

# A Case Investigation of Product Structure Complexity in Mass Customization Using a Data Mining Approach

Peter Nielsen, Thomas D. Brunoe and Kjeld Nielsen

**Abstract** This paper presents a data mining method for analyzing historical configuration data providing a number of opportunities for improving mass customization capabilities. The overall objective of this paper is to investigate how specific quantitative analyses, more specifically the association rule Apriori, can support the development within the three fundamental mass customization capabilities. The results of the Apriori analysis can be utilized for improving the configuration process by introducing soft constraints and consolidating the product structure by joining components or modules and finally for improving production planning and control.

**Keywords** Mass customization capabilities · Data mining product architecture · Apriori

## 1 Introduction

In any company, it is essential to offer products which match the needs and desires of customers to achieve sales and profit. This is true for mass producers as well as mass customizers; however, in mass customization, this issue is somewhat more complex than mass production due to a much higher variety and a more complex product structure. As pointed out by Salvador et al., mass customizers need three fundamental capabilities to be successful: (1) solution space development—identifying the attributes along which customer needs diverge, (2) robust process design—reusing or recombining existing organizational and value chain resources

---

P. Nielsen · T. D. Brunoe (✉) · K. Nielsen  
Department of Mechanical and Manufacturing Engineering, Aalborg University,  
Fibigerstraede 16, DK 9220 Aalborg, Denmark  
e-mail: tdp@m-tech.aau.dk

P. Nielsen  
e-mail: pn@m-tech.aau.dk  
URL: <http://www.m-tech.aau.dk>

to fulfill a stream of differentiated customer needs, and (3) choice navigation—supporting customers in identifying their own solutions while minimizing complexity and the burden of choice [1, 2].

To support companies in developing their capabilities as mass customizers, some research has focused on assessing the fundamental capabilities, to evaluate within which areas the companies should strengthen their efforts. All three capabilities relate strongly to the product variety, which is also the main element which differentiates mass customization from other business strategies. This implies that tools and methods for continuously assessing, adjusting, and communicating the product variety are needed in order to improve the three capabilities. Mass customizers often utilize product configuration software for implementing a choice navigation process. A product configuration is a piece of software which allows customers or sales people to configure a product from a set of predefined variety. During a configuration process, a large amount of data are generated. These data can be utilized for different kinds of analyses with the purpose of improving performance within the three fundamental capabilities; Salvador et al. [3] address this specific as one of several approaches to achieve mass customizing capabilities. The data generated by a product configuration may include the following kind of data:

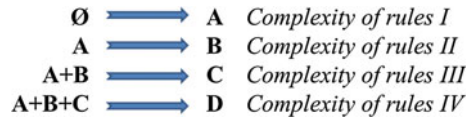
- Information about the customer
- Selected product options, e.g., parametric dimensions, optional modules, colors, and functional requirements.
- BOM information, i.e., specific components for manufacturing the product and quantities
- Sales process and manufacturing costs
- Lead times, quality data, etc.

The overall objective of this paper is to investigate how specific quantitative analyses, more specifically the association rule Apriori, can support the development within the three fundamental mass customization capabilities.

## 2 State of the Art

Data mining and association rules are a well-known field and have seen some application in the area of mass customization [4]. The data mining methods applied in this research have similarly been used in the domain of mass customization by among others: Geng et al. [5], Hong et al. [6], and Zhou et al. [7]. However, this paper contains two novel contributions compared with the current state of research. First, it presents a specific case study where the number of association rules is studied as a function of the support and confidence levels chosen. Second, it suggests a method to exploit this knowledge to choose (in the specific case) a reasonable combination of the complexity of association rules and support and confidence levels.

**Fig. 1** An illustration of the complexity of the mining rules and their definition in this paper



### 3 Method

The paper uses the association rule Apriori, which is a widely used data mining rule [8]. The Apriori association rule is based on determining two parameters: confidence and support. Support is “the ratio of the number of transactions that contain the item set to the total number of transactions” [8], so support is the indicator for the frequency with which a certain combination occurs.

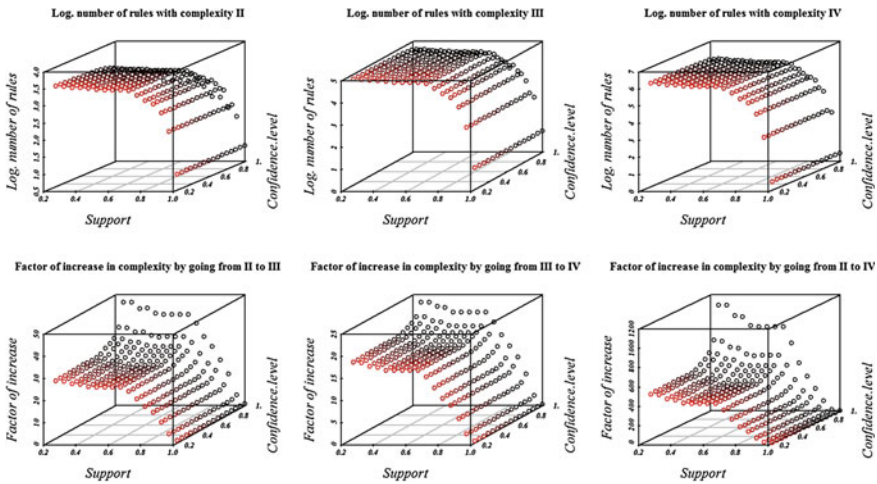
Confidence is the support for a given combination (e.g.,  $A \rightarrow B$ ) divided by the number of occurrences of  $A$ . So high support indicates a frequent occurrence of a given combination ( $A + B$ ), while high confidence indicates that, e.g.,  $A$  often is found with  $B$ . So the confidence becomes a proxy for the likelihood of observing  $B$  given  $A$ .

Figure 1 illustrates the different levels of complexity of association rules defined and addressed in this research. Where the most simple association rule assumes that lack of input leads to  $A$  occurring, the next level assumes that if  $A$  occurs,  $B$  occurs with a given support and confidence and so forth. Note that links are unidirectional, and thus,  $A$  leads to  $B$  does not imply that  $B$  leads to  $A$ . In the context of mass customization, these complexities can be directly related to the configuration choices or bill-of-material of the configured products. As an example assume that customers choosing to use component  $A$  in their configuration also chooses with some confidence and support to include component  $B$ . Of these complexities of rules II–IV are investigated in this research to limit the scope.

This paper investigates two issues:

1. How the complexity of the association rules influences the number of rules that can/should be taken into account.
2. How the support and confidence influence the number of rules at a given complexity level of associations and compared to a higher level of complexity, i.e., investigate what is driving increases in number of rules, confidence levels, and/or support levels.

The resulting method investigates how complex a solution space is and how difficult it can be to, e.g., guide a customer through the customization process. The first point is investigated through identification of the number of additional rules created through adding a level of complexity at given level of confidence and support. The second point is investigated through ANOVA with confidence and support levels as independent variables and the number of rules and the factorial increase in the number of rules as dependent variables.



**Fig. 2** First row: the logarithmic increase in the number of rules when going from one complexity level of rules to the next. Second row: factorial increase in complexity from one complexity level to another

The experimental setup is as follows: Confidence and support are varied from 0.25 to 1.00 in steps of 0.05, and all combinations of confidence and support levels for rule complexity II–IV are investigated. This gives  $16 \times 16 \times 3$  (16 confidence levels  $\times$  16 support levels  $\times$  3 complexity levels) = 768 investigations that are carried out.

## 4 Results

The studied case contains 180 unique orders with 178 bill-of-material parameters that vary in inclusion/exclusion of a configured product. BOM items always included have been removed from the study, so that only BOM items that are actually configured are included. The method is implemented and tested in the open source software R using the package *arules* [9].

The results of these investigations on the case data are shown in Fig. 2. From Fig. 2, it is shown that the number of rules increases from in thousands, to the hundreds of thousands and to millions as one goes from a complexity of rules of II, to III and III to IV, respectively. A simple experiment of adding a further level of complexity to the rules indicates that tens of millions of additional rules are created by adding this complexity. This underlines that the even a limited number of orders can contain a very large number of association rules, when the complexity of these rules increases from simple rules (i.e.,  $A \rightarrow B$ ) to more complicated rules.

By using ANOVA analysis, treat support and confidence levels as independent variables and the logarithmic number of rules (due to exponential behavior of the

**Table 1** Overview of best ANOVA models excluding any non-significant variables

	Impact on complexity level II		Impact on complexity level III		Impact on complexity level IV	
	Estimate	Pr (> t )	Estimate	Pr (> t )	Estimate	Pr (> t )
Intercept	4.53	0.000	6.32	0.000	8.24	0.000
Support level	-1.99	0.000	-3.29	0.000	-4.75	0.000
Confidence level	-0.39	0.001	-	-	-	-
$R^2$		0.53		0.55		0.55
Adj. $R^2$		0.52		0.54		0.55
	From to II to III in rule complexity		From to III to IV in rule complexity		From to II to IV in rule complexity	
	Estimate	Pr (> t )	Estimate	Pr (> t )	Estimate	Pr (> t )
Intercept	38.30	0.0000	26.03	0.0000	703.68	0.0000
Support level	-36.35	0.0000	-23.20	0.0000	-824.22	0.0000
Confidence level	9.61	0.0000	2.85	0.0001	252.18	0.0000
$R^2$		0.76		0.80		0.78
Adj. $R^2$		0.76		0.80		0.78

number of rules) and the factorial increase in the number of rules as dependent variables. The best fitted models are given in Table 1.

It should be noted that in the case of the factorial increase from going from II to III and from II to IV in the complexity of the rules, the interaction between support and confidence levels is in fact significant. However, removing the interaction only lowers  $R^2$  from 0.77 to 0.76, so the interaction has limited explanatory value and has been excluded. From the  $R^2$  and adjusted  $R^2$  values shown in Table 1, it is clear that a large part of variation in the factorial increase in the number of rules and the actual logarithmic values of the number of rules can be explained by a combination of the support and confidence levels.

For all three levels of complexity of association rules, the support level is the most significant variable in determining the number of rules created by using the association rules. In general, the analysis shows that the higher the support (i.e., the higher the frequency of occurrence), the lower the number of rules. This seems intuitively correct and can indicate a number of issues for the case. First, as the number of association rules is quite low when support is 0.90 or above (for II-IV, respectively, 117-545, 1,675-4,987, and 11,886-26,615, depending on the confidence level), the number of fixed BOM combinations with high use frequencies is low (taking 178 BOM items into account). Second, when support levels are low (0.25), the number of rules increases dramatically (for II-IV, respectively, 904-3,322, 42,315-92,583, and 989,015-1,676,487, depending on the confidence level). This implies that there are in fact a very large number of BOM combinations that are frequently used (even with confidence level 1.00) and implies a large degree of dependence inside the BOM structure in the particular case. For the complexity of association rules at II, the confidence level is also significant in explaining the number of rules. However, removing the confidence level only

lowers  $R^2$  from 0.53 to 0.52, so like the interaction discussed previously, it can be discounted from the discussion.

Of equal interest is how the number of association rules increases (for a fixed support and confidence level) when the complexity of the association rules increases. Interestingly enough, this depends both on the support level and on the confidence level, though again with the largest contribution from the support level. The  $R^2$  values are *c.* 0.80 for the fitted response models, indicating a strong explanatory value of support and confidence levels. This is not necessarily intuitive. However, it implies that when increasing the complexity of the association rules to investigate the BOM dependencies, the support and confidence levels are non-trivial. Specifically, in the particular case, the response models imply that low support levels tend to have high increases in the number of rules when the complexity of the association rules is increased. In the sense of managing a solution space, the number of rules to consider is thus very much higher if the support levels are low. This could also indicate for the case that there are a large number of constraints in the solution space for combinations that are seldom used.

## 5 Implications

The knowledge obtained from the Apriori analysis can be utilized in a number of different ways. In this section, it is described for each mass customization capability, how utilization of the results can benefit mass customizers.

### 5.1 *Solution Space Development*

Solution space development concerns the identification and development of product variety. This implies also to revise a company's current product portfolio in order to consolidate it if necessary over time. This is typically done by removing unnecessary components or modules, which are seldom sold or which have a function that can be incorporated into other modules or components. This will generally imply lower manufacturing costs, similar to general design for manufacturing principles [10]. In this context, the results of the Apriori analysis could be utilized to identify candidates for combining two modules into one module. Often, mass-customized products are modular, and the modules are used to accommodate product variety by allowing different variants of a certain module type. However, this variety comes at a cost and reducing the number of module variants will usually imply lower manufacturing costs. The results of the Apriori analysis will indicate which modules are often sold together. If certain modules are always sold together, it would be natural to consider joining these two modules. There are, however, other considerations which could limit the possibilities of joining two modules, e.g., if the two modules are supplied by different suppliers,

utilize different manufacturing technologies, have different life cycles, etc., which needs to be taken into account. If the module, however, could be combined, this could imply a faster assembly process, as fewer modules would need to be assembled, lower fabrication cost of modules, a simpler product structure with lower administration costs as well as a simpler product family model and a simpler configuration system.

## ***5.2 Robust Process Design***

Another natural approach to using the information gained from the Apriori method would be to use this to design both planning processes and inventory management. In production planning [11], the typical approach is to batch similar products/components for planning purposes [12]. However, in mass customization, this is typically one of the main planning challenges, as there are by definition no standard products to group for production, so products are either made-to-order or assembled-to-order from a central inventory of components/modules. Standard inventory management would then imply that components/modules are grouped based on their individual characteristics (price, demand profile, lead time, supplier, etc.) [13]. However, this implies that items can be managed individually. The study presented in this paper clearly concludes that a large number of components directly or indirectly are always sold together. This strongly implies that both the inventory management approach and the subsequent planning approach need to take these dependencies into account.

## ***5.3 Choice Navigation***

The capability choice navigation is defined by Salvador et al. [3] as “Support customers in identifying their own solutions while minimizing complexity and the burden of choice.” Hence, this capability is related primarily to the capabilities of the configuration system and its ability to configure a variety of products. The ideal product configuration system should, after a customer has finished a configuration, leave the customer with the experience that the process has not been unnecessarily difficult to perform and the customer has been able to match his or her needs exactly to a specific configuration of a product [3].

The knowledge obtained from the Apriori analysis can be applied almost directly to improve the product configuration process. If it is determined that customers nearly always select component B given component A, it would be obvious to introduce a “soft constraint” in the product configuration, which would imply that once a customer selects component A, then component B is automatically chosen as default. Contrary to a hard constraint, a soft constraint would allow the customer to choose a different component than B following this.

However, selecting B automatically would imply less effort from the customer for performing the whole configuration process, yet leaving the flexibility to alter the automated selection.

The introduction of soft constraints should, however, only be made if there is a high confidence for a certain rule, i.e., component B is almost always chosen given component A. If the confidence is lower, a different approach could be taken. The Apriori analysis may in some cases indicate that if a component A is selected, then components B, C, or D are chosen equally frequent; however, other components may also be chosen. In this case, in an actual configuration, the customer is usually presented with a number of different components to choose from (e.g., B, C, D, E, F, and G). However, if this list is sorted according to confidence, then the most likely component would be on top of the list and the least likely in the bottom. This would, like the introduction of soft constraints, improve the configuration process by reducing the necessary effort to perform the configuration process.

## 6 Conclusions

Establishing the links between components in configured products can potentially significantly improve the ability to control the solution space and choice navigation for customers. Previous research has applied the Apriori data mining technique to establish these links. This paper focuses on a specific case and analyzes 180 sold configurations and their link to support and confidence levels used in the data mining.

From the case study, it can be seen that number of rules to consider increases dramatically as a function of the complexity of the rules. Furthermore, it can be concluded that in particular, the support levels are critical when investigating the rules. In general, low support levels lead to many rules and to disproportional large increases in the number of rules as the complexity of rules applied is increased.

## References

1. Lyons, A.C., Mondragon, A.E.C., Piller, F., et al.: mass customisation: a strategy for customer-centric enterprises. *Custom. Driven Supply Chains* 71–94 (2012)
2. Salvador, F., Forza, C.: Configuring products to address the customization-responsiveness squeeze: a survey of management issues and opportunities. *Int. J. Prod. Econ.* **91**, 273–291 (2004)
3. Salvador, F., De Holan, P.M., Piller, F.: Cracking the code of mass customization. *MIT Sloan Manage. Rev.* **50**, 71–78 (2009)
4. Choudhary, A.K., Harding, J.A., Tiwari, M.K.: Data mining in manufacturing: a review based on the kind of knowledge. *J. Intell. Manuf.* **20**, 501–521 (2009)
5. Geng, X., Chu, X., Zhang, Z.: An association rule mining and maintaining approach in dynamic database for aiding product–service system conceptual design. *Int. J. Adv. Manuf. Technol.* **62**, 1–13 (2012)



6. Hong, G., Hu, L., Xue, D., et al.: Identification of the optimal product configuration and parameters based on individual customer requirements on performance and costs in one-of-a-kind production. *Int. J. Prod. Res.* **46**, 3297–3326 (2008)
7. Yang, D., Miao, R., Wu, H., et al.: Product configuration knowledge modeling using ontology web language. *Expert Syst. Appl.* **36**, 4399–4411 (2009)
8. Wu, X., Kumar, V.: *The Top Ten Algorithms in Data Mining*. CRC Press, USA (2010)
9. R Project: 2013 (2013)
10. Ulrich, K.T., Eppinger, S.D.: *Product Design and Development* (2008)
11. Vollman, T.E., Berry, W.L., Whybark, D.C., et al.: *Manufacturing Planning and Control for Supply Chain Management*. McGrawHill/Irwin Series in Operations and Decision Sciences (2005)
12. Burbridge, J.L.: *Production Flow Analysis—for Planning Group Technology*. Oxford Series on Advanced Manufacturing, vol. 8 (1996)
13. Silver, E.A., Pyke, D.F., Peterson, R.: *Inventory Management and Production Planning and Scheduling*, vol. 3. Wiley, New York (1998)