Introduction to the Special Issue on Visual Analytics and Knowledge Discovery

Kai Puolamäki
Department of Media Technology
Helsinki Institute for Information Technology HIIT
Helsinki University of Technology (TKK)
P.O.Box 5400, FI-02015 TKK, Finland
kai.puolamaki@tkk.fi

Alessio Bertone
Department of Information and Knowledge Engineering
Danube University Krems
Dr.-Karl-Dorrek-Strasse 30, AT-3500 Krems, Austria
alessio.bertone@donau-uni.ac.at

ABSTRACT

The papers in this Special Issue present the state of the art in Visual Analytics and Knowledge Discovery and Data Mining (KDD), as well as propose potential extensions and research questions to further advance and integrate these two fields.

1. BACKGROUND AND MOTIVATION

Visual Analytics is the science of analytical reasoning supported by interactive visual interfaces [3]. It is a relatively new multidisciplinary field that attempts to combine various research areas including knowledge discovery and data mining (KDD), visualization, human-computer interaction, data management, spatiotemporal data processing, and statistics.[1]

The focus of the Visual Analytics community has been on interactive visual representation and exploration of data. The aim of the KDD community has, on the other hand, been in developing computational methods that can be used to extract knowledge from data. The paths of these two communities have not really crossed, even though the methods and goals of the two are complementary. Our argument is that there exist lots of problems for which it would be advantageous to combine these two approaches, and that these two communities should co-operate.

Last summer, we organized a full-day Workshop on Visual Analytics and Knowledge Discovery (VAKD '09)[2] in conjunction with the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. This Special Issue was targeted especially for the revised versions of the high quality papers presented in the workshop, even though the call for papers was open for all.

There were two invited talks in the workshop, given by Jim Thomas and Daniel Keim, who are also co-authors of the invited paper in this Special Issue. A total of five papers were presented in oral and four in poster sessions [2]. The workshop concluded with a working session where the results of the workshop were summarized and the KDD re-

---

[1]See the papers by Keim et al. and Bertini et al. in this Special Issue for discussion about the scope and definition of Visual Analytics.
[2]http://www.hiit.fi/vakd09/

search plan in visual analytics was discussed. There were about 50 participants and the workshop received a total of 19 submissions.

The goal of the workshop was to raise the awareness of the KDD community for the importance of Visual Analytics and to bring together researchers from the underlying fields in order to identify existing research gaps and overlaps—with the goal of producing a KDD research roadmap on Visual Analytics.

It is clear from the discussions in the workshop and elsewhere that there is a need for the integration of Visual Analytics and KDD. The efficient exploration and analysis of large and heterogeneous data masses requires a combination of automated KDD methods and visual interaction. Many standard data analysis environments however lack even the most fundamental interactive visualization tools, such as brushing. The use and interpretation of the results of the advanced KDD methods often require time and expertize; many of the present KDD methods are as such not applicable to fast, intuitive and interactive use. There is a need for *visually controllable* KDD methods, that is, algorithms that are (i) fast enough for visual interaction and whose model structure can be (ii) represented visually and (iii) controlled using visual interaction.

There is also the difficult question of evaluation. A typical KDD contribution can be evaluated, for example, by classification accuracies on some benchmark data sets. The question of how to judge a visual analytics contribution is still open for discussion, because there may be no obvious and easily measurable objective function. One approach is the one taken by the IEEE VAST 2008 Challenge [1], used by some of the papers in this Special Issue, where the performance criterion include—in addition to the accuracy—process description, analysis, and visualizations.

We expect and hope that this workshop and this Special Issue was not a last of this topic, but a beginning of a new field of research in the KDD community. The best is yet to come.

2. PAPERS OF THE SPECIAL ISSUE

The papers in the special issue represent the current state of the art of the research in this field. There is one invited paper by Keim et al., and five regular papers which were selected through a peer-review process.

The invited paper by *Keim, Mansmann, and Thomas* outlines the scope of Visual Analytics. The advantages and disadvantages of automated and visual analysis methods are explained. The authors argue that Visual Analytics methods that combine automated and visual approaches will enable us to go beyond what is possible when separately using the two methods.

The thought-provoking paper by *Bertini and Lalanne* uses bibliographic research to categorize the techniques, trends, gaps and potential future directions for research. The article analyzes strengths and weaknesses of information visualization and KDD approaches, and proposes potential extensions and research questions to further advance and integrate these fields.

Andrienko and Andrienko propose an approach to exploratory analysis of diverse types of spatiotemporal data with the use of clustering and interactive visual displays. The authors demonstrate their approach using tasks and datasets from a visual analytics challenge.

The paper by *Johansson and Johansson* suggests a solution to deal with mixed data sets, that is, data sets which combine categorical and numerical values. They propose a tool to perform exploratory analysis using common visual representations in multiple coordinated views and algorithmic analysis that facilitates detection of potentially interesting patterns within combinations of categorical and numerical variables.

Leung and Carmichael present a visual analytic tool for frequent pattern mining. Their tool is equipped with several interactive features for effective visual support in the data analysis and KDD process for various real-life applications.

Finally, the paper by *Piringer, Buchetis, Hauser, and Gröller* presents an approach called Hierarchical Difference Scatterplots to analyze data cubes. Their approach allows to relate multiple hierarchy levels and explicitly visualize differences between them in the context of the absolute position of pivoted values. The authors propose the integration in a setup of multiple views, which are linked by interactive queries on the data.

3. ACKNOWLEDGEMENTS

4. REFERENCES

[1] G. Grinstein, C. Plaisant, S. Laskowski, T. O'Connell, J. Scholtz, and M. Whiting. VAST 2008 challenge: Introducing mini-challenges. In *IEEE Symposium on Visual Analytics Science and Technology (VAST '08)*, pages 195–196, 2008.

[2] K. Puolamäki, editor. *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery*, 2009.

[3] J. Thomas and K. Cook, editors. *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE Press, 2005.

---

[3]http://www.vismaster.eu/
[4]http://www.pascal-network.org/