18th International Conference on Production Research - ICPR18, University of Salerno, Fisciamo, Italy, July 31 – August 4, 2005.

IMPROVING MANUFACTURING QUALITY BY RE-SEQUENCING ASSEMBLY OPERATIONS: A DATA-MINING APPROACH

C. da Cunha[1], B. Agard[1], A. Kusiak[2]

[1]Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, C.P. 6079, succ. Centre-ville, Montréal (Québec), H3C3A7, Canada

{catherine.dacunha, bruno.agard}@polymtl.ca

[2]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

An assemble-to-order policy delays the final assembly operations of a product until a customer order is received. The quality of the final product is determined by the quality of the assembly operations. The data-mining approach presented in this paper uses information extracted from production history to determine the sequence of assemblies that minimize the risk of producing faulty products. The extracted knowledge plays important role in sequencing modules and forming product families that minimize the cost of non-quality. The concepts introduced in the paper are illustrated with numerical results.

Keywords: Assemble-to-order, quality, data mining, mass customization.

1 INTRODUCTION

Faulty products lead to unnecessary expense due to rework, repairing, recycling, and wasted time. “Zero fault” is an objective that industries are eager to reach. A variety of methods aim to achieve such goal, e.g., six sigma [1, 2] and total productive maintenance (TPM) [3]. Analysis of past performance of production systems is necessary. The difficulty is in finding pertinent information as the data is stored in numerous forms and at different locations. Data mining aims at extracting knowledge from large data sets. Using this new approach for improvement of production quality is quite natural. The goal of this paper is to emphasize the role of knowledge extraction in manufacturing quality in an assemble-to-order (ATO) context.

This paper is structured following six sections. First the background of the study is provided. Then the information use in assembly sequencing is addressed. The methodology used is described in Section 4. The paper concludes with computational results.

2 APPLICATION CONTEXT

2.1 Diversity: An industrial example

To meet the customers’ needs, product diversity tends to grow and therefore a management strategy is needed. The cost of offering a large portfolio should not exceed the gains obtained by satisfying the wide range of customer needs [4]. It is then essential to find the level of diversity that minimizes the total cost (Figure 1).

Diverse strategies could be considered in designing a product line. The major issue is to be able to offer large product diversity while managing a limited diversity of components, operations, and packaging. Different approaches have been used to address this challenge, e.g., design of product families, modular design and product delayed differentiation. Assemble-to-order is a policy that links modular design and product delayed differentiation. Indeed with this policy, modules are built from basic parts and stocked, the final assembly is done after an order has been confirmed.

The large apparent diversity for the customers is enabled by a combinatorial association of basic parts.

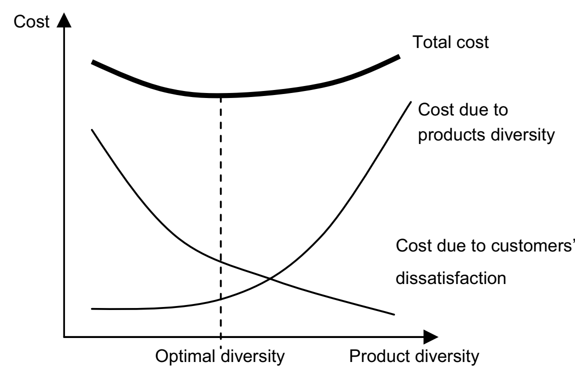


Figure 1: Diversity costs (Tarondeau [5]).

Surprisingly, the major part of diversity is not visible to the customers. It is actually created by the evolution of components (changes in technology) or the creation of new versions (upgrades).

In this paper, an industrial example of the electrical wire harness is discussed. This product possesses many of the previously described characteristics. Indeed, it is a major component of a vehicle. This set of wires and connectors transmits electricity and information between different devices all over the car (see Figure 2).

The functions (airbag, electrical windows, headlights’ control, etc.) are performed by combination of different wires and connectors. To illustrate the diversity of this product, consider a standard wire harness in a middle range car. This wire harness performs 15 different functions. Depending on the silhouette and the motor, these functions appear in different versions (up to 9). Potential diversity is then about 7 millions of different wire harnesses for a unique car model [6].

In addition, there are inclusive and exclusive relations between the functions, e.g., the function “passenger air-bag” requires the function “driver air-bag”. Those relations reduce the actual diversity.

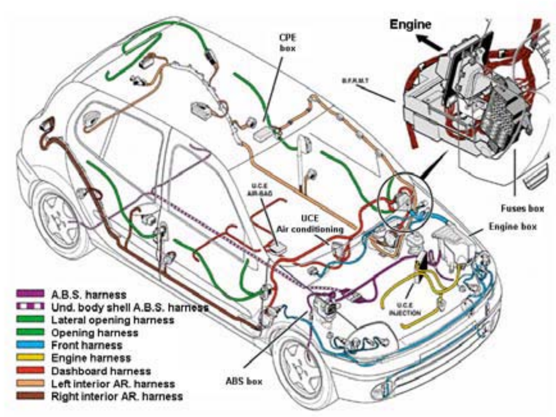


Figure 2: Wire harnesses in a car.

2.2 Costs

Evaluating the diversity cost is a problematic task. Even if direct costs, like investment in new equipment or material costs, can be measured, indirect costs are difficult to estimate. Martin and Ishii [7] proposed metrics to compare design alternatives based on the costs they induce, however this evaluation is difficult to perform for industrial cases.

Product quality can be impacted by its complexity. As the workers have to perform more tasks, the number of errors may increase. The negative impact of product complexity seems unavoidable, unless the diversity is controlled.

McDuffie *et al.* [8] presented results of an international study in automotive industry. This statistical study stressed the relationship between product diversity, productivity, and quality. This analysis indicates that when plants are adequately equipped to manage diversity, the scope of the product mix does not have a large impact on the productivity.

Similar to the diversity cost, which is difficult to evaluate, savings due to process redesign are not easy to quantify. Actually, there are criteria (e.g., reduction of the time-to-market or flexibility improvement) for which a non-qualitative evaluation is not recommended. Furthermore, if cost savings due to quality improvements can be partially measured, in terms of reduction of the mean assembly time, reduction of needed materials, the savings (or even gains) due to the improvement of the products' image for the customers can not be directly evaluated.

2.3 Diversity management: Modularity and ATO

A postponement strategy aims at reducing the risk associated with product diversity. It uses similarity between objects to delay their differentiation [9, 10].

The modularity concept has been used in different areas to manage diversity. Modular production is defined by the APICS as the capacity to design and product sets of modules that can be combined in a maximal numbers of ways [11].

The choice of a modular design implies a rethinking about the design process within the company [12]. The modules created can be independent or not (i.e., they can be assembled without requiring another module or not). Figure 3 illustrates this concept. Actually Figure 3 (a) shows modules that are independent: modules 1 and 3 can not be assembled unless module 2 is already installed. On the contrary the connections of the modules described in Figure 3 (b) are such that any module can be assembled with any other one.

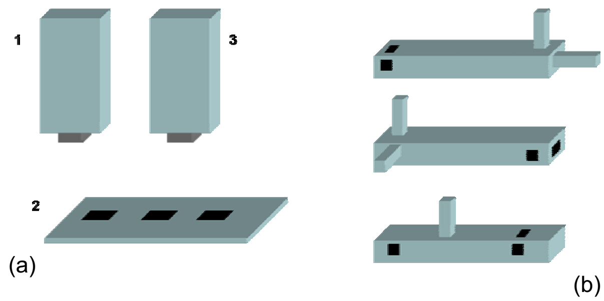


Figure 3: Example of modules compatibility.

One of the advantages of having independent components is that re-sequencing of the assembly sequence can be done without having to re-design the different modules.

The example considered in Section 5 is constituted by independent modules.

3 QUALITY ANALYSIS

3.1 Fault identification

The links between information system, control system and management control exist irrespective of production policy. Nevertheless, the lack of previsions (workload, production volume, etc.) induced by the awaiting of the customers' real demand endangers the relations between production and control. Performances have to be measured *ex-post*. The role of the control system is then to analyze the causes of non-respect of the performances, the objective being to improve them by modifying the existent organization.

A product is considered as non-quality when its characteristics do not meet the specifications defined by the designers.

A necessary first step is to identify the non-quality. This can be done at the end of the assembly process. Redoing the finished product will have an adverse economic effect.

If inspection was carried throughout the assembly process, non-quality would be detected sooner. Thus, the rework cost would decrease.

Nevertheless, inspection is expensive and it may not be possible to test the product after each assembly operation. It is then essential to carefully determine the location and timing of tests.

3.2 Data mining applications

Anand and Büchner [13] defined data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets. Data-mining algorithms have been applied in many different areas such as marketing [14], medicine (identification of genes impacting cure/drug development [15]), and industrial design [16].

The patterns extracted from production data can assist production managers following the assemble-to-order principle.

However, the difficulty is to find a data representation that enables identifying interesting patterns. This representation should also be understandable for the data miner and the field expert, so that the results could be interpreted.

4 GOAL AND METHODS

The main goal of this study is to apply data mining to improve production quality in an assembly line. The challenge is in the extraction of associations between non-quality and assembly sequence in presence of noise. Actually, non-quality may come from sources other than the operation sequence, for example, power outage or non-quality of raw material. Furthermore, random phenomena can also be responsible of fault.

When faulty patterns are identified, it becomes then possible to re-sequence the assembly operation or to rethink the tests' policy [17] in order to reduce the number of operations to be redone. It is therefore important to keep in mind that operations of rework are more complicated (i.e. they last longer and cost more) when they applied on a more complex product.

The method used is described in the following steps:

1. Identification of assembly sequences having an impact on quality
2. Generation of a new sequence
3. Generation of a new test policy

5 COMPUTATIONAL RESULTS

5.1 Methodology

The data-mining approach to be used in this research was prototyped on a randomly generated data set. The data was randomly generated with respect to the following constraints:

- An operation can be performed as a normal task (i.e., non-rework task) at most once per product

$$\forall i \in [1;6]\ \forall j,k \in [1;6]$$
$$(task_j = i \text{ and } task_k = i) \Rightarrow (j = k)$$

- To reproduce the behavior of a real production system, random faults are generated. Because of a randomly phenomenon, 5% of the products need rework. When the first test proved the product to be non-quality, rework has to be done.
- Moreover the rework operation can be not sufficient to reach the quality characteristics; in this case the product will be considered as faulty and destroyed.

Furthermore, systematic faulty sequences are considered. The challenge of the data mining process will be to extract those patterns.

The two always faulty operation sequences are:

- If operation 5 is not the last task, this operation has to be redone
- If operation 2 precedes operation 4, operation 4 has to be redone

5.2 Input data

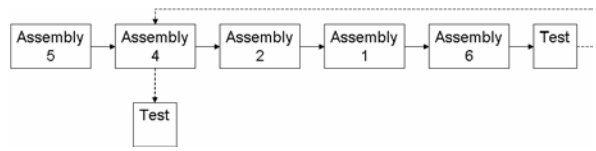
Consider here a unique example, which consists of 6 assembly operations. This instance includes 250000[1] product's operation routes. The data store in the production history records are the operations' sequence. It encloses assembly operations as well as quality tests. Depending on the results of those tests some assembly operations could have to be redone.

An example of a product route would be:

5-4-2-1-6-Tf-4-Tok

This route (represented by Figure 4) tells us that this product after assembly operations 5-4-2-1-6 was detected as faulty (Tf). It was then necessary to redo assembly task 4 (4). The final test proved the product to conform to quality standards (Tok).

Figure 4: Example of a product route.



The historical data is represented in a tabular form, where each row represents the route of a product. An example of such a route is stored as the first row (in bold) of Table 1.

Table 1: Input data.

Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Test	Rework	Final test
5	**4**	**2**	**1**	**6**		**Tf**	**4**	**Tok**
1	4	5				Tok		
6	2	5	3			Tf	5	Tok
2	1	4				Tok		
2	4					Tf	4	Tok

The first 6 columns in Table 1 represent different assembly operations needed for the product. The sequence of the operations is indicated by the sequence of columns'. Column 7 represents the status of the first quality test, column 8 the possible rework operation and the last column the final quality status of the product.

5.3 Data mining implementation

The freeware TANAGRA[2] software was chosen as the knowledge extraction tool. Agrawal's [18] algorithm was used to find the association rules. The cross-tabulation parameters describe the contingence of variable pairs; obtained following Agresti's method [19].

One of the challenges of data mining is to manage the difficulty of handling different types of data [20].

Here the difficulty was to find a technique to identify the sequences causing non-quality. Therefore, the structure of the input data has to be modified. The frequency of faulty products being low, the support of the rules can be rather low too. Therefore it can be interesting to study two sets of data: one with all routes and one describing only faulty products' routes. The first set allows identifying sequences conducting to quality products and the second to understand the reasons of faults. The data of Table 1 are then separated in two tables (Table 2 and Table 3).

[1] TANAGRA's limit (see Section 5.3).

[2] http://eric.univ-lyon2.fr/~ricco/tanagra/index.html

Table 2: Quality products.

Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Test	Rework	Final test
1	4	5				Tok		
2	1	4				Tok		

Patterns extracted from Table 2 could be used to improve the motivation of the assembly line workers. For example, if an operation never needs a rework, the workers performing it could be awarded.

Nevertheless, this article focused on the identification of non-quality patterns. Therefore in the following section Table 3 will be considered for further analysis.

Table 3: Faulty products.

Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Test	Rework	Final test
5	**4**	**2**	**1**	**6**		**Tf**	**4**	**Tok**
6	2	5	3			Tf	5	Tok
2	4					Tf	4	Tok

Sequences such as: 1-4-2 and 4-2-5 should be recognized for sharing a sub-sequence of 2 operations. This recognition is not possible when dealing directly with the data presented as in Table 1, Table 2 or Table 3.

To perform this identification, a different structure for the routes is required.

The example sequence **5-4-2-1-6-Tf-4-Tok** represented in the first line of Table 3, is then represented by the set **{B5, 5_4, 4_2, 2_1, 1_6, 6F, R4}**. The precedence information is contained in the data itself. The routes can then be encoded as binaries data. Table 4 represents the same information as in Table 3. Column B1 states that operation 1 is the first task (beginning of the assembly process), respectively column B5 states that operation 5 is the first task. Column 1F states that operation 1 is the final task. The latest column rework represents the rework operations, when the product is: R4 states that operation 4 has to be redone.

Note that the fact that the first test indicates that the product is faulty, is included in the route in Table 4 containing faulty products routes.

Table 4: New form of faulty product routes.

B1	B2	B3	B4	B5	B6	1_2	1_3	1_4	1_5	1_6	1F	2_1	2_3	2_4	2_5	2_6	2F	3_1	…	3F	4_1	4_2	…	4F	…	5_3	5_4	…	6_2	…	6F	Rework
				1						**1**		**1**										**1**					**1**				**1**	**4**
					1										1					1						1			1			5
	1													1										1								4

A post-processing of the extracted rules is required to eliminate redundancies. Furthermore, as the quality of process is of interest, only the rules with the “test” as an outcome were considered.

It is common in data mining that human expertise is needed to assess and validate the extracted knowledge. The latter is possible due to the world knowledge [21] acquired through experience. Thus an expert may be able to detect rules that can not be used out of context.

Consider the following two rules:

Rule a: Any product requiring operation 6 is reworked more often than any other product.

Rule b: Operation 6 never needs rework.

An expert could interpret these rules using her/his perception-based information: The worker performing operation 6 is highly skilled; however, he distracts his co-workers by singing when working.

5.4 Results: Extracted rules

Association rules

Data mining algorithms identify patterns in the data. These results validate the representation chosen for the routes in Section 5.3.

The rules extracted have the format shown in Table 5.

A human expert could make the rules more comprehensive, e.g.:

Rules 2, 3, 4, 5, and 6 have similar meaning: If operation 5 is not the last task, it needs to be redone.

Rule 1: If operation 2 precedes operation 4, the product has to be reworked.

Rule 11: If operation 2 is the only operation, it needs to be redone.

Table 5: Extracted rules.

No.	Antecedent	Consequent	Support	Confidence
1	"2_4=1"	"rework=4"	0,165	0,961
2	"5_4=1"	"rework=5"	0,217	1
3	"5_1=1"	"rework=5"	0,162	0,98
4	"5_3=1"	"rework=5"	0,158	0,979
5	"B5=1"	"rework=5"	0,368	0,948
6	"5_2=1"	"rework=5"	0,185	0,895
7	"B2=1"	"rework=5"	0,195	0,886
8	"2_4=0"	"rework=5"	0,725	0,877
9	"1F=1"	"rework=5"	0,182	0,832
10	"3F=1"	"rework=5"	0,163	0,817
11	"2F=1" and "B2=1"	"rework=2"	0,012	1

The rules with confidence lower than 1 may unveil patterns that are not always true. This is important when improving the quality of a system not to be limited to the search for systematic failures but also to look for sources of non-quality. Indeed, quality faults often have multiple sources and finding the exact cause requires a synthesis of information residing at various data bases. “Partial truth” may be the source of potential improvements.

Cross Tabular

Consider an example where a data miner finds two major links between rework operations and the assembly operations. The pertinence of the rules was evaluated using the Tschuprow's T indicator[3] as shown in Table 6.

Table 6: Summary results.

Row	Column	Tschuprow's T	Cross-tab
Rework	2_4	0,717	Rework = 4 and 2_4 = 1 99 occurrences
Rework	6F	0,657	Rework = 6 and 6F = 1 1 occurrence

Contingence links were found between rework of operation 4 and the operations sequence **operation 2-operation 4.** This rule validates the associate rule 1 listed in Section 5.1.

The source of the rework of operation 6 was also clarified as operation 6 is performed at the end of the assembly line. Nevertheless, this information should be handled carefully because of the rare frequency of this event.

[3]Tschuprow's T varies between 0 and 1, T = 0 states the independence (in the mathematical sense) of the 2 variables, T = 1 states that the link between the variables explicates the whole observation.

5.5 Analysis and decisions

The rules can be used to determine a new sequence of assembly operations. For example, **Rule 1** indicates re-sequencing of operations 2 and 4:

The product requiring operations 2 and 4 should always pass through operation 4 first.

Rules 2, 3, 4, 5, and 6 point to re-sequencing: operation 5 to be done at last.

The incorporation of **Rule 11** will not lead to re-sequencing of operations, however, this operation should be dealt with as it often leads to non-quality, e.g., different tools could be used.

Introducing an additional test operation after the second task could be considered. Products needing rework would be detected sooner and the rework would be less costly.

Of course, any reorganization of the process has its consequences, and to predict them it is not easy. Therefore the whole process of data-analysis and system reorganization should be on a regular basis to determine the best configuration (Figure 5).

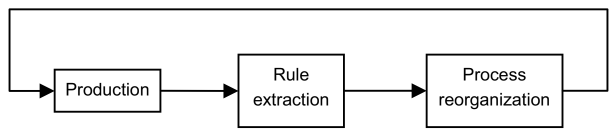


Figure 5: Improvement scheme.

Furthermore, production data could be temporal, e.g. due to seasons. Therefore, it is critical to consider a possible link between production quality and time period.

In order to integrate this characteristic, data, other than operations and quality status should be considered, e.g. studying the impact of the work shift could lead to a redesign of break schedule (e.g. 2 breaks of 15 minutes instead of 1 break of 30 minutes).

6 CONCLUSION

This paper discussed the application of data mining for improvement of manufacturing quality of assembly operations. The data considered in this research included random events that occur in production systems. The computational results confirmed that sources of faults can be detected with association rules, even in presence of noise. The rules extracted with data-mining algorithms can be used to improve production quality.

The selection of an assemble-to-order policy is not limited to technical solutions. Such a change in production policy has to be completed with organizational rethinking [22].

Further research should consider operation sequences and other non-quality sources including human factors and material.

REFERENCES

[1] Harry M., Schroeder R., Linsenmann D.R., 2000, Six Sigma: The Breakthrough Management Strategy Revolutionizing the World’s Top Corporations, Currency, New York, NY.

[2] Breyfogle F.W., 1999, Implementing Six Sigma: Smarter Solutions Using Statistical Methods, John Wiley, New York, NY.

[3] Nakajima S., 1988, Introduction to TPM: Total Productive Maintenance, Productivity Press, Cambridge, MA.

[4] Child P., Diederichs R., Sanders F.-H., Wisniowski S., 1991, The management of complexity, Sloan Management Review, Vol. 33, No. 1, pp. 73-80.

[5] Tarondeau J-C., 1998, Stratégie industrielle, 2nd Edition, Vuibert, France.

[6] Agard B., Tollenaere M., 2002, Design of wire harnesses for mass customization, 4th international conference on integrated design and manufacturing in technical engineering (IDMME 2002), Clermont-Ferrand, France.

[7] Martin M., Ishii K., 1997, Design for Variety: Development of Complexity Indices and Design Charts, ASME Design Engineering Technical Conferences, DETC.

[8] MacDuffie J., Sethuraman K., Fisher M., 1996, Product variety and manufacturing performance: evidence from the international automotive assembly plant study, Management Science, Vol. 42, No. 3, pp. 350-369.

[9] Aviv Y., Federgruen A., 1999, The Benefits of Design for Postponement, Chap. 19 in Quantitative Models for Supply Chain Management, pp. 553-584. Tayur S., Ganeshan R., and Magazine M. (eds), Kluwer Academic Publishers, Boston, MA.

[10] Zinn W., 1990, Should you assemble products before an order is received? Business Horizons, No. 5, pp. 70-73.

[11] APICS, 1998, APICS Dictionary, 9th Edition, Falls Church, VA.

[12] Kusiak A., 1999, Engineering Design: Products, Processes, and Systems, Academic Press, San Diego, CA.

[13] Anand S., Büchner A., 1998, Decision Support Using Data Mining, Financial Times Pitman Publishers, London, UK.

[14] Berry M., Linoff G., 1997, Data Mining Techniques: For Marketing, Sales, and Customer Support, Wiley, New York, NY.

[15] Shah S., Kusiak A., 2004, Data Mining and Genetic Programming Based Gene/SNP Selection, Artificial Intelligence in Medicine, Vol. 31, No. 3, pp. 183-258.

[16] Agard B., Kusiak A., 2004, A Data-Mining Based Methodology for the Design of Product Families, International Journal of Production Research, Vol. 42, No. 15, pp. 2955-2969.

[17] Kusiak A., Huang C.C., 1997, Design of Modular Digital Circuits for Testability, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, Vol. 20, No. 1, pp. 48-57.

[18] Agrawal R., Srikant R., 1994, Fast algorithms for mining association rules in large databases, in Proc. International Conference in Large Databases, pp. 478-499.

[19] Agresti, A., 1990, Categorical Data Analysis, Wiley, New York, NY.

[20] Chen M.-S., Han J., Yu P., 1996, Data mining: An overview from a database perspective, IEEE transactions on knowledge and data engineering Vol. 8, No. 6, pp. 866-883.

[21] Craven M., DiPasquo D., Freitag D., McCallum A., Mitchell T., Nigam K., Slattery S., 2000, Learning to construct knowledge bases for the world wide web, Artificial Intelligence, Vol. 118, Nos 1-2, pp. 69-113.

[22] Benson P., Saraph J., Schroeder R., 1991, The effects of organizational context on quality management: an empirical investigation, Management Science, Vol. 37, No. 9, pp. 1107–1124.