Waymo Fleet Profitability Optimizer

Aksel Kretsinger-Walters   Alena Chan   Andrey Aksyutkin   Blake Sisson *

Problem Statement

One of the most promising and revolutionary applications of reinforcement learning is in the domain of autonomous robots, specifically self-driving cars logistics. The challenges facing the autonomous car community span multiple dimensions: including intellectual, ethical, financial, and technical hurdles. For our project, we narrow our focus on optimizing the profitability of robo-taxi fleets.

Monitoring, maintaining, and optimizing a large fleet of self driving cars is a complex problem, and one can quickly think of many dimensions that the problem takes on. Predicting demand, scheduling maintenance, recharging vehicles, setting competitive prices, maximizing coverage, minimizing wait times, and more are all separately non-trivial problems. Jointly optimizing across all of these dimensions and adapting to distribution shifts is an even more challenging problem, and the interactive nature of the problem makes it a natural fit for reinforcement learning and agentic approaches.

RL formulation

We plan to simulate a fleet of self-driving cars as a Markov Decision Process (MDP), and develop reinforcement learning algorithms to optimize the fleet's operations and profitability.

The state includes the number of cars available in each city region, the number of current rides, weather conditions, high-low intraday demand conditions, and fuel reserve for each car, distance to the nearest charge station. Action includes where to send the car when it is not with a user inside, when to go to gas station, and what price to demand for each user/district to balance demand.

Interest and Relevance

This is an area of active research and development, with many companies investing heavily in self-driving technology. For this technology to reach the market, it is critical to resolve the safety and reliability challenges that currently exist; however, the economic viability must also be solved for self-driving to reach its full potential. Given rapid introduction of Waymo in multiple cities, this problem is increasingly relevant.

For the purposes of this course, we believe that the interactive nature of the ride-sharing environment, the high dimensionality of the state and action spaces, the data and research publicly available, and the potential of performing far better than a heuristic algorithm make the robotaxi optimization challenge a good fit for our semester project.

*Collaborator(s): Aksel Kretsinger-Walters (`adk2164`), Alena Chan (`ac5477`), Andrey Aksyutkin (`aa5499`), Blake Sisson (`mbs2246`)

Previous Research

Ride-sharing in general is an active area of research in the reinforcement learning community [5]. There are existing data sets and open simulation environments.

Environment Data

Our research is based on publicly available datasets to model crowd behavior, weather conditions, and seasonal effects.

To model crowd behavior, we will utilize data from the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau. The LEHD data provide valuable information on workforce dynamics, which can serve as a proxy for crowd movement and density. Specifically, we will leverage the following public-use data products:

- **LEHD Origin-Destination Employment Statistics (LODES):** This data set provides detailed spatial information on job locations and worker residential locations, which is crucial to understanding commuting patterns and daytime population distributions.
- **Quarterly Workforce Indicators (QWI):** QWIs offer statistics on employment, job creation, and earnings for various demographic groups and industries. This data will be used to analyze the composition and temporal dynamics of the workforce.

For weather conditions, we will use historical meteorological data from the National Weather Service. This data will include variables such as temperature, precipitation, and wind speed.

Proposed solution

We use a simulation environment to produce states (traffic, demand, weather conditions) and develop reward model to balance wait times, profit and coverage. We plan to compare multiple algorithms: policy gradient, mulit-agent RL and other. We will evaluate efficiency of our approach against heuristic distribution policies.

References

[1] U.S. Census Bureau. "LEHD Origin-Destination Employment Statistics (LODES)." https://lehd.ces.census.gov/data/

[2] Sutton, R. S., & Barto, A. G. "Reinforcement Learning: An Introduction." MIT Press, 2018.

[3] Waymo. "Waymo Open Dataset." https://waymo.com/open/

[4] Xiao, I. "Reinforcement Learning Project: CitiBike." https://github.com/ianxxiao/reinforcement_learning_project/blob/master/Reports/Presentation_RL_citiBike_20180514.pdf

[5] Li, J., Li, X., & Wang, F. "Reinforcement Learning for Ridesharing: An Extended Survey." arXiv preprint arXiv:2102.11896, 2021.

[6] Lin, K., Zhou, M., Wang, J., & Li, L. (2018). *Efficient Large-Scale Fleet Management via Multi-Agent Deep Reinforcement Learning*. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018).

[7] Kim, H., & Kim, Y. (2020). *Optimizing Large-Scale Fleet Management on a Road Network using Multi-Agent Deep Reinforcement Learning with Graph Neural Network*. arXiv preprint arXiv:2011.06175. https://arxiv.org/abs/2011.06175

[8] O'Mahony, E., & Shmoys, D. (2015). *Data Analysis and Optimization for (Citi)Bike Sharing*. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015).

[9] Xu, Y., Zhang, R., Li, X., & Wang, F. (2021). *Towards More Efficient Shared Autonomous Mobility: A Learning-Based Fleet Repositioning Approach*. Proceedings of the AAAI Conference on Artificial Intelligence, 35(1), 1166-1174. https://ojs.aaai.org/index.php/AAAI/article/view/16033