Calibration of Multi-Agent Simulations through a Participatory Experiment

Kévin Darty, Julien Saunier, Nicolas Sabouret

Calibration of Multi-Agent Simulations through a Participatory Experiment

(Extended Abstract)

Kévin Darty
LIMSI-CNRS, Univ. Paris-Sud, Orsay, France
kevin.darty@limsi.fr

Julien Saunier
LITIS, INSA de Rouen, France
julien.saunier@insa-rouen.fr

Nicolas Sabouret
LIMSI-CNRS, Univ. Paris-Sud, Orsay, France
nicolas.sabouret@limsi.fr

ABSTRACT

In the context of agent-based simulation, a major issue is to define relevant parameters of the agent model and calibrate them. In this paper, we propose to log and analyse agents' behaviours to evaluate their similarity to humans behaviours in an immersive virtual environment. The behaviour archetypes are studied in terms of cluster members in order to identify missing agent behaviours, capacities and errors. This study enables to (1) dismiss invalid parameter sets, (2) calibrate valid simulations and (3) explain lacks in the agent models for further improvement.

Categories and Subject Descriptors

I.6.3 [**Modelling and simulation**]: Agent models

Keywords

Simulation, Behaviour analysis, Calibration

1. INTRODUCTION

Multi-agent simulation is used in numerous fields such as artificial economics or social simulation. The agents have to produce behaviours that are similar to human ones. A major issue is to identify relevant sets of parameter values to produce valid behaviours and calibrate the agent population. On the one hand, the agents' behaviours must be believable, *i.e.* the parameter values must lead to behaviours that a human being could adopt. On the other hand, the agents' behaviours must be characteristic of the simulated population, *i.e.* the parameter values must produce sufficient variation in the agents' behaviours.

Most approaches focus on the first aspect and tend to set values that correspond to average or normative behaviours. While this is relevant for macro-simulation, the *MAS* approach considers complex microscopic phenomena for which normative behaviours are not well-suited [1]. We have proposed in [2] a semi-automatic analysis of agents' behaviour through their comparison to human behaviour clustering. This method combines human expertise and simulation logs analysis for the evaluation of the agents' credibility. In the specific case of agents aiming at reproducing human behaviours in an immersive virtual environment, individual behaviours of the agents and their credibility are examined in terms of capacities, lacks (*i.e.* missing behaviours), and errors with respect to human beings. In this paper, we extend this work for the validation and the calibration of a *MAS* simulation.

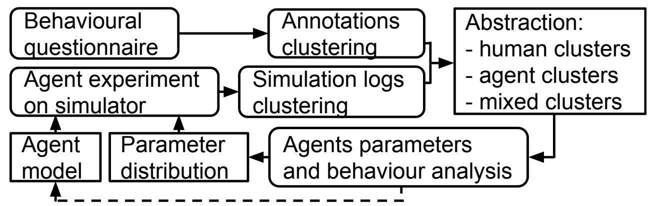

Figure 1: Behaviour analysis and calibration.

2. METHOD

The method presented in [2] enables to evaluate agents' behaviours in the context of virtual environment. It is based on the combination of simulation logs analysis (objective part) and answers to a behaviour questionnaire (subjective part). The general method consists of 5 main steps: 1. Collection of agent behaviour data in simulation; 2. Collection of human data in the same situations through participatory simulation; 3. Annotation of these data by human participants; 4. Data preprocessing and automatic clustering, which leads to clusters of both humans and agents; 5. Clusters comparison: composition analysis and behaviour explicitation.

Three types of clusters may be found. The clusters $\mathbb{C}_M$ containing both humans and agents correspond to high-level behaviours that are correctly reproduced by the agents. The clusters containing only agents $\mathbb{C}_A$ correspond to behaviours that were produced by the agents only, and are thus simulation errors. The clusters containing only participants $\mathbb{C}_H$ correspond to behaviours that have not been replicated by the agents, and are thus lacks in the agent model.

We extend this method in the following directions (Fig. 1): the abstraction step is used to compute scores that assess the proportion of agents capabilities, errors and lacks, as well as the behaviour reproduction level, *i.e.* what agents proportion do reproduce human behaviour, taking into account the occurrence rate of these behaviour in human simulation logs; These correct behaviours are correlated with the agent parameters to propose a new parameters distribution; Finally, we cycle again through the evaluation method to test these new parameters in terms of behaviour reproduction level. Optionally, the new cycle can be used to explore the parameter space, or if the agent model can be modified, to correct the erroneous agent behaviours and to add missing ones.

Calibration

Model calibration means tuning the parameters so that some desired (global) society goal(s) or behaviour(s) are achieved [3]. Calibration of model parameters for detailed agent-based models is a problem for standard calibration techniques due to the large parameter spaces, and long simulation run times. In participatory simulations, another requirement is that individual behaviours are *believable*.

In our case, we rely on human behaviour logs collected during the experimentation process: they define the set of valid behaviours. Considering the agent model as a "black box" that can produce different behaviours depending on its parameters, the calibration process must ensure that: (1) each agent behaviour is believable, hence that their parameter set $P_a = \{p_1, ..., p_i\}$ (with i the number of parameters of the model) is individually valid; and (2) the population globally reproduces the same behaviours as humans, in equivalent proportions, hence the distribution of parameter sets $\mathcal{P} = \{P_1, ..., P_n\}$ with n the number of agents.

Agent members of "errors" clusters have shown behaviours that were not displayed by humans. Their parameter sets are withdrawn from the group of valid parameter sets.

Defining a new parameter set

Depending on the agent model, it is possible to generate the whole spectrum of agents' behaviours. In this case, one cycle through the method is enough to determine valid parameters and their proportion in the agent population; otherwise the calibration is done only on valid behaviours.

Let $\mathcal{P}_v$ be the valid parameter sets corresponding to the valid behaviours $\mathcal{B}_v$, with $simul(P_i) = b \in \mathcal{B}$ the set of possible behaviours, and $p(b)$ the proportion of humans displaying this behaviour. Since several parameter sets may produce the same behaviour, the production of a parameter set $P(a_i)$ with $i \in \{1, ..., n\}$ for n agents implies to choose between several parameter sets.

We propose to select the parameter sets as follows: $P(a_i) = P_i \in \mathcal{P}_v$ with the probability $p(P_i)$ depending on the proportion of observed behaviours b and the number of parameter sets P_j leading to b. Hence, $p(P_i) = n \cdot \frac{p(b)}{|P_j|}$ with $P_j \in \mathcal{P}_v \mid simul(P_j) = b$. In this way, the behaviours that were under-represented have more probability to be produced, since the probability to select a parameter set compatible with it is increased ; and the reverse is true for over-represented behaviours. By using only $\mathcal{P}_v$ and not $\mathcal{P}$, all invalid parameter sets are withdrawn.

Parameter space exploration

All possible behaviours may not be produced in the first cycle of the *MAS* evaluation method. In the case where missing behaviours are found, we include an exploration function that chooses the parameters in non-explored areas of the parameter set: $P(a_i) = P_i \in \mathcal{P}_v$, `if` $p > \gamma$; `else` $P(a_i) = P_k \notin \mathcal{P}$.

The exploration parameter γ allows to iteratively search new behaviours, and p is a uniform random value. P_k must never have been chosen in any previous steps' $\mathcal{P}$. If P_k leads to a valid behaviour, then it is added to the set of valid parameters $\mathcal{P}_v$, and otherwise it is discarded.

Cycling through the method

If all the target behaviours (determined by the human logs clustering) are reproduced, only one step of calibration is necessary. When behaviours are missing, exploring the parameter space may permit to discover new agents' capabilities. Lacks and errors may also be fixed through the agent designer intervention. In this case, the information from the annotation step enables to identify the missing behaviours and errors in a semantic way.

3. DISCUSSION

An originality of this calibration process is the context of the participatory simulations. The goal function of the calibration process is not as usual [3] at the macroscopic level but at the individual level. Virtual reality requires each agent to propose believable behaviours. In this context, we first withdraw parameter sets which do not produce valid behaviours, and then calibrate agents proportions with the remaining parameter sets according to human participant data. Only one cycle of our method ensures that invalid behaviours are detected and that correct proportions are produced, notwithstanding the agent lacks.

Working in the "black box" case where the agent model is unknown and cannot be modified, if there are missing behaviours, a solution is to explore the parameter space to find new agent behaviours. Let us note that these new steps do not require another experimentation with human participants, since the reference data are already available. Each new cycle will hence potentially enable to find new valid parameter sets, either in already mixed clusters, or in previously "lacks" clusters. Working in a "white box" case where the agent model is known and can potentially be modified, the annotation data explain missing behaviours and errors, hence allowing to improve the agent model. Furthermore, the parameter sets exploration can be guided by the knowledge of the model [3].

Several extensions need to be considered. Firstly, the aggregation method depends on a tolerance rate whose value might impact the results quality: This impact should be studied. Secondly, the model convergence has not been evaluated. Our clustering algorithm leads to scores that could be used to stop the process, but a proof of convergence is required when the cycle is not used to explore new parameters.

REFERENCES

[1] T. Bosse, M. Hoogendoorn, M. C. Klein, J. Treur, and C. N. Van Der Wal. Agent-based analysis of patterns in crowd behaviour involving contagion of mental states. In *Modern Approaches in Applied Intelligence*, pages 566–577. Springer, 2011.

[2] K. Darty, J. Saunier, and N. Sabouret. Agents behavior semi-automatic analysis through their comparison to human behavior clustering. In *Intelligent Virtual Agents*, pages 154–163. Springer, 2014.

[3] M. Fehler, F. Klügl, and F. Puppe. Techniques for analysis and calibration of multi-agent simulations. In *Engineering Societies in the Agents World V*, pages 305–321. Springer, 2005.