POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Knowledge-based Model Enhancement through Conformance Checking techniques in Manufacturing Systems

LAUREA MAGISTRALE IN MECHANICAL ENGINEERING - INGEGNERIA MECCANICA

Author: LORENZO SPILOTROS - DANIELE URBINATI

Advisor: PROF. TULLIO ANTONIO MARIA TOLIO

Co-advisor: DR. MARIA CHIARA MAGNANINI

Academic year: 2021-2022

1. Introduction and Purpose

Over the last decades, digitization in manufacturing domain has become a driving force for new process analysis methods and technologies such as simulations or digital twins. Companies are moving from being simple production sites to smart factories, i.e. manufacturing facilities are being equipped with sensors, state-of-the-art integrated software, and robotics that collect data, analyze them and help on decision-making. These digital technologies lead to greater automation, predictive maintenance, optimization of process and, most importantly, a new level of efficiency and responsiveness to customers that were not previously possible.

When the digital universe and the real world are aligned, higher value is created: new levels of visibility and understanding can be created from previously isolated information. The current challenge in manufacturing systems is to harness the data efficiently and effectively; identification of bottlenecks, prediction of problems, and detection of system deviations are examples of how data can be used to improve processes.

The wide availability of data allows the integration of Process Mining techniques into production systems, whose purpose is to map processes to improve them [2]. Process Mining is a relatively new discipline that lies between Business Process Management and Data Mining, focusing on process analysis through event logs. Starting from logs, Process Mining techniques allow processes to be discovered, monitored, and improved by extracting relevant knowledge [4]. As shown in Figure 1, Process Mining is divided into three areas: Process Discovery [1], Conformance Checking, and Model Enhancement.

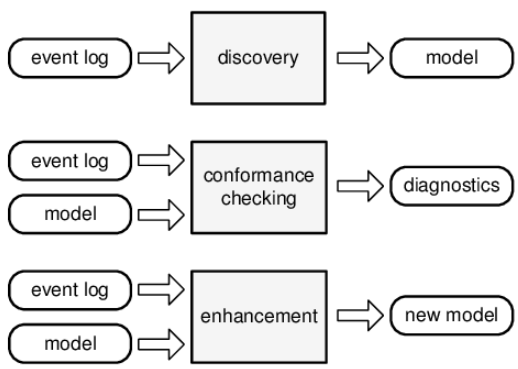

Figure 1: Process Mining techniques explained in terms of input and output: Discovery, Conformance Checking and Enhancement [3]

Process Discovery is a fundamental tool when the flow of instances are unknown, which is the

case of almost all business applications. An example is the healthcare domain, where the implications of some clinical treatments are not known and depend on individual diseases or biological interactions of patients and, therefore, the study of historical data allows the discovery of unknown clinical pathways. The information technology and social networking or education domain also follow the same logic line: the process to be discovered is not known a priori, and Process Mining techniques allow it to be precisely discovered and improved.
Manufacturing systems, however, do not follow the same rules as the aforementioned fields. In fact, Process Discovery may be considered unnecessary and redundant in manufacturing: this is the idea behind the methodology developed in this work. The reason why Process Discovery is redundant is justified by the fact that every company has a structured and available knowledge of its production processes spread across different sources, such as Enterprise Resource Planning system (ERP), Manufacturing Execution System (MES), Business Process Modeling Notation (BPMN) Diagram, senior workers, as well as technical knowledge about the technologies used. Therefore, it is no longer necessary to discover the process from event logs, but this step can be replaced by a translation of the already available and structured knowledge into an appropriate model.
This study proposes the development of a novel methodology for knowledge-based model enhancement through Conformance Checking techniques in manufacturing systems with the aim to guarantee continuous alignment between system digital model and real physical system and the related information. In addition, the proposed methodology allows the implementation of What-if Analyses on strategic production scenarios.
In order to highlight the advantages of the developed methodology, it has been applied to a case study using data collected over 65 months by a leading company in innovative technologies and complete life-cycle solutions for the marine and energy markets. The application of the proposed methodology to a real case allows to highlight the advantages of knowledge-based model building. Lastly, in order to objectively compare the two methodologies, a comparison is proposed based on the results of an experiment conducted with 8 working groups performing the two methodologies in parallel.

The results obtained both from the case study and from the experiment prove that the knowledge-based model construction enables structured identification of all the information not explained by the model, such as Recurrent Patterns that are the primary source of knowledge upgrades and, consequently, of Model Enhancement; instead, Outliers will be able to be identified and then studied individually. A further benefit proven through the application to a real case is the possibility of integrating the methodology with forward-looking techniques allowing the application of the new methodology as a decision-making tool.

2. Criticalities of the Existing Methodology

Process Mining is currently employed to discovery, monitor, and improve processes by emphasizing compliance and performance issues. The starting point of the existing methodology are event logs recorded in a production systems. Therefore, *Process Discovery* aims to create process models starting from the event logs. Subsequently, *Conformance Checking* techniques study deviations between the behavior described by the event logs and the behavior modeled through Process Discovery. Once a deviation is identified, *Model Enhancement* can be performed by redesigning the model through Process Discovery or by repairing the discovered model. The application of Process Mining to production systems has shown some criticalities, if applied to the manufacturing systems, described below:

1. The choice of initial parameters of the Process Discovery algorithm greatly influences the model discovered, affecting the results of all subsequent analyses.
2. Conformance Checking techniques return the measure in which the discovered model fits data, not the measure in which data fit the actual production system, because the discovered model is not representative of the production system.
3. The enhancement of the discovered model is

not coupled with the knowledge update. In addition, not all the extracted information have to change the model: if an item has performed a wrong route, it is important to identify and study it, but it does not have to mandatory correct the model.

4. Process Mining could be improved integrating it with forward-looking techniques as a decision making tool.

3. Proposed Methodology

This study develops a novel methodology for knowledge-based model enhancement through Conformance Checking techniques in manufacturing systems to easily estimate current system performance, allow What-if Analysis on strategic production scenarios and guarantee the continuous alignment between the system digital model and the real physical system. The knowledge-based approach makes it possible to avoid the main criticalities of Process Mining in manufacturing systems. The approach is based on constructing a model from knowledge fragments distributed among the various information sources. Since the model is the translation of all process-related corporate knowledge, the proposed methodology enables nominal knowledge to be compared with the actual behavior of the production system. The proposed methodology is summarized in the block diagram in Figure 2.
The methodology is designed to be general, and therefore can be applied to different production realities. The application requirements are:

1. The processes of the production system have to consist of several operations, tracked by an information system. In practice, one must be able to reconstruct the real path of the part within the production system.
2. Each part within the system must be traced with a unique identifier number.
3. In order to carry out what if analyses, it is necessary to have information on the average processing time of products. This implies that the production monitoring system, must have a clock-on and clock-off function in work orders.
4. Last but not least, information on how processes are to be executed nominally must be present in the company.

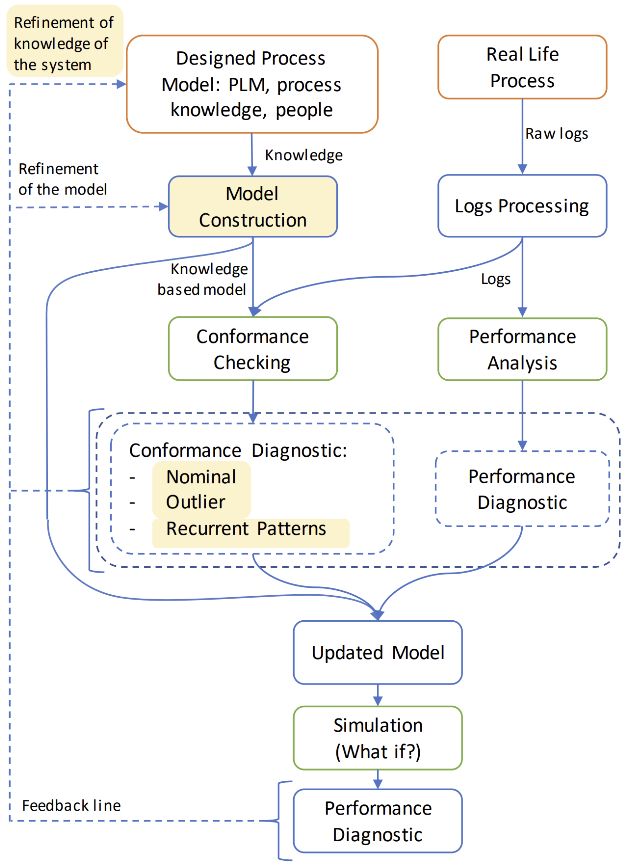

Figure 2: Proposed Methodology.

As can be observed from the block diagram, the proposed methodology replaces Process Discovery with a priori *Model Construction*. The a priori knowledge-based model, in the form of Petri Net, allows to avoid the aforementioned problems. In particular, in Figure 2 the novelties of the proposed methodology with respect to the existing one are highlighted.
Event logs still play a key role as input into the *Conformance Checking* together with the knowledge-based model. This is because event logs keep information about the actual process. Therefore, their integration into Conformance Checking with the knowledge-based model permit the identification of deviations of the actual production process from the nominal one.
In addition, since the nominal process has been translated into the model, it is possible through alignments to classify in a structured way *Nominal Flows*, *Outliers* and *Recurrent Patterns*. Respectively Nominal Flows are paths consistent with the model and they are used to refine the model itself; Outliers are exceptional cases that deviates from the nominal path, they shall be

identified and studied individually; finally Recurrent Patterns are paths not in line with the nominal process, but executed by a high number of items. Recurrent Patterns are the main source of knowledge upgrades representing a gap, a lack in the initial knowledge on which the model was built.
Once the lack of knowledge is identified, and an assignable cause is discovered, it is possible to update both the model and the knowledge about the system. The update is easy to implement because the model is knowledge-based, and it is therefore possible to quickly identify where deviations occur modifying the Petri Net model and updating the associate source of incomplete knowledge.
Filling knowledge gap brings many advantages, chief among them the possibility of making predictions about the production system. What-if Analysis permits the application of the proposed methodology at a strategic level as a support tool for decision-making. It can answer the question "what if...?" and it support managers to better reflect on possible changes of any kind, such as changing routing policy by aiming for flow balancing across the various machines or thinking about adopting a CONWIP policy.

4. Results and Validation of the Proposed Methodology: Case Study

As already mentioned, a case study is analysed to highlight the advantages of the novel methodology. After an initial in-depth study of the company knowledge and a Data Processing phase, it is possible to construct the production process model in the form of a Petri Net. For sake of brevity, only the knowledge-based model of one type of product is reported in Figure 3.

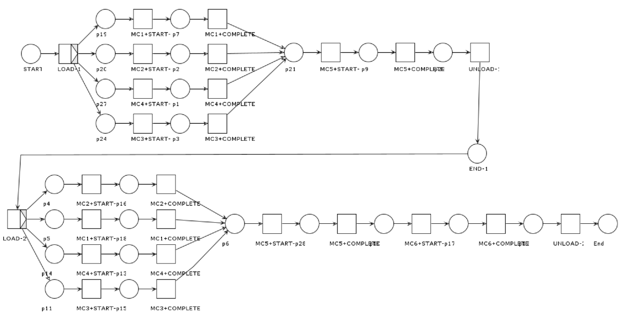

Figure 3: Petri Net model for product PROD1.

The knowledge-based model and the event logs make possible to perform the Conformance Checking. The results reveal misalignments in the traces resulting in partial behavior not mapped by the nominal knowledge of the company. The Figure 4 shows the result of Conformance Checking. Yellow events are wrong events, meaning events present in the data and not in the model. This mismatch emphasizes that the model built on the basis of nominal knowledge is not representative of what actually occurs in the production process.

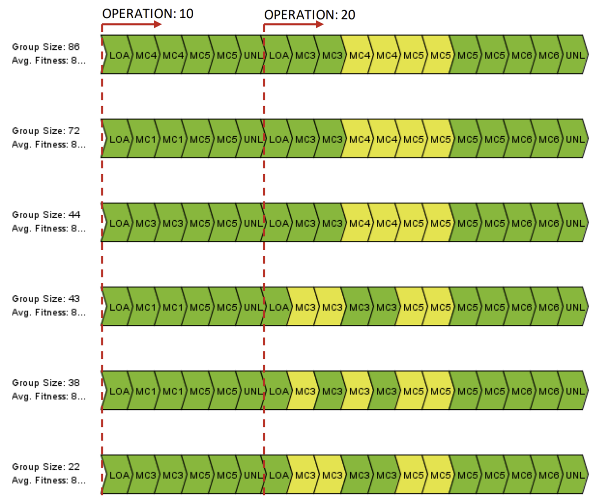

Figure 4: Trace View of the product PROD1 X.

In this specific case, these divergences highlight that, during operation 20, parts pass twice over the machines MC4 or MC3 and MC5. Since the data provider could not be questioned about the reason of this double processing, only hypotheses could be made. The most plausible hypothesis is a rework due to features not respected, such as tolerance.
Once an assignable cause is considered correct, the model is enhanced and the initial knowledge has been updated according to the discovered information. Subsequently, the new model will be the input of a second Conformance Checking. The trace view of the second Conformance Checking is reported in Figure 5.

Due to the model update, most of the traces are aligned with the constructed model. This means that the constructed model and what happens in the production system are now aligned. Some Wrong (yellow) and Missing (purple) Events are present; they can be studied punctually by tak-

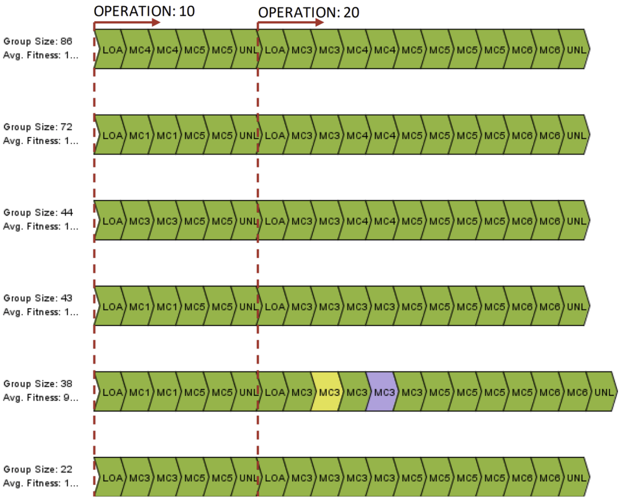

Figure 5: Trace View with model enhanced of the product PROD1 X.

ing a single case under consideration at a time. A key aspect of this way of proceeding is the possibility of focusing on missing events, meaning events present in the model but not performed by the parts. They are, in fact really significant since, if a critical operation or a measurement of important features is not performed on an item, it is possible to quickly identify the item. Figure 6 is an example.

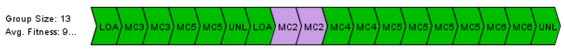

Figure 6: Detail of the Trace View of the product PROD1 X with missing events.

Finally, it is possible to perform What-if practices with the obtained model that accurately represents the process knowledge in terms of routing and cycle times. For this purpose, Jackson Network, a class of Queuing Networks, is used. Firstly, the assumptions of the Jackson Network were verified. Subsequently, Jackson Network was developed in `Matlab` using as input the current performance of the system, such as cycle times and frequency of parts machined. In Table 1 are shown the saturation ρ of the machines, Lead Time and Waiting Time computed in two different ways: from the event logs and from the Jackson Network.

	Data	Model
ρ_{MC1}	78.94%	77.66%
ρ_{MC2}	75.99%	74.89%
ρ_{MC3}	69.30%	67.68%
ρ_{MC4}	94.60%	92.42%
ρ_{MC5}	27.33%	26.78%
ρ_{MC6}	29.36%	28.92%
Lead Time	16.42 h	17.40 h
Waiting Time	13.43 h	14.43 h

Table 1: Saturation, Lead Time and Waiting Time computed from the event logs and from the Jackson Network.

The comparison in Table 1, confirms the validity of the model built through the Jackson Network. This means that it is employable to perform What-if Analysis. By changing input data it is possible to simulate the future behavior of the system. Table 2 shows how the key performance indicators of the system change if parts of product PROD1 processed by the machine MC1 are moved to machine MC3.

	Actual	Scenario
ρ_{MC1}	77.66%	69.45%
ρ_{MC2}	74.89%	74.89%
ρ_{MC3}	67.68%	74.87%
ρ_{MC4}	92.42%	92.42%
ρ_{MC5}	26.78%	26.78%
ρ_{MC6}	28.92%	28.92%
Lead Time	17.40 h	17.14 h
Waiting Time	14.43 h	14.19 h

Table 2: Comparison of actual and modified saturation, Lead Time and Waiting Time.

This model assists managers in decision-making, in fact the consequences of decisions can be predicted. Specifically, as shown in the Table 2, Lead Time decreased as a consequence of the routing change.
In conclusion, by applying the methodology deviations from the nominal information initially provided by the company are discovered,

hence both the model and the knowledge source were updated appropriately. This update allows alignment between the digital model and the actual system. In addition, it is possible to extract system performance indicators such as lead time, saturation of the machines, and up-to-date process parameters, such as cycle time and part routing. These indicators allow the Jackson Network to be constructed with the purpose of employing it as a What-if tool.

5. Comparison between the Existing and Proposed Methodologies: Experiment

In order to compare the two methodologies, the concept behind the Value Stream Mapping, i.e. the distinction between time spent for value-added and non-value-added activities, is applied. Concretely, an experiment is carried out with 8 working groups running the two methodologies in parallel with a simplified manufacturing system case. During the experiment, the groups were asked to keep track of the time spent performing the various steps of the methodologies. In this way, it was subsequently possible to construct a map similar to the Value Stream Map identifying the activities that add value for the ultimate goal of the methodology, i.e. knowledge extraction, with their respective timeframes. Knowledge extraction is translated into directed questions equal for all the working groups. These questions were aimed at investigating the performances of the simplified case study, such as processing time, part routing and Lead Time.

The results obtained were summarized according to a parameter called accuracy. This parameter indicates how complete the answer to the questions is: a value of 100% indicates a completely correct answer, a value of 50% indicates an half complete answer. Table 3 shows, averaging the results by methodology, the accuracy of the questions to the different answers. For sake of simplicity, the questions have not been listed since they are not interesting for comparison.

	1^{st} **Q**	2^{nd} **Q**	3^{rd} **Q**	4^{th} **Q**	5^{th} **Q**
Old	100%	71.5%	100%	75%	50%
New	89.5%	92.5%	100%	100%	100%

Table 3: Accuracy of answers divided by methodologies and questions.

The knowledge-based model allows a more comprehensive and accurate overview of the process. Therefore, the answers of the groups working on the new methodology were complete and detailed, with a higher accuracy than the answers of the other groups.

Finally, it is possible to construct a Value Stream Map of the two methodologies, shown in Figure 7, using the recording of the times taken by different groups in performing the steps required by each methodology.

The Value Stream Map highlights that the methodologies have similar adding value times but different non-adding value times. In particular in the new methodology less time is spent to accomplish tasks useless for knowledge extraction.

Summarising, groups performing the proposed methodology reported answers with higher accuracy and more details, spending the majority of time in doing adding value tasks.

6. Conclusion

The developed methodology, described in details in Section 3, is an entirely different approach to enhance the processes with respect to the existing methodology. In fact, it allows to point out divergences that can occur between the nominal process consisting in company knowledge and the production system data. These divergences are the primary source of knowledge updates.

The results obtained by applying the proposed methodology to a real case prove the advantages it can bring in manufacturing systems. Building the process model based on the company knowledge allows the extraction of detailed and complete information from the event logs and, consequently, a conscious and structured knowledge update. In addition, the developed methodology includes the possibility of integrating it with forward-looking techniques.

The main results obtained from the experiment

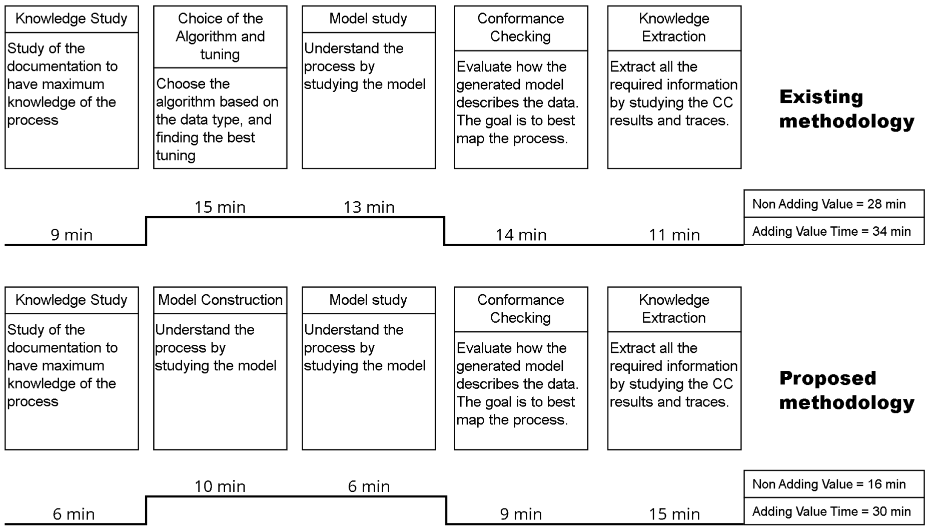

Figure 7: Value Stream Map of the two methodology.

confirm the case study results. In fact, the new approach requires less time and it allows more detailed and comprehensive information to be extracted. In addition, the Value Stream Map, in Figure 7, shows that the percentage of time spent on value-adding tasks is higher in the proposed methodology (65%) with respect to the existing one (55%).
In conclusion, the knowledge-based model enhancement through Conformance Checking will provide significant benefits in the manufacturing systems highlighting the divergences between the nominal process and data from the production system resulting in continuous alignment between system digital model and real physical system. The proposed methodology could further be improved by taking care of different aspects. Particularly, it would be interesting to:

1. Automatically translate BPMN into Petri nets through ad-hoc tools. This possibility is very interesting from a business point of view, as companies widely adopt the BPMN language due to its simplicity of construction.
2. Automatic integration of extracted knowledge into business documentation, such as flow charts or BPMN Diagram.
3. Automatic integration of the Petri Net model with tools for performing what-if analyses. The work shows that Petri Net model used for knowledge extraction can be directly translated and simulated via Jackson Network. This integration could be automated.

Acknowledgements

First of all, we would like to thank our advisor Prof. Tullio Tolio and our co-advisor, Dr. Maria Chiara Magnanini of the Department of mechanical engineering Politecnico di Milano, who was always ready to give us the right guidance at every stage in the creation of the paper. Thanks to you, we have increased our knowledge and competence. Special gratitude goes to the 16 students who actively participated in the experiment by spending their precious time with us.

References

[1] Felix Mannhardt, Massimiliano de Leoni, Hajo A. Reijers, and Wil M. P. van der Aalst. Data-driven process discovery - revealing conditional infrequent behavior from event logs. In *Advanced Information Systems Engineering*, 2017.

[2] Wil Van Der Aalst. Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems*, 2012.

[3] Wil van der Aalst and et al. Process mining manifesto. In *Business Process Management Workshops*, 2012.

[4] Wil M. P. van der Aalst. Distributed process discovery and conformance checking. In *Fundamental Approaches to Software Engineering*, 2012.