

Multidimensional Process Model Forecasting (MuDiPMF)

Yongbo Yu

KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

Abstract

Process analytics aims to improve processes based on event logs generated by information systems by, among others, automatically discovering models representing the current system. This discovery, however, is typically based on static models, ignoring its underlying trends and shifts. Recently, the modeling and prediction of the full system have been proposed as process model forecasting (PMF). However, the current SOTA lacks the ability to model intricate control flow constructs while also not incorporating extra information, such as resources tied to the process. Besides, by using univariate models, the underlying relations between the different elements of the system are ignored. This proposal addresses these issues by firstly extending PMF to richer control flow models that are able to capture relationships between activities in workflows. Secondly, the current PMF techniques will be replaced by a multivariate framework based on state-of-the-art deep learning techniques such as graph neural networks, which form a natural fit for graph-based models such as workflow diagrams and capture both temporal, structural, and multiscale patterns. Next, additional perspectives, such as resources, will be added to obtain fully object-centric process model forecasts, which can incorporate any data related to a process through the forecasting of event knowledge graphs. Finally, two industry cases in finance and logistics will be used to validate the findings in a real-life setting.

Keywords

Process Model Forecasting, Time Series Forecasting, Graph Neural Networks

1. Positioning and Motivation

Within the field of Process Mining, Predictive Process Monitoring (PPM) entails forecasting future elements of ongoing process instances or cases, including the most probable next activities, outcomes, and remaining runtime. Notably, the integration of machine learning and deep learning solutions into this domain has been extensively explored in academic literature [1]. While real-time insights at the individual case level allow process owners to intervene in specific instances, they often lack the capacity to provide end-users with information regarding the future trajectory of the entire process. Consequently, a new paradigm known as Process Model Forecasting (PMF) has emerged [2], focusing on predicting future states of the overall process model over a long-term horizon, drawing information from historical event data [3]. This Multidimensional Process Model Forecasting (MuDiPMF) project aims to develop and validate a set of tailored and integrated multi-dimensional forecasting models using multivariate predictive methods for multi-perspective business process models.

The current state-of-the-art in PMF involves depicting the evolution of process behavior through time series analysis of individual Directly-Follows relations (DFs), which track the frequency of one activity following another within cases, over a predefined timeframe [3]. These DFs collectively form a Directly-Follows Graph (DFG), a widely utilized process visualization tool offering a clear representation of the flow of the process. The individual DFs are forecasted using univariate time series forecasting techniques, overlooking correlations between different DFs induced by the underlying relations between process elements within the information system. Additionally, DFGs lack the capability to model more nuanced process constructs, such as parallel behavior, in contrast to more semantically rich process model notations like Petri nets and BPMN models.

Therefore the first objective of MuDiPMF is to forecast more semantically rich process models. This entails expanding the feature set beyond DFs, such as the constructs utilized in advanced process discovery techniques. Furthermore, multimodal predictive models can be used to forecast various

ICPM 2024 Doctoral Consortium, October 14–18, 2024, Kongens Lyngby, Denmark

✉ yongbo.yu@kuleuven.be (Y. Yu)

🆔 0009-0004-2964-6611 (Y. Yu)



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

time series of multidimensional feature sets simultaneously, thereby accounting for process related dependencies and correlations. Next to this, MuDiPMF aims to enhance the forecasted process models beyond the control-flow aspect, by incorporating additional dimensions such as resource occupation, execution times, and decision point analyses. This would, among others, allow process owners to perform timed interventions regarding bottlenecks, and optimize resource allocations. Furthermore, in recent years Object-Centric Process Mining (OCPM) has emerged as a new family of approaches tailored to handle event data from processes involving different interconnected objects such as orders, items, and shipments, garnering widespread attention in both academia and industry [4]. Given the rapid emergence and relevance of these object-centric process models, a third objective of MuDiPMF is to develop a framework extending forecasting capabilities to such process models.

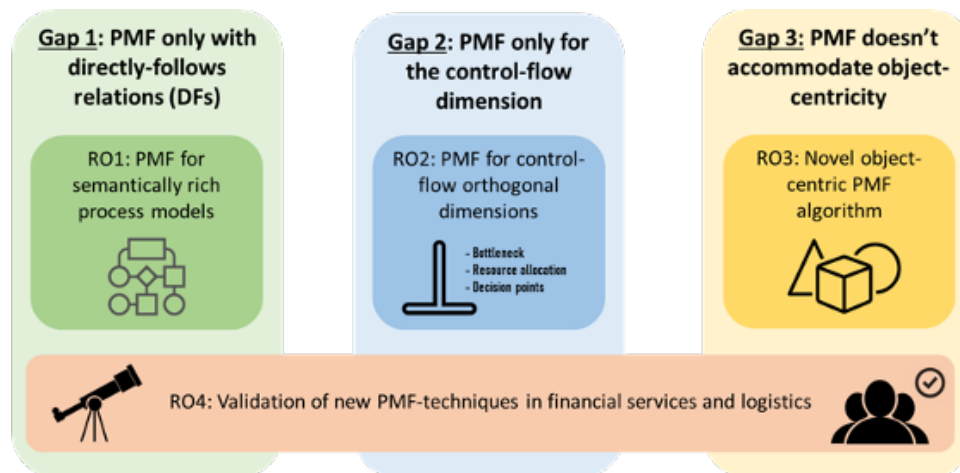


Figure 1: Overview of research gaps and objectives

In summary, the Multidimensional Process Model Forecasting (MuDiPMF) project will significantly contribute to the current state-of-the-art in Business Process Management and Process Mining by enhancing the recently proposed PMF framework by broadening the feature set to be predicted with additional dimensions, while exploring the application of more suitable multivariate predictive methods. Finally, the project aims to demonstrate the practical utility of the different PMF enhancements using real-life process data from two different domains: financial services and logistics.

2. Current Solutions and Research Objectives

Figure 1 illustrates the research gaps and research objectives (ROs). Given the recent inception of the PMF paradigm, the state-of-the-art is currently confined to univariate forecasting of distinct DFs [3]. More specifically, current solutions rely on auto-regressive time-series forecasting techniques such as ARIMA and GARCH. The forecasted DFs can collectively represent a process model (DFG), but are not sufficient to discover more complex process model structures such as parallelism. In contrast, the literature on automated process discovery is more developed, with numerous approaches proposed over the years to discover, e.g., Petri Nets or BPMN models [5]. Other examples include Heuristics Miner and its extension Fodina, which utilizes various heuristics and formalisms to automatically discover among others, concurrency, exclusive choices, and loops in a process from event logs [6], [7]. Another approach involves a top-down strategy, exemplified by methods like Inductive Miner, which partitions larger event logs into more manageable segments for analysis [8]. RO1 discusses expanding PMF towards forecasting shifts in processes expressed by more semantically rich process model representations.

Next, the growing literature on PPM remains relevant despite its focus on single objectives such as suffix prediction [9], or case outcome prediction [10]. Many of the predictive approaches have assumed deep learning models such as long short-term memory networks and even graph neural networks. Many operate a multivariate, but not a multitarget approach, as envisioned for RO2. Finally, a growing

interest in the object-centric perspective of process mining has been evident in recent years [11]. The representation of these object-centric event logs as an event knowledge graph is especially of interest for this project given the similarity of object-centric process models changing over time according to a temporal graph-based structure of the data [12]. This will be addressed in RO3.

From an algorithmic perspective underpinning these applications, various data-driven and deep learning approaches for multivariate time series forecasting have been proposed [13]. Given the graph-based structure of process models, together with their emergence as powerful predictors in different tasks, Graph Neural Networks (GNNs) are a natural match to learn both temporal and structural properties of process models. Particularly, work incorporating the time dimension into GNNs to investigate both spatial and temporal dependency together could provide PMF with more powerful and flexible predictors capable of taking into account process-specific dependencies. Different approaches for spatial-temporal graph neural networks (STGNNs), such as STGCN [14] and StemGNN [15], have shown potential in domains such as traffic forecasting.

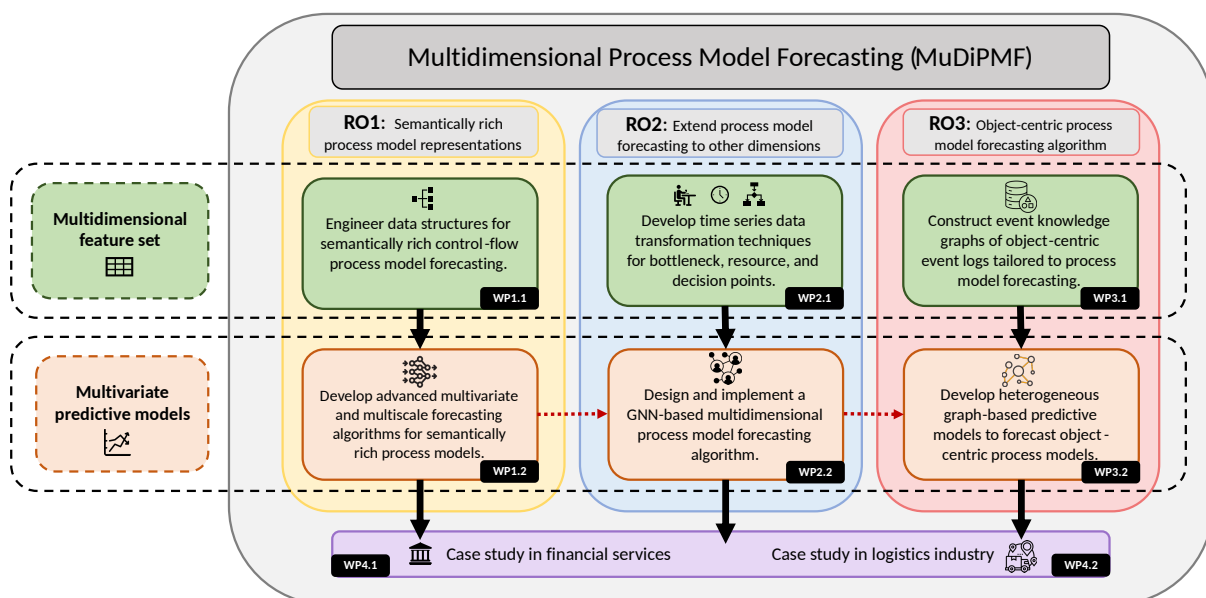


Figure 2: Overview of the work packages

3. Planned Research Methodology

Figure 2 presents a schematic overview of the proposed work plan designed based on [16], illustrating the alignment of different work packages (WPs) with the four research objectives (ROs).

RO1 aims to extend the feature set of the forecasting techniques beyond DFs by incorporating process representations and dependencies utilized in various widely used process discovery methodologies. For example, we can forecast full dependency graphs, as this would allow us to discover parallel activities or forecast the required metrics for the partition creation used by top-down discovery algorithms. Correspondingly, existing predictive models will be replaced with multivariate models capable of simultaneously forecasting all time series while accounting for cross-dependencies. One promising avenue involves the exploration of multi-scale spatial-temporal graph neural networks.

RO2 extends process model forecasting capabilities to incorporate control-flow orthogonal dimensions, including bottlenecks, resource allocation, and decision points. We aim to integrate resource information into the feature set, considering multiple granularity levels from overall resource occupancy to allocations at specific activities. Finally, we will incorporate the attention mechanism in GNNs to extract more reliable and efficient patterns and leverage the multitask learning (MTL) framework to implement a multidimensional PMF algorithm.

RO3 aims to design and implement a comprehensive object-centric process model forecasting algorithm. To account for the complexity of object-centric event logs, we will develop an appropriate event knowledge graph (EKG) structure on top of which a new process model forecasting algorithm can be built. This would entail the construction of time series features for heterogeneous graph elements within EKGs. Next to it, we aim to explore extra architectures tailored toward heterogeneous graph forecasting. The initial avenue that will be pursued focuses on the use of heterogeneous temporal graph neural networks.

RO4 aims to extend the impact of the developed techniques within MuDiPMF by deploying them in practical applications across diverse industries, specifically targeting the financial services and logistics sectors. The research group's network will be leveraged to collaborate with two partnering companies. Through these case studies, the objective is to demonstrate how the advancements made can substantially improve the state-of-the-art in Process Model Forecasting (PMF) and highlight their practical effectiveness. By validating our algorithms using real-world problems and data, we do not only aim to emphasize their added value, but refinement and adaptation strategies will be developed to make MuDiPMF algorithms extensible to other application domains.

Acknowledgments

This study was financed by the Research Foundation Flanders under grant number G039923N and Internal Funds KU Leuven under grant number C14/23/031.

This Ph.D. thesis is supervised by Prof. dr. Johannes De Smedt and Prof. dr. Jochen De Weerd.

References

- [1] F. M. Maggi, C. Di Francescomarino, M. Dumas, C. Ghidini, Predictive monitoring of business processes, in: *Advanced Information Systems Engineering: 26th International Conference, CAiSE 2014*, Thessaloniki, Greece, June 16-20, 2014. *Proceedings 26*, Springer, 2014, pp. 457–472.
- [2] R. Poll, A. Polyvyanyy, M. Rosemann, M. Röglinger, L. Rupprecht, Process forecasting: Towards proactive business process management, in: *Business Process Management: 16th International Conference, BPM 2018*, Sydney, NSW, Australia, September 9–14, 2018, *Proceedings 16*, Springer, 2018, pp. 496–512.
- [3] J. De Smedt, A. Yeshchenko, A. Polyvyanyy, J. De Weerd, J. Mendling, Process model forecasting and change exploration using time series analysis of event sequence data, *Data & Knowledge Engineering* 145 (2023) 102145.
- [4] W. M. van der Aalst, Object-centric process mining: Dealing with divergence and convergence in event data, in: *Software Engineering and Formal Methods: 17th International Conference, SEFM 2019*, Oslo, Norway, September 18–20, 2019, *Proceedings 17*, Springer, 2019, pp. 3–25.
- [5] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, F. M. Maggi, A. Marrella, M. Mecella, A. Soo, Automated discovery of process models from event logs: Review and benchmark, *IEEE transactions on knowledge and data engineering* 31 (2018) 686–705.
- [6] A. J. Weijters, W. M. van Der Aalst, A. A. De Medeiros, *Process mining with the heuristicsminer algorithm* (2006).
- [7] S. K. vanden Broucke, J. De Weerd, Fodina: A robust and flexible heuristic process discovery technique, *decision support systems* 100 (2017) 109–118.
- [8] S. J. Leemans, D. Fahland, W. M. Van Der Aalst, Discovering block-structured process models from event logs—a constructive approach, in: *Application and Theory of Petri Nets and Concurrency: 34th International Conference, PETRI NETS 2013*, Milan, Italy, June 24–28, 2013. *Proceedings 34*, Springer, 2013, pp. 311–329.
- [9] M. Camargo, M. Dumas, O. González-Rojas, Learning accurate lstm models of business processes, in: *Business Process Management: 17th International Conference, BPM 2019*, Vienna, Austria, September 1–6, 2019, *Proceedings 17*, Springer, 2019, pp. 286–302.

- [10] I. Teinemaa, M. Dumas, M. L. Rosa, F. M. Maggi, Outcome-oriented predictive process monitoring: Review and benchmark, *ACM Transactions on Knowledge Discovery from Data (TKDD)* 13 (2019) 1–57.
- [11] R. Galanti, M. De Leoni, N. Navarin, A. Marazzi, Object-centric process predictive analytics, *Expert Systems with Applications* 213 (2023) 119173.
- [12] D. Fahland, Process mining over multiple behavioral dimensions with event knowledge graphs, in: *Process mining handbook*, Springer, 2022, pp. 274–319.
- [13] B. Lim, S. Zohren, Time-series forecasting with deep learning: a survey, *Philosophical Transactions of the Royal Society A* 379 (2021) 20200209.
- [14] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, *arXiv preprint arXiv:1709.04875* (2017).
- [15] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang, Y. Tong, B. Xu, J. Bai, J. Tong, et al., Spectral temporal graph neural network for multivariate time-series forecasting, *Advances in neural information processing systems* 33 (2020) 17766–17778.
- [16] J. Mendling, H. Leopold, H. Meyerhenke, B. Depaire, Methodology of algorithm engineering, *arXiv preprint arXiv:2310.18979* (2023).