Automated Multi-agent Simulation Generation and Validation

Philippe Caillou

LRI, Universite Paris Sud, F-91405 Orsay France
caillou@lri.fr
http://www.lri.fr/~caillou

Abstract. Multi-agent based simulation (MABS) is increasingly used for social science studies. However, few methodologies and tools exist. A strong issue is the choice of the number of simulation runs and the validation of the results by statistical methods. In this article, we propose a model of tool which automatically generates and runs new simulations until the results are statistically valid using a chi-square test. The choice of the test configuration allows both a general overview of the variable links and a more specific independence analysis. We present a generic tool for any RePast-based simulation and apply it on an Academic Labor Market economic simulation.

Keywords: Multi-Agent Based Simulation, Simulation Validation, Simulation Tool, Chi-square test, statistical test.

1 Introduction

Multi-Agent Based Simulations (MABS) are increasingly being considered as flexible and versatile modeling frameworks, enabling positive and normative investigations of phenomena out of reach when one uses analytical studies[1,2]. However, few methodologies exist on MABS usage. The main problem of MABS is validation: since simulations are by definition too complex to be validated analytically (otherwise they are only useful to inspire analytical analysis), other methods have to be considered. The result of a simulation is a set of observations (for example a set of evacuation times for a simulation of a stadium fire evacuation). As for empirical observations, statistical tools can be used to validate results obtained by the simulation[1]. Their usage is growing, even if the expert validation is still mainly used [3]. One important condition to be able to apply most statistical tests is to have a large enough number of observations. Compared to empirical observations or biological experimentations, MABS has a big advantage: it is easy and almost free to generate new simulation results. Our goal here is to use this advantage to generate automatically new simulations until observed results are statistically valid.

An ideal tool would work with any simulation framework and would select the best statistical test considering the experimenter goal. As a first step, we will however begin

[1] We consider here the validation of the results considering the model is sound. For example, statistical tests can validate the fact that the most important variable for the evacuation time is the number of exits in the simulation. However the model in itself - the agent behaviors, the stadium model - needs to be validated separately.

N. Desai, A. Liu, and M. Winikoff (Eds.): PRIMA 2010, LNAI 7057, pp. 398–412, 2012.

with a single test and a single framework. The chosen framework is RePast[4] as one of the most used framework for MABS. To be generic, we aim to keep the interaction with the framework minimal: the tool reads and write parameter files, starts the simulation program and analyzes result logs.

One of the first results of any MABS is usually the link between the observed variables and the parameters, which ones are the most important parameters and which ones have an influence on the final result. As our goal is to propose a generic tool, we consider here a test with a small number of hypotheses on the tested variables: the Pearson's Chi-square test. It tests if two observed variables/parameters are independent or not with a known percentage of error (by comparing the observed distribution of data with an expected distribution obtained from the distribution of the two tested variables considered independently).

Our objective here is to propose a model for a generic tool which automatically run new simulations from any RePast simulation until the independence results (obtained by chi-square tests) on selected variables are statistically valid. To test and illustrate our method, we apply a corresponding tool on a simulation of the French Academic Labor Market (presented - without statistical validation - in [5,6]). The main parameters describe the hiring system properties and the candidates and universities utility functions, while the observed variables measure the quality of the hiring and the rate of jobs fulfilled by local candidates.

In the following, we present the state of the art in simulation methodology and tools in section 2. Then we describe the method, the chi-square test, propose some heuristics to increase the analysis efficiency and discuss the possible applications of our tool in section 3. The simulation of the French academic labor market is introduced in section 4, some results are presented in section 5 and finally we conclude.

2 Related Work

A variety of social and economic problems have been investigated using multi-agent systems (MAS) [1]. MAS have demonstrated their ability to represent (cognitive) agents and constrained interaction rules, and provide insightful pictures of the dynamics of the system [7]. Several frameworks are available, such as RePast[4], NetLogo[8] and ModulEco[9] (see a review in [10]). The use of automated simulation generation and analysis is not yet integrated in these frameworks. NetLogo has the *BehaviorSpace* tool (and its corresponding API), but it is mainly the equivalent of RePast parameter files: it allows the user to choose parameter ranges to launch multiple experiments. An alternative approach is the LEIA tool from the IODA Framework[11], which is able to reverse-engineer agents and to explore there parameter space[12].

The closest approach to our work is the "robot scientist"[13] developed to achieve biological experiments autonomously. Our approach represents the equivalent for Multi-Agent Simulation, and improves it with efficiency heuristics and the statistical tests both for analysis and as a termination criterion. Another similar approach is the SimExplorer project [14]. Its goal is to manage simulations parameters and results with a generic framework. It is an ongoing project, with many limitations on parameter values, no possibility to "program" simulation runs with stopping criteria. Their work is very complementary to our objective, since they have no specific analysis tool in their