume some of the FR variables,$v_k \in V|\forall k=1,\dots,Q\le N$, are of numerical type. It means that their values, $v_{kr}^* \in V_k^*|\forall r=1,\dots,n_k$, are numerical, where n_k refers to the number of values that v_k can assume. Applying the max–min method, each individual value of v_k, v_{kr}^*, can be normalized to become a dimensionless number ranged between 0 and 1, that is,

$$N\_v_{kr}^* = \frac{v_{kr}^* - \min\{v_{kj}^*|\forall j=1,\dots,n_k\}}{\max\{v_{kj}^*|\forall j=1,\dots,n_k\} - \min\{v_{kj}^*|\forall j=1,\dots,n_k\}}, \tag{1}$$

where $N\_v_{kr}^*$ denotes the normalized value for the rth value of FR v_k, v_{kr}^* is the original value of v_k, and $\max\{v_{kj}^*|\forall j=1,\dots,n_k\}$ and $\min\{v_{kj}^*|\forall j=1,\dots,n_k\}$ are the maximum and minimum values among all values of v_k with size-n_k, respectively.

In some cases, those non-numerical FR instances, such as nominal FRs, should be transformed into normalized numerical values. For instance, the data type of FR 'coating material' is originally of nominal type (i.e. character strings). A scaling transformation can be applied such that, for example, 'Au coating' is supplanted by 0.2, 'Alloy coating' becomes 0.4, and so on. When all FR instances possess the same measurements and ranges, we can proceed to the FR clustering process.

4.2. FR clustering module

Within the ARMS, FR clustering includes two steps: distance measure and fuzzy clustering. As a preparatory stage for fuzzy clustering, the distance measure module measures the dissimilarity between FR instances in order to define the fuzzy compatible relations among such data objects.

(1) Distance measure. In general, each FR instance, $\overline{v_t^*} = [v_{1t}^*, v_{2t}^*, \dots, v_{qt}^*, \dots, v_{Nt}^*] \in V^*$, where $\forall v_{qt}^* \equiv v_{qr}^*, \exists v_{qr}^* \in V_q^*, \forall r=1,\dots,n_q$, may involve three types of FR variables: numerical, binary, and nominal FRs. For example, v_{1t}^* may be a numerical value whilst v_{2t}^* may be a binary or nominal value. The distance between any two FR instances means the dissimilarity of them and thus is measured as a composite distance of three distance components corresponding to these three types of FR variables.

Numerical FRs—a number of methods of distance measure have been proposed for purpose of numerical clustering, including the Euclidean distance, Manhattan distance, Minkowski distance and weighted Euclidean distance measure [60]. The ARMS employs the weighted Euclidean distance. It is computed as the following,

$$d_{\text{numerical}}(\overline{v_i^*}, \overline{v_j^*}) = \sqrt{\sum_{q=1}^{Q} (w_q(N\_v_{qi}^* - N\_v_{qj}^*))^2}, \tag{2}$$

where $d_{\text{numerical}}(\overline{v_i^*}, \overline{v_j^*})$ indicates the numerical distance between two FR instances, $\overline{v_i^*}$ and $\overline{v_j^*}, \forall \overline{v_i^*}, \overline{v_j^*} \in V^*, w_q$ means the relative importance of the qth numerical FR variable, $v_q \in V^{\text{numerical}} \subseteq V$, Q represents the total number of numerical FR variables among the total size-N FR variables ($Q \le N$), and $N\_v_{qi}^*$ and $N\_v_{qj}^*$ denote the normalized values of original v_{qi}^* and v_{qj}^* according to Eq. (1), respectively,

Binary FRs—a binary variable assumes only two states: 0 or 1, where 0 means the variable is absent and 1 means it is present. The ARMS uses a well-accepted coefficient for assessing the distance between symmetric binary variables, called the simple matching coefficient [60]. It is calculated as the following,

$$d_{\text{binary}}(\overline{v_i^*}, \overline{v_j^*}) = \frac{\alpha_2 + \alpha_3}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}, \tag{3}$$

where $d_{\text{binary}}(\overline{v_i^*}, \overline{v_j^*})$ indicates the binary distance between two FR instances, $\overline{v_i^*}$ and $\overline{v_j^*}, \forall \overline{v_i^*}, \overline{v_j^*} \in V^*$, α_1 is the total number of binary FR variables in V (i.e. $v_q \in V^{\text{binary}} \subseteq V$) that equal to 1 for both $\overline{v_i^*}$ and $\overline{v_j^*}$, α_2 is the total number of binary FR variables that equal to 1 for $\overline{v_i^*}$ but 0 for $\overline{v_j^*}$, α_3 is the total number of binary FR variables that equal to 0 for $\overline{v_i^*}$ but 1 for $\overline{v_j^*}$, and α_4 is the total number of binary FR variables that equal to 0 for both $\overline{v_i^*}$ and $\overline{v_j^*}$.

Nominal FRs—a nominal variable can be regarded as a generalization of a binary variable in that it can take on more than two states. This type of variables cannot be expressed by numerical values but by qualitative expressions with more than one option. Therefore, the simple matching coefficient can also be used here to measure the nominal distance between two FR instances containing nominal FR variables [60]:

$$d_{\text{nominal}}(\overline{v_i^*}, \overline{v_j^*}) = \frac{\beta - \gamma}{\beta}, \tag{4}$$

where $d_{\text{nomical}}(\overline{v_i^*}, \overline{v_j^*})$ indicates the nominal distance between two FR instances, $\overline{v_i^*}$ and $\overline{v_j^*}, \forall \overline{v_i^*}, \overline{v_j^*} \in V^*$, γ means the total number of nominal FR variables in V (i.e. $v_q \in V^{\text{nomical}} \subseteq V$) that assume the same states for $\overline{v_i^*}$ and $\overline{v_j^*}$; and β is the total number of nominal variables among total size-N FR variables ($\beta \le N$).

Given a set of FR variables, $V \equiv \{v_1, v_2, \ldots, v_N\}$, every FR instance assumes a certain value for each of the FR variable, and thus consists of a combination of numerical, binary and/or nominal FR values, that is, $V^{\text{numerical}} \cup V^{\text{binary}} \cup V^{\text{nominal}} = V$. As a result, the overall distance between $\overline{v_i^*}$ and $\overline{v_j^*}$ comprises three components: the numerical, binary and nominal distances. A composite distance can thus be obtained by the weighted sum:

$$\begin{aligned} d(\overline{v_i^*}, \overline{v_j^*}) = {} & W_{\text{numerical}} d_{\text{numerical}}(\overline{v_i^*}, \overline{v_j^*}) \\ & + W_{\text{binary}} d_{\text{binary}}(\overline{v_i^*}, \overline{v_j^*}) \\ & + W_{\text{nominal}} d_{\text{nominal}}(\overline{v_i^*}, \overline{v_j^*}), \end{aligned} \tag{5}$$

$$\sum W_{\text{numerical}} + W_{\text{binary}} + W_{\text{nominal}} = 1, \tag{6}$$

where $W_{\text{numerical}}$, W_{binary} and W_{nominal} refer to the relative importance of numerical, binary and nominal distances, respectively. These weights can be determined in the similar way as that of FR variables—applying the AHP.

(2) Fuzzy clustering. The first step of fuzzy clustering is to define a fuzzy compatible relation, R, for a given set of FR instances, $V^* = \{\overline{v_1^*}, \overline{v_2^*}, \ldots, \overline{v_T^*}\}$. The R is constructed in a matrix form, that is, $R = [\rho(\overline{v_i^*}, \overline{v_j^*})]_{T\times T} | \forall (\overline{v_i^*}, \overline{v_j^*}) \in V^* \times V^*$, where $(\overline{v_i^*}, \overline{v_j^*})$ suggests pair-wise relationships among FR instances. Within the context of FR clustering, R is called the compatible matrix. A matrix element $\rho(\overline{v_i^*}, \overline{v_j^*})$ indicates the similarity grade between any two FR instances, $\overline{v_i^*}$ and $\overline{v_j^*}$. As a measure of similarity, it can be derived from the aforementioned dissimilarity measure that is determined by the distance between FR instances. Then we have the following:

(a) Normalize the distance measure between $\overline{v_i^*}$ and $\overline{v_j^*}$ based on Eqs. (1) and (5), i.e.,

$$N\_d(\overline{v_i^*}, \overline{v_j^*}) = \frac{d(\overline{v_i^*}, \overline{v_j^*}) - \min\{d(\overline{v_x^*}, \overline{v_y^*}) | \forall x, y = 1, \ldots, T\}}{\max\{d(\overline{v_x^*}, \overline{v_y^*}) | \forall x, y = 1, \ldots, T\} - \min\{d(\overline{v_x^*}, \overline{v_y^*}) | \forall x, y = 1, \ldots, T\}} \tag{7}$$

where $N\_d(\overline{v_i^*}, \overline{v_j^*}) \in [0,1]$ is the normalized value of original distance $d(\overline{v_i^*}, \overline{v_j^*})$, and $d(\overline{v_x^*}, \overline{v_y^*}) | \forall \overline{v_x^*}, \overline{v_y^*} \in V^*$ stands for a distance measure between any two FR instance based on pair-wise comparisons, $(x,y) \in T \times T$; and

(b) Derive the similarity grade $\rho(\overline{v_i^*}, \overline{v_j^*})$ from normalized distance measure $N\_d(\overline{v_i^*}, \overline{v_j^*})$, since it indicates the dissimilarity, i.e.,

$$\rho(\overline{v_i^*}, \overline{v_j^*}) = 1 - N\_d(\overline{v_i^*}, \overline{v_j^*}). \tag{8}$$

Hence, we have $0 \le \rho(\overline{v_i^*}, \overline{v_j^*}) \le 1$. In addition, we can infer that $\rho(\overline{v_i^*}, \overline{v_i^*}) = 1 | \forall i = 1, \ldots, T$, suggesting that R is reflexive, and $\rho(\overline{v_i^*}, \overline{v_j^*}) = \rho(\overline{v_j^*}, \overline{v_i^*}) | \forall i, j = 1, \ldots, T$, suggesting R is symmetrical. As a result, matrix $R = [\rho(\overline{v_i^*}, \overline{v_j^*})]_{T\times T} | \rho(\overline{v_i^*}, \overline{v_j^*}) \in [0,1]$ becomes a fuzzy compatible relation defined on V^*. Representing a subset of Cartesian product $V^* \times V^*$, matrix R is called a fuzzy compatible matrix.

The second step is to construct a fuzzy equivalence relation for V^* with transitive closure of the fuzzy compatible relation defined above. The fuzzy compatible matrix R is a fuzzy equivalence matrix if and only if the transitive condition can be met, i.e.,

$$\rho(\overline{v_i^*}, \overline{v_j^*}) \ge \max\{\min\{\rho(\overline{v_i^*}, \overline{v_z^*}), \rho(\overline{v_z^*}, \overline{v_j^*}) | \forall \overline{v_i^*}, \overline{v_z^*}, \overline{v_j^*} \in V^*\}\}. \tag{9}$$

To convert a compatible matrix to an equivalence matrix, the 'continuous multiplication' method is often used. Multiplication in fuzzy relations is also known as max–min composition [62]. Let $R(\overline{v_i^*}, \overline{v_z^*})$ and $R(\overline{v_z^*}, \overline{v_j^*})$ be two fuzzy compatible relations, then $R \circ R = [(\overline{v_i^*}, \overline{v_j^*}), \max\{\min\{\rho(\overline{v_i^*}, \overline{v_z^*}), \rho(\overline{v_z^*}, \overline{v_j^*})\}\}]$ is also a fuzzy compatible relation. To achieve the max–min-transitive closure of R, the flowchart of max–min composition is shown in Fig. 7.

The third step is to determine λ-cut of the equivalence matrix. The λ-cut is a crisp set, R_λ, that contains all the elements of the universe, V^*, such that the similarity grade of R is no less than λ. that is,

$$R_\lambda = [\tau(\overline{v_i^*}, \overline{v_j^*})]_{T\times T}, \tag{10}$$

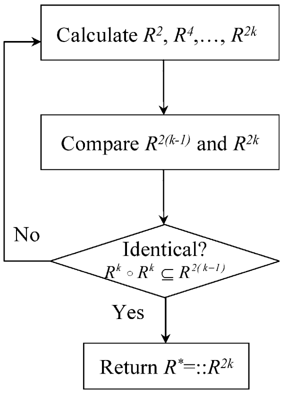


Fig. 7. The flowchart of converting a compatible matrix to an equivalent matrix.

where

$$\tau(\overline{v_i^*}, \overline{v_j^*}) = \begin{cases} 1 & \text{if } \rho(\overline{v_i^*}, \overline{v_j^*}) \geq \lambda \\ 0 & \text{if } \rho(\overline{v_i^*}, \overline{v_j^*}) < \lambda \end{cases}, \tag{11}$$

$\rho(\overline{v_i^*}, \overline{v_j^*}) \in [0, 1]$.

Then each λ-cut, R_λ, is an equivalence relation representing the presence of similarity among FR instances to the degree λ. For this equivalence matrix, there exists a partition on V^*, $\psi(R_\lambda)$, such that each compatible matrix is associated with a set, $\psi(R) = \{\psi(R_\lambda)\}$. The ARMS applies a netting method [65] to identify partitions of FR instances with respect to a given equivalence matrix. The procedure of generating a fuzzy netting graph is summarized as the following,

(a) Fill the signals of the elements in the diagonal;
(b) Replace element 1 as signal $*$ and element 0 as blank;
(c) Connect longitude and latitude to the nodes where the signals $*$ are located; and
(d) Assign the elements that are connected through the nodes into the same cluster.

The value of $\lambda \in [0,1]$ indicates the similarity threshold of a λ-cut. Given an equivalence matrix, different clustering results can be obtained according to individual similarity thresholds, as shown in Fig. 4(c). In practice, the value of λ is often determined by domain experts with many practical considerations [62]. Furthermore, latent and future customer needs, trends of product and process technologies, repeatability in design and manufacturing, ease of configuration, core competencies, and many others, are also important dimensions of decision making for the threshold.

Finally, with the hierarchy of partitions of objects, k-clusters of objects can be identified. The ARMS adopts a straightforward algorithm introduced by Wang and McCauley-Bell [63] for hierarchical clustering. Each FR cluster, $\chi_l = (\mu_l, \Delta_l)|\forall l = 1, \ldots, L$, is described by a vector of its mean, $\mu_l = [x_q^l]_N$, and a vector of its variation range, $\Delta_l = [\delta_q^l]_N$.

For a numerical FR value (i.e. $v_{qt}^* \sim v_q \in V^{\text{numerical}}$), the mean value and the variation range are calculated as the following,

$$x_q^l = \sum_{t=1}^{n_l} v_{qt}^* / n_i, \tag{12}$$

$$\delta_q^l = \max|v_{qt}^* - x_q^l|, \forall t = 1, \ldots, n_i, \tag{13}$$

where $\forall q \in [1, \ldots, N], \forall \overline{v_t^*} = [v_{qt}^*]_N \in V^*$, and n_l refers to the number of FR instances associated with the lth FR cluster, i.e. $\forall \overline{v_t^*} \sim \chi_l | \forall t = 1, \ldots, n_l \leq T$.

For a binary FR value (i.e. $v_{qt}^* \sim v_q \in V^{\text{binary}}$), the mean value and the variation range are determined as the following,

$$x_q^l = \begin{cases} 1 & \text{if } \alpha_Y \geq \alpha_N \\ 0 & \text{if } \alpha_Y < \alpha_N \end{cases}, \tag{14}$$

$$\delta_q^l = 0, \tag{15}$$

where $\forall q \in [1, \ldots, N]$, $\forall \overline{v_t^*} = [v_{qt}^*]_N \in V^*$, $\alpha_Y + \alpha_N = n_l$, n_l refers to the number of FR instances associated with the lth FR cluster, i.e. $\forall \overline{v_t^*} \sim \chi_l | \forall t = 1, \ldots, n_l \leq T$, α_Y is the total number of FR instances that assume a 1-state for v_q, and α_N is the total number of FR instances that assume a 0-state for v_q.

For a nominal FR value (i.e. $v_{qt}^* \sim v_q \in V^{\text{nominal}}$), the mean value and the variation range are determined as the following,

$$x_q^l = v_{qr}^* | r = \max(\alpha_r), \tag{16}$$

$$\delta_q^l = 0, \tag{17}$$

where $\forall q \in [1, \ldots, N]$, $\forall \overline{v_t^*} = [v_{qt}^*]_N \in V^*$, v_{qr}^* represents the rth state of v_q that possesses n_q possible states, i.e. $\exists r \in [1, n_q]$, n_l $\forall \overline{v_t^*} \sim \chi_l | \forall t = 1, \ldots, n_l \leq T$ and α_r is the total number of FR instances that assume a v_{qr}^*-state for v_q.

4.3. Association rule mining module

As reviewed in Section 3.3, traditional association rule mining ($Z \Rightarrow Y$) conforms to the general model of market basket analysis, where all items are assumed to belong to one itemset of transaction data ($Z \subseteq I$ and $Y \subseteq I$). In the ARMS scenario, rule mining involves two different itemsets, that is, $Z \subseteq A^*$ and $Y \subseteq V^*$, corresponding to the customer and functional domains, respectively. Based on the clustered FR instances, association rules regarding the mappings between individual A^* and V^* turn out to be the association rules mapping A^* to FR clusters, X, that is, $A^* \Rightarrow X$. Therefore, the ARMS's transaction data comprises these two itemsets, i.e. $\text{DB} \sim \langle A^*, X \rangle$, where $A^* = \{\overline{a_s^*} | \forall s = 1, \ldots, S\}$ and $X = \{\chi_l | \forall l = 1, \ldots, L\}$. Itemset A^* consists of a number of sales records of CNs embodied in various combinations of customer choices for diverse options of features, i.e. $\{a_{ij}^*|\forall i=1,\ldots,M,\ \forall j=1,\ldots,n_i\}$, where a_{ij}^* corresponds to the jth option of feature a_i, which possesses n_i possible options. Each customer's order indicates a particular combination of these options, i.e. $\overline{a_s^*}=[a_{ij}^*]_M$. Itemset X comprises a set of FR clusters in the form of mean-variation tuples, i.e. $\{(\mu_l,\Delta_l)=([x_q^l]_N,[\delta_q^l]_N)|\forall l=1,\ldots,L\}$. As a result, the general form of an association rule in the ARMS is given as the following,

$$\alpha_1\wedge\alpha_2\cdots\wedge\alpha_e\cdots\wedge\alpha_E \Rightarrow \beta_1\wedge\beta_2\cdots\wedge\beta_f\cdots\wedge\beta_F \text{ [Support} = s\%; \text{ Confidence} = c\%], \tag{18}$$

where $\exists\alpha_e\in\{a_{ij}^*\}_{\sum_{i=1}^M n_i}|\forall e=1,\ldots,E\le M, \exists\beta_f\in\{(x_q^l,\delta_q^l)\}_{N\times L}|\forall f=1,\ldots,F\le N$, and $s\%$ and $c\%$ refer to the support and confidence levels for this rule, respectively. They are calculated based on the following,

$$s\%=\frac{\text{count}(\alpha_1\wedge\alpha_2\cdots\wedge\alpha_E\wedge\beta_1\wedge\beta_2\cdots\wedge\beta_F)}{\text{count(DB)}}\times 100\%, \tag{19}$$

$$c\%=\frac{\text{count}(\alpha_1\wedge\alpha_2\cdots\wedge\alpha_E\wedge\beta_1\wedge\beta_2\cdots\wedge\beta_F)}{\text{count}(\alpha_1\wedge\alpha_2\cdots\wedge\alpha_E)}\times 100\%, \tag{20}$$

where count $(\alpha_1\wedge\alpha_2\cdots\wedge\alpha_E\wedge\beta_1\wedge\beta_2\cdots\wedge\beta_F)$ is the number of transaction records in DB containing all items $\alpha_1,\alpha_2,\ldots,$ and α_E as well as $\beta_1,\beta_2,\ldots,$ and β_F, count(DB) is the total number of data records contained in DB, and count$(\alpha_1\wedge\alpha_2\cdots\wedge\alpha_E)$ is the number of transaction records in DB containing all items $\alpha_1,\alpha_2,\ldots,$ and α_E. In general, count$(DB)=S$, because each *TID* corresponds to a $s-t$ pair. In addition, the set $\{\alpha_1,\alpha_2,\ldots,\alpha_e,\ldots,\alpha_E\}$ embodies a non-empty subset of $\{a_{ij}^*|\forall i\in[1,M];\exists j\in[1,n_i]\}$, whereas the set $\{\beta_1,\beta_2,\ldots,\beta_f,\ldots,\beta_F\}$ exhibits a non-empty subset of $\{(x_q^l,\delta_q^l)|\forall q\in[1,N];\exists l\in[1,L]\}$. The association rule in Eq. (18) means that the data occurrence of $\alpha_1,\alpha_2,\ldots,$ and α_E will most likely (at a s%-support and with a c%-confidence) associate with the data occurrence of $\beta_1,\beta_2,\ldots,$ and β_F.

A large number of efficient algorithms for mining association rules have been proposed [27]. The ARMS adopts a well-known algorithm, called Apriori algorithm [66] to determine frequent itemsets. Once the frequent itemsets are identified from DB, it is straightforward to generate strong association rules from them. For a large volume of source relations, the performance of rule generation may be slow. Rather than updating the association rule base continuously, the ARMS derives association rules incrementally by storing the record counts of previous computing data into the existing rule set and adding the new record counts during the new data computing process. Table 1 shows the procedure of such an incremental strategy for rule mining.

4.4. Rule evaluation and presentation module

Based on all the association rules created, the evaluation and presentation module comes into play to refine these rules in order to keep the most relevant and valuable rules in the knowledge base in the form of either case bases or rule bases. The characteristics of each FR cluster should also be explored based on the rules and the related support and confidence levels. Moreover, the causality of original association rules are defined for single feature options, as the precedent of each rule is a subset of $\{a_{ij}^*\}$ and the consequence of each rule is a subset of $\{(x_q^l,\delta_q^l)\}$ per se. Nevertheless, inference relationships do exist in various combinations of more feature options. This means a need for

Table 1
Algorithm of incremental mining of association rules in the ARMS

```
Begin
Let N = count(DB);  /* Total data record count */
Let S_m = min_sup;   /* Minimum support threshold specified by the user */
Let C_m = min_conf;  /* Minimum confidence threshold specified by the user */
   For i = 1 to N do
       Begin
       Let S = count(α1 ∧ α2 ··· ∧ αE ∧ β1 ∧ β2 ··· ∧ βF);   /* Call the Apriori algorithm */
       Let C = count(α1 ∧ α2 ··· ∧ αE);       /* Call the Apriori algorithm */
       Let s = (S/N) × 100%;
       Let c = (S/C) × 100%;
       If s ≥ S_m and c ≥ C_m
            Then Rule_i is derived;
       End if;
   End;
End;
```

Table 2
List of CNs

Feature		Option		
$a_i \mid \forall i = 1,\ldots,M$	Description	$a_{ij}^* \mid \forall j = 1,\ldots,n_i$	Code	Description
a_1	Feel of vibration	a_{11}^*	A11	Feel the vibration very strongly
		a_{12}^*	A12	Alarmed by vibration without vibrating suddenly
		a_{13}^*	A13	Sensitive to the vibration
a_2	Price	a_{21}^*	A21	Buy an expensive mobile phone with desire for a long time use
		a_{22}^*	A22	Catch up the mobile phone style occasionally at a low price
		a_{23}^*	A23	Try latest fashion of mobile phones at a moderate price
a_3	Size	a_{31}^*	A31	Portable
		a_{32}^*	A32	Comfortable to hold
		a_{33}^*	A33	Not easy to lose
a_4	Volume of sound	a_{41}^*	A41	Little noise
		a_{42}^*	A42	Alarmed independent of vibration
		a_{43}^*	A43	Alarmed by both vibration and sound
a_5	Material	a_{51}^*	A51	Green material for environment friendliness
a_6	Weight	a_{61}^*	A61	As light as possible

generating combinatorial rules. To solve such a rule refinement problem, the ARMS adopts an equivalence class method proposed by ChangChien and Lu [67]. Finally, users can retrieve all the rules stored in the knowledge base to understand the mappings of CNs to FRs clearly, to gain insights into the consequences of diverse customer preferences on the product fulfillment, and thus to justify the proper specification of product offerings in a portfolio.

5. Case study

The potential of ARMS has been tested in an electronics company that produces a large variety of vibration motors for major world-leading mobile phone manufacturers. The company had conducted extensive market studies and derived data of customer expressions of various functionality related to mobile phones. These data have been collected from market surveys and analyzed based on natural language processing. As far as the 'Alarm' function is concerned, the related features and their options are summarized in Table 2. Those CNs listed in Table 2 provide the ground for diverse specifications of the Alarm function as perceived by different mobile phone users. A variety of the Alarm functions correspond to different vibration motor designs. In other words, the Alarm-related CNs of mobile phones are fulfilled by the FRs of vibration motors. Based on existing product documentation and consultation with design engineers, we know that the functional specification of vibration motors is described by a set of FRs and their values, as shown in Table 3. Among these nine FRs, the 'Pbfree' is of binary type and the 'Coating' is of nominal type, while all the rest are numerical variables.

It is interesting to observe the difference between CNs and FRs in this case. What customers really perceive is how they feel about the Alarm function of mobile phones. Customers have no idea of the implications of this functionality in engineering–vibration motors. From the company's viewpoint, CNs refer to mobile phones, whereas FRs are related to vibration motors. When the company makes decisions about its vibration motor portfolio, it has to understand the mapping mechanisms between the customer and functional domains, as well as the tradeoffs of requirement specification between mobile phones and vibration motors.

Based on the sales records, target data are identified and organized into a transaction database, as shown in Table 4.

Table 3
List of FRs

FR			FR value		
$v_q \mid \forall q = 1,\ldots,N$	Description	Type	$v_{qr}^* \mid \forall r = 1,\ldots,n_q$	Code	Description
v_1	Current	Numerical	v_{11}^*	V11	100 mA
			v_{12}^*	V12	80 mA
			v_{13}^*	V13	60 mA
v_2	Pbfree	Binary	v_{21}^*	V21	1 (Yes)
			v_{22}^*	V22	0 (No)
v_3	Length	Numerical	v_{31}^*	V31	8 mm
			v_{32}^*	V32	12 mm
			v_{33}^*	V33	10 mm
v_4	Diameter	Numerical	v_{41}^*	V41	5 mm
			v_{42}^*	V42	4 mm
			v_{43}^*	V43	6 mm
v_5	Coating	Nominal	v_{51}^*	V51	Au
			v_{52}^*	V52	Alloy
			v_{53}^*	V53	None
v_6	Angle	Numerical	v_{61}^*	V61	40°
			v_{62}^*	V62	55°
v_7	Strength	Numerical	v_{71}^*	V71	7 kg
			v_{72}^*	V72	4 kg
v_8	Weight	Numerical	v_{81}^*	V81	2 g
			v_{82}^*	V82	3 g
v_9	Hardness	Numerical	v_{91}^*	V91	40 HB
			v_{92}^*	V92	70 HB

Table 4
Transaction database

Record (TID)	CNs ($\overline{a_s^*}\vert\forall s = 1,\dots,S$)	FRs ($\overline{v_t^*}\vert\forall t = 1,\dots,T$)
T001	A11, A21, A31, A43, A51, A61	V11, V21, V31, V42, V53, V62, V71, V82, V92
T002	A11, A21, A43, A51	V11, V21, V31, V41, V51, V61, V71, V81, V92
T003	A12, A22, A33, A61	V12, V21, V33, V43, V51, V61, V72, V82, V91
⋮	⋮	⋮
T028	A13, A22, A33, A41, A61	V13, V22, V31, V42, V52, V61, V72, V81, V91
T029	A11, A21, A31, A43, A51, A61	V12, V22, V33, V43, V52, V62, V72, V81, V92
T030	A12, A22, A33, A42, A61	V11, V22, V33, V42, V53, V61, V72, V82, V91

For illustrative simplicity, only 30 out of hundreds of transaction records are used in the case study here. As shown in Table 4, each customer order indicates the customer's choice of certain feature options related to the Alarm function of mobile phones, which is presented as a specific instance of a subset of $A = \{a_i\}_M$. Corresponding to the 30 customers (end-users of mobile phones), there are 30 vibration motors provided, whose requirement information are described as particular instances of FR vector, $[v_{qr}^*]_N$.

To prioritize nine FR variables, the AHP is applied. A 9-scale rating system is used to provide subjective judgments of preference, as shown in Table 5. The result of each weight associated with each FR variable is given in Table 6.

Due to different metrics used for FR variables, all FR instances in Table 4 need to be standardized based on the max-min normalization method. After that, the distances between every two FR instances are calculated to suggest the dissimilarity among them. The SPSS software package (SPSS 12.0 for Windows, http://www.spss.com/) is used to obtain the weighted Euclidean distance measures. The 30 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. The pair-wise measures of distances are

Table 5
Scale for subjective judgment

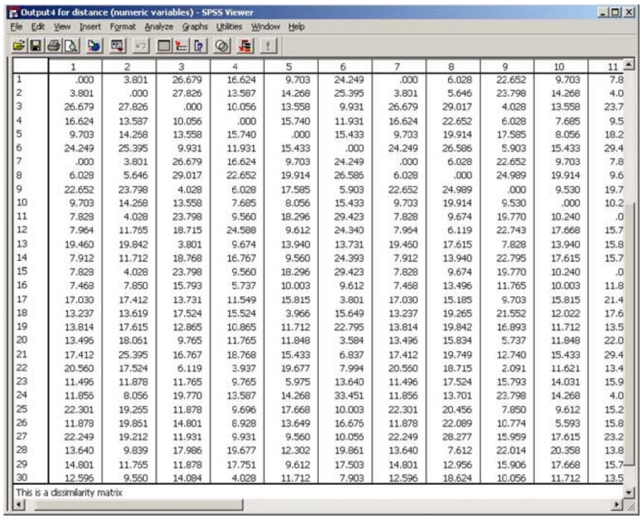
Verbal judgment of preference	Numerical rating
Extremely preferred	1
Very strong to extremely	2
Very strongly preferred	3
Strongly to very strongly	4
Strongly preferred	5
Moderately to strongly	6
Moderately preferred	7
Equally to moderately	8
Equally preferred	9

Table 6
Relative importance among FR variables

FR (v_q)	Weight (w_q)
v_1	0.219
v_2	0.304
v_3	0.046
v_4	0.031
v_5	0.019
v_6	0.066
v_7	0.157
v_8	0.095
v_9	0.083
	$\sum w_q = 1$

presented as a 30×30 matrix. Fig. 8 shows the raw data for distance measures of numerical FR instances before the normalization. The normalized distance measures of numerical FR instances are presented in a matrix form, $[N\_d_{\text{numerical}}(\overline{v_i^*}, \overline{v_j^*})]_{30\times30}$, as shown in Fig. 9. The results of distance measures for binary and nominal FR instances, $[N\_d_{\text{binary}}(\overline{v_i^*}, \overline{v_j^*})]_{30\times30}$ and $[N\_d_{\text{nominal}}(\overline{v_i^*}, \overline{v_j^*})]_{30\times30}$, are shown in Figs. 10 and 11, respectively. Based on these three distance components, the composite distances are calculated and presented as a dissimilarity matrix, $[d(\overline{v_i^*}, \overline{v_j^*})]_{30\times30}$, for all FR instances, as shown in Fig. 12. Based on the relative importance of FR variables, the weights associated with numerical, binary and nominal distance components are determined as $W_{\text{numerical}} = w_1 + w_3 + w_4 + w_6 + w_7 + w_8 + w_9 = 0.677$, $W_{\text{binary}} = w_2 = 0.304$ and $W_{\text{nominal}} = w_5 = 0.019$, respectively.

Based on the dissimilarity matrix, a fuzzy compatible matrix, R, is determined, as shown in Fig. 13. Obviously, R meets both the reflexive and symmetric characteristics. To obtain a fuzzy equivalence matrix, the max–min composition is applied. The result of $R^2 = R \circ R$ is shown in Fig. 14. As $R^2 \neq R$, we know R^2 is not a fuzzy equivalence matrix yet. Continuing to apply the max–min


Output4 for distance (numeric variables) - SPSS Viewer

	1	2	3	4	5	6	7	8	9	10	11
1	.000	3.801	26.679	16.624	9.703	24.249	.000	6.028	22.652	9.703	7.8
2	3.801	.000	27.826	13.587	14.268	25.395	3.801	5.646	23.798	14.268	4.0
3	26.679	27.826	.000	10.056	13.558	9.931	26.679	29.017	4.028	13.558	23.7
4	16.624	13.587	10.056	.000	15.740	11.931	16.624	22.652	6.028	7.685	9.5
5	9.703	14.268	13.558	15.740	.000	15.433	9.703	19.914	17.585	8.056	18.2
6	24.249	25.395	9.931	11.931	15.433	.000	24.249	26.586	5.903	15.433	29.4
7	.000	3.801	26.679	16.624	9.703	24.249	.000	6.028	22.652	9.703	7.8
8	6.028	5.646	29.017	22.652	19.914	26.586	6.028	.000	24.989	19.914	9.6
9	22.652	23.798	4.028	6.028	17.585	5.903	22.652	24.989	.000	9.530	19.7
10	9.703	14.268	13.558	7.685	8.056	15.433	9.703	19.914	9.530	.000	10.2
11	7.828	4.028	23.798	9.560	18.296	29.423	7.828	9.674	19.770	10.240	.0
12	7.964	11.765	18.715	24.588	9.612	24.340	7.964	6.119	22.743	17.668	15.7
13	19.460	19.842	3.801	9.674	13.940	13.731	19.460	17.615	7.828	13.940	15.8
14	7.912	11.712	18.768	16.767	9.560	24.393	7.912	13.940	22.795	17.615	15.7
15	7.828	4.028	23.798	9.560	18.296	29.423	7.828	9.674	19.770	10.240	.0
16	7.468	7.850	15.793	5.737	10.003	9.612	7.468	13.496	11.765	10.003	11.8
17	17.030	17.412	13.731	11.549	15.815	3.801	17.030	15.185	9.703	15.815	21.4
18	13.237	13.619	17.524	15.524	3.966	15.649	13.237	19.265	21.552	12.022	17.6
19	13.814	17.615	12.865	10.865	11.712	22.795	13.814	19.842	16.893	11.712	13.5
20	13.496	18.061	9.765	11.765	11.848	3.584	13.496	15.834	5.737	11.848	22.0
21	17.412	25.395	16.767	18.768	15.433	6.837	17.412	19.749	12.740	15.433	29.4
22	20.560	17.524	6.119	3.937	19.677	7.994	20.560	18.715	2.091	11.621	13.4
23	11.496	11.878	11.765	9.765	5.975	13.640	11.496	17.524	15.793	14.031	15.9
24	11.856	8.056	19.770	13.587	14.268	33.451	11.856	13.701	23.798	14.268	4.0
25	22.301	19.255	11.878	9.696	17.668	10.003	22.301	20.456	7.850	9.612	15.2
26	11.878	19.861	14.801	8.928	13.649	16.676	11.878	22.089	10.774	5.593	15.8
27	22.249	19.212	11.931	9.931	9.560	10.056	22.249	28.277	15.959	17.615	23.2
28	13.640	9.839	17.986	19.677	12.302	19.861	13.640	7.612	22.014	20.358	13.8
29	14.801	11.765	11.878	17.751	9.612	17.503	14.801	12.956	15.906	17.668	15.7
30	12.596	9.560	14.084	4.028	11.712	7.903	12.596	18.624	10.056	11.712	13.5

This is a dissimilarity matrix

Fig. 8. Raw data for distance measures of numerical FR instances.

$$\begin{bmatrix}
0 \\
.076 & 0 \\
.534 & .557 & 0 \\
.332 & .272 & .2 & 0 \\
.194 & .285 & .271 & .315 & 0 \\
.485 & .508 & .2 & .24 & .309 & 0 \\
0 & .076 & .534 & .332 & .194 & .485 & 0 \\
.121 & .113 & .58 & .453 & .398 & .532 & .12 & 0 \\
.453 & .476 & .08 & .121 & .352 & .118 & .453 & .5 & 0 \\
.194 & .285 & .271 & .154 & .161 & .309 & .194 & .398 & .19 & 0 \\
.157 & .08 & .476 & .191 & .366 & .588 & .157 & .193 & .4 & .2 & 0 \\
.16 & .235 & .374 & .492 & .192 & .487 & .159 & .122 & .455 & .353 & .316 & 0 \\
.39 & .397 & .076 & .193 & .279 & .275 & .39 & .352 & .157 & .279 & .316 & .23 & 0 \\
.158 & .234 & .375 & .335 & .191 & .488 & .158 & .279 & .456 & .352 & .315 & .156 & .23 & 0 \\
.157 & .081 & .476 & .191 & .366 & .588 & .157 & .193 & .396 & .2 & 0 & .316 & .316 & .315 & 0 \\
.149 & .157 & .316 & .115 & .2 & .192 & .149 & .27 & .235 & .2 & .238 & .309 & .24 & .152 & .238 & 0 \\
.341 & .348 & .275 & .231 & .316 & .076 & .341 & .304 & .194 & .316 & .429 & .342 & .2 & .343 & .43 & .116 & 0 \\
.265 & .272 & .35 & .31 & .079 & .313 & .265 & .385 & .431 & .24 & .353 & .263 & .274 & .262 & .353 & .196 & .237 & 0 \\
.277 & .352 & .257 & .217 & .234 & .456 & .276 & .397 & .338 & .234 & .272 & .274 & .113 & .118 & .272 & .195 & .312 & .23 & 0 \\
.27 & .361 & .195 & .235 & .237 & .072 & .27 & .317 & .115 & .237 & .442 & .272 & .2 & .273 & .442 & .12 & .08 & .316 & .316 & 0 \\
.348 & .508 & .335 & .375 & .309 & .137 & .348 & .395 & .255 & .309 & .588 & .27 & .275 & .35 & .588 & .192 & .076 & .313 & .319 & .07 & 0 \\
.411 & .35 & .122 & .079 & .394 & .16 & .411 & .374 & .042 & .232 & .27 & .413 & .115 & .414 & .27 & .913 & .152 & .389 & .296 & .157 & .297 & 0 \\
.23 & .238 & .235 & .195 & .12 & .273 & .23 & .35 & .316 & .28 & .318 & .228 & .16 & .072 & .318 & .08 & .197 & .115 & .115 & .2 & .273 & .274 & 0 \\
.237 & .161 & .395 & .272 & .285 & .669 & .237 & .274 & .476 & .285 & .08 & .235 & .236 & .234 & .08 & .318 & .509 & .272 & .191 & .522 & .669 & .35 & ..238 & 0 \\
.446 & .385 & .238 & .194 & .353 & .2 & .446 & .409 & .157 & .192 & .305 & .448 & .23 & .604 & .305 & .309 & .192 & .274 & .411 & .272 & .337 & .115 & .389 & .385 & 0 \\
.238 & .397 & .296 & .179 & .273 & .334 & .238 & .442 & .215 & .112 & .317 & .4 & .235 & .24 & .371 & .157 & .273 & .352 & .122 & .193 & .197 & .257 & .237 & .397 & .372 & 0 \\
.445 & .384 & .239 & .2 & .191 & .2 & .445 & .566 & .319 & .352 & .465 & .443 & .23 & .287 & .465 & .152 & .193 & .112 & .255 & .273 & .338 & .277 & .07 & .384 & .318 & .377 & 0 \\
.273 & .197 & .36 & .394 & .246 & .397 & .273 & .152 & .44 & .407 & .277 & .114 & .2 & .27 & .277 & .279 & .238 & .158 & .313 & .326 & .397 & .315 & .198 & .197 & .275 & .519 & .27 & 0 \\
.296 & .235 & .238 & .355 & .192 & .35 & .296 & .259 & .318 & .353 & .316 & .137 & .23 & .293 & .316 & .309 & .342 & .263 & .411 & .272 & .487 & .276 & .228 & .235 & .311 & .534 & .306 & .114 & 0 \\
.252 & .191 & .282 & .081 & .234 & .158 & .252 & .372 & .2 & .234 & .272 & .411 & .274 & .255 & .272 & .03 & .15 & .23 & .298 & .155 & .295 & .159 & .115 & .352 & .274 & .259 & .118 & .313 & .274 & 0
\end{bmatrix}_{30\times 30}$$

Fig. 9. Result of distance measures for numerical FR instances.

composition, $R^4 = (R \circ R) \circ (R \circ R)$ is obtained, which equals to R^2. The result of R^4 is also shown in Fig. 14. As a result, R^4 turns out to be a fuzzy equivalence matrix. Based on R^4, the λ-cut is derived with a similarity threshold setting at 0.84. The result of the λ-cut is shown in Fig. 15.

With the obtained λ-cut, a fuzzy netting graph is constructed, as shown in Fig. 16. Based on the partitions derived from the fuzzy netting graph, three clusters of FR instances are identified. The mean value and variation range for each FR cluster are calculated based on those FR instances that are grouped into this cluster. The result of FR clustering is given in Table 7, in which, for example, FR cluster, χ_1, is associated with its mean, $\mu_l = [100, \mathrm{Y}, 9.2, 4.5, \mathrm{Au}, 44.5, 6.7, 2.4, 49]$, and variation range, $\delta_l = [0, 0, 1.2, 0.5, 0, 10.5, 2.7, 0.6, 21]$, and contains 10 FR instances, including $\overline{v_1^*}, \overline{v_2^*}, \overline{v_7^*}, \overline{v_8^*}, \overline{v_{11}^*}, \overline{v_{12}^*}, \overline{v_{14}^*}, \overline{v_{15}^*}, \overline{v_{24}^*}$, and $\overline{v_{29}^*}$.

The resulted FR clusters comprise an itemset, $X = \{(x_q^l, \delta_q^l) | \forall q \in [1, 9]; \exists l \in [1, 3]\}$, as shown in Table 8. The characteristics of each FR cluster entail the specification of

$$\begin{bmatrix}
0 \\
0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}_{30\times 30}$$

Fig. 10. Result of distance measures for binary FR instances.

$$\begin{bmatrix}
0 \\
0 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0
\end{bmatrix}_{30\times 30}$$

Fig. 11. Result of distance measures for nominal FR instances.

$$\left[\begin{array}{*{30}{l}}
0 \\
.076 & 0 \\
.554 & .58 & 0 \\
.352 & .29 & .2 & 0 \\
.52 & .6 & .57 & .62 & 0 \\
.51 & .53 & .22 & .26 & .53 & 0 \\
.3 & .38 & .85 & .65 & .22 & .81 & 0 \\
.12 & .11 & .6 & .47 & .82 & .55 & .42 & 0 \\
.47 & .5 & .08 & .12 & .65 & .14 & .77 & .52 & 0 \\
.22 & .31 & .29 & .17 & .48 & .31 & .52 & .42 & .21 & 0 \\
.48 & .4 & .78 & .49 & .37 & .91 & .18 & .51 & .7 & .52 & 0 \\
.16 & .24 & .39 & .51 & .51 & .51 & .46 & .12 & .48 & .37 & .64 & 0 \\
.41 & .42 & .08 & .19 & .58 & .3 & .71 & .37 & .16 & .3 & .62 & .25 & 0 \\
.16 & .24 & .4 & .34 & .51 & .51 & .46 & .28 & .48 & .37 & .64 & .26 & .25 & 0 \\
.46 & .38 & .74 & .51 & .39 & .91 & .16 & .49 & .72 & .52 & .02 & .62 & .64 & .62 & 0 \\
.17 & .18 & .32 & .12 & .5 & .21 & .47 & .29 & .24 & .22 & .54 & .33 & .24 & .17 & .44 & 0 \\
.36 & .37 & .3 & .25 & .64 & .08 & .66 & .32 & .21 & .32 & .75 & .36 & .22 & .36 & .75 & .14 & 0 \\
.59 & .6 & .67 & .63 & .01 & .62 & .29 & .71 & .75 & .54 & .37 & .58 & .6 & .58 & .37 & .52 & .54 & 0 \\
.3 & .32 & .26 & .22 & .53 & .48 & .6 & .42 & .34 & .25 & .57 & .29 & .11 & .14 & .6 & .2 & .33 & .55 & 0 \\
.29 & .38 & .2 & .24 & .54 & .09 & .59 & .34 & .12 & .26 & .74 & .29 & .2 & .29 & .76 & .12 & .1 & .64 & .32 & 0 \\
.27 & .53 & .36 & .4 & .53 & .14 & .67 & .42 & .28 & .31 & .91 & .29 & .3 & .37 & .91 & .21 & .01 & .63 & .34 & .09 & 0 \\
.43 & .37 & .12 & .08 & .69 & .18 & .73 & .39 & .04 & .25 & .57 & .43 & .12 & .43 & .59 & .93 & .15 & .69 & .3 & .16 & .32 & 0 \\
.25 & .26 & .24 & .2 & .42 & .25 & .55 & .37 & .32 & .3 & .62 & .24 & .16 & .09 & .64 & .08 & .22 & .44 & .12 & .2 & .29 & .27 & 0 \\
.24 & .16 & .42 & .29 & .61 & .69 & .44 & .27 & .5 & .31 & .4 & .24 & .26 & .23 & .38 & .34 & .53 & .59 & .21 & .54 & .69 & .37 & .27 & 0 \\
.77 & .7 & .56 & .51 & .37 & .5 & .47 & .73 & .48 & .49 & .23 & .77 & .55 & .62 & .23 & .63 & .49 & .27 & .73 & .6 & .64 & .44 & .71 & .71 & 0 \\
.56 & .72 & .62 & .5 & .29 & .54 & .26 & .76 & .54 & .41 & .34 & .72 & .56 & .56 & .34 & .48 & .57 & .35 & .44 & .51 & .5 & .58 & .56 & .72 & .69 & 0 \\
.47 & .4 & .26 & .22 & .51 & .2 & .77 & .59 & .34 & .35 & .79 & .46 & .25 & .25 & .79 & .17 & .19 & .41 & .28 & .29 & .34 & .3 & .09 & .4 & .34 & .68 & 0 \\
.29 & .22 & .36 & .39 & .55 & .42 & .59 & .17 & .44 & .43 & .58 & .13 & .2 & .2 & .6 & .28 & .26 & .48 & .31 & .33 & .42 & .32 & .2 & .22 & .3 & .84 & .29 & 0 \\
.3 & .24 & .26 & .38 & .51 & .37 & .6 & .26 & .34 & .37 & .54 & .14 & .25 & .25 & .62 & .33 & .36 & .58 & .43 & .29 & .51 & .3 & .25 & .24 & .31 & .85 & .23 & .13 & 1 \\
.27 & .21 & .28 & .08 & .53 & .16 & .57 & .39 & .2 & .25 & .57 & .43 & .27 & .27 & .59 & .03 & .17 & .55 & .3 & .16 & .32 & .16 & .18 & .37 & .29 & .58 & .14 & .31 & .29 & 0
\end{array}\right]_{30\times 30}$$

Fig. 12. Dissimilarity matrix based on distance measures for all FR instances.

a product platform—a set of base values together with the related variation ranges, and therefore can be used to suggest standard settings for vibration motor portfolio. These items are added to the transaction database. The link of each customer order to a FR instance is then replaced with the link to the items of the FR cluster that this FR instance belongs to. To mining rules between itemsets A^* and X, a data mining tool, called Magnum Opus (Version 2.0, http://www.rulequest.com/), is employed. All data are extracted from the transaction database and input as a text file to the Magnum Opus. The system allows data to be input as identifier-item files that list customers to be analyzed in the identifier-item format. Each customer has a unique identifier consisting of two columns: one for the identifier and one for the item. The Magnum Opus provides five association metrics: leverage, lift, strength, coverage, and support, each of which is supported by a search mode. The case study only uses the support and strength modes for the handling of support and confidence measures, respectively. This is because the coverage, lift and leverage criteria are

$$\left[\begin{array}{*{30}{l}}
1 \\
.924 & 1 \\
.446 & .42 & 1 \\
.648 & .71 & .8 & 1 \\
.48 & .4 & .43 & .38 & 1 \\
.49 & .47 & .78 & .74 & .47 & 1 \\
.7 & .62 & .15 & .35 & .78 & .19 & 1 \\
.88 & .89 & .4 & .53 & .18 & .45 & .58 & 1 \\
.53 & .5 & .92 & .88 & .35 & .86 & .23 & .48 & 1 \\
.78 & .69 & .71 & .83 & .52 & .69 & .48 & .58 & .79 & 1 \\
.52 & .6 & .22 & .51 & .63 & .09 & .82 & .49 & .3 & .48 & 1 \\
.84 & .76 & .61 & .49 & .49 & .49 & .54 & .88 & .52 & .63 & .36 & 1 \\
.59 & .58 & .92 & .81 & .42 & .7 & .29 & .63 & .84 & .7 & .38 & .75 & 1 \\
.84 & .76 & .6 & .66 & .49 & .49 & .54 & .72 & .52 & .63 & .36 & .74 & .75 & 1 \\
.54 & .62 & .26 & .49 & .61 & .09 & .84 & .51 & .28 & .48 & .98 & .38 & .36 & .38 & 1 \\
.83 & .82 & .68 & .88 & .5 & .79 & .53 & .71 & .76 & .78 & .46 & .67 & .76 & .83 & .56 & 1 \\
.64 & .63 & .7 & .75 & .36 & .92 & .34 & .68 & .79 & .68 & .25 & .64 & .78 & .64 & .25 & .86 & 1 \\
.41 & .4 & .33 & .37 & .99 & .38 & .71 & .29 & .25 & .46 & .63 & .42 & .4 & .42 & .63 & .48 & .46 & 1 \\
.7 & .68 & .74 & .78 & .47 & .52 & .4 & .58 & .66 & .75 & .43 & .71 & .89 & .86 & .4 & .8 & .67 & .45 & 1 \\
.71 & .62 & .8 & .76 & .46 & .91 & .41 & .66 & .88 & .74 & .26 & .71 & .8 & .71 & .24 & .88 & .9 & .36 & .68 & 1 \\
.73 & .47 & .64 & .6 & .47 & .86 & .33 & .58 & .72 & .69 & .09 & .71 & .7 & .63 & .09 & .79 & .99 & .37 & .66 & .91 & 1 \\
.57 & .63 & .88 & .92 & .31 & .82 & .27 & .61 & .96 & .75 & .43 & .57 & .88 & .57 & .41 & .07 & .85 & .31 & .7 & .84 & .68 & 1 \\
.75 & .74 & .76 & .8 & .58 & .75 & .45 & .63 & .68 & .7 & .38 & .76 & .84 & .91 & .36 & .92 & .78 & .56 & .88 & .8 & .71 & .73 & 1 \\
.76 & .84 & .58 & .71 & .39 & .31 & .56 & .73 & .5 & .69 & .6 & .76 & .74 & .77 & .62 & .66 & .47 & .41 & .79 & .46 & .31 & .63 & .73 & 1 \\
.23 & .3 & .44 & .49 & .63 & .5 & .53 & .27 & .52 & .51 & .77 & .23 & .45 & .38 & .77 & .37 & .51 & .73 & .27 & .4 & .36 & .56 & .29 & .29 & 1 \\
.44 & .28 & .38 & .5 & .71 & .46 & .74 & .24 & .46 & .59 & .66 & .28 & .44 & .44 & .66 & .52 & .43 & .65 & .56 & .49 & .5 & .42 & .44 & .28 & .31 & 1 \\
.53 & .6 & .74 & .78 & .49 & .8 & .23 & .41 & .66 & .65 & .21 & .54 & .75 & .75 & .21 & .83 & .81 & .59 & .72 & .71 & .66 & .7 & .91 & .6 & .66 & .32 & 1 \\
.71 & .78 & .64 & .61 & .45 & .58 & .41 & .83 & .56 & .57 & .42 & .87 & .8 & .8 & .4 & .72 & .74 & .52 & .69 & .67 & .58 & .68 & .8 & .78 & .7 & .16 & .71 & 1 \\
.7 & .76 & .74 & .62 & .49 & .63 & .4 & .74 & .66 & .63 & .46 & .86 & .75 & .75 & .38 & .67 & .64 & .42 & .57 & .71 & .49 & .7 & .75 & .76 & .69 & .15 & .77 & .87 & 1 \\
.73 & .79 & .72 & .92 & .47 & .84 & .43 & .61 & .8 & .75 & .43 & .57 & .73 & .73 & .41 & .97 & .83 & .45 & .7 & .84 & .68 & .84 & .82 & .63 & .71 & .42 & .86 & .69 & .71 & 1
\end{array}\right]_{30\times 30}$$

Fig. 13. Result of R.

$$\left[\begin{array}{*{30}{l}}
1 \\
.7 & 1 \\
.64 & .74 & 1 \\
.52 & .82 & .8 & 1 \\
.7 & .62 & .85 & .85 & 1 \\
.75 & .71 & .76 & .76 & .78 & 1 \\
.7 & .7 & .56 & .648 & .78 & .79 & 1 \\
.88 & .88 & .73 & .71 & .58 & .71 & .9 & 1 \\
.76 & .76 & .92 & .88 & .58 & .76 & .53 & .71 & 1 \\
.78 & .78 & .8 & .83 & .58 & .79 & .7 & .78 & .85 & 1 \\
.7 & .62 & .58 & .6 & .78 & .51 & .82 & .6 & .52 & .6 & 1 \\
.84 & .84 & .82 & .75 & .58 & .75 & .7 & .78 & .75 & .78 & .6 & 1 \\
.76 & .75 & .92 & .8 & .52 & .78 & .59 & .75 & .82 & .81 & .8 & .75 & 1 \\
.84 & .74 & .8 & .75 & .58 & .79 & .7 & .84 & .76 & .78 & .6 & .84 & .76 & 1 \\
.7 & .62 & .58 & .62 & .78 & .56 & .84 & .62 & .53 & .78 & .82 & .62 & .62 & .62 & 1 \\
.83 & .83 & .8 & .78 & .58 & .89 & .7 & .83 & .78 & .83 & .83 & .83 & .81 & .83 & .62 & 1 \\
.68 & .82 & .79 & .79 & .58 & .72 & .64 & .71 & .86 & .79 & .81 & .76 & .75 & .83 & .62 & .79 & 1 \\
.7 & .62 & .56 & .49 & .79 & .89 & .78 & .58 & .59 & .56 & .71 & .65 & .59 & .59 & .73 & .59 & .59 & 1 \\
.74 & .8 & .89 & .81 & .58 & .79 & .7 & .73 & .84 & .78 & .72 & .76 & .78 & .75 & .62 & .83 & .8 & .59 & 1 \\
.83 & .82 & .8 & .8 & .58 & .81 & .7 & .71 & .8 & .76 & .6 & .75 & .8 & .75 & .62 & .79 & .91 & .59 & .8 & 1 \\
.79 & .79 & .78 & .79 & .58 & .86 & .7 & .73 & .82 & .78 & .68 & .73 & .78 & .79 & .56 & .86 & .76 & .59 & .71 & .76 & 1 \\
.77 & .71 & .78 & .8 & .63 & .86 & .62 & .73 & .78 & .83 & .7 & .75 & .78 & .75 & .62 & .88 & .82 & .59 & .78 & .78 & .82 & 1 \\
.83 & .82 & .8 & .8 & .58 & .79 & .7 & .76 & .8 & .8 & .71 & .76 & .8 & .81 & .62 & .83 & .78 & .58 & .78 & .88 & .79 & .8 & 1 \\
.84 & .84 & .74 & .71 & .61 & .73 & .7 & .74 & .74 & .76 & .88 & .76 & .75 & .76 & .62 & .82 & .74 & .62 & .79 & .74 & .73 & .74 & .79 & 1 \\
.71 & .71 & .71 & .71 & .63 & .56 & .77 & .7 & .74 & .74 & .76 & .88 & .76 & .75 & .76 & .62 & .82 & .74 & .62 & .79 & .74 & .73 & .74 & .79 & 1 \\
.7 & .62 & .59 & .56 & .71 & .59 & .74 & .58 & .89 & .59 & .74 & .59 & .56 & .59 & .74 & .59 & .59 & .71 & .59 & .59 & .59 & .59 & .59 & .62 & .66 & 1 \\
.83 & .75 & .78 & .78 & .6 & .8 & .6 & .71 & .8 & .78 & .6 & .75 & .78 & .75 & .6 & .79 & .8 & .89 & .78 & .8 & .8 & .8 & .83 & .74 & .59 & .59 & 1 \\
.78 & .83 & .8 & .81 & .58 & .75 & .7 & .78 & .8 & .72 & .7 & .76 & .8 & .76 & .62 & .88 & .78 & .73 & .8 & .8 & .71 & .8 & .8 & .78 & .7 & .57 & .75 & 1 \\
.84 & .76 & .75 & .75 & .58 & .74 & .7 & .76 & .74 & .71 & .6 & .76 & .75 & .76 & .62 & .76 & .75 & .69 & .75 & .75 & .71 & .75 & .76 & .76 & .71 & .59 & .75 & .86 & 1 \\
.83 & .82 & .85 & .92 & .58 & .8 & .7 & .79 & .88 & .83 & .71 & .76 & .81 & .76 & .62 & .78 & .84 & .71 & .78 & .84 & .82 & .82 & .8 & .79 & .71 & .59 & .8 & .78 & .76 & 1
\end{array}\right]_{30\times 30}$$

Fig. 14. Result of R^2 and R^4.

not considered in the Apriori algorithm. Under either search mode, the Magnum Opus finds a number of association rules specified by the user. The search guarantees that only those rules with the highest values on the specified metric are found according to user specified search settings.

The Magnum Opus will find fewer than the specified number of association rules if the search is terminated by the user or there are fewer than the specified number of associations that satisfy user specified search settings. In our case, the maximum number of associations is set to be 10,000 to make sure that the association rules can be derived completely. The minimum leverage, minimum lift, minimum strength, minimum coverage, and minimum support are set as 0, 1.0 (default value required by the system), 0.6, 0, and 0.5, respectively. Fig. 17 shows the setting of search modes and their metrics as well as the rule induction process in the Magnum Opus.

At the end of mining, the system generates 37 association rules, as shown in Table 9. These rules serve as the basis of knowledge discovery. Some rules, for example, Rules 31, 32 and 33, are coupled and should be aggregated into one. The possibility of some rule combinations is also considered to discover more implicit rules. For example, Rules 15, 16 and 17 together with Rules 23, 24 and 25 can give more hints to optimize the size of motors. In addition to such rule refinement, the characteristics of each FR cluster and implicit relationships among them are explored to gain more understanding of vibration motor design specifications, so as to identify prominent settings of particular FR variables, to analyze the tradeoffs between different customer perceptions on mobile phones and the relevant FR values of vibration motors, and so on. All the identified patterns of CNs, FRs and the mapping are built into the knowledge base and are utilized to assist users in portfolio decision making based on the generated portfolio (Table 8).

6. Sensitivity analysis

To evaluate the performance of the ARMS, the sensitivity of the identified product portfolio is studied with respect to varying values of data mining parameters, including the similarity threshold, and the minimum support and confidence levels. These parameters involve two

$$\left[\begin{array}{*{30}{l}}
1 \\
0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]_{30\times 30}$$

Fig. 15. Result of a λ-cut with $\lambda = 0.84$.

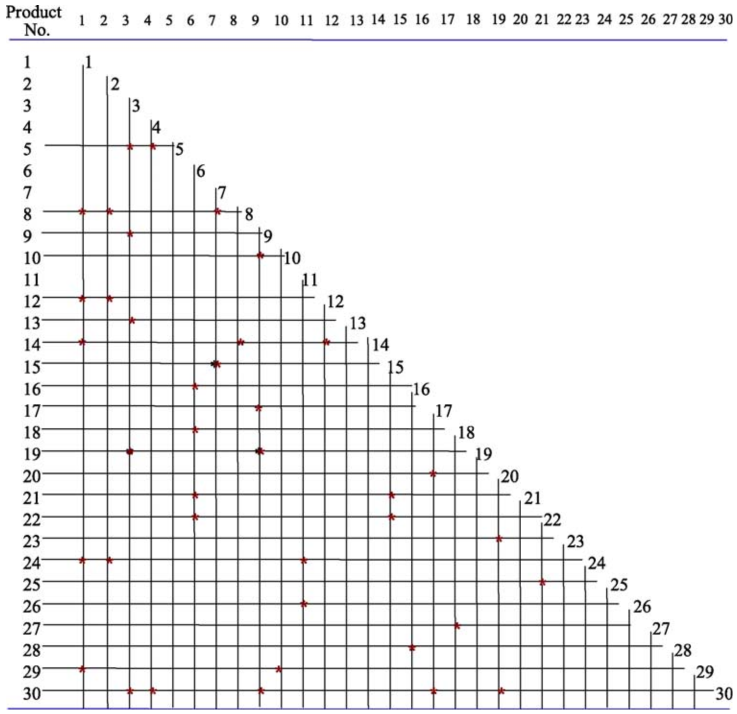


Fig. 16. Fuzzy netting graph.

modules of the ARMS: FR clustering and association rule mining, respectively.

The FR clustering module entails the specification of an optimal value of similarity threshold of λ-cut. Essentially, it gives rise to a tradeoff issue of FR granularity inherent in mass customization [58]. With a large (small) value of λ-cut, more (less) FR clusters will be identified. These FR clusters affect the downstream planning of the product and process platforms. In terms of the economic latitude, the cost of introducing more FRs (i.e. finer FR clustering) 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 FRs is too spread or too low a level of aggregation, such as at the nuts and bolts level, then the number of DPs and PVs may be too many and the product fulfillment becomes difficult to leverage the investments. On the contrary, if the FR aggregation is at a very high level, such as complete subassemblies, then the repetition may not be sufficient to take advantage of mass production efficiency.

An optimal granularity can normally be determined by assessing the performance of the product and process platforms in accordance with the resulted FR clusters. Jiao et al. [68] apply the real option theory to the valuation of flexibility enabled by the product and process platforms. On the other hand, the construction of the product and process platforms embodies a type of fixed costs [12,59]. Therefore, we introduce a performance measure of λ-cut, Ψ^{λ}, as the following

$$\Psi^{\lambda} = \frac{E[V]}{C^{F}}, \tag{21}$$

where $E[V]$ denotes the expected value of the product and process platforms, which is determined based on a real option framework [68,69], and C^{F} stands for the fix cost of the product and process platforms. Furthermore, Jiao and

Table 7
Result of FR clustering

FR cluster			Clustered FR instances ($\{\overline{v_t^*} \sim \chi_l \mid \forall t = 1, \ldots, n_l \le T\}$)
χ_l	Mean value (μ_l)	Variation range (Δ_l)	
χ_1	[100, Y, 9.2, 4.5, Au, 44.5, 6.7, 2.4, 49]	[0, 0, 1.2, 0.5, 0, 10.5, 2.7, 0.6, 21]	$\{\overline{v_1^*}, \overline{v_2^*}, \overline{v_7^*}, \overline{v_8^*}, \overline{v_{11}^*}, \overline{v_{12}^*}, \overline{v_{14}^*}, \overline{v_{15}^*}, \overline{v_{24}^*}, \overline{v_{29}^*}\}$
χ_2	[78.3, Y, 11.17, 5.5, Alloy, 47, 4.5, 2.42, 57.5]	[21.7, 0, 1.17, 0.5, 0, 8, 2.5, 0.58, 17.5]	$\{\overline{v_3^*}, \overline{v_4^*}, \overline{v_5^*}, \overline{v_9^*}, \overline{v_{10}^*}, \overline{v_{13}^*}, \overline{v_{17}^*}, \overline{v_{19}^*}, \overline{v_{20}^*}, \overline{v_{23}^*}, \overline{v_{26}^*}, \overline{v_{30}^*}\}$
χ_3	[67.5, Y, 10.75, 5.13, None, 42.5, 5.13, 2.38, 47.5]	[12.5, 0, 1.25, 0.87, 0, 12.5, 1.87, 0.62, 22.5]	$\{\overline{v_6^*}, \overline{v_{16}^*}, \overline{v_{18}^*}, \overline{v_{21}^*}, \overline{v_{22}^*}, \overline{v_{25}^*}, \overline{v_{27}^*}, \overline{v_{28}^*}\}$

Table 8
Specification of vibration motor portfolio based on FR clusters

FR variable	FR value	
	Base value	Variation range
Current (mA)	100	±0
	78.3	±21.7
	67.5	±12.5
Pbfree	1 (Yes)	±0
Length (mm)	9.2	±1.2
	11.17	±1.17
	10.75	±1.25
Diameter (mm)	4.5	±0.5
	5.5	±0.5
	5.13	±0.87
Coating	Au	±0
	Alloy	±0
	None	±0
Angle (°)	44.5	±10.5
	47	±8
	42.5	±12.5
Strength (kg)	6.7	±2.7
	4.5	±2.5
	5.13	±1.87
Weight (g)	2.4	±0.6
	2.42	±0.58
	2.38	±0.62
Hardness (HB)	49	±21
	57.5	±17.5
	47.5	±22.5

Tseng [16] posit the rationale of justifying cost implications of the product and process platforms based on process variations. Following [16,68], we employ a process capability index to measure the above fixed cost, as the following,

$$C^F = \beta^F \mathrm{e} \frac{1}{\mathrm{PCI}} = \beta^F \mathrm{e} \frac{6\sigma}{\mathrm{USL}-\mathrm{LSL}}, \tag{22}$$

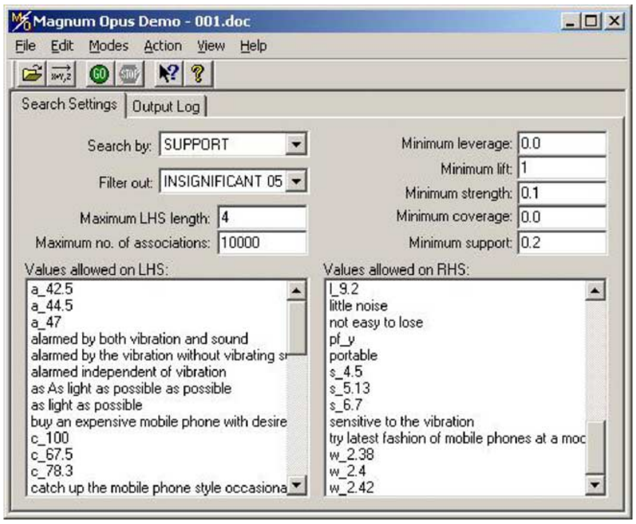


Fig. 17. Association rule induction in the Magnum Opus.

Table 9
Result of association rule mining

Rule 1: Green material for environment friendliness\=>pf_y\[Support = 0.882; Strength = 1.000];

Rule 2: Alarmed independent of vibration\&Not easy to lose\&Catch up the mobile phone style occasionally at a low price\=>h_57.5[±17.5]\[Support = 0.265; Strength = 0.900];

Rule 3: Alarmed independent of vibration\&Try latest fashion of mobile phones at a moderate price\&Not easy to lose\=> c_78.3[±21.7]\[Support = 0.265; Strength = 0.900];

Rule 4: Alarmed independent of vibration\&Buy an expensive mobile phone with desire for a long time use\=>l_11.17[±1.17]\[Support = 0.265; Strength = 0.900];

Rule 5: Alarmed independent of vibration\=>h_57.5[±17.5]\[Support = 0.294; Strength = 0.833];

Rule 6: Not easy to lose\&Alarmed independent of vibration\=>a_47[±8]\[Support = 0.265; Strength = 0.900];

Rule 7: Not easy to lose\&Comfortable to hold\&Catch up the mobile phone style occasionally at a low price\ = > w_2.42[±0.58]\[Support = 0.265; Strength = 0.750];

Rule 8: Not easy to lose\=>w_2.42[±0.58]\&h_57.5[±17.5]\[Support = 0.265; Strength = 0.900];

Rule 9: Catch up the mobile phone style occasionally at a low price\=>co_None\[Support = 0.324; Strength = 0.688];

Rule 10: Buy an expensive mobile phone with desire for a long time use\&Feel the vibration very strongly\=>h_49[±21]\[Support = 0.206; Strength = 1.000];

Rule 11: Buy an expensive mobile phone with desire for a long time use\=>s_6.7[±2.7]\[Support = 0.206; Strength = 1.000];

Rule 12: Buy an expensive mobile phone with desire for a long time use\&Alarmed by both vibration and sound\ = >a_44.5[±10.5]\[Support = 0.206; Strength = 1.000];

Rule 13: Buy an expensive mobile phone with desire for a long time use\&Portable\=>a_44.5[±10.5\[Support = 0.265; Strength = 0.818];

Rule 14: Buy an expensive mobile phone with desire for a long time use\=>co_Au\[Support = 0.206; Strength = 1.000];

Rule 15: Feel the vibration very strongly\&Portable\=>l_9.2[±1.2]\[Support = 0.206; Strength = 0.875];

Rule 16: Feel the vibration very strongly\=>c_100[±0]\[Support = 0.206; Strength = 0.875];

Rule 17: Feel the vibration very strongly\&As light as possible\=>d_4.5[±0.5]\[Support = 0.265; Strength = 0.750];

Rule 18: As light as possible\=>a_42.5[±12.5]\[Support = 0.206; Strength = 0.875];

Rule 19: As light as possible\&Little Noise=>w_2.38[±0.62]\[Support = 0.206; Strength = 0.875];

Rule 20: As light as possible\=>co_None\[Support = 0.206; Strength = 0.875]

Rule 21: Alarmed by the vibration without vibrating suddenly\=>s_4.5[±2.5]\[Support = 0.294; Strength = 0.833];

Rule 22: Alarmed by the vibration without vibrating suddenly\=>l_11.17[±1.17]\[Support = 0.294; Strength = 0.833];

Rule 23: Portable\&As light as possible\=>d_4.5[±0.5]\[Support = 0.265; Strength = 0.818];

Rule 24: Portable\&Feel the vibration very strongly\=>l_9.2[±1.2]\[Support = 0.265; Strength = 0.818];

Rule 25: Portable\=>a_44.5[±10.5]\[Support = 0.294; Strength = 0.833];

(continued on next page)

Table 9 *(Continued)*

Rule 26: Sensitive to the vibration\=>d_5.13[±0.87]\[Support = 0.235; Strength = 0.800];
Rule 27: Sensitive to the vibration\&Little noise\=>c_67.5[±12.5]\[Support = 0.235; Strength = 0.800];
Rule 28: Sensitive to the vibration\&Little noise\&As light as possible\=>h_47.5[±22.5]\[Support = 0.235; Strength = 0.727];
Rule 29: Little noise\&As light as possible\=>s_5.13[±1.87]\[Support = 0.206; Strength = 0.700];
Rule 30: Little noise\=>c_67.5[±12.5]\[Support = 0.206; Strength = 0.700];
Rule 31: Alarmed by both vibration and sound\=>a_44.5[±10.5]\[Support = 0.206; Strength = 0.700];
Rule 32: Alarmed by both vibration and sound\=>d_4.5[±0.5]\[Support = 0.206; Strength = 0.700];
Rule 33: Alarmed by both vibration and sound\=>l_10.75[±1.25]\[Support = 0.206; Strength = 0.700];
Rule 34: Comfortable to hold\=>w_2.40[±0.6]\[Support = 0.206; Strength = 0.700];
Rule 35: Try latest fashion of mobile phones at a moderate price\&Alarmed by both vibration and sound\=>c_78.3[±21.7]\[Support = 0.206; Strength = 0.700];
Rule 36: Try latest fashion of mobile phones at a moderate price\=>d_5.5[±0.5]\[Support = 0.206; Strength = 0.700];
Rule 37: Try latest fashion of mobile phones at a moderate price\=>co_Alloy\[Support = 0.294; Strength = 0.833];

where β^F is a constant indicating the average dollar cost per variation of process capabilities, USL, LSL and σ are the upper specification limit, lower specification limit and standard deviation of part-worth cost estimates corresponding to individual FR clusters, respectively. The part-worth cost estimates are determined using a pragmatic approach based on standard time estimation [68].

To analyze the sensitivity of product portfolio identification, a total number of 17 runs of FR clustering are generated by changing λ value from 0.1 to 0.95 with an increment of 0.05. Using process data of vibration motors in Ref. [11] and flexibility valuation data of vibration motors in Ref. [70], the result of sensitivity analysis is obtained. As shown in Fig. 18, the performance measure in Eq. (21) is presented as a normalized comparison. The result clearly shows that a λ value of 0.84 yields the best performance of FR clustering for product portfolio identification.

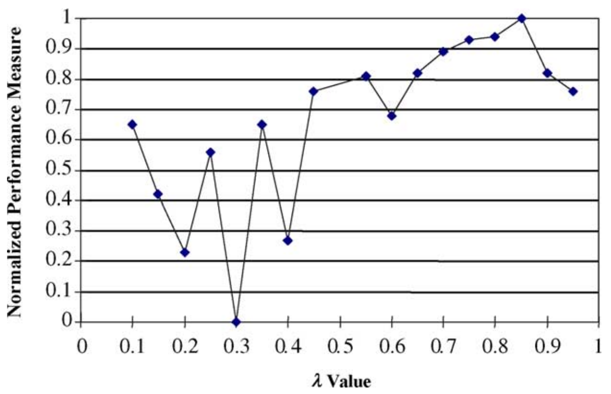


Fig. 18. Sensitivity analysis of product portfolio identification w.r.t. similarity threshold.

The difficulty in association rule mining originates from the need for determining appropriate thresholds for the support and confidence levels. If the support and confidence thresholds are planned with low values, useful information may be overwhelmed in excessive rules. On the contrary, certain relationship patterns that are of interest may be ignored if the support and confidence criteria are specified very strict.

Association rules basically suggest the mapping relationships between CNs and FRs. To meet the required CNs, the associated FRs must be fulfilled through configuration of DPs and PVs within the existing product and process platforms—a process of product variant derivation [59]. Such a variant derivation exhibits the accounting of a type of variable costs [12]. Jiao et al. [68] review the implications of customer-perceived value per unit cost in regard to the measure of profitability. Therefore, we introduce a performance measure of association rule mining, ψ^{AR}, based on the ratio of utility and the variable cost, as the following

$$\Psi^{AR} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{U_{ij}}{C_j^V}, \tag{23}$$

where the resulted product portfolio comprises $j = 1, \ldots, J$ products that are offered to meet a target market segment with $i = 1, \ldots, I$ customers, U_{ij} denotes the utility of the ith customer with respect to the jth product, and C_j^V is the related variable cost of producing this product variant. As suggested in Ref. [68], product level utilities, $\{U_{ij}\}_{I \cdot J}$, are derived from part-worth utilities of individual CNs based on conjoint analysis [32]. Likewise, product costs, $\{C_j^V\}_J$, are determined by the regression of part-worth cost estimates of individual FRs. The association rules indicate what FRs are to be used to satisfy what CNs. Such customer choice and product instantiation can be implemented by introducing binary variables to the part-worth regressions [68].

To analyze the sensitivity of association rule mining, a total number of $18 \times 18 = 324$ runs of ARMS are set up by enumerating all combinations of the min_sup and min_conf values, where both the min_sup and min_conf values are changed from 0.05 to 0.95 with an increment of 0.05. Using utility data of vibration motors in Ref. [70] and process data of vibration motors in Ref. [11], the result of sensitivity analysis is obtained. As shown in Fig. 19, the performance measure in Eq. (23) is presented as a normalized comparison. The result of sensitivity analysis suggests that the optimal criteria of association rule mining are given as the support and confidence thresholds of 0.5 and 0.6, respectively.

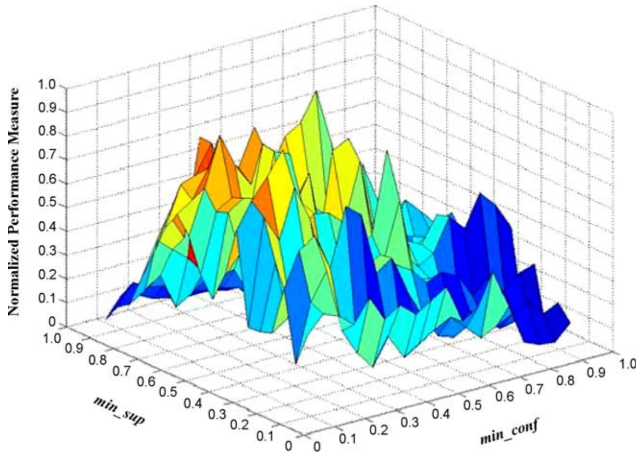


Fig. 19. Sensitivity analysis of product portfolio identification w.r.t. minimum support and confidence levels.

7. Discussions

As witnessed in the case study, it is profound to discern CNs from FRs in the respective customer and functional domains. Such a contextual difference in requirement information, as a matter of fact, constitutes the major tradeoffs inherent in the product definition process. While customers concern about the Alarm function of a mobile phone, designers have to interpret the implications of these CNs in terms of the functional specification of a vibration motor. During this process, engineering concerns play different roles in analyzing CNs and FRs. In accordance, product portfolio identification should seek for a synergy of these two sets of requirement information so as to achieve the desired 'dynamic' functional variety while keeping 'stability' in technical variety [59]. Therefore, the ARMS specifies a portfolio in terms of clusters of FRs while bearing correspondence to CNs. We believe this is more reasonable than most models in market research and requirement management, in which customer groups, market segments, or requirement patterns are all built upon the assumption that CNs and FRs connote the same semantic set of requirements. In this sense, the ARMS is more applicable to those consumer products than capital products (industrial products, e.g. power supplies). Consumer products usually involve more explicit interfaces between customers and engineering, whereas capital products involve less explicit customer involvement in engineering. In addition, knowledge recovery by data mining should be more useful for variant designs rather than new designs. Moreover, we advocate the importance of reusing knowledge from past data in order to deliver mass customization within the existing capabilities. In this regard, the portfolio identification has to conform to the product and process platforms that have been installed in the company. Hence, the specification of product offerings in a portfolio indeed represents the functional view of the product and process platforms.

In terms of requirement pattern recognition, association rule mining is advantageous over the tradition method based on decision trees. The key difference between the two techniques lies in that the decision tree method can only produce rules that are mutually exclusive, whilst association rule mining can produce rules that may not be mutually exclusive [71]. The reason behind this originates from the way they operate. Association rule mining seeks to go from the bottom up and collect all possible patterns that are of interest, and then use these patterns for some prediction targets. Decision trees, on the other hand, work from a prediction target downward in a manner known as 'greedy' search. They look for the best possible split at the next step. Furthermore, decision trees deal with data records that belong to the same category, whereas association rule mining can handle data records from different itemsets.

Nevertheless, the applicability of ARMS requires intensive collaboration with domain experts and considerations of particular problem contexts. Decisions on the proper similarity threshold and reasonable support and confidence levels may be too complex, and tricky as well, for enterprise managers. In practice, this can be alleviated through iterative interactions between portfolio identification and portfolio evaluation, as what we have done in the sensitivity analysis. Usually, a few scenarios with different settings of these parameters are identified and then input to the ARMS. Based on the running results, the performances of them are evaluated against a few pre-defined business objectives. Then the best setup is determined and the portfolio specification is refined. Hence, portfolio identification and its evaluation are iterative in implementation and thereby should be integrated within a unified framework of product portfolio planning.

While data mining techniques excels in identifying hidden patterns of mapping relationships between CNs and FRs, a practical data mining application is often complex, involving a number of interactive and iterative steps [60]. The processing of data throughout the data mining process deserves a particular attention for the achievement of good results. This is, however, often neglected and difficult to implement in practice. Pyle [72] provides a comprehensive coverage of existing data preparation techniques, including discretization, dimensionality reduction, normalization, etc. Treatment of missing values and data cleaning are important exercises for the implementation of data mining. The post-processing of discovered patterns is also important. This may involve interpreting association rules, analyzing the patterns automatically or semi-automatically, or identifying those truly interesting and useful patterns for the user. Also important is to extract target data sets from transaction records based on thorough understanding of the application domain and the application goals.

As for association rule mining, the support-confidence framework has been the subject of several criticisms. The confidence measure does not adequately capture the intuitive and natural semantics of direct associations, in which the associations are obvious [73]. To improve this, Brin et al. [74] propose an alternative measure, called conviction, to account for the strength of direct associations. In addition, the support-confidence framework tends to favor those rules with dense consequent. As a result, the rule generation process inclines to overstress those rules with a high consequent support. For instance, certain biased rules involving negated attributes are likely to appear in the outcome, making it contain many spurious rules [75]. Towards this end, a number of improvements have been proposed, including improvement-based rule pruning, collective strength, correlated attribute-set enumeration, intensity measure, and so on [73]. Moreover, traditional association rule mining adopts only a single minimum support in rule generation. However, classification data often contains a huge number of rules, which may cause combinatorial exploration. To tackle such an unbalanced data class distribution, Liu et al. [76] introduce the use of multiple class minimum supports to rule generation by assigning a different minimum support for each class. By incorporating appropriate measures into the association rule mining process, the quality of rules could be improved dramatically. For example, the Magnum Opus data mining tool employed in this study provides five instruments: coverage, support, strength, lift, and leverage. In the current mining process, we have only used two of them: support and strength. Conjoint use of all these five measures could improve the predictive accuracy of association rule mining substantially (http://www.rulequest.com/MOnew.html). However, the challenge lies in how to apply appropriate measures in accordance with the specific problem context of domain applications.

Another weakness inherent in data mining is that most rule induction methods perform a local, greedy search in the space of candidate rules. Intuitively, a global search can discover interesting rules and patterns that would be missed by the greedy search. Along this line, a number of efforts have been added into the research agenda. For example, evolutionary algorithms are applied to data mining to enable a robust search method that performs a global search in the solution space [77]. The amalgamation of statistical and data mining techniques has also attracted much attention [78].

8. Conclusions

This article presents a domain independent inference system for analyzing and organizing requirement information to support product portfolio identification. The methodology is based on the mining of association rules so as to provide an integration of requirement information from both customer and design viewpoints within a coherent framework.

Product portfolio identification entails a mapping process from customer needs in the customer domain to functional requirements in the functional domain. The specification of product offerings in a portfolio is embodied in a set of functional requirement clusters in conjunction with a set of associations of customer needs and the clusters. Each functional requirement cluster performs as a functional platform to satisfy a group of customers by enabling a certain range of variation with respect to a base value. For most variant product designs, where market segments have been established and product platforms have been installed, the association rule mining methodology can improve the efficiency and quality of portfolio identification by alleviating the tedious, ambiguous and error-prone process of requirement analysis enacted among customers, marketing folks, and designers. Generating the portfolio based on knowledge discovery from past data avails to maintain the integrity of existing product and process platforms, as well as the continuity of the infrastructure and core competencies, hence leveraging existing design and manufacturing investments. The application of data mining opens opportunities for incorporating experts' experiences into the projection of portfolio patterns from historical data, thereby enhancing the ability to explore and utilize domain knowledge more effectively.

Acknowledgements

This research is supported by Singapore NTU-Gintic Collaborative Research Project (U01-A-130B). The authors would like to express their sincere thanks to Mr Qingwen Sun, Engineer of the company, for his technical support to the case study. Professors Mitchell M. Tseng and Martin Helander, and colleagues of Global Manufacturing and Logistics Forum at Nanyang are appreciated for their valuable advices. Gratitude is also extended to the anonymous referees and the Editor for their constructive comments.

References

[1] Pine BJ. Mass customization: the new frontier in business competition. Boston: Harvard Business School Press; 1993.
[2] Ho TH, Tang CS. Product variety management. Research advances. Boston/Dordrecht/London: Kluwer Academic Publishers; 1998.
[3] Child P, Diederichs R, Sanders FH, Wisniowski S. SMR forum: The management of complexity. Sloan Manage Rev 1991;33(1): 73–80.
[4] Huffman C, Kahn B. Variety for sale: mass customization or mass confusion. J Retail 1998;74(4):491–513.
[5] Henderson BD. The product portfolio. Boston, MA: Boston Consulting Group; 1970.
[6] Warren AA. Optimal control of the product portfolio. PhD Thesis. The University of Texas at Austin; 1983. 335 pp.
[7] Cooper R, Edgett S, Kleinschmidt E. Portfolio management for new product development: results of an industry practices study. R & D Manage 2001;31(4):361–81.

[8] Li H, Azarm S. An approach for product line design selection under uncertainty and competition. Trans ASME. J Mech Des 2002;124(3): 385–92.
[9] Prasad B. Concurrent engineering fundamentals. New York: Prentice Hall PTR; 1996.
[10] Simpson TW. Product platform design and customization. Status and promise. AIEDAM. Special Issue on Platform Product Development for Mass Customization 2004;18(2).
[11] Jiao J, Zhang L, Pokharel S. Process platform planning for mass customization. Second Interdisciplinary World Congress on Mass Customization and Personalization. CD-ROM Proceedings, Munich, Germany; 2003.
[12] Meyer M, Lehnerd AP. The power of product platform-building value and cost leadership. New York: Free Press; 1997.
[13] Jiao J, Tseng MM, Ma Q, Zou Y. Generic bill of materials and operations for high-variety production management. Concurr Eng Res Appl 2000;8(4):297–322.
[14] Suh NP. Axiomatic design—advances and applications. New York: Oxford University Press; 2001.
[15] Jiao J, Tseng MM. A methodology of developing product family architecture for mass customization. J Intell Manuf 1999;10(1): 3–20.
[16] Jiao J, Tseng MM. Customizability analysis in design for mass customization. Comput-Aided Des 2004;36(8):745–57.
[17] Tseng MM, Jiao J. Computer-aided requirement management for product definition: A methodology and implementation. Concurr Eng Res Appl 1998;6(2):145–60.
[18] Tarasewich P, Nair SK. Designer-moderated product design. IEEE Trans Eng Manage 2001;48(2):175–88.
[19] Yan W, Chen CH, Khoo LP. An integrated approach to the elicitation of customer requirements for engineering design using picture sorts and fuzzy evaluation. AIEDAM 2002;16(2):59–71.
[20] Hauge PL, Stauffer LA. ELK. A method for eliciting knowledge from customers. Design and methodology, DE-vol. 53.; 1993. p. 73–81.
[21] Yan W, Chen CH, Khoo LP. A radial basis function neural network multicultural factors evaluation engine for product concept development. Expert Syst 2001;8(5):219–32.
[22] Shaw MLG, Gaines BR. Requirements acquisition. Software Eng J 1996;11(3):149–65.
[23] Christopher MC, McDonald M, Wills G. Introducing marketing. London: Pan; 1980.
[24] Du X, Jiao J, Tseng MM. Identifying customer need patterns for customization and personalization. Integr Manuf Syst 2003;14(5): 387–96.
[25] Prudhomme G, Zwolinski P, Brissaud D. Integrating into the design process the needs of those involved in the product life-cycle. J Eng Des 2003;14(3):333–53.
[26] Chen C-H, Khoo LP, Yan W. A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network. Adv Eng Inf 2002;16(3):229–40.
[27] Chen M, Han J, Yu P. Data mining. An overview from database perspective. IEEE Trans Knowledge Data Eng 1996;8(6):866–83.
[28] McKay A, de Pennington A, Baxter J. Requirements management: a representation scheme for product. Comput-Aided Des 2001;33(7): 511–20.
[29] Shoji S, Graham A, Walden D. A New American TQM. Portland: Productivity Press; 1993.
[30] Kano N, Seraku N, Takahashi F, Tsuji S. Attractive and must-be quality (in Japanese). Hinshitsu 1984;14(2):39–48.
[31] Jenkins, S., 1995. Modeling a perfect profile. Marketing. (July 13): 6, London.
[32] Green PE, DeSarbo WS. Additive decomposition of perceptions data via conjoint analysis. J Consumer Res 1978;5(1):58–65.
[33] Louviere J, Anderson D, White JB, Eagle TC. Predicting preferences for new product configurations. A high-tech example. Proceedings of the IFIP TC 7 Conference, Modeling the Innovation: Communications, Automation and Information Systems, Rome, Italy; 1990. p. 53–61.
[34] Turksen IB, Willson IA. Customer preferences models: fuzzy theory approach. Proceedings of the SPIE—International Society for Optical Engineering, Boston, MA; 1992. p. 203–11.
[35] LaChance-Porter S. Impact of user focus groups on the design of new products. Proceedings of the 14th National On-line Meeting, New York; 1993. p. 265–71.
[36] Griffin A, Hauser JR. The voice of the customer. Market Sci 1992; 12(1):1–27.
[37] Veryzer RW. Aesthetic response and the influence of design principles on product performance. In: McAllister L, Rothschele M, editors. Advances in customer research, vol. 20. Provo, UT: Association for Customer Research; 1993. p. 224–31.
[38] Louder D, Bitta AJD. Consumer behavior: concepts and applications. London: McGraw-Hill; 1988.
[39] Nagamachi M. Kansei engineering. Tokyo: Kaibundo Publisher; 1989.
[40] JSKE. Japan Society of Kansei engineering. http://www.jske.org/
[41] Kelly GA. The psychology of personal constructs. New York: W.W. Norton; 1955.
[42] Rugg G, McGeorge P. Laddering. Expert Syst 1995;12(4):279–91.
[43] Cortazzi D, Roote S. Illuminative incident analysis. London: McGraw-Hill; 1975.
[44] Mead M. Coming of age in Samoa. New York: William Morrow; 1928.
[45] Shaw MLG. Recent advances in personal construct technology. London: Academic Press; 1980.
[46] Maiden NAM, Rugg G. ACRE: selection methods for requirements acquisition. Software Eng J 1996;11(3):183–92.
[47] Chen C-H, Occeña LG. An expert system for wood head golf clubs design. Proceedings of the Fourth Industrial Engineering Research Conference. Nashville, TN, USA, IIE; 1995. p. 926–32.
[48] Chen C-H, Khoo LP, Yan W. An investigation to the elicitation of customer requirements using sorting techniques and fuzzy evaluation. Proceedings of the Sixth Asia Pacific Management Conference, Tainan, Taiwan; 2000. p. 45–55.
[49] Clausing D. Total quality development. A step-by-step guide to world class concurrent engineering. New York: ASME Press; 1994.
[50] Fung RTK, Popplewell K. The analysis of customer requirements for effective rationalization of product attributes in manufacturing. Proceedings of Third International Conference on Manufacturing Technology, Hong Kong; 1995. p. 287–96.
[51] Fung RYK, Popplewell K, Xie J. An intelligent hybrid system for customer requirements analysis and product attribute targets determination. Int J Prod Res 1998;36(1):13–34.
[52] Saaty T. The analytic hierarchy process. New York: McGraw-Hill; 1980.
[53] Fung RYK, Tang J, Tu Y, Wang D. Product design resources optimization using a non-linear fuzzy quality function deployment model. Int J Prod Res 2002;40(3):585–99.
[54] Fukuda S, Matsuura Y. Prioritizing the customer's requirements by AHP for concurrent design. Proceedings of Design for Manufacturability, DE-vol. 52.; 1993. p. 13–9.
[55] Byrne JG, Barlow T. Structured brainstorming. a method for collecting user requirements. Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society, Seattle, WA; 1993. p. 427–31.
[56] McAdams DA, Stone RB, Wood KL. Functional interdependence and product similarity based on customer needs. Res Eng Des 1999;11(1): 1–19.
[57] Hotelling HH. Stability in competition. Econ J 1929;39(1):47–51.
[58] Tseng MM, Jiao J. Design for mass customization. CIRP Ann 1996; 45(1):153–6.
[59] Du X, Jiao J, Tseng MM. Architecture of product family. Fundamentals and methodology. Concurr Eng: Res Appl 2001;9(4): 309–25.
[60] Han J, Kamber M. Data mining. Concepts and techniques. San Francisco: Morgan Kaufmann Publishers; 2001.

[61] Zimmermann HJ. Fuzzy set theory and its applications. Boston: Kluwer-Nijhoff Publishing; 1985.
[62] Lin CT, Lee CSG. Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems. NJ, USA: Prentice Hall; 1996.
[63] Wang H, McCauley-Bell P. Fuzzy clustering analysis and multi-factorial evaluation for students' imaginative power in physics problem solving. Fuzzy Sets Syst 1996;78(1):95–105.
[64] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in massive database. Proceedings of the ACM/SIGMOD International Conference on Management of Data; 1993. p. 207–16.
[65] Yang L, Gao Y. Fuzzy mathematics: theory and applications. 1996. ISBN. 7-5623-0440-8.
[66] Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. Proceedings of 20th International Conference on Very Large Data Bases, Santiago de Chile, Chile; 1994. p. 487–99.
[67] ChangChien SW, Lu TC. Mining association rules procedure to support on-line recommendation by customers and products fragmentation. Expert Syst Appl 2001;20(4):325–35.
[68] Jiao J, Zhang Y, Wang Y. Product portfolio planning considering customer-engineering interaction: problem formulation. The Seventh International Conference on Work with Computing Systems, Kuala Lumpur, Malaysia; 2004.
[69] Gonzalez-Zugasti JP, Otto KN, Baker JD. Assessing value for platformed product family design. Res Eng Des 2001;13(1):30–41.
[70] Jiao J, Kumar A, Lim CM. Flexibility study of product and process platforms. A real-option approach. The 11th ISPE International Conference on Concurrent Engineering, Beijing, China; 2004.
[71] Berson A, Smith S, Thearling K. Building data mining applications for CRM. New York: McGraw-Hill; 1999.
[72] Pyle D. Data preparation for data mining. San Francisco: Morgan Kaufmann; 1999.
[73] Adamo J-M. Data mining for association rules and sequential patterns. Sequential and parallel algorithms. New York: Springer; 2001.
[74] Brin S, Motwani R, Ullman JD, Tsur S. Dynamic itemset counting and implication rules for market basket data. Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada; 1997. p. 255–64.
[75] Aggarwal CC, Yu PS. A new framework for itemset generation. Proceedings of the 17th ACM SIGACT–SIGMOD–SIGART Symposium on Principles of Database Systems, Seattle, WA; 1998. p. 18–24.
[76] Liu B, Hsu W, Ma Y. Integrating classification and association rule mining. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York; 1998.
[77] Freitas AA. Data mining and knowledge discovery with evolutionary algorithms. New York: Springer; 2002.
[78] Mani DR, Drew J, Betz A, Datta P. Amalgamation of statistics and data mining techniques: explorations in customer lifetime value modeling. In: Abramowicz W, Zurada J, editors. Knowledge discovery for business information systems. Boston: Kluwer Academic Publishers; 2001. p. 229–50.

Jianxin Jiao is Assistant Professor of Systems and Engineering Management, School of Mechanical and Production Engineering, Nanyang Technological University, Singapore. He is convener and coordinator of Global Manufacturing & Logistics Forum at Nanyang. He received a PhD from Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology. He holds a Bachelor degree in Mechanical Engineering from Tianjin University of Science & Technology in China, and a Master degree in Mechanical Engineering from Tianjin University in China. He has worked as a Lecturer in Department of Management at Tianjin University. His research interests include mass customization, design theory and methodology, reconfigurable manufacturing systems, engineering logistics, and intelligent systems.

Yiyang Zhang is a PhD candidate in School of Mechanical and Production Engineering at Nanyang Technological University, Singapore. She received her BBA and MBA degrees from School of Management at Northeastern University, China, in 1999 and 2002, respectively. Her current research interests are customer decision-making process, product portfolio planning and customer requirement management.