

Using Event Log Knowledge to Support Business Process Simulation Model Construction

Niels Martin

Hasselt University, Agoralaan Building D, 3590 Diepenbeek, Belgium
`niels.martin@uhasselt.be`

Abstract. My dissertation focuses on the use of event log knowledge, i.e. process mining, to support the development of business process simulation models. Despite the fact that the Process Mining Manifesto highlights this topic as a key research challenge, prior research efforts tend to have a proof-of-concept nature. To this end, this dissertation contributes towards fundamentally bridging the gap between these domains by (i) providing the required conceptualization and (ii) developing a set of methods that extract knowledge from event logs to support specific business process simulation modeling tasks.

Keywords: Business process simulation · Process mining

1 Positioning and dissertation objectives

Every organization is comprised of a set of business processes such as the production process, and the transportation process. When decision-makers analyze these processes and particular issues appear, several ideas for potential process changes are likely to be generated.

To evaluate the effects of policy measures, managers can experiment with the real-life process by, e.g., changing the staffing policy and measure its operational effect. However, this is a high-risk approach as the implementation of a measure does not guarantee the desired outcomes. This is a setting in which business process simulation can be a valuable instrument. Business process simulation (BPS) refers to the imitation of business process behavior through the use of a simulation model. Using a BPS model, an organization can verify the consequences of proposed process modifications prior to implementation [5].

The use of BPS requires the construction of a simulation model, which, in its turn, necessitates a profound insight in the business process. To this end, extensive information needs to be collected. Typical information sources include business documents, interviews with business experts and observations of the process. Despite the valuable insights that can stem from these information sources and their common use, their limitations should also be recognized. Business documents can contain information deviating from real-life process behavior [6]. Interviews, in their turn, can result in contradictory information [9] and can, e.g., be heavily influenced by recent experiences within the business process [3].

Collecting observational data requires a significant time investment and can suffer from the Hawthorne effect, which refers to the performance increase of staff members due to the mere fact that their actions are observed [7].

The aforementioned limitations stress the need for information sources that are more readily available and less influenced by human perception. In this respect, an important trend is that business processes are increasingly supported by process-aware information systems such as Enterprise Resource Planning systems. These systems record process execution information in an event log, of which the structure is illustrated in Table 1. Each row in the event log represents ‘something’ that happens in the process. For instance: the first row indicates that Sue started to register application 143 on April 3rd, 2018 at 08:52:41. She completes this registration at 09:04:04, as shown in the second row.

Table 1. Illustration of event log structure

case id	timestamp	activity	transaction type	resource	...
...
143	03/04/2018 08:52:41	Register application start		Sue	...
143	03/04/2018 09:04:04	Register application complete		Sue	...
144	03/04/2018 09:04:04	Register application start		Sue	...
107	03/04/2018 09:06:22	Judge application	start	Mike	...
...

As an event log is automatically recorded and captures real-life process behavior, it is well-suited as an additional information source for the construction of a BPS model. My dissertation focuses on the use of event log knowledge, i.e. process mining, to support simulation model construction. This is marked as one of the key research challenges in the Process Mining Manifesto [1].

Literature on this topic includes [2], [4], and [6], with [8] providing the most comprehensive support. In [8], a stepwise method to mine a simulation model from an event log is provided together with suggestions for suitable ProM plugins to perform mining operations. However, simplifying assumptions were made such as equating case arrival time to the first activity’s start time. Moreover, modeling tasks such as the queue discipline and resource schedules are excluded.

From a thorough analysis of existing literature, it follows that, despite the intrinsic potential of process mining to support BPS model construction, limited insights exist in the systematic use of event log knowledge within this context. A significant research gap is present as literature on this matter tends to have a proof-of-concept nature. This implies that important simplifying assumptions are made and that various modeling tasks that need to be performed when building a real-life BPS model are not taken into consideration. Moreover, the required conceptual foundation to fundamentally integrate event log knowledge in BPS model development is not present, marking another research challenge.

Consistent with these research gaps, the two main objectives of this work are (i) providing the required conceptualization to fundamentally integrate event log

knowledge in BPS model construction and (ii) developing a set of methods that extract knowledge from event logs to support specific BPS modeling tasks.

2 Overview of the dissertation

Given the limited research insights in the systematic use of event log knowledge within this context, the first part of the dissertation is situated at a conceptual level (Section 2.1). Taking this conceptualization as an input, the second part of the thesis develops methods to mine relevant insights related to specific BPS modeling tasks from an event log (Section 2.2).

2.1 Conceptualization of the use of event log knowledge in BPS model construction

A first stage of conceptualization relates to defining the key steps of a simulation study. This enables positioning the use of event log knowledge in the wider picture of a BPS study. Simulation literature proposes a multitude of methods describing the key steps in a simulation study. Even though key components such as computer modeling are included in most of them, differences can also be observed. However, existing methods tend to be defined in isolation. Given this observation, a new method for conducting a simulation study is developed, based on a critical analysis of 14 existing methods. The developed method consists of 9 steps with continuous assessment as a central feedback mechanism.

After elaborating on the wider perspective of a BPS study, the second stage of conceptualization uses a generic representation of a BPS model to define a series of modeling tasks. For each of them, the potential of event log knowledge to support their specification is outlined. These insights are compared to the state of the art in process mining literature in general and research on the use of process mining for BPS purposes in particular. From this comparison, it follows that few process mining algorithms are directly applicable to support BPS model construction. This can be attributed to the differences between the underlying paradigms in process mining and simulation. While many process mining techniques treat an event log as a series of independent cases, simulation mimics the behavior of an operational business process in which the interaction between cases that are simultaneously present in the process influences process behavior. Hence, tailored methods need to be developed for a wide range of BPS modeling tasks, which demonstrates that additional research efforts are required. The structured set of research challenges provided in the dissertation provides clear directions for future research and, in this way, is a solid foundation for the future development of the domain.

2.2 Developed methods to support specific BPS modeling tasks

Using the conceptualization outlined in Section 2.1 as an input, the dissertation also develops a series of new algorithms to retrieve event log insights to support

specific BPS modeling tasks. Three modeling tasks are selected based on criteria including the potential consequences of inaccurately performing it on the simulation study results and the absence of process mining algorithms which are directly applicable or easily adjustable. Based on this assessment, new methods are developed to support the specification of (i) the entity arrival rate, (ii) batch processing, and (iii) resource schedules. All algorithms are rigorously evaluated using artificial data and, whenever relevant, real-life event logs.

Entity arrival rate An entity is a dynamic object (e.g., a patient, a parcel) that moves through the process. The entity arrival rate models the pace at which new entities arrive, which is often expressed using a parameterized probability distribution for the time between two consecutive arrivals. As it defines the inflow at the beginning of the process, inaccurately modeling the arrival rate can have a major influence on simulation outputs.

Process mining literature on this topic implicitly assumes that an entity's arrival time corresponds to its first recorded timestamp. However, the developed taxonomy shows that this assumption is only appropriate when no queues are formed at the start activity. Given this observation, an Arrival Rate Parameter Retrieval Algorithm (ARPRA) is created, which is the first algorithm that explicitly takes queue formation into account during arrival rate mining.

Batch processing Batch processing refers to a resource's tendency to accumulate entities in order to process them simultaneously, concurrently, or sequentially. Regarding batch processing, two distinct topics are treated: batch identification and batch activation rules. For both topics, a new mining algorithm is developed, implemented and evaluated. Firstly, for batch identification, the Batch Organization of Work Identification algorithm (BOWI) is introduced, which is based on a distinction between simultaneous, sequential and concurrent batching. It is the first algorithm that systematically identifies batches in an event log and calculates a set of batch processing metrics.

Secondly, related to batch activation rules, the Batch Activation Rule Identification algorithm (BARI) is created. A batch activation rule captures the circumstances under which a resource starts processing a batch, e.g. when a particular number of entities has been collected. Using explicit and implicit event log information, BARI transforms the problem into a classification problem, which is solved using decision tree analysis to obtain batch activation rules.

Resource schedules A resource schedule expresses the availability of a resource for the process under consideration. When focusing on human resources, staff members might not work full-time or might be involved in multiple processes. Consequently, it can be difficult to deduce their availability for one specific process from generic schedules provided by the HR-department. This is especially the case when work organization is less rigid, leaving significant freedom for resources to divide their attention between processes. As resource availability is critical information when building a simulation model, a Resource Schedule Identification Method (RSIM) is developed, which supports the specification of resource schedules. More specifically, daily availability records are mined from

the event log, which express a resource's allocation to a particular process on a particular day. It is the first algorithm in its kind as the daily availability records take into account (i) the temporal dimension of availability, i.e. the time of day at which a resource is available, and (ii) intermediate availability interruptions.

3 Key contribution of the dissertation

My dissertation presents an important step towards the integration of process mining in BPS model construction. Despite the fact that the Process Mining Manifesto [1] highlights this integration as a key research challenge, prior research efforts tend to have a proof-of-concept nature. Consequently, from a scientific perspective, the developed conceptualization in this dissertation provides the much needed structure to this research field. Moreover, the developed algorithms demonstrate how an event log can be used in previously unexplored directions to, e.g., mine batching behavior or resource schedules. From a managerial perspective, the improved BPS models due to the use of process mining will provide more accurate insights in the effect of particular policy alternatives on process performance. In this way, BPS will become a more powerful decision support tool, which will translate in an increased use of BPS in practice. Even though further research is required, e.g. by developing algorithms for other modeling tasks, this dissertation lays important foundations towards more powerful BPS models through the use of process mining.

References

1. van der Aalst, W.M.P., Adriansyah, A., de Medeiros, A.K.A., ..., Wen, L., Westergaard, M., Wynn, M.: Process mining manifesto. *Lecture Notes in Business Information Processing* **99**, 169–194 (2012)
2. Aguirre, S., Parra, C., Alvarado, J.: Combination of process mining and simulation techniques for business process redesign: a methodological approach. *Lecture Notes in Business Information Processing* **162**, 24–43 (2013)
3. Dickey, D., Pearson, C.: Recency effect in college student course evaluations. *Practical Assessment, Research and Evaluation* **10**(6), 1–10 (2005)
4. Leyer, M., Moormann, J.: Comparing concepts for shop floor control of information-processing services in a job shop setting: a case from the financial services sector. *International Journal of Production Research* **53**(4), 1168–1179 (2015)
5. Melão, N., Pidd, M.: Use of business process simulation: a survey of practitioners. *Journal of the Operational Research Society* **54**(1), 2–10 (2003)
6. Märuşter, L., van Beest, N.R.T.P.: Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems* **21**(3), 267–297 (2009)
7. Robinson, S.: *Simulation: the practice of model development and use*. Wiley, Chichester (2004)
8. Rozinat, A., Mans, R.S., Song, M., van der Aalst, W.M.P.: Discovering simulation models. *Information Systems* **34**(3), 305–327 (2009)
9. Vincent, S.: Input data analysis. In: Banks, J. (ed.) *Handbook of simulation: principles, advances, applications, and practice*, pp. 3–30. John Wiley & Sons, Hoboken (1998)