Scaling up Cooperative Multi-agent Reinforcement Learning Systems

Doctoral Consortium

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

ABSTRACT

Cooperative multi-agent reinforcement learning methods aim to learn effective collaborative behaviours of multiple agents performing complex tasks. However, existing MARL methods are commonly proposed for fairly small-scale multi-agent benchmark problems, wherein both the number of agents and the length of the time horizons are typically restricted. My initial work investigates hierarchical controls of multi-agent systems, where a unified overarching framework coordinates multiple smaller multi-agent subsystems, tackling complex, long-horizon tasks that involve multiple objectives. Addressing another critical need in the field, my research introduces a comprehensive benchmark for evaluating MARL methods in long-horizon, multi-agent, and multi-objective scenarios. This benchmark aims to fill the current gap in the MARL community for assessing methodologies in more complex and realistic scenarios. My dissertation would focus on proposing and evaluating methods for scaling up multi-agent systems in two aspects: structural-wise increasing the number of reinforcement learning agents and temporal-wise extending the planning horizon and complexity of problem domains that agents are deployed in.

KEYWORDS

Multi-agent Reinforcement Learning; Scaling up MARL; Long-horizon MARL; Hierarchical Multi-agent Systems; Task Decomposition

ACM Reference Format:
Minghong Geng. 2024. Scaling up Cooperative Multi-agent Reinforcement Learning Systems: Doctoral Consortium. 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

Multi-agent reinforcement learning (MARL) enables autonomous agents to develop collaborative strategies for addressing complex challenges. Over the last decade, this domain has advanced significantly, with numerous methods achieving notable success across various benchmarks and applications. A critical focus within the MARL community is the scalability of multi-agent systems (MAS), especially in terms of the number of agents and the time horizons for decision-making. Conventionally, MARL methods focus on small-scale problems, where the number of agents and the decision-making steps per agent taken within each episode are constrained. Significant challenges still lie in effectively scaling up MAS, i.e., structurally increasing the number of reinforcement learning (RL) agents and temporally extending the time horizons of multi-agent decision-making.

Increasing the number of RL agents in MAS, along with long-horizon planning, significantly boosts their ability to tackle complex problems unmanageable by smaller-scale systems. This is particularly evident in real-world scenarios where extensive, sustained efforts from numerous agents are essential. For example, managing urban traffic often involves thousands of traffic lights working continuously to optimize city-level traffic networks [23]. However, scaling up MAS introduces many challenges, such as structural and temporal credit assignment [1], non-stationarity [12], and the curse of dimensionality [7]. Furthermore, despite the growing popularity of research on scaling up MARL, unified benchmarks for evaluating these methods remain scarce.

My dissertation focuses on scaling multi-agent reinforcement learning systems through *task decomposition*, which segments large, complex multi-agent problems into smaller, manageable subproblems. Specifically, my research involves two facets of task decomposition. The first approach, *structural task decomposition*, divides large-scale MAS into smaller autonomous modules called *mini-MAS* and develops effective cooperation mechanisms among mini-MAS. The second, *temporal task decomposition*, focuses on dividing long-horizon, complex problems into shorter, simpler subproblems. Importantly, these two approaches are interconnected: mini-MAS, structured through structural task decomposition, are ideally suited for addressing subtasks generated by temporal task decomposition.

My initial study presents HiSOMA [5], a hierarchical multi-agent reinforcement learning system with task decomposition. It features a three-level hierarchy, utilizing mini-MAS as intermediary function modules to learn policies through various MARL algorithms to address subproblems. In my subsequent study, I develop MOSMAC [6], a challenging MARL benchmark aimed at assessing state-of-the-art MARL methods in complex long-horizon scenarios involving multiple objectives.

2 MARL WITH HIERARCHICAL CONTROL

HiSOMA [5] is a hybrid hierarchical MARL model that combines a class of Self-Organizing Neural Network (SONN), named Fusion Architecture for Learning, Cognition, and Navigation (FALCON) [20, 21], with state-of-the-art non-hierarchical MARL methods to navigate complex, long-horizon decision-making problems. HiSOMA implements two task decomposition strategies: clustering

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).

agents into mini-MAS via structural task decomposition to establish a hierarchical control architecture and dividing long-horizon domain-specific *tasks* into short-horizon *subtasks* using temporal task decomposition. Different from conventional hierarchical MARL approaches [2, 9, 10, 17], where the focus is the cooperation and communication among agents, HiSOMA emphasizes efficient coordination among mini-MAS. At its core, the adaptability of HiSOMA is anchored in mini-MAS — fully functional autonomous multi-agent modules that interact with environments through information exchanges and primitive actions, compatible with many state-of-the-art MARL methods. Such a configuration enables deeper hierarchical control and task decomposition, enhancing HiSOMA's ability to address complex problems.

In HiSOMA, while mini-MAS focus on specific subtasks, the central controller, FALCON, is pivotal for global task decomposition and sequential subtasks allocation. It monitors the global states and assigns suitable subtasks to each middle-level controller, i.e., mini-MAS, which subsequently distributes subtasks as *intrinsic goals* to its lower-level agents. The lower-level agents execute primitive actions and directly interact with the environments according to the received intrinsic goals. The experiments on the MOSMAC benchmark (see Section 3) demonstrate HiSOMA's effectiveness in scaled-up scenarios with long-horizon planning compared to non-hierarchical MARL methods like QMIX [14]. As HiSOMA adopts FALCON as the central controller, characteristics like *cognitive codes* could be utilized so that the learning progress of HiSOMA can be effortlessly tracked and analyzed. Moreover, HiSOMA allows for controllers to be pre-trained on subtasks and subsequently fine-tuned upon integration. This approach significantly reduces the training costs for long-horizon MARL, addressing a practical challenge of end-to-end training over long trajectories.

3 MOSMAC: A BENCHMARK WITH VARYING HORIZON AND MULTI-OBJECTIVES

There is a notable gap in the literature concerning MARL benchmarks tailored for scaled-up, long-horizon challenges. Moving beyond the HiSOMA model, another focus of my research is to benchmark existing state-of-the-art MARL algorithms against complex, long-horizon multi-agent tasks. To this end, I propose a new benchmark named *multi-objective SMAC* (MOSMAC) [6], which offers a variety of multi-objective tasks scalable across different time horizons. MOSMAC is characterized by its combination of multiple objectives, diverse temporal scales, and complex navigation terrains.

Specifically, MOSMAC's tasks involve dual objectives: engaging adversarial units and navigating to *strategic positions*. Agents must balance these objectives to complete tasks with varying horizons. To further mirror realistic scenarios in long-horizon tasks, MOSMAC challenges agents with tasks involving sequences of multi-objective subtasks targeting different strategic positions. By changing the target strategic positions and relative positions of enemies, agents can be exposed to a large variation of tasks in short-horizon cases and subtask sequences in long-horizon cases. These variations in objectives and task sequences provide a wide range of scenarios, which are further complicated by complex terrain features like plains, canyons, ramps, and high/low grounds, which are rarely considered by existing benchmarks in StarCraft II [3, 8, 15].

We evaluated nine popular MARL algorithms on MOSMAC using the EPyMARL framework [13]. These algorithms include IA2C [13], IPPO [16], COMA [4], MAA2C [13], MAPPO [22], IQL [19], MADDPG [11], VDN [18], and QMIX [14]. Our findings reveal that while current MARL algorithms excel in scenarios with lower stochasticity, they face challenges in more generalized tasks involving multiple objectives over extended horizons. Interestingly, the results reveal that while centralized training with decentralized execution (CTDE) algorithms typically surpasses independent learning methods in many benchmarks, independent learning algorithms showed superior performance in highly stochastic scenarios with multi-objective cooperation, especially with complex terrain features, as seen in MOSMAC. This observation indicates that although CTDE approaches mitigate MARL issues like non-stationarity and the curse of dimensionality, these benefits are often offset by the high cost associated with centralized training in large-scale MAS. Consequently, decomposing large-scale MAS into independent CTDE modules is a potential strategy for scaling up MARL methods.

4 TOWARDS LARGE AGENT TEAM AND LONG-HORIZON PLANNING

In this paper, I have presented HiSOMA, a hybrid hierarchical MARL model that integrates SONN with MARL to scale up multi-agent reinforcement learning systems for long-horizon problems, and MOSMAC, a challenging MARL benchmark featuring long-horizon multi-objective MARL problems. Moving forward, my dissertation aims to broaden the scope by exploring two pivotal research directions of multi-agent reinforcement learning systems, namely, towards large agent teams and long-horizon planning. These directions are intrinsically linked, with the former concentrating on increasing the number of agents through novel MARL algorithms and MAS architectures and the latter focusing on applying and evaluating these MARL approaches on long-horizon problems.

Structurally scaling up MARL methods with hierarchical strategies, such as employing hierarchical MARL algorithms and hierarchical MAS structures like HiSOMA, is an interesting avenue of research. Temporally scaling up MARL methods for long-horizon planning necessitates a certain level of *temporal abstraction* to transform the long-horizon learning problems into multiple short-horizon subproblems, also inherently calling for a hierarchical framework to decompose and allocate subproblems efficiently. Therefore, adopting hierarchical architectures in MARL is a promising research direction for scaling up MARL methods.

Successfully scaling up MARL methods significantly hinges on the availability of appropriate benchmarks and environments. While a number of MARL benchmarks have recently emerged in the MARL community, there is a notable lack of environments for structurally and temporally scaled-up MARL methods. Consequently, it is imperative to develop benchmarks and tasks tailored for scaled-up MARL methods and to establish standard evaluation criteria.

ACKNOWLEDGMENTS

This research was supported by the DSO National Laboratories, Singapore (Agreement No. DSOCL20200).

REFERENCES

[1] Adrian K. Agogino and Kagan Tumer. 2004. Unifying Temporal and Structural Credit Assignment Problems. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2 (AAMAS '04)*. IEEE Computer Society, USA, 980–987. https://dl.acm.org/doi/10.5555/1018410.1018852

[2] Sanjeevan Ahilan and Peter Dayan. 2019. Feudal Multi-Agent Hierarchies for Cooperative Reinforcement Learning. *arXiv preprint arXiv:1901.08492 [cs.MA]* (Jan. 2019). https://doi.org/10.48550/arXiv.1901.08492

[3] 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

[4] Jakob N. Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, New Orleans, Louisiana, USA, 2974–2982. https://dl.acm.org/doi/10.5555/3504035.3504398

[5] Minghong Geng, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2023. HiSOMA: A Hierarchical Multi-Agent Model Integrating Self-Organizing Neural Networks with Multi-Agent Deep Reinforcement Learning. *Available at SSRN* (Dec. 2023). https://doi.org/10.2139/ssrn.4666277

[6] Minghong Geng, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2024. Benchmarking MARL on Long Horizon Sequential Multi-Objective Tasks. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024)*. International Foundation for Autonomous Agents and Multiagent Systems, Auckland, New Zealand.

[7] Pablo Hernandez-Leal, Michael Kaisers, Tim Baarslag, and Enrique Munoz de Cote. 2019. A Survey of Learning in Multiagent Environments: Dealing with Non-Stationarity. *arXiv preprint arXiv:1707.09183v2 [cs.MA]* (March 2019). https://doi.org/10.48550/arXiv.1707.09183

[8] 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

[9] Xiangyu Kong, Bo Xin, Fangchen Liu, and Yizhou Wang. 2017. Revisiting the Master-Slave Architecture in Multi-Agent Deep Reinforcement Learning. *arXiv preprint arXiv:1712.07305v1 [cs.AI]* (Dec. 2017). https://doi.org/10.48550/arXiv.1712.07305

[10] Saurabh Kumar, Pararth Shah, Dilek Hakkani-Tur, and Larry Heck. 2017. Federated Control with Hierarchical Multi-Agent Deep Reinforcement Learning. *arXiv preprint arXiv:1712.08266v1 [cs.AI]* (Dec. 2017). https://doi.org/10.48550/arXiv.1712.08266

[11] 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 (NIPS'17)*. Curran Associates Inc., Long Beach, California, USA, 6382–6393. https://dl.acm.org/doi/10.5555/3295222.3295385

[12] Georgios Papoudakis, Filippos Christianos, Arrasy Rahman, and Stefano V. Albrecht. 2019. Dealing with Non-Stationarity in Multi-Agent Deep Reinforcement Learning. *arXiv preprint arXiv:1906.04737v1 [cs.LG]* (June 2019). https://doi.org/10.48550/arXiv.1906.04737

[13] 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. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/a8baa56554f96369ab93e4f3bb068c22-Abstract-round1.html

[14] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, Stockholm, Sweden, 4295–4304. https://proceedings.mlr.press/v80/rashid18a.html

[15] 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

[16] 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

[17] Jianzhun Shao, Zhiqiang Lou, Hongchang Zhang, Yuhang Jiang, Shuncheng He, and Xiangyang Ji. 2022. Self-Organized Group for Cooperative Multi-agent Reinforcement Learning. In *Advances in Neural Information Processing Systems 35*, Vol. 35. Curran Associates, Inc., New Orleans, Louisiana, USA, 5711–5723. https://proceedings.neurips.cc/paper_files/paper/2022/hash/25b040c97a75021e57100648a20b1e10-Abstract-Conference.html

[18] 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

[19] 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

[20] Ah-Hwee Tan. 2004. FALCON: a fusion architecture for learning, cognition, and navigation. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Vol. 4. IEEE, Budapest, Hungary, 3297–3302. https://doi.org/10.1109/IJCNN.2004.1381208

[21] Ah-Hwee Tan, Budhitama Subagdja, Di Wang, and Lei Meng. 2019. Self-organizing neural networks for universal learning and multimodal memory encoding. *Neural Networks* 120 (Dec. 2019), 58–73. https://doi.org/10.1016/j.neunet.2019.08.020

[22] 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 35*, Vol. 32. Curran Associates, Inc., New Orleans, Louisiana, USA, 24611–24624. https://proceedings.neurips.cc/paper_files/paper/2022/hash/9c1535a02f0ce079433344e14d910597-Abstract-Datasets_and_Benchmarks.html

[23] Huichu Zhang, Siyuan Feng, Chang Liu, Yaoyao Ding, Yichen Zhu, Zihan Zhou, Weinan Zhang, Yong Yu, Haiming Jin, and Zhenhui Li. 2019. CityFlow: A Multi-Agent Reinforcement Learning Environment for Large Scale City Traffic Scenario. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 3620–3624. https://doi.org/10.1145/3308558.3314139