Stochastic Object-Centric Process Mining: Analysing Object Interaction Patterns

Jan Niklas van Detten[1,2]

[1]RWTH Aachen, Aachen
[2]Celonis, Munich

Abstract
The research area of process mining provides techniques to model and analyze business processes based on digital execution track records, called event logs. Traditionally, such logs contain isolated process instances that describe the behaviour of a single business object. However, real-life processes rarely exist in such a singe-object-vacuum. Instead, they often consist of multiple, interacting business objects of different types. Interactions between such object include, for example, the formation of object groups and the usage of resources. In practice, interactions between objects are not guaranteed to be deterministic. That is, the number of participating objects per interaction, and the number of interactions per object, is not fixed, but subject to a probability distribution. These distributions, if modelled and analysed correctly, could significantly enhance the understanding of the underlying business process. Therefore, we plan to combine object-centric process mining techniques, in which interacting objects are analysed from a control flow perspective, with stochastic process mining principles, in which probabilistic concepts are applied. Our work includes modelling formalisms to describe object-centric processes with probabilistic properties, algorithms to automatically construct such models from object-centric logs, and conformance checking techniques to evaluate their quality.

Keywords
object-centric, stochastic, process mining

1. Proposal

The execution of modern business processes is often traced by automated digital systems. The research area of process mining attempts to optimize business processes based on *event logs* extracted from such systems. A common optimization strategy is to first construct a process model from an event log with *process discovery* algorithms. Subsequently, the model is qualitatively validated with *conformance checking* techniques and analysed for inefficient or unintended behaviour. Corresponding insights, ideally, translate to optimization potential in the underlying business process in terms of measurable quality metrics, such as time and money spent during execution.

Traditionally, event logs only describe the behaviour of a single business object in isolation, like an order. However, real-life processes often contain interacting business objects of different types, such as orders, items and machines. *Object-centric* process mining techniques attempt to address this reality by accounting for the intertwined control flows of different object types [1]. This enables the analysis of interactions between multiple, previously considered independent, business objects.

For analytical purposes, the nature of object interactions is of particular interest. The execution of an activity might for example refer to fixed object groups or require access to shared resources. These high-level interaction patterns imply different analysis technique in a real-life setting. A machine that interacts with arbitrary many, otherwise independent, items over time by performing a production step on them requires a different analytical focus than an order that encapsulates a fixed set of items. For the machine, resource-oriented performance measures, such as throughput or utilization, are relevant, while the adherence to a shared control flow is of key interest for the items of an order.

In addition to the nature of object interactions, the often non-deterministic extent to which they are present is relevant as well. In case of the machine, the binary question on whether it acts as a resource or not is of less business interest, than how often it does so. Similarly, stating that an order

ICPM 2024 Doctoral Consortium, October 14–18, 2024, Kongens Lyngby, Denmark
n.vandetten@bpm.rwth-aachen.de (J. N. v. Detten)

might involve multiple items is less expressive than describing the distribution of the actual number of items per order. For a set of machines or items, or business objects in general, such interaction numbers follow probability distributions that provide relevant information about the underlying process.

However, state-of-the-art approaches for object-centric process mining suffer from gaps in that regard. In particular, high-level object interaction patterns are not explicitly captured in process models. Additionally, properties of these interactions that are subject to probability distributions are not considered as such. In our work, we fill this gap by proposing modelling formalisms that explicitly capture interaction patterns and their stochastic properties. Additionally, we propose corresponding discovery techniques and conformance checking methods for them, to enable an enhanced analysis of object interaction patterns in object-centric processes with stochastic properties. Therefore, we address the following research questions:

- (**RQ1**) What are the most important high-level object interaction patterns in processes?
- (**RQ2**) How can these high-level interaction patterns be explicitly captured in process models?
- (**RQ3**) Which stochastic properties do these patterns exhibit and how can they be modelled?
- (**RQ4**) Which insights can be generated from object-centric models with probabilistic properties?

Our work covers a timeframe of three years, out of which ten months have passed. As a first step, we performed a literature review to identify the most relevant high-level interaction patterns (**RQ1**). This phase did not require evaluative methods, but served as a literature foundation for further research phases. We identified five key interaction patterns, only briefly described here. If any of these patterns are present in a business process, traditional techniques might lead to flawed insights.

- Divergence [2], which is associated to the presence of resource-like business objects. Formally it is defined by the same object of one type interacting with different objects of another type across multiple events. Traditional method can induce false orders between such events.
- Convergence [2], which indicates the presence of arbitrary object groups or batches in the process. It is formally defined as multiple objects of the same type being involved in the same event. The application of traditional techniques can lead to the duplication of such events.
- Deficiency [2], which indicates the independence of the process from the involvement of certain objects. In an event log, this pattern is present if some events do not involve any object of an object type. These events might hence not be considered by traditional techniques.
- Synchronization [3], which expresses that multiple objects remain in stable combinations across the process. Traditional techniques do not consider dependencies between such objects.
- Specialization [3], which denotes that certain objects specialize on different activities based on a property.

Existing object-centric modelling formalisms suffer from gaps with regards to representing these patterns. They either do not capture the patterns above explicitly, or lack process discovery and conformance checking techniques for them. For example, object-centric Petri nets [1] only implicitly represent divergence, convergence and deficiency and do not capture synchronizations and specialization at all. Consequently, conformance checking measures for these nets, such as [4] or [5] hence do not account for these patterns. Some specialized techniques for object synchronizations have been proposed in [6], but without an explicit representation of the remaining patterns or discovery techniques. Typed Petri nets with identifiers [7] offer the option to represent some of these patterns, but no discovery algorithm or conformance checking technique exists for them. Similar problems apply to other object-centric formalisms as well, since they either lack explicit representations of some of these patterns or fail to account for them at all. Additionally, no formalism considers the probabilistic properties of object interactions. Stochastic process mining techniques so far mostly focus on probability distributions of control flow aspects of traditional, single object, processes or additional properties such as time.

Therefore, we designed new modelling formalisms that explicitly cover all of the identified interaction patterns (**RQ2**) to subsequently extend them with probability distributions. The evaluation of these formalisms focused on formal guarantees. We addressed the two patterns of synchronization and specialization by introducing the notion of silent objects [3]. We proved that the inclusion of these patterns can only improve process models in terms of existing quality criteria. Additionally, we included automated methods to detect synchronizations and specializations in event logs. We covered the remaining patterns of divergence, convergence and deficiency by introducing the notion of object-centric process trees [8]. We showed that these interaction patterns can be explicitly captured, while providing a set of desirable behavioral guarantees. Additionally, we introduced a discovery algorithm to construct models with these guarantees from event logs.

Currently, we are combining our work on silent objects and object-centric process trees into a joint formalism. Additionally, we are working on conformance checking measures that explicitly account for the five interactions patterns above. We will use those to optimize the discovery techniques that we introduced so far, by extending previously proposed work into the object-centric setting [9].

Afterwards, we will extend all of our techniques with stochastic properties of the identified interaction patterns. This includes extending our formalisms with probability distributions, discovering them in event logs and adapting our conformance checking measures to account for them. In particular, we intend to focus on incorporating distributions that describe the following aspects (**RQ3**):

- How likely is each high-level interaction pattern to occur?
- In how many interactions is a given object involved?
- How many objects participate in a given interaction?

Technically, we address these aspect by introducing a stochastic extension of object-centric process trees with silent objects. These trees already offer an explicit binary notion on whether each interaction pattern is present for each object type and activity. We will exchange this binary notion with a probability distribution that quantifies how likely each interaction pattern is present and to which extent. Subsequently, we will extend our discovery techniques and conformance checking to account for these distributions. We expect that existing techniques from stochastic process mining can be extended towards the interaction patterns for that purpose and combined with our work. Some work has, for example, already been done on comparing probability distributions for conformance checking.

As a result of our research, we will be able to answer multiple relevant questions for analytical purposes that have so far not been addressed. In particular, we will be able to give probabilistic insights on the likelihood of certain behaviour in a given object-centric model. Probabilistic properties of the interaction patterns could for example be used to determine how likely and often an object acts as a resource (divergence), how many objects are most likely needed to execute an activity (convergence) and how likely an execution of an event has to wait for objects of a given type (deficiency).

In parallel to the research described above we perform additional practical projects with our industrial partner Celonis to test the real-life applicability of our concepts (**RQ4**). While these projects are of course confidential, they give us valuable insights based on large, real-life data sets, user experiences and the opinions of domain experts. So far, the industrial applications have produced promising results, in particular on the object interaction pattern of divergence. A corresponding paper has been accepted and will be published soon [10]. Additional, application oriented, papers will follow as our approaches are integrated and applied in real-life use cases.

References

[1] W. M. P. van der Aalst, A. Berti, Discovering object-centric petri nets, Fundam. Informaticae 175 (2020) 1–40. URL: https://doi.org/10.3233/FI-2020-1946. doi:10.3233/FI-2020-1946.

[2] J. N. Adams, W. M. P. van der Aalst, Addressing convergence, divergence, and deficiency issues, in: Business Process Management Workshops - BPM International Workshops, volume 492 of *Lecture Notes in Business Information Processing*, Springer, 2023, pp. 496–507. URL: https://doi.org/10.1007/978-3-031-50974-2_37. doi:10.1007/978-3-031-50974-2_37.

[3] J. N. van Detten, P. Schumacher, S. J. Leemans, Object synchronization and specialization with silent objects in object-centric petri nets, in: Business Process Management Proceedings - BPM (in press), 2024.

[4] L. Liss, J. N. Adams, W. M. P. van der Aalst, Object-centric alignments, in: J. P. A. Almeida, J. Borbinha, G. Guizzardi, S. Link, J. Zdravkovic (Eds.), Conceptual Modeling - 42nd International Conference, ER 2023, Lisbon, Portugal, November 6-9, 2023, Proceedings, volume 14320 of *Lecture Notes in Computer Science*, Springer, 2023. URL: https://doi.org/10.1007/978-3-031-47262-6_11. doi:10.1007/978-3-031-47262-6_11.

[5] J. N. Adams, W. M. P. van der Aalst, Precision and fitness in object-centric process mining, in: C. D. Ciccio, C. D. Francescomarino, P. Soffer (Eds.), 3rd International Conference on Process Mining, ICPM 2021, Eindhoven, The Netherlands, October 31 - Nov. 4, 2021, 2021. URL: https://doi.org/10.1109/ICPM53251.2021.9576886. doi:10.1109/ICPM53251.2021.9576886.

[6] A. Gianola, M. Montali, S. Winkler, Object-centric conformance alignments with synchronization, in: G. Guizzardi, F. M. Santoro, H. Mouratidis, P. Soffer (Eds.), Advanced Information Systems Engineering - 36th International Conference, volume 14663 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 3–19. URL: https://doi.org/10.1007/978-3-031-61057-8_1. doi:10.1007/978-3-031-61057-8_1.

[7] J. M. E. M. v.d. Werf, A. Rivkin, A. Polyvyanyy, M. Montali, Data and process resonance - identifier soundness for models of information systems, in: Application and Theory of Petri Nets - 43rd International Conference„ volume 13288 of *Lecture Notes in Computer Science*, 2022. URL: https://doi.org/10.1007/978-3-031-06653-5_19. doi:10.1007/978-3-031-06653-5_19.

[8] J. N. van Detten, P. Schumacher, S. J. Leemans, Discpovering compact, live and identifier-sound object-centric models, in: International Conference On Process Mining Proceedings - ICPM (in press), 2024.

[9] J. N. van Detten, P. Schumacher, S. J. J. Leemans, An approximate inductive miner, in: 5th International Conference on Process Mining, ICPM 2023, Rome, Italy, October 23-27, 2023, IEEE, 2023, pp. 129–136. URL: https://doi.org/10.1109/ICPM60904.2023.10271971. doi:10.1109/ICPM60904.2023.10271971.

[10] J. N. van Detten, P. Schumacher, S. J. Leemans, A framework for advanced case notions in object-centric process mining, in: International Conference On Process Mining Proceedings - Workshops ICPM (in press), 2024.