

Predicting River Water Communities with Logical Decision Trees

Sašo Džeroski*, Jasna Grbovic** and Hendrik Blockeel***

**Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia*

***Hydrometeorological Institute, Ljubljana, Slovenia*

**** Katholieke Universiteit Leuven, Belgium*

We apply the machine learning approach of induction of decision trees to biological and chemical data collected through regular monitoring of rivers in Slovenia. We address the problem of finding relationships between the physical and chemical properties of river water and the biological community present in that water.

Typical predictive approaches to this problem use artificial neural networks. A separate neural network is learned to predict the occurrence of each of a number of selected taxa. A similar approach can be taken with decision tree induction, where a separate tree is learned for predicting each taxon. The decision trees have the advantage of expressing regularities explicitly and being easy to inspect and check for ecological validity. Logical decision trees, as implemented in the system TILDE, enable us to address the task of predictive clustering and thus predict the occurrence of a number of taxa simultaneously with a single decision tree, rather than having a separate tree for each taxon.

In this work, we use data about Slovenian that cover a six year period, from 1990 to 1995. Biological samples are taken twice a year, once in summer and once in winter, while physical and chemical analyses are performed several times a year for each sampling site. The physical and chemical samples include the measured values of 15 different parameters, such as biological oxygen demand and chlorine concentration. We apply TILDE to induce two decision trees, one for predicting selected plant taxa and the other for predicting selected animal taxa. Different subsets of the chemical parameters are used for each of the trees, as different parameters are relevant for plant and animal taxa. We examine the predictive power and ecological validity of the learned logical decision trees.