



I Seminari su “Information Technology Outlook” – Dottorato di Ricerca in Informatica e Matematica

Dr. Dragi Kocev

Post-doctoral researcher, Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

Visiting fellow, Dipartimento di Informatica University of Bari Aldo Moro, Bari, Italy

Venerdì 14 novembre 2014, ore 10.00, Aula Goedel, 2° piano, Dipartimento di Informatica

Tree ensembles for predicting structured outputs

In this talk, we will address the task of learning models for predicting structured outputs - the input is a tuple of attribute values and the output is a structured object. We consider both global and local prediction of structured outputs, the first based on a single model that predicts the entire output structure and the latter based on a collection of models, each predicting a component of the output structure. In particular, we consider predictive clustering trees and ensembles thereof for predicting structured outputs. We will present methods for learning bagging and random forests of PCTs for global and local prediction of different types of structured outputs. The types of considered outputs correspond to different predictive modeling tasks: predicting tuples of continuous variables corresponds to multi-target regression, predicting tuples of discrete variables corresponds to multi-target classification, predicting tuples of binary variables corresponds to multi-label classification and predicting tuples of binary variables organized into a hierarchy corresponds to hierarchical multi-label classification.

We give the results from an extensive experimental evaluation of the ensembles for structured prediction across a range of benchmark datasets for each of the four types of structured outputs. We compare ensembles for global and local prediction, as well as single trees for global prediction and tree collections for local prediction, both in terms of predictive performance and efficiency (running times and model complexity). Both global and local ensembles perform better than the single model counterparts in terms of predictive power. Global and local ensembles perform equally well, with global ensembles being more efficient and producing smaller models, as well as needing fewer trees in the ensemble to achieve the maximal performance.

Finally, we will present successful applications of the ensembles for predicting structured outputs in three real-life domains: life sciences, ecological modelling and image analysis. In life sciences, we will focus on the task of drug repositioning. In ecological modelling, we will show the use of the ensembles for habitat modelling, modelling the vegetation condition and modelling the weed and crop cover in an agro-ecosystem. In image analysis, we will discuss the construction of visual codebooks for image retrieval and annotation.

Dragi Kocev received his Ph.D. in computer science in 2011 from the Jožef Stefan International Postgraduate School, Ljubljana, Slovenia. Since 2011, he has been post-doctoral Researcher of the Jožef Stefan Institute. His current research interests include the study and development of different data mining algorithms and their applications in machine vision, life sciences and ecological modelling. He has a long publication record in well-known international journals and has been Program committee co-chair of Discovery Science 2014. He has collaborated on the EU funded projects: IQ and PHAGOSYS. Currently, he is involved in the Human Brain Project and is co-coordinator of the FP7 FET Open project MAESTRA.