Data Analytics for Industrial Process Improvement A Vision Paper

(Workshop Paper)

Stefan Thalmann*, Juergen Mangler†, Tobias Schreck*, Christian Huemer†, Marc Streit*, Florian Pauker†, Georg Weichhart*, Stefan Schulte†, Christian Kittl*, Christoph Pollak†, Matej Vukovic*, Gerti Kappel†, Milot Gashi*, Stefanie Rinderle-Ma†, Josef Suschnigg*, Nikolina Jekic*, Stefanie Lindstaedt*
*Pro2Future
Graz/Linz/Steyr, Austria
Email: firstname.lastname@pro2future.at
†(Austrian) Center For Digital Production
Vienna, Austria
Email: firstname.lastname@acdp.at

***Abstract*—Nowadays, manufacturers are increasingly able to collect and analyze data from sensors on manufacturing equipment, and also from other types of machinery, such as smart meters, pipelines, delivery trucks, etc. Nevertheless, many manufacturers are not yet ready to use analytics beyond a tool to track historical performance. However, just knowing what happened and why it happened does not use the full potential of the data and is not sufficient anymore. Manufacturers need to know what happens next and what actions to take in order to get optimal results. It is a challenge to develop advanced analytics techniques including machine learning and predictive algorithms to transform data into relevant information for gaining useful insights to take appropriate action. In the proposed research we target new analytic methods and tools that make insights not only more understandable but also actionable by decision makers. The latter requires that the results of data analytics have an immediate effect on the processes of the manufacturer. Thereby, data analytics has the potential to improve industrial processes by reducing maintenance costs, avoiding equipment failures and improving business operations. Accordingly, the overall objective of this project is to develop a set of tools — including algorithms, analytic machinery, planning approaches and visualizations — for industrial process improvements based on data analytics.**

I. Introduction

Advanced data analytics in manufacturing has the goal to gain new insights about the production process and to provide better decision support. Therefore, the new insights should be added to the already existing knowledge base. In production, this knowledge base is mainly automated by using low level rules in the process automation system. However, not all insights from data analytics can be broken down to this level of detail and hence a more general approach is needed. For this purpose, approaches from business process management (BPM) are perfectly suitable. Hence, the goal of our paper is to show how advanced data analytics can be used to derive useful insights of production processes and how these insights can be linked to the business processes of the company. The general engineering approach on data is based on elaborated knowledge-based models. These models support efficient data understanding. With new technical tools and techniques, it is possible to analyze data with no or little knowledge of the underlying domain specifics. By combining those approaches a better knowledge of the data and underlying processes can be achieved.

In engineering approaches the manufacturing execution systems (MES) is an intermediate layer between the Enterprise Resource Planning (ERP) and the Supervisory Control and Data Acquisition (SCADA) level. The SCADA system fetches real time data from machines and controls theme via the Programmable Logic Controller (PLC/SPS) with strict rule based functions. The rule set for the SCADA system is provided via the MES. These systems capture the data produced in the underlying layers and set up individual product records. Therefore, these systems are also well integrated in the product life cycle. Aggregated reporting from MES is done towards the ERP and business information is transferred back to MES — as needed for scheduling and resource planning. Along all vertical levels in production data is created, transferred and stored. The heterogeneous data is linked to processes on the higher levels. These processes are engineered knowledge driven. Data-driven approaches perceive knowledge as something that is built on the interpretation of information. Information is created by analysis of data. Starting from simple descriptive approaches to more advanced algorithmic analytics, data is transferred into information and by cognitive analytics knowledge is created. In advanced models wisdom is built out of proven knowledge.

Our goal is to improve industrial processes by means of data analytics which will be achieved by a combination of data-driven knowledge generation and corresponding redesign and reconfiguration of business processes models. To support the decision makers (responsible for changing the manufacturing processes), it is essential to visualize process designs, running process instances, recognized defects or inefficiencies, and possible changes. Furthermore, it is critical that the decision maker is aware of potential effects of process changes. Thus, this consortium wants to deliver a simulation tool to evaluate

2378-1971/18/$31.00 ©2018 IEEE
DOI 10.1109/CBI.2018.10051

IEEE computer society

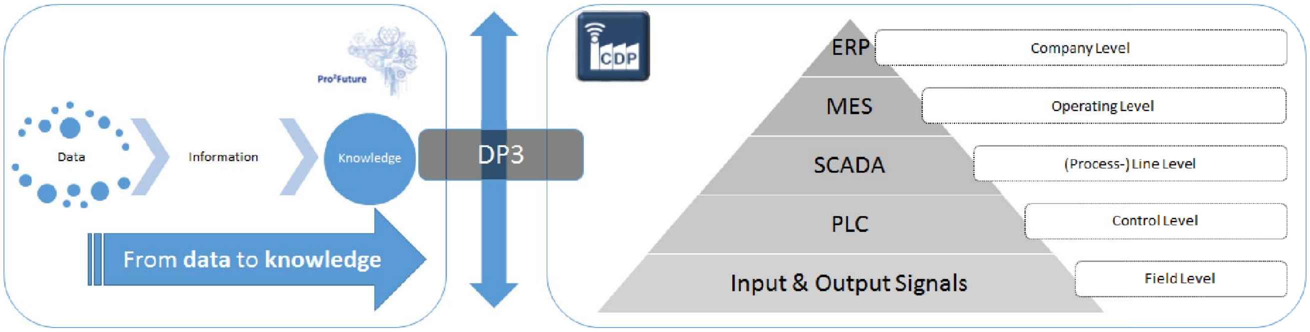


Fig. 1: From data to information to knowledge on all levels of the automation pyramid

different alternative shop floor process changes. If a process is changed, the corresponding effects may be local to a specific level of the automation pyramid or it may have cascading effects on other levels of the pyramid that will be supported by the developed tools. In addition, we will provide a secure data communication protocol which will enable the communication of data to be analyzed, e.g., between production sites. Finally, we will evaluate the framework and prototypes by a common use case which is independently demonstrated at two different smart factories (in Vienna and in Graz).

In summary, we want to achieve our main goal by six components:

1) Process redesign and reconfiguration by means of data analytics
2) Visualization of processes and process changes
3) Planning/scheduling and simulation of process alternatives
4) Support of cascading process changes
5) Secure data connections
6) Use case based evaluation in smart factories

II. State of the Art

In engineering approaches the Manufacturing Execution Systems (MES) is an intermediate layer between the Enterprise Resource Planning (ERP) and the Supervisory Control and Data Acquisition (SCADA) level [1-4]. The SCADA system fetches real time data from machines and controls them via the Programmable Logic Controller (PLC/SPS) with strict rule based functions. The rule set for the SCADA system is provided via the MES. These systems capture the data produced in the underlying layers and set up individual product records. Therefore, these systems are also well integrated in the product life cycle. Aggregated reporting from MES is done towards the ERP and business information is transferred back to MES as needed for process scheduling and resource planning [5-8].

Whereas "traditional" Business Process Management (BPM) orchestration assumes stable and well-defined processes, aspects such as dynamics, efficiency, the increasing complexity and the large amount of real-time data, specific to the manufacturing domain, calls for new solutions received little recognition. Especially, to the best of our knowledge, the integration of data analytics into business process design and enactment has not been systematically discussed and suitable solutions are still missing.

Production data is created, transferred and stored along all vertical levels (see Figure 1). Heterogeneous data is linked to processes on higher levels. These processes are mainly driven by the knowledge of engineers. Data-driven approaches perceive knowledge as something that is built on the interpretation of information which is created by visual analysis of data [9-12]. Starting from simple descriptive approaches to more advanced algorithmic analytics, data is transferred into information and cognitive analytics knowledge is created. In advanced models wisdom is built on proven knowledge [13-16].

The field of interactive data visualization is concerned with methods to find appropriate visual representations and interactions which allow users to understand complex data, search for patterns, verify assumptions, and to monitor streaming data, among others. A general introduction can be found in [9], [10]. In previous work, we have developed visualizations for different types of data and analysis tasks. For example, we proposed a method for pattern detection in time-dependent high-dimensional based on dimensionality reduction [11]. Such methods could be helpful to detect regularities and anomalies in streaming production sensor data. Another type of visualization techniques support network-oriented data, which is applicable in our context e.g., for networks of production nodes [12]. Also, our previous work on the development of an interactive visualization technique for ranking multi-attribute data [19] is relevant in the context of this project, as it is applicable to a wide range of prioritization tasks.

This joint project aims towards creating new knowledge for improving and adapting already existing processes by use of data analytics and interactive visualization. The simultaneous and systematic examination of data and processes will support production at all levels of the automation pyramid as shown in Figure 1.

III. Approaches and Methods

We investigate on applying data analytics concepts to a process-oriented view. Accordingly, we envision flexible production processes that benefit from data analytics. Thus, we will gather and analyze (including visualizing) production data, interpret it in the context of production processes in order to support the decisions on process design and execution as well as for delivering the corresponding process enactment.

We will follow a process-oriented view in order to facilitate production processes on the one hand. It is envisioned that the necessary process steps are (semi-automatically) derived from the product design, although manual process design is also supported to give the process owner a complete set of design choices. On the other hand, we also aim to (1) monitor horizontal and vertical (production) activities and to make the (production) progress visible, (2) develop predictive models based on the collected data, and (3) analyze data across automation pyramid levels in order to generate data-based models and knowledge.

For this purpose, we plan to develop conceptual approaches and methodologies for the desired functionalities. These conceptual approaches and methodologies will be supported by fully functional research prototypes and demonstrated in our demonstration scenario. These prototypes will be used in case studies that serve both as a proof-of-concept evaluation and as demonstrators that facilitate the technology transfer.

In order to reach the goals and to demonstrate our entire approach, we have developed an application scenario in the Industry 4.0 Pilot Factory of the Vienna University of Technology.

IV. Data Collection Scenario Based On Processes

As shown in Fig. 2, and discussed in previous sections, for our data analysis scenario we rely on BPMN modelled processes that span the automation pyramid. For example the business logic at the top level (ERP) is represented by a process that calls a sub-process which models plant level logic (MES), which in turn calls a sub-process that controls and monitors machines and collects data from sensors.

For our initial example scenario we decided, to concentrate on the MES level and below (highlighted by the red square in Fig. 2), to keep the complexity low enough for easy understanding, while still being close to shop-floor reality.

Based on the decisions outline above, we can thus list the types of data that will accrue:

Control flow data: this may include how how raw materials and tools are transported to machines and between machines, how machines are set up, and which parameters are used when machining is triggered.

Sensor data: this includes sensor readings that happened during the steps described above, especially during machining.

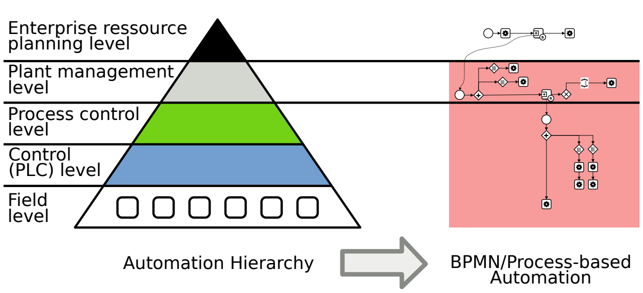


Fig. 2: Automation Pyramid vs. Processes

During a production run in the Industry 4.0 Pilot Factory of the Technical University of Vienna[1], the generic process models for data collection depicted in Fig. 4 have been enacted. In total the amount of 2 GiB of data for the creation of a ballpoint pen in an EMCO MaxxTurn 45 milling machine has been collected.

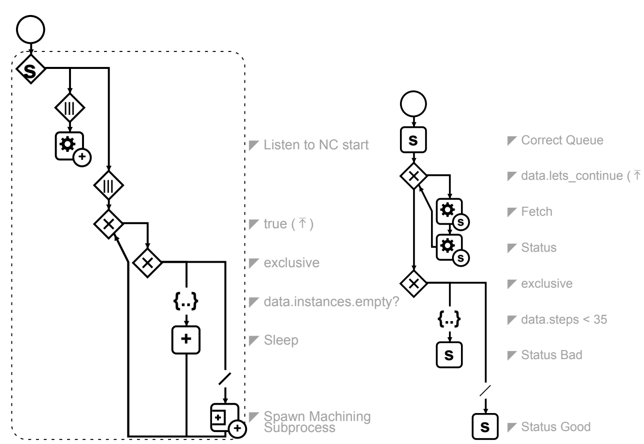
The process depicted in Fig. 4a is permanently listening to the start of a certain NC program (left-hand branch) from one or many machines. Whenever such a program is started, the task ”Listen to NC start” modifies a data structure in the process, which in turn allows the task ”Sleep” to continue. Eventually a sub-process will be spawned and the task ”Sleep” becomes active again until the next start event.

Fig. 4b shows a generic process for monitoring a machine during production. For the scenario used in this paper a total of 34 parameters have been monitored, ranging from

- Information about the tools used in the milling process.
- Information about the NC program used in milling process.
- Information about the progress in the NC program, especially which block/line was used.
- Information about the spindle movement and speed.
- Information about axis movement and speed.
- Information about the amount of electricity used.

The values during machining (SCADA level and below) have been collected from the OPC-UA interface of the machine at a rate of about 20Hz. An example data set in the IEEE 1849-2016 XES format [2] can be found under https://bit.ly/2smbQsx.

[1] http://pilotfabrik.tuwien.ac.at/
[2] http://www.xes-standard.org



(a) MES Level (b) SCADA Level and Below

Fig. 3: Generic Data Collection BPMN Models

V. EXPECTED RESULTS

We will deliver and evaluate a reference architecture, tools and demonstrator for production process analytics. Therefore, we will mainly contribute to six areas which are described in the following subsections.

1) Process Engine: The process engine is responsible for storing the business process models and for integrating the new insights from the analytics component. Thereby, the following tasks are tackled:

- *Data Analytics Aware Processes* This component allows the integration of data analytics as regular steps of a production process that provide relevant context information for production steps.
- *Data-Driven Process Design Improvement* This component will analyze production processes happened in the past in order to improve the production process design.
- *Data-Driven Process Reconfiguration* This component envisions flexible production processes that may be changed during runtime if problems are encountered. In other words, live data analysis of production data streams is used to provide self-healing capabilities for production processes.

2) Secure Data Connection Framework: This component supports cascading effects of changed process designs or process instances at different levels (of the automation pyramid). It consists of a technological infrastructure which allows for a highly secure transmission of data in a distributed production environment, and respective data integration tools for efficiently handling design changes over multiple process levels. The technical solution for guaranteeing a high security level when connecting remote sites over the Internet applies hardware-based encryption and the MQTT protocol as messaging standard. The latter is a lightweight publish/subscribe messaging protocol which works well also with low-bandwidth, high-latency or unreliable networks. The secure data connection framework supports the data extraction and transfer on the filed level.

3) Data Analytics Component: This component allows analyzing the data from the production systems in depth applying machine learning algorithms and by supporting the users in analyzing and interpreting the data. The goal is to support the control level (PLC) and to identify critical events. In this regard predictive maintenance is in the primary focus to ensure a high quality production of goods. The predictive models from the data analytics component should be used to improve the control level and to add identified rules to the process models. Core of the component is a library of suitable algorithms including lessons learned for application. The data analytic component is closely related with the process engine to enable analytics aware processes and to recommend process re-configurations.

4) Tools for Visual Data Analytics: This component allows the interactive visualization of processes based on process definitions and also, historic and (near) real-time measurement data obtained from nodes of the processes (i.e., machines in a production line) to support the process control level. To this end, a set of tools to appropriately visualize the data at hand will be realized. The set of tools will support a process expert to monitor processes in (near) real-time and hence support the decision making process on the process control level, based on appropriately selected and tailored methods from interactive data visualization, e.g., [19] Insights gained by the experts should be directly influence the business process design. Hence, the visual analytics component is closely related to the process engine and allows the visual exploration of processes and implementation of process re-configurations.

5) Planning/Scheduling and Simulation of Process Alternatives: This component will support decision making by elaborating on process alternatives based on planing and scheduling algorithms and simulation of production systems. This components takes not only a process point of view, but additionally takes the production system into account. The effects of decisions by human and artificial agents on various process levels will be shown and can be evaluated by the experts and insights can be used to adapt the process models. This component supporting the decision about process alternatives, links to the data analytics component and to the process engine allowing to deploy the required process re-configurations.

6) Demonstration Scenario in Smart Factory: We define demonstration scenarios for smart factories. The first scenario will be implemented in the Industry 4.0 Pilot Factory of the Vienna University of Technology. Later on similar scenarios will be demonstrated in the Pilot Factory in Graz. In both pilot factories we will capture large amounts of data from various sources in order to demonstrate and evaluate our approach towards data analytics for industrial process improvement. The demonstrations in both pilot factories will provide local small and medium sized companies the opportunity to gather new ideas for their digital transformation and define new projects.

VI. DISCUSSION AND OUTLOOK

In this paper we described the challenge of bridging advanced data analytics in manufacturing with the management of business processes. For this purpose, we proposed an approach building on low level machine data and finally informing business process design decisions.

We proposed technical components needed for this task (see. Fig. 4): (1) a process engine for managing and executing business process models on the ERP level, (2) a secure data

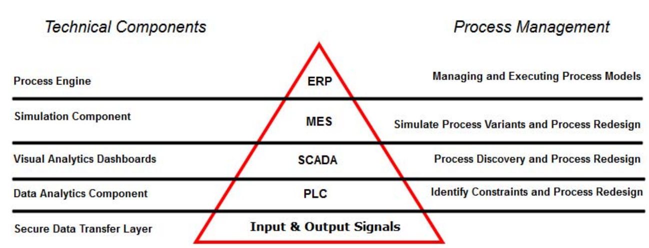


Fig. 4: Discussion Overview

transport layer allowing a secure transfer of the sensor data from the field level to our framework, (3) an analytics component for analyzing data on the control level and for building suitable predictive models, (4) visual analytics dashboards enabling experts to explore data on the process control level and (5) a simulation component to investigate the impact on the plant management level. The components 2-4 are all linked to the process engine enabling analytics aware processes by integrated process re-configurations. Finally, we presented our demonstration scenario in which we will show the applicability of the integrated tool set.

By using visualization tools for a human-centered data analysis, this project will enable cognitive decision making in real life production environments, allowing to keep the human and their expert heuristics in the loop of production processes. This can also support future workers in the new production environments and make it more secure. With these tools the full potential of human beings will be used in the future factory.

Our demonstrators implemented in the pilot factories in Vienna and Graz will help SMEs to understand the potential of Industry 4.0 with respect to data analytics for industrial process improvement.

Acknowledgment

This research is done truly joint by the two Comet K1-Research Centers ProFuture and Center for Digital Production (CDP). ProFuture (Contract Nr. 854184) and CDP (Contract Nr. 854187) are funded within the Austrian COMET ProgramCompetence Centers for Excellent Technologies under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry for Digital and Economic Affairs and the Provinces of Upper Austria and Styria (for ProFuture) and the Provinces of Vienna, Lower Austria, and Vorarlberg (for CDP). COMET is managed by the Austrian Research Promotion Agency FFG.

References

[1] Hollender, M., Collaborative process automation systems. ISA, (2010)

[2] Iiro, H., Nystrm, R., and Horch, A., "Integration of scheduling and controlTheory or practice?." Computers & Chemical Engineering 33.12: pp. 1909-1918, (2008)

[3] Cohn, M., "The forgotten layer of the test automation pyramid." Mike Cohns Blog ucceeding with Agile, http://blog. mountaingoatsoftware. com/the-forgotten-layerof-the-testautomation- pyramid (2009).

[4] Rmann, M., et al. "Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries." Boston Consulting Group (2015).

[5] Hoenisch, P., Schuller, D., Schulte, S., Hochreiner, C., Dustdar, S., Optimization of Complex Elastic Processes. IEEE Transactions on Services Computing, Volume 9, Number 5, 700-713, (2016)

[6] Schulte, S., Janiesch, C., Venugopal, S., Weber, I., Hoenisch, P., Elastic Business Process Management: State of the Art and Open Challenges for BPM in the Cloud. Future Generation Computer Systems, Volume 46, 36-50, (2015)

[7] Schulte, S., Borkowski, M., Hochreiner, C., Klusch, M., Murguzur, A., Skarlat, O., Waibel, P., Bringing Cloud-based Rapid Elastic Manufacturing to Reality with CREMA (2016). In Workshop on Intelligent Systems Configuration Services for Flexible Dynamic Global Production Networks (FLEXINET) at the 8th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA 2016)

[8] Schulte S., Hoenisch, P., Hochreiner, C., Dustdar, S., Klusch, M., Schuller, D., Towards Process Support for Cloud Manufacturing. In 18th IEEE International Enterprise Distributed Object Computing Conference, pages 142-149. IEEE Computer Society, Washington, DC, USA, (2014)

[9] Matthew O. Ward, Georges Grinstein and Daniel Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. A K Peters/CRC Press, 2015.

[10] Tamara Munzner, Visualization Analysis and Design. A K Peters, 2014.

[11] Jaeckle, D., Fischer, F., Schreck, T., Keim, D., Temporal MDS plots for analysis of multivariate data. IEEE Transactions on Visualization and Computer Graphics (Proc. VAST), 22(1):141-150, (2015)

[12] von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk J., Fekete, J.- D., and Fellner D., Visual analysis of large graphs: State-of-the-art and future research challenges. Wiley-Blackwell Computer Graphics Forum, 30 (6):17191749, (2011)

[13] Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hanspeter Pfister, Marc Streit: Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets. IEEE Trans. Vis. Comput. Graph. 20(12): 2023-2032 (2014)

[14] Partl, C., Gratzl, S., Streit, M., Wassermann, A. M., Pfister, H., Schmalstieg, D., Lex, A., Pathfinder: Visual Analysis of Paths in Graphs. Comput. Graph. Forum 35(3): 71-80 (2016)

[15] Alexander, S., Levashova, T., and Shilov N., "Semantic-oriented support of interoperability between production information systems." International Journal of Product Development 4.3-4, 225-240, (2007)

[16] Yu, Q., and Wang, K., "Applications of IoT in production logistics: Opportunities and challenges." WIT Transactions on Engineering Sciences, 113, pp 233-240, (2016)

[17] Zoier, M., Rosenberger, M., Marko, NC., Softic, S., Kaiser, C., Fachbach, B. et al. Konzeption einer Architektur fr flexible Suche und Informationsvernetzung in der Fahrzeugentwicklung. In Proceedings, pp 1-11, (2014)

[18] Rexha, A., Dragoni, M, Kern, R, Krll, M. An Information Retrieval Based Approach for Multilingual Ontology Matching, in International Conference on Applications of Natural Language to Information Systems, pp 433-439, (2016)

[19] Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H.P., Streit, M.: LineUp: Visual Analysis of Multi-Attribute Rankings, IEEE Transactions on Visualization and Computer Graphics, 19(12), pp 22772286 (2013)

[20] Maritsch, M., Kittl, C., Ebner, T.: Sichere Vernetzung von Gerten in Smart Factories mit MQTT (Secure Connection of Devices in Smart Factories with MQTT), in Koch, M., Butz, A., Schlichter, J. (Eds.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, pp 141-148 (2015)