Industrial Production Process Improvement by a Process Engine Visual Analytics Dashboard

Josef Suschnigg
Pro2Future Graz, Austria
josef.suschnigg@pro2future.at

Florian Ziessler
Pro2Future Graz, Austria
florian.ziessler@pro2future.at

Markus Brillinger
Pro2Future Graz, Austria
markus.brillinger@pro2future.at

Matej Vukovic
Pro2Future Graz, Austria
matej.vukovic@pro2future.at

Jürgen Mangler
CDP Vienna, Austria
juergen.mangler@univie.ac.at

Tobias Schreck
Graz University of Technology, Austria
tobias.schreck@cgv.tugraz.at

Stefan Thalmann
University of Graz, Austria
stefan.thalmann@uni-graz.at

Abstract

Digitalization reshapes production in a sense that production processes are required to be more flexible and more interconnected to produce products in smaller lot sizes. This makes the process improvement much more challenging, as traditional approaches, which are based on the learning curve, are difficult to apply. Data-driven technologies promise help in learning faster by making use of the massive data volumes collected in production environments. Visual analytics approaches are particularly promising in this regard as they aim to enable engineers with their rich domain knowledge to identify opportunities for process improvements. Based on the assumption that process improvement should be connected with the process engine managing the process execution, we propose a visual analytics dashboard which integrates process models. Based on a case study in the smart factory of Vienna, we conducted two pair analytics sessions. The first results seem promising, whereas domain experts articulate their wish for improvements and future work.

1. Introduction

Digitalization is currently changing the way manufacturing is done and it is promoted as Industry 4.0 or smart manufacturing. Digitalized manufacturing lines are designed to fulfill dynamic customer demands with high variability and flexibility designed for manufacturing of small lot sizes [1]. This increased flexibility and especially the higher fluctuation make the realization of learning curve effects in a digitalized manufacturing much more difficult [2]. Therefore, this poses new challenges for the production management, as most approaches focus on improving quality and efficiency in manufacturing processes that rely on the learning curve effects. Hence, new approaches are needed to make the learning curve steeper and the learning thus faster.

Digitalization promotes the application of information systems (IS) in manufacturing, leading to vertically and horizontally integrated production systems [3]. Due to this more integrated nature of production the root cause analysis and thus the improvement of such complex manufacturing systems becomes more challenging. To manage this increased complexity, more sensors are installed, and even more complex and comprehensive data sets are collected. As a consequence of these trends, the amounts of information and data generated is growing at a fast pace and poses the challenge of identifying relevant issues required for managing manufacturing in a connected supply chain [4]. In this regard, data-driven technologies and data-driven decision support seems to be a suitable answer on how to cope with this complexity [5].

Production management has so far been organized in a central and hierarchical manner, implementing the production pyramid from the Enterprise Resource Planning Systems (ERP Systems) down to the production processes. This is no longer suitable, however when it comes to more flexible production processes and customized production [6]. In such environments bottom-up approaches are needed to enable employees to analyze and optimize production processes. Hence, approaches are needed which

URI: https://hdl.handle.net/10125/63902
978-0-9981331-3-3

are suitable for analyzing huge data sets created in digitalized manufacturing environments, in such a way, that employees can quickly identify improvement possibilities and thus learn faster.

To tackle this challenge we propose a visual analytics approach using the process model of the production process [7] as main navigation element. Visual analytics would appear suitable as it aims at presenting large data sets that employees who possess the domain knowledge are empowered to draw suitable conclusions on the production process. A decentralized and bottom-up approach can be realized by this means, which is also suitable for fulfilling high flexibility demands. The process model appears to be the natural exploration model dealing with big data amounts as it represents the steps of the production process [8].

2. Related work

In this section an overview of relevant research work either in the domains of production science, process management and visual analytics is given.

2.1. Production science

In manufacturing processes, the quality requirements by the customer (e.g. a measurement within a tolerance range) must be ensured. Inadequate fulfillment of these quality requirements is often decisive for competition [9]. In the case of an inadequate fulfillment of the quality requirements, the focus lies on finding the causes in order to remedy. Since the beginning of industrialization, various methods have been developed for this purpose, which are based on experience learning, i.e. Taguchi [10], Deming [11], Ishikawa and others.

The main limitation of these methods is, that root cause analysis usually requires in-depth expert and process knowledge, but often also implicit knowledge in the form of experience. Due to the mostly complex interdisciplinary field of the problem, root cause analysis can often be very extensive and time consuming, as shown by the example of a machining process: surface defects and dimensional inaccuracies on the desired product, for example, can be traced back to vibrations. These can have causes such as the geometry, the duration of use and the wear of the tool or also by the chemical composition and the microstructure of the material to be machined. Further causes of dimensional inaccuracies on the product may have their cause in thermal expansion of the processing machine and the tool due to temperature changes. However, the influence of other internal effects, such as a cold machine vs. machine on operating temperature, and external effects, such as the variation of the shopfloor air temperature, are often neglected or ignored. Dimensional inaccuracies can also be due to elastic deformations of the part, the tool and machine components. This depends strongly on the material and the production strategy (e.g. cutting depth or feed rate) of the machining process.

During the ongoing mass customization and diversification of products, the production process needs to be more flexible. In the case of an inadequate fulfillment of high quality requirements of one customized part, the importance of a simple and low-effort method for root cause analysis emerge. The root cause analysis of quality issues using classical methods, as mentioned above, has proven to be successful, but can become very extensive and demands larger samples, especially for complex problems. However, this is challenging in the fast-moving digitalized manufacturing promising "lot size one" (single item produced).

As a possible avenue for improvement, modern methods of data analysis and visual analytics promise to reduce the effort of this process significantly [6]. This reduction of the effort is based on faster generation of knowledge (steeper learning curve) regarding the process and parameters influencing the quality requirements. A steep learning curve is crucial for a mass customized manufacturing process due to the lot size being too small to optimize the process over a long period [12].

2.2. Process management

Process management has so far mainly been used in office environments in the context of work flow management systems to span inter-organizational supply chains. However, process management offers many advantages which appear suitable for responding to the challenges of digitalization. Integrating process management and low-level process automation in an industrial context can enable a better process understanding by extracting actionable knowledge out of the collected manufacturing data [13].

Modern automated manufacturing environments comprise a vast amount of IS, which have been well researched and applied over the past few decades in several models such as computer integrated manufacturing [14] or the automation pyramid [6]. Such models explain the relations of higher level organizational IS, such as ERP, to lower level systems, such as operational machine data collection.

Pauker et al. developed a process engine "centurio.work" [7] for managing and executing production processes and also collecting and storing all related sensor data. As a process engine, it brings the flexibility of business process management systems to the shopfloor, by offering the interoperability of different IS to one holistic framework. Accordingly, the process engine collects data from different and diverse systems and facilitates the problem of structuring and combining heterogeneous data, by offering a process-based interface. This enables the application of process mining enhancements [15] by domain experts to trigger process improvements on the production directly on the shopfloor by flexible production processes.

2.3. Visual analytics

Industrial contexts have a strong need for domain experts working with generated data to include their domain knowledge and to interpret data analysis results. For this purpose, visualization and visual analytics research effectively satisfy multiple demands of the new production and management models in Industry 4.0 [16]. In general, visual analytics is the combination of automated data analysis algorithms and user interactions through information visualization [17].

The knowledge of domain experts is added by user interactions on the information visualization component and is included to knowledge as the outcome of a data analysis iteration. Recently, an extensive survey study on visualization and visual analytics research in industrial environments, more specifically in smart manufacturing, has been published [16]. It demonstrates the need engineers working with industrial data have for support and also the breath and diversity of industrial applications. The requirement of industrial data analysis for extensive professional and domain-specific knowledge has also been emphasized. Another research question has been raised due to complex manufacturing systems and the need of coupling heterogeneous data sources. Therefore, our research work shows an approach to integrate (1) production data from production machines and (2) data on a higher level of production: process management data.

A number of studies have been published on the field of information visualization and visual analytics of business process models. In the work of Van der Aalst et al. [18] the need of a combination of the approaches of data mining, process mining, visualization and visual analytics has been identified. An approach is given on how to “breathe life“ into otherwise static business process models by visualizing the change of business process states over times. The need for user guidance in executing activities in business processes has been defined for relevant future work, which is also currently investigated in visual analytics research. The position paper presented by Gschwandtner [19] explores the opportunities for combining visual data exploration with process mining algorithms. As an opportunity it makes complex information structures more comprehensible, but also the challenges regarding the combination of those fields are identified.

Production machine data, in general, consists either of time-oriented continuous data or discrete event data. A vast number of visualization techniques are available for visualizing time-dependent data [20]. In this work, however, time series are extracted to scalar values by calculating statistically significant characteristics to features [21], creating time-independent multivariate data records for production items. A scatterplot matrix (SPLOM) [22] based visualization for comparing production entities and interactively modeling [23] of relations between extracted features has thus been identified as suitable for visual analysis of production data.

3. Approach

In this section, first the approach on how to access data of several data sources from the process engine is given. After the data has been acquired, the data is prepared, and relevant features are extracted and selected to be visualized in the visual analytics dashboard. As the last step, the concept on how to modify the flexible production process model by knowledge of visual analytics is presented.

3.1. Visual analytics process dashboard

Sarikaya et al. [24] defined an extensive design space of dashboards. This provides a clear structure on visual and functional aspects of dashboards and therefore is taken as the framework for introducing our visual analytics process engine dashboard approach. As a consequence, we introduce the dashboard according to the four categories: (1) purpose, (2) audience, (3) visual & interactive features and (4) data semantics as follows.

3.1.1. Purpose

The main purpose of our dashboard is to support shopfloor engineers in their operational decision making on how to adapt a flexible production process according to knowledge of interactive visual data analysis and production related domain knowledge. It is used in an operational manner and connects knowledge of current production data to the production process. In further consequence, the production process can be adapted and therefore optimized in the context of several indicators (i.e., costs, quality and time). Such modifications on the production process can be either, inserting subprocesses, removing subprocesses or BPMN (business process model notation) gateway decision making on a running process instance.

3.1.2. Audience

The dashboard exclusively addresses domain experts, or more specifically production engineers working on the production shopfloor. These domain experts have an understanding of the production process from a production science point of view, and also of production and quality measurement machines. Such users thus also have a great understanding of the generated raw data, but due to the large amount and the complexity of it, the request for a tool to support the usage of data has been identified. The dashboard is an expert system and the required visualization literacy can be high. Therefore, visualization techniques such as SPLOMs and BPMN do not limit the comprehensibility of the dashboard.

3.1.3. Visual & interactive features

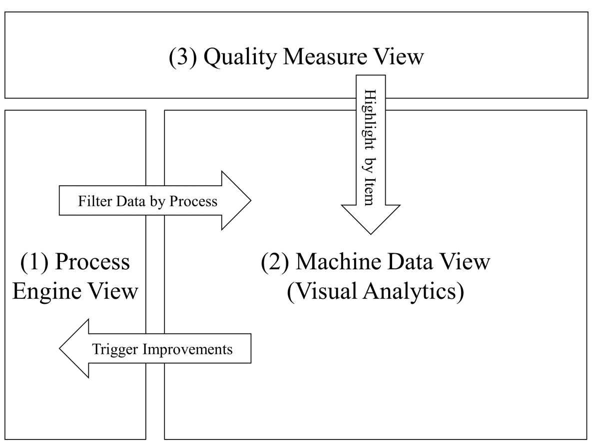

Figure 1. Scheme of dashboard views and their interactions

The dashboard consists of a single page and non-customizable user interface. It connects three different views, each representing a different data source and allows interaction between these sources. Each of the three views is related to a different entity of production process data: (1) Process engine view, (2) Machine data view and (3) Quality measure view. Their interactions are illustrated in a schematic overview in Figure 1 and the prototypical realization is illustrated in Figure 4.

The process engine view (1) visualizes the business process model. Clicking on subprocesses acts as a filter to visualize only the machine data from that specific subprocess in the machine data view. The machine data view (2) consists of a visual analytics tool enabling interactive data analysis.

To analyze machining time series data, features are extracted from machine data channels for each produced item. In the following, statistically significant features in regard to the quality are selected by the method introduced by Christ et al. [21]. Consequently, the most significant features are visualized in an interactive SPLOM, enabling the visual analysis of features. One analysis goal is to identify multivariate data points that do not fit to a specific distribution, familiar from previous data or domain knowledge. If an interesting data point has been identified with the context of a specific problem, the expert user can adapt the flexible production process in the process engine view to trigger counteractions. The (3) quality measure view, shows the current progress of the quality of items in a line plot. It also acts as a selection, for the machine data view. When clicking on an item in the line plot, the data points of that item are highlighted in color in the machine data view.

3.1.4. Data semantics

Data semantics of the type “Updateable” can be provided by the dashboard: The dashboard is updateable, since it visualizes data of the last produced item in the process engine view and the last 500 items in the machine data view and the quality measure view.

4. Demonstration case: pilot factory

To develop, test and improve our research work on smart production, it was conducted in the pilot factory of the TU Wien [25]. One goal of the pilot factory is to create a realistic environment for research on real-time communication infrastructure and distributed production systems. Thus, items are not only produced for demonstration purposes, but rather for real industrial use. The product, which has been produced for our industry partner in our demonstration case, has the dimensions of approximately 1 inch in length, width and height, and the production machining takes about 3 minutes in a turning machine. Given that produced items are used in a complex security mechanism of another product, the challenge of production was able to achieve accuracy and low production tolerances of down to $6\mu m$.

The quality measurements take twice the time required for the actual production. Therefore, the intention of our work is either to (1) anticipate a good quality through visual data analysis and consequently skip the costly quality measure routine or (2) anticipate a bad quality, skip the costly quality measure routine and trigger counteractions. Both of these aims are strongly related to the flexible production process, which can be adapted by the machine operator in real time through the process engine. Such adaptations are either related to removing or including subprocesses or decision making on a BPMN gateway of the production process.

The entire production process is planned and executed within the "centurio.work" [7] process engine and enables the data collection for several data sources, but also the adaption of the production process. As a consequence, our dashboard approach receives all data from a single process engine instance for each item.

In the following we describe how to acquire and parse data from the process engine, how to process data for visualization purposes and how we designed a prototypical dashboard with production data for evaluation and discussion purposes.

4.1. Process engine data pre-processing

All of the business and production tasks, from the ERP systems, all the way down to the single machine instance, are orchestrated using the process engine. The process engine collects and structures the data from different sources and provides an interface for data exchange to other systems. It stores data in a form of YAML [26] stream files. Each file is isolated based on: (1) a task within production process and (2) position of the task in the automation pyramid. A file is mapped one-to-one with the activity/event in the BPMN diagram within a level of automation pyramid. The files contain streams of log entries, whereas a log entry is a YAML object that describes a single execution step of the BPMN activity or BPMN event. For example, an execution step is: production start, creation of a subprocess, start of a subprocess or reception of machining data. A sample file structure with log entries is shown in Figure 2.

Since observing the data from one task is not enough and each level in the automation pyramid contains the valuable information, it is necessary to make the connections between these isolated log files and provide a more holistic view of the production.

Finding direct connections between files on the same level in an automation pyramid is not possible in this context. A connection between files is established through certain types of execution steps found in the log files. For example, in the top level of the automation pyramid, the activity that processes orders can be found. This activity creates a task for each production item and


```yaml
---
event:
  trace:id: '395'
  concept:name: GV12 Turn
  concept:endpoint: https://centurio.work/flow/start/url/
  id:id: a1
  cpee:uuid: 45d4f9fe-5a00-42cf-9360-b2a0f40080b6
  lifecycle:transition: complete
  cpee:lifecycle:transition: activity/done
  time:timestamp: '2018-10-17T11:26:25+02:00'
---
event:
  trace:id: '395'
  concept:name: QC Shop Floor
  concept:endpoint: https://centurio.work/out/measure/tasks/
  id:id: a2
  cpee:uuid: 45d4f9fe-5a00-42cf-9360-b2a0f40080b6
  lifecycle:transition: unknown
  cpee:lifecycle:transition: activity/receiving
  list:
    data_receiver:
    - name: info
      mimetype: application/json
      data:
        task: https://centurio.work/out/measure/?task=588d8d3b
  time:timestamp: '2018-10-17T11:26:26+02:00'
```


Figure 2. Sample YAML file structure

within the log files of this activity, an execution step with an identifier of this sub-task can be found. A similar approach is used for establishing connections between files on the lower levels of the automation pyramid. Once all files are connected, a tree structure is created connecting all of the files created within a production run. This structure (production tree) is visible in the Figure 3.

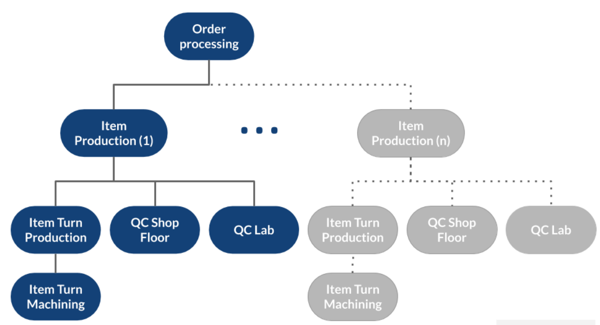

Figure 3. Tree structure generated by connecting files through the automation pyramid

During the process of the production tree generation, data is adjusted to a format suitable for the next data analysis tasks. The production of each product generates around 10MB of data that is retrieved through an interface with the process engine through an application program interface (API). During the parsing and processing of the files, data is extracted and stored in a relational database. A relational database is used as the production data in the files is represented in a complex form and the process of parsing the files and establishing connections between them is time consuming.

Once the data is available in a more suitable format, respectively stored in a relational database, it can be easily extracted by querying the database using a database query language. In this extraction step the data is cleaned. Products with incomplete data samples are removed and sparse data samples are completed if possible (values are interpolated or replicated). After this the data is ready for information extraction.

For our purpose we extract the following types of data: (1) machining data, (2) quality data, (3) meta data. Machining data (1) contain time series data of the production turn machine from several channels and three axes (x, y, z) like, torque, tool movement, energy consumption, spindle speed or tool related data like the tool length of the current tool. Overall 14 machining data channels are used. Quality data (2) is related to the production process (see Figure 4 on the left) and contain two different quality control measurements. "Quality Control 1" which is related to a manual quality measure by the machine operator and "Quality Control 2", which is related to an automated measurement process by a measurement machine. Extracted meta data (3) are for example production duration for each item or also pause between the machining of two successive items.

4.2. Feature extraction

In this section we discuss further preparation of the data to analyze and to feed it to a feature extraction algorithm. Since we are dealing with time series we must find a way that allows extraction of the important information and at the same time leads to a reduction of the dimension that comes along with time series. This is where feature extraction and as a concluding step feature selection is needed. According to Meyer-Baese et al. [27] the foundations of these methods are mathematical functions that describe the data through features (or attributes). Feature extraction and of even more important finding the right features are the key factors to a successful machine learning project [28] and also the most time consuming part [29]. As this is a crucial task we decided to use "tsfresh" [21], a library for Python, that both extracts and selects statistically significant features of time series data. The library calculates 63 different well-known features and their variations for a single time series that result in 794 features overall. In our experiments this led to over 11000 features calculated for each of the 14 machining data channels.

After the relevant features have been selected for each machining subprocess (or tool) and each produced item, for which "tsfresh" uses statistical hypothesis tests [21], up to ten features for each subprocess remained. Those include in our case simple generic functions like "minimum" and "maximum" of a series, as well as features which were added by hand to the feature selection process, like pauses between produced items, as they turned out to be relevant. "tsfresh" also contains more sophisticated functions such as "coefficients of the one-dimensional discrete Fourier Transformation", which returns a complex value, the real part, the imaginary part, the absolute value or the angle. We emphasize that complex features extracted by feature extraction methods contain information about parameters of interest, but may not be comprehensible by the user. However, "tsfresh" enables a fast solution for extracting and selecting relevant features for a given time series and may work for the demonstration case. In other use cases, a more extensive feature engineering, also including domain knowledge, may be more expedient.

5. Demonstration - pair analytics study

To demonstrate an exploratory evaluation of our approach, we conducted a pair analytics study [30] to capture the reasoning processes in our visual analytics approach. In general, pair analytics is an in-vivo study, generating verbal data in a naturalistic human-to-human dialog, to capture the reasoning behind the interactions taken with a visual analytics tool. Pair analytics requires two participants with different roles. The first role is the Subject Matter Expert (SME), which has expertise in the domain of the analysis task but may be unfamiliar to visual analytics tools. The second role, Visual Analytics Expert (VAE), may lack of domain knowledge, but has technical expertise in operating visual analytic applications. During a pair analytics session both participants attempt to fulfill an analytical task, whereas according to their roles the SME acts as the "navigator" and the VAE acts as the "driver", while performing the task.

We performed two pair analytics sessions of 90 minutes each with two experienced mechanical engineers as SMEs. Both SMEs are directly involved in our demonstration case (i.e., one as production machine operator and one as researcher in the project) and therefore having good knowledge about the production process, the production machine and the data. In addtio to the investigation of reasoning in visual analytics, we are also interested in the following questions in pair analytics sessions: (1) Does our approach support domain experts in conducting root cause analysis faster? (2) Is the proposed visual analytics approach suitable for improving the process understanding? (3) Can new knowledge about the production processes be generated by exploratory visual analysis?

5.1. Visual analytics process engine dashboard demonstration case

As described in Section 3.1 the proposed dashboard consists of three different views. The application case, as we discovered in the sessions, is illustrated in Figure 4. The dashboard enables user interactions between views, whereas the focus stays at the visual analytics view: On the left side the process engine view visualizes the production process, which contains two groups of subprocesses: (1) machining subprocesses and (2) quality control subprocesses.

The production process view acts as filter for the visual analytics tool in the machining data view. In the machining data view on the right of the dashboard an interactive SPLOM is shown. It enables the interactive visual analysis of bivariate relations of extracted features for each item. As an overview of the current production quality measures, the quality measure view on the right top the progress of the last item's quality is visualized. The view offers the capability of selecting items of interest and to highlight data points accordingly in the machining data view. Default-wise the last produced item is selected.

To minimize the impact of high cognitive load caused by complex visualization techniques, we decided to choose basic visualization methods in the dashboard. First of all, the BPMN diagram is a standard for process modelling and visualization, and thus seems suitable for the process engine view. For the quality measure view a line plot visualizing an univariate time series is reasonable and also highly comprehensible. The SPLOM in the machining data view adds complexity to the dashboard, but should be comprehensible to the target group (engineers and machine operators). Our preliminary experiments showed that only a small number of up to 10 features contain relevant information about the target variable and therefore the SPLOM is still scalable for our use.

The workflow of the application case in Figure 4 is marked by ascending letters from (a) to (f) and can be interpreted as follows: (a) In the quality measure view an outlier on the last measured item has been visually detected. Consequently, as an analysis goal the responsible subprocess needs to be identified, by exploring machining subprocess data separately. Therefore, data in the visual analytics view can be filtered by subprocesses and by selecting the according subprocess in the process view, which is highlighted by the yellow background color (b). Single SPLOMs can be selected in the interactive SPLOM on the left side of the visual analytics view. Selected scatterplots are visualized on a bigger scale on the right side of the visual analytics view (c). In (d), the actual production item data point is highlighted by its square shape in the scatterplot, comparing the production quality (y-axis) with the 'tool length' feature (x-axis). It can be visually identified as an outlier.

No item with a similar tool length with a high quality has been produced to intepret this finding, and therefore one can assume, that an issue in the production process has been identified. As a result, the decision to skip the costly quality control (f) on items, which have been produced during the occurrence of the issue, can be made by the BPMN gateway in (e). As last step, the machine operator can fix the tool length issue and thus continue production.

The exemplary application case shows that a machine operator can quickly identify problems in the running production process. As the visual analytics dashboard is directly linked to the process engine, changes can directly be performed.

6. Evaluation results and discussion

Participants of the pair analytics study agree, that the connection of the process engine with visual analytics techniques in the dashboard could be of great benefit. One SME said:

"The connection between the process engine and the visual analytics dashboard is not only beneficial, it is essential. Because otherwise data is way too abstract - you need something to navigate which is related to the real process."

The participant is aware, that modern digitalized manufacturing systems generate huge amounts of diverse and abstract data. Both participants emphasized that they need support to make sense out of the data and to interpret the data. The production experts have the currently running production machining process in mind, in which its subprocesses are assigned to specific tools. In order to be able to monitor the process and to identify opportunities for process improvements, the visualization of the product quality is key as one participant explained:

"During the production run I am responsible that the quality is perfect. For this reason, the chart showing the production quality is key. If I see some deviations, I know something is wrong and I can start exploring - I think this is a good start to trigger a root cause analysis."

One major motivation for our approach was to speed up the learning curve in production processes. Due to limited repetitions and small lot sizes, analyzing production processes ex-post makes less sense. Hence, visualizing the data in time is key to speed up the process

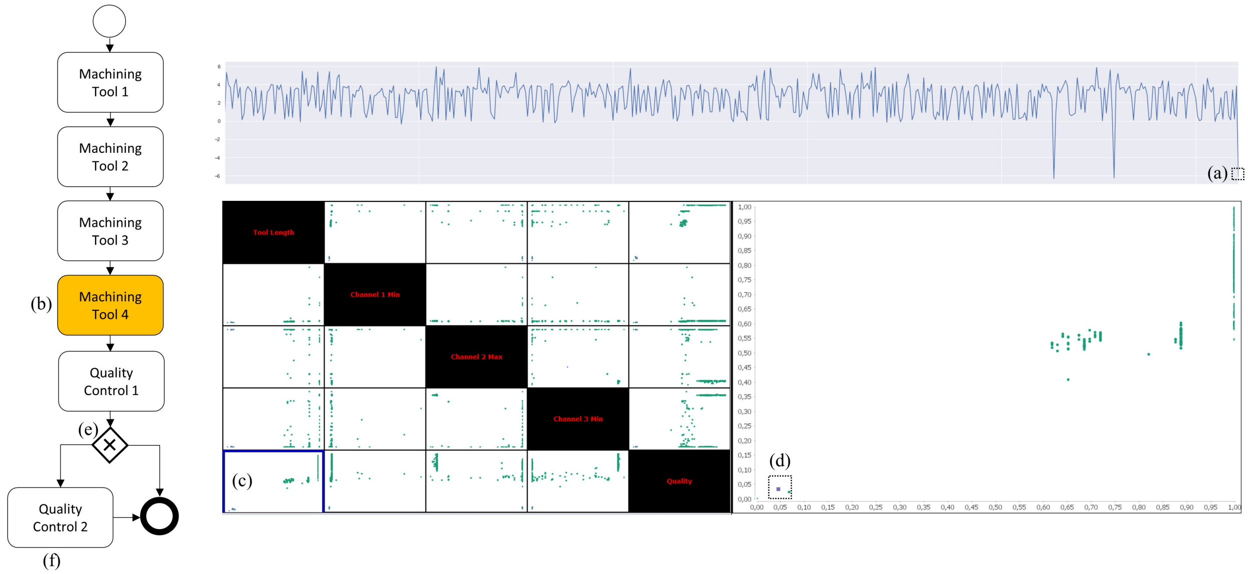

Figure 4. Visual analytics process engine dashboard

improvement process. The second important aspect in this regard is, that the engineer is quickly able to detect problems.

Even if the visualization helps to find problems, it does not provide a guidance. Both SMEs would like to have explanations for patterns and for some of the complex features, extracted by the feature extraction library. The demonstration case is an example on how to identify the root cause of a currently occurring problem in a running production process. In complex cases the engineer must search for and evaluate different possibilities. This requires experience and domain knowledge, whereas the visualization supports in this process. However, an automation and delivery of advanced explanations currently appears to be difficult to realize, although this would be desirable from an engineer's point of view.

Another aspect is also relevant for the application of our prototype. Production workers and engineers are no longer responsible for one machine only in which they can use their senses intuitively to monitor its operation. Among the trends of digitalization is that one single employee is responsible for ever more machines. An SME resonates about this issue:

"If you have several machines to monitor and different products, you do not always know the reasons for why things go wrong. If you have only one machine producing the same item for 20 years this is different. You may be able to hear what is wrong from the sound it is making."

The exploration of quality issues is the major driver to identify possibilities for improvement. The SMEs also mentioned knowledge about the machine independently from the current production process which seems relatively stable and knowledge which depends on the currently produced product. Our approach focuses on the second type of knowledge and this is also one motivation to connect the visualization to the process model of the currently performed production process. One SME said:

"If I have to explore a problem in the production process, this tool could be helpful. I can find the reason and then I know what is happening and what I have to do. If it is something serious and if it has to do with the production process I can change the process model."

Data exploration in general also has the purpose of generating hypotheses. One participant explored a specific scatterplot and said: *"I see, that threshold on this feature is at* 0.8 *for good parts."* or *"I see in this interesting area, that no bad item has been produced."*. Therefore, we can assume, that the data exploration generates hypotheses. However, those two statements were followed by: *"But I don't know if it is statically significant"* and *"I need to be careful, I might possibly be seeing something that is not there"*. Hence, the participants generated hypotheses, but collecting evidence to confirm or reject these is not an easy task, especially when the participants lack an understanding of the complex features involved.

One the one hand, a participant appreciates the link to the process engine and the stored models. On the other hand, he remarks that most identified problems will not result in process changes. Including an option to capture such minor insights also seems useful and is a valuable direction for future work. Apart from this our SMEs had further ideas for how to improve our work. One participant stated that *"temperature and vibration sensors on different locations on the production machine"*, could enable better insights to the current machines condition. Another request, we also faced in other research projects, is that there is a need for processing meta data about the machine, i.e. cooling liquid level, tool change or oil change records.

The concept still needs to be proved, however, by other more complex datasets. The production process has been designed by requirements of the industry partner and therefore participants had a broad understanding of the subprocesses, regarding the machining and quality control. They agree that our approach can be helpful for better process understanding, especially if anomalous behavior in comparison to a production that is performing well occurs. Hence, new knowledge can well be generated and subsequently the production process can be optimized based on this.

During our work we encountered challenges explored by Gschwandter in [19], however our main motivation is the intertwining of production process analysis and visual analytics. The scheme presented in Figure 1 works for our use case and also possibly for other production environments, but the different views of the scheme strongly depend on the needs of a specific application and its tasks. For example, the design choice for the SPLOM is suitable for the use case and its small feature space after relevant features have been selected. Other use cases may require different visualization techniques and therefore the visualization methods in the presented approach should be considered as exchangeable.

Zhou et al. identified data integration as an ongoing challenge for industrial data visualization [16]. Our work thus demonstrated an approach on how to integrate data of varied data sources in a single dashboard. In this regard, we noticed that the use of the process engine "centurio.work" [7] in the pilot factory is helpful for data acquisition and subsequent data prepossessing.

For future work we plan to investigate how users can be guided in their use of the presented dashboard. For example, visually highlighting interesting data points with visual quality metrics [31] is a suggestion for further exploration that could be beneficial to the user. Another challenge we address for future work is the reduction of complexity of extracted features by the feature extraction library. For this purpose a simple solution, such as adding drill-down capabilities to explore and compare time series directly, after these have been identified as interesting in the SPLOM, can be helpful to the user. We see that this approach needs additional effort input by working on other use cases and more costly evaluations. In this context we plan to continue our work in another pilot factory to research more significant results on other use cases, production environments and evaluations with different SMEs.

7. Conclusion

Digitalization of production requires approaches which allow a quick and timely investigation of production processes. Therefore, we investigated an approach on how to analyze production data through a process engine visual analytics dashboard. The contribution made by our research work is (1) showing the intertwining of production process analysis and visual analytics and (2) industrial data integration for data visualization, both investigated on an industrial use case. We have collected first constructive and positive feedback from the SMEs. These both also expressed wishes and ideas for further future improvement. Most prominent among these was the wish that the approach should provide a specific level of explainability and guidance during the root cause analysis. In particular, some extracted features are too complex and incomprehensible to the target group. This is also related to the current research topic on the explainability of artificial intelligence. Our case study was conducted in a pilot factory with real machines and real production tasks. Even though the work has been done in an environment close to a real production setting, the next step for future work, that would appear to be promising, should be the evaluation in real industrial production settings over a longer time period.

Acknowledgements

This research is done joint by the two Comet K1-Research Centers Pro^{2}Future and Center for Digital Production (CDP). Pro^{2}Future (Contract no. 854184) and CDP (Contract no. 854187) are funded within the Austrian COMET Program "Competence 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 Pro^{2}Future) and the Provinces of Vienna, Lower Austria and Vorarlberg (for CDP). COMET is managed by the Austrian Research Promotion Agency FFG.

References

[1] T. Niesen, C. Houy, P. Fettke, and P. Loos, “Towards an integrative big data analysis framework for data-driven risk management in industry 4.0,” in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 5065–5074, IEEE, 2016.

[2] F. Longo, L. Nicoletti, and A. Padovano, “Smart operators in industry 4.0: A human-centered approach to enhance operators capabilities and competencies within the new smart factory context,” *Computers & industrial engineering*, vol. 113, pp. 144–159, 2017.

[3] K.-D. Thoben, S. Wiesner, and T. Wuest, “Industrie 4.0 and smart manufacturing-a review of research issues and application examples,” *International Journal of Automation Technology*, vol. 11, no. 1, pp. 4–16, 2017.

[4] F. Kache and S. Seuring, “Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management,” *International Journal of Operations & Production Management*, vol. 37, no. 1, pp. 10–36, 2017.

[5] S. Thalmann, “Data driven decision support,” *it-Information Technology*, vol. 60, no. 4, pp. 179–181, 2018.

[6] S. Thalmann, J. Mangler, T. Schreck, C. Huemer, M. Streit, F. Pauker, G. Weichhart, S. Schulte, C. Kittl, C. Pollak, *et al.*, “Data analytics for industrial process improvement a vision paper,” in *20th Conference on Business Informatics (CBI)*, vol. 2, pp. 92–96, IEEE, 2018.

[7] F. Pauker, J. Mangler, S. Rinderle-Ma, and C. Pollak, “centurio.work - modular secure manufacturing orchestration,” in *Proceedings of the Dissertation Award, Demonstration, and Industrial Track at BPM 2018 co-located with 16th International Conference on Business Process Management (BPM 2018)*, pp. 164–171, Springer, 2018.

[8] W. M. Van Der Aalst, “Business process management: a comprehensive survey,” *Software Engineering*, vol. 2013, 2013.

[9] D. I. Prajogo, “The relationship between competitive strategies and product quality,” *Industrial Management & Data Systems*, vol. 107, no. 1, pp. 69–83, 2007.

[10] M. Nalbant, H. Gkkaya, and G. Sur, “Application of taguchi method in the optimization of cutting parameters for surface roughness in turning,” *Materials & Design*, vol. 28, no. 4, pp. 1379 – 1385, 2007.

[11] M. Rungtusanatham, J. A. Ogden, and B. Wu, “Advancing theory development in total quality management: A deming management method perspective,” *International Journal of Operations & Production Management*, vol. 23, no. 8, pp. 918–936, 2003.

[12] M. Lohmann, M. J. Anzanello, F. S. Fogliatto, and G. C. da Silveira, “Grouping workers with similar learning profiles in mass customization production lines,” *Computers and Industrial Engineering*, vol. 131, pp. 542–551, 2019.

[13] S. Thalmann, H. Gursch, J. Suschnigg, M. Gashi, H. Ennsbrunner, A. Fuchs, T. Schreck, B. Mutlu, J. Mangler, G. Kappl, C. Huemer, and S. Lindstaedt, “Cognitive decision support for industrial product life cycles: A position paper,” in *COGNITIVE 2019*, vol. 11, pp. 3 – 9, IARIA, 2019.

[14] M. P. Groover, *Automation, Production Systems, and Computer-Integrated Manufacturing*. Prentice Hall Press, 3rd ed., 2007.

[15] W. M. P. Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, vol. 136. Springer, 2011.

[16] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren, “A survey of visualization for smart manufacturing,” *Journal of Visualization*, 2018.

[17] D. Keim, G. Andrienko, J. D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, “Visual analytics: Definition, process, and challenges,” in *Lecture Notes in Computer Science*, Springer, 2008.

[18] W. M. P. Van Der Aalst, M. De Leoni, and A. H. M. Ter Hofstede, “Process mining and visual analytics: Breathing life into business process models,” in *Computational Intelligence*, Wiley, 2012.

[19] T. Gschwandtner, “Visual analytics meets process mining: Challenges and opportunities,” in *Lecture Notes in Business Information Processing*, Springer, 2017.

[20] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*. Springer, 2011.

[21] M. Christ, N. Braun, J. Neuffer, and A. Kempa-Liehr, “Time series feature extraction on basis of scalable hypothesis tests (tsfresh a python package),” *Neurocomputing*, 2018.

[22] W. S. Cleveland, *Visualizing data*. Hobart Press, 1993.

[23] L. Shao, A. Mahajan, T. Schreck, and D. J. Lehmann, “Interactive regression lens for exploring scatter plots,” *Comput. Graph. Forum*, vol. 36, pp. 157–166, June 2017.

[24] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher, “What do we talk about when we talk about dashboards?,” *Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 682–692, 2019.

[25] TU Wien, “Pilot factory of the tu wien.” "`http://pilotfabrik.tuwien.ac.at/en/`. [Online; accessed 2019-06-04].

[26] O. Ben-Kiki, C. Evans, and B. Ingerson, “Yaml ain’t markup language (yaml) version 1.1,” *yaml. org, Tech. Rep*, p. 23, 2005.

[27] A. Meyer-Baese and V. Schmid, “Chapter 2 - feature selection and extraction,” in *Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition)* (A. Meyer-Baese and V. Schmid, eds.), pp. 21 – 69, Academic Press, second edition ed., 2014.

[28] E. Y. Ahn, T. Mullen, and J. Yen, “Finding feature transformation functions using genetic algorithm,” in *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO ’10, pp. 2061–2062, ACM, 2010.

[29] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[30] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher, “Pair analytics: Capturing reasoning processes in collaborative visual analytics,” in *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10, IEEE, 2011.

[31] M. Behrisch, M. Blumenschein, N. W. Kim, L. Shao, M. El-Assady, J. Fuchs, D. Seebacher, A. Diehl, U. Brandes, H. Pfister, T. Schreck, D. Weiskopf, and D. A. Keim, “Quality metrics for information visualization,” *Computer Graphics Forum*, vol. 37, no. 3, pp. 625–662, 2018.