Enhancing Digital Twins for Production through Process Mining Techniques: A Literature Review

Marcel Schumacher[1], Ralf Buschermöhle[1], Liane Haak[1], Max Höfinghoff[1], Arne Seipolt[1], Goy-Hinrich Korn[2]

Abstract: A digital twin (DT) plays a vital role in the advancement of manufacturers towards Industry 4.0. However, the creation and maintenance of DTs can be time-consuming. One approach to streamline this process is the utilization of process mining (PM) methods and techniques, which can automatically generate valuable information for DTs. Therefore, this paper aims to examine different approaches that augment DTs with PM and explore their effects. The review categorizes these approaches into three groups: theoretical approaches, approaches with laboratory case studies, and approaches with real-world case studies conducted by manufacturers. The review reveals that the use of PM can enhance the flexibility and sustainability of DTs. However, this improvement comes at the cost of requiring high-quality data and more data preparation efforts.

Keywords: Process Mining, Digital Twin, Production

1 Introduction

The creation of a digital twin (DT) is time-consuming [Br21] but at the same time the DT is an essential component for manufacturers advancing to industry 4.0 [LM21]. In order to accurately represent real-world objects, such as the production process, a DT could relie on a process model [Br21]. Traditionally, the manual creation or updating of process models is both time-consuming and susceptible to errors. To overcome these challenges, Process Mining (PM) techniques can be employed to automatically create and update process models using event data. Event data is generated by performing the process and stored in various information systems like enterprise resource planning systems (ERP-Systems), manufacturing execution system (MES), programmable logic controller (PLC), etc. [va16].

The objective of this paper is to provide an overview of different approaches for augmenting DTs using PM and examine their effects on DTs. To achieve this, a comprehensive review of the existing literature will be conducted, categorizing the approaches into three distinct categories. The first category comprises theoretical approaches that have not been tested in any case studies. The second category includes

[1] Hochschule Osnabrück, Fakultät Management, Kultur und Technik, Kaiserstraße 10c, 49809 Lingen (Ems), marcel.schumacher@hs-osnabrueck.de; r.buschermoehle@hs-osnabrueck.de; l.haak@hs-osnabrueck.de; m.hoefinghoff@hs-osnabrueck.de; a.seipolt@hs-osnabrueck.de

[2] Bernard Krone Holding GmbH & Co. KG, CIO & CDO, 48480 Spelle, Heinrich-Krone-Straße 10, goy-hinrich.korn@krone.de

approaches that have been tested in a laboratory environment, and the third category consists of approaches that have been tested in real-world manufacturing companies.

The paper is structured as follows: Chapter 2 begins by introducing a definition of DTs. Chapter 3 outlines the fundamentals of PM. Chapter 4 explains the research methodology employed in this study. Chapter 5 presents the review itself, providing an analysis of the different approaches. The final chapter initiates the discussion based on the findings and concludes the paper by summarizing the key insights gained from the review.

2 Digital Twin

The concept of a digital twin (DT) encompasses various definitions that vary depending on the specific domain they originate from [Be21]. One definition that encapsulates the essence of a DT in the manufacturing context is as follows: " a digital twin is an integrated multi-physics, multi-scale, probabilistic simulation of a complex product and uses the best available physical models, sensor updates, etc., to mirror the life of its corresponding twin." [Ta18, p. 3564] This definition highlights the key characteristics of a DT in the production domain. In manufacturing, the primary technologies employed for digital twins include simulation techniques such as discrete event simulation and continuous simulation, communication protocols, and core technologies associated with Industry 4.0 [Kr18]. These technologies enable the creation of a virtual representation that closely mirrors the behavior and lifecycle of the physical counterpart, allowing for advanced analysis and optimization within the manufacturing process. There are multiple benefits from using DTs in the production. For Example, the DT supports the planning and controlling [Ro15], mainting [DUM17] and the layout planning of the production [ULS17]. The next chapter explains the theoretical knowledge of PM and the different ways PM can be used to enhance DTs.

3 Process Mining

Process mining (PM) is focused on capturing and analyzing the workflow of a process based on its digital footprints. When processes are executed in the real world, they often leave behind data that can be considered as event data. This event data can be transformed into an event log, which is a compilation of event data organized by cases, sequences, and activities. For instance, let's consider an example of an order being processed for a customer. All the events associated with that order, such as order placement, order confirmation, production, packaging, and shipping, would be grouped together as a case in the event log. Each event would have a specific order in which it occurred within the overall process. By analyzing this event log, process mining techniques can extract valuable insights about the process, such as the sequence of activities, bottlenecks, and variations. This allows for a comprehensive understanding of the process and facilitates process improvement and optimization efforts [vWM04]. PM can be categorized into three types, each serving a distinct purpose within the analysis of event logs: The first type is process discovery, which aims to create a process model based on a given event log, utilizing discovery algorithms [va10]. By analyzing the event log, these algorithms uncover the underlying process flow and dependencies, resulting in a process model that represents the observed behavior. The second type is conformance checking, which involves comparing a process model against an event log to evaluate its fitness and precision [va16]. This comparison helps identify any deviations or non-conformances between the model and the actual process execution. Furthermore, conformance checking can also be used to compare two process models, enabling the detection of differences between them [va05]. The third and last type of PM is process enhancement. The process enhancement aims to repair or extend the process model with new perspectives [va12]. The most common ones are the time, organizational and case. All of them show a different view of the process[va12]. The organizational represents the organizational structure, the social network, the role and the behavior of the resources [Sv08]. The timley exposes frequencies, process time, service time, waiting time, etc. of the process and allows to find bottlenecks in the process [va12]. The case perspective uncovers the rules behind a decision in the process and makes them visible as a decision tree [Rv06]. The next chapter explains the used research methodology for the paper and which paper were chosen for the review.

4 Research Methodology

The objective of the papers is to provide an overview of the various approaches of process mining in digital twins specifically within the manufacturing domain. The review follows the structure recommended by Kitchenham et al. in 2013 [KB13] to address the following research question: How could PM enhance DTs in the production and which effects result subsequently?

In order to answer this research question, a systematic approach was taken to gather relevant scientific papers from diverse scientific databases. The details of the search string and library databases used are presented in Table 1. It is important to note that the total number of papers includes the results obtained from multiple databases.

Library	Search in	Search string	Result
IEEE	Metadata	“process mining”	17
Science Direct	Metadata	AND “digital twin”	4
Google Scholar	Title & Keywords	AND “production”	87

Tab. 1: Overview library databases

The total of 108 papers were manual sorted with the help of exclusion (EC) and inclusion criteria (IC) like Kitchenham suggested [KB13].

- EC1: Papers aren't available in English.
- EC2: Papers aren't accessible by the authors.
- EC3: Papers aren't in the context of production or manufacturing
- IC1: Papers discuses a DT with PM as a component

Out of the 108, only 13 satisfy the IC and EC and are checked for their quality. The quality check secures that only paper with a high quality are used for the review, several quality criteria (QC) are introduced to give each paper a quality score. The score reflects if a paper fully (1), partly (0,5), or not (0) meets the criteria.

- QC1: Does the paper explain how PM enhances the DT?
- QC2: Does the paper explain why PM is used for the DT?
- QC3: Does the paper use its own approach for PM and DT?
- QC4: Does the paper explain the effects of PM for the augmentation of the DT?

Paper	QC1	QC2	QC3	QC4	Total
[PLN22]	1	0	0	0	1
[RA20]	1	1	1	1	4
[Tr21]	1	1	1	1	4
[KNB22]	1	1	1	1	4
[Fr22]	1	1	1	1	4
[MLC22]	1	1	1	0	3
[LM21a]	0	0	0	0	0
[NV22]	1	1	1	1	4
[Br21]	1	1	1	1	4
[CAP21]	1	1	1	0	3
[Ya22]	0	0	0	0	0
[Fu23]	1	1	0	0	2

Tab. 2: Quality criteria results

The score of each paper after applying the quality criteria can be viewed in the table 2. If a paper doesn't reach a total score of at least 3, it will be excluded from the review. Only 9 paper could reach the needed quality score of at least three. These papers are reviewed and presented in the next chapter.

5 Literature Review

This chapter explains in three subchapters the different ways PM is used to augment a DT.

5.1 Theoretical approaches

Brockhoff presents in his paper [Br21] an approach for the creation of a digital twin. According to their interpretation, the DT serves the purpose of predicting problems while a process is in operation. They adopt a model-driven approach to generate the DT, and PM techniques are employed to create process models from the stored information within the DT. The discovered process models are stored in the DT and utilized in conjunction with conformance checking methods to determine any kind of deviation between the discovered model and the event-log. If any deviations are found, they are reported to a reasoner, which suggests a reaction. This action is executed by the executer.

A similar approach is used by Chiò et al. to detect changes in the production line. By utilizing process mining techniques, they generate a graphical representation of the event log, which allows for visual analysis by a human operator. Based on this analysis, necessary actions can be taken if any changes are identified. [CAP21].

5.2 Approaches with case study in the lab

The paper by Friedrichs et al. presents a framework for data-driven DTs. This approach uses PM to extract process models from the event logs using process discovery. The discovered model creates the simulation model of the DT. The discovered process model serves as the simulation model for the DT, ensuring that it is always up-to-date and accurately reflects the mirrored production environment [Fr22].

Another approach from Mayr et al. [MLC22] focusses on creating process models from process-state data. This involves generating an event log of the process using clustering techniques. The generated event log is then used to discover the process model, which monitors the control-flow of the production process.

A completely different approach by Novák and Vyskočil [NV22] uses PM to mine the operational data of time for the DT. The time data itself is dived into duration, waiting time and activity times of the process. This allows the DW to be flexible and gives him the ability to react to changes in the production.

5.3 Approaches with case study by manufactures

In the paper from Kumbhar et al., the objective is to create a digital twin (DT) of the production process to identify bottlenecks based on operational data. Process mining (PM) is employed to construct an accurate process map of the as-is process. This involves creating a comprehensive process map, which is then abstracted and tested for conformance in subsequent steps. The outcome is an as-is process map that serves as the foundation for simulating the production process in the DT. The study demonstrates that PM facilitates the creation of the simulation at the expense of data preparation efforts. [KNB22].

Tran et al. [Tr21] adopt a similar approach to create a DT that optimizes the parameters of the real-world system using simulation. PM is utilized to extract a process model from data, which in turn generates and updates the autonomy of the DT. This automated update process ensures that the simulation remains accurate and closely reflects the actual system at all times.

A different approach for PM and DTs is used by Ruppert and Abonyi [RA20]. This conception has the goal of creating a DT for real time tracking of products in the production. The concept involves utilizing several sensors and the manufacturing execution system (MES) to monitor and track products in real time. PM uses this information to extract the process flow and to keep the simulation adaptive.

6 Discussion & Conclusion

The objective of this paper is to present an overview of how process mining is employed to enhance digital twins (DTs) and the resulting effects on DT. Based on a limited number of papers reviewed, it is evident that the majority of approaches focus on achieving information retrieval goals for DTs. Currently, the retrieved information primarily includes the process model, processing time, waiting time, and total process time. In addition to the process model and times, process mining could also generate information about the individuals involved in the process. This is generally not the case. Reasons for this may be the insufficient availability of such information for extraction or the absence of a need for it.

The papers show that the enhancement of DT with PM brings a few positive effects: The DT reflects the actual process, can be adjusted to a newly planned production and the generation of simulation models is supported.

Indeed, the utilization of process mining in enhancing digital twins (DTs) does present certain challenges: The availability and the completeness of data for PM as well as the event data needs to be prepared for PM.

Up to now, there are only a few studies using DTs to enhance with PM. Nevertheless, a clear picture emerges. The approach of a sustainable and flexible DTs can be supported by PM. The ability to extract and prepare information from existing information systems and process it within the digital twin framework is a crucial step in advancing production towards the state of Industry 4.0.

Acknowledgement

This work is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant No. 01MD22001C as part of the edge data economy initiative.

Bibliography

[Be21] Bergs, T. et al.: The Concept of Digital Twin and Digital Shadow in Manufacturing. Procedia CIRP 101, pp. 81–84, 2021.

[Br21] Brockhoff, T. et al.: Process Prediction with Digital Twins: 2021 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C). IEEE, pp. 182–187, 2021.

[CAP21] Chiò, E.; Alfieri, A.; Pastore, E.: Change-point visualization and variation analysis in a simple production line: a process mining application in manufacturing. Procedia CIRP 99, pp. 573–579, 2021.

[DUM17] D'Addona, D. M.; Ullah, A. M. M. S.; Matarazzo, D.: Tool-wear prediction and pattern-recognition using artificial neural network and DNA-based computing. Journal of Intelligent Manufacturing 6/28, pp. 1285–1301, 2017.

[Fr22] Friederich, J. et al.: A framework for data-driven digital twins of smart manufacturing systems. Computers in Industry 136, pp. 103586, 2022.

[Fu23] Fur, S. et al.: Sustainable Digital Twin Engineering for the Internet of Production. In Digital Twin Driven Intelligent Systems and Emerging Metaverse. Springer Nature, Singapore, pp. 101–121, 2023.

[KB13] Kitchenham, B.; Brereton, P.: A systematic review of systematic review process research in software engineering. Information and Software Technology 12/55, pp. 2049–2075, 2013.

[KNB22] Kumbhar, M.; Ng, A. H.; Bandaru, S.: Bottleneck Detection Through Data Integration, Process Mining and Factory Physics-Based Analytics. In SPS2022: Proceedings of the 10th Swedish Production Symposium. IOS Press, 2022.

[Kr18] Kritzinger, W. et al.: Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine 11/51, pp. 1016–1022, 2018.

[LM21] Lugaresi, G.; Matta, A.: Automated Digital Twins Generation for Manufacturing Systems: a Case Study. IFAC-PapersOnLine 1/54, pp. 749–754, 2021.

[MLC22] Mayr, M.; Luftensteiner, S.; Chasparis, G. C.: Abstracting Process Mining Event Logs From Process-State Data To Monitor Control-Flow Of Industrial Manufacturing Processes. Procedia Computer Science 200, pp. 1442–1450, 2022.

[NV22] Novák, P.; Vyskočil, J.: Digitalized Automation Engineering of Industry 4.0 Production Systems and Their Tight Cooperation with Digital Twins. Processes 2/10, pp. 404, 2022.

[PLN22] Park, K. T.; Lee, S. H.; Noh, S. D.: Information fusion and systematic logic library-generation methods for self-configuration of autonomous digital twin. Journal of Intelligent Manufacturing 8/33, pp. 2409–2439, 2022.

[RA20] Ruppert, T.; Abonyi, J.: Integration of real-time locating systems into digital twins. Journal of Industrial Information Integration 20, pp. 100174, 2020.

[Ro15] Rosen, R. et al.: About The Importance of Autonomy and Digital Twins for the Future of Manufacturing. IFAC-PapersOnLine 3/48, pp. 567–572, 2015.

[Rv06] Rozinat, A.; van der Aalst, W. M. P.: Decision Mining in ProM. In (Hutchison, D. et al. Hrsg.): Business Process Management. Springer, Berlin, Heidelberg, pp. 420–425, 2006.

[Sv08] Song, M.; van der Aalst, W. M.: Towards comprehensive support for organizational mining. Decision Support Systems 1/46, pp. 300–317, 2008.

[Ta18] Tao, F. et al.: Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology 9/94, pp. 3563–3576, 2018.

[Tr21] Tran, T. et al.: Real-time locating system and digital twin in Lean 4.0: 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI). IEEE, pp. 369–374, 2021.

[ULS17] Uhlemann, T. H.-J.; Lehmann, C.; Steinhilper, R.: The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0. Procedia CIRP 61, pp. 335–340, 2017.

[va05] van der Aalst, W. M. P.: Business alignment: using process mining as a tool for Delta analysis and conformance testing. Requirements Engineering 3/10, pp. 198–211, 2005.

[va10] van der Aalst, W.: Process Discovery: Capturing the Invisible. IEEE Computational Intelligence Magazine 1/5, pp. 28–41, 2010.

[va12] van der Aalst, W. et al.: Process Mining Manifesto. In (Daniel, F.; Barkaoui, K.; Dustdar, S. Hrsg.): Business Process Management Workshops. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–194, 2012.

[va16] van der Aalst, W.: Process Mining. Data Science in Action. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[vWM04] van der Aalst, W.; Weijters, T.; Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering 9/16, pp. 1128–1142, 2004.

[Ya22] Yang, M. et al.: A Novel Embedding Model Based on a Transition System for Building Industry-Collaborative Digital Twin. Applied Sciences 2/12, pp. 553, 2022.