ELSEVIER

Ecological Modelling 146 (2001) 263–273

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Applications of symbolic machine learning to ecological modelling

Sašo Džeroski *

Department of Intelligent Systems, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Abstract

Symbolic machine learning methods induce explicitly represented symbolic models from data. The models can thus be inspected, modified, used and verified by human experts and have the potential to become part of the knowledge in the respective application domain. Applications of symbolic machine learning methods to ecological modelling problems are numerous and varied, ranging from modelling algal growth in lagoons and lakes (e.g. in the Venice lagoon) to predicting biodegradation rates for chemicals. This paper gives an overview of machine learning applications to ecological modelling, focussing on applications of symbolic machine learning and giving more detailed accounts of several such applications.

Keywords: Machine learning; Decision trees; Equation discovery; Population dynamics; Habitat suitability; Environmental monitoring

1. Introduction

Symbolic machine learning methods induce explicitly represented symbolic models from data. This is in contrast to methods like neural networks, which most often produce black-box models. Models constructed by symbolic machine learning can be inspected, modified, used and verified by human experts and have the potential to become part of the knowledge in the respective application domain. Symbolic machine learning methods include the induction of decision trees for classification and regression, algebraic, differential and partial differential equations, and classification rules, including relational rules (also known as inductive logic programming). They heuristically search the space of possible models to identify models appropriate for the data at hand.

Applications of symbolic machine learning methods to ecological modelling problems are numerous and varied. They range from modelling algal growth in lagoons and lakes (e.g. in the Venice lagoon), through modelling interactions of a red deer population with the new growth in a forest, to habitat-suitability modelling for pronghorn antelopes, brown bears, and Collembola.

This paper gives an overview of machine learning applications to ecological modelling, focussing on applications of symbolic machine learning. It

* Fax: +386-1-4251038.
E-mail address: saso.dzeroski@ijs.si (S. Džeroski).

0304-3800/01/$ - see front matter
PII: S0304-3800(01)00312-X

gives more detailed accounts of several such applications. Symbolic machine learning methods are not described in detail but rather illustrated through the presented applications.

2. Applications in modelling population dynamics

Ecological modelling is concerned with the development of models of the relationships among members of living communities and between those communities and their abiotic environment. These models can then be used to better understand the domain at hand or to predict the behavior of the studied communities and thus support decision making for environmental management. Typical modelling topics are population dynamics of several interacting species and habitat suitability for a given species (or higher taxonomic unit).

Population dynamics studies the behavior of a given community of living organisms (population) over time, usually taking into account abiotic factors and other living communities in the environment. For example, one might study the population of phytoplankton in a given lake (Todorovski et al., 1998) and its relation to water temperature, concentrations of nutrients/pollutants (such as nitrogen and phosphorus) and the biomass of zooplankton (which feeds on phytoplankton). The modelling formalism most often used by ecological experts is the formalism of differential equations, which describe the change of state of a dynamic system over time. A typical approach to modelling population dynamics is as follows: an ecological expert writes a set of differential equations that capture the most important relationships in the domain. These are often linear differential equations. The coefficients of these equations are then determined (calibrated) using measured data.

Relationships among living communities and their abiotic environment can be highly nonlinear. Population dynamics (and other ecological) models have to reflect this to be realistic. This has caused a surge of interest in the use of techniques such as neural networks for ecological modelling (Lek and Guegan, 1999). Measured data are used to train a neural network which can then be used to predict future behavior of the studied population. In this fashion, population dynamics of algae (Recknagel et al., 1997), aquatic fauna (Schleiter et al., 1999), fish (Brosse et al., 1999), phytoplankton (Scardi and Harding, 1999) and zooplankton (Aoki et al., 1999) — among others — have been modelled.

Symbolic machine learning has also been used to model population dynamics. Systems for discovery of differential equations have proved most useful in this respect (Džeroski et al., 1999b), since differential equations are the prevailing formalism used for ecological modelling. Algal growth has been modelled for the Lagoon of Venice (Kompare and Džeroski, 1995; Kompare et al., 1997b) and the Slovenian Lake of Bled (Kompare et al., 1997a), as well as phytoplankton growth for the Danish Lake Glumsoe (Todorovski et al., 1998).

2.1. Case study: modelling algal growth in the Lagoon of Venice

The beautiful and shallow Lagoon of Venice is under heavy pollution stress due to agricultural activities (use of fertilizers) on the neighboring mainland. Pollutants are food (nutrients) for algae, which have on occasions grown excessively to the point of suffocating themselves, then decayed and caused unpleasant odors (noticed also by the tourists). Models of algal growth are needed to support environmental management decisions and answer questions such as: 'would a reduction in the use of phosphorus-rich fertilizers reduce algal growth?'

Kompare and Džeroski (1995) and Kompare et al. (1997b) use regression trees and equation discovery to model the growth of the dominant species of algae (*Ulva rigida*) in 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. Fig. 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.

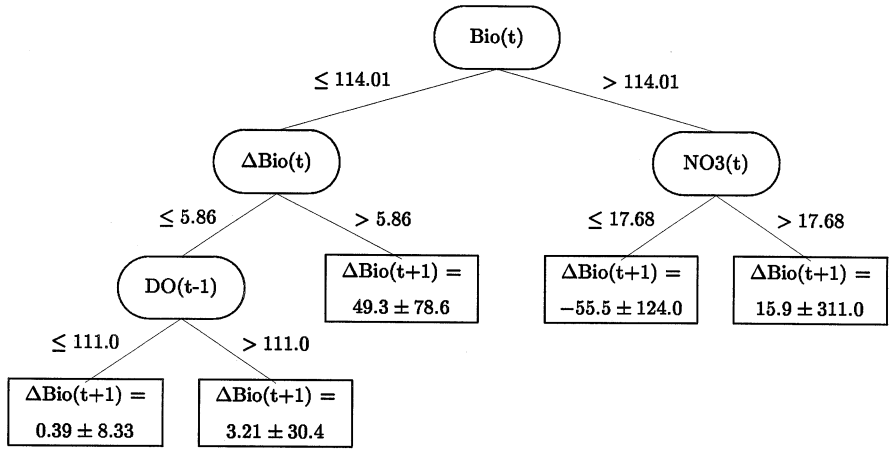

Fig. 1. A regression tree for predicting algal growth, i.e. change in biomass. $Bio(t)$, $DO(t)$ and $NO_3(t)$ stand for the concentrations of biomass, dissolved oxygen and nitrates at time t. The change in biomass $\Delta\text{Bio}(t)$ is defined as $\text{Bio}(t) - \text{Bio}(t-1)$.

Severe problems of data quality were encountered in this application.

1. Dissolved oxygen, for example, was measured at the water surface approximately at noon (when oxygen is produced by photosynthesis and is plentiful) and does not reveal potential anoxic conditions (which might occur at night) — which it was supposed to reveal.
2. Measurement errors of algal biomass were estimated to be quite large by the domain experts (up to 50% relative error).
3. Finally, winds were not taken into account: these might move algae away from the sampling stations and cause huge variations in the observed biomass values.

2.2. Case study: phytoplankton growth in Lake Glumsoe

The shallow Lake Glumsoe is situated in a sub-glacial valley in Denmark. It has received mechanically-biologically treated waste water, as well as non-point source pollution due to agricultural activities in the surrounding area. High concentration of pollutants (food for phytoplankton) lead to excessive growth of phytoplankton and consequently no submerged vegetation, due to low transparency of the water and oxygen deficit (anoxia) at the bottom of the lake. It was thus important to have a good model of phytoplankton growth to support environmental management decisions.

We used machine learning methods for the discovery of differential equations (Džeroski et al., 1999b) to relate phytoplankton (*phyt*) growth to water temperature (*temp*), nutrient concentrations (nitrogen (*nitro*) and phosphorus (*phosp*)) and zooplankton concentration (*zoo*) (Todorovski et al., 1998). Some elementary knowledge on population dynamics modelling was taken into account during the discovery process. This domain knowledge tells us that a term called Monod's term, which has the form $Nutrient/(Nutrient + constant)$ is a reasonable term to be expected in differential equations describing the growth of an organism that feeds on *Nutrient*. It describes the saturation of the population of organisms with the nutrient.

The discovered model is given in Table 1. Here $\dot{p}hyt$ denotes the rate of change of phytoplankton concentration. The model reveals that phosphorus is the limiting nutrient for phytoplankton growth, as it includes a Monod term with phosphorus as a nutrient. This model made better predictions than a linear model, which has the form

$$\dot{p}hyt = -5.41 - 0.0439 \cdot phyt - 13.5 \cdot nitro - 38.2 \cdot zoo + 93.9 \cdot phosp + 3.20 \cdot temp.$$

It was also more understandable to domain experts: the first term describes phytoplankton

Table 1
A model for phytoplankton growth in Lake Glumsoe

$$\dot{phyt} = 0.553 \cdot temp \cdot phyt \cdot \frac{phosp}{0.0264 + phosp} - 4.35 \cdot phyt - 8.67 \cdot phyt \cdot zoo$$

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 necessary to generate enough data for discovering differential equations.
2. Data quality: ecological experts often have poor understanding of modelling concepts, which strongly influences the way data are collected. An electrical engineer with knowledge of control theory would know much better that sampling frequency has to be increased at times when the system under study has faster dynamics (e.g. at peaks of phytoplankton growth).
3. The need for taking into account domain knowledge during the machine learning process: this can compensate to a certain extent for poor data quality and quantity (as was the case in this application). This issue is of great importance, yet few machine learning methods allow for the provision of domain knowledge by experts.

2.3. Case study: modelling the interactions of a red deer population

Here we studied the interactions among a population of red deer and new forest growth in a natural regenerated forest in Slovenia. Ideally, foresters would like to keep in balance the size of the deer population and the rate of regeneration of the forest: if the deer population is large, so are the browsing rates of new forest growth and regeneration slows down. Understanding the relationship between the two is crucial for managing the balance. Our study has shown that meteorological parameters strongly influence this relationship and have to be taken into account.

A preliminary study using regression trees to model the interactions was performed by Stankovski et al. (1998). Here we summarize the results of a follow-up study that used a slightly larger dataset, cleaner data, and more reliable methods of regression tree induction (Debeljak et al., 1999). The induced models show that the degree of browsing for maple (the preferred browse species of red deer) depends directly on the size of the population. The degree of beech browsing, on the other hand, was most strongly influenced by meteorological parameters, i.e. winter monthly quantity of precipitation (snow) and average monthly minimal diurnal air temperature (cf. Fig. 2). While beech is not the preferred browse species of red deer, it is consumed yearlong; it is also elastic and snow-resistant and thus more exposed to the reach of red deer even in deeper snow.

The following issues were raised by this application:

1. Data quantity: the size of the deer population and browsing rates are only estimated once a year. Even though we were dealing with 18 years worth of data, these were still only 18 data points.

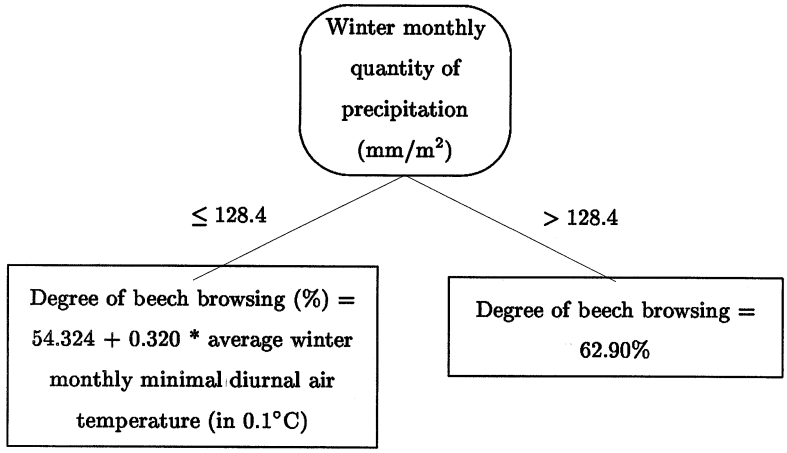

Fig. 2. A regression tree for predicting the degree of beech browsing.

2. Data quality: some of the data collected in this domain were unreliable and had to be cleaned/corrected/removed before obtaining reasonable results.
3. Missing information: the outcome of the data analysis process suggested that measuring winter and summer browsing rates separately would greatly improve the models. This information was not measured and it couldn't be reconstructed from the currently measured data, but should be measured in the future.

3. Applications in habitat-suitability modelling

Habitat-suitability modelling is closely related to population dynamics modelling. Typically, the effect of the abiotic characteristics of the habitat on the presence, abundance or diversity of a given taxonomic group of organisms is studied. For example, one might study the influence of soil characteristics, such as soil temperature, water content, and proportion of mineral soil on the abundance and species richness of *Collembola* (springtails), the most abundant insects in soil (Lek-Ang et al., 1999). The study uses neural networks to build a number of predictive models for collembolan diversity. Another study of habitat suitability modelling by neural networks is given by Ozesmi and Ozesmi (1999).

Several habitat-suitability modelling applications of other data mining methods are surveyed by Fielding (1999b). Fielding (1999a) applies a number of methods, including discriminant analysis, logistic regression, neural networks and genetic algorithms, to predict nesting sites for golden eagles. Bell (1999) uses decision trees to describe the winter habitat of pronghorn antelope. Jeffers (1999) uses a genetic algorithm to discover rules that describe habitat preferences for aquatic species in British rivers.

The author has been involved in a number of habitat suitability studies using rule induction and decision trees. Rule induction was used to relate the presence or absence of a number of species in Slovenian rivers to physical and chemical properties of river water, such as temperature, dissolved oxygen, pollutant concentrations, chemical oxygen demand, etc. (Džeroski and Grbović, 1995). Regression trees were used to study the influence of soil characteristics, such as soil texture, moisture and acidity on the abundance (total number of individuals) and diversity (number of species) of *Collembola* (springtails) (Kampichler et al., 2000). We have also used decision trees to model habitat suitability for red deer in Slovenian forests using GIS data, such as elevation, slope, and forest composition (Debeljak et al., 2001). Finally, decision trees that model habitat suitability for brown bears have been induced from GIS data and data

on brown bear sightings (Kobler and Adamič, 1999). The model has then been used to identify the most suitable locations for the construction of wildlife bridges/underpasses that would enable the bears to safely cross the highway passing through the bear habitat.

4. Applications in 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. To this end, it performs environmental monitoring, i.e. 'periodic or continuous surveillance or testing to determine the level of compliance with statutory requirements and/or pollutant levels in various media or in humans, plants, and animals' (US EPA Terms, 2000).

Given this context, environmental protection includes, e.g. biological and chemical monitoring of river water quality, which further includes regular sampling (field work) and analysis/interpretation (typically laboratory work) of the samples in terms of, e.g. water quality classes. It also includes testing chemical compounds for toxicity and biodegradability. Finally, it includes the study of effects of various pollutants on the health of the population in a given region (environmental epidemiology).

4.1. Environmental monitoring

Several machine learning methods have been used to interpret and classify samples of river water into quality classes. Walley et al. (1992) and Ruck et al. (1993) used Bayesian methods and neural networks respectively to classify river water quality, and Walley and Džeroski (1996) compared Bayesian classification, neural networks and regression trees to classify biological samples taken from British rivers. Džeroski and Grbović (1995) apply rule induction to classify biological, as well as chemical, samples taken from Slovenian rivers in terms of water quality classes. Walley et al. (2000) use unsupervised neural networks to diagnose river quality from biological and environmental data.

4.2. Case study: from biological communities to chemical properties of river water

Physical and chemical properties give a specific picture of river water quality at a particular point in time, while the biota (living organisms) act as continuous monitors and give a more general picture of water quality over a period of time. This has increased the relative importance of biological methods for monitoring water quality. The problem of inferring the chemical properties from the biota is practically relevant, especially in 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 Slovenian rivers collected through the monitoring program of the Hydrometeorological Institute of Slovenia (Džeroski et al., 2000). Pairs of biological and chemical samples that were taken at the same site at approximately the same time were used: there were 1061 such pairs, collected over 6 years. Data on biological samples list all the species/taxa present at the site and their abundances. Chemical samples contain the measured values of 16 physical an chemical parameters: biological oxygen demand (BOD), chlorine concentration (Cl), CO_2 concentration, electrical conductivity, chemical oxygen demand COD ($K_2Cr_2O_7$ and $KMnO_4$), concentrations of ammonia (NH_4), NO_2, NO_3 and dissolved oxygen (O_2), alkalinity (pH), PO_4, oxygen saturation, SiO_2, water temperature, and total hardness.

We used regression tree induction to learn predictive models for each of the 16 parameters (Džeroski et al., 2000) separately. The models for the most important indicators of pollution (ammonia, biological oxygen demand, chemical oxygen demand) had the best predictive power. We also used clustering trees (Blockeel et al., 1998) to predict the values for all 16 parameters at the same time (Blockeel et al., 1999): this actually improved the accuracy as compared to individual predictions for each of the 16 parameters.

An example clustering tree is shown in Fig. 3. In

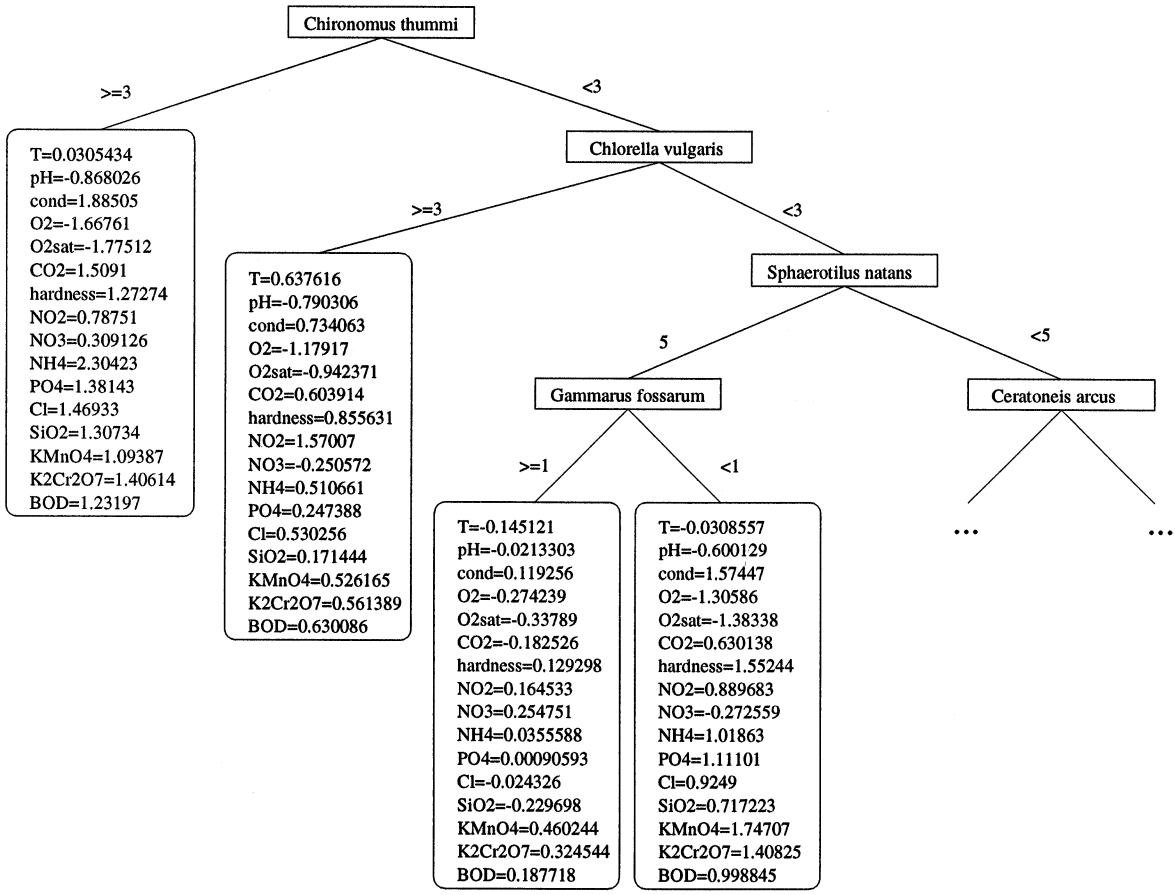

Fig. 3. A clustering tree for simultaneous prediction of multiple chemical parameters for biological data.

the leftmost leaf (Chironomus tummi ≥ 3), highly increased values are predicted for NH_4 (2.3 S.D. above the average), as well as for BOD, COD ($K_2Cr_2O_7$ and $KMnO_4$), SiO_2, Cl, PO_4, CO_2, and conductivity. On the other hand, highly decreased values are predicted for dissolved oxygen concentration and oxygen saturation (1.66 and 1.77 S.D. below average). This indicates heavy pollution and is consistent with expert knowledge, as Chironomus tummi is an indicator of heavily polluted waters.

The following issues were raised in this application:

1. Varying length data records: biological samples list all the species present. Depending on the site and water quality, the number of taxa present can vary. Methods for handling structural information or careful feature selection are thus needed.
2. Aggregating data: we used detailed data, where organisms were identified to species level, and aggregated data, where species from the same family were grouped together. Domain knowledge on the taxonomy of river water organisms was used.
3. Making multiple predictions: most machine learning methods for prediction only deal with one target variable. In many cases, however, it might be beneficial to try to predict several interrelated variables simultaneously.

4.3. Environmental impact of chemicals

Large numbers of hazardous organic chemicals

are emitted into the environment from anthropogenic and natural sources. Extensive tests of the impact of chemicals are expensive. For example, to test chemicals for carcinogenicity (the capacity to cause cancer), trials are typically performed on rodents (long term rodent bioassays) that take several years and hundreds 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 (Peijnenburg and Damborsky, 1996).

Typically, linear regression would be used to develop a so-called quantitative structure-activity relationship (QSAR) model. A domain expert chooses the features that are relevant to describe the group of compounds studied (which is often very small). As the group of compounds increases and diversifies in terms of chemical structure, this traditional approach becomes less and less appropriate.

A number of machine learning methods has been used to derive QSAR models for mutagenicity, carcinogenicity and biodegradability. Here methods of inductive logic programming (ILP) are suitable, since they can use structural information and existing domain knowledge. ILP has been used to build a predictive theory for mutagenicity, the capacity to cause genetic change which is relevant to carcinogenicity (Srinivasan et al., 1996). A set of structurally diverse aromatic and heteroaromatic nitro-compounds (some of which are present in car exhaust gasses) was studied, and a new structural alert for high mutagenicity was discovered. ILP was also used to predict the carcinogenicity of a diverse set of chemical compounds (Srinivasan et al., 1997), yielding the most accurate predictor that did not use data from biological tests on rodents in an open competition conducted within the US National Toxicology Program.

ILP has been also used to predict the biodegradability of compounds (Džeroski et al., 1999a, see below). Neural networks have been used to develop QSAR models for predicting biodegradability of organic contaminants in soil systems (Govind et al., 1996). Damborsky et al. (1996) use clustering in combination with linear regression to develop models for the dehalogenation of haloaliphatic compounds. Finally, rule induction was used by Gamberger et al. (1996) to develop biodegradation models for two sets of chemicals.

Applications of machine learning methods to relate exposure to pollution and human health are also starting to appear. Kontić and Džeroski (1997) study the influence of exposure to polluted air (as a consequence of coal mining) and other environmental/social factors on acute respiratory diseases in children in Slovakia. Rajkumar et al. (2000) use neural networks to assess health risk through inhalation exposure to benzene from vehicular emissions (car exhaust gasses).

4.4. Case study: predicting the biodegradability of compounds

We used a database of 328 structurally diverse and widely used (commercial) chemicals described in a handbook of degradation rates. Complete data on the structure of the chemicals (SMILES notation) was available, as well as data on the overall, biotic and abiotic degradation rates in four environmental compartments (soil, air, surface water and ground water). We built models for biotic degradation in surface water, predicting the logarithm of the half-life time of aqueous biodegradation. Half-life times were measured for some compounds and estimated by experts for others: in the latter case, an upper and a lower bound were given and we took the arithmetic mean of these.

We used several propositional and ILP methods for decision tree, regression tree and rule induction. In addition to a few global features, such as molecular weight, the main information used for learning was the data on the structure of compounds, i.e. the atoms within a molecule and the connections/bonds between them. Domain knowledge about a variety of functional groups and substructures was used. ILP systems use this data directly, while propositional systems use features derived from it, which represent the compounds' structure approximately, but not completely. Several of the derived models perform better than a state-of-the-art biodegradability prediction system based on linear regression.

An example rule for predicting biodegradability is given in Table 2. Note that this rule is relational, since it makes use of the relations 'contains' be-

Table 2
A relational rule for predicting the biodegradability of a compound

A compound M degrades fast IF
M contains an atom A1 AND
Atom A1 is a nitrogen atom AND
Atom A1 is connected to atom A2 with bond B AND
Bond B is an aromatic bond AND
The molecular weight of M is less than 110 units AND
The log *P* value (hydrophobicity) of M is positive.

tween a compound and its components (in this case an atom) and 'is connected to' between atoms.

Two important issues were raised here:

1. The need to handle structural information, i.e. information on the structure of chemicals. The natural representations of chemical structures are not straightforward to squeeze into a fixed-width-table.
2. The need for prior/domain knowledge: chunks of knowledge defining functional groups and substructures are essential for good performance.

ILP methods provide facilities for using both types of information directly.

5. Conclusion

Symbolic methods have been successfully applied to many problems in environmental science, engineering and management. Of the case studies presented here, the results of applying symbolic machine learning methods are most directly relevant to practice in the red deer population modelling case.

These applications have also raised a number of issues of concern to developers of symbolic machine learning methods. These include problems of data quantity (handling large datasets, generalize from small datasets), data quality (handling missing data, noisy data), and handling nonstandard learning tasks (where the data do not necessarily reside in a single fixed-width table, e.g. information on chemical structures, images). Especially important is the problem of using existing domain knowledge in the learning process.

As demonstrated by the case studies presented, models constructed by symbolic machine learning can be inspected, modified, used and verified by human experts. They have the potential to become part of the knowledge in the respective application domain. Symbolic machine learning methods thus hold much promise for further applications in ecological modelling.

References

Aoki, I., Komatsu, T., Hwang, K., 1999. Prediction of response of zooplankton biomass to climatic and oceanic changes. Ecol. Model. 120 (2–3), 261–270.

Bell, J.F., 1999. Tree based methods. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, pp. 89–105.

Blockeel, H., De Raedt, L., Ramon, J., 1998. Top-down induction of clustering trees. In: Proc. 15th International Conference on Machine Learning. Morgan Kaufmann, San Francisco, CA, pp. 55–63.

Blockeel, H., Džeroski, S., Grbović, J., 1999. Simultaneous prediction of multiple chemical parameters of river water quality with TILDE. In: Proc. Third European Conference on Principles of Data Mining and Knowledge Discovery. Springer, Berlin, pp. 15–18.

Brosse, S., Guegan, J.-F., Tourenq, J.-N., Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 120 (2–3), 299–311.

Damborsky, J., Manova, K., Kuty, M., 1996. A mechanistic approach to deriving quantitative structure biodegradability relationships. A case study: dehalogenation of haloaliphatic compounds. In: Peijnenburg, W., Damborsky, J. (Eds.), Biodegradability Prediction. Kluwer, Dordrecht, pp. 75–92.

Debeljak, M., Džeroski, S., Adamič, M., 1999. Interactions among the red deer (*Cervus elaphus*, L.) population, meteorological parameters and new growth of the natural regenerated forest in Snenik, Slovenia. Ecol. Model. 121 (1), 51–61.

Debeljak, M., Džeroski, S., Jerina, K., Kobler, A., Adamič, M., 2001. Habitat suitability modelling of red deer (Cervus elaphus, L.) in South-Central Slovenia with classification trees. Ecol. Model. 138 (1–3), 321–330.

Džeroski, S., Grbović, J., 1995. Knowledge discovery in a water quality database. In: Proc. First International Conference on Knowledge Discovery and Data Mining. AAAI Press, Menlo Park, CA, pp. 81–86.

Džeroski, S., Blockeel, H., Kompare, B., Kramer, S., Pfahringer, B., Van Laer, W., 1999a. Experiments in predicting biodegradability. In: Proc. Ninth International Conference on Inductive Logic Programming. Springer, Berlin, pp. 80–91.

Džeroski, S., Todorovski, L., Bratko, I., Kompare, B., Križman, V., 1999b. Equation discovery with ecological applications. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, pp. 185–207.

Džeroski, S., Grbović, J., Demšar, D., 2000. Predicting chemical parameters of river water quality from bioindicator data. Appl. Intell. 13, 7–17.

Fielding, A.H., 1999a. An introduction to machine learning methods. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, pp. 1–35.

Fielding, A.H. (Ed.), 1999. Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht.

Gamberger, D., Sekusak, S., Medven, Z., Sabljic, A., 1996. Application of artificial intelligence in biodegradation modelling. In: Peijnenburg, W., Damborsky, J. (Eds.), Biodegradability Prediction. Kluwer, Dordrecht, pp. 27–40.

Govind, R., Lei, L., Tabak, H., 1996. Development of structure-biodegradability relationships (SBRs) for estimating half-lifes of organic contaminants in soil systems. In: Peijnenburg, W., Damborsky, J. (Eds.), Biodegradability Prediction. Kluwer, Dordrecht, pp. 115–138.

Jeffers, J.N.R., 1999. Genetic algorithms I. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, pp. 107–121.

Kampichler, C., Džeroski, S., Wieland, R., 2000. The application of machine learning techniques to the analysis of soil ecological data bases: relationships between habitat features and Collembola community characteristics. Soil Biol. Biochem. 32, 197–209.

Kobler, A., Adamič, M., 1999. Brown bears in Slovenia: identifying locations for construction of wildlife bridges across highways. In: Proc. Third Int. Conf. on Wildlife Ecology and Transportation. Florida Department of Transportation, Tallahassee, FL, pp. 29–38.

Kompare, B., Džeroski, S., 1995. Getting more out of data: automated modelling of algal growth with machine learning. In: Proc. International Conference on Coastal Ocean Space Utilization. University of Hawaii, Hawaii, pp. 209–220.

Kompare, B., Džeroski, S., Karalič, A., 1997a. Identification of the Lake of Bled ecosystem with the artificial intelligence tools M5 and FORS. In: Proc. Fourth International Conference on Water Pollution. Computational Mechanics Publications, Southampton, pp. 789–798.

Kompare, B., Džeroski, S., Križman, V., 1997b. Modelling the growth of algae in the Lagoon of Venice with the artificial intelligence tool GoldHorn. In: Proc. Fourth International Conference on Water Pollution. Computational Mechanics Publications, Southampton, pp. 799–808.

Kontić, B., Džeroski, S., 1997. Perspective of machine learning in epidemiological studies. In: Proc. International Symposium on Environmental Epidemiology in Central and Eastern Europe. International Institute for Rural and Environmental Health, Bratislava, Slovakia, pp. 27–30.

Lek, S., Guegan, J.F., Guest Editors, 1999. Application of artificial neural networks in ecological modelling. Special issue of Ecol. Model. 120 (2–3).

Lek-Ang, S., Deharveng, L., Lek, S., 1999. Predictive models of collembolan diversity and abundance in a riparian habitat. Ecol. Model. 120 (2–3), 247–260.

Ozesmi, S.L., Ozesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecol. Model. 116 (1), 15–31.

Peijnenburg, W., Damborsky, J. (Eds.), 1996. Biodegradability Prediction. Kluwer, Dordrecht.

Rajkumar, T., Guesgen, H.W., Robinson, S., Fisher, G.W., 2000. A new dose model for assessment of health risk due to contaminants in air. J. Air Waste Manage. Assoc. 50 (1), 3–20.

Recknagel, F., French, M., Harkonen, P., Yabunaka, K., 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Model. 96 (1–3), 11–28.

Ruck, B.M., Walley, W.J., Hawkes, H.A., 1993. Biological classification of river water quality using neural networks. In: Rzevski, G., Pastor, J., Adey, R.A. (Eds.), Applications of Artificial Intelligence in Engineering: Applications and Techniques, vol. 2. Computational Mechanics Publications, Southampton, pp. 361–372.

Scardi, M., Harding, L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120 (2–3), 213–223.

Schleiter, I.M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K.-D., Schmidt, H.-H., Werner, H., 1999. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. Ecol. Model. 120 (2–3), 271–286.

Srinivasan, A., Muggleton, S.H., King, R.D., Sternberg, M.J.E., 1996. Theories for mutagenicity: a study of first–order and feature based induction. Artif. Intell. 85, 277–299.

Srinivasan, A., King, R.D., Muggleton, S.H., Sternberg, M.J.E., 1997. Carcinogenesis prediction using inductive logic programming. In: Lavrač, N., Keravnou, E., Zupan, B. (Eds.), Intelligent Data Analysis in Medicine and Pharmacology. Kluwer, Dordrecht, pp. 243–260.

Stankovski, V., Debeljak, M., Bratko, I., Adamič, M., 1998. Modelling the population dynamics of red deer (Cervus elaphus L.) with regard to forest development. Ecol. Model. 108 (1–3), 145–153.

Todorovski, L., Džeroski, S., Kompare, B., 1998. Modelling and prediction of phytoplankton growth with equation discovery. Ecol. Model. 113, 71–81.

US EPA Terms, 2000. Terms of the Environment. Document order number EPA175B97001, National Service Center for Environmental Publications. Also available at http://www.epa.gov/OCEPAterms/

Walley, W.J., Džeroski, S., 1996. Biological monitoring: A comparison between Bayesian, neural and machine learning methods of water quality classification. In: Proc. International Symposium on Environmental Software Systems, 1995. Chapman and Hall, London, pp. 229–240.

Walley, W.J., Hawkes, H.A., Boyd, M., 1992. Application of Bayesian inference to river water quality surveillance. In: Grierson, D.E., Rzevski, G., Adey, R.A. (Eds.), Applications of Artificial Intelligence in Engineering. Computational Mechanics Publications, Southampton, pp. 1030–1047.

Walley, W.J., Martin, R.W., O'Connor, M.A., 2000. Self-organizing maps for the classification and diagnosis of river quality from biological and environmental data. In: Denzer, R., Swayne, D.A., Purvis, M., Schimak, G. (Eds.), Environmental Software Systems: Environmental Information and Decision Support. Kluwer, Dordrecht, pp. 27–41.