Pattern-based progressive analytics on interaction traces

Vincent Raveneau, Julien Blanchard, and Yannick Prié

Abstract—Due to the ever increasing use of data analysis tools, their performance has become an important scientific question, notably regarding the tools' ability to assist the analyst using them. We present our first proposals toward building a data analysis tool allowing an analyst to interact with his data mining algorithm, while working with interaction traces. First, we present an architecture for such a system, based on the ideas of progressive analytics. We then present the challenges that will need to be faced in future work.

Index Terms—Interaction traces, pattern mining, progressive analytics, user behavior

1 Introduction

Visual analytics originate from the willingness to integrate the analyst in the knowledge discovery process, rather than just using an algorithm to extract it from the data [3]. According to this vision, work in the domain has mainly been focused on how to visually explore data rather than on the algorithm used in the process. In the meantime, work in the data mining community has been focused on the corresponding algorithms without considering the person executing the algorithm as an important element of the process.

Our work is about integrating data mining elements into the visual analytics approach, allowing the analyst to interact both with the visualization and the data mining algorithms. The importance of merging visual analytics and pattern mining as already been expressed [2, 6], with the expected benefit of obtaining tools that benefit from both visual analytics' consideration of the user and pattern mining's fast and efficient algorithms.

The application domain considered in our work is the study and discovery of user behavior from interaction traces. By interacting with digital tools, everyone builds up a set of behaviors and practices. These behaviors are specific to each user, and continue to evolve with time. In order to produce tools capable of assisting their users in the best possible way, it is necessary to have a better understanding of how human-computer interaction and appropriation of digital tools work, as exposed by Belin and al. [1].

Interaction traces are a sequence of events recorded while a user was interacting with a digital system. These events are obtained through probes integrated into the digital tools, and are associated with a time stamp. The nature of these events and the additional information they contain may vary.

Many aspects of the search for behaviors in these traces make it a difficult problem. Firstly, the exploratory nature of the process implies that the analyst will bring his skills, goals, curiosity, intuition and beliefs, thus adding a strong subjective aspect to the task. Secondly, the events are usually logged at a far too precise level to directly infer user behaviors from them.

In this poster we present our first proposals towards producing tools allowing an analyst to interact with pattern mining algorithms in order to discover user behaviors in interaction traces. We first propose a global architecture for such tools, then identify the challenges it implies on the visualization, algorithmic and interaction levels.

- *Vincent Raveneau is with University of Nantes, France. E-mail: vincent.raveneau@univ-nantes.fr.*
- *Julien Blanchard is with University of Nantes, France. E-mail: julien.blanchard@univ-nantes.fr.*
- *Yannick Prié is with University of Nantes, France. E-mail: yannick.prie@univ-nantes.fr.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Relevant work

2.1 Using interactive pattern mining

Pattern mining is known for producing understandable knowledge, since patterns can be seen as data parts directly expressed in the same terminology than data [6]. As it aims to be exhaustive, the pattern mining process has the advantage to discover patterns the analyst wouldn't have thought could exist. In his paper, Van Leeuwen proposes a list of four challenges inherent to interactive data mining: being able to learn and model user- and task-specific interestingness of patterns, being able to mine subjectively interesting pattern-based models virtually instantly, having principled evaluation methods for pattern mining algorithms and being able to visualize and interact with the discovered patterns.

The need for interaction between visualization and data mining is also expressed by Bertini and Lalanne [2], as they highlight the fact that very few work really using both has been done. They offer some suggestions for data mining and visualization, if one would like to more easily involve them both in a system.

2.2 Progressive analysis of the data

To get a truly interactive trace analysis tool, the analyst must be able to constantly interact with his data. Thus, it is necessary that the algorithmic analysis of the data (such as pattern mining) has the ability to produce results at any moment. This way, the analyst always has elements to work on without having to wait. Depending on his findings in some results, he must be able to steer the remaining analysis process without always restarting from scratch. Based on these considerations, Stolper and al. [5] proposed the notion of *progressive analytics*. Their objective is to design an alternative to the usual batch analysis of the data, where the analyst has to wait for the analysis to be completed before working on its results. Instead, they favor an analysis constantly providing partial results the analyst can work on. Their proposal being recent, the available tools to design such systems are very few, such as Jean-Daniel Fekete's ProgressiVis [4], a framework for the implementation of progressive analytics workflows.

3 An architecture proposal

We aim at providing tools to be used by an analyst in his work on data. While using pattern mining algorithm, they are meant to comply with the vision of both visual analytics and progressive analytics. The underlying architecture we propose is illustrated in Fig. 1. From the interaction traces, a progressive pattern mining algorithm outputs patterns and statistical measures about them (such as their number of occurrences, occurrence durations or temporal distributions). Both the patterns and the measures are then stored inside a pattern manager, while the algorithm keeps iterating and producing partial results in the background. Patterns and statistical measures are represented in a visualization, alongside the events from the interaction traces. By exploring this visualization, an analyst is able to request additional patterns to continue his work. Such request is handled by the pattern search engine of the pattern manager, that will answer it in one or both of the following ways : by retrieving and sending already discovered

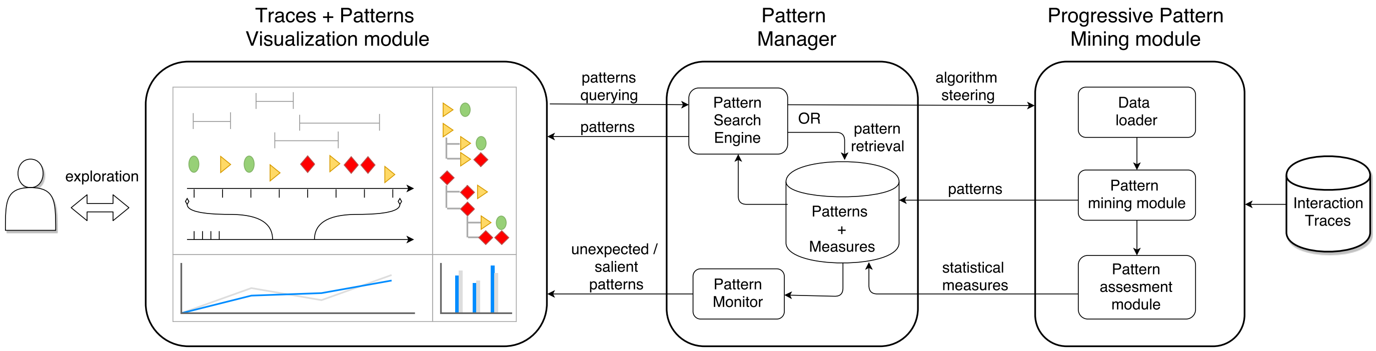

Fig. 1. Our proposed architecture

patterns relevant to the query, or by steering the pattern mining algorithm's execution in order to prioritize subparts of the data relevant to the query. Besides, some patterns are pushed in the visualization on the initiative of the pattern manager, thanks to a module which monitors the generated pattern stream. The module detects the unexpected or salient patterns on the basis of their statistical characteristics and of the user's exploration behavior.

4 Challenges

From our proposal, we identify some challenges that will have to be addressed in future work.

Visualization challenges

- **Designing a visualization that enables to see both the raw data and the patterns.** This will allow the analyst to capture the context in which the discovered patterns occur, while giving the opportunity to navigate through various observation levels.

- **Being able to dynamically integrate new patterns in the visualization as soon as they are discovered, without cluttering.** Constantly updating the visualization the analyst is working on will hamper more than help his work. We suggest that new patterns are added to a queue, only to be integrated into the visualization when the analyst deems it interesting and not a nuisance.

Algorithmic challenges

- **Designing a progressive algorithm for pattern mining.** The algorithm will have to produce partial results (i.e. sets of patterns) quickly enough, which raises the question of the order in which the patterns are mined. We suggest that this order depends on the parameters specified by the analyst during the exploration on the one hand, and on interestingness measures used to assess patterns on the second hand.

- **Giving the analyst the possibility to module the order in which the patterns are mined**. This should be addressed by providing the analyst to specify a strategy lying between local search (to focus on a certain pattern template) and random search (to discover unexpected patterns and improve serendipity).

Interaction challenges

- **Offering meaningful ways to steer the algorithm in a subpart of the exploration space by specifying constraints.** Steering the algorithm can either be done implicitly, while the analyst explores the visualization, or explicitly through a menu. The implicit approach would bring a more seamless knowledge discovery process, but starting with the explicit approach will allow us to more easily work towards efficient steering techniques.

- **Allowing the understanding of parameter change's consequences.** The analyst needs to understand if the changes he made to the algorithm's parameters had the intended consequences. Depending on how he can make these changes, information about these consequences could be displayed before the change (for example, an estimation of how many patterns will be found), or the patterns discovered with the new parameters could be distinguished in the visualization from those who where found before.

5 Future Work

Now that we have a broad view of what we want our system to be able to do, we will soon propose a first implementation of it. If the place of the visualization and of the pattern mining algorithm are rather straightforward, we will need to make choices regarding the pattern manager. The advantages and disadvantages of it being either a separate component or integrated with one of the others will be considered, the pattern manager being a key component of our architecture. For a first prototype, we wish to experiment with timeline-based visualizations to represent our data. On the algorithmic side, we will start by adapting existing pattern mining algorithms to the progressive analytics paradigm (similarly to what Stolper did with the SPAM algorithm [5]). Providing additional and efficient ways of steering these algorithms' execution will be considered when we will have a working system to experiment on.

References

[1] A. Belin, Y. Prié, and A. Tabard. Supporting the Development of Digital Skills. In *Digital Intelligence*. Nantes, France, Sept. 2014.

[2] E. Bertini and D. Lalanne. Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery. *ACM SIGKDD Explorations Newsletter*, 11(2):9–18, 2010.

[3] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2005.

[4] J.-D. Fekete. ProgressiVis: a Toolkit for Steerable Progressive Analytics and Visualization. In *1st Workshop on Data Systems for Interactive Analysis*, p. 5. Chicago, United States, Oct. 2015.

[5] C. Stolper, A. Perer, and D. Gotz. Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1653–1662, 2014. doi: 10.1109/TVCG.2014.2346574

[6] M. van Leeuwen. Interactive data exploration using pattern mining. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, pp. 169–182. Springer, 2014.