# An Association Rule-Based Inference System for Customer Need Recognition

Yiyang Zhang, Jianxin Jiao
School of Mechanical and Production Engineering
Nanyang Technological University
Singapore
zhang_yi_yang@hotmail.com, jiao@pmail.ntu.edu.sg

## Abstract

Facing intensive competition in the marketplace, manufacturing companies are forced to react to the growing individualization of demands. To meet this challenge, more and more companies have realized the importance of understanding the voice of customers and recognizing customer needs. Customer needs are inherently implicit and ambiguous, resulting in barriers for customers, marketers and the designers to communicate one another and cooperate throughout the product development process. This paper studies the mapping mechanism between the customer domain and the functional domain of design. An Association Rule-Based Inference System (ARIS) is presented for the recognition of customer needs.

## 1. Introduction

Manufacturing companies are facing the intensive competition in the marketplace. Under such a competitive environment, more and more companies shift to the customer needs. Product definition has been considered as one of the key factors in the early stage of the product design. The success or failure of the product definition directly influences the whole product development.

Based on the Axiomatic Design theory [18], the products are specified by several functional requirements $(FRs)$; and for each $FR$, there are different value options $(FRVs)$. The class-member relationships for variety representation theory [12] has illuminated that for each type of product alternative (class) with a specific set of $FRs$ (class attributes), product variety can further result from different $FRVs$ (members). The class-member relationship for variety representation is shown as Fig. 1.

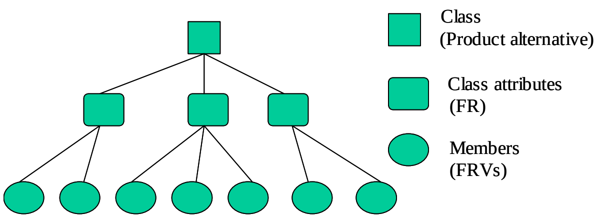


**Fig. 1. Class-member relationship for variety representation**

Axiomatic Design theory also indicates that the design world is composed of four distinct domains. The customer needs and product specifications are stated in $CN$ domain (customer needs domain) and functional domain [18] respectively. The customer domain and the functional domain comprise the product definition phase, from the abstract customer needs to the concrete product requirement specifications [15]. Under the philosophy of "design for customers", the product specifications $(FRs)$ should be evolved from the real needs of customers $(CNs)$.

There are many difficulties when translate the genuine customer needs $(CNs)$ into $FRs$ because $CNs$ are always originated in a verbatim format and possess the non-technical nature. On the contrary, $FRs$ are always expressed with the explicit technical parameters which can be easily understood by the design engineers. Traditionally, design engineers use their experience and intuition to translate imprecise customer requirements into a clearly defined design specification, but this is a less than ideal approach, as it leads to imperfect designs, which satisfy a design engineer's interpretation of customer requirements, rather than satisfying the genuine customer requirements [1].

Data mining is a rapidly evolving area of data analysis that attempts to identify trends or unexpected pattern in large pools of data by using mathematical tools. It could help extract useful information from large database. Association rule mining is a data mining technique that discovers strong associations or correlation relationships among data without constraining the data type and is very suited to mine the mapping relationships between the customers' needs with the linguistic expression and functional requirements specified by the technical parameters.

This paper presents an association rule-based inference system to provide the mapping mechanism between customer domain and functional domain. Section 2 describes related research work in the field of requirement management. In section 3, an association rule-based inference methodology is represented. Section 4 represents a case study and section 5 summarizes the research work in this paper.

## 2. Literature review

The voice of customers $(VoC)$ has received amount of interest of both academicians and practitioners [8]. The traditional method to translate the customer needs into the corresponding functional requirement is Quality Function Deployment ($QFD$) [3] [7] [12]. Though $QFD$ can help improve the product quality by providing the information about the customers [16], it is a document with a list of customer needs and functional requirements rather than a mapping tool to recognize the relationship pattern between the two.

From a design perspective, Hauge and Stauffer develop a taxonomy method of product requirements as a way to organize information [7]. Though the taxonomy method does recognize the topology of the functional requirements by classifying large bodies of information, it is too general to be a domain-independent framework. Also, because the taxonomy method starts from the engineering design perspective, it can't provide the pattern information about the customer domain and the functional domain.

In recent years, several researchers have gained achievement in the field of requirement management. Fung establishes an intelligent inference model in which the customer needs, the product specifications are both expressed in Fuzzy Sets characterized by their relevant membership functions, and the relationship between them is represented in the form of Fuzzy Propositions [5]. Though Fuzzy Set Theory excels in treating the notions that can't be clearly defined, it is not good at solving the quantitative variables because the variables must be fuzzified before they enter the inference process. It is not appropriate to translate the quantitative variables into fuzzy values.

Tseng and Jiao adopt Computer-Aided Requirement Management methodology for product definition noted as $PDFR$ [19]. This method adopts functional requirement ($FR$) patterns from previous product designs to map $FR$ into desired value ( $FR$ instance). The values of $FR$ variables are characterized by the representative center vectors. Their work focus on the FR domain, but the mapping between the customer domain and functional domain is still not recognized. A mapping mechanism to recognize the relationship between the customers' needs and product functional specifications is needed to be established to help companies provide the products to meet the customers' needs.

## 3. Mapping mechanism between customer needs and functional requirements

### 3.1 Product definition

Supposing $P=\{FRV_{jz_j}\}(j=1,2...m; z_j=1,2...n_j)$ as the generic representation of a product definition, where the product definition ($P$) is embodied by $FRV_{jz_j}$ ( $j = 1,2...m$; $z_j=1,2...n_j$) and it refers to the $j^{th}$ $FR$ taken with the $z_j{}^{th}$ value; $j$ refers to the numbers of $FRs$ that the product alternatives consist of; and $n_j$ refers to the numbers of $FRVs$ for each $FR_j$.

The recognition of the mapping mechanism between $CNs$ and $FRs$ to help develop the product definition then is identical to the mapping recognition problem between $CNs$ and $FRVs$.

This paper proposes an ARIS system to establish the mapping mechanism between $CNs$ and $FRVs$ . A vivid model of customers' needs recognition is shown as Fig. 2. The $FRVs$ are clustered and then the mapping mechanism is established between the customers' needs and $FRVs$. The definitions of products are described by the combinations of $FRVs$.

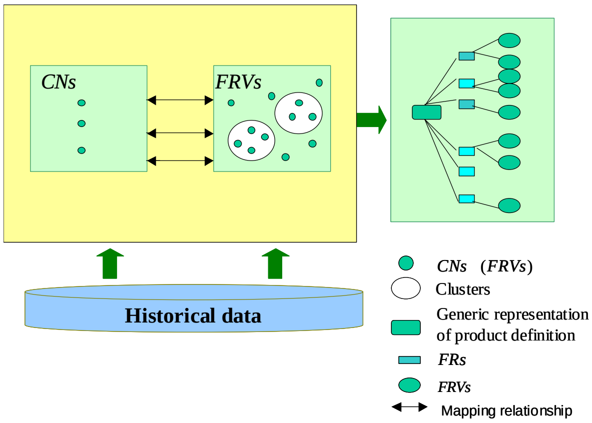


**Fig.2. A vivid model of customers' needs recognition**

### 3.2 ARIS architecture

The ARIS system consists of three parts: the user interface, the knowledge base and the inference engine. The knowledge base captures the data of $CNs$ and $FRVs$. It can be partitioned into two parts:

1. $K-CN$ : The knowledge about customer needs;
2. $K-FRV$ : The knowledge about functional requirement values.

The architecture of ARIS is showed as Fig. 3.

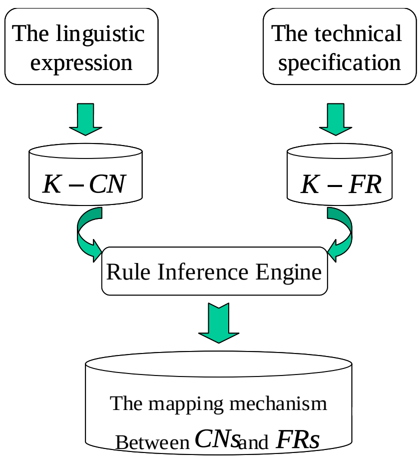


**Fig. 3. Illustration of the ARIS architecture**

The relational database can be established, in which $CNs$ and $FRVs$ are stored in two tables which are related by means of a common file. In this system, the common file is the customer ID.

*CNs* are represented as $CN_{ip_i}(\forall i=1,2...n;\ p_i=1,2...k_i)$, which denotes the $i^{th}$ customer's the ${p_i}^{th}$ needs, where $i$ refers to the numbers of the customers and $k_i$ refers to the numbers of the needs of the $i^{th}$ customer. Similarly, $FRV_{ijz_j}$ denotes the ${z_j}^{th}$ value of the $j^{th}$ $FR$ $(\forall j=1,2...m;$ $z_j=1,2...n_j)$, and these are what the designers design in order to meet the requirement of the $i^{th}$ customer, where $j$ refers to the numbers of $FRs$, and $n_j$ refers to the numbers of the $FRVs$ candidates for the $j^{th}$ $FR$.

The schematic representation of ARIS can be shown as Fig. 4.

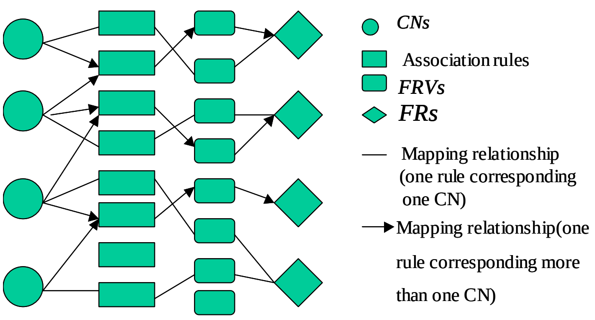


**Fig. 4. The schematic representation of ARIS**

## 3.3 Methodology

The methodology to establish the mapping pattern between *CNs* and *FRVs* by using association rules will be illustrated as following. The whole process involves two steps: data clustering and mining association rules.

### 3.3.1 Data clustering

Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Different customers desire various $FRVs$. For almost all the manufacturing companies, it is impossible to meet every individual need because of the high cost. Based on the Chamberlin-Dixit-Stiglitz $(CDS)$ Model [13], the products close together on the spectrum are better substitutes than those further apart. That is, the customers are willing to choose those products that have *FRVs* which are closest to their desired values if they can't find the products with their desired values are on the markets. It is wise to cluster the customers' desired values and provide the product family not only to meet the needs of customers but also reduce the cost and design time.

1. Measuring Distance

There are three types of the *FRVs* variables: the numerical variables, the binary variables and the nominal variables [9] [10]. The methods to measure the distances of different types of variables are different.

- Numerical variables distance measure

The distance measures of the numerical variables are not single. In general, the clustering methods for numerical variables include Eluclidean distance, Manhattan distance, Minkowski distance and weighted Eluclidean distance measure. In this paper, the weighted Eluclidean distance measure is adopted:

$$d(i,v)_{NUMERICAL} = \sqrt{\sum_{j,z_j} W_j (FRV_{ijz_j} - FRV_{vjz_j})^2} \qquad (1)$$

subject to

$$FRV_{i(v)jz_j} = \frac{FRV_{i(v)jz_j} - \min FRV_{i(v)jz_j}}{\max FRV_{i(v)jz_j} - \min FRV_{i(v)jz_j}}, \qquad (2)$$

where $d(i,v)_{NUMERICAL}$ : the distance between customer $i$ and customer $v$ based on the numerical variables;

$FRV_{ijz_j}$ : the standardized ${z_j}^{th}$ value for the $j^{th}$ $FR$ $(\forall j=1,2...m; z_j=1,2...n_j)$ held by the $i^{th}$ customer;

$W_j$ : the relative importance of the $j^{th} FR$.

Equation (2) aims at normalize the $FRV$ variables to eliminate the effect of the unit difference.

- Binary variables distance measure

A binary variable has only two states: 0 or 1, where 0 means that the variable is absent and 1 means that it is present. The binary variables can be grouped into two types: the symmetric and asymmetric binary variables. A binary variable is symmetric if both of its states are equally valuable and carry the same weight. The most well-known coefficient for assessing the distance measure of symmetric binary variables is the simple matching coefficient. A binary variable is asymmetric if the outcomes of the states are not equally important. For the distance measure of the asymmetric binary variables, the most well-known coefficient is the Jaccard coefficient. In this research, the binary variables in *FRVs* are treated as symmetric variables and thus a simple matching coefficient is adopted:

$$d(i,v)_{BINARY}=(r+s)/(q+r+s+t) \qquad (3)$$

where $d(i,v)_{BINARY}$ : the distance between customer $i$ and customer $v$ based on the binary variables;

$q$ : the total numbers of binary variables that equal to 1 for both customer $i$ and $v$;

$r$ : the numbers of variables that equal to 1 for customer $i$ but equal to 0 for customer $v$;

$s$ : the numbers of variables that equal to 0 for customer $i$ but equal to 1 for customer $v$; and

$t$ : the numbers of variables that equal to 0 for both customer $i$ and $v$.

- Nominal variable distance measure

A nominal variable is a generalization of the binary variables in that it can take on more than two states. A discrete ordinal variable resembles a nominal variable, except that the states of the ordinal value are ordered in a meaning sequence. In this paper, the ordering sequence doesn't exist, so the simple matching approach is employed to measure the distance of the nominal variables:

$$d(i,v)_{NOMINAL} = (p-m)/p \quad (4)$$

where $d(i,v)_{NOMINAL}$ : the distance between customer $i$ and customer $v$ based on the nominal variables;

$m$ : the numbers of variables for customer $i$ and $v$ that are in the same state; and

$p$ : the total numbers of variables.

2. Fuzzy Clustering Analysis

- Forming a fuzzy similar matrix

$$z(i,v)=\{d(i,v)-\min d(i,v)\}/\{\max d(i,v)-\min d(i,v)\} \quad (5)$$

$$d(i,v)' = W_{NUMERICAL}\, z(i,v)_{NUMERICAL} + W_{BINARY}\, z(i,v)_{BINARY} + W_{NOMINAL} z(i,v)_{NOMINAL} \quad (6)$$

$$R=[r(i,v)]_{n*n}(i,v=1,2...n) \quad (7)$$

$$r(i,v)=1-d(i,v)' \quad (8)$$

where $R=[r(i,v)_{n*n}](i,v=1,2...n)$ : the similarity matrix;

$r(i,v)$ : the degree of similarity between customer $i$ and customer $v$;

$W_{NUMERICAL}$ : the relative importance of the numerical values;

$W_{BINARY}$ : the relative importance of the binary values;

$W_{NOMINAL}$ : the relative importance of the nominal values; and

$d(i,v)'$ : the dissimilarity between customer $i$ and customer $v$ based on all the types of variables.

Equation (5) is used to normalize the $FRV$ variables because the values in the similarity matrix are within the interval 0 and 1and it is applicable to each type of $FRVs$. Equation (6) calculates the distance between customer $i$ and customer $v$ with the consideration of all types of $FRVs$.

- Getting the fuzzy equivalent matrix

Equivalent matrix has the following characteristics:

a. Reflexive: $r_{iv}=1(i,v=1,2...n)$;

b. Symmetrical: $r_{iv}=r_{vi}(i,v=1,2...n)$; and

c. Transitive: $R\circ R=R$.

The similarity matrix generally meets the first two requests. To meet the third, convert it into an equivalent matrix by using the "continuous multiplication" method.

Multiplication in fuzzy relations is also known as the maxmin composition. Let $R(i,v),(i,v)I*V$ and $R(i,v),(i,v)I*V$ be two fuzzy relations, then $R\circ R=\{[(i,v),\max\{\min r_R(i,v),r_R(i,v)\}]/i,I,v,V\}$ is also a fuzzy relation where $r_R$ is an element in a fuzzy relation.

- Choosing a threshold and perform clustering

After the equivalent matrix has been obtained, a threshold is chosen noted as $l_n$ above which the $r(i,v)$ values validate that the customer $i$ and customer $v$ hold the similarity and can be classified into the same cluster. Based on the equivalent matrix, given different thresholds, different clustering results can be obtained. The clustering graph is shown as Fig. 5.

Denoting $l$ as the $l^{th}$ cluster $(l=1,2...s)$ where $l[(i,v)\setminus i,v,(r_{iv}>l_n)](i,v=1,2...n)$, then we have:

$$FRV_{jl} = \sum_i FRV_{ijl}/N_l \quad (9)$$

where $N_l$ : the customers' numbers that fall into $l$; and

$FRV_{ijl}$ : the $FRV$ for the $j^{th}FR$ that falls into the $l^{th}$ cluster desired by customer $i$.

Equation (9) calculates the average values of $j^{th}FR$ desired by those customers falling into the same cluster. And $FRV_{ijl}$ are then stored into the database [2].

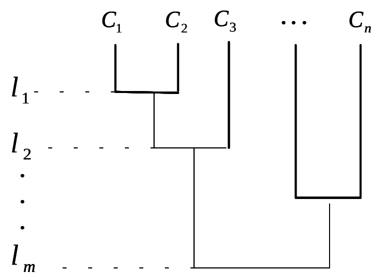


**Fig.5. The clustering graph**

### 3.3.2 Mining association rules

1. Transaction and Itemsets in ARIS

The itemsets originally refer to the itemsets that are purchased by the customers in a transaction. As for the frequent itemsets, it means those itemsets whose counts exceed the min_sup [4] [17].

In the ARIS system, a transaction is denoted as $L_i$ which refers to the complete collection of $CN_{ip_i}$ and $FRV_{jl}{}^i$ for the $i^{th}$ customer $(i=1,2...n)$, and the itemsets are those $CN_{ip_i}$ and $FRV_{jl}{}^i$ contained in $L_i$ where $FRV_{jl}{}^i$ represents $FRV_{ijz_j}$ falling into $FRV_{jl}$. The structure of the transactional database is shown as Tab. 1.

**Tab. 1. Transactional data structure for ARIS**

| Transaction No. | Itemsets |
|---|---|
| $L_1$ | $FRV_{jl}{}^1$ $CN_{1p_1}$ |
| $L_2$ | $FRV_{jl}{}^2$ $CN_{2p_2}$ |
| . | . |
| . | . |
| . | . |
| $L_n$ | $FRV_{jl}{}^n$ $CN_{np_n}$ |

2. Finding the Frequent Itemsets

Finding the frequent itemsets is an iteration process. The k-frequent-itemset is denoted as $A_k$ and the candidate k-frequent-itemsets that is generated by joining $A_{k-1}$ with itself is denoted as $C_k$. The join, $A_{k-1}\cap A_{k-1}$ is performed, where members of $A_{k-1}$ are joinable if their first (k-2) items are in common. Denoting the notation $l_i(j)$ as the

$j^{th}$ item in $l_i$ where $l_i$ refers to the $i^{th}$ member in $A_{k-1}$, then members $l_1$ and $l_2$ of $A_{k-1}$ are joined if

$$(l_1[1] = l_2[2]) \wedge (l_1[2] = l_2[2]) \wedge ... \wedge (l_1[k-2] = l_2[k-2]) \wedge (l_1[k-1] \neq l_2[k-1])$$

In the first iteration, each item is a member of the set of candidate 1-itemsets, $C_1$. Then the frequent itemset $A_1$ can be determined by scanning all the $C_1$ and selecting those items that reach the min_sup. The second candidate $C_2$ then is evolved by joining $A_1$ with $A_1$ and thus $A_2$ is determined by scanning all the $C_2$ and selecting those items that reach the min_sup. The iteration terminates until $C_{k+1} = \Phi$ and $A_k$ is determined.

3. Generating Association Rules from Frequent Itemsets

Once the frequent itemsets from the database have been found, it is straightforward to generate strong association rules from them [14] [20].

- The format of association rule

Association rules are displayed by using the following format:

$CN_{ip_i} -> FRV_{jl}$ [Coverage=value; Support=value; Strength=value; Lift=value; Leverage=value]

The left-hand-side of this rule - $CN_{ip_i}$ is presented before the -> arrow. The right-hand-side- $FRV_{jl}$ follows the arrow and precedes the opening bracket. It indicates that item $CN_{ip_i}$ associates with item $FRV_{jl}$. That is, if item $CN_{ip_i}$ exists, item $FRV_{jl}$ also occurs.

- Criteria

Only when the criteria have been met, it validates there are strong associations between the items. There are five criteria are needed to be met during the association rule inference process. They are shown as the following.

1. Coverage

Coverage is the proportion of transactions covered by $CN_{ip_i}$:

$$\text{Coverage} = P(CN_{ip_i}) \tag{10}$$

2. Support

Support is the proportion of transactions covered by both the $CN_{ip_i}$ and the $FRV_{jl}$:

$$\text{Support} = P(CN_{ip_i} \cap FRV_{jl}) \tag{11}$$

3. Strength

Strength is the proportion of transactions covered by the $CN_{ip_i}$ that are also covered by the $FRV_{jl}$:

$$\text{Strength} = P(FRV_{jl} / CN_{ip_i}) \tag{12}$$

4. Lift

Lift is the strength divided by the proportion of all transactions that are covered by the $FRV_{jl}$. Lift provides an indication of how much more likely is the right hand given the left hand than normal:

$$\text{Lift} = P(FRV_{jl}) \setminus P(FRV_{jl} / CN_{ip_i}) \tag{13}$$

5. Leverage

Leverage is the proportion of additional transactions covered by both the $CN_{ip_i}$ and $FRV_{jl}$ above those expected if the $CN_{ip_i}$ and $FRV_{jl}$ were independent of each other:

$$\text{Leverage} = P(FRV_{jl} \cap CN_{ip_i}) - P(FRV_{jl}) \times P(CN_{ip_i}) \tag{14}$$

# 4. Case study

## 4.1 Background

So far, the ARIS system representation, the mathematical formulation and the algorithm have been discussed. This section presents one case that is greatly related to the above aspects to present the application of the proposed methodology. The case adopts the high variety product – the vibrating motor used as the absolutely necessary accessory of mobilephones - as an example. This is because (1) The motor products are not complex for the customers and it is easy to understand the customers' requirements for the motor products expressed in nature language; and (2) The motor products are standardized customization products which can reflect the customers' different individual needs so as to reveal the nature of ARIS system.

Since different customers have their own requirements for the mobilephone products, the motor products matched with these different mobilephones are typical customized products under mass customization. The difference in motor product design requirements results in the products characterized by a number of variations including the customers' requirements which expressed in their nature language and the functional requirement values.

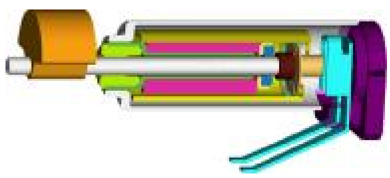

**Fig.6. Illustration of a motor product**

The motor product is a collection of nine $FRs$. These nine $FRs$ are current, pbfree, length, diameter, coating, angle, strength, weight and hardness respectively. The illustration of the real motor product is shown in Fig. 6. The description of $FR_j$ is shown in Tab. 2.

**Tab.2. Description of FR$_j$**

| $FR_i$ | Description |
|---|---|
| $FR_1$ | Current |
| $FR_2$ | Pbfree |
| $FR_3$ | Length |
| $FR_4$ | Diameter |
| $FR_5$ | Coating |
| $FR_6$ | Angle |
| $FR_7$ | Strength |
| $FR_8$ | Weight |
| $FR_9$ | Hardness |

**Tab. 3(a) Description of FRVijz$_j$**

| $FR_j$ | $FR_j$ Description | $FRV_{ijz_j}$ (Encoded with $jz_j$) | $FRV_{ijz_j}$ Description |
|---|---|---|---|
| $FR_1$ | Current | 11 | 100 |
| | | 12 | 80 |
| | | 13 | 60 |
| $FR_2$ | Pbfree | 21 | Y |
| | | 22 | N |
| $FR_3$ | Length | 31 | 8 |
| | | 32 | 12 |
| | | 33 | 10 |
| $FR_4$ | Diameter | 41 | 5 |
| | | 42 | 4 |
| | | 43 | 6 |
| $FR_5$ | Coating | 51 | Au |
| | | 52 | Alloy |
| | | 53 | None |
| $FR_6$ | Angle | 61 | 40 |
| | | 62 | 55 |
| $FR_7$ | Strength | 71 | 7 |
| | | 72 | 4 |
| $FR_8$ | Weight | 81 | 2 |
| | | 82 | 3 |
| $FR_9$ | Hardness | 91 | 40 |
| | | 92 | 70 |

**Tab. 3(b) Description of CNs**

| CN No. | CN Description |
|---|---|
| 1 | Feel the vibration very strongly |
| 2 | Alarmed by the vibration without vibrating suddenly |
| 3 | Sensitive to the vibration |
| 4 | Buy the expensive mobilphone to use for a long time |
| 5 | Change the style of the mobilephone occasionally at a lower price |
| 6 | Try other style of mobilephones every now and then with a moderate price |
| 7 | Portable |
| 8 | Comfortable to hold |
| 9 | Not easy to lose |
| 10 | Little noise |
| 11 | Alarmed when can't depend on vibration |
| 12 | Alarmed by both vibration and sound |
| 13 | Green environment |
| 14 | Light |

## 4.2 Applying ARIS for the mapping mechanism

### 4.2.1 The data list about CN and FRV

It has been mentioned above that the motor product is a collection of nine $FRs$. Among these $FRs$, $FR$-pbfree is of binary type, $FR$-coating is of nominal type and the remaining seven $FRs$ are of numerical types. There are 30 customers' needs derived from the company's database.

It is not surprising that there are different requirements for each $FR$. The historical data are composed of both customers' requirements expressed in their language and the corresponding $FR$ values ($FRV_{ijz_j}$) which meet their requirements. The requirements and values bound with each customer differ from each other and are stored in the database identified by an identifier – the customer No. The description of the customer requirements ($CN$) and $FRV$ are partly shown as Tab. 3.

### 4.2.2 Weight determination

Analytical Hierarchy Process ($AHP$) is used to calculate the relative importance of each $FR$. The pairwise comparison scale is shown as Tab. 4. The weights ($W_j$) for $FRs$ are shown as Tab. 5.

**Tab. 4 The scale for pairwise comparison**

| Verbal judgement 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 |

**Tab. 5 The preferences for FR$_j$**

| $FR_j$ | $W_j$ |
|---|---|
| Current | 0.219 |
| Pbfree | 0.304 |
| Length | 0.046 |
| Diameter | 0.031 |
| Coating | 0.019 |
| Angle | 0.066 |
| Strength | 0.157 |
| Weight | 0.095 |
| Hardness | 0.083 |

### 4.2.3 Customer clustering

1. Distance Measurement

- The numerical FRs distance measurement

There are seven numerical $FRs$ including current, length, diameter, angle, strength, weight and hardness. The first three $FRs$ have three levels of $FRVs$, the last four $FRs$ have two levels of $FRVs$. That is:

$FR_1 = \{FRV_{11}, FRV_{12}, FRV_{13}\}$,

$FR_2 = \{FRV_{21}, FRV_{22}, FRV_{23}\}$'

...

$FR_7\{FRV_{71}, FRV_{72}\}$.

The weighted Eluclidean distance measure is adopted. The original data are input into SPSS software to be processed during which the original data are normalized automatically by computer and then the results are calculated.

The distance between customer $i$ and customer $v$, i.e., $d(i,v)_{NUMEROCAL}$, $(\forall i,v = 1,2...30)$, is presented as a $30 \times 30$ matrix whose elements represent the distance between two customers measured based on numerical variables.
The result is shown as Tab. 6.

- The binary FRs distance measure

There is only one binary $FR$ namely pbfree function. It has only two states: 0 or 1, where 0 means that pbfree function is required and 1 means that it is not required. The simple matching coefficient is adopted. The distance between customer $i$ and customer $v$, $d(i,v)_{BINARY}, (\forall i,v = 1,2...30)$, is presented as a 30×30 matrix and the result is shown as Tab. 7.

**Tab. 6 Result of numerical variables distance measure**

$$\begin{bmatrix} 0 & 0.076 & 0.534 & \cdots & 0.273 & 0.296 & 0.252 \\ 0.076 & 0 & 0.557 & \cdots & 0.197 & 0.235 & 0.191 \\ 0.534 & 0.557 & 0 & \cdots & 0.36 & 0.238 & 0.282 \\ \vdots & & & \ddots & & & \vdots \\ 0.273 & 0.197 & 0.36 & & 0 & 0.114 & 0.313 \\ 0.296 & 0.235 & 0.238 & & 0.114 & 0 & 0.274 \\ 0.252 & 0.191 & 0.282 & & 0.313 & 0.274 & 0 \end{bmatrix}_{30\times30}$$

**Tab. 7 The result of binary variables distance measure**

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}_{30\times30}$$

- The nominal FRs distance measure

There is only one nominal *FR* namely coating. It has three states: Au, Alloy and None. The simple matching approach is adopted to measure the distance of the nominal variables. The distance between customer $i$ and customer $v$, $d(i,v)_{NOMINAL}, (\forall i,v = 1,2...30)$, is presented as a 30×30 matrix and the result is shown as Tab. 8.

**Tab. 8 The result of nominal variables distance measure**

$$\begin{bmatrix} 0 & 0 & 1 & \cdots & 1 & 0 & 1 \\ 0 & 0 & 1 & \cdots & 1 & 0 & 1 \\ 1 & 1 & 0 & \cdots & 1 & 0 & 1 \\ \vdots & & & \ddots & & & \vdots \\ 1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 1 & & 0 & 0 & 1 \\ 1 & 1 & 0 & & 0 & 1 & 0 \end{bmatrix}_{30\times30}$$

2. Fuzzy Clustering

The similarity matrix is represented as $R$. The elements in $R$ represent the similarity between every two customers. The result of R is shown as Tab. 9 and the equivalent matrix $R^2$ is shown as Tab. 10.

Because $R^2$ doesn't equal to $R$, the $R^2$ is not fuzzy equivalent matrix. The algorithm is required to iterate again. The result of $R^4$ is shown as Tab. 11.

As $R^4 = R$, $R^4$ is fuzzy equivalent matrix. Setting the threshold $l_n$ as 0.84, the clustering graph is derived. It is shown as Fig. 7. Based on the clustering graph, the customers are clustered into three clusters. That is:

$l_1 = \{1,2,7,8,11,12,14,15,24,29\}$,

$l_2 = \{3,4,5,9,10,13,17,19,20,23,26,30\}$,

$l_3 = \{6,16,18,21,22,25,27,28\}$.

Then the $FRV_{jl}\,(j = 1,2...9; l = 1,2,3)$ are shown in Tab. 12.

**Tab. 9 The result of R**

$$\begin{bmatrix} 1 & 0.924 & 0.446 & \cdots & 0.71 & 0.7 & 0.73 \\ 0.924 & 1 & 0.42 & \cdots & 0.78 & 0.76 & 0.79 \\ 0.446 & 0.42 & 1 & \cdots & 0.64 & 0.74 & 0.72 \\ \vdots & & & \ddots & & & \vdots \\ 0.71 & 0.78 & 0.64 & & 1 & 0.87 & 0.69 \\ 0.7 & 0.76 & 0.74 & & 0.87 & 1 & 0.71 \\ 0.73 & 0.79 & 0.72 & & 0.69 & 0.71 & 1 \end{bmatrix}_{30\times30}$$

**Tab. 10 The result of R²**

$$\begin{bmatrix} 1 & 0.7 & 0.64 & \cdots & 0.78 & 0.84 & 0.83 \\ 0.7 & 1 & 0.74 & & 0.83 & 0.76 & 0.82 \\ 0.64 & 0.74 & 1 & & 0.8 & 0.75 & 0.85 \\ \vdots & & & \ddots & & & \vdots \\ 0.78 & 0.83 & 0.8 & & 1 & 0.86 & 0.78 \\ 0.84 & 0.76 & 0.75 & & 0.86 & 1 & 0.76 \\ 0.83 & 0.82 & 0.85 & & 0.78 & 0.76 & 1 \end{bmatrix}_{30\times30}$$

**Tab. 11 The result of R⁴**

$$\begin{bmatrix} 1 & 0.924 & 0.446 & \cdots & 0.71 & 0.7 & 0.73 \\ 0.924 & 1 & 0.42 & \cdots & 0.78 & 0.76 & 0.79 \\ 0.446 & 0.42 & 1 & \cdots & 0.64 & 0.74 & 0.72 \\ \vdots & & & \ddots & & & \vdots \\ 0.71 & 0.78 & 0.64 & & 1 & 0.87 & 0.69 \\ 0.7 & 0.76 & 0.74 & & 0.87 & 1 & 0.71 \\ 0.73 & 0.79 & 0.72 & & 0.69 & 0.71 & 1 \end{bmatrix}_{30\times30}$$

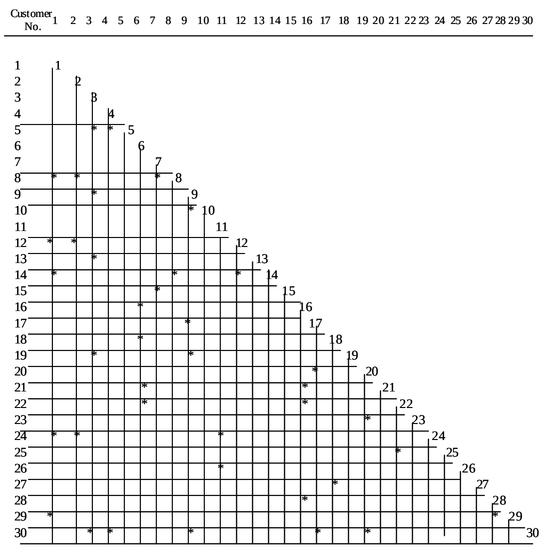


**Fig.7. Clustering graph**

**Tab. 12 List of $FR_j$ and $FRV_{jl}$**

| $FR_j$ | $FR_j$ Description | $FRV_{jl}$ (Encoded with jl) | $FRV_{jl}$ Description |
|---|---|---|---|
| $FR_1$ | Current | 11 | 100 |
| | | 12 | 78.3 |
| | | 13 | 67.5 |
| $FR_2$ | Pbfree | 21 | Y |
| $FR_3$ | Length | 31 | 9.2 |
| | | 32 | 11.17 |
| | | 33 | 10.75 |
| $FR_4$ | Diameter | 41 | 4.5 |
| | | 42 | 5.5 |
| | | 43 | 5.13 |
| $FR_5$ | Coating | 51 | Au |
| | | 52 | Alloy |
| | | 53 | None |
| $FR_6$ | Angle | 61 | 44.5 |
| | | 62 | 47 |
| | | 63 | 42.5 |
| $FR_7$ | Strength | 71 | 6.7 |
| | | 72 | 4.5 |
| | | 73 | 5.13 |
| $FR_8$ | Weight | 81 | 2.4 |
| | | 82 | 2.42 |
| | | 83 | 2.38 |
| $FR_9$ | Hardness | 91 | 49 |
| | | 92 | 57.5 |
| | | 93 | 47.5 |

### 4.2.4 Mapping mechanism between CN and FRV

Since the customer requirements $CN_{ip_i}$ and $FRV_{jl}\,(j=1,2...9; l=1,2,3)$ have been obtained as above, the data list about the $CN_{ip_i}$ and the corresponding $FRV_{jl}{}^i$ are shown as Tab. 13.

$CN_{ip_i}$ and $FRV_{jl}{}^i$ are input as identifier-item file which is a text file that lists customers' information which is to be analyzed in identifier-item format.

Each customer has a unique identifier within which are at least two columns, one for the identifier and one for the items.

**Tab. 13 The data list of $CNip_i$ and $FRV_{jl}{}^i$**

| Customer ID | $CN_{ip_i}$ (Represented by CN No) | $FRV_{jl}{}^i$ (Encoded as jl$^{(i)}$) |
|---|---|---|
| 001 | 1,4,7,8,12,13,14 | $21^{(1)},11^{(1)},31^{(1)},71^{(1)},93^{(1)},63^{(1)}$… |
| 002 | 1,4,7,8,9,12,13 | $21^{(2)},11^{(2)},31^{(2)},41^{(2)},71^{(2)},51^{(2)}$… |
| 003 | 2,5,9,11,13 | $21^{(3)},12^{(3)},33^{(3)},43^{(3)},73^{(3)},52^{(3)}$… |
| . | | . |
| . | | . |
| . | | . |
| 028 | 3,5,9,10,13 | $21^{(28)},13^{(28)},32^{(28)},42^{(28)},72^{(28)}$… |
| 029 | 1,4,7,12,13,14 | $11^{(29)},31^{(29)},41^{(29)},71^{(29)},51^{(29)}$… |
| 030 | 2.5.9.11.13 | $12^{(30)},21^{(30)},32^{(30)},43^{(30)},52^{(30)}$… |

In this case, the search mode is set as searching for strength, that is, the association rules are generated according to the strength value as a descendent order. The maximum number of association is set to 10000 so that the association rules can be derived completely. The minimum leverage, minimum lift, minimum strength, minimum coverage and minimum support are set as 0.001, 1.0, 0.1, 0.2 and 0.2 respectively. There are 37 rules generated. The results for the associations between $CN_{ip_i}$ and $FRV_{jl}\,(j=1,2...9; l=1,2,3)$ are partly shown as Tab. 14 (For illustrative simplicity, only 7 rules are listed in the Table).

**Tab. 14 Association rules between $CN_{ipi}$ and $FRV_{jl}$**

Rule1:
green environment\par -> pf_y\par
[Coverage=0.882 ; Support=0.882 ; Strength=1.000; Lift=1.13; Leverage=0.1038 ]

Rule 2:
alarmed when can't depend on vibration\par &not easy to lose\par&
change the style of the mobilephone occasionally at a lower price\par-> h_57.5\par
[Coverage=0.294; Support=0.265; Strength=0.900; Lift=2.55; Leverage=0.1609]

Rule 3:
alarmed when can't depend on vibration\par&
change the style of the mobilephone occasionally at a lower price\par &
not easy to lose\par -> c_78.3\par
[Coverage=0.294 ; Support=0.265 ; Strength=0.900; Lift=2.55; Leverage=0.1609 ]

Rule 4:
alarmed when can't depend on vibration\par &
change the style of the mobilephone occasionally at a lower price\par -> l_11.17\par
[Coverage=0.294 ; Support=0.265 ; Strength=0.900; Lift=2.55; Leverage=0.1609 ]

Rule 5:
alarmed when can't depend on vibration\par->h_57.5
[Coverage=0.353 ; Support=0.294 ; Strength=0.833; Lift=2.36; Leverage=0.1696 ]

Rule 6:
not easy to lose\par &alarmed when can't depend on vibration\par -> a_47\par
[Coverage=0.294 ; Support=0.265 ; Strength=0.900; Lift=2.55; Leverage=0.1609 ]

Rule 7:
not easy to lose\par&comfortable to hold&
change the style of the mobilephone occasionally at a lower price\par ->w_2.42\par
[Coverage=0.353 ; Support=0.265 ; Strength=0.750; Lift=1.96; Leverage=0.1298 ]

.
.
.

## 5. Summary

This paper proposes the ARIS system to facilitate communications between customers, marketers and designers. In evolutionary design, the historical data about the customers' needs and the functional specifications can be used to mine valuable information. The ARIS system can improve design efficiency and quality by providing the clear mapping mechanism between the customers' requirements and product specifications and enabling the design of products based on customers' needs. It also establishes a coherent framework for cooperation among marketers, designers and customers.

## References

[1] Bailetti, J. "Integrating Customer Requirements into Product Designs," *Journal of Product Innovation Management,* 1995,

12(1), 3-15.
[2] Chen, Y. "Application of cluster analysis to fabric classification," *International Journal of Clothing Science and Technology*, 1999, 11(4), 206-215.
[3] Cristiano, J. "Customer-Driven Development Through Quality Function Deployment in the U.S. and Japan," *Journal of Product Innovation Management*, 2000, 17(4), 286-308.
[4] Demir, Y. & Ucar, A. "Modeling and simulation with neural and fuzzy-neural networks of classifiers switched circuits," *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering,* 2003, 22(2), 253-272.
[5] Fung, Y.K. & Law, S.T. "Design targets determination for inter-dependent product attributes in QFD using fuzzy inference," *Integrated Manufacturing Systems*, MCB University Press, ISSN 0957-6061, 1999.
[6] Gause, D.C. & Weinberg, G.M. *Exploring Requirements: Quality Before Design*, New York: Dorset House Publishers, 1989.
[7] Gershenson, J.K. & Stauffer, L.A. "A Taxonomy for Design Requirements from Corporate Customers," *Research in Engineering Design*, 1999, 11(2), 103-115.
[8] Goffin. K. "Evaluating Customer Support During New Product Development-An Exploratory Study," *Journal of Product Innovation Management,* 1998, 15(1), 42-56.
[9] Han, J. & Kamber, M. *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 335-387.
[10] Heggere, S. & Rushing, J.A. "Using Association Rules as Texture Features," *IEEE Transactions on pattern Analysis and Machine Intelligence*, 2001, 23(8), 845-858.
[11] Herrmann, A. & Huber, F. "Market-driven product and service design: Bridging the gap between customer needs, quality management, and customer satisfaction," *International Journal of Production Economics,* 2000, 66(1), 77-96.
[12] Jiao, J. & Tseng, M.M. "Fundamentals of Product Family Architecture," *Integrated Manufacturing Systems,* 2000, 11(7), 469-483.
[13] Lancaster, K. "Markets and Product Variety Management," *Product Variety Management Research Advances,* Kluwer Academic Publishers, 1998, 1-15.
[14] Min, H. "A data mining approach to developing the profiles of hotel customers," *International Journal of Contemporary Hospitality Management,* 2002, 14(6), 274-285.
[15] Nayak, U. "Product Standardization using Axiomatic Design," *Second International Conference on Axiomatic Design,* 2002.
[16] Pullman, E. "A comparison of quality function deployment and conjoint analysis in new product design," *Journal of Product Innovation Management,* 2002, 19(5), 354-364.
[17] Rushing, A. "Using Association Rule as Texture Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2002, 23(8), 845-858.
[18] Suh, N. *The principle of Design*, New York: Oxford University Press, ISBN 0-19-504345-6, 1990.
[19] Tseng, M.M. & Jiao, J. "Computer-Aided Requirement Management for Product Definition: A Methodology and Implementation," *Concurrent Engineering: Research and Application*, 1998, 6(3), 145-160.
[20] Wei, Q. & Chen, G. "Simple Association Rules (SAR) and the SAR-Based Rule Discovery," *Journal of Computer & Industrial Engineering,* 2002, In Press.