Learning from Massive, Incompletely annotated & Structured Data

Sašo Džeroski*

*Department of Knowledge Technologies, Jožef Stefan Institute
Jamova cesta 39, Ljubljana, Slovenia.
http://www-ai.ijs.si/SasoDzeroski/

Biography

Sašo Džeroski is a scientific councillor at the Jozef Stefan Institute and the Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, both in Ljubljana, Slovenia. He is also a full professor at the Jozef Stefan International Postgraduate School. His research is mainly in the area of machine learning and data mining (including structured output prediction and automated modeling of dynamic systems) and their applications (mainly in environmental sciences, incl. ecology, and life sciences, incl. systems biology). He is co-author/co-editor of more than ten books/volumes, including "Inductive Logic Programming", "Relational Data Mining", "Learning Language in Logic", "Computational Discovery of Scientific Knowledge" and "Inductive Databases and Constraint-Based Data Mining". He has participated in many international research projects (mostly EU-funded) and coordinated two of them in the past. He is currently the coordinator of the FET XTrack project MAESTRA (Learning from Massive, Incompletely annotated, and Structured Data) and one of the principal investigators in the FET Flagship Human Brain Project.

Summary

The MAESTRA project (http://maestra-project.eu/) addresses the ambitious task of predicting different types of structured outputs in several challenging settings, such as semi-supervised learning, mining data streams and mining network data. It develops machine learning methods that work in each of these settings, as well as combinations thereof. The techniques developed are applied to problems from the area of biology and bioinformatics, sensor data analysis, multimedia annotation and retrieval, and social network analysis. The talk will give an introduction to the project and the topics it addresses, an overview of the results of the project, and a detailed description of selected techniques and applications: Semi-supervised learning for structured-output prediction (SOP) and SOP on data streams will be discussed for the task of multi-target regression (MTR), as well as applications of MTR for the annotation/retrieval of images.