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

doi:10.3233/978-1-61499-440-4-261

Intelligent Utilization of Digital Manufacturing Data in Modern Product Emergence Processes

Regina WALLIS [a,1], Josip STJEPANDIC[b], Stefan RULHOFF[b], Frank STROMBERGER [a] and Jochen DEUSE[c]

[a]*Daimler AG, Digital Planning Methods Daimler Trucks, 68299 Mannheim, Germany*
[b]*PROSTEP AG, 3D Product Creation, 64293 Darmstadt, Germany*
[c]*TU Dortmund University, Institute of Production Systems, 44227 Dortmund, Germany*

Abstract. The application of digital manufacturing tools has been continuously increasing in order to deal with product and process complexity in shortened product lifecycles. The resulting comprehensive digital documentation of the product emergence process provides an opportunity to support concurrent engineering processes. By identifying correlations and recurrent patterns with the aid of data mining techniques, tacit planning knowledge can be revealed and reintegrated into new process planning workflows in order to enhance planning efficiency and facilitate decision making. Based on the classification and clustering of both product and process data and the determination of their respective linkages, this paper presents a novel approach for the knowledge-based support of product emergence processes.

Keywords. Digital manufacturing, digital process planning, data mining, concurrent engineering, planning support.

Introduction and Motivation

Manufacturing companies are facing the challenge of developing and producing a continuously rising number of product variants in shortened product lifecycles to satisfy customer demands. Especially in assembly planning the resulting process complexity becomes apparent [1] and is difficult to keep under control, as all product and process variants have to be managed at the same time. The application of digital manufacturing tools has been increased in order to approach this complexity.

As a result, a digital documentation of product design and assembly planning results is available [2]. These data may include patterns, trends, associations and dependencies [3]. Thus Knowledge Discovery in Databases (KDD) methods can be used to identify these patterns representing tacit planning knowledge. By increasing the transparency in planning processes, planning efficiency can be enhanced and decision making facilitated.

In this context, the research and development project “Prospective Determination of Assembly Work Content in Digital Manufacturing (Pro Mondi)” has been initiated to develop a concept using KDD methods to extract useful assembly information from

[1]Corresponding Author.

digital manufacturing databases. Aim of this project is the accurate estimation of the expected assembly work content and the resulting costs of a newly designed product at an early stage of the product emergence process.

The approach to achieve this support function comprises a detailed data analysis from the areas of product design and assembly planning. The identification of correlations in the combined datasets allows the product-specific evaluation of expected assembly process complexity, time and cost. The evaluation results can be used to support the assembly planner on the one hand with a first estimation of the required assembly processes as initial starting point for following planning activities. On the other, product development can be supported with information regarding the assembly process complexity for the current product design alternative enabling to optimize the product in terms of assembly quality.

1. Data Mining to Support Product Emergence Processes

The approach to support product emergence processes with processed assembly knowledge requires a detailed data analysis from the areas of product design and assembly planning. After a description of the existing data structures in this context, the overall concept of KDD will be highlighted. Yet, the application in the manufacturing field is connected with a number of challenges, which will be depicted at the end of this chapter.

1.1. Product and Process Data in Assembly Planning

The product-related input data for assembly planning contains information such as e.g. the shape, size and material of the work piece. It is typically represented by the bill of materials and the 3D shape representations of individual parts. The engineering product structure – represented as a hierarchical tree with the final product being the root node – forms the basis for the systematic storage and management of this data [4], commonly in product data management (PDM) systems. These have been developed in the context of Computer Aided Design to support product design and construction processes. Product Lifecycle Management (PLM) solutions, which focus on the system application alongside the entire product lifecycle including process planning, production and after sales management, are seen to be the enhancements of PDM systems [4].

First support functions for assembly planning have been developed in the context of the Computer Aided Process Planning systems. These are able to generate work plans based on the product description. The required planning knowledge is provided in an organized and formalized way [5], e.g. in the form of “if-then”-rules. However, with an increasing number of rules, these systems often lack transparency and reach their limits with regard to efficiency and maintainability. The idea of an integrated product- and process planning and consequently of the central storage of product, process and resource information has been focused in digital manufacturing systems [6]. Product, process and resource data are stored in separate tree data structures, whose elements are interlinked by references. The databases composing the backbones of these IT tools provide a comprehensive documentation of product design and assembly planning results. They form the basis for the subsequent KDD analysis.

1.2. Knowledge Discovery in Databases

KDD describes the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [7]. One of the leading models describing the procedure of knowledge discovery with high practical relevance is the Cross-Industry Standard Process for Data Mining (CRISP-DM), which consists of the six following steps [8]: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment (Figure 1).

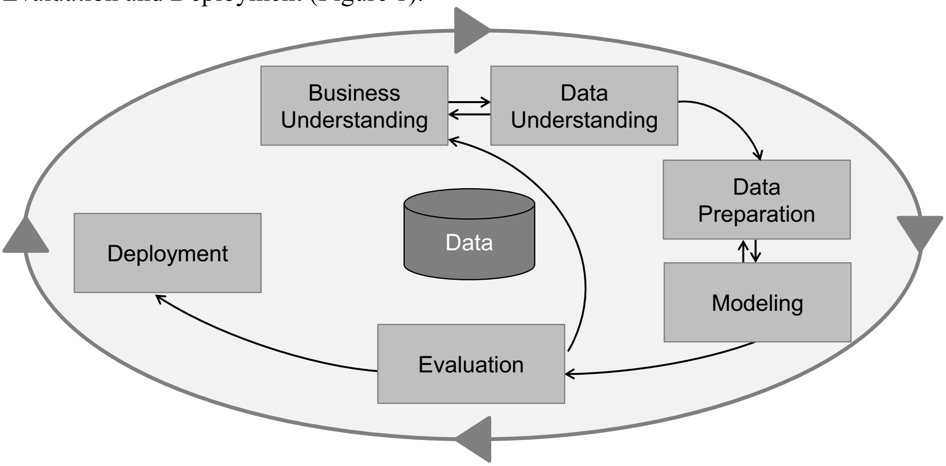

Figure 1. CRISP-DM process steps [8].

During the phase of Business Understanding, both the objectives of knowledge discovery and the nature of the data mining task are defined. In this context, data mining describes the concept of applying data analysis and discovery algorithms to produce a particular enumeration of patterns or models over data [9]. In general, predictive and descriptive data mining tasks can be distinguished. While predictive data mining is applied to find relationships between a dependent (target) variable and the independent variables in the dataset in order to make predictions on new and unlabeled data, descriptive data mining serves to produce understandable and useful patterns describing a complex dataset, yet without any prior knowledge of what patterns exist [10].

Availability, quality and possibilities of access of relevant data are assessed within Data Understanding. In the subsequent phase of Data Preparation, the often time-consuming transformation of data to a suitable format for data mining is performed. The transformation comprises e.g. the breaking down of hierarchical data structures to flat data tables, the selection of specific attributes or the removal of noisy data. The application of specific data mining techniques to extract patterns from data is performed in the Modeling step.

Within the Evaluation phase, results are assessed and interpreted by experts of the future application area. Assuming positive and valid results, the developed model can be deployed in the considered field and consequently be used e.g. for planning- and decision-making support.

1.3. KDD Challenges in the Context of Product Design and Assembly Planning

In the approach to provide a support function for future product emergence processes, several challenges have to be taken into account: The above mentioned heterogeneous data sources contain data characterized by differences in level of measurement and hierarchical structuring. With regard to the data mining activities, the transformation of object-oriented and hierarchical data structures into flat data tables is necessary. A unique identifier is required to map product structures with their corresponding assembly processes. Hereby, a consistent 1:1 mapping of parts to assembly work steps cannot be taken for granted.

A model for Knowledge Discovery in Industrial Databases (KDID) has been developed, which incorporates in particular the issues related to industrial databases and the utilization of domain knowledge provided by experts [11].

2. Approach for the Intelligent Utilization of Digital Manufacturing Data

The application of KDD including the usage of linked data of different application contexts enables additional benefits for involved departments in the product emergence process. In the following approach assembly planning and product design gain additional and automatically generated information as an input to their particular tasks in form of a support function.

2.1. Provision of a Support Function for Production Planning and Product Design Departments

The general concept for the intelligent utilization of digital manufacturing data is illustrated in Figure 2. It contains the automatic identification of similar product design alternatives in existing product data and the retrieval of associated assembly work plans in order to evaluate the complexity and costs of the assembly process.

Two use cases have been identified to support both assembly planning and product design departments. The first use case represents a support function for the assembly process planner. As soon as a new subassembly is created by the product design department, the assembly planner can trigger the automated generation of the product-specific assembly process plan. Since the new subassembly already contains characteristic attributes in geometry and model structure, it can be initially classified into a defined subassembly type. These predefined types are provided by a data mining model based on previously designed and analyzed product data. After this step the new subassembly is assigned to a product cluster within the specified assembly type. The identified product cluster is linked to the associated process cluster. A product-process-mapping, which has been trained on the linkages between existing subassemblies and their assembly work plans, retrieves the corresponding process cluster for the given new subassembly. For each process cluster a specific assembly work plan template can be generated containing typical operation sequences for this type of process cluster. This automatically generated assembly process represents a first estimation of the assembly plan for the new subassembly still in the product design process. The assembly planner obtains a starting point for his following planning activities with comparably little planning effort. With the help of further manual adjustment and completion, the planner determines the exact assembly process, time and costs.

The product design process can be supported by means of the second use case. The required assembly processes for the newly constructed subassembly can be determined as described above. By comparing the resulting assembly process complexity and time of the current product alternative to process complexity and time of previously designed product alternatives, the product designer obtains feedback about the assembly quality, which, in turn, can be used to optimize the current design.

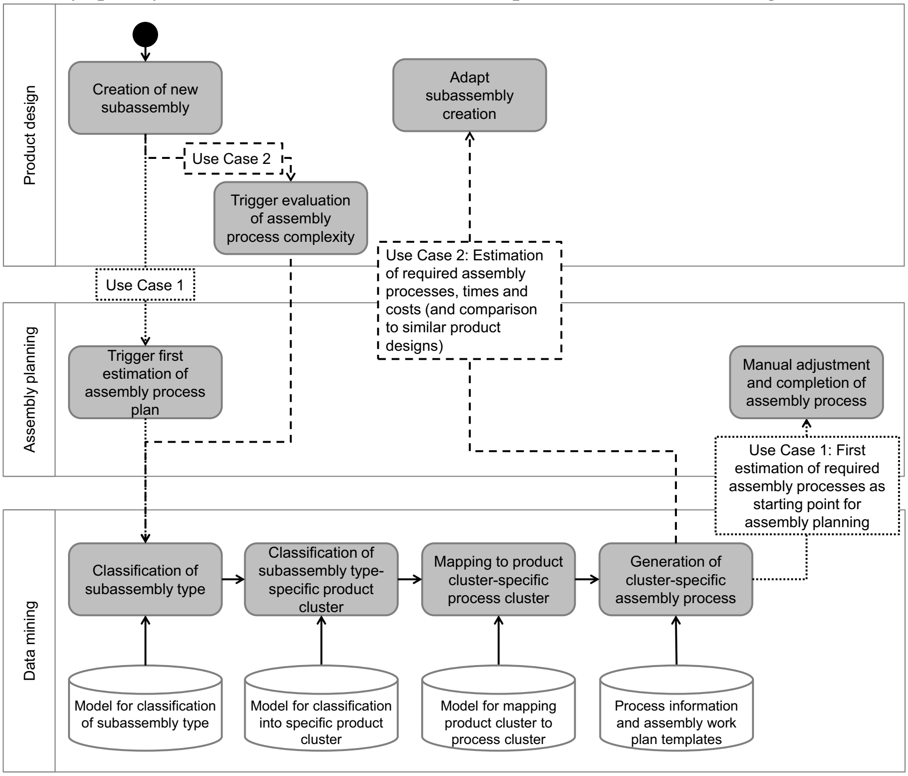

Figure 2. Concept for the intelligent utilization of manufacturing data

2.2. Learning of a Data Mining Model to Estimate Assembly Processes

The learning of a data mining model to estimate suitable assembly processes for a given product structure is set up in seven main processes. Thereby, four different classification models are created, which are required for the model application in the above mentioned use cases (see Figure 3). In the first step, the hierarchical product structures are transformed into flat data tables and aggregated to subassembly-specific feature vectors. These are used to train the model for the classification of the subassembly type in step 2. Regarding e.g. engine assembly, cylinder head, crankcase and electric generator are just a few examples for subassembly types.

Once being differentiated, a clustering algorithm is applied to all variants of each subassembly type in order to find similar subassemblies within one type and to reduce variant complexity (step 3). The resulting cluster assignments serve as target variables for the learning of the classification model into specific process clusters. In the fourth step, the process data extracted from the digital manufacturing system are brought into a suitable form for data mining. Hereby, the textual descriptions of the work plans are transformed into word vectors with the help of text mining algorithms. Additionally, building blocks from predetermined motion time systems (PMTS) are used to characterize the operations in detail. These building blocks are machine readable and serve as additional attributes for the subsequent clustering algorithm (step 5). The required process information such as characteristic operation sequences and the assembly time is stored with the process clusters in step 6.

Due to the existing single linkages between elements of the product structure and their required assembly processes, which are available in the digital manufacturing system, a mapping model can be computed to assign product clusters containing subassemblies with similar characteristics to their respective process clusters (step 7).

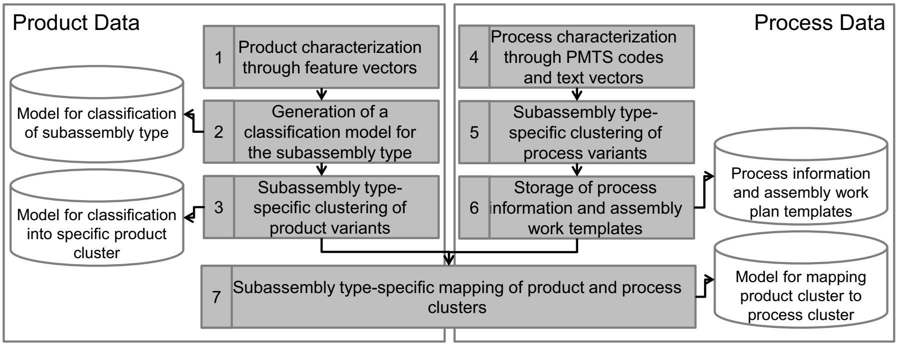

Figure 3. Data mining approach

2.3. Enhanced Data Model for Product and Assembly Process Data

An enhanced data model is defined in order to support the KDD process as well as the exchange and persistent storage of linked product and process data. An overview of the data model including attributes relevant for both the analysis and the results is presented in Figure 4.

Modern digital manufacturing systems allow the direct assignment of processes to product data. Some of them also provide additional attributes describing the assembly connection, which concern e.g. the modeling of joining or quality control processes. However, these attributes lack the ability to store detailed information regarding the specific assembly connection. Thus, the class *ProductAssemblyInformation* has been set as a central element in this data scheme and represents an assembly connection realized in the product assembly. References to the corresponding time analysis, assembly requirements, designed parts or products as well as to a wide range of meta data including the assembly department are stored in the data model. The detailed concept for modelling product assembly information can be found in [12].

The class *ProductAssemblyInformation* is supplemented with attributes of different connection types, such as e.g. screw or welded connections. The class *CustomAssemblyInformation* allows the instantiation of all possible connection type objects not specifically defined in the data model. Furthermore, the class *AssemblyConditionList* contains information regarding the assembly situation of the particular assembly connection at the product. This information usually represents tacit knowledge of the production planner and is implicitly included in the resulting assembly process.

Item represents the second fundamental class in the data model. It contains references to existing subassembly units, geometrical characteristics and further meta data. Each *Item* refers to *ProductAssemblyInformation*, which can in turn refer to further *Items*. This construct is chosen to enable data mining methods to determine exact similarities between new parts and/or products and other existing parts. Furthermore, it allows the comparison between new and existing parts and products in any order and combination.

As to represent assembly process data in the data model on hand, parts of another existing data model are used. Application-specific data models have been developed in the research and development project "ADiFa" (Process Harmonization based on Application Protocols). The so-called "ADiFa Application Protocols" offer the integration of processes and time-related data for different digital factory systems [13], [14]. Two basic classes in this data model are *OperationDefinition* and *TimeModulOccurrence*, which allow the representation of generic process definitions as well as hierarchical assembly process structures with instantiated operations in a work schedule.

The requirement to support the data mining process and to store the data mining models results in a fourth part of the data model. This part comprising the classes *ClusterModel* and *ClassificationModel* allows the persistent storage of training sets, the resulting product and process cluster as well as the deduced product-process-mapping.

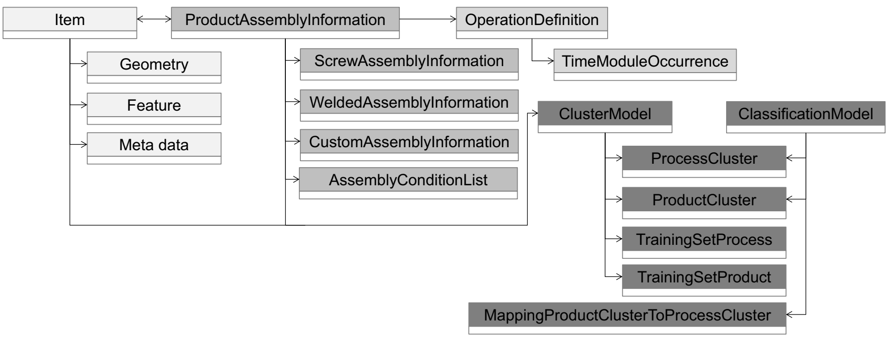

Figure 4. Data model overview

3. Case Study on Daimler Data

The described concept for the intelligent utilization of digital manufacturing data to support future product emergence processes has been applied to three multi-variant subassembly types of the automotive industry according to the CRISP-DM process steps. RapidMiner 5.3 [15] has been used for the implementation.

3.1. Data Preparation

The required input data for the described concept are extracted from the PDM and digital manufacturing system. The hierarchical data structures of both product and process data are transformed into flat data tables first.

Regarding product data, relevant characteristics such as the outer dimensions, the weight and the center of gravity for the assembly situation are extracted from the 3D shape representation and stored in a subassembly-specific feature vector. This feature vector is enriched with additional information describing the content of the bill of materials, gathered from the various hierarchical levels of the engineering product structure.

As to process data, the textual descriptions of the assembly operations are enriched with the detailed itemization of process steps from a predetermined motion time system. In this way, a flat data structure containing detailed information about the assembly operations can be obtained.

3.2. Modeling

In order to distinguish between the various subassembly types automatically, a classification model is trained. Therefore, the value differences of the feature attributes for different subassembly types are analyzed and exemplarily visualized in Figure 5.

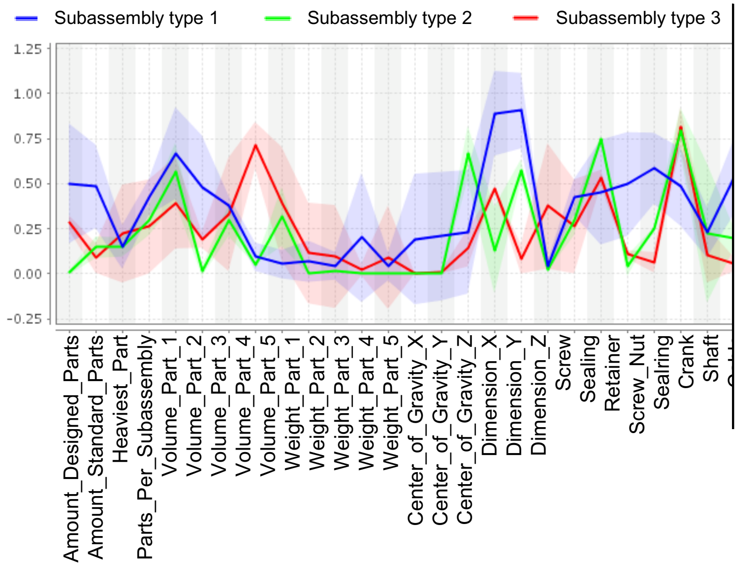

Figure 5. Visualization of feature vectors

The lines show the mean attribute values for each subassembly type and the deviations are added in order to demonstrate the attribute value ranges. Based on this input data, a naïve Bayes classification model is trained to extract and formalize the differences in the attribute values and to use this knowledge to classify new and unseen subassembly types. Hereafter, a k-means clustering algorithm is applied to the classified data in order to reduce variant complexity. The resulting partition is checked for plausibility with domain experts. Valid subject-specific groups, e.g. concerning the distinct country or operation editions of the subassembly, have been accomplished.

The data describing the assembly process operations is clustered with a k-means algorithm in parallel to the clustering of product data. Each of the resulting assembly work plan clusters contains similar assembly processes. The information characterizing the process clusters, e.g. the assembly process steps and the required assembly time is stored with the process clusters in order to be able to generate the work plan templates during model application. In the last step, a mapping function between product and process clusters is trained. A naïve Bayes classifier has been chosen to assign the most likely process clusters to the given product clusters. The resulting product-process-mapping indicates the matching probability of a process cluster given a certain product cluster and shows a comprehensive connection between these two.

3.3. Evaluation

In order to validate the described concept, the total dataset has been evenly divided into a training and a testing dataset. After the training has been performed as described in section 2.2, the consecutive classification models have been applied to the product data of the testing dataset. Due to the modular composition of RapidMiner and the possibility to trigger the execution of the defined data mining processes, the generation of assembly operation sequences could be performed in an automated way. The results generated for the subassemblies of the testing dataset have been compared with the assembly operation sequences actually developed by assembly planners. On average, 87.8% of the assembly operations have been determined correctly. This shows that data mining techniques allow the segmentation of product data in valid subject-specific groups and the subsequent mapping of adequate assembly operations.

Yet the worth achieved by the generated assembly operation sequences is directly associated to the quality of the training data: If the classification models are applied to complex product subassemblies and the training data does not comprise comparable datasets, the model will not be able to “predict” the required (complex) assembly operations. In this case the model generates the basic (routine) assembly operations and missing operations need to be added manually. However, the adjusted planning data will be added to the analysis datasets for the recalculation of product and process clusters so that mapping model can be refined with every iteration loop.

4. Conclusion and Further Developments

The developed approach for the intelligent utilization of digital manufacturing data provides a new support function for modern product emergence processes. As the presented approach is based on planning data compiled during preceding product emergence processes, products can be evaluated more easily concerning their assembly process complexity, which leads to a faster and easier attainment of planning and construction levels. At present, the realization of the first use case is advanced. The feasibility to segment product data in valid subject-specific groups and to map adequate product-specific assembly operations has been shown. These findings form the basis for the realization of the second use case.

The created models as well as the obtained results represent the processing of implicit planning knowledge. The persistent storage in the presented data model helps to enhance transparency and promote reusability. The enhancement of current digital manufacturing systems with the presented data model allows the efficient provision of feedback. By means of a consequent feedback concerning the assembly process complexity, product designers can profit from experiences gained in series production. This in turn can significantly accelerate the overall design and assembly planning process and thereby reduce the number of planning iterations. In order to make the information available to the product designer, a further integration of the support function into Computer-Aided Design (CAD) systems, as regular working environment of the product designer, might be contemplated.

Acknowledgements

This paper represents the background, objectives and first results of the research project “Prospective Determination of assembly work content in Digital Factory (Pro Mondi)”. This research and development project is funded 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

[1] H. Bley, C. Zenner, Variant-oriented Assembly Planning, *CIRP Annals Manufacturing Technology***55** (2006), 23-28.
[2] O. Erohin, P. Kuhlang, J. Schallow, J. Deuse, Intelligent Utilisation of Digital Databases for Assembly Time Determination in Early Phases of Product Emergence, *Procedia CIRP - 45th CIRP Conference on Manufacturing Systems***3** (2012), 424-429.
[3] 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***2** (2008) 1/2, 147-162.
[4] M. Eigner, R. Stelzer, *Product Lifecycle Management – EinLeitfadenfür Product Development und Lifecycle Management*, Springer, Berlin, 2009.
[5] P. Mognol, B. Anselmetti, Evaluation Criteria: A Method to Represent and Compute Technological Knowledge in CAPP, *IEEE Symposium on Emerging Technologies and Factory Automation***2** (1995), 419-426.
[6] Digital Factory – Fundamentals. Düsseldorf: VDI; 2008.
[7] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From Data Mining to Knowledge Discovery in Databases, *AI Magazine***17** (1996), 37-54.
[8] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinhartz, C. Shearer, R. Wirth, *CRISP-DM 1.0 – Step-by-step data mining guide*, SPSS, 2000.
[9] U.Fayyad, G. Piatetsky-Shapiro, P. Smyth, From Data Mining to Knowledge Discovery: An Overview, *Advances in Knowledge Discovery and Data Mining*, AAAI Press,Mensolo Park, 1997.
[10] K. Wang, Applying data mining to manufacturing: the nature and implications, *J IntellManuf***18** (2007), 457-495.
[11] D. Lieber, O. Erohin, J. Deuse, Wissensentdeckung im industriellen Kontext, *Zeitschrift für wirtschaftlichen Fabrikbetrieb* **108** (2013), 388-393.
[12] M. Eigner, J. Ernst, D. Roubanov, J. Deuse, J. Schallow, O. Erohin, Product Assembly Information to Improve Virtual Product Development, *Proceedings of the 23rd CIRP Design Conference - Smart Product Engineering* (2013), 303-313.
[13] D. Petzelt, J. Schallow, J. Deuse, S. Rulhoff, S., Anwendungsspezifische Datenmodelle in der Digitalen Fabrik, *ProduktDaten Journal***16** (2009), 45-48.
[14] J. Schallow, K. Magenheimer, J. Deuse, G. Reinhart, Application Protocols for Standardising of Processes and Datain Digital Manufacturing, *Enabling Manufacturing Competitiveness and Economic Sustainability - Proceedings of 4th CIRP Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2011)*,Springer, Berlin / Heidelberg / New York, 2011.
[15] I. Mierswa, M. Scholz, R. Klinkenberg, M. Wurst, T. Euler, YALE: Rapid Prototyping for Complex Data Mining Tasks, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006), 935–940.