

A Time-Series Process Event Log Preprocessing Approach for Data-Intensive and Predictive Operationalization of Smart Factories

Minjae Park*, Sohyang Park**, Thanh-Hai Nguyen**, Dinh-Lam Pham***, Dongkeun Oh****, Kyoung-Sook Kim****, In-Kyoo Chun****, Kwanghoon Pio Kim**

*Div. of Computer Information, DAELIM UNIVERSITY COLLEGE, Anyang Gyeonggi, 13916, Republic of Korea

**Div. of AI Computer Science & Engineering, KYONGGI UNIVERSITY, Suwon Gyeonggi, 16227, Republic of Korea

***Contents Convergence Software Research Institute, KYONGGI UNIVERSITY, Suwon Gyeonggi, 16227, Republic of Korea

****Smart Factory Research Institute, IC&IT Co., Guro Seoul, 08377, Republic of Korea

mjpgark@daelim.ac.kr, {sa7291, nguyenthanhhai, phamdinhlam, kwang,}@kgu.ac.kr, {dkoh, kskim, taz}@icnit.co.kr

Abstract—This paper proposes a time-series process event log preprocessing approach applied to realize data-intensive and predictive operationalization of IoT-supported smart factories. The advanced conceptual approach of the data-intensive and predictive operationalization processes is newly proposed for effectively and predictively operating very large-scale smart factories that are supported by the Internet of Things platforms with temporal operationalization processes and their time-series event log datasets. The authors' research group is trying to develop a predictive smart factory operationalization system, which is characterized by the advanced conceptual approach, and the starting point of which is at acquiring the time-series process event log datasets with a proper data preprocessing approach. Hence, this paper devises a time-series process event log preprocessing approach with its algorithmic details to arrange the proper datasets to be used for concretizing the data-intensive and predictive operationalization in the very large-scale smart factories. Finally, we develop an XES-formatted event log preprocessing system based upon the devised preprocessing algorithms and apply to a real and nation-wide production smart factory in order to validate the operability and feasibility of the developed system.

Index Terms—Smart Factory, Internet of Things, Data-Intensive Smart Factory, Operationalization Process Automation, Digital Twin Platform, IoT-Sensed Dataset, Time-Series Dataset

I. INTRODUCTION

In recent, the smart factory arena has been faced with a requirement of great urgency that fosters equipping such IoT-centric intelligent sensors and actuators [1], installing the operational processes automation [2] and their event log functionality, and adopting the creative big-data mining and deep learning technologies for realizing the proactive production and operationalization intelligence [1] on the corresponding smart factories. In accordance with gratifying the urgent requirement, the research group of the co-authors has built a challengeable R&D project through the ATC+ program of KEIT, and the goal of which is to adopt the deep-learning-based predictive production and operationalization process

methodology as an impeccable silver-bullet for realizing the production level-4 smart factory automation onto the textile industries. The essential functions of the methodology are about modeling, operating, managing, monitoring, and logging the production and operationalization processes and their execution histories in smart factories. The eventual goal of these functions is for organizing an intelligent and predictive smart factory platform being fortified with diverse pieces of data-intensive equipment and their time-series process mining and deep-learning technologies. This paper is particularly about the logging and preprocessing approach to be applied to the IEEE XES-formatted event log datasets [5] collected from the execution histories of the smart factory operationalization processes.

So far, there are several research works considering the log structures and formats of smart factory process operationalization and enactment event log datasets, such as XWELL¹ [3], BPAF² [4], and XES³ [5]. The most appropriate format ought to be the IEEE XES standardized event log stream format. In recent, IEEE has released a standard tag-based language of XES that aims at providing a unified and extensible event streaming methodology for capturing systems' behaviors by means of event logs and event streams. The IEEE XES ought to be now the most popular format. In other words, almost all the event stream datasets, which are now available in the literature, are based upon the "XES Schema [5]" describing the structure of an XES event log/stream and the "XES extension" describing the structure of an extension of such a log/stream. Based upon the format of the extended XES schema, this paper proposes a schema structure and its preprocessing algorithms of the production and operationalization processes' enactment event logs in those data-intensive IoT-based smart factories.

¹XWELL stands for XML-based workflow execution logging mechanism and language.

²BPAF stands for business process audit format, which is supported by the Workflow Management Coalition (WfMC).

³XES stands for extensible event stream, which is supported by the extensible event stream working group of the IEEE standard committee.

This research was supported by the ATC+ RD Project of KEIT (Korea Evaluations Institute of Industrial Technology) from the Ministry of Trade, Industry and Energy in 2020 (Grant No. 20009228).

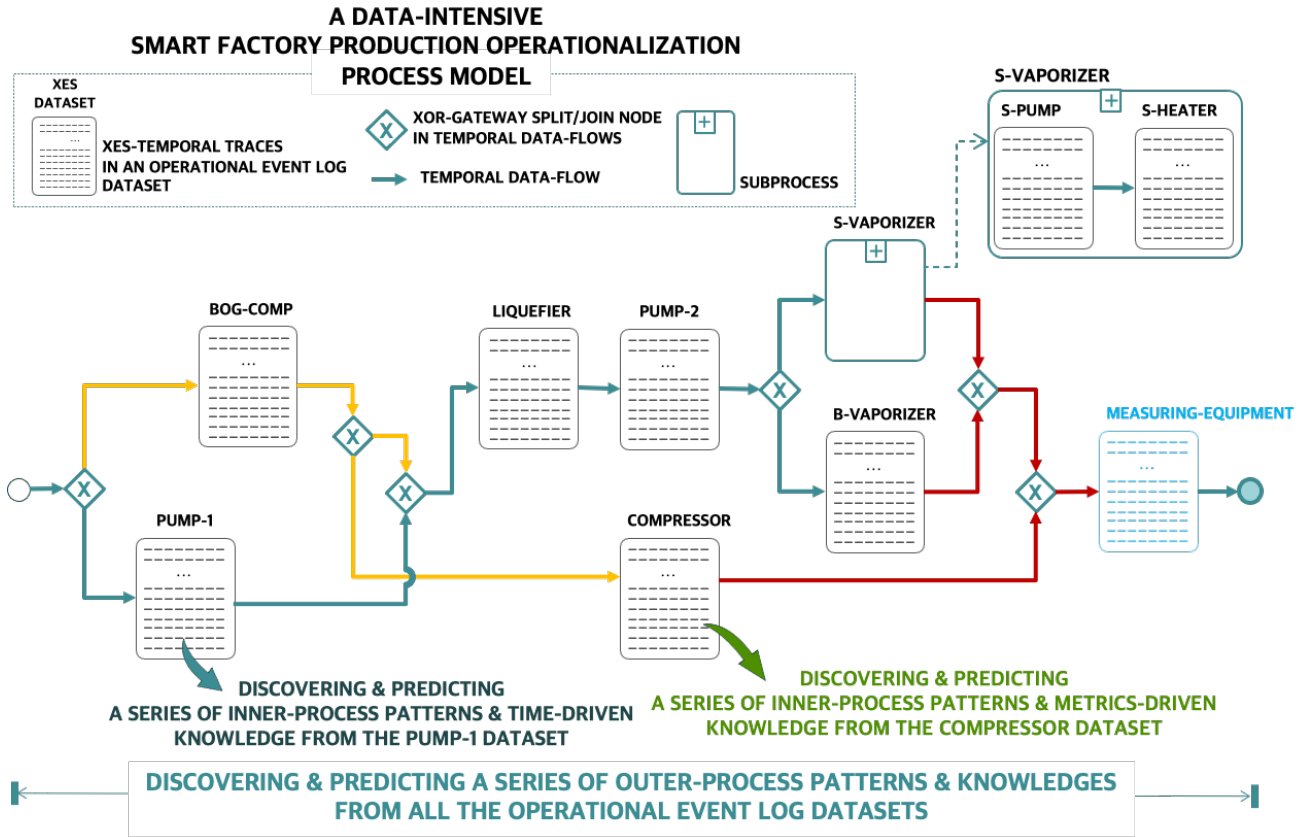


Fig. 1. A Conceptual Formation of Process-Aware Datacubes from Workflow Enactment Event Log Data Centers

II. DATA-INTENSIVE OPERATIONALIZATION PROCESSES

The smart factory productional and operational processes are theoretically supported by the data-intensive concept of operationalization process modeling methodology that is theoretically supported by the information control net model [2] [6]. There are two types of operationalization process patterns in the data-intensive methodology: Inner-Process Pattern and Outer-Process Pattern. The inner-process pattern is for modeling the time-series event stream traces based upon the timestamp inside of the dataset collected from a smart factory IoT-sensor, while on the other hand the outer-process pattern is for modeling the factory-wide production and operationalization processes associated with the whole of equipment and its equipped IoT-sensors in a smart factory [7], as shown in Fig. 1.

In the figure, the smart factory production and operationalization outer-process model is theoretically based upon the graph model of the information control net and follow the business process modeling notation standardized by OMG's BMI-DTF in terms of the graphical notation. The small circles positioned as the initiation node and the final node represent the events of the start and the end of the model. The boxes represent the sensor-datasets collected from the devices and facilities of equipment built in the corresponding smart factory. The blank-box with containing a +-marked small-box especially represents the subprocess model containing other model. Finally, these various types of nodes are connected

by the directed arcs with three types of gateway nodes (sequential, disjunctive exclusive-OR, and conjunctive parallel-AND) implying the production-flow. Particularly, the special gateway nodes of disjunctive and conjunctive production-flow are graphically represented by shapes of diamond containing X-mark and +-mark that are exclusive-OR production-flow and parallel-AND production-flow, respectively. Also, it is assumed that the data schema structure of each sensor-dataset is formatted in the IEEE- XES standardized event stream format and that various time-series groups of the inner-processes can be discovered from each sensor-dataset. Note that the detailed schema format is not described in the paper due to the page limitation.

III. OPERATIONALIZATION INNER-PROCESS BIGDATA PREPROCESSING ALGORITHMS

Each sensor-dataset collected from a corresponding device of the operationalization process is structured in the IEEE-XES [5] standardized event log format. We develop a preprocessing algorithm that is able to discover two types of inner-processes with temporal traces from the XES-formatted sensor-dataset, as illustrated in Fig. 2. One is the time-driven temporal traces of inner-processes and the other is the metrics-driven temporal traces of inner-processes. The left-hand side of the figure depicts the schema structure of the IEEE-XES format. The operational process event log consists of a group of temporal traces, and each temporal trace consists of a group

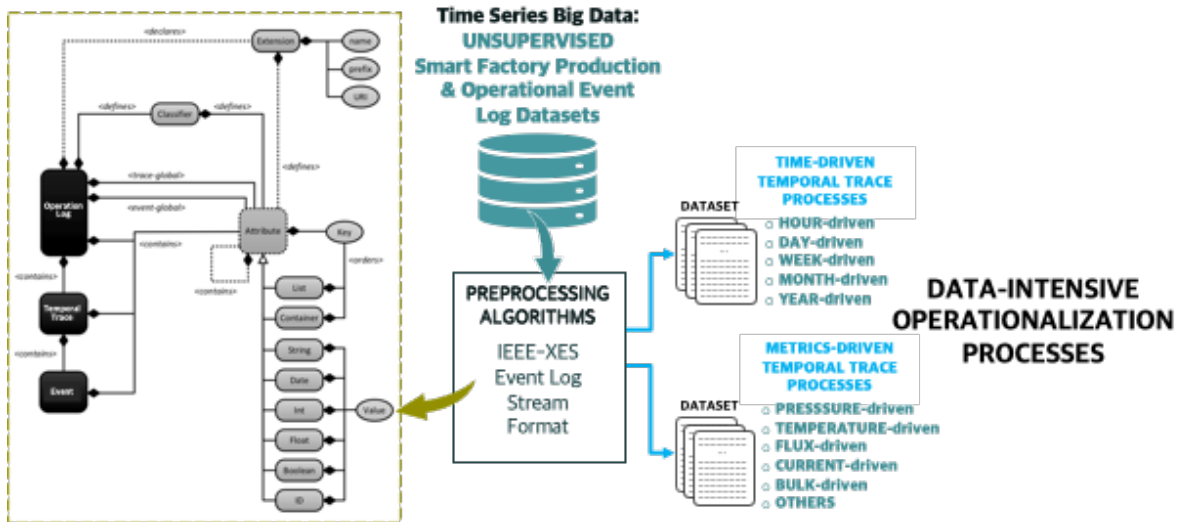


Fig. 2. The Data-Intensive Operationalization Temporal Trace (Inner-Processes) Datasets and their Preprocessing Algorithms

of events that contains various attributes characterized by the typical elementary data types and the compound data types, as well. The preprocessing algorithms devised in this paper are classified into two classes of discovery functions. One class is able to discover the time-driven inner-processes such as hour-driven, day-driven, week-driven, month-driven, and year-driven temporal traces, the other class is able to discover the metrics-driven inner-processes such as pressure-driven, temperature-driven, flux-driven, current-driven, bulk-driven, and other sensor types of temporal traces.

The **Algorithm 1** represents the inner-process discovery and visualization algorithm that is able to discover a class of inner-processes and visualize their temporal traces from a single IEEE-XES formatted sensor-dataset. The sensor-dataset is basically a sort of time-series dataset and each temporal trace so has the timestamp property with its own device's characteristics, such as pressure, flux, current, temperature values, installed in the corresponding smart factory. Each of these properties are structured with a pair of event-tags within a single trace-tags.

Algorithm 1: Discovery and Visualization of the Inner-Processes from XES Files

Input: XES file Ψ , device property μ

Output: Discovered inner-processes and visualization of each temporal trace Θ out of the inner-processes

```

1 Initialize  $\theta = \emptyset$ , listOfDevice =  $\emptyset$ , dataType =  $\emptyset$ 
2 globalTrace =  $\Psi.ReadGlobalTraceInfo()$ 
3 globalEvent =  $\Psi.ReadGlobalEventInfo()$ 
4 for  $\forall \eta$  in globalTrace do
5   currentDeviceInfo =  $\eta.GetDevice()$ 
6   listOfDevice.AddDevice(currentDeviceInfo)
7 for  $\forall \eta$  in dataType do
8   currentTypeInfo =  $\eta.TypeInfo()$ 
9   dataType.AddType(currentTypeInfo)
```

```

while (row =  $\Psi.readline()$ ) != null do
  isTag = false
  eventInfo =  $\emptyset$ 
  if row = < trace > then
    beginTrace = true
    isTag = true
     $\theta.AddNewTrace()$ 
  if row = < /trace > then
    endTrace = true
    isTag = true
  if row = < event > then
    beginEvent = true
    isTag = true
     $\theta.AddNewEvent()$ 
  if row = < /event > then
    endEvent = true
    isTag = true
  if isTag == false then
    eventInfo = row.GetValue()
     $\theta.AddEventInfo(eventInfo)$ 
```

```

for  $\forall \zeta$  in  $\theta$  do
  currentTrace =  $\zeta$ 
  currentDevice = currentTrace.GetDeviceInfo()
  if currentDevice ==  $\mu$  then
     $\Theta.AddTrace()$ 
    for  $\forall \vartheta$  in currentTrace do
      currentEvent =  $\vartheta.GetEventInfo()$ 
      currentProperties = currentEvent.GetProperties()
       $\Theta.AddNode(currentEvent)$ 
       $\Theta.AddSubNode(currentProperties)$ 
       $\Theta.AddEdge(currentProperties$ 
       $\Rightarrow$ 
       $currentEvent)$ 
    break
  Return  $\Theta$ 
```

IV. CONCLUSION

In this paper, we have proposed a time-series event log preprocessing approach for discovering a group of operationalization inner-processes from an XES-formatted sensor-dataset collected from executing a smart factory operationalization process. Especially, the proposed preprocessing approach with its algorithms can discover two types of inner-processes with temporal traces; One is the time-driven temporal traces of inner-processes and the other is the metrics-driven temporal traces of inner-processes. Conclusively, through the proposed preprocessing approach, we can realize the novel concept of data-intensive and predictive operationalization of smart factories, which ought to be an eventual solution for concretizing the data-intensive digital twin of smart factories.

ACKNOWLEDGMENT

This research was supported by the ATC+ RD Project of KEIT (Korea Evaluations Institute of Industrial Technology) from the Ministry of Trade, Industry and Energy in 2020 (Grant No. 20009228, Development of Deep Learning-based Predictive Production Quality and Field Safety Management Technology to Support Level-4 Smart Factories in the Textile Industries).

REFERENCES

- [1] M. Kim, H. Ahn and K. P. Kim, "Process-Aware Internet of Things: A Conceptual Extension of the Internet of Things Framework and Architecture," *KSII Transactions on Internet and Information Systems*, Vol. 10, No. 8, pp. 4008-4022, 2016. DOI: 10.3837/tiis.2016.08.032.
- [2] Ellis, Clarence A., Kim, K., Rembert, A., Wainer, J., "Investigations on Stochastic Information Control Nets," *INFORMATION SCIENCES*, Vol. 194, pp. 120-137, 2012. DOI: 10.1016/j.ins.2011.07.031
- [3] Minjae Park and Kwanghoon Kim, "XWELL: A XML-Based Workflow Event Logging Mechanism and Language for Workflow Mining Systems," *Lecture Notes in Computer Science*, Vol. 4707, pp. 900-909, 2007. DOI: 10.1007/978-3-540-74484-9_76
- [4] Michael zur Muehlen and Keith D. Swenson, "BPAF: A Standard for the Interchange of Process Analytics Data," *Lecture Notes in Business Information Processing*, Vol. 66, pp. 170-181, 2011.
- [5] IEEE, "IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams," IEEE 1849-2016, 2016. DOI: 10.1109/IEEESTD.2016.7740858
- [6] Kim, Kwanghoon and Ellis, Clarence A., "Section II / Chapter VII. An ICN-based Workflow Model and Its Advances," *HANDBOOK OF RESEARCH ON BUSINESS PROCESS MODELING*, pp. 142-172, IGI Global, ISR, 2009.
- [7] W. Saeed, Z. Ahmad, A. I. Jehangiri, N. Mohamed, A. I. Umar and J. Ahmad, "A Fault Tolerant Data Management Scheme for Healthcare Internet of Things in Fog Computing," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 1, pp. 35-57, 2021. DOI: 10.3837/tiis.2021.01.003.



Minjae Park is an assistant professor in the division of computer and information at Daelim University College, South Korea. He received B.S. degree, M.S. degree, and Ph. D. degree in computer science from Kyonggi University in 2004, 2006, and 2009, respectively. He had worked as a researcher and developer at BISTel, Inc. in South Korea. His research interests include semiconductor manufacturing and fault-detection systems, large-scale database systems and applications, enterprise information systems, the Internet of Things collaboration architectures and services, process-aware knowledge discovery, process mining, and process monitoring and predictive monitoring.



supported social and affiliation networks discovery and analysis, and process-aware Internet of Things architecture and services.

Dinh-Lam Pham received the B.S. and M.S. degrees in computer science from Thai Nguyen University, Vietnam, in 2008 and 2010, respectively, and the Ph.D. degree in computer science from Kyonggi University, South Korea, in 2021. He is currently a Research Professor with the Contents Convergence Software Research Institute and a Research Member with the Data and Process Engineering Research Laboratory, Kyonggi University. His research interests include process mining, deep learning-based process predicting, workflow systems, workflow-



engineering, process-aware enterprise information systems, business process intelligence, process mining, and deep learning and knowledge engineering using active contents big data.

Kyoung-Sook Kim received the B.S. degree in computer science from Kyonggi University, South Korea, in 1984. She also received the M.S. degree and the Ph.D. degree in computer science and engineering from Kyunghee University, South Korea, in 2004 and 2019, respectively. She is currently a lecturer with the Division of AI Computer Science and Engineering, Kyonggi University. She is also a research member of the Contents Convergence Software Research Institute in Kyonggi University. Her research interests include data and knowledge



Software Research Institute, Kyonggi University, where he has been leading and fulfilling a multi-million dollars research project that will be continuously supported and funded by the National Research Foundation of Korea, from 2020 to 2029, ever since the institute was officially designated as the National Science and Engineering Research Institute by the Ministry of Education, South Korea, in 2020. He had worked as a Researcher and a Developer at Aztek Engineering, American Educational Products Inc., and IBM, USA, and a Research Member of Staff at the Electronics and Telecommunications Research Institute (ETRI), South Korea. He is currently a Full Professor with the Division of AI Computer Science and Engineering and the Founder and the Supervisor of the Data and Process Engineering Research Laboratory, Kyonggi University, South Korea. His research interests include groupware, workflow and business process management systems, process warehousing and mining, predictive process monitoring, workflow-supported social networks discovery and analysis, process-aware Internet of Things, predictive process modeling, process deep-learning.

Kwanghoon Pio Kim (Member, IEEE) received the B.S. degree in computer science from Kyonggi University, South Korea, in 1984, the M.S. degree in computer science from Chungang University, South Korea, in 1986, and the M.S. and Ph.D. degrees in computer science from the University of Colorado Boulder, USA, in 1994 and 1998, respectively. He had been the Dean of the Computerization and Informatics Institute, Kyonggi University, from 2017 to 2021. Since 2007, he has been the Founder and the Director in charge of the Contents Convergence