Guest editors' introduction: special issue of selected papers from ECML PKDD 2011

**Dimitrios Gunopulos · Donato Malerba ·
Michalis Vazirgiannis**

Received: 26 June 2012 / Revised: 18 July 2012 / Accepted: 18 July 2012 / Published online: 26 July 2012
© The Author(s) 2012

The 2011 edition of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) was held in Athens, Greece, during September 5–9, 2011. Ten years after the first edition of this joint conference, ECML PKDD 2011 continued to provide a common forum for the closely related fields of machine learning and data mining. Apart from six plenary invited talks, four invited talks for the industrial session, a demo session, six tutorials and eleven co-located workshops, the main technical sessions comprised the presentation of 121 peer-reviewed papers selected by the program committee from 599 full-paper submissions. ECML PKDD 2011 was a highly selective conference and the proceedings were published in three volumes of the Springer's Lecture Notes in Artificial Intelligence series (Gunopulos et al. 2011a, 2011b, 2011c).

Authors of the best ten machine learning papers presented at the conference were invited to submit a significantly extended version of their paper to this special issue. The selection was made by the Program Chairs on the basis of their exceptional scientific quality and high impact on the field, as indicated by conference reviewers.

In this special issue you will find seven papers which have been accepted after two or three rounds of peer-reviewing according to the journal criteria. The diversity of topics addressed in these papers reflects the significant progress being made by the machine learning community in the theoretical understanding of the principles underlying knowledge discov-

D. Gunopulos
University of Athens, Panepistimiopolis, 15784 Athens, Greece
e-mail: dg@di.uoa.gr

D. Malerba (✉)
Università degli Studi di Bari "Aldo Moro", Dipartimento di Informatica, via Orabona 4, 70125 Bari, Italy
e-mail: donato.malerba@uniba.it

M. Vazirgiannis
Athens University of Economics & Business, 12 Kodrigktonos, 11257 Athens, Greece
e-mail: mvazirg@aub.gr

ery in databases, as well as in the application of tools and techniques to real-world problems. We believe these works have the potential to spur new research in the field.

The paper “Good edit similarity learning by loss minimization” (Bellet et al. 2012) deals with the problem of learning similarity functions for string data (with extensions to tree-structured data). The similarity measure considered in this work is based on the edit distance between two strings, which is defined as the cost of the best sequence of operations (insertion, deletion, substitution of a symbol) required to transform one string into another. An edit cost is assigned to each operation and the authors propose a new framework to learn the edit costs, so as to optimize the “goodness” of the resulting similarity function. Goodness is here based on the recent notion of (ϵ, γ, τ) -good similarity functions.

The paper “Novel high intrinsic dimensionality estimators” (Rozza et al. 2012) advances the state of the art on estimators of the intrinsic dimensionality of a dataset, which is defined as the minimum number of parameters needed to represent the data without information loss. The authors provide us a theoretical motivation for the bias that causes the underestimation of intrinsic dimensionality, when this is sufficiently high. Based on these theoretical considerations, they propose two new estimators which are less affected by the bias and are based on statistical properties of manifold neighborhoods.

The paper “Generating feature spaces for linear algorithms with regularized sparse kernel slow feature analysis” (Böhmer et al. 2012) introduces a kernelized version of Slow Feature Analysis (SFA). This is an unsupervised feature extraction method used in multidimensional time series, which look for feature spaces with a slow variation between consecutive values of the time series. By kernelizing SFA it is possible to generate a new representation of time series which are linear functions in the feature space induced by the kernel, and non-linear in the original feature space. Thus, linear algorithms can be applied to non-linear problems.

The paper “Sequential approaches for learning datum-wise sparse representations” (Dulac-Arnold et al. 2012) proposes a novel classification technique for selecting an appropriate representation for each data point, in contrast to the usual approach of selecting a representation encompassing the whole dataset. This datum-wise representation is found by using a sparsity inducing empirical risk, a relaxation of the standard L_0 regularized risk. The sparsity level is configured for each instance depending on how difficult it is to classify: easy-to-classify instances are represented with few features, while hard-to-classify instances may employ more features. Potential applications of this datum-wise learning framework to a wide range of sparsity-relevant problems are reported.

The main motivation of the paper “Towards preference-based reinforcement learning” (Fürrnkranz et al. 2012) is that conventional reinforcement learning methods are essentially confined to dealing with numerical rewards, while there are many applications in which this type of information is not available and in which only qualitative reward signals are provided instead. Therefore, it is important to investigate a reinforcement learning algorithm that leverages qualitative feedback. In this work, instead of assigning a numerical utility to actions, feedback is given as preference statements in the form of ordered pairs of actions. The medical application domain described in this paper is a good example of the natural application of this preference-based reinforcement learning.

The paper “Focused multi-task learning in a Gaussian process framework” (Leen et al. 2012) is concerned with the application of Gaussian process models to multi-task learning scenarios. In contrast to previous works, where all tasks have been assumed to be of equal importance, in this paper an asymmetry is introduced by making a clear distinction between a primary task, for which performance is the most important, and auxiliary tasks. Although transfer learning is inherently asymmetric, the novelty here is in the taking of a symmetric

learning approach and adjusting its focus to a particular task. This approach offers a nice conceptual bridge between multi-task learning and other methods of knowledge transfer.

The paper “Learning monotone nonlinear models using the Choquet integral” (Tehrani et al. 2012) investigates the problem of learning predictive models which guarantee the monotonicity in the input variables, i.e., *ceteris paribus*, by increasing or decreasing an input variable, the output variable can only increase. In several application domains, the conformance of the learned predictive models to this monotonicity constraint is a desirable property since it leads to more easily interpretable results. The authors observe that a solution to this constrained learning problem can be found by using the discrete Choquet integral. Moreover, by analyzing the Choquet integral from a classification perspective, they provide us with a detailed analysis of the upper and lower bounds on the VC-dimension. As a concrete application of the Choquet integral, a generalization of logistic regression is proposed, called “Choquistic regression.”

We hope the readers will enjoy these articles and will find in them a source of inspiration for their work.

Acknowledgements We wish to warmly thank all authors who submitted their papers to this special issue. We are indebted to the reviewers for their careful work and for their constructive comments to the authors. We are grateful to the journal Editor-in-Chief Dr. Peter Flach for his great help and support in organizing this special issue. A special thanks to Melissa Fearon from the Springer editorial staff for her tireless support and troubleshooting in the production of this special issue.

References

- Bellet, A., Habrard, A., & Sebban, M. (2012). Good edit similarity learning by loss minimization. *Machine Learning*. doi:10.1007/s10994-012-5293-8.
- Böhrer, W., Grünwälder, S., Nickisch, H., & Obermayer, K. (2012). Generating feature spaces for linear algorithms with regularized sparse kernel slow feature analysis. *Machine Learning*. doi:10.1007/s10994-012-5300-0.
- Dulac-Arnold, G., Denoyer, L., Preux, P., & Gallinari, P. (2012). Sequential approaches for learning datum-wise sparse representations. *Machine Learning*. doi:10.1007/s10994-012-5306-7.
- Fürnkranz, J., Hüllermeier, E., Cheng, W., & Park, S. H. (2012). Towards preference-based reinforcement learning. *Machine Learning*. doi:10.1007/s10994-012-5313-8.
- Gunopulos, D., Hofmann, T., Malerba, D., & Vazirgiannis, M. (Eds.) (2011a). *Lecture notes in computer science: Vol. 6911. Machine learning and knowledge discovery in databases—European conference, ECML PKDD 2011, Athens, Greece, September 5–9, 2011. Proceedings, Part I*. Berlin: Springer.
- Gunopulos, D., Hofmann, T., Malerba, D., & Vazirgiannis, M. (Eds.) (2011b). *Lecture notes in computer science: Vol. 6912. Machine learning and knowledge discovery in databases—European conference, ECML PKDD 2011, Athens, Greece, September 5–9, 2011. Proceedings, Part II*. Berlin: Springer.
- Gunopulos, D., Hofmann, T., Malerba, D., & Vazirgiannis, M. (Eds.) (2011c). *Lecture notes in computer science: Vol. 6913. Machine learning and knowledge discovery in databases—European conference, ECML PKDD 2011, Athens, Greece, September 5–9, 2011. Proceedings, Part III*. Berlin: Springer.
- Leen, G., Peltonen, J., & Kaski, S. (2012). Focused multi-task learning in a Gaussian process framework. *Machine Learning*. doi:10.1007/s10994-012-5302-y.
- Rozza, A., Lombardi, G., Ceruti, C., Casiraghi, E., & Campadelli, P. (2012). Novel high intrinsic dimensionality estimators. *Machine Learning*. doi:10.1007/s10994-012-5294-7.
- Tehrani, A. F., Cheng, W., Dembczyński, K., & Hüllermeier, E. (2012). Learning monotone nonlinear models using the Choquet integral. *Machine Learning*. doi:10.1007/s10994-012-5318-3.