ENHANCED DESIGN FOR ASSEMBLY IN SERIES PRODUCTION BY USING DATA MINING METHODS

R. Kretschmer, S. Rulhoff and J. Stjepandić

Keywords: product realization, manufacturing, digital factory, assembly, process planning, data mining

1. Introduction

Today globally operating companies face additional challenges due to the increasing variability of products, shortened product lifecycles and corresponding complexity of processes. Due to the market fluctuations, there are increasing demands on the flexibility in the production system on the economic dispatch of new products in an existing production line. In the modern product emergence process (PEP), production planning gains in importance and has to be executed as parallel as possible to the product development according to concurrent engineering principles [Bracht and Masurat 2005]. In this early phase of the product creation, a first step for planning processes is a cost calculation for the industrialization of the product in existing production lines regarding basic conditions [Bley and Franke 2004]. The economic feasibility of series production must be assured with vague information on the product and given general conditions, e.g. shift model [Boothroyd 2005]. This is a great challenge especially to the planning of the cost-intensive assembly of the product [Rekiek and Delchambre 2006], [Lotter and Wiendahl 2013].

The research and development project “Prospective Determination of Assembly Work Content in Digital Manufacturing (Pro Mondi)” was initiated to develop a concept using methods of data modeling and Data Mining (DM) to generate information with focus on the product assembly planning for new products in early production planning phases. Aim of this project is the accurate estimation of the expected assembly work content and the resulting costs in an early stage of the product development as well as the additional support of the design process with assembly knowledge for the specific design. The approach contains the reuse of existing planning data in order to extrapolate assembly processes. Especially linked product and process data allow the innovative usage of Data Mining methods. As proof of concept this approach will be evaluated with different manufacturing companies.

Data mining is a process of discovering valuable information from observational data sets which is an interdisciplinary field bringing together techniques from databases, machine learning, optimization theory, statistics, pattern recognition, and visualization. Data mining has been widely used in various areas such as business, medicine, science, and engineering. Introduction of data mining concepts, implementation procedures and application cases in engineering is described by [Yin et al. 2011], [Talia and Trunfio 2013]. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. One of the greatest expected benefits of DM methods is the ability to link seemingly disparate disciplines, for the purpose of developing and testing hypotheses that cannot be approached within a single knowledge domain. Methods are reviewed by which analysts can navigate through different data resources (e.g. historical
data] to create new, merged data sets [Chu 2014]. Significant factors are efficient knowledge utilization and knowledge exchange on an interdisciplinary level.

New processes suitable to assemble the given new product shall be designed based on this existing, historical data (product linked to corresponding process). Automatic analysis with a specific DM method shall be used to create a first draft of the assembly process and estimate the expected costs. Following production planning processes can be supported by automatic proposals of adequate assembly processes, which then can be customized [Rao 2011]. Moreover, the design engineer can be supported at the selection of appropriate joining elements. With this approach, an assembly knowledge based support of the designer in series production can be achieved using innovative DM methods.

This paper, as partial results achieved in this project, describes the innovative methods of PROSTEP AG facilitating the use cases of Miele Cie& KG, one of the leading manufacturers of domestic appliances.

2. Business Requirements

Available on five continents, Miele is the global premium brand of domestic appliances and commercial machines in the field of laundry care, dishwashing and disinfection. A continuous stream of innovations is the foundation of Miele's business success since 1899. In terms of quality, Miele appliances are considerably better than those of the competition, otherwise they would not have been able to compete successfully on such a fiercely competitive market. In order to address the challenges of data mining, the integration of various planning tasks within the PEP, new concepts are necessary. Though, as a part of integrated product and process development there are different definitions for various phases and aspects of planning activities along the PEP [Whitney 2004]. Regardless of the specific definition of these phases and aspects, however, based on the analysis it is certain that great amount of their containing information and knowledge are either utilized insufficiently and ineffectively or remain unused [Erohin et al. 2012]. In this regard, the presented concept focuses on product design and production assembly planning [Magrab et al. 2010]. Subsequently, for the product designer and production planer, there are varieties of applications which can assist the design or the planning process through information gathered by data mining [Rulhoff et al. 2013].

2.1 Preparation and Requirements

The proposed approach is shown in Figure 1 and runs through the following steps:

**Enrich CAD data with assembling information:** Derived from previously similar constructed products, assembly information such as time data about the actual design situation can be identified and provided in order to support the designer. These additional information can be used to enrich the CAD data to support the current design and be updated in later assembly planning processes.

**Suggesting assembly connections:** An assisting option for the designer is a suggestion list of similar previously constructed assembly connection variations. These lists give a quick overview of possible and already implemented connection types in the assembly.

**Assembly process estimation:** The focus is on the creation of an assembly process for a new product. Based on existing product and process data compilation of a first approximated assembly process for a new product could be developed. From this, the production planner can specify further details and thus determine a first estimation regarding assembly time. Based on the assembly time and associated calculation scheme, the planner can perform the first cost estimation in a very early planning phase. The information in production planning and engineering processes can mutually enrich each other. Additionally, intelligent interconnecting information from both areas creates added value. The newly obtained information supports the workflow throughout the PEP. Therefore, as part of this concept some requirements need to be met. Thus, the pre-conditions attached to both systems as well as their respective processes have to be fulfilled [Schallow et al. 2011].
2.2 Attributes and Data Sources

Data Mining methods can be used for data clustering and classification, however criteria for comparison of data sets have to be identified [Han et al. 2012]. To determine these criteria, within the scope of ProMondi project, a survey of users as well as an analysis of various tools of the DM was performed. The objective of this analysis was to identify attributes relevant for assembly processes that could be assigned to products and parts in CAD [Hartung et al. 2012], PDM and production planning systems. In CAD systems, attributes assigned to parts contain mainly geometric information including volume and weight. The PDM systems contain organizational information, such as creator, version and revision as well as the mentioned parts information form CAD [Eigner and Stelzer 2009]. In addition to the conventional systems for design and stacking product parts and assemblies, systems for process planning and time measurement were also taken into account. They sustain a comprehensive portfolio of information and therefore can be used to distinguish different product parts and assemblies. The results of this analysing are capsulated as an object oriented data model, further described in chapter 3.

The necessary enrichment of product and process data on the fly for the presented concept requires additional efforts in the design. This additional expenditure also relates to the assembly connections and includes the acquisition of new information form the designer’s know how. The designer usually defines assembly connections either implicitly through formed locked joints by the shaping of the parts or explicitly by connecting elements such as in screwed fasteners.

2.3 Data Collection and Availability

The designer of the assembly connection considers all these information in the design but cannot store them in the CAD model because the CAD tools for the most part are not able to define the necessary attributes.

To overcome this problem as part of the concept presented in this paper, the designer will be provided with an additional tool in the CAD system. It can be used to create assembly connections and gives additional information and explicit design possibilities. These additional assembly information are named below as “product assembly information”. Thus, data will be collected in the source system, the CAD system in particular. Since the defined objects are not part of PDM systems, an extension is
necessary in order to implement connections as objects and to store them after the transfer in the PDM system persistently. In the further processing, the product information will be linked to the planning processes. Unless the storage of product data are in the same system as for production planning, the information flow from the PDM system to the planning system as well as the DM tool, for further analysis, has to be ensured.

In current planning systems it is often possible to directly link processes to products [Petzelt et al. 2009]. Thus, an allocation of to be assembled product and the associated assembly processes can be realized. In the assembly, however, parts are joined with other parts or products. These assembly connections with their additional information have no digital equivalent object yet. However by means of an object such as the “product assembly information”, it is possible to store useful additional connection information, which relates directly to the respective assembly connection. As part of this concept, the combination of the products and processes does not take place directly but through this object “product assembly information”. The linking of product and process does not necessarily need to occur at the part level.

3. Solution Approach

The concept presented in this paper describes an assisting workflow to support the designer (Figure 2). As part of a new or modified design, the designer creates new product data. In creating the assembly connections, a software assistant supports the designer and enriches the CAD model with product assembly information for each connection. This product assembly information includes additional connection information including e.g. the torque screwed fasteners or the type and the form of a welded joint and information about other connection types. In the on going design process, the designer can trigger an evaluation regarding the assembly connections in the model.

For this purpose, the characteristics of the CAD model are first prepared and analyzed with Data Mining. The analysis focuses on the product assembly information and their properties. The parts associated with the product assembly information and their geometric properties are also included in the analysis as additional information set. Furthermore, an extended database is also provided and consists of product and process data of existing products which are linked via the product assembly information. The characteristics of the product assembly information of the new product are compared with the properties of the product assembly information of the existing product in the extended database. Then, the most similar product assembly information is determined from the existing products. This analysis can be restricted by a class of the connection types (screwed, weld, rivet) or deliberately left open to widen the solution space and to provide the designer with information about other assembly connections.

Figure 2. Design optimization with additional time data
A limitation on the particular type of connection yields as a result of the closest realized assembly connection of the same kind. Depending on the properties of the parts, other mounting connections can also be found and proposed to the designer in a proposal list.

The presented application for the support of the design process uses the product assembly information identified in the analysis of the PDM database to determine the respective associated and related sub processes. Each product assembly information represents an assembly connection. By multiple connections within the assemblies, multiple sub-processes for the assembly can be determined. These processes contain the time data relevant for the new product design. Therefore, the corresponding time information of the existing products and, if requested, an alternative proposal list is transferred in the CAD System and displayed. This assembly time information of the existing product represents a first approximated assembly time for the new product. So the designer is provided with this additional information regarding the assembly time and with an enterprise specific factor the corresponding cost of the current design solution. In the final step, the designer is able to optimize the product iteratively on the basis of anticipated assembly time and costs for each design.

### 3.1 Data Model

Based on determined assembly characteristics, a range of attributes is derived to classify the assembly of the parts. Figure 3 shows an overview of the generated data model for the data mining analysis.

The ProductAssemblyInformation (PAI) is the central element in this data scheme and represents the assembly of the product parts. References for time analysis, assembly requirements, designed parts or assembly units, geometrical characteristics as well as ProductAssemblyInformation. Each of the parts. Figure 3 shows an overview of the generated data model for the data mining analysis. The ProductAssemblyInformation is the central element in this data scheme and represents the assembly of the product parts. References for time analysis, assembly requirements, designed parts or assembly units, geometrical characteristics as well as ProductAssemblyInformation. Each Item refers to the ProductAssemblyInformation, which also refers to further used Items. This construct is chosen to enable Data Mining methods to determine exact similarities between new parts and/or products and the existing ones, in any order and combination interchangeably possible.

In the first approach the attributes for screw connections are clustered and evaluated regarding the relevance for assembly operation. The identified attributes are classified in the categories fasteners, installation / assembly situation, tools / equipment, installation regulations and additional assembly elements. These attributes are represented in the data model in different object categories (here illustrated by colour, c.f. Figure 4). The evaluation regarding the influence on the assembly time provides a first indication for the relevance in the data mining analysis. Production planner weighted
the attributes with low medium and high in order to determine a first selection of attributes relevant for the data mining. Which attributes really are significant for the similarity of assembly connection have to be determined in a data mining analysis with a large quantity of product data. The columns construction and work planning shall indicate in which department the attributes can be collected.

<table>
<thead>
<tr>
<th>Category</th>
<th>Product Assembly Information (PAI)</th>
<th>values</th>
<th>Influence on the assembly time</th>
<th>construction system</th>
<th>knowhow system</th>
<th>work planning system</th>
<th>knowhow system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasteners</td>
<td>screw head diameter</td>
<td>mm</td>
<td>nm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>thread type</td>
<td>fine thread / coarse thread</td>
<td>mm, cm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td></td>
<td>number of thread transitions (used)</td>
<td>yes/no</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>screw diameter</td>
<td>mm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>screw length (thread)</td>
<td>mm</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>screw type</td>
<td>tapping screw, metric screw</td>
<td>medium</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>material</td>
<td>metal, steel, wood, etc.</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>magnetic screw</td>
<td>yes/no</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The columns with a list are generated and the most related one can be preallocated some parameters with estimated values which can be reviewed later. Other parameter and the corresponding value data can be extracted out of other systems e.g. the attributes of standard parts.

3.2 Data Mapping and Data Mining

After aggregating and appending the data subsets from different sources and systems, it is necessary to remove redundant data sets [Ohno-Machado et al. 1998] for the DM process. The next step is converting and porting data in the presented data model. Depending on data source the conversion is either fully automated or partially automated with further manual adjustment. Often, value and scale of different attributes are heterogeneous. In these cases, a normalization of ratings prevents the undesired high or low impact of certain attributes on the results and evaluation process. In this regard, a [0, 1] linear normalization has been used. Additionally, a further attribute prioritizing via weighting can be necessary to define the importance of each attribute for the evaluation. An automated learning of the weights via machine learning methods depends on the existing data sets and their quality. Otherwise, they are determined based on expert knowledge or a combination of both methods. To prevent further expansion of scope and the complexity of existing problem, expert knowledge was applied to determine the attribute weights. Furthermore, it is possible to have more than a single weight vector. This approach is useful, if there are various object types or parts, which have different prioritization for their attributes [Zhang et al. 1992]. To identify the objects with most similar product assembly information for a new object, the classification algorithm k-nearest neighbour (kNN) [Dhanabal and Chandramathi 2011] with Euclidean distance as evaluation function is used. From the identified objects, a list is generated and the most related one can be manually chosen, which passes its assembly
process data to new object. To assure the reliability of the presented method and prevent over fitting, a cross validation [Kohavi 1995] is used.

The implementation of the presented approach is challenging due to high requirement for interconnection and the overall quality of the existing data in different source system. In particular the pure number of realized and existing assembly connections and, thus, necessary instances of an product assembly information as well as the quality of the data regarding their attributes are important.

![Diagram of Design Support Tools]

**Figure 5. Formation of product clusters and process agglomerations**

Momentarily, the fulfilment of these high requirements have to be verified. In particular, the quality of the linkage of product data with the corresponding assembly processes poses a challenge. Methods to improve the for this concept important quality will be evaluated. Is this task solved, the selection of the properties and attributes for the DM analysis also has to be determined based on production data to ensure the reliability of generated results (Figure 5). In this scope a special focus is on the characteristics of the parts and of the connection itself. In conformity with the presented objectives and concept, a utilization of the methodology is described as follows.

**Suggesting assembly connections and enrichment of CAD data:** The designer creates a new module with already known and new assembly connections in the CAD system. He designs assembly connections and complements these connection properties in the context of the new module. Via the automated DM process, he is provided with various information about the assembly connections. Moreover, for each assembly connection a list of alternative or ever realized connections can be created. Depending on the product properties, the five most similar product assembly information are made available to the designer as a prepared proposal list, which is generated through cluster analysis of existing product data. These information can be used directly and enhance the CAD model in order to use it later or in an extended context of the product. If the analysis is dispensed with the filtering of associated connections with the product assembly information, the designer can also be provided with other not associated connections as alternatives.

**Estimation of assembly process and information:** The production planner drafts an initial assembly process for a new assembly at an early stage of product development. Analogous to the use case of the designer, for known assembly connections that are implemented in the new product as well as in the old product data, the right product assembly information and, thus, the assembly processes are found.
For new unknown connections, the most similar product assembly information and related assembly processes from the database are determined and duplicated. Each of the founded product assembly information represents a single connection and the linked process represents precisely the assembly work content for this connection. The sum of the individual connections for the new product is its first assembly process. Thereby, an initial draft of an assembly process of the new module can be generated. The found individual connections, the individual process, as well as the overall process can be used in different ways to assist the designer and the production planner. The planner and designer also get a first estimation for the expected assembly time and costs in the automated process. In addition, the production planner can increase the quality of the process by manual intervention. On the one hand, he adapts the product assembly information created by the designer before the DM analysis. On the other hand, he can complete the product assembly information in the attributes with practical knowledge. Thus, he has an impact on the input of the DM analysis and increases the quality of the result thereby. Furthermore, the designer has a first draft for the assembly process at one’s disposal and a first estimated assembly time in the current CAD system. By a company-specific factor, the designer receives also information about the cost of the connection in the assembly. By verifying this information, the designer can evaluate and compare the alternatives for different connections.

4. Use Case Evaluation

4.1 Building of Product Clusters

Currently, several individual Use Cases are considered for validation purposes. First, the formation of the product cluster was considered. To estimate a suitable number for the product cluster structural data from part lists of assemblies were used. Then, the attributes such as size, material and weight which were shown previously were used in the first approach. In this case, the calculation of the component dimensions of the bounding box proved to be sufficiently accurate. This way, 5 product clusters with respective reference to the component category could be formed. With regard to the associated assembly processes, a first result was that components with similar design characteristics also need similar assembly processes. This seemingly trivial statement however validates an important requirement for the applicability of existing assembly processes for current products to future assembly processes for new-developed products.

4.2 Building of Process Clusters

The clustering for the processes is based on time analysis in the MTM method (Methods Time Measurement). Basic principle is the determination of target times by combining the time measurement units. Depending on the containing time blocks, 7 process cluster with explicit reference to the component category were formed. Additional accuracy of clustering can be achieved with different similarity searches within the process data. The different similarity searches are used:

1. Similarity search via process parameter: Single time blocks such as „pick and place“ are composed of individual movements. Each single time block has attributes like „distance to pick“. These attributes are used in the data mining method K-Means to divide similar assembly processes into clusters.

2. Similarity search via description text: The similarity of description texted are evaluated based on the text mining. In particular, key words as “switch panel base” or „steering“ obtain a high weighting.

3. Similarity search via sequences: Structural characteristics of individual assembly processes are considered. In addition, the question of how many same sequences of individual time blocks are used in a parent sequence is analyzed. The more identical or similar sequences of time blocks, the more similar the considered assembly processes.

Figure 6 presents the automated similarity search with the tool “Data Miner“. In different process blocks, data are read, processed and analyzed with different similarity searches.
5. Conclusion and outlook

Through utilization of Data Mining tools, the efficient design of assembly connection, the quality of planning results and planning processes can be increased, while simultaneously time and cost reduction can be realized. With this approach working schedules as planning results are based on field-tested assembly processes and contain the implicit knowledge used in similar assembly planning processes. The automatic generation of an adapted assembly process enables the fast customization to the concrete setting at the shopfloor. In this regard, the presented approach contributes an important added value to production design and planning through usage of knowledge in the existing systems. Thus, reduction of planning time, increasing availability of information in product design as well as making the cooperation between the designer and product planning teams easier are the consequences. The feasibility of this concept with productive data is in evaluation. New approaches with clustered data to improve data quality are in an assessment. However, to produce reliable outcomes, the product data have to fulfil high requirements in regard to connection elements. Concurrently, the necessary data model and some tool sets are provided to make the data integration easier. In the future, further development of tool sets and methods could help to reduce the high initial effort for adjustment of the data even more. Besides the evaluation of the results based on product data, it is important to investigate the behaviour and results of the methodology for new and innovative assembly technologies. Furthermore, for analysing more complex data sets as well as obtaining better results, it is important to develop and refine the concept and to apply further Data Mining methods.

Acknowledgement

The research project “Prospective Determination of Assembly Work Content in Digital Manufacturing (ProMondi)” is supported by the German Federal Ministry of Education and Research (BMBF) within the Framework Concept “Research for Tomorrow’s Production” (funding number 02PJ1110) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

References
