In Proceedings of the Sixth International Conference on Cognitive Modeling, 410-411.
Mahwah, NJ: Lawrence Earlbaum.

Methodology for Comparing Agent-based Models of Land-use Decisions

Tei Laine (telaine@cs.indiana.edu)
Indiana University
Computer Science Department and the Cognitive Science Program
Bloomington, IN 47405-7104 USA

Abstract

The focus of the research is mainly methodological. The goal is to develop a framework for comparing computational, agent-based models for land-use decision making. The framework will allow studying spatial patterns emerging from different distributions of agent characteristics, learning and communications schemes, initial spatial configurations, and varying spatial suitabilities. The framework will also facilitate the assessment of model complexity, thus enabling a rigorous practice of model comparison and selection.

Keywords Model selection, agent-based modeling, learning, decision making, land-use

Introduction

A common problem in computational modeling of cognitive, natural or social phenomena is *validation* of the model; how can one tell the model is the right one, i.e., it captures the essential mechanisms behind the modeled empirical phenomenon? The solution is to compare model predictions to the empirical data and measure the fit between them. One can then use this measure draw conclusions about the plausibility of model assumptions.

The problem of this approach is that the fit does not tell us anything about the model's goodness with respect to unseen data or alternative models designed to explain the same phenomenon. Once the validation issue is settled, the next question is the problem of model *selection*; how can one tell the model is the best one, i.e., one that explains the target phenomenon most accurately? One way is to compare several models with each other and report their relative goodnesses. However, this approach does not necessarily treat the models fairly, since some of them may achieve their superior performance due to their inherent complexity or number of ad hoc assumptions.

This research project's goal is to design and implement a framework for comparing complex real world agent-based decision making models, which accounts for the trade-off between the quality of models' performance and the complexity with which they achieve the performance. In the next chapter an overview on study's task domain, land-use decision making, is given. Then some issues regarding model selection and model complexity are addressed. The extended abstract concludes by briefly describing the proposed modeling framework.

Land-use decision making

This research focuses on a specific type of agent-based models (Janssen, 2004); models that are designed to explain how human land-owners make land-use decisions, and to explore land-cover patterns emerging from these decisions. The decision maker's goal is to find an effective way of using his/her assets — size and quality of land, technology, education and experience — in allocation of available resources — labour and land — to different uses. There are a number of factors to be considered ranging from the suitability of the land, dictated by various bio-physical variables, to the expected monetary or non-pecuniary returns from the uses. The optimal or good decision does not solely depend on careful consideration of the above factors, but also on the decisions of neighboring owners and the use of their land. Modeling land-use decision making is challenging, because the domain is inherently multidisciplinary. It involves human factors, environmental dynamics, and a number of ecological, economic, social and institutional variables and several kinds of interactions between them. Finally, the models are used to understand human cognitive processes, but seldom in this domain does any empirical data exist about them; yet inferences need to be made from land-cover data.

Model comparison and complexity

Computational modeling is a relatively established method of studying human cognition. The philosophical and theoretical issues of designing and implementing models have been discussed (see for instance (Cooper, Fox, & Farringdon, 1996)), but the question of what should be done once the model exists is practically unaddressed. The discipline lacks efficient methods of comparing model performance and assessing their complexity.

Good scientific practice prefers simple models, which explain the target phenomenon with few well-founded theoretical principles. The goal of the modeling endeavor is to increase understanding of the underlying mechanisms of the phenomenon, not to fit data. A model that fits the data perfectly does not necessarily capture the essential mechanisms behind the modeled phenomenon, but due to excess complexity is flexible enough to fit random noise. Model complexity is defined as a characteristic of the model that makes it fit varying patterns of data (Myung & Pitt, 1997; Pitt, Myung, & Zhang, 2002).

In general, model selection methods proposed for computational modeling of cognition are rare. Pitt *et al.* (Pitt et al., 2002) propose a selection method that emphasizes the generalizability of the model, and is based on *minimum description length (MDL)* (Grünwald, 2000). The MDL measure combines the maximum likelihood factor and penalty terms that account for the number of free parameters and the functional form. Busemeyer and Wang (Busemeyer & Wang, 2000) propose a *generalization criterion* method for model selection. The criterion is related to *cross-validation*, but instead of using different data sets in calibration and validation stages, it uses different designs. Myung and Pitt (Myung & Pitt, 1997) propose a *Bayesian method* that is based on maximizing the marginal likelihood, which can be obtained by normalizing goodness of fit with model complexity term that accounts for number of parameters, their functional form and the range of parameter values.

Proposed modeling framework

Most of the techniques discussed above apply to mathematical models only, but the models of interest in this research are not necessarily easily formulated in equations. They may also introduce other potential sources of complexity that enhance the fit but not necessarily the explanatory power. Thus, the goal of this research is to develop a methodology for comparing complex agent-based models of land-use, which allows the assessment of model complexity and quality of performance along various dimensions. Within the framework the performance can be evaluated both qualitatively and quantitatively (e.g., focus on amount of changes in land-cover or the patterns of changes), and either on global or individual level (i.e., assess prediction accuracy at the level of individual decision maker or the whole modeled area). The complexity assessment can take into account the number of parameters (i.e., degrees of freedom) and the algorithmic complexity (i.e., the requirements for time and space).

The main components of the general framework are *agents*, that represent land-owners, and the *landscape*, an abstract rectangular area divided into equal-size cells. A cell is the basic unit in which land-related information is encoded and it also defines the resolution at which land-use decisions are made.

In the general modeling framework a number of characteristics can be defined for these components. Agent characteristics are the size and location of the land they own and make decisions about, size of the household, initial wealth, number of neighbors and the relationship to them, learning and decision strategy they use, the activities (land-uses) they may take part in, and finally a list of individual parameters, that give a preference for each of these activities. Landscape characteristics are the area of land, number of possible land-covers and a list of suitability parameters associated to each land-use. Exogenous parameters to the model include monetary returns from different land-uses and costs associated to changes in uses.

The specific models implemented in this framework define possible land-covers (e.g., primary forest, farmed land, residential areas) together with the initial cover, the actual available land-uses (e.g., farming, development), relationship between uses and resulting covers, agent learning and decision strategies (e.g., expected utility maximizer, reinforcement learner), and the model specific parameters related to each of these (e.g., learning rate, number of hidden units).

Summary

Within the proposed framework several models, assuming different agent and landscape characteristics, will be implemented and compared. The framework will allow assessment of model performance by measuring prediction accuracy in various resolutions of decisions and output metrics. The framework will also enable the quantification of model complexity, by taking into account both architectural and computational factors, such as degrees of freedom, and time and space requirements. Finally, the framework will facilitate understanding of the models by making explicit their intuitive and hidden assumptions.

Acknowledgments

The research is partially funded by NSF grant SES-0083511.

References

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*, 171-189.

Cooper, R., Fox, J., & Farringdon, J. (1996). A systematic methodology for cognitive modeling. *Artificial Intelligence*, *85*(1), 3-44.

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

Janssen, M. A. (2004). Agent-based modelling. In *Modelling in ecological economics.* Edward Elgar.

Myung, I. J., & Pitt, M. A. (1997). Applying occam's razor in modeling cognition: A bayesian approach. *Psychonomic Bulletin & Review*, *4*(1), 79-95.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psyhcological Review*, *109*(3), 472-491.