Research profile

The Jožef Stefan Institute (www.ijs.si), JSI, is the leading Slovenian research institute, covering a broad spectrum of basic and applied research in the area of natural sciences, life sciences and engineering. Most of its 960 employees hold a PhD degree. Based on the last European Research Ranking, the institute iss ranked among the top ten research institutes in Europe.

The Department of Knowledge Technologies performs research in advanced information technologies aimed at acquiring, storing and managing knowledge, to be used in the development of an information and knowledge-based society. Established areas include intelligent data analysis (machine learning, data mining, and knowledge discovery in databases), semantic data mining and the semantic web, language technologies and computational linguistics, decision support and knowledge management, while computational creativity is a novel research area. Apart from research in knowledge technologies, mostly based on artificial intelligence methods, we are also developing applications in environmental sciences, medicine and health care, biomedicine and bioinformatics, economy and marketing, linguistics and digital humanities. The Department's Knowledge Technologies research programme has been evaluated as the best research programme in ICT by the Slovenian Research Agency.

The research group of prof. Dr. Sašo Džeroski is consisted of approximately 20 researchers and students. The focus of the group is on machine learning and its application to practical problems from environmental and life sciences. The major contributions of the group members is to the machine learning areas of relational learning, computational scientific discovery, and the data mining areas of inductive databases and constraint-based data mining. Recent contributions of group members include development of methods for structured output prediction, methods for automated modelling of dynamic systems and development of scientific ontologies. Group members have been involved in various domestic and EU funded projects, such as network of excellence ILP, ILP2, ILPnet, METAL, ECOGEN, cInQ, SIGMEA, EET Pipeline, PHAGOSYS, REWIRE, SUMO; as well as coordinated several projects, such as ILPnet2, and the projects IQ (Inductive Queries for Mining Patterns and Models, FP6) and MAESTRA (Learning from Massive, Incompletely annotated, and Structured Data,FP7), both FET Open.

Software

CLUS

https://sourceforge.net/projects/clus/

Clus is a decision tree and rule learning system that works in the predictive clustering framework. While most decision tree learners induce classification or regression trees, Clus generalizes this approach by learning trees that are interpreted as cluster hierarchies. We call such trees predictive clustering trees or PCTs. Depending on the learning task at hand, different goal criteria are to be optimized while creating the clusters, and different heuristics will be suitable to achieve this. Classification and regression trees are special cases of PCTs, and by choosing the right parameter settings Clus can closely mimic the behavior of tree learners such as CART or C4.5. However, its applicability goes well beyond classical classification or regression tasks: Clus has been successfully applied to many different tasks including multi-task learning (multi-target classification and regression), structured output learning, multi-label classification, hierarchical classification, hierarchical regression, and time series prediction. Next to these supervised learning tasks, PCTs are also applicable to semi-supervised learning, subgroup discovery, and clustering. In a similar way, predictive clustering rules (PCRs) generalize classification rule sets and also apply to the aforementioned learning tasks.

ProBMot - Process-based modeling tool
http://probmot.ijs.si/
ProBMoT is an implementation of the process-based modeling approach to modeling dynamical systems.

In the core of the process-based modeling approach is a formalism for representing models of dynamical systems as well as knowledge for modeling dynamical systems in a particular domain of interest. The process-based model formalism allows for representing models of dynamical systems at two levels of abstraction. At the higher level, the model is represented as sets of processes that govern the dynamics of the observed system and entities involved in the processes. At the lower level, each process includes a model of its dynamical influence on the variables of the observed system. The process-based modeling software can automatically combine the models of individual processes into a set of coupled differential equations used to simulate the behavior of the observed system. Thus, process-based models at the higher abstraction level reveal the structure of the observed systems in terms of entities and process interactions among them, providing explanation of the model behavior obtained by a lower-level declaration of the model equations.

To start establishing process-based models, we first have to formalize the modeling knowledge by establishing templates of generic entities that appear in the generic processes that govern the dynamics of systems in the particular domain. This modular knowledge representation allows for automated modeling of an observed system following a compositional approach. For a given modeling task, the generic templates are being instantiated into specific components (entities and processes) that can be used as building blocks for process-based models. Combinations of these building blocks represent candidate process-based models of the observed system. Automated modeling tool than searches for a process-based model with an optimal fit between the simulated and observed behavior of the system at hand.