

Vladimir Dzyuba

Department Computer Science
PhD defence 09 June 2017
Supervisor Prof. dr. Luc De Raedt
Co-supervisor Dr. Matthijs van Leeuwen
Funding FWO
E-mail dzyuba.v@gmail.com



Mine, Interact, Learn, Repeat: Interactive pattern-based data exploration

Objective

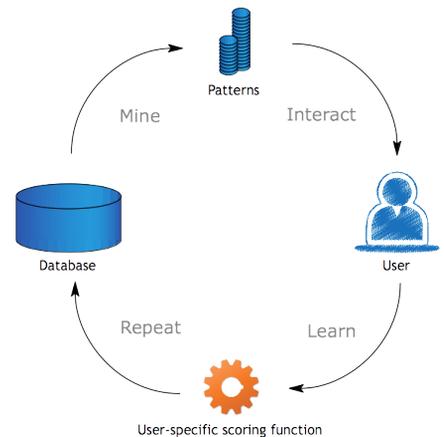
The rapid growth of the amount of available data creates the need for automated tools to assist analysts in discovering useful knowledge in these data. Unfortunately, many advanced approaches are “black boxes” that require considerable expertise to fine-tune. Our primary research goal is to develop data analysis tools that allow its users to provide feedback with respect to intermediate output, which is used to tailor the results to their *subjective* interests. This makes advanced data analysis tools accessible to non-experts, who do not need to understand how these tools work “under the hood”.

Research Methodology

We focus on *pattern mining*, a well-studied exploratory data analysis technique. Its output is a collection of *patterns*, i.e., interpretable descriptions of recurring regularities in the given dataset. The key challenge is to use the user’s feedback to identify genuinely interesting patterns among many possible regularities.

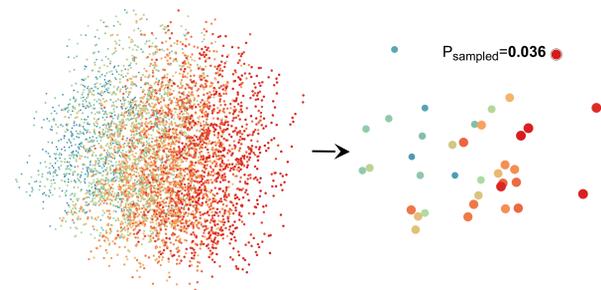
We frame this problem as learning a user-specific pattern scoring function, which assigns higher scores to subjectively more interesting patterns, by means of a four-step interactive mining and learning loop:

- **Mine:** Search for patterns with high scores
- **Interact:** Show few patterns & let user order them by interestingness
- **Learn:** Use the feedback to update the scoring function
- **Repeat:** Mine & interact again relying on the updated score

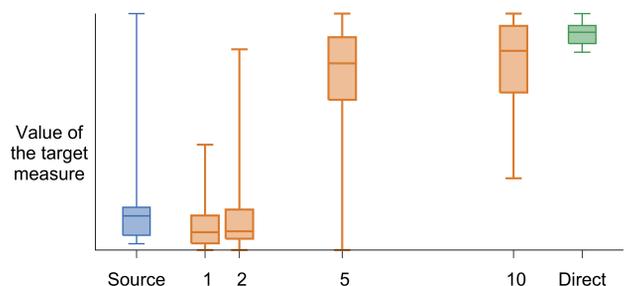


Results & Conclusions

Pattern sets shown to the user should be compact, diverse—so that they can reveal different aspects of user interests—and contain interesting patterns. To build such sets in an anytime manner, we developed a flexible pattern sampling algorithm based on insights from *SAT sampling*.



We learn subjective scoring functions with *preference learning* techniques. In experiments we emulate user feedback using complex *objective* scoring functions. The learned functions allow discovering novel high-scoring patterns. This only requires few loop iterations.



Major publications

- V. Dzyuba, M. van Leeuwen, L. De Raedt (2017). *Flexible constrained sampling with guarantees for pattern mining*. Data Mining and Knowledge Discovery, in press, DOI: 10.1007/s10618-017-0501-6.
- V. Dzyuba, M. van Leeuwen, S. Nijssen, L. De Raedt (2014). *Interactive learning of pattern rankings*. International Journal on Artificial Intelligence Tools, 23 (6), art.nr. 1460026.