Learning Multi-Agent Action Coordination via Electing First-Move Agent

Jingqing Ruan, Linghui Meng, Xuantang Xiong, Dengpeng Xing, Bo Xu

Abstract

Learning to coordinate actions among agents is essential. Prior works are constrained mainly by the assumption that all agents act simultaneously, and asynchronous action coordination between agents is rarely considered. This paper introduces a bi-level multi-agent decision hierarchy for coordinated behavior planning. We propose a novel election mechanism and elect a first-move agent for asynchronous guidance. This work is the first to explicitly model the asynchronous multi-agent action coordination. The results demonstrate that the proposed algorithm can achieve superior performance in cooperative environments. Our code is available at *https://github.com/Amanda-1997/EFA-DWM*.

Contributions

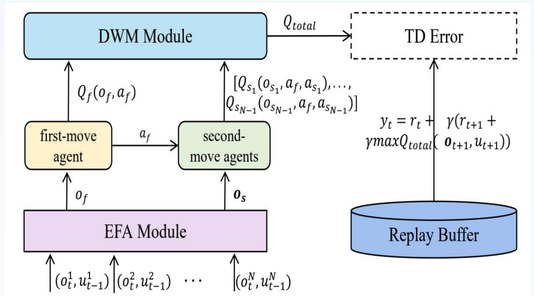
- We introduce a novel framework to construct a bi-level decision hierarchy to promote asynchronous action coordination for multiple agents.
- We propose to use a GCN-based election mechanism to select the optimal first-move agent and adopt the dynam ically weighted mixing network to alleviate the problem of misestimation of the value function.
  Empir
- cal evaluations on several challenging MARL benchmarks demonstrate the significant performance of the proposed algorithm.

Methods

1. The optimization process

$$a_f \leftarrow \arg\max_{a_{f'}} Q_f(o_f, a_{f'}; \theta_f)$$
$$a_{s_j} \leftarrow \arg\max_{a_{s_j'}} Q_{s_j}(o_{s_j}, a_f, a_{s_j'}; \theta_{s_j})$$
$$\theta_f \leftarrow r_f + \gamma \max Q_f(o_f', a_f'; \theta_i) - Q_f(o_f, a_f; \theta_f)$$
$$\theta_{s_j} \leftarrow r_{s_j} + \gamma \max Q_{s_j}(o_{s_j}', a_f, a_{s_j}'; \theta_{s_j}) - Q_{s_j}(o_{s_j}, a_f, a_{s_j}; \theta_f), j = 1, ..., N-1$$

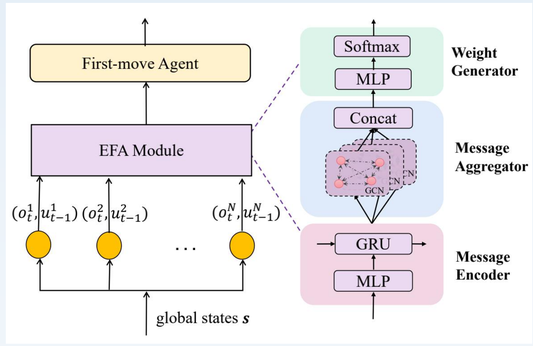
2. The overall framework



- The proposed approach EFA-DWM combines the Electing First-move Agent (EFA) module with a Dynamically Weighted Mixing (DWM) module.
- The EFA module elects the first-move agent based on the current observations and previous actions. We adopt the improved value decomposition network (VDN) (Sunehag et al. 2018) as the DWM module.

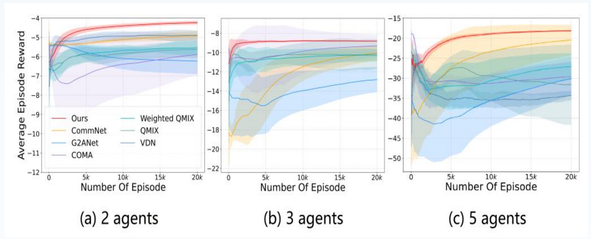
3. The EFA module

- It consists of a triple of the following networks: the message encoder, the message aggregator and the weight generator, shown as next page.



Results

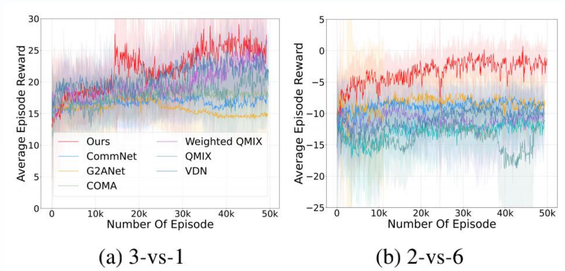
1. Results on Cooperative Navigation



(a) 2 agents (b) 3 agents (c) 5 agents

- As the number increases, ours maintains a stable performance improvement.

2. Results on Google Football



(a) 3-vs-1 (b) 2-vs-6

- GF requires more efficient coordination, the results demonstrate that ours can better grasp the stochasticity and complexity.

Conclusion

We propose a novel hierarchical framework to explicitly model the election of the optimal first-move agent for coor dinated behaviour planning in MARL. The election module brings together the benefits of graph convolutional network and attention mechanism for message aggregation, and we design the weight-based scheduler to elect the optimal first move agent. Then the dynamically weighted mixing network can alleviate the problem of misestimation and put more emphasis on better joint actions. Empirical results show that our algorithm can achieve higher rewards, faster convergence, and lower variance.

References

[1]. Sunehag, P.; Lever, G.; et al. 2018. Value-decomposition networks for cooperative multi-agent learning based on team reward. International Conference on Autonomous Agents and Multi-Agent Systems, 2085–2087.

[2]. Ruan, J.; Du, Y.; et al. 2022. GCS: Graph based Coordination Strategy for Multi-Agent Reinforcement Learning. International Conference on Autonomous Agents and Multi-Agent Systems.

[3]. Velickovic, P.; Cucurull, G.; et al. 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.