INT. J. PROD. RES., 1 AUGUST 2004, VOL. 42, NO. 15, 2955–2969

Data-mining-based methodology for the design of product families

B. AGARD† and A. KUSIAK‡*

Companies design and manufacture widely diversified products to satisfy the needs of their customers and markets. Two issues important to achieving this aim are discussed. The first concerns adequate diversity for a particular market. The second concerns the management and manufacture of products within an acceptable lead time and an acceptable cost. The two issues are examined with a methodology for the design of products families. This methodology is based on a data-mining approach and it focuses on the analysis of functional requirements.

1. Introduction

Manufacturers compete on price, adaptability and variety of products, which are all driven by customer satisfaction. To satisfy customers' needs, customer-specific products should be produced. However, the latter increases production costs and the product market price. Manufacturing cost can be reduced by standardizing products to realize the benefits of the economy of scale.

Many manufacturers use some degree of standardization by defining product families. These product families could be further partitioned into subfamilies to better match distinct market segments. Each subfamily can be customized according to the needs and preferences of the specific customer segment. Such an attempt to match products with market segments results in a large number of options and alternatives. For producers, this commercial diversity must be controlled, otherwise it could lead to expensive diversification process (Martin and Ishii 1997). The necessary market diversity of products can be provided when it is supported by a low technical diversity, which guarantees acceptable product development and manufacturing costs (Child *et al.* 1991).

Managing diversity is key to the proper interaction between different departments in any company:

- Sales department is interested in a wide diversity of products to satisfy each customer.
- Purchasing department prefers a low diversity of products for simplicity of interactions with suppliers.
- Design department supports product families due to the likely reuse of previous designs, and the reduction in the amount of information to be handled, and reduced product development cost.

Revision received February 2004.

†Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal, Québec H3C3A7, Canada

‡Intelligent Systems Laboratory, 2139 Seamans Center, Department of Mechanical and Industrial Engineering, The University of Iowa, IA 52242-1527, USA.

*To whom correspondence should be addressed. e-mail: andrew-kusiak@uiowa.edu.

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540410001691929

It is important to consider the relationship between the product market and technical diversity early in the product life cycle, ideally at the product development stage. This paper proposes a methodology for the design of product families based on the analysis of the customers' requirements using a data-mining approach.

The paper is organized as follows. Section 2 discuses the issues arising in the design of product families, standardization, modular design and the existing methodologies for the design of product families. Data-mining techniques are introduced. In Section 3, a methodology for the design of product families with data-mining algorithms is proposed. Section 4 concludes the paper.

2. Problem statement

2.1. *Design of product families*

Numerous product design methodologies have been developed in recent years (Pahl and Beitz 1988). Many of these developments have been captured as Design-for-X (DFX), where X takes different meanings, e.g. A = Assembly, M = Manufacturing, R = Recycling and V = Variety.

This paper emphasizes the design of products for variety (DFV), in particular the timing of the product diversification.

To meet diversified product requirements, numerous strategies are available. It is conceivable that a standardized product would satisfy many customers as well as the requirements of a single customer. A cost-based compromise between these two strategies is of interest.

The following sections discuss standardization, modular design and methodologies for the design of product families.

2.1.1. *Standardization*

Standardization of products has been discussed in the literature from different view points. Tarondeau (1998) analysed the benefits of standardization due to the decrease in the number of reference points to be managed, an increase in the quantity of components and the reduction of complexity of the manufacturing system.

Lee and Tang (1997) developed a mathematical model to determine the best compromise between the investment necessary for the standardization and the profit resulting from the economy of scale.

Erol (1999) proposed a mathematical formulation for the standardization of low value components that was solved by Dupont *et al.* (1991).

Fouque (1999) examined several scenarios for the standardization of two components (C1 and C2) into one (Cs), namely an increase in the service level of the component C1 and/or C2, a decrease in the correlation between the demand for C1 and C2, an increase in the uncertainty of the demand for C1 and/or C2, similar costs of the components C1 and C2, and a low demand for the two components. Standardization aggregates the risk and makes the uncertainty of the standardized component Cs smaller than the uncertainty of each component C1 and C2. In addition, the size of the buffers may be reduced and the productivity and service level may increase Dupont (1998).

Lee and Tang (1997) discussed an industrial example of printer standardization.

Kota *et al.* (2000) proposed a measure that captures the level of commonality in a product family, i.e. the potential of the part family to divide the elements and to reduce the total number of parts. This measure allows for the comparison of design alternatives.

Thoteman and Brandeau (2000) presented an approach for determining the optimal level of commonality in a sub product that does not differentiate models from the customer's point of view. Using an automobile case study, they showed that an optimal design (from a cost point of view) can be obtained by optimizing the commonality between components.

For highly diversified products, standardization does not appear to be a relevant solution. Standardization implies that the design and production of generic products (or components) satisfy a large set of functions. The latter might result in unnecessary cost due to functional redundancy present in the final product. If the requirements are too highly diversified, the standardization cost of the final product may not be acceptable to the customer.

2.1.2. *Modular design*

Another way to design products for highly diversified requirements is to apply modular design principles.

Even a limited number of modular components can result in a large number of final products. Note that when using m modular components, it is possible to realize $2^m - 1$ different products. In reality, functional and physical constraints between the modules may limit this product diversity.

Modular design enables delayed differentiation between products (Lee and Tang 1997). It leads to optimized products, processes, and improved productivity (Child *et al.* 1991, MacDuffie *et al.* 1996).

Modular components necessitate standard interfaces allowing for their use across different products. To implement the modular concept, it is necessary to partition the product into semi-independent or mutually separable elements. It then becomes possible to design, manufacture and service the modules independently. The differentiation of products is accomplished at the assembly stage by the selection of modules and their location in a product.

Kusiak and Huang (1996) and Huang and Kusiak (1998) discussed modular design aimed at the production of a wide variety of products at lower cost. A matrix representation of the product allowed the identification of modules sharing different characteristics.

Numerous applications of the product modular concept in are presented by Kusiak (1999) and the engineering design literature: however, none of the papers published on this subject discusses the design of product families.

2.1.3. *Methodologies for the design of product families*

The design of product families calls for methodologies leading to:

- stable environment where each component can be developed independently;
- simplified risk management;
- ease of replacement of a faulty component by another one with an identical interface; and
- new versions of a product developed by replacing marginal components.

Product families create a stable product architecture. A design methodology should separate fixed and variable elements to ensure robustness of the product architecture.

Erens and Verhulst (1997) and Jiao and Tseng (1999) recommended that the product structure be based on the coupling between functional and technical domains. The authors considered a physical domain, which corresponds to the physical realization, but the structure of the product family was not based on this domain.

Erens and Verhulst (2004) proposed handling the stable aspect of the product architecture by increasing the performance/cost ratio and the variable aspect by increasing the variety/cost ratio. Jiao and Tseng (1999) suggested that functional variety (related to consumer satisfaction) should be promoted at the product development stage, whereas technical variety (related to manufacturing complexity and production cost) should be reduced.

The Engineering Design Research Laboratory (Yu *et al.* 1998, Zamirowski and Otto 1999a,b) has distinguished between the origin and diversity of requirements. Variation in the requirements and evolution in time are expressed by the following design rules:

- If the variations are small, design a standardized product.
- If the variation originates from heterogeneous customer requirements, design a product family with variants. The variants absorb heterogeneity.
- If the requirements evolve in time, isolate the evolving characteristics and design a module for them. This module will have different versions in time.

Gonzalez-Zugasti *et al.* (1999, 2000) proposed a methodology based on negotiation. A common platform is negotiated with the total cost of the product family. The alternatives are derived from the platform based on the individual technical performances. It is an optimization problem where the advantages from the common platform are balanced with the constraints of the alternatives.

Martin and Ishii (1999, 2000) discussed a methodology for the development of a robust architecture for a product family. The goal was to decrease the design effort and reduce the time to market of various alternatives. The methodology is based on a measure that identifies the redesign effort of adapting a component to various alternatives. A distinction is made between external reasons and the (internal) propagation for redesign.

Agard and Tollenaere (2003) proposed a methodology for design of a product family based on the analysis of consumers' needs. The methodology is illustrated with the design of electrical wire harnesses (Agard 2002, Agard and Tollenaere 2002).

Fujita (2002) optimized the content of modules and their mix in a fixed modular architecture. Yigit *et al.* (2002) solved a similar problem by determining the best subsets of modules that minimized the reconfiguration cost. Hamou (2002) proposed a methodology for the selection of the best alternative with a focus on logistics.

2.1.4. *Summary*

Most of the published methodologies for the design of product families consider cases where the set of functional requirements is well defined before the actual design is created. They are concerned with a small-to-medium number of different products. However, in the case of changing design requirements and a wide diversity of products, data-mining algorithms could provide additional information useful in the definition of the product family structure.

The following section introduces basic concepts for data mining.

2.2. *Data mining*

Before a product is designed, most companies perform marketing studies. The goal of these studies is to understand the customers' expectations. Different metrics are used to extract relevant information from databases.

Classical statistical tools are used to compute various models (e.g. regression models) and parameters (e.g. mean, confidence intervals) based on the collected data. Hypotheses can be validated in support of decision-making.

The goal of the research discussed in this paper is to extract unknown information and knowledge from databases rather than validate a hypothesis. This makes the classical statistical tools insufficient. Moreover, the focus is on an individual product or a group of products (subpopulation) rather than the entire population of products.

Some of the data-mining applications of interest to the research presented are reviewed next.

Anand and Büchner (1998) defined data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets. They classified data-mining tasks as predictive and descriptive. Predictive tasks are those that produce models that can be used for classification. Descriptive tasks produce understandable and useful patterns and relationships describing a complex data set.

Westphal and Blaxton (1998) identified four functions of data mining: classification, estimation, segmentation and description. Classification involves assigning labels to previously unseen data records based on the knowledge extracted from historical data. Estimation is the task of filing in missing values in the fields of an incoming record as a function of fields in other records. Segmentation (called also clustering) divides a population into smaller subpopulations with similar behaviour. Clustering methods maximize homogeneity within a group and maximize heterogeneity between the groups. The description task focuses on explaining the relationships among the data.

Apté (1997) described different classes of algorithms for mining data.

In predictive modelling, a function (model) links an output with the input. If the output variable is discrete valued, classification is employed. If the output variable is continuous, prediction modelling is used, e.g. by using neural networks. A clustering algorithm identifies sets of similar examples according to the predefined metric. To obtain these sets, different methods are used, e.g. the k-means algorithm, hierarchical algorithms, pattern recognition, Bayesian statistics and neural networks.

Extraction of frequent patterns (called association rules mining) leads to data patterns with some predefined level of regularity. Measures such as support and confidence permit the evaluation of the quality of the extracted rules (patterns). For example, for the association rule $A \Rightarrow B$, support expresses the number of times A occurs as a fraction of the total number of examples. Confidence is the fraction of the number of times B exists in the data when A is present. An association with high confidence and support is called strong and could be potentially useful.

Fayyad *et al.* (1996) defined a data-mining process for the extraction of knowledge from a data set. Several steps are considered with frequent iterations aimed at the extraction of valuable knowledge. They begin with the development of an understanding of the application domain, the relevant prior knowledge and the goals of the end user. The next steps deal with the creation and preparation of the data to be mined (selection, cleaning, preprocessing, reduction and projection of the data).

Then, the most suitable data-mining algorithm is selected to search for patterns in a particular representation form or a set of such representations. Knowledge is then extracted, interpreted, and validated.

Berry and Linoff (1997) presented many examples and applications of data mining in marketing, sales, and customer support.

More specifically in engineering, many applications have been managed. Büchner *et al.* (1997) describe the concepts of data mining and their synergy with manufacturing environments. A generic process is introduced and applications are shown.

In engineering design, Leu *et al.* (2001) present a data-mining approach for the prediction of tunnel support stability. This approach is used to determine the size of the tunnel support knowing rock type and construction parameters. Kusiak (2000) used data mining (the rough set theory approach) to extract rules for making predictions in the semiconductor industry. Numerous methods for the decomposition of data sets are discussed in Kusiak (2001). The decomposition is used to enhance the quality of knowledge extracted from large databases by simplifying the data-mining task. This decomposition is used for the prediction and prevention of manufacturing faults in wafers.

Gertosio and Dussauchoy (2004) applied data mining to extract knowledge from an industrial truck manufacturer database. The goal was to discover knowledge in the data of the test engine process in order to significantly reduce the processing time.

The following section outlines a methodology for the application of data mining in the design of product families.

3. Methodology for the design of product families

Consider a scenario where a set of product features dictated by certain technical requirements is to be satisfied by a number of products. Assume that a description of the customers and their requirements is available for these products for different times. An example of the customers' descriptions and requirements for features F1, ..., F7 at time t_1 is shown in table 1.

Row 1 indicates that customer 1 is 35 years old, is female (F) with a low income (L) and has insurance (yes). With regard to the product to be designed, she is interested in features F1, F3 and F4. For feature F7, she expects the value 40. Row 2 states that customer 2 is 55 years old, is male (M) with a high income (H)

	Customer description				Product requirement						
Time t_1	Age	Gender	Income	Insurance	F1	F2	F3	F4	F5	F6	F7
Customer 1	35	F	L	Yes	1		1	1			40
Customer 2	55	M	H	Yes		1		1	1		40
Customer 3	40	M	M	Yes	1	1		1		1	70
Customer 4	28	F	H	No	1		1				65
Customer 5	29	F	L	Yes	1	1	1	1	1	1	75
Customer 6	50	M	H	Yes		1			1	1	75
Customer 7	32	F	M	No	1			1		1	45
Customer 8	37	M	L	No	1		1				45
Customer 9	48	M	M	No	1	1	1	1	1	1	60
Customer 10	65	M	H	Yes		1	1		1		65

Table 1. Description of customers' and requirements at time t_1.

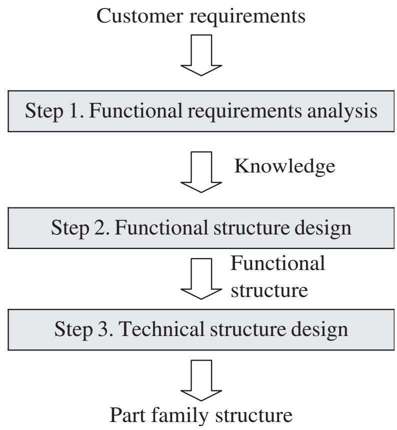


Figure 1. Three-step methodology for the design of product families.

and has insurance. With regard to the product to be designed, he is interested in features F2, F4 and F5. For feature F7, he expects the value 40.

The methodology presented below uses data mining to design a product family based on customer descriptions and requirements.

Large databases similar to that in table 1 are used for the extraction of relevant information needed for the design of product families.

The methodology is based on three steps (figure 1). It begins with the definition of customer requirements and ends with a structure for the product family.

The methodology defines a product structure with a stable architecture as well as the necessary options and variants. Each of the three steps of the methodology is discussed next.

3.1. *Step 1. Functional requirements' analysis*

In this step, customer expectations for the product's (re)design are formulated. Two points of view are used: the first (customer centered) is based on an analysis of the customers' description, the second (product centered) is based on an analysis of the customers' expectations to be delivered by a product.

3.1.1. *Customer centered*

The diversity considered here is dictated by the heterogeneous needs of the customers as well as some internal parameters of the company. For example, the company's strategy might aim at manufacturing products with strong diversity, while a commonly used commercial approach may be to compete. Functional diversity considers at the same time the company's strategy as well as the customers' requirements. Analysing customers' requirements allows one to:

- develop of a model of a representative customer;
- define several groups of similar customers. It could be judicious to propose several groups of products for each group of customers; and
- select and prioritize a set of customers in order to produce, for example, a top-of-the-range product for a certain category of customers.

3.1.1.1. *How to build a representative customer model.* Different techniques are available to define an 'average' customer based on the customers' expectations, as opposed to a 'population'-based customer.

An 'averaging' technique produces a virtual customer that does not directly correspond to any of the actual customers. For example, if two customers want a 'black car' and another one wants a 'white car', an 'average customer' wants a 'gray car'. A population-based customer follows the majority of customers and selects a 'black car'. It can be created with a neural network.

Once a 'representative customer' has been created, a product (in particular a standardized one) can be developed. The design of such a product is not discussed in this paper, but see Pahl and Beitz (1988).

3.1.1.2. *How to define relevant set of customers.* Clustering is one of data-mining techniques for grouping data objects in subsets. These subsets contain objects that are similar according to predefined metrics.

In this paper, clustering is used to identify similar customers that share the same or highly similar behaviors.

When applying clustering on a table similar to table 1, customers are isolated and regrouped in clusters (table 2) according to selected features.

The first row indicates that customers 1, 3, 4, 7 and 8 have similar behaviour. Row 2 indicates that customers 2, 6 and 10 share other similar characteristics, and so on.

Table 2 also provides useful information about the size of each cluster: 50% of the customer population for set 1, 30% for set 2 and 20% for set 3. This information may be used to determine the capacity of the manufacturing process, to size productions and to select design alternatives.

Since customers from the same cluster share similar requirements, it could be judicious to propose a specific product design for each cluster of customers.

Questions of a different nature could be asked based on the data stored in each cluster. For example, one could consider the following:

- Rejecting the design: if the size of the cluster is too small to assure sufficient benefits, it could be the case of set 3.
- Accepting the design and in turn considering:
  - Standardized product: the same product could be designed for all the customers in the same set.
  - Product platform based on a common structure and modules: a product platform will be shared by all the customers in the set, and modules will be designed to better fit some specific requirements (this alternative is detailed in Steps 2 and 3).

	Customer
Set 1: (50%)	1, 3, 4, 7, 8
Set 2: (30%)	2, 6, 10
Set 3: (20%)	5, 9

Table 2. Clusters of customers based of features F1, ..., F6.

3.1.1.3. *How to select customers likely to buy a product.* A company may follow a strategy of focusing on specific categories of customers (young people, females with cars, insured males). In this case, it might be beneficial to identify such customers in the overall population. Classification tools allow differentiating and classifying the customers based on their characteristics (customer's description). Having identified all relevant characteristics that differentiate the customers, it becomes possible to select the most relevant characteristics and the corresponding set (cluster) of customers. For a specific set of customers, it may be possible to design a standardized product, or a product family in the same way as described below.

Each product family is specific to a customer group. According to the diversity of the requirements and the company's strategy, it is necessary to select the most suitable model, i.e. average customer, subset of customers or preferential subset of customers. The same company may select different models for different products or for the same product but at different times. The competition, partnerships and commercial opportunities might impact this decision.

For the demonstration, according to the strategic perspective of the company, set 1 (from table 2) will be considered the set of customers to satisfy (the other sets would not provide enough benefits). The design of a product platform will be performed for this set of customers.

3.1.2. *Product centered*

This section provides information about the product. Consider that the decision has been made to focus on a product platform to meet the heterogeneous requirements of a subset of customers. To design a product platform, the set of requirements for a product is analysed with data-mining algorithms. Knowledge is extracted from the data for selected customers, e.g. customers 1, 3, 4, 7 and 8 from table 1.

The next section discusses analysis that might provide useful information for the design of a product family. First, a distinction is made between variations of customers' requirements and then the associations among customers' requirements are identified.

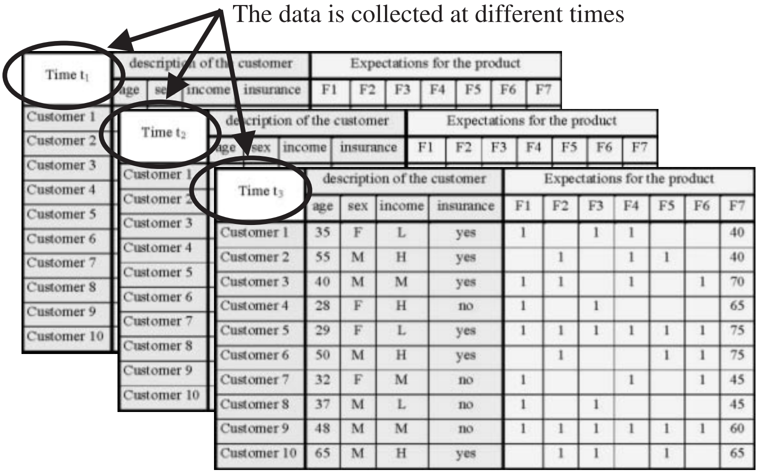
3.1.2.1. *Differentiate variation in customers' requirements.* Recognizing the origin of the variation of customers' requirements allows designing and maintaining different modules as a function of the origin of the distribution they support (Yu *et al.* 1998, Zamirowski and Otto 1999a,b).

Suppose the customer expectations for the product have been collected at different times (figure 2).

It is then possible to represent for each feature at each time the number of customers that requires a certain value (figure 3). For example, it could be feature F7 (maximum speed of a car). It is then possible to distinguish, for each feature, variation in customers' expectations at a fixed time, and the evolution in customers' expectations.

To represent the relationship between a product's performance and the evolution in product performance, two types of variation are identified from the customers' requirements at different times. Two measures are used to characterize the variations:

- σ_a (customer-dependence) represents the variation at a given time, e.g. $\sigma_a(t_1)$, $\sigma_a(t_2)$, $\sigma_a(t_3)$. It means that at time t_1 the dispersion for the maximum speed of the car is $\sigma_a(t_1)$, at time t_2 it is $\sigma_a(t_2)$, and so on.



Time t_3	description of the customer				Expectations for the product						
	age	sex	income	insurance	F1	F2	F3	F4	F5	F6	F7
Customer 1	35	F	L	yes	1		1	1			40
Customer 2	55	M	H	yes		1		1	1		40
Customer 3	40	M	M	yes	1	1		1		1	70
Customer 4	28	F	H	no	1		1				65
Customer 5	29	F	L	yes	1	1	1	1	1	1	75
Customer 6	50	M	H	yes		1			1	1	75
Customer 7	32	F	M	no	1			1		1	45
Customer 8	37	M	L	no	1		1				45
Customer 9	48	M	M	no	1	1	1	1	1	1	60
Customer 10	65	M	H	yes		1	1		1		65

Figure 2. Customer requirements collected at different times t_1, t_2 and t_3.

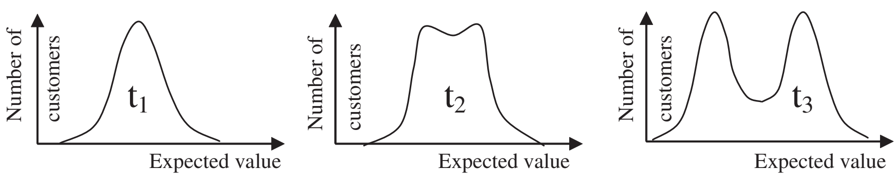


Figure 3. Variations in customer requirements at time t_1, t_2 and t_3.

- σ_t (time-dependence) represents the evolution of the variation in time. For the maximum speed of the car, it is possible to discover that the evolution distinguishes two different maximum speeds required (even if the medium speed is still the same), then perhaps two different kinds or power levels of engines are to be commercialized.

These two measures are used in Step 2 to differentiate the modules.

3.1.2.2. *Derive associations among requirements.* Once a set of customers has been selected and the diversity between the products has been identified, associations between customers' requirements are quantified in this phase.

Assume that set 1 of table 2 is selected for the design of a product family as shown in table 3. Table 3 provides the percentage of customers for each feature, F1, …, F6. This measure can be used to determine the capacity of the manufacturing process, select design alternatives, select material, production equipment, suppliers, and so on.

The percentages for sets 1 and 3 could be aggregated to provide a global value of the production (when different subsets of customers are considered) (figure 4).

The association among requirements may provide additional information useful for the design. For example, technical specifications might state that a car that has two doors and a diesel engine requires a specific speed transmission. In such a case, knowing the number of cars with two doors and the number of cars with a diesel

	F1	F2	F3	F4	F5	F6
Customer 1	1		1	1		
Customer 3	1	1		1		1
Customer 4	1		1			
Customer 7	1			1		1
Customer 8	1		1			
Total: five customers	100%	20%	60%	60%	0%	40%

Table 3. Subset of customers from table 1.

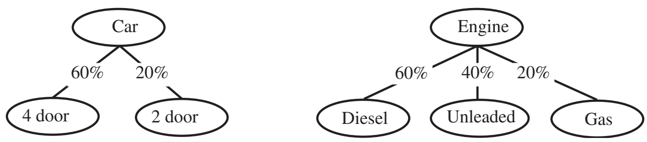


Figure 4. Probability of the appearance of different functions.

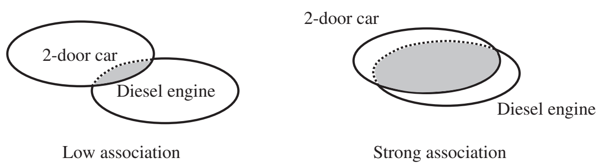


Figure 5. Associations among customers' requirements.

engine is not relevant. To determine the capacity of the manufacturing process, it is necessary to know the number of cars with two doors and a diesel engine.

Associations of this nature can be extracted with data-mining algorithms based on the database. Moreover, these associations may have pre-specified strength and confidence (figure 5).

Relationships among customers' requirements discovered by the association-rule algorithms might suggest specific designs. Step 3 emphasizes the use of this knowledge.

An example of an association rule extracted from table 3:

$$\text{F1 and F4} \Rightarrow \text{F6 with support} = 3/5 \text{ and confidence} = 2/3.$$

This rule may suggest considering the design for F1, F4 and F6 not independently.

3.2. *Step 2. Functional structure design*

In this step, a functional structure is designed for the product family based on the analysis of the customers' requirements. The following three questions help to distinguish between relevant cases. Each function has to be considered.

(1) Is the function stable (i.e. are σ_a and σ_t small)? This means that most customers want the same function with almost the same characteristic/performance. Such functions are called 'stable functions.'

① May be treated as 'stable functions' with a robust design		σ_a (Customer-dependence)	
		Small	Large
σ_t (time-dependence)	Small	Stable functions	Options and alternatives ①
	Large	Functions to be versioned ①	Options and alternatives to be versioned ①

Table 4. Design of a functional structure.

(2) Can the variation of the function be supported by a robust and inexpensive design? The design of a function is robust if the same design satisfies various requirements. Such functions are treated as 'stable functions.'

(3) Is the variation customer-dependent and/or time-dependent (are σ_a and/or σ_t large?) If the variation is customer-dependent (σ_a is large), it means that all customers have different requirements. Variations among products for the same function are considered. These functions are called 'options and alternatives.' Alternatives are always considered for final products (products with modifiable parameters, e.g. the color of a car), while options are not always considered for the final products. If the variation is time-dependent (σ_t is large), then the function calls for an evolution in time to meet the variable customers' requirements, i.e. these functions need to be versioned.

Table 4 summarizes all different cases with stable functions, functions to be versioned, options, and identified alternatives.

3.3. *Step 3. Technical structure design*

The focus of this step is on the design of a technical structure for the product family that supports these different types of functions.

It is necessary to make a strategic decision about whether or not each function is 0to be designed. This choice needs to be argued through a cost analysis that needs to be as thorough as possible concerning all possible solutions by answering the following questions:

- What is the cost/pay-off of implementation of such an option/alternative?
- What is the cost/pay-off of producing such an option/alternative?
- How much is the cost/pay-off of not producing such an option/alternative?

The above considerations may lead to the rejection or the introduction of some functions.

After this strategic decision has been made, it is necessary to answer the question how to manufacture a wide (functional) diversity of products at a lower cost? Consider that the decision to design and manufacture a product family based on a common platform with options and alternatives has been made.

A technical structure is necessary for the product family. This structure identifies functions that are common for the family and the ones that are specific to the adaptation of the family to specific requirements.

Stable aspects of the design are integrated in a common platform to the degree possible. They are developed to increase the performance/cost ratio by integrating the maximum number of functions. The latter allows for the rationalization of manufacturing resources.

① *Q*1 *is true* ② *Q*1 *is wrong and Q*2 *is true* ③ *Q*1 *is wrong and Q*2 *is wrong* ④ *Q*2 *is true* ⑤ *Q*2 *is wrong*		Common platform/ permanent technical solutions	Standardized module	Specific module
Stable functions		✓		
Customer-dependent functions	options	①	②	③
	alternatives		④	⑤
Time-dependent functions			④	⑤

Q1: Is there a high demand and is the standardization cost low?
Q2: Is there a high association in the requirements and is the standardization cost low?

Table 5. Design of a technical structure.

The variable aspects are designed according to the optimal variety/cost ratio. The components without alternatives are a stable factor in the design. Negotiating variable elements makes it possible to include some of these elements in the common platform (Gonzalez-Zugasti *et al.* 2000).

The technical structure separates stable and variable aspects from the functional structure. The creation of the technical structure is summarized in table 5.

Variable elements are designed around modules with standard interfaces to facilitate exchanging modules. This allows for the possibility of taking advantage of the product-delayed differentiation by using the same module for several uses.

4. Conclusions

In this paper, a methodology for using data-mining algorithms in the design of product families was introduced. In the first step of this methodology, data-mining algorithms were used for customer segmentation. Once a set of customers was selected, an analysis of the requirements for the product design was performed and association rules extracted.

The second step created a functional structure that identified the source of the requirements' variability. Options and variants are designed to satisfy the diversified requirements based on a common platform. The last step elaborated on a product structure and distinguished modules to support the product variability. Finally, the paper showed that data-mining techniques could be applied efficiently in the design of product families.

References

Agard, B., 2002, Contribution à une méthodologie de conception de produits à forte diversité. PhD thesis, Institut National Polytechnique de Grenoble.

Agard, B. and Tollenaere, M., 2002, Design of Wire Harnesses for Mass Customization. *Proceedings of the 4th International Conference on Integrated Design and Manufacturing in Mechanical Engineering – IDMME 2002*, Clermont-Ferrand, France.

Agard, B. and Tollenaere, M., 2003, Design of product families: methodology and application. *Proceedings of the 14th International Conference on Engineering Design – ICED'03*, Stockholm, Sweden.

Anand, S. S. and Büchner, A. G., 1998, *Decision Support Using Data Mining* (London: Financial Times Pitman).

Apté, C., 1997, Data mining: an industrial research perspective. *IEEE Computational Science and Engineering*, **April-June**, 6–9.

Berry, M. J. A. and Linoff, G., 2004, *Data Mining Techniques: For Marketing, Sales, and Customer Support* (New York: Wiley).

Büchner, A. G., Anand, S. S. and Hugues, J. G., 1997, Data mining in manufacturing environments: goals, techniques and applications. *Studies in Informatics and Control*, **6**, 319–328.

Child, P., Diederichs, R., Sanders, F.-H. and Wisniowski, S., 1991, The management of complexity. *Sloan Management Review*, **Fall**, 73–80.

Dupont, L., 1998, *La Gestion Industrielle: Concepts et Outils* (Paris: Hermès).

Dupont, L., Erol, M., Cormier, G. and Turkkan, N., 1999, La standardisation des composants: modèles et algorithmes. *3ème Congrès International de Génie Industriel*, Montréal, Canada, May, pp. 671–680.

Erol, M., 1999, Prise en compte de la flexibilité dans la planification dynamique. PhD thesis, Institut National Polytechnique de Grenoble.

Erens, F. and Verhulst, K., 1997, Architectures for product families. *Computers in Industry*, **33**, 165–178.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R., 1996, *Advances in Knowledge Discovery and Data Mining* (Cambridge, MA: AAAI Press/MIT Press).

Fouque, T., 1999, A la recherche des produits flexibles. *Revue Française de Gestion*, mars-avril-mai, **no. 123**, 80–87.

Fujita, K., 2002, Product variety optimization under modular architecture. *Computer-Aided Design*, 953–965.

Gertosio, C. and Dussauchoy, A., 2004, Knowledge discovery from industrial databases. *Journal of Intelligent Manufacturing*, **15**, 29–37.

Gonzalez-Zugasti, J., Otto, K. and Baker, J., 2000, A method for architecting product platforms. *Research in Engineering Design*, **12**, 61–72.

Hamou, H., 2002, Contribution à la conception de produits à forte diversité et de leur chaîne logistique: une approche par contraintes. PhD thesis, Institut National Polytechnique de Toulouse.

Huang, C. C. and Kusiak, A., 1998, Modularity in design of products and systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, **28**, 66–77.

Jiao, J. and Tseng, M., 1999, A methodology of developing product family architecture for mass customization. *Journal of Intelligent Manufacturing*, **10**, 3–20.

Kota, S., Sethuraman, K. and Miller, R., 2000, A metric for evaluating design commonality in product families. *ASME Transactions: Journal of Mechanical Design*, **122**, 403–410.

Kusiak, A. and Huang, C. C., 1996, Development of modular products. *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, *Part A*, **19**, 523–538.

Kusiak, A., 1999, *Engineering Design: Products, Processes and Systems* (San Diego: Academic Press).

Kusiak, A., 2000, Decomposition in data mining: an industrial case study. *IEEE Transactions on Electronics Packaging Manufacturing*, **23**, 345–353.

Kusiak, A., 2001, Rough set theory: a data mining tool for semiconductor manufacturing. *IEEE Transactions on Electronics Packaging Manufacturing*, **24(1)**.

Lee, H. L. and Tang, C. S., 1997, Modeling the costs and benefits of delayed product differentiation. *Management Science*, **43**, 40–53.

Leu, S.-S., Chen, C.-N. and Chang, S.-L., 2001, Data mining for tunnel support stability: neural network approach. *Automation in Construction*, **10**, 429–441.

MacDuffie, J.-P., Sethuraman, K. and Fisher, M.-L., 1996, Product variety and manufacturing performance: evidence from the international automotive assembly plant study. *Management Science*, **42**, 350–369.

Martin, M. V. and Ishii, K., 1997, Design for variety: development of complexity indices and design charts. *Proceedings of 1997 ASME Design Engineering Technical Conferences*, 14–17 September, Sacramento, CA.

Martin, M. V., 1999, Design for variety: a methodology for developing product platform architectures. PhD thesis, Stanford University, Stanford, CA.

Martin, M. V. and Ishii, K., 2000, Design for variety: a Methodology for developing product platform architectures. *Proceedings of DETC2000, 2000 ASME Design Engineering Technical Conferences*, 10–13 September, Baltimore, MD.

Pahl, G. and Beitz, W., 1988, *Engineering Design* (Berlin: Springer).

Tarondeau, J. C., 1998, *Stratégie Industrielle*, 2nd edn (Vuibert: Collection Gestion).

Thonemann, U. W. and Brandeau, M., 2000, Optimal commonality in component design. *Operations Research*, **48**, 1–19.

Yigit, A. S., Galip-Ulsoy, A. G. and Allahverdi, A., 2002, Optimizing modular product design for reconfigurable manufacturing. *Journal of Intelligent Manufacturing*, **13**, 309–316.

Yu, J. S., Gonzalez-Zugasti, J. P. and Otto, K., 1998, Product architecture definition based upon customer demands. *Proceedings of 1998 DETC ASME Design Theory and Methodology Conference*, Atlanta, GA, pp. 271–277.

Westphal, C. and Blaxton, T., 1998, *Data Mining Solutions* (New York: Wiley).

Zamirowski, E. and Otto, K., 1999a, Product portfolio architecture definition and selection. *Proceedings of the International Conference on Engineering Design, Proceedings of International Conference on Engineering Design — ICED 99*, Munich, 24–26 August.

Zamirowski, E. and Otto, K., 1999b, Identifying product portfolio architecture modularity using function and variety heuristics. *Proceedings of the 11th International Conference on Design Theory and Methodology: ASME Design Engineering Technical Conferences*, Las Vegas, Nevada, 12–15 September.