Concept of an Inference Procedure for Fault Detection in Production Planning

Jan Michael Spoor
Team Digital Factory Sindelfingen
Mercedes-Benz Group AG
Sindelfingen, Germany
jan_michael.spoor@mercedes-benz.com

Jens Weber
Faculty of Business and Economics
University of Applied Sciences Zwickau
Zwickau, Germany
jens.weber@fh-zwickau.de

Simon Hagemann
Team Digital Factory Sindelfingen
Mercedes-Benz Group AG
Sindelfingen, Germany
simon.hagemann@mercedes-benz.com

Frederik S. Bäumer
Faculty of Business
Bielefeld University of Applied Sciences
Bielefeld, Germany
frederik.baeumer@fh-bielefeld.de

***Abstract*—To date, no implemented solution in manufacturing, i.e., in automotive industry, exists to support production planning with insights from production. A structured feedback loop from operations to planning is required to further improve production planning. This contribution discusses the limitations of an existing concept for an inference procedure from operations to new planning tasks using the findings from previous implementation studies. Using the constraints found in these studies, six principles for inference procedures are derived. Thus, the existing concept is renewed and a structured and specific approach in providing an inference procedure for planning activities of similar manufacturing systems is proposed. This approach is split into the different sub-tasks of data acquisition, fault detection, knowledge representation, and knowledge inference. Each sub-task has its unique state-of-the-art solutions, challenges, and limitations which have to be examined during further implementations. Most notably, the concept requires a definition of a normal model to derive fault events and error patterns, an embedding of the fault events in an ontology to create a knowledge base, and the definition of a metric to measure similarity between the current configuration in operation and new configurations of the production planning.**

***Index Terms*—Production Planning, Fault Detection, Knowledge Engineering, Data Mining, Case-Based Reasoning**

I. Introduction

The automation of production has a long tradition in the automotive industry. This industry has therefore been leading in the field of automation. Since the first initiatives of Henry Ford in the early 20th century, the assembly stages have especially been in the focus of automation [1]. Assembly processes require different levels of flexibility and in turn have led to different automation levels. The final assembly in automotive industry is still characterized by a high proportion of manual processes, while the Body-In-White (BIW) assembly is nowadays considered fully automated. The major step towards automation of production has been achieved as part of the third industrial revolution and is based on the use of computers, robotics, and electronics.

However, the ongoing fourth industrial revolution, which in Germany is considered as Industry 4.0, is based on increasing system connectivity and therefore the use of Cyber-Physical Systems (CPS) [2]. Cyber-Physical Production Systems are defined as "systems that integrate computation and physical processes [...]". The use of CPS continuously generates large amounts of sensor data. In automotive factories several terabytes of raw data are collected on a daily basis. However, the German high-tech strategy Industry 4.0 (I4.0) comprises more than the application of CPS. I4.0 targets the data continuity and autonomous orchestration of processes along the whole product-creation process [3].

In this contribution, we focus on a concept which assists the early phases of production planning in which the configuration of new assembly systems takes place. Therefore, real production data of operations is supposed to be a basis of the assistance system and requires the data continuity from the early phases of production configuration to the operational production. We conducted interviews with production planners in multiple European automotive Original Equipment Manufacturers (OEM), suppliers, and research institutes which have shown that there is no digital feedback loop from production back to the production planning.

In general, the computer-assisted assembly system configuration has for long been a discipline which has not been sufficiently regarded by research and industry [4]. Recently, research projects have been tackling the automation or at least the assistance of the early planning phases. Hagemann and Stark [5] provide an algorithmic approach in which the configuration is processed fully automated. The approach uses combinatorial optimization algorithms and determines the best production system configuration with the aim of minimizing investment costs. Other authors, such as Michalos et al. [6] and Michels et al. [7], published automated approaches for the design of assembly lines. However, to the best of the authors' knowledge, there are as of today no implemented solutions which aim at assisting the production planner based on the usage of real production data.

The following paragraphs describe a novel concept tackling this research gap. This novel concept introduces four sequential steps in the inference procedure: data acquisition, fault detection, knowledge representation, and knowledge inference. Through this approach, a comprehensive analysis and feedback of faults from operations to production planning is enabled. The focus of our contribution is on the automotive industry, but the concept is discussed in a universal manner, enabling the use in manufacturing in general.

In section II, a preceding concept for an inference procedure is presented, and its limitations found by our conducted implementation studies are discussed and analyzed. Thereupon in section III, the determined limitations are used to derive relevant principles, which must be taken into account when developing or implementing future concepts of inference procedures for fault detection. Subsequently, these principles are further aggregated to derive a mathematical problem description. Concluding in section IV, applying the problem description and principles, a novel concept for inference procedures using the four sequential steps is introduced. Furthermore, the state-of-the-art solutions and methods for each step of the introduced concept are discussed by conducting a comprehensive review of the methods in the current literature.

II. Derivations from the Preceding Concept

This contribution enhances the first concept by Gelwer et al. [8] and attempts to improve the approach by addressing limitations along the stages of the concept.

A. Description of Preceding Concept

The concept by Gelwer et al. [8] is a specific model to target a real manufacturing application at Mercedes-Benz Group AG (former Mercedes-Benz AG). The concepts sketch is given in Figure 1. The approach by Gelwer et al. [8] is summarized by following three stages:

(i) Identification of faults in the manufacturing production system during operations, the knowledge creation.
(ii) Set-up of a knowledge base.
(iii) Feedback of faults by identifying similar manufacturing systems in planning, an inference process applied in production planning procedures.

Stage (i) uses two approaches to identify faults. Firstly, an anomaly detection is conducted using data from Internet of Things (IoT) devices provided by a Manufacturing Service Bus (MSB), see Minguez [10], i.e., real-time data from devices, processes, and conditions. Secondly, Natural Language Processing (NLP) is applied for analyzing the error documentation. Input of the NLP can be plain text for a classification of errors, e.g., documented in an Enterprise Resource Planning (ERP) system or other third party sources. Described faults within shift logs should then be classified using standardized error codes. If both error detection and classification approaches work successfully, the results correspond, since documented errors by maintenance workers should also

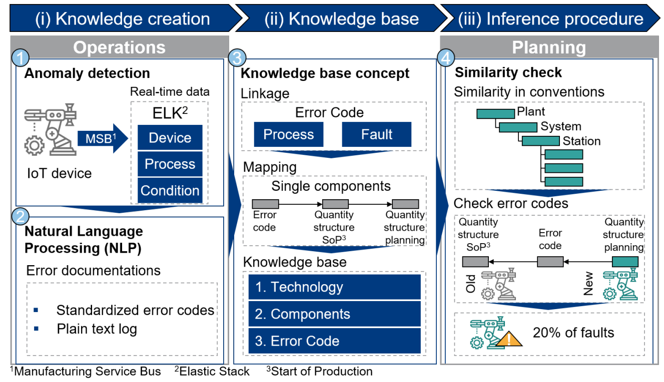

Fig. 1. Concept for data consistency checks between operation and production planning enabling an improved knowledge of past errors in planning by Gelwer et al. [8].

be visible in the data, and *vice versa* the detected anomalies should appear in the necessary documentation of the shift log.

Stage (ii) is the set-up of the knowledge base by linking the corresponding technical description of the occurred faults and the affected processes within the error codes. These linked error codes are mapped with the hierarchical quantity structure of the manufacturing system after start of production and also in the stages of production planning, i.e., within the used library on the single component level. Further contextual information about the errors are stored, i.e., used technologies and parts numbers. The error and the additional contextual information are documented within the knowledge base.

Using the knowledge model, stage (iii) conducts an inference procedure in the case of a production planning process of a new manufacturing system. The new defined quantity structure in production planning is compared to the documented faults occurred in similar quantity structures after start of production within the library of stage (ii). Using the documented error code within the context of the quantity structure enables enriched information about possible problems or faults with the suggested component of the new planned manufacturing system. The proposed comparison using documented anomalies and faults of the past should be applied to the part and component level.

The concept by Gelwer et al. [8] was tested after its publication and further evaluated within the organizational structure of the Mercedes-Benz Group AG. By this practical application multiple limitations were detected, resulting in the necessity of a renewal of the approach. The limitations are ordered by corresponding stages not by importance.

B. Limitations of the Knowledge Creation

In stage (i), it is concluded that no so-called jack-of-all-trades algorithm or method for a consistent anomaly detection exists. This is not surprising since the free-lunch theorem implies that, considering all possible data, different anomalies, and targets of the detection process itself, no single algorithm is expected to solve all tasks [11]. This problem is directly relevant for the Mercedes-Benz Group AG and assumed for most organizations within manufacturing and automotive industry, as a wide variety of data types are available and in use. Using the proposed framework by Foorthuis [12], out of the 9 types with 63 subtypes of anomalies, 38 different subtypes from all 9 types of anomalies are expected within the data of the Mercedes-Benz Group AG. Although theoretically feasible, within an efficient organization or a framework of a business case individual implementations of all solutions per subtype are difficult to achieve. During the tests performed, the used algorithms, i.e., Isolation Forest, Multi-Layer-Perception, and K-Means, heavily relied on well-labeled data, test datasets, or required an extensive amount of prior investigation for setting up valid parameters. Thus, more generic approaches need to be defined, lessening the requirements for anomaly detection.

While the amount of data accessible is large enough, the error states are only occasionally and not consistently labeled. Furthermore, errors are quite rare. We estimate more than an additional decade of runtime using same configurations, as comparability is necessary, for creating sufficient error instances at the facilities of Mercedes-Benz Group AG. Many data of so called normal states [13] of the manufacturing system model exist, but real incidents of errors are rarely found in the available dataset or are often not documented enough in a consistent manner to draw structured conclusions.

Furthermore, the requirement to use real-time streaming data should be dropped since the proposed usage of real-time streaming data is technically complicated to implement [14]. More importantly, for a planning procedure with prior analysis of past data, real-time information processing is not necessary since no acute, short-term, and quick call for action is given.

The problem of stage (i) is exacerbated in the area of NLP. Only a limited amount of shift log entries exist, and from these only a limited number are related to specific errors. It is assumed that a higher training dataset size is highly beneficial for increasing the accuracy of NLP [15]. Furthermore, the documentation of manufacturing workers filling the shift books often lacks the required details in delimitation of the different types of faults or error codes. The insufficient documentation can be attributed to implicit knowledge of the workers, which is not known to the NLP algorithm. One current shortcoming of NLP and the development of artificial intelligence in general is the inclusion of implicit knowledge and human 'common sense' [16]. Therefore, shift logs could be used to determine if an error occurred but not what error occurred. Also, this still requires a larger amount of shift logs since the current entries are still too few to perform analysis.

To summarize, in stage (i) the classification of errors is a challenge of the concept and is not solvable with the current state-of-the-art tools proposed by Gelwer et al. [8].

C. Limitations of the Knowledge Base

While the linkage of the error code to the process is an important step in setting up and understanding the context of error messages, it is often not sufficient for later inferences. This procedure might correctly identify critical combinations of components and processes within the planning process, which lead to the described error states, but offers no sufficient information about the cause of the error that occurred and does not enable countermeasures except to dispense with the combination of component and process. Since faults are often foreshadowed by certain patterns and comparable faults can occur in different processes, linking these patterns might help to identify the specific error more precisely. This linkage enables a comparison it with similar faults, a comparison of solutions for these similar faults, and in conclusion enables targeted countermeasures. Therefore, the context of usage might also be an important factor in comparing the error with other occurring faults and their corresponding solutions. This enables a more detailed measure of criticality of the error and guides a decision on how to handle the error, if cost efficient, instead of avoiding it.

In addition, if the patterns are transferred and reused in stage (i), this additional context becomes an important part of the error classification and important to document within the error messages. Underlying faults, i.e., currently researched unwanted cold welding processes in holding pins, might be detectable by an overall pattern in the data not only by single faults and error messages. The feedback and usage of fault patterns is a useful addition to the knowledge base.

A helpful approach in stage (ii) is to identify affected components within their position in start of production as well as in the production planning libraries. The differentiation between start of production and planning might often be important since position, usage, and linked processes are changing during the production planning process in a manner that renders the reasoning behind the choice unclear. Nevertheless, using only the structural context of a resource within the quantity structure offers little information about the component and its use. Important contextual information is not documented within the quantity structure during production planning and start of production. A component might cause comparable errors within different quantity structures and contextual information about technologies, parts, usage, processes, and products might offer more explanatory value in describing errors.

D. Limitations of the Inference Process

This missing context within the quantity structure is even more important in setting up a similarity check in the inference process. The quantity structure itself, even if tracked within start of production and production planning, is not enough to detect similar set-ups. Very different quantity structures share comparable faults, and solving the faults in these different quantity structures might offer very important insights and enable solutions. While the quantity structure is certainly a part of the similarity measure, it must be enriched with more context. Similar quantity structures might behave very differently, and *vice versa* different quantity structures might be more comparable regarding documentation and detection of faults. Therefore, a fleshed-out ontology is needed to provide additional information about types, linkages, relations, and the interaction of product, process, and resources planned and deployed in this structure.

Furthermore, it is unclear how the similarity measure is set up since the quantity structure alone offers too little information. Even if the quantity structure is quite similar, it is difficult and unclear how to transform this similarity into a quantified measurement. Therefore, the proposed ontology must also offer the possibility to apply a quantifiable similarity measure. Based on the quantified measure, more similar set-ups and their respective faults should be given more weight when the planner is informed of potential errors by the inference process. Faults, risks, and solutions should be weighted by similarity. Therefore, the similarity, based on a metric quantifying the distance between the past and new planned configuration, becomes a measure on how likely a similar fault, which occurred prior in the compared past configuration, will occur in the new configuration. The predicted error-proneness of the new configuration is assumed to be correlated to the distance measure between the new and past configuration and the prior measured error-proneness of the past configuration. This metric needs to be developed and embedded within the proposed ontology.

III. Derived Requirements and Problem Definition

Considering the discussed limitations, relevant principles for future concepts can be derived and a mathematical problem formulation can be set up.

A. Requirements for Future Concepts

The relevant findings from the discussion of the preceding concept can be expressed by the following six principles:

1) Since faults are rare in the data, an approach using labeled faults requires more labeled training data for a valid classification of errors than currently available. As an alternative, a normal model needs to be defined, and all data deviating from the normal model should be classified as generic faults. The use of only supervised approaches is not recommended.
2) Since shift logs contain no information about the exact errors but can be used to identify if any error occurred, they enable spotting of time frames of interest for finding error patterns. Not all data are analyzed but data occurring during days with entries in the shift logs are.
3) Using the deviations from the normal data, these findings can then be compared regarding their unique patterns and segmented for building a new fault classification structure. The classified patterns are then the classification criteria for all anomalies.
4) As these fault patterns might be highly individual for each configuration, the configurations need to be described in a more meaningful way. A simple description within the quantity structure of production planning is not sufficient. Each configuration must be enriched with contextual data which then enables a deeper contextual anomaly detection and a real causality analysis.
5) Furthermore, because configurations are solely dependent on their quantity structure, an additional ontology must be created to make configurations more specific and comparable beyond the quantity structure.
6) Using this ontology, a metric must be developed, capable of comparing the similarity of configurations independently of their hierarchical position within the quantity structure of production planning and start of production.

If the proposed principles are considered, a risk assessment of a new planned configuration can be conducted by a comparison with the past configurations. By applying a similarity score based on a metric using ontologies and combining this information with the risk of a fault event, the risk of the observed new configuration in the production planning process is derived. This problem description and the resulting approach is related to case-based reasoning [9].

B. Mathematical Problem Formulation

To address the requirements discussed, we build a fundamental logic on how to feed errors back.

First, each resource is assumed to have a certain and known configuration θ_k out of a finite set of all possible configurations for these resource types. The configuration depends strongly on the Products, Processes, and Resource (PPR) model.

$$\theta_k \in \Theta = \{\theta_1, ..., \theta_K\} \tag{1}$$

Each resource and its specific configuration k have a finite set of possible faults or error states. Each error state j is defined as follows:

$$e_j \in E_k = \{e_1, ..., e_J\} \tag{2}$$

For each error j in configuration k there exists a certain probability $r_{j,k}$ that the error occurs.

$$r_{j,k} = P(e_j \mid \theta_k) \tag{3}$$

The risk of any error occurring in configuration k is then given as following expression:

$$r_k = \sum_{e_j \in E_k} P(e_j \mid \theta_k) \tag{4}$$

If each risk does not contribute equally to the perceived economic risk, a weight w_j of each error can be applied.

If additionally a second configuration k^* exists, there exists an amount of errors which are both present in the configuration k and k^*.

$$E_{k \cap k^*} \in E_k \cap E_{k^*} \neq \emptyset \tag{5}$$

The risk of any error occurring in configuration k is analogously given by following expression:

$$r_{k^*} = \sum_{e_j \in E_{k \cap k^*}} P(e_j \mid \theta_{k^*}) + \sum_{e_j \notin E_{k \cap k^*}} P(e_j \mid \theta_{k^*}) \tag{6}$$

If configuration k^* is not within operation and currently just a configuration during production planning, no estimation of $P(e_j \mid \theta_{k^*})$ can be conducted. But if configuration k^* and k are similar enough, it is assumed that the set of errors within

 ISBN: 978-1-61208-953-9

configuration k^* but not within configuration k is very small. The occurrence of completely new errors is unlikely.

$$\sum_{e_j \notin E_{k \cap k^*}} P(e_j \mid \theta_{k^*}) \approx 0 \tag{7}$$

To conduct a valid inference procedure, a metric defining a distance measure $\Delta(\theta_k, \theta_{k^*})$ is necessary to compare k and k^*. This metric should then give an approximation of the possible error states using the configuration k as base.

$$r_{k^*} \approx \sum_{e_j \in E_{k \cap k^*}} P(e_j \mid \theta_{k^*}) \approx \sum_{e_j \in E_k} P(e_j \mid \Delta(\theta_k, \theta_{k^*}), \theta_k) \tag{8}$$

For each error, a relation between configuration k and k^* dependent on the distance measure between them is assumed.

$$P(e_j \mid \Delta(\theta_k, \theta_{k^*}), \theta_k) \sim P(e_j \mid \theta_k) \circ \Delta(\theta_k, \theta_{k^*}) \tag{9}$$

If this relation is measurable by, e.g., a correlation analysis, the inference is then a useful risk measure for the error-proneness of configuration k^*. Therefore, in order to conduct a risk assessment of a new configuration k^*, the following challenges need to be addressed:

1) The risk assessment of base configuration k is necessary.
2) There needs to be a valid definition of a metric $\Delta(\theta_k, \theta_{k^*})$.
3) Using the metric and risk assessment of k, a risk assessment of k^* must be derived.

IV. Proposed Concept of Inference Procedure

Based on the existing concept and the derived principles and mathematical problem formulation, a new concept is developed and presented in this section. This concept is then discussed along its proposed steps.

A. Concept Overview

First, the concept overview and core ideas are presented. Since the proposed concept needs to include additional information about ontologies, a relevant comparable fault diagnostics method is proposed by Zhou et al. [17]. The structural set-up of the proposed concept by Zhou et al. [17] is included into the proposal by Gelwer et al. [8] and adapted to meet all defined principles. Our proposed concept for an inference process is sketched in Figure 2.

The concept is split into four constructive steps and three input models:

The (1) data acquisition describes the process of collecting, processing, storing, and providing the data in order to conduct a fault detection.

The (2) fault detection accesses the normal model to use it as base for a detection if any kind of event happened and to describe the event patterns. The normal model utilizes the insights of principle 1) since more normal data are available and a normal model can be set up using training datasets. Also, this utilizes principle 2) by first detecting if any event happened before classifying or describing the event.

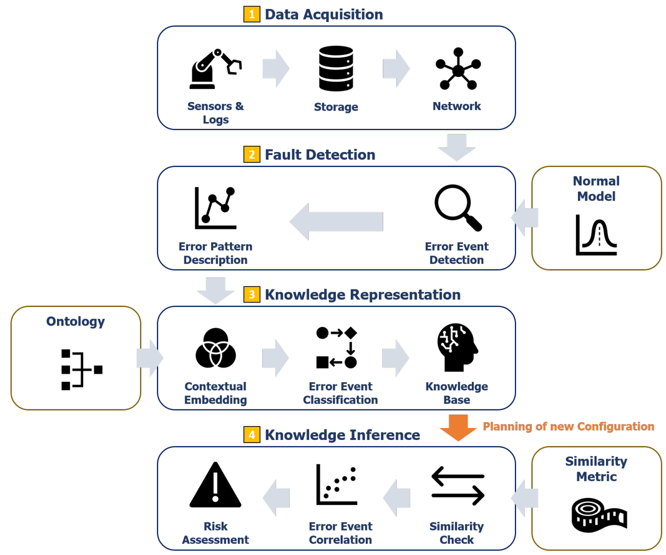

Fig. 2. Proposed concept for an inference process based on knowledge representation. The model is enabled by four constructive steps and three inputs: An ontology of the configurations, a similarity metric, and a normal model of the data.

In the (3) knowledge representation, the pattern and error events are then embedded in the ontology of the configuration, in which the event took place, and used for an event classification. A notable ontology in a similar use case is defined by Ming et al. [18] using, adapted to the discussed use case, components' taxonomy, properties, and relationships regarding features, operation, and quantity structure. This kind of ontology fulfills principle 5). If the event is classified by comparing the error pattern with similar events, as requested by principle 3), the knowledge base represents specific types of error events. These events are documented by well-defined patterns and are occurring within delimited areas and applications according to the ontology, as principle 4) suggests. For each error, the probability of occurrence $r_{j,k}$ can be determined by predictive pattern mining of the specific error event.

The (4) knowledge inference is then conducted when a new configuration is planned. The new planned configuration is compared to the configuration in the knowledge base by applying the defined metric as per principle 6). This similarity is the expression of the term $\Delta(\theta_k, \theta_{k^*})$ and is then used in the error correlation. The correlation process itself is represented by equation (9). This correlation is then used to calculate a risk assessment r_{k^*}.

The constructive steps and their challenges, current state-of-the-art solutions, requirements, and derived research questions are discussed in more detail in the following subsections.

B. Data Acquisition

The data acquisition is similar to the proposal by Gelwer et al. [8] and also proposes the usage of a MSB as described by Minguez [10]. The MSB uses a multitude of interfaces, which need an implementation within the overall IT infrastructure of a manufacturing company. The MSB acts therefore as a universal communication layer enabling the integration of all data from the shop floor for our proposed method [19]. The MSB solves the challenge of integrating a multitude of IoT data [1] and can be, as proposed by Gelwer et al. [8], transferred by Message Queuing Telemetry Transport (MQTT) protocols in JavaScript Object Notation (JSON) using process parameters and information from the Programmable Logic Controller (PLC) combined with informations from the ERP.

An additional part of the data acquisition is the provision of shift logs as possibility to detect error events. The application of an NLP is then necessary to detect if error events happened and mark these dates within the data for the error event detection research. The marked dates are primarily important within the error event detection of the fault detection step.

C. Fault Detection

While it is assumed that in most cases a supervised anomaly detection might offer more insights due to the incorporation of application-specific knowledge, the rare occurrence of faults within the application in production planning makes it difficult to create robust and generalized methods in this case [13].

Therefore, the fault detection process becomes for supervised methods a case of one-class variation since mostly normal data are available. A possible method would be the application of a one-class Support Vector Machine (SVM) as described by Schölkopf et al. [20]. Alternatively, unsupervised approaches can be used but risk the falsely positive detection of noise in the data as faults [13].

One challenge is the false inclusion of anomalies, despite being rare, into the normal data. An anomaly could easily be misclassified as normal. These rare instances could result in a more sensitive one-class SVM. A robust method for fault detection using one-class SVM is given by Yin et al. [21].

Another possible method to apply in these cases are Kernel Principal Component Analysis (PCA) for novelty detection. Kernel PCAs map the mostly normal instances containing training dataset into a feature space. The squared distance to the corresponding principal subspace is then a measure for anomalous data [22].

Different from an approach using one-class SVM or Kernel PCA methods, would be the full modeling of the normal machinery behavior to deduct anomalies by comparing the delta between prediction and measurement. These methods are more complex since they require a set-up of a complete normal model of the planned system but are very robust and rely on simpler distance-based comparisons between predictions and measurements. The quality assessment of such models is not the novelty detection method but the quality of the model itself. These models are achieved by a comprehensive Digital Twin defined as a simulation using physical models [23].

A possible model, which does not require a full Digital Twin, is an autoregressive time series AR(p) of an order p comparing the measurements with distance-based metrics, i.e., the Mahalanobis distance [24].

For rare instances of anomalies, the model might not be necessary to set up, but the state signal itself is the base of comparison, i.e., by comparing different windows of a signal. The cross correlation entropy between two windows can be measured and windows containing anomalies will result in a larger entropy of cross-correlation [25].

Further research is necessary to determine the best model or approach for the fault event detection since no clear recommendation for one specific approach is currently possible.

For a pattern identification of fault events, the rarity of faults must be taken into account. The pattern identification task becomes a problem of infrequent patterns. Therefore, descriptive tasks should be used to identify comprehensible patterns which are later labeled in the knowledge representation step.

Since it is expected that faults occur in very specific scenarios and are foreshadowed by co-occurrences prior or after the fault event, these co-occurrences can be used to describe the pattern and delimit it from other faults. Therefore, methods of descriptive association rule mining might be most useful in the pattern identification [26].

D. Knowledge Representation

Important input for the set up of the knowledge representation is the prior set-up of an ontology.

An ontology is defined as the model representing the semantics of the domain model. A knowledge graph is the result if data instances are acquired, integrated into an ontology, and additionally a reasoning is applied to derive new knowledge [27]. This differentiates an ontology and the resulting knowledge graph. The ontology itself offers little specific insight on the domain, but the domain becomes relevant when applied in the knowledge graph [28]. Therefore, the ontology is the input model and the application and contextual embedding of the detected faults integrated into the ontology becomes a knowledge graph which acts as the knowledge base for the following inference procedure.

The main challenge of a valid knowledge representation is the definition of a useful ontology, since for the presented application two obstacles are limiting the usage of currently available semantic model-based ontologies in Industry 4.0 applications as described by Yahya et al. [28]:

1) Production models do not fully follow the linked data principles and require a new vocabulary instead of the re-usage of current used vocabularies.
2) The scope of currently used ontologies is too application-specific and not applicable in all areas of the production.

A notable contribution in defining a possibly relevant ontology for contextual Industry 4.0. systems is given by Giustozzi et al. [29]. Besides the already defined necessary relationships (see subsection A), the ontology by Giustozzi et al. [29] uses dedicated resource, situation, process, time, location, and sensor ontologies.

Most relevant for production planning in the automotive industry are the domains of Product, Process, and Resources, bundled in the PPR concept [30]. A detailed exemplary set-up of the ontology of the PPR concept is given by Agyapong-Kodua et al. [31]. An applicable ontology for the proposed concept must therefore combine the aspects of the PPR concept as well as the aspects of ontologies for a contextual anomaly detection. Most beneficial would be a smooth integration into the existing PPR ontology models which are currently in use by manufactures and automotive.

The usage of more refined ontologies is enabled by the progressive efforts of companies in the holistic implementation of a Digital Twin, defined as a comprehensive physical but also functional description of components, products, and systems which enables insights in later lifecycle phases [32]. This also requires a Digital Twin definition as the sum of logically related data represented by semantic data models [33]. While the call for a more refined ontology seems to be a difficult requirement at first glance, it might be solved parenthetically due to the set-up of Digital Twins.

In the contextual embedding step of the detected faults, only the linked and semantic description of the faults are capable of setting up contextual error identifications, thus enabling a contextual error classification. By classifying the errors in the knowledge representation step, this enables a contextual anomaly detection using the error patterns and the domain knowledge. A notable exemplary method for sensor data is given by Hayes and Capretz [34] using cluster analysis to define sensor profiles and enabling a contextual analysis within the profiles. This application could be relevant for the discussed measurements of streaming sensor data in the data acquisition step. Also, the contextual anomaly detection might enable the previously unsuccessful NLP of shift logs for error classification. An exemplary application is described by Mahapatra et al. [35].

After the fault is described within the ontology and the fault patterns are delimited, the fault j is uniquely classified and combined with the corresponding pattern description which is the error event classification step. After classification, the past data are searched for the fault patterns. The found instances of faults of the same type within the specific configuration k are then counted. This pattern mining enables a calculation of the error-proneness probability $r_{j,k}$.

This task can be conducted by applying different predictive pattern mining algorithms [26]. Which algorithm for the prediction and classification of faults performs most usefully in the described use case must be further researched.

The error-proneness probability and error pattern embedded in the ontological description of the configuration build up the knowledge base.

E. Knowledge Inference

Since the concept is set within the idea of the digital factory, the product development and the production planning are parallelized. Therefore, the concept should also support a rough planning as the first step of the production planning process. Furthermore, in the applied planning within manufacturing companies, focus of the rough planning is often more on resources than processes since resources are main part of the cost calculation [36]. Within the rough planning a quantity structure and a 3D layout based on the used library containing the single components and parts of the production system is set up. Only in the detailed planning the supplier is involved, optimizing the quantity structure up until the start of production [37]. The higher the degree of maturity, the more information and usefulness a rough production planning provides [38].

Therefore, the similarity measure can only be as good as the rough production planning. Since more information is added during planning, more ontology types are also added and then enable better similarity measures. Conversely, this also means that the ontology must be imposed to planners, suppliers, and operation since a documented ontology in the production planning and start of production state benefits the significance of the analysis.

If the ontology is documented diligently, the main challenge is the set-up of a useful metric. Already in the definition of the metric a contrary objective arises: the metric must be set up in a way to properly describe the error-proneness of planned configurations k^* based on current configurations k, but the error-proneness of the planned configurations is itself derived from the distance measure of the metric. This conflict of goals makes an objective definition difficult. It is assumed that the metric to be defined is more likely a fuzzy similarity assignment, i.e., a probability that the configurations are similar, than a hard assignment listing the most similar configurations. This is the case, since similarity is often context-specific and different from equality measured in degrees [9]. Even then, a fuzzy assignment still needs to be quantified and must be tenable even under a generous error interval in real-world applications considering domain knowledge of the planners. The set-up of a metric is one of the current challenges and open research questions of the concept.

If a valid metric is defined, the correlation analysis following equation (9) is conducted to calculate a risk assignment of the new configuration k^* in production planning. Commonly-used tools in case-based reasoning include, e.g., regressions, bayesian learning, and Artificial Neural Networks, which might also be applicable in the presented use case [9].

V. Conclusion

Though purposefully improving the former concept proposed by Gelwer et al. [8] through the six principles established, our new proposed concept still has open challenges and needs further efforts to address these issues, namely the set-up of a valid ontology within the manufacturing system description and the derivation of a useful metric to determine similarity between configurations. Further challenges are the selection of a useful fault detection method and the set-up of a use case oriented pattern mining.

Nevertheless, within this contribution we were able to determine requirements of inference procedures and make a targeted proposal for future research in this area. In particular, the six defined principles and the proposed mathematical correlation definition between the error-proneness of planned configurations k^* based on current configurations k contribute to the current efforts towards building an inference procedure.

 ISBN: 978-1-61208-953-9

Furthermore, our proposed concept acknowledges the shortcomings of the former concept and proposes an advanced structure. Our proposal uses the stages of data acquisition, fault detection, knowledge representation, and knowledge inference. These stages are enabled by the definition of a normal model as a basis for fault detection, an ontology for a valid representation, and a similarity metric in order to be able to carry out target-oriented comparisons.

The authors plan to examine the proposed concept in more detail and implement use-case oriented applications of the concept in production planning in future studies.

ACKNOWLEDGMENT

We thank all colleagues and former colleagues who provided data, concepts and development support of the research work within Mercedes-Benz Group AG and the University of Applied Sciences Zwickau.

The Mercedes-Benz Group AG funds this research.

REFERENCES

[1] T. Bauernhansl, M. Hompel, and B. Vogel-Heuser (eds), Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung, Technologien, Migration. SpringerLink, Springer Vieweg, Wiesbaden, 2014.
[2] Plattform Industrie 4.0, Plattform Industrie 4.0 - Digital Transformation "Made in Germany". 2019.
[3] acatech - Deutsche Akademie der Technikwissenschaften, Key themes of Industrie 4.0. 2019.
[4] S. Hagemann and R. Stark, "Automated Body-in-White Production System Design: Data-Based Generation of Production System Configurations," In: Proceedings of the 4th International Conference on Frontiers of Educational Technologies, ACM, New York NY, pp. 192-196, 2018.
[5] S. Hagemann and R. Stark, "An optimal algorithm for the robotic assembly system design problem: An industrial case study," In: CIRP Journal of Manufacturing Science and Technology, vol. 31, pp. 500-513, 2020.
[6] G. Michalos, A. Fysikopoulos, S. Makris, D. Mourtzis, and G. Chryssolouris, "Multi criteria assembly line design and configuration - An automotive case study," In: CIRP Journal of Manufacturing Science and Technology, vol. 9, pp. 69-87, 2015.
[7] A. S. Michels, T. C. Lopes, C. G. Sikora, and L. Magatão, "The Robotic Assembly Line Design (RALD) Problem: Model and case studies with practical extensions," In: Computers and Industrial Engineering, vol. 120, pp. 320-333, 2018.
[8] E. Gelwer, J. Weber, and F. Bäumer, "A Concept of Enabling Data Consistency Checks Between Production and Production Planning Using AI," In: Proceedings of the 17th International Conference on Applied Computing, pp. 139-142, 2020.
[9] M. M. Richter and R. O. Weber, Case-Based Reasoning. Springer, Berlin, Heidelberg, 2013.
[10] J. Minguez, "Der Manufacturing Service Bus," In: E. Westkämper, D. Spath, C. Constantinescu, and J. Lentes (eds), Digitale Produktion. Springer, Berlin, Heidelberg, 2013.
[11] Y. Ho and D. Pepyne, "Simple Explanation of the No-Free-Lunch Theorem and Its Implications," In: Journal of Optimization Theory and Applications, vol. 115, pp. 549-570, 2002.
[12] R. Foorthuis, "On the Nature and Types of Anomalies: A Review of Deviations in Data," In: Int J Data Sci Anal, vol. 12, pp. 297-331, 2021.
[13] C. C. Aggarwal, Outlier Analysis. Springer Science+Business Media, New York, 2013.
[14] A. Lavin and S. Ahmad, "Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark," In: IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 38-44, 2015.
[15] M. Banko and E. Brill, "Mitigating the Paucity-of-Data Problem: Exploring the Effect of Training Corpus Size on Classifier Performance for Natural Language Processing," In: Proceedings of the first international conference on Human language technology research, pp. 1-5, 2001.
[16] N. J. Nilsson, The Quest for Artificial Intelligence. Cambridge University Press, 2009.
[17] Q. Zhou, P. Yan, H. Liu, and Y. Xin, "A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis," In: J Intell Manuf, vol. 30, pp. 1693-1715, 2019.
[18] Z. Ming, C. Zeng, G. Wang, J. Hao, and Y. Yan, "Ontology-based module selection in the design of reconfigurable machine tools," In: J Intell Manuf, vol. 31, pp. 301-317, 2020.
[19] D. Schel, C. Henkel, D. Stock, O. Meyer, G. Rauhöft, P. Einberger, M. Stöhr, M. A. Daxer, and J. Seidelmann, "Manufacturing Service Bus: An Implementation," In: Procedia CIRP, vol. 67, pp. 179-184, 2018.
[20] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating Support of a High-Dimensional Distribution," In: Neural Comput, vol. 13, no. 7, pp. 1443-1471, 2001.
[21] S. Yin, X. Zhu, and C. Jing, "Fault detection based on a robust one class support vector machine," In: Neurocomputing, vol. 145, pp. 263-268, 2014.
[22] H. Hoffmann, "Kernel PCA for novelty detection," In: Pattern Recogn, vol. 40, no. 3, pp. 863-874, 2007.
[23] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," In: Int J Adv Manuf Technol, vol. 94, pp. 3563-3576, 2018.
[24] M. Hau and H. Tong, "A practical method for outlier detection in autoregressive time series modelling," In: Stochastic Hydrol Hydraul, vol. 3, pp. 241-260, 1989.
[25] T. Wang, W. Cheng, J. Li, W. Wen, and H. Wang, "Anomaly detection for equipment condition via cross-correlation approximate entropy," In: MSIE 2011, pp. 52-55, 2011.
[26] S. Ventura and J. M. Luna, Supervised Descriptive Pattern Mining. Springer, Cham, 2018.
[27] L. Ehrlinger and W. Wöß, "Towards a Definition of Knowledge Graphs," In: SEMANTiCS, vol. 48, pp. 1-4, 2016.
[28] M. Yahya, J. G. Breslin, and M. I. Ali, "Semantic Web and Knowledge Graphs for Industry 4.0," In: Appl. Sci. 2021, vol.11, article 5110, 2021.
[29] F. Giustozzi, J. Saunier, and C. Zanni-Merk, "Context Modeling for Industry 4.0: an Ontology-Based Proposal," In: Procedia Computer Science, vol. 126, pp. 675-684, 2018.
[30] R. B. Ferrer, B. Achmad, D. Vera, A. Lobov, R. Harrison, and J. L. Martínez Lastra, "Product, process and resource model coupling for knowledge-driven assembly automation," In: Automatisierungstechnik, vol. 64, no. 3, pp. 231-243, 2016.
[31] K. Agyapong-Kodua, C. Haraszkó, and I. Németh, "Recipe-based Integrated Semantic Product, Process, Resource (PPR) Digital Modelling Methodology," In: Procedia CIRP, vol. 17, pp. 112-117, 2014.
[32] S. Boschert and R. Rosen, "Digital Twin - The Simulation Aspect," In: P. Hehenberger, D. Bradley (eds): Mechatronic Futures. Springer, Cham, 2016.
[33] M. Kunath and H. Winkler, "Integrating the Digital Twin of the manufacturing system into a decision support system for improving the order management process," In: Procedia CIRP, vol. 72, pp. 225-231, 2018.
[34] M. A. Hayes and M. A. M. Capretz, "Contextual anomaly detection framework for big sensor data," In: Journal of Big Data, vol. 2, article 2, 2015.
[35] A. Mahapatra, N. Srivastava, and J. Srivastava, "Contextual Anomaly Detection in Text Data," In: Algorithms, vol. 5, no. 4, pp. 469-489, 2012.
[36] S. Hagemann, A. Sünnetcioglu, and R. Stark, "Hybrid Artificial Intelligence System for the Design of Highly-Automated Production Systems," In: Procedia Manufacturing, vol. 28, pp. 160-166, 2019.
[37] D. Weidemann and R. Drath, "Einleitung," In: R. Draht (eds), Datenaustausch in der Anlagenplanung mit AutomationML. VDI-Buch. Springer, Berlin, Heidelberg, 2010.
[38] W. Walla, Standard- und Modulbasierte digitale Rohbauprozesskette. Frühzeitige Produktbeeinflussung bezüglich Produktionsanforderungen im Karosserierohbau der Automobilindustrie. Dissertation, Karlsruhe Institute of Technology, Karlsruhe, 2015.