An Interactive Framework to Facilitate Behavioural Pattern Exploration in Event Data (Extended Abstract)

Leen Jooken
Faculty of Business Economics
UHasselt - Hasselt University
Hasselt, Belgium
leen.jooken@uhasselt.be

***Index Terms*—event data behavioural analytics, pattern mining, visual analytics, exploratory data analysis**

I. Problem Statement and Positioning with Regard to the State of the Art

A lot of data that is collected by systems can be categorized or treated as event data. Event data is data that describes events in which the state of the subject to which the event relates, changes. This subject can be a person, an object or a process. Data must meet three important characteristics in order to be categorized as event data: (1) an event happens instantaneously, therefore we do not consider begin or end timestamps, nor the duration, (2) it is possible to (partially) order events in time, (3) an event describes a change of state or context. The availability of this type of data has witnessed an increase because of two characteristics of this digital age: (1) the capturing has become easier, and (2) it has become easier to deal with huge amounts of data as a result of cheaper data storage opportunities, improved database technology and the availability of big data technology.

Event data can provide a different type of insights than traditional 'rectangular' data, because of how the information is stored. Traditional data typically describes some type of business object by means of a set of attributes [1], which means that it only provides a snapshot of the state of that business object. It cannot provide insights into the exact behaviour that led to that state. The ability to learn behavioural patterns is what makes event data so interesting. After all, understanding, predicting and correctly reacting to changes in behaviour are crucial business capabilities. Furthermore, including the behaviour perspective to complement traditional analysis can help you build a multi-dimensional viewpoint to better solve certain business problems [1].

Behaviour, however, is an abstract concept that consists of many attributes and properties, for instance: the subject that carries out the action, the object on which a behaviour is imposed, the context in which behaviour manifests itself, the goal of the behaviour and the impact on the object or context [1]. Events, on the other hand, are the raw observation of behaviour and very hard to interpret in their raw form. The fact that this data is so low-level and fine-grained is what makes analyzing it so challenging. In order to analyze the behaviour, the low-level event data first needs to be transformed into higher-level behavioural patterns.

This is where our research problem arises: because this data is more complex than conventional transactional data, traditional data analysis techniques are not always appropriate to extract these insights. There is a wide variety of pattern mining techniques available, and although some could be used on (preprocessed) event data, they were not developed with event data in mind and often rely upon a set of assumptions which are not universal for event data, for example: process mining [2] expects each event to be related to an instance of the process. However, exploring event data is not only challenging because this data is generally interrelated in a multidimensional manner, but also because of the large amounts of data, which creates challenges for current techniques both in terms of feasibility and in terms of interpretability of the results. When the data consists of a lot of observations and many different types of items, these techniques tend to produce a long list of possibly interesting patterns, which is difficult to act on by a user who wants to explore the data. Furthermore, insights can only be learned from patterns that are meaningful with respect to the practitioner's use case. Hence it is crucial that understanding of behavioural structures, semantics and dynamics of the event data is incorporated in the pattern discovery and analysis stage [1].

So we identified a set of needs related to exploring behavioural patterns in event data. Firstly there is a need for clarity on which techniques can be used, and how they can be used, to mine for behavioural patterns in event data. Secondly there is a need to effectively present the found patterns in a way that is intuitively comprehensible to the end user. Lastly there is a need to incorporate domain knowledge in an interactive way while exploring the data to ensure the quality, correctness and relevance of the found patterns with respect to the practitioner's use case.

II. Proposed Solution

The final envisioned solution is an exploration framework to guide the practitioner in interactively exploring behavioural patterns in their event data. The framework must meet the following requirements: (1) it should take raw event data as an input, (2) it should uncover insights in behaviour of people, objects or systems, (3) it should enable interactive exploration, and (4) it must be possible to incorporate domain knowledge.

The development of this framework is an application of the behaviour informatics approach [1]. The scientific field of Behaviour Informatics focuses on the development of methodologies, techniques and practical tools for representing, modeling, analyzing, understanding and utilizing behaviour [1].

The PhD Project is divided into three work packages. Each work package meets one of the needs we identified in Sect. I, with the following deliverables:

1)
   a) a conceptual translation from the field of behavioral informatics [1] to a set of analytical challenges that describe different types of behavioural insights that can be learned
   b) an overview on how different pattern mining techniques, or underlying concepts of these techniques, can address these challenges and be used to answer related research questions
2) a framework that incorporates the most useful concepts from these pattern mining algorithms to facilitate the exploration of event data from a behavioural analytics perspective and visualizes the found patterns in a way that is intuitively comprehensible to the end user
3) a way in which the end user can interact with the framework and the feedback is reentered into the iterative analysis process

III. Research Methodology

A. Work Package 1

Since behaviour is an abstract concept that consists of many low-level aspects, there are many viewpoints on behaviour that can be investigated [1], and many different insights that can be mined. For this first work package we aim to map out the different domains of behaviour topics of interest and make a classification of which pattern mining techniques can provide answers to questions from each domain. We will explore pattern mining techniques from different domains to get an overview of the kind of behavioural insights that each technique can discover. Different techniques from the following domains, among others, will be examined: sequential pattern mining [3], [4], sequential rule mining [4], itemset mining [3], [4], episode mining [3], [4], periodic pattern mining [3], [4], high-utility pattern mining [4], association rule mining [3], [4], graph pattern mining [3], [4] and process mining [2].

B. Work Package 2

Using the results of work package 1 we aim to build a framework that incorporates underlying concepts of the previously mentioned pattern mining techniques to facilitate the easy exploration of behavioural patterns in the practitioner's event data. This artifact will be designed and implemented following a Design Science approach [5], to safeguard that it will meet the predetermined requirements and the actual needs of the initial problem. Next, available visualization techniques need to be explored and put to the test, in order to present the resulting patterns in a intuitively comprehensible way to the end user.

C. Work Package 3

The focus of this last work package lies on incorporating expert knowledge into the analysis process. We plan to do this in an interactive way based on a visual analytics approach [6]. Visual analytics is described as the "science of analytical reasoning facilitated by interactive visual interfaces" [7]. At its base it is an iterative process in which automatic discovery and visualization gets refined by feedback from the end user [6]–[8]. The interactivity of visual analytics makes it extremely well suited to be used for exploratory analysis, and places it in the broader area of hybrid intelligence, aimed at creating synergies by letting machines and humans work together [9].

IV. Result Validation

Validation of the results of the different work packages is a great challenge. To validate our framework's ability to find meaningful patterns and its user experience, one or more case studies will be set up in which a practitioner can evaluate the performance of the framework on their data. However, it is important to keep in mind that this evaluation is subjective and highly influenced by the end result the practitioner already has in mind. A strategy needs to be worked out to minimize this influence and to evaluate not only patterns that are expected to be found but also patterns that are unusual or overlooked. A combination of a structured interview of the practitioner's expectations beforehand, the capturing of the framework's user experience and an in-dept interview afterwards, seems to be the best approach.

References

[1] L. Cao, "In-depth Behavior Understanding and Use: The Behavior Informatics Approach," *Information Sciences*, vol. 180, no. 17, Sep 2010.

[2] W. Van der Aalst, *Process Mining: Data Science in Action*, 2nd ed. Springer-Verlag Berlin Heidelberg, 2016, 978-3-662-49851-4.

[3] C. C. Aggarwal and J. Han, Eds., *Frequent Pattern Mining*, 1st ed. Springer International Publishing, 2014, 978-3-319-07821-2.

[4] P. Fournier-Viger, C.-W. Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, "A Survey of Sequential Pattern Mining," *Data Science and Pattern Recognition*, vol. 1, no. 1, pp. 54–77, 2017.

[5] P. Johannesson and E. Perjons, *An Introduction to Design Science*, 1st ed. Springer International Publishing, 2014, 978-3-319-10632-8.

[6] D. Keim, G. Andrienko, J. D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual Analytics: Definition, Process, and Challenges*. Springer Berlin Heidelberg, 2008, pp. 154–175.

[7] J. J. Thomas and K. A. Cook, Eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005, 978-0-769-52323-1.

[8] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim, "Knowledge Generation Model for Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1604–1613, 2014.

[9] E. Kamar, "Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, March 2016.