CONFERENCE ON PRODUCTION SYSTEMS AND LOGISTICS
CPSL 2023-2

5th Conference on Production Systems and Logistics

A Framework For The Domain-Driven Utilization Of Manufacturing Sensor Data In Process Mining: An Action Design Approach

Jonathan Brock[1], Niclas Rempe[1], Sebastian von Enzberg[1], Arno Kühn[1], Roman Dumitrescu[2]

[1]*Fraunhofer Institute for Mechatronic System Design IEM, Zukunftsmeile 1, 33102 Paderborn, Germany*
[2]*Heinz Nixdorf Institute, University of Paderborn, Fürstenallee 11, 33102 Paderborn, Germany*

Abstract

Manufacturers install and rely on a large number of sensors to operate and control their processes. However, the collected sensor data is rarely used to analyse and improve the higher-level, aggregated business processes. Process mining (PM) appears to be a promising solution, with the ability to automatically generate and analyse business process models based on data. However, the atomic events of sensor measurements need to be refined, aggregated, and enriched to properly represent a business process. In this paper, we propose a novel framework to make manufacturing sensor data analysable with PM. The framework allows manufacturers with batch and continuous processes (BCP) to systematically enrich their sensor data to use it for optimization purposes. Following the action design research, we demonstrate the applicability of the framework in a use case study using sensor data from a BCP beverage production.

Keywords

Process Mining; Manufacturing; Sensor Data; Framework; Real-World

1. Introduction

Manufacturers need to constantly analyse and optimize their value generating processes, in order to save costs and stay competitive [1]. Traditionally, they deploy a variety of tools and methodologies to achieve this, such as lean management [1,2]. More recently, digitalization has proven to be a viable optimization possibility [3]. While this digitalization often had no effect on the sensor and actor IT-layer, recent trends like the industrial internet of things steadily increase the availability of data from this layer [4,5]. In the intersection between the need for optimization and the availability of data, process mining (PM) has emerged in recent years as a promising technology [6]. PM describes "[…] techniques, tool and methods to discover, monitor and improve real processes […]." [7]. Manufacturers can particularly profit from PM through increased transparency, measurement of process performance, or the creation of digital twins [8,9]. The minimum requirement for data in PM is a case ID (e.g. an order number), an activity name (e.g. drilling) and a timestamp [7]. Data with at least these three features is called an event log [7]. PM is typically applied to data generated from process aware IT-systems, such as Enterprise Resource Planning (ERP), where for example workflows easily provide the necessary data structure [6].

However, in many real-world scenarios, sensors and actors are not necessarily aware of the current case that is being processed [10]. This has to do with the prevalent, classical pyramid style IT-architecture in

DOI: https://doi.org/10.15488/15267
ISSN: 2701-6277

manufacturing companies, where machine sensors and actors are often decoupled from higher hierarchy IT-systems in order to ensure real-time capability [4,5]. Consequently, when working with sensor data in PM, problems like a lack of process notion (i.e., missing case IDs or activity names), the mapping of fine granular sensor data to (human) activities, and the aggregation of sensor data to process activities, arise [11,10,12]. Various authors call for further research on the utilization on sensor data for PM [13–15]. In this paper, we derive a framework to make manufacturing data from the sensor and actor layer usable in PM. The framework is designed through Action Design Research (ADR), meaning that we iteratively work on the manufacturing sensor data from a real-world organization [16]. Our main contribution is the framework consisting of six phases with an emphasis on applicability in industry, meaning that it draws from existing domain expertise. Additionally, we contribute an activity grouping scheme and case ID inheritance algorithm, which allows other organizations to apply PM on their manufacturing sensor data.

The remainder is structured as follows. In section 2 we introduce the application scenario and existing solutions for sensor abstraction. Section 3 will explain our research methodology ADR. In section 4, the framework for will be introduced and applied to the application scenario. In section 5 we will discuss our results, before concluding the paper in section 6.

2. Need for action on manufacturing sensor data utilization in process mining

In this section we will introduce the application scenario (section 2.1) and existing solutions to it in the state of the art (section 2.2). In section 2.3 these two will be compared to motivate the need for further research.

2.1 Application scenario and objective of Big Beverage Inc.

Big Beverage Inc. is a family owned and run manufacturer of drinks, with around 400 employees in Germany. Their production resolves around mixing raw fluids, which are then later combined, and finalized with water. Their machinery setup as well as the collected data are displayed in Figure 1. The machine layer displays how the organization utilizes different tanks and pipes to produce their drinks.

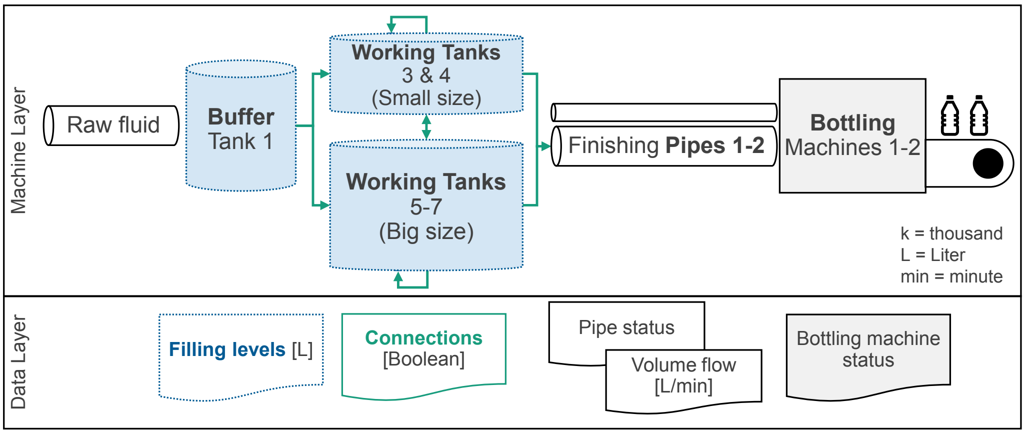

Figure 1 The machinery and data at the Big Beverage Inc.

Figure 1 shows that initially raw fluids run into a buffer tank, which then distributes the material into one or multiple working tanks (No. 4 to 7). The working tanks have different volumes. Before sending the final good to one of two bottling machines through a finishing pipe, the raw fluid might be re-distributed to other tanks or combined with other raw fluids. The raw fluid will always be finalized by adding water before they are bottled. From a production perspective, some elements of this process are a batch process (e.g. moving an entire tank filling), while others are a continuous process (e.g. the finishing pipe) [17]. From a data perspective, the filling levels of all tanks are measured. Additionally, it is registered if any of the tanks are connected to one another or to any of the two finishing pipes as a boolean value (i.e. True/False). For the finishing pipes and bottling machines, the status (e.g. production) is registered. Additionally, the current volume flow for the pipes is measured. All sensor data is measured every 30 seconds. The manufacturer is faced with a rapid increase in orders, while keeping their production resources constant in the past years. Therefore, the objective of Big Beverage Inc. is to increase their resource availability, by identify if and how they are unnecessarily using tanks to produce final products. Their solution approach is to understand the frequency of shuffling between tanks to derive meaningful actions using evidence-based business processes discovered with PM. To achieve this, the existing sensor data needs to be aggregated to an event log.

2.2 State of the art

PM uses data to analyse processes [7]. Different techniques for this analysis exist. The first technique *process discovery* uses data to produce a process model, such as a petri net [6]. Different algorithms such as the alpha miner [18] or the inductive miner have been proposed for this [19]. The second type *conformance checking* compares data from as-is executions with (normative) process models, in order to identify deviations [6]. For this, tokens, rules or alignments are used [20]. The third type *process enhancement* adds additional information to process models, for example time information [6]. While these three are the traditional main techniques, recent years have added comparative, predictive and action-oriented PM [21]. The minimum data requirement for PM are a case ID, a timestamp and an activity name [22,7]. Additionally, a lifecycle information (e.g. start) helps to aggregate atomic events to a coherent activity [22].

Matching atomic events to higher hierarchy business processes is a general challenge in PM, not only limited to manufacturing sensor data. Van Zelst et al. [23] recently proposed a taxonomy for event abstraction techniques in PM. In their taxonomy, techniques are classified based on their supervision strategy, relation of case and activity, or used data. Only one of the 21 identified publications deals with continuous data (i.e. sensor data). The research of van Zelst et al. [23] focuses on very specific approaches, often on an algorithmic level. Consequently, the authors exclude higher-level procedures.

However, in practice, abstracting sensor data to higher level event logs is not necessarily an algorithmic challenge but also a procedural because data and domain might vary. Hence, we identify five higher-level approaches in the literature that explicitly deal with the aggregation of sensor data for PM [24,10,25,26,12]. We exclude approaches like [8,27], because PM is solely used to analyze machine behavior (i.e. no aggregation to superordinate business processes is performed) or the approach is too generic, respectively. By comparing and correlating the objectives of the authors' explicitly mentioned steps, we derive nine general phases that are described in literature. These nine general phases are shown on the left in Table 1. In *Collection*, necessary data is acquired. In the second phase *Identification*, the case ID needed to conduct PM is determined. *Segmentation* deals with the division of the sensor data following some logic. In *Characterization*, relevant features to distinguish possible activities are identified and calculated. Based on these features, the phase *Clustering* deals with the (automated) grouping of the sensor values. Afterwards, these clusters need some *Interpretation*, often using domain expert input. Once sensor data is clustered and interpreted, activities with human readable names can be derived in the *Generation* phase. In *Creation*, the final event log is generated, followed by the actual process *Mining*.

Van Eck et al. [12] propose a six step transformation approach to use sensor data from smart products in PM. Compared to the other approaches, the approach has a detailed description of activity generation. Koschmider et al. [10] develop a four step framework to derive process models from any kind of sensor data. Prathama et al. [26] derive a three step framework that utilizes sensor data from wearable devices in PM. The authors have a dedicated step for data acquisition. Brzyhczy and Trzcionkowska [24] report on their four step experience creating an event log from underground mining machinery logs. In contrast to the other authors, a detailed discussion on the challenges of case ID identification is reported. Lastly, de Leoni and Dündar [25] propose a four step abstraction technique based on clustering, that functions with little domain knowledge. Their approach is rather algorithmic and general. The two publications [25,12] also appear in the taxonomy in [23]. A general observation of Table 1 is that none of the procedures are applied to industrial manufacturing settings within a factory, albeit [24] describe a underground mining setting. Possibly therefore, they are the only ones discussing the challenge of identifying the case ID. The approaches [24,10,25,12] additionally have a strong focus on the computation of clusters by using distances measures or machine learning. De Leoni and Dündar [25] highlight this aspect as a key contribution.

Table 1 Overview of literature dealing with sensor data in the context of PM [24,10,25,26,12]

		Source				
		van Eck et al. [12]	Koschmider et al. [10]	Prathama et al. [26]	Brzyhczy and Trzcionkowska [24]	de Leoni and Dündar [25]
Phases	**Collection**			(1) Event data gathering		
	Identification				(1) Case ID Identification	
	Segmentation	(1) Segmentation of sensor data		(2.1) data pre-processing	(2) Data preprocessing	(1) Create sessions
	Characterization	(2) Feature selection and calculation to characterise segments		(2.2) data integration		
	Clustering	(3) Clustering of segments to identify time periods with similar behaviour	(1) Event correlation		(3) Data clustering	(2) Clustering of sessions
	Interpretation	(4) Interpretation of clustering results to label behaviour	(2) Activity Discovery		(4) Classification for labelling activity names in the raw data	(3) Visualize heat map
	Generation	(5) Activity creation by grouping labelled segments				
	Creation	(6) Process instance creation	(3) Event Abstraction	(2) Event log creation		(4) Creation of abstract event log
	Mining		(4) Discovery of process	(3) Multi-case perspective pm		

2.3 Comparison of Big Beverage Inc. needs and the state of the art

In summary, abstracting fine granular event data to higher level is a known challenge both in practice and research. Their overall goal is to create an event log with the columns case ID, timestamp, activity, and lifecycle information from raw sensor data. Even though Big Beverage Inc. has data from their machines at hand and goals in mind, the existing knowledge base about data abstraction does not prove to be applicable for them. This has to do with two aspects. First, the existing solutions rarely focus on manufacturing challenges, especially for BCP. These challenges include the difficulties of identifying activities and the lack of a case ID. This is complicated by the high parallelism and nesting of the manufacturing steps which leads to the challenge of associating activities and cases. Second, the algorithmic heavy approaches do not fit into the daily work of the company, because experts on clustering techniques from machine learning are sparse. As a result, we derive the following research question (RQ):

RQ: What is a general approach for manufacturers to utilize their machine sensor data in process mining?

Given the application scenario, the following prerequisites can be assumed. First, the needed data sources are known (i.e. if and where machine data exists). Second, the data has been extracted and is complete (e.g. queries for data bases are written and it is not a streaming PM use case). Third, the existing data is of adequate quality (e.g. no missing data). Various guiding publications can be found in literature concerning these issues [22,28,29,12], and are hence not explicitly detailed within this paper.

3. Research methodology to close the need for action

To answer our research question and achieve the goal of Big Beverage Inc., we followed the ADR approach proposed by [16]. The ADR team consists of the ADR researcher from academia, and various departments on the practitioner's side, such as the digital transformation, IT and production. Our research procedure, which took half a year in total, the design iterations, and the final generalization are shown in Table 2.

4. Framework for the domain driven utilization of manufacturing sensor data

The outcome of stage 4 of the research methodology is a framework that guides practitioners to utilize PM on their manufacturing sensor data. The framework consists of six phases with 17 sub-steps. Section 4.1

introduces the general procedure of the framework. Because phase (5) of the framework utilizes a grouping scheme and a dedicated case-inheritance algorithm, details about these two contributions will be given in a separate section 4.2. The framework is applied to Big Beverage Inc. in section 4.3.

Table 2 Summary of the ADR process and generated artefacts according to [16]

Stages and Principles		Artefact
Stage 1: Problem Formulation		
Principle 1: Practice-Inspired Research	The research was driven by the need for machine data to be analyzed with PM, and the challenges faced while generating an event log.	**Recognition:** While many approaches for event abstraction exist, they do not fit the needs of Big Beverage Inc. Many organizations have similar processes. A generalized framework is beneficial for many organizations.
Principle 2: Theory-Ingrained Artefact	We use the existing literature base depicted Table 1 and the event abstraction taxonomy by van Zelst et al. as a guiding principle.	
Stage 2: Building, Intervention, and Evaluation		
Principle 3: Reciprocal Shaping	The initial utilization was not as straight forward as the final framework may suggest. Instead, phases were constantly redone and generalized together with the domain experts. Especially the case definition proofed to be challenging, and multiple approaches were tried.	The **first analysis** approach resulted in an event log that only captured individual machine behavior but could not relate multiple machine to the manufacturing of the same product. A **second analysis** approach with a refinement of the machine data and a collaborative breakthrough in how cases are associated to machine behavior resulted in an event log that reflected all machine efforts necessary to produce a product. By the end of stage 2, we had the **technological capability** to transform the data to an event log.
Principle 4: Mutually Influential Roles	The ADR team included researchers, domain experts, and process analysts, both from academia and practice. The lead for the ADR project lays within the academia.	
Principle 5: Authentic and Concurrent Evaluation	A technical solution for the utilization was derived to work on the specific, real world machine data provided by Big Beverage Inc. A usable event log could be generated.	
Stage 3: Reflection and Learning		
Principle 6: Guided Emergence	The manufacturer realized that a standardization of the procedure was necessary to reproduce results in different machine areas. The ADR team conducted various brainstorming sessions to **analyze** the most relevant actions taken to achieve the project goal.	By the end of stage 3, a **schematic description of our technical solution** was derived.
Stage 4: Formalization of Learning		
Principle 7: Generalized Outcomes	We **synthesized** the schematic description of our actions by aggregating steps and cutting individual solutions.	The **artefact** of stage 3 is the **framework** for utilizing machine sensor data in process mining

4.1 Procedure model of the framework

The first phase **(1) Envision the desired outcome of the process mining analysis** has two sub-steps. In (1.1) the desired outcome needs to be determined. By outcome, we refer to the analysis-artifact that needs to be generated to achieve the higher hierarchy project goal. In (1.2) relevant stakeholders are listed. We differentiate between a shortlist team, i.e. the core project team, and a longlist team, i.e. experts who might be relevant for the process, IT or the sponsoring of the project.

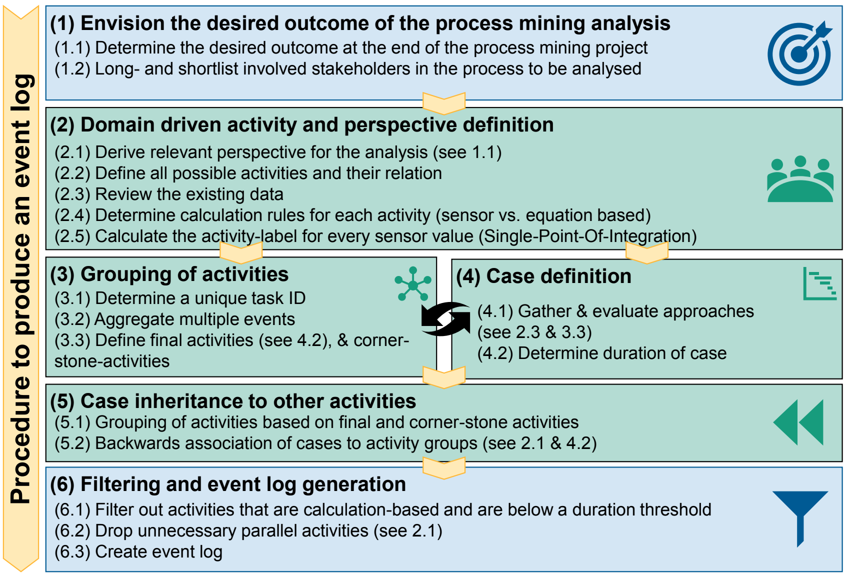

Figure 2 Framework for the utilization of manufacturing sensor data in PM projects

The second phase **(2) Domain driven activity and perspective** uses the initially defined outcome targets as an input. The outcome of this phase is a data frame, where every sensor value has an activity label. In **(2.1)** the relevant perspective for the analysis is chosen, based on the defined goals. This step has a crucial impact on the case definition in later steps. The perspective directly relates to the analysis goal. In **(2.2)** the core project team determines the possible activities and their relation. This can be done in a brainstorming session. Alternatively, a first sketch of the expected process or old process model can be made and discussed, e.g. by using process modeling techniques such as BPMN. The relation aspect refers to the possibility of activities having direct relation, e.g. when one tank fills, another one empties. In **(2.3)** initial data drops are reviewed, possibility leading to a first refinement of the needed data. In step **(2.4)** for each possible activity, calculation rules are determined. We differentiate between sensor- and equation-based rules. Sensor-based rules refer to calculations made using other sensor information, e.g. using machine coupling information. Equation-based rules are mathematical calculations performed to the existing data, e.g. a gradient. Lastly, in **(2.5)** the equations are applied to the sensor data. We advise to use one existing data file that fits the perspective the best, and then associate further sensor information to it. We call this the single-point-of-integration.

The third phase **(3) Grouping of activities** and the fourth phase (4) case definition are performed in parallel. In **(3.1),** every sensor response is given a unique task ID by finding the first and the last consecutive activity-label of the same kind. This task ID can be artificial, e.g. a combination of a consecutive number and the first letter of the activity-label. Then, in (3.2) the same events can be grouped to activity-label. Lastly, in **(3.3)** activities can be marked as a special activity based on the results from phase (4). Two kinds of activities are important: activities that can be the final activities in the process and activities which are corner-stone activities. Corner-stone activities are characteristic to the business process and are the primary activities in a process, which other sub-process activities work towards to.

In phase **(4) Case definition** the notion of a case is determined. In **(4.1)** possibilities for the case definition are gathered. This step is dependent on the initially defined perspective. While ideally, some business level information, e.g. a manufacturing order number, is available, our methodology is specifically designed for manufacturers who do not have these information on their machine data level (see section 1). Therefore, we propose a workaround that allows for the artificial generation of a case. We suggest looking for distinct characteristics in the machine data. This might be a certain machine status, sequences of machine states, or other characteristics that define the end of the case relevant to the analysis perspective. In **(4.2)** the start and end time of the case are determined as a preparation for the fifth phase.

In the fifth phase **(5) Case inheritance to other activities**, every activity is iteratively associated to a case, based on the backwards oriented business perspective of a final product. Because of this iterative passing down of case association, we call this inheritance. Like the previous phase, this phase is also tailored to organizations who have a more classical IT-architecture (see section 1). To achieve this, in **(5.1)** activities are grouped based on a special scheme utilizing the corner-stone activities. Afterwards, in **(5.2)** the final case ID can be iteratively associated to these groups. Remarkably, this allows to integrate BCP. The activity grouping scheme and the case inheritance algorithm are described in more detail in section 4.2.

Lastly, in **(6) Filtering and event log generation**, final preparations are conducted. In **(6.1)** activities below a certain duration threshold are dropped. In this way, the to-be discovered process model becomes clearer. For the same reason, we suggest dropping parallel, related activities in **(6.2)**. Parallelism is a complex driver in PM, especially when using out-of-the-box directly follows graphs, as used by most software vendors. Finally, in (6.3), the event log is generated in a CSV, XES, OCEL, etc. standard.

4.2 Activity grouping scheme and case inheritance algorithm

Phase (5) of the framework introduces the concept of activity grouping and iterative case inheritance. In (5.1), all activities are associated to the same group until a relevant corner-stone or final activity occurs. Then, a new group is build following the same scheme. This scheme is shown in Figure 3. The corner-stone and final activities are in bold letters. The group building scheme is displayed on the left side. Groups are necessary to cut down parallelism, and to associate a wide range of activities to a case. This association is done in (5.2), where activity-groups are now iteratively connected to the interval of the case using information about activity relations from step (2.1) of the framework. This association is performed on the three dimensions relation, time, and case-association-status. The association concept is shown in Figure 3 on the right side. We start by looking at the first defined case, which is case ID 1. We find an activity that has a relation to other activities ("Filled by Tank" is related to "Filling to Tank"). Based on the duration (dimension time) of that activity, we can search in the pool of all activity groups for a related activity (dimension relation) that has no case association yet (dimension case-association-status), for a closely overlapping activity. If we find one, like the activity in group 1, we associate all activities in that group to case ID 1. The same procedure is performed to the case ID 2, associating activity group 2 to the case.

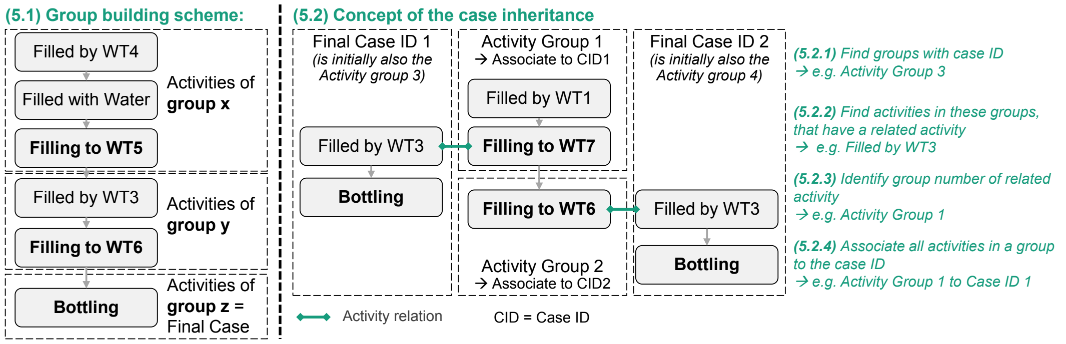

Figure 3 The schemes for the group building in step (5.1) and the (5.2) backwards association scheme

This algorithm is then performed again. While we will not find new activities for association in the activity groups 3 and 4, the activity group 1 is now associated to case ID 1, and has an activity “Filled by Tank 1” which possibly has a related activity, which is not associated to any case ID yet (not displayed in Figure 3). The iteration terminates, when no more associations can be made. This algorithmic approach is displayed in Figure 4, and can be executed in most programming languages.

Case-inheritance-algorithm

Input: dataframe, dictionary of activity relations
Output: dataframe

```
initialize previous_length with -1
while previous_length != length(activities with case_id):
    previous_length = length(activities with case_id)
    for every case_id:
        find activities that can have a relation
            get timestamps of activities
            search relating, undefined activities, where timestamps are close
                set every case_id in the activity_group to current case_id
```

Figure 4 Algorithm to iteratively associate case IDs

4.3 Application example at Big Beverage Inc.

As mentioned in section 3, our methodology involved multiple iterations to come up with the final solution. In this section, we briefly want to highlight the key application steps of the framework and the results achieved. In the beginning, the product batch was chosen as the analysis perspective to determine the needed interactions between tanks (2.1). Then, the activities were defined using a BPMN sketch of the process (2.2). In a workshop, each activity was related to one another (e.g. bottling and pipe flow need to be in parallel) and was given a calculation rule (2.4). While the interaction between the tanks could be determined using the connection data, other activities needed to be calculated using the gradient of filling level (e.g. filling by pipes). In the grouping of activities phase (3), we used the labelled sensor data to derive actual activities. The first and last sensor value of the same kind can easily be determined using a helper column that identifies changes to new labels. The data state after phase (3) is shown on the left side in Figure 5. The label “Fill” is reported from 2023-07-03 10:57 to 2023-07-03 12:13. These two timestamps can be taken and aggregated based on the task ID to the table structure on the right side. Depending on the analysis goal, the sensor data in between those two timestamps can be dropped or aggregated. For example, we kept the start and end payloads. The labels are now called activity because they have a start and end. It is now possible to mark final and primary activities, which can be used to form activity groups. In phase (4) a case ID needs to be determined. For example, we found that the final bottling machines saves data on the product change. This can be used to determine start and end timestamps for products. However, the bottling machine is not linked to other machines. But because these production times and the filling to the pipes are related activities, the case inheritance algorithm can be applied, iteratively associating more activity groups to a case ID. The data wrangling was conducted using the programming language Python.

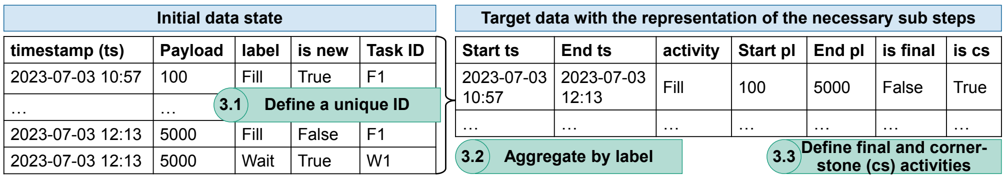

Figure 5 Schematic visualisation of the data transformation in phase (3)

5. Discussion

The application of the framework in section 4.3 shows that our framework produces viable event logs for PM, fulfilling the requirements given in section 2.3. From a managerial perspective, we contribute an applicable, easy to use framework for practitioners. With our framework, data scientists at Big Beverage Inc. were able to identify and quantify the frequency of tank utilization patters needed to produce a final beverage. In the future, the organization wants to incorporate the framework into a data pipeline, to reproduce the analysis more frequently. Practical limitations are twofold. First, the prerequisites discussed in section 2.3 need to be fulfilled. Second, the framework is designed with and for BCP. While this limits the contribution, the phenomenon of BCP is very frequent, e.g. in pharmaceutical processes [17]. In summary, our approach is applicable to manufacturing companies, especially with BCP, where data (sources) are known, extracted and of adequate quality. Additionally, no streaming PM use cases are possible. Therefore, our framework is relevant for practice and not only limited to beverage manufacturing scenarios. From a scientific perspective, we contribute a framework, the grouping scheme, and the case inheritance algorithm to the knowledge base. All three artefacts have been iteratively designed, generalized, and proven valid in a real-world setting. We briefly want to highlight the key differentiations to the existing state of the art. In contrast to [25], our framework is less algorithmic. Unlike [10,12] where sensor data is clustered, then labelled, we identify the activities domain-driven, and then determine them in the data. In that way, we are more applicable for practitioners. By particularly addressing BCP, we address shortcomings of [26,12], who focus on smart products. Lastly, both [24] and our framework address the problem of identifying a case ID. However, unlike [24] we do not have the possibility to identify the case IDs based on the workers positions. We therefore propose the case inheritance algorithm, whereby we address the high parallelism and nesting of typical BCP. In summary, we answer our initially defined research question, as well as the call for research in [13–15].

6. Conclusion and outlook

In this paper, we derive a framework for the utilization of manufacturing sensor data to be used in PM projects by conducting ADR. The framework consists of six phases that guide practitioners in transforming raw manufacturing sensor data to PM usable event logs. By using the framework, domain experts and process scientists can systematically solve challenges inherent in sensor data from batch and continuous manufacturing processes, such as a lack of case notion. We further contribute an activity grouping scheme and a case inheritance algorithm to the knowledge base. In contrast to other publications, we put a strong emphasis on involving domain expertise into the activity and case ID definition, because our ADR revealed this to be more practical in real world settings. Additionally, this allows for a wider range of application scenarios, such as pharmaceutical BCP. By collaborating with an organization from the beverage industry throughout the ADR, we demonstrate and validate the usability of our framework in real world scenarios. Limitations of our research are a possible over-specificity to the application scenario used in the ADR cycle. Further research should therefore identify, refine, and adjust our framework to more domains, process types and application scenarios.

Acknowledgements

This work is funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy (MWIKE) of the State of North Rhine-Westphalia as part of the Leading-Edge Cluster, Intelligente Technische Systeme OstwestfalenLippe (it's OWL) and supervised by the project administration in Jülich (PtJ). The responsibility for the content of this publication lies with the authors.

References

[1] Joppen, R., Enzberg, S. von, Kühn, A., Dumitrescu, R., 2019. A practical Framework for the Optimization of Production Management Processes. Procedia Manufacturing 33, 406–413.

[2] Pirrone, L., Bernasconi, M., Friedli, T., 2023. Investigating The Role Of Digitalization In OPEX Programs – A Case Study From The Pharmaceutical Industry. Hannover : publish-Ing.

[3] Harmon, P., 2010. The Scope and Evolution of Business Process Management, in: , Handbook on Business Process Management 1. Springer, Berlin, Heidelberg, pp. 37–81.

[4] Boyes, H., Hallaq, B., Cunningham, J., Watson, T., 2018. The industrial internet of things (IIoT): An analysis framework. Computers in Industry 101, 1–12.

[5] Ismail, A., Truong, H.-L., Kastner, W., 2019. Manufacturing process data analysis pipelines: a requirements analysis and survey. J Big Data 6 (1).

[6] van der Aalst, W., 2016. Process Mining: Data science in action, 2. edtion ed. Springer, Berlin, Heidelberg, 467 pp.

[7] van der Aalst et al., 2012. Process Mining Manifesto, in: Daniel, Daniel, F., Dustdar, S., Barkaoui, K. (Eds.), Business Process Management Workshops, vol. 99. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–194.

[8] Bayomie, D., Revoredo, K., Bachhofner, S., Kurniawan, K., Kiesling, E., Mendling, J., 2022. Analyzing Manufacturing Process By Enabling Process Mining on Sensor Data.

[9] Martin, N., Fischer, D.A., Kerpedzhiev, G.D., Goel, K., Leemans, S.J.J., Röglinger, M., van der Aalst, W.M.P., Dumas, M., La Rosa, M., Wynn, M.T., 2021. Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi Study. Bus Inf Syst Eng 63 (5), 511–527.

[10] Koschmider, A., Janssen, D., Mannhardt, F., 2020. Framework for process discovery from sensor data. International Workshop on Enterprise Modeling and Information Systems Architectures, EMISA 10, 32–38.

[11] Bano, D., Michael, J., Rumpe, B., Varga, S., Weske, M., 2022. Process-aware digital twin cockpit synthesis from event logs. Journal of Computer Languages 70, 101121.

[12] van Eck, M.L., Sidorova, N., van der Aalst, W.M.P., 2016. Enabling process mining on sensor data from smart products, in: 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS). 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS), Grenoble, France. 01.06.2016 - 03.06.2016. IEEE, pp. 1–12.

[13] Dreher, S., Reimann, P., Gröger, C., 2021. GI Edition Proceedings Band 307 "INFORMATIK 2020" Back to the Future: 28. September - 2. Oktober 2020 Karlsruhe, Deutschland. Köllen, Bonn, CD-ROM.

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

[15] Thiede, M., Fuerstenau, D., Bezerra Barquet, A.P., 2018. How is process mining technology used by organizations? A systematic literature review of empirical studies. BPMJ 24 (4), 900–922.

[16] Sein, Henfridsson, Purao, Rossi, Lindgren, 2011. Action Design Research. MIS Quarterly 35 (1), 37.

[17] Lee, S.L., O'Connor, T.F., Yang, X., Cruz, C.N., Chatterjee, S., Madurawe, R.D., Moore, C.M.V., Yu, L.X., Woodcock, J., 2015. Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. J Pharm Innov 10 (3), 191–199.

[18] van der Aalst, W., Weijters, T., Maruster, L., 2004. Workflow mining: discovering process models from event logs. IEEE Trans. Knowl. Data Eng. 16 (9), 1128–1142.

[19] Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P., 2013. Discovering Block-Structured Process Models from Event Logs - A Constructive Approach, in: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Colom, J.-M., Desel, J. (Eds.), Application and Theory of Petri Nets and Concurrency, vol. 7927. Springer Nature, Berlin, Heidelberg, pp. 311–329.

[20] Carmona, J., van Dongen, B., Weidlich, M., 2022. Conformance Checking: Foundations, Milestones and Challenges, in: van der Aalst, W.M.P., Carmona, J. (Eds.), Process Mining Handbook, vol. 448. Springer International Publishing, Cham, pp. 155–190.

[21] van der Aalst, W.M.P., Carmona, J. (Eds.), 2022. Process Mining Handbook. Springer International Publishing, Cham.

[22] Bertrand, Y., van Belle, R., Weerdt, J. de, Serral, E., 2023. Defining Data Quality Issues in Process Mining with IoT Data, in: Montali, M., Senderovich, A., Weidlich, M. (Eds.), PROCESS MINING WORKSHOPS. Icpm 2022 international workshops, bozen-bolzano, vol. 468. SPRINGER INTERNATIONAL PU, [S.l.], pp. 422–434.

[23] van Zelst, S.J., Mannhardt, F., Leoni, M. de, Koschmider, A., 2021. Event abstraction in process mining: literature review and taxonomy. Granul. Comput. 6 (3), 719–736.

[24] Brzychczy, E., Trzcionkowska, A., 2018. Creation of an Event Log from a Low-Level Machinery Monitoring System for Process Mining Purposes, in: Yin, H., Camacho, D., Novais, P., Tallón-Ballesteros, A.J. (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2018, vol. 11315. Springer International Publishing, Cham, pp. 54–63.

[25] Leoni, M. de, Dündar, S., 2020. Event-log abstraction using batch session identification and clustering, in: Proceedings of the 35th Annual ACM Symposium on Applied Computing, New York, NY, USA. ACM, New York, NY, USA.

[26] Prathama, F., Nugroho Yahya, B., Lee, S.-L., 2021. A Multi-case Perspective Analytical Framework for Discovering Human Daily Behavior from Sensors using Process Mining, in: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain. 12.07.2021 - 16.07.2021. IEEE, pp. 638–644.

[27] Iman, E., Laanaoui, M.D., Sbai, H., 2023. Applying Process Mining to Sensor Data in Smart Environment: A Comparative Study, in: Ben Ahmed, M., Boudhir, A.A., Santos, D., Dionisio, R., Benaya, N. (Eds.), Innovations in Smart Cities Applications Volume 6, vol. 629. Springer International Publishing, Cham, pp. 511–522.

[28] Bose, R.J.C., Mans, R.S., van der Aalst, W.M., 2013. Wanna improve process mining results?, in: 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). IEEE.

[29] Brock, J., Enzberg, D.-I.S. von, Kühn, D.-I.A., Dumitrescu, D.-I.R., 2023. Process Mining Data Canvas: A method to identify data and process knowledge for data collection and preparation in process mining projects. Procedia CIRP 119, 602–607.

Biography

Jonathan Brock (*1995) is a research associate at Fraunhofer Institute for Mechatronics Design IEM in Paderborn, Germany. His research focuses on using process mining in manufacturing settings. He studied industrial engineering in Bielefeld and Magdeburg, acquiring his M.Sc. in 2020.

Niclas Rempe (*1987) is a senior expert at Fraunhofer IEM. His research focuses on digital transformation and process optimization in manufacturing companies. He studied industrial engineering at the University of Paderborn and his doctoral thesis focused on the optimization of purchasing processes in SAP material management and KPI development including process mining in the railway industry.

Sebastian von Enzberg (*1984) leads the Industrial Data Analytics group at Fraunhofer IEM. His research focuses on the application of Data Analytics and Machine Learning in industrial companies. After studying electrical engineering at Otto-von-Guericke University Magdeburg, he completed his PhD on 3D data processing for metrology applications at the Chair of Neuro-Information Technology in December 2018.

Arno Kühn (*1985) is the head of strategic product & organization design at Fraunhofer IEM. With his team, he focuses on guiding medium-sized companies in digitalization and digital transformation. The topics range from digital strategies for product and business to digital platforms and Industry 4.0. Arno Kühn is also part of the management team of the technology network it's OWL - Intelligent Technical Systems OstWestfalenLippe.

Roman Dumitrescu (*1981) is director at the Fraunhofer IEM and head of the department »Advanced Systems Engineering« at the University of Paderborn. His research focuses on the product development of intelligent technical systems. Prof. Dumitrescu is also one of the managing director of it’s OWL. Among others, he is a member of acatech, the expert group Innovation Dialogue of the German Federal Government and the steering committee of the NRW state government’s Initiative Wirtschaft & Arbeit 4.0.