Reinforcement Learning in agent-based modeling to reduce carbon emissions in transportation

Yuhao Yuan*
Systems Engineering
University of California, Berkeley

Felipe Leno da Silva*
Computational Engineering Division
Lawrence Livermore National Laboratory

Ruben Glatt
Computational Engineering Division
Lawrence Livermore National Laboratory
glatt1@llnl.gov

Abstract

This paper explores the integration of reinforcement learning (RL) into transportation simulations to explore system interventions to reduce greenhouse gas emissions. The study leverages the *Behavior, Energy, Automation, and Mobility (BEAM)* transportation simulation framework in conjunction with the *Berkeley Integrated System for Transportation Optimization (BISTRO)* for scenario development. The main objective is to determine optimal parameters for transportation simulations to increase public transport usage and reduce individual vehicle reliance. Initial experiments were conducted on a simplified transportation scenario, and results indicate that RL can effectively find system interventions that increase public transit usage and decrease transportation emissions.

1 Introduction

The transportation sector is a primary driver of greenhouse gas (GHG) emissions [6]. Presently, technological advancements are catalyzing a monumental shift in transportation, transitioning the energy source from mainly fossil fuels to electric power. As a consequence, there's an imminent need to reevaluate our current infrastructure, build new capabilities, and formulate progressive policies. In the last years, extensive data gathering and advances in modeling have enhanced our capability to forecast mobility trends. Techniques such as agent-based modeling (ABM) and simulation are now increasingly used for transportation scenario evaluations [5].

Public transportation will play a pivotal role in reducing GHG emissions in any future transportation scenario. Modern public transit vehicles tend to be more energy-efficient per passenger compared to private vehicles. If more individuals opt for public transit, it could also lead to a notable reduction in traffic congestion, subsequently leading to fewer cars idling and releasing GHGs [15]. The reduced dependence on private vehicles not only lowers the emissions from their direct use but also from their production, maintenance, and disposal. Additionally, public transport systems foster more compact, pedestrian-friendly urban developments, facilitating shorter, less carbon-intensive trips [13]. Infrastructure-wise, transportation methods like subways and light rail require less land than roads meant for personal vehicles, offering potential carbon sequestration if the saved land is employed for environmental purposes. Furthermore, it is more feasible to transition large-scale public transport systems, such as buses or trains, to renewable energy sources like solar or wind than transitioning millions of individual cars, making public transport even greener as our energy grids evolve.

*equal contribution

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

In this paper, we are interested in the question of how we can utilize reinforcement learning (RL) [14] to build a tool that can ultimately serve to assist urban planners. For this purpose, we used the *Behavior, Energy, Automation, and Mobility (BEAM)* [12] transportation simulation framework developed at the Lawrence Berkeley National Laboratory combined with the *Berkeley Integrated System for Transportation Optimization (BISTRO)* [2] scenario and optimization wrapper, and built an RL environment that allowed us to perform system interventions in parallel, de-centralized simulation runs. The resulting framework allows us to train agents that are capable of optimizing transportation objectives based on a given metric under fixed behavior and transit profiles for individual agents. In this initial experiments, we are limiting our focus on a simplified transportation scenario with few action dimensions to show the feasibility of the approach.

2 Agent-based modeling with BEAM and BISTRO

ABM frameworks [1] have evolved as a valuable tool with growing capabilities and versatility to interactively estimate the effect of different actions and interventions in communities. Based on real-world data, ABMs allow city planners to model a given city or district based around its existing transportation and energy infrastructure, population, and vehicular make-up. Starting from an existing model, different interventions to the existing infrastructure can be applied and simulations of how the population is affected by these interventions are carried out. Those tools allow us to anticipate several aspects of policy or infrastructure interventions, e.g., how people adjust their mode preferences and movement patterns, how EV charging affects the energy grid and transit, and in general anticipate demand changes for public transport, road usage and the electrical grid.

For our experiments, we are using BEAM and BISTRO; BEAM is an agent-based transportation simulation framework which can investigate the impacts of new and emerging forms of transportation and BISTRO provides a Python based interface to BEAM to enable scenario development and an optimization framework to identify system interventions that best align with policy and planning objectives. BISTRO can be leveraged for activity-based travel models that integrates optimization algorithms to automate the search for potential optimal policy parameters evaluated on sets of defined Key Performance Indicators (KPI) over the diverse transportation system, such as congestion, travel cost, and GHG emissions. It enables the simulation capabilities of hundreds of thousands of synthetic agents of urban commuters with unique survey-based socio-demographic characteristics and commute plan to be executed during the day. Agents can select desired travel modes with re-planning capabilities to complete their trips to and from activities and improve their mode choice decision at each iteration until desired convergence condition. Consequently, BISTRO facilitates the algorithmic optimization of transportation system intervention strategies (e.g., public transit scheduling, changing vehicle fleet mixes, and altering rate structures for various modes).

Each BISTRO run environment is configured using a set of fixed, scenario-dependent data that defines the transportation supply elements (street network, vehicle configuration, transit schedule, etc) and demand elements (agent characteristics and commute plans). Once a metropolitan scenario is developed and calibrated based on travel demand and survey data, users can deploy BISTRO for distributed development of algorithms that optimize a feasible set of policy and investment decisions.

We used the Sioux Faux scenario for our experiments, see Figure 1. It is based on Sioux Falls scenario commonly used as benchmark in ABM research introduced already in 1973 [10]. We used a population sample size of 15,000 agents as a trade off among population size, behavioral realism, and computational complexity The transportation network for the Sioux Faux scenario includes a road network accessible for walking, private and shared vehicles, and public transit vehicles operated by a single transit agency with 12 transit routes throughout the city. The shared vehicles implemented for Sioux Faux only allowed single-passenger rides from a fleet of on-demand ride vehicles that was distributed randomly across the road network at the start of each simulation run.

3 Reinforcement Learning environment

As BISTRO does not have a natural interface to interact with RL algorithms, we relied on RLlib [8] to build an RL environment capable of adjusting simulation run parameters and starting simulations in parallel to generate sufficient samples for the RL agent to learn. The multi-process distributed training, as well as the contained RL algorithms are making RLlib an excellent tool for quickly

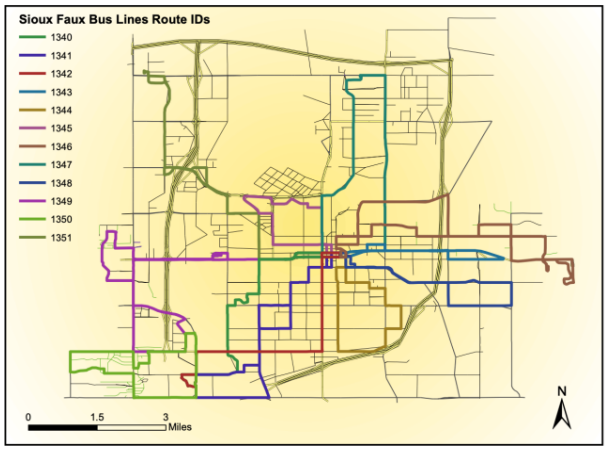

Figure 1: The Sioux Faux scenario is based on a simplified version of the city of Sioux Falls, South Dakota, transit and road network.

building versatile learning environments. The RL environment was set up in a way to enable either single agent [7] or multi-agent [16] experiments. In the single agent version, an RL algorithm is used to find a universal setting for all controllable entities in the environment, for example, a new bus fare for all lines at the same time, while the multi-agent setting would allow the setting of individual fares for each line. In this paper, we are only considering the single agent approach.

Each simulation as described above has an approximate run-time of 15 minutes, requiring the RL agent to learn from few samples even if many simulations can be run in parallel. For this reason, we selected the Proximal Policy Optimization (PPO) algorithm [11] as underlying learning algorithm as it has a high sample efficiency and can learn from fewer samples. PPO is a policy gradient method for RL [4] that stabilizes learning by clipping policy updates and avoiding large changes of a policy between updates. Its simplicity and ease of implementation have mode it one of the most used RL algorithms in the existing literature. PPO also allows to use either a discrete or continuous action space. This enabled us to run two experiments with slightly different settings while leaving much of the environment identical. We trained the algorithm with 5 samples per batch for 24 hours, where each sample represented a single simulation with 10 optimization cycles.

First, we formulated the problem as a state-less problem, where every simulation starts with the same configuration and the agents uses an action to set some parameters. Agents run a single simulation that ends the episode similar as in the bandit problem [9]. In this case, our actionable parameters are discrete values for the *public transit seat capacity* in the range between 0 and 40 and the *public transit standing capacity* in the range from 0 to 10. Actions are chosen jointly. Second, we allow the agent to adjust parameters sequentially. After the first action is chosen, the agent then sees the value for the first action as part of the state space and can make a better informed decision. Here, parameters for the simulation fall in the same range of discrete values, however, the agents select a value in continuous space between 0.0 and 1.0 which is then extrapolated and rounded to the closest discrete value.

In both cases, we used the same reward function. Every BISTRO simulation provides an output with scores for a number of different metrics. As our objective was to reduce the number of passengers in individual travel modes and increase the number of passengers in public transport, we used the *total number of public transport passengers* as a key metric for the reward calculation. For this purpose, we ran the simulation under the original settings to establish a reasonable baseline and note this target passenger number as p_T. At every simulation, we then collect the current passenger number p_c and define the reward for each simulation step t as $r(t) = \frac{p_c}{p_T}$. Under this formulation, a reward higher than 1.0 indicates a higher usage of public transport and is the desired behavior. It is important to note that the simulation optimization is multi-faceted , so just setting a high capacity would not necessarily lead to optimal results as different transit participants have different preferences in terms of travel duration, destination, and cost.

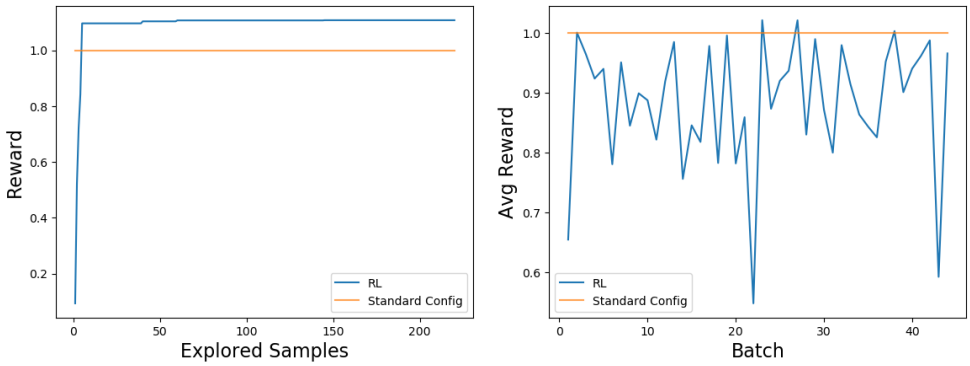

Figure 2: Reward development for PPO with discrete action space.

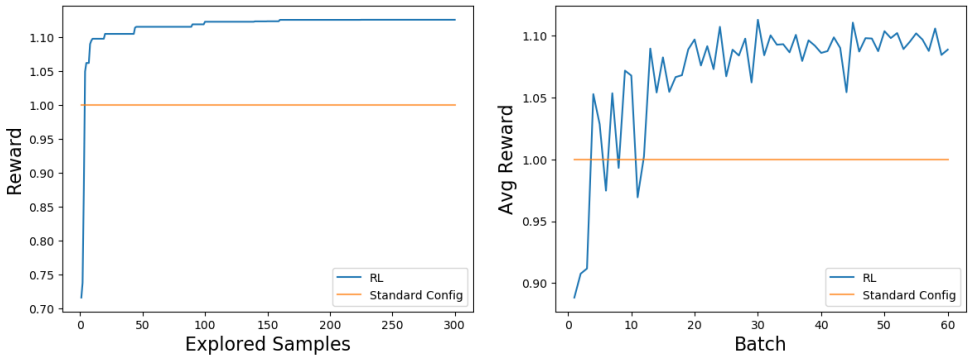

Figure 3: Reward development for PPO with continuous action space.

4 Results and Discussion

In the first setting, the algorithm was able to find better configurations than the standard settings but was not able to continuously improve the policy throughout the learning process as can be seen in Figure 2. However, with the continuous action space in the same range and the additional state information for the second action, PPO successfully learned to improve parameters compared to the original setting reliably after learning for a few episodes as seen in Figure 3.

Our Results indicate that RL is able to explore and find good simulation parameters to optimize given transit metrics that can ultimately help reduce transportation emissions, confirming our main hypothesis in this feasibility study. While we were only able to use a limited scenario and a restricted action set due to budget reasons, these results confirmed the feasibility of using an RL approach in ABM that could potentially support human-in-the-loop decision making by finding parameter settings that optimize a desired KPI in the simulation environment. Future work will focus on working out more localized KPIs to improve the reward signal in multi-agent learning. This will enable agents to better optimize their own objectives while keeping system aspects relevant. This could be further improved by leveraging transfer learning to speed up learning in more complex environments [3]. Eventually, this kind of simulation may be used as support tool for human planners to improve decision making in infrastructure planning by proposing policies or interventions that can effectively reduce GHG emissions.

Acknowledgments and Disclosure of Funding

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC. This work was supported by the Laboratory Directed Research and Development (LDRD) program under project tracking code 23-FS-025. LLNL-CONF-855111.

References

[1] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl_3):7280–7287, 2002.

[2] S. A. Feygin, J. R. Lazarus, E. H. Forscher, V. Golfier-Vetterli, J. W. Lee, A. Gupta, R. A. Waraich, C. J. Sheppard, and A. M. Bayen. Bistro: Berkeley integrated system for transportation optimization. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(4):1–27, 2020.

[3] R. Glatt, F. L. da Silva, R. A. da Costa Bianchi, and A. H. R. Costa. A study on efficient reinforcement learning through knowledge transfer. In *Federated and Transfer Learning*, pages 329–356. Springer, 2022.

[4] I. Grondman, L. Busoniu, G. A. Lopes, and R. Babuska. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):1291–1307, 2012.

[5] D. Harris, F. L. D. Silva, W. Su, and R. Glatt. A review on simulation platforms for agent-based modeling in electrified transportation. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–17, 2023.

[6] International Energy Agency (IEA), Paris. Greenhouse gas emissions from energy data explorer. *https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer*, 2023.

[7] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[8] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, and I. Stoica. Rllib: Abstractions for distributed reinforcement learning. In *International conference on machine learning*, pages 3053–3062. PMLR, 2018.

[9] A. Mahajan and D. Teneketzis. Multi-armed bandit problems. In *Foundations and applications of sensor management*, pages 121–151. Springer, 2008.

[10] E. Morlok, J. Schofer, W. Pierskalla, R. Marsten, S. Agarwal, J. Stoner, J. Edwards, L. LeBlanc, and D. Spacek. Development and application of a highway network design model, vols. 1 and 2. *Northwestern University. Final Report: FHWA Contract Number DOT-PH-11*, 1973.

[11] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[12] C. Sheppard, R. Waraich, A. Campbell, A. Pozdnukov, and A. R. Gopal. Modeling plug-in electric vehicle charging demand with beam: The framework for behavior energy autonomy mobility. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2017.

[13] M. Stevenson, J. Thompson, T. H. de Sá, R. Ewing, D. Mohan, R. McClure, I. Roberts, G. Tiwari, B. Giles-Corti, X. Sun, et al. Land use, transport, and population health: estimating the health benefits of compact cities. *The Lancet*, 388(10062):2925–2935, 2016.

[14] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[15] J. Woodcock, P. Edwards, C. Tonne, B. G. Armstrong, O. Ashiru, D. Banister, S. Beevers, Z. Chalabi, Z. Chowdhury, A. Cohen, et al. Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *The Lancet*, 374(9705):1930–1943, 2009.

[16] K. Zhang, Z. Yang, and T. Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pages 321–384, 2021.