ProcessProfiler3D: A Tool for Visualising Performance Differences Between Process Cohorts and Process Instances

E. Poppe[1], M.T. Wynn[1], A.H.M. ter Hofstede[12], R. Brown[1], A. Pini[3], and W.M.P. van der Aalst[21]

[1] Queensland University of Technology, Queensland, Australia
[2] Eindhoven University of Technology, Eindhoven, The Netherlands
[3] DensityDesign Research Lab, Politecnico di Milano, Milan, Italy

Abstract. An organisation's event logs can give great insight into factors that affect the execution of their business processes by comparing different process cohorts. We have recently presented ProcessProfiler3D, a novel tool for such comparisons that supports interactive data exploration, automatic calculation of performance data and visual comparison of multiple cohorts. The approach enables the intuitive discovery of differences and trends in cohort performance. To better support the interpretation of these differences in the context of process execution we now extended the tool with a novel visualisation technique that enables the visualisation of case execution and timing in a way that provides context to such a performance analysis.

1 Introduction

Analysing process data in event logs to identify problems and opportunities with existing processes can be of great value for improving the processes of an organisation. Process mining [1], a specialised field of research in business process management, develops tools and techniques to support this. By splitting an event log into process cohorts, i.e. a *group of process instances that have one or more shared characteristics*, one can analyse how different case characteristics (often called context factors) affect the execution of a process. We have recently identified that despite continued industry interest [6,4,2], there is a lack of tools to support such analyses effectively [7]. None of the existing academic or commercial tools provided both support for interactive data exploration, by supporting interactive splitting of the event log, as well as an integrated comparison of multiple process cohorts, by supporting the visualisation of performance data for more than two cohorts in one view. Consequently, we presented ProcessProfiler3D, a framework to solve this issue [7]. We now present a complementary novel visualisation technique that covers additional performance analysis scenarios by providing additional context to the presented performance data.
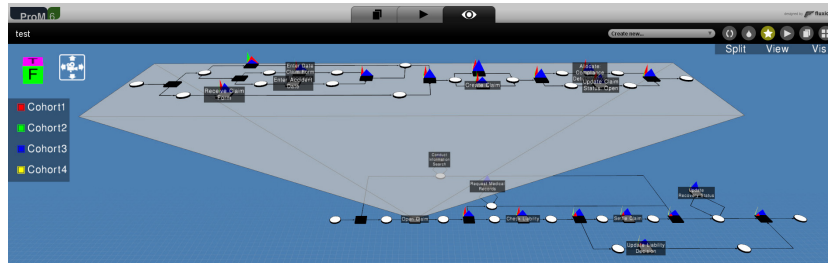
Fig. 1: Example of comparative process performance visualisation for four cohorts on a hierarchical process model (across two dimensions - time and frequency) using ProcessProfiler3D

2   Overview of the tool

ProcessProfiler3D enables comparing the performance of multiple process cohorts by

- aligning an event log with a process model
- calculating common node level process performance indicators such as activity duration, activity throughput time and waiting times between activities
- storing performance data in a data cube
- interactively splitting the event log by defining cohorts
- visualising performance data in a third dimension on top of the process model at multiple levels of process abstraction
- visualising data related to activities using either one of three different types of bar charts or a triangle chart (see [5])
- visualising data related to activity pairs can be visualised using coloured arcs between the two activities (see [7])

The framework was implemented in two plugins for the process mining framework ProM. Figure 1 shows an example of comparative performance analysis using this tool.

However, we note that some scenarios are still not well covered by existing performance analysis techniques and in the remainder of this paper we will discuss one of these scenarios and present a novel visualisation technique that we have added to ProcessProfiler3D to address this issue.

3   Problem statement

One issue with existing techniques for process performance analysis is the loss of context that occurs when performance data are a) localised and b) summarised as is usually the case with activity duration, throughput time and waiting time calculations. Both problems have the potential to affect our understanding of performance analysis results and can complicate finding root causes.

Firstly, the analysis results are currently localised to one point in the process model. For example, an activity C may be preceded by either activity A or B. By looking at performance indicators of these activities we cannot tell if cases that first executed A on average take longer to execute C than cases that executed B. So by localising the analysis results per activity we lose the context of how preceding activities affected the case and how subsequent activities were impacted.

Secondly, the statistical summary of performance indicators by minimum, median, mean and maximum also means that we are losing context in the results. It is, for example, hard to tell whether a few extreme cases skewed the results or what the general distribution of cases is. Furthermore, if the same case executes an activity multiple times, it is impossible to identify differences between the individual execution times (e.g. the activity took much time on the first execution, but finished really quickly on every following execution). Some absolute indicators, such as the average case runtime at an activity, also get distorted by loops. Consequently, while existing process performance analysis techniques already provide valuable insights into the execution of a process, additional analysis techniques are required to add context to the results of existing techniques.

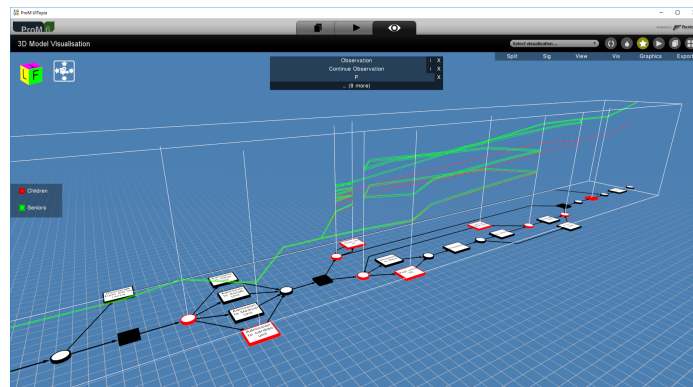## 4 Trajectory Visualisation

Fig. 2: Example of the proposed visualisation for process performance comparisons based on case trajectories (variant 3)

We therefore propose a novel visualisation technique inspired by geo-spatial data visualisations (e.g. [3]) to present performance data in the context of both history and future execution of a process instance. This visualisation presents the path of individual cases through a process model, while showing timing

information in a third dimension, orthogonal to the process model. An example of this technique can be seen in Figure 2.

We construct this visualisation by replaying a token-game on a given Petri net and recording each token move as a line in two dimensions. We then use the time of each event that triggered the token move to calculate the height of the start and end point of each line. Our implementation provides three different configurations of the trajectory visualisation. The first variant visualises token paths from one activity straight to the next activity. The second variant visualises the token paths from the activity through the place to the next activity. The third variant visualises the token path from the activity along the edge connecting it to the place and then along the edge to the next activity. Each variant increases the complexity of the visualisation, but often lines following the model layout more precisely make it easier to relate them back to the underlying process model and therefore easier to understand. To further facilitate this, vertical support lines can be displayed by selecting nodes in the process model, as shown in Figure 2.

In addition to the shape of case trajectories, colours can encode additional information in the visualisation. By default, case trajectories are coloured to indicate the cohort a case belongs to (see Figure 3). However, our implementation can also colour the trajectory to display relative completion of the case as a colour gradient. This can facilitate finding bottlenecks in large event logs. Furthermore, the cohort classification can be used to filter the visualisation, by hiding trajectories belonging to a particular cohort. Lastly, the vertical scale of the visualisation can be changed by clicking on the white frame surrounding the trajectories and pulling it upwards or downwards. This can make it easier to see differences between otherwise densely packed trajectories.

Seeing both control-flow and time perspective in one view enables users to identify interactions between control-flow constructs such as loops and process execution times. Using this technique together with the previously presented techniques for comparative performance visualisation (see [7]) therefore facilitates the understanding of performance analysis results.
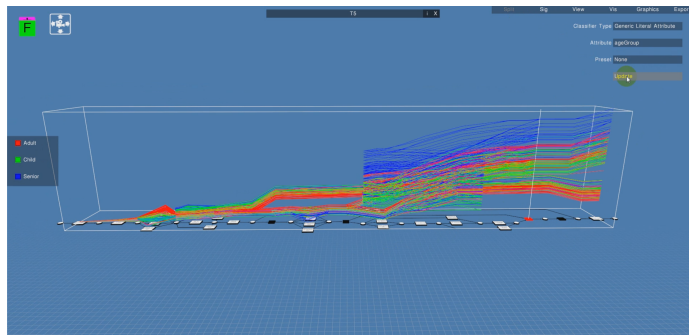

Fig. 3: Trajectory view of a complex event log with colour encoding the cohort a trajectory belongs to

5 Conclusion

We have presented ProcessProfiler3D, a framework that can be used to analyse and compare the performance of multiple process cohorts. The usefulness of this framework has previously been demonstrated by analysing two industry data sets and evaluating the tool with two industry partners [7]. In this paper we have added a novel visualisation technique, the trajectory visualisation, to this framework, to address the loss of context in the existing performance analysis approaches.

The framework is available as a package (called "ProcessProfiler3D") for the process mining framework ProM. In addition, the complete source code for the tool including the trajectory visualisation is available in the ProM repository[4].

A screencast of the tool including the new technique is available at: https://www.youtube.com/watch?v=CkgBTFk6MXY

6 References

References

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer (2016)
2. Bolt, A., de Leoni, M., van der Aalst, W.M.P., Gorissen, P.: Exploiting process cubes, analytic workflows and process mining for business process reporting: A case study in education. In: International Symposium on Data-driven Process Discovery and Analysis. pp. 33–47. CEUR-WS.org (2015)
3. Kraak, M.J.: The space-time cube revisited from a geovisualization perspective. In: Proc. 21st International Cartographic Conference. pp. 1988–1996 (2003)
4. Partington, A., Wynn, M., Suriadi, S., Ouyang, C., Karnon, J.: Process mining for clinical processes: A comparative analysis of four Australian hospitals. ACM Transactions on Management Information Systems 5(4), 19:1–19:18 (Jan 2015)
5. Pini, A., Brown, R., Wynn, M.T.: Process visualization techniques for multi-perspective process comparisons. In: Bae, J., Suriadi, S., Wen, L. (eds.) Asia Pacific Business Process Management. Lecture Notes in Business Information Processing, vol. 219, pp. 183–197. Springer, Busan, Korea (March 2015)
6. Suriadi, S., Wynn, M.T., Ouyang, C., ter Hofstede, A.H.M., van Dijk, N.J.: Understanding process behaviours in a large insurance company in Australia: A case study. In: Salinesi, C., Norrie, M.C., Pastor, O. (eds.) Advanced Information Systems Engineering, Lecture Notes in Computer Science, vol. 7908, pp. 449–464. Springer (2013)
7. Wynn, M.T., Poppe, E., Xu, J., ter Hofstede, A.H.M., Brown, R.A., Pini, A., van der Aalst, W.M.P.: ProcessProfiler3D: A visualisation framework for log-based process performance comparison. Decision Support Systems (2017, in press), `https://doi.org/10.1016/j.dss.2017.04.004`

---

[4] https://svn.win.tue.nl/repos/prom/Packages/ProcessProfiler3D/