Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 23 (2014) 178 – 183

www.elsevier.com/locate/procedia

Conference on Assembly Technologies and Systems

Data Mining-supported Generation of Assembly Process Plans

R. Wallis[a]*, O. Erohin[b], R. Klinkenberg[c], J. Deuse[b], F. Stromberger[a]

[a]*Daimler AG, Hanns-Martin-Schleyer-Str. 21-57, 68305 Mannheim, Germany*
[b]*TU Dortmund University, Institute of Production Systems, Leonhard-Euler-Str. 5, 44227 Dortmund, Germany*
[c]*RapidMiner GmbH, Stockumer Str. 475, 44227 Dortmund, Germany*

* Corresponding author. Tel.: +49-621-393-7734; fax: +49-711-3052115938. *E-mail address:* regina.wallis@daimler.com

Abstract

The application and functional scope of digital assembly planning tools have been permanently increasing in order to deal with product and process complexity. Consequently a large amount of assembly-related data is stored in different systems alongside the product emergence process. By means of data mining techniques an intelligent utilization of this data can be accomplished for future assembly planning. This paper presents an approach for data mining-supported generation of assembly process plans to enhance planning efficiency. The approach is based on the classification and clustering of both product and process data as well as on the identification of their correlations.

Selection and peer-review under responsibility of the International Scientific Committee of 5th CATS 2014 in the person of the Conference Chair Prof. Dr. Matthias Putz matthias.putz@iwu.fraunhofer.de

Keywords: assembly; digital process planning; data mining; variant management; planning support

1. Introduction

Manufacturing companies are facing the challenge of developing and producing a continuously rising number of product variants in shorter periods of time in order to be competitive. 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 kept at the same time. Manufacturing industry and the automotive industry as a pioneer, has consequently been concentrating on the application of digital manufacturing systems in order to counteract these challenges as they allow to support simultaneous engineering processes and to cope with the multitude of product variants. As a result, big digital databases concerning product and assembly process planning are available nowadays. These databases possess the potential for an improvement of the assembly planning process: By identifying correlations and recurrent patterns, data mining techniques provide an opportunity to discover tacit planning knowledge in these databases, which subsequently can be reintegrated into new assembly planning workflows to enhance planning efficiency and facilitate decision making.

This paper presents a novel approach for the data mining-supported generation of assembly process plans based on the data compiled during the product emergence process. Therefore Section 2 reviews the current state of the art in assembly planning, classification and clustering methods as well as their integration in the procedure of knowledge discovery in databases (KDD). Section 3 subsequently summarizes the challenges and requirements for the improvement of modern assembly process planning. Sections 4 and 5 present the concept of data mining-supported assembly planning and demonstrate first application results in the automotive industry. Section 6 summarizes the results and identifies future research activities.

2. Knowledge discovery based on assembly-related data

The intensified application of digital manufacturing and product data management (PDM) systems results in a large amount of data allowing to document the product emergence

Selection and peer-review under responsibility of the International Scientific Committee of 5th CATS 2014 in the person of the Conference Chair Prof. Dr. Matthias Putz matthias.putz@iwu.fraunhofer.de
doi:10.1016/j.procir.2014.10.095

process thoroughly. The utilization of this product and process data by application of data mining techniques provides a basis for the discovery of the hidden assembly relevant knowledge.

2.1. Product and process data in assembly planning

Product data represents a relevant input dataset for the assembly planning process and is typically stored in PDM systems, which have been developed in the context of Computer Aided Design to support product development and construction. They provide the central storage and management of product-related data [2]. The enhancements of PDM systems are nowadays represented by Product Lifecycle Management solutions, which focus on the system application alongside the entire product lifecycle including process planning, production and after sales management.

First support functions for assembly planning have been developed in the context of the Computer Aided Process Planning systems, which are able to generate work plans based on the product description. The required planning knowledge is provided in an organized and formalized way, e.g. in the form of "if-then-else"-rules [3] or in form of decision trees [4]. However, with an increasing number of rules, these systems often lack transparency and reach their limits with regard to efficiency and maintainability. Since the 1990s 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 [5]. Thereby, product, process and resource data is interlinked in order to describe the assembly processes thoroughly in the phases of assembly planning and production.

The databases composing the backbones of the modern IT tools mentioned above can be utilized to support the assembly planning process. Depending on the nature of the new planning task, [6] distinguishes the repetitive, adaptive, variant or new planning approach. Repetitive planning is applied in the case a former process plan for the exact same part and manufacturing process already exists. Adaptive planning is used, if a similar assembly process plan for the same part is available, but the planning premises have been different at the time. In the case of production-technical similarity, variant planning can be applied. Completely new products imply the generation of a novel process plan [6]. Especially in the case of adaptive and variant planning data mining approaches are potentially useful to identify similar structures in product and process data, e.g. in order to form a part family [7, 8].

2.2. Reduction of product and process complexity by data mining application

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]. To reuse existing knowledge for future product generations and to reduce product and process complexity in assembly planning, classification and clustering as representatives of predictive resp. descriptive data mining methods can be used.

In terms of classification, the input dataset consists of n examples $x=(x_1,...,x_n)$ with each i – th example representing a vector of observed variables $x_i=(x_{i1},...,x_{im})$ of the total set of independent variables $X=(X_1,...,X_m)$ and having a class label g_i [11]. The input dataset serves as the basis for the learning of the function $f(x)=\hat{g}$, which describes the relationship between the input variables x and their correspondent class label g. $\hat{g}$ represents the target value predicted by the learning function. The extracted relationships between attribute values and class labels are used to assign labels to previously unseen data.

A wide spread classification method is the naive Bayes classifier. Hereby, the learning function f is interpreted as a classifier which predicts the class of a given observation x_i based on the probability $p(g_i)$ of the class and on the likelihood $p(x_{i1},...,x_{im}\mid g_i)$ of the feature values given the class g_i [12]. It assumes independency of the attributes $p(x_{i1},...,x_{im})$:

$$p(g_i \mid x_{i1},...,x_{im}) = p(g_i)\prod_{j=1}^{m} p(x_{ij}\mid g_i) \quad (1)$$

In contrast, clustering is a typical approach for the reduction of data complexity by dividing a population of n examples $x=(x_1,...,x_n)$ into smaller subpopulations (clusters), so that the pairwise dissimilarities between the examples in one cluster tend to be smaller than those in different clusters [11].

The k-means clustering algorithm optimizes this criterion and is one of the most popular iterative clustering methods [11]. The parameter k representing the number of clusters must be specified in advance [12]. Being applied to datasets with quantitative variables, the algorithm begins to choose k random data points as cluster centroids. Subsequently, all examples of the input dataset $(x_1,...,x_n)$ are assigned to their closest cluster center using the Euclidean distance as dissimilarity measure. Then the centroid (mean)

$$\mu_k = \sum_{i\in S_k}\frac{x_i}{n_k} \quad (2)$$

of each newly formed cluster S_k is calculated, with n_k being the number of points in S_k. The centroids are set to be the new cluster centers. This process of minimizing the sum of squared errors [13]

$$J_K = \sum_{k=1}^{K}\sum_{i\in S_k}(x_i-\mu_k)^2 \quad (3)$$