Data-Driven Agent-based Modeling of Innovation Diffusion (Doctoral Consortium)

Haifeng Zhang
Supervisor: Yevgeniy Vorobeychik
Electrical Engineering and Computer Science
Vanderbilt University
Nashville, USA
haifeng.zhang@vanderbilt.edu

ABSTRACT

We present a novel data-driven agent-based modeling framework to study innovation diffusion. Our first step is to learn a model of individual agent behavior from individual adoption characteristics. We then construct an agent-based simulation with the learned model embedded in artificial agents, and proceed to validate it using a holdout sequence of collective adoption decisions. Finally, we exemplify the proposed method can be used to explore and analyze a broad class of policies aimed at spurring innovation adoption.

Keywords

Machine Learning; Agent-based Modeling; Innovation Diffusion; Policy Optimization

1. INTRODUCTION

Rogers' [13] theory of innovation diffusion aims to explain how, why, and at what rate new ideas and technology spread through social systems. Bass outlined essence the theory and proposed one of the most influential diffusion models [2]. However, models of this kind treat diffusion at aggregate-level. It hardly handle individual-level data missing the key to understand innovation adoption.

Agent-based modeling (ABM) is introduced to study aggregate properties of complex systems arising from micro behaviors [3, 10]. Moreover, the emergence of "Big Data" offers new opportunities to develop agent-based models entirely data-driven. Data from various sources can be combined to make a high-fidelity dataset and train agent behavior models using machine learning techniques. We propose a novel data-driven agent-based modeling framework for study of innovation diffusion, which can be quantitatively validated and reliably used for policy analysis.

2. RELATED WORK

While typical "agent-base" approach uses simple agent models to derive complexity from individual interactions, our method departures from this treatment to developing sophisticated predictive agent models based on empirical data entirely. It is novel in the field of innovation diffusion, i.e., none of state-of-the-art agent-based models is developed by rigorous machine learning techniques [4, 8, 12, 11, 9, 17].

Three related efforts are somewhat closer in spirit to our method. Kearns and Wortman [6] develop a theoretical model of learning from collective behavior, which however does not address the general problem of learning from a single observed sequence of collective behavior. Judd et al. [5] use machine learning to predict behavior of participants in social network coordination experiments, but are only able to match the behavior qualitatively. Torrens [14] uses machine learning to calibrate individual walking models from real and synthetic data, which however does not consider the subsequent problem of policy evaluation and optimization.

3. DATA-DRIVEN AGENT-BASED MODELING

We now propose a general framework, *data-driven agent-based modeling (DDABM)*, which is introduced to efficiently learn agent models from a sequence of individual behaviors. The method explicitly divides data into "calibration" and "validation" to ensure sound and reliable model validation and automates agent model training by cross-validation. [1]

We start with a data set of individual agent behavior over time, $D = \{(x_{it}, y_{it})\}_{i,t=0,\ldots,T}$, where i indexes agents, t time through some horizon T and y_{it} indicates agent i's decision, i.e., 1 for "adopted" and 0 for "did not adopt" at time t.

1. Split dataset D into *calibration* D_c and *validation* D_v parts along the time horizon: $D_c = \{(x_{it}, y_{it})\}_{i,t \le T_c}$ and $D_v = \{(x_{it}, y_{it})\}_{i,t>T_c}$ where T_c is a threshold.
2. Learn a model of agent behavior h on D_c. Use cross-validation on D_c for model (e.g., feature) selection.
3. Instantiate agents in ABM using h learned in step 2.
4. Initialize the ABM to state x_{jT_c} for all agents j.
5. Validate the ABM by running it from x_{T_c} using D_v.

We applied the DDABM in the context of spatial-temporal solar adoption dynamics in San Diego county [16]. Figure 1 (left) illustrates that the agent-based model successfully

[1] We assume: a) discrete time, b) homogeneous agent and c) independent decision-making at any time t, conditional on state x.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

forecasts solar adoption trends and provides a meaningful quantification of uncertainty about its predictions. Moreover, *likelihood ratio* in Figure 1 (right) shows that our model significantly outperforms a baseline model.

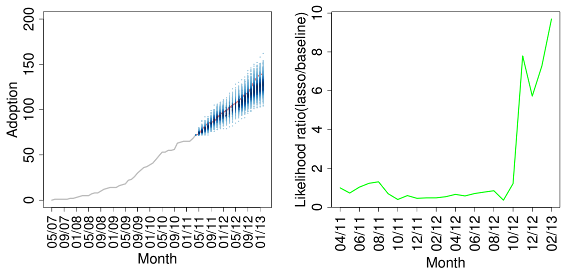

Figure 1: Left: likelihood ratio R of our model (lasso) relative to the baseline. Right: spread of sample runs of our model, with heavier colored regions corresponding to higher density, and the observed average adoption trend.

4. POLICY ANALYSIS

The proposed DDABM framework can support a variety of policy experiments. Generally, agent model would include features, such as, temporary economic variable, peer measures, individual characteristics etc. A policy could leverage these economic variables, i.e., subsidy programs, group buy discount etc. Based on peer effect, seeding policy, i.e. giving away free systems, can be designed and evaluated. Targeted marketing strategies, that aim to target influential subpopulation based on demographics is also testable. In addition, finding optimal policy can be highly complex, not only because the model is data-driven, but also multi-agent simulation is heterogeneous and nonlinear. Our work reveal that simple algorithm can be developed to find optimal seeding policy in a general dynamic influence maximization setting, but however it loses efficacy to other heuristics subject to a more realistic model [15].

5. CONCLUSIONS

We introduced a DDABM framework demonstrating its efficacy in modeling rooftop solar adoption. The model was validated quantitatively and shown to support analysis of a variety of policy schemes. In future, graphical models, i.e. Bayesian networks, can be a remedy to avoid estimation of multiple unknown variables to fit a logistic regression model. Thus, a real-time decision support system based upon probabilistic inference and influence diagrams can be envisioned [7]. Moreover, design of efficient algorithms to find optimal or near-optimal policy is indeed necessary. Reinforcement learning algorithms might be used to estimate action utilities and speed up the search [1]. Finally, we would like to apply the developed DDABM framework in a different domain of innovation diffusion.

Acknowledgments

This work was partially supported by the U.S. Department of Energy (DOE) office of Energy Efficiency and Renewable Energy, under the Solar Energy Evolution and Diffusion Studies (SEEDS) program.

REFERENCES

[1] A. G. Barto. *Reinforcement learning: An introduction.* MIT press, 1998.

[2] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[3] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Supp 3):7280–7287, 2002.

[4] A. Borghesi, M. Milano, M. Gavanelli, and T. Woods. Simulation of incentive mechanisms for renewable energy policies. In *European Conference on Modeling and Simulation*, 2013.

[5] S. Judd, M. Kearns, and Y. Vorobeychik. Behavioral dynamics and influence in networked coloring and consensus. *Proceedings of the National Academy of Sciences*, 107(34):14978–14982, 2010.

[6] M. Kearns and J. Wortman. Learning from collective behavior. In *Conference on Learning Theory*, 2008.

[7] T. D. Nielsen and F. V. Jensen. *Bayesian networks and decision graphs.* Springer Science & Business Media, 2009.

[8] J. Palmer, G. Sorda, and R. Madlener. Modeling the diffusion of residential photovoltaic systems in italy: An agent-based simulation. Working paper, 2013.

[9] V. Rai and S. Robinson. Determinants of spatio-temporal patterns of energytechnology adoption: An agent-based modelingapproach. Working paper, 2014.

[10] W. Rand and R. Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3):181–193, 2011.

[11] S. Robinson and V. Rai. Agent-based modeling of energy technology adoption:empirical integration of social, behavioral, economic,and environmental factors. Working paper, 2014.

[12] S. Robinson, M. Stringer, V. Rai, and A. Tondon. GIS-integrated agent-based model of residential solar pv diffusion. Working paper, 2013.

[13] E. M. Rogers. *Diffusion of Innovations.* Free Press, 5th edition, 2003.

[14] P. Torrens, X. Li, and W. A. Griffin. Building agent-based walking models by machine-learning on diverse databases of space-time trajectory samples. *Transactions in GIS*, 15(s1):67–94, 2011.

[15] H. Zhang, A. Procaccia, and Y. Vorobeychik. Dynamic influence maximization under increasing returns to scale. In *Proceedings of the 13th Intl. Joint Conference on Autonomous Agents and Multiagent Systems*, 2015, to appear.

[16] H. Zhang, Y. Vorobeychik, J. Letchford, and K. Lakkaraju. Data-driven agent-based modeling, with application to rooftop solar adoption. In *Proceedings of the 13th Intl. Joint Conference on Autonomous Agents and Multiagent Systems*, 2015, to appear.

[17] J. Zhao, E. Mazhari, N. Celik, and Y.-J. Son. Hybrid agent-based simulation for policy evaluation of solar power generation systems. *Simulation Modelling Practice and Theory*, 19:2189–2205, 2011.