N-agent Ad Hoc Teamwork

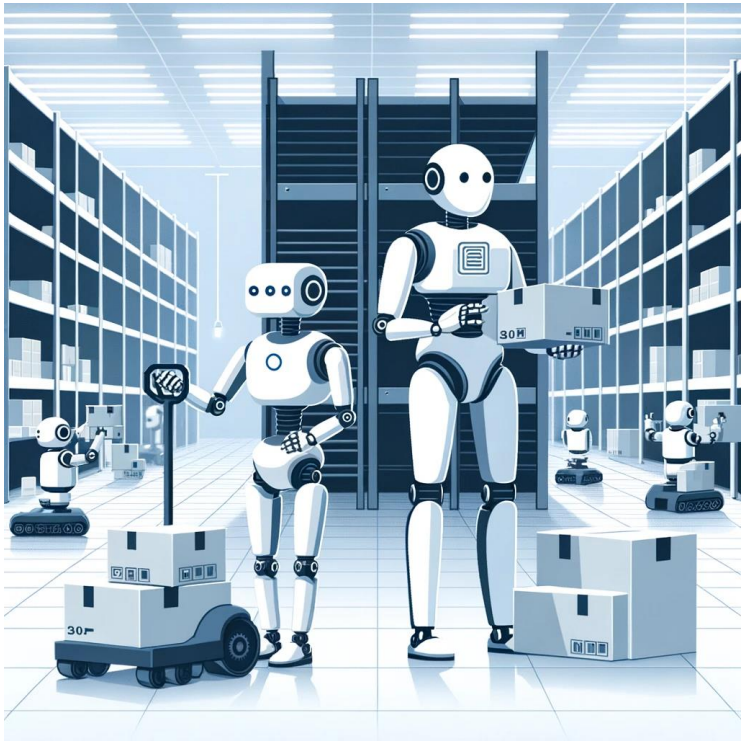
Caroline Wang, Arrasy Rahman, Ishan Durugkar,
Elad Liebman, Peter Stone

Neurips 2024

NEURAL INFORMATION PROCESSING SYSTEMS

LARG Learning Agents Research Group
The University of Texas at Austin

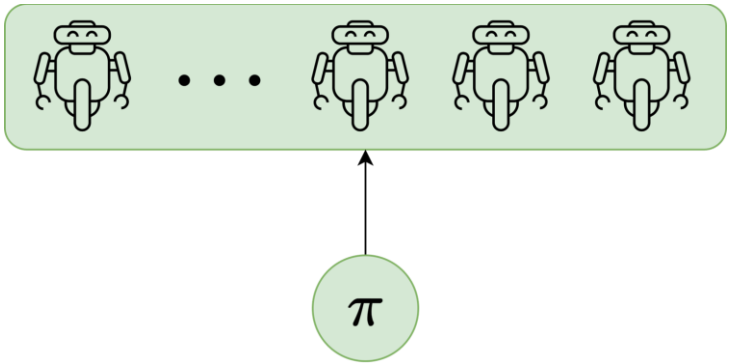
Sony AI



*generated by ChatGPT and DALL-E

Current multi-agent learning paradigms are not flexible enough.

- Cooperative Multi-Agent Reinforcement Learning[1] (C-MARL) assumes **all** agents are under control of learning algorithm
- Ad Hoc Teamwork[2] (AHT) & Zero Shot Coordination[3] (ZSC): assumes a **single** agent under control of learning algorithm

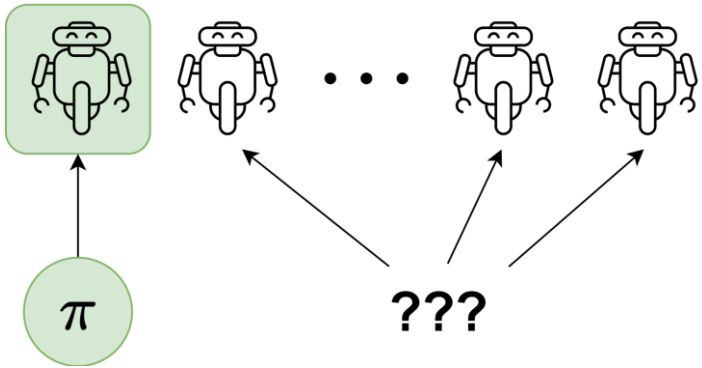


How can *sets* of agents coordinate with each other?

[1] Sunehag et al., Value Decomposition Networks for Cooperative Multiagent learning, AAMAS 2018.
[2] Mirsky et al. A Survey of Ad Hoc Teamwork Research. EUMAS 2022.
[3] Hu et al. “Other-Play” for Zero-Shot Coordination. ICML 2020

Current multi-agent learning paradigms are not flexible enough.

- Cooperative Multi-Agent Reinforcement Learning[1] (C-MARL) assumes **all** agents are under control of learning algorithm
- Ad Hoc Teamwork[2] (AHT) & Zero Shot Coordination[3] (ZSC): assumes a **single** agent under control of learning algorithm

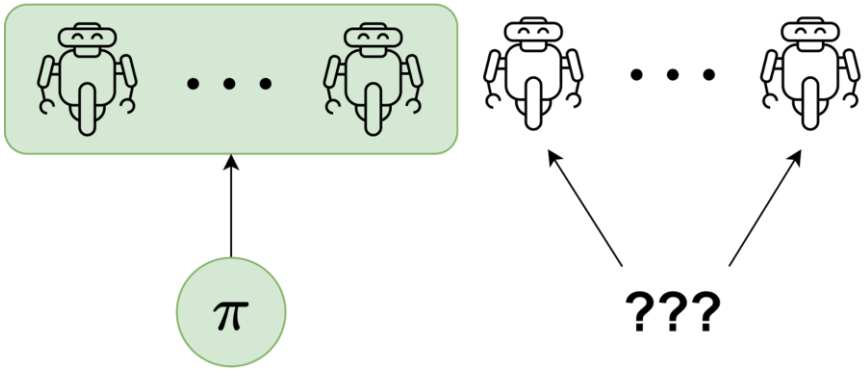


How can *sets* of agents coordinate with each other?

[1] Sunehag et al., Value Decomposition Networks for Cooperative Multiagent learning, AAMAS 2018.
[2] Mirsky et al. A Survey of Ad Hoc Teamwork Research. EUMAS 2022.
[3] Hu et al. “Other-Play” for Zero-Shot Coordination. ICML 2020

Current multi-agent learning paradigms are not flexible enough.

- Cooperative Multi-Agent Reinforcement Learning[1] (C-MARL) assumes **all** agents are under control of learning algorithm
- Ad Hoc Teamwork[2] (AHT) & Zero Shot Coordination[3] (ZSC): assumes a **single** agent under control of learning algorithm



How can *sets* of agents coordinate with each other?

[1] Sunehag et al., Value Decomposition Networks for Cooperative Multiagent learning, AAMAS 2018.
[2] Mirsky et al. A Survey of Ad Hoc Teamwork Research. EUMAS 2022.
[3] Hu et al. “Other-Play” for Zero-Shot Coordination. ICML 2020

Problem Statement

N-agent ad hoc teamwork (NAHT):

To create a **set** of autonomous agents that are able to efficiently and robustly collaborate with previously unknown teammates on tasks to which they are all individually capable of contributing as team members.

[1] Stone et al., Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination, AAAI 2018.

N-agent Ad Hoc Teamwork (NAHT)

Dec-POMDP

$$(M, \mathcal{S}, \mathcal{A}, \mathcal{O}_i, \mathcal{T}, r, T)$$

team size

joint state space

$$\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_M$$

joint action space

per-agent observation function

transition function

cooperative task reward

horizon

$$(\mathcal{C}, \mathcal{U}, \mathcal{X})$$

N-agent Ad Hoc Teamwork

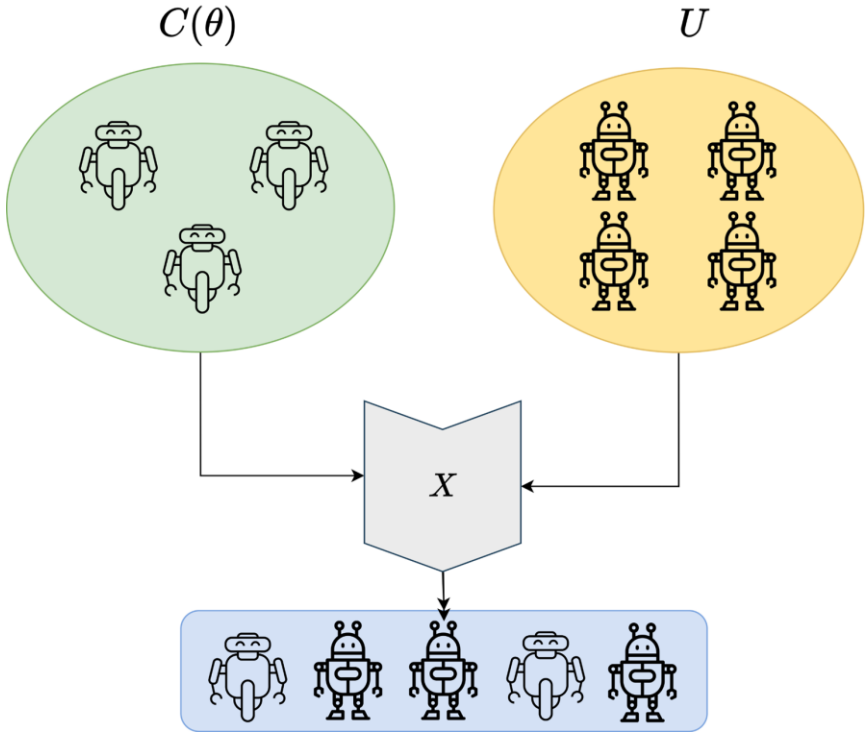
C(θ) - set of controlled agents, parameterized by *θ*

U - set of non-controlled agents

X - team sampling procedure

Objective:

$$\max_{\theta} \left(\mathbb{E}_{\boldsymbol{\pi}^{(M)} \sim X(U, C(\theta))} \left[\sum_{t=0}^{T} \gamma^t r_t \right] \right)$$



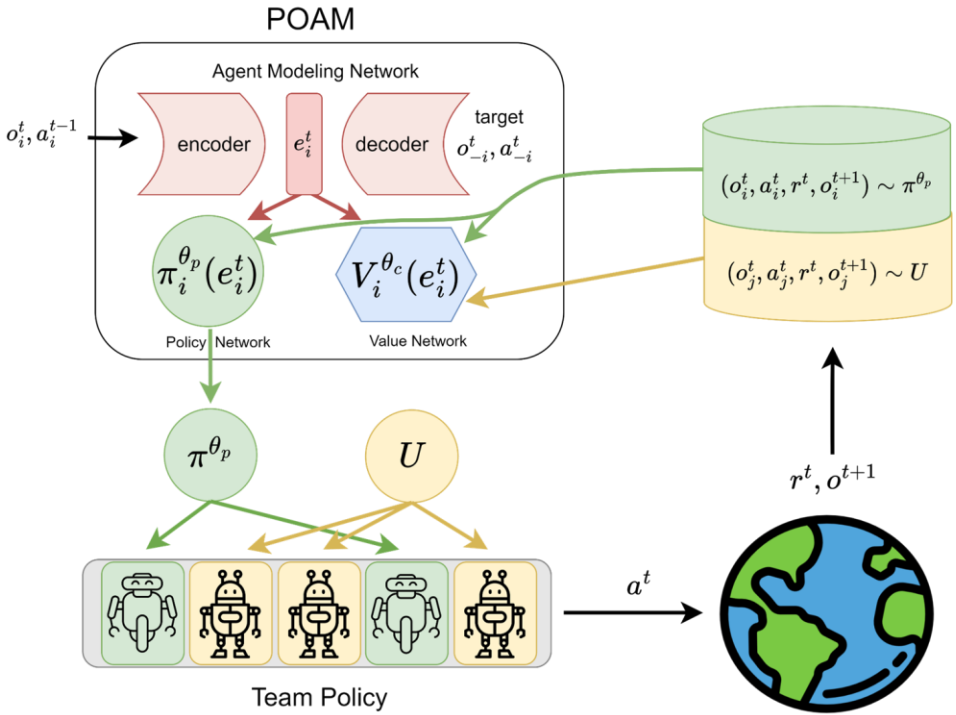
Challenges:

1) Generalization: Coordinating with non-controlled and potentially unknown teammates
2) Openness: coping with an unknown number of controlled teammates

Policy Optimization with Agent Modeling (POAM)

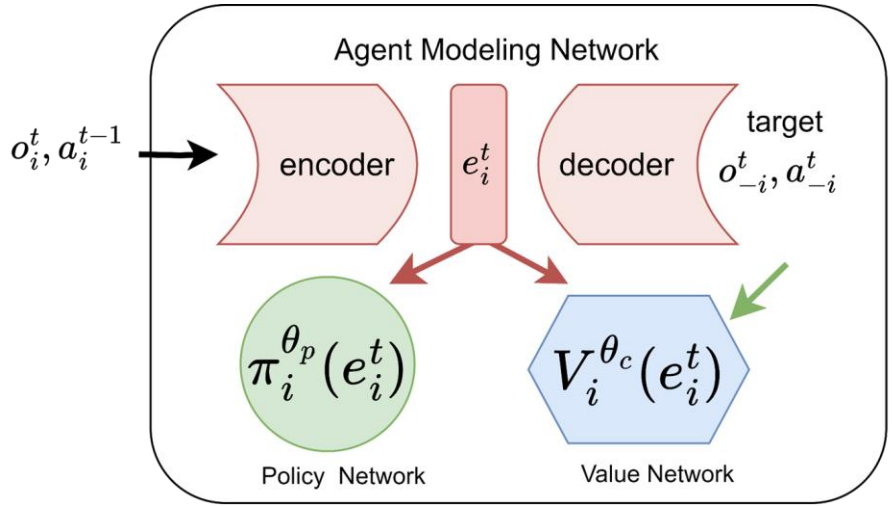
Key Ideas

- Independent PPO with parameter sharing
- Training critic with data from controlled and noncontrolled agents
- Teammate modelling



Policy Optimization with Agent Modelling (POAM)

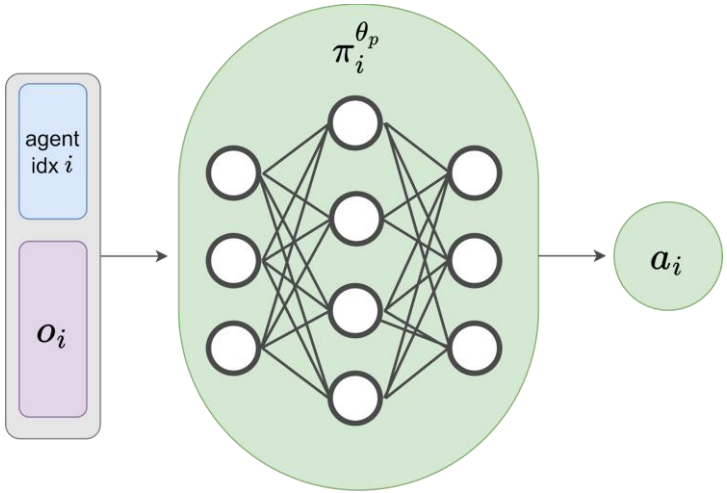
Teammate modelling via recurrent encoder-decoder architecture

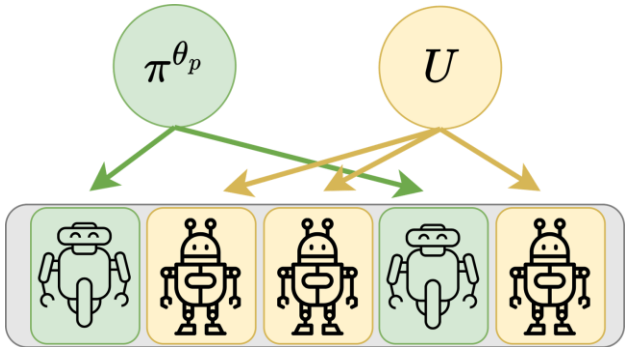


Policy Optimization with Agent Modelling (POAM)

Independent PPO with parameter sharing

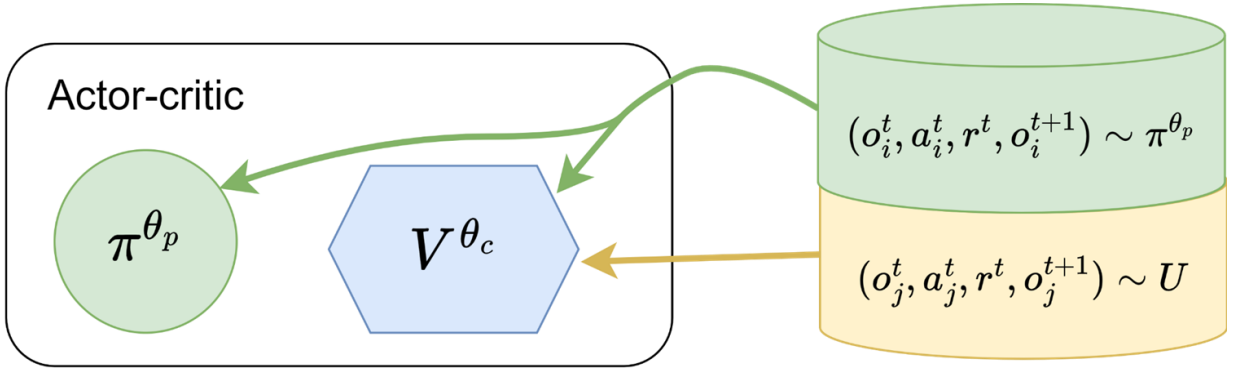
enables dealing with a changing number of teammates during training





Policy Optimization with Agent Modelling (POAM)

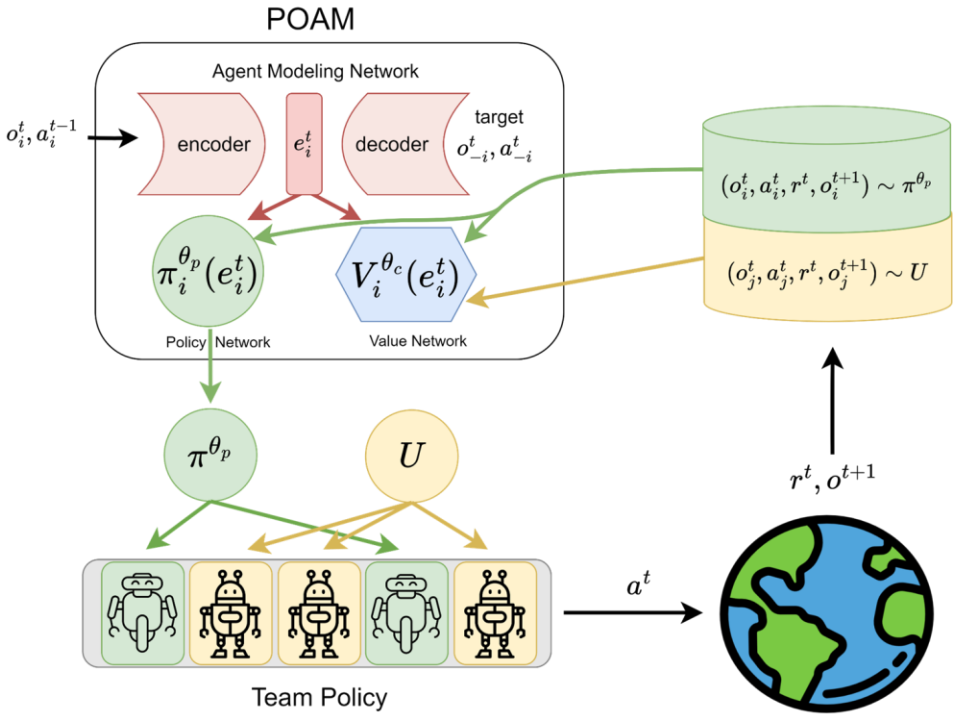
Training critic with data from controlled and uncontrolled agents



Policy Optimization with Agent Modelling (POAM)

Key Ideas

- Independent PPO with parameter sharing
- Training critic with data from controlled and noncontrolled agents
- Teammate modelling



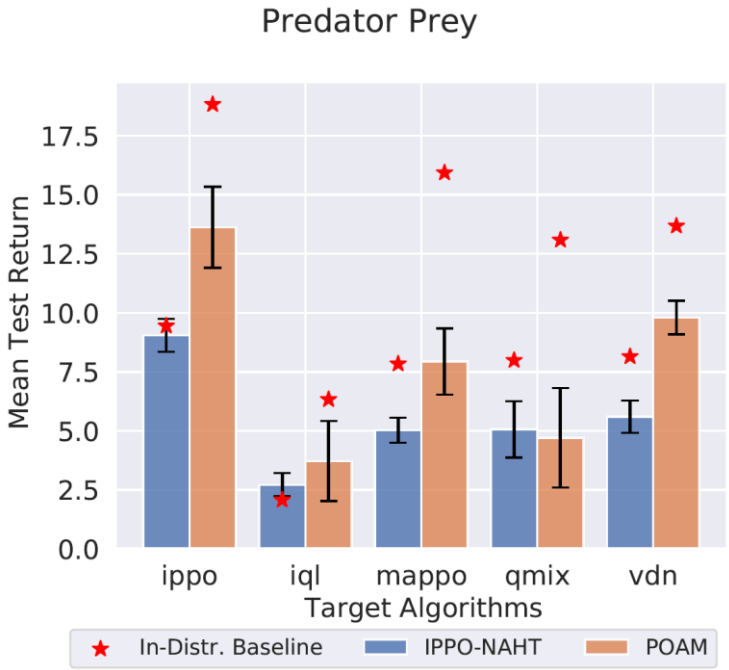
Experiments

- Domains:
  - StarCraft II Multi-Agent Challenge[1] : 5v6, 8v9, 3s5z, 10v11
  - Multi-agent particle environment[2] : Predator-prey task (MPE-PP)

- Uncontrolled teammates: IPPO, QMIX, VDN, IQL, MAPPO
- Baseline: IPPO-NAHT

[1] Samvelyan et al., The StarCraft Multi-Agent Challenge, AAMAS 2019.
[2] Lowe et al., Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments, NeurIPS 2017.

Generalization to Unseen Agents



*Pictured: mean and 95% confidence intervals, computed over 5 trials

Thanks for listening!

Caroline Wang
caroline.l.wang@utexas.edu

Scan to view paper

Collaborators:

Arrasy Rahman

Ishan Durugkar

Elad Liebman

Peter Stone

LARG Learning Agents Research Group
The University of Texas at Austin

Sony AI