INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED'09
24 - 27 AUGUST 2009, STANFORD UNIVERSITY, STANFORD, CA, USA

COMPUTER AIDED SIZE RANGE DEVELOPMENT – DATA MINING VS. OPTIMIZATION

Thomas Kipp and Dieter Krause
Hamburg University of Technology

ABSTRACT
Due to the growing number of demanded product variants, product family design gets more and more in the focus of research. Beside the development of advantageous product family structures the definition of size ranges is an important aspect of this topic. This contribution analyzes, which Data Mining and optimization methods are useful to support the development of size ranges and how these methods can be applied to this task. Finally on basis of two examples it is examined, if these methods are useful to adapted size ranges better to customer demands and therefore help to reduce undesired overdesign.

Keywords: Product family design, size ranges, Cluster Analysis, Evolutionary Algorithms

INTRODUCTION
The megatrend of globalization causes an increasing reduction of trade restriction, which leads to an unprecedented abundance of choice [1]. At the same time modern communication and information technologies enable most vendors to offer their products world-wide with reasonable costs. The result is a profound change of markets. So far they were dominated by particularly successive products, so called hit products. Now there is a growing relevance of niche products. Formerly was the 80/20-rule an appropriate description of markets, according to Anderson now the following paradigm is valid: “Endless choice is creating unlimited demand [2].”

Business companies face this profound change by constantly enlarging their product variety. This again requires a new alignment of product development. Formerly it was an appropriate approach to develop a singular product and to derive afterwards a few product variants. Nowadays ab initio the development of the whole product family has to be in the focus to achieve maximum commonality between product family members. That way internal complexity of the company can be minimized while the offered product variants meet the manifold costumer demands (see Figure 1).

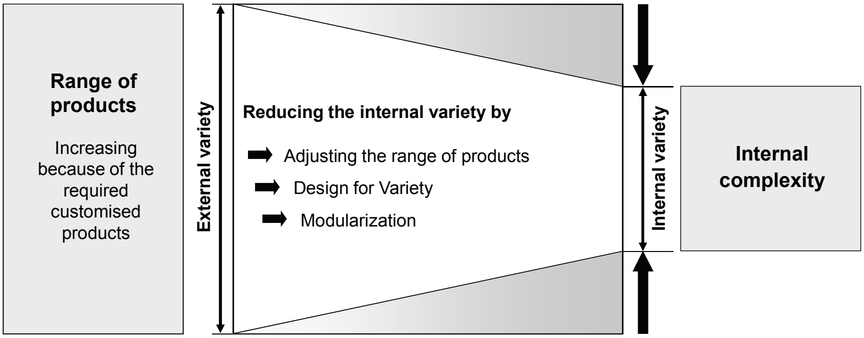

Figure 1. Challenges of product family design

A basic task, which has to be performed at the beginning of a product family development, is definition the variant product characteristics. Subsequently the different values of these characteristics have to be specified for each product variant. In mechanical engineering a common kind of these tasks is the definition of size ranges. In practice these sizes ranges are sometimes barely adapted to the customer demands. This leads either to a significant overdesign of the sold products or to a huge number of sizes and therefore to enhanced internal complexity.

In order solve this problem the following hypothesis is stated and examined in this paper: *Data mining or optimization tools can be used to define size ranges, which significantly better meet customer demands with the same number of different sizes.*

In the following it is shown for both alternatives tools, how they can be used as design method to define size ranges with respect to the customer demands. Subsequently both alternative methods are applied to practical examples. On basis of the results the benefit of applying these tools to develop size ranges is discussed.

CONTEXT OF PRODUCT FAMILY DESIGN

Due to the previous described profound change many contributions in the area of product family design were published in recent years. Focus of most of these publications was the development of adequate product structures from the functional or production point of view. For this a detailed state of science is described in previous papers [3].

The presented research examines an integrating approach of product family design. Basic ideas are a separate development of ideal product structures for the product planning, purchase, assembly and after-sales view (see Figure 2) as well as a subsequently methodical integration of the resulting four different concepts [5].

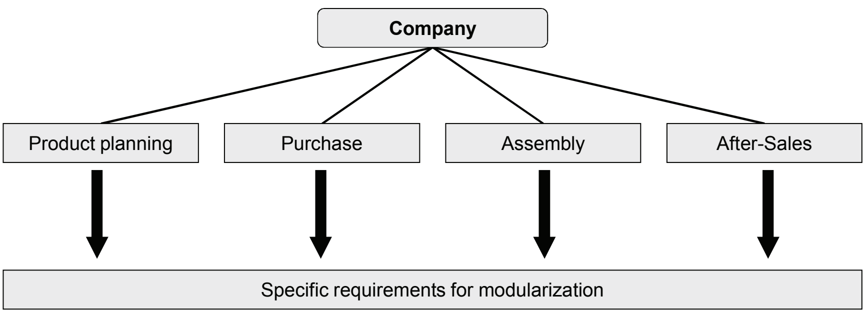

Figure 2. Business perspectives effecting product family design

This approach clearly reveals conflicts between these four different perspectives. Beside tradeoffs the use of innovative solutions is supported to solve these conflicts. Therefore a resulting improved quality of the final integrated concept is expected.

For the further development of this approach it is an important aim, to develop for each of these company perspectives a specific method for the product family design. From the product planning point of view the product family structure should enable the company to meet customer demands properly with minimal expenses. Product family design from this perspective is often named as Design for Variety.

A reference model for a product family structure designed for variety was presented in preceding papers [3]. This reference model is characterized by the following attributes:

- Explicit differentiation between standard and variant components
- Reduction of variant components to carriers of variant product characteristics
- One to one mapping between variant product characteristics and variant components
- Decoupling between the components

In order to apply this reference model to real products the so called Four-Layer-Model was introduced (see Figure 3). This model visualizes, how variant customer demands effecting the technical parameters, the working-geometry and finally the variant components [4]. While developing product families this model reveals clearly differences in how alternative concepts provide the demanded variety. So the design engineer should be supported in performing a Design for Variety.

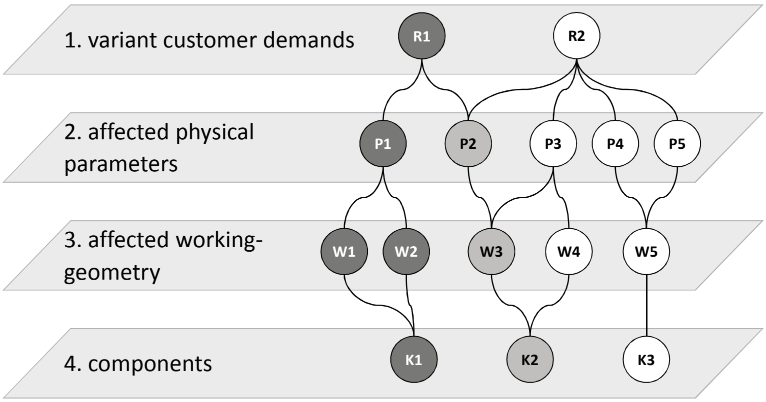

Figure 3. Four-Layer-Model to support product family design [4]

Thereby every Design for Variety is based on the first layer, the layer of demanded variant characteristics. In this stage the design engineer is confronted with the following two questions: Which product characteristics should be used to adapt the product to various customer demands? Which specifications of these variant product characteristics should be offered?

As mentioned before in mechanical and plant engineering the definition of size ranges is a common kind of this task. In generally the stepping of the determining product characteristic is defined by geometric series, i.e. the next bigger size in this series is always defined by a constant step factor φ. Based on the number of sizes z this step factor can be calculated in the following way [6]:

$$\varphi = \sqrt[(n)]{\frac{\text{Greatest term of the range}}{\text{Smallest term of the range}}}\text{, where } z = n + 1$$

When defining such a size range customer and production demands have to be balanced carefully. Customer and technical demands on the one hand side require small increments. Production on the other side requires large batches and therefore a coarse step factor [6].

As seen in practice it is not always possible to meet customer demands using size ranges based on geometric series. Looking at complex components with multiple independent variant characteristics this behavior can be often observed. Here a size range definition based on geometric series leads to an exaggerated number of sizes or an overdesign of the offered products, which is in many cases unacceptable for customers. As observed in practice Data Mining tools such as Cluster Analyses are already used to support the definition of size ranges based on customer demands. No reference could be found for the usage of optimization tools to develop size ranges, though the nature of optimization promises at least equal results. Based on these observations it was hypothesized, that *Data Mining or optimization tools can be used to define size ranges, which significantly better meet customer demands with the same number of different sizes.*

To verify this hypothesis in the following both alternative approaches are separately analyzed and used to define two exemplary size ranges based on customer demands. Subsequently the benefit of these approaches is discussed based on a comparison between the resulting and manually defined size ranges.

DEFINITION OF SIZE RANGES SUPPORTED BY DATA MINING TOOLS

The first alternative is the use of Data Mining tools to support size range definition. Below the systematic choice of an appropriate data mining tool is described. Subsequently the size range development using the selected tool is presented.

Selection of an useful data mining tool

As illustrated below Data Mining tools can be assigned to five categories based on their function:

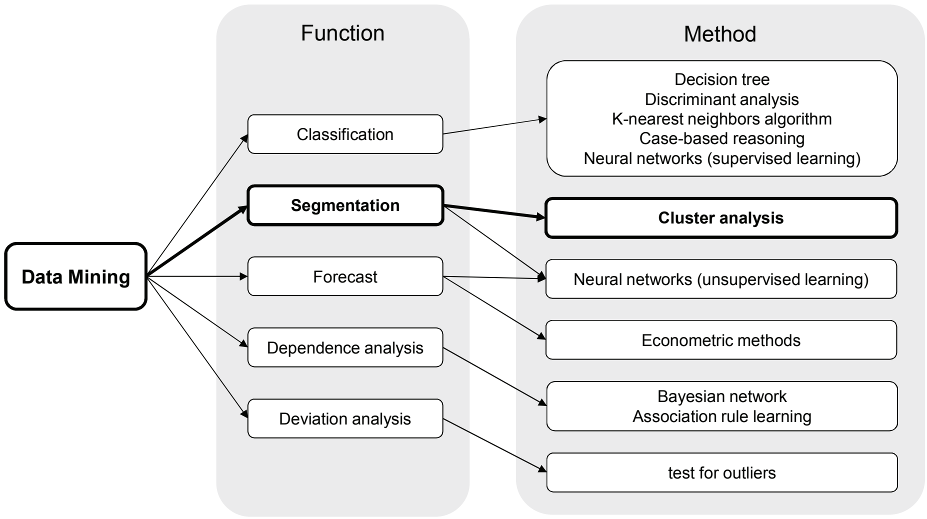

Figure 4. Classification of Data Mining Methods [7]

To identify an appropriate Data Mining method the task of defining size ranges based on customer demands has to be assigned to one of these five categories. To perform this task the whole of customer requests has to be split up into a defined number of groups of similar requests. In mathematic terms *Segmentation* can be defined as splitting up a heterogeneous quantity into homogenous and previously unknown groups of similar objects [7]. Therefore the present task can be clearly identified as *Segmentation*, which identifies *Neural Networks (unsupervised learning)* and *Cluster Analysis* as appropriate methods.

The structure of *Artificial Neural Networks* is based on the human brain. Neural Networks learn and adapt through interaction with external stimuli [8]. This method is a so called black box method, which is an important disadvantage. It means that the user can only observe the input and output layer. This way the generation of the results is hardly understandable. The lacking transparency also impedes the interpretation of results [9].

In contrast the *Cluster Analysis* is a statistical technique for the segmentation of heterogeneous quantities. Thereby the quantity is segmented in a way that the members of the resulting groups are as similar as possible, while similarities between members of different groups are minimized [10]. How similar or dissimilar two objects are is evaluated by a defined distance measure.

Any Cluster Analysis is structured in two important steps. The first step is the choice and calculation of a distance measure. The second step is the choice and performance of a clustering algorithm. If the groups (clusters) are previously unknown, like in the present case, a hierarchical clustering algorithm has to be used. Such an algorithm groups step by step always the most similar object until only one group is left. As shown below the results of hierarchical clustering are typically represented in dendrograms.

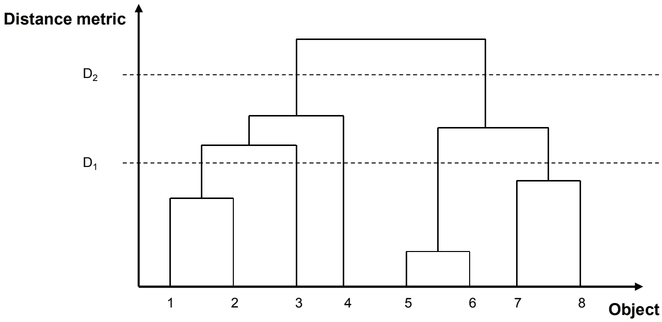

Figure 5. Dendrogram based on a cluster analysis

The dendrogram makes it easy to identify, which cluster are merging on which step. Thereby is the homogeneity of clusters decreasing with an increase of the distance measure. This representation supports the interpretation of results as well as an identification of a reasonable number of clusters.

Cluster Analysis is here compared to *Artificial Neural Networks* the preferred method, due to its easier use, better transparency and comprehensibility of results. The following paragraph describes how Cluster Analysis can be used as design method to develop size ranges.

Development of size ranges supported by Cluster Analysis

Basis for the definition of size ranges are the customer request data from several years. The first step is the transformation of each customer request into ideal physical characteristics of the product.

The following second step is the choice of the distance measure. Within this work both the City Block Distance (simple example) and the Quadratic Euclidean Distance (complex example) were used. The selection of a distance measure is highly linked to product specific considerations; therefore no general suggestion can be made.

Subsequently outliers have to be eliminated from the data set. For simple size ranges this can be performed manually, for complex data sets Cluster Analysis can also be used to support outlier elimination. The Single Linkage clustering algorithm can be applied to this task. This algorithm creates quickly one big cluster and adds in the following always the next nearest neighbor. At last the most dissimilar objects are added, so these outliers are easy to identify in the resulting dendrogram.

The following fourth step is the real clustering, whereto the Complete Linkage algorithm is used. To measure the distance between two clusters this algorithm always uses the biggest distance between two objects of these clusters. This algorithm creates some very homogenous clusters and such homogenous groups are useful for the definition of sizes.

The final step is the definition of the product sizes based on the dendrogram. Therefore the segmentation of the dendrogram for the favored number sizes is used. For instance in the above presented dendrogram a size range with five different sizes was defined based on the D_1-line. Subsequently for each resulting group the maximum value of the variant product characteristic has to be determined and used to define the related product size.

The whole approach for the Definition of size ranges supported by a Cluster Analysis based on customer requests is subsumed in the following flow chart.

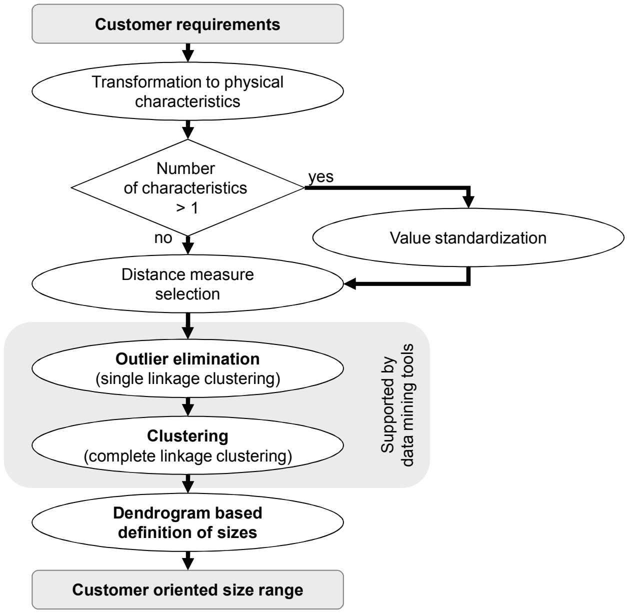

Figure 6. Approach for size range development supported by Cluster Analysis

DEVELOPMENT OF SIZE RANGES USING OPTIMIZATION TOOLS

Commonly the application of a computer supported optimization can be structured in three steps. It starts with the definition of the target function based on the real problem. Afterwards an optimization method has to be selected, which is applicable to the target function. Finally the optimization has to be performed. In the following the use of optimization tools for defining size ranges is described on basis of these three steps.

Definition of the target function

Target functions are utilized in optimizations to evaluate the solution quality. They have to be defined carefully, because the target function widely determines, which optimization algorithms are applicable. A strong simplification of the target function often helps to achieve desired characteristics of the function like linearity or continuity. This holds the risk that the real optimization task is unsatisfactory represented. In contrast an optimal representation of the task easily leads to a non-linear and discontinuous target function and in result most optimization algorithms are inapplicable.

In the present case the aim of the target function is to evaluate the quality different size ranges with a defined number of sizes *(n)*. The optimization is again based on a quantity of customer request from several years, which are transformed to the ideal values r_1 to r_m of variant product characteristic. The sizes x_1 to x_n, which have to be defined, are the parameters of the target function.

Quantity of customer request: $R = \{r_1, .., r_m\}$, *where* $m \in \mathrm{N}$

Undefined sizes: $x_1, x_2, ..., x_n$, *where* $n \in \mathrm{N}$

The quality of a potential size range should be evaluated on basis of the resulting overdesign. For each customer request the overdesign *(Y)* can be calculated as difference between the next bigger size step and the ideal value of the variant product characteristic. Therefore the overdesign can be described in mathematic terms as minimum nonnegative difference between the sizes x_1 to x_n and the ideal product characteristic r:

$$Y_0(x_1, x_2, ..., x_n) = \min_i (x_i - r_0 \geq 0) \tag{1}$$

Here the accumulated overdesign based on all customer requests is defined as measure for the quality of a size range. The outcome of this is the following target function, which has to be minimized:

$$f(x_1, x_2, ..., x_n) = \sum_{j=1}^{m} Y_j(x_1, x_2, ..., x_n) = \sum_{j=1}^{m} \min_i (x_i - r_j \geq 0) \rightarrow \min \tag{2}$$

In the case of multiple independent variant product characteristics the target function can be defined analogous, though the scalar parameters have to be replaced by vectors.

Selection and Presentation of the Optimization method

As illustrated in Figure 7 the selection of the optimization method is a result of the target function. Thereby the target function is non-linear and discontinuous caused by the inevitably nonnegative constraint. Thus for this task only Evolutionary Algorithms seem to be applicable.

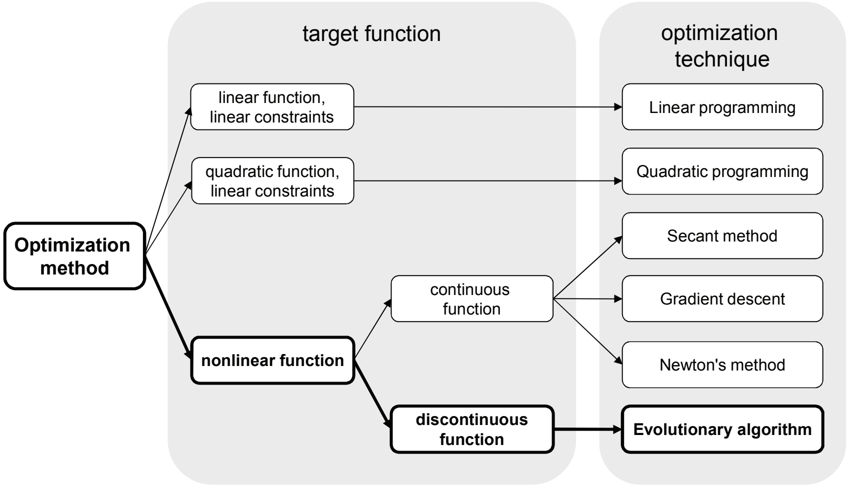

Figure 7. Selection of optimization methods [11]

These Evolutionary Algorithms try to adapt Darwin's principle of "the survival of the fittest" as shown in Figure 8. Before the beginning of such an optimization an initial population of potential solutions is created. The parameter values of these individuals are randomized.

At the beginning of the optimization loop it is checked whether a defined abort criterion is already fulfilled or not. If this is not the case so called sets of parents are selected, thereby individuals with better fitness values are preferred. Subsequently children are created through random recombination and mutation of the parameter values of their parents. These children represent the next generation of the population. With this new generation the optimization loop starts all over again until the abort criterion is fulfilled. The best individual of the final generation represents the determined optimum.

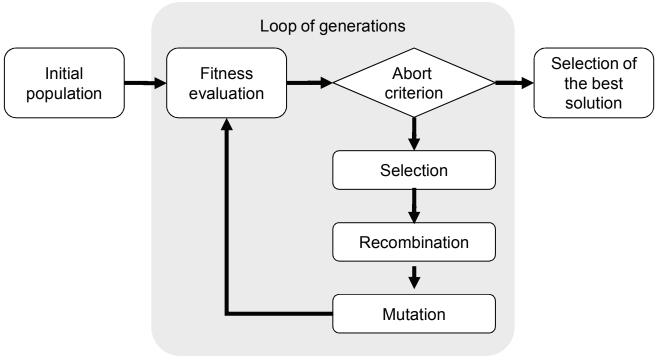

Figure 8. Evolutionary algorithm flow chart

Evolutionary Algorithms represent a stochastic search technique. Analog to the example of the real evolution the search proceeds in an intelligent way. The primary advantage of this optimization method is the universal applicability. On the other hand the black box behavior and the lack of repeatability are significant disadvantages.

Development of size ranges using Evolutionary Algorithms

To develop size ranges using Evolutionary Algorithms a procedure similar quite to one presented for the Cluster Analysis (see Figure 6) can be used. Therefore only the steps “Clustering” and “Dendrogram based definition of sizes” have to be replaced through the step of evolutionary optimization. In the following only this step is described.

In order to perform this optimization a special Evolutionary Algorithm was implemented instead of using an available tool. This allows a better adaption of the algorithm to its task and leads to a better understanding of the optimization tool and its functionality. Finally this significantly simplifies the interpretation of results.

According to the number of independent variant product characteristics, i.e. to the complexity of the task population sizes between 150 and 7000 individuals were used. Thereby always 15 generation loops were performed.

The selection algorithm assigns the individuals a probability of selection inversely proportional to the rank of an individual according to the fitness values. This selection algorithm was chosen, because it less favors the best individuals compared to fitness value proportional selection. Thus a heterogeneous population is longer maintained.

Subsequently four children per set of parents are created. Therefore each parameter is randomly picked from one of the parents. In order to diversify the population a small percentage of the children’s parameters is randomly mutated with a defined maximum change.

Finally the replacement algorithm, which is necessary to define the next generation, starts with creating a new bigger population including the last generation and all children. This extended new population is ranked using an additional fitness evaluation. To create the next generation the best individuals are picked according to their rank until the normal population size is reached. This replacement algorithm avoids, that already existing optimal solutions get lost in the optimization process.

EVALUATING BOTH METHODS ON BASIS OF TWO EXAMPLES

In order to compare and to evaluate both presented alternative methods they are applied to two examples. Both examples are real data sets from two different companies. With regard to these companies the resulting sizes are not mentioned here.

Both examples have in common, that the distributions of customer demands were nearly constant in the observed period and that it is hardly feasible to approximate these distributions using geometric series.

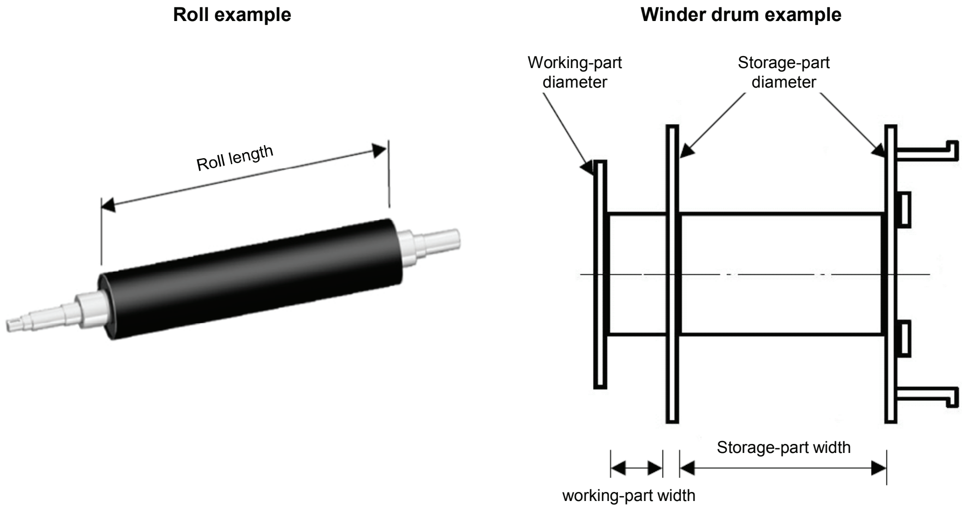

Figure 9. Examples to evaluate the examined methods

The first of these examples is a roll which is part of a printing machine. In the examined range of sizes the only variant product characteristic of this roll is its width, therefore it is used as example for a simple size range. In this example an overdesign up to 100 mm was definitely acceptable for the customer, furthermore the usability cannot be guaranteed in each case.

In Figure 10 the size ranges, which are defined using the presented methods, are compared with a manually developed size range. Increasing maximum values of overdesign are assigned to the x-axis. The y-axis shows the accumulated percentage of customer request, which can be satisfied with less than the defined maximum overdesign.

Looking at the point of a maximum overdesign of 100 mm obviously these three size ranges reach an almost similar quality. The only difference between these size ranges is the part of customer request, which can be satisfied without any overdesign. Here the size range defined by Evolutionary Algorithm gains the best result.

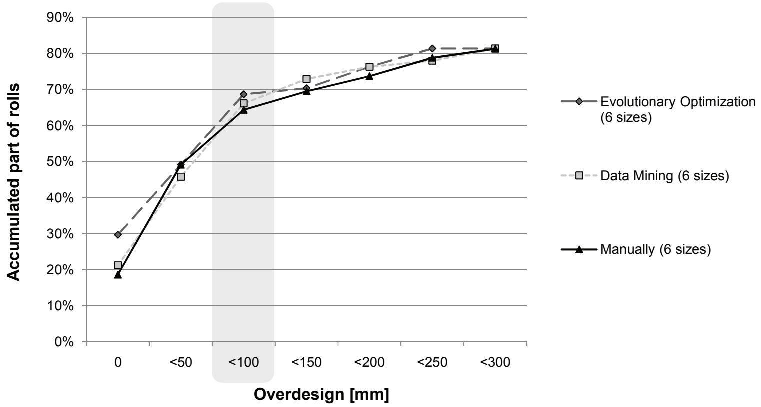

Figure 10. Comparing the size ranges of the roll example

The second example a winder drum size range has the following four independent variant product characteristics:

- Width of working part
- Diameter of the flanged wheel of the working part
- Width of storage part
- Diameter of the flanged wheel of the storage part

For each of these product characteristics a maximum overdesign of 50 mm is acceptable. For overdesign exceeding this limit the acceptance of the customer has to be checked in the particular case. In Figure 11 the differently developed size ranges are compared exemplary on basis of the *width and diameter of the storage part* analog to the preceding example.

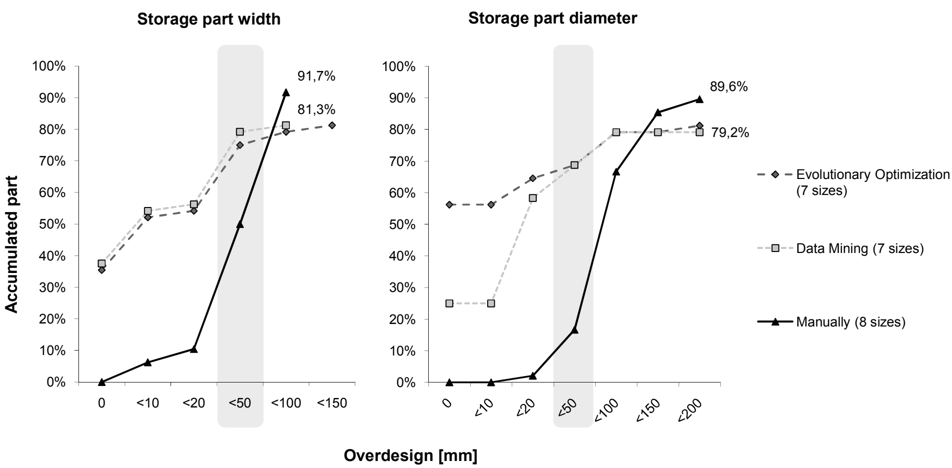

Figure 11. Comparing the size ranges of the winder drum example

It becomes obvious, that the manually developed size ranges satisfies a significantly smaller part of the customer requests with the defined maximum overdesign of 50 mm. In comparison to the Data Mining size range the optimized size range achieves the same standard (*width*) or is clearly better in the area of minor overdesign *(diameter)*.

It can be summarized, that the development of size ranges supported by Cluster Analysis or Evolutionary Optimization is an interesting alternative, if size ranges based on geometric series are inappropriate to meet the customer demands. Furthermore it can be stated, that the quality of manually developed size ranges is decreasing with an increasing complexity of the task. Comparing both presented tool Evolutionary Algorithm gain better results in the area of minor overdesign.

CONCLUSION

The examined hypothesis, that *Data Mining or optimization tools can be used to define size ranges, which significantly better meet customer demands with the same number of different sizes* is valid under specific conditions. These conditions are fulfilled, if size ranges on basis of geometric series meet customer demands unsatisfactory and the size range cannot be developed manually due to complexity of the task. In this case the examined methods gain significantly better results, i.e. the developed size ranges met the customer demands better and therefore undesired overdesign is minimized.

Comparing both methods it becomes obvious, that size ranges developed by Evolutionary Algorithm satisfy a considerably higher percentage of customer demand without any overdesign. Disadvantages are in comparison the difficult application of Evolutionary Algorithm and the black box behavior, which impedes the interpretation of results.

Finally it can be stated, that Evolutionary Algorithm bear a huge potential used as a design method for the development of size ranges, but there is considerably need for further research to create a user-friendly on basis of this method.

REFERENCES

[1] Naisbitt J. and Aburdene P. *Megatrends 2000: Zehn Perspektiven für den Weg ins nächste Jahrtausend*, Düsseldorf, 1992 (Econ).

[2] Anderson C. *The long tail: Why the future of business is selling less of more*, New York, 2006 (Hyperion)

[3] Kipp T. and Krause D. Design for Variety - Ein Ansatz der variantengerechten Produktstrukturierung. *6. Gemeinsames Kolloquium Konstruktionstechnik*, Aachen, 2008 (Shaker)

[4] Kipp T. and Krause D. Methodischer Ansatz der variantengerechten Produktstrukturierung. *Berliner Kreis: Tagungsband – Jahrestagung 2008.* Kaiserslautern, 2008 (Technische Universität Kaiserslautern)

[5] Blees C., Jonas H. and Krause D. Entwurf von modularen Produktarchitekturen unter Betrachtung unterschiedlicher Unternehmenssichten. *Design for X : Beiträge zum 19. Symposium*, Neukirchen, 2008, pp.149-158.

[6] Pahl G., Beitz W., Feldhusen J. and Grote K.H. *Engineering Design: A Systematic Approach*, London, 2007 (Springer).

[7] Alpar P. and Niedereichholz J. *Data Mining im praktischen Einsatz*, Wiesbaden, 2000 (Vieweg).

[8] Bigus J.P. *Data Mining with Neural Networks*, New York, 1996 (McGraw-Hill).

[9] Schinzer H., Bange C. and Mertens H. *Data Warehouse und Data Mining: Marktführende Produkte im Vergleich*, München, 1999 (Vahlen).

[10] Backhaus K., Erichson B., Plinke W. and Weiber R. *Multivariate Analysenmethoden: eine anwendungsorientierte Einführung*, Berlin, 2006 (Springer).

[11] Poddig T. *Statistik, Ökonometrie, Optimierung*, Bad Soden, 2001 (Uhlenbruch)

Contact: Thomas Kipp
Hamburg University of Technology
Product Development and Mechanical Engineering Design
Denickestr. 17
21073 Hamburg
Germany
Tel: +49 40 42878 2951
Fax: +49 40 42878 2296
Email: kipp@tuhh.de
URL: http://www.tu-harburg.de/pkt

Thomas studied mechanical engineering at the Technische Universität München. Since 2006 he works as a scientific assistant at the Hamburg University of Technology. He researches in the field of engineering design. Thereby in the focus of his research are new methods and approaches for the design of highly variant products.

Prof. Dieter Krause is head of the Institute Product Development and Mechanical Engineering Design and dean of mechanical engineering at Hamburg University of Technology. He is member of the Berliner Kreis and the Design Society. Main topics of his research are new design methods with respect to product variety and modularization as well as lightweight design for aircraft interiors.