CrossMark

Guest editors' introduction to the ECMLPKDD 2016 journal track special issue of Machine Learning

Thomas Gärtner[1] · **Mirco Nanni[2]** ·
Andrea Passerini[3] · **Céline Robardet[4]**

This special issue contains papers accepted by the ECMLPKDD 2016 journal track for publication in the Data Mining and Knowledge Discovery journal. Since 2013, the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery has been running the journal track for four consecutive years. To cover the full scope of the conference, journal track papers are published either in this special issue of Data Mining and Knowledge Discovery or in a special issue of Machine Learning. Apart from being published in the respective journal, papers accepted for the journal track are also presented at the conference, just like the papers published in the regular conference proceedings. All accepted papers will be presented by their authors at ECMLPKDD 2016 in Riva del Garda, Italy, during September 19–23, 2016. To be accepted in the ECMLPKDD journal track, papers need to be as novel and intriguing as conference papers and at the same time as substantial and mature as journal papers. In total, 62 original manuscripts were submitted to the ECMLPKDD 2016 special issue of Data Mining and Knowledge Discovery, out of which 8 were accepted in time for

✉ Mirco Nanni
   mirco.nanni@isti.cnr.it

   Thomas Gärtner
   Thomas.Gaertner@nottingham.ac.uk

   Andrea Passerini
   andrea.passerini@unitn.it

   Céline Robardet
   celine.robardet@insa-lyon.fr

[1] University of Nottingham, Nottingham, UK

[2] ISTI - CNR, Pisa, Italy

[3] Università degli Studi di Trento, Trento, Italy

[4] Institut national des sciences appliquées de Lyon, Lyon, France

⚛ Springer

printing of this special issue. In addition, 8 papers from the previous ECMLPKDD journal track were included in this special issue. This was caused by the longer reviewing and revision process required by these papers, which made them too late for printing in the previous special issue.

In their paper *Irrevocable-Choice Algorithms for Sampling from a Stream*, Yan Zhu and Eamonn Keogh show how to generalize diversification sampling strategies to the case of stream data sampling where the decision to take a (physical) sample can be irrevocable. In *A Distributed Approach for Graph Mining in Massive Networks*, Nilothpal Talukder and Mohammed J. Zaki propose a distributed algorithm to mine frequent subgraphs from a single labeled network that is too large to fit in the memory of any individual compute node. The paper *Generalized Random Shapelet Forests*, by Isak Karlsson, Panagiotis Papapetrou and Henrik Bostrom, introduces a shapelet-based classification method for large (multivariate) time series databases. In *SkOPUS: Mining top-k sequential patterns under leverage*, Francois Petitjean, Tao Li, Niko-laj Tatti and Geoffrey I. Webb focus on the extraction of top-k interesting sequential patterns based on a new notion of expected support of a pattern. In *Using regression makes extraction of shared variation in multiple datasets easy*, Jussi Korpela, Andreas Henelius, Lauri Ahonen, Arto Klami and Kai Puolamäki extend traditional redundancy analysis to compare more than two datasets at a time, and using also non-linear regression functions. The work *Top-k overlapping densest subgraphs*, by Esther Galbrun, Aristides Gionis and Nikolaj Tatti, proposes a heuristics with constant-factor approximation guarantees for extracting multiple, possibly overlapping, dense subgraphs from a large network. In *Bayesian Wishart Matrix Factorization*, by Cheng Luo and Xiongcai Cai, a matrix factorization method is introduced to model the temporal dynamics of variations among user preferences and item attractiveness. *Ensembles of label noise filters: a ranking approach*, by Luìs P. F. Garcia, Ana C. Lorena, Stan Matwin and André C. P. L. F. de Carvalho, investigates how label noise detection can be improved by using an ensemble of noise filtering techniques, and when doing that is computationally convenient. In *Locating the contagion source in networks with partial timestamps*, Kai Zhu, Zhen Chen and Lei Ying study the problem of identifying a single contagion source when partial timestamps of a contagion process are available, proposing two ranking algorithms. *Mining rooted ordered trees under subtree homeomorphism*, by Mostafa Haghir Chehreghani and Maurice Bruynooghe, presents an efficient algorithm for subtree homeomorphism with application to frequent pattern mining, based on novel compact data-structures and associated join operations. In *Scalable time series classification*, Patrick Schäfer proposes a novel algorithm for classifying time series that is orders of magnitude more efficient than baseline solutions and robust to noise. *C-BiLDA: extracting cross-lingual topics from non-parallel texts by distinguishing shared from unshared content*, by Geert Heyman, Ivan Vulić and Marie-Francine Moens, introduces a new bilingual probabilistic topic model which is able to deal with non-parallel multilingual text datasets with partially overlapping thematic content. In *ClusPath: a temporal-driven clustering to infer typical evolution paths*, Marian-Andrei Rizoiu, Julien Velcin, Stéphane Bonnevay and Stéphane Lallich propose a novel algorithm for detecting high-level evolution tendencies in a population of entities, inferred from low-level descriptive features. The paper *An efficient exact algorithm for triangle listing in large graphs*, by Sofiane Lagraa and Hamida

Seba, tackles the problem mentioned in the title by using a compressed copy of the input graph without decompressing it, yielding improvements in both storage requirement of the graphs and their time processing. In *Exact and efficient top-K inference for multi-target prediction by querying separable linear relational models*, Michiel Stock, Krzysztof Dembczyński, Bernard De Baets and Willem Waegeman analyse efficient and exact algorithms for computing the top-K predictions in complex multi-target prediction problems on large target spaces, using a general class of models referred as separable linear relational models. *Optimizing network robustness by edge rewiring: a general framework*, by Hau Chan and Leman Akoglu, addresses the problem of modifying a graphs structure under a given budget so as to maximally improve its robustness, as quantified by spectral measures.