Mach Learn (2014) 97:1–3
DOI 10.1007/s10994-014-5468-6

EDITORIAL

Guest Editors' introduction: special issue of the ECML/PKDD 2014 journal track

Toon Calders · Floriana Esposito · Eyke Hüllermeier · Rosa Meo

Received: 10 July 2014 / Accepted: 16 July 2014 / Published online: 19 August 2014

This special issue is a collection of papers submitted to the ECML/PKDD 2013 and 2014 journal tracks and accepted for publication in "Machine Learning".

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML/PKDD, launched its journal track in 2013. In order to cover the full scope of the conference, which is a merger of the formerly independent conferences ECML and PKDD, two journals were involved: "Machine Learning" and "Data Mining and Knowledge Discovery". In addition to being published in the respective journal, papers accepted in the journal track are also presented at the conference, just like the contributions to the regular conference proceedings. Thus, all papers of this special issue are also presented by their authors at the ECML/PKDD 2014 conference in Nancy, France, from September 15th to 19th.

Given the special nature of the ECML/PKDD journal track, submissions are supposed to meet specific criteria. First and foremost, like any other submission, they are of course expected to comply with the high scientific standards of the two journals. Not less importantly, however, they should naturally lend themselves to conference presentations. In other words, they are supposed to be conference and journal papers at the same time: as novel and intriguing

T. Calders (✉)
Department of Computer and Decision Engineering, Faculty of Applied Sciences, Université Libre de Bruxelles, 50, Av. F. Roosevelt, CP 165/15, 1050 Brussels, Belgium
e-mail: toon.calders@ulb.ac.be

F. Esposito
Dipartimento di Informatica, Universitá di Bari, Via Orabona 4, 70126 Bari, Italy
e-mail: floriana.esposito@uniba.it

E. Hüllermeier
Department of Computer Science, University of Paderborn, Paderborn, Germany
e-mail: eyke@upb.de

R. Meo
Dipartimento di Informatica, Facoltà di Scienze della Natura, Università di Torino, Corso Svizzera 185, 10149 Turin, Italy
e-mail: meo@di.unito.it

as the former, and as substantial and mature as the latter. These requirements exclude, for example, journal versions of previously published conference papers as well as survey papers, which were not considered for the special issue.

In total, 40 original manuscripts were submitted to the ECML/PKDD 2014 special issue of "Machine Learning" in the course of the year 2014, of which 4 were eventually accepted in time for inclusion in this special issue. In addition, 5 papers from the 2013 journal track are included here. Due to sometimes inevitable delays in the reviewing and revision process, papers can not always be included in the special issue for the year they were originally submitted. In what follows, we briefly summarize the contents of the accepted papers.

In "The bane of skew: Uncertain ranks and unrepresentative precision", Thomas Lampert and Pierre Gançarski analyze how the skew of a classification problem can influence performance measures such as precision and area under a precision-recall curve (AUCPR). In this regard, they show several interesting theoretical results and provide new insights. They show, for example, that changing the skew may change the relative ranking of algorithms with regard to AUCPR.

In their paper "On combining machine learning with decision making", Theja Tulabandhula and Cynthia Rudin introduce the *Machine Learning with Operational Costs* (MLOC) framework, which is considered as an exploratory form of decision theory and incorporates knowledge about how a predictive model will be used for a subsequent task. They make use of this framework for studying the "machine learning and traveling repairman" problem.

In "Leave-one-out cross-validation is risk consistent for lasso", Darren Homrighausen and Daniel McDonald analyze the behavior of the lasso when the smoothing parameter is chosen in a data-dependent way. In particular, they show that, under some restrictions on the design matrix, the lasso estimator is still risk consistent if this parameter is chosen via cross-validation.

In "A theoretical and empirical analysis of support vector machine methods for multiple-instance classification", Gary Doran and Soumya Ray analyze extensions of support vector machines for the multiple-instance (MI) classification problem. They show that all existing algoroithms are lacking at least some of the (desirable) properties that are valid for SVMs in standard classification. They provide theoretical results showing that this problem is of fundamental nature, as well as experimental observations supporting these results.

In their paper "A constrained matrix-variate Gaussian process for transposable data", Oluwasanmi Koyejo, Cheng Lee, and Joydeep Ghosh propose a novel approach for modeling so-called transposable data with missing interactions given additional side information. To this end, they combine the matrix-variate Gaussian process model with low rank constraints. The approach is applied to the prediction of hidden associations between genes and diseases. Another application is presented in the filed of recommender systems.

In "Learning a priori constrained weighted majority votes", Aurélien Bellet, Amaury Habrard, Emilie Morvant, and Marc Sebban propose an extension of the MinCq algorithm for finding posterior weights in a majority voting rule. Their algorithm, called P-MinCq, extends MinCq to the case of non-uniform prior probabilities. Moreover, the authors offer an extension of a proof of convergence to the data-dependent setting, a new approach for learning weights of a set of k-nearest neighbor classifiers, and an extensive experimentation and comparison with alternative approaches.

In "Regression conformal prediction with random forests", Ulf Johansson, Henrik Boström, Tuve Löfström, and Henrik Linusson introduce a novel approach to conformal prediction for regression using random forests. Apart from their robust predictive performance, the authors show that random forests allow for determining the size of prediction intervals

by using out-of-bag estimates instead of requiring a separate calibration set. Moreover, they provide an extensive empirical evaluation of their approach.

In "Collaborative filtering with information-rich and information-sparse entities", Kai Zhu, Rui Wu, Lei Ying, and R. Srikant consider a popular model for collaborative filtering in recommender systems, where only users (or items) are clustered, and the co-clustering model, where both users and items are clustered. The authors extended an existing approach to the more realistic case in which only a smaller number of users rated a large number of items. For this new setting, the authors analyze the sample complexity for rating matrix recovery.

In "SAGA: Sparse and geometry-aware non-negative matrix factorization through non-linear local embedding", Nicolas Courty, Xing Gong, Jimmy Vandel, and Thomas Burger present a variant of the non-negative matrix factorization procedure, which accounts for the geometrical structure of the manifold embedding of the data, allows for controlling the level of sparsity and has an overall linear complexity. The authors show several advantages of their approach in comparison to existing methods.

This special issue would not have been possible without the help of many people. In particular, we would like to thank the members of the ECML/PKDD 2014 "guest editorial board", as well as the additional referees for their hard work and timely reviewing of the papers submitted to the special issue.