Decision and Learning Model Selection for Complex Adaptive Systems

Tei Laine (tei@iki.fi)
Department of Computer Science, Gustaf Hällströmin katu 2b
00014 University of Helsinki FINLAND

Keywords: Model selection; agent-based modeling; decision making; minimum description length principle.

Introduction

Computer modeling is gaining popularity in study of systems whose underlying processes are hard to measure directly, or controlled experimentation is impossible. Since many real-world phenomena, for instance psychological or ecological, are often complicated, and the models trying to capture their essence relatively complex, selecting the best model from the candidates is a challenge. In this presentation I address model selection in the context of complex adaptive systems, and particularly among classes of models that are used to gain understanding in the most significant single factor behind the global climate change, namely human land-use. In order to understand the impact of the land-use change, not only its consequences but also the underlying mechanisms and socio-economical, political, psychological, and historical factors driving the change need to be explained (Parker, Manson, Janssen, Hoffman, & Deadman, 2003).

I focus on *agent-based models* of human learning and decision making in the domain of land-use, and propose a criterion to select between these models. The candidate models constitute a set of relatively straightforward reinforcement-based strategies familiar from psychology and economics (Camerer & Ho, 1999; Watkins & Dayan, 1992).

Model Selection Framework

What distinguishes agent-based land-use and land-cover change models from more traditional cognitive models is that they are often validated against land-use data rather than experimental human data. Furthermore, model validation and subsequent selection is challenged by the fact that usually not many data samples are readily available for adequate generalization tests. Finally, the class of agent-based land-use change models does not easily lend itself to probabilistic interpretation, but can be best characterized a complex adaptive system — a system consisting of multiple autonomous components and processes that interact at multiple spatial levels and temporal scales.

For these reasons, traditional methods such as AIC (Akaike, 1973), BIC (Schwarz, 1978), or cross-validation (Lendasse, Wertz, & Verleysen, 2003) do not apply. The void left by these methods is filled with a selection criterion proposed in this presentation. The method is based on a practical interpretation of the *Minimum Description Length Principle* (Grünwald, 2000) first introduced in Laine (2006). The method makes the following three assumptions: first, no 'true' model exists; secondly, the complexity of the model class is based on the performance of the models belonging to the class, not some predetermined structural property; and thirdly, the model itself does not determine its fit to data, but an error function is required.

While the last two points address the trade-off between goodness-of-fit and the model class complexity, the first one takes a more ideological standpoint on what is tried to achieve with the model selection criterion, namely that the goal is to find the best model to explain the data rather than a model that approximates some 'true' state of the world, which cannot necessarily be verified.

Experimental Results

The proposed selection criterion is tested with an extensive set of artificial data in multiple experimental conditions varying the agent and the landscape characteristics, and a representative case of real land-cover change data.

The major findings are: the selection criterion tends to select the generating class if it is among the candidates, it prefers model classes with fewer free parameters, and finally, the error function plays an equally pivotal role as the selection criterion in finding the best model class.

Acknowledgments

This study was funded by the Biocomplexity grant (NSF SES0083511) for the Center for the Study of Institutions, Population, and Environmental Change (CIPEC) at Indiana University, and was also supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrox & F. Caski (Eds.), *Second International Symposium on Information Theory* (p. 267-281). Akademiai Kiado, Budapest, Hungary.

Camerer, C., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827-874.

Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, *44*, 133-152.

Laine, T. (2006). *Agent-based model selection framework for complex adaptive systems*. Doctoral dissertation, Indiana University.

Lendasse, A., Wertz, V., & Verleysen, M. (2003). Model selection with cross-validation and bootstraps — application to time series prediction with RBFN models. In (p. 573-580). Berlin, Germany: Springer-Verlag.

Parker, D. C., Manson, S., Janssen, M., Hoffman, M., & Deadman, P. (2003). Multi-agent system models for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, *93*(2), 316-340.

Schwarz, G. (1978). Estimating the dimension of the model. *The Annals of Statistics*, *6*, 461-464.

Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3/4), 279-292.