Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)
Doctoral Consortium Track

N-Agent Ad Hoc Teamwork

Caroline Wang

Department of Computer Science
The University of Texas at Austin
caroline.l.wang@utexas.edu

Abstract

Current approaches to learning cooperative multi-agent behaviors assume relatively restrictive settings. In fully cooperative multi-agent reinforcement learning, the learning algorithm controls *all* agents in the scenario, while in ad hoc teamwork, the learning algorithm usually assumes control over only a *single* agent in the scenario. However, many cooperative settings in the real world are much less restrictive. For example, in an autonomous driving scenario, a company might train its cars to cooperate with each other, yet once on the road, these cars must additionally cooperate with cars from other companies. Towards expanding the class of scenarios that cooperative learning methods may optimally address, this research agenda introduces and proposes to study N*-agent ad hoc teamwork* (NAHT), where a set of autonomous agents must interact and cooperate with dynamically varying numbers and types of teammates.

1 Introduction and Research Challenge

As the capabilities of artificial intelligence systems advance, artificial agents will inevitably become pervasive in everyday settings. In some settings, it will be necessary for artificial agents to interact and coordinate with each other. As such, there is a growing need to develop methods to enable agents to cooperate with unfamiliar agents, and deal with behaviors that were not anticipated at development time.

In recent years, the research community has examined two frameworks for learning cooperative, multi-agent behaviors. Cooperative multi-agent reinforcement learning (CMARL) is a paradigm for training agent teams to solve a common task via interaction with each other and the environment [Littman, 1994]. On the other hand, ad hoc teamwork (AHT) focuses on formulating policies to enable individual agents to collaborate with previously unknown teammates, towards solving a shared task [Stone *et al.*, 2010].

While impressive examples of cooperative team behaviors have been learned using both paradigms, these advances have naturally been confined to restrictive scenarios in which either complete control over all agents is assumed, or only a single agent is adapted for cooperation [Rashid *et al.*, 2018; Papoudakis *et al.*, 2021]. However, real-world collaborative scenarios might demand agent *subteams* that are able to collaborate with unfamiliar teammates that follow different coordination conventions. Towards expanding the class of scenarios that cooperative learning methods can address to match real-world scenarios more closely, the goal of the proposed research agenda is to answer the following question:

> How can we design autonomous agent teams that are able to efficiently collaborate with known and unknown teammates to maximize return on cooperative tasks?

2 Contributions

To answer this research question, the following contributions are proposed and described below:

1. **Defining the Problem** (published at the Workshop on Ad Hoc Teamwork at AAAI 2024): The research question is formalized as the problem of N**-agent ad hoc teamwork** (NAHT). When there is only a single AHT agent, NAHT is equivalent to AHT. On the other hand, when all AHT agents were produced by some CMARL algorithm and there are no uncontrolled teammates, NAHT is equivalent to CMARL. Thus, the NAHT problem generalizes both CMARL and AHT. The formalism is described below.

Let C denote a set of AHT agents. If the policies of the agents in C are generated by an algorithm, we say that the algorithm controls agents in C. Since our intention is to develop algorithms for generating the policies of agents in C, we refer to agents in C as *controlled*. Let U denote a set of *uncontrolled* agents, which we define as all agents in the environment not included in C.

We model an open system of interaction, in which a random selection of M agents from sets C and U must coordinate to perform task r. For illustration, consider a warehouse staffed by robots developed by companies A and B, where there is a box-lifting task that requires three robots to accomplish. If Company A's robots are AHT agents (corresponding to C), then some robots from Company A could collaborate with robots from Company B (corresponding to U) to accomplish the task, rather than requiring that all three robots come exclusively from A or B. Motivated thus, we introduce a *team sampling procedure* X. At the beginning of each episode, X samples a team of M agents by first sam-

pling $N < M$, then sampling N agents from the set C and $M - N$ agents from U. We define X more formally below.

Let $g(k, \mu, s)$ represent a function for randomly selecting k elements from a generic set μ, conditioned on a state $s \in \mathcal{S}$. For instance, g might be the distribution, *Multinomial*$(k, |\mu|, \boldsymbol{\phi}(s))$, for selecting k elements with replacement from the set μ, parameterized by the state-dependent probability logits $\boldsymbol{\phi}(s) \in [0, 1]^{|\mu|}$. Allowing g to depend on an initial state enriches the representable open interactions. Returning to our robot warehouse example, a state-dependent g might enable us to model that robots suitable for heavy loads are more likely to be present near the loading dock area. For $s \in \mathcal{S}$, X is a sampling procedure parameterized by the tuple $(s, M, U, C, \boldsymbol{\phi}_M, g_U, g_C)$, where $\boldsymbol{\phi}_M \in [0, 1]^M$ represents the logits of a categorical distribution:

(a) Sample an integer N from *Cat*$(\boldsymbol{\phi}_M)$.
(b) Sample N controlled agents via g_U from U.
(c) Sample $M - N$ uncontrolled agents via g_C from C.

Without loss of generality, let $C(\theta) = \{\pi^i_\theta(.|s)\}^N_{i=1}$ denote a set of N controlled agent policies parameterized by θ, such that a learning algorithm might optimize for θ. Let the $\boldsymbol{\pi}^{(M)}$ indicate a *team* of M agents and $\boldsymbol{\pi}^{(M)} \sim X(U, C)$ indicate sampling such a team from U and C via the team sampling procedure. The *objective* of the NAHT problem is to find parameters θ, such that $C(\theta)$ maximizes the expected return in the presence of teammates from U:

$$\max_\theta \left(\mathbb{E}_{\boldsymbol{\pi}^{(M)} \sim X(U, C(\theta))} \left[\sum_{t=0}^{T} \gamma^t r_t \right] \right). \tag{1}$$

The central challenges of the NAHT problem include (1) coordinating with potentially unknown teammates (*generalization*), and (2) coping with a varying number of uncontrolled teammates (*openness*).

2. **An Agent-Modelling Algorithm for NAHT** (under review): Modelling the behaviors and intentions of teammates is a strategy that has proven to be successful for enabling generalization to unknown teammates in the related problem of ad hoc teamwork [Papoudakis *et al.*, 2021]. Simultaneously, independent policy gradient approaches have proven capable of learning sophisticated, coordinated team behaviors for CMARL [Yu *et al.*, 2022]. These algorithmic ideas are combined in a first algorithm for NAHT, called Policy Optimization with Agent Modelling (POAM). POAM is a policy-gradient based approach for learning cooperative multi-agent team behaviors in the presence of varying numbers and types of teammate behaviors. The approach enables adaptation to diverse teammate behaviors by training policies conditioned on learned representations of teammate behaviors. Empirical evaluation on a predator-prey task and various StarCraft II tasks shows that POAM improves cooperative task returns compared to baseline approaches, and enables out-of-distribution generalization to unseen teammates.

3. **A Teammate Generation Algorithm for NAHT** (proposed): As an initial algorithm for NAHT, Contribution 2 examines the easier setting where access to a limited set of teammate policies is assumed for training purposes. To move beyond relying on hand-crafted teammate policies, a critical area to explore is *generating* teammate policies [Papoudakis *et al.*, 2021]. Teammate generation approaches have proven to be successful for AHT, but in practice, have largely been limited to settings with only two agents—for which only a single teammate needs to be generated. An initial approach for this direction would consist of starting from existing teammate generation algorithms, and analyzing their ability to scale to scenarios with more than two agents.

4. **A Benchmark for NAHT** (proposed): As a newly proposed problem, there are no benchmark suites designed for training and evaluating NAHT methods. To facilitate research on NAHT, this contribution aims to provide a unified codebase containing suitable experimental domains, implementations of baseline algorithms, and evaluation protocols.

3 Conclusion

This document has laid out a research agenda around the novel problem of N-agent ad hoc teamwork (NAHT). By pointing out the limitations of current paradigms, developing solution methods, and providing a benchmark, the ultimate aim of this research agenda is to lay the groundwork for integrating autonomous agents into society, in an ecosystem where there are multiple developers producing such agents. If successful, the NAHT research agenda would expand the set of scenarios that current multi-agent learning methods are able to address, representing a significant contribution to the field of multi-agent systems.

References

[Littman, 1994] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine Learning*, pages 157–163. Elsevier, 1994.

[Papoudakis *et al.*, 2021] Georgios Papoudakis, Filippos Christianos, and Stefano V. Albrecht. Agent modelling under partial observability for deep reinforcement learning. In *Advances in Neural Information Processing Systems*, 2021.

[Rashid *et al.*, 2018] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, 2018.

[Stone *et al.*, 2010] Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2010.

[Yu *et al.*, 2022] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of mappo in cooperative multi-agent games. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2022.