B. Agard and A. Kusiak, Standardization of Components, Products and Processes with Data Mining, *International Conference on Production Research Americas 2004*, Santiago, Chile, August 1-4, 2004.

Standardization of Components, Products and Processes with Data Mining

Bruno AGARD
Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal,
C.P. 6079, succ. Centre-ville, Montréal (Québec), H3C3A7, Canada
bruno.agard@polymtl.ca
and
Andrew KUSIAK
Intelligent Systems Laboratory, 2139 Seamans Center,
Department of Mechanical and Industrial Engineering, The University of Iowa,
IA 52242 - 1527, USA
andrew-kusiak@uiowa.edu

ABSTRACT
Data mining offers tools for extracting knowledge from databases. This paper discusses applications of data mining in standardization of components, products, and processes. Standardization of components is accomplished using association rules derived from customers' requirements. A design process is proposed for the standardization of products. The design of a unique standardized product, different standardized products, and a standardized product for a selected group of customers are considered. A methodology for process standardization is proposed. Each approach discussed in the paper is illustrated with an industrial scenario.

Keywords: Data mining, standardization, component, product, process.

1. INTRODUCTION

Engineering design, manufacturing, and marketing activities generate large volumes of data. The data collection efforts are often driven by productivity improvement efforts. This large quantity of data makes it almost impossible for a person to develop a complete understanding of it without using any tools.

Examples of data sources growing in size include (see Figure 1):

- Data from marketing studies describing customers' expectations.
- Data describing actual designs.
- Data characterizing manufacturing processes used in a company.

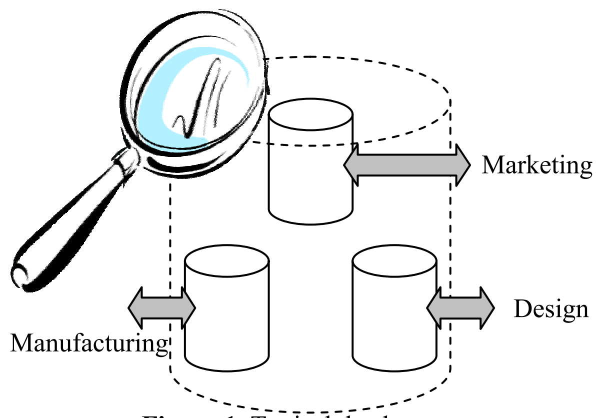

Figure 1. Typical databases

Researches have demonstrated that standardization may result in improved efficiency by taking advantage of the economy of scale and simplifying manufacturing processes. Different levels of standardization are considered (see Figure 2):

- Use of common components in different products.
- Design of standardized products for comparable requirements.
- Standardization of a manufacturing process for similar products.

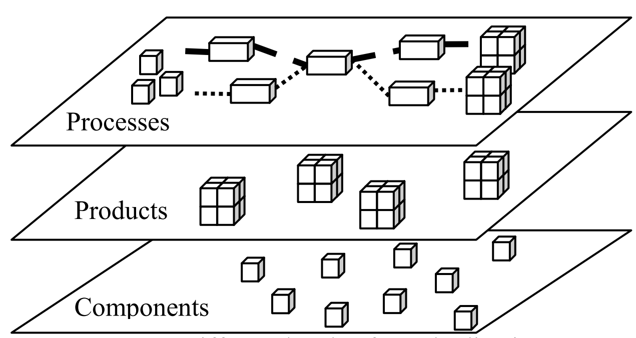

Figure 2. Different levels of standardization

The focus of this paper is on extracting knowledge from industrial databases in support of standardisation.

Section 2 discusses advantages of standardization of components, products, and processes. Some of the data-mining applications of interest to the research presented in this paper are reviewed in Section 3. Section 4 presents opportunities for data mining in standardization. Section 5 concludes the paper.

2. STANDARDIZATION

Standardization is concerned with the use of common components, products, or processes to satisfy heterogeneous needs. It necessitates designing an overly robust product or the use of a robust process (often a more flexible process). Different aspects of standardization have been discussed in the literature.

Tarondeau [18] argued that standardization results in higher productivity, larger lot sizes, decrease in the number of reference points to be managed, decrease in the stock level, and the reduction of complexity of a manufacturing system.

The unnecessary cost of robustness may be balanced by increased productivity and decrease in product and process control.

Lee and Tang [15] 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. They considered the standardization of products and processes.

Erol [6] proposed a mathematical model for standardization of low value components that was solved by Dupont *et al.* [5].

Fouque [8] identified favourable scenarios for the standardization of two components into one. The uncertainty of the components demand increases, the service level of the components increases, the components have similar costs, and the demand for the two components is small. Standardization aggregates the risk and the uncertainty of the standardized component is smaller than the uncertainty of each individual component. Then the buffer sizes may be reduced and the productivity and service level may increase [4].

Thoteman and Brandeau [19] determined the optimal level of internal standardization of products with characteristics that do not differentiate models from the customer's point of view. They concluded that an optimal design (for cost) may be also obtained.

The group technology [11] approach involves grouping parts, products, and processes that are similar, and therefore it can be applied to standardization. Numerous algorithms are available to define these groups according to various criteria, e.g., production cost, setup time, and in-process inventory.

3. DATA MINING IN ENGINEERING

Anand and Büchner [1] defined data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets.

Available volume of data in engineering provides many opportunities for knowledge extraction with data mining.

Fayyad *et al.* described the iterative process of extracting knowledge from data (data mining). The first step deals with understanding of the application domain and of the goal of the end user. The selected data is pre-processed, reduced, and presented in a suitable format for the mining. According to the goal of the analysis, an appropriate data-mining algorithm is selected. Finally, knowledge is extracted. The end user has to interpret and validate the knowledge for using it or for better specifying the problem for iteration.

Westphal and Blaxton [20] categorized data mining functions as classification, segmentation, and description.

- Classification involves assigning labels to new data based on the knowledge extracted from historical data.
- Segmentation (called also clustering) divides a population into smaller sub-populations with similar behaviour according to a predefined metric. It maximizes homogeneity within a group and maximizes heterogeneity between the groups.
- Description and visualization are used to explain the relationships among the data. Frequent patterns may be extracted in the form of A => B rules with two measures of quality: the support which represents the number of times A occurs as a fraction of the total number of examples and confidence which expresses the number of times B exists in the data when A is present.

Berry and Linoff [2] presented numerous examples and applications of data mining in marketing, sales, and customer support. Büchner *et al.* [3] focused on the applications of data mining in manufacturing. Leu *et al.* [16] applied data mining in support of the tunnel design. Prediction of the quality products in the semiconductor industry was discussed in Kusiak [13]. Gertosio and Dussauchoy [9] applied data mining to reduce the processing time for checking and adjusting electronically controlled truck diesel engines. Huang *et al.* [10] acquired knowledge from data automatically to generate IF-THEN rules for a knowledge-based expert system in the design and manufacturing of micromachined atomizer.

Many algorithms are available for extracting knowledge from databases [7]. The efficiency of these algorithms may be improved with decomposition methods [12] and feature transformation methods [14].

4. DATA MINING AND STANDARDIZATION

Data mining techniques are useful in providing answers to the following questions:

- Which components could be standardized?
- Which products could be standardized?
- Which processes could be standardized?

Different application scenarios for standardization of components, products, and processes are presented in Section 4.1, Section 4.2, and Section 4.3, respectively.

4.1 Data mining in standardization of components

Consider a company manufacturing personalized products with numerous options and alternatives. Each personalized product meets the requirements of a specific customer.

To simplify manufacturing of personalized products and meeting an acceptable lead time, standardization of the components is considered.

Assume the availably of historical data about the features selected by the customers. Table 1 represents the data relating six options and alternatives selected by eight customers.

Table 1. The relationship between customers, options, and alternatives

	F1	F2	F3	F4	F5	F6
Customer 1	1	1	1	1		1
Customer 2		1		1		1
Customer 3	1		1		1	
Customer 4		1		1		1
Customer 5	1		1		1	
Customer 6					1	
Customer 7	1	1		1		1
Customer 8	1		1			

The data in Table 1 show that Customer 1 has selected features F1, F2, F3, F4 and F6. Customer 2 has selected features F2, F4 and F6, and so on.

Information contained in Table 1 may lead to useful patterns for the company. The similarities in customers' requirements may imply relationship between the corresponding features.

The data in Table 1 could be mined to derive associations between customers' requirements (see Figure 3).

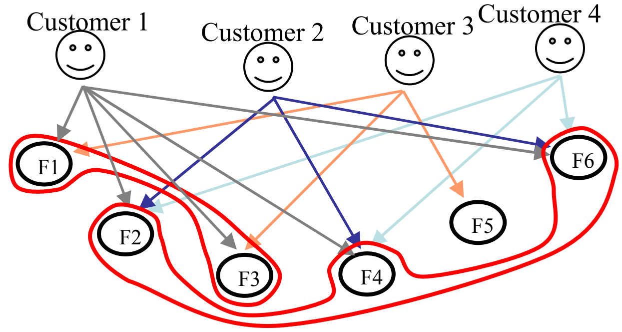

Figure 3. Associations between customers' requirements

Example rules extracted from the data in Table 1 are presented next:

Rule 1. F1 => F3, with support = 62.5% (= 5/8) and confidence = 80 % (= 4/5)

Rule 2. (F2 & F4) => F6, with support = 50 % (= 4/8) and confidence = 100% (= 4/4)

Rule 3. F5 => NOT F2, with support = 37.5 % (= 3/8) and confidence = 100% (= 4/4)

Associations among requirements with high support and confidence suggest standardization, for example:

Rule 1. Suggests a standardized component supporting features F1 and F3. The standardized component {F1, F3} would meet the requirements {F1}, {F3} and {F1, F3}.

Rule 2. This rule suggests the design of a standardized component {F2, F4, F6}.

Rule 3. This rule implies that whenever F5 is present, F2 is not included, i.e., F2 and F5 should not be realized as one component.

The associations among customers' requirements provide indications for component standardization. The design team needs to evaluate the value of these indications.

4.2 Data mining in standardization of products

Consider a company performing a marketing study preceding the design of a new product. The goal of the study is to obtain information about customers' expectations. Sample data from a marketing study is shown in Table 2.

Table 2. Description of customers and their requirements

	Customer Description				Product Requirement					
	Age	Gender	Income	Insurance	F1	F2	F3	F4	F5	F6
Customer 1	35	F	L	Yes	1		1	1		
Customer 2	55	M	H	Yes		1		1	1	
Customer 3	40	M	M	Yes	1	1		1		
Customer 4	28	F	H	No	1		1		1	
Customer 5	29	F	L	Yes	1	1	1	1	1	1
Customer 6	50	M	H	Yes		1			1	1
Customer 7	32	F	M	No	1		1	1		1
Customer 8	37	M	L	No	1		1	1		
Customer 9	48	M	M	No	1	1	1	1	1	1
Customer 10	65	M	H	Yes		1	1		1	

Table 2 contains data about customers, e.g., Customer 1 is 35 years old, a female with low income and insurance. Moreover, Customer 1 expects the features F1, F3 and F4 to be present in the product.
The similarity among customers' requirements may be analyzed. Any cluster identified in Table 1 may indicate opportunities for product standardization. Thus analyzing the customers' data may emphasize customers of special interest.
The following alternatives are considered in this section:

- The design of a standardized product for a large population of customers.
- The design of different standardized products (How many? Which ones?).
- The design of a standardized product for selected customers.

The three alternatives are discussed next.

Design of a standardized product for a large population of customers

To design a standardized product, the functions to be integrated into the product need to be specified. A standardized product does not need to support all the functions. Some functions may be integrated within the design while others can be excluded. A representative (target) customer model can be constructed using an averaging technique. However, such a customer could not represent any actual customers.
A neural network is able to learn customers' preferences to construct a target customer that could be used in the design of the standardized product (Figure 4).

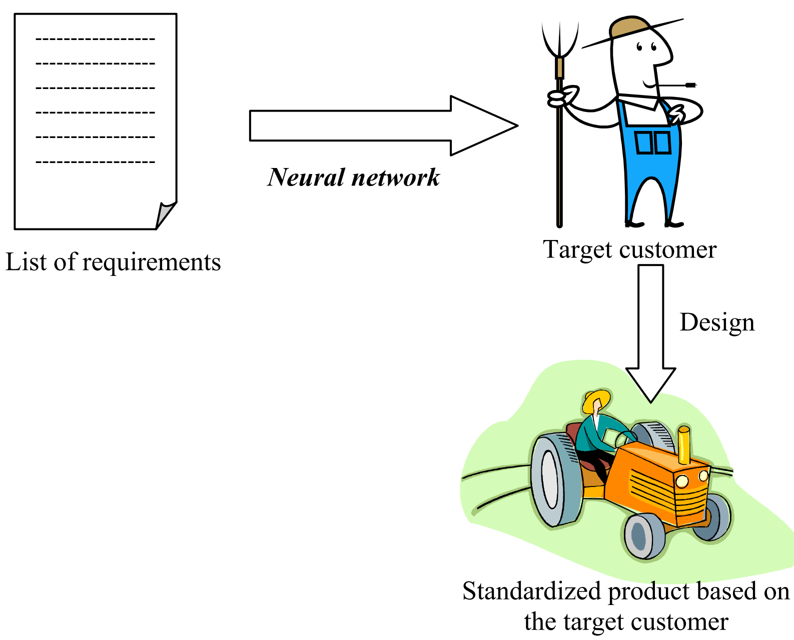

Figure 4. Design process for a standardized product

Once the requirements for the target customer are identified, economic evaluation of each feature may be performed. To determine its inclusion or exclusion in the standardized product two questions have to be answered:

- How much does it cost to design such an option/alternative?
- What is the loss of NOT designing it?

For more details on the design of standardized products see [17].

Design of different standardized products

To better match the customers' requirements, it may be judicious to design different standardized products rather than a unique one. Each standardized product could be designed for a specific subset of customers. The latter calls for determining the number of products to be designed.

Clustering is useful here as it defines groups of data objects that are similar according to the predefined metrics. In this paper, clustering is used to identify similar customers that share the same or highly similar behaviours (see Figure 5). The following process is suggested:

1. Define the expected number of clusters. The initial value could be the number of expected standardized products.
2. Form clusters of customers.
3. Analyze each cluster for suitability for standardization. Some clusters may not be relevant (not enough customers to ensure benefits).
4. If some clusters are disqualified, customers from these clusters could be considered or not in the next iteration.

Elements from the same cluster have close requirements; it implies the design of a dedicated product for each cluster (Figure 5).

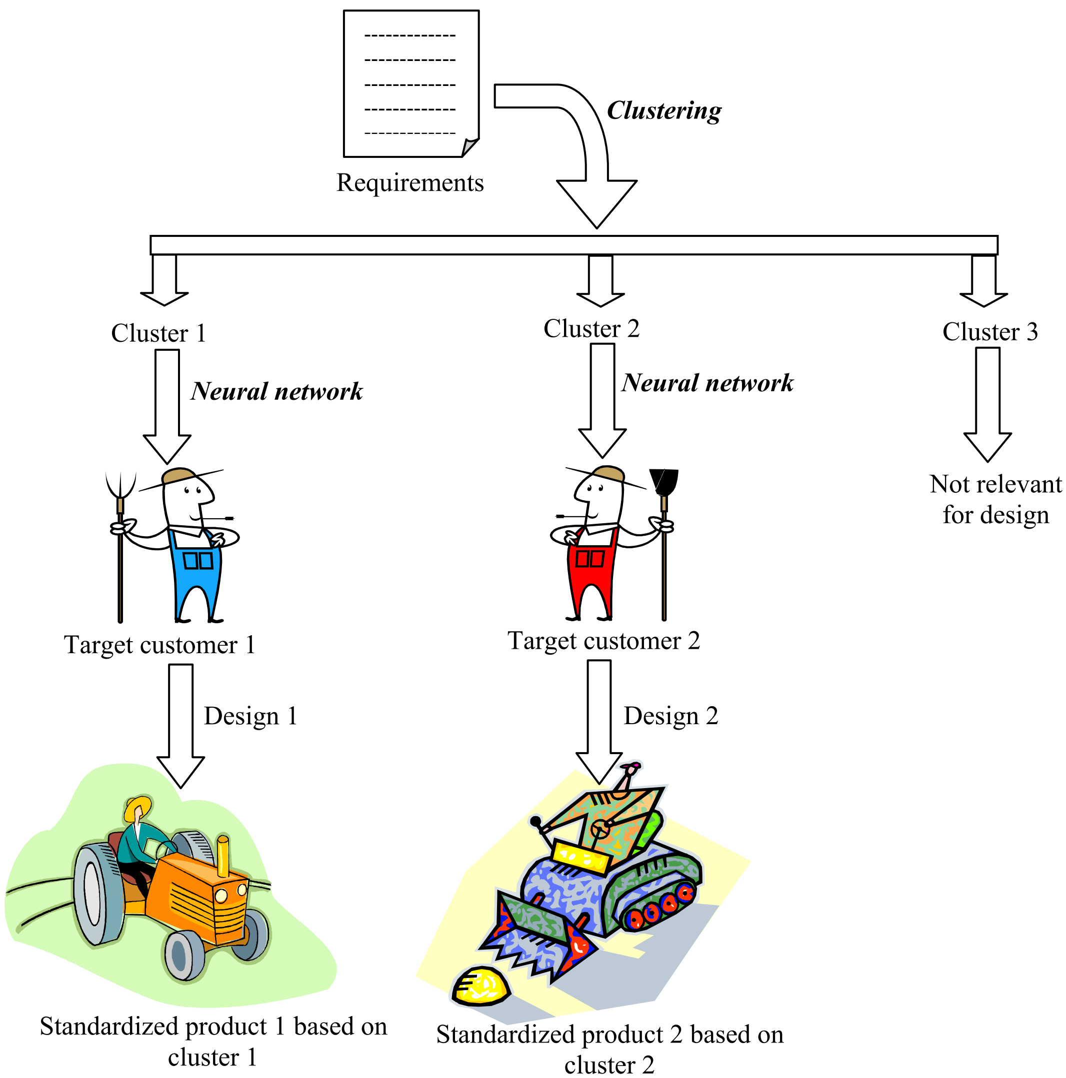

Figure 5. Design process for different standardized products

Applying clustering to Table 2 based on features F1,…, F6 for extracting three clusters; the following result is obtained (see Table 3):

Table 3. Three clusters of customers

		Customer Requirement					
		F1	F2	F3	F4	F5	F6
Cluster 1 (50%)	Customer 1	1		1	1		
	Customer 3	1	1		1		
	Customer 4	1		1		1	
	Customer 7	1		1	1		1
	Customer 8	1		1	1		
Cluster 2 (30%)	Customer 2		1		1	1	
	Customer 6		1			1	1
	Customer 10		1	1		1	
Cluster 3 (20%)	Customer 5	1	1	1	1	1	1
	Customer 9	1	1	1	1	1	1

The data in Table 3 shows that customers 1, 3, 4, 7 and 8 share similar behaviours expressed with features F1,..., F6. In the same way customers 2, 6 and 10 are similar.
The opportunities for standardization to be considered here are:

- Cluster 1: The design of a product with features F1, F3 and F4.
- Cluster 2: The design of a product with features F2 and F5.
- Cluster 3: The design of a product with features F1, F2, F3, F4, F5 and F6.

Table 3 also provides information about the size of each cluster, which may be useful in a cost-based evaluation. If Cluster 3 was too small to provide enough benefits, it would not be considered for the design of a product. The company may concentrate on the two other clusters or alternatively another iteration of the clustering algorithm could be considered. Some customers could be deleted from the clustering process. Different clusters could be tested and the resulting designs evaluated to select the best alternative.

Design a unique standardized product for selected customers

For a company that focuses on specific categories of customers for strategic reasons (fashion, reputation, test of new markets), it may be helpful to identify relevant sets of customers.
Using the customers' data in Table 2, classification separates the customers in different categories based on the relevant characteristics identified by the company. From the selected set of customers the company may learn their preferences and design a dedicated product.
For example, it is possible to differentiate the preferences of different categories of customers (young people, female, old men) and design specific products or adapt specific services (Figure 6).

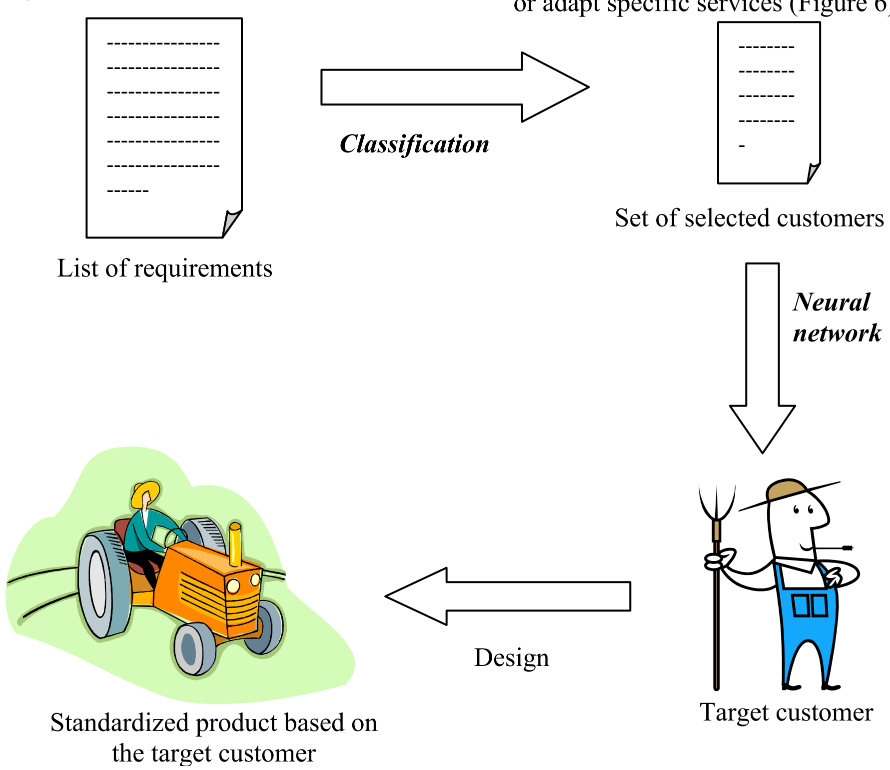

Figure 6. Design process of a standardized product for selected customers

4.3 Data mining in process standardization

A company manufacturing diverse parts could establish a database describing all previously manufactured parts, their characteristics, and the processes used for their manufacturing (see Table 4).

Table 4. Description of parts and the corresponding manufacturing process

	Part Description						Manufacturing Process					
	C1	C2	C3	C4	C5	C6	P1	P2	P3	P4	P5	P6
Part 1	1	1		1		1	M1	M4	M7	M2		
Part 2		1	1		1	1	M5	M2	M1	M4		
Part 3	1			1		1	M4	M3	M8			
Part 4		1			1		M2	M7				
Part 5		1	1		1		M5	M2	M1	M4		
Part 6		1		1		1	M4	M7	M2			
Part 7	1		1		1		M1	M4	M7	M5	M2	

The data in Table 4 shows that Part 1 has characteristics C1, C2, C4 and C6. The manufacturing process of this part is Machine 1, Machine 4, Machine 7 and Machine 2. Part 2 shares characteristics C2, C3, C5 and C6 and involves Machine 5, Machine 2, Machine 1 and Machine 4.

A data mining algorithm may link the description of parts to the manufacturing processes. The goal is to generate standardized processes for manufacturing new parts. The following methodology is proposed (see Figure 7):

1. Similar parts and processes are grouped into clusters.
2. Standardized processes are affected to each cluster.
3. When a new part is to be manufactured according to the identified cluster it impacts the standardized process.
4. The standardized process is adapted to the new part and the new data is incorporated into the database to enrich it.

Step 1 and 2 represent the learning stage. Step 3 exploits the knowledge extracted. Step 4 adjusts the knowledge provided for a new item and provides one more element in the database (Figure 7).

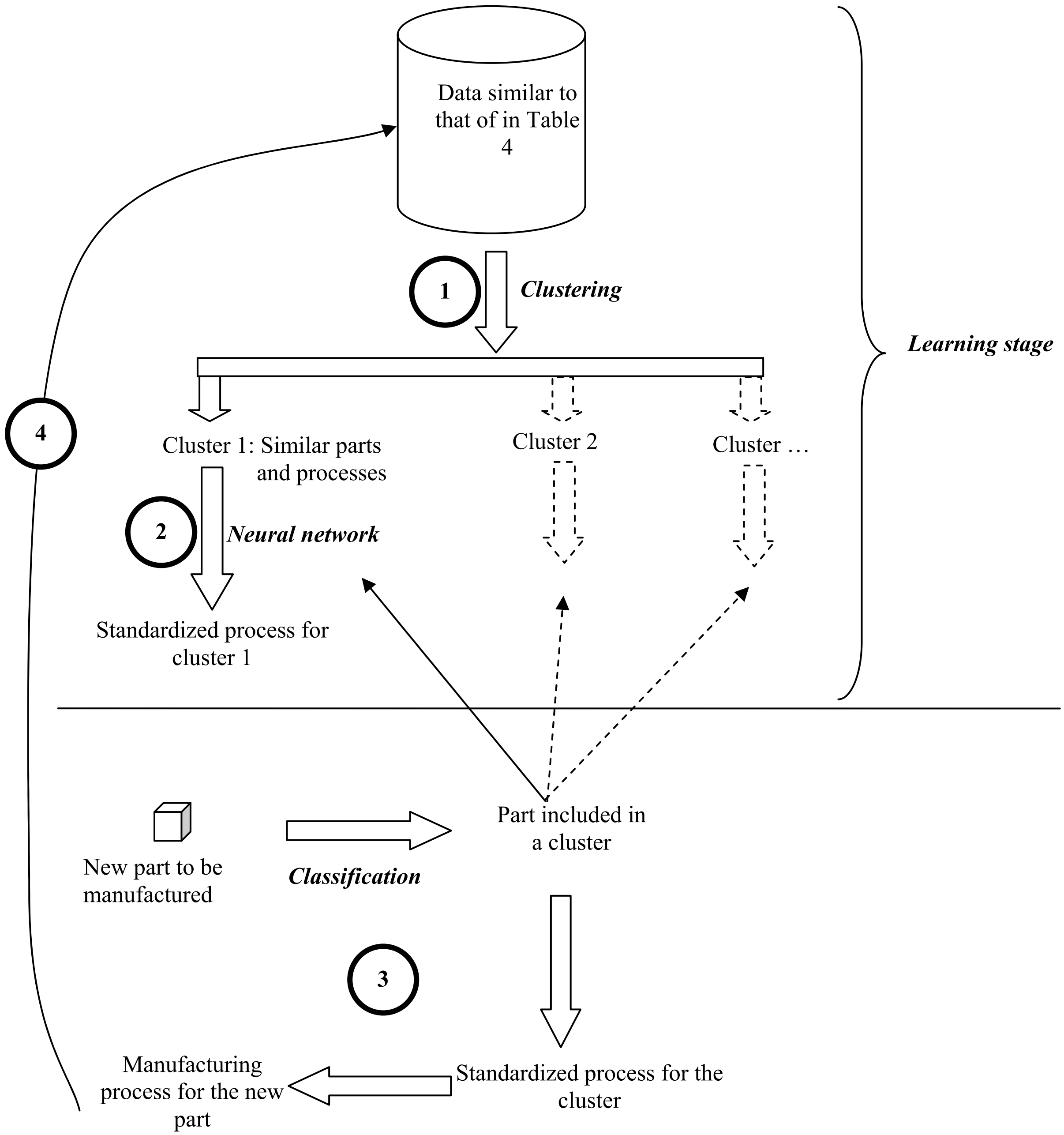

Figure 7. Process standardization with a data mining approach

The more complete the database is, the more exhaustive is the standardized process suggested in Step 3. For example, the database may contain the sequence of machines as well as manufacturing parameters (speed, temperatures, times, settings).

5. CONCLUSION

This paper discussed the application of data mining in standardization of components, products and processes. The standardization of components is realized through the extraction of association rules from the historical data representing customers' requirements.

Various options for the application of data mining in standardization of products were discussed. The design of a standardized product for a large population of customers was presented as well as the design of different standardized products. A data mining approach was used to select and prioritize a set of customers, e.g., a top-of-the-range product for high income customers.

Finally, the standardization of processes was considered for improving the efficiency of the manufacturing process to produce new parts.

This paper has illustrated that data mining is a useful tool for standardization. The existing industrial databases should be explored more systematically to extract new knowledge. The latter is key to the improvement of products, processes, and systems.

6. REFERENCES

1. Anand, S.S. and Büchner, A.G., *Decision Support Using Data Mining*, Financial Times Pitman Publishers, London, UK, 1998.
2. Berry, M.J.A. and Linoff, G., *Data Mining Techniques: For Marketing, Sales, and Customer Support*, New York: Wiley, 2004.
3. Büchner, A.G., Anand, S.S. and Hugues, J.G., *Data Mining in Manufacturing Environments: Goals, Techniques and Applications*, in Studies in Informatics and Control, Vol. 6, No 4, 1997, pp. 319-328.
4. Dupont, L., *La Gestion Industrielle : Concepts et Outils*, Hermès, Paris, 1998.
5. Dupont, L., Erol, M., Cormier, G. and Turkkan, N., *La standardisation des composants: modèles et algorithmes*, 3ème Congrès International de Génie Industriel, Montréal, Canada, May, 1999, pp. 671-680.
6. Erol, M., *Prise en compte de la flexibilité dans la planification dynamique*, PhD Thesis, Institut National Polytechnique de Grenoble, France, 1999.
7. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R., *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, Cambridge, MA, 1996.
8. Fouque, T., *A la recherche des produits flexibles*, Revue Française de Gestion, mars-avril-mai, 1999, pp. 80-87.
9. Gertosio, C. and Dussauchoy, A., *Knowledge discovery from industrial databases*, Journal of Intelligent Manufacturing, Vol. 15, 2004, pp. 29-37.
10. Huang, S.H., Hao, X. and Benjamin, M., *Automated knowledge acquisition for design and manufacturing: The case of micromachined atomizer*, Journal of Intelligent Manufacturing, Vol. 12, 2001, pp. 377-391.
11. Hyer, N. and Wemmerlöv, U., *Group technology and productivity*, Harvard Business Review, July-August, 1984, pp. 140-149.
12. Kusiak, A., *Decomposition in data mining: An industrial case study*, IEEE Transactions on Electronics Packaging Manufacturing, Vol. 23, No. 4, 2000, pp. 345-353.
13. Kusiak, A., *Rough set theory: A data mining tool for semiconductor manufacturing*, IEEE Transactions on Electronics Packaging Manufacturing, Vol. 24, No. 1, 2001, pp. 44-50.
14. Kusiak, A., *Feature Transformation Methods in Data Mining*, IEEE Transactions on Electronics Packaging Manufacturing, Vol. 24, No. 3, 2001, pp. 214-221.
15. Lee, H.L. and Tang, C.S., *Modeling the costs and benefits of delayed product differentiation*, Management Science, Vol. 43, No. 1, 1997, pp. 40-53.
16. Leu, S.S., Chen, C.N. and Chang, S.L., *Data mining for tunnel support stability: neural network approach*, Automation in Construction, Vol. 10, No. 4, 2001, pp. 429-441.
17. Pahl, G. and Beitz, W., *Engineering Design*, Berlin, Springer-Verlag, 1998.
18. Tarondeau, J.C., *Stratégie Industrielle: Second Edition*, Collection Gestion, Vuibert, 1998.
19. Thonemann, U.W. and Brandeau, M., *Optimal commonality in component design*, Operations Research, Vol. 48, No. 1, 2000, pp. 1-19.
20. Westphal, C. and Blaxton, T., *Data Mining Solutions*, John Wiley, New York, 1998.