Explainable Process Predictions (xPP): A Holistic Framework and Applications (Extended Abstract)

Nijat Mehdiyev
German Research Center for Artificial Intelligence (DFKI)
Saarland University
Saarbrücken, Germany
nijat.mehdiyev@dfki.de

I. Introduction

Business process prediction also referred to as predictive process monitoring or predictive business process management is a branch of process mining that pursues the objective to predict the target of interest by using the activities from the process traces. Recently, several studies have been conducted to explore the applicability of various machine learning approaches for different problems in the process prediction context such as next event prediction, process outcome prediction, prediction of service level agreement violations, remaining time prediction, risk prediction, cost prediction, prediction of activity delays etc. The recent research also suggests that the black-box machine learning approaches especially deep learning methods provide superior results for process prediction problems compared to conventional approaches. However, these opaque, non-transparent models lack the capabilities to provide explanations about their reasoning trace or delivered outcomes. This in turn introduces the barriers to operationalizing data-driven decision-making since the users tend not to use the outcomes by such artificial advice givers due to the lack of understanding or justification.

II. Thesis Contribution

Explainable Artificial Intelligence (XAI) has recently reemerged as an important research domain with the purpose to establish the trust between human users and AI systems by making the communication understandable and transparent. Although, the predictive process analytics has been recently emerging as an important research area, the explainability issues in this domain have only been partially addressed. To fill this research gap, this thesis makes *three major contributions*. **First**, this study explores the applicability of black-box machine learning approaches, particularly deep neural networks, for different process prediction problems in various domains. **Second**, this study attempts to propose a theory-driven conceptual framework which is presumed to guide developing explainable process prediction solutions by providing an overview of important aspects of decision-making situations. **Third**, the applicability of explainable process predictions is illustrated by adopting well-recognized or developing new explanation methods for different use-cases.

A. *Black-Box Machine Learning for Process Prediction*

The predictive strength of the adopted machine learning models is one of the most important prerequisites for generating reliable robust, and consistent explanations. For this purpose, we have investigated various black box approaches for different process prediction problems. In one of our earlier studies we applied pre-trained stacked autoencoder based deep neural networks to address next event prediction problem after carrying out an intensive data-preprocessing procedure on the event log data including n-gram encoding, extracting data flow and resource features and feature hashing [1]. Consequently, this study was extended by applying hyperparameter optimization of the adopted deep learning approach and by addressing imbalanced classification problem [2]. In another study, a deep LSTM approach was applied on sensor time series data from the process industry to detect the quality of the semi-finished products and accordingly to predict the next production process step [3]. The necessity of this approach and an overview of other relevant methods and systems for industrial predictive process analytics were presented in [4]. Furthermore, a discussion of predictive process analytics based on the log data generated by Manufacturing Execution Systems (MES) was introduced in [5] which also presented a use-case from individual manufacturing.

B. *A Framework for Explainable Process Predictions*

Making process predictions delivered by black-box machine learning models explainable is a multi-dimensional and multi-faceted issue that requires to consider the context of explanation situation, the users‘ preferences and backgrounds, the nature of underlying processes, the defined technical and economic objectives, the organizational factors, etc. when developing explanation systems. Hence, it conceivable to suggest there is no "one-fits-all" explainable process prediction solution and a systematic approach is required in generating adequate explanations by incorporating the implications from various dimensions of the decision-making environment. To fill this research gap, this thesis aims to propose a ***conceptual framework*** which is supposed to guide the process mining practitioners and researchers in designing and developing the explanation solutions for predictive process monitoring problems. For this purpose, **our study [6]** proposed an initial holistic framework by analyzing, combining and adapting the propositions from the explainable artificial intelligence research domain. The constructs of the proposed framework include ***subjects, objectives, instruments/techniques, context, generation time*** and other various elements. The ***subjects*** of explainable process prediction solutions are various stakeholders with different levels of knowledge background, expertise levels and explanation preferences. *Process owners, process/data analysts, process/data engineers, domain experts, regulatory authorities, supervisory bodies* etc. are some of the key users for each of them customized explanation methods have to be designed that conform to these users' mental models. These users pursue various ***objectives*** such as *justification, ratification, verification, duplication, debugging, learning, satisfaction, effectiveness/efficiency* etc. For instance, the knowledge engineers such as process engineers are more interested in the reasoning mechanism of the black-box approaches, follow mainly the verification objective and prefer the explanation methods that provide algorithmic transparency. On the other side, the domain experts or process owners with limited machine learning background opt to justify the goodness of the models' individual decisions. There have been numerous studies to investigate various XAI ***techniques*** by proposing different taxonomies. E.g. the explanation techniques can be classified as methods related to *transparency* which aim to examine how the model works or as *post-hoc* explanation approaches that attempt to extract the explanation from learned model without investigating the reasoning mechanism of the model. Regarding the relationship with the underlying black-box model the explanations are categorized as *model-specific* or *model-agnostic approaches.* A further categorization refers to the scope of the generated explanations. The *global* explanation approaches attempt to deliver explanations for the whole dataset whereas the *local* explanation techniques make the individual observations explainable. ***Generation time*** informs the users when the explanations are generated. They can be obtained before developing the black-box machine learning models *(pre-model),* during the model implementation *(in-model)* and finally after the model is trained *(post-model).*

C. Applying/Developing Explainable Process Prediction Solutions

In this thesis, the proposed conceptual framework is instantiated to develop various explainable process prediction solutions. For illustrative purposes, three use-cases are presented which we have already examined in various studies [6]–[8]. ***In the firs use-case***, the main users of the explainable process prediction solutions are the process owners of a Dutch autonomous administrative authority [7]. They are interested to understand the user behavior globally that leads to using more expensive channels such as sending messages. We applied first a deep neural network to predict when the users tend to send messages. Since the objective in this use-case was defined as understanding the global user behavior to make a strategic process enhancement decision, we have adopted the Partial Dependence Plots (PDP), a global, model-agnostic post-hoc explanation technique to generate the relevant causal explanations. ***In the second use case***, we explored the potential of explainable process predictions to enable data-driven process planning in manufacturing [8]. For this purpose, two complementary local post-hoc explanation approaches, Shapley Values, and Individual Conditional Expectation (ICE) plots are applied which facilitate the domain experts to examine the explanations from different perspectives. ***In the third use-case***, we explored the applicability of the explainable process predictions for an incident management with the purpose to make model decisions for domain experts justifiable [6]. Different from previous two studies, which adopted known XAI methods, in this study we have proposed a novel local post-hoc explanation approach by defining the local regions from the validation data by using the intermediate learned representations of the applied black box approach. Particularly, the learned neural codes from the last hidden layer of the applied neural network were used as input to clustering algorithm to define the local regions. Finally, a local surrogate decision tree was fitted. By adopting this approach, we aimed to avoid the shortcomings related to perturbation-based approaches such as LIME or Shapley. Furthermore, our other studies on explainable process analytics address various use-cases and approaches [9], [10].

III. Outlook

The future work comprises activities to enhance the proposed framework for explainable process predictions through the lens of information systems theory, particularly activity theory combined with the design science research. Furthermore, a thorough integration of explanation methods will be examined by ensuring the robust transitions among them with the goal to facilitate the users in examining the explanations from different perspectives. An investigation of xPP for other use-cases such as explaining the reasons for algorithmic fairness violations is another future research direction.

References

[1] N. Mehdiyev, J. Evermann, and P. Fettke, "A multi-stage deep learning approach for business process event prediction," IEEE 19th Conference on Business Informatics, (CBI) 2017, vol. 1, pp. 119–128.

[2] N. Mehdiyev, J. Evermann, and P. Fettke, "A Novel Business Process Prediction Model Using a Deep Learning Method," Bus. Inf. Syst. Eng., pp. 1–15, 2018.

[3] N. Mehdiyev, J. Lahann, A. Emrich, P. Fettke, and P. Loos, "Time Series Classification using Deep Learning for Process Planning: A Case from Process Industry," in Procedia Computer Science, 2017, pp. 114, 242-249.

[4] N. Mehdiyev, A. Emrich, B. Stahmer, P. Fettke, and P. Loos, "iPRODICT - Intelligent process prediction based on big data analytics," in CEUR Workshop Proceedings (BPM Industry Track), 2017, vol. 1985.

[5] P. Fettke, L. Mayer, and N. Mehdiyev, "Big-Prozess-Analytik für Fertigungsmanagementsysteme (MES)," in Big Data: Anwendung und Nutzungspotenziale in der Produktion, 1st ed., M. Steven and T. Klünder, Eds. Stuttgart: Kohlhammer, 2020, pp. 215–239.

[6] N. Mehdiyev and P. Fettke, "Explainable Artificial Intelligence for Process Mining: A General Overview and Application of a Novel Local Explanation Approach for Predictive Process Monitoring," arXiv2009.02098, 2020. – accepted to the edited volume "Interpretable Artificial Intelligence: A perspective of Granular Computing" (published by Springer)

[7] N. Mehdiyev and P. Fettke, "Prescriptive Process Analytics with Deep Learning and Explainable Artificial Intelligence," in 28th European Conference on Information Systems (ECIS), 2020.

[8] N. Mehdiyev and P. Fettke, "Local Post-Hoc Explanations for Predictive Process Monitoring in Manufacturing," arXiv2009.10513, 2020

[9] J.-R. Rehse, N. Mehdiyev, and P. Fettke, "Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory," KI-Künstliche Intelligenz, pp. 1–7, 2019.

[10] P. Lübbecke, N. Mehdiyev, and P. Fettke, "Substitution of hazardous chemical substances using Deep Learning and t-SNE," in Wirtschaftinformatik (WI), 2019.