Agent-Based Modeling

Dan Herweg

January 2019

MADAS

Prepared for: Pietro Terna

Summary

This memo lays the groundwork for a forthcoming framework I am developing to find and evaluate use cases for agent-based modeling (ABM). This document discusses key points that differentiate ABM as a methodological technique, outlines the basic process of creating an ABM, and details how one might go about the identification of questions that are best addressed by an ABM. [The framework that emerges from this memo will be presented in my final project for Morini, and I would love to share it with you when I am finished.]

Agent-based modeling

This section will discuss what agent-based modeling (ABM) is, how this technique relates to complex adaptive systems (CAS), why this makes it a uniquely useful tool and why it is especially relevant now.

What is ABM?

An agent-based model is a bottom-up approach to modeling systems that allows the simulator to specify local, micro-level attributes of a system's individual parts and observe their simulated effects over time on the micro and macro-level behavior of a system. The phenomenon of local actions yielding macro-level characteristics is called emergence, and the relationship between the two levels is often difficult to predict or counterintuitive. Using agents to generate system characteristics at the macro level obviates the need for a causal explanation for them at the aggregate level, which we often do not have. [1]

How do ABMs capture Complex Adaptive Systems?

ABMs mirror the structure that is used to describe complex adaptive systems (CAS). CAS are characterized by many agents that act and interact over time, adapting or coevolving with the environment in a path dependent way and producing emergent features at the global level. Emergent features of systems are both the result of local actions and can influence or constrain agents, and this self-reference can help model the endogenous self-organization in systems. [2]

This also highlights that systems, as well as the computers that simulate them, are embedded in time and thus have historical contingency. This fact is not well-captured by models that might treat time as reversible. I cannot improve upon these excerpts from the book Modeling Cities and Regions as Complex Systems [2]:

> *The system state is no longer predictable by a simple law. Rather it depends on a history of choices made at bifurcation points, that is, points where one possible system state splits in two…The final state is the result of a particular contingent history, the history of our choice at every bifurcation point. Consequently, explanation by physical law must be replaced by historical explanation. (p. 19)*

And later:

> *…the contradiction-free status of math and logic is possible only because they are ideal, formal systems without time; thus the only systems that can be mapped into them are those*

> *in which time can be eliminated… evolutionary epistemology – its basic position is that the creation of knowledge is an evolutionary process, a process embedded in time, and therefore one whose description cannot be mapped into a timeless formal system like mathematics or logic… biological evolution itself is knowledge creation… the appearance of new species represents the accumulation of new knowledge about the system as well as a change in the system itself. As the system evolves, some knowledge becomes false and is eliminated through extinction. (p. 232-3)*

Why ABM?

ABMs may be seen as a response to the inadequacies of the reductionist hypothesis, which has been useful in describing the world but does not give guidance on how to construct it. [3] ABM is truer to the nature of the systems it simulates: it is full of individuals with bounded knowledge and information processing capacity who adapt to their changing environment in discrete, irreversible time steps, generated algorithmically. [4] [5] [6] [7] Reductionist techniques do not support the processing of heterogeneity the way ABM does, using aggregates to characterize systems. While aggregation makes problems more tractable, vital information is lost in summary statistics that could be vitally important to system outcomes. [7]

Additionally, ABM can use algorithmic computation to find local solutions to computationally irreducible questions. [8] [I will be leaving this issue aside for now, as I need to research it more before commenting.]

Another benefit of ABM is it's ability to complement other analysis, providing intuitive construction and data visualizations that improve communication. [1]

Finally, for systems that are closed, predictable, homogenous, or linear, the benefits of producing an ABM may compare less favorably with the costs of building such a model, which may be more expensive than other techniques. [7] [1]

Why now?

Stephen Hawking has famously said the next century will be the century of complexity. Thankfully, the feasibility of conducting analysis with ABMs to address complexity has never been greater. This is due to the prevalence of object-oriented programming, increased computational power, and the massive collection of data.

ABMs are most reasonably written in object-oriented programming languages to facilitate the efficient creation of large numbers of agents as objects. [6] The code of an ABM is generative of a much larger computational representation of a system much in the way human DNA contains the instructions for a fertilized cell to create a person (that most modelled agents are people might suggest that it is generative instructions all the way down). This highlights the impact of increased silicon computing capacity – ABM requires multiple runs with code that generates large amounts of computation. [2]

Finally, recent accumulation of large, granular data sets in the Big Data era has enabled better descriptions of agents in the wild and the impending Internet of Things promises to extend this trend by attaching sensors to traditional agents (people) as well as injecting "agency" into traditionally non-agentic things like refrigerators by giving them the ability to interact with the world around them and anticipate, learn and predict based on machine learning techniques.

Making an agent-based model

This section provides a general outline for the steps of writing an ABM: Creation, Verification, Validation and Replication.

Creation

An ABM should begin with the goal of building the simplest model that captures the behavior of the target system. [1] To begin to make an ABM, one should set out to code agents, environment, actions and interactions. Here is a set of initial questions to consider when building out each component:

COMPONENT	KEY QUESTIONS
AGENTS	What types of agents exist? What characteristics (states, knowledge, etc.) do the agents have? What are their goals and strategies? Do they learn or anticipate based on the past? How do these dimensions change over time?
ENVIRONMENT	What are the relevant variables to characterize the environment? How do these dimensions change over time?
ACTIONS	What are the behaviors and rules of behavior? How are these actions ordered in time?
INTERACTIONS	How are agents connected? What do agents know about other agents locally and globally? How do agents interact with other agents? How do they change their behavior in response to local and global agent populations? How are the different pieces of the environment connected? How do agents interact with the environment? What do agents know about the environment locally and globally? How does the environment change in response to agents acting upon it?

Getting values for parameters that are unknown to the modeler can come from academic research, interviews with subject matter experts or practitioners, observation, or industry heuristics. A final consideration might be to code the model such as to collect the data that is relevant to your question.

Verification

An ABM should then be verified. Verification is ensuring the computational model faithfully represents the target conceptual model. [1] This means checking the code does what is intended at multiple stages of execution and under a range of parameter values to find unexpected results that require rework.

Validation

> *A different kind of science requires a different methodology, with different methodological standards. Instead of a few simple, elegant experiments, we will need many messy tests. And, instead of simple but strict standards, we will need multiple but loose ones [2] (p. 217)*

Validation is comparing the model to the real target system using multiple runs. What it means to validate a generative model is tricky. Often, being able to generate system level phenomena – thus proving that such a system can be generated from simple rules and initial conditions - is a success in itself. [7] However, if a systems history is characterized by bifurcations, there is more thinking to be done. Reality has a unique historical path, and there is considerable speculation in quantifying where the counterfactual roads not taken would have led. For example, when predicting a city's development from its unique initial conditions, there is only one example to validate against. A model that yields results close to reality on a fraction of its runs has shown that the present was a possible future, but the other predicted states will be more difficult to prove as valid bifurcations and not evidence of an erroneous model. Conversely, a model that generates a good result almost every time could mean that the system's past had few potential bifurcations, or that the model is overfit to the validation criteria. [2]

Ideally, validation should be done at micro and macro levels and for multiple output variables. Calibrating to multiple output variables can allow models to be judged on which aspects of the model correspond best to reality and will protect against overfitting. With increasing validation criteria, it is less likely that all will be satisfied, and trade-offs will need to be made to prioritize those most relevant to the modeler's purpose. [2] One important check might be that the parameters we have selected as important to the model impact outputs more than the sources of noise in the model. Outputs that can be validated can be patterns, proportions, distributions, values, or even that the model gives visual patterns that appear normal (such as the perception of "valid" flocking behavior in the famous boids model). [1]

Replication

Replication of an experiment helps the modeler ensure that the observed results are not artifacts of idiosyncrasies in the creation and implementation of the model. There are many layers of abstraction between the conceptual model (which is itself an abstraction!) and the results of an ABM. The assumptions and algorithms in these layers are often hidden, and it is prudent to determine whether they matter to enhance the epistemic status of the modeler's claims. Wilenski lists a non-exhaustive list of test areas of replication robustness as time, hardware, language, toolkits, algorithms, and authors. Each of these can be varied to see if the result agrees with the original model by the standard of numerical identity, distributional equivalence, or relational alignment. [1]

When and how we can use ABM

Here the application of ABMs will be discussed in more detail. First, the types of systems to which ABM are most fruitfully applied are discussed and how to characterize them. Then a partial list of the types of answers and outcomes we can expect from ABM are enumerated.

Characteristics of ABM-friendly systems

To help us determine whether ABM might be the appropriate tool to analyze a problem, proposed below are characteristics to serve as a heuristic checklist for when one might use ABM:

COMPONENT	CHARACTERISTICS
SYSTEM	Open futures, but a finite amount of factors can characterize its evolution Hard to predict in aggregate, but local processes are known or knowable Does not converge to a predictable equilibrium state, or exists in distinct regimes Has many heterogenous agents and links between them Has an environment that is topographically diverse Can reasonably be assumed to exhibit path dependence and the bifurcations are of interest The tails and outlier results of the system dominate what it is relevant to know about it and average characterizations are less important (power laws) Exhibits patterns that are of interest to the modeler
AGENTS	Interact with each other and the environment Learn or act strategically, anticipating future states from the current context Exhibit bounded rationality, locally constrained perception, use heuristics, have limited information processing capacity Are embedded spatially or in a network with nontrivial implications Are subject to selection effects
ACTIONS	Change over time and have non-linearities Are only known locally or do not intuitively scale to observed macro behaviors Are embedded in time with nontrivial implications of ordering

What to ask an ABM

When using a model to test a hypothesis, one must consider what evidence one would look to for falsification of that hypothesis. CAS and ABM have a double effect on how causality is addressed - emergence and interaction in CAS appear to break or obscure simple causality links, yet ABM allows us to essentially run simulated randomized controlled trials, the gold standard for empirically determining causality. [6]

The potential independent variables could be interventions involving any of the classes of elements considered in the coding of the model as described in the Creation subsection. Changing answers to any of the suggested questions could be coded and tested against its effect on the dependent variable.

Simulations yield rich data that can be collected in runtime, from simply “the number of agents” to “the frequency distribution of a temporal motif among agent interactions as a function of environmental parameters.” Aggregation and statistical analysis of these data give us even more permutations of potential dependent variables. One initial way to think about the space for variables of interest is the following:

1) Does it concern agents, environment, actions, interactions, or a combination?
2) Is it an individual agent, space, action or interaction? A subset? The entirety?
3) Is a value directly from the model, a relationship between variables, or is an aggregate metric needed (average, variance, distribution, proportion, etc.)?
4) Is it a longitudinal data set or cross-sectional?

Some examples of what we might look for:

- The impact of an intervention, or characteristics of an intervention to produce a given impact
- History of an instance, subset, or population of agents
- The distribution of the states, characteristics, stocks or flows
- Winning/losing strategies for agents
- Spatial patterns over time, frequency of specific patterns, and the existence of equilibrium
- Time to percolation or some other threshold
- Parameter values where percolation effects or patterns emerge
- Values within which system variables are bounded for typical inputs
- Sensitivity analysis to weigh trade-offs in achieving an effect through different means

Other motivations for agent-based modeling

While the primary goal of most modeling endeavors is prediction, the results can be used in many contexts that go beyond prediction. Even the process of creating an ABM can yield benefits for the modeling team. Epstein [8] and Wilensky [1] have enumerated these, and drawing from their examples I will illustrate a few benefits here:

- Explain a phenomenon: a better picture of causality in a system, especially one connecting local actions to global effects, may emerge in the building of the model as subject matter experts interact and make parameters explicit. Even analogies that emerge in approximating an interaction or parameter could be instructive, inspiring new insights and hypotheses
- Data collection: the unique data requirements to build an ABM may broaden the data collected by an organization, yielding new insights and questions
- Illuminate uncertainties: The modeling process may uncover gaps in knowledge to investigate

Conclusion and next steps

ABMs are a novel modeling tool, especially in the realm of complex systems where emergent effects are of most interest. Using this tool properly means understanding the epistemological and methodological issues that accompany it to build robust models on explicit assumptions with reasonable expectations.

ABMs generate rich data that can support a variety of inquiries, and practitioners should be on the lookout for new contexts where such a tool will add unique value. The next steps in this analysis will be to build a framework to systematically find and evaluate use cases for agent-based modeling.

References

[1] U. WIlensky and W. Rand, An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo, Cambridge, Massachusetts: The MIT Press, 2015.

[2] R. White, G. Engelen and I. Uljee, Modeling Cities and Regions as Complex Systems: From Theory to Planning Applications, Cambridge, Massachusetts: The MIT Press, 2015.

[3] P. W. Anderson, "More is Different," *Science,* vol. 177, no. 4047, pp. 393-396, 1972.

[4] A. Kirman, "Complexity and Economic Policy: A Paradigm Shift or a Change in Perspective? A Review Essay on David Colander and Roland Kupers's Complexity and the Art of Public Policy.," *Journal of Economic Literature,* vol. 54, no. 2, pp. 534-572, 2016.

[5] J. Axtell, "Why agents? on the varied motivations for agent computing in the social sciences," Center on Social and Economic Dynamics , 2000.

[6] R. M. M. S. M. T. P. Boero, Agent-based Models of the Economy: From Theories to Applications, New York, New York: Palgrave MacMillan, 2015.

[7] F. D. V. D. a. C. J. A. Ghorbani, "Enhancing ABM into an Inevitable Tool for Policy Analysis," *Policy and Complex Systems,* vol. 1, no. 1, pp. 61-76, 2014.

[8] J. Epstein., "Why model?," *Journal of Artificial Societies and Social Simulation,* vol. 11 (4), 2008.

[9] L. T. Paul L. Borrill, "Agent-Based Modeling: The Right Mathematics for the Social Sciences?," Iowa State University Deepartment of Economics, Ames, Iowa, 2010.