Benchmarking MARL on Long Horizon Sequential Multi-Objective Tasks

Extended Abstract

Minghong Geng
Singapore Management University
Singapore
mhgeng.2021@phdcs.smu.edu.sg

Shubham Pateria
Singapore Management University
Singapore
shbhamp@smu.edu.sg

Budhitama Subagdja
Singapore Management University
Singapore
budhitamas@smu.edu.sg

Ah-Hwee Tan
Singapore Management University
Singapore
ahtan@smu.edu.sg

ABSTRACT

Current MARL benchmarks fall short in simulating realistic scenarios, particularly those involving long action sequences with sequential tasks and multiple conflicting objectives. Addressing this gap, we introduce Multi-Objective SMAC (MOSMAC)

1, a novel MARL benchmark tailored to assess MARL methods on tasks with varying time horizons and multiple objectives. Each MOSMAC task contains one or multiple sequential subtasks. Agents are required to simultaneously balance between two objectives — combat and navigation — to successfully complete each subtask. Our evaluation of nine state-of-the-art MARL algorithms reveals that MOSMAC presents substantial challenges to many state-of-the-art MARL methods and effectively fills a critical gap in existing benchmarks for both single-objective and multi-objective MARL research.

KEYWORDS

Multi-agent Reinforcement Learning; Multi-Objective Multi-agent Reinforcement Learning; Benchmark

1 Code is available at https://github.com/smu-ncc/mosmac

1 INTRODUCTION

Studies on multi-agent reinforcement learning (MARL) have recently garnered significant achievements in various fields, including traffic signal control [4], game-playing [20], and stock-trading [1]. Despite the achievements, these applications commonly entail tasks with short horizons and single objectives [20]. In fact, learning over long horizons is a non-trivial challenge of MARL. In such scenarios, challenges like the exploration and temporal credit assignment become increasingly complex compared to their short-horizon counterparts [9]. In addition, the complexity of the hypothesis space for optimal value functions scales with the planning horizon [12], leading to the convergence of action-gaps and trap agents in local optima. However, currently there is still a scarcity of benchmarks for examining methods in long-horizon MARL contexts.

This paper presents a MARL benchmark named Multi-Objective SMAC (MOSMAC), which provides a set of multi-objective MARL (MOMARL) tasks that scale to various temporal horizons. Building upon the foundations laid by SMAC [20], SMACv2 [5], and SMAC-Exp [11], MOSMAC differentiates itself with three distinct features: varying temporal horizons, multiple objectives, and sequential sub-task assignments. MOSMAC also incorporates scenarios featuring complex terrains including plains, canyons, ramps, and high/low grounds, mirroring real-world scenarios and significantly challenging multi-agent exploration in a large state-action space. As a result, MOSMAC provides various interesting scenarios covering the aspects that are not included in most of the existing MARL tasks [2, 3, 18] and benchmarks [5, 11, 20], making it challenging for both MARL and MOMARL [7, 8, 10, 15, 24] domains.

We evaluate nine MARL algorithms [6, 14, 16, 17, 21–23, 25] on MOSMAC with the EPyMARL framework [16]. We find that while several methods exhibit good performance on addressing short-horizon MOMARL tasks, the long-horizon ones are still challenging, highlighting the need for more efficient MARL methods.

2 MULTI-OBJECTIVE SMAC (MOSMAC)

The short-horizon MOSMAC contains a set of MOMARL tasks with stochastic target placements. It contains scenarios with 3, 4, 8, and 12 Siege Tank units in both the ally and adversarial teams. Figure 1(a) shows an example scenario, named 4t, with four ally units, each controlled by a learning agent. Agents share the winning criteria of occupying a system-selected strategic position. The ally team wins the game if all remaining agents can reach the strategic position. The adversarial units are symmetric to ally units, controlled by the built-in controller of the StarCraft II game with a difficulty level of 7. Adversarial units are configured to guard the strategic position and will attack ally units when they are in close proximity. Similar to SMACv2 [5], units have their default sight and attack ranges, as
Our evaluation takes a single-policy approach [13, 19], where the utility of multiple objectives is represented by a scalar reward value, while multi-policy methods [8] can also be applied. Specifically, the short-horizon MOSMAC contains the following two objectives:

1. **Objective 1 (combat):** To maximize the damages to the enemy units.
2. **Objective 2 (navigate):** To minimize the distance between agents and the target strategic position.

Therefore, the reward functions for Objective 1 and 2 are:

\[ r_{obj1} = \sum_{i=1}^{n} (r_{a1}^i + r_{d1}^i) \]  \hspace{1cm} (1)

and

\[ r_{obj2} = \sum_{i=1}^{n} r_{f}^i \]  \hspace{1cm} (2)

respectively, where \( r_{a1}^i \) and \( r_{d1}^i \) are the rewards for attacking and destroying enemy units by agent \( i \), \( r_{f}^i \) is the reward for reducing the Euclidean distance to the strategic position by agent \( i \), and \( n \) is the total number of agents. The complete step-wise intermediate reward function \( r \) for short-horizon MOSMAC is as follows:

\[ r = \alpha \times r_{obj1} + (1 - \alpha) \times r_{obj2} \]  \hspace{1cm} (3)

where \( \alpha \) is a weight of preference that indicates the priority [8] given to Objective 1. Besides \( r \), agents will receive \( r_w \) as the terminal reward for winning the game by occupying the strategic position.

3 RESULTS AND CONCLUSION

This paper introduces MOSMAC, a new MARL benchmark aimed at challenging MARL algorithms with multi-objective long-horizon tasks. Through our experiments, we found that 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. This shows the utility of the proposed benchmark in pushing the performance boundary of the MARL algorithms.

Going forward, we aim to extend MOSMAC with new challenging scenarios with a more diverse set of units and provide more evaluation results of MARL methods, particularly in areas such as MARL with hierarchical learning paradigms and MOMARL.

ACKNOWLEDGMENTS

This research was supported by the DSO National Laboratories, Singapore (Agreement No. DSOCL20209) and the Jubilee Technology Fellowship awarded to Ah-Hwee Tan by Singapore Management University.