A Hybrid Sensitivity Analysis Approach for Agent-based Disease Spread Models

Laura L. Pullum
Computational Data Analytics Group
Oak Ridge National Laboratory
Oak Ridge, USA
pullumll@ornl.gov

Xiaohui Cui
Department of Computer Science
New York Institute of Technology
Manhattan, USA
xcui@nyit.edu

Keywords: Agent-based models, hybrid models, validation, disease spread model

I. INTRODUCTION

Agent-based models (ABM) have been widely deployed in different fields for studying the collective behavior of large numbers of interacting agents. Of particular interest lately is the application of agent-based and hybrid models to epidemiology, specifically Agent-based Disease Spread Models (ABDSM). Validation (one aspect of the means to achieve dependability) of ABDSM simulation models is extremely important. It ensures that the right model has been built and lends confidence to the use of that model to inform critical decisions. In this report, we describe our preliminary efforts in ABDSM validation by using hybrid model fusion technology.

II. ABM V&V APPROACHES

In general, ABM belong to a class of software sometimes referred to as “non-testable programs” and described by Weyuker [3] as “programs which were written in order to determine the answer in the first place. There would be no need to write such programs if the correct answer were known.” Since there are no oracles for these programs, it is generally impossible to know *a priori* what the correct or expected output should be for a given input. There exist many criticisms about using ABM to study complex systems [1]. Recently discussed ABM V&V approaches cite several V&V frameworks and techniques as the bases to build upon, e.g., [4]. In this paper, we categorize them as historical, predictive and sensitivity validation techniques. *Historical data validation* is a technique applied when historical data exists or can be collected. *Predictive validation* is used to compare a model’s prediction with actual system behavior. *Sensitivity analysis* is a method used to evaluate variability in the model’s parameters.

III. ABDSM VALIDATION

Some ABM V&V techniques presented in the literature may be used for ABDSM validation. For example, sensitivity analysis is often used for validating disease-spread simulations based on ABM [6, 7]. When constructing the agent rules for ABM, which affects the dynamics of the model, we may turn to experts to hypothesize about the relevant factors driving disease spread levels. Still a lack of quantitative data on epidemic levels prevents us from comparing model output to data from the real world, and limits historical and predictive validation. Therefore, using sensitivity analysis can support validation by showing whether factors have the effects they are expected to have [8]. A large parameter space is characteristic of multi-agent models, and they permit potentially large response surfaces. The challenge then becomes determining over which ranges and sets of parameters the model is capable of producing valid results. Sensitivity analysis can be applied to complex dynamic models to provide insight on how uncertainty in the input variables affect the model outputs, and which input variables tend to drive variation in the outputs. For models such as infectious disease models meant to inform decision makers, uncertainty in the output can be disconcerting as a single- valued result is not provided. However, the benefit is that a range of output values can reveal a suite of possible model outcomes [2].

In our ABDSM validation project, we conducted a new hybrid model sensitivity analysis based validation approach for ABDSM validation. The validation system integrated several sub-models, including a population model, transportation models and social network model to develop the base validation framework. Testing these various elements is analogous to unit, subsystem, and system testing as indicated in Fig. 1. Unit, subsystem, and system testing are well-known processes included in software V&V standards such as [9-11]. In the remainder of this section, we present how the sub-models are fused in our hybrid model sensitivity analysis framework.

A. *Synthetic Population Model*

When developing an ABDSM, the initial step is the definition of agents – the people. A snapshot of the entire population of the study area is needed as an initial condition of the ABDSM. Due to privacy and cost constraints, such data is often not available. To tackle this issue, current ABDSM combine different data sources to derive a disaggregate representation of the people, matching given criteria like correlation structure and marginal sums. This process is referred to as population synthesis. One feasible source for such disaggregate data is the national census that is collected on a regular basis for many countries. The generated synthetic population is a set of geographically located people and households (referred to as a proto-population), each associated with demographic variables drawn from any of the demographics available in the census. Each synthetic individual is placed in a household with other synthetic people and each household is located geographically in such a way that a census of our synthetic population yields results that are statistically indistinguishable from the original census data, if they are both aggregated to the block group level.

B. ABDSM Social Network Model

In ABDSM, social networks constrain possible agent behaviors, while agent behaviors shape the social networks. Many early ABDSM models used simple random graph techniques for their synthetic population and social network generation. It has been proven that realistic social networks in most urban regions are structurally different than synthetic networks generated using simple random processes. Current ABDSM, such as EpiSims, use real world data sources and combine them with behavioral and social theories to generate a synthetic population and social networks. The model uses a set of activity templates for households based on several thousand survey responses. These activity templates include the sort of activities each household member performs and the time of day they are performed. Thus a minute-by-minute schedule of each person's activities and the locations where these activities take place can be generated by a combination of simulation and data fusion techniques; and eventually a dynamic social contact network can capture this information.

The synthetic population model and ABDSM social network model discussed herein are the two major sub-models fused into our ABDSM validation framework. The transportation sub-model is another model we have integrated in the framework. Because of the space limitation, we did not present this model in detail in this report. Our framework used bottom up approach for analyzing the sensitive variations in these three sub-models and their impact on the whole framework output.

IV. Conclusion

In our ABDSM validation project, we proposed a modified sensitivity analysis validation technique for ABDSM validation. Compared to existing ABM validation methods presented in recent literature, e.g., [2], this technique involved more pre-validated sub-system models as the foundation of the validation framework. By converting the sensitive variation in the model into the variations in the sub-system models, we successfully move the challenges of ABM validation into the solutions for validations of the sub-system models, such as the transportation model, social network model and population model. And such challenge transference allows us to use existing validation techniques on sub-system models to validate the ABDSM model. Some issues impede the use of population models that use untreated census data as input for ABDSM simulation. First, the complete census is not always available in many countries; only a small subsample, the so-called public-use sample, can be accessed. Second, the census is collected rather infrequently - as much as 10 years can pass between two consecutive surveys. The primary solution [12] is to combine the census data with readily available up-to-date aggregate data. We will use this approach to validate population data.

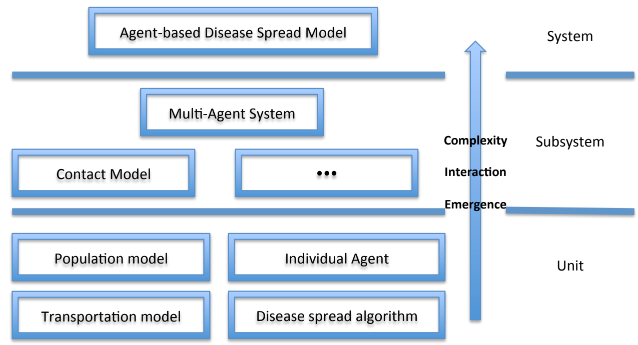

Figure 1. The elements of an agent-based disease spread model.

Acknowledgment

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The research described in this paper was conducted at ORNL where Dr. Pullum is, and while Dr. Cui was, an ORNL employee.

References

[1] G. Fortino, et al. A discrete-event simulation framework for the validation of agent-based and multi-agent systems, *Proceedings WOA 2005 Conference*, Italy, November 2005.

[2] K. C. Ryan, X. Xiang, and T. F. Cosimano. V&V of agent-based and equation-based simulations: a comparison, *Spring Sim'06 Conference*, Huntsville, AL, April 2006.

[3] E. J. Weyuker. On testing non-testable programs. *Computer Journal*, 25(4):465-470, November 1982.

[4] M. A. Louie and K. M. Carley. Balancing the criticisms: validating multi-agent models of social systems, *Sim Modeling Practice and Theory*, 16(2), Elsevier, pp. 242-256, 2008.

[5] M. A. Niazi, et al. Verification and validation of an agent-based forest fire simulation model, *SpringSim'09 - Agent-Directed Simulation (ADS)*, San Diego, CA, March 2009.

[6] R. Garcia, et al. Validating agent-based marketing models using conjoint analysis, *J. of Business Research*, (60), 2007, pp. 848-857.

[7] N. Shillingford, et al. Verification and validation of agent-based scientific simulation models, *Agent-Directed Simulation Conference*, San Diego, CA, April 2005, pp. 27-34.

[8] D. Powell, J. Fair, R. LeClaire, and L. Moore. Sensitivity analysis of an infectious disease model, *2005 International System Dynamics Conference*, Boston, MA, July 2005.

[9] IEEE Std 1012™-2004. IEEE Std for Software Verification and Validation, 2004.

[10] ISO/IEC 12207, Systems & software engineering - Software Life Cycle Processes, 2008.

[11] ISO/IEC 15288, Systems & software engineering - System Life Cycle Processes, 2008.

[12] R. J. Beckman, et al. Creating synthetic baseline populations, Trans. Research Part A: Policy & Practice, 30 (6) 415–429, 1996.