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
shubhamp@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

ACM Reference Format:
Minghong Geng, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2024. Benchmarking MARL on Long Horizon Sequential Multi-Objective Tasks: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

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 subtask 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.

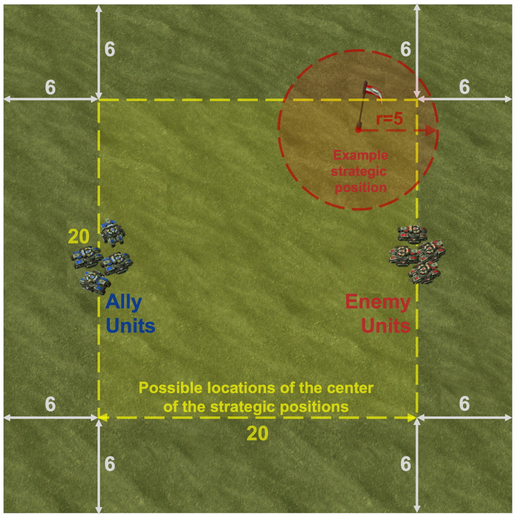
2 MULTI-OBJECTIVE SMAC (MOSMAC)

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

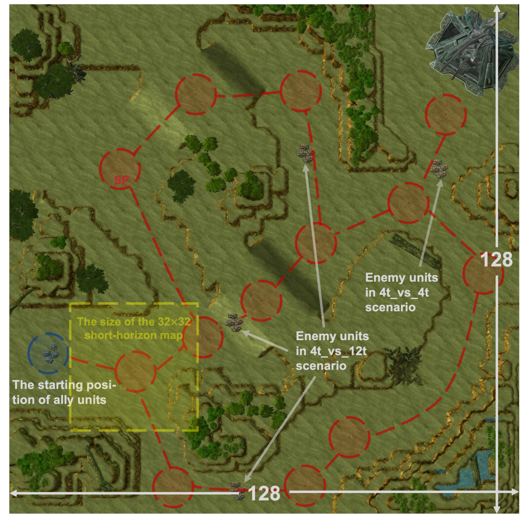
[1] Code is available at https://github.com/smu-ncc/mosmac

This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).



(a) An illustration of the short-horizon MOSMAC, 4t scenario.



(b) An illustration of the long-horizon MOSMAC with terrain features.

Figure 1: Illustrations of short-horizon and long-horizon MOSMAC. (a) The strategic position is marked by the dotted red circle, with a center drawn from a uniform distribution over the 20×20 area, marked by the dotted yellow square. The full map size is 32×32. (b) The yellow area shows the size of the short-horizon scenarios' map, which is depicted in Figure 1(a). 4t_vs_4t and 4t_vs_12t are the names of scenarios with 4 and 12 adversarial units.

in the StarCraft II games. In addition to the default environment information as in SMAC [20] and SMACv2 [5], i.e., units' information and optional terrain features, agents also perceive the relative direction and distance towards the strategic position to navigate effectively. The action space is discrete and contains four *movement* actions, one *attack* action, and one *stop* action. Agents can execute up to 50 decision-making and action cycles in 3t and 4t games, while this limit extends to 100 in 8t and 12t scenarios. The games will be forced to be terminated once agents reach this limit.

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:

$$r_{obj1} = \sum_{i=1}^{n} (r_a^i + r_d^i) \tag{1}$$

and

$$r_{obj2} = \sum_{i=1}^{n} r_r^i \tag{2}$$

respectively, 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 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:

$$r = \alpha \times r_{obj1} + (1 - \alpha) \times r_{obj2} \tag{3}$$

where α 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.

The long-horizon MOSMAC features three sets of subtasks, as illustrated in Figure 1(b). Each set of subtasks is derived by dissecting a path that commences at the starting position and ends at the final position, employing segmentation points as intermediate targets. Consequently, each subtask becomes a short-horizon MOMARL task akin to short-horizon MOSMAC. Agents need to address a series of interconnected subtasks, where the completion of one subtask triggers the beginning of the next. Each episode uniformly selects a path with a set of subtasks. In total, a full long-horizon MOSMAC task entails 6-8 subtasks. We expand the map to 128×128 and provide variations including fully flat terrain scenarios and settings with intricate topographical features. The ally team comprises four units, whereas the adversarial team encompasses 0, 4, or 12 units. To maintain parity in combat capabilities with the ally team, enemy units are organized into clusters, each with four units. Ally agents encounter at most one enemy cluster in each episode.

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. DSOCL20200) and the Jubilee Technology Fellowship awarded to Ah-Hwee Tan by Singapore Management University.

REFERENCES

[1] Wenhang Bao and Xiao-yang Liu. 2019. Multi-Agent Deep Reinforcement Learning for Liquidation Strategy Analysis. *arXiv preprint arXiv:1906.11046v1 [q-fin.TR]* (June 2019). https://doi.org/10.48550/arXiv.1906.11046

[2] Micah Carroll, Rohin Shah, Mark K. Ho, Thomas L. Griffiths, Sanjit A. Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-AI coordination. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 5174–5185. https://dl.acm.org/doi/10.5555/3454287.3454752

[3] Filippos Christianos, Georgios Papoudakis, Muhammad A. Rahman, and Stefano V. Albrecht. 2021. Scaling Multi-Agent Reinforcement Learning with Selective Parameter Sharing. In *Proceedings of the 38th International Conference on Machine Learning*. Proceedings of Machine Learning Research, Vol. 139. PMLR, 1989–1998. https://proceedings.mlr.press/v139/christianos21a.html

[4] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. 2020. Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (March 2020), 1086–1095. https://doi.org/10.1109/tits.2019.2901791

[5] Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob N. Foerster, and Shimon Whiteson. 2023. SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2212.07489v2 [cs.LG]* (Oct. 2023). https://doi.org/10.48550/arXiv.2212.07489

[6] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual Multi-Agent Policy Gradients. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (April 2018). https://doi.org/10.1609/aaai.v32i1.11794

[7] Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. 2022. A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems* 36, 1 (April 2022), 26. https://doi.org/10.1007/s10458-022-09552-y

[8] Tianmeng Hu, Biao Luo, Chunhua Yang, and Tingwen Huang. 2023. MO-MIX: Multi-Objective Multi-Agent Cooperative Decision-Making With Deep Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 10 (Oct. 2023), 1–15. https://doi.org/10.1109/TPAMI.2023.3283537

[9] Nan Jiang and Alekh Agarwal. 2018. Open Problem: The Dependence of Sample Complexity Lower Bounds on Planning Horizon. In *Proceedings of the 31st Conference On Learning Theory*. Proceedings of Machine Learning Research, Vol. 75. PMLR, 3395–3398. https://proceedings.mlr.press/v75/jiang18a.html ISSN: 2640-3498.

[10] Mohamed A. Khamis and Walid Gomaa. 2014. Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework. *Engineering Applications of Artificial Intelligence* 29 (2014), 134–151. https://doi.org/10.1016/j.engappai.2014.01.007

[11] Mingyu Kim, Jihwan Oh, Yongsik Lee, Joonkee Kim, Seonghwan Kim, Song Chong, and Seyoung Yun. 2023. The StarCraft Multi-Agent Exploration Challenges: Learning Multi-Stage Tasks and Environmental Factors Without Precise Reward Functions. *IEEE Access* 11 (2023), 37854–37868. https://doi.org/10.1109/ACCESS.2023.3266652

[12] Lucas Lehnert, Romain Laroche, and Harm van Seijen. 2018. On Value Function Representation of Long Horizon Problems. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (April 2018). https://doi.org/10.1609/aaai.v32i1.11646

[13] Chunming Liu, Xin Xu, and Dewen Hu. 2015. Multiobjective Reinforcement Learning: A Comprehensive Overview. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 3 (March 2015), 385–398. https://doi.org/10.1109/TSMC.2014.2358639

[14] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 6382–6393. https://dl.acm.org/doi/10.5555/3295222.3295385

[15] Patrick Mannion, Karl Mason, Sam Devlin, Jim Duggan, and Enda Howley. 2016. Multi-Objective Dynamic Dispatch Optimisation using Multi-Agent Reinforcement Learning: (Extended Abstract). In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, USA, 1345–1346. http://dl.acm.org/citation.cfm?id=2937152

[16] Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V. Albrecht. 2021. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*. Vol. 1. Curran Associates Inc., Red Hook, NY, USA. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/a8baa56554f96369ab93e4f3bb068c22-Abstract-round1.html

[17] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic value function factorisation for deep multi-agent reinforcement learning. *The Journal of Machine Learning Research* 21, 1 (2020), 7234–7284. https://dl.acm.org/doi/abs/10.5555/3455716.3455894

[18] Cinjon Resnick, Wes Eldridge, David Ha, Denny Britz, Jakob Foerster, Julian Togelius, Kyunghyun Cho, and Joan Bruna. 2018. PommerMan: A multi-agent playground. *CEUR Workshop Proceedings* 2282 (2018). http://www.scopus.com/inward/record.url?scp=85059815518&partnerID=8YFLogxK

[19] Diederik Marijn Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A Survey of Multi-Objective Sequential Decision-Making. *Journal of Artificial Intelligence Research* 48, 1 (Oct. 2013), 67–113. https://doi.org/10.1613/jair.3987

[20] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. *arXiv preprint arXiv:1902.04043v5 [cs.LG]* (Dec. 2019). http://arxiv.org/abs/1902.04043

[21] Christian Schroeder de Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. 2020. Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge? *arXiv preprint arXiv:2011.09533v1 [cs.AI]* (Nov. 2020). https://doi.org/10.48550/arXiv.2011.09533

[22] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z. Leibo, Karl Tuyls, and Thore Graepel. 2017. Value-Decomposition Networks For Cooperative Multi-Agent Learning. *arXiv preprint arXiv:1706.05296v1 [cs.AI]* (June 2017). https://doi.org/10.48550/arXiv.1706.05296

[23] Ardi Tampuu, Tambet Matiisen, Dorian Kodelja, Ilya Kuzovkin, Kristjan Korjus, Juhan Aru, Jaan Aru, and Raul Vicente. 2017. Multiagent cooperation and competition with deep reinforcement learning. *PLoS ONE* 12, 4 (April 2017). https://doi.org/10.1371/journal.pone.0172395

[24] Kristof Van Moffaert, Tim Brys, Arjun Chandra, Lukas Esterle, Peter R. Lewis, and Ann Nowe. 2014. A novel adaptive weight selection algorithm for multi-objective multi-agent reinforcement learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Beijing, China, 2306–2314. https://doi.org/10.1109/IJCNN.2014.6889637

[25] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and YI WU. 2022. The surprising effectiveness of PPO in cooperative multi-agent games. In *Advances in neural information processing systems*. Vol. 35. Curran Associates, Inc., Red Hook, NY, USA, 24611–24624. https://proceedings.neurips.cc/paper_files/paper/2022/file/9c1535a02f0ce079433344e14d910597-Paper-Datasets_and_Benchmarks.pdf