

Text-Aware Predictive Monitoring of Business Processes

Marco Pegoraro¹[\[https://orcid.org/0000-0002-8997-7517\]](https://orcid.org/0000-0002-8997-7517), Merih Seran Uysal¹[\[https://orcid.org/0000-0003-1115-6601\]](https://orcid.org/0000-0003-1115-6601), David Benedikt Georgi¹, and Wil M.P. van der Aalst¹[\[https://orcid.org/0000-0002-0955-6940\]](https://orcid.org/0000-0002-0955-6940)

¹Process and Data Science chair, RWTH Aachen University, Aachen, Germany
{pegoraro,uysal,wvdaalst}@pads.rwth-aachen.de david.georgi@rwth-aachen.de

Abstract. The real-time prediction of business processes using historical event data is an important capability of modern business process monitoring systems. Existing process prediction methods are able to also exploit the data perspective of recorded events, in addition to the control-flow perspective. However, while well-structured numerical or categorical attributes are considered in many prediction techniques, almost no technique is able to utilize text documents written in natural language, which can hold information critical to the prediction task. In this paper, we illustrate the design, implementation, and evaluation of a novel *text-aware process prediction model* based on *Long Short-Term Memory* (LSTM) neural networks and natural language models. The proposed model can take categorical, numerical and textual attributes in event data into account to predict the activity and timestamp of the next event, the outcome, and the cycle time of a running process instance. Experiments show that the text-aware model is able to outperform state-of-the-art process prediction methods on simulated and real-world event logs containing textual data.

Keywords: Predictive Monitoring, Process Mining, Natural Language Processing, LSTM Neural Networks

1 Introduction

In recent years, a progressive and rapid tendency to digital transformation has become apparent in most aspects of industrial production, provision of services, science, education, and leisure. This has, in turn, caused the widespread adoption of new technologies to support human activities. A significant number of these technologies specialize in the management of enterprise business processes.

The need of analysis and compliance in business processes, united to a larger and larger availability of historical event data have stimulated the birth and growth of the scientific discipline of *process mining*. Process mining enables the discovery of process models from historical execution data, the measurement of compliance between data and a process model, and the enhancement of process models with additional information extracted from complete process cases.

Advancements in process mining and other branches of data science have also enabled the possibility of adopting *prediction* techniques, algorithms that train a mathematical model from known data instances and are able to perform accurate estimates of various features of future instances. In the specific context of process mining, *predictive monitoring* is the task of

- Python (PM4Py): Bridging the Gap Between Process-and Data Science.” In *International Conference on Process Mining (ICPM) Demo Track (CEUR 2374)*. pp. 13–16 (2019)
4. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation." *the Journal of Machine Learning Research* 3 (2003): 993-1022.
 5. Brown, Peter F., Vincent J. Della Pietra, Peter V. Desouza, Jennifer C. Lai, and Robert L. Mercer. "Class-based n-gram models of natural language." *Computational linguistics* 18.4 (1992): 467-480.
 6. Ceci, Michelangelo, Pasqua Fabiana Lanotte, Fabio Fumarola, Dario Pietro Cavallo, and Donato Malerba. "Completion time and next activity prediction of processes using sequential pattern mining." In *International Conference on Discovery Science*, pp. 49-61. Springer, Cham, 2014.
 7. Evermann, Joerg, Jana-Rebecca Rehse, and Peter Fettke. "A deep learning approach for predicting process behaviour at runtime." In *International Conference on Business Process Management*, pp. 327-338. Springer, Cham, 2016.
 8. Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." In *International Conference on Machine Learning*, pp. 1188-1196. PMLR, 2014.
 9. Navarin, Nicolò, Beatrice Vincenzi, Mirko Polato, and Alessandro Sperduti. "LSTM networks for data-aware remaining time prediction of business process instances." In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-7. IEEE, 2017.
 10. Park, Gyunam, and Minseok Song. "Predicting performances in business processes using deep neural networks." *Decision Support Systems* 129 (2020): 113191.
 11. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-learn: Machine learning in Python." *the Journal of Machine Learning Research* 12 (2011): 2825-2830.
 12. Polato, Mirko, Alessandro Sperduti, Andrea Burattin, and Massimiliano de Leoni. "Time and activity sequence prediction of business process instances." *Computing* 100.9 (2018): 1005-1031.
 13. Rogge-Solti, Andreas, and Mathias Weske. "Prediction of remaining service execution time using stochastic Petri nets with arbitrary firing delays." In *International Conference on Service-Oriented Computing*, pp. 389-403. Springer, Berlin, Heidelberg, 2013.
 14. Tax, Niek, Irene Teinemaa, and Sebastiaan J. van Zelst. "An interdisciplinary comparison of sequence modeling methods for next-element prediction." *Software and Systems Modeling* 19.6 (2020): 1345-1365.
 15. Tax, Niek, Ilya Verenich, Marcello La Rosa, and Marlon Dumas. "Predictive business process monitoring with LSTM neural networks." In *International Conference on Advanced Information Systems Engineering*, pp. 477-492. Springer, Cham, 2017.
 16. Teinemaa, Irene, Marlon Dumas, Fabrizio Maria Maggi, and Chiara Di Francescomarino. "Predictive business process monitoring with structured and unstructured data." In *International Conference on Business Process Management*, pp. 401-417. Springer, Cham, 2016.
 17. Uysal, Merih Seran, Sebastiaan J. van Zelst and Tobias Brockhoff and Anahita Farhang Ghahfarokhi and Mahsa Pourbafrani and Ruben Schumacher and Sebastian Junglas and Günther Schuh and Wil M.P. van der Aalst. "Process Mining for Production Processes in the Automotive Industry" In *Industry Forum at BPM 2020 co-located with International Conference on Business Process Management*, Springer, 2020.