Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 104 (2021) 636–640

www.elsevier.com/locate/procedia

54th CIRP Conference on Manufacturing Systems

Data Acquisition and Preparation – Enabling Data Analytics Projects within Production

Christoph Schock[a,*], Jonas Dumler[b], Prof. Dr.-Ing. Frank Doepper[a,b]

[a]*Fraunhofer IPA Project Group Process Innovation, Universitaetsstr. 9, 95447 Bayreuth, Germany*
[b]*Chair of Manufacturing and Remanufacuring Technology, University of Bayreuth, Universitaetsstr. 9, 95447 Bayreuth, Germany*

* Corresponding author. Tel.: +49 921-78516-314; fax: +49-921-78516-105. *E-mail address:* christoph.schock@ipa.fraunhofer.de

Abstract

The increasing amount of available data in production systems is associated with great potential for process optimization. Due to lack of a data analytics methodology and low data quality within production these potentials often remain unused. Therefore, in this paper we present a model for data acquisition and data preparation including feature engineering for characteristic sensor signals of production machines. The model allows the extraction of relevant process information from the signal, which can be used for monitoring, KPI tracking, trend analysis and anomaly detection. The approach is evaluated on an industrial turning process.

Keywords: Data Analytics, CRISP-DM, Data Acquisition, Data Preparation, Feature Engineering, Process Monitoring; Condition Monitoring

1. Motivation

The fourth industrial revolution causes a multitude of challenges for manufacturing companies. To ensure business success, workflows and processes within the value chain have to be optimized. [1-3]

However, the rapidly growing volume of data and the use of information and communication technologies and cyber-physical systems are creating promising opportunities.

Data analytics, for example, offers the possibility of deriving knowledge from data and thus increasing process efficiency. [4-6] Even though, the proportion of data utilized for analysis purposes in companies is very low (see Fig. 1) [7]. In particular, collected data within production environments is often incomplete and heterogeneous, as operating parameters and process states are frequently not transparent. [8,9]

As a result, optimization potentials regarding prediction, the recognition of cause-effect relationships and the proactive control of processes in particular are not exploited and efforts made for data collection remain wasted.

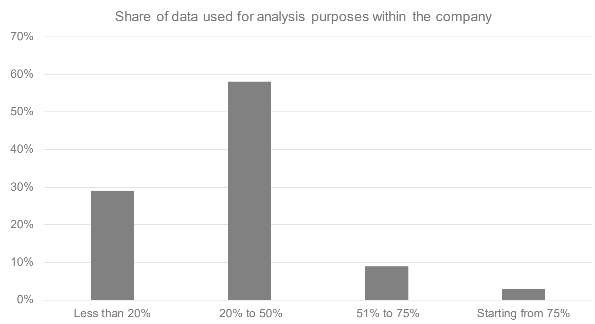

Fig. 1. Share of data used for analysis purposes within the company

To overcome these barriers, this paper presents a step-by-step model for addressing heterogeneous and incomplete data bases through data acquisition and data preparation.

10.1016/j.procir.2021.11.107

Therefore, the state of scientific knowledge in section two is followed by the description of a concept for data acquisition and processing in section 3. Both topics are described by using a schematic diagram, to be evaluated in section 4 by presenting an industrial case study.

2. State of scientific knowledge

Extracting information from data is accompanied with great potential regarding maintenance, fault detection and quality control. [10] Hence various applications of data analytics and data mining within manufacturing, engineering design and logistics were discussed since 2006. [10,11]

Nevertheless, studies show that the level of data utilization within companies is still low, since companies are confronted with poor data quality, large data volumes and a lack of methodology. [2,3,12,13]

These theses are supported by a survey among experts from manufacturing companies, conducted by the Fraunhofer Project Group Process Innovation in 2018, 2019 and 2020. It shows that companies (from small and medium sized to multinational enterprises) currently focus on data collection activities within production – and disregard the holistic analysis to gain benefits form collected data. As a result, most of the surveyed companies utilize only a small fraction of their data available and efforts for data collection often remain obsolete. This is primarily for two reasons: The lack of a methodology for data based projects and a poor data quality within production environments.

For deducing information from data in general, data mining methods like the Knowledge Discovery in Databases (KDD [14]), the Sample, Explore, Modify, Model and Assess Process (SEMMA [15]) or the Cross Industry Standard Process for Data Mining (CRISP-DM [16]) provide a basis. Yet, data mining methods are not often utilized in manufacturing. Only 17 % of users solve existing potentials by applying data mining, according to a survey conducted by Rexer Analytics. [17]

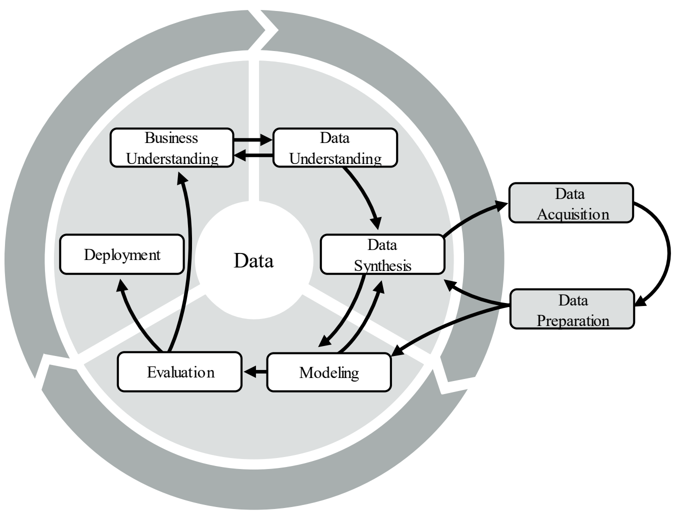

Fig. 2. Enhanced CRISP-DM based on Schock (2019) [9]

Therefore, the CRISP-DM was enhanced to suit engineering (DMME [18]) and production needs [9] (see Fig. 2). However, both enhanced methodologies do not provide further information to address poor data quality.

3. Data Acquisition and Preparation

Since the second phase of the CRISP-DM only considers existing data sources, we extend it with the sub step "Derivation of Data Needs". Thereby, we distinct between needs regarding attribute data (features) and event data (labels). Features are the descriptive attributes, i.e. data coming from the machine and process operation. The labels represent data for the assessment of the features, such as data from maintenance, product quality or tool wear. Accordingly, the label corresponds to a possible output variable of a supervised machine learning model.

3.1. Data Acquisition

In case data needs are derived and defined, the phase "Data Acquisition" starts. The phase is shown in Fig. 3.

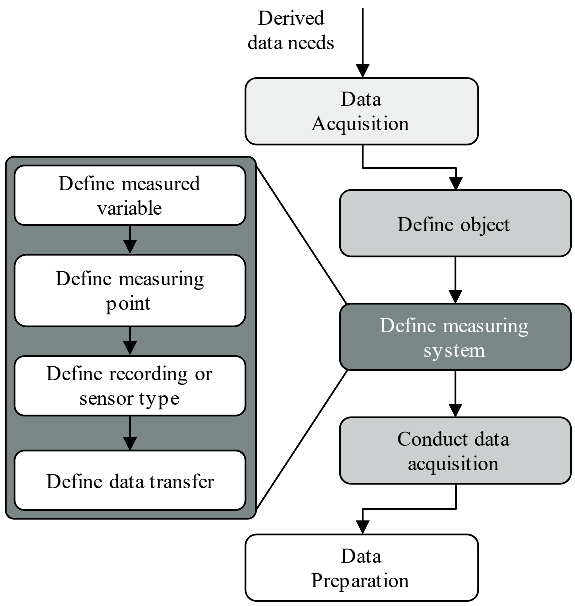

Fig. 3. Process for Data Acquisition

First, the measurement object is defined. Potential objects are processes (consisting of several steps), machines, machine components or products.

Subsequent, the measurement system is specified. First, suitable measured variables are defined. Regarding data mining projects within manufacturing, some of the most important measurands are: electric current and power, acceleration, temperature, pressure as well as airborne and structure-borne sound. The selection can be based on a benchmark of similar projects and in consideration of physical principles.

Then the measuring points are defined. Next to the defined object itself, accessibility and the effort required for installation must be taken into account.

Afterwards, the selection of the corresponding measuring equipment and sensors takes place. The selection of sensor types must include criteria such as resolution and bandwidth, environmental conditions and spatial dimensions, as well as mounting options and connection types. In terms of measurement equipment, there can be significant differences between temporary test setups and long-term installations.

Finally, the data format and the transmission must be described. All activities of the phase are documented in a protocol.

Output of the phase “Data Acquisition” is a data set, consisting of

- raw sensor signals as attribute data, matching existing event data,
- raw event data matching existing attribute data, or
- raw attribute and matching raw event data.

3.2. Data Preparation

In order to generate a final data set of high quality, “Data Acquisition” is followed by “Data Preparation”, which includes the sub step “Feature Engineering”. Furthermore, this phase covers standard activities like data exploration, cleansing, formatting, integration of different data sources, and data reduction.

In the illustrated case of acquiring new data, raw signals from sensors in particular have to be processed. Usually, these signals are time series, often defined by a high temporal resolution, high data volume and low information density [19]. In order to cope with these characteristics, the sub-steps of the phase “Data Preparation” have to be adapted accordingly. The phase is shown in Fig. 4.

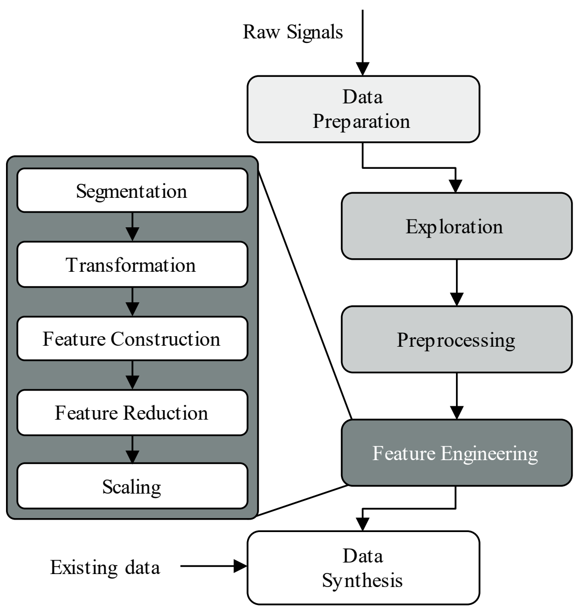

Fig. 4. Redesigned Process for Data Preparation

The first sub-step is “Data Exploration”, using methods of visualization and statistics. To initially interpret and enable further target-oriented processing, basic data properties are visualized.

This is followed by the sub-step “Data Preprocessing”, which includes standard activities of data cleansing, such as the removal of erroneous measurements, unassignable and redundant data sets as well as the correction of formatting errors. In case of processing sensor signals, filtering is explicitly added in order to reduce signal noise or remove unintentional frequency components. Thereby, the signal quality is increased.

“Feature Engineering” is used to compress information. At first, depending on the analytics goal and physical measurement objective, the signal has to be segmented. In continuous processes, a step-by-step segmentation into equal-sized time segments can be performed, whereas in batch production a semantic segmentation is desired. Thereby, a continuous time signal can be subdivided and explicitly assigned to a specific work piece or a production process step.

For this purpose, possibilities supporting the segmentation are iteratively assessed and integrated in the phase “Data Acquisition”, e.g. the recording of trigger signals.

In order to derive specific features describing properties of the oscillation behavior, transformations can be applied. The most common methods are the Fast Fourier Transform (FFT), the Short-Time Fourier Transform (STFT) and wavelet transformations.

The sub-step “Feature Construction” (or extraction) describes the extraction of new features through functional mapping. The objective is to create new attributes that contain more information than existing features or are easier to interpret with respect to the application. This can be achieved, for example, by ratio indicators, such as energy consumption per production unit. In General, both statistical key figures and features specifically designed for the application are calculated. For instance, physical correlations or empirical evaluations from process experts can serve as a basis. Redundancies within the feature set should be avoided.

In order to improve the performance of the data analysis model in a later phase, the sub-step “Feature Reduction” is applied. The most common technique is feature selection. It is used to reduce the range of features to a specific subset with high discriminatory information and variance by removing redundant and non-relevant features. [19]

Finally, all features are scaled, since the range of features can vary widely. Normalization is the most common method.

In case of existing data, the new constructed, rich data set is merged with existing data from other sources. The phase “Data Synthesis” includes further steps of formatting data type structures.

4. Evaluation

For the purpose of incremental adjustment and evaluation, the presented model was applied step-by-step in several industrial case studies regarding predictive analytics within production.

The evaluation case study objective is a multi-stage process, consisting of a turning machine and a subsequent honing machine (see Fig. 5). With this production process, barrel-shaped work pieces are produced. Variations in tool wear during turning result in different surface roughness and finishes (Rz). This in turns has an influence on the machine parameters required for the honing process. To ensure highest quality standards, turning tools are changed after a defined life time with a priced in safety factor, even though tool cost as well as setup time and costs are high. In order to compensate for variations in roughness after turning, a safety factor is also applied during honing, which results in increased machining times.

Therefore, the goal of this evaluation case study was to detect tool wear automatically, increase the tool life cycle and thereby the availability of the turning machine, while maintaining the quality level of work pieces produced.

The existing data consisted of both attribute and event data. Attribute data from machine operation was available batchwise with a time-delay. The structure of the existing data batches was not comprehensible. Regarding event data, specific surface characteristics of every produced work piece as well as the width of the wear mark on the tool surface were available.

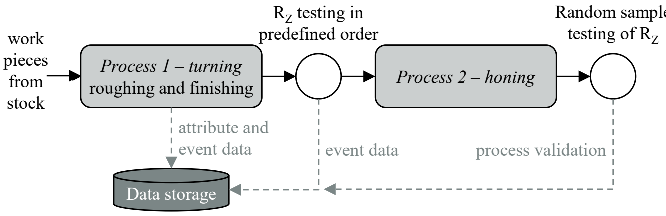

Fig. 5. Flowchart of evaluated production process and data acquisition

4.1. Data Acquisition

Since existing data was only partially reliable, data needs were defined. In order to detect tool wear automatically, attribute data for the model training had to be acquired. Therefore, the object – the first in line turning machine Hembrug Mikroturn® 100 – was defined and cross linked with a sensor-based infrastructure.

For the purpose of attribute data acquisition, a NI cDAQ-9188 Ethernet-Chassis with a simultaneous analog input module NI 9215 (4 channels, 50 kS/s, 16 bit, ±10 V) was connected to a laptop with a LabVIEW program. The electric current of the spindle drive was measured with Fluke i30s AC/DC Current clamps based on Hall Effect technology with BNC connector. They provide a measurement range up to 30 A at 100 mV/A. The measuring sequences were triggered by a digital signal from the machine control.

During data acquisition, an overall number of 310 work pieces were produced. Flank wear of the tool was measured after the predefined tool life.

4.2. Data Preparation

Visual exploration was used to verify the plausibility of the signal amplitudes and of the measurement setup. "Data Preprocessing" included the elimination of faulty and incomplete measurements and renaming of files to ensure correct mapping between signals and wear data.

Since the turning process consisted of several steps, segmentation was implemented using peak detection in the electric current signal of the internal tool changer. As a result, we were able to separate the single machining process for every produced work piece with the same signal length due to the constant processing speed over all measurements.

In order to obtain features not only from time domain, but also from frequency and time-frequency domain, Fourier and wavelet transformations were conducted. This allowed a targeted analysis of specific frequency ranges.

During feature construction statistical metrics were calculated from both the time domain signals and the transformed spectral data. In order to compress the data efficiently and to obtain more detailed insights, features were calculated over the entire machining process on the one hand, and over smaller time segments on the other hand. This procedure resulted in a set of 72 features in total for each machining process and for each of the smaller time segments.

Feature reduction included a selection based on empirical evidence from similar applications and a selection based on a correlation analysis.

During data synthesis we mapped the acquired event data to the calculated feature set. The final high quality data set consisted of 12 features including RMS, peak-to-peak amplitude and mean of band power spectrum for specific frequency ranges, as well as the attributes flank wear, tool life and R_z as corresponding event data.

4.3. Case study results

Fig. 6 shows that our model consisting of data acquisition and preparation including feature engineering was effective, since the correlation between attributes and tool life can be successfully demonstrated and modeled.

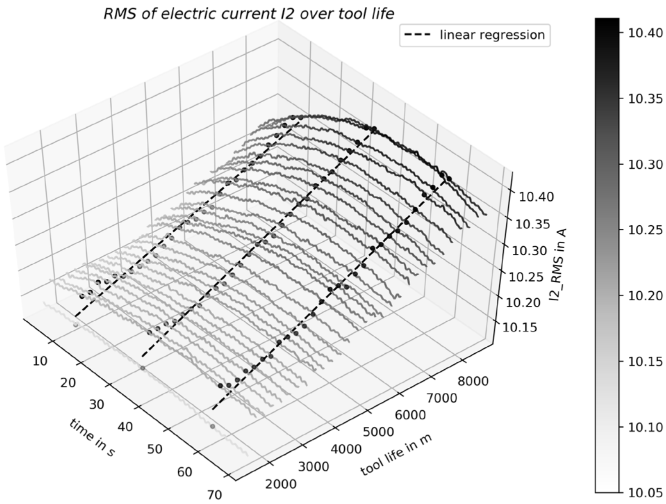

Fig. 6. Linear regression of electric current over tool life

Therefore, Fig. 6 illustrates the time course of the moving RMS (window size 1 s, overlap 0.9 s) of the spindle drive current during the finishing process plotted against the cumulated tool life after the respective turning process. The turning process takes 68 seconds. To illustrate the increasing current consumption, the RMS values are highlighted at three different positions of the process. The increasing power consumption can be modeled with a linear regression. The average R^2-value for the shown measurement series is 0.95.

In this case study we were able to demonstrate that data acquisition and data preparation including feature engineering enables the definition of a RMS-based threshold value for tool wear and the automated initiation of tool change.

5. Conclusion and further research

The model represents a guideline, which consists of different sub-steps in subsequent order, but with an iterative approach. By applying the model, we were able to enrich and improve the quality of heterogeneous and incomplete data bases and thereby enable data analytics projects in several production environments, from small and medium sized to multinational enterprises. During the case studies, cooperation among specialized personnel was one of the success factors. The roles involved were subject matter experts regarding production technology, data science and analytics. To evaluate the beneficial use of data analytics, business experts were involved as well.

In the future, we will work on a holistic methodology that empowers project teams in an end-to-end approach through data analysis projects. The research focus will be on project initiation, in which optimization potentials are identified and prioritized, the selection and evaluation of data bases, and the target- and hypothesis-based selection of analysis methods. Furthermore, a procedure for the derivation of hardware and software requirements for the implementation of a permanent application must be specified.

Acknowledgements

The research work presented is partly funded by the Bavarian Research Foundation. The evaluation case study is anonymized in part with regard to the descriptions and values in order not to allow any conclusions to be drawn about the industry partner. The authors are responsible for the content of this publication.

References

[1] Schröder C. Herausforderungen von Industrie 4.0 für den Mittelstand. In: Gute Gesellschaft - soziale Demokratie #2017plus. Bonn: Friedrich-Ebert-Stiftung; 2016. p. 9-10.

[2] BITKOM. Potenziale und Einsatz von Big Data. Ergebnisse einer repräsentativen Befragung von Unternehmen in Deutschland. In: BITKOM; 2014.

[3] BITKOM. Mit Daten Werte schaffen. Report 2015. In: BITKOM and KPMG; 2015.

[4] BITKOM. Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele. Leitfaden. In: BITKOM-Arbeitskreis Big Data; 2015.

[5] Vazan P, Janikova D, Tanuska P, Kebisek M, Cervenanska Z. Using data mining methods for manufacturing process control. In: IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd; 2017. p. 6178-6183.

[6] Schebek L, Sauer A, Steinhilper R, Schock C, et al. Ressourceneffizienz durch Industrie 4.0. Potenziale für KMU des verarbeitenden Gewerbes. In: VDI Zentrum Ressourceneffizienz; 2017.

[7] Gronau N, Fohrholz C, Weber N. Wettbewerbsfaktor Analytics - Reifegrad ermitteln, Wirtschaftlichkeitspotenziale entdecken. Ergebnisse einer explorativen Studie zur Nutzung von Business Analytics in Unternehmen der DACH-Region. Universität Potsdam; 2013.

[8] Steinhilper R, Kuebler F, et al. Minimierung nicht wertschöpfender Energieaufwände durch energetische Anlagenoptimierung. Fraunhofer Publica; 2015.

[9] Schock C. CRISP-DM - Ein Ansatz zur Systematisierung von Digitalisierungsprojekten in Produktionsumfeldern. Fraunhofer Publica; 2019.

[10] Choudhary AK, Harding JA, Tiwari MK. Data mining in manufacturing: a review based on the kind of knowledge. In: Journal of Intelligent Manufacturing; 2008. p. 6-23.

[11] Feng JCX, Kusiak A. Data mining applications in engineering design, manufacturing and logistics. In: International Journal of Production Research 44; 2006. p. 2689-2694.

[12] Vossen G, Lechtenbörger J, Fekete D. Big Data in kleinen und mittleren Unternehmen – eine empirische Bestandsaufnahme. Institut für WirtschaftsinformatikWestfälische Wilhelms-Universität Münster; 2015.

[13] Schalla P, Niemann F, Flug M. Predictive Analytics in der Fertigungsindustrie. Nieschenthema oder Mainstream. Pierre Audoin Consultants; 2014.

[14] Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases; In: AI Magazine Volume 17 Number 3; 1996. p. 37-54.

[15] SAS Institute Inc. Sampling, Exploring, Modifying, Modeling, and Assessing (SEMMA); 2017.

[16] Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. CRISP-DM 1.0, Step-by.step data mining guide. In: SPSS; 2000.

[17] Rexer K, Gearan P, Allen H. Data Science Survey. In: Rexter Analytics; 2015. p. 1-5.

[18] Huber S, Wiemer H, Schneider D, Ihlenfeldt S. DMME: Data mining methodology for engineering applications - a holistic extension to the CRISP-DM model. In: Procedia CIRP Volume 79; 2019. p. 403-408.

[19] Aremu OO, Palau AS, Parlikad AK, Hyland-Wood D, McAree PR. Structuring Data for Intelligent Predictive Maintenance in Asset Management. In: IFAC-PapersOnLine 51 (11); 2018. p. 514-519.