Moving Integrated Product Development to Service Clouds in the Global Economy
J. Cha et al. (Eds.)

doi:10.3233/978-1-61499-440-4-379

Design for Assembly in Series Production by Using Data Mining Methods

Ralf KRETSCHMER [a], Stefan RULHOFF [b] and Josip STJEPANDIĆ [b,1]
[a] *Miele & Cie. KG, Germany*
[b] *PROSTEP AG, Germany*

Abstract. Decision making in early production planning phases is often based on vague expert knowledge due to lack of a reliable knowledge base. Virtual planning has been prevailed as a method used to evaluate risks and costs before the concrete realization of production processes. This paper introduces a new concept and the corresponding data model for Design for Assembly by using Data Mining (DM) methods in the field of series production. The approach adopts the usage of existing planning data in order to extrapolate assembly processes. Especially linked product and process data allow the innovative usage of Data Mining methods. The concept presents assistance potentials for development of new products variants along the product emergence process (PEP). With this approach an early cost estimation of assembly processes in series production can be achieved using innovative Data Mining methods as shown in an industrial use case. Furthermore, design and planning processes can be supported effectively.

Keywords. Product Realization, Manufacturing, Digital Factory, Assembly, Process Planning, Data Mining

Introduction

Increasing variability of products, shortened product lifecycles and corresponding complexity of processes as well as huge market fluctuations are the main challenges for modern manufacturers [1]. Production planning gains in importance and has to run as parallel as possible to the product development according to concurrent engineering principles [2]. In this early phase of the product creation, a first step is a cost calculation for the industrialization of the product in existing production lines regarding basic conditions [3]. The economic feasibility of series production must be assured with vague information on the product and given general conditions, e.g. shift model [4] where the cost-intensive assembly is a big challenge among others [5, 6].

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 [7]. 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

[1] Corresponding Author, Mail: Josip.stjepandic@opendesc.com.

data in order to extrapolate assembly processes. Especially linked product and process data allow the innovative usage of DM methods. Facilitating such an interconnection of highly interdependent models and historical data requires the identification and assessment of the character of interdependencies between the models. 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 has been widely used in various areas such as business, medicine, science, and engineering [8, 9]. 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 [10]. 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 [11]. 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. In previous report [7] authors have described the initial approach.

1. 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 product emergence process (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 [12]. 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 [13]. In this regard, the presented concept focuses on product design and production assembly planning [14]. 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 [15].

1.1. Preparation and Requirements

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

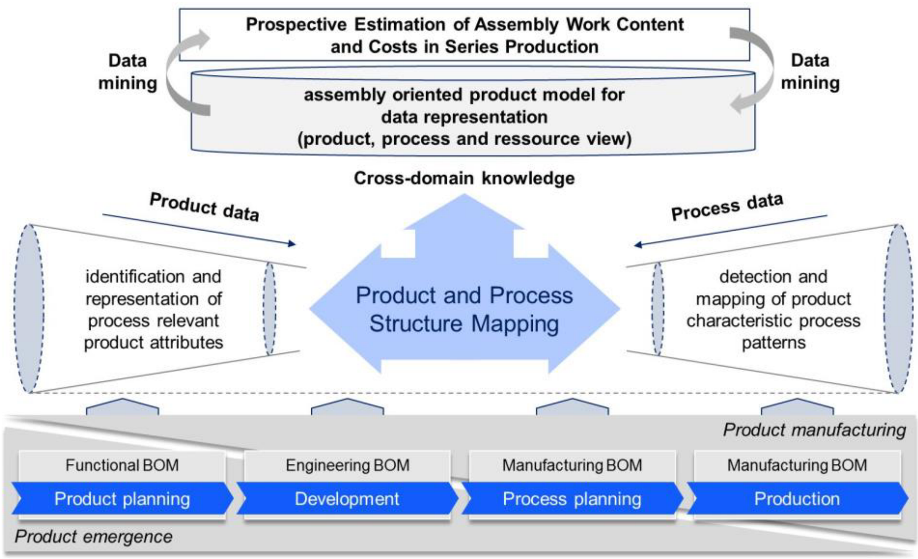

Figure 1. Design optimization with additional time data

Enrich CAD data with assembling information: Derived from the similarly designed products in the past, assembly information such as time data for singular process steps about the actual design situation can be identified, extracted and provided in order to give support for the designer. This additional information can be used to enrich the CAD data to assist 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. Subsequently, 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 product development processes can mutually enrich each other and creates significant added value. The newly obtained information supports the entire workflow throughout the PEP. Therefore, as part of this concept, certain requirements need to be fulfilled. Therefore, the pre-conditions assigned with both systems as well as their respective processes have to be fulfilled [16].

1.2. 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 [17]. Typical context of Data Mining procedure is shown in Figure 2 [18]. 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 [19], 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 [20]. 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 analysis are capsulated as an object oriented data model, further described in chapter 3.

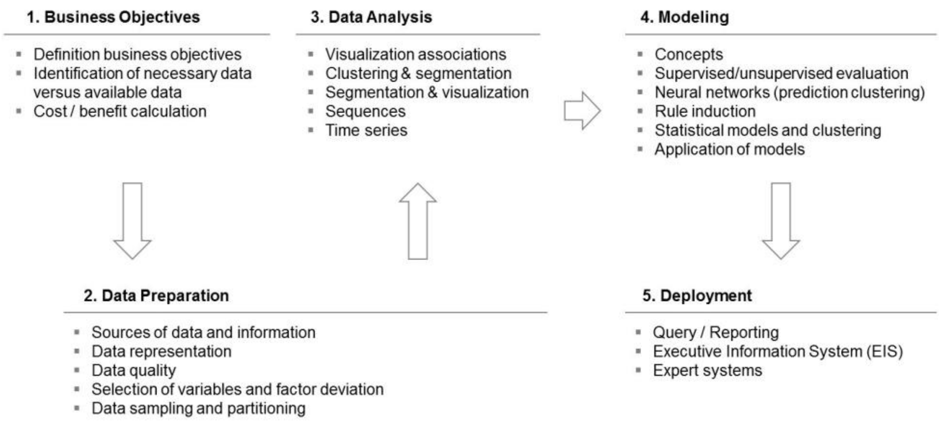

Figure 2. Typical context of data mining procedure

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.

1.3. 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 informations 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.

Basically, in current planning systems it is often possible to directly link processes to the corresponding products [21]. Therefore, an allocation of 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 special, newly created object "product assembly information". The linking of product and process does not necessarily need to occur at the part level.

2. Solution Concept

The solution concept encompasses an assisting workflow to support the designer. As part of a new, modified or variant design, the designer generates new product data. In case of assembly connections, a software assistant shall support the designer on-the-fly with product assembly information for each connection (e.g. the torque screwed fasteners or the type and the form of a welded joint and information about other connection types). The designer can trigger an evaluation regarding the assembly connections in the model.

For this purpose, the product assembly information of the CAD model are first prepared, analyzed with Data Mining, and compared with 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. A limitation on the particular type of connection yields as a result of the closest realized assembly connection of the same kind.

The software assistant uses the product assembly information identified in the analysis of the PDM database to determine the respective associated and related sub processes. Therefore, the corresponding time information of the existing products and, if requested, an alternative proposal list are 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.

2.1. Data Model

Based on determined assembly characteristics, a range of attributes is derived to classify the assembly of the parts. The connections between parts gained a particular importance. An overview of the generated data model for the data mining analysis is given in the previous report [11]. Further connection types can be added to the data model. To provide the required information for the time analysis, a standardized data model is applied. In this regard, ADiFa project's "Application-specific data models", so called ADiFa Application Protocols, were used, which offer the integration of processes and data for different DM systems [20].

Not all of these attributes can be identified in the CAD system. Some can be determined in production planning workshops in order to optimize the current design. Experienced designers and production planner can pre-allocate 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.

2.2. Data Mapping and Data Mining

After aggregating and appending the data subsets from different sources, it is necessary to remove redundant data sets [21] 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. Value and scale of different attributes are often 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. 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. Weights 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. 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 [22]. To identify the objects with most similar product assembly information for a new object, the classification algorithm k-nearest neighbour (kNN) [23] 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 [24] 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 a product assembly information as well as the quality of the data regarding their attributes are important. The fulfillment of those high requirements has to be verified. Methods to improve the quality of the linkage of product data with the corresponding assembly processes will be evaluated. Is this task solved, the selection of the properties and attributes for the DM analysis has to be determined based on production data to ensure the reliability of generated results (Figure 3). In this scope a special focus is on the characteristics of the parts and of the connection itself. Utilization of the methodology is described as follows.

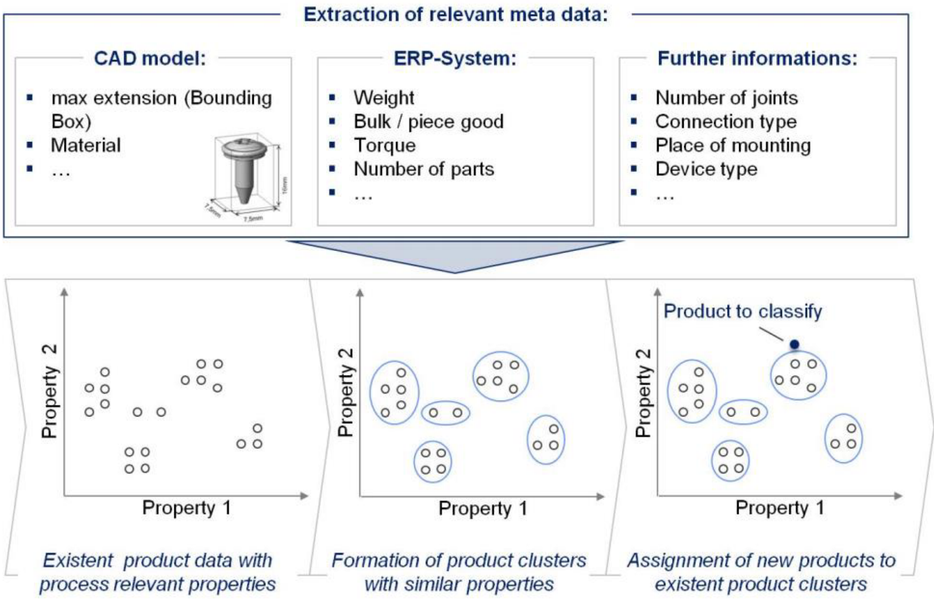

Figure 3. Formation of product clusters and process agglomerations

Suggesting assembly connections and enrichment of CAD data: Designer creates a new module with already known and new assembly connections in CAD system. He defines assembly connections and complements their properties in the context of the new module. Via the automated DM process, he is provided with 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 informations are made available to the designer by 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. 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 an assembly at an early stage of product development. 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 initial draft of first assembly process. The founded individual connections, the individual process, as well as the overall process can be used 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. The production planner can increase the quality of the process by manual intervention. He adapts the product assembly information created by the designer before the DM analysis and 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.

3. Use Case Evaluation

3.1. 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.

3.2. 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 clusters 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 is 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 4 presents the automated similarity search with the tool "Rapid Miner". In different process blocks, data are read, processed and analyzed with different similarity searches.

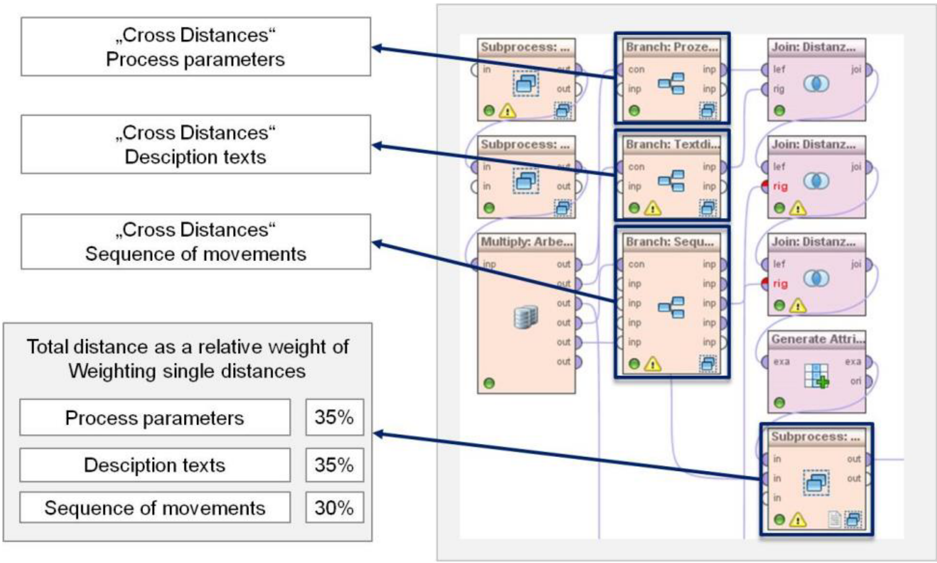

Figure 4. Similarity search for process data

4. Conclusions and Outlook

Through the efficient design of assembly connections supported by Data Mining tools, the quality of planning results and planning processes can be increased, while simultaneously reduce time and cost. 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 shop-floor. The presented approach contributes an important added value to design and production planning through usage of knowledge in the existing systems. Consequences are reduction of planning time, increasing availability of information as well as the better collaboration between design and production planning. The technical feasibility of the proposed solution has been shown by a prototypical implementation of the concept in CAD and PDM systems. New approaches with clustered data to improve data quality are in an assessment. Further development of tool sets and methods could help to reduce the high initial effort for adjustment of the data.

Acknowledgements

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). Authors are responsible for the contents of this publication.

References

[1] A. McLay, Re-reengineering the dream: agility as competitive adaptability, *Int. J. Agile Systems and Management, Vol. 7, No. 2, 2014, pp. 101–115*.

[2] U. Bracht, T. Masurat, The Digital Factory between vision and reality*, Computers in Industry 56, pp. 325-333*, 2005.

[3] H. Bley, C. Franke, Integration of Product Design and Assembly Planning in the Digital Factory, *Annals of the CIRP, Vol. **53/1**, pp. 25-30*, 2004.

[4] G. Boothroyd, *Assembly Automation and Product Design*, Second Edition, Taylor & Francis Group, Boca Raton, 2005.

[5] B. Lotter, H.-P. Wiendahl, *Montage in der industriellen Produktion, Ein Handbuch für die Praxis*, 2. Auflage, Springer-Verlag Berlin-Heidelberg, 2013.

[6] A. Bryan, J. Ko, S. J. Hu, Y. Koren, *Co-Evolution of Product Families and Assembly Systems*, Annals of the CIRP Vol. 56/1/2007, pp 41 - 44.

[7] R. Kretschmer, S. Rulhoff, J. Stjepandić, Prospective Evaluation of Assembly Work Content and Costs in Series Production, in: *Bil, C. (ed..): Proceedings of 20th ISPE International Conference on Concurrent Engineering (CE2013), Sep, 2 - 5 2013*, Melbourne, Australia, IOS Press, Amsterdam, pp. 421- 430, 2013.

[8] Y. Yin, I. Kaku, J. Tang, J. M. Zhu, *Data Mining: Concepts, Methods and Applications in Management and Engineering Design*, Springer-Verlag, London, 2011.

[9] D. Talia, P. Trunfio, *Service-Oriented Distributed Knowledge Discovery*, CRC Press, Boca Raton, 2013.

[10] W. W. Chu, *Data Mining and Knowledge Discovery for Big Data. Methodologies, Challenge and Opportunities*, Springer-Verlag, Berlin Heidelberg, 2014.

[11] H. Al-Mubaid, E. S. Abouel Nasr, A.K. Kamrani, Using data mining in the manufacturing systems for CAD model analysis and classification. *Int. J. Agile Systems and Management, Vol. 3, Nos. 1/2, 2008, pp. 147-162*.

[12] F. Demoly, A. Matsokis, D. Kirisits, A mereotopological product relationship description approach for assembly oriented design, *Robotics and Computer-Integrated Manufacturing*, 28 (2012), pp. 681–693.

[13] O. Erohin, P. Kuhlang, J. Schallow, J. Deuse, Intelligent Utilisation of Digital Databases for Assembly Time Determination in Early Phases of Product Emergence, *45th CIRP Conference on Manufacturing Systems 2012, Vol. 3, pp. 424-429,* 2012.

[14] E. B.Magrab, S. K. Gupta, F. P. McCluskey, P. A. Sandborn, *Integrated product and process design and development: the product realization process*, second edition, Taylor & Francis, Boca Raton, 2010.

[15] S. Rulhoff, R. Jalali Sousanabady, J. Deuse, C. Emmer, Concept and Data Model for Assembly Work Content Determination, *Enabling Manufacturing Competitiveness and Economic Sustainability Proceedings of 5th CIRP Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2013), Zäh, J. (ed.), München, Germany, Springer, Berlin, Heidelberg, New York, 2012, pp. 353-360*.

[16] J. Schallow, K. Magenheimer, J. Deuse, G. Reinhart, Application Protocols for Standardising of Processes and Data in Digital Manufacturing, in: *ElMaraghy, H. A. (Hrsg.): Enabling Manufacturing Competitiveness and Economic Sustainability - Proceedings of 4th CIRP Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2011), 2.-5. October 2011*, Montreal, Canada, Springer, Berlin, Heidelberg, New York, pp. 648-653, 2011.

[17] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, third edition, Morgan Kaufmann Publishers, Waltham, 2012.

[18] D. Nettleton, *Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects*, Morgan Kaufmann, Waltham, 2014.

[19] J. Hartung, J. Schallow; S. Rulhoff, Moderne Produktionsplanung - Integration in der Produktentstehung, *ProduktDaten Journal 19 1, pp. 20-21*, 2012.

[20] D. Petzelt, J. Schallow, J. Deuse, S. Rulhoff, Anwendungsspezifische Datenmodelle in der Digitalen Fabrik, in: *ProduktDaten Journal 16 1, pp. 45-48*, 2009.

[21] L. Ohno-Machado, H. S. Fraser HS; A. Øhrn, Improving Machine Learning Performance by Removing Redundant Cases in Medical Data Sets, *AMIA Fall Symposium, pp. 523-527*, 1998.

[22] D. Zhang, P. L. Yu, P. Z. Wang, State-dependent weights in multicriteria value functions, *Journal of Optimization Theory and Applications, Vol.74, No.1, pp. 1-21*, 1992.

[23] S. Dhanabal, S. Chandramathi, Review of various k-Nearest Neighbor Query Processing Techniques, *International Journal of Computer Applications Vol. 31, No.7*, 2011

[24] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *14th international joint conference on Artificial intelligence, Vol. 2 (IJCAI'95*), Morgan Kaufmann Publishers Inc., San Francisco, pp. 1137-1143, 1995.