Essay

Not peer-reviewed version

Perspectives: Two Approaches in Computational Social Sciences

Shiping Tang *

Posted Date: 22 October 2024

doi: 10.20944/preprints202410.1654.v1

Keywords: computational social sciences; machine learning; artificial intelligence; agent-based modeling doi:10.20944/preprints202410.1654.v1

Essay

Perspectives: Two Approaches in Computational Social Sciences

Shiping Tang

Center for Complex Decision Analysis, Fudan University, Shanghai, China; twukong@yahoo.com

Summary: Machine learning and agent-based modeling are fundamentally different but can complement each other.

Abstract: Big data-driven machine learning and artificial intelligence (ML/AI) is not all computational social sciences (CSS) have. Agent-based modeling (ABM) or multi-agent system (MAS) is another fundamentally different but equally useful approach in CSS. In fact, the two approaches start from very different orientations and their differences have deeper root in ontology. ML/AI aims to imitate and then surpass human capacities, from sensing to perceiving, reasoning, calculating, and acting. In contrast, ABM seeks to simulate how social outcomes emerge from the complex interactions of agents' actions within a specific environment. Yet, precisely because these two technologies are different, they can be complementary to each other. There is a bright future for integrating ABM with ML/AI for tackling real world challenges.

Keywords: computational social sciences; machine learning; artificial intelligence; agent-based modeling

---

Big data-driven machine learning (ML) and artificial intelligence (AI) has dominated much of the discussion on computational social sciences (CSS).[1] In fact, many have essentially equated CSS with ML/AI as if other approaches do not exist or are mostly irrelevant, not only in popular forum but also in academic discourse by prominent computational social scientists (e.g., Athey & Imbens 2019; Brady 2019; Lazer et al. 2009; 2020; Lazer & Radford 2017; Molina & Garip 2019; Edelmann et al. 2020; Grimmer et al. 2021; Hofman et al., 2021; *Nature*, Vol. 595, No. 7866). Unfortunately, this understanding is one-sided, to say the least.

In reality, CSS contains another fundamentally different but equally useful approach, that is, agent-based modeling (ABM) or multi-agent system (MAS). Evidently, the second approach has been lost in the discussion on CSS centered upon big data and ML/AI.

This essay argues that the two approaches start from very different orientations and their differences have deeper root in ontology. Yet, precisely because these two technologies are different, they can be complementary to each other. The key is to first grasp the differences of these two approaches and then understand why they can be complementary to each other.

---

[1] There is no consensus regarding the definition of CSS, and existing definitions tend to be somewhat narrow (e.g., Lazer et al 2009; Cioffi-Revilla 2014). I venture to propose a broad and inclusive definition: CSS is an interdisciplinary and rapidly expanding field that seeks to better understand and forecast social actions and social outcome with intensive computation. Drawing theoretical and empirical knowledge from social sciences for modeling and programming, CSS relies on mathematics, computer sciences, and data sciences as key technological tools. CSS is almost always based on data and computation at a large scale, and the aim of CSS is usually, but not exclusively, to uncover and detect patterns and forecast actions and outcomes.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2024 doi:10.20944/preprints202410.1654.v1

ML/AI vs. ABM: Imitation vs. Simulation

In a nutshell, ML/AI aims to imitate and then surpass human capacities, from sensing to perceiving, reasoning, calculating, and acting. In contrast, ABM seeks to simulate how social outcomes emerge from the complex interactions of agents' actions within a specific environment (see Table 1 below).

Table 1. ML/AL vs. ABM, or, imitation vs. simulation.

Item	Imitation: ML+AI	Simulation: ABM
Foundation	Statistical learning	Multiple foundations, with statistical learning being one of them.
Mathematical forms of key equations	A single type of equation: $y = f(X) + \epsilon$.	An ABM system almost always contains more than one type of equation.
Principle of model improvement: Connection with social sciences	Train models with train set data, improve models, and then use refined models to forecast outcomes (y) with causing variables (X). The key is to reduce the error term (ϵ) to as little as possible. Hence, model improvement mostly depends on more and better data. Connection with social sciences is loose.	Use simulation with historical data to approximate historical outcomes and refine the system. The refined system is then deployed for forecasting based on projected data. Besides more and better data, model improvement depends on insights from social sciences. Connection with social sciences is tight.
Sources of data	Mostly big data	Both basic data and big data
Integration of data from different levels	Low	High
Integration with other platforms	Less easy	Easy
System: linear or evolutionary	Mostly linear	More evolutionary.

Since the coming of deep learning, ML/AI has been making tremendous progresses and found wide applications in the real world (reviewed by LeCun, Bengio, & Hinton 2015). Unsurprisingly, ML/AI has come to dominate discourses within CSS.

In contrast, partly due to the limitation imposed by its critical dependence on social sciences theories and empirics plus computational power when working with complex , ABM has been mostly deployed for verifying factors and mechanisms (including interactions) behind complex emergent social outcomes such as segregation, electoral outcomes, civil war, interstate competition, the evolution of social networks, and the evolution of industrial clusters etc. (for a review, see Cioffi-Revilla 2014, chap. 10). Rarely has ABM been utilized for forecasting complex social outcomes in the real world.

With the accumulation of knowledge by social sciences and the recent advancement in computational power, ABM can now be deployed for simulating and then forecasting complex social outcomes in the real world. For example, Magliocca et al. (2019) simulate the ongoing cat-and-mouse

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2024 doi:10.20944/preprints202410.1654.v1

game between drug-trafficking networks and U.S. drug enforcement agency (DEA) and other law enforcement agencies. Their ABM simulations may help U.S. agencies design better countermeasures against drug trafficking. From our own lab, we have been deploying an ABM-based platform from forecasting elections in U.S. and Taiwan, without relying on public opinion data and social media data, and what we forecast is the relative share of votes rather than simply which party or candidate will win. Through five experiments of forecasting with results publicly released ahead of the actual elections, our forecasting has been shown to be remarkably accurate (Gao et al., 2021).[2] These recent advancements showcase ABM's promises in tackling forecasting in and for the real world.

ML/AI vs. ABM: Action/Behavior vs. Outcome

The differences between ML/AI and ABM have even deeper roots. Ontologically, there are three broad objects in social sciences: idea, action, and outcome. Whether a potential voter likes a candidate is an idea; whether s/he votes for the candidate is action, and whether the candidate wins is an outcome. Similarly, whether a country's leader wants to engineer economic development is an idea, whether s/he and her/his people work hard is action, and whether the country achieves development is an outcome.

Apparently, idea drives action, action drives outcome (and ideas), and outcomes shapes both ideas and actions. Very critically, while action drives any outcome, it does so only partly: whether a candidate wins or not depends only partly upon whether s/he has campaigned effectively. In short, the three objects are ontologically different. As a result, understanding, explaining, and forecasting these three objects face quite different epistemological and methodological challenges (for a more detailed discussion, see Tang 2016).

Many leading authors of CSS have not explicitly differentiated these three objects, especially action and outcome. Worse, they may have essentially equated (forecasting) human action with social outcome or assumed that actions are all CSS is about. For example, the word "outcome" does not appear in Lazer et al. (2009) and Lazer & Radford (2017) at all. Likewise, Hofman et al. (2017, 487) asked "how predictable is human behavior?" and then discussed both actions (e.g., whether a user likes or retweets a tweet) and outcomes (e.g., "black swan" events). Most recently, Hofman et al. (2021) have again lumped together action and outcome.

Once we admit that action and outcome are different, we can immediately admit that ML/AI and ABM may have different strengths and weaknesses in forecasting the two objects. Straightforwardly, because ABM explicitly admits that agents' actions only partly drive the emergent outcomes within a social system, it is better equipped for forecasting outcomes, at least in principle. In contrast, because ML/AI aims to imitate and then surpass human capacities, it is not well equipped for forecasting emergent social outcomes, as many have pointed out (e.g., Cederman & Weidmann 2017; Chenoweth & Ulfelder 2017; Bowlsby et al. 2020). Rather, ML/AL may be better equipped for gauging and forecasting human actions, especially in relatively stable settings, such as consumer behavior, social media behavior, and perhaps some social outcomes directly driven by actions (e.g., war driven by one state attacking another state).

ML/AI vs. ABM: Strength and Weakness

The two approaches also have different strengths and weaknesses in seven different areas.

Foundations and forms of equation. The foundation of ML/AI is statistical learning. The core model of ML/AI is a single type of equation: $y = f(X) + \epsilon$. In contrast, statistical learning is only one of ABM's foundations. Moreover, because an ABM system contains multiple equations (and parameters) that capture agents' actions, how agents interact with each other, how agents interact with the environment, and how the eventual output is derived, an ABM system almost always contains equations with different mathematical forms (e.g., Magliocca et al. 2019).

---

[2] Indeed, we completed our forecasting several months ahead of the actual elections. To avoid impacting the actual voting, however, we withheld our forecasting results until several days before the actual election date.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 October 2024 doi:10.20944/preprints202410.1654.v1

Implementation. Once a ML/AI algorithm is developed, it is relatively easy to implement it, partly because it is less dependent on social sciences theory and empirical results, including expert knowledge. In fact, most ML/AI algorithms have been developed exclusively by computer scientists and mathematicians, with little input from social scientists. After all, a key goal of ML/AI is to minimize, if not replace, human judgment when it comes to detecting and forecasting. Hence, even for quite different problems, the basic underlying principles of models are quite similar. In short, the connection of ML/AI with social sciences is loose, or at least not as tight as that of ABM with social sciences.

In contrast, an ABM system is almost always more complex and more difficult to implement than ML/AI, because ABM requires extensive input from social sciences theory and empirical results, including expert knowledge. Indeed, without input from social sciences theory and empirical results, no real world-relevant (or meaningful) ABM platform can be developed. Moreover, it is often necessary to design quite different simulation systems for different problems. In short, the connection of ABM with social sciences is much tighter.

Model improvement: data-driven vs. theoretical driven. Once a ML/AI algorithm is developed, the improvement of its performance is mostly driven by more and better data. In contrast, improving the performance of an ABM systems depends on not only more and better data, but also upon better theoretical assumptions and empirics. A better ABM model is usually developed by conceiving a new solution of a complex emergent outcome and then deploying theories and empirics to build upon the model.

Source of data: dependence on big data. Most data for training ML/AI algorithms come from big data. In fact, without the availability of big data, ML/AI could not have possibly progressed as it has today. In contrast, almost all ABM systems require input from basic data such as agent properties, parameters capturing the environment, etc., and more often than not, these data are not big data. Hence, ABM depends on big data less than ML/AI does.

Data integration across different levels. Even for ensembles and hierarchical models, the degree of data integration across different levels (i.e., micro, meso, and macro) by ML/AI is relatively low. The best hierarchical models can do is to use interaction terms, but interaction terms with more than three variables are very difficult to interpret. In fact, most ML/AI models do not deploy interaction terms, at least not widely, and certainly not rigorously.

In contrast, ABM can integrate data from different levels into the same ABM system more readily because one of ABM's key principle is to simulate how micro-level inputs (i.e., behaviors) interact with each other but also other environmental constraints (i.e., meso and macro factors) to drive emergent outcomes.

Integration with other platforms. Different ML/AI algorithms can certainly be combined (as an ensemble) for solving a problem, but they do not integrate with other technological systems that easily. Here, by integration, we do not mean that data from other technologies are input as mere data, but rather that ML/AI and other technologies platforms work together to solve a problem. Understood as such, ML/AI cannot be easily integrated with other technologies, such as game theory and qualitative comparative analysis (QCA). Partly because ABM is a simulation system, it can be easily integrated with other technologies, such as dynamic social network analysis (D-SNA), game theory, QCA, and of course, ML/AI (de Marchi & Page 2014). In fact, we see a bright future for integrating ABM with ML/AI (Lamperti et al. 2018; van der Hoog 2019).

Social system: evolutionary or not? Although ML/AI algorithms can definitely improve (thus can be understood metaphorically and loosely as "evolve"), one of ML/AI's central assumptions is that the social system is a mostly linear system. In contrast, from the very beginning, ABM is an evolutionary enterprise. In fact, the whole ABM enterprise has been built upon the assumption that the human society is an evolutionary system. Unsurprisingly, the two earliest applications of ABM were all about evolutionary systems: Thomas Schelling's (1971) "dynamic segregation" system and Robert Axelrod's (1984) *Evolution of Cooperation*.

Conclusion

CSS is undoubtedly a key direction for the future of social sciences. Yet, the notion that big data-driven ML/AI is all what CSS is about is one sided: ABM is another different but equally useful approach. Once we grasp that ML/AI and ABM are two major different approaches in CSS, it is easy to conclude that ML/AI and ABM can be fruitfully combined. More concretely, when setting agents' behavioral rules, and to a less extent, when setting rules for the system evolution (or environment) in ABM, ML/AI can contribute significantly.[3] ABMs with inputs from ML/AI can perform better than ABM alone or ML/AI alone, especially when it comes to understanding and forecasting important social outcomes in the real world. But first, we have to explicitly grasp the key differences of the two approaches, including their respective strengths and weaknesses.

References

Athey, S., and Imbens, G. W. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11: 685–725 (2019).

Brady, H. E. "The Challenge of Big Data and Data Science." *Annual Review of Political Science* 22: 297-323 (2019).

Lazer, D. et al. Computational Social Science. *Science* 323: 721-723 (2009).

Lazer, D. et al. Computational Social Science: Obstacles and opportunities. *Science* 369: 1060-1062 (2020).

Lazer, D., & Radford, J. Data ex Machina: Introduction to Big Data. *Annual Review of Sociology* 43: 19-39 (2017).

Molina, M. and Garip, F. "Machine Learning for Sociology," *Annual Review of Sociology* 45: 27-45 (2019).

Edelmann, Achim, Wolff, T., Montagne, D., & Bail, C. A. "Computational Social Science and Sociology," *Annual Review of Sociology* vol. 46 (2020)

Grimmer, J., Roberts, M. E., and Stewart, B. M. "Machine Learning for Social Science: An Agnostic Approach," *Annual Review of Political Science*, vol. 24: 395–419 (2021).

Hofman, J. M. et al. "Integrating explanation and prediction in Computational Social Science," *Nature* 595: 181-188 (2021).

LeCun, Yann, Yoshua Bengio and Geoffrey Hinton, "Deep Learning," *Nature* 521: 436-444 (2015).

Cioffi-Revilla, C. *Introduction to Computational Social Science*. Springer (2014).

Magliocca, N. R. et al., "Modeling Cocaine Traffickers and Counterdrug Interdiction Forces as a Complex Adaptive System," *Proceedings of the National Academy of Sciences of the United States of America* 116: 7784-7792 (2019).

Gao, M. et al. "Forecasting elections with ABM: two live experiments," unpublished manuscript, DOI: 10.13140/RG.2.2.14605.87528 (2021)

Tang, S. 2016. "Idea, Action, and Outcome: The Objects and Tasks of Social Sciences", working paper, DOI: 10.2139/ssrn.2790615

Hofman, J. M., Sharman, A., and Watts, D. J. "Prediction and explanation in social system," *Sciences* 355: 486-488 (2017).

Cederman, L-E. & Weidmann, N. B. "Predicting Armed Conflict: Time to Adjust Our Expectations?" *Science* 355: 474-476 (2017).

Chenoweth, E. & Ulfelder, J. "Can Structural Conditions Explain the Onset of Nonviolent Uprisings?" *J. Conflict Resolution* 61: 298-324 (2017).

Bowlsby, D., Chenoweth, E., Hendrix, C., Moyer, J. D. The Future is a Moving Target: Predicting Political Instability. *British Journal of Political Science* 50: 1405-1417 (2020).

[3] In dynamic social network analysis, stochastic agent-oriented modeling (SAOM) can be understood as a method that combines ML/AI with simulation (though not exactly ABM). See Snijders 1996.

doi:10.20944/preprints202410.1654.v1

de Marchi, S., & Page, S. E. "Agent-Based Models," *Annual Review of Political Science*, vol. 17: 1-20 (2014).

Lamperti, F., et al., "Agent-based model calibration using machine learning surrogates," *Journal of Economic Dynamics and Control*, 90: 366-389 (2018).

van der Hoog, S. "Surrogate Modelling in (and of) Agent-Based Models: A Prospectus," *Computational Economics*, 53 (3): 1245-1263 (2019).

Schelling, T. C. "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1 (2): 143-186 (1971).

Axelrod, R. *The Evolution of Cooperation*. New York: Basic Books (1984).

Snijders, T. AB. "Stochastic actor-oriented models for network change." *Journal of Mathematical Sociology*, 21 (1–2): 149– 172 (1996).