Using Surrogate Models to Calibrate Agent-based Model Parameters Under Data Scarcity

Extended Abstract

Priscilla Avegliano
IBM Research
LTI / EP / University of São Paulo (USP)
São Paulo, SP
pba@br.ibm.com

Jaime Simão Sichman
LTI / EP / University of São Paulo (USP)
São Paulo, SP
jaime.sichman@usp.br

ABSTRACT

Social Simulation is one of the most prominent uses of Multiagent Systems, but it requires the costly task of fitting parameters to assure the credibility of the model. As, to date, there is no consensus on how to calibrate parameters of agent-based models, we have investigated other knowledge domains to develop an efficient method for this task. Our proposal is based on the definition of a *surrogate model*, that reduces search space dimension. We have tested our method in the housing market scenario, using real data. We achieved satisfactory results, that corroborate the idea that it is important to reduce the search space for an efficient parameter calibration.

KEYWORDS

Social simulation; Simulation of complex systems; Modelling for agent based simulation

ACM Reference Format:
Priscilla Avegliano and Jaime Simão Sichman. 2019. Using Surrogate Models to Calibrate Agent-based Model Parameters Under Data Scarcity. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

Agent-based models are notably applied in the context of social simulations, as they allow the replication of emergent behaviour by modelling the constituents parts of a system [7]; Additionally, they provide an endogenous explanation for the simulated phenomena, which is specially desirable in scenarios of decision-making, enabling one to foresee the consequences of new politics. The challenges faced in the design of an agent-based model have roots in its own advantages: as it requires the definition of each constituent part of the system, several parameters that are omitted by other modelling approaches must be defined. This means that more detailed data is required, which is rare to happen, mainly because of two aspects: (1) even in the era of big data, we usually have access to aggregated data sets, that do not provide information at the agent level; (2) the very nature of the parameters are challenging: modelling agents in social simulations involves quantifying feelings and mental states that are not easily translated into a numerical scale.

Although the parameter calibration task presents some peculiarities in the agent-based models domain, other knowledge areas face similar problems, such as high dimensionality. Techniques from these areas were used as an inspiration for the proposition of an efficient method for parameter calibration for agent-based models, applying the definition of a *surrogate model* to reduce the associated computational cost. Our proposal consists of using a model library with canonical behaviours to generate a surrogate function $\hat{f}$, that describes, in a simplified way, the behaviour of the original agent-based model. Our method presents several advantages. Namely: (1) the simplification brought by the surrogate model allows a consistent convergence for satisfactory approximations of parameter values, due to the reduction in search space size and in the stochasticity of agent-based models; (2) with an inexpensive algorithm, we are able to generate a function $\hat{f}$ that could be translated into the numerous parameters of the agent-based models.

2 CALIBRATING PARAMETERS

Agent-based models devoted to social systems are basically the quintessence of an ill-posed problem for the task of parameter calibration: they are mainly nonlinear, data is scarce and they present a great number of parameters. Despite the clear importance to the field, still there is no golden standard for parameter calibration in agent-based models [15]. Some works used the strategy of calculating an approximation of the *Maximum Likelihood* [13] [8]. Parameter calibration is frequently addressed in the literature via metaheuristics, specially by evolutionary/genetic algorithms. This strategy was applied in [1], [2], [9], and [6]. Three different parameter calibration methods were compared in [10].

Several other knowledge domains face the problem of parameter calibration in large search spaces. One common strategy adopted in situations in which data is scarce allied with a large search space is the creation of a **surrogate model** [14]. A surrogate model is an approximated representation of the model output in the parameter space. Generally speaking, a model can be seen as a function. Given x, that can be input data and model parameters, we apply function f (the model) to obtain y, the output. In mathematical terms, we have: $y = f(x)$. Therefore, the surrogate model can be defined as $\hat{f}$ in the equation $y = f(x) \approx \hat{f}(x')$, with $x \in X$, $x' \in X'$, and $X \gg X'$.

3 PROPOSAL

In this work, we investigate the usage of a surrogate model to reduce computational cost of parameter calibration for social agent-based models. Our method can be seen as a two-phase process. In the

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

first one, the surrogate function $\hat{f}$, that requires a smaller number of parameters than the original model, is generated and calibrated using real data. On the second phase, the reduced set of parameters previously calculated is used as a reference, and the whole set of parameters of the original agent-based model is then calibrated.

The first challenge of our method is to define the strategy to choose function $\hat{f}$ that describes the surrogate model. As we are in the domain of social systems, we are mainly dealing with non-linear functions. We opted to create a model library with the most common patterns of behaviours. Our inspiration came from System Dynamics, created by Forrester [4]. In particular, we applied canonical behaviours (*fundamental modes*) [11]. In our approach, collected data is compared against such template curves in the model library to infer which one better suits the observed phenomena.

The strategy adopted for calibrating the parameters is another cornerstone of our method. This task is performed twice during its execution: first, to obtain the adjusted parameters of the ordinary differential equation (ODE) associated to the selected model template for the surrogate function $\hat{f}$; those macro-parameters are then used as the initial guess values of the parameters of f of the agent-based model. This is important because finding zones of interest in a highly dimensional search space is hard [12]. With the usage of the simplification brought by function $\hat{f}$, it is easier to find them. The second calibration acts as a local search, incorporating the stochasticy of the agent-based model and providing satisfactory results. The *CMA-ES* algorithm [5], a population-based metaheuristics, was selected to perform the task of parameter calibration due to its performance with ill-conditioned problems (nonlinear, non-convex, with high dimensionality and non-separable), its capacity to auto-regulate parameters, such as step size, and for having already been applied successfully for the purpose of parameter calibration [6].

4 EXPERIMENT

Data from the housing market in Brazil was chosen for our experiment and was obtained from a publicly available repository at [3], regarding the price p of m^2 in São Paulo (Brazil) from January, 2008 to March, 2018. The time-series behaviour resembles the *S-shaped growth* fundamental mode, that can be modelled as the following ODE, used as surrogate function $\hat{f}$: $\frac{dp}{dt} = \alpha p - \beta \frac{p^2}{limit}$.

We have instantiated a simulation, containing two types of agents (sellers and buyers), that repeatedly negotiate the price p of houses. At each step, sellers increase the asked price by factor β and their accepted $lowestPrice$. Buyers also update what they are willing to pay: they verify the average price of the last step, and increase this value by factor α to define the offered price (if it is inferior to his savings, denoted by $limit$). If the offered price is above $lowestPrice$, the transaction occurs; if not, the buyer offers the same amount to the next seller. If a seller was successful on the previous step, he increases the price as described earlier. Otherwise, he decreases the asked price down to his lower limit ($lowestPrice$).

In our experiment, we first performed the task of identifying the surrogate function $\hat{f}$, providing calibrated values for variables $alpha$, $beta$, $limit$ and p_0. We then defined a population of 100 agents of each type, whose properties were modelled by following a normal distribution around the original values of parameters α, β and $limit$, thus generating heterogeneity in the society. The result of the simulation (agent-based curve in Figure 1(a)) shows that by using the previously calculated parameters, we can obtain a reasonably good fit of parameters. This is an important feature, as parameter calibration can make part of the modelling process of agent-based simulation. By quickly reaching the desired behaviour, one can see if the definitions and assumptions of the model are correct. We then executed another calibration of the parameters of the simulation, providing the previously calculated values of the variables as the initial point in the search space. This provided a refined solution for our simulation, closer to the observed real data, shown as calibrated in Figure 1(a).

Finally, to compare the performance of our method, we executed the calibration of the parameters directly into the agent-based simulation model, without using the concept of a surrogate model, whose result is shown at Figure 1(b). We called this as the raw calibration of parameters. Using the same method for parameter calibration (*CMA-ES*) and the exact same model of simulation can result in very discrepant outcomes. The raw calibration was not able to find suitable parameters to satisfactory replicate important features of the real system. This is quite relevant as it can be a misleading result, that could invalidate a viable agent-based model.

As the model presents some stochasticity, we executed the calibration process both with our proposed method and with the raw calibration 40 times. The mean *RMSE* and the associated standard deviation are, respectively, 4.66 ± 0.46 for the method with the surrogate model and 23.37 ± 23.26 for the raw calibration. Given the randomness of the model and of the calibration method, we can see that sometimes the raw calibration provided good results. This is not a guaranteed behaviour, though. On the other hand, our method provided satisfactory and consistent results, always converging to the desired behaviour. This happens because the first step of generating the surrogate model and performing a calibration on a reduced space search made it easier to find suitable values.

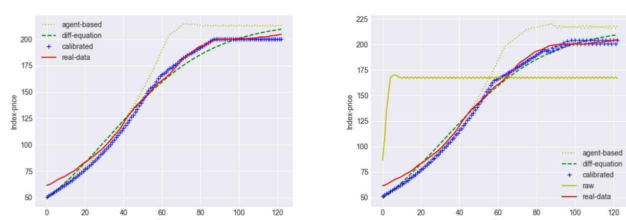

Figure 1: (a) Using surrogate function, (b) Raw calibration

5 CONCLUSION

We have presented a new method for parameter calibration for social agent-based models. Our contribution is twofold: the first and more important one is the use of a surrogate model to reduce space search dimensionability, assuring the convergence of the result; additionally, we presented an efficient and light method to identify the surrogate model, based on a library composed of common fundamental modes of behaviour. We have validated both concepts with real world data from the housing market scenario.

In future work, we believe it will be well worth exploring further other calibration methods, to investigate if the stochasticity of the *CMA-ES* influenced somehow the results.

REFERENCES

[1] Benoît Calvez and Guillaume Hutzler. 2005. Automatic tuning of agent-based models using genetic algorithms. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 41–57.

[2] Raphaël Duboz, David Versmisse, Morgane Travers, Eric Ramat, and Yunne-Jai Shin. 2010. Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. *Ecological modelling* 221, 5 (2010), 840–849.

[3] FIPE-ZAP. 2018. Varia cão do índice FIPE-ZAP. (2018). http://fipezap.zapimoveis.com.br.

[4] Jay W Forrester. 1994. System dynamics, systems thinking, and soft OR. *System Dynamics Review* 10, 2-3 (1994), 245–256.

[5] Nikolaus Hansen. 2006. The CMA evolution strategy: a comparing review. In *Towards a new evolutionary computation*. Springer, 75–102.

[6] Zach Lewkovicz, Daniel Domingue, and Jean-Daniel Kant. 2009. An agent-based simulation of the French labour market: studying age discrimination. In *The 6th Conference of the European Social Simulation Association, eds., Bruce Edmonds and Nigel Gilbert,(9 2009)*.

[7] J. H. Miller and S. E. Page. 2007. *Complex Adaptive Systems*. Princeton University Press, New Jersey, United States.

[8] Matteo G Richiardi, Roberto Leombruni, Nicole J Saam, and Michele Sonnessa. 2006. A common protocol for agent-based social simulation. *Journal of artificial societies and social simulation* 9 (2006).

[9] Alex Rogers and Peter Von Tessin. 2004. Multi-objective calibration for agent-based models. In *Proceedings of 5th Workshop on Agent-Based Simulation*.

[10] Meghendra Singh, Achla Marathe, Madhav V Marathe, and Samarth Swarup. 2018. Behavior Model Calibration for Epidemic Simulations. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1640–1648.

[11] John D. Sterman. 2000. *Business Dynamics - Systems thinking and modeling for a complex world*. Irwin McGraw-Hill, United States.

[12] Albert Tarantola. 2005. *Inverse problem theory and methods for model parameter estimation*. siam.

[13] Jan C Thiele, Winfried Kurth, and Volker Grimm. 2014. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation* 17, 3 (2014), 11.

[14] Rohit Tripathy and Ilias Bilionis. 2018. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *arXiv preprint arXiv:1802.00850* (2018).

[15] Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. 2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10, 2 (2007), 8.