Process and Resource-Aware Responsible Recommender Systems (Extended Abstract)

Alessandro Padella[1]

[1]University of Padua, Dipartimento di Matematica "Tullio Levi-Civita", Via Trieste, 63, 35131 Padova PD

Abstract

In recent years, predictive and prescriptive analytics have become increasingly valuable in optimizing business processes by enabling data-driven decision-making. This thesis focuses on the enhancement of these systems in terms of fairness, accuracy, and transparency. First, a fairness-aware predictive framework is proposed, leveraging adversarial learning techniques to ensure that attributes that have to be protected, such as gender or ethnicity do not influence prediction outcomes. Experimental results demonstrate a significant reduction in biased predictions and recommendations. The thesis also addresses the issue of limited or imbalanced event logs, which can affect the training of reliable recommendation models. A comparative evaluation of current event-log augmentation methods is conducted, followed by the introduction of a novel augmentation approach based on statistical sampling. This method is shown to outperform state-of-the-art techniques in generating synthetic event logs that closely resemble real-world data distributions. Furthermore, the thesis presents two resource allocation frameworks to improve the global efficiency of business processes. The first generates recommendations that are globally optimal while allowing flexibility for local decisions of resources, about which task to perform as next. The second framework ensures a balanced workload distribution among process participants, addressing practical constraints in real-world resource management. Lastly, to support transparency and user trust, an explainable recommender system that recommends which task to perform as next is developed to accompany recommendations with interpretable justifications. By incorporating Shapley values into the recommendation model, the framework can provide meaningful insights into the rationale behind specific suggestions across different process domains.

Keywords

Process Mining, Deep Learning, Machine Learning, Resource Allocation, Explainability, Data Augmentation

1. Introduction

This PhD thesis focuses on the design and enhancement of Process-Aware Recommender Systems (PARs), a class of information systems that aims to monitor business processes, predict their outcomes, and eventually recommend corrective actions to improve performance. These systems are increasingly data-driven and rely on techniques from machine-and-deep learning to optimize the so called Key Performance Indicators (KPIs) such as cost, execution time, and customer satisfaction [1]. PAR systems are composed of three main components: Process Monitoring, Process Predictive Analytics, and Process Prescriptive Analytics. Process Monitoring enables real-time tracking of ongoing processes, Process Predictive Analytics leverages historical and real-time data to forecast future process behaviour, and Process Prescriptive Analytics provides actionable recommendations to provide feasible and effective interventions.

While these systems offer significant potential for process optimization, their development raises critical challenges related to the ethical, accurate, and transparent use of data. To address these challenges, this thesis focuses on a PAR system's predictive and prescriptive components following the **Responsible Data Science** concepts pointed out by Van der Aalst et al. in [2], that if applied to PARs, translates into four challenging research questions, three of which have been identified as the subject of study in the thesis:

Proceedings of the Best BPM Dissertation Award, Doctoral Consortium, and Demonstrations & Resources Forum co-located with 23rd International Conference on Business Process Management (BPM 2025), Seville, Spain, August 31st to September 5th, 2025.

alessandro.padella@unipd.it (A. Padella)
https://pado123.github.io/ (A. Padella)

- **RQ1 – Fairness**: *How can process predictions and recommendations avoid unfair conclusions, even when such conclusions are supported by historical data?*
  This question explores the integration of fairness-aware techniques into predictive models to mitigate bias toward protected attributes (e.g., gender, ethnicity). The goal is to prevent discriminatory outcomes in recommender systems by ensuring that predictions are not influenced by variables that should be ethically irrelevant. Recommender systems may inherit biases from training data, especially when demographic attributes are involved, leading to unfair outcomes [3]. Studies have shown that such biases can reinforce discrimination, prompting legal and ethical concerns (e.g., Article 2 of GDPR [4]). In Process-Aware Recommender Systems, these biases can influence corrective actions, disproportionately affecting certain groups. To prevent this, models must be trained and evaluated using fairness-aware techniques. Ensuring unbiased predictions is essential for ethical and responsible decision support.
- **RQ2 – Accuracy**: *How can we ensure process predictions and recommendations are accurate and reliable, especially when data is imbalanced or when resources must be shared across concurrent process instances?*
  Predictive models in business processes often struggle with rare but critical events due to imbalanced datasets, leading to inaccurate forecasts and suboptimal interventions. In scenarios like loan processing, this can affect the prediction of the occurrence of exceptional activities that are often related to problems and inefficiencies. Additionally, in a recommender system, resources are shared among multiple process instances, each typically performing one activity at a time. If resources are assigned to each instance independently, without considering others, overall efficiency suffers. For example, if resource R1 is given to instance P1, it cannot assist instance P2. However, assigning R2 to P1—even if it's slightly less effective—could be better if R1 is uniquely suitable for P2. This highlights the importance of evaluating resource allocation globally, not in isolation. Effective intervention decisions must consider all running instances to optimize overall outcomes.
- **RQ3 – Transparency**: *How can we provide transparent and understandable explanations for process predictions and recommendations to facilitate trust and human oversight?*
  Process Prescriptive Analytics often neglects interaction with human decision-makers, limiting its practical adoption. Despite high predictive accuracy, users may distrust recommendations without clear explanations. Explainable AI (XAI) is crucial for bridging this gap, enhancing transparency and trust. Studies show that understanding the reasoning behind predictions increases user confidence. Therefore, integrating interpretability into PAR systems is essential for effective human-AI collaboration.

The fourth question, i.e. Confidentiality, although relevant, is beyond the scope of this work and it has been identified as future work.

2. Research Approach and Outcomes

In light of the identified research questions, this PhD thesis outlines frameworks for extending the state-of-the-art of Predictive and Prescriptive Analytics following the identified research questions.

RQ1 – Fairness How can process predictions and recommendations avoid unfair conclusions, even when such conclusions are supported by historical data?

For addressing this research question in the thesis, a framework is developed to address unfair predictions in process analytics caused by bias from certain variables that should not influence the outcome and should therefore be considered as protected (e.g., gender, citizenship). It is important to state that solely removing the protected variables is insufficient, as bias can shift to correlated features, becoming a hidden variable, which is even harder to handle. The framework introduced in the thesis uses adversarial debiasing on a fully connected neural network [5]: the idea below adversarial debiasing is to train a predictive model based on a neural network while simultaneously training an adversary network to prevent learning from protected variables. This ensures the model focuses on legitimate, fair patterns. In the thesis, the protected variables have been defined on a case-by-case basis, resulting in 4 case studies on both real and synthetic datasets with 6 different protected variables. The experiments showed that the proposed model not only significantly reduces the influence of the protected variables, but is also able to maintain the same accuracy of the model in which the protected variables influence the KPI prediction. Furthermore, the predictive model is also shown to reduce the influence of the variables that are correlated with the protected ones, proving the initial statement. This work also resulted in the publication of a paper on the International Conference on Cooperative Information Systems [6].

RQ2 – Accuracy: How can we ensure process predictions and recommendations are accurate and reliable, especially when data is imbalanced or when resources must be shared across concurrent process instances?

This research question has been addressed under two points of view: i) Data Augmentation and ii) Resource Allocation. In this thesis, a framework for event-log augmentation to improve predictions of rare process behaviours has been developed. The approach addresses the challenge of imbalanced data by generating synthetic traces that reflect rare but important events. The thesis provide two contributions: (1) independently evaluating existing augmentation methods using consistent datasets and criteria from [7], and (2) proposing a novel technique that outperforms existing ones in quality and speed. The evaluation incorporates both established and new metrics. The Train-on-Synthetic-Test-on-Real [8] metric is finally used to assess how well synthetic logs replicate real patterns by measuring model performance trained on synthetic and tested on real data. The development of this framework led to the publication of a paper that is currently submitted to the Business & Information Systems Engineering journal.

ii) From the point of view of resource allocation, this thesis puts forward 2 different frameworks for resource allocation in Process Prescriptive Analytics. The first framework focuses on optimizing recommendations by jointly assigning both the next activity and the most appropriate resource for all ongoing process instances. Unlike traditional methods that make isolated decisions for each instance, this approach adopts a global optimisation perspective, aiming to maximize key performance indicators (KPIs) across the entire system. It takes into account real-world constraints, such as the fact that a resource can only work on one activity at a time, and it allows some flexibility in assigning sub-optimal local resources if doing so improves the overall system outcome, with the goal of leaving the resources free to choice the next task on which they will be working on. The development of this framework led to the publication of a paper at the Business Process Management conference [9].

The second framework integrates the concept of worker experience into the allocation process. Rather than defining experience based on resumes or generic seniority metrics, it uses historical process execution data as task execution times for the single resources for the respective tasks, to create a more data-driven measure of expertise. The goal here is twofold: not only to maintain high performance in terms of global KPIs, but also to ensure a fair distribution of workload and promote the development of less experienced workers by gradually assigning them more complex tasks. This contributes to long-term organizational learning and balances short-term efficiency with human-centric growth. The development of this framework led to the publication of a paper at the Business Process Management conference [10].

RQ3 – Transparency: How can we provide transparent and understandable explanations for process predictions and recommendations to facilitate trust and human oversight?

The thesis also proposes a framework for adding explanations to Process-aware Recommender Systems (PAR systems) to improve trust and understanding of recommendations. While existing systems focus on predicting and recommending activities for processes at risk, they often lack explanations for their suggestions. The proposed framework uses Shapley Values from game theory [11], a technique that is independent of the machine-learning technique used, and allows to decompose the input variables' influence to explain the rationale behind predictions and recommendations. Applied to a Process Prescriptive Analytics system using gradient boosting with decision trees, the framework was evaluated on real-life datasets, demonstrating its effectiveness in both improving recommendations and providing clear explanations. The development of this framework led to the publication of a paper [12].

3. Conclusions and Post-doctoral Research Directions

Process-Aware Recommender Systems analyses historical process execution data to evaluate performance against predefined objectives for ongoing process instances. Their primary goal is to provide real-time insights into process performance and offer actionable recommendations to correct instances likely to produce suboptimal outcomes, based on KPIs. These systems address three key research questions within the field of Responsible Data Science. Despite their performance benefits, these techniques face challenges such as explainability issues, the need for large datasets, and potential bias in training processes. These limitations have been noted across various machine learning frameworks used in Process Predictive Analytics. Each framework developed in this thesis has been experimentally validated across multiple case studies and event logs, considering various KPIs. All the code is publicly available for free use. Nevertheless, several future research paths remain, particularly those that involve integrating the individual approaches examined in this study. These will be considered by the post-doctoral researcher. One challenge is developing a system that addresses the Confidentiality research question: How can process predictions and recommendations be made while maintaining confidentiality? Blockchain techniques [13, 14] could be a potential solution by ensuring data integrity without a central authority. Next, integrating fairness, synthetic event log generation, and explainability into a unified framework would improve predictive accuracy, fairness, and interpretability. This could prevent unfair influences from protected variables and correct imbalances in underrepresented activities. Extending explainability further involves using large language models (LLMs) to develop a dynamic explanation engine that provides process actors with insights into factors contributing to suboptimal outcomes and recommendations for corrective actions, this research path has been already took into account in my post-doctoral period, leading to a publication in the Business Process Management Forum 2025. Additionally, future research could explore advanced graph-neural-network-based approaches to dynamically adapt recommendations over time based on resource performance. This would create an adaptive recommender system that continuously improves process and resource efficiency. Finally, A/B testing and surveys could be used to assess the quality of recommendations provided by the system.

Declaration on Generative AI

The authors declare that Generative AI tools were used solely to assist with language refinement and not for the generation of scientific content or analysis.

References

[1] P. Ceravolo, M. Comuzzi, J. De Weerdt, et al., Predictive process monitoring: Concepts, challenges, and future research directions, Process Science 1 (2024) 2. URL: https://doi.org/10.1007/s44311-024-00002-4. doi:10.1007/s44311-024-00002-4.

[2] W. M. P. van der Aalst, M. Bichler, A. Heinzl, Responsible data science, Business & Information Systems Engineering 59 (2017) 311–313. doi:10.1007/s12599-017-0487-z.

[3] F. Mannhardt, Responsible Process Mining, Springer, Cham, 2022, pp. 373–401.

[4] European Union, General data protection regulation (gdpr), 2018. URL: https://gdpr-info.eu/, accessed: 2024-11-26.

[5] J. Yoon, D. Jarrett, M. van der Schaar, Time-series generative adversarial networks, Curran Associates Inc., Red Hook, NY, USA, 2019, p. 11.

[6] M. de Leoni, A. Padella, Achieving Fairness in Predictive Process Analytics via Adversarial Learning, volume 15506 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 346–354.

[7] D. Chapela-Campa, I. Benchekroun, O. Baron, M. Dumas, D. Krass, A. Senderovich, Can i trust my simulation model? measuring the quality of business process simulation models, in: 21th International Conference on Business Process Management, BPM 2023, Proceedings, 2023.

[8] P. Conen, et al., Train on synthetic - test on real: Domain adaptation for strain-based damage detection on an aircraft wing, in: ICAS 2024, International Council of the Aeronautical Sciences, liely conference proceedings, 2024. URL: https://elib.dlr.de/204979/2/ICAS2024_Paper_PhilippConen.pdf, available in DLR eLibrary.
[9] A. Padella, M. de Leoni, Resource allocation in recommender systems for global kpi improvement, in: C. Di Francescomarino, A. Burattin, C. Janiesch, S. Sadiq (Eds.), Business Process Management Forum, Springer Nature Switzerland, Cham, 2023, pp. 249–266.
[10] A. Padella, F. Mannhardt, F. Vinci, M. de Leoni, I. Vanderfeesten, Experience-based resource allocation for remaining time optimization, in: A. Marrella, M. Resinas, M. Jans, M. Rosemann (Eds.), Business Process Management - 22nd International Conference, BPM 2024, Krakow, Poland, September 1-6, 2024, Proceedings, volume 14940 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 345–362.
[11] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in neural information processing systems, 2017, pp. 4765–4774.
[12] A. Padella, M. de Leoni, O. Dogan, R. Galanti, Explainable process prescriptive analytics, in: 2022 4th International Conference on Process Mining (ICPM), 2022, pp. 16–23. doi:10.1109/ICPM57379.2022.9980535.
[13] I. Weber, X. Xu, R. Riveret, G. Governatori, A. Ponomarev, J. Mendling, Untrusted business process monitoring and execution using blockchain, in: M. La Rosa, P. Loos, O. Pastor (Eds.), Business Process Management, Springer International Publishing, Cham, 2016, pp. 329–347.
[14] C. Cabanillas, C. Di Ciccio, J. Mendling, A. Baumgrass, Predictive task monitoring for business processes, in: S. Sadiq, P. Soffer, H. Völzer (Eds.), Business Process Management, Springer International Publishing, Cham, 2014, pp. 424–432.