Deep Reinforcement Learning for Ride-sharing Dispatching and Repositioning

Zhiwei (Tony) Qin[1,2*], **Xiaocheng Tang**[1,2], **Yan Jiao**[1,2], **Fan Zhang**[2], **Chenxi Wang**[1,3] and **Qun (Tracy) Li**[3]
[1]DiDi Research America, Mountain View, CA, USA
[2]DiDi AI Labs, Beijing, China
[3]DiDi Research, Beijing, China
{qinzhiwei, xiaochengtang, yanjiao, feynmanzhangfan, wangchenxi, liquntracy}@didiglobal.com

Abstract

In this demo, we will present a simulation-based human-computer interaction of deep reinforcement learning in action on order dispatching and driver repositioning for ride-sharing. Specifically, we will demonstrate through several specially designed domains how we use deep reinforcement learning to train agents (drivers) to have longer optimization horizon and to cooperate to achieve higher objective values collectively.

1 Introduction

Real-time single-passenger ride-sharing platform matches passengers and drivers instantly. Such technology has greatly transformed the way people travel nowadays. By optimizing decisions in both space and time dimensions, it offers more efficiency on the traffic management, and the traffic congestion can be further alleviated as well.

The decision system at a ride-sharing platform must make decisions both for assigning available drivers to nearby unassigned passengers (hereby called orders) over a large spatial decision-making region (e.g., a city) and for repositioning (dispatching) drivers who have no nearby orders. Such decisions not only have immediate to short-term impact in the form of revenue from assigned orders and driver availability, but also long-term effects on the distribution of available drivers across the city. This distribution critically affects how well future orders can be served. The need to address the exploration-exploitation dilemma as well as the delayed consequences of assignment actions makes this a Reinforcement Learning (RL) problem.

2 Technology

The core technology backing our demonstration is deep RL (DRL). In this demo, we showcase the ride-sharing applications of two DRL models through simulation.

2.1 Spatiotemporal Contextual Value Network

This work focuses on building a novel neural network for learning the value function of a driver, building upon our previous work [Wang *et al.*, 2018]. We model the ride dispatching problem as a Semi Markov Decision Process to account for the temporal aspect of the dispatching actions. Specifically, we tackle the policy evaluation problem for a fixed policy through proposing a value network with a novel distributed state representation layer that combines multi-scale grid tiling, quantization, and embedding. The resulting value network is called Cerebellar Value Networks (CVNet) [Tang *et al.*, 2019]. We further derive a regularized policy evaluation scheme for CVNet that penalizes large Lipschitz constant of the value network for additional robustness against adversarial perturbation and noises. To apply this new value network to online order dispatching at production scale, we adopt a two-stage learning and planning approach, where the learning part is accomplished by training the CVNet on static historical trip data with specific techniques to ensure convergence and successful learning. Then, the planning part is done by solving a bipartite matching [Xu *et al.*, 2018]. We have conducted real-world AB tests in production for this method. Part of this demo shows this method in action on a city-scale order dispatching simulator backed by real trip data.

2.2 Global-view Attention-enabled DQN

In this model, the agent (driver) maintains a global view of the supply and demand condition of a given region as its state, including the statuses of all the orders and drivers. Figure 1 is a visualization of such state information. The agent computes its state-action values through embedding and attention mechanisms and executes an optimal action (i.e. picking up an order or repositioning) based on this value function. We will demonstrate through several specially designed domains how we use this DRL method to train agents to have longer optimization horizon and to cooperate to achieve higher objective values collectively on a ride-share platform.

The novelty of this intelligent dispatching technology comes in three folds: global-view state representation, a novel neural network architecture that enables end-to-end training, and the support for mixed order and driver dispatching. See [Holler *et al.*, 2018] for the paper that describes the technology that part of this demonstration is based on.

*Contact Author

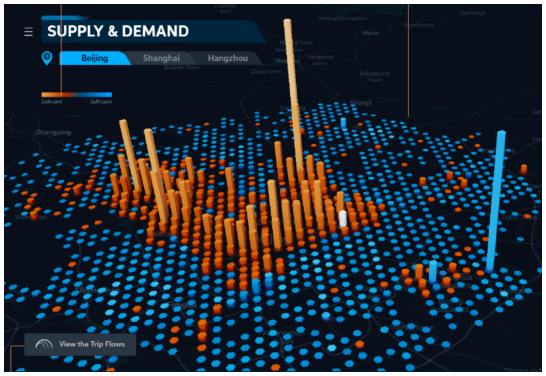

Figure 1: Setting supply and demand over a grid map.

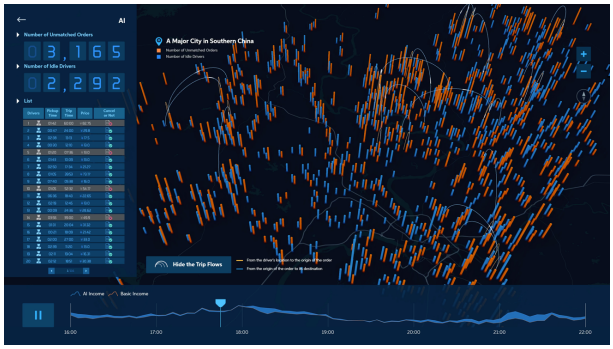

Figure 2: City-scale order dispatching.

Figure 3: Distribute.

Figure 4: Hot/cold domain.

3 Interaction

Our demo consists of two parts. The first part showcases city-scale dispatching in action by the CVNet model. A screenshot is shown in Figure 2. The map is implicitly divided into hex grids, each of which shows the numbers of open trip orders and free drivers. The main screen provides a holistic view of the supply-demand context within a five-min interval. The map can be fully zoomed and panned. We can stop at any step to examine the dispatching details: The paginated list on the left-hand-side shows the specifics of every dispatched trips within the current five-minute window. Clicking on any of them will highlight the trip in the center map. The two curves plotted along the timeline track the platform revenue for each step generated by DRL and the baseline combinatorial optimization method.

The second part features multiple specifically designed ride-share domains that demonstrate the ability of the model in [Holler *et al.*, 2018] to learn good dispatching and repositioning policies. One domain (Distribute domain) is to test if the agents can learn non-uniform repositioning behavior based on the implicitly learned demand distribution over spatiotemporal space. A screenshot is shown in Figure 3. The agents have to separate and head to the correction direction before the orders are realized at the two corners. Equivalently, the system is required to dispatch drivers to certain locations (without active orders) to anticipate future demand. Another domain (Hot/cold domain) verifies that the agents are able to learn subtle temporal advantage between orders going to hot and cold regions. See Figure 4. It simulates the scenario of downtown and suburb. The agents are expected to reposition themselves from cold areas after being brought there by orders. All the domains in this demo are served live by the policy generated from the DRL models at the backend service.

User interaction will be carried out in several illustrative domains for order and driver dispatching to demonstrate the advantage of DRL. The user can set passenger demand and driver supply to a certain extent. The user can control the progress of the simulation through a button and can pause at any transition point. This way, the user can compare the different dispatching or repositioning actions made by the DRL policy based on different global state.

4 Contribution

The system in this demo is a prototype of a large-scale ride-share marketplace engine. Part of the technology has been deployed and tested in production. As far as we know, this is the first industry-scale demo of DRL methods for order dispatching and driver repositioning. Users are able to see DRL-based policies in action both at a holistic level and within specific scenarios.

References

[Holler *et al.*, 2018] John Holler, Zhiwei Qin, Xiaocheng Tang, Yan Jiao, Tianchen Jin, Satinder Singh, Chenxi

Wang, and Jieping Ye. Deep q-learning approaches to dynamic multi-driver dispatching and repositioning. In *NeurIPS 2018 Deep Reinforcement Learning Workshop*, 2018.

[Wang *et al.*, 2018] Zhaodong Wang, Zhiwei Qin, Xiaocheng Tang, Jieping Ye, and Hongtu Zhu. Deep reinforcement learning with knowledge transfer for online rides order dispatching. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 617–626, Nov 2018.

[Xu *et al.*, 2018] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 905–913. ACM, 2018.