Mutual Information as Intrinsic Reward of Reinforcement Learning Agents for On-demand Ride Pooling

Extended Abstract

Xianjie Zhang
Dalian University of Technology
Dalian, China
xj_zh@foxmail.com

Jiahao Sun
Dalian University of Technology
Dalian, China
Start1w@mail.dlut.edu.cn

Chen Gong
University of Virginia
Virginia, USA
chengong@virginia.edu

Kai Wang
Nanyang Technological University
Singapore
kai_wang@ntu.edu.sg

Yifei Cao
Dalian University of Technology
Dalian, China
yfcao@mail.dlut.edu.cn

Hao Chen
Institute of Automation, Chinese Academy of Sciencec
Beijing, China
hao_universe@163.com

Yu Liu*
Dalian University of Technology
Dalian, China
yuliu@dlut.edu.cn

ABSTRACT

The emergence of on-demand ride pooling services allows each vehicle to serve multiple passengers at a time, thus increasing drivers' income and enabling passengers to travel at lower prices than taxi/car on-demand services. Although on-demand ride pooling services can bring so many benefits, ride pooling services need a well-defined matching strategy to maximize the benefits for all parties (passengers, drivers, aggregation companies and environment), especially the regional dispatching of vehicles has a significant impact on matching and revenue. Existing algorithms often only consider revenue maximization, which makes it difficult for requests with unusual distribution to get rides. How to increase revenue while ensuring a reasonable assignment of requests brings a challenge to ride pooling service companies (aggregation companies). In this paper, we propose a framework for vehicle dispatching for ride pooling tasks, which splits the city into discrete dispatching regions and uses the reinforcement learning (RL) algorithm to dispatch vehicles in these regions. We also consider the mutual information (MI) between vehicle and request distribution as the intrinsic reward of the RL algorithm to improve the correlation between their distributions, thus ensuring the possibility of getting a ride for unusually distributed requests. In experimental results on a real-world taxi dataset, we demonstrate that our framework can significantly increase revenue up to an average of 3% over the existing best on-demand ride pooling method.

*Corresponding author

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

KEYWORDS

Transportation, Reinforcement learning, Multiple agents, Mutual information

ACM Reference Format:
Xianjie Zhang, Jiahao Sun, Chen Gong, Kai Wang, Yifei Cao, Hao Chen, and Yu Liu. 2024. Mutual Information as Intrinsic Reward of Reinforcement Learning Agents for On-demand Ride Pooling: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

With the development of the mobile internet and sharing economy, it is becoming more and more accepted for people to hail a ride anytime by using mobile devices. On-demand ride pooling is one of the most popular services among them, where service providers like UberPool, LyftLine, and GrabShare allow multiple passengers traveling in the same direction to be matched with the same vehicle through intelligent algorithms and smart terminal devices [4, 6, 7, 14]. The emergence of ride pooling services not only reduces energy consumption and emissions, but also reduces traffic congestion, and lowers the cost of individual passenger's taxi fares. At the same time, it also brings economic benefits to drivers and aggregation companies [10, 17].

However, there are still some challenges in city-scale on-demand ride pooling systems. First, travel demand is not uniformly distributed over different regions of the city, and ride pooling systems often face problems such as an imbalance between supply and demand in different regions. Secondly, there is a dependency relationship between dispatching and matching. Finally, unlike usual ride sharing, an on-demand ride pooling system requires combining passengers on the same route into a "trip" (a combination of requests) and matching them to the same vehicle [2].

To make the algorithm applicable to on-demand ride pooling, and to consider future benefits and the vehicle-requset distribution differences while dispatching and matching decisions, we propose a reinforcement learning-based (RL-based) vehicle dispatch framework (Figure 2). Specifically, our contributions are as follows: (1) We provide a precise definition of the dispatching and matching problem for ride pooling systems. Based on this definition, for the dispatching problem, the city map is divided into an appropriate number of hexagonal regions (As shown in Figure 1) to facilitate near-field dispatching in these hexagonal regions; (2) We propose a reinforcement learning-based regional dispatching algorithm that uses a mean field Q-learning (MFQL) [13], allowing the algorithm to scale up to city-scale ride pooling tasks; (3) Optimizing the mutual information (MI) between the distribution of vehicles and requests can enable the ride pooling system to adjust the distribution of vehicles according to the distribution of requests, thereby improving the overall revenue. By using MI as the intrinsic reward value in reinforcement learning (RL) [15, 16], we can optimize the value of MI; (4) In experiments, we use a simulation of an on-demand pooling task with a real-world dataset to show that our framework represents a 3% relative improvement over the best available on-demand ride pooling approach.

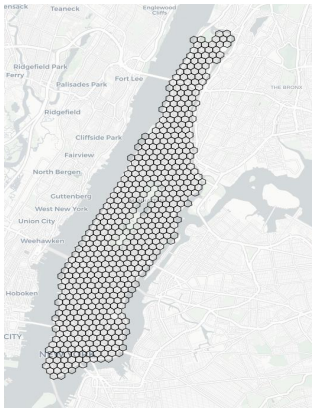

Figure 1: Manhattan is divided into hexagonal regions.

2 METHOD

The simple matching algorithm can only get the vehicle-request matching policy with the highest revenue at the current moment, which is a rather myopic algorithm. To avoid this myopia, in our framework, vehicles are modeled as RL agents to obtain the long-term future impact of the vehicle dispatching strategy. The dispatched vehicles are then involved in vehicle-request matching, so that both the long-term view of the dispatching strategy and the maximum revenue of timely matching are taken into account in our algorithm.

Specifically, the framework (Figure 2) includes the following components: (1) The Q-learning is used to obtain the vehicle dispatching strategy. The Q function calculates the Q-value $Q(\omega_v, a_v)$ of all possible actions of vehicle v. The Q-value is used to select the action of v. (2) Once the action a_v is determined, the vehicle v will dispatch to the region d_v, and the vehicle v is matched with the

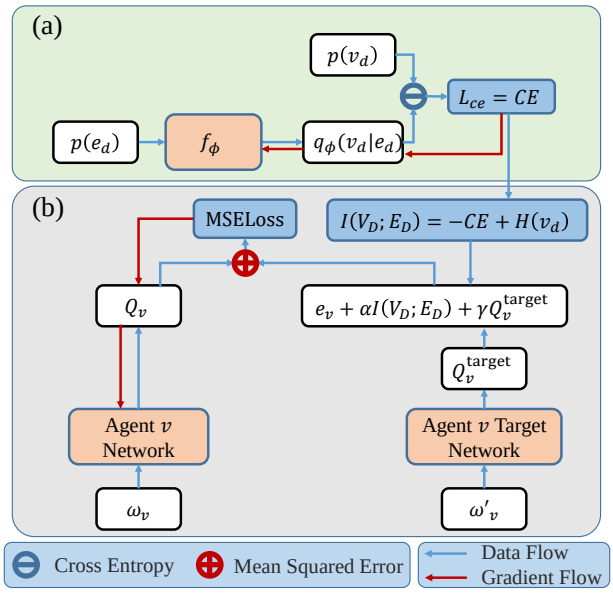

Figure 2: The overall framework.

requests in the region d_v. (3) To further capture the dependence of dispatching actions on neighboring vehicles/agents, we use MFQL in the RL dispatching framework. (4) To better serve the unusually distributed requests, the mutual information between the request and vehicle distribution is added as a reinforcement learning intrinsic reward value. However, estimating and maximizing MI is usually intractable. Taking inspiration from the literature on variational inference [1, 5, 11], the variational posterior estimator is introduced to derive a tractable lower bound on the MI for each time step t.

Varying	Parameters			NeurADP [10]	MFQL+MI	Improve
	NV	PD	C			
PD	600	300	4	106180.23	110196.19	3.78%
NV	500	300	4	89003.86	92292.44	3.69%
C	600	300	8	117115.35	120660.29	3.03%

Table 1: Compare MFQL+MI with NeurADP [10].

3 EXPERIMENTAL RESULTS

We chose the street network of Manhattan, New York as the source of dependency for vehicle operation. The New York yellow taxi dataset [9] is also obtained from the open network. We use Boeing [3]'s work osmnx to obtain the city's street network from openstreetmap using "drivers". Our experimental setup is similar to papers [10, 17].

Compared with the existing best method for on-demand ride pooling NeurADP [10] in Table 1, our algorithm has an average performance improvement of 3%. Considering that even a 1% improvement in revenue is considered by the industry to be a large improvement on similar transportation problems [8, 12].

REFERENCES

[1] Alex Alemi, Ian Fischer, Josh Dillon, and Kevin Murphy. 2017. Deep Variational Information Bottleneck. In *ICLR*. https://arxiv.org/abs/1612.00410

[2] Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences* 114, 3 (2017), 462–467. https://doi.org/10.1073/pnas.1611675114 arXiv:https://www.pnas.org/content/114/3/462.full.pdf

[3] Geoff Boeing. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65 (2017), 126–139.

[4] Kathryn Gessner. 2019. Uber vs. Lyft: Who's tops in the battle of U.S. rideshare companies. https://www.uber.com/en-GB/newsroom/company-info/.

[5] Chen Gong, Yunpeng Bai, Xinwen Hou, and Xiaohui Ji. 2020. Stable training of bellman error in reinforcement learning. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18–22, 2020, Proceedings, Part V 27*. Springer, 439–448.

[6] Chen Gong, Zhou Yang, Yunpeng Bai, Junda He, Jieke Shi, Arunesh Sinha, Bowen Xu, Xinwen Hou, Guoliang Fan, and David Lo. 2022. Mind your data! hiding backdoors in offline reinforcement learning datasets. *arXiv preprint arXiv:2210.04688* (2022).

[7] Alex Heath. 2016. Inside Uber's quest to get more people in fewer cars. https://www.businessinsider.com/uberpool-ride-sharing-could-be-the-future-of-uber-2016-6/.

[8] Meghna Lowalekar, Pradeep Varakantham, and Patrick Jaillet. 2019. ZAC: A Zone Path Construction Approach for Effective Real-Time Ridesharing. In *ICAPS*.

[9] NYYellowTaxi. 2016. New York Yellow Taxi DataSet. http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

[10] Sanket Shah, Meghna Lowalekar, and Pradeep Varakantham. 2020. Neural Approximate Dynamic Programming for On-Demand Ride-Pooling. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*. AAAI Press, 507–515.

[11] Martin J. Wainwright and Michael I. Jordan. 2008. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning* 1, 1–2 (2008), 1–305. https://doi.org/10.1561/2200000001

[12] Zhe Xu, Zhixin Li, Qingwen Guan, Dingshui Zhang, Qiang Li, Junxiao Nan, Chunyang Liu, Wei Bian, and Jieping Ye. 2018. 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 and Data Mining (KDD '18)*. 905–913.

[13] Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. 2018. Mean Field Multi-Agent Reinforcement Learning. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, Vol. 80. 5571–5580.

[14] Xianjie Zhang, Yu Liu, Wenjun Li, and Chen Gong. 2022. Pruning the Communication Bandwidth between Reinforcement Learning Agents through Causal Inference: An Innovative Approach to Designing a Smart Grid Power System. *Sensors* 22, 20 (2022). https://www.mdpi.com/1424-8220/22/20/7785

[15] Xianjie Zhang, Yu Liu, Hangyu Mao, and Chao Yu. 2022. Common belief multi-agent reinforcement learning based on variational recurrent models. *Neurocomputing* 513 (2022), 341–350. https://doi.org/10.1016/j.neucom.2022.09.144

[16] Xianjie Zhang, Yu Liu, Xiujuan Xu, Qiong Huang, Hangyu Mao, and Anil Carie. 2021. Structural relational inference actor-critic for multi-agent reinforcement learning. *Neurocomputing* 459 (2021), 383–394.

[17] Xianjie Zhang, Pradeep Varakantham, and Hao Jiang. 2023. Future Aware Pricing and Matching for Sustainable On-Demand Ride Pooling. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, (AAAI)*. 14628–14636.