Prediction of Business Process Instances with Dynamic Bayesian Networks

Jens Brunk[1], Kate Revoredo[2], Matthias Stierle[3], Martin Matzner[3], Patrick Delfmann[4], and Jörg Becker[1]

[1] European Research Center for Information Systems (ERCIS), University of Münster, {jens.brunk,becker}@ercis.uni-muenster.de
[2] Federal University of Rio de Janeiro katerevoredo@ppgi.ufrj.br
[3] Chair of Digital Industrial Service Systems, Friedrich-Alexander Universität Nürnberg-Erlangen, {matthias.stierle,martin.matzner}@fau.de
[4] Institute for Information Systems Research, University of Koblenz-Landau, delfmann@uni-koblenz.de

Abstract. Predicting undesirable events during a process instance execution provides the process participants with an opportunity to intervene and keep the process aligned with its goals. Few approaches for tackling this challenge consider a multi-perspective view, where the flow perspective of the process is combined with its surrounding context. In particular, the dynamism of this context over time has been ignored in most prediction approaches. In this paper we tackle this issue by leveraging previous work on probabilistic finite automata to develop a Dynamic Bayesian Network (DBN). The DBN includes different kinds of context information in customizable process models and then predicts the next event of a process instance. The initial results reveal two major challenges: choosing the optimal DBN structure and using meaningful contextual information.

1 Introduction

Monitoring a business process instance or rather anticipating an unexpected event, provides an opportunity for decision-makers to evaluate the best way to overcome obstacles and to continuously achieve the process goals. The task of learning a process model that is also able to predict future behavior of a process instance is gaining increased attention [4, 11]. In [2], an approach, henceforth called RegPFA, was proposed, where a probabilistic predictive model is learned through process log data and used to predict future events given an observed sequence of events. This probabilistic model also represents the process model, thus RegPFA is a process-aware approach The evaluation of the method showed that an improvement in sensitivity is necessary, i.e., an improvement on the rate of correct predictions of a next event. The hypothesis we investigate in this paper is that solely looking at the control-flow data is not enough towards learning an accurate model. For example, when predicting whether the next event in a procurement process after sending out the order is the goods receipt or cancellation of the order. Then we can expect that the prediction is highly influenced by whether the ordered goods were cheap or expensive and/or which vendor we ordered from. The data attributes related to an activity during process execution are called contextual elements [18]. They define the context in which the ongoing instance is running. Thus, a certain event may occur conditioned on the current state of the process and a set of contextual elements. For example, if in a procurement process an order is currently in the *goods and invoice received* state and the order value is below 500, then the next event will be the payment of the invoice (whereas if the goods were not received yet or the invoice has a high value, another event would occur). Furthermore, the contextual elements in this set may have their values determined by the current state (e.g. if the state is *goods received* the order is not blocked for payment) and additionally they can, together with the current event, determine the next state of a process (e.g. if the number of goods received are less than expected, the next state could be *Quality Check*). A model that represents these dependencies, as well as uncertainty in terms of changing context, may be more adequate for prediction.

In this work, we investigate the benefits of considering dynamic contextual information to process-aware prediction. More specifically, we extend the RegPFA approach towards considering contextual information that can dynamically change over time. i.e. throughout process execution. Therefore, we propose a technique for learning a Dynamic Bayesian Network (DBN). We investigate the hypothesis that the combination of historical events and dynamically changing contextual information improves the prediction of unknown future events.

Following the Design Science Research (DSR) paradigm defined by Hevner [8] we build two artifacts which is, on the one hand, a theoretical model of the DBN as well as its instantiation. Our method is guided by the DSR process proposed by Peffers et al. [15] of which we finished the first iteration of *build, demonstrate and evaluate* leading us to the challenges of how to choose the optimal network structure and how to identify meaningful contextual information.

2 Related Work

In this section, we describe works that perform prediction in the presence of contextual data and thus are related to our proposal. These works were evaluated from two perspectives: *process-aware* [11] and *attribute-dynamic-aware*. In the former, an approach is considered *process-aware* if the process model is used as input for the prediction. In the latter, an approach is *attribute-dynamic-aware* if the change over time of the attribute is considered for the prediction.

Fifteen related works were found and we could observe that none of them considered the process model during prediction. On regards of *attribute-dynamic-aware* the works [20, 22, 9, 12, 10, 5] considered the dynamism of the attributes for the prediction while the works [19, 21, 1, 7, 3, 14, 17, 16, 6] did not.

The limitation that we address in this paper is that the minority of works consider the dynamism of the attributes and that they generally neglect the process model for prediction.

3 Proposal

Our proposal follows the general idea behind the RegPFA [2], which is based on probabilistic finite automata (PFA). They represent the model through hidden and observed nodes for which parametrization is key to modeling a process and execute predictions for it. Figure 1 (left part) depicts a PFA, where z_0 and z_1 represent the hidden state variables and $x_0 ... x_n$ the activity sequence event. Since apart from the events (observed nodes) our approach will also model the contextual information, we will make use of Dynamic Bayesian Networks (DBN) [13]. In our case, we consider *two-time-slice DBNs*. This means that the visual representation consists of an initial first time slice and a second slice, which is repeated indefinitely to model the steps of the time dimension. Although there are only two time slices, the value of the variables change over time and in this way the dynamism of the context is modeled.

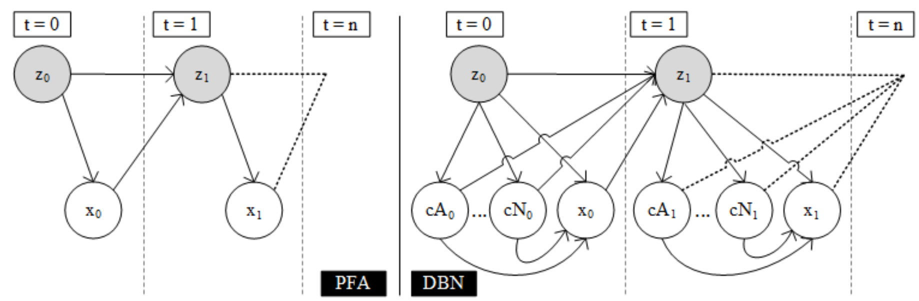

Fig. 1. PFA and DBN Model Structure

In Figure 1 (right side), nodes $cA_0 ... cN_1$ represent contextual elements, which are dependent on the hidden state and influence both, the event and the next state. Similarly, more context variables and many different DBN structures can be imagined.

The major challenge of log data based approaches is learning and running inference on the created models. As the idea behind using DBNs is to generalize and be able to construct many different structures, it is important that the available learning and inference algorithms are able to support this. To date, there exist various frameworks for probabilistic graphical models [5] and each of them implements a different set of inference algorithms. These algorithms can generally be distinguished by doing exact/approximate inference, the support of different node types (e.g. discrete and/or continue) and the type of topology that they support for the DBN. To our knowledge, only the Bayes Net Toolbox includes an inference algorithm, which supports any topology, different node types and that can be used for inference as well as prediction. For this reason it was our choice for implementing our solution. Our implementation supports the creation

[5] AMIDST Toolbox: https://github.com/amidst/toolbox; Graphical Models Toolkit: https://melodi.ee.washington.edu/gmtk/; Bayes Net Toolbox: https://github.com/bayesnet/bnt

of HMM, PFA and freely customizable DBNs and the prediction of the next event of process instances.

4 Evaluation and Discussion

To assure the correctness of the implementation we benchmark our PFA implementation against the RegPFA with the dataset of the BPI challenge 2012[6] and yield similar results. The slightly better performance of RegPFA is expected, as we implemented only a basic version, especially in terms of optimizing input parameters. As initial evaluation of our context-aware prediction approach we considered the *resource* column of the BPI2012 dataset as a single context attribute. The results displayed in Table 1 show the impact of considering contextual information and its dynamic over time. The preliminary results indicate that our approach is underperforming for all three performance measurements and it is far from the results yielded by the PFA/RegPFA.

Table 1. Predictor Performance Measures

Event log	Predictor	Accuracy	Sensitivity	Specificity
BPI2012_A	RegPFA	0,801	0,723	0,980
	PFA	0,6981	0,6997	0,6986
	DBN	0,2083	0,1594	0,2074

We can identify two major reasons for the bad results. First the *quality* or rather *suitability* of the context data. Looking at the content of the *resource* column that was used, we have to notice that it is hardly correlated to the process flow. Thus, choosing the context is a challenge. Second the network structure that we chose (see Figure 1). Currently, the contextual elements are both influencing the observation as well as the future state and they themselves are influenced by the current state. Especially the latter seems to introduce a high degree of uncertainty into the model as the context in $t + 1$ is unknown when predicting the observation in $t + 1$.

Given these observations, we plan to continue our research by running further evaluations with various datasets and different combinations of context attributes as well as by varying the network structure.

5 Conclusion and Future Research

We implemented a DBN based business process prediction approach that is based on the Bayes Net Toolbox for MATLAB. Our approach contributes to the body of knowledge by introducing a process-aware business process prediction approach that can handle dynamic context attributes and is based on probabilistic graphical models. It enables and calls for future research on which types of dynamic context

[6] https://data.4tu.nl/repository/uuid:3926db30-f712-4394-aebc-75976070e91f

attributes can be used for business process instance predictions and which structures work best. This is also the major limitation of our work, as we developed and initially evaluated the approach, but in-depth discussions and studies, as well as suitable data logs, are needed to improve the accuracy of our DBN predictions, until they outperform e.g. the RegPFA.

Acknowledgments: This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skodowska-Curie Grant agreement No 645751.

References

1. Borkowski, M., Fdhila, W., Nardelli, M., Rinderle-Ma, S., Schulte, S.: Event-based failure prediction in distributed business processes. Information Systems (2017)
2. Breuker, D., Matzner, M., Delfmann, P., Becker, J.: Comprehensible predictive models for business processes. MIS Quarterly **40**(4), 1009–1034 (2016)
3. Di Francescomarino, C., Dumas, M., Maggi, F.M., Teinemaa, I.: Clustering-based predictive process monitoring. IEEE Transactions on Services Computing (2016)
4. Di Francescomarino, C., Ghidini, C., Maggi, F.M., Milani, F.: Predictive process monitoring methods: Which one suits me best? In: BPM. pp. 462–479 (2018)
5. 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. pp. 252–268. Springer, Cham (2017)
6. Folino, F., Guarascio, M., Pontieri, L.: Discovering context-aware models for predicting business process performances. In: OTM. pp. 287–304. Springer (2012)
7. Frey, M., Emrich, A., Fettke, P., Loos, P.: Event entry time prediction in financial business processes using machine learning: A use case from loan applications. In: HICSS. pp. 1386–1394. IEEE Computer Society (2018)
8. Hevner, A.R.: Design science in information systems research. MIS Quarterly **28**(1), 75–105 (2004)
9. Leontjeva, A., Conforti, R., Di Francescomarino, C., Dumas, M., Maggi, F.M.: Complex Symbolic Sequence Encodings for Predictive Monitoring of Business Processes. pp. 297–313. Springer, Cham (2015)
10. Maggi, F.M., Di Francescomarino, C., Dumas, M., Ghidini, C.: Predictive monitoring of business processes. In: CAISE. pp. 457–472. Springer (2014)
11. Marquez-Chamorro, A.E., Resinas, M., Ruiz-Cortes, A.: Predictive monitoring of business processes: a survey. IEEE Transactions on Services Computing (2017)
12. Marquez-Chamorro, A.E., Resinas, M., Ruiz-Corts, A., Toro, M.: Run-time prediction of business process indicators using evolutionary decision rules. Expert Syst. Appl. **87**(C), 1–14 (2017)
13. Murphy, K.P., Russell, S.: Dynamic bayesian networks: representation, inference and learning (2002)
14. Navarin, N., Vincenzi, B., Polato, M., Sperduti, A.: LSTM networks for data-aware remaining time prediction of business process instances. CoRR (2017)
15. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems **24**(3), 45–77 (2007)
16. Polato, M., Sperduti, A., Burattin, A., de Leoni, M.: Data-aware remaining time prediction of business process instances. In: In IJCNN. pp. 816–823 (2014)
17. Polato, M., Sperduti, A., Burattin, A., de Leoni, M.: Time and activity sequence prediction of business process instances. CoRR **abs/1602.07566** (2016)
18. Rosemann, M., Recker, J.C., Flender, C., Ansell, P.D.: Understanding context-awareness in business process design (2006)
19. Senderovich, A., Francescomarino, C.D., Ghidini, C., Jorbina, K., Maggi, F.M.: Intra and inter-case features in predictive process monitoring: A tale of two dimensions. In: BPM. vol. 10445, pp. 306–323. Springer (2017)
20. Teinemaa, I., Dumas, M., Maggi, F.M., Francescomarino, C.D.: Predictive business process monitoring with structured and unstructured data. In: BPM. vol. 9850, pp. 401–417. Springer (2016)
21. Unuvar, M., Lakshmanan, G.T., Doganata, Y.N.: Leveraging path information to generate predictions for parallel business processes. Knowledge and Information Systems **47**(2), 433–461 (2016)
22. Verenich, I., Dumas, M., Rosa, M.L., Maggi, F.M., Francescomarino, C.D.: Complex symbolic sequence clustering and multiple classifiers for predictive process monitoring. In: BPM Workshops. vol. 256, pp. 218–229. Springer (2015)