Internet of Processes and Things: A Repository for IoT-Enriched Event Logs in Smart Environments

Matthias Ehrendorfer[1], Yannis Bertrand[2], Lukas Malburg[3,4], Juergen Mangler[1], Joscha Grüger[3,4], Stefanie Rinderle-Ma[1], Ralph Bergmann[3,4] and Estefania Serral Asensio[2]

[1]*Department of Computer Science, School of Computation, Information and Technology, Technical University of Munich, 85748 Garching, Germany*

[2]*Research Centre for Information Systems Engineering (LIRIS), KU Leuven, Warmoesberg 26, 1000 Brussels, Belgium*

[3]*Artificial Intelligence and Intelligent Information Systems, University of Trier, 54296 Trier, Germany*

[4]*German Research Center for Artificial Intelligence (DFKI), Branch University of Trier, 54296 Trier, Germany*

Abstract

Current research efforts from the Business Processes Management and IoT communities, touching novel techniques such as object-centric process mining, predictive process monitoring or IoT based process enhancement suffer from a lack of real-world open data-sets from different domains to evaluate and refine their approaches. In this paper we present an online resource for the collection of IoT-enhanced process logs, hoping to inspire companies, organisations and researchers to cooperate in the release of well-understood high-quality data sets.

Keywords

IoT-Enriched Event Logs, Business Process Management, IoT

1. Introduction

The rise of the *Internet of Things (IoT)* is leading more and more organizations to use IoT devices to monitor and automate their Business *Processes.* Prominent domains increasingly embracing process-based automation include the manufacturing domain (i.e., Manufacturing Processes) [1, 2], the Health Care domain (i.e, lab automation) [3], and the transportation and logistics domain [4]. In all of these domains IoT devices, i.e. sensors or machines, easily generate big amounts of data: e.g., 10 sensors at 1 measurement per second record 315 million data points over one year. While the automation of business processes typically generates (control-flow) data documenting how and in which order machines (tasks) are invoked, as well as results and errors produced by this invocation, IoT sensors often produce data which is complementary.

BPM 2023: Demos and Resources

matthias.ehrendorfer@tum.de (M. Ehrendorfer); yannis.bertrand@kuleuven.be (Y. Bertrand); malburgl@uni-trier.de (L. Malburg); juergen.mangler@tum.de (J. Mangler); grueger@uni-trier.de (J. Grüger); stefanie.rinderle-ma@tum.de (S. Rinderle-Ma); bergmann@uni-trier.de (R. Bergmann); estefania.serralasensio@kuleuven.be (E. Serral Asensio)

0000-0002-7739-9123 (M. Ehrendorfer); 0000-0002-6407-7221 (Y. Bertrand); 0000-0002-6866-0799 (L. Malburg); 0000-0002-6332-5801 (J. Mangler); 0000-0001-7538-1248 (J. Grüger); 0000-0001-5656-6108 (S. Rinderle-Ma); 0000-0002-5515-7158 (R. Bergmann); 0000-0001-7579-910X (E. Serral Asensio)

IoT sensors document the progression of states while process tasks or instances of processes are running. Due to their granularity and loose link with the control-flow [5], these data cannot be analyzed with traditional *Process Mining (PM)* techniques, requiring the development of specific *IoT-enhanced* PM approaches [6].

This being said, researchers require openly available IoT-enhanced diverse real-world data sets to benchmark and ensure the generalizability of new techniques. So far, most researchers have either used publicly available smart spaces data sets, e.g., data sets from the CASAS project[1], van Kasteren data set[2] or proprietary data sets. The first suffers from the caveats that smart spaces logs do not track proper BPs (most of the time, human habits are observed) and, as a result, they do not always contain control-flow information. The latter often have the advantage of tracking real-life processes, e.g., in manufacturing [7, 8], but they cannot be used by the community as benchmark and are restricted to one domain of application.

Data sets which include traditional event based process logs, fine-grained IoT sensor data, as well as well understood and tight coupling between those data sets are rare, both, because of the (1) reluctance of organisations to release real-world data sets, and (2) because of the high effort needed to link the different types of information together. We strive to form a community of like-minded people, to share their tightly coupled real-world data sets, which contain both, event based process data, as well as IoT data. In this paper, we present a curated public resource repository ...

https://zenodo.org/communities/iopt/

... containing IoT-enhanced real-world data sets following a published data model grounded on the well-established XES standard [9]. It contains a set of logs from the above mentioned domains, to allow researchers to explore novel analysis approaches.

In the following, Sect. 2 describes how the IoT-enriched event logs can be stored to improve the analysability of the data, including minimal requirements. A detailed description of one currently contained data set is given in Sect. 3. A discussion regarding the usefulness of the data sets is provided in Sect. 4, which also summarizes our efforts and aspirations for the resource repository.

2. Foundations

In this section, we briefly present the foundations for this work. As we focus on using the DataStream XES extension [8], we present the data structure used for event logs in our Internet of Processes and Things repository[3] in Sect. 2.1. In addition, we present the minimal requirements which must be satisfied to submit data sets to the repository (see Sect. 2.2). In this context, we do not restrict the data set to be represented in the DataStream format. Thus, other suitable representation formats can be used, such as XES+DataStream [7], OCEL[4], XES [9], OCED[5], provided they satisfy the requirements listed in Sect. 2.2.

[1]https://casas.wsu.edu/datasets/
[2]https://ailab.wsu.edu/mavhome/research.html
[3]https://zenodo.org/communities/iopt
[4]https://ocel-standard.org/
[5]https://www.tf-pm.org/resources/oced-standard

2.1. DataStream Representation Format

The DataStream extension [8] enables the recording of IoT data in connection with process events. To do this, the eXtensible Event Stream (XES) is extended with the following concepts:

- `stream:point`: Represents a single IoT data artefact (measurement) including its ID, timestamp, value, source and meta-data.
- `stream:multipoint`: Compresses information from multiple points sharing a timestamp, source, ...
- `stream:datastream`: Connects multiple `stream:points` and/or `stream:multipoints` to events or traces.
- `stream:datacontext`: Connects multiple `stream:datastreams` and/or `stream:datacontexts` to a set of traces and/or events.

Overall, this allows integration of IoT data into process logs in a structured way and enable analysis of the data in the context of the process in which it was collected. In the next section, we introduce a sample data set included in the presented repository that provides process logs of a process where chess pieces are manufactured and measured using a machine tool and a measuring machine. Besides this process log, there are several other logs available from other application scenarios and domains, such as public transport or production of sheet metals.

2.2. Minimal Requirements for Submitted Data Sets

Data sets in the repository do not necessarily have to be in the DataStream format (see Sect. 2.1) or contain only real-world data. However, they must fulfill the following requirements:

- The data set must contain information about the steps performed in the process (i.e., include the process events which occurred during the execution of the process).
- The data set must contain IoT data collected during the process execution or which is relevant in the context of the process.
- The IoT data must be related to the process execution. Therefore, IoT data should be linked with the corresponding process activities, i. e., IoT data might contribute to one or more process events.
- The data set must contain a description of the process model either explicitly (e.g., as a BPMN model etc.) or implicitly (e.g., the log contains the initial process model as well as potential changes to this model).
- It must be clearly stated if the data set represents real-world or artificial data.
- The data set must be made available under the Creative Commons Attribution 4.0 International license[6] to allow open and free research for the community.

3. Description of the Manufacturing Data Set

The "XES Chess Pieces Production" data set[7] contains the log of a manufacturing process for a chess piece. It includes (1) event data obtained when enacting the process model and (2) data

[6]https://creativecommons.org/licenses/by/4.0/
[7]https://doi.org/10.5281/zenodo.7958478

streams encountered during the execution of this process. The log is created according to the DataStream XES extension [8], which enables the recording of IoT data in connection with process events. Data analysis can then be performed on this fully contextualized data without the need to perform a pre-processing step where the collected data is connected to the tasks performed in the process.

In the process chess pieces are produced, measured directly afterwards, and put on a tray (see Fig. 1). The recorded data streams are collected from (1) an EMCO MT45 Lathe providing data from its standard internal sensors as well as custom power measurements, (2) a Keyence LS-7000 High-speed, High-accuracy Optical Digital Micrometer providing measurement data for the diameter of the part while it is moved through the measuring machine, and (3) an ABB IRB-2600 Industrial Robot providing movement coordinates and the states of the pneumatic valves controlling the gripper.

Figure 1: Example Data Set Work Items

The data set contains (1) a simple human readable list of contained (sub-)process instances (“index.txt”) showing the structure of the logs (i.e., sub-process instances are indented in relation to their parents - see Lst. 1), (2) a file “pallet.jpeg” which is an image of the 9 manufactured parts where some are wrapped in chips from the turning process, and (3) several “[UUID].xes.yaml” files which are YAML[8] logs of the executed process models conforming to the eXtensible Event Stream (XES) format and containing events as described in the DataStream XES extension [8]. Furthermore, logs contain the actual process model in the CPEE RPST format in the “cpee:lifecycle:transition” “description/change” events.

Listing 1: Extract of “index.txt” file

```
Turm Batch Processing V2 (ca2328b4-2831-431f-af1d-e187ff267f72) - 5541
  X MT45 Control Getter (1ba78e0f-acfd-4fd5-a59a-da5b575abead) - 5564
  Turm Single wr04 (14eb245e-9aba-446f-9624-2f110624dc65) - 5565
    X Generate NC (f9b0528d-cd74-47a3-83d9-2eb2fc129dcb) - 5567
  [...]
```

4. Discussion of Use-Cases, Upcoming Data Sets and Conclusion

The information contained in the data sets provided in the resource repository allows applying PM [10] techniques (such as process discovery and conformance checking) to smart environments. Analysis which has been performed on earlier versions of the resources contained in the presented repository include the work by Stertz et al. [11] which analyzes process concept drifts based on sensor event streams and the work by Grüger et al. [7] that investigates data quality issues during event log generation. Moreover, PM [10] techniques can be applied in smart environments [12] such as manufacturing [13] to check conformance w. r. t. the given process model or to adapt and optimize processes when runtime failures occur.

[8] https://yaml.org/

We hope that our efforts will spark the publication of more real-world data sets. We also hope to inspire a move from the publication of collections of unrelated pieces of data linked together by documentation, towards the publication of tightly coupled and formally linked data sets to decrease the hurdle of using such data sets for the demonstration and evaluation of novel algorithms handling connected IoT and control-flow data, i.e., considering and integrating both perspectives for improved performance and quality of mining and analysis results. A second focus is to provide data sets conforming to the DataStream XES extension [8] in the OCEL format to ease the transition for scientific and industrial researchers.

References

[1] F. Pauker, J. Mangler, S. Rinderle-Ma, M. Ehrendorfer, Industry 4.0 Integration Assessment and Evolution at EVVA GmbH: Process-Driven Automation Through centurio.work, Springer, 2021, pp. 81–91.
[2] L. Malburg, M. Hoffmann, R. Bergmann, Applying MAPE-K control loops for adaptive workflow management in smart factories, J. Intell. Inf. Syst. (2023) 1–29.
[3] S. Holzmüller-Laue, B. Göde, K. Thurow, Model-driven complex workflow automation for laboratories, in: 2013 IEEE International Conference on Automation Science and Engineering (CASE), IEEE, 2013, pp. 758–763.
[4] P. Grefen, H. Ludwig, S. Tata, R. Dijkman, N. Baracaldo, A. Wilbik, T. D'hondt, Complex collaborative physical process management: A position on the trinity of bpm, iot and da, in: 19th IFIP WG 5.5 Working Conference on Virtual Enterprise, Springer, 2018, pp. 244–253.
[5] C. Janiesch, et al., The Internet of Things Meets Business Process Management: A Manifesto, IEEE Syst. Man Cybern. Mag. 6 (2020) 34–44.
[6] Y. Bertrand, J. De Weerdt, E. Serral, Assessing the suitability of traditional event log standards for iot-enhanced event logs, in: Business Process Management Workshops: BPM 2022 International Workshops, Münster, Germany, September 11–16, 2022, Revised Selected Papers, Springer, 2023, pp. 63–75.
[7] J. Grüger, L. Malburg, R. Bergmann, IoT-enriched event log generation and quality analytics: a case study, it - Information Technology (2023).
[8] J. Mangler, J. Grüger, L. Malburg, M. Ehrendorfer, Y. Bertrand, J.-V. Benzin, S. Rinderle-Ma, E. Serral Asensio, R. Bergmann, DataStream XES Extension: Embedding IoT Sensor Data into Extensible Event Stream Logs, Future Internet 15 (2023).
[9] C. W. Günther, E. Verbeek, XES Standard Definition - Version 2.0, 2014.
[10] W. M. P. van der Aalst et al., Process Mining Manifesto, in: BPM Workshops, Springer, 2012, pp. 169–194.
[11] F. Stertz, S. Rinderle-Ma, J. Mangler, Analyzing process concept drifts based on sensor event streams during runtime, in: 18th BPM, Springer, 2020, pp. 202–219.
[12] F. Leotta, M. Mecella, J. Mendling, Applying Process Mining to Smart Spaces: Perspectives and Research Challenges, in: 27th CAiSE Workshops, LNBIP, Springer, 2015, pp. 298–304.
[13] S. Rinderle-Ma, J. Mangler, Process Automation and Process Mining in Manufacturing, in: BPM, volume 12875, Springer, 2021, pp. 3–14.