

Forward-Looking Process Mining

Mahsa Pourbafrani, Wil M.P. van der Aalst

Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany

Abstract

Most process mining techniques are *backward-looking*. Based on historic data, they produce descriptive process models and reveal performance and compliance problems. *Forward-looking* process mining focuses on turning the results of backward-looking techniques into prediction and actions. The current techniques use detailed event data to provide insights. However, different angles of event data enable capturing process behavior and underlying relationships between process variables for further future analyses, e.g., daily arrival rate and resource efficiency. In this project, we aim to provide a forward-looking framework for business processes to replay their processes and assess the effect of the actions performed based on process mining insights. To do so, we use both detailed event data and aggregated event data of processes over time, i.e., *fine-grained* event logs and *coarse-grained* process logs. Using alternative simulation approaches such as System Dynamics (SD) it is possible to incorporate external factors into the model using more coarse-grained logs. Furthermore, we focus on connecting the effects of strategic decisions and detailed business processes by providing a comprehensive simulation framework, i.e., hybrid simulations.

Keywords

process mining, fine-grained event logs, coarse-grained process logs, scenario-based simulation, system dynamics

1. Introduction and Problem Definition


Historical data on executions of business processes may be used to support the business owners to analyze their processes. These event data have a wealth of information about the processes, where backward-looking process mining techniques help to uncover these insights [1]. Forward-looking process mining supports process owners in taking actions based on the provided knowledge [2]. Forward-looking process mining is generally categorized in two categories: prediction models using machine learning techniques, such as [3], and simulation techniques. For instance, process mining capability to describe a process is used to enrich simulation models and make foreseeing their future possible [4]. The current forward-looking techniques for assessing the future state of a process do not fully cover the following aspects: (1) The majority of approaches are at a fine-grained level (detailed) and do not take into account the impact of quality-based factors on the process. Discrete Event Simulations (DES), for instance, are incapable of capturing the impact of resource training on the efficiency of processes. (2) Various cause and effect relations are invisible in the fine-grained event logs. Aggregating the event data

Proceedings of the Demonstration & Resources Track, Best BPM Dissertation Award, and Doctoral Consortium at BPM 2021 co-located with the 19th International Conference on Business Process Management, BPM 2021, Rome, Italy, September 6-10, 2021

✉ mahsa.bafrani@pads.rwth-aachen.de (M. Pourbafrani); wvdaalst@pads.rwth-aachen.de (W. M.P. v. d. Aalst)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

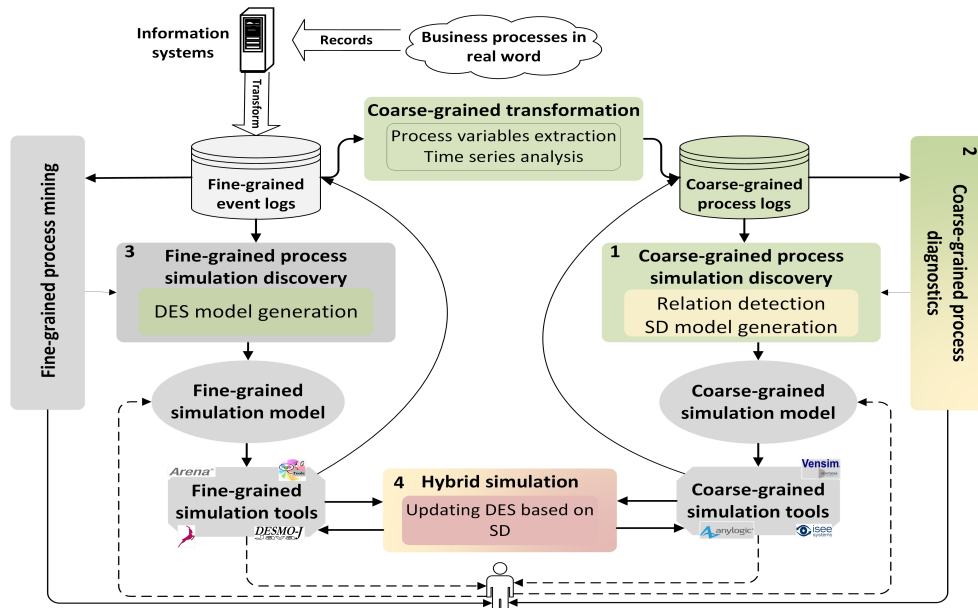


Figure 1: The overview of the existing (gray) and the proposed approaches for process simulation. We use process mining techniques to make simulations more data-driven. Moreover, we combine classical discrete event simulation (fine-grained models) with coarse-grained simulation models.

can reveal the different states of processes, e.g., the effect of daily workload on the resources' efficiency. (3) The interaction between coarse-grained and fine-grained simulations of business processes is not provided, e.g., the effect of an advertisement after three months on the idle time of process activities and the arrival rate. Dotted lines in Figure 1 represent conventional paths for designing simulation models of processes.

Our goal is to define and generate coarse-grained process logs at different levels and from multiple aspects, shown in Figure 1. In this context, standard process mining techniques and aggregated process analyses are referred to as *Fine-grained process mining* and *Coarse-grained process diagnostics*, respectively. The process aggregated state and its behavior at that level directly affect every single instance in the process. Following this step, data-supported process diagnostics, as well as fine-grained and coarse-grained simulation models are derived. As a result, a set of simulation models capable of performing the prescribing role for processes are generated. The generated simulation models can be designed and verified to reflect the processes. For instance, DES is a form of fine-grained process simulations using fine-grained event logs and SD is a type of coarse-grained process simulations using coarse-grained process logs. System dynamics, which is an aggregated simulation technique, models a system using variables that describe the system over time [5]. For processes, instead of simulating each event including single case arrival and execution, we simulate the process using variables such as the daily arrival rate, or daily average service time.

Contribution to BPM Research Providing what-if analyses is an important step in Business Processes Management (BPM) [6]. Exploiting historical event data of organizations supports

business process simulation in BPM [7]. The simulation parameters are extracted from the processes, and they make simulation models more realistic. However, the provided models are always detailed and try to mimic the exact behavior of the business processes, i.e., at the detailed events. The goal of this project is to pave the way for comprehensive future analyses of business processes by providing diagnostics on top of the coarse-grained business process historical data. These insights and diagnostics are used to form data-driven simulations with two levels of granularity: fine-grained and coarse-grained (strategical what-if analysis) models.

2. Proposed Solution (Framework)

We propose an approach to transform fine-grained event logs into coarse-grained process logs. The transformed logs enable further analyses such as discovering hidden relations and supporting simulation models generation of business processes at aggregated levels. Figure 1 illustrates the designed steps from a fine-grained event log to generate ultimate results, i.e., SD models (1), process diagnostics (2), DES models (3), and hybrid process simulations (4). The purpose of the forward-looking analysis of business processes determines different paths inside the proposed approach. Given the process diagnostics at different levels such as discovering the effect of workload on the resources per day, the coarse-grained simulation models in the form of SD can be extracted [8].

The prototype of the main steps of the project is implemented as a tool (*PMSD*) in [9]. Our proposed solution includes 7 different yet connected frameworks. The output of each framework is based on a specific purpose in forward-looking process mining. The designed/implemented frameworks are as follows:

- Preprocessing (coarse-grained process log): a set of possible process variables are defined based on the process aspects in fine-grained event logs [10] and these process variables are calculated over specific steps of time, e.g., hourly, or daily. To design accurate simulation/prediction models, the time window for extracting the process variables highly affects the quality of the simulation/prediction models. In [11], we used *Time Series Analysis* techniques such as *ARIMA* [12] for finding the best window of time.
- Coarse-grained simulation model generation: we used linear/nonlinear correlations between process variables over time to discover existing relationships at higher levels and design the system dynamics models [13].
- Process diagnostics (aggregated levels): coarse-grained process logs represent the process at different steps of time using different aspects, e.g., the daily arrival rate of cases and the average waiting time in the process. Techniques such as *Granger Causality* [14] and *Curve Fitting* [15] are used to discover underlying causes and effects relations.
- Simulation and validation: after the simulation, the accuracy of the simulation models can be evaluated by comparing the results to the values of process variables at each time step, e.g., [16] performs an evaluation for a car production line.
- Simulation model refinement: by adding external and qualitative factors to the validated SD models, strategic analyses will be possible, e.g., effects of a new advertisement strategy.
- Fine-grained simulation model: [4] is the pioneering work that introduces data-driven simulation in process mining. This module presented in [17] automatically discovers the

process activity flow and enriches that with resources, capacities, and time aspects of the process. These models using multiple techniques such as *Earth-Mover's distance* [18] and *Performance Spectrum* [19] are validated in [20], and can be used for regenerating processes.

- Hybrid simulation of processes (SD and DES): the connection between two types of simulation models enables applying the effect of high-level what-if analyses on the fine-grained simulation of business processes.

3. Current Status and Challenges

The green steps in Figure 1 have been designed and evaluated in the forward-looking process mining project. Designing and implementing the framework for process diagnostics at higher levels of aggregation is the next step. We do so by exploiting customized methods to identify the relationships between process variables on coarse-grained process logs, such as *Vector Autoregressive*. The project's second focus is on the automatic discovery of mathematical equations for the SD models using statistical and machine learning methods, and the final step is to implement hybrid simulations of processes, which are highlighted in yellow in Figure 1.

The current challenges to be addressed are mainly in providing use cases, determining underlying equations for all the process variables in the SD models, and connecting two types of simulations (red steps in Figure 1). There is always a trade-off between adding external factors and the accuracy of simulation results. Therefore, real-world case studies with known changes and effects, as well as process domain knowledge in the form of external variables are required. Since these external variables are not quantifiable, the validity of the designed simulation model cannot be easily assessed. Furthermore, using only the generated process variables in the SD-Logs will limit scenarios and the main purpose of system dynamics modeling. As a result, strategic decision-making involving quality-based variables is not entirely possible.

When connecting fine-grained simulations (DES) and coarse-grained simulations (SD) in the hybrid simulation step, the following questions should be addressed. How can both simulation models be synchronized? How should interaction points be defined and discovered in practice? Which DES parameters, for example, are updated as a result of SD simulation? How should the execution of two models be handled in practice? For instance, for generating CPN models and updating them while simulating SD models at the same time. Furthermore, the user interaction and designing the scenarios are the project's open challenges.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy- EXC 2023 Internet of Production- Project ID: 390621612. We also thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

References

- [1] W. M. P. van der Aalst, *Process mining - Data science in action*, second edition, Springer (2016). doi:10.1007/978-3-662-49851-4.
- [2] W. M. P. van der Aalst, *Process mining and simulation: A match made in heaven!*, in: *Proceedings of the 50th Computer Simulation Conference, SummerSim 2018*, 2018, pp. 4:1–4:12.
- [3] N. Tax, I. Verenich, M. La Rosa, M. Dumas, *Predictive business process monitoring with lstm neural networks*, in: E. Dubois, K. Pohl (Eds.), *Advanced Information Systems Engineering*, Springer International Publishing, Cham, 2017, pp. 477–492.
- [4] A. Rozinat, R. S. Mans, M. Song, W. M. P. van der Aalst, *Discovering simulation models*, *Inf. Syst.* 34 (2009) 305–327. doi:10.1016/j.is.2008.09.002.
- [5] J. D. Sterman, *Business dynamics: Systems thinking and modeling for a complex world*, McGraw-Hill (2000).
- [6] K. Tumay, *Business process simulation*, in: *Proceedings Winter Simulation Conference*, 1996, pp. 93–98. doi:10.1109/WSC.1996.873265.
- [7] M. Camargo, M. Dumas, O. González, *Automated discovery of business process simulation models from event logs*, *Decis. Support Syst.* 134 (2020) 113284. doi:10.1016/j.dss.2020.113284.
- [8] M. Pourbafrani, S. J. van Zelst, W. M. P. van der Aalst, *Scenario-based prediction of business processes using system dynamics*, in: *On the Move to Meaningful Internet Systems: COOPIS 2019 Conferences*, 2019, pp. 422–439.
- [9] M. Pourbafrani, W. M. P. van der Aalst, *PMSD: Data-driven simulation using system dynamics and process mining*, in: *Proceedings of Demonstration at the 18th International Conference on Business Process Management*, 2020, pp. 77–81. URL: <http://ceur-ws.org/Vol-2673/paperDR03.pdf>.
- [10] M. Pourbafrani, W. M. P. van der Aalst, *Extracting process features from event logs to learn coarse-grained simulation models*, in: *Advanced Information Systems Engineering - 33rd International Conference, CAiSE 2021, Melbourne, VIC, Australia, June 28 - July 2, 2021, Proceedings*, volume 12751 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 125–140. URL: https://doi.org/10.1007/978-3-030-79382-1_8. doi:10.1007/978-3-030-79382-1_8.
- [11] M. Pourbafrani, S. J. van Zelst, W. M. P. van der Aalst, *Semi-automated time-granularity detection for data-driven simulation using process mining and system dynamics*, in: *Conceptual Modeling - 39th International Conference, ER 2020, Proceedings*, 2020, pp. 77–91. doi:10.1007/978-3-030-62522-1_6.
- [12] G. Box, G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.
- [13] M. Pourbafrani, S. J. van Zelst, W. M. P. van der Aalst, *Supporting automatic system dynamics model generation for simulation in the context of process mining*, in: *Business Information Systems - 23rd International Conference*, 2020, pp. 249–263. doi:10.1007/978-3-030-53337-3_19.
- [14] C. W. Granger, *Investigating causal relations by econometric models and cross-spectral methods*, *Econometrica: journal of the Econometric Society* (1969) 424–438.
- [15] A. Zielesny, *From Curve Fitting to Machine Learning*, volume 18, Springer, 2011.

- [16] M. Pourbafrani, S. J. van Zelst, W. M. P. van der Aalst, Supporting decisions in production line processes by combining process mining and system dynamics, in: *Intelligent Human Systems Integration 2020*, 2020, pp. 461–467. doi:10.1007/978-3-030-39512-4_72.
- [17] M. Pourbafrani, S. Jiao, W. M. P. van der Aalst, SIMPT: Process improvement using interactive simulation of time-aware process trees, in: *Research Challenges in Information Science*, Springer International Publishing, Cham, 2021, pp. 588–594. doi:10.1007/978-3-030-75018-3_40.
- [18] S. J. J. Leemans, A. F. Syring, W. M. P. van der Aalst, Earth movers' stochastic conformance checking, in: T. Hildebrandt, B. F. van Dongen, M. Röglinger, J. Mendling (Eds.), *Business Process Management Forum*, Springer International Publishing, Cham, 2019, pp. 127–143.
- [19] V. Denisov, D. Fahland, W. M. P. van der Aalst, Unbiased, fine-grained description of processes performance from event data, in: M. Weske, M. Montali, I. Weber, J. vom Brocke (Eds.), *Business Process Management*, Springer International Publishing, Cham, 2018, pp. 139–157.
- [20] M. Pourbafrani, W. M. P. van der Aalst, Interactive process improvement using enriched process trees, in: *4th International Workshop on Empirical Methods in Conceptual Modeling EmpER 2021 Conceptual Modeling Co-located with the 40th International Conference, ER 2021, Proceedings*, 2021.