Multiagent Learning via Dynamic Skill Selection

Enna Sachdeva, Shauharda Khadka, Somdeb Majumdar, Kagan Tumer

Oregon State University

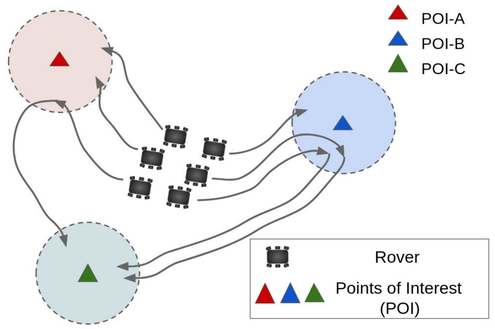
intel

Multiagent Learning with sparse rewards

Learning coordination strategies with sparse rewards is challenging.

High spatial and temporal coupling makes the problem harder

Sequential Tasks



Search and Rescue Task:

Spatial Coupling: Requires more than one robot to simultaneously lift a person/ object. eg.- 2 agents need to observe a POI.

Temporal Coupling: Agents need to sequentially complete different components of task. First execute a "search" policy in a team before executing a "rescue" policy in another team. For eg.- Observe

POI A->POI-B->POI-C

An agent receives reward when the team finishes the task.

Thanks

This work was partially supported by the NSF (iis-1815886), afosr(fa9550-19-1-0195), and Intel.

MADyS

Bi-level optimization framework that leverages a portfolio of semantically meaningful local rewards to optimize team reward.

Local reward represents a basic skill, hand-designed based on domain knowledge.

Coordinates agents across time and space.

Augments CCEA with policy gradient methods, to address structural credit assignment problem.

1. **Local Reward Optimization:** Policy gradient methods optimize for dense local rewards to learn local skills.
2. **Dynamic Skill Selection:** CCEA searches the most optimal local skill for each agent at each time, to optimize team reward.
3. **Shared Replay Buffer:** Enables sharing of information across evolutionary population and the skill learners.

Concurrent learning of local skills and team reward optimization, via shared replay buffer, achieves better sample efficiency.

MADyS outperforms Prior methods

Outperforms baselines like CCEA and MFL:

CCEA: Operates directly on low level actions

MFL: CCEA searches the most optimal skills, which are pre-trained separately in simple environment without access to team objective.

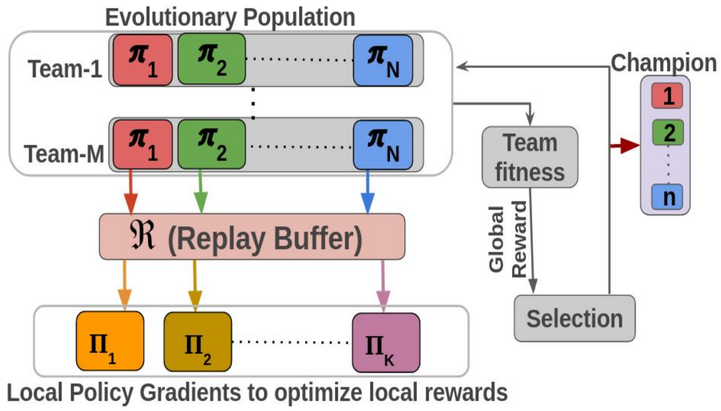
MADyS learns to solve long horizon task by solving subtasks, by dynamically forming sub-teams.

Results demonstrate MADyS is more sample efficient than MFL.

Conclusions and Future Work

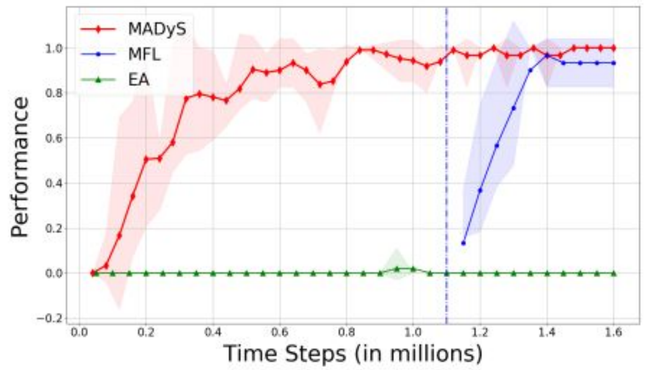
MADyS leverages local skills to solve complex coordination tasks.

Solves long horizon tasks with sparse rewards by allowing multiagent teams to dynamically select from local policies trained on different dense local objectives.

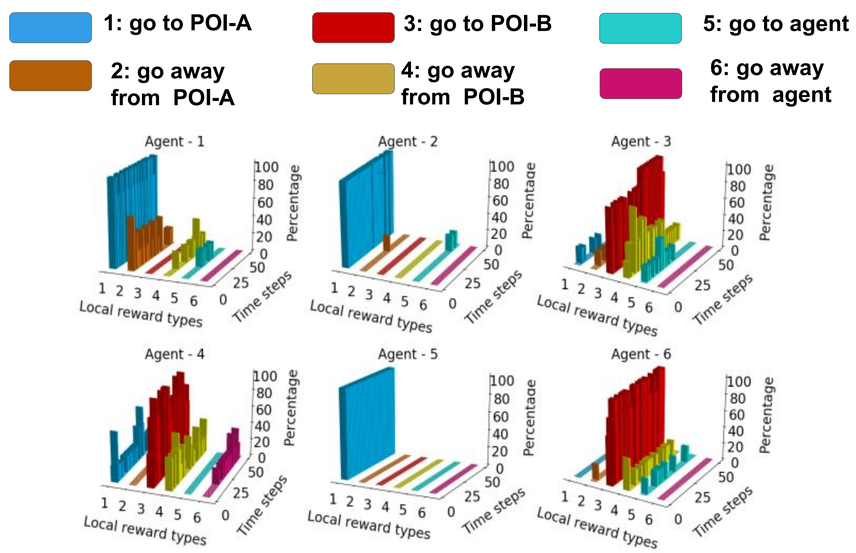


MADyS Architecture

π_n : Policy of an n^{th} agent, Π_k: Policy to learn a k^{th} skill



Training curves of MADyS, MFL and CCEA



Histograms showing distribution shift of local skills with time