Human-Computer Interaction & Knowledge Discovery from Big Data: Introduction to the Special Session

Andreas Holzinger[1]& Andreas Auinger[2]

[1]Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation (IMI)
Research Unit Human-Computer Interaction
`a.holzinger@hci4all.at`

[2]University of Applied Sciences Upper Austria, A-4400 Steyr, Austria
School of Management, Research Unit Digital Business
`andreas.auinger@fh-steyr.at`

Abstract. We are confronted with increased masses of data, across all domains (e.g. in biomedicine, health care, digital business, social media etc.) especially on the Web. Research in Human-Computer Interaction (HCI) and Knowledge Discovery in Databases and Data Mining (KDD), has long been working to develop methods that help end users to identify, extract, visualize and understand useful information from these huge masses of high dimensional and often weakly structured and/or non-standardized data. Our goal is to combine those efforts to support end users to interactively analyze information properties and to visualize the most relevant parts without getting overwhelmed. Ideally we speak of HI-CI – Human intelligence (HI) "meets" Computational intelligence (CI) by making computational methodologies and approaches interactively accessible to the domain expert by supporting them to solve complex problems in the real world. The challenge is to enable effective human control over powerful machine algorithms and to integrate statistical methods and visual analytics, to support insight, discoveries and decision making.

1 Introduction and Motivation

Data exploration is called the *fourth paradigm* in the investigation of nature, after empiricism, theory and computation [1], [2]. Big data has always been an issue in the computational sciences [3], and is still of increasing interest [4]. The increasing data deluge [5] requires sophisticated methods of handling, e.g., researchers in bioinformatics are confronted with masses of complex data sets [6],[7], [8]. Whereas HCI concentrates on human intelligence, and KDD concentrates on machine

intelligence, the grand challenge is to combine these diverse fields to support the expert end users in learning to interactively analyze information properties thus enabling them to visualize the relevant parts of their data. For example, the broad application of business enterprise hospital information systems amasses large amounts of medical data, which must be reviewed, observed, and analyzed by human experts [9], or big data from the Web, which is highly relevant for digital business, e.g. opinion mining [10].

We are working towards enabling effective human control over powerful machine intelligence and on the integration of statistical methods and visual analytics, to support human insight and decision making [11] following the notion "Science is testing ideas – Engineering is to put these ideas into practice [12]". Additionally, mobile, ubiquitous computing, sensors everywhere, computational power and storage at extremely low cost will accelerate this avalanche of data [13] and there will be definitely the danger of drowning in data, but starving for knowledge. Herbert Simon pointed out 40 years ago, when medical informatics and digital business was in its infancy: "A wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it" [14].

Whether in the life sciences or in digital business, as a matter of fact the future will be data-centric, and we believe that the cross-disciplinary combination of both KDD and HCI will help to gain useful knowledge from these masses of data, to gain added values for the end users – to make data useable and useful [7], [15], [16].

2   The Special Session Program

From a total of 14 papers submitted to this special thematic session, 8 were carefully selected after a rigorous peer review, each paper assigning a minimum of three international reviewers. The papers will be presented in the following order, 4 papers in each part of the session (presenters in bold type):

Patrice Clemente, Bangaly Kaba, Jonathan Rouzaud-Cornabas, Marc Alexandre, Guillaume Aujay : *SPTrack: Visual Analysis of Information Flows Within SELinux Policies and Attack Logs*

Shuyan Xie, Markus Helfert, Lukasz Ostrowski: *Human Involvement in Designing an Information Quality Assessment Technique – Demonstrated in a Healthcare Setting*

Andreas Holzinger, Christof Stocker, Manuel Bruschi, Andreas Auinger, Hugo Silva, Hugo Gamboa, Ana Fred: *On Applying Approximate Entropy to ECG Signals for Knowledge Discovery on the Example of Big Sensor Data*

Martin Billinger, Clemens Brunner, **Reinhold Scherer**, Andreas Holzinger, and Gernot R. Müller-Putz: *Towards a framework based on single trial connectivity for enhancing knowledge discovery in BCI*

Gerald Petz, **Michał Karpowicz**, Harald Fürschuß, Andreas Auinger, Stephan M. Winkler, Susanne Schaller, Andreas Holzinger: *On Text Preprocessing for Opinion Mining outside of Laboratory Environments*

Rüdiger Heimgärtner, Harald Kindermann: *Revealing Cultural Influences in Human Computer Interaction by Analyzing Big Data in Interactions*

Michael Fire, Gilad Katz, Yuval Elovici, Bracha Shapira, and Lior Rokach: *Predicting Student Exam's Scores by Analyzing Social Network Data*

André Calero Valdez, Anne Kathrin Schaar, Martina Ziefle, Andreas Holzinger, Sabina Jeschke, Christian Brecher: *Using Mixed Node Publication Network Graphs for Analyzing Success in Interdisciplinary Teams*

3 Acknowledgements

References

1. Bell, G., Hey, T., Szalay, A.: Beyond the data deluge. Science 323, 1297-1298 (2009)
2. Buxton, B., Hayward, V., Pearson, I., Kärkkäinen, L., Greiner, H., Dyson, E., Ito, J., Chung, A., Kelly, K., Schillace, S.: Big data: the next Google. Interview by Duncan Graham-Rowe. Nature 455, 8 (2008)
3. Wing, J.M.: Computational thinking. Communications of the ACM 49, 33-35 (2006)
4. Mervis, J.: Agencies Rally to Tackle Big Data. Science 336, 22 (2012)
5. Szalay, A., Gray, J.: 2020 Computing: Science in an exponential world. Nature 440, 413-414 (2006)
6. Hey, T., Tansley, S., Tolle, K.: The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond (WA) (2009)
7. Holzinger, A.: Biomedical Informatics: Computational Sciences meets Life Sciences. BoD, Norderstedt (2012)
8. Akil, H., Martone, M.E., Van Essen, D.C.: Challenges and opportunities in mining neuroscience data. Science 331, 708-712 (2011)
9. Holzinger, A., Geierhofer, R., Modritscher, F., Tatzl, R.: Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses. J. Univers. Comput. Sci. 14, 3781-3795 (2008)
10. Auinger, A., Fischer, M.: Mining consumers' opinions on the Web. In: FH Science Day, pp. 410-419. (Year)
11. Holzinger, A.: Interacting with Information: Challenges in Human-Computer Interaction and Information Retrieval (HCI-IR). IADIS Multiconference on Computer Science and Information Systems (MCCSIS), Interfaces and Human-Computer Interaction, pp. 13-17. IADIS, Rome (2011)
12. Holzinger, A.: Successful Management of Research and Development. BoD, Norderstedt (2011)
13. Butler, D.: 2020 computing: Everything, everywhere. Nature 440, 402-405 (2006)
14. Simon, H.A.: Designing Organizations for an Information-Rich World. In: Greenberger, M. (ed.) Computers, Communication, and the Public Interest, pp. 37-72. The Johns Hopkins Press, Baltimore (MD) (1971)
15. Auinger, A., Aistleithner, A., Kindermann, H., Holzinger, A.: Conformity with User Expectations on the Web: Are There Cultural Differences for Design Principles? In: Marcus, A. (ed.) Design, User Experience, and Usability. Theory, Methods, Tools and Practice. Lecture Notes in Computer Science LNCS 6769, vol. 6769, pp. 3-12. Springer, Heidelberg, Berlin, New York (2011)
16. Holzinger, A.: On Knowledge Discovery and Interactive Intelligent Visualization of Biomedical Data - Challenges in Human–Computer Interaction & Biomedical Informatics. Conference on e-Business and Telecommunications (ICETE 2012), pp. IS9-IS20, Rome, Italy (2012)