Event Data and Process Model Forecasting

Wenjun Zhou[0000−0002−2262−1428], Artem Polyvyanyy[0000−0002−7672−1643], and James Bailey

The University of Melbourne, Victoria VIC 3010, Australia
{wenjun.zhou,artem.polyvyanyy,baileyj}@unimelb.edu.au

Abstract. Process mining studies ways to use event data generated by information systems to understand and improve the business processes of organizations. One of the core problems in process mining is *process discovery*. A process discovery algorithm takes event data as input and constructs a process model that describes the processes the system that generated the data can execute. The discovered model, hence, aims to represent both historical processes with traces in the data and the yet unseen processes of the system (*total generalization*). In this paper, we introduce *process forecasting* as an alternative approach to process discovery. First, given historical event data, the corresponding future event data is forecasted for a requested period in the future (*event data forecasting*). Then, a process model is constructed from the forecasted data to describe the processes the system is anticipated to execute during the target future period (*process model forecasting*). The benefits of this alternative approach are at least twofold. Firstly, it divides the problem into two fundamentally different sub-problems that can be studied and mastered separately. Secondly, a forecasted model that describes the processes of the system from a given period rather than in general (*tailored generalization*) can help organizations plan future operations and process improvement initiatives.

Keywords: Process mining · Process forecasting · Process model forecasting · Event log forecasting.

1 Introduction

Through studying the event log generated by organization information systems, *process mining* bridges the gap between data science and process science. Process mining can be used to identify process bottlenecks and noncompliance during the process execution to improve the process using its visual representations, such as Directly-Follows Graphs (DFGs), Petri Nets, and Process Trees [2]. These models are abstractions of the processes they represent, where event logs collected through a software system are used as the input for generating the models. This approach of transforming event logs into *process models* is also known as *process discovery*, a sub-field of process mining.

With the business demand for prediction of the future and the prospering of machine learning in recent years, researchers have started a trend of predicting process elements. For example, Cardoso, J., and Lenič, M. [5] proposed

Postprint, July 2024

an approach to business activity prediction. A line of research has focused on predicting time aspects in processes [3]. Existing techniques, however, focus on predicting case-level process elements, with the prerequisite that a process case has executed a few activities [22]. Very little research has focused on forecasting future process models. With the concept being proposed by Poll et. al. [15], one technique has been devised to forecast future process models [6,7].

A discovered process model aims to describe both the historical traces of the system found in the input event log and the unseen traces of the system that generated the data. Such *total generalization* of the discovered models helps to understand the system that generated the data. However, it is less useful for business planning, as it does not relate the described traces to the period when the system is expected to execute them. If one can get in possession of a process model or an event log that accurately describes the processes the system will execute in a given period in the future, they can use this knowledge to plan future operations or prepare for upcoming process drifts. In this paper, we introduce *process forecasting* as an alternative approach to process discovery that, given historical data, aims to construct event data or process models that describe future processes, hence implementing such a tailored generalization over the input data. Specifically, this paper makes these contributions:

- Definitions of event data and process model forecasting problems;
- Comparison of existing techniques for forecasting of process elements;
- Discussion of the challenges posed by process forecasting.

The next section presents the terms and background knowledge that supports the understanding of the subsequent sections. Section 3 discusses related work and compares the differences between process forecasting and other prediction techniques proposed by the Business Process Management (BPM) community. Section 4 presents event data forecasting and process model forecasting and discusses ideas for tackling these problems. Finally, Section 5 concludes the paper.

2 Preliminaries

Event Logs. In process mining, an *event log*, or a *log*, is a collection of events executed and recorded during the execution of multiple instances, or cases, of a business process. At least three compulsory attributes are recorded for each event: *activity*, *case identifier* (case ID), and *timestamp* [2]. Other common event attributes include *cost*, *duration*, and *resource*. The availability of these additional attributes depends on a particular dataset. The *activity* attribute refers to the executed activity that triggered the event. The *timestamp* attribute records the time of the occurrence of the event. Finally, all activities from the same business process instance share the same *case ID* attribute. In this paper, we use *trace* to refer to the sequence of *activities* that stem from the events with the same *case ID* ordered by the timestamps of these events.

Process Model. A *business process model*, or a *process model*, is an abstraction of a running business [17]. There are different levels of abstraction for constructing a process model. Process models at different levels of abstraction serve different purposes, such as understanding the true process compared to the designed ideal process and improving the process. A common way to construct a true process model is by using event logs collected from the running business software systems.

Process Discovery. Process discovery is a problem in process mining that studies ways to construct process models from historical traces recorded in an event log of a system [1]. A good discovered model should describe the traces in the event log (*good recall*), not describe traces not in the event log (*good precision*), be as *simple* as possible, and capture the traces the system can generate but are not in the event log (*good generalization*) [4]. By describing process traces of the system beyond those in the event log, the constructed process model encodes possible future traces of the system.

Process Simulation. Process simulation, especially business process simulation, involves creating a model that mimics the operations of a hypothetical business and analyzing its properties [16]. Process simulation can be used for various purposes, such as optimizing resource allocation, identifying bottlenecks, testing alternative scenarios, and assessing the impact of process variations. It is often used to evaluate and compare alternative process redesign solutions.

Process Drift Detection. Process drift in the business process context is a change in the operations of the business process, including sudden, reoccurring, incremental, and gradual changes [23]. These changes are identified using dedicated detection algorithms that incorporate the ability to predict future operations given the current data or identify the point of drift as soon as possible while minimizing false positives.

Predictive Process Monitoring. Predictive Process Monitoring (PPM) [14] focuses on case- or micro-level predictions; examples of the main use cases include outcome prediction, next activity prediction, and process duration prediction. PPM, as in all the listed examples, often operates within the scope of a single process case.

Process Forecasting. Process Forecasting is an umbrella term for predictions of process elements [15]. In this paper, the term *process forecasting* is used to refer to the problem of forecasting macro-level process elements, such as process models and event logs.

3 Related Work

In this section, we summarize works in process mining that tackle prediction tasks and discuss the differences between these existing works and our work.

Specifically, we review works in process discovery, process simulation, process drift detection, PPM [14], and process forecasting.

A discovered model encodes the historical event log as well as possible while also being simple and general to reflect the possible future traces. The benefits of process discovery are speed and diversity. It usually takes a relatively short time to construct a model from a large event log. In addition, many configuration parameters can be explored to discover models of different complexities and accuracies. The downsides are mainly two. First, process discovery is not specifically designed for prediction tasks. Second, there is no specific guideline for discovery to anticipate the future; in other words, how much generalization in discovered models is good enough is unclear.

Like process discovery, process simulation can be used for different purposes. It is often used in what-if analysis. For example, one may use process simulation to analyze "What is the extra cost and time if activity B is introduced between activities A and C." Process simulation can predict and evaluate the impact of changes, improvements, or disruptions without affecting the actual system. Nevertheless, the simulation system may depend on the quality of the discovered models, and hence, it inherits the downsides of the discovery algorithms.

Process drifts refer to changes in the way a business process is executed or in the environment in which it operates, leading to deviations from the expected or desired behavior [23]. Process drift analysis can inform process improvement initiatives, regulatory compliance efforts, and strategic decision-making. Identifying and addressing business process drift can improve operational efficiency, compliance with regulations, and customer satisfaction. It can also enable organizations to proactively respond to changing market conditions and emerging risks. One can use process drift detection techniques to analyze frequent drift patterns and predict the next drifts by projecting historical drifts into the future.

The vast majority of PPM techniques take historical event logs as input and learn models that encode possible process traces/cases. Then, the newly lodged process case is monitored, and the prefix of that case observed so far is used to generate predictions for that particular case. The benefit of PPM is that the prediction for a particular case can often be made in real-time, and the newly observed data can be used immediately to update the learned models to improve future predictions. PPM techniques are often deployed before they get used to allow sufficient training before the techniques get productive. The state-of-the-art PPM techniques often demonstrate high prediction accuracy [3,13,19,20] and are grounded in conventional statistical and process analysis techniques, while several existing works also explore deep learning approaches [9,8,10,19]. As mentioned in Section 2, PPM operates within the micro-, or case-, level. In contrast, Process Forecasting, including process model forecasting and event log forecasting, focuses on the model-, or macro-, level predictions.

Process Forecasting studies changes in process models over time. The concept was proposed in 2018; since then, only a few techniques have been proposed. The work by De Smedt et. al. [6,7], explores how statistical methods over time series help in forecasting the directly-follows relationships of real-life business

Table 1: Comparison of process prediction/forecasting techniques.

Techniques	**Use cases**	**Pros**	**Cons**
Process Discovery	Discover process models from event data	Simple, diverse, often fast	Unclear generalization level for prediction
Process Simulation	Evaluate business redesigns	Cost-effective	Prediction is scenario-based
Process Drift Detection	Detect process changes	Optimized for change detection	Limited to change detection
Predictive Process Monitoring	Predict case-level process elements	Prediction accuracy, diverse techniques, supports prediction of multiple process elements	Limited to case-level predictions
Process Forecasting	Predict macro-level process elements	Specialized for prediction	Can be complex

processes. The lesson learned from this work is that no method works well across all datasets. However, the quality of achieved forecasts can enable proactive business process planning, including process drift and change predictions.

Table 1 summarizes the advantages and disadvantages of the above-discussed approaches.

4 Process Forecasting

This section presents two problems in process forecasting: *event log forecasting* and *process model forecasting*. Figure 1 schematically shows the contexts of the two problems. Given an event log, process model forecasting (Fig. 1a) constructs process models that describe the processes the system that generated the input log will execute in a given period in the future, while event log forecasting (Fig. 1b) generates future event logs of the system for the requested period. For instance, process model forecasting can proceed in two steps. First, a process model can be discovered, or a future event log can be forecasted from the input event log. Then, the obtained artifacts can be used to induce a forecasted process model. Event log forecasting, in turn, aims to generate future event logs directly from the input log. One can then use a forecasted event log to construct a forecasted model, for instance, using process discovery techniques.

4.1 Process Model Forecasting

Given an event log of a system and a time period, the *process model forecasting* problem consists of constructing a process model that accurately describes the processes the system will generate in the given period. A solution to the process model forecasting problem can involve a process discovery from the log and then the use of the discovered model constructs to anticipate the constructs in the forecasted model. For example, one can break the event log into multiple time windows, discover process models for the different windows, and then project the trend in observed model constructs into the future. Smedt et al. [6] demonstrated that this approach could result in useful forecasted models. Depending on the employed abstraction level, one can operationalize this approach using time series forecasting.

4.2 Event Log Forecasting

Given an event log and a period in the future, the *event log forecasting* problem studies ways to generate an accurate log that the system that generated the input log will generate in the given period. If one succeeds in forecasting the genuine future log, they can prepare to support the corresponding process, for instance, by planning sufficient resources to ensure successful operations. As indicated in Fig. 1b, this approach deals with historical event logs directly. Given a forecasted event log, one can then construct multiple models that aim to represent this log. To this end, one can employ different discovery techniques. Such discovered models, if faithfully describing the forecasted log, can be accepted as forecasted process models. Event logs contain more information than process models, as they constitute the raw data, while models can be seen as aggregations of the data. Finally, if there are missing entries in the historical event log, the forecasts can be available for those fields for better process analysis and planning.

Despite the advantages described above, event log forecasting is associated with challenges. To generate an accurate forecasted event log, sophisticated techniques such as deep learning may be required, which may be costly in terms of resources and time required for training. Yet, there has been no successful demonstration that deep learning is indeed helpful for event log forecasting. Additional challenges include a lack of appropriate measurements for the forecasted results, as it may be insufficient to compare the forecasted log to ground truth on a

(a) (b)

Fig. 1: Contexts of (a) process model forecasting and (b) event log forecasting.

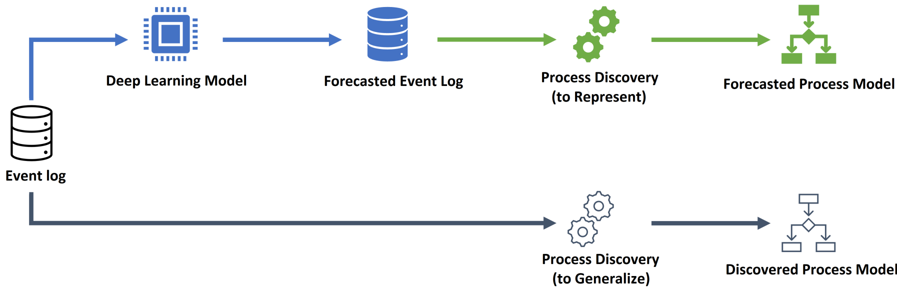

Fig. 2: Process forecasting and process discovery.

trace-by-trace basis. In addition, if a forecasted log is aggregated as a forecasted model for evaluation, the effect of the discovery algorithm on the forecasting results may be unknown.

Fig. 2 illustrates a hypothetical process forecasting pipeline and its relationship to process discovery. The upper branch describes an event log forecasting approach grounded in a deep learning technique, while the bottom branch captures the conventional process discovery steps. A potential benefit of using deep learning techniques for event log forecasting is that the problem can fit into the current deep learning architectures, such as Seq2Seq [18] and transformer [21]. The success of Large Language Models (LLM) [11] has proved that the architecture is capable of handling sequence forecasting and generation tasks. Similar to LLM, the BPM community explores ways to build a Large Process Model (LPM) to solve forecasting problems [12]. However, such models are associated with challenges. Specifically, LLM suffers from long training time, and the training outcome is not guaranteed. In addition to that, it also consumes significant computing resources to build a model. Similar challenges will likely manifest if these models are used to solve the event log forecasting problem.

5 Conclusion

This paper presents and discusses problems of event data and process model forecasting, which are alternative approaches to process discovery studied in process mining that aim to construct artifacts that describe processes of the system for a specified period in the future. It is envisaged that accurate solutions to these problems will support organizations in planning their future operational processes. Several ideas for solving these new problems are discussed, pointing to deep learning models as a promising approach for tackling them.

References

1. van der Aalst, W.M.P.: Process discovery: Capturing the invisible. IEEE Comput. Intell. Mag. (2010)

2. van der Aalst, W.M.P.: Process Mining - Data Science in Action, Second Edition. Springer (2016)
3. van der Aalst, W.M.P., Schonenberg, M.H., Song, M.: Time prediction based on process mining. Inf. Syst. (2011)
4. Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. Int. J. Cooperative Inf. Syst. (2014)
5. Cardoso, J., Lenic, M.: Web process and workflow path mining using the multi-method approach. Int. J. Bus. Intell. Data Min. (2006)
6. De Smedt, J., Yeshchenko, A., Polyvyanyy, A., De Weerdt, J., Mendling, J.: Process model forecasting using time series analysis of event sequence data. In: ER (2021)
7. De Smedt, J., Yeshchenko, A., Polyvyanyy, A., De Weerdt, J., Mendling, J.: Process model forecasting and change exploration using time series analysis of event sequence data. Data Knowl. Eng. (2023)
8. Di Francescomarino, C., Ghidini, C., Maggi, F.M., Petrucci, G., Yeshchenko, A.: An eye into the future: Leveraging a-priori knowledge in predictive business process monitoring. In: BPM (2017)
9. Evermann, J., Rehse, J., Fettke, P.: Predicting process behaviour using deep learning. Decis. Support Syst. (2017)
10. Jalayer, A., Kahani, M., Beheshti, A., Pourmasoumi, A., Motahari-Nezhad, H.R.: Attention mechanism in predictive business process monitoring. In: EDOC (2020)
11. Kalyan, K.S.: A survey of GPT-3 family large language models including chatgpt and GPT-4. CoRR (2023)
12. Kampik, T., Warmuth, C., Rebmann, A., Agam, R., Egger, L.N.P., Gerber, A., Hoffart, J., Kolk, J., Herzig, P., Decker, G., van der Aa, H., Polyvyanyy, A., Rinderle-Ma, S., Weber, I., Weidlich, M.: Large process models: Business process management in the age of generative AI. CoRR **abs/2309.00900** (2023)
13. Le, M., Gabrys, B., Nauck, D.D.: A hybrid model for business process event prediction. In: SGAI Conf (2012)
14. Metzger, A., Leitner, P., Ivanovic, D., Schmieders, E., Franklin, R., Carro, M., Dustdar, S., Pohl, K.: Comparing and combining predictive business process monitoring techniques. IEEE Trans. Syst. Man Cybern. Syst. (2015)
15. Poll, R., Polyvyanyy, A., Rosemann, M., Röglinger, M., Rupprecht, L.: Process forecasting: Towards proactive business process management. In: BPM (2018)
16. Robinson, W.N., Ding, Y.: A survey of customization support in agent-based business process simulation tools. ACM Trans. Model. Comput. Simul. (2010)
17. Smirnov, S., Reijers, H.A., Weske, M., Nugteren, T.: Business process model abstraction: a definition, catalog, and survey. Distributed Parallel Databases (2012)
18. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: NIPS (2014)
19. Tax, N., Verenich, I., La Rosa, M., Dumas, M.: Predictive business process monitoring with LSTM neural networks. In: CAiSE (2017)
20. Tschumitschew, K., Nauck, D.D., Klawonn, F.: A classification algorithm for process sequences based on markov chains and bayesian networks. In: KES (1) (2010)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS (2017)
22. Verenich, I.: A general framework for predictive business process monitoring. In: CAiSE (Doctoral Consortium) (2016)
23. Yeshchenko, A., Di Ciccio, C., Mendling, J., Polyvyanyy, A.: Visual drift detection for event sequence data of business processes. IEEE Trans. Vis. Comput. Graph. (2022)