

G.6.1 - Applications of KDD Methods in Environmental Sciences

Sašo Džeroski

Department of Intelligent Systems, Jožef Stefan Institute

Jalova 39, 1000 Ljubljana, Slovenia

in the research area of Artificial Intelligence

the lagoon of Venice in relation to water temperature, dissolved nitrogen and phosphorus and dissolved oxygen. The trees give a rough picture of the relative importance of the factors influencing algal growth (cf. Figure 1), revealing that nitrogen is the limiting factor (and thus providing a negative answer to the question in the above paragraph). The equations discovered, on the other hand, give better prediction of the peaks and crashes of algal biomass.



Table 1: The discovered mode for phytoplankton

It was also more understandable to domain experts: the first term describes phytoplankton growth, where temperature and phosphorus are limiting factors. The last two terms describe phytoplankton death and the feeding of zooplankton on phytoplankton.

The following issues were raised in this application:

1. Data quantity and preprocessing: measurements were only made at 14 time points during two months (once weekly). Some preprocessing/interpolation was thus

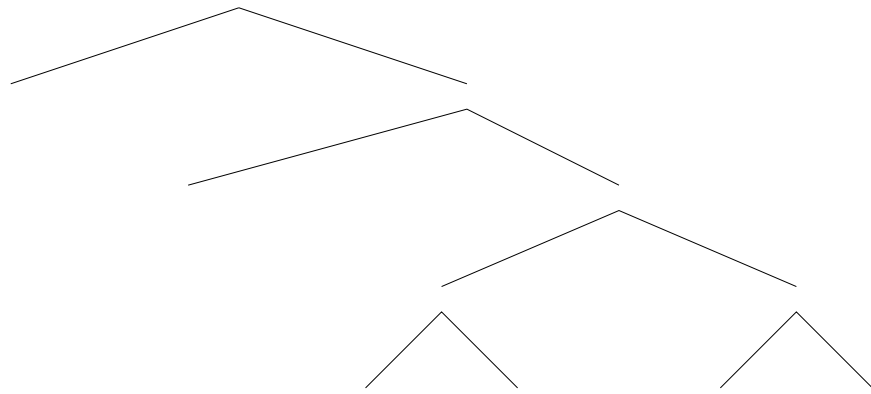
Ha

Environmental monitoring and protection

A typical national environmental protection agency aims to "protect public health and to safeguard and improve the natural environment." It sets and enforces national pollution-control standards.

countries where extensive biological monitoring is conducted. Regular monitoring for a very wide range of chemical pollutants would be very expensive, if not impossible. On the other hand, the state of the biota can reflect an increase in pollution and indicate likely causes/sources.

We used data on biological and chemical samples from S



of animals. It is thus highly desirable to have reliable models that can be used to both quantitatively and qualitatively describe the fate and behavior of compounds in the environment

Table 2:

A more traditional remote sensing application of data mining is land cover classification.

G

airia stereo photographs.

Unsupervised Result

Proximity to Lake



4. Allowing experts to provide domain knowledge by directing the KDD process: in this application it has proven crucial for the domain expert to be able to guide the discovery of patterns (decision trees);

- Hansen, M., Defries, R., Townshend, J., and Sohlberg, R. (2000) Global land cover classification at 1km spatial resolution using a classification tree approach. *Int. J. Remote Sensing*. Forthcoming.
- Jeffers, J.N.R. (1999) Genetic algorithms I. In A.H. Fielding, editor, *Machine Learning Methods for Ecological Applications*, pages 107–121. Kluwer Academic Publishers, Dordrecht.
- Jones, E., and Roydhouse, A. (1995) Intelligent retrieval of archived meteorological data. *IEEE Intelligent Systems* 10(6).
- Kampichler, C., Džeroski, S., and Wieland, R. (2000) The application of machine learning techniques to the analysis of soil ecological data bases

- Todorovski, L., Džeroski, S