Data-driven Agent-based Modeling

Experimenting with the Schelling's Model

Jamali, Ruhollah; Vermeiren, Wannes; Lazarova-Molnar, Sanja

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2024.06.028

Publication date:
2024

Document version:
Final published version

Citation for pulished version (APA):
Jamali, R., Vermeiren, W., & Lazarova-Molnar, S. (2024). Data-driven Agent-based Modeling: Experimenting with the Schelling's Model. *Procedia Computer Science*, *238*, 298-305.
https://doi.org/10.1016/j.procs.2024.06.028

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 238 (2024) 298–305

Procedia Computer Science

www.elsevier.com/locate/procedia

The 15th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 23-25, 2024, Hasselt, Belgium

Data-driven Agent-based Modeling: Experimenting with the Schelling's Model

Ruhollah Jamali[a,*], Wannes Vermeiren[b], Sanja Lazarova-Molnar[c,a]

[a] *Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark*
[b] *Faculty of Engineering Technology, Catholic University of Leuven, Gebroeders De Smetstraat 1, 9000 Ghent, Belgium*
[c] *Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Kaiserstr. 89, 76133 Karlsruhe, Germany*

Abstract

Agent-based modeling and simulation is a practical computational technique for studying complex systems of autonomous agents in various disciplines. Agent-based models facilitate the study of emergent phenomena by simulating heterogeneous agents and their flexible behaviors and interactions. However, developing an agent-based model of a complex system is often time-consuming and vulnerable to the modeler's biases. Addressing this challenge requires a paradigm shift from knowledge-driven modeling to data-driven modeling. In this research, we initiate and experiment with automating the process of composing agent-based models by developing data-driven model extraction. To achieve this objective, we conduct experiments employing different variations of Schelling's segregation model, a well-known agent-based model, each featuring different parameter sets and complexity levels.

Keywords: Agent-based modeling; Schelling Model; Data-driven modeling; Data analysis;

1. Introduction

Models have been instrumental in understanding, predicting, and simulating various phenomena across different disciplines. Whether applied to natural ecosystems, social dynamics, or economic systems, models provide a structured lens through which we can observe and comprehend complex interactions. In the field of modeling and simulation, Agent-Based Modeling (ABM) is known as a robust approach to modeling complex systems by modeling individual entities (agents) and their interactions within a defined environment. Generally, models represent objects, events, or processes wherein certain attributes are simplified or highlighted while some might be disregarded. However, the creation of highly detailed models is challenging, as it involves navigating the balance between capturing informative features and maintaining computational efficiency. Achieving this equilibrium demands innovative

* Corresponding author. Tel.: +45-92-99-94-11.
E-mail address: ruja@mmmi.sdu.dk

10.1016/j.procs.2024.06.028

approaches and can lead to the development of diverse models on the same subject. For instance, a person can be modeled to study obesity levels or social behaviors, each model focusing on different attributes. The process of developing models demands significant effort and background knowledge in the context of the model's subjective. Modelers have been tasked with defining interaction rules within the model while defining and handling numerous attributes and thresholds. Developing a model is a laborious iteration of trial and experimentation, which relies on the modeler's expertise and familiarity with the subject.

ABM utilizes mathematical equations or computational algorithms to represent entities of a process or system, their relationships, and the resulting dynamics within the system [26]. ABM's unique capability lies in its ability to capture the bottom-up dynamics of systems, allowing for the study of emergent phenomena that arise from heterogeneous agent interactions, and, thus, enabling the representation of complex, non-linear interactions and emergent phenomena. ABM allows us to explore how individual-level decisions lead to system-level outcomes, making it valuable in understanding and predicting in various areas such as economics, social sciences, ecology, and more.

Applications of ABM span a broad range of disciplines. The most well-known application of ABM is studying social behaviors, as agents are suitable for representing humans and their complex and heterogeneous behaviors. Schelling's segregation model [25] is an example of a social behavior model that is useful for studying residential segregation of ethnic groups. ABMs can be constructed to study the impact of new guidelines or legal decrees. For instance, a model was constructed to understand the EU ban on incandescent lamps [5]. Other ABM applications range from predicting the spread of epidemics with the Susceptible-Infectious-Recovered (SIR) model [23], to the closely related bio-warfare [2], to more peaceful economic models. Economic models include the wealth distribution model [7] and more advanced models to study early warning measures for economic crises [4].

Developing an agent-based model presents a set of significant challenges, including defining agent characteristics and interaction rules among them to ensure the model's accuracy and reliability. Additionally, defining realistic agent behaviors and interactions while managing computational complexity remains a fundamental challenge in constructing agent-based models. This paper aims to demonstrate the potential for improving the model development process by automation of agent-based model development. For this, we begin with extracting agent-based model parameters from synthetic data extracted from simulated scenarios using the same model. Next, we advance this approach by integrating machine learning methodologies to extract model logic from the synthetic data. Here, we primarily focus on Schelling's Segregation model and its extensions.

The rest of the paper is structured as follows: In Section 2, we present the related works on data-driven agent-based modeling. Section 3 describes the prerequisites and use case models for our research. Our approach for developing data-driven agent-based models is presented in Section 4. Lastly, Section 5 provides a summary, discussion on results, and outlook of the paper.

2. Background and Related Work

Data-Driven Agent-Based Modeling (DD-ABM) has received scant attention in the research literature. While agent-based models have evolved to a higher maturity, the development of agent-based models in an automated and data-driven manner remains an understudied domain [15]. By "data-driven", we refer to the generation of the constituent elements of an agent-based model from data associated with the system of interest. Data-driven methods in ABM can be applied for three main purposes: firstly, to estimate the parameters of an available agent-based model of an event based on the available data of the event; secondly, to construct an agent-based model that accurately represents an event based on the observed event data; and lastly, to develop a fully functional agent-based model in which both the model's structure and its associated parameters(model's logic) are derived from the available data sources of the event.

The varied methods for parameter estimation in agent-based models arise from the complex and heterogeneous nature of these simulations, requiring adaptable approaches to capture diverse agent characteristics and interactions. The choice of a specific methodology is dependent on factors such as model complexity, data availability, and the desired balance between realism and computational feasibility. Traditional brute force methods, which involve exploring all potential parameter combinations, are computationally intensive and often impractical. Manual tuning, on the other hand, tends to be suboptimal compared to more analytical techniques. In a comprehensive review by Hazelbag et al. [10], the authors reviewed parameter search strategies and goodness-of-fit (GOF) measures in

ABMs, focusing on modeling infectious disease spread subject. They reviewed 84 different articles to standardize the parameter search strategy. However, their analysis revealed that parameter search is frequently neglected, with 52% of the surveyed articles lacking identifiable or replicable search strategies. This underlines the limited adoption of rigorous model calibration practices in the field. As presented by Hazelbag et al., the most common optimization algorithms are grid search and local search algorithms. Sampling optimization algorithms are more refined, considering they return a distribution of parameter values and not just a single combination. Besides the noted methods, evolutionary algorithms [22, 27], machine learning algorithms [3, 14], and Bayesian methods [13, 9] are also employed for parameter estimation in ABM. Furthermore, there have been numerous comparative efforts to assess agent-based calibration methods, tailored to specific modeling domains. For instance, in the fields of economics [21], macroeconomics [20], and the stock market [28].

In most cases, agent-based models are composed based on the domain expertise of modelers. However, there are some notable efforts to design agent-based models in a data-driven manner. Keller and Hu [12] present a compelling study on mobile agent-based models, wherein their main goal is to reduce the effects of all possible kinds of bias from the modeler. They define a broad space of possible models, and then an algorithm decides which model is the most applicable one. In their approach, the modeler only defines the desired behaviors and not the underlying rules. Their approach is in between knowledge-driven modeling and data-driven modeling. There are also other research efforts where authors combined knowledge-driven and data-driven approaches to develop an agent-based model. E.g., Zhang et al. [30] developed an agent-based model for rooftop solar adoption, Bell et al. [1] developed an agent-based model to understand the model's parameters and customer behaviors, and Jamali et al. [11] investigated discovering agents' behavior patterns, all in a data-driven manner.

To create a fully functional agent-based model, we need both a model and the corresponding parameter values. Whilst some approaches focus on one aspect, some more comprehensive approaches cover both. Developing domain-specific simulation models for dynamic processes is not only time-consuming, but also, the robustness of the model could be questionable, resulting in unreliable simulation results [17]. Furthermore, a considerable amount of time is often spent modifying the foundational elements of a general model to serve specific purposes in different contexts when, ideally, these elements should serve as a unified basis for the model across diverse applications. To address this, Loo et al. have started formulating a generic agent-based model in another publication, but it still requires more research [18].

3. Prerequisites and Use Case Models

To comprehensively engage with our proposed data-driven procedure, it is essential to establish a foundational understanding of several key concepts and methodologies that underpin our study, which we cover in this section. First, we elaborate on Schelling's model as a case study for our approach. Then, we provide a short description of the methods we employ in our data-driven modeling approach.

3.1. Schelling's Agent-Based Model

Thomas Schelling developed the Schelling segregation model in the late 1960s [24, 25] for exploring the emergent dynamics of segregation in human societies, with its core principles rooted in the individual-level decision-making processes that underlie spatial patterns of residential choice. Schelling's segregation model demonstrates that local preferences can lead to global segregation patterns without the need for extreme bias in individuals. Schelling's model offers an intuitive entry point into the area of complex system modeling due to its simplicity, interactivity, real-world relevance, and its capacity to introduce fundamental concepts in complex systems. Schelling's model employs a simplified representation of a spatial grid populated by agents, each characterized by a personal preference for the composition of their immediate neighborhood. Agents continually assess their satisfaction with their surroundings, which, in turn, influences their decision to relocate. A key feature of this model is its straightforward rule-based nature, where agents merely seek to be surrounded by a certain threshold of like-minded neighbors. This simple yet insightful approach serves as a great example of ABM, making it an ideal starting point for exploring and understanding data-driven modeling techniques.

According to Macal and North's tutorial, [19] on developing agent-based models, three components are needed: environment, agents, and interaction rules. For a basic Schelling's model [25], the environment is a grid where each cell represents a location in a neighborhood and all adjacent cells are the neighbors. Each agent has an ethnicity attribute, which can take two values. In every step of the simulation, an agent checks all its immediate neighbors' ethnicities, and if the number of neighbors of the same ethnicity divided by the number of all neighbors is over a threshold τ_h (termed - happy threshold), it stays in the same location, and otherwise the agent moves to a new random location. We term the happy threshold, a model's parameter. In every step of the model, each agent chooses between relocating or staying in the same location, based on the happy threshold. The simulation might reach an equilibrium if all the agents are happy, but this situation cannot always be achieved.

In our research, we also considered a multidimensional segregation model proposed by Liu et al. [16]. We chose this model because it adds valuable additional insights whilst staying reasonably faithful to the original working of Schelling's segregation model. The multidimensional segregation model adds complexity by allowing agents to be a part of multiple groups based on multiple attributes reflecting groups, e.g., ethnicity, age, and education level. This enables the model to more accurately represent real-life situations where humans consider multiple aspects or characteristics of their neighbors to determine their similarity levels. Along with the increase in the number of attributes, a different threshold is introduced. The similarity threshold τ_s determines how many groups two neighbors must have in common to be considered similar enough. For example, if the similarity threshold is at 50%, agents with different ethnicity groups with the same age group and education level will still recognize each other as similar enough since two out of their three attributes match.

We also considered a multidimensional Schelling model with weights, where we extended the multidimensional Schelling model further by adding weights to each attribute when calculating similarities. Using weights in a model helps to highlight the importance of certain attributes over others. For instance, the weight of the ethnicity might have a factor 3 times greater than the age group, as this could be much more important to the agents in their decision-making process.

3.2. *Data-driven modeling*

As noted in Section 2, data-driven methods can be used to estimate model parameters, rules, or a combination of both. In our research, we apply two methods to estimate model parameters and rules: decision trees (DTs) and support vector machines (SVMs).

DTs are supervised learning methods widely employed for classification and regression tasks in machine learning and data mining. Their purpose is to partition data into subsets based on the values of particular features, ultimately leading to the classification of instances into predefined classes. DTs demonstrated broad application in the field of computer science as they offer an interpretable structure, do not need data preparation, can handle multi-output problems, and have the possibility to validate models using statistical tests. In general, DTs are expressive and easy to understand, and they have particular appeal to computer scientists due to their recursive divide-and-conquer nature [8].

In addition to decision trees, we also employ SVMs as our second method. SVMs are black-box algorithms, implying that they do not provide a human-readable explanation for their classification decisions. However, in some use cases, SVMs compensate for this by providing superior classification accuracy. For an extensive mathematical study of SVMs, Rozenberg et al. [29] provides a comprehensive tutorial.

4. Data-driven Agent-based Modeling

In the following, we describe how we performed three experiments to develop data-driven models from synthetic data generated by variations of Schelling's model. These variations are made by adding multiple agent attributes and weights to the basic model for presenting agents' preferences, as noted in Section 3.1. We run the first experiment using data from a basic Schelling's model with irregular behaviors of agents, where we focus on estimating agents' happiness threshold by employing decision trees. Irregular behavior in Schelling's model is when an agent has random behaviors during a given percentage of time. We applied the second experiment on data from multidimensional Schelling's models with irregular behaviors again to estimate the model's parameter, where agents have multiple attributes and the model's agents consider two thresholds to make their decisions. In the third and last experiment, we employ DTs and SVMs to extract the agents' logic instead of only determining model parameters. In all experiments, we implemented the models using Julia programming language, employing the Agent.jl library [6]. Subsequently, we produced

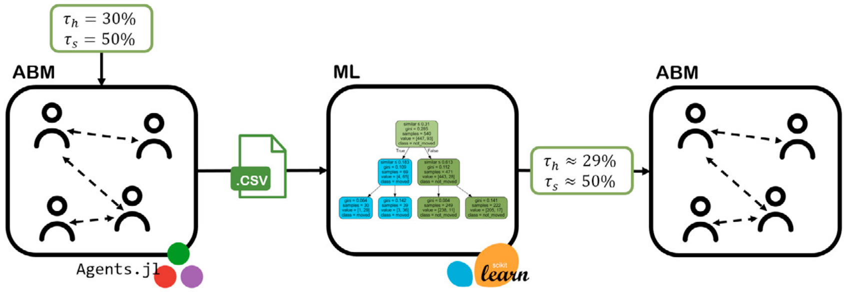

Fig. 1. High-level overview of the data-driven ABM workflow

synthetic data, which is essentially a log of the agent's movements within the environment, using the models we implemented. Finally, we attempted to reconstruct the original model using the synthetic data that was generated. This procedure enabled us to have control over models' internal mechanisms and curate datasets devoid of any extraneous variables to facilitate early-stage testing. Moreover, using synthetic data enabled us to eliminate data-driven methods that were not reliable for datasets without anomalies. Figure 1 demonstrates an overview of our data-driven procedure. Given that all agent-based models inherently exhibit stochastic behavior, deviations may occur based on the initial parameters of the model. In our experiments, we added noise into the synthetic data to evaluate how our data-driven approach performs under conditions that more closely resemble real-world scenarios. For the introduction of noise, we used a fixed seed in all experiments. Further, we investigated the impact of the seed by testing ten different seeds. Our observations indicated that variations in the seed had only a minor impact on the results, which was negligible.

In the first experiment, we generated synthetic data using the basic Schelling's model, presented in Subsection 3.1. Without loss of generality, we considered the environment to be a 15 by 15 grid, resulting in a total of 225 locations, where 80% of the grid is populated by agents. We collected data from the first five iterations of the model, excluding the initialization step. We define one iteration as cycling through all agents and executing their agent steps. The agent step function allows agents to relocate or stay in place. Using this approach, we collected 900 data points for each model. We collect two features for each agent at all steps of running the simulation: the group number and the percentage of similar neighbors. The label for each data point is a Boolean value, indicating whether the agent with those two features moved or not. For the first experiment, we ran the basic Schelling's model with a happy threshold of 30% where 10% of the agents display irregular behavior. We trained a classification tree on the synthetic data we generated, and we observed that the threshold of 30% was discovered from the data by training a classification tree on the data. We observed the split condition in the root node of the tree had a 30 value as the split threshold, indicating the value for the happy threshold.

For the second experiment, we focused on multidimensional Schelling's model. In this model variant, each agent has three attributes: ethnicity, age, and education, each characterized by binary values. Given the multiple agents' attributes, we have the flexibility to employ different presentations of input features set for our data-driven approach. We chose the following three representations of input features, considering their ability to capture different aspects of agents' interactions:

Feature Set A For each attribute, we determine the percentage of neighbors that share the same attribute. For instance, 37.5% of neighbors have the same education level.

Feature Set B We examine the number of neighbors who share specific combinations of attributes within our three characteristics. For example, two neighbors share the same ethnicity.

Feature Set C We calculate the percentage of neighbors who share a common characteristic, considering the number of characteristics in common. For instance, 50% of the neighbors have only one common characteristic.

In multidimensional Schelling's models, agents possess three attributes, offering them the choice of considering a neighbor as similar based on one, two, or all three attributes. To capture this agent preference, we employed the similarity threshold, denoted as τ_s. In our experiment, we set τ_s at 50%, indicating that a neighbor is considered similar if they share the same group with respect to at least two attributes. We added $0, 1, 2, 5, 10, 20, and 30$ percentages of irregular behavior to the data for all possible values of the happy threshold, and we examined the predicted happy threshold values. The decision tree performed comparable to the basic Schelling model in most scenarios. However, under the conditions of 30% irregular behavior and a happy threshold exceeding 80%, the decision tree's performance was suboptimal.

In our data-driven approach to estimate both thresholds, we added $0, 1, 2, 5, 10$, and 20 percentages of irregular behavior to the synthetic data. We initially employed a decision tree, which performed well in all cases. Afterwards, for the similarity threshold, we employed SelectKBest, a statistical method for feature selection, and for the happy threshold, we implemented a linear classifier. However, this procedure did not yield any notable improvements in the results compared to the decision tree, hence we decided not to pursue it further. Another option we considered was to employ Support Vector Machines (SVMs) to estimate both thresholds. We investigated C-Support Vector Classification (SVC) method considering different kernels. The only kernel that demonstrated comparable results to the decision tree performance was the radial basis function kernel. Accordingly, we will use SVC for further investigation.

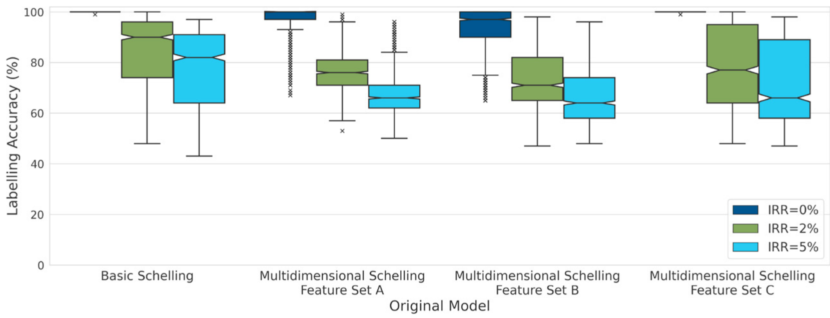

Fig. 2. Labelling accuracy using decision trees

So far, we have focused on determining the model's parameters, namely the happy and similarity thresholds for uni- and multi-dimensional Schelling's models. However, with weighted Schelling's and tree-based Schelling's models, we introduced models that are too complex to extract the parameters from directly. In the third experiment, we focus on learning the model logic instead of only determining model parameters. In many cases, this approach is more applicable since the underlying model is usually unknown, making it challenging to recognize the target parameters. We examined two different approaches for reconstructing model logic. The first approach focuses on understanding the agents' behavior with a human-interpretable logic (DTs). The second approach focuses on mimicking the behavior of the original model as accurately as possible with less emphasis on the explainability of the logic. The similarity criteria in the model we used to generate the synthetic data could be defined using different feature sets (A, B, or C) we explained earlier. We execute our experiments considering each of these feature sets.

For the third experiment, we developed two phases. In the first phase, we replace the logic of the basic Schelling model and multidimensional Schelling model presented by Feature sets A, B, and C with a DT trained on the synthetic data generated by the original model. The DT approach allows for a step-by-step understanding of agent decision-making by observing the tree structure. However, the effectiveness of replicating the original model's behavior varied in our experiments. To assess the performance of the DTs, we calculated labeling accuracy, representing the accuracy of the decision tree in predicting agents' decisions, for each feature set and level of irregularity ($0, 2$, and 5 percent). Employing feature set C outperformed the model that was using feature set B, particularly with zero irregularity. This performance difference was attributed to the simplicity of the model's feature set, enabling accurate replication of the original model. In contrast, the feature set B model, which examines neighbors one at a time, posed a challenge due to its complex logic and the impracticality of dealing with many features. Figure 2 shows the box plot of labeling accuracy of predicted agents' decisions using a decision tree considering different feature sets. Running multiple experiments, we observed that balanced datasets, where all categories (move or stay) have approximately the same number of samples in the dataset, prevent bias in the training process compared to imbalanced datasets.

The second phase is using SVMs, which represent a black-box logic method. SVMs operate mathematically and don't offer interpretable explanations for human understanding. Instead, they separate data points of different

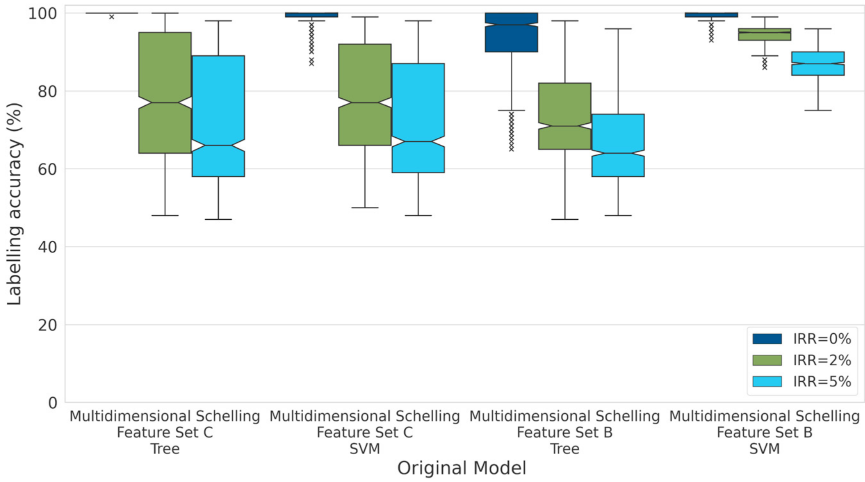

Fig. 3. Labelling accuracy comparison of features and algorithms

classes using hyperplanes in high-dimensional space. However, SVMs often excel in terms of labeling accuracy when compared to DTs. Figure 3 directly compares predicted labeling accuracy between DTs and SVMs in the context of the multidimensional Schelling's model for each feature set. When replacing the original model logic with a DT, the highest accuracy is achieved with the feature set C, especially for models without irregular behavior. This outcome aligns with expectations, given that the feature set C was designed specifically for the multidimensional Schelling's model. Comparing DTs and SVMs as model logic, results are very similar when using feature set C, with a slight advantage for decision trees. However, using the feature set B yields markedly different results, with SVMs paired with feature set B significantly outperforming other options. This illustrates the advantage of SVM in the case of the availability of more features.

5. Summary and Outlook

The objective of our research was to develop a data-driven approach for reconstructing variations of Schelling's models. We investigated discovering model parameters and learning model logic from available data generated by the original model. Our experiments confirmed that we can accurately determine model parameters for uni-dimensional and multidimensional Schelling models with high levels of irregular behavior. Furthermore, by employing machine learning techniques, we can learn a model logic that can mimic the behavior of the original model, independent of the original model's logic. Finally, we demonstrated how irregular behavior levels should also be replicated in the data-driven model. Taken together, these findings suggest that we can reconstruct Schelling-based models with reasonable accuracy from their data. The work is the first comprehensive investigation of a data-driven approach applied to Schelling's model. The comparison between the five Schelling-based models helps to illustrate the strengths and weaknesses of different data-driven approaches.

In this paper, we focused on a small subset of agent-based models, and future work should be done to turn our results into a more generic approach that can be used to develop agent-based models from the available data. Furthermore, additional methods should be developed to cover gaps in the dataset, such as a lack of information on a specific dimension. Finally, experiments with real-world data are required to verify the results we found using synthetic data. Despite its exploratory nature, this study showed that data-driven agent-based modeling has strong potential. Data-driven modeling proposes multiple applications in various fields, such as constructing simulations of real cities, traffic flow, or even market movement, where simulations can assist policymakers in their decisions. Data-driven modeling possesses excellent potential for modelers to construct their models in a more data-oriented way rather than relying on knowledge-driven modeling. This will not only make the modeling process more efficient but also improve the model quality.

Acknowledgements

This work is partly funded by the Innovation Fund Denmark (IFD) under File No. 9065-00207B

References

[1] Bell, D., Mgbemena, C., 2018. Data-driven agent-based exploration of customer behavior. Simulation 94, 195–212.
[2] Carley, K., Fridsma, D., Casman, E., Yahja, A., Altman, N., Chen, L.C., Kaminsky, B., Nave, D., 2006. Biowar: scalable agent-based model of bioattacks. IEEE transactions on systems, man and cybernetics. Part A, Systems and humans 36, 252–265.
[3] Carrella, E., Bailey, R., Madsen, J., 2020. Calibrating agent-based models with linear regressions. Journal of Artificial Societies and Social Simulation 23.
[4] Catullo, E., 2017. Chapter 5 - early warning indicator for crises in an agent-based macromodel, in: Introduction to Agent-Based Economics. Elsevier Inc, pp. 105–124.
[5] Chappin, E.J.L., Afman, M.R., 2013. An agent-based model of consumer lighting, in: Agent-Based Modelling of Socio-Technical Systems. Springer Netherlands, Dordrecht. Agent-Based Social Systems, pp. 181–200.
[6] Datseris, G., Vahdati, A.R., DuBois, T.C., 2022. Agents. jl: a performant and feature-full agent-based modeling software of minimal code complexity. Simulation , 00375497211068820.
[7] Dragulescu, A.A., Yakovenko, V.M., 2002. Statistical mechanics of money, income, and wealth: A short survey. arXiv.org .
[8] Flach, P., 2012. Machine learning: the art and science of algorithms that make sense of data. Cambridge university press.
[9] Grazzini, J., Richiardi, M.G., Tsionas, M., 2017. Bayesian estimation of agent-based models. Journal of Economic Dynamics and Control 77, 26–47.
[10] Hazelbag, C.M., Dushoff, J., Dominic, E.M., Mthombothi, Z.E., Delva, W., 2020. Calibration of individual-based models to epidemiological data: A systematic review. PLoS computational biology 16, e1007893.
[11] Jamali, R., Lazarova-Molnar, S., 2023. Uncovering competitor pricing patterns in the danish pharmaceutical market via subsequence time series clustering: A case study, in: 2023 Winter Simulation Conference (WSC), IEEE. pp. 793–804.
[12] Keller, N., Hu, X., 2019. Towards data-driven simulation modeling for mobile agent-based systems. ACM transactions on modeling and computer simulation 29, 1–26.
[13] Kim, D., Yun, T.S., Moon, I.C., Bae, J.W., 2021. Automatic calibration of dynamic and heterogeneous parameters in agent-based models. Autonomous Agents and Multi-Agent Systems 35, 46.
[14] Lamperti, F., Roventini, A., Sani, A., 2018. Agent-based model calibration using machine learning surrogates. Journal of Economic Dynamics and Control 90, 366–389.
[15] Lazarova-Molnar, S., Li, X., 2019. Deriving simulation models from data: steps of simulation studies revisited, in: 2019 Winter Simulation Conference (WSC), IEEE. pp. 2771–2782.
[16] Liu, Z., Li, X., Khojandi, A., Lazarova-Molnar, S., 2019. On the extension of schelling's segregation model, in: 2019 Winter Simulation Conference (WSC), IEEE Press. pp. 285–296.
[17] Loo, Y.L., Ahmad, A., Mustapha, A., Mostafa, S.A., 2021. A self-adaptive agent-based dynamic processes simulation modelling framework, in: 2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR), IEEE. pp. 124–129.
[18] Loo, Y.L., Tang, A.Y., Ahmad, A., 2015. The gap of current agent based simulation modeling practices and feasibility of a generic agent based simulation model. International Journal of Advanced Computer Research (IJACR) 5, 115–123.
[19] Macal, C.M., North, M.J., 2005. Tutorial on agent-based modeling and simulation, in: Proceedings of the Winter Simulation Conference, 2005., IEEE. pp. 14–pp.
[20] Naumann-Woleske, K., Knicker, M.S., Benzaquen, M., Bouchaud, J.P., 2021. Exploration of the parameter space in macroeconomic agent-based models. arXiv preprint arXiv:2111.08654 .
[21] Platt, D., 2020. A comparison of economic agent-based model calibration methods. Journal of Economic Dynamics and Control 113, 103859.
[22] Robles, J.F., Bermejo, E., Chica, M., Cordón, Ó., 2021. Multimodal evolutionary algorithms for easing the complexity of agent-based model calibration. Journal of Artificial Societies and Social Simulation 24.
[23] Saxena, R.N., 2022. Exploring Susceptible-Infectious-Recovered (SIR) model for COVID-19 investigation. SpringerBriefs in applied sciences and technology, Springer, Singapore.
[24] Schelling, T.C., 1969. Models of segregation. The American economic review 59, 488–493.
[25] Schelling, T.C., 1971. Dynamic models of segregation. Journal of Mathematical Sociology 1, 143–186. doi:10.1080/0022250X.1971.9989794.
[26] Smaldino, P.E., Calanchini, J., Pickett, C.L., 2015. Theory development with agent-based models. Organizational Psychology Review 5, 300–317.
[27] Stonedahl, F., Wilensky, U., 2011. Finding forms of flocking: Evolutionary search in abm parameter-spaces, in: Multi-Agent-Based Simulation XI: International Workshop, MABS 2010, Toronto, Canada, May 11, 2010, Revised Selected Papers 11, Springer. pp. 61–75.
[28] Tran, M., Ngo, M., Pham-Hi, D., Bui, M., 2020. Bayesian calibration of hyperparameters in agent-based stock market, in: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE. pp. 1–6.
[29] Yu, H., Kim, S., 2012. Svm tutorial-classification, regression and ranking. Handbook of Natural computing 1, 479–506.
[30] Zhang, H., Vorobeychik, Y., Letchford, J., Lakkaraju, K., 2016. Data-driven agent-based modeling, with application to rooftop solar adoption. Autonomous Agents and Multi-Agent Systems 30, 1023–1049.