Sequen-C Explorer for Process Mining

Luis Montana[1,*], Maria-Cruz Villa-Uriol[1,2]

[1]School of Computer Science, University of Sheffield, Sheffield, United Kingdom
[2]INSIGNEO Institute for in silico Medicine, University of Sheffield, Sheffield, United Kingdom

Abstract

We present Sequen-C for Process Mining, a visual analytics tool based on Sequen-C, that allows multilevel and detail-on-demand exploration of traces, and the inspection of attributes at process model, variant, trace or activity level. Users can interactively explore the data, adjusting the granularity at which process models are displayed (*multilevel overviews*). To obtain the process models, the tool comes with a hierarchical agglomerative clustering technique by default; however, users can provide their own clusterings. Process models are presented using a timeline-based visualization that can be interactively modified by users using a combination of alignment by activities and simplification strategies.

Keywords

process mining, temporal event sequences, event logs, visualization, visual process analytics

1. Introduction

Visual process analytics [1] has emerged as an opportunity to bring together the Visual Analytics (VA) and Process Mining (PM) communities. In this demonstration, we present Sequen-C for Process Mining, a tool originating from the VA domain, whose terminology has been adapted to the one used in process mining (PM) (see Table 1).

2. Main characteristics and Innovations

Sequen-C for process mining uses a timeline-based visualization approach to represent visual overviews of temporal event sequences. Users can interactively explore the data at multiple levels of detail, from coarser to finer, to manipulate the visualizations to easily reveal frequent and infrequent patterns in the data, and to obtain detail-on-demand at process model, variant, trace or activity level. The tool relies on the use of hierarchical clustering, alignment and other simplification approaches, together with classical strategies in VA such as interactivity and the use of coordinated views. For full technical details about Sequen-C and the techniques it uses, we refer the reader to Magallanes et al. [2] and an earlier work not covered in this demonstration [3]. Here, we provide an overview of the possibilities offered for process mining.

System's overview. Figure 1 shows the layout of Sequen-C for Process Mining, which includes seven panels. Panel A (General settings and controls) groups settings and controls used to perform alignment, sorting or zooming operations. Panel B (Activities) provides details on demand about the activities in the analyzed dataset, allowing their selection or deselection. Panel C (Process models) allows users to select alternative process models that offer different levels of analysis detail. Panel D (Variants) includes the variants for the selected process models or traces, together with their frequency of occurrence. Panel E (Traces) enables the exploration of the original traces for various user selections, e.g. a process model or a variant. Panel F (Attribute analysis) enables detailed analysis of attributes at the trace or activity level for a user-selected set of traces, variants, or process models. Finally, panel G

Proceedings of the Best BPM Dissertation Award, Doctoral Consortium, and Demonstrations & Resources Forum co-located with 23rd International Conference on Business Process Management (BPM 2025), Seville, Spain, August 31st to September 5th, 2025.
*Corresponding author.
l.montanagonzalez@sheffield.ac.uk (L. Montana); m.villa-uriol@sheffield.ac.uk (M. Villa-Uriol)
0000-0002-0511-0466 (L. Montana); 0000-0002-3345-539X (M. Villa-Uriol)

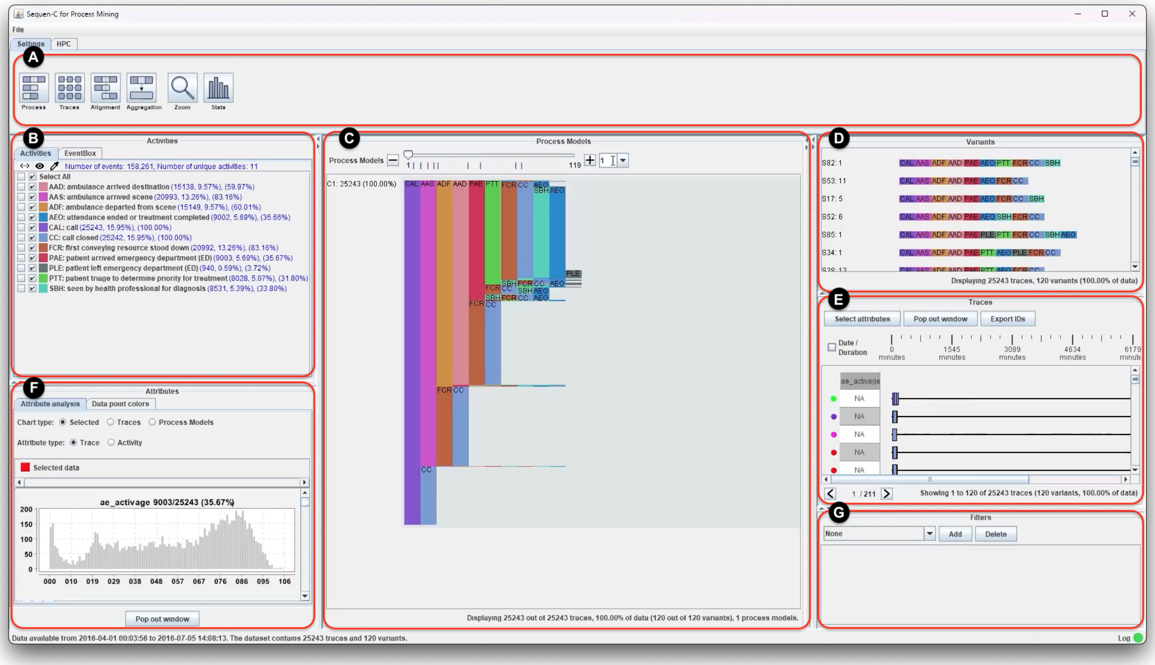


Figure 1: Sequen-C for Process Mining graphical user interface showing one process model.

(Filters) provides a guided filters creation feature to identify subsets of traces that match a user-defined query containing specific subsequences of activities or attribute values, or have a minimum number of occurrences. All panels are coordinated, meaning that user interactions and selections in one panel are always propagated to the other ones. All panels are resizable, allowing panels E and F to be opened in a separate window, which enables more detailed exploration while maintaining coordination with the other views.

Table 1
Terminology typically used in the VA and PM domains when analyzing event logs

Sequen-C (VA)	Sequen-C for process mining (PM)
Event type	Activity
Event occurrence or event	Event
Individual sequences	Traces
Unique sequences	Variants
Sequential patterns or Clusters	Process models

Visual encoding of process models. The process models in Panel C are represented using a timeline-based representation. Figure 2 shows how the set of traces in (a) is visually encoded as shown in (b). Figure 2(c) illustrates how the use of alignment by activities inserts artificial gaps that help gain a better understanding of the structure of the processes. Sequen-C for process mining does not include Directed-Follow-Graphs, but Figure 2(d) shows how the same model would be represented as a DFG.

Creation of multilevel overviews of process models. To build the multilevel overview of process models shown in Panel C, we use a bottom-up hierarchical aggregation [4] on all the traces in the dataset. The resulting *aggregate tree* enables the control of the granularity of the visualized process models (clusters in the aggregate tree), ranging from coarser to finer. By default, this panel displays the optimal number of process models, as determined by the average silhouette width metric [5]. Users can then explore alternative combinations of process models by changing k. For example, Figure 3(a) shows one process model representing all the traces in the dataset ($k = 1$), and compares it to Figure 3(c), which

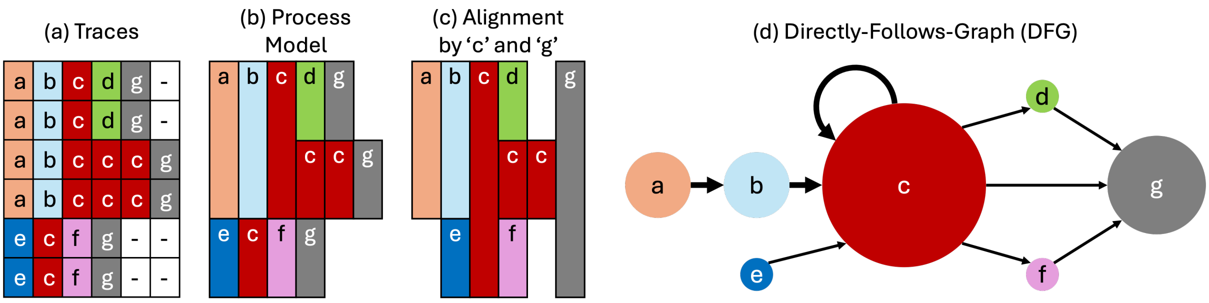
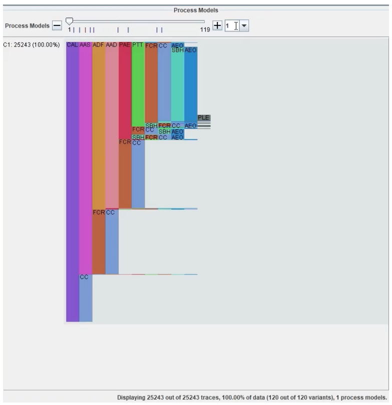
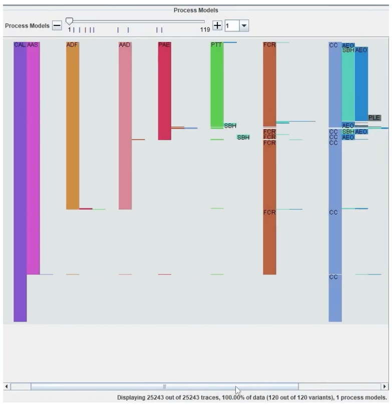


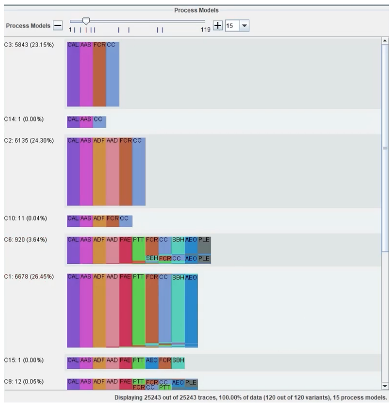
Figure 2: Timeline visualizations in (b,c) and the DFG in (d) represent the set of traces in (a).



(a) One process model



(b) One process model with manual alignment



(c) 15 process models

Figure 3: Timeline visualization of process models at multiple levels of detail with and without alignment by several activities.

shows 15 process models ($k = 15$). This shows that as the number of clusters (process models) increases, more detailed patterns within the process emerge, enabling the user to explore process variation at multiple levels of detail.

Alignment by activities and simplification. Sequen-C for process mining offers the option of using manual and automatic alignment by activities. Figure 3(b) illustrates an example where the activities in a process model have been manually aligned by a selection of activities, which helps in understanding the underlying structure of the processes captured and better understanding how the order of activities and frequency of traces relate. The automatic alignment option is available only if users use the default hierarchical clustering.

Implementation details. Sequen-C for process mining is implemented in Java and uses Python and R for various operations involving the manipulation of the original traces and hierarchical clustering.

3. Case studies and maturity

We have successfully used Sequen-C with a range of datasets including up to 150,000 traces. We include a description of some of them (Table 2) and the type of findings that have been obtained. Some of the results have already been published and, where possible, we include the relevant citation.

Scalability and maturity. Sequen-C for process mining suffers from classical scalability visualization problems when working with large volumes of data, including color availability, interactive experience degradation, among others. The tool has been used and evaluated by a diverse range of users, primarily with a background in data science. The tool has been well accepted and considered relevant to the PM community.

Table 2
Datasets

Dataset Name	Number of traces	Number of activities	Number of attributes
CUREd	25,243	11	57
Outpatient clinics	26,455	18	5
Road traffic fines	150,370	11	12

3.1. Case study 1: Calls to emergency services (CUREd)

The CUREd research database [6] contains timestamped events and demographic data related to telephone calls made to the emergency service (calls to 999 or 111), throughout Yorkshire and the Humber region, United Kingdom. Calls can lead to different pathways, including ambulance conveyance to the Emergency department (ED) and admissions to inpatient facilities. A three-month subset of the dataset was used, containing 25,243 calls relating to 21,805 unique patients, and 57 data attributes. Some of the findings obtained in this case study [2] included the identification of four process models that contain 85% of the data, showing that a remarkable 16.8% of phone calls to emergency services do not result in an ambulance dispatch. Interesting patterns emerged when exploring how attributes such as age and recorded symptoms at the time of the call influenced the process flows. For instance, one of the process models exclusively covered children, while another was more frequently observed among callers in their 80s. Additionally, it was observed that 59.7% of calls reporting chest pain (in one of the process models) resulted in the patient attending the emergency department.

3.2. Case study 2: Outpatient clinics

This case study [3] uses one year of real-world patient flow data from a Rheumatology outpatient clinic (Sheffield Teaching Hospitals - NHS Foundation Trust, Sheffield, United Kingdom). On average, the Department has an approximate annual workload of 9,000 patients and 25,000 appointments. Patient visits at this clinic are routinely tracked using an in-house workflow tracking system, where clinical staff (e.g., nurses, receptionists, consultants) input the current state of a patient according to the service being provided. The hospital uses the produced event logs to obtain basic statistics about the quality of care being delivered, particularly focusing on the study of waiting times and lengths of visits. Our analyses offered the possibility of delving into the raw event logs to extract key insights about patient flow within the clinic. These are being used to gain a better understanding of how the department operates and to suggest strategies for optimizing the delivery of care.

3.3. Case study 3: Road traffic fines

We have used the Road traffic fines dataset [7]. This log captures over a period of 13 years the management of road traffic fines by a local police force in Italy. It contains 150,370 cases and 561,480 events. During the study of this dataset, we realized the benefits that the current process models' visualizations offer when compared to Directed Follow Graphs (DFGs).

For example, we observed that 30.84% of the traces follow the process model *Create Fine – Payment*. Additionally, 37.64% of the traces conclude with the activity *Send for Credit Collection*, typically after the individual was notified and failed to make any payment. Another finding is that one of the process models reveals variants with multiple consecutive *Payment* events (up to seven in some cases).

4. Conclusion and Future work

We have presented Sequen-C for Process Mining, a tool adapted for process mining from a VA tool called Sequen-C [2]. We believe that the tool offers a range of features that would benefit process miners in their analyses. One is the use of a timeline-based visualization to represent process models.

They enable a quick understanding of how processes are structured, and used in combination with the visualization of variants, traces, and attributes, the analyses can be more detailed. The use of alignment operations, which automatically insert gaps to align by interesting activities, offers the possibility of better understanding order of activities, repetition of activities, and frequency. The use of filtering enables focussing on interesting traces, variants, or attributes. All this using coordinated views where selections in one panel are automatically propagated to the other panels to allow for more focussed analyses. Our tool does not include some commonly used PM representations, such as DFGs, but we include as future work the integration of DFGs as well as other types of visualizations to cover the multifaceted nature of the data used in PM [1, 8]. Another key area for improvement is support for additional data formats: at present, Sequen-C for PM only accepts CSV files with a predefined structure. Extending support to standard formats such as XES would enable wider adoption and easier integration. More details about Sequen-C for Process Mining are available at http://bit.ly/3IzN2ba.

Acknowledgments

We gratefully acknowledge the contribution of the Sheffield Teaching Hospital NHS Trust, and the NHS Trusts in the Yorkshire and the Humber region, which provided the original data to the CUREd Research Database. This report is independent research funded by the National Institute for Health and Care Research, Yorkshire and Humber Applied Research Collaborations NIHR200166. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health and Care Research or the Department of Health and Social Care. This study is also supported by the European Union's H2020 programme under grant agreements CompBioMed, CompBioMed2, and Sano (Nos. 675451, 823712, and 857533).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. van den Elzen, M. Jans, N. Martin, F. Pieters, C. Tominski, M.-C. Villa-Uriol, S. J. van Zelst, Towards multi-faceted visual process analytics, Information Systems 133 (2025) 102560.

[2] J. Magallanes, T. Stone, P. D. Morris, S. Mason, S. Wood, M.-C. Villa-Uriol, Sequen-C: A Multilevel Overview of Temporal Event Sequences, IEEE Trans. Vis. Comput. Graph. 28 (2022) 901 – 911.

[3] J. Magallanes, L. van Gemeren, S. Wood, M.-C. Villa-Uriol, Analyzing time attributes in temporal event sequences, in: 2019 IEEE Visualization Conference (VIS), IEEE, 2019, pp. 1–5. doi:10.1109/VISUAL.2019.8933770.

[4] C. C. Aggarwal, C. K. Reddy, Data clustering: algorithms and applications, CRC Press, 2014.

[5] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, volume 344, John Wiley & Sons, 2009.

[6] S. Mason, T. Stone, R. Jacques, J. Lewis, R. Simpson, M. Kuczawski, M. Franklin, Creating a real-world linked research platform for analyzing the urgent and emergency care system, Medical Decision Making 42 (2022) 999–1009. doi:10.1177/0272989X221098699, pMID: 35574663.

[7] M. de Leoni, F. Mannhardt, Road traffic fine management process, 2015. doi:10.4121/uuid:270fd440-1057-4fb9-89a9-b699b47990f5.

[8] A. Yeshchenko, J. Mendling, A survey of approaches for event sequence analysis and visualization, Information Systems 120 (2024) 102283.