Data-driven Management of Interconnected Business Processes Contributions to Predictive and Prescriptive Mining

Wolfgang Kratsch [1,2]

[1] *University of Applied Sciences, Augsburg, Germany*
[2] *Project Group Business & Information Systems Engineering of the Fraunhofer FIT,Bayreuth, Germany*

Business process management (BPM) is an accepted paradigm of organizational design and a source of corporate performance [1]. Due to substantial progress in process identification, analysis, implementation, and improvement [2, 3], BPM receives constant attention from industry [4]. In times of market consolidation and increasing competition, operational excellence (i.e., continuously optimizing an organization's processes in terms of effectiveness and efficiency) is key to staying competitive. While traditional research in BPM focused on process models and model-based information systems (e.g., workflow management systems), recently, the focus has shifted to data-driven methods such as process mining [5]. In contrast to model-driven BPM, process mining uses execution data in the form of events arising during process enactment, which may be exploited in several ways [6]. Process mining strives to discover, monitor, and improve processes by extracting knowledge from event logs available in information systems [7]. The most commonly applied use case in process mining is discovering as-is process models that also serve as a starting point for more detailed analysis [8]. Based on the mined as-is-process, the use case of conformance checking helps to point out deviations from normative, predefined process models and actual process enactments (e.g., unintended handover of tasks, skipped activities, missed performance goals). As process mining analyzes information on an event-level, it also helps evaluate the actual process performance (e.g., measuring cycle times, interruptions, exceptions). In sum, process mining can help ensure process hygiene, constituting a fundamental requirement to achieve operational excellence [8].

As process mining is one of the most active streams in BPM, numerous approaches have been proposed in the last decade, and various commercial vendors transferred these methods into practice, substantially facilitating event data analysis [9]. At the tip of the iceberg, Celonis expanded in only seven years from start-up to a unicorn, indicating the enormous cross-industry business potential of process mining [10]. By 2023, Markets and Markets predicts a market potential of 1.42 billion US$ for process mining technologies [11]. However, there are still numerous unsolved challenges that hinder the further adoption and usage of process mining at the enterprise level [12]. First, finding, extracting, and preprocessing relevant event data is still challenging and requires a significant amount of time in a process mining project and, thus, remains a bottleneck without providing appropriate support [13]. Second, most process mining approaches operate on a single-process level, but organizations are confronted with a process network covering hundreds of interdependent processes [12]. Third, process managers strongly require forward-directed operational support, but most process mining approaches provide only descriptive ex-post insights, e.g., discovered models or performance analysis of a past period [8]. Since these challenges mainly drive this doctoral thesis, they will be discussed in detail below.

First, finding, extracting, and preprocessing relevant event data is still challenging. This is most frequently due to the lack of domain knowledge about the process, the distributed storage of required data in different databases and tables, and the requirement of advanced data engineering skills [13]. Most recent process mining approaches assume high-quality event logs without describing how such logs can be extracted from process-aware (PAIS) and particularly non-process-aware information

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
EMAIL: wolfgang.kratsch@fim-rc.de (W. Kratsch)
ORCID: 0000-0001-9815-0653 (W. Kratsch)

systems (non-PAIS). In case of solely relying on process-aware information systems (PAIS) that directly output minable event logs, the risks of neglecting process-relevant information arise, and so-called blind spots can occur. For instance, if processes contain activities enacted by physical resources or software bots that are not directly connected to PAIS, details of these enactments cannot be explored using classical PAIS-based event logs. Due to increasingly digitized organizations, a growing part of the available data is highly unstructured (e.g., text, video, or audio files) and requires the application of novel concepts [8]. To sum up, although process mining approaches significantly matured in the last decade, the step of data extraction is still too weekly supported and often results in bottlenecks that negatively affect the quality of process mining analysis.

Second, most process mining approaches operate on a single-process level. However, process mining currently evolves from project-based single-process analysis to an enterprise-wide ongoing task [8]. Thus, methods for scaling process mining approaches on an enterprise level are needed [12]. One of the most challenging topics relies on applying process mining methods that operate primarily on a single-process perspective to enterprise-wide process networks, frequently covering hundreds of highly interconnected processes. Typically, process mining initiatives consume substantial resources, such as computing resources, but also expensive experts such as process owners or business analysts. Event-data-driven process prioritization approaches considering process interdependencies can be the missing part of the puzzle to ensure allocating these scarce resources to the most critical and central processes.

Third, process managers strongly require forward-directed operational support. Traditionally, process mining approaches focused on historical data for backward-looking, descriptive process mining (e.g., discovering process models). Descriptive process mining is an excellent starting point to improve processes. However, process managers need operational in their forward-looking day-to-day business [8]. As exemplary forward-looking predictive process mining use cases, predicting the behavior, performance, and outcomes of process instances help organizations act proactively in fast-changing environments. By combining process predictions with the decision area from normative process data (e.g., performance thresholds), prescriptive process mining approaches are able to trigger actions autonomously, e.g., by scheduling improvement projects [8]. The increasing volume of data (i.e., event records and event properties) offers new opportunities and poses significant challenges for predictive monitoring methods.

Visualized in Figure 1, BPM strives for connecting the real-world – physical in nature – with the digital world enabling value co-creation between human beings and machines (i.e., physical machines or software systems). The physical world consists of actors interacting with physical resources. Commonly, actors and resources are orchestrated through processes that relate to PAIS, creating a digital footprint (i.e., events) of each performed process activity. As the digital world's central element, the event log can be seen as a digital twin of the actual processes. Physical actors might also interact with non-PAIS or perform manual activities that are not connected with the digital world and, consequently, are not covered by PAIS-generated logs. Inspired by the three challenges introduced above, this cumulative doctoral thesis consists of six research papers that are assigned to the field of BPM and process mining, as indicated in Figure 1.

To lower the barriers for non-data-engineers to extract appropriate event logs, research paper #1 [14] presents RDB2Log, a semi-automated, quality-informed approach to event log generation from relational databases. RDB2Log takes a relational database as input and assesses its data quality based on standard data quality dimensions. Thus, RDB2Log supports mapping data columns to event log attributes and generating an appropriate event log. The artifact proposed in this research paper #1 is envisioned as a step towards a process data quality lifecycle: systematic detection, repair, and tracking of data quality issues. By providing a graphical interface that helps users extract high-quality event logs, research paper #1 also strives to improve usability for non-experts.

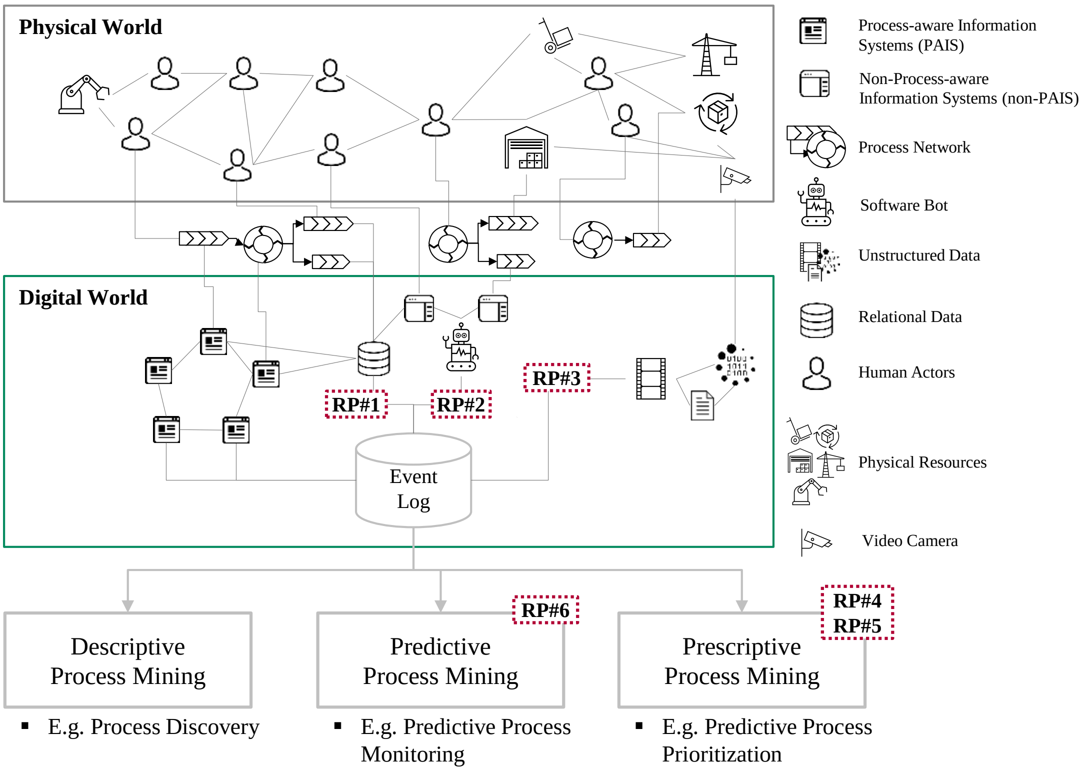

Figure 1: Assignment of individual Research Papers to forward-directed Process Mining

Research paper #2 [15] proposes an approach enabling integrated analysis using bot and process logs that provides new insights into bot-human interaction. An integrated analysis of bot and process data can also show the effects of bots on business processes and explore how exceptions are handled. Joint data analysis of bot and process data might also benefit the redesign of bots used in business processes. As a central artifact, research paper #2 proposes an integrated conceptual data model specifying the relations between bots and business processes. Based on this data model, it is possible to merge bot logs and process logs, allowing for integrated analysis.

Research paper #3 [16] focuses on analyzing manual processes that are not supported by process-aware information systems. In the case of solely relying on process-aware information systems (PAIS) that directly output minable event logs, the risks of neglecting process-relevant information arise, and so-called blind spots can occur. By providing an initial idea of how video data can be leveraged for process mining purposes, research paper #3 strives to exploit valuable process-relevant information beyond structured data sources bearing the potential to broaden the coverage of process mining analysis substantially.

To decide which processes should be in focus of process mining initiatives, process prioritization can be applied. Research paper #4 [17] proposes the Data-driven Process Prioritization approach (D2P2), leveraging performance and dependency data from process logs to determine the risky performance of all involved processes. Thereby, the D2P2 accounts for structural dependencies (e.g., processes that use other processes) and stochastic dependencies (e.g., instances that affect other instances of the same process). Based on the dependency-adjusted risky process performance, the D2P2 predicts when each process is likely to violate predefined performance thresholds and schedules it for in-depth analysis to future planning periods. Process analysts can then check whether the process under consideration requires improvement. Based on event log data, the D2P2's output is more reliable and detailed than other process prioritization approaches.

While D2P2 ends up providing a prioritized list of process candidates for an in-depth analysis, research paper #5 [18] expands the scope of process prioritization to schedule improvement projects providing even more prescriptive support. To do so, research paper #5 proposes the PMP2 drawing on the main concepts of D2P2 and extends an economic decision model optimizing the assignment of improvement project alternatives. By combining Markov reward models and normative analytical modeling, PMP2 helps organizations determine business process improvement roadmaps (i.e., sequential implementation of improvement projects on business processes), which maximize an organization's long-term firm value while catering for process dependencies and interactions among projects. Thereby, PMP2 takes a multi-period, multi-process, and multi-project perspective. Thus, the PMP2 considers dependencies between processes and improvement projects and thus schedules improvement projects to optimize an organization's long-term firm value.

Research paper #6 [19] explores the third challenge of providing operational, forward-directed support to process managers by extensively comparing the performance of different ML (i.e., Random Forests and Support Vector Machines) and DL (i.e., simple feedforward Deep Neural Networks and Long Short Term Memory Networks) techniques for a diverse set of five publicly available logs in terms of established evaluation metrics (i.e., Accuracy, F-Score, and ROC AUC). To provide generalizable results, research paper #6 combines data-to-description and description-to-theory strategies [20]. Also referred to as Level-1 inference, data-to-description generalization takes empirical data as input, condensed into higher-level yet still empirical observations or descriptions [20]. In a nutshell, the observations led to conclude that the application of DL is specifically promising when it comes to variant-rich processes producing a vast amount of data during runtime.

In sum, the thesis contributes to the existing body of knowledge on data-driven management of interconnected business processes. Hence, this thesis provides a basis for applying process mining in a forward-looking view and, thus, supports researchers and practitioners on the journey of converting project-based and isolated process mining initiatives to an ongoing supplement to the core of traditional BPM methods.

References

[1] Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer, Berlin, Heidelberg (2018)

[2] Recker, J., Mendling, J.: The State of the Art of Business Process Management Research as Published in the BPM Conference. Business & Information Systems Engineering, vol. 58, 55–72 (2016). doi: 10.1007/s12599-015-0411-3

[3] Vanwersch, R.J.B., Shahzad, K., Vanderfeesten, I., Vanhaecht, K., Grefen, P., Pintelon, L., Mendling, J., van Merode, G.G., Reijers, H.A.: A Critical Evaluation and Framework of Business Process Improvement Methods. Business & Information Systems Engineering, vol. 58, 43–53 (2016).
doi: 10.1007/s12599-015-0417-x

[4] Harmon, P.: The state of business process management 2020, vol. (2020)

[5] Diba, K., Batoulis, K., Weidlich, M., Weske, M.: Extraction, correlation, and abstraction of event data for process mining. WIREs Data Mining Knowl Discov, vol. 10 (2020). doi: 10.1002/widm.1346

[6] van der Aalst, W. (ed.): Process Mining. Data Science in Action, vol. . Springer Berlin Heidelberg, Berlin, Heidelberg (2016). doi: 10.1007/978-3-662-49851-4

[7] van der Aalst, W., Adriansyah, A., Medeiros, A.K.A. de, Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., et al.: Process Mining Manifesto. In: BPM 2011 Workshops Proceedings, vol. 99, pp. 169–194 (2011). doi: 10.1007/978-3-642-28108-2_19

[8] van der Aalst, W.: Academic View: Development of the Process Mining Discipline. In: Reinkemeyer, L. (ed.) Process mining in action. Principles, use cases and outlook, pp. 181–196. Springer International Publishing, Cham (2020). doi: 10.1007/978-3-030-40172-6_21

[9] Viner, D., Stierle, M., Matzner, M.: A Process Mining Software Comparison. In: ICPM 2020 Proceedings (2020)

[10] Browne, R.: How three friends turned a college project into a $2.5 billion software unicorn. CNBC, vol. (2019)

[11] Research and Markets: Process Analytics Market by Process Mining Type (Process Discovery, Process Conformance & Process Enhancement), Deployment Type, Organization Size, Application (Business Process, It Process, & Customer Interaction) & Region - Global Forecast to

2023 (2020), https://www.researchandmarkets.com/reports/4576970/process-analytics-market-by-process-mining-type
[12] vom Brocke, J., Jans, M., Mendling, J., Reijers, H.A.: Process Mining at the Enterprise Level. Bus Inf Syst Eng, vol. 62, 185–187 (2020). doi: 10.1007/s12599-020-00630-7
[13] Li, J., Wang, H.J., Bai, X.: An intelligent approach to data extraction and task identification for process mining. Inf Syst Front, vol. 17, 1195–1208 (2015). doi: 10.1007/s10796-015-9564-3
[14] Andrews, R., van Dun, C., Wynn, M.T., Kratsch, W., Röglinger, M., ter Hofstede, A.: Quality-informed semi-automated event log generation for process mining. Decision Support Systems, vol. 132, 113265 (2020). doi: 10.1016/j.dss.2020.113265
[15] Egger, A., ter Hofstede, A.H.M., Kratsch, W., Leemans, S.J.J., Röglinger, M., Wynn, M.T.: Bot Log Mining: Using Logs from Robotic Process Automation for Process Mining. In: ER 2020 Proceedings, vol. 12400, pp. 51–61 (2020). doi: 10.1007/978-3-030-62522-1_4
[16] Kratsch, W., König, F., Röglinger, M.: Shedding Light on Blind Spots: Developing a Reference Architecture to Leverage Video Data for Process Mining. Working Paper submitted to Information Systems and currently facing major revisions, preprint available on https://arxiv.org/pdf/2010.11289 (2020).
[17] Kratsch, W., Manderscheid, J., Reißner, D., Röglinger, M.: Data-driven Process Prioritization in Process Networks. Decision Support Systems, vol. 100, 27–40 (2017). doi: 10.1016/j.dss.2017.02.011
[18] Bitomsky, L., Huhn, J., Kratsch, W., Röglinger, M.: Process Meets Project Prioritization – A Decision Model for Developing Process Improvement Roadmaps. In: ECIS 2019 Proceedings (2019)
[19] Kratsch, W., Manderscheid, J., Röglinger, M., Seyfried, J.: Machine Learning in Business Process Monitoring: A Comparison of Deep Learning and Classical Approaches Used for Outcome Prediction. Bus Inf Syst Eng, vol. (2020). doi: 10.1007/s12599-020-00645-0
[20] Yin RK (1994). Case study research: design and methods (2nd ed). Sage, Thousand Oaks, Calif.