Benchmarking MARL on Long Horizon, Sequential, Multi-Objective Tasks

Minghong Geng, Shubham Pateria, Budhitama Subagdja, Ah-Hwee Tan

Neural and Cognitive Computing Group
Singapore Management University

Research Group

GitHub Repo

Introduction

- Current MARL benchmarks primarily focus on short-horizon, single-objective tasks, lacking realistic multi-objective scenarios.
- We present ***Multi-Objective SMAC* (MOSMAC)**, a new benchmark with multiple objectives, sequential task allocation, and varying horizons.
- MOSMAC scenarios are designed to evaluate the ability of agents to make strategic trade-offs between objectives and adapt to different task lengths and difficulties.

Multi-objective MARL (MOMARL)

- Many real-world problems involve agents dealing with multiple objectives while collaborating on a single task.
- Such learning problems can be categorized as multi-objective MARL (**MOMARL**) [3] problems.
- MOMARL methods can be broadly classified into two categories:

Single-policy methods: Agents learn a single policy that optimizes a particular utility u.

Multi-policy methods: Agents learn a set of policies to approximate the Pareto front.

- In this work, we report the preliminary results of using SOTA MARL methods as *single-policy* methods for two objs.
- The reward functions for the combat and navigate objectives are:

Objective 1 (combat) $$r_{obj_1} = \sum_{i=1}^{n} (r_a^i + r_d^i)$$

Objective 2 (navigate) $$r_{obj_2} = \sum_{i=1}^{n} (r_r^i)$$

where r_a^i and r_d^i are the rewards for attacking and destroying enemy units by agent i; r_r^i is the reward for reducing distance to the strategic position by agent i; n is the total number of agents.

- The complete reward function is:

$$r = \alpha \times r_{obj_1} + (1 - \alpha) \times r_{obj_2}$$

where α is a weight of preference that indicates the *priority* [3] for Objective 1.

MOSMAC Overview

❖ MOSMAC features three characteristics:

1 **M**ultiple **O**bjectives

Objective 1 (combat): The damage to enemy units should be as much as possible.

Objective 2 (navigate): The distance agents from agents to strategic positions should be as minimal as possible.

2 **S**equential **T**ask **A**llocation

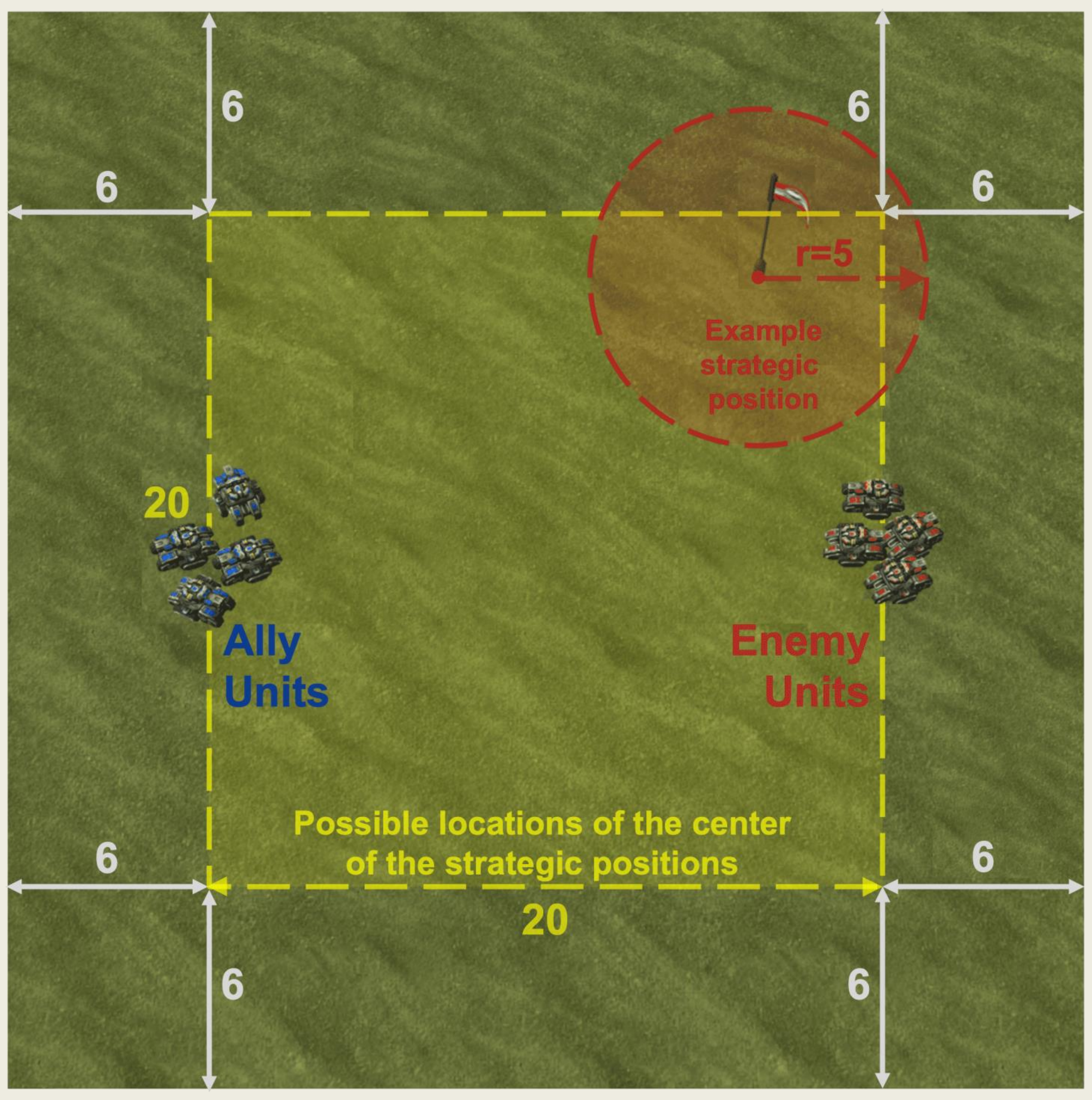
Agents are allocated with tasks sequentially in an episode, where the completion of the previous task triggers the next task allocation. As such, the long horizon tasks are decomposed into many subtasks.

3 **V**arying **H**orizons

The total timesteps of each long-horizon task vary based on the target locations and paths. The environment generates random directed acyclic paths for agents to navigate from the starting to the target locations.

❖ MOSMAC contains two sets of scenarios:

Short-horizon MOSMAC

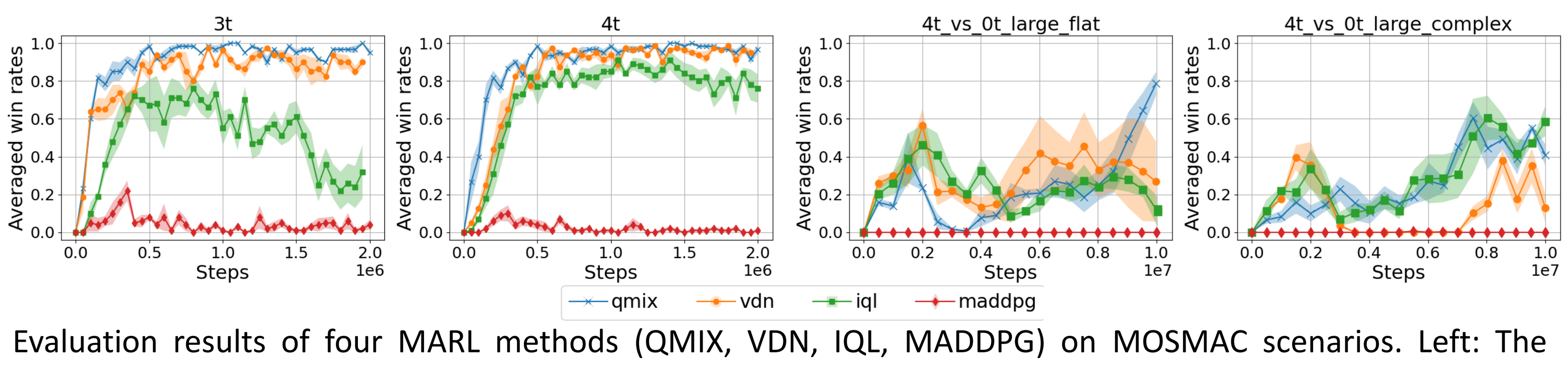


An illustration of a short-horizon MOSMAC scenario named 4t. The winning condition is all alive agents arrive at the strategic position, motivated by two objectives. The target locations are randomly selected for each episode, inspired by SMACv2 [2].

Long-horizon MOSMAC

An illustration of a long-horizon MOSMAC scenario, named 4t_vs_12t. The length of the horizons is configurable by moving the final targets to any of the red areas. Enemies are placed in three positions to deter the advancement of multi-objective agents.

Analysis and Findings in Preliminary Results



Evaluation results of four MARL methods (QMIX, VDN, IQL, MADDPG) on MOSMAC scenarios. Left: The results on two short-horizon scenarios. Right: The results on two long-horizon scenarios.

- We evaluate nine MARL methods with the EPyMARL [1] framework.
- Existing MARL methods are able to address short-horizon tasks but struggle when dealing with sequential tasks that involve multiple objectives over a longer horizon.
- Independent methods show superior results over CTDE algorithms in both short-horizon and long-horizon MOSMAC scenarios, suggesting the benefits of independent learning in complex multi-objective tasks.
- Long-horizon multi-objective learning poses significant challenges for MARL algorithms, as evident from the performance drops in long-horizon scenarios compared to short-horizon ones.

Conclusion

- MOSMAC fills the gap in current MARL and MOMARL benchmarks by providing a challenging testbed for evaluating algorithms in multi-objective tasks with varying horizons, which is applicable to both single-objective MARL and multi-objective MOMARL research.
- Future work includes expanding MOSMAC with additional objectives, scenarios, and algorithms, as well as exploring hierarchical learning and domain knowledge-based task decomposition to improve performance on long-horizon multi-objective tasks.

References

[1] G. Papoudakis, F. Christianos, L. Schäfer, and S. V. Albrecht, “Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks”, in Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (2020).

[2] B. Ellis et al., “SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning”, arXiv preprint arXiv:2212.07489v2 [cs.LG].

[3] T. Hu, B. Luo, C. Yang, and T. Huang, “MO-MIX: Multi-Objective Multi-Agent Cooperative Decision-Making With Deep Reinforcement Learning”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 10, pp. 1–15, Oct. 2023.