Visual Process Mining: Event Data Exploration and Analysis

Peter Bodesinsky Bilal Alsallakh Theresia Gschwandtner Silvia Miksch *

ABSTRACT

The research field of Process Mining deals with the extraction of information from event logs. Since large amounts of event data are generated every day, analyzing these often unstructured event logs poses a research challenge. To cope with the complexity of the data and the associated mining tasks, appropriate visualizations and interactive means are needed. We present our early prototype for the visual exploration of event sequences and patterns. Our approach combines (1) a visual representation of event sequences emphasizing recurring event patterns, (2) automated pattern mining methods, as well as (3) interactive means for exploration. We provide first steps to support browsing event logs in an interactive environment and to facilitate the inspection of recurring pattern locations within the context of the surrounding events.

1 INTRODUCTION

Process Mining [4] aims to analyze event data, which is often generated automatically by information systems or sensors. The data can, therefore, be overwhelming in its amount as well as in its complexity. Data analysis techniques in Process Mining are categorized into three different types [4]: *Process Discovery* aims to generate an appropriate process model from given event data. *Process Conformance* compares a process model with cases. *Process Enhancement* makes use of additional data (attributes) of event logs, to enrich process models with information (e.g., task execution time), or to generate alternative views on the data. Figure 1 outlines the basic idea behind process mining.

Three challenges need to be solved to involve domain experts in the process and to make effective use of their domain knowledge: (1) The complex and manifold event data has to be made comprehensible for the user, (2) the actual process of mining the data has to be steered, and (3) the output has to be presented in an understandable manner. Visual process mining is the use of Visual Analytics (VA) methods in process mining to achieve these goals.

The first steps of an analysis process require visualization to get an overview and to allow to explore event data in an effective manner. Dotted charts [2] and EventTunnel [3] are examples for these kind of tools. Another common task is the identification of recurring patterns and subsequences among one or more process instances. In addition it is often necessary to perform grouping and clustering of cases (i.e. process instances) in order to obtain homogeneous sublogs. Bose and van der Aalst [1] present an approach for this task, which is based on biological sequence alignment. In contrast to other approaches, we show individual event sequences augmented with information computed by automated pattern analysis.

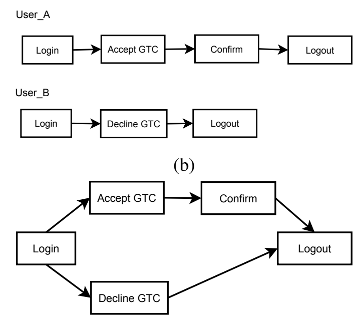
Interactive visual support is needed for all these issues, which is an open research challenge for the VA community. Our prototype aims to support several tasks: The initial visual exploration of the data. The identification irregularities in the data by browsing through the individual event sequences. The search for recurring patterns and the inspection of the pattern's location within an event sequence.

*All authors are with Vienna University of Technology, Vienna, Austria, E-Mail: {peter.bodesinsky, bilal.alsallakh, theresia.gschwandtner, silvia.miksch}@tuwien.ac.at

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

Session	Event	Time
User_A	Login	17:22:44
User_B	Login	17:23:10
User_B	Decline GTC	17:24:30
User_A	Accept GTC	17:24:40
User_B	Logout	17:24:50
User_A	Confirm	17:24:52
User_A	Logout	17:25:07

(a)



(c)

Figure 1: Simplified example of the concept behind process mining. A process is defined in an event log (a). A case (i.e. sequence of events) is the data of one user in this example. Process tasks (i.e. events) are login, accept GTC (general terms and conditions), confirm, decline GTC, and logout. With the information in the log, event sequences (b) and a control-flow/process model (c) can be constructed, which describes the behaviour. Real event logs contain noisy, unstructured data and much more cases and events.

2 APPROACH

Our research prototype is meant to help the user to determine the characteristics of an event log in advance. This aids in selecting specific mining methods that are suited to the log. The mining and visualization of recurrent patterns aims to give a first clue about the structural properties of the cases and similarities among cases.

Visual Encoding In our prototype, cases (i.e. process instances) are shown horizontally as a sequence of bars, each bar represents an event (see Figure 2). The color of the rectangle encodes the event type. Cases are arranged vertically beneath each other. The prototype supports different sorting modes for the vertical arrangement of the cases: the default order (sorting by execution time), sorting by the count of pattern occurrence and sorting by case length. Recurring patterns inside a case, i.e. a recurring sequence of events, are connected by semi-circular arcs, similar to the arc diagram visualization, proposed by Wattenberg [5]. To save space, arcs which exceed a certain height are compactly represented with flattened tops. If a pattern occurs only once within a case, it is marked by a rectangle. This visual encoding aims to convey that the pattern does not recur. Optionally it is also possible to draw connections between different cases. The first occurrence of the pattern in two neighbouring cases is highlighted by a connecting line (not shown in Figure 2).

By using this kind of visual encoding, we provide a visual flow and guidance for the users eye. This way, we support the inspection of the pattern's location within a case in combination with contextual information of surrounding events. Moreover, the visualization also highlights the frequency of a given pattern within the whole data set.

Pattern Detection A pattern of interest can be defined either interactively by selecting the sequence of events within the visual-

Figure 2: Design prototype. The screenshot shows a part of the example event log. (a) The visualization of event sequences together with a recurring event pattern. (b) List of mined and manually entered event patterns. (c) Pattern frequency shown by horizontal bars. (d) Patterns of a defined length and support value can be mined automatically.

ization or by entering the pattern directly manually as a regular expression. All previously defined patterns are stored in a list and can be selected for display. We show the pattern frequencies using a bar chart. Our approach currently supports the display of one pattern at a time to avoid visual clutter. However, we consider to show multiple patterns simultaneously in the future, for example by coloring the arcs accordingly. In addition, simple sequential event patterns of a defined length n can be computed automatically. In this way the user is allowed to inspect cases and pattern occurrences of the whole event log visually, supported by automatic pattern search and mining during iterative exploration.

Figure 2 shows parts of an example event log, which contains the recorded requests of a webservice connected to a webshop. Each user session (i.e. one user browsing the shop) represents a case. The type of the event can either be a request for related webshop items (r), a view request for an item (v), or a buy request (b). Inspecting event sequences of multiple cases and identifying common sequential patterns and their occurrences gives insights into user behaviour or might help to debug a webservice. In this example the user has searched for all patterns of length two with a support above 60 percent. The pattern *vr*, i.e. item view request (green) followed by a request for related items (red) is selected for visualization. It is visible that this pattern appears multiple times for some user sessions, but not for all, furthermore they rather appear in the beginning or middle and buy requests happen in the end (blue).

The approach currently has some limitations we will adress in the future. The limited number of well-distinguishable colors limits the number of types of events that can be represented simultaneously. Another interesting question is how to display multiple patterns simultaneously without cluttering the visualization. Appropriate scaling and aggregation mechanisms are needed to be found for logs containing thousands or millions of cases (and events). We plan to provide an initial statistical overview of the data set with means to select groups of cases of interest for further investigation with the arc diagrams. Furthermore we consider a pixel based overview navigation and a visual representation of patterns sequences.

3 Conclusion and Future Work

Our research approach supports the exploration and analysis of event log data. We pursue an approach based on VA and user-centered design. Following these ideas, we gathered requiremets of our prototype and designed an interactive visualization augmented with means for automatic pattern mining. Selected cases of an event log can be inspected individually, while at the same time it is possible to manually browse the whole event log, in order to gain an overview of existing patterns. Our approach supports the search for event sequence patterns and exploring the results visually. In the future we plan to integrate a time scale to provide more information on individual event execution time. We also consider clustering for arrangement of similar cases and the application of pattern mining algorithms, which can cope with noisy data. In addition, we aim to mine and display additional attributes of the events.

Acknowledgement This work was supported by Centre for Visual Analytics Science and Technology CVAST (funded by Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative, project number: 822746)).

References

[1] R. Bose and W. van der Aalst. Trace alignment in process mining: Opportunities for process diagnostics. In *Proceedings of the 8th International Conference on Business Process Management*, BPM'10, pages 227–242, Berlin, Heidelberg, 2010. Springer-Verlag.

[2] M. Song and W. van der Aalst. Supporting process mining by showing events at a glance. In *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)*, pages 139–145, 2007.

[3] M. Suntinger, H. Obweger, J. Schiefer, and M. Gröller. Event tunnel: Exploring event-driven business processes. *Computer Graphics and Applications, IEEE*, 28(5):46–55, 2008.

[4] W. van der Aalst et al. Process mining manifesto. In F. Daniel, K. Barkaoui, and S. Dustdar, editors, *Business Process Management Workshops*, volume 99 of *Lecture Notes in Business Information Processing*, pages 169–194. Springer Berlin Heidelberg, 2012.

[5] M. Wattenberg. Arc diagrams: Visualizing structure in strings. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 110–116. IEEE, 2002.