

Extending predictive process monitoring for collaborative processes

Daniel Calegari^{1,2} and Andrea Delgado¹

¹ Instituto de Computación, Facultad de Ingeniería, Universidad de la República
{adelgado,dcalegar}@fing.edu.uy
² Universidad ORT Uruguay
calegari@ort.edu.uy

Abstract. Process mining on business process execution data has focused primarily on orchestration-type processes performed in a single organization (intra-organizational). Collaborative (inter-organizational) processes, unlike those of orchestration type, expand several organizations (for example, in e-Government), adding complexity and various challenges both for their implementation and for their discovery, prediction, and analysis of their execution. Predictive process monitoring is based on exploiting execution data from past instances to predict the execution of current cases. It is possible to make predictions on the next activity and remaining time, among others, to anticipate possible deviations, violations, and delays in the processes to take preventive measures (e.g., re-allocation of resources). In this work, we propose an extension for collaborative processes of traditional process prediction, considering particularities of this type of process, which add information of interest in this context, for example, the next activity of which participant or the following message to be exchanged between two participants.

Keywords: Process mining, inter-organizational collaborative processes, predictive process monitoring

1 Introduction

Business processes are coordinated sets of activities designed to achieve specific business objectives [18]. Their execution generates a wealth of data for process evaluation and continuous improvement, whether in traditional or process-oriented information systems. Process Mining techniques [2]—encompassing discovery, conformance, and enhancement—enable complex, in-depth analyses of actual process executions. These techniques provide organizations with crucial insights into efficiency, quality, and regulatory compliance, thereby uncovering opportunities for evidence-based improvements. Such post-mortem analyses are invaluable, forming the foundation for systematic process enhancement and supporting informed decision-making.

Nevertheless, organizations demand the ability to exploit historical execution data and real-time observations to forecast the future states or outcomes of

ongoing business process instances. Predictive Process Monitoring [9] is a subfield of Process Mining that focuses on this aspect. Using event logs as input makes it possible to make accurate predictions regarding process execution, such as anticipating potential deviations, violations, or delays. These predictions enable proactive interventions—such as resource reallocation or adjustments in time management—to mitigate risks and ensure smooth process execution.

Traditionally, Process Mining research has concentrated on orchestration-type processes contained within a single organization (intra-organizational). In contrast, collaborative processes (inter-organizational), which involve multiple organizations (e.g., in e-Government), present additional complexity. Due to their inherently distributed and collaborative nature, these processes introduce challenges in implementation, discovery, prediction, and analysis [1,15].

Considering their unique characteristics, addressing the prediction needs for collaborative processes is critically important for organizations. For instance, beyond merely applying and evaluating existing “as-is” techniques for predicting the remaining execution time of a trace or the next event, it is essential to integrate or extend these methods. Incorporating elements such as the execution history of the involved participants can enable more accurate predictions of interest tailored to the collaborative context.

The primary objective of this work is to analyze existing techniques and approaches for predicting the execution of inter-organizational collaborative business processes using Process Mining. Moreover, it seeks to define, extend, or adapt predictive methods for collaborative business processes, emphasizing e-Government environments. This research aims to bridge the gap in predictive analytics for collaborative processes, providing organizations with tools to better manage and optimize their cross-organizational workflows.

The rest of the paper is structured as follows. We discuss related work in Section 2. In Section 3, we present collaboration event logs used as input for predictions. In Section 4, we present how prediction can be conceived in the context of collaborative processes, and in Section 5, we present an example application. Finally, in Section 6, we conclude and outline future work.

2 Related work

In recent years, different approaches to predictive process monitoring have been applied and evaluated [9,14]. These approaches use different encodings of the problem, exploiting an explicit representation of the process model to make the prediction or not, and also addressing classification or regression problems based on the type of predicted value (categorical or numerical). In this context, works are allowing remaining time (e.g., [3]), outcome (e.g., [13]), next event (e.g., [4]), and deviation (e.g., [12]) predictions, among others, that can be applied to business processes, as well as other types of industrial processes such as production, manufacturing or case-handling processes. More recently, works have spread into analyzing object-centric event logs, e.g., [10,17].

Most of the predictive methods are focused on the predicted value of an individual process instance [9,14]. Few works propose predictive process monitoring in the context of multiple process instances. It poses new challenges since predictions heavily depend on other cases running simultaneously, e.g. when cases share limited resources. In [7], the authors describe a recommendation system determining the probability of a risk in a system. It deals with the interplay between risks relative to multiple process instances running concurrently. As another example, in [16], the authors propose a method for feature encoding within a bi-dimensional space characterized by intra- and inter-case features.

Although these works considered multiple process instances, they did not consider such instances within a collaborative environment, as in the case of collaborative business processes. Few works consider this collaborative context. In [6], the authors propose a strategy for predicting remaining time in collaborative business processes while preserving organizations' privacy, i.e., by sharing hidden state information between sub-models. As far as we know, this is the only work that analyzes predictive process monitoring for collaborative processes.

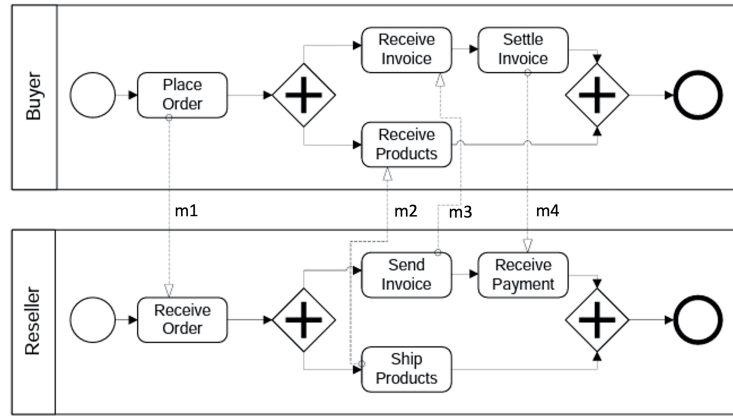
3 Event logs for collaborations

Making process predictions requires training, in which a predictive model is learned from historical (complete) execution traces in the form of an event log. Such models are then queried to predict the future of an ongoing case [9].

Within a collaborative environment, several organizations (or participants of an organization) interact with each other to carry out a global process. Single orchestrations exist for each participant, and messages allow them to coordinate.

Traditional event logs only sometimes register data regarding which participant enacts the activity of the corresponding role/person. To cope with this limitation, we have defined an extension [11] of the XES format for collaborative business process, which was tested for discovering collaborative processes [15]. A collaboration event log comprises collaborative cases involving several participants whose events come from different participants and include mandatory attributes: **participant** to identify the participant that enacts the event, **elemType** with the type of the event (user, message), and if the event is of type **message**, from which participant it is being received (**fromParticipant**) or to which participant it is being sent (**toParticipant**). As described in [15], a collaboration event log can be built from the participants' logs by merging them.

In Figure 1a, a reseller company's ordering process is depicted as a collaborative business process between a Buyer and a Reseller. Figure 1b shows an example of an event log for each participant, and Figure 1b shows an extended event log for the collaboration. The collaboration event log merges each participant's event log, considering the timestamp to maintain the order of execution in the general view of the collaborative case.



(a) Example collaborative business process with messages from [18]

Buyer event Log		Collaboration event Log	
Event = Place Order		Event = Place Order	
* timestamp = 01/04/23 10-07-22		* Participant = Buyer	
Receive Invoice		* toParticipant = Reseller	
* timestamp = 01/04/23 12-21-05		* timestamp = 01/04/23 10-07-22	
Receive Products		Event = Receive Order	
* timestamp = 05/04/23 18-11-02		* Participant = Reseller	
Settle Invoice		* fromParticipant = Buyer	
* timestamp = 06/04/23 09-10-01		* timestamp = 01/04/23 10-10-09	
		Event = Send Invoice	
		* Participant = Reseller	
		* toParticipant = Buyer	
		* timestamp = 01/04/23 12-11-06	
		Event = Receive Invoice	
		* Participant = Buyer	
		* fromParticipant = Reseller	
		* timestamp = 01/04/23 12-21-05	
		...	
Reseller event Log			
Receive Order			
* timestamp = 01/04/23 10-10-09			
Send Invoice			
* timestamp = 01/04/23 12-11-06			
Ship Products			
* timestamp = 05/04/23 16-12-03			
Receive Payment			
* timestamp = 07/04/23 13-08-03			

(b) Participants' reduced event logs (orchestration)

(c) Extended event log for collaborations

Fig. 1: Example collaborative process and associated event logs from [15]

4 Predictive collaborative process monitoring

According to the literature, it is possible to classify the existing prediction types into three categories [9]:

1. Outcome-based: predictions related to predefined categorical or boolean outcome values, e.g., predicting if a particular process path will occur.

2. Numeric value: predictions related to measures of interest taking numeric or continuous values, e.g., time prediction such as estimating how long it will take for a currently running process instance to complete.
3. Next event: predictions related to sequences of future activities and related data payloads, e.g., forecasting the next step or event likely to occur.

Considering the importance of main collaborative concepts (participants and messages), the kind of predictions could be extended, as summarized in Table 1.

Table 1: Non-exhaustive type of predictions for collaborative processes

Category	Predictions
Outcome-based	If a participant will participate in a case. If a particular message will be sent/received.
Numeric value	Number of remaining/total messages of a participant. Number of remaining/total messages in the process. Participant remaining/duration time (other could be active). Process remaining/duration time (every participant finishes). Time until the next message to send/receive.
Next Event	Next event that is likely to occur in a participant. Next event that is likely to occur in the process. Next participant that is likely to act. Next participant that is likely to send/receive a message. Next message (send/receive) that is likely to occur in the process. Next message (send/receive) that is likely to occur in a participant.

Broadly speaking, the possibilities are divided between predictions linked to the process as a whole or a specific participant (a problem reduced to forecasts within a single organization). For example, concerning the model in Figure 1a, it could be possible to focus predictions from the perspective of the Buyer and estimate what will be the next event if the case is on the place order activity or the remaining time of such a participant. If considering the process as a whole, the next activity can be within the Buyer or the Reseller (e.g., receiving an order), and the remaining time needs to consider both participants. Moreover, it is possible to identify messages as events that occur and are registered, similar to process activities. Again, in the Buyer's example, it could be possible to estimate whether the following message to receive will be the receipt of the product (m2) or the invoice (m3).

4.1 Adapting predictive methods

As explained next, the problem of making predictions within a collaborative environment could be reduced to the problem of making predictions for single

orchestrations. The critical point is the form of the collaboration event log itself. Since it merges each participant’s event log, it is like an orchestration log. Moreover, for each event, it is possible to identify the participant who enacts it and whether the event concerns the sending/reception of a message. This information takes the form of attributes that can be used for outcome-based and numeric value predictions.

Notice that many encodings and prediction methods are possible. For example, to predict the following message that is likely to occur within a participant, it is possible to take all the events of a case, only the events of a particular participant, or only the message events of the case or participant. Each one of these options offers more or less information concerning the problem to resolve, which is later used to train the predictive model. As a result, it could yield more or less precise results. This aspect must be evaluated in depth in future work.

Outcome-based predictions could be addressed by taking the participant or message names as the actual (categorical) values of a variable that we aim to predict, similar to any outcome-based prediction in the case of an orchestration. The encoding of traces could consider the whole process, e.g., in the case of predicting the participant, or a concrete participant, e.g., in the case of predicting a particular message.

Numeric value predictions such as the number of remaining/total messages can be addressed similarly to the number of remaining/total events in an orchestration. Event logs for training could consider the events or only messages of the whole process or a particular participant. The case of predicting a participant’s remaining/duration time is precisely the orchestration case. To predict the same in the case of the whole process, we can take the entire event log as being of one participant. Finally, predicting the time until the following message is similar to predicting the time until the next event of a particular type.

Next event predictions are the most straightforward and only require taking the event log as a whole or reducing it for a given participant. The case of predicting the following message is similar to predicting the next event of a particular type (if any). Finally, predicting the next participant is similar to predicting the next event with the desired participant as an attribute.

5 Assessment

We have performed a preliminary assessment of the ideas by implementing a prediction tool based on the work in [5]. In such work, the authors propose ProcessTransformer, an approach for learning high-level representations from event logs with an attention-based network. The transformer allows for predicting the next activity, the event time, and the remaining time for a running case. We have

used such a transformer as a basis and performed the transformations mentioned in the last section to make collaborative predictions.

As an application example, we use the healthcare collaborative process introduced in [8] and depicted in Figure 2. The process illustrates a healthcare scenario about the treatment of gynecological diseases.

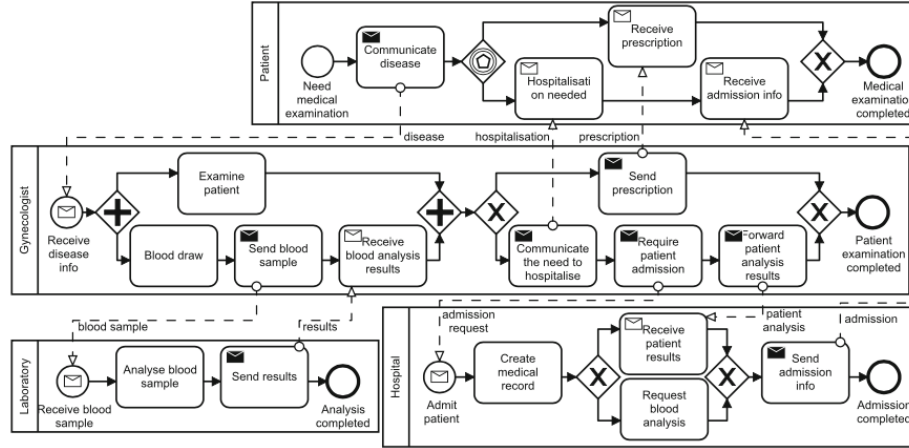


Fig. 2: A healthcare business process collaboration from [8]

We took several collaborative logs conforming to such a process. The tool allowed the upload of a log with the complete traces to be used for the training phase. Figure 3a depicts how the log is uploaded and the type of prediction selected. In this case, we were interested in the following send message/participant that is likely to occur. We also selected the column that stores the information for training the model, i.e., the message information. Every other event, not being a message, was ignored for training (just for the example).

After training, a second page allows uploading a new log (of incomplete traces), and for each trace, the transformer predicts the following send message. There are three cases in the example depicted in Figure 3b. The first case (case_44) corresponds to a trace where the Patient participant already communicated the disease to the Gynecologist (start event Receive disease info), and the Gynecologist already sent the Send blood sample message to the Laboratory-Laboratory (from the second path in the parallel gateway) and its corresponding reception (Receive blood sample). Based on this information, the prediction correctly retrieves the message to be sent back to the Gynecologist (Send results).

In the other two traces, predictions correspond to messages to be sent from the Gynecologist participant, e.g., the second trace (case_9) corresponds to the Send blood sample message. All predictions can be downloaded as a .csv file.

Create and train your model

Log:
 collectivelog_healthcare_collab_2.csv

Column
 MP

Type of prediction
 Select a type of prediction
 Select a type of prediction
 Next message to send/participant
 Next activity/participant
 Next participant

(a) Training phase

Prediction results

Case	Prediction
case_44	sendtask-laboratory
case_9	sendtask-gynecologist
case_209	sendtask-gynecologist

Download CSV

(b) Prediction result

Fig. 3: Web tool training and prediction results

6 Conclusions

In this work, we analyzed how traditional predictive process monitoring could be extended to inter-organizational collaborative business processes. Although concepts such as participant and message become more relevant in this context, conventional prediction techniques can be adapted to consider them.

We theoretically analyzed how this adaptation could be performed and then showed preliminary results of how some of the predictions could be achieved practically. Since our analysis is preliminary, other prediction types could be defined, and new techniques can be explored and tailored in this context. As mentioned, many possible encodings and prediction methods exist for the same problem. In this context, a more rigorous experimental evaluation is necessary to assess the quality of the predictions and the suitability of methods for predicting certain specific aspects of interest.

7 Acknowledgment

We want to thank the undergraduate students who worked on the practical application: Carolina Espino and Nicolás Ribero. Supported by project “Minería de procesos y datos para la mejora de procesos colaborativos aplicada a e-Government” funded by Agencia Nacional de Investigación e Innovación (ANII), Fondo María Viñas (FMV) ”Proyecto ANII N° FMV_1_2021_1_167483”, Uruguay

References

1. van der Aalst, W.M.P.: Intra- and inter-organizational process mining: Discovering processes within and between organizations. In: *The Practice of Enterprise Modeling (PoEM 2011)*. LNCS, vol. 92, pp. 1–11. Springer (2011)
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.* **36**(2), 450–475 (2011)
4. Breuker, D., Delfmann, P., Matzner, M., Becker, J.: Designing and evaluating an interpretable predictive modeling technique for business processes. In: *Business Process Management Workshops*. pp. 541–553. Springer, Cham (2015)
5. Bukhsh, Z.A., Saeed, A., Dijkman, R.M.: Processtransformer: Predictive business process monitoring with transformer network (2021), <https://arxiv.org/abs/2104.00721>
6. Cao, J., Wang, C., Guan, W., Qian, S., Zhao, H.: Remaining time prediction for collaborative business processes with privacy preservation. In: *Service-Oriented Computing*. pp. 38–53. Springer, Cham (2023)
7. Conforti, R., de Leoni, M., La Rosa, M., van der Aalst, W.M., ter Hofstede, A.H.: A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems* **69**, 1–19 (2015)
8. Corradini, F., Re, B., Rossi, L., Tiezzi, F.: A technique for collaboration discovery. In: *Enterprise, Business-Process and Information Systems Modeling*. pp. 63–78. Springer International Publishing (2022)
9. Di Francescomarino, C., Ghidini, C.: *Predictive Process Monitoring*, pp. 320–346. Springer, Cham (2022)
10. Galanti, R., de Leoni, M., Navarin, N., Marazzi, A.: Object-centric process predictive analytics. *Expert Syst. Appl.* **213**(Part), 119173 (2023)
11. González, L., Delgado, A.: Compliance requirements model for collaborative business process and evaluation with process mining. In: *XLVII Latin American Computing Conference (CLEI)*. IEEE (2021)
12. Grohs, M., Pfeiffer, P., Rehse, J.: Business process deviation prediction: Predicting non-conforming process behavior. In: *5th Intl. Conf. on Process Mining, ICPM*. pp. 113–120. IEEE (2023)
13. Hong Tu, T.B., Song, M.: Analysis and prediction cost of manufacturing process based on process mining. In: *2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*. pp. 1–5 (2016)
14. Márquez-Chamorro, A.E., Resinas, M., Ruiz-Cortés, A.: Predictive monitoring of business processes: A survey. *IEEE Trans. Serv. Comput.* **11**(6), 962–977 (2018)

15. Peña, L., Andrade, D., Delgado, A., Calegari, D.: An approach for discovering inter-organizational collaborative business processes in BPMN 2.0. In: Smedt, J.D., Soffer, P. (eds.) *Process Mining Workshops - ICPM 2023 Intl. Workshops, Revised Selected Papers*. LNBIP, vol. 503, pp. 487–498. Springer (2023)
16. Senderovich, A., Francescomarino, C.D., Maggi, F.M.: From knowledge-driven to data-driven inter-case feature encoding in predictive process monitoring. *Information Systems* **84**, 255–264 (2019)
17. Smit, T.K., Reijers, H.A., Lu, X.: HOEG: A new approach for object-centric predictive process monitoring. In: *Advanced Information Systems Engineering - 36th Intl. Conf., CAiSE, Proceedings*. LNCS, vol. 14663, pp. 231–247. Springer (2024)
18. Weske, M.: *Business Process Management - Concepts, Languages, Architectures, Third Edition*. Springer (2019)