

Framework for Process Discovery from Sensor Data

Agnes Koschmider,¹ Dominik Janssen,² Felix Mannhardt³

Abstract: Process mining can give valuable insights into how real-life activities are performed when extracting meaningful activities instances from raw sensor events. However, in many cases, the event data generated during the execution of a process is at a much lower level of abstraction, and, in some cases, even continuous, e.g., sensor data. This paper presents a framework to discover activities and process models from event location sensor data. The framework is flexible enough to be applied to any data set from raw sensor data.

Keywords: Process Mining, Raw Sensor Data, Event Abstraction, Framework.

1 Introduction

There is a clear need to give deeper insights beyond the execution sequence and performance of formal process activities that are easily extracted from an organization's' enterprise resource planning (ERP) system. In the domain of screen or computer-based work, commercial process mining tools have already introduced task mining ⁴, which combines screen recording and optical character recognition with topic modeling techniques, to identify not only the existence of process problems but automatically give insights into the underlying root causes. However, many processes span both the digital world and the physical world. For example, some parts of a make-to-order manufacturing process will be traceable by extracting events from the ERP and manufacturing execution systems and can be analyzed well with current techniques. Other crucial parts of the process with high impact on the performance, e.g., the activities on the shop-floor, will remain hidden since often there is insufficient data on the individual steps. Thus, inferring what actually happened in the physical world based on raw sensor data is required to fully understand such processes.

This paper presents our framework for mapping of event data to the different levels from event (location) sensor data to a process model in order to understand processes that can be derived from raw sensor data. The framework consists of four modular steps starting with the correlation of events to cases, followed by activity discovery and abstracting activities

¹ Kiel University, Group Process Analytics, Hermann-Rodewald-Str. 3, 24118 Kiel, Germany ak@informatik.uni-kiel.de

² Kiel University, Group Process Analytics, Hermann-Rodewald-Str. 3, 24118 Kiel, Germany dominik.janssen@informatik.uni-kiel.de

³ Dept. of Technology Management, SINTEF Digital, Norway felix.mannhardt@sintef.no

⁴ <https://www.celonis.com/process-mining/what-is-task-mining/>

from events and finally the process discovery step. Each step of the framework will be sketched in this paper.

The paper is structured as follows. The next section lists challenges for process discovery from raw sensor data that should be considered by a framework for process discovery from raw sensor data. Section 3 presents our framework to meet some of these challenges. Related works are summarized in Section 4. The paper ends with a conclusion and discusses future research directions.

2 Challenges for Process Discovery from Raw Sensor Data

To support process discovery from raw sensor data requires to bridge several assumptions that traditional process mining techniques enforce on the structure and semantics of recorded event data. In traditional process mining, i.e., process mining working on recorded events or real-time data of information systems and not with low-level raw data, events shall be (i) available as part of a single, comprehensive dataset (commonly called event log), be (ii) totally ordered, (iii) discrete, (iv) correct and accurate, (v) show an isomorphic relation to individual activity executions and (vi) relatively high level (i.e., close to the business-level like "truck loading ended"). In the context of raw sensor data events can be any kind of observations (e.g., a sensor value changed) with relevance to a certain process. To apply process discovery from raw sensor data requires to cope with these assumptions of traditional process mining. For this, new techniques are required that put emphasis of the following:

1. *no single, comprehensive dataset*: interconnected devices like large-scale sensor networks or mobile devices emit events at distributed, potentially not even stationary sources representing low-level raw data (e.g., raw RFID readings). Thus, events are distributed. Solutions would help to overcome assumptions (i) and (ii).
2. *no discrete events*: usually, process mining works with recorded, discrete events or real-time data of information systems and not with low-level raw data [KMH18]. Process mining algorithms expect event logs, that organizes the events in terms of traces at a certain level of abstraction, which does not match the low-level concept of raw sensor event data [So19]. Related solution would help to overcome the assumption (iii).
3. *no correct event log*: events that are composed of an observation of an activity accompanied by time information extracted from sensor data are mostly affected by noise, inaccurate measurements and ambiguous information resulting in an uncertain event log. This point is related to the assumption (iv).
4. *no isomorphic relation to activity executions*: sensing or observations of the physical world generate low-level events. The translation of low-level data into high-level data, known as *event abstraction* requires appropriate data aggregations. Beside

this, the low-level event data is tracked in continuous form requiring new techniques for event abstraction. For further reading we refer to [KMH18], which sketches a schematic overview of the hierarchy of low-level events to process model discovery. The state-of-the-art of event abstraction is summarized in [Ze20]. Solutions shall help to overcome (v) and (vi).

In the next section we present our approach coping with the fourth issue.

3 Framework for Process Discovery from Raw Sensor Data

The discovery of process models from raw sensor data requires several data aggregations. Our framework for mapping to the different levels of event data is presented in Fig. 1, which incorporates such data aggregations. Our framework assumes a sensor event log as input

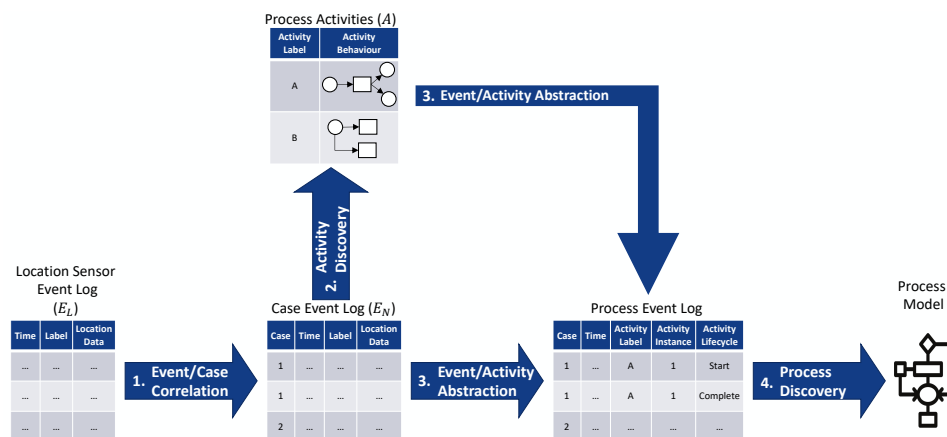


Fig. 1: Mapping to the different levels of event data.

derived from e.g., networks of WiFi-access points, or motion sensors in smart homes and smart factories. Event data like this does not necessarily have a unique identifier attached to it to identify who triggered the event. One component of the framework allows to detect multiple entities (i.e., people or objects) at the same time. For this, it has to be ensured that all activities are associated with the correct entity. To abstract activities from events machine learning techniques are used. To provide a solution for process discovery from raw sensor events, our framework consists of four steps:

1. *Event Correlation:* In this step raw sensor events are grouped to a set of (unlabelled) activity instances by correlating each individual sensor event to an activity instance. The output of this first module is a *log* in which beyond the requirements for a sensor event log each event is additionally assigned an activity instance that can be retrieved within the set of instances.

2. *Activity Discovery*: In the second step process activities are discovered together with their labels and sensor-level process models describing the expected behaviour on a sensor level. To use our approach in scenarios with multiple entities, a mechanism is incorporated distinguishing persons from objects. The approach bases on the relative distance of the event (location) data.
3. *Event Abstraction*: Abstraction of the log (see step 1) to a process event log where events are directly related to the *start* or *completion* of process activities. In particular, sub-traces from the *event / case correlation* (step 1) are combined with activity clusters detected within the second step *Activity Discovery*. Once the combination is complete, the result of this step is a process event log. It specifies every event that has been recorded in the location sensor event log as an activity. These activities have the following tags allocated to them: the name of the activity (label), the entity that is executing the activity (instance) and the information if the activity has been started or completed (life cycle).
4. *Discovery of Processes*: Having promoted the raw sensor events to the level of activity instances, we still need to identify which process cases are meaningful to our analysis target. We have to assume that regular and routine behaviour can be observed. As a starting point an activity has to be selected that most likely will be the origin of the routine behaviour. Finally, a process model is discovered using a standard technique. The final output is a process model reflecting the observed behaviour of the entities aggregated only from raw sensor data.

We evaluated our approach on the publicly available CASAS data-set, which contains raw sensor data from a smart home environment [Ra07]. The data set refers to a smart home test-bed with two residents and a house equipped with 51 motion sensors. To discover activities and process models from motion sensors, a self-organizing map (SOM) has been used as unsupervised learning in form of clustering and process mining. Our framework can be applied to any other arbitrary data set if the minimum requirement of time and location data is met.

4 Related Work

A large body of research exists that addresses the discovery of events and activities at different levels (see Fig. 1). Particularly, approaches can be found for activity recognition from low-level events addressing recognition on different sensor types like motion or video [Di18; NS10; RMM16; ZS12] or analyzing data from wearable [Ci19; Sz16] or reference sensors [LRP15] where machine learning is used for data aggregation. Our framework allows a process model view on raw sensor data, which advances existing approaches and is beneficial in terms of evaluating the quality of data aggregations, which machine learning is not capable of.

Related to our framework are approaches for activity recognition for process mining. The technique of [ESA16] assumes the low-level event data to be discrete and expects particular representations of traces. This does not apply for our framework.

5 Conclusion

The rise of Internet-of-Things (IoT) gradually changes the granularity level at which *event data*, captured during the execution of operational processes, is stored. While in the past process models have been discovered from documents or interviews with domain experts, which is error-prone and time-consuming, the challenge will be to automatically discover the *de-jure* process model coined under the umbrella of process mining from large-scale sensor networks or wearable devices [Ja17]. Process mining can give valuable and new insights into how real-life activities in the physical world are performed when meaningful activities instances are extracted from raw sensor data. In this paper we presented a framework for process discovery from raw sensor data consisting of data aggregations steps. The framework gives guidelines how to discover high-level events from raw (location) event data and how to discover processes from activity instances that were derived from high-level events.

Further research tasks for this paper are:

1. We plan to conduct empirical studies that compare various clustering techniques for data aggregations. We plan to apply different machine learning and deep learning techniques for abstracting activity instances from event data and comparing them in terms of accuracy, scalability and efficiency.
2. The recognition of entities (person vs. object) from raw sensor data also requires improvement in order to discover accurate process models. We plan to use different approaches for entity recognition and allocation to activities and compare them. Hidden Markov Models and other machine learning-based approaches could remedy the situation.
3. To foster a responsible use of event logs we plan to consider privacy issues when abstracting activities from event data [Vo20].

References

- [Ci19] Civitarese, G.: Human Activity Recognition in Smart-Home Environments for Health-Care Applications. In: 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). Pp. 1–1, Mar. 2019.

- [Di18] Diète, A.; Szt Tyler, T.; Weiland, L.; Stuckenschmidt, H.: Improving Motion-based Activity Recognition with Ego-centric Vision. In: PerCom Workshops 2018. IEEE Computer Society, pp. 488–491, 2018, ISBN: 978-1-5386-3227-7, URL: <https://doi.org/10.1109/PERCOMW.2018.8480334>.
- [ESA16] van Eck, M. L.; Sidorova, N.; van der Aalst, W. M. P.: Enabling process mining on sensor data from smart products. In: RCIS 2016. Pp. 1–12, 2016, URL: <https://doi.org/10.1109/RCIS.2016.7549355>.
- [Ja17] Janiesch, C.; Koschmider, A.; Mecella, M.; Weber, B.; et al.: The Internet-of-Things Meets Business Process Management: Mutual Benefits and Challenges. CoRR abs/1709.03628/, 2017, arXiv: 1709.03628, URL: <http://arxiv.org/abs/1709.03628>.
- [KMH18] Koschmider, A.; Mannhardt, F.; Heuser, T.: On the Contextualization of Event-Activity Mappings. In: BPM 2018 Workshops. LNBIP, Springer, Sept. 2018.
- [LRP15] Larue, G. S.; Rakotonirainy, A.; Pettitt, A. N.: Predicting Reduced Driver Alertness on Monotonous Highways. IEEE Pervasive Computing 14/2, pp. 78–85, Apr. 2015, ISSN: 1536-1268.
- [NS10] Nentwig, M.; Stamminger, M.: A method for the reproduction of vehicle test drives for the simulation based evaluation of image processing algorithms. In: 13th International IEEE Conference on Intelligent Transportation Systems. Pp. 1307–1312, 2010.
- [Ra07] Rashidi, P.; Youngblood, G. M.; Cook, D. J.; Das, S. K.: Inhabitant Guidance of Smart Environments. In: HCI (2). Vol. 4551. LNCS, Springer, pp. 910–919, 2007.
- [RMM16] Ranasinghe, S.; Machot, F. A.; Mayr, H. C.: A review on applications of activity recognition systems with regard to performance and evaluation. International Journal of Distributed Sensor Networks 12/8, p. 1550147716665520, 2016, eprint: <https://doi.org/10.1177/1550147716665520>, URL: <https://doi.org/10.1177/1550147716665520>.
- [So19] Soffer, P.; Hinze, A.; Koschmider, A.; Ziekow, H.; Ciccio, C. D.; Koldehofe, B.; Kopp, O.; Jacobsen, A.; Sürmeli, J.; Song, W.: From event streams to process models and back: Challenges and opportunities. Information Systems 81/, pp. 181–200, 2019, ISSN: 0306-4379, URL: <http://www.sciencedirect.com/science/article/pii/S0306437917300145>.
- [Sz16] Szt Tyler, T.; Carmona, J.; Völker, J.; Stuckenschmidt, H.: Self-tracking Reloaded: Applying Process Mining to Personalized Health Care from Labeled Sensor Data. T. Petri Nets and Other Models of Concurrency 11/, pp. 160–180, 2016.

- [Vo20] von Voigt, S. N.; Fahrenkrog-Petersen, S. A.; Janssen, D.; Koschmider, A.; Tschorsch, F.; Mannhardt, F.; Landsiedel, O.; Weidlich, M.: Quantifying the Re-identification Risk of Event Logs for Process Mining - Empirical Evaluation Paper. In (Dustdar, S.; Yu, E.; Salinesi, C.; Rieu, D.; Pant, V., eds.): Advanced Information Systems Engineering - 32nd International Conference, CAiSE 2020, Grenoble, France, June 8-12, 2020, Proceedings. Vol. 12127. Lecture Notes in Computer Science, Springer, pp. 252–267, 2020, URL: https://doi.org/10.1007/978-3-030-49435-3%5C_16.
- [Ze20] van Zelst, S.; Mannhardt, F.; de Leoni, M.; Koschmider, A.: Event abstraction in process mining: literature review and taxonomy. *Granular Computing*, 2020, URL: <https://doi.org/10.1007/s41066-020-00226-2>.
- [ZS12] Zhang, M.; Sawchuk, A. A.: Motion Primitive-based Human Activity Recognition Using a Bag-of-features Approach. In: Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium. IHI '12, ACM, pp. 631–640, 2012.