Chapter XXI

Metadata-Supported Automated Ecological Modelling

Virginia Brilhante
University of Edinburgh and University of Amazonas

Dave Robertson
University of Edinburgh

INTRODUCTION

Ecological models should be rooted in data derived from observation, allowing methodical model construction and clear accounts of model results with respect to the data. Unfortunately, many models are retrospectively fitted to data because in practice it is difficult to bridge the gap between concrete data and abstract models. Our research is on automated methods to support bridging this gap. The approach proposed consists of raising the data level of abstraction via an ecological metadata ontology and from that, through logic-based knowledge representation and inference, to automatically generate prototypical partial models to be further improved by the modeler.

In this chapter we aim to: 1) give an overview of current automated modelling approaches applied to ecology, and relate them to our metadata-based approach under investigation; and 2) explain and demonstrate how it is realized using logic-based formalisms.

We give the overview of current automated modelling approaches in the section "Ecological Modeling and Automation: Current Approaches," focusing on compositional modelling and model induction. The contrast between these and our approach, where we adopt metadata descriptions through an ontology and logic-based modelling, is discussed in the section "Our Automated Ecological Modelling Avenue." The next section, "Towards a System for Metadata–Supported Automated Modeling," makes ideas more concrete, starting with further details on the Ecolingua ontology, followed by examples of automated model structuring and parameter estimation. In the concluding section, "A Look Ahead and Conclusion," we comment briefly on the ontologies trend and on the outlook of our research.

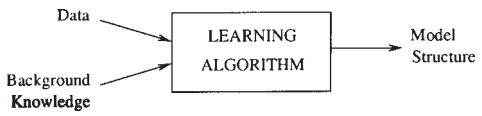
ECOLOGICAL MODELLING AND AUTOMATION: CURRENT APPROACHES

Ecological models are a widely used and powerful tool in environmental studies. Through the design of such models, modelers are able to give shape to their understanding of ecological systems and thus discover possible knowledge gaps. Once built, models can be used for ecological simulation, giving support to prediction and decision making. Initiatives exist to make accessible the large body of knowledge generated by ecological modelling efforts worldwide. The Register of Ecological Models, REM, is a comprehensive repository of ecological models, equipped with search facilities, available at http://dino.wiz.uni-kassel.de/ecobas.html (Benz and Voigt, 1996; Benz and Knorrenschild, 1997). The models in REM are documented under a unifying documentation system called ECOBAS (Hoch et al., 1998).

Automating the construction of models in general is the ambition of the automated modelling field. Modelling processes of various disciplines are investigated and the modelling knowledge is explicitly represented, aiming at developing computer tools able to automatically follow modelling principles (Xia and Smith, 1996). Compositional modelling, the dominant automated modelling approach, consists of assembling basic model fragments, usually gathered in a library, into a well-defined model. In Rickel and Porter (1997), we find an application of compositional modelling to an ecology-related domain, namely plant physiology, chosen to demonstrate automated modelling of complex systems. The domain knowledge used by TRIPEL, the automated modelling system, is part of a large multipurpose biology knowledge base. The core of the system is an algorithm that given a prediction (what if) question, the variables of the physical system, the influences among them, and other domain knowledge, outputs the simplest differential equation that can adequately answer the prediction question.

In the model induction approach, investigated within the machine learning field, models are derived from data. Figure 1 summarizes the machine learning task viewed from a modelling perspective. Data, together with background knowledge, are used to provide examples from which a general model structure can be induced (learned) and described in some language such as if-then rules, equations, decision and regression trees.

Figure 1. The Machine Learning task



One important motivation for this form of knowledge induction is that it is a promising alternative for ill-defined or poorly understood problems that lack algorithmic solutions (Kubat et al., 1997), characteristic of the environmental sciences. In other words, given that we fail to formally state the problem and all we have are factual examples, machine learning attempts to extract concepts that are embedded in the examples.

The series (Kompare et al., 1994; Dzeroski et al., 1997) describes the use of machine learning techniques on automated modelling of ecological systems. The first paper discusses early results of machine learning systems running over ecological