Increasing performance of electric vehicles in ride-hailing services using deep reinforcement learning

Jacob F. Pettit, Ruben Glatt, Jonathan R. Donadee, Brenden K. Petersen
Computational Engineering Division
Lawrence Livermore National Laboratory
{pettit8, glatt1, donadee1, petersen33}@llnl.gov

Abstract

New forms of on-demand transportation such as ride-hailing and connected autonomous vehicles are proliferating, yet are a challenging use case for electric vehicles (EV). This paper explores the feasibility of using *deep reinforcement learning (DRL)* to optimize a driving and charging policy for a ride-hailing EV agent, with the goal of reducing costs and emissions while increasing transportation service provided. We introduce a data-driven simulation of a ride-hailing EV agent that provides transportation service and charges energy at congested charging infrastructure. We then formulate a test case for the sequential driving and charging decision making problem of the agent and apply DRL to optimize the agent's decision making policy. We evaluate the performance against hand-written policies and show that our agent learns to act competitively without any prior knowledge.

1 Introduction

The transportation sector is responsible for a significant share of all CO_2 emissions from human activity, representing 37% of emissions produced by California [1] and 29% of U.S. emissions [5]. Looking toward the future, the U.S. Department of Energy research shows that the proliferation of new forms of mobility, such as connected and autonomous vehicles (CAVs), could result in a 200% increase in energy consumption in the U.S. transportation sector by 2050 relative to their baseline [14]. The state of California has identified transportation electrification in tandem with transitioning to carbon free sources of electric power generation as a key strategy for reducing emissions from the transportation sector [1].

However, electric vehicles (EV) face many obstacles to reaching high adoption levels, especially for use in ride-hailing for transportation network companies (TNCs) such as Lyft and Uber or in future CAV based services. EVs have less driving range and longer refueling times than internal combustion powered vehicles, as well as limited charging infrastructure. A ride-hailing EV driver should consider many spatially and temporally varying factors when deciding when and where to charge. This includes time-varying electric energy prices and emissions, ride-hailing surge pricing, demand for transportation service, traffic conditions, and queues at charging stations. Each of these items can exhibit strong seasonality with time of day, geographic patterns, and randomness, creating opportunities to optimize timing and location of charging. An optimal driving and charging strategy (known as a *policy*) could help ride-hailing EVs to charge at the times of day when electric energy costs and emissions are lowest, to reduce time spent waiting in queues, and maximize a driver's revenues, reducing the economic barrier to adoption of EVs in ride-hailing and CAV fleets. Such a policy could be implemented as a decision support tool for human drivers, or as part of a CAV fleet management system. In this feasibility study, we focus on the case of optimizing the policy of a single agent as a first step, with its environment taken as exogenous and given.

33rd NeurIPS 2019, Workshop on Tackling Climate Change with Machine Learning, Vancouver, Canada.

2 Approach & Results

To investigate the potential to optimize policies for operation of ride-hailing EVs, we developed a data-driven simulation of a ride-hailing EV driving for a TNC. The simulation includes modules representing the TNC, EV, the network of EV charging stations, and the electric grid. The TNC randomly generates requested trips depending on the time of day and the EV's current location. Trip information includes destination, distance driven, time elapsed while driving, and transportation revenue earned. The probabilistic model of trips is estimated from the open source New York City taxicab dataset for the year 2015[1]. In our simulation, we consider trips within Manhattan, and have discretized the map into a grid of 120 zones. The EV tracks the battery state of charge, depleting the battery when driving or recharging the battery when charging, and has a battery capacity of 100kWh. The EV charging station network defines the locations of charging stations and models exogenous EV use of charging stations with an assumed pattern for queuing wait times. Charging stations are assumed to have a fixed charging power of 100kW. The electric grid determines the energy prices and CO_2 emissions produced per energy unit charged, depending on the time of day. California values are used for electric energy prices[2] and electric grid emissions[3].

The driving and charging decision making of a ride-hailing EV can be formulated as a Markov Decision Process (MDP) [9]. In the MDP framework, an agent receives some observation about the state of their environment, chooses an action, and receives some reward. Solving an MDP is equivalent to finding the decision making policy that maximizes an expected reward over the lifetime of the agent.

Our simulation environment is based on OpenAI Gym [3] with a discrete action space, where the EV agent chooses between charging the vehicle or accepting a new ride at every step. The agent's observation space is a vector consisting of the EV battery level, time of day, expected battery usage if the agent were to perform a ride, expected charging cost if the agent were to charge, expected emissions produced by choosing to charge, expected time when the vehicle would finish charging, and the expected queuing duration at the nearest charging station.

The reward function is defined as

$$r = \begin{cases} -(c + \epsilon E) & \text{if choosing to charge} \\ \xi & \text{if successfully provide a ride} \\ -3(c + \epsilon E) & \text{if attempting to provide a ride with insufficient battery} \end{cases}, \tag{1}$$

where c is the cost paid to charge the vehicle, ϵ is the emissions produced from charging the vehicle, E is a positive coefficient used to alter how much the agent considers the emissions produced from charging the vehicle, and ξ is revenue from completing a ride. E can be considered analagous to a carbon tax or carbon price in dollars per ton of CO_2 produced.

If the agent attempts to provide a ride with insufficient battery energy, it receives the reward $(-3(c + \epsilon E))$ and is forced to charge the vehicle. After the vehicle is charged, a new ride is generated and the episode continues. This penalty term encourages the agent to learn to charge the vehicle before the battery is depleted. Should the agent choose to charge the car, then it is assumed to drive to the nearest charging station and the car is charged until full. Finally, if the agent successfully provides a ride, it drives the passenger to the dropoff point and is paid for providing transportation service.

Reinforcement Learning (RL) [15] approaches to solving MDPs have been extended in recent years to what are known as Deep Reinforcement Learning (DRL) algorithms [2]. DRL has demonstrated superhuman performance on computer games [8], was used to beat the world's strongest players in Go [13], and has made progress towards knowledge sharing [6] and multiagent learning [12]. We believe that DRL is well suited to exploit the spatial and temporal arbitrage opportunities presented in the ride-hailing EV's driving and charging environment.

We applied the Trust Region Policy Optimization (TRPO) [11] DRL algorithm, as implemented by OpenAI baselines [4], to train an EV agent's policy for our environment. The policy and value

[1] Source ride data: `https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page`

[2] Source commercial energy prices: `https://caltransit.org/cta/assets/File/WebinarElements/WEBINAR-PGERateDesign11-20-18.pdf`

[3] Source emission data: `http://www.caiso.com/TodaysOutlook/Pages/Emissions.aspx`

[4] Algorithm implementation from `https://github.com/openai/baselines`

function networks are each two-layer perceptrons with 64 units per layer and hyperbolic tangent activation functions. Training batch size was set to 4096, and the discount factor was set to 0.8. Each simulation episode consists of one week of simulated time, and we assume that the vehicle provides uninterrupted service for the duration of the week.

For comparison, we also created heuristic policies that select when to charge based solely on the current battery level. Given a threshold λ, the handcrafted policies decide to charge when the battery level falls below λ, otherwise they accept the ride. We found $\lambda = 10\%$ to yield the highest performance, and we included baselines for $\lambda = 25\%$ and $\lambda = 50\%$ for additional comparisons.

Our preliminary experiments show that our agent was able to achieve favorable results over the best performing heuristic policies. In Figure 1 *(a)* we see that the trained agent is able to achieve higher average reward than the heuristic policies while being much more consistent, which has a clear positive impact on dollar per miles as shown in Figure 1 *(b)*. The development of the reward during training is shown in Figure 1 *(c)* where we see the average total episode reward over 2500 training episodes where each episode represents one week of driving. Figure 1 *(d)* shows the frequency of the agent's choice to charge by time of day. The agent learns to charge most frequently in the middle of the night when energy prices and congestion at charging stations are low; or occassionally around noon, when people are at work, energy prices are at their lowest daily point, and wait times at charging stations are short. The agent chooses not to charge at all during peak commute times, when surge pricing, charging wait times, and ride demand are greatest.

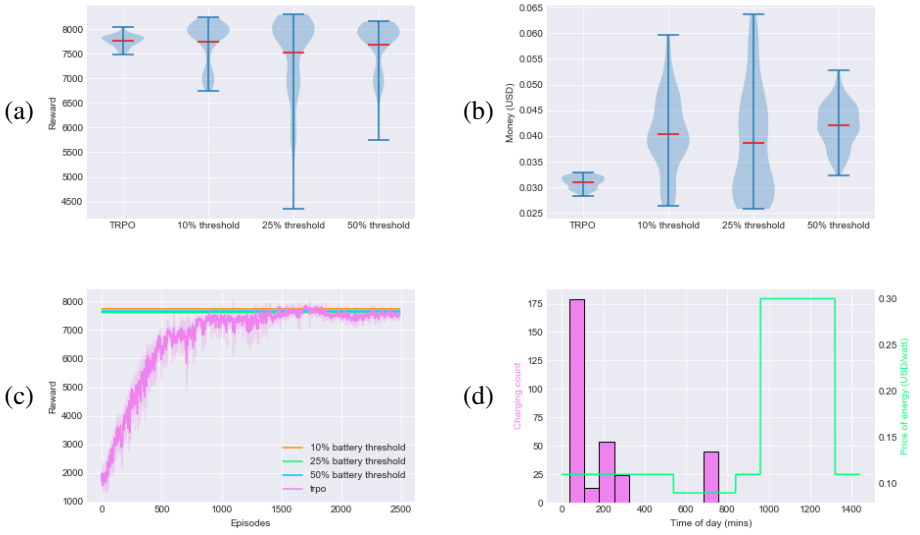

Figure 1: Experiment results: *(a)* Reward earned by the trained policy (TRPO) in evaluation runs. *(b)* Dollars spent on energy per mile driven in evaluation runs. *(c)* Moving average of episodic rewards experienced by the agent over training. Error bar represents one standard deviation. *(d)* Distribution of choosing to charge over the day peaks when energy prices are low.

3 Discussion

Use of electric vehicles in existing and emerging forms of transportation, such as ride-hailing, public transit, and CAVs, presents numerous planning and operational challenges. Lin et al. [7] explore multi-agent DRL agorithms for optimizing operations of ride-hailing fleets, but without considering the many unique challenges related to EV charging. Rossi et al. [10] studys the centralized operation of a joint electric-grid and EV transportation system according to meso-scale average rates of demand for trips. This approach doesn't address on-line operations in response to stochastic events. Chen et al. [4] follow a heuristic procedure to identify location and number of charging stations, and then evaluate the performance of ride-hailing EV agents following threshold policies. Our approach focuses on solving the immediate issues facing today's independent ride-hailing drivers that might adopt EVs, and intend to extend our work to future multi-agent CAV EV fleets. We believe that DRL based

approaches to optimizing the operation of EV fleets in transportation services remain an important research opportunity to be pursued. With further work on design of agent's observations, policy neural network architecture, and hyper parameter selection, performance can be greatly increased beyond what we have demonstrated here.

Acknowledgments

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. LLNL-CONF-789379.

References

[1] California's 2017 climate change scoping plan: The strategy for achieving california's 2030 greenhouse gas target. Technical report, California Air Resources Board, 2017.

[2] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

[3] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

[4] T Donna Chen, Kara M Kockelman, and Josiah P Hanna. Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice*, 94:243–254, 2016.

[5] United States Environmental Protection Agency (EPA). Sources of greenhouse gas emissions, 2019. URL `https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions`.

[6] Ruben Glatt, Felipe Leno da Silva, and Anna Helena Reali Costa. Towards knowledge transfer in deep reinforcement learning. In *Proceedings of the 5th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 91–96. IEEE, 2016.

[7] Kaixiang Lin, Renyu Zhao, Zhe Xu, and Jiayu Zhou. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1774–1783. ACM, 2018.

[8] V. Mnih, D. Silver, A. A. Rusu, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[9] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, New Yor, NY, USA, 2014.

[10] F. Rossi, R. D. Iglesias, M. Alizadeh, and M. Pavone. On the interaction between autonomous mobility-on-demand systems and the power network: models and coordination algorithms. *IEEE Transactions on Control of Network Systems*, pages 1–1, 2019.

[11] John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897, 2015.

[12] Felipe Leno da Silva, Ruben Glatt, and Anna Helena Reali Costa. Simultaneously learning and advising in multiagent reinforcement learning. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2017.

[13] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[14] T.S. Stephens, J. Gonder, Y. Chen, Z. Lin, C. Liu, and D Gohlke. Estimated bounds and important factors for fuel use and consumer costs of connected and automated vehicles. Technical Report NREL/TP-5400-67216, National Renewable Energy Laboratory, 2016.

[15] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.