Queue Mining: Service Perspectives in Process Mining (Extended Abstract)

Arik Senderovich

1 Introduction

Modern business processes are supported by information systems that record process-related events in event logs. *Process mining* is a maturing research field that aims at discovering useful information about the business process from these event logs [1]. Process mining can be viewed as the link that connects process analysis fields (e.g. business process management and operations research) to data analysis fields (e.g. machine learning and data mining) [2]. This thesis is mainly concerned with process mining techniques that aim at answering operational questions such as 'does the executed process as observed in the event log correspond to what was planned?', 'how long will it take for a running case to finish?' and 'how should resource capacity or staffing levels change to improve the process with respect to some cost criteria? [1, Ch. 7, Ch. 8]. We refer to process mining solutions of such questions as *operational process mining*. Other types of process mining subfields are beyond the scope of the present work.

To the best of our knowledge, and prior to this thesis, process mining techniques overlooked dependencies between cases when answering operational questions. For example, prior to our work, state-of-the-art methods for predicting remaining times of running cases considered only historical data of the case itself [3], while the interactions among cases (e.g. through queueing for shared resources) were neglected. The independence assumption is plausible in processes where capacity always exceeds demand. However, in service processes, which are prevalent in numerous application domains (e.g., healthcare, banking, transportation), multiple customer-resource interactions occur, and customers often compete over scarce resources. Consequently, the central argument of this thesis is that for service-oriented processes, process mining solutions *must* consider case interactions when answering operational questions.

The main contribution of this research thesis is the start of bridging a noticeable gap between process mining and queueing theory. To this end, we introduce *queue mining* (a term coined in this thesis), which is a set of data-driven methods (models and algorithms) for queueing analysis of business processes. The remainder of this extended abstract starts with an overview of operational process mining where we emphasize gaps in the existing literature (Section 2). We then provide an outline of the main contributions of our research (Section 3).[1]

[1] The thesis is comprised of three journal papers (sequence of papers thesis), thus we refer to the relevant paper when presenting the contribution.

2 Operational Process Mining

In this section, we provide a literature survey of operational process mining.[2] Specifically, we go over three streams of work, namely *simulation mining*, *supervised process mining*, and *multi-perspective conformance checking*. The former two lines of work focus on performance-oriented operational questions, while the latter focuses on the conformance between model and log.

Simulation Mining. Early works on performance-oriented process mining were based on the enhancement of control-flow models with further data-based information such as decisions, resources and time [6]. The work of Măruşter and van Beest [7] extends this approach from modeling and analysis of an existing process to its improvement based on the discovered model. Due to the richness of these discovered models they can only be analyzed via simulation. Thus, we refer to these methods as simulation mining or model-based process mining, interchangeably.

The main benefit of these model-based techniques is their ability to solve a large set of performance problems. For example, one can use simulation to predict waiting times and propose resource staffing. However, current simulation mining techniques overlook queueing effects by simulating process cases independently of each other and drawing waiting times (and other measures that stem from case dependencies) from fitted distributions. Furthermore, model-based mining techniques tend to overfit reality as it is observed in the event log and therefore it was shown that simpler performance models often provide accurate results and are more efficient [8]. The expressiveness problem is partially addressed in [4], where authors mine non-Markovian stochastic Petri nets [9], which are simple formalisms that account for control-flow and exponential durations. In [8], the best of both worlds is combined by a method for simulation mining that captures queueing effects and simplifies the resulting model. This simplification results both in improved efficiency (run-time complexity of the solution) and accuracy.[3]

Supervised Process Mining. Another prevalent stream of work in process mining is the encoding of event logs into well-established supervised learning methods. Examples for such methods can be found in [3,10,11,12]. In these works, both the behavior and the operational aspects of the process are encoded: the former representing the control-flow perspective and the latter capturing activity durations and resources. While these approaches are often accurate in solving a given mining problem with respect to a set of performance measures observed in the event log (e.g. total time in process), it has two major drawbacks.

First, one cannot use these techniques for quantifying measures that are not directly observable in the log. For example, information that allows one to calculate resource utilization might be missing from the event log, and thus supervised approaches cannot be applied. Second, exploring process improvement directions and sensitivity of process parameters (e.g. durations and arrival rates) is impossible due to the lack of data that

[2] Solving operational problems based on data is not limited to process mining. For references to works outside the process mining discipline, see [4] (for operations research), and [5] (for data mining).

[3] The work is a result of the current research. It is not included in this thesis.

describes the effect of changing these parameters. In [5] and [13], we propose supervised process mining solutions that bring together the benefits of model-based mining (or simulation mining) and supervised learning. Specifically, these solutions comprise machine learning methods (e.g. decision trees) and queueing models. This combination enables quantifying measures that were not observed in the event log due to the existence of the model. For measures that appear in the event log we get accurate solutions by exploiting the strengths of supervised learning [14]. Furthermore, we are able to have what-if reasoning based on the queueing model, without the need to acquire further data from the changed process.

Multi-Perspective Conformance Checking. In this thesis, we call operational process mining to works on multi-perspective conformance analysis that verify alignment of process models and event logs with respect to time and resources [15,16]. The approach in the literature is to compare deterministic data values according to a given model to those available in the log. For example, activity durations are considered to be known values (according to the model) and the alignment procedure checks whether or not event log durations are equal to model durations. However, the reality in processes is often random (random durations, random arrival process, etc.). Therefore, a methodology for comparing distributions (that come from the model) and deterministic values (coming from the log) is required. We show such an approach in our work on conformance checking from the queueing perspective [17]. Further, our work proposes a methodology that considers queueing interactions that were neglected in existing methods, and recommends improvement based on the result of the conformance checking procedure [17].

3 Thesis Contributions

The main overall contribution of this thesis is the development of queue mining, which is a set of operational process mining solutions based on the discovery of queueing models from event data. These solutions explicitly capture delays due to case-resource interactions. Below, we detail the specific contributions of this research and outline their impact in two research fields, namely process mining and queueing theory (or, more broadly, operations research).

3.1 Contributions to Process Mining

The thesis makes the following contributions to process mining:

- The work integrates data-driven queueing modeling and analysis into operational process mining [13,17,5,23,8,22], namely queue mining.
- Our work links operational conformance analysis to performance improvement in business processes [17]. Specifically, we develop a methodology for the detection of problematic areas in a process (e.g., due to synchronization or resource delays) and propose proper adjustments to queueing policies (selection of the next case to enter service) in these problematic parts.
- We make advances in the area of multi-dimensional conformance analysis, by adding the queueing perspective, structuring the approach, and demonstrating how conformance checking can lead to performance improvement [17].

– We propose hybrid methods that combine supervised learning and model-based process mining techniques to improve the accuracy of answering operational questions [13,5]. This combination improves accuracy with respect to applying the two approaches separately.

Three of the works mentioned in the list above, [13,17,5] are part of the thesis. Other works were published in papers that were not included in the thesis. The proposed methods were validated based on three real-world datasets coming from service domains. In [22,13] we evaluate queue mining against a bank's call center; in [17,23,8] we mine the dataset of the Dana-Farber Cancer Institute (a large outpatient cancer hospital in the United States); and in [5] we evaluate our techniques on GPS bus data that comes from the city of Dublin.

3.2 Contributions to Queueing Theory

Queueing theory has been mainly a theoretical research field that aims at modeling and analyzing (typically) stochastic and dynamic service or manufacturing systems [24,25]. Some works in queueing theory provide numerical experiments that justify the practicality of theoretical results. The underlying data is typically simulated based on systems that adhere to various assumptions that fit the models. Only a small sample of works in queueing theory include real-world data validation (see [26,27,28] and references within). In this work, we validate multiple results from queueing theory, thus showing their practical applicability to real-world data. Further, we provide a platform for testing new and (other) existing results in queueing theory based on event data, and the queueing model is well-defined). Specifically, all our techniques were implemented as prototype systems that are available as open-source. For example, the methods that we use for delay prediction [13] can be extended, replaced, and tested against, when new prediction methods become available. In other words, this research creates a playground for queueing theorists to empirically evaluate their developed theories and results.

References

1. van der Aalst, W.M.P.: Process Mining - Discovery, Conformance and Enhancement of Business Processes. Springer (2011)
2. van der Aalst, W.M.P.: Process mining: overview and opportunities. ACM Trans. Management Inf. Syst. **3**(2) (2012) 7
3. van der Aalst, W.M.P., Schonenberg, M.H., Song, M.: Time prediction based on process mining. Information Systems **36**(2) (2011) 450–475
4. Rogge-Solti, A., Weske, M.: Prediction of business process durations using non-markovian stochastic petri nets. Inf. Syst. **54** (2015) 1–14
5. Gal, A., Mandelbaum, A., Schnitzler, F., Senderovich, A., Weidlich, M.: Traveling time prediction in scheduled transportation with journey segments. Information Systems (2015)
6. Rozinat, A., Mans, R., Song, M., van der Aalst, W.M.P.: Discovering simulation models. Information Systems **34**(3) (2009) 305–327
7. Măruşter, L., van Beest, N.R.: Redesigning business processes: a methodology based on simulation and process mining techniques. Knowledge and Information Systems **21**(3) (2009) 267–297
8. Senderovich, A., Rogge-Solti, A., Gal, A., Mendling, J., Mandelbaum, A., Kadish, S., Bunnell, C.A.: Data-driven performance analysis of scheduled processes. In: Business Process Management. Springer (2015) 35–52

9. Haas, P.J.: Stochastic Petri Nets: Modelling, Stability, Simulation. Springer (2002)
10. de Leoni, M., van der Aalst, W.M.P., Dees, M.: A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. Information Systems **56** (2016) 235–257
11. Polato, M., Sperduti, A., Burattin, A., de Leoni, M.: Time and activity sequence prediction of business process instances. Technical Report, (2016)
12. Leontjeva, A., Conforti, R., Francescomarino, C.D., Dumas, M., Maggi, F.M.: Complex symbolic sequence encodings for predictive monitoring of business processes. In: Business Process Management - 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 - September 3, 2015, Proceedings. (2015) 297–313
13. Senderovich, A., Weidlich, M., Gal, A., Mandelbaum, A.: Queue mining for delay prediction in multi-class service processes. Information Systems **53** (2015) 278–295
14. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA (2001)
15. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Balanced multi-perspective checking of process conformance. Computing **98**(4) (2016) 407–437
16. de Leoni, M., van der Aalst, W.M.P.: Data-aware process mining: discovering decisions in processes using alignments. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, ACM (2013) 1454–1461
17. Senderovich, A., Weidlich, M., Yedidsion, L., Gal, A., Mandelbaum, A., Kadish, S., Bunnell, C.A.: Conformance checking and performance improvement in scheduled processes: A queueing-network perspective. Information Systems (2016)
18. Folino, F., Guarascio, M., Pontieri, L.: Discovering context-aware models for predicting business process performances. In: On the Move to Meaningful Internet Systems: OTM 2012. Springer (2012) 287–304
19. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **2**(2) (2012) 182–192
20. De Leoni, M., van der Aalst, W.M.P.: Aligning event logs and process models for multi-perspective conformance checking: An approach based on integer linear programming. In: Business Process Management. Springer (2013) 113–129
21. De Leoni, M., van der Aalst, W.M.P., van Dongen, B.F.: Data-and resource-aware conformance checking of business processes. In: Business Information Systems, Springer (2012) 48–59
22. Senderovich, A., Weidlich, M., Gal, A., Mandelbaum, A.: Mining resource scheduling protocols. In: Business Process Management. Springer (2014) 200–216
23. Senderovich, A., Weidlich, M., Gal, A., Mandelbaum, A., Kadish, S., Bunnell, C.A.: Discovery and validation of queueing networks in scheduled processes. In: Advanced Information Systems Engineering, Springer (2015) 417–433
24. Buzacott, J.A., Shanthikumar, J.G.: Stochastic Models of Manufacturing Systems. Volume 4. Prentice Hall Englewood Cliffs, NJ (1993)
25. Mandelbaum, A.: Service engineering (science, management): A subjective view. Technical report, Technical report, Technion-Israel Institute of Technology (2007)
26. Azriel, D., Feigin, P.D., Mandelbaum, A.: Erlang S: A data-based model of servers in queueing networks. Technical report, Working paper (2014)
27. Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y.N., Tseytlin, Y., Yom-Tov, G.B.: On patient flow in hospitals: A data-based queueing-science perspective. Stochastic Systems **5**(1) (2015) 146–194
28. Mandelbaum, A., Zeltyn, S.: Data-stories about (im)patient customers in tele-queues. Queueing Systems **75**(2-4) (2013) 115–146