Process Diagnostics at Coarse-grained Levels

Mahsa Pourbafrani[a], Firas Gharbi[b] and Wil M. P. van der Aalst[c]
Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany

Keywords: Process Mining, Event Logs, Time Series Analysis, Process Diagnostics, Performance Analysis, Coarse-Grained Process Logs, Concept Drift.

Abstract: Process mining enables the discovery of actionable insights from event data of organizations. Process analysis techniques typically focus on process executions at detailed, i.e., fine-grained levels, which might lead to missed insights. For instance, the relation between the waiting time of process instances and the current states of the process including resources workload is hidden at fine-grained level analysis. We propose an approach for coarse-grained diagnostics of processes while decreasing user dependency and ad hoc decisions compared to the current approaches. Our approach begins with the analysis of processes at fine-grained levels focusing on performance and compliance and proceeds with an automated translation of processes to the time series format, i.e., coarse-grained process logs. We exploit time series analysis techniques to uncover the underlying patterns and potential causes and effects in processes. The evaluation using real and synthetic event logs indicates the efficiency of our approach to discover overlooked insights at fine-grained levels.

1 INTRODUCTION

Event data in process mining are presented in the form of event logs, in which process executions for a process instance are captured, such as a sequence of manufacturing steps with respect to time for a product in a production line. Event logs are the primary means of performing performance and bottleneck analysis for enhancing/improving processes (van der Aalst, 2016). Current performance analyses techniques in process mining highly rely on event-based data, and they are mainly in a detailed manner (Leemans et al., 2014; Mannhardt et al., 2015). For instance, the effects of resources workload on their speed is not detectable using fine-grained event logs, i.e., detailed event logs, and it requires daily performance metrics.

The effectiveness of projecting processes over time is proven. Consider for example *Dotted Charts* (Song and van der Aalst, 2007) that represent the existing patterns at fine-grained levels inside the process over time. Therefore, aggregating processes w.r.t time is a practical approach for process diagnostics. For instance, the idea of using the sliding window for concept drift detection in process mining is proposed in (Bose et al., 2011). Time-related diagnostics such as dotted charts or concept drift detection as well as anomaly detections that use time series analysis are mainly dependent on the user for defining and extracting the aggregated process variables and analyzing the results.

In this paper, we introduce a generic framework that benefits from both fine- and coarse-grained process analyses. The fine-grained event logs (steps are events) are transformed into coarse-grained process logs, i.e., a collection of measurable aspects of a process over steps of time when steps are time windows. All the aspects are defined and extracted systematically on top of standard event logs, as explained in (Pourbafrani and van der Aalst, 2021). For instance, *arrival rate* is an aspect that can be measured daily, and different aspects at every time step are considered as a *process state*, e.g., arrival rate, waiting time, and the number of unique resources in one day. We define three categories of change points in process aspects and a module to indicate potential cause and effects among process aspects, e.g., detecting the effects of the number of engaged resources per week on the idle time of a single resource after three weeks. It should be noted that throughout the paper, we refer to event logs and diagnostics techniques utilizing the standard event logs in which the steps are events as fine-grained event logs, and diagnostics to stress the distinctions in the granularity of data and approaches.

[a] https://orcid.org/0000-0002-7883-1627
[b] https://orcid.org/0000-0001-6849-7939
[c] https://orcid.org/0000-0002-0955-6940

Pourbafrani, M., Gharbi, F. and van der Aalst, W.
Process Diagnostics at Coarse-grained Levels.
DOI: 10.5220/0011035000003179
In *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022) - Volume 1*, pages 484-491
ISBN: 978-989-758-569-2; ISSN: 2184-4992

The designed synthetic event log of an insurance company (INS) and two public event logs, i.e., BPI Challenge 2017 (BPIC'17) (van Dongen, B.F., 2017) and BPI Challenge 2015 (BPIC'15), are used for providing examples explaining the approach and performing the evaluation.

We present the related work in Section 2 and Section 3 covers the process mining and time series analysis concepts. We introduce our approach in Section 4 and evaluate it in Section 5. Section 6 concludes this work with challenges and future work.

2 RELATED WORK

Presenting processes over time will reveal process behavior, including compliance and performance problems. Time perspective is considered for diagnostics at various levels, fine- or coarse-grained levels. General time-related diagnostics such as (Hornix, 2007) are at aggregated levels, where they calculate a set of predefined KPIs, e.g., average waiting time in processes, for the whole process. Dotted charts are fine-grained process diagnostics techniques (Song and van der Aalst, 2007) that depend on the user to spot the insights over time. Time series analysis is used in existing process behavior analysis for a variety of objectives. In (Pourbafrani et al., 2020a), we use the time series models such as *ARIMA* for detecting the best window size to extract process variables for the purpose of simulation. Concept drift in processes, their type, and the use of time series for their detection are proposed in (Bose et al., 2011). The concept drift detection in (Bose et al., 2011) is based on using different periods of time inside processes.

Detecting anomalies in processes is the other purpose of utilizing time series as presented by (Bezerra et al., 2009). Moreover, in (Pourbafrani et al., 2020b), the relations between process aspects are discovered to form a simulation model which (Adams et al., 2021) used the same idea to detect the cause and effect relations among process variables. The purpose is to capture new insights using time series. Authors in (Yeshchenko et al., 2019) propose to employ the time series for concept drifts by applying the PELT algorithm. The results are clustered and are visually prepared for the user. We refer to (Sato et al., 2021) as a survey of concept drift detection in process mining. In the causal and relation detection between process variables, multiple researchers exploited time series analyses. In (Hompes et al., 2017), cause and effect relations between a business process characteristics and process performance are detected. Authors in (Adams et al., 2021) employ time series analysis to determine the potential cause and effects between process variables. However, similar to previous approaches, they are rather too much reliant on the user, or the variables are extracted in an ad hoc manner. The user domain knowledge is used to define the variable, which makes the approach process specific.

Fine- and coarse-grained analyses are required to detect process behavior. As a result, there is a gap in providing an integrated and general framework for defining and extracting process measurable aspects while also having a comprehensive approach for applying time series analysis to processes. By increasing the granularity of process event data, we can represent a process from various perspectives using its aspects over time, as presented in (Pourbafrani and van der Aalst, 2021) and implemented in (Pourbafrani and van der Aalst, 2020).

3 PRELIMINARIES

In this section, we define coarse-grained process logs and introduce time series concepts used in our approach.

Process Mining.

Definition 1 (Event Log). *An event $e=(c,a,r,t_s,t_c)$, where $c \in \mathcal{C}$ is the case identifier, $a \in \mathcal{A}$ is the activity in e, $r \in \mathcal{R}$ is the resource, $t_s \in \mathcal{T}$ is the start time, and $t_c \in \mathcal{T}$ is the complete time of the event e. $\xi=\mathcal{C} \times \mathcal{A} \times \mathcal{R} \times \mathcal{T} \times \mathcal{T}$ is the universe of events. We define projection functions for e as follows: $\pi_{\mathcal{C}} : \xi \rightarrow \mathcal{C}$, $\pi_{\mathcal{A}} : \xi \rightarrow \mathcal{A}$, $\pi_{\mathcal{R}} : \xi \rightarrow \mathcal{R}$, $\pi_{\mathcal{T}_S} : \xi \rightarrow \mathcal{T}$ and $\pi_{\mathcal{T}_C} : \xi \rightarrow \mathcal{T}$. Event log $L \subseteq \xi$ is a set of events in which events are unique.*

The start and complete timestamps of an event log $L \subseteq \xi$, are obtained using p_s and p_c, respectively. $p_s(L)=\min\limits_{e \in L} \pi_{\mathcal{T}_S}(e)$ and $p_c(L) = \max\limits_{e \in L} \pi_{\mathcal{T}_C}(e)$. A sequence of events w.r.t. timestamp with the same case identifier represents a process instance, i.e., a trace. In the event log of a production line, the first event $e=(c,a,r,t_s,t_c)$ is for the first item with $c=1$, the activity is $a=welding$ which was started at timestamp t_s=08:30:25 02.01.2021 by resource $r=employee1$ and was completed at timestamp t_c=10:02:47 02.01.2021.

Coarse-grained Process Logs. Coarse-grained process logs are the collections of measurable aspects of a process over a specific time window, e.g., Table 1 shows a sample coarse-grained log. The time window is δ=1 *day*. Each column describes the process in a time step (*process state*), e.g., 1 day, and each row

Table 1: A sample coarse-grained process log in which the time window is one day. It includes six process aspects that values are represented in the cells.

Process aspects \ Time window (day)	1	2	3	4	5	6
Arrival rate	180	147	160	116	94	...
Finish rate	180	147	160	116	94	...
Num of unique resources	6	6	6	6	6	...
Avg service time	0.35	0.41	0.40	0.44	0.52	...
Avg time in process	0.96	0.95	0.99	0.93	0.82	...
Avg waiting time	0.60	0.54	0.59	0.49	0.32	...

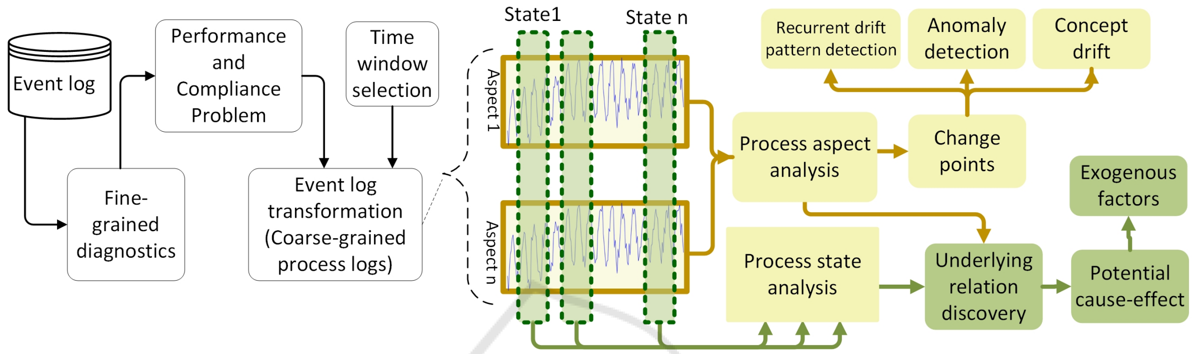

Figure 1: The proposed coarse-grained process diagnostic framework and the main modules.

is a *process aspect*. Cells are the values of process aspects, e.g., at the second time window, the value of process aspects *Arrival rate* is 147, which means in this time window, 147 cases arrived.

Definition 2 (Coarse-Grained Process Log). *Let $L \subseteq \xi$ be an event log, $\mathcal{V}$ be a set of process aspects, $\delta \in \mathbb{N}$ be the selected time window, and $k = \lceil \frac{(p_c(L) - p_s(L))}{\delta} \rceil$ be the number of time steps in the event log w.r.t. δ. The coarse-grained process log of given L and δ is $PL_{L,\delta}: \{1,\ldots,k\} \times \mathcal{V} \to \mathbb{R}_{\geq 0}$, such that $PL_{L,\delta}(i,v)$ represents the value of aspect $v \in \mathcal{V}$ in the i^{th}-time window ($1 \leq i \leq k$).*

We use PL throughout the paper instead of $PL_{L,\delta}$ when it is clear from the context. We also define $\Pi_v(PL) \in \mathbb{R}^*$ to return the sequence of values $\langle x_1,\ldots,x_k \rangle$ for aspect $v \in \mathcal{V}$. To access each value of the process aspect over time, we define π_i that returns the i^{th} value in a sequence, i.e., $\pi_i(\Pi_v(PL)) = x_i$.

Time Series. The analysis of sequences of real values and/or sequences of tuples of real values is often referred to as *time series analysis* (Hamilton, 1994). Pattern detection, next value prediction, or relation detection are the examples of time series analysis. Given a time series $\sigma = \langle x_1,\ldots,x_k \rangle \in \mathbb{R}^*$ with length of $k \in \mathbb{N}$, i.e., $|\sigma| = k$, $S \in \mathbb{R}^*$ is a subsequence of continuous values of σ with lengths $m \leq k$. For instance, $S = \langle x_i, x_{i+1}, \ldots, x_{i+m-1} \rangle$ is a subsequence of σ where $1 \leq i \leq k-m+1$.

Consider the given example in Table 1 as PL, where *Finish rate* is process aspect v. Applying function $\Pi_v(PL)$ results in a time series, i.e., $\Pi_v(PL) = \langle 180, 147, 160, 116, \ldots \rangle$ where $\langle 180, 147 \rangle$ is a subsequence of that.

4 APPROACH

Figure 1 presents the introduced coarse-grained process diagnostics approach. First, in Section 4.1, the existing process mining techniques are applied to event logs. Given the results, e.g., a specific activity or a set of resources, and the time window, coarse-grained process logs including process aspects and states are generated, e.g., times series presenting the process. Afterward, in Section 4.2, we explain the techniques designed for process diagnostics given the process logs. Considering multiple aspects in a similar window of time in relation to each other provides the process states analysis. Discovering the relationship between different aspects over time is the purpose of process state diagnostics.

4.1 Fine-grained Diagnostics

We defined and implemented the problem detection techniques w.r.t. deviation and performance, where it can be considered for activity flow, resources, and organizations. These problems are identified based on

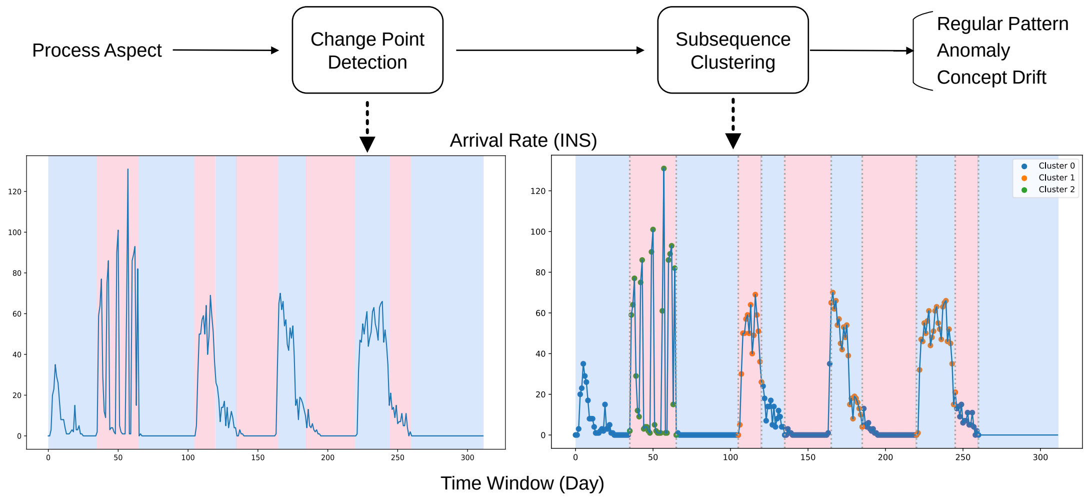

Figure 2: The overview of our process aspect analysis for the arrival rate of activity *Review request for rejection* in the INS event log.

performance and bottleneck analysis, deviation detection, or social network analyses. We form the problematic part of the process in the form of a set of activities A, resources R, or cases C. Note that organizations with performance and compliance issues can be shown as a set of resources. Consider that in event log L, bottleneck analysis as one of the fine-grained diagnostics returns activity a as a process bottleneck, i.e., long execution time. Therefore, $A = \{a\}$, $R = \{\pi_r(e) \in \mathcal{R} | e \in L\}$, and $C = \{\pi_C(e) \in \mathcal{R} | e \in L\}$. Another example can be the detected role, which is the set of resources performing activity a as a potential bottleneck in the process using social network analysis.

In the event log transformation step, in Fig. 1, event logs are projected using the provided sets of resources, activities and cases. For instance, for event log L, the projected event log is $L' = \{e \in L | \pi_{\mathcal{A}}(e){=}a \wedge \pi_{\mathcal{R}}(e) \in R \wedge \pi_C(e) \in C\}$. The coarse-grained process logs are created for the time window $\delta{=}1$ *day* by using $PL_{L',\delta}$ in Definition 2. The description of defining and extracting all possible aspects of the process over time is derived from (Pourbafrani and van der Aalst, 2021).

4.2 Diagnostics in a Process Aspect

By focusing on a single aspect of a process with event log L over time in the generated PL, e.g., $\Pi_{arrival\ rate}(PL)$, we are able to identify the existing change points. Given the fact that we are looking for the behavior of process aspects over time, there are three types of change points that we are interested in discovering: regular patterns (recurrent concept drifts), anomaly behavior, and concept drifts. To do so, we define the change point detection function in Definition 3, where the input is a process aspect and the output is change points in the values over time. In the implementation, we used *Pruned Exact Linear Time Method (PELT)* since it has a linear computational time. PELT detects change points by recursively splitting the time series into subsequences (Killick et al., 2012).

Definition 3 (Change Point Detection)**.** *Function $CPD : \mathbb{R}^* \to \mathbb{N}^*$ such that for $\sigma{=}\langle x_1, x_2, ..., x_k\rangle \in \mathbb{R}^*$ if $CPD(\sigma){=}\langle \tau_1, ..., \tau_m\rangle \in \mathbb{N}^*$ then for all $1 \leq j \leq m$, $1 < \tau_j < \tau_{j+1} < k$.*

Figure 2 depicts the general steps of the process aspect analysis module in our approach, as exemplified by the values of the daily arrival rate for activity *review request for rejection* in an insurance company's generated synthetic event log (INS). The activity is identified as a deviation in the fine-grained diagnostic module, i.e., a low fitness value. Function $CPD(\Pi_{arrival\ rate}(PL)){=}\langle 43, 65, ...\rangle$, returns 10 change points. Given the detected change points for process aspects, the corresponding subsequences are generated. For instance, in Fig. 2, $\Pi_{arrival\ rate}(PL)$ is transformed into 11 subsequences, e.g., $s_1 = \langle \pi_1(\Pi_{arrival\ rate}(PL)), ..., \pi_{43}(\Pi_{arrival\ rate}(PL))\rangle$, and $s_2 = \langle \pi_{44}(\Pi_{arrival\ rate}(PL)), ..., \pi_{65}(\Pi_{arrival\ rate}(PL))\rangle$. We refer to the set of corresponding subsequences of a process aspect as $S_{(CPD(\sigma))}$.

The ultimate goal in single process aspect analysis is to discover aspects' behavior over time. Having the subsequences of an aspect, we investigate the similarities of these subsequences to find whether there is recurrent concept drift, anomaly, or concept drift in the process. To do so, we cluster these subsequences using the defined function in Definition 4. To implement the module, our technique is based on (Niennattrakul and Ratanamahatana, 2007). Since our clustering is based on *k-means*, the silhouette metric (Shahapure and Nicholas, 2020) is used for automatic selection of k. For other potential clustering techniques of time series, we refer to (Warren Liao, 2005).

Definition 4 (Subsequence Clustering)**.** *Let* $\sigma = \langle x_1, x_2, \ldots, x_k \rangle \in \mathbb{R}^k$ *be a time series,* $CPD(\sigma)$ *be the detected change points, and* $S_{CPD(\sigma)}$ *be the set of the subsequences of* σ *w.r.t. change points. We define function* $CS(S_{CPD(\sigma)}) \subseteq 2^{S_{CPD(\sigma)}}$ *such that it returns the set of clusters including the subsequences inside.*

In Fig. 2, for the set of subsequence $S_{CPD(\Pi_{arrival\ rate}(PL))}$, the output of function subsequence clustering CS is presented. There are 3 clusters, where s_4, s_7, and s_9 belong to one cluster, and s_2 belongs to a second cluster. Afterward, the detected behaviors using the change points and clustering results are categorized such that:

- *Regular Pattern* (*Recurrent Concept Drift*): when a cluster has at least two subsequences that are not consecutive, there is a potential for regular patterns in the aspect. For cluster $C = \{s_1 = \langle x_i, ..x_j \rangle, s_2 = \langle x_p, \ldots, x_q \rangle\}$, and $p \neq j+1$, i.e., s_1 and s_2 are not consecutive. Note that if there is a period of time between every two subsequences in the cluster that does not match them, then, there is a regular pattern in the data. For instance, cluster 0 and 1 in Fig. 2 represent regular patterns that appear over time in the arrival rate.
- *Anomaly*: when a cluster has only one subsequence and the aspect has more than one subsequence, the time duration in the cluster is a potential anomaly. For cluster $C = \{s_1 = \langle x_i, ..x_j \rangle\}$, s_1 indicates an anomaly in the period of time steps i until j. For instance, cluster 2 in Fig. 2 is a detected period of an anomaly.
- *Concept Drift*: a cluster with consecutive sequences is a concept drift. For instance, $C = \{s_1 = \langle x_i, ..x_j \rangle, s_2 = \langle x_{j+1}, \ldots, x_p \rangle\}$, then, s_1 and s_2 are matches and directly following each other over time, therefore, x_i shows a concept drift at time step i. Note that an aspect with two subsequences, i.e., one change point, and two clusters is also considered as a potential concept drift.

4.3 Diagnostics in Process States

In addition to analyzing the behavior of a single aspect of a process, the relationships between process aspects provide an insight into the states of the process through time. Performing diagnostics for process states among the process aspects is able to reveal the potential effects of other aspects on the identified behavior of a single aspect. The extracted process states are in the form of *Multivariate Time Series*, therefore, techniques such as *Multivariate Time Series Analysis* and *Granger Causality* are applicable.

Analyzing whether the previous/current values of a variable plays a role in the future values of other variable is performed using *Granger Causality* (Granger, 1988). Considering both linear and non-linear Granger Causality, we define relation detection function as Definition 5. The output of this function only indicates whether there is a relation between the current/previous values of the first process aspect and the future values of the second process aspects.

Definition 5 (Cause and Effect Relation)**.** *Let* $\sigma_1, \sigma_2 \in \mathbb{R}^k$*, and* $k \in \mathbb{N}$*, function* $GC : \mathbb{R}^k \times \mathbb{R}^k \to \{\mathbb{R}^k \times \mathbb{R}^k \times \mathbb{N}, 0\}$. *For two time series, the function returns whether one of the time series causes on the other one with the shift in time, i.e.,* $GC(\sigma_1, \sigma_2) = (\sigma_2, \sigma_1, 2)$ *represents that the values of aspect* σ_2 *in each time window has effects on the next value of* σ_1 *in the next two time windows.*

For a detected bottleneck in *WValidateApplication* in BPI Challenge 2017 (L) with $\delta = 1$ *week*, process aspects including the number of cases in process per week (v_1), and the number of unique engaged resources (v_2) per week are extracted. $GC(\Pi_{v_1}(PL), \Pi_{v_2}(PL)) = (\Pi_{v_2}(PL), \Pi_{v_1}(PL), 1)$ indicates that the change in the number of cases in the process per week causes a drift in the number of engaged resources per week within 1 time step as presented in Fig. 3.

4.3.1 Exogenous Aspects

The developed cause and effect module determines the potential effects of aspects on the value of each other over time. We refer to the process aspects that are not affected by any of the aspects as exogenous aspects. Consider v=*arrival rate*, we identify it as an exogenous aspect if $\forall v_i \in \mathcal{V}$, $Corr(v_i, v) = 0 \wedge GC(v_i, v) = 0$, where $Corr$ is the correlation among two aspects, and GC is the cause and effect function. In processes, the exogenous factors are generally ignored. The influence of the environment and aspects that can (or cannot) be adjusted during the process is a determining factor in process improvement and predictive process monitoring.

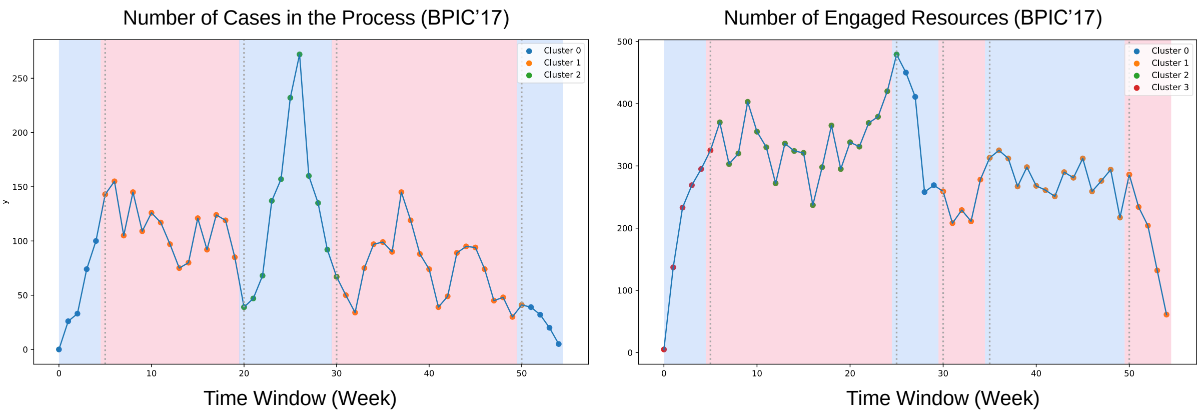

Figure 3: The change point in the number of engaged resources for activity *WValidateApplication* in BPIC'17 depends on the anomaly in the number of cases in process, i.e., after one week.

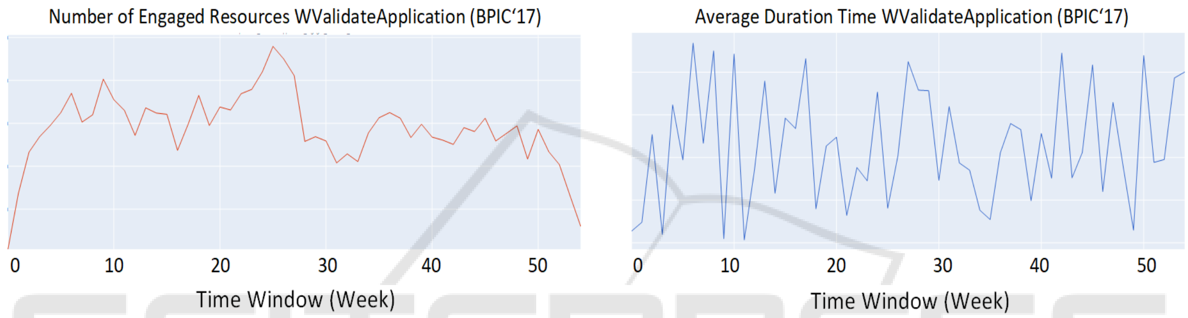

Figure 4: Two aspects for activity *WValidateApplication* as a detected bottleneck in BPIC'17.

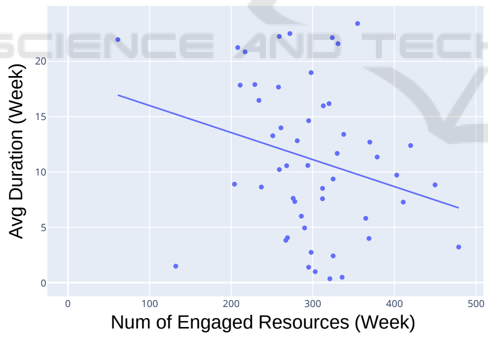

Figure 5: Revealed relations Fig. 4 using shifted time window of two aspects.

5 EVALUATION

To assess our approach, we used real and public event logs such as BPI Challenge 2017 (BPIC'17), BPI Challenge 2015 (BPIC'15), and a synthetic event log of an insurance company (INS) generated using the simulation tool in (Pouarbafrani et al., 2021). We begin by identifying and applying performance, compliance, and social network techniques to event logs. The results show that the approach is able to identify the patterns inside processes including the potential cause and effect relations among process aspects with time lags and without the user domain knowledge. Activity *WValidateApplication* in BPIC'17 is detected as a bottleneck, hence we extracted all the process aspects w.r.t. that activity. As shown in Fig. 4, no relation can be found at the same time window. However, using our approach, we find the potential cause and effect relations between the number of unique resources per week and the average duration of the activity with a shifted time window automatically, i.e., after 4 weeks, see, Fig. 5.

It should be noted that the size of the time window used to extract values of process aspects is critical in detecting process behavior. Given that we apply a time window selection module presented in (Pourbafrani et al., 2020a), the approach may determine the time window size in which the aspects have more stable behavior, such as stationary time series. However, this does not imply that other window sizes do not provide insights and information. For example, there could be a pattern in the process daily arrival rate as well as different patterns in the weekly arrival rate. As a result, these various insights resulting from the different time windows should be taken into account.

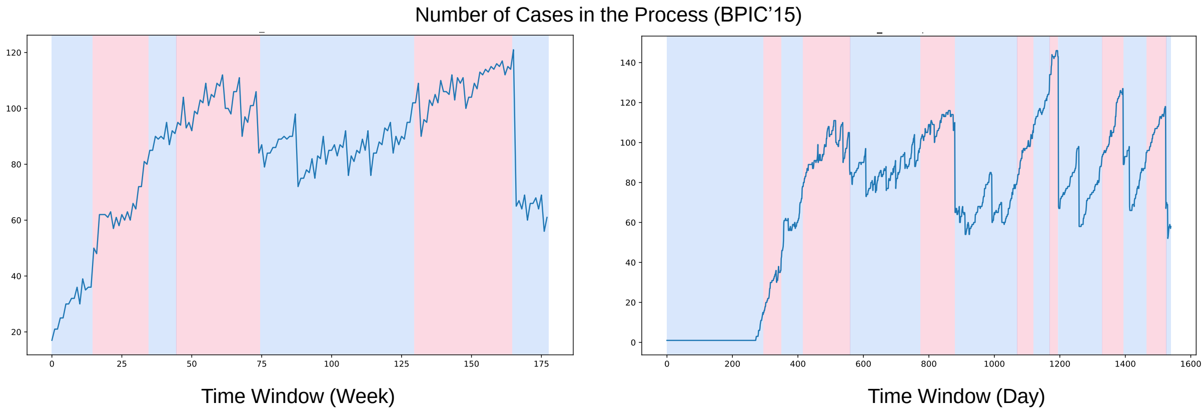

Figure 6: The number of cases in the process per day (right), and week (left) for BPIC'15. The detected patterns, change points and general behavior of event logs are affected by the size of time window to generate coarse-grained process logs.

We take the BPIC'15 event log as an example, where different patterns in weekly and daily values of process aspect the *number of cases in the process* can be seen. The different insights include different change points followed by different detected behavior, as presented in Fig. 6.

All the steps including the fine-grained process diagnostics, coarse-grained process logs generation, and coarse-grained analysis are integrated into one publicly available tool[1]. We also provided the used data sets, techniques, and evaluation results.

6 CONCLUSION

One of the primary goals of process mining is to provide process diagnostics that can be used to enhance processes. We designed a new and generic diagnostic approach in this paper. The approach unifies fine- and coarse-grained analysis by increasing the time granularity of event logs. Our approach overcomes the limitation of current techniques in extracting process variables from event logs, as well as selecting the time window for analysis, such as concept drift detection. We detect change points in process aspects and propose the potential causes. Process state analysis includes relation detection between process aspects across time. Distinguishing between the types of concept drifts, e.g., sudden drift or gradual drift, is one of the challenges and open problems for future work. Furthermore, the discovered insights in the proposed framework allow for future analysis, e.g., prediction and what-if analysis.

[1]https://github.com/mbafrani/PMSD

ACKNOWLEDGEMENTS

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.

REFERENCES

Adams, J. N., van Zelst, S. J., Quack, L., Hausmann, K., van der Aalst, W. M. P., and Rose, T. (2021). A framework for explainable concept drift detection in process mining. In *Business Process Management*, pages 400–416.

Bezerra, F., Wainer, J., and van der Aalst, W. M. P. (2009). Anomaly detection using process mining. volume 29 LNBIP of *Lecture Notes in Business Information Processing*, pages 149–161.

Bose, R. P. J. C., van der Aalst, W. M. P., Žliobaitė, I., and Pechenizkiy, M. (2011). Handling concept drift in process mining. In Mouratidis, H. and Rolland, C., editors, *Advanced Information Systems Engineering*, pages 391–405.

Granger, C. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, 39(1):199–211.

Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton New Jersey.

Hompes, B. F., Maaradji, A., La Rosa, M., Dumas, M., Buijs, J. C., and van der Aalst, W. M. P. (2017). Discovering causal factors explaining business process performance variation. In *CAISE*, pages 177–192. Springer.

Hornix, P. T. (2007). Performance analysis of business processes through process mining.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Leemans, S. J. J., Fahland, D., and van der Aalst, W. M. P. (2014). Process and deviation exploration with inductive visual miner. In *Proceedings of the BPM Demo Sessions 2014 Co-located with the 12th International Conference on Business Process Management (BPM 2014),*, page 46.

Mannhardt, F., de Leoni, M., and Reijers, H. A. (2015). The multi-perspective process explorer. In *Proceedings of the BPM Demo Session 2015*, pages 130–134.

Niennattrakul, V. and Ratanamahatana, C. (2007). On clustering multimedia time series data using k-means and dynamic time warping. pages 733–738.

Pouarbafrani, M., Balyan, S., Ahmed, M., Chugh, S., and van der Aalst, W. M. (2021). Gencpn:automatic cpn model generation of processes. In *3rd International Conference ICPM 2021, Proceedings (Demo Track)*.

Pourbafrani, M. and van der Aalst, W. M. P. (2020). PMSD: data-driven simulation using system dynamics and process mining. In *BPM Demo*, pages 77–81.

Pourbafrani, M. and van der Aalst, W. M. P. (2021). Extracting process features from event logs to learn coarse-grained simulation models. In *CAiSE*, volume 12751 of *Lecture Notes in Computer Science*, pages 125–140. Springer.

Pourbafrani, M., van Zelst, S. J., and van der Aalst, W. M. P. (2020a). Semi-automated time-granularity detection for data-driven simulation using process mining and system dynamics. In *ER*, pages 77–91.

Pourbafrani, M., van Zelst, S. J., and van der Aalst, W. M. P. (2020b). Supporting automatic system dynamics model generation for simulation in the context of process mining. In *BIS*, pages 249–263.

Sato, D. M. V., De Freitas, S. C., Barddal, J. P., and Scalabrin, E. E. (2021). A survey on concept drift in process mining. *ACM Computing Surveys (CSUR)*, 54(9):1–38.

Shahapure, K. R. and Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *DSAA*, pages 747–748. IEEE.

Song, M. and van der Aalst, W. M. P. (2007). Supporting process mining by showing events at a glance. In *WITS 2007*, pages 139–145.

van der Aalst, W. M. P. (2016). *Process Mining - Data Science in Action, Second Edition*. Springer.

van Dongen, B.F. (2017). BPIC 2017. Eindhoven University of Technology.

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857–1874.

Yeshchenko, A., Di Ciccio, C., Mendling, J., and Polyvyanyy, A. (2019). Comprehensive process drift detection with visual analytics. In *ER*, pages 119–135. Springer.