Scaling Up Cooperative Multi-Agent Reinforcement Learning Through Hierarchical Heterogeneous Modular Architectures

AAAI-26 / IAAI-26 / EAAI-26

Minghong Geng

Neural and Cognitive Computing Group
Singapore Management University
Email: mhgeng.2021@phdcs.smu.edu.sg
Homepage: https://gengminghong.github.io

My Homepage

GitHub Repo

Motivation

- Multi-agent systems (MAS) have emerged as a powerful paradigm for solving complex real-world problems that require coordination among multiple autonomous agents.
- Multi-agent reinforcement learning (MARL) faces challenges in scaling to complex scenarios w. sustained planning and coordination across long horizons.
- To overcome the scalability challenges, this work explores heterogeneous modular systems.
- The core idea is to leverage the complementary strengths of different computational paradigms.

Research Questions

This work attempts to answer the following questions.

1: How is the scalability of current MARL methods?

2: What design principles and metrics are needed to benchmark scaled-up multi-agent systems?

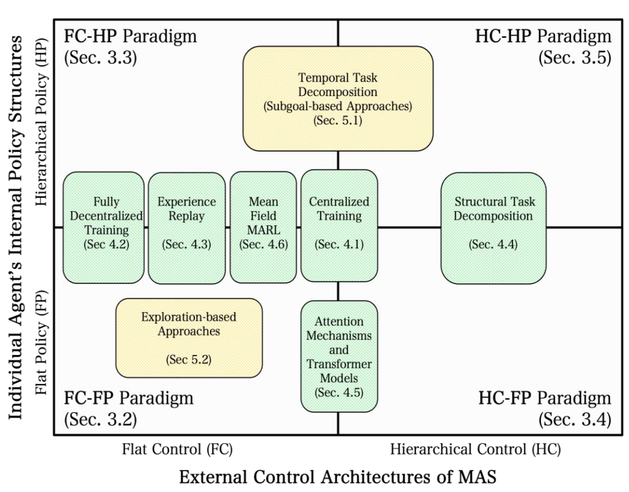
3: How should hierarchical frameworks address scaled multi-agent problems?

4: How can we further enhance the capabilities of high-level agents in hierarchical multi-agent frameworks?

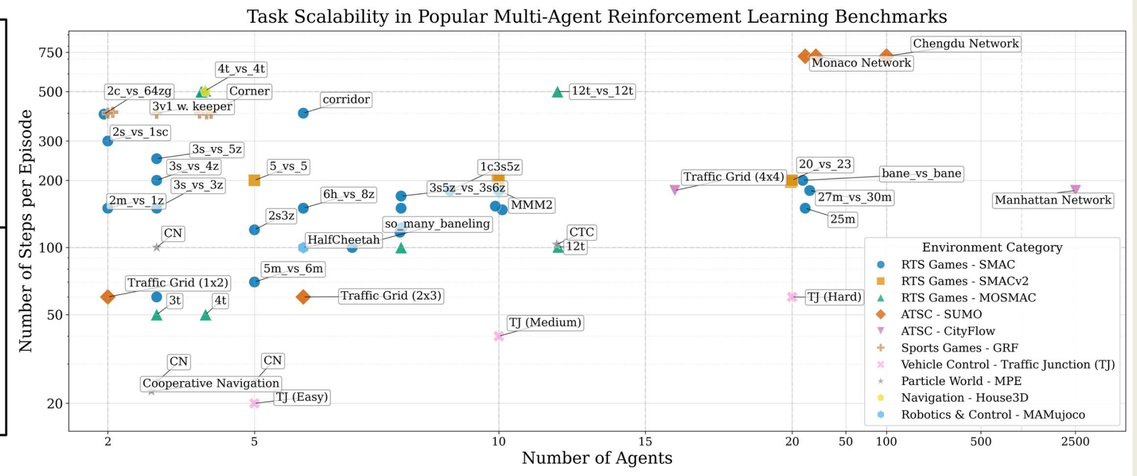
5: How to interpret the learned policies and behaviors of MARL agents?

Part 1 - A comprehensive *Survey* on Scaling up MARL

We analyze 120+ papers on MARL published in top venues and present a novel taxonomy based on the design principles of MAS.



The taxonomy on MARL methods.



The scale of MARL tasks in representative environments.

The results indicate that hierarchical approaches represent promising avenues for scaling MARL systems. It also highlights a critical shortage of suitable MARL benchmarks that better reflect real-world scaling requirements.

[1] **Minghong Geng**, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. Scaling up MARL for Large Agent Teams and Long-horizon Tasks: A Survey. In Press.

Part 2 - MOSMAC: A Multi-objective MARL *Benchmark*

MOSMAC (Multi-Objective SMAC) is a new benchmark featuring multiple objectives, sequential task allocation, and dynamic horizons.

1 Multiple Objectives

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

Objective 2 (safety): The damage to ally agents should be as low as possible.

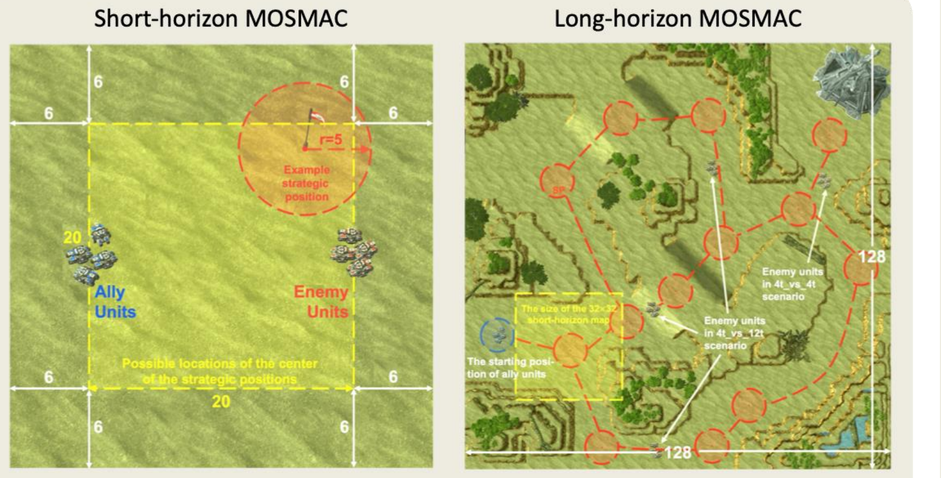
Objective 3 (navigation): The distance from agents to strategic positions should be as low as possible.

2 Sequential Task Allocation

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 Varying Horizons

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.



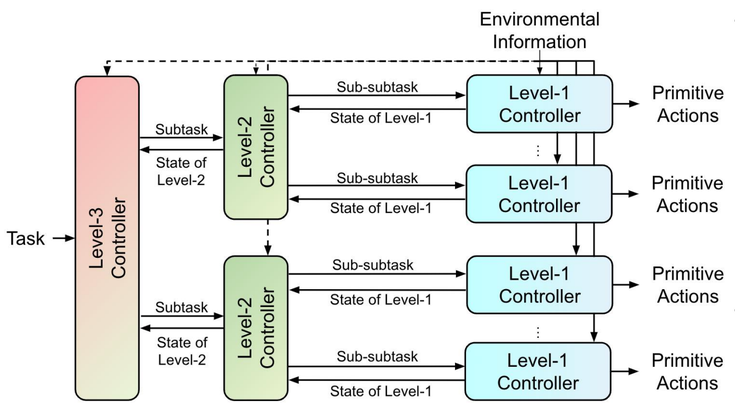
An illustration of a short-horizon MOSMAC scenario named 4t. The winning condition is all surviving agents arrive at the strategic position, motivated by multiple objectives. The locations are randomly selected in each episode.

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.

[2] **Minghong Geng**, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. MOSMAC: A Multi-agent Reinforcement Learning Benchmark on Sequential Multi-objective Tasks. In AAMAS '25.
[3] **Minghong Geng**, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. Benchmarking MARL on Long Horizon Sequential Multi-Objective Tasks. In AAMAS '24.

Part 3 - HiSOMA *Framework*: SONN Integrating MARL

HiSOMA is a hybrid three-level hierarchical MARL framework that integrates self-organizing neural network (SONN) and MARL.

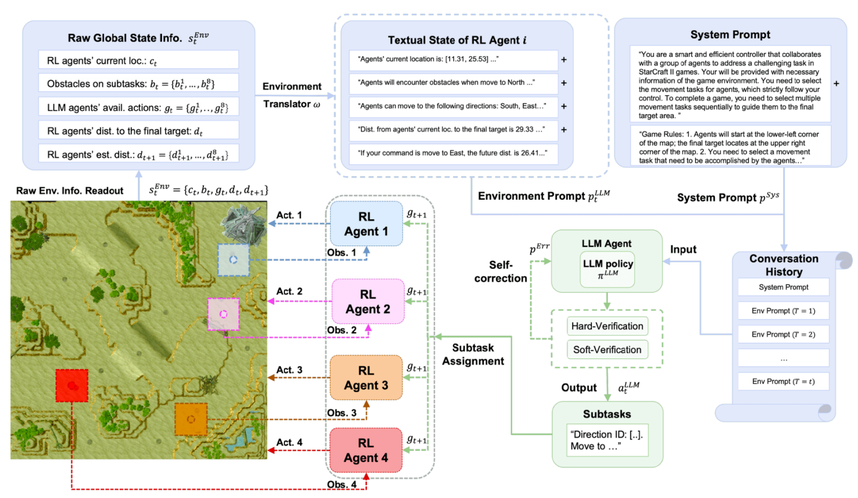


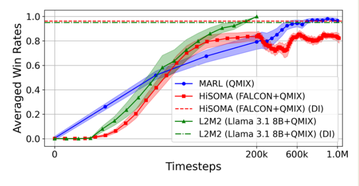
- HiSOMA is a hybrid hierarchical MARL framework that combines a class of Self-Organizing Neural Network (SONN), named Fusion Architecture for Learning, Cognition, and Navigation (FALCON) and state-of-the-art non-hierarchical MARL methods to navigate long-horizon decision-making problems.
- HiSOMA is a novel attempt to incorporate heterogenous learning algorithms into a unified framework and is able to generalized to abroad-range of MARL methods.

[4] **Minghong Geng**, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2024. HiSOMA: A hierarchical multi-agent model integrating Self-Organizing Neural Networks with multi-agent deep reinforcement learning. ESwA.

Part 4 - L2M2: An LLM-MARL *Framework*

L2M2 integrates MARL with LLM for automated hierarchical control without generating pre-defined subgoals with domain knowledge.



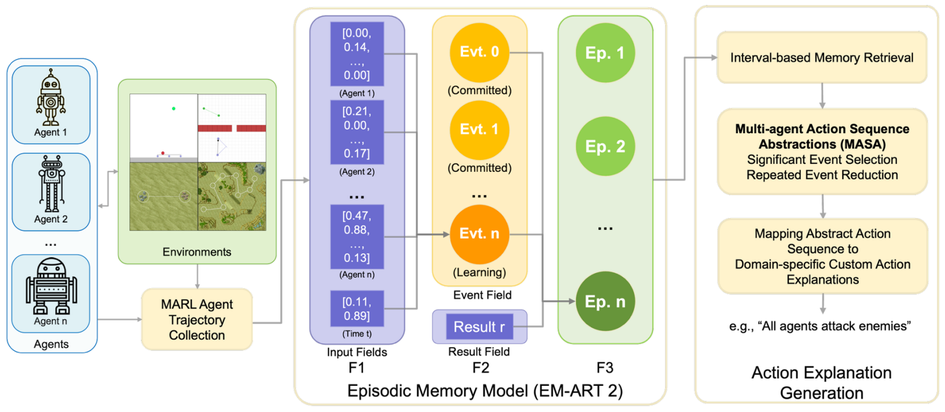


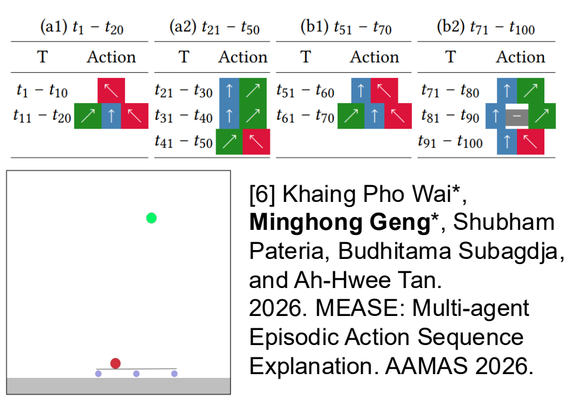
L2M2 demonstrates superior *performance* and *sample efficiency (20%)*.

[5] **Minghong Geng**, Shubham Pateria, Budhitama Subagdja, Lin Li, Xin Zhao, and Ah-Hwee Tan. 2025. L2M2: A Hierarchical Framework Integrating Large Language Model and Multi-agent Reinforcement Learning. IJCAI '25.

Part 5 – MEASE: A *Framework* Explaining MARL Policies

MEASE is an explainable MARL (XMARL) framework that explains MARL policies by learning episodic memories from multi-agent behavioral data and explaining them as interpretable coordination behaviors.





[6] Khaing Pho Wai*, **Minghong Geng***, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2026. MEASE: Multi-agent Episodic Action Sequence Explanation. AAMAS 2026.

Conclusions and Future Work

- The ability to decompose problems, coordinate capabilities, and maintain interpretable decision-making are crucial for developing MAS systems to address pressing challenges.
- Broader future research directions include unified intelligence architecture to conceptualize MAS as components of a single, unified artificial intelligence entity.
- Such approach integrates heterogeneous modules, such as self-organizing neural networks, large language models, and deep (reinforcement) learning model, to construct a comprehensive MAS in which these modules serve specialized cognitive functions.