

Concept-drift-aware Prescriptive Analytics for Object-centric Processes

Ngoc-Diem Le¹

¹University of Padua, Italy

Abstract

In the field of Process Mining, Process-aware recommender systems are designed to monitor process executions, predict future behavior, and recommend effective interventions based on the knowledge learned from historical event logs to reduce the risk of failure or to maximize a given reference Key Performance Indicator. In reality, concept drift in business processes involves changes in the underlying process dynamics over time, which can invalidate prediction models. Additionally, in complex business environments that involve multiple events and objects interacting, traditional recommender systems may struggle to accurately model these interactions, leading to less precise recommendations. Due to the intricate nature of the event data, existing methods for generating recommendations may be inadequate as they fail to account for the dependencies and relationships between object types. This Ph.D. project aims to develop a recommendation module that can generate meaningful and applicable recommendations. The objective is to apply counterfactual reasoning to provide clear insights into how particular changes from input can lead to desired results. Additionally, the project intends to adapt the system to handle concept drift effectively through continuous learning. This adjustment ensures that the recommendations maintain their accuracy and effectiveness even when the underlying process dynamics alter. Finally, the project seeks to provide recommendations for object-centric processes. This paradigm is gaining popularity for its ability to model processes more accurately and detailedly.

Keywords

Process Prescriptive Analytics, Process-aware Recommendation systems, Concept drift detection, Counterfactuals, Object-centric Process Mining,

1. Research Problem and Motivation

In recent years, process mining has emerged as a research area bridging between data science and process science. Its goal is to discover, monitor, and enhance actual processes by extracting information from event logs, helping businesses identify inefficiencies and bottlenecks and optimize overall performance [1].

Recommender systems are a particular class of Information Systems that aim to analyze user data and behavior to provide personalized recommendations, thereby enhancing user experience and decision-making. In the context of Process Mining, there is a growing trend towards the development of Process-aware recommender systems (hereafter referred to as PAR systems). Conceptually, a PAR system comprises three sub-systems: monitoring, predictive

Proceedings of the Best BPM Dissertation Award, Doctoral Consortium, and Demonstrations & Resources Forum co-located with 22nd International Conference on Business Process Management (BPM 2024), Krakow, Poland, September 1st to 6th, 2024.

✉ ngocdiem.le@phd.unipd.it (N. Le)

🌐 <https://github.com/ngocdiemle296> (N. Le)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

analytics block, and prescriptive analytics block. These systems are intended to (i) forecast process execution outcomes, (ii) identify high-risk executions that may fail to meet performance goals, such as costs, deadlines, and customer satisfaction, and (iii) recommend contingency actions to improve the executions [2]. Generally, an outcome is defined through a so-called process-specific KPI (Key Performance Indicator) that measures a process' performance and highlights improvement areas.

Nevertheless, there is a lack of comprehensive studies on the application of prescriptive analytics for dynamically generating the optimal combination of the next activity and resource allocation in real time. Current recommendation modules predominantly focus on recommending the next activity, with limited capabilities in simultaneously recommending the appropriate resources for these activities. This gap is significant, as the impact of the recommended activity on the process outcome can be highly dependent on the specific resources assigned to perform it.

The purpose of this Ph.D. project is to construct a recommendation module that suggests the optimal combinations of the next activity and resource in a workflow. The current work of Padella et al. [3] recommends the next activity by testing the improvement for every possible combination of activity and resource is unfeasible in practice, especially for processes involving hundreds of potential resources. This leads to our primary research question.

Research Question 1: *How can we build a recommendation module that can efficiently help improve a given reference KPI?*

Next, in reality, processes often change due to various factors such as market dynamics, new regulations, or efforts toward process enhancement and repair [4]. These changes, as known as concept drift, challenge the underlying assumptions of classical methods, which typically assume that processes are in a steady state [5]. The occurrence of concept drift in the recommendation module can result in inadequate recommendations. This is because the prediction and recommendation models may no longer accurately capture the current dynamics of the workflow. As the process evolves, previously effective models may become outdated, resulting in suboptimal or incorrect recommendations. This issue drives us to our second research question.

Research Question 2: *How do we maintain the effectiveness of recommendations through continuous learning as processes evolve?*

Traditional process mining methods usually assume that each event is associated with a single object. These considerations motivate the introduction of object-centric processes. This paradigm has been gaining attention for its ability to model inter-organizational processes naturally [6]. However, the transition from traditional single-flow processes to object-centric ones poses significant challenges, especially regarding the complexity of the input data. This makes generating recommendations particularly challenging when multiple object types and varying numbers of objects are involved with events. This brings us to the third research question.

Research Question 3: *How to expand the use of our framework beyond conventional single-flow processes?*

2. Literature Analysis

In the last decade, a significant amount of research has focused on process predictive analytics. These methods have been developed to predict various KPIs, tackle challenges from diverse perspectives, and have been applied across multiple domains. There are comprehensive reviews of predictive monitoring works before 2022 by Márquez-Chamorro et al. [7], Di Francescomarino et al. [8], and Rama-Maneiro et al. [9]. In addition, some recent papers [10, 11, 3] focus on predicting time-related outcomes, total costs, and activity occurrences, where the latter involves forecasting whether a specific activity is likely to happen in the future. Galanti et al. [11] proposed a predictive analytics framework to compare the effectiveness of gradient boosting (specifically, Catboost) and Long Short-Term Memory (LSTM) models on several real-life case studies for predicting various KPIs. While both models provided similar prediction quality, the experiments demonstrated that CatBoost significantly reduced the model's training time.

Once predictive analytics provides forecasts, prescriptive analytics uses these insights to suggest specific actions to optimize outcomes. However, the benefits of prescriptive process monitoring can only be fully realized if these methods prescribe effective interventions that are followed [12]. Kubrak et al. [13] proposed a thorough study on prescriptive process monitoring, discussing its evolution and future prospects and focusing on how predictive analytics can be combined with optimization techniques to recommend actionable interventions in business processes. However, the paper also mentions the lack of methods for explainability and feedback loops between a prescriptive monitoring system and its end-users. In response to this, Padella et al. [3] proposed a framework to accompany recommendations with sensible explanations based on process-related characteristics, using Shapley values [14] to generate explanations for the selected recommendations.

Sato et al. [4] conducted a comprehensive survey on concept drift detection in process mining, categorizing approaches into two main branches: explicit drift detection and adapting process mining techniques to handle event streams in evolving environments. However, none of the papers reviewed in [4] addressed the resource perspective. Adams et al. [15] generalized the concept drift detection framework using PELT (Pruned Exact Linear Time) proposed in [16] by allowing for the testing of non-linear relationships, which can be potentially applied to resource allocation within information systems. Moreover, their approach supports object-centric event logs. A recent study [17] extended control-flow drift detection to multi-perspective drift detection by extracting features from a multi-layered event knowledge graph (EKG). The idea is to aggregate information in an EKG with actor and case paths to gain new insights into actor behavior and task handovers and utilize an existing change point detection technique proposed in [15]. However, it is worth noting that concept drift in process mining is still primarily being focused on offline analysis. Accordingly, Hassani [18] introduced a method for detecting concept drift in event streams by employing ADWIN (Adaptive Windowing) [19], an adaptive window technique. This approach dynamically adjusts the window size by focusing on short intervals for highly deviating periods or increasing its width in case of uniform observations. Additionally, by integrating the advantages of both Heuristic Miner and Fuzzy Miner, it enhances the detection and adaptation to changes in process behavior. Nevertheless, there remains a lack of research that focuses on the problem of concept drift in terms of optimizing the project-specific KPIs and their impact on the recommendations

generated by the PAR system.

The object-centric process paradigm is increasingly gaining traction in both academia and industry [20]. While existing discovery algorithms focus on modeling the control flow of processes, they often struggle to effectively represent the interactions and dependencies between different objects in event data. Tour et al. [21] introduced Agent Miner, a divide-and-conquer algorithm that constructs models of agents and their interactions from event data, providing a new perspective on how agents work together to perform activities. Besides, Klijn et al. [22] proposed a new aggregation method that focuses on analyzing task executions in event knowledge graphs. Regarding predictive analytics, some studies [20, 23, 24] have proposed approaches to incorporate information about object interactions into the predictive model. The first two papers [20, 23] showed promising results when using techniques based on gradient boosting, particularly the Catboost model. In addition, Adams et al. [24] examined the impact of flattening and the potential benefits of object-centric innovations in predictive process monitoring. Despite these advancements, in the context of prescriptive analytics, no known research has been conducted so far on providing recommendations for object-centric processes.

3. Project Roadmap

The project aims to develop a recommendation module within the PAR system that suggests optimal combinations of the next activity and resource in a workflow, specifically targeting administrative processes. Initially, the construction of the recommendation module requires the development of a predictive analytics component. Due to its ease of implementation and significant advantage in training time, we chose the CatBoost model [25] to predict the KPIs of interest.

In recent years, counterfactual reasoning has become increasingly crucial in many aspects of process mining. This method seeks to demonstrate how modifying a real input instance can lead to significant changes in the output based on what the machine learning prediction model has learned. It is important to distinguish our approach from counterfactuals used in causal machine learning. In our case, we are not estimating causal effects but rather aiming to understand how changes in input affect the output. In the PAR system, counterfactual reasoning is mostly applied in the field of predictive analytics [26, 27, 28]. However, to the best of our knowledge, this approach has yet to be explored in the context of prescriptive analytics to recommend the next activities and resources while optimizing process-specific KPIs. Therefore, to address research question 1, we have developed a recommendation module that uses counterfactuals to generate a list of recommendations that contains potential next activities and resources. Specifically, we employ the DiCE (Diverse Counterfactual Explanations) algorithm [29] to explore the potential KPIs that would occur if a different combination of activity and resource was taken at a decision point in the process. Within DiCE, the Catboost model predicts potential KPIs for these alternative scenarios. The combinations of the next activity and resource are considered hypothetical alternatives to the actual events that have happened. By using the data about historical events and features, DiCE can create a series of “what-if” scenarios by altering features related to the next activity and resource while keeping others constant. By assessing these “what-if” scenarios, the system can infer potential KPIs without requiring exhaustive

testing, making it more feasible in practice. Through this methodology, the recommendation module can effectively suggest optimal activity-resource combinations, thereby enhancing workflow efficiency and productivity. The advantage of DiCE is that it uses genetic approaches that prevent testing all combinations of activities and resources. We are aware that the solutions proposed are not necessarily optimal, but we expect DiCE to give very good solutions while keeping the problem tractable.

To assess the effectiveness of our recommendation module, we need to compare the KPI values for the process instances being executed, when the recommendations are followed versus when they are ignored. The best would be to have an A/B testing with the system in production, but that is unfeasible in practice because companies do not want to put their business at risk. In [3], Padella et al. use the event log of the past executions to find traces that are similar to those subjected to recommendations and assume that the latter ones would behave similarly under the same recommendations. Unfortunately, it is often the case that you cannot find similar traces, invalidating the proposal. In this project, we aim to use business process simulation to generate executions that do or do not follow the recommendations, similar to what was proposed by Padella et al. [30]. Simulated data can help generate a wide range of hypothetical scenarios that may not have occurred in the past. We have already achieved preliminary results at this step in our project. We will soon compile a paper on our findings, which we aim to submit to a workshop at an upcoming conference.

The concept drift problem in Process Mining underscores the importance of continuously monitoring and updating the model to maintain system effectiveness. In addressing research question 2, we intend to detect changes in data patterns while keeping the predictive model up-to-date. Since our goal is to recommend optimal combinations of the next activity and resource, it is crucial to consider both actor behavior and control flow together when detecting concept drift. Therefore, we intend to leverage the framework proposed in [17] for multi-perspective concept-drift detection by extracting features from a multi-layered EKG. To enhance the recommendation module against concept drift, we plan to leverage CatBoost's support for continuous learning. By utilizing CatBoost's capabilities to adapt incrementally to new data, we aim to maintain the system's predictive accuracy and effectiveness as patterns evolve.

Regarding research question 3, we aim to enhance our concept-drift-aware prescriptive analytics framework with object-centric processes. Specifically, we plan to adapt the Catboost model with object-centric processes within the predictive analytics block to achieve optimal performance. As highlighted by [20], the Catboost model provides similar results to LSTM and graph-based neural networks in terms of accuracy. However, it is significantly faster than the methods relying on graph-based neural networks when applied in object-centric processes. However, a significant challenge we encounter involves determining what constitutes a recommendation for object-centric processes. Due to the intricate nature of the event data, a recommendation for object-centric processes should consider the multiple dimensions and interactions of the various objects involved. Therefore, the recommendations generated by our initial framework may be inadequate as they fail to account for the dependencies and relationships between different object types. To address this, it is crucial to develop a more advanced framework capable of comprehensively identifying and analyzing these interactions. This approach will ensure that the recommendations are well-informed and tailored to meet the particular requirements and conditions of the interacting objects.

References

- [1] W. M. P. van der Aalst, *Process Mining: Data Science in Action*, Springer, 2016.
- [2] M. de Leoni, M. Dees, L. Reulink, Design and evaluation of a process-aware recommender system based on prescriptive analytics, in: *Proceedings of the 2nd International Conference on Process Mining (ICPM 2020)*, IEEE, 2020.
- [3] A. Padella, M. de Leoni, O. Dogan, R. Galanti, Explainable process prescriptive analytics, in: *Proceedings of the 4th International Conference on Process Mining (ICPM 2022)*, IEEE, 2022.
- [4] D. M. V. Sato, S. C. De Freitas, J. P. Barddal, E. E. Scalabrin, A survey on concept drift in process mining, *ACM Computing Surveys* 54 (2022) 189:1–189:38.
- [5] R. J. C. Bose, W. M. P. van der Aalst, I. Žliobaitė, M. Pechenizkiy, Dealing with concept drifts in process mining, *IEEE Transactions on neural networks and learning systems* 25 (2013) 154–171.
- [6] W. M. P. van der Aalst, Object-centric process mining: Dealing with divergence and convergence in event data, in: *Proceedings of the 17th International Conference on Software Engineering and Formal Methods (SEFM 2019)*, Springer, 2019, pp. 3–25.
- [7] A. E. Márquez-Chamorro, M. Resinas, A. Ruiz-Cortés, Predictive monitoring of business processes: a survey, *IEEE Transactions on Services Computing* 11 (2017) 962–977.
- [8] C. Di Francescomarino, C. Ghidini, F. M. Maggi, F. Milani, Predictive process monitoring methods: Which one suits me best?, in: *Proceedings of the 16th International Conference on Business Process Management (BPM 2018)*, Springer, 2018, pp. 462–479.
- [9] E. Rama-Maneiro, J. C. Vidal, M. Lama, Deep learning for predictive business process monitoring: Review and benchmark, *IEEE Transactions on Services Computing* 16 (2021) 739–756.
- [10] R. Galanti, B. Coma-Puig, M. de Leoni, J. Carmona, N. Navarin, Explainable predictive process monitoring, in: *Proceedings of the 2nd International Conference on Process Mining (ICPM 2020)*, IEEE, 2020, pp. 1–8.
- [11] R. Galanti, M. de Leoni, M. Monaro, N. Navarin, A. Marazzi, B. Di Stasi, S. Maldera, An explainable decision support system for predictive process analytics, *Engineering Applications of Artificial Intelligence* 120 (2023) 105904.
- [12] M. Dees, M. de Leoni, W. M. van der Aalst, H. A. Reijers, What if process predictions are not followed by good recommendations?(technical report), *arXiv preprint arXiv:1905.10173* (2019).
- [13] K. Kubrak, F. Milani, A. Nolte, M. Dumas, Prescriptive process monitoring: Quo vadis?, *PeerJ Comput. Sci.* 8 (2022) e1097.
- [14] L. S. Shapley, et al., A value for n-person games (1953).
- [15] J. N. Adams, S. J. van Zelst, T. Rose, W. M. van der Aalst, Explainable concept drift in process mining, *Information Systems* 114 (2023) 102177.
- [16] J. Adams, S. van Zelst, L. Quack, K. Hausmann, W. M. P. van der Aalst, T. Rose, A framework for explainable concept drift detection in process mining, in: *Proceedings of the 19th International Conference on Business Process Management (BPM 2021)*, 2021.
- [17] E. L. Klijn, F. Mannhardt, D. Fahland, Multi-perspective concept drift detection: Including the actor perspective, in: *Proceedings of the International Conference on Advanced*

- Information Systems Engineering, Springer, 2024, pp. 141–157.
- [18] M. Hassani, Concept drift detection of event streams using an adaptive window, in: Proceedings of the 33rd International ECMS Conference on Modelling and Simulation, ECMS 2019, 2019, pp. 230–239.
 - [19] A. Bifet, R. Gavaldà, Learning from time-changing data with adaptive windowing, in: Proceedings of the 7th International Conference on Data Mining (2007), SIAM, 2007, pp. 443–448.
 - [20] R. Galanti, M. de Leoni, Predictive analytics for object-centric processes: Do graph neural networks really help?, in: Proceedings of the 21st International Conference on Business Process Management (BPM 2023), Springer, 2023, pp. 521–533.
 - [21] A. Tour, A. Polyvyanyy, A. Kalenkova, A. Senderovich, Agent miner: an algorithm for discovering agent systems from event data, in: Proceedings of the International Conference on Business Process Management, Springer, 2023, pp. 284–302.
 - [22] E. L. Klijn, F. Mannhardt, D. Fahland, Aggregating event knowledge graphs for task analysis, in: Proceedings of the International Conference on Process Mining, Springer, 2022, pp. 493–505.
 - [23] R. Galanti, M. de Leoni, N. Navarin, A. Marazzi, Object-centric process predictive analytics, *Expert Systems with Applications* 213 (2023) 119173.
 - [24] J. N. Adams, H. Drescher, A. Swoboda, N. Günnemann, G. Park, W. M. P. van der Aalst, Improving predictive process monitoring using object-centric process mining, in: Proceedings of the 32nd European Conference on Information Systems (ECIS), 2024.
 - [25] A. V. Dorogush, V. Ershov, A. Gulin, Catboost: Gradient boosting with categorical features support, *arXiv preprint arXiv:1810.11363* (2018).
 - [26] C. Hsieh, C. Moreira, C. Ouyang, DiCE4EL: Interpreting process predictions using a milestone-aware counterfactual approach, in: Proceedings of the 3rd International Conference on Process Mining (ICPM 2021), IEEE, 2021, pp. 88–95.
 - [27] A. Buliga, C. Di Francescomarino, C. Ghidini, F. M. Maggi, Counterfactuals and ways to build them: Evaluating approaches in predictive process monitoring, in: Proceedings of the 35th International Conference on Advanced Information Systems Engineering (CAiSE 2023), Springer, 2023, pp. 558–574.
 - [28] A. Stevens, C. Ouyang, J. De Smedt, C. Moreira, Generating feasible and plausible counterfactual explanations for outcome prediction of business processes, *arXiv preprint arXiv:2403.09232* (2024).
 - [29] R. K. Mothilal, A. Sharma, C. Tan, Explaining machine learning classifiers through diverse counterfactual explanations, in: Proceedings of the 3rd ACM Conference on Fairness, Accountability, and Transparency (ACM FAT 2020), 2020, pp. 607–617.
 - [30] A. Padella, F. Mannhardt, F. Vinci, M. de Leoni, I. Vanderfeesten, Experience-based resource allocation for remaining time optimization, in: Proceedings of the 22nd International Conference on Business Process Management (BPM 2024), 2024. Accepted. In press.