52nd CIRP Conference on Manufacturing Systems

Data Analytics for Manufacturing Systems – A Data-Driven Approach for Process Optimization

Florian Ungermann*a, Andreas Kuhnlea, Nicole Strickera, Gisela Lanzaa

* wbk Institute of Production Science, Kaiserstrasse 12, 76185 Karlsruhe, Germany

Abstract

In the course of digitalization many small and medium-sized companies face the challenge of using the existing database for process optimization in manufacturing. Furthermore, the demand-oriented expansion of the database is a great challenge. A lack of competencies, limited financial resources and historically grown data structures, which show a strong heterogeneity and lack of transparency, are the central obstacles. A specific approach, how data analytics projects for process optimization should be carried out in manufacturing, is presented. In particular, the question which sensors should be implemented to expand the database is answered. The approach is applied exemplarily for a manufacturing line.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/)

Peer-review under responsibility of the scientific committee of the 52nd CIRP Conference on Manufacturing Systems.

Keywords: data-analytics; process optimization; manufacturing systems; procedure models

1. Introduction

The ongoing digitalization enables new optimization potentials for manufacturing companies. With a growing number of sensors implemented in manufacturing systems a large amount of data is generated. This data is an important resource to maintain competitiveness in volatile global markets [1]. However, the existing data is often insufficiently used for information acquisition and process optimization in production management [2]. Among manufacturing companies only 5.5% of the available database is used for process optimization applying data-analytics [1]. Central reasons for this are historically grown data structures, which show a large heterogeneity and lack of transparency, limited data access and missing product tracing information [3, 4]. Furthermore, the data sets used for data-analytics are often not of sufficiently high quality and it is difficult for companies to assess their problem-specific adequacy [5, 1]. The demand-oriented enhancement of databases driven by a specific use case in order to achieve a higher quality level and amount of data is a great challenge. It is associated with financial expenditures [7] (e.g. for additional sensors and data preprocessing) and organizational barriers (e.g. series release of a manufacturing line might be lost, when additional sensors are implemented). Therefore, production engineers have to estimate which additionally gathered data could potentially contribute to the intended process optimization before changing the production equipment or resources.

There are established process models on how to approach data-analytics projects, such as the CRISP-DM model [8, 9]. However, these models are neither backed by practical and production-related methods and tools, nor answer the question of the potential of additional data. Furthermore, the design of a problem-specific adequate database, which integrates not only sensor data but also the knowledge of manufacturing process experts, is not explicitly formulated [10]. The approach presented in this paper addresses these deficits. A methodology is developed, which can be applied as a guideline for data-analytics projects within process optimization of manufacturing systems. Based on expert knowledge, a
selection of additional sensors is part of the approach. Furthermore, the evaluation and selection of process optimization measures are included.

The paper is structured as follows: the next section deals with existing process models for implementing data-analytics projects, approaches for expert-based process modelling and Key Performance Indicators (KPIs) for the evaluation of optimization measures. In Section 3 the developed approach is presented. An exemplary application using the example of a real manufacturing line is given in Section 4. The last section provides a summary.

2. State of the Art

An overview of existing process models for data-analytics projects is presented subsequently. Furthermore, approaches for expert-based process modeling are discussed. In Section 2.3 KPIs are considered as a basis for evaluating process improvement measures.

2.1. Implementing Data-Analytics Projects for Process Optimization

According to Russom [11] data-analytics is defined as advanced analytic techniques that operate on a large database. Data-analytics is therefore generally the statistical and mathematical analysis of data, also including data mining. Data Mining is used to extract knowledge out of data that is generally applicable, not trivial, new, useful and understandable. For this knowledge retrieval process, machine learning algorithms are nowadays broadly and successfully used. [12, 13] For the application of data-analytics, different holistic end-to-end process models exist. The KDD-Model is one process model for data mining [12]. It is an interactive and iterative process following nine steps. In addition to the actual data mining phase, the procedure also includes the steps of preparing the data and evaluating the results. The KDD model forms the basis for quite a few more recent process models like for instance the KDD-Roadmap [14], which additionally determines resources required for the project or marketing specific developments [15, 16]. The well-established CRISP-DM procedure model for data-analytics projects by Shearer [8] was developed based on the KDD-Model [17]. It includes the iterative steps business understanding, data understanding, data preparation, modeling, evaluation and deployment [8]. Nevertheless, neither KDD nor CRISP-DM are backed by well-defined methods and tools that facilitate a practical domain-specific application in manufacturing. In particular, there is no description of how the results of the data-analytics model can be used to derive and evaluate technical optimization measures.

The process model according to Marbín et al. [20] combines software engineering methods with CRISP-DM to extend it. In addition to the actual data-analytics process, aspects of project management and organizational processes are taken into account (e.g. training of employees and maintenance of the data-analytics model). However, in many cases in practice there is no perfect database available in the first place. So, Reinhart [18] motivates the necessity of an additional data acquisition step, but does not explain how it could be implemented. Bahrepour [19] adds a step called “data validation” between the phases of data preprocessing and modeling of the classical CRISP-DM approach, to avoid the input of faulty data into the data mining model. Nevertheless, the validation of the database is only based on an expert discussion and there is no description of how and which measures could be taken.

Even though the need for evaluation and adjustment of an existing database is recognized, a design of a problem-specific adequate database and also considering manufacturing expert knowledge is not explicitly formulated in existing approaches.

2.2. Approaches for Expert-Based Process Modelling

Manufacturing processes are naturally characterized by complex cause-effect relationships and a not negligible underlying degree of uncertainty. The control and monitoring of these processes are therefore often highly dependent on extensive expert knowledge. Besides the process data, which is gathered by sensors, the human experience is of equal importance in order to improve manufacturing processes [9]. Hence, it is essential to take this knowledge into account, when applying data-analytics in manufacturing systems [18].

Tools like Ishikawa analysis, FMEA or DoE are supportive and powerful techniques when expert knowledge and influences on the manufacturing process need to be recorded systematically. In order to map processes and thus prepare and arrange existing process data, various classical approaches exist. For example, Value-Stream Mapping (VSM) focuses on the entire value chain from raw material to the customer and captures the material and information flows [21]. Enhancements of VSM include the following: mapping with IT-landscape [22], informational wastes regarding data processing [23] and labeling of self-regulating processes [24]. The details of the actual manufacturing processes, however, remain a black box. Marbín et al. [9] define a modelling language to formalize manufacturing process chains as a pre-phase of CRISP-DM, however existing control loops and the information on sensors are not considered. Effect plans can be used to represent control loops within and between the individual machines and formalize knowledge regarding the manufacturing processes [25].

All in all there is no holistic process representation of value streams, control loops and sensors in manufacturing.

2.3. KPI Models for the Evaluation of Optimization Measures

KPIs are used to measure the performance of a manufacturing system and to evaluate improvement measures [26]. KPI systems were initially used just from a financial perspective (e.g. the well-known Du-Pont System). Later, KPIs were applied to include also quality aspects and other non-financial systems [27]. Kang et al. [28] proposed a hierarchical structure for KPIs and enables the identification of mutual relationships. For this purpose they separated basic and aggregated KPIs and analyze pairwise relationships among these KPIs. Today many other KPI systems can be found, generally focusing on either a special area of manufacturing or providing a general framework without domain-specific KPIs.
In order to obtain a comprehensive understanding of a discrete manufacturing system, a generic KPI model was developed by Stricker et al. [29]. This KPI model consists of 142 KPIs and is hierarchically structured. All KPIs are interconnected by their mathematical interdependencies. Base KPIs make up the bottom of the hierarchy (level 0). They are fundamental KPIs that can directly be measured on the shop floor (e.g. process times, number of defect parts). Aggregated KPIs on higher levels are derived from these base KPIs [29]. Machine-specific process parameters (e.g. torque, speed) are not part of the network.

An analytical extension of this KPI model by process parameters is complex and time-consuming, since all technical interdependencies must be explicitly, mathematically formulated. Alternatively, machine learning methods can be used to “learn” the connections between process parameters and KPIs based on historical data. Jinsong et al. [30] and Shin et al. [31] apply neural networks to predict the development of single KPIs based on process parameters. Ding et al. [32] use the partial least squares method in this context to identify correlations between process parameters and key figures.

In summary, existing methods are often used to predict individual KPIs, e.g. in quality control. However, they are not used in a holistic network of KPIs, which helps to identify and evaluate measures to improve processes.

3. Methodology

By combining process parameters with a KPI model for the evaluation of optimization measures, a specific process model for data-analytics projects in manufacturing is introduced in this section. It addresses the shortcomings of existing approaches outlined in Section 2. The method is intended for the application in existing manufacturing systems. It consists of five iterative steps, wherein the following key questions are addressed: Which overall KPIs, from the manufacturing system point of view, are the target values of the data analytics project? How to evaluate and if needed extend an existing database? What additional data requirements can be met and implemented when involving expert knowledge? Which process improvement measures can be derived and evaluated based on the data analytics outcome? The approach is shown in Fig. 1 and explained in the following subsections.

3.1. Process Landscape and Target Values

In the first phase Process Landscape and Target Values a comprehensive current state analysis of the considered manufacturing system is performed. The aim is to gain a comprehensive understanding of the process. First, the system boundary needs to be defined. Based on an advanced VSM technique, value stream, data landscape, process control loops and sensor technologies are mapped in an integrated process map. This integrated process map consists of four linked layers, which are arranged hierarchically.

The first and superordinate layer has the highest aggregation level. It is similar to the VSM and compromises the manufacturing processes and the material flows. The single process steps and machines are explicitly stated.

In the second layer the process steps and machines with respect to the control loops are specified. All existing quality-related control loops, including the associated machine parameters, are mapped. The identification of the quality-critical control loops is carried out in a FMEA. The control loops can refer to just one single process step or across several steps. Each control loop has defined incoming data streams, which originate from sensors.

The so-called landscape of sensors is described in the third layer. This layer includes the location of sensors. For each sensor characteristic attributes are recorded such as: measurement value, frequency of measurement and output data format. A sensor does not necessarily have to be a technical unit, but can also be represented by a person that manually records and transmits measured values (e.g. control panel).

The fourth and last layer compromises the data landscape. All databases in which the sensor data is stored and processed are listed in this layer. Furthermore, the data connections are illustrated in order to be able to identify system breaks.

Based on the integrated process map, critical process steps can easily be identified (e.g. high reject rates or long processing times). These are in the focus of the subsequent optimization phase. In addition, the target KPI (e.g. OEE of the entire manufacturing system) is defined based on the KPI model (see Section 2.3).

In this first phase of the approach it is crucial that all participants, e.g. manufacturing process experts, production planner and data-analytics experts, are involved. In this way a comprehensive understanding of the considered manufacturing process is likely to be achieved and the objectives are selected.

3.2. Data Requirements and Collection

The aim of the second phase Data Requirements and Collection is the examination and selective extension of the existing database. First, the Ishikawa method is used to determine the most relevant influencing factors on the manufacturing process. Based on the Ishikawa analysis, the integrated process map is used to determine which influencing factors are already captured by sensors and which are also included in existing control loops. For this purpose, an evaluation of the proportion of recorded influencing factors for the respective classes of the Ishikawa method is performed.
In a next step the quality and information content of the already collected data is examined. In addition to an explorative analysis and visualization of the data, the variance in the data is investigated. If the variance is low with regard to the optimization target defined in the first phase, which means the parameter do not vary or there is no statistical relationship, an increase of the variance can be achieved by a Design of Experiments (DoE). The result of a DoE is a well-defined dataset with high variance and eventually also additional process knowledge confirming the experiences of the experts. If this is not possible, additional data can alternatively be generated from existing process knowledge or with the help of a simulation model.

The identified influencing factors, which are not recorded, serve as a starting point for the integration of additional sensors. The relevance of the influencing factors is evaluated by a utility analysis and the feasibility of the recording is evaluated based on a feasibility analysis. A distinction is made between target and ideal data requirements. Target data requirements are those that must be met. Ideal data requirements represent an option for the future.

In the last step, suitable sensor systems are selected and implemented. The main questions in this phase are: Which measurands have to be recorded? Which installation space and interfaces are available? Which specifications should the measurement signal with respect to the data interpretation? The sensor capability is investigated by a prototypical implementation. Sensor capability means that the sensor directly measures the addressed measurand, i.e. the influencing factor, and that the measurement uncertainty is sufficient according to the requirements. If so, the sensors are implemented and the process map is adapted accordingly.

### 3.3. Data Analysis

In the third phase Data Analysis the resulting database is transformed, cleaned and a plausibility check of the measured values is carried out, e.g. physically impossible values or incorrect data formats are identified. Identified outliers are replaced and missing data points are integrated via estimation procedures. The available series of process data is then linked to the product-related quality data and the targeted KPIs. In doing so, a time series of process data values is assigned to each workpiece. For further investigations, a suitable aggregation, e.g. standard deviation, median, mean etc., of these time series is identified, which represents the influence on the quality data and targeted KPIs. A correlation matrix is used to identify those parameters that have the greatest influence on the KPIs. In this way, the characteristic features of the measured process parameters are identified. The influencing factors determined in the second phase can thus be confirmed in their importance and new correlations can be discovered. It is also possible that no significant correlation can be found. Then it is necessary to jump back to the phase Data Requirements and Collection and adapt according to additional influencing factors of the Ishikawa analysis.

As a result, a set of process parameters is identified that provides a high potential for deriving improvement measures.

### 3.4. Process-Specific Extension of the KPI Model

In the fourth phase of the approach, the KPI model (see Section 2.3) is extended by the set of process parameters defined before. As shown in Fig. 2, the process parameters form an additional hierarchy level (level -1) below the base KPIs (level 0). While the KPIs of the KPI model are all analytically linked with their mathematical relations, the connections between process parameters and the base KPIs are computed by a learning method (e.g. neural network). For this purpose, the process parameters data is linked to the characteristics of the base KPIs via their time stamp over a period of time. The connections are first learned based on a training data set and then validated on a test data set. This extended KPI model forms the basis for the selection of measures in the next phase.

### 3.5. Selection and Evaluation of Improvement Measures

For the selection of improvement measures, a distinction is made between process-specific measures and measures that are generally valid for interlinked manufacturing lines. The latter are listed in a catalogue which is derived from general knowledge on manufacturing performance improvement measures such as integration of a buffer or employee training. The catalogue is categorized in four different categories: personal-, organizational-, material- and equipment-related measures. Each measure is described and the prerequisites as well as assumptions for implementation are recorded. While the generic catalogue can be used for all interlinked manufacturing lines, it is necessary to derive process-specific measures with a direct effect on the identified process parameters in expert workshops. Process-specific measures are, for example, the design or adaptation of control loops or a new set of fixed manufacturing parameters. The selection is based on the integrated process map (see Section 3.1). Weak points are systematically analyzed in the map, by a similar approach as the seven types of waste in lean management, and suitable optimization measures are derived. Useful generic measures are also preselected in this way. The resulting list of measures is supplemented by a cost estimation and forms the basis for the following evaluation step.

For each measure, the effect on the target KPIs is examined using the expanded KPI model (see Fig. 2). This is done qualitatively for the generic measures. This means that a statement is made if the measure contributes to improving the target value or not. For example, an additional training of employees (M3) leads to an improvement of the OEE. The effect of a process-specific measure (M4) can be quantitatively described by changing the process parameters value. Using the learning method and the mathematical connections in the KPI model, the effect on the target KPIs can be calculated (see Fig. 2). As a result, the measures are ranked according to their cost-
benefit ratio. The promising measures are recommended for implementation. The new data points serve as further train data for the learning method. The ongoing operation of the manufacturing system continuously adjusts the KPI model.

4. Application to Industrial Use Case

The introduced approach is applied exemplarily to a manufacturing line. On this line, a thermal, ablative process is applied to metallic parts. The line consists of five stations, is highly automated and rigidly interlinked.

4.1. Process Landscape and Target Values

The primary goal of the process improvement is the OEE improvement by reducing the number of rejects. The integrated process map includes all five stations. At the first station the key product features are defined. At the second and third station fine machining is performed and at the fourth station, multiple product quality tests are carried out. At the fifth and last station, a data matrix code is applied. A total of five quality-critical control loops can be identified, including the sensors from the second layer of the integrated process map. Most of the sensor data is stored in the internal machine control system. Access is possible via a process data acquisition system. Since the reject and rework rate at the first station is the highest, the focus of further analysis is on this process step. The high reject rate implies a large potential for improvement by optimizing the process parameters.

4.2. Data Requirements and Collection

A total number of 42 influencing factors can be identified within the Ishikawa analysis on the first station based on expert knowledge. However, just 13 of these influencing factors are currently recorded and only one quality critical control loop making use of them is implemented. The utility and feasibility analysis result in a desirable enhancement of the database by eight parameters (+19%). The ideal database represents 19 additional parameters. It is recommended to record these influencing factors by means of additional sensors. In the following, the data requirement liquid pressure is focused as an example. The aim is to achieve a constant pressure within the process to ensure a higher robustness of the considered thermal, ablative manufacturing process. The liquid pressure can be recorded with a pressure sensor, which can easily be integrated in the manufacturing line.

4.3. Data Analysis

For analysis, an example data record is generated from the new and existing sensors. Since the new sensors are not yet integrated into the manufacturing system, two data sets are created over a period of four weeks. The time stamp is used to merge both data sets. To identify a suitable representation of the time series, a correlation analysis is performed. The correlation matrix for the liquid pressure is shown in Fig. 3. Q1 represents the product quality data. The mean value has the highest correlation with the quality date Q1 (-0.35). The mean value is therefore a suitable indicator to evaluate the level of the pressure and thus forms the additional input variable for the KPI model. The correlation analysis furthermore confirms the relevance of all additional sensors, which were identified based on expert knowledge.

4.4. Extension of the KPI Model and Derivation of Improvement Measures

The extension of the KPI model and derivation of improvement measures results first in the extension of the KPI model. The additional process parameter liquid pressure is determined analytically as a learning method is not yet required for just one single parameter. The selection of optimization measures leads to the outcome that an additional control loop, which keeps the pressure of the liquid constant at a high level, should be installed. In addition to the already integrated sensors, the required control logic is built up with a separate microcontroller. Targeted employee training is another selected measure to improve the OEE, an influencing factor related to a human operation. A positive effect of these measures on the OEE can be shown with the help of the KPI model and has been confirmed in first prototypical implementations. In order to derive further measures, a neural network will be set up in a next step to identify complex and non-linear dependencies.

5. Summary

The objective of this paper is to present a methodology for data-driven process optimization in manufacturing. In particular, the question how to use and extend an existing database.
The proposed methodology includes a five step procedure. First a comprehensive current state analysis is performed. Based on this analysis further data requirements are derived and suitable sensors are implemented. These additional data are used to identify improvement measures, which are evaluated by a process-specific KPI model. The approach is validated using the example of a manufacturing line. Measures to improve OEE were derived for a thermal, ablative process step.

Acknowledgements

This research is funded by the German Federal Ministry of Education and Research (BMBF): "Produktionsbezogene Dienstleistungssysteme auf Basis von Big-Data-Analysen (ProData)" (funding code 02K16C08). Furthermore, the authors are indebted to the Baden-Württemberg Stiftung for the financial support of this research project by the “Eliteprogramme für Postdocs”.

References