EDITORIAL

Guest editors' introduction: Special Issue from ECML PKDD 2009

**Aleksander Kołcz · Dunja Mladenić · Wray Buntine ·
Marko Grobelnik · John Shawe-Taylor**

In 2009 the joint ECML PKDD event, combining the European Conference on Machine Learning (ECML) and the European Conference on Principles of Data Mining and Knowledge Discovery (PKDD), attracted 422 strong paper submissions. ECML PKKD 2009 is a continuation of the successful series of 19 past ECML and 12 past PKDD conferences, which have been collocated since 2001. After a rigorous review process involving the General Chair, 3 Program Chairs, 26 Area Chairs and 254 Program Committee Members, 105 papers were accepted for publication at the conference. Continuing with the tradition initiated in 2008, a number the best papers have been selected to be published in the special issues of the journal of Machine Learning and the journal of Data Mining and Knowledge Discovery.

A. Kołcz (✉)
Microsoft, 1 Microsoft Way, Redmond, WA 98052, USA
e-mail: alek@ir.iit.edu

D. Mladenić · M. Grobelnik
Jozef Stefan Institute, Ljubljana, Slovenia

D. Mladenić
e-mail: dunja.mladenic@ijs.si

M. Grobelnik
e-mail: Marko.Grobelnik@ijs.si

W. Buntine
Helsinki Institute of IT, Helsinki, Finland
e-mail: wray.buntine@nicta.com.au

W. Buntine
NICTA, Canberra, Australia

J. Shawe-Taylor
University College London, London, UK
e-mail: J.Shawe-Taylor@cs.ucl.ac.uk

Below we provide a brief overview of the 14 papers selected for the journal special issues. The papers cover a broad spectrum of topics relevant to the machine learning and data mining communities.

Papers appearing in the journal of *Machine Learning*:

In *"Sparse kernel SVMs via cutting plane training"* Joachims and Yu (2009) advance the state of the art in efficient induction and evaluation of SVM models. Model sparsity is enhanced by allowing for support vectors from outside of the training set.

In *"Learning multilinear representations for efficient inference"* Roth and Samdani (2009) present a method of speeding up probabilistic inference by representing discrete probability distributions using multilinear forms.

In *"Combing instance-based learning and logistic regression for multi-label classification"* Cheng and Huellermeier (2009) propose to use a Nearest Neighbor technique to locally condition logistic regression when learning multi-label problems.

In *"On structured output training: hard cases and an efficient alternative"* Gärtner and Vembu (2009) propose an efficient algorithm for avoiding excessive computations in problems involving structured-output prediction.

In *"Hybrid least-squared algorithms for approximate policy evaluation"* Johns et al. (2009) combine two key methods of approximate policy evaluation into a stronger alternative.

In *"A self-training approach to cost sensitive uncertainty sampling"* Liu et al. (2009) demonstrate how traditional forms of uncertainty sampling in active learning fail in asymmetric misclassification cost scenarios and propose an remedy technique based on self-training.

In *"Cost-sensitive learning based on Bregman divergences"* Santos-Rodríguez et al. (2009) propose the use of Bregman divergences to learn accurate posterior probabilities near the optimum decision boundary.

Papers appearing in the journal of *Data Mining and Knowledge Discovery*:

In *"Identifying components"* van Leeuwen et al. (2009) propose two parameter-free MDL-based methods of identifying mixture components.

In *"A recursive realistic graph generator using random typing"* Akoglu and Faloutsos (2009) propose an efficient method of simulating graphs that satisfy the statistical properties of real data.

In *"Subgroup discovery in numerical domains"* Grosskreutz and Rüping (2009) propose an improved way of identifying subgroups for continuous-attribute data with an extensive comparison to existing approaches.

In *"Taxonomy-driven lumping for sequence mining"* Bonchi et al. (2009) introduce a first-order Markov method of describing sequence data by algorithm based on a user-provided taxonomy.

In *"Harnessing the strengths of anytime algorithms for constant data streams"* Kranen and Seidl (2009) compare the performance of several approaches to budgeted classification for several popular learners.

In *"Two-way analysis of high-dimensional collinear data"* Huopaniemi et al. (2009) propose a hierarchical approach to the ANOVA analysis, which addresses the challenges of small-sample data of high dimensionality.

In *"A fast ensemble pruning algorithm based on pattern mining process"* Zhao et al. (2009) apply the techniques from the area of associative rule mining to perform pruning of an ensemble of predictors.

References

Akoglu, L., & Faloutsos, C. (2009). A recursive realistic graph generator using random typing. *Data Mining and Knowledge Discovery*, *19*(2), 194–209.

Bonchi, F., Castillo, C., & Donato, D. (2009). Taxonomy-driven lumping for sequence mining. *Data Mining and Knowledge Discovery*, *19*(2), 227–244.

Cheng, W., & Huellermeier, E. (2009). Combing instance-based learning and logistic regression for multi-label classification. *Machine Learning*, *76*(2–3), 211–225.

Gärtner, T., & Vembu, S. (2009). On structured output training: hard cases and an efficient alternative. *Machine Learning*, *76*(2–3), 227–242.

Grosskreutz, H., & Rüping, S. (2009). Subgroup discovery in numerical domains. *Data Mining and Knowledge Discovery*, *19*(2), 210–226.

Huopaniemi, I., Suvitaival, T., Nikkila, J., Oresic, M., & Kaski, S. (2009). Two-way analysis of high-dimensional collinear data. *Data Mining and Knowledge Discovery*, *19*(2), 261–276.

Joachims, T., & Yu, C.-N.J. (2009). Sparse kernel SVMs via cutting plane training. *Machine Learning*, *76*(2–3), 179–193.

Johns, J., Petrik, M., & Mahadevan, S. (2009). Hybrid least-squared algorithms for approximate policy evaluation. *Machine Learning*, *76*(2–3), 243–256.

Kranen, P., & Seidl, T. (2009). Harnessing the strengths of anytime algorithms for constant data streams. *Data Mining and Knowledge Discovery*, *19*(2), 245–260.

Liu, A., Jun, G., & Ghosh, J. (2009). A self-training approach to cost sensitive uncertainty sampling. *Machine Learning*, *76*(2–3), 257–270.

Roth, D., & Samdani, R. (2009). Learning multilinear representations for efficient inference. *Machine Learning*, *76*(2–3), 195–209.

Santos-Rodríguez, R., Alaiz-Rodríguez, R., Guerrero-Curieses, A., & Cid-Sueiro, J. (2009). Cost-sensitive learning based on Bregman divergences. *Machine Learning*, *76*(2–3), 271–285.

van Leeuwen, M., Vreeken, J., & Siebes, A. (2009). Identifying components. *Data Mining and Knowledge Discovery*, *19*(2), 176–193.

Zhao, Q.-L., Jiang, Y.-H., & Xu, M. (2009). A fast ensemble pruning algorithm based on pattern mining process. *Data Mining and Knowledge Discovery*, *19*(2), 277–292.