Deep Reinforcement Learning for Crowdsourced Urban Delivery: System States Characterization, Heuristics-guided Action Choice, and Rule-Interposing Integration

Tanvir Ahamed[1] (presenter), Bo Zou[1], Nahid Parvez Farazi[1], Theja Tulabandhula[2]

[1] Department of Civil, Materials, and Environmental Engineering, University of Illinois at Chicago

[2] Department of Information and Decision Sciences, University of Illinois at Chicago

Keywords

Crowdshipping, deep Q network, state representation, heuristics-guided action choice, rule-interposing

Introduction

Urban delivery is undergoing an exciting and challenging time with ever-growing online shopping demand. On the other hand, the time-sensitive nature of urban delivery has imposed considerable pressure on delivery service providers (DSP) to control shipping cost while meeting customer expectations, for which delivery within one or two hours after an order is placed has become increasingly common. In this environment, crowdshipping has emerged as an attractive new form of delivery. In crowdshipping, a DSP solicits ordinary people, termed crowdsourcees, who have some available time and may walk, bike, or drive a car to perform delivery to earn payment. Many companies including Postmates, Deliv, Piggy Baggy, Amazon Flex, Uber Eats, DoorDash, and Instacart are rapidly expanding their crowdshipping businesses.

This paper focuses on the crowdshipping problem with spatially distributed request pickup and delivery locations, using "dedicated crowdsourcees" who are also spatially distributed. Distributed pickup and delivery locations are relevant to delivery from restaurants, grocery stores, and retail shops to customers. Dedicated crowdsourcees inform the DSP about their available time for performing delivery. Each crowdsourcee has a limited carrying capacity and gets paid with a fixed rate whenever carrying a request. Each shipping request has a narrow time window defined by the time between the earliest time of pickup and the latest time of delivery.

Methodology

Given the above setup, the central research question is how a DSP efficiently assigns requests to crowdsourcees by minimizing total shipping cost, at the same time respecting constraints arising from the time availability and carrying capacity limits of crowdsourcees and pickup and delivery time windows of the requests. While this problem may be viewed as a specific type of pickup-and-delivery problem and integer programming models and heuristic algorithms could be considered, the novelty of this work is that we propose an approach that leverages deep Q learning (DQN) (Minh et al., 2015), a relatively new deep reinforcement learning (DRL) algorithm, to frame and solve the constrained crowdsourcee-shipping request assignment problem. More importantly, our work goes beyond simple adaptation of DQN to crowdshipping (which actually is difficult), but introduces and integrates three methodological advancements.

1. System state representation. Given the combinatorial nature of the crowdshipping problem and heterogeneity of both requests and crowdsourcees in terms of time and carrying capacity, the states of a crowdshipping system cannot be represented by one or a handful of metrics. A comprehensive representation needs to capture the sequence of pickup and delivery nodes on each crowdsourcee route, as well as time information of requests and crowdsourcees: each request has a limited time window between earliest possible pickup and latest delivery; each crowdsourcee also limited time availability. The time information dynamically changes as crowdsourcee routes are constructed, and is critical to the DRL agent routing decision-making with respect to request delivery urgency and crowdsourcee route priority. By leveraging the notation of information array, a multi-tuple state representation is proposed encompassing not only static information of request pickup and delivery locations but also information on crowdsourcee routing sequences, request-specific time availability, and crowdsourcee-specific time and capacity availability.

2. Embedding heuristics-guided action choice in DRL. The combinatorial nature of the problem means that a very large number of actions can be taken to construct and improve crowdsourcee routing. However, enumerating all possible actions would not be practical in DRL training. For this reason, existing related work avoids considering system-level DRL training, but treats each "crowdsourcee"-equivalent entity as an independent agent and trains the agents separately (Al-Abbasi et al., 2019; Chen et al., 2019; Yu et al., 2019) to reduce the individual DRL problem size, which yields routing that is system suboptimum. To address this challenge while preserving training tractability, we propose to abstract the action space into five types of actions (insertion, intra-route move, inter-route move, 1-exchange, and do-nothing). DRL will be not only trained to direct the type of actions to take based on past actions and outcomes, but embedded with heuristics which guide the ultimate specific action to take, using information provided by the system state as well as intuitive reasoning.

3. Integrating rule-interposing into DRL. The rules aim to prevent certain routes or node sequences from being visited repeatedly during neighborhood moves (intra-route move, inter-route move, and 1-exchange) within a period of time, as repeated visiting discourages exploring more actions and may get the routing sequence trapped in local optimum compromising the efficiency of DRL training. Two rules are developed that: 1) set up and update a priority list of crowdsourcee routes for each neighborhood move, based on criteria in line with the nature of the neighborhood moves. A crowdsourcee route that is chosen for a neighborhood move will be removed from the priority list and not considered for a period of time; 2) introduce Tabu tenure for the relative positions of pickup and delivery nodes. With the two rules, repeated visiting of routes or node sequences during neighborhood moves are largely spared, thereby enhancing the training efficiency by accelerating learning.

Results

The effectiveness of the proposed approach is demonstrated by solving a large number of crowdshipping problem instances. Our results show non-trivial benefits brought by heuristics-guided action choice and rule-imposing (Fig. 1-2). In addition, the trained DRL model outperforms existing heuristics in both solution quality (Fig. 3) and computation time (DRL can solve problems in a matter of seconds, as opposed to tens of minutes by heuristics).

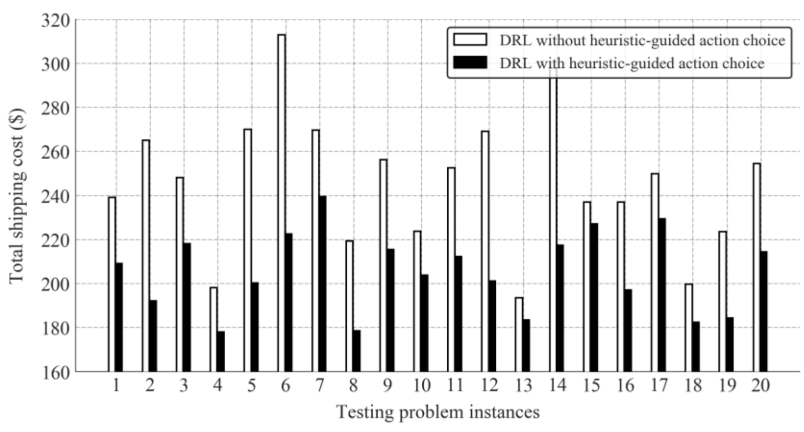

Fig. 1. Total shipping cost with and without heuristics-guided action choice

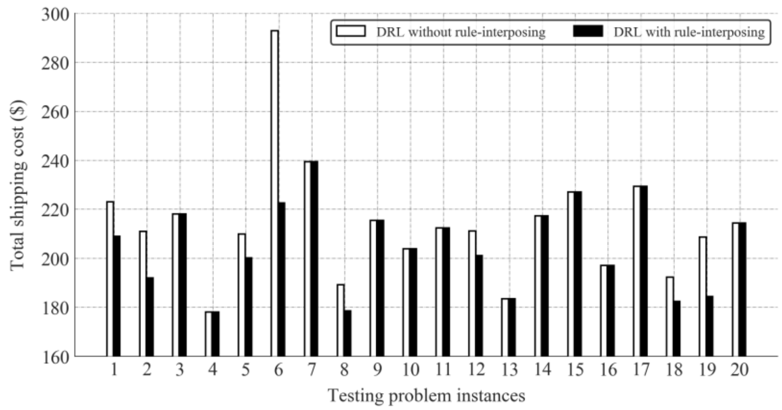

Fig. 2. Total shipping cost with and without rule-interposing

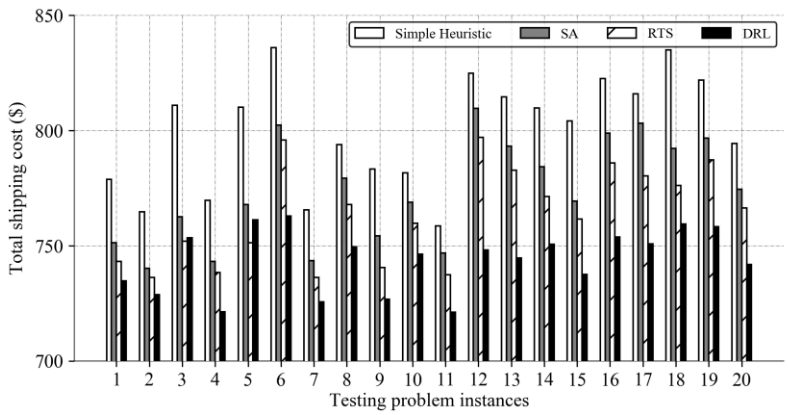

Fig. 3. Comparison of DRL with existing heuristics in terms of total shipping cost

Conclusion

In this paper, we propose a new DRL-based approach to seek high-quality and computationally efficient assignment of requests to crowdsourcees. The novelty of the proposed approach lies on the introduction and integration of novel state representation, heuristics-guided action choice, and rule-interposing in DQN. The approach has potential for practical crowdshipping operation planning and even real-time decision-support. The methodology framework also provides a promising new avenue for solving general pickup and delivery and VRP problems. Future research can be extended to solving dynamic versions of the problem, adding crowdsourcee relocation, and considering the possibility that crowdsourcees may reject assigned requests.

References

1. Al-Abbasi, A.O., Ghosh, A. and Aggarwal, V., 2019. Deeppool: Distributed model-free algorithm for ride-sharing using deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 20(12), pp.4714-4727.
2. Chen, X., Ulmer, M.W. and Thomas, B.W., 2019. Deep Q-Learning for Same-Day Delivery with a Heterogeneous Fleet of Vehicles and Drones. arXiv preprint arXiv:1910.11901.
3. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G. and Petersen, S., 2015. Human-level control through deep reinforcement learning. nature, 518(7540), pp.529-533.
4. Yu, J., Yu, W. and Gu, J., 2019. Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 20(10), pp.3806-3817.