Week 11 Paper Review

Niharika Shrivastava
School of Computing
National University of Singapore
Singapore, 119077
niharika@comp.nus.edu.sg

Abstract

This is a brief review of [1], [2].

1 Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments

The authors introduce a novel actor-critic algorithm called Multi-Agent DDPG (MADDPG) for both cooperative and competitive multi-agent problems. The action-value functions are learnt by considering the actions of all players making the critic a centralized one. However, during test time, the actors are decentralized making each agent act individually. This setup helps to evaluate the rewards for each agent in a shared setting. Moreover, it learns an ensemble of policies for each agent. During training at each epoch, the agent samples a policy from this ensemble thereby reducing the non-stationarity caused by multiple agents learning at the same time. Several interesting, both cooperative and competitive, are considered to evaluate the proposed algorithm. Simulation results show the benefit of using ensembles and improvement with respect to independent DDPG agents. The authors show that MADDPG outperforms state-of-the-art MARL algorithms in terms of both sample efficiency and final performance.

However, the algorithm is less effective when the number of possible joint actions in the environment is very large, making it computationally difficult to learn the optimal policies for all agents. Moreover, the centralized critic components require agents to share information and coordinate their actions during training, which may not be feasible or desirable in all real-world multi-agent scenarios.

2 Mingling Foresight with Imagination: Model-Based Cooperative Multi-Agent Reinforcement Learning

This paper introduces the model-based value decomposition (MBVD) framework, which extends a value decomposition method (e.g., QMIX) with the imagination module. Specifically, the imagination module consists of two VAEs and outputs the rollout of imagined future states in the latent space. Then, this rollout information is concatenated with the current state and becomes an input to the mixing network of QMIX. MBVD assumes that these imagined states contain helpful information for agents to evaluate the current state's value more accurately. Empirical results show that MBVD outperforms baselines in various domains of SMAC, Football, and multi-agent MuJoCo in terms of efficiency and generalization.

As the authors have mentioned, the limitation of the paper is that the rollout horizon is manually chosen. However, further explanation could have been provided on existing literature where how others are solving this problem or how learnt horizon values can improve the existing framework. Overall, its a significant contribution since model-based techniques for POMDP is considered to have high complexity.

Preprint. Under review.

References

[1] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. ArXiv. /abs/1706.02275

[2] Xu, Z., Li, D., Zhang, B., Zhan, Y., Bai, Y., & Fan, G. (2022). Mingling Foresight with Imagination: Model-Based Cooperative Multi-Agent Reinforcement Learning. ArXiv. /abs/2204.09418