Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 93 (2020) 449–454

www.elsevier.com/locate/procedia

53rd CIRP Conference on Manufacturing Systems

Data-driven Analysis of Product State Propagation in Manufacturing Systems Using Visual Analytics and Machine Learning

Marc-André Filz[a,*], Sebastian Gellrich[a], Christoph Herrmann[a], Sebastian Thiede[a]

[a]*Institute of Machine Tools and Production Technology, Chair of Sustainable Manufacturing and Life Cycle Engineering, Technische Universität Braunschweig, Langer Kamp 19 b, 38106 Braunschweig, Germany*

* Corresponding author. Tel+49-531-390-65032; fax: +49-531-391-5842. *E-mail address:* m.filz@tu-braunschweig.de

Abstract

The importance of quality and efficiency has increased in recent years. Moreover, the rise of computational power and the development of advanced analytics has enabled the industry to enhance the performance of manufacturing systems. Therefore, further transparency of intermediate product states is necessary to derive appropriate actions. The goal of this paper is to develop a framework to enable the data-driven analysis of product state propagation within manufacturing systems to improve the transparency of product quality related cause-effect relationships. Based on their intermediate product features, machine learning algorithms assign products to classes of similar characteristics. This approach is practically applied to a case study from the electronic production industry. By using visual analytics tools, the propagation of product states along the manufacturing process chain is exemplarily analyzed.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 53rd CIRP Conference on Manufacturing Systems

Keywords: Intermediate Product States; Product State Classes; Machine Learning; Product Propagation; Visual Analytics; Manufacturing Systems

1. Introduction

Today, manufacturing companies are exposed to increased complexity and competition. To stay competitive, manufacturers need to increase their efficiency by ensuring high quality standards, low cost and short lead-times. In this environment, successive companies have to improve their quality constantly to meet customers' demands [1].

Moreover, digitalization is a major driver for change in manufacturing. In this context, there are already numerous technologies, tools and methodologies that have already been implemented or are in the process of implementation [2]. One example are cyber-physical production systems that include approaches and techniques to gather and process data from the physical production environment, modeling these data with data analytics in combination with simulation for forecasting or real time control approaches and providing decision support to the shop floor [3].

To improve the overall quality and reduce scrap, it is necessary to consider the complete manufacturing system instead of single and isolated processes. Therefore, it is important to understand dynamics and interdependencies within the manufacturing system. To derive proper quality management strategies, knowledge about the propagation and the detection of intermediate product feature changes along manufacturing processes is essential [4].

Against this background, it is necessary to gather information about the product features along the process chain. The aim of this paper is to develop a framework which characterizes different product states based on manufacturing related data. Moreover, based on machine learning algorithms, product states with similar features are supposed to be assigned to classes. By visualizing the propagation of these classes

2212-8271 © 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 53rd CIRP Conference on Manufacturing Systems
10.1016/j.procir.2020.03.065

through the complete manufacturing system, the process and product quality transparency can be improved. In addition, the framework is supposed to allow the analysis of interactions and the product state propagation within the manufacturing system.

2. Characterization of Intermediate Product States

The aim of manufacturing processes is to transform raw material into final products by value adding to meet the customers' requirements. This value adding consists of the physical transformation of the products (e.g. transforming geometry, aggregate state, chemical composition). Consequently, product characteristics or states change with every process step [5].

Observing or monitoring products by their intermediate states allows describing and highlighting the product changes or transformations over time. Thus, monitoring product states along the whole manufacturing processes enable an overall view of the realized changes, leading to specific product features [5]. Figure 1 gives an overview of product states and their intermediate features within a manufacturing process chain.

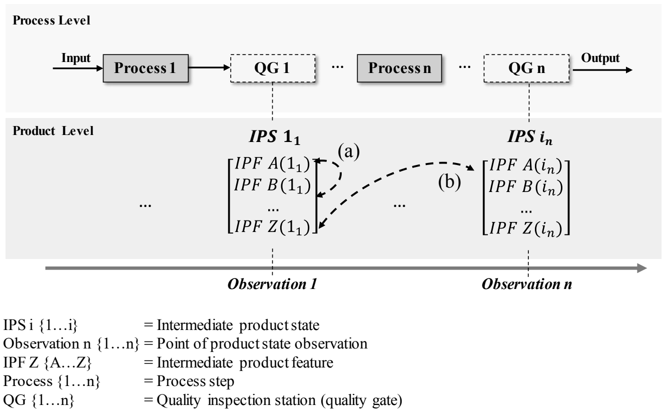

Figure 1: Overview of product states and intermediate product features and states (adopted from [5])

An intermediate product state (IPS) characterizes a product at a certain observation point along the manufacturing process chain by a combination of intermediate product features (IPF) [5]. IPF are defined as a quantitative (e.g. temperature) or qualitative (e.g. material composition) description of the product itself as well as definable and deterministic measurements [6,5].

A product state changes due to external influences like manufacturing processing from one observation to another when at least one IPF is changed [6,4].

Theoretically, product features change constantly within the manufacturing process chain, leading to a strict time dependency for product states. Therefore, it is of high importance to define observation points that are similar in the sense of comparability of different product states. Here, quality inspection stations are useful, since they are installed in predefined places in the process chain. Moreover, due to the use of digital methods and tools, virtual inspection stations can be installed at almost every possible observation point. This supports, for example, the definition of product states that cannot be achieved with conventional measurement technologies, e.g. within a process or during a time-dependent process (reflow oven). The use and allocation of inspection stations to individual processes is again very process- and product-dependent.

In addition, it is necessary to describe the IPS in a comprehensive way. Therefore, it is important to choose relevant IPF for certain observations. To identify these, it is useful to consider the whole manufacturing system [7]. In this context, data analytics tools can be very helpful to identify relevant product features.

Within a manufacturing system, different dependencies or interdependencies can occur. Regarding the IPS, bi-directional relations between IPF within one product state may exist (see Figure 1 (a)). However, a forward relation can appear between a product state and their successor (see Figure 1 (b)). Forward relations can occur between two following product states or between random product states [6].

Applying this approach to quality management, product states could be used to forecast final states along the process chain. Moreover, deviations in the product states can be detected and appropriate actions can be taken to ensure the required quality.

3. Conceptual Framework for Data-Driven Analysis of Product State Propagation

For analyzing the behavior and interactions of IPS within the manufacturing system, the proposed framework strives at building classes of similar IPS at defined observations, so-called intermediate product classes (IPC). To define IPC, IPF can be used as representatives for one observation to identify state characteristics. Figure 2 demonstrates the idea of IPC and their propagation within a manufacturing system.

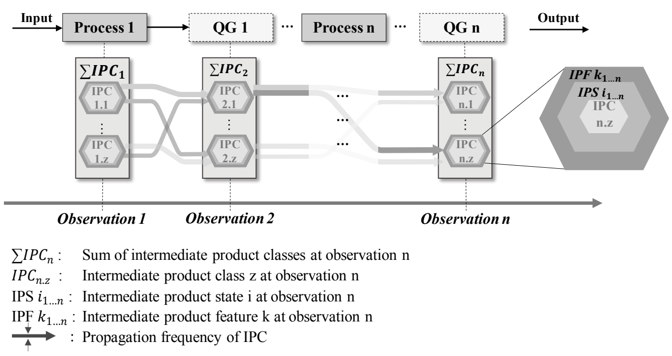

Figure 2: Overview and propagation of IPC within a manufacturing system

Each box represents an observation, whereas $IPC_{n.z}$ characterizes one IPC for one observation based on clustering of IPF. The amount of different IPC is therefore dependent on the individual observation and the corresponding available IPF as input data (e.g. inspection results).

In order to derive characteristic IPC, a framework for a data driven analysis using machine learning is proposed in Figure 3.

The approach draws on clustering methods that need less parameterization effort and supports the resulting models' interplay with visualizations for model analysis and knowledge building.

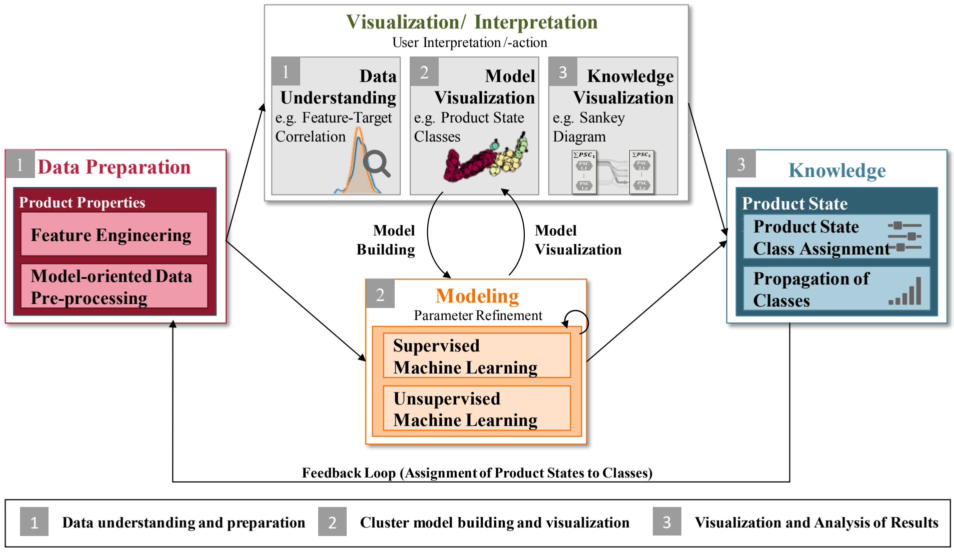

Figure 3: Framework for data-based analysis of product state propagation through the eyes of visual analytics (adapted from [13])

Visualizations play a vital role in industrial data analysis, e.g. facilitating data understanding, model evaluation and knowledge extraction from models. In this context, visual analytics (VA) is a multidisciplinary approach that combines data analysis with visualizations. VA can be described as "the science of analytical reasoning facilitated by interactive visual interfaces" [8]. The process of VA is characterized by a human-centered approach, facilitating hypothesizing and an informed decision-making through complexity reduction of huge data sets [9]. Through VA the user is enabled to deal with massive, dynamically changing data and detect anomalies, changes, patterns and relationships, in order to gain new knowledge [10]. Keim et al. (2008) proposed a framework that structures elements and processes of VA by describing the interplay of data acquisition, models, visualizations and knowledge extraction [11]. The framework of visual analytics delivers an appropriate foundation for a structured mapping of the proposed modeling procedure (see Figure 3).

To analyze the behavior of IPC along the processes within a manufacturing system, the proposed framework is based on VA. The framework is graphically shown in Figure 3.

The framework can be separated into the three consecutive steps: *data understanding and preparation* (1), *model building and visualization* (2) and finally, *visualization and knowledge building* (3). The methods and considerations to be taken within these steps are described in the following.

3.1. Data understanding and preparation

Following the task of deriving knowledge on IPC at different observation points and their propagation throughout the process chain, a data set that describes the IPF at different observation points is needed. The first step is to become familiar with the underlying data set. During this phase, known as *data understanding*, methods of exploratory data analysis are widely utilized. For example, the data set is examined with regard to distributions (e.g. median, standard deviation, skewness), correlations, data quality (e.g. missing values, wrong values) and outliers. *Data understanding* provides an important basis for the selection of relevant features as well as for the creation of further meaningful features derived from the existing ones (feature engineering). In addition, the data set is pre-processed in a model-oriented manner during *data preparation*. An essential task of the proposed approach is the separation of the data set into observation-specific sub-sets. This ensures that the IPF at a given observation point is only generated based on the previous processing of the product. Furthermore, the features are to be standardized, especially when partitioning methods are to be used. Furthermore, depending on the machine learning method, only numerical features may be applicable.

3.2. Model building and visualization

The proposed approach pursues the deviation of IPC at different observation points n (see Figure 2). For this purpose, *model building* through machine learning has to be carried out. The focus of this paper is on unsupervised machine learning algorithms. The basic assumption here to identify different IPC is that similar states are characterized by a smaller distance from each other than dissimilar ones. With the help of the collected data per observation, the similarity of states can be quantified [8].

One main principle in unsupervised machine learning is clustering that can be divided in two main approaches. Hierarchical clustering is defined as an algorithm which builds clusters based on minimal distances between the individual observations. Within this category of clustering algorithms, two main perspectives which lead to the final hierarchical structure are differentiated. Agglomerative clustering represents the bottom-up approach and starts with creating clusters with exactly one object each. Based on the minimal distance, two clusters are then merged into a higher order cluster. Divisive clustering, on the other hand, is the top-down approach of hierarchical clustering. At the beginning of the algorithm, one big cluster comprising all observations is formed and then gradually divided into clusters of a lower order. [8,9]

Another common clustering methodology is the partitioning algorithm, which focuses on the number of observations within an area. Its main goal is to divide a set of data points into a predefined number of k clusters. This can be achieved using either centroids or medoids which are calculated by the mean of all data points or the median of all data points, respectively. The most widely used partitioning algorithm is the k-means algorithm. [8,9]

The proposed approach aims at identifying similar or characteristic classes of IPS within the IPF (see Figure 2, right side). Therefore, the number of classes is not known in advance. In addition, for each individual observation a separated model has to be built. For each specific model, the parameters are iteratively refined. Further, supervised machine learning approaches can for example be used to assign product states to already known IPC (e.g. good or defect product quality).

Model visualization (e.g. scatter-plot) supports in evaluating the model with respect to its meaningfulness for the use case. The iterative cycle of model building and model visualization can be supported and partly automated by analyzing the elbow curve (e.g. model parameter against the silhouette coefficient) of selected model parameters (see section 4.3). Deciding on which machine learning approach to choose strongly depends on the examined use case and its data structure and may not be the same method for each observation of the process chain. Moreover, visualization and evaluation of model results strongly depend on the business and data understanding.

3.3. Visualization and Knowledge Building

To complete the visual analytics cycle, process knowledge is to be gained from the observation-specific models obtained previously (*knowledge building*). Among other insights, a key statement should be made regarding the propagation of IPC along the process chain. As an appropriate *visualization* for displaying the propagation of IPC through the manufacturing system, a Sankey diagram can be used (see Figure 2). In manufacturing, Sankey diagrams are used to reduce complexity while showing important interactions within the system. Moreover, it is a comprehensive way to show quantitative and qualitative information within a customized visualization, showing different levels of detail.

On the one hand, this visualization can help to identify preferences regarding the propagation to certain classes. On the other hand, the width of the arrows indicates how likely it is for an IPC to result in a particular IPC in a subsequent observation (see Figure 2). This information can be used, for example, to identify IPC combinations that lead to particularly faulty parts. In this case, manufacturing control strategies can be derived. For example, these parts could be extracted early in the production chain to save the related costs, better (and probably cheaper) rework methods could be assigned due to the early finding of the defects or the following production process parameters might be adjusted to the special conditions of the parts.

Moreover, based on the propagation of IPC, class specific quality inspection strategies can be defined to reduce inspection effort (e.g. skipping of inspection).

4. Case Study

In order to analyze the behavior of the products along the manufacturing process chain, product state classes are formed within this case study on the basis of quality inspection data. With help of these classes, interactions may be analyzed and forecasts made with regard to future quality states.

4.1. Description of Case Study

The case study within this paper is covers the printed circuit board (PCB) assembly in electronic production. The PCB production line adopted for this case study focusses on surface mounted technology (SMT). Here, the individual components are placed on the surface of the PCB and soldered conductively. During the printing process, soldering paste is dispersed onto the surface of the PCB. This is followed by the solder paste inspection (SPI). To ensure a reliable connection of the components, a defined amount of solder paste is needed and variations in the solder paste volume have to be minimized.

The products are then inserted into the pick & place machine (P&P). The P&P places the individual components on the liquid solder in several sequential steps. After placing all components on the board, the PCB pass through a reflow oven in which the solder is melted and the connection pins are firmly soldered to the PCB. The automated optical inspection (AOI) checks the PCB after the reflow oven as a final examination at the end of line.

Observed defects are often a combination of effects from several processes. For instance, due to offset component placement during P&P and excessive soak times within the reflow oven, solder shorts can occur. Further, solder shorts can also result from too much solder volume disposition during printing [12].

Therefore, it is important to analyze the behavior from different product state classes over the whole process chain to understand failure states and derive specific strategies.

4.2. Data understanding and preparation

Within this case study, the focus is on quality inspection data from SPI and AOI. The SPI inspection results offer information about solder paste disposition on the PCB. By laser scanning technology, a 3D profile of the solder paste on the PCB is provided. Here, measured values for defined sample areas can be solder area, volume or height [12,13].

The AOI represents an image processing technology that is very often used in the assembly of PCB. It evaluates the quality of components on the PCB by machine vision recorded with a high speed and precision camera system, here, quantitative measurement results like angle, tilt, x- and y-direction. These measured values are used as IPF at the respective observations.

The respective dataset consists of quantitative inspection results from SPI and AOI (e.g. height or angle) as features, whereas each sample represents one inspection.

Within this paper, all existing numerical features are used and no feature selection or engineering is applied on the dataset. To achieve comparability of the data, all features are standardized by removing the mean and scaling to unit variance. In the context of this case study, extensive and complicated preprocessing was deliberately avoided in order to show which results could be achieved based on available data.

4.3. Cluster model building and visualization

The aim of this paper is to enhance the clustering of IPS to identify different IPC. Within this case study, the focus is on partitioning clustering algorithm. Here in particular, one promising approach is density based clustering. This type of clustering separates areas with a high density of data points from areas low in density. In a next step, areas with high density are assigned to a cluster, whereas areas with low density are excluded as so-called noise. In order to be able to make a clear distinction between the two categories, a threshold for the minimum density within a cluster has to be defined [15,14]. The most common density-based algorithm is DBSCAN (Density-Based Spatial Clustering of Applications with Noise) which is also the underlying algorithm applied in the case study.

DBSCAN has three major advantages.. Firstly, it can discover clusters of different shapes unlike other clustering algorithms like k-means. Furthermore, outliers are reliably detected and treated as noise points, hence are not being taken into consideration anymore. Another main benefit is that there is no need to define a number of clusters [14].

The inspection results of SPI and AOI are treated as two different observations. Hence, the observation specific data is clustered separately. To build an appropriate clustering model, an iterative parameter refinement is performed based on the shown framework in Figure 3. Hence, different clustering models are calculated with changing eps-values. Moreover, for each model the corresponding silhouette coefficient is calculated as an evaluation score.

The silhouette coefficient measures the distance between resulting clusters and shows how close data points in one cluster are to points in the neighboring clusters. The coefficient has a range between [-1, 1], where 1 indicates that the data point is far away from the neighboring clusters, 0 indicates that the sample is close to the decision boundary and -1 indicates that the data point is assigned to the wrong cluster [16].

Since the silhouette coefficient cannot be interpreted as an absolute recommendation to choose the eps-value, different clustering models need to be built to evaluate the results based on the business understanding (see Figure 3). The models with an eps-value around 3.5 show one huge cluster. Therefore, starting from the highest silhouette coefficient, different models are created and validated by the number of clusters and associated samples.

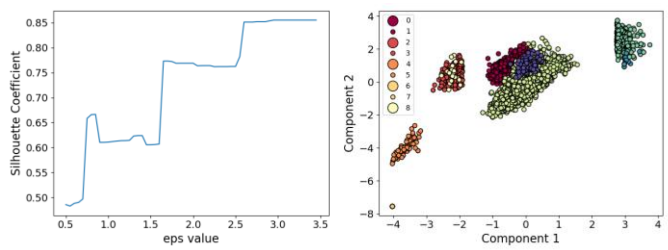

Figure 4: SPI plot of silhouette coefficient based on eps-values (left) and clustering results with eps-value of 0.7 (right)

For the SPI observation, an eps-value of 0.7 with a silhouette coefficient of about 68 % was chosen with a resulting clustering model of nine clusters (see Figure 4 right side). These clusters are interpreted as PSC. Overall, it can be seen that the clusters are clearly separated, even in dense areas. Only one area on the upper left shows some overlapping clusters. In this case, this can be explained by the two-dimensional representation.

In case of the AOI observation, the same procedure was chosen. Hence, the silhouette coefficients for different eps-values were calculated. Based on this, various models were created and checked on the basis of and with regard to the meaningfulness of the results in order to refine the parameter of the DBSCAN model. Figure 5 graphically shows the results for clustering the IPS.

On the left side, the eps-value with respect to the silhouette coefficient is plotted. The highest silhouette coefficients are reached with an eps-value over 3.3. However, cluster models with this parameter lead to huge clusters or not assigned data points. Therefore, the model parameters are refined stepwise by visualizing the model results with decreasing eps-values. Best results are achieved with the selected algorithm at an eps-value of 3.2 and a silhouette coefficient of about 57 %. The cluster results plotted with two principal components are shown in Figure 5 on the right side.

The model assigns one large cluster in the center, which corresponds to the high proportion of parts classified as good. However, it can be seen that there are different clusters located on *cluster 0*. Moreover, it can be stated that even outliers are assigned to *cluster 0*. Therefore, compared to the results of the SPI, it can be concluded that the results of the DBSCAN clustering on AOI inspection data are significantly more difficult to follow from the data and business understanding perspective.

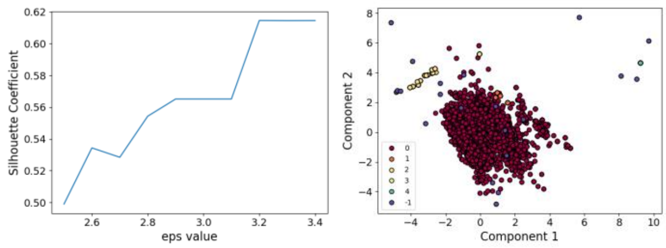

Figure 5: AOI plot of silhouette coefficient based on eps-values (left) and clustering results with eps-value of 3.2 (right)

4.4. Visualization and Analysis of Results

In order to describe the propagation and analyze interdependencies of product state classes within a manufacturing system, it is important to visualize existing IPC flows in the system. To gain even more insight into the characteristics of IPC within electronic manufacturing process chains, the propagation from the inspection stations SPI to AOI is shown by a Sankey diagram in Figure 6.

In the upper part of the figure, the respective processes and inspection stations are shown. In the lower part of the figure, a Sankey diagram is presented. The left side of the Sankey diagram shows existing IPC of SPI, based on the clustering in chapter 4.3. These classes are labeled with the prefix of the corresponding observation (e.g. SPI or AOI). This is followed on the right side with the classes of the AOI. The width of each arrow is proportional to its frequency. The classes *"-1"* represent outliers, meaning that the density-based clustering algorithm could not find enough close neighboring points to meet necessary criteria.

For example, the results show one large product class propagation from *SPI_5* to *AOI_0*. Moreover, it can be analyzed that all flows of SPI end up in *AOI_0* at least to some share. Here, further analysis is necessary to check why these classes differ from *SPI_5* if they finally end up in a conforming AOI PCB. However, based on the size of the classes, it can be assumed that these are potential "good" PCB.

In addition, it can be stated that the class *AOI_2* is only connected to *SPI_-1* and *SPI_8*. A clear connection and interaction between these classes can be identified. However, it needs to be analyzed why PCB in *SPI_-1* cannot be assigned to different clusters and whether, for example, *AOI_2* is a faulty class. In this case, products that are assigned to *SPI_-1* and *SPI_8* could be removed directly from the production line to avoid waste. As shown in Figure 6, these classes can be seen as potential "error prone" PCB.

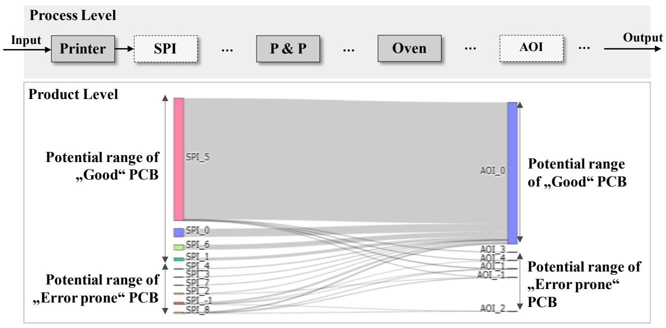

Figure 6: Resulting Sankey diagram showing the propagation of product state classes from SPI to AOI

5. Discussion and Outlook

The developed framework offers a structured approach to analyze the propagation of product states along the manufacturing process chain. Based on this, the behavior of different products within a manufacturing system can be tracked and product class specific control strategies can be derived. The approach is iterative with multiple visualizations in order to achieve valid model parameters.

The case study has demonstrated the applicability and use of the developed framework. However, the results have shown that extensive data preprocessing is necessary in order to achieve good clustering results. In particular, the results of the AOI have shown that the selection of the algorithm based on the available data should be case-specific.

Further work is necessary to apply the framework on a larger manufacturing system with several observations and a final quality inspection. On this basis, further investigations can be carried out with regard to the prediction of product properties based on the IPC and to analyze and quantify the behavior to derive meaningful manufacturing control strategies. In addition to the quantitative evaluation of the model performance, the IPC must be validated with regard to its usefulness and significance. Moreover, machine learning approaches have to be tested and validated to find optimal IPC combinations and reduce the effort for manual pre-processing of data.

Acknowledgements

This work was carried out within the "QU4LITY" project. The project has received funding from the European Union's Horizon 2020 research and innovation program with the grant agreement NO 825030.

References

[1] Gebauer, H., Gustafsson, A., Witell, L., 2011. Competitive advantage through service differentiation by manufacturing companies. Journal of Business Research 64 (12), 1270–1280.
[2] Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H., Noh, S.D., 2016. Smart manufacturing: Past research, present findings, and future directions. Int. J. of Precis. Eng. and Manuf.-Green Tech. 3 (1), 111–128.
[3] Thiede, S., 2018. Environmental Sustainability of Cyber Physical Production Systems. Procedia CIRP 69, 644–649.
[4] Wuest, T., Klein, D., Thoben, K.-D., 2011. State of steel products in industrial production processes. Procedia Engineering 10, 2220–2225.
[5] Wuest, T., Irgens, C., Thoben, K.-D., 2014. An approach to monitoring quality in manufacturing using supervised machine learning on product state data. J Intell Manuf 25 (5), 1167–1180.
[6] Wuest, T., Irgens, C., Thoben, K.-D., 2013. Analysis of Manufacturing Process Sequences, Using Machine Learning on Intermediate Product States (as Process Proxy Data), in: Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services. IFIP WG 5. 7 International Conference, APMS 2012, Rhodes, Greece, September 24-26, 2012, Revised Selected Papers, Part II, Berlin, Heidelberg. 2013. Springer Berlin Heidelberg, Berlin/Heidelberg, pp. 1–8.
[7] Wuest, T., Klein, D., Seifert, M., Thoben, K.-D., 2012. Method to describe interdependencies of state characteristics related to distortion. Mat.-wiss. u. Werkstofftech. 43 (1-2), 186–191.
[8] Thomas, J.J. (Ed.), 2005. Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Soc, Los Alamitos, Calif., 186 pp.
[9] Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., Andrienko, G., 2011. Solving Problems with Visual Analytics. Procedia Computer Science 7, 117–120.
[10] Cook, K., Earnshaw, R., Stasko, J., 2007. Discovering the unexpected. IEEE computer graphics and applications 27 (5), 15–19.
[11] Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H., 2008. Visual Analytics: Scope and Challenges, in: Simoff, S.J., Böhlen, M.H., Mazeika, A. (Eds.), Visual data mining. Theory, techniques and tools for visual analytics, vol. 4404. Springer, Berlin, pp. 76–90.
[12] Huang, J.C.Y., 2010. Reducing Solder Paste Inspection in Surface-Mount Assembly Through Mahalanobis–Taguchi Analysis. IEEE Trans. Electron. Packag. Manufact. 33 (4), 265–274.
[13] Wu, H., Zhang, X., Kuang, Y., Lu, S., 2008. A real-time machine vision system for solder paste inspection, in: The 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. AIM 2008 : July 2-5, 2008, Xi'an, China. 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Xian, China. 7/2/2008 - 7/5/2008. IEEE Xplore, [Piscataway, N.J.], pp. 205–210.
[14] Kotu, V., Deshpande, B. (Eds.), 2014. Predictive Analytics and Data Mining, 1st edition ed. Morgan Kaufmann, Boston, 446 pp.
[15] Cleve, J., Lämmel, U., 2016. Data Mining, 2nd ed. De Gruyter, Berlin, 328 pp.
[16] Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.