Prescriptive Monitoring of Business Processes Under Uncertainty and Resource Constraints

Mahmoud Shoush[1]

[1]University of Tartu, Narva mnt 18, 51009 Tartu, Estonia

Abstract

Prescriptive process monitoring is a set of techniques that optimize business processes' performance by constructing intervention policies. Such policy monitors processes to identify undesired outcomes (e.g., customer dissatisfaction) and prescribe actions or interventions (e.g., proposing a customer discount) to prevent their costly occurrence. However, existing techniques predominantly rely on pure predictive models to predict undesired outcomes without considering the level of uncertainty associated with these predictions. Consequently, interventions are triggered based on uncertain predictions, assuming they are effective and can be immediately executed for all cases. The underpinning assumption is that a process worker always exists to execute the intervention, e.g., calling a customer to propose a discount. If executed, it reduces the cost of undesired outcomes. However, in practice, the resources available to organizations, such as personnel, may be limited, and we cannot observe the effectiveness of utilizing the intervention beforehand. To address the mentioned gaps, in this doctoral project, we aim to design and develop an intervention policy that optimally learns with confidence, during the execution time of the process, how to select and prioritize business cases to allocate resources to execute interventions and thus improve a given process objective, e.g., gain. To achieve this, we integrate machine learning methodologies. Initially, we use a predictive model with a confidence estimation method to identify cases with a higher likelihood of undesired outcomes. Then, we use a causal inference model to estimate the expected impact of an intervention on a case's outcome. Finally, we develop a resource allocator that monitors and assigns resources to selected cases when necessary.

Keywords

Prescriptive Process Monitoring, Machine Learning, Causal Inference

CEUR-WS.org/Vol-3407/paper6.pdf

1. Introduction

Prescriptive process monitoring(PrPM) is a process mining subfield that uses business process execution records (aka *event logs*) and machine learning techniques to monitor and improve business processes during runtime. Hence, it is precious for businesses seeking to improve operational efficiency, reduce costs, and enhance the quality of their products or services. By monitoring and analyzing process data during runtime, PrPM can identify areas for improvement and provide actionable recommendations for optimizing the process. For example, a customer churn process where customers stop doing business with a company can lead to lost revenue

CAiSE '23: Doctoral Consortium co-located with the 35th International Conference on Advanced Information Systems Engineering, CAiSE '23, June 12-16, 2023, Zaragoza, Spain
mahmoud.shoush@ut.ee (M. Shoush)
https://github.com/mshoush (M. Shoush)
0000-0002-7423-9909 (M. Shoush)
© 2022 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)

and reduced profitability. PrPM can identify churn customers (or *cases*) and provide prescriptive recommendations for reducing churn, such as offering targeted promotions.

Recently, several PrPM techniques have been proposed [1, 2, 3, 4]. These techniques consist of two components. The first is a predictive model that estimates the probability that a given case will end up with an undesired outcome. The second is an intervention policy, which determines if an intervention should be triggered for a given case to optimize a gain function. However, these techniques lack two main things. Firstly, they trigger interventions when cases are likely to end undesirably, assuming that predictions will be continuously generated, even when wholly uncertain and regardless of the consequences of lower-quality predictions. Secondly, these techniques assume that triggered interventions are effective and that it is possible to trigger any number of interventions at any time. However, business processes are dynamic in reality, and uncertainty levels may remain high. Furthermore, interventions may come with costs and require resources, such as an employee's time, which are limited and expensive.

In this Ph.D. project, we aim to address the abovementioned gap. We consider a problem that includes monitoring business cases to provide timely intervention during runtime to avoid the costly occurrence of undesired outcomes when cases unfold. We assume that resources are bounded, and process interventions with associated costs are predefined based on domain expertise. Specifically, we seek to answer the following research question:

Whom? When? Which?
How can we confidently identify and prioritize business cases (whom?) and provide timely (when?) interventions (which?) considering resources are bounded to optimize a business value, e.g., gain?

In order to address the research question, this study utilizes a comprehensive approach involving integrating three distinct machine learning algorithms: *predictive, conformal, and causal. The predictive algorithm* estimates the likelihood of future outcomes for individual cases, while also providing a measure of uncertainty associated with these predictions. By doing so, it quantifies how certain the predictive model is with its predictions. In comparison, *the conformal algorithm* is applied to convert the predictive model's output into predictions with a guaranteed confidence level. *The causal model* is then implemented to estimate the potential effectiveness of interventions in mitigating negative outcomes for cases identified as requiring intervention based on the preceding predictive and conformal analyses. Moreover, a resources allocator is developed and utilized to monitor the availability of resources, allocate, and release them. The resources allocator is used to learn which cases should be treated during runtime and when given the available resources.

2. Related Work

A number of PrPM techniques have been proposed in the past decade. These techniques can be organized into three groups based on how the intervention policy prescribes interventions to improve a business value [5]. The first focuses on control flow to recommend the next best action or activity to improve a pre-defined KPI [6, 7, 3]. The second focuses on a resource view to guide resource allocation decisions [8, 9]. Finally, the third focuses on triggering interventions to avoid or mitigate the effect of undesired outcomes and considers both control flow and resources perspectives [1, 4, 2]. Our proposal fits in the latter; It seeks to learn policies for triggering interventions to avoid undesired outcomes under uncertainty and finite resources.

The works in [4] and [1] utilize an outcome-oriented predictive model that focuses solely on generating predictions related to the outcome. However, these models do not consider the quality of the predictions or provide any measure of confidence in the predictions. In [4], they trigger an intervention when the probability of an undesired outcome exceeds a threshold determined manually or empirically. While in [1], a reinforcement learning agent learns when to trigger an intervention according to parameters such as prediction scores and reliability estimates. The work in [2] introduces a causal inference model to estimate the expected effect of triggering an intervention to reduce the cycle time of a process, which is another problem compared to us. However, the works in [1, 4, 2] assume that interventions with a positive impact occur immediately even when the level of uncertainty is high and do not examine the finite capacity of resources.

3. Proposed Solutions

To address the research problems outlined, we introduce a prescriptive process monitoring approach, illustrated in Fig. 1. Our proposed method breaks down the research problems into three key questions, which we will answer in the subsequent sections.

- Whom to intervene? Exploring how to identify and prioritize candidate cases for interventions with confidence.
- When to intervene? Discussing the tradeoff between triggering an intervention now versus later.
- Which intervention? Studying which intervention is suitable for which case.

3.1. Whom to intervene?

The present study aims to develop a method for identifying cases that require intervention by utilizing predictive, conformal, and causal models, as outlined in previous studies [10, 11]. During the training phase, the aforementioned models are trained to filter ongoing cases and identify suitable candidates for intervention. Subsequently, the predictive model is utilized during the testing phase to calculate the probability of an undesirable outcome for a given ongoing case. To provide a level of confidence in the predicted outcome, a conformal algorithm is employed to transform the predictive model outputs into a prediction set. This algorithm returns a prediction set containing a single outcome with a confidence of $1 - \alpha$, based on a user-defined significance level (α) and a predictive model. If the predicted outcome is uncertain, the algorithm will return a prediction set containing both desired and undesired outcomes or an empty set. Finally, the causal model is utilized to estimate the potential impact of an intervention on an ongoing case, by assessing the increase in the probability of a positive outcome, commonly known as the Conditional Average Treatment Effect ($CATE$).

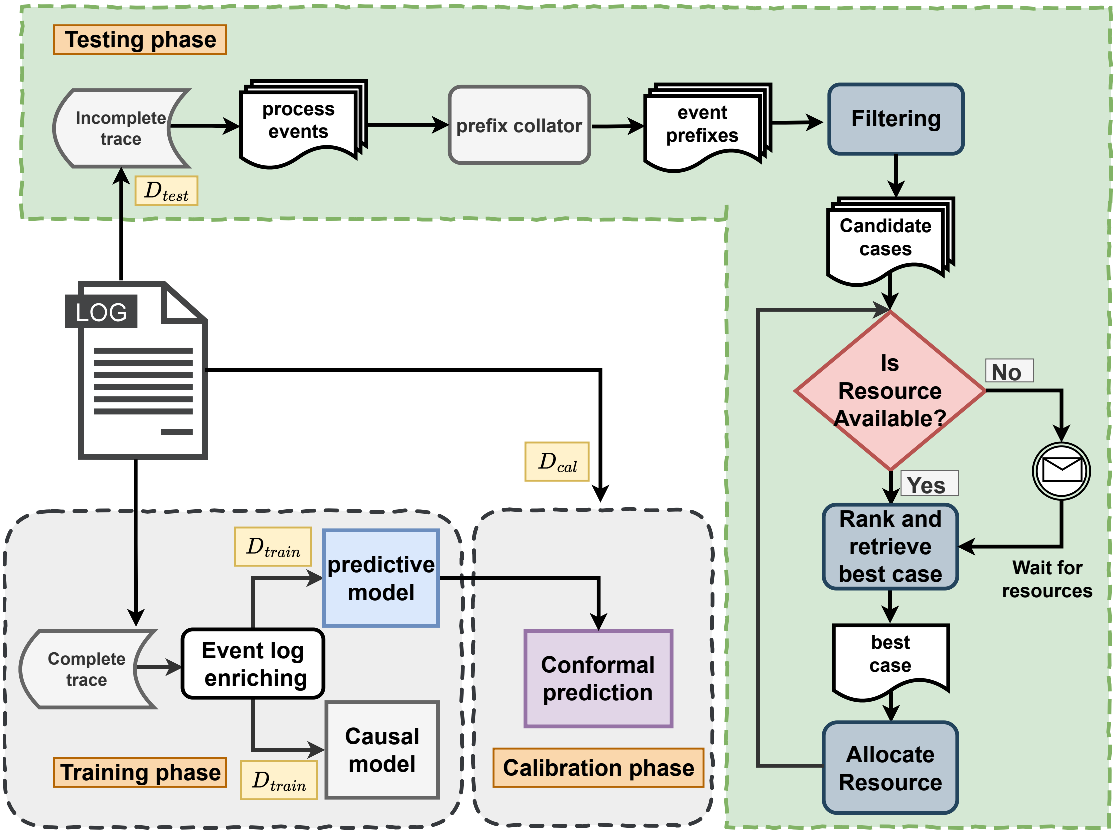

Figure 1: An overview of the proposed solutions.

3.2. When to intervene?

The current study involves prioritizing candidate cases from a preceding step and constructing a resource allocator to manage and allocate resources to the most profitable cases, thereby determining the appropriate time to trigger interventions.

We propose utilizing two measures, namely *gain* [10] and *adjusted gain* [12], to prioritize candidate cases and determine the appropriate timing of interventions. The gain represents the benefits acquired at the current state of an ongoing case, while the adjusted gain considers the current and future states of ongoing cases. Both measures incorporate the costs of the intervention and the possibility of undesired outcomes. To operationalize these measures, we developed a resource allocator [10], which assesses the availability of resources and assigns them to the most profitable cases. Upon selecting the most profitable case, and once a free resource is available, we assign it to the selected case and block it for a specific duration, i.e., the *treatment duration*. The number of available resources and the time required to perform the intervention can be determined based on domain knowledge.

3.3. Which intervention?

Our previous proposal assumed that each ongoing case would require a single intervention. However, in reality, cases may require one or more interventions, each with their own costs and benefits. Furthermore, a case may need to be intervened upon multiple times at various points during its execution. This process of allocating resources to ongoing cases multiple times with different interventions can be costly, making the study challenging. Additionally, identifying a set of interventions that would positively impact outcomes in advance presents yet another challenge. To address these challenges, we plan to extend the definition of gain and adjusted gain from our previous works to account for multiple interventions.

4. Methodology

The research methodology employed for this Ph.D. project will be the Design Science approach, as outlined in [13]. This approach ensures accuracy through the use of extensive literature review and the creation of a complete evaluation benchmark with clear selection and assessment measures. To verify the relevance of the proposed solutions, the developed techniques will undergo a comprehensive evaluation using both real-life and simulated logs. Furthermore, where possible, the project will include case studies with relevant organizations to ensure the practicality and effectiveness of the proposed solutions.

5. Challenges and Future Work

Prescriptive process monitoring approaches are often confronted with difficulties in identifying interventions that could lead to improved outcomes. Discovering or defining interventions from execution data can be challenging, as it necessitates the identification of a causal relationship between a specific variable and the outcome. For example, identifying a correlation between loan approval and monthly interest rates.

Moreover, our previous proposal encounters the challenge of optimizing intervention triggers. While our rule-based model assigns resources to ongoing cases, these resources may only be well-utilized if the potential gain is high. Therefore, to make informed decisions on when to allocate and when not to, we plan to develop a more intelligent resource allocator based on the anticipated gain at each point.

To advance our approach, we aim to move away from the rule-based model and adopt a learning-based approach in the future. To accomplish this, we plan to leverage reinforcement learning to learn an optimal policy at runtime for deciding when to trigger interventions and how to efficiently allocate resources.

Acknowledgement

This project is supported by the European Research Council (PIX Project). The project is under the supervision of Prof. Marlon Dumas, who is affiliated with the University of Tartu in Estonia.

References

[1] A. Metzger, T. Kley, A. Palm, Triggering proactive business process adaptations via online reinforcement learning, in: BPM, volume 12168 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 273–290.

[2] Z. D. Bozorgi, I. Teinemaa, M. Dumas, M. La Rosa, Prescriptive process monitoring for cost-aware cycle time reduction, ICPM (2021).

[3] M. de Leoni, M. Dees, L. Reulink, Design and evaluation of a process-aware recommender system based on prescriptive analytics, in: ICPM, IEEE, 2020.

[4] S. A. Fahrenkrog-Petersen, N. Tax, I. Teinemaa, M. Dumas, M. de Leoni, F. M. Maggi, M. Weidlich, Fire now, fire later: alarm-based systems for prescriptive process monitoring, Knowl. Inf. Syst. 64 (2022) 559–587.

[5] K. Kubrak, F. Milani, A. Nolte, M. Dumas, Prescriptive process monitoring: *Quo vadis*?, PeerJ Comput. Sci. 8 (2022) e1097.

[6] S. Weinzierl, S. Dunzer, S. Zilker, M. Matzner, Prescriptive business process monitoring for recommending next best actions, in: BPM (Forum), volume 392 of *Lecture Notes in Business Information Processing*, Springer, 2020, pp. 193–209.

[7] P. Agarwal, A. Gupta, R. Sindhgatta, S. Dechu, Goal-oriented next best activity recommendation using reinforcement learning, CoRR abs/2205.03219 (2022).

[8] R. Sindhgatta, A. K. Ghose, H. K. Dam, Context-aware analysis of past process executions to aid resource allocation decisions, in: CAiSE, volume 9694 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 575–589.

[9] G. Park, M. Song, Prediction-based resource allocation using LSTM and minimum cost and maximum flow algorithm, in: ICPM, IEEE, 2019, pp. 121–128.

[10] M. Shoush, M. Dumas, Prescriptive process monitoring under resource constraints: A causal inference approach, in: ICPM Workshops, volume 433 of *Lecture Notes in Business Information Processing*, Springer, 2021, pp. 180–193.

[11] M. Shoush, M. Dumas, Intervening with confidence: Conformal prescriptive monitoring of business processes, CoRR abs/2212.03710 (2022). URL: https://doi.org/10.48550/arXiv.2212.03710. doi:10.48550/arXiv.2212.03710. arXiv:2212.03710.

[12] M. Shoush, M. Dumas, When to intervene? prescriptive process monitoring under uncertainty and resource constraints, in: BPM (Forum), 2022.

[13] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS Q. 28 (2004) 75–105. URL: http://misq.org/design-science-in-information-systems-research.html.