A Tool for Business Processes Diagnostics

Mahsa Pourbafrani(✉), Firas Gharbi, and Wil M. P. van der Aalst

Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany
{mahsa.bafrani,firas.gharbi,wvdaalst}@pads.rwth-aachen.de

Abstract. Recorded event data of processes inside organizations is a valuable source for providing insights and information using process mining. Most techniques analyze process executions at detailed levels, e.g., process instances, which may result in missing insights. Techniques at detailed levels using detailed event data should be complemented by techniques at aggregated levels. We designed and developed a standalone tool for diagnostics in event data of business processes based on both detailed and aggregated data and techniques. The data-driven framework first analyzes the event data of processes for possible compliance and performance problems, e.g., bottlenecks in processes. The results are used for aggregating the event data per window of time, i.e., extracting features in the time series format. The tool is able to uncover hidden insights in an explainable manner using time series analysis. The focus of the tool is to provide a data-driven business process analysis at different levels while reducing the dependencies on the user's domain knowledge for interpretation and feature engineering steps. The tool is applied to both real-world and synthetic event data.

Keywords: Process mining · Change point · Event logs · Time series

1 Introduction

A specific activity being performed for a process instance (case) at a specific point in time is an event. These stored events inside processes form event logs, i.e., the data source of data-driven process analysis. Tools and techniques to find such problems, find recurrent patterns, concept drifts, and anomalies, as well as predict the future state of processes, are of high interest. These diagnostics affect process performance and understanding of the processes for future improvement. Techniques such as Dotted Charts [8], visualize the detailed event data of processes and patterns such as the arrival rate of cases and potential concept drifts. Current tools and techniques mostly rely on the user for interpretation. Moreover, the patterns and insights at higher levels are ignored. In [1], it is discussed that the drift in the execution of processes can be captured while features are

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy-EXC-2023 Internet of Production-390621612. We also thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

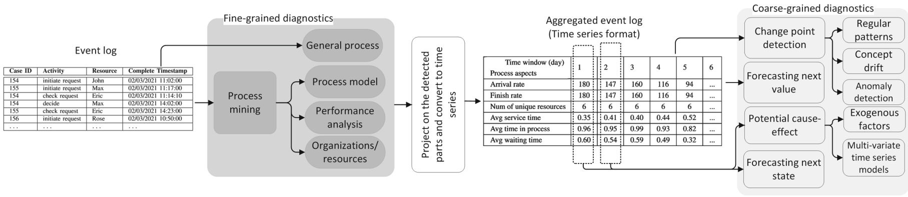

Fig. 1. The framework of the designed tool for data-driven process diagnostics.

defined and the data are transformed into different levels, e.g., calculating the daily arrival rate shows that the order of execution of activities in the process has changed (a concept drift occurred). However, such techniques are limited in feature extraction, dependent on specific diagnostics, and highly subject to the user for both forming features and interpreting the results.

In this paper, we focus on using the idea of systematic transformation of event data to time series data regardless of the process context in [5]. The authors use that for aggregated simulation. Time series and data science techniques can be applied to the aggregated event data. Automatic feature extraction per window of time out of event data and implemented techniques in a user-friendly interface enable the extended analysis of business processes. The data analysis techniques enable the revealing of the hidden patterns inside business processes, e.g., concept drifts, and anomalies. Furthermore, it enables prediction models based on time series models. The tool can be applied to any generic form of event data. The detailed techniques for the application of the time series data of processes' event data have been proposed in [6]. Our tool includes three modules, diagnostics insights at detailed levels, single perspective analysis (process variables), and multiple perspective analysis. The two last modules are based on the generated time series data from the detailed event data at a higher level, e.g., daily, or weekly. Figure 1 shows the framework of the developed tool, including the techniques and modules. A process variable, e.g., arrival rate per window of time, is a process aspect, and multiple process aspects form a process state at a point in time, e.g., the state of the process in a day.

2 Coarse-Grained Diagnostics Architecture and Features

Interesting parts based on process mining techniques, such as an activity with a bottleneck, are pointed out to the user, and the user is redirected to the time series generator modules. Then the generated aggregated event logs are fed into the single aspect analysis, and by selecting each of the aspects, the change points, patterns, and potential anomalies inside a process variable are detected. Change point detection [4], i.e., the discovery of points in a time when the values of the process variables are not consistent with the previous values, represents the behavior of the process w.r.t. that process aspect, e.g., an increase in daily arrival rate every Monday. Moreover, the relationship of each process variable with other variables is assessed. For instance, the increase in the number of cases waiting in the process per week causes an increase in the number of assigned resources. The last module enables uploading different generated time series data sets of processes to assess their relations with each other and the possibility of training time series models based on other process variables. The three main modules of the tool are as follows:

– *Diagnostics Insights* is for detailed performance analysis of activities, resources, and organization.
– *Single Perspective Analysis* is for detecting the change points in a single process aspect, i.e., patterns, concept drifts, and anomalies. The implemented techniques are PELT [3], Binary Segmentation, and Kolmogorov Smirnov [4]. The differences are in the search of the subsequences for the change points [6]. For relation detection, the Granger Causality test is implemented [2].
– *Multiple Perspective Analysis* is for the relations among different process aspects and whether they have causal effect relations are assessed in this module. This module employs multivariate time series analysis [7].

Technical Implementation. The tool is implemented in Python and uses the Flask library and is available as a web application. Each module is available for separate usage and integration with different platforms.

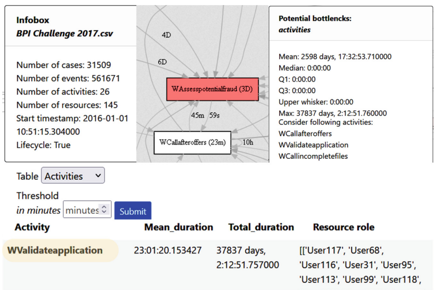

Fig. 2. A Screenshot of coarse-grained diagnostics tool using BPIC'17. (Color figure online)

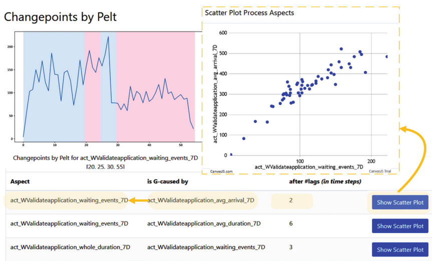

Fig. 3. A Screenshot of the aspect analysis results for the process aspect analysis using BPIC'17.

3 Application Domain and Scenarios in Practice

The source code and the event logs are publicly available.[1] We show the presented tabs and modules using screenshots of their output in practice. Figure 2 shows the diagnostic insights tab for the real-world event log, BPIC'17 [9]. The performance

[1] https://github.com/mbafrani/Coarse-grained-Process-Diagnostics.

analysis reveals a bottleneck in activity *W Validate Application* shown in red in Fig. 2. The process model, deviation, and performance information are provided as the results of the first step. The results are also provided for resources and organizations.

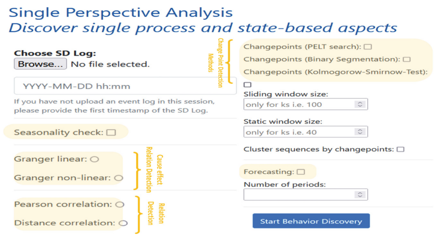

Fig. 4. A Screenshot of coarse-grained diagnostics tool for the process aspect analysis.

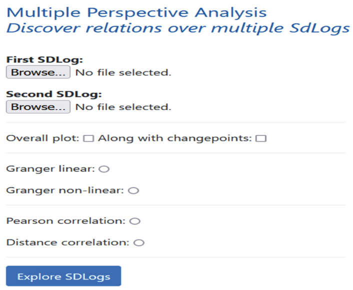

Fig. 5. A screenshot of the analyzing multiple aspects module in the coarse-grained analysis tool.

Using these detailed insights, we transform the event log into time series data sets for the detected activity over the steps of time to look for undiscovered insights. The parameters and different implemented techniques in single aspect analysis are presented in Fig. 4. The seasonality, linear and nonlinear Granger Causality, and different introduced change point detection techniques, can be assessed on the transformed data. For forecasting and training time series models for a single variable, the number of periods to predict is input. For instance, the cause and effect relation between the number of waiting cases and the average arrival rate in the detected activities in BPIC'17 has been detected with lag 2. The scatter plot of the relation, and the change points can be seen in Fig. 3. The same can be investigated for all the variables and their relations, among other variables. In Fig. 5, a screenshot for multiple perspective analysis is shown. This module performs the analysis among different aspects and process logs, such as using multivariate time series analysis [7]. Our tool is designed to support users in analyzing their processes with less dependency. The aim is to fill the gap by transforming event logs into a higher level of time granularity where the hidden insights at the detailed level are discovered.

References

1. Bose, R.P.J.C., van der Aalst, W.M.P., Žliobaitė, I., Pechenizkiy, M.: Handling concept drift in process mining. In: Mouratidis, H., Rolland, C. (eds.) Advanced Information Systems Engineering, pp. 391–405 (2011)
2. Granger, C.: Some recent development in a concept of causality. J. Econometrics **39**(1), 199–211 (1988)

3. Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. **107**(500), 1590–1598 (2012)
4. Martjushev, J., Bose, R.P.J.C., van der Aalst, W.M.P.: Change point detection and dealing with gradual and multi-order dynamics in process mining. In: Matulevičius, R., Dumas, M. (eds.) BIR 2015. LNBIP, vol. 229, pp. 161–178. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21915-8_11
5. Pourbafrani, M., van der Aalst, W.M.P.: PMSD: data-driven simulation using system dynamics and process mining. In: Proceedings of Demonstration & Resources Track at BPM 2020, vol. 2673, pp. 77–81. CEUR-WS.org (2020)
6. Pourbafrani, M., Gharbi, F., van der Aalst, W.M.P.: Process diagnostics at coarse-grained levels. In: Proceedings of the 24th International Conference on Enterprise Information Systems, ICEIS 2022, vol. 1, pp. 484–491 (2022)
7. Reinsel, G.C.: Elements of multivariate time series analysis. In: Elements of Multivariate Time Series Analysis (1995)
8. Song, M., van der Aalst, W.M.P.: Supporting process mining by showing events at a glance. In: WITS, pp. 139–145 (2007)
9. van Dongen, B.F.: BPIC 2017. Eindhoven University of Technology (2017)