

Guest editor's introduction: special issue of the ECML PKDD 2013 journal track

Hendrik Blockeel · Kristian Kersting ·
Siegfried Nijssen · Filip Železný

Received: 24 June 2013 / Accepted: 27 June 2013 / Published online: 10 July 2013
© The Author(s) 2013

The 2013 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD) was held in Prague, September 2013. Following the long-standing tradition of the series, it brought together researchers in Machine Learning and Data Mining for an exciting program of cutting-edge research results, aiming in particular at a cross-fertilization between these two areas. The 2013 edition featured a record-breaking 138 full presentations of novel research results, selected through a careful peer review process from 629 submissions.

H. Blockeel (✉) · S. Nijssen
Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium
e-mail: hendrik.blockeel@cs.kuleuven.be

K. Kersting
Institute for Geodesy and Geoinformation, University of Bonn,
Meckenheimer Allee 172, 53115 Bonn, Germany
e-mail: kersting@igg.uni-bonn.de

K. Kersting
Department of Knowledge Discovery, Fraunhofer IAIS,
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
e-mail: kristian.kersting@iais.fraunhofer.de

S. Nijssen
Leiden Institute for Advanced Computer Science, Universiteit Leiden,
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
e-mail: siegfried.nijssen@cs.kuleuven.be

F. Železný
Faculty of Electrical Engineering, Czech Technical University in Prague,
Technická 2, 16627 Prague, Czech Republic
e-mail: zelezny@fel.cvut.cz

For the first time, the conference used a mixed submission model. Work could be submitted as a journal article to one of two journals participating in this model (*Data Mining and Knowledge Discovery* and *Machine Learning*), or it could be submitted for publication in the conference proceedings. A total of 96 original manuscripts were submitted to *Data Mining and Knowledge Discovery*. Six of these were accepted for inclusion in this issue, and for presentation at the conference. We briefly summarize them in alphabetical order of authors.

ABACUS: Frequent Pattern Mining Based Community Discovery in Multidimensional Networks, by Michele Berlingerio, Fabio Pinelli, and Francesco Calabrese, studies the problem of community detection in multidimensional networks. Starting off with a novel definition of multidimensional community, grouping together nodes sharing memberships to the same monodimensional communities in different single dimensions, it shows that such communities are meaningful and able to group highly correlated nodes, even if they might not be connected in any of the monodimensional networks. Then, ABACUS is presented, an algorithm for extracting such multidimensional communities based on the apriori itemset miner applied to monodimensional community memberships.

Francesco Bonchi, Gianmarco De Francisci Morales, Aristides Gionis, and Antti Ukkonen tackle in *Activity Preserving Graph Simplification* the problem of simplifying a given directed graph such that a given set of observed traces of information propagation across the graph can still be explained. Unlike previous approaches, no assumption on the information propagation model is made. Instead, the problem is treated and analysed from a combinatorial point of view.

A Framework for Semi-Supervised and Unsupervised Optimal Extraction of Clusters from Hierarchies by Ricardo J.G.B. Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander presents a framework for the optimal extraction of flat clusterings from local cuts through cluster hierarchies. The extraction is formulated as an optimization problem, and a linear complexity algorithm is presented that provides the globally optimal solution to this problem in semi-supervised as well as in unsupervised scenarios.

In *Growing a list*, Benjamin Letham, Cynthia Rudin, and Katherine A Heller show how to extract overviews of topics from expert knowledge on the Internet. Starting from a small seed of example list of things posted by experts on the Internet, they grow a list of relevant items using mining techniques to find the experts and to aggregate their lists into single complete and meaningful list.

Yi-Chen Lo, Jhao-Yin Li, Mi-Yen Yeh, Shou-De Lin, and Jian Pei pose and answer the question *What Distinguishes One from Its Peers in Social Networks?*. Specifically, they consider two variants of this question, namely, how to identify the uniqueness of a given query vertex, and how to identify a group of vertices that can mutually identify each other. Several algorithms for solving these questions are developed and empirically evaluated on several networks.

Finally, in *Fast Sequence Segmentation using Log-Linear Models*, Nikolaj Tatti presents an efficient way to segment data sequences. In general, this is a quadratic problem, but Tatti shows that under a mild assumption one can do much better. Specifically, he modifies a dynamic programming approach to segmentation by pruning candidates which cannot possibly exist in an optimal segmentation. A sufficient condition is given

for segmentation in general and explored fully for the 1-d case, where checking the condition is doable. A number of experiments are done to show that the promised speed-ups occur in practice.

The double-track publication model was introduced in an attempt to bring the thorough and efficient reviewing process of journals to the conference context, while safeguarding the possibility to have innovative work presented at the conference at an early stage. It enables authors to immediately publish their newest results in a journal, without giving up the opportunity of presenting them at a conference. We believe that this model can result in a faster, more efficient and higher-quality review process, of which all benefit: journals, conferences, authors, reviewers, and ultimately the reader.

This special issue would not have been possible without the help of many people. We thank the members of the ECML PKDD 2013 “guest editorial board”, as well as the additional reviewers, for the hundreds of reviews that they have written. Their reviewing work was exceptional in terms of both quality and timeliness, and many of the articles in this issue have improved significantly through their efforts. We thank the editor-in-chief and Springer’s staff for being open to this new publication model, which, at times, put their submission procedures under a substantial amount of stress. Finally, we thank the authors for choosing Data Mining and Knowledge Discovery and ECML PKDD 2013 to publish their work.