Professorship of Business Analytics & Intelligent Systems
TUM School of Management
Technical University of Munich

Joint rewards in hybrid multi-agent deep reinforcement learning for autonomous mobility on demand

Topic and motivation

Services offering autonomous mobility on demand (AMoD) will likely transform urban mobility in future years. In contrast to classical mobility on demand services like Uber, Lyft or DiDi, operators of autonomous MoD fleets can exercise full control over their vehicles. This full control can enable better vehicle dispatching but increases the dispatching problem´s action space at the same time.

Solving the vehicle dispatching problem requires a method suitable for online decision making in a stochastic environment. A particularly well-suited method for such problems is deep reinforcement learning (DRL), but the large action space makes single-agent DRL infeasible. Enders et al. (2022) therefore propose using multi-agent DRL in combination with optimization-based central decision-making to assign vehicles to orders in large-scale settings. They successfully show that this approach can find better dispatching policies than both greedy and model predictive control algorithms. An important component of DRL algorithms is the structure of the rewards; Enders et al. (2022) use egoistic rewards for each agent. However, agents should contribute to the overall performance and maximize the profit of the entire system, which is not necessarily the case when they receive egoistic rewards.

This research project therefore proposes to use joint rewards to align all agents with the target of creating maximum value for the entire system, extending the work by Enders et al. (2022). However, when assigning rewards to agents based on the overall system performance, agents cannot easily derive their own contribution from the joint reward, as the actions of other agents are seen as being part of the environment. This can lead to inferior solutions and is known as the credit assignment problem (Chang et al. 2004, Weiß 1995, Wolpert and Tumer 1999). Accordingly, this project aims to find a solution for the credit assignment problem in the AMoD setting, to enable agents to infer their own contribution to a joint reward.

Research gap

This project addresses two open issues in current research. On one side, it contributes to the literature on dispatching decisions in AMoD using multi-agent DRL. Previous approaches to vehicle dispatching in (A)MoD using DRL are often based on optimization objectives not suitable for autonomous vehicles (e.g., Tang et al. 2019, Zhou et al. 2019, Xu et al. 2018). In addition, none of these approaches uses joint rewards successfully. By introducing joint rewards into the literature on AMoD, this project can therefore fill a relevant research gap.

On the other side, the project contributes to a better understanding of joint rewards in multi-agent DRL. The above-mentioned credit assignment problem is a major obstacle in settings which require agents to cooperate. Possible solution approaches include decomposing the joint reward into individual rewards (e.g., Kok and Vlassis 2006, Sunehag et al. 2018), inverse reinforcement learning (e.g., Ng and Russell 2000, Lin et al. 2018), or marginalizing out the effect of single actions (e.g., Nguyen et al. 2018, Foerster et al. 2018). Especially the approach of Foerster et al. (2018) appears to be promising for the setting of this project. However, the credit assignment problem in multi-agent DRL is not fully solved yet (cf. Gronauer and Diepold 2022) and none of the mentioned solution approaches have been applied to the AMoD setting so far.

Time plan

The project shall start in April 2023 and end in December 2023. It consists of an individual research phase and a master´s thesis by Heiko Hoppe, supervised by Prof. Maximilian Schiffer and PhD student Tobias Enders from TU Munich, and Prof. Quentin Cappart from Polytechnique Montréal. The project includes a research stay in Montréal by Heiko Hoppe from beginning of July to end of November 2023.

The project focuses on integrating joint rewards into the AMoD setting, comprising the following working steps:

1. Motivation of the topic and problem setting

2. Review of literature on using joint rewards in DRL (focus on solving the credit assignment problem)
3. Choosing one or multiple joint reward integration structures
4. Mathematically formulating the chosen structures
5. Implementing the structures in Python
6. Testing the implemented structures against the algorithm with egoistic rewards, a greedy model and a model predictive control algorithm in the New York taxi case
7. Deduction of technical conclusions on how joint rewards can be used in DRL
8. Derivation of managerial insights

References

Chang Y, Ho T, Kaelbling LP (2004): All learning is Local: Multi-agent Learning in Global Reward Games. In: Advances in Neural Information Processing Systems 16: 807–814.

Enders T, Harrison J, Pavone M, Schiffer M (2022): Hybrid Multi-agent Deep Reinforcement Learning for Autonomous Mobility on Demand Systems. In: arXiv:2212.07313.

Foerster J, Farquhar G, Afouras T, Nardelli N, Whiteson S (2018): Counterfactual Multi-Agent Policy Gradients. In: Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

Gronauer S and Diepold K (2022): Multi-agent deep reinforcement learning: a survey. In: Artificial Intelligence Review, 55(2): 895–943.

Kok JR, Vlassis N (2006): Collaborative Multiagent Reinforcement Learning by Payoff Propagation. In: Journal of Machine Learning Research 7: 1789–1828.

Lin X, Beling PA, Cogill R (2018): Multiagent Inverse Reinforcement Learning for Two-Person Zero-Sum Games. In: IEEE Transactions on Games 10(1): 56–68.

Ng AY, Russell SJ (2000): Algorithms for Inverse Reinforcement Learning. In: Proceedings of the Seventeenth International Conference on Machine Learning: 663–670.

Nguyen DT, Kumar A, Lau HC (2018): Credit Assignment For Collective Multiagent RL With Global Rewards. In: Advances in Neural Information Processing Systems 31: 8102–8113.

Sunehag P, Lever G, Gruslys A, Czarnecki WM, Zambaldi V, Jaderberg M, Lanctot M, Sonnerat N, Leibo JZ, Tuyls K, Graepel T (2018): Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems: 2085–2087.

Tang X, Qin Z, Zhang Z, Wang Z, Xu Z, Ma Y, Zhu H, Ye J (2019): A Deep Value-network Based Approach for Multi-Driver Order Dispatching. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining: 1780–1790.

Weiß G (1995): Distributed Reinforcement Learning. In: Steels L (ed) The Biology and Technology of Intelligent Autonomous Agents. Springer, Berlin: 415–428.

Wolpert DH, Tumer K (1999): An Introduction to Collective Intelligence. In: arXiv:cs/9908014.

Xu Z, Li Z, Guan Q, Zhang D, Li Q, Nan J, Liu C, Bian W, Ye J (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 & Data Mining: 905–913.

Zhou M, Jin J, Zhang W, Qin Z, Jiao Y, Wang C, Wu G, Yu Y, Ye J (2019): Multi-Agent Reinforcement Learning for Order-dispatching via Order-Vehicle Distribution Matching. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management: 2645–2653.