Multi-Team Fitness Critics For Robust Teaming

Extended Abstract

Joshua Cook
Oregon State University
Corvallis, Oregon, USA
cookjos@oregonstate.edu

Tristan Scheiner
Rose-Hulman Institute of Technology
Terre Haute, Indiana, USA
scheintc@rose-hulman.edu

Kagan Tumer
Oregon State University
Corvallis, Oregon, USA
kagan.tumer@oregonstate.edu

ABSTRACT

Many multiagent systems, such as search and rescue or underwater exploration, rely on generalizable teamwork abilities to achieve complex tasks. Though many ad-hoc teaming algorithms focus on finding an agent's *best* fit with static team members, domains with high degrees of uncertainty and dynamic teammates require an agent to cooperate with *arbitrary* teams. Prior work views this as an issue of uninformative rewards, providing high-quality but potentially expensive evaluation methods to isolate an agent's contribution. In this work, we provide a local-evaluation-based approach that leverages state trajectories of agents to better identify their impact across multiple teams. The key insight that enables this approach is that agent trajectories and previous experiences carry sufficient information to map agent abilities to team performance. As a result, we are able to train multiple agents to cooperate across arbitrary teams as well as, if not better than, current methods, while only using local information and significantly fewer team evaluations.

KEYWORDS

Multiagent Learning, Evaluation Mechanisms, General Teaming

ACM Reference Format:
Joshua Cook, Tristan Scheiner, and Kagan Tumer. 2023. Multi-Team Fitness Critics For Robust Teaming: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Learning to cooperate with a variety of team members is vital to the success of many multiagent systems, especially those where the structure of the team is uncertain or dynamic. Domains such as robot soccer [1, 6, 9], poker [4, 5], and robust multi-robot systems [2, 8, 10] all require interaction across varying groups of agents. Working with multiple teams compounds the difficulties of credit assignment as agents must not just learn to cooperate with a single team, but adapt to team variety.

This problem is exemplified by autonomous search and rescue, where multiple robots need to form multiple different teams at different times to efficiently search multiple areas. Agents must learn general teamwork from an uninformative feedback signal describing the success of the team as a whole. Recent work frames this as a reward-shaping issue, providing a global and local evaluation method to train agents to cooperate across teams [7]. The local evaluation method provides the highest quality learning signal through the evaluation of counterfactual states. However, this signal requires a significant number of team evaluations, which can be prohibitively expensive or even impossible to calculate in many domains.

In this work, we present Multi-Team Fitness Critics (MTFC) that learn a mapping from local state information to an agent's contribution to multiple teams. This method trains local models of each agent's contribution to the global objective based solely on the states the agents visit. Evaluating an agent in each team using this model, agents receive a learning signal that describes their impact on the performance of multiple teams. Agents that optimize this signal will complete the team objective effectively across a variety of teams.

The contribution of this work is a method of learning local rewards to train agents to cooperate with multiple teams. Experimental results in a simulated multiagent exploration domain show that these learned reward functions train agents to cooperate in teams as effectively or even better than previous methods while requiring fewer team evaluations to train. Further results show that MTFC effectively learns a dense, agent-specific signal representing an agent's contribution to multiple teams.

2 MULTI-TEAM FITNESS CRITICS

We present Multi-Team Fitness Critics (MTFC) to train multiple agents that cooperate with numerous teams through learned local objective functions. We concurrently train critics to map local states to team performance, and the agents using the fitness critics. We define the MTFC function as follows:

$$MTFC_i(Z_1, Z_2, ..., Z_n) = \sum_{j=1}^{n} \max_{\forall z \in Z_j} F_i(z_i) \quad (1)$$

Here, MTFC evaluates joint-state trajectories $(Z_1, Z_2, ..., Z_n)$ for n teams. To evaluate agent i, we use the fitness critic of agent i (F_i) evaluate the agent's local states in each team. The final fitness of the agent is the sum of the highest fitness critic scores in each team. This provides an agent with the sum of their optimistic contribution to a variety of teams.

To train each fitness critic, we train a neural network as a regression model that maps a state to the global objective. As the agents interact with the environment, we store tuples of local states along with the global reward received at the end of the episode. By training the critic on these samples, the model learns to map local information to global performance.

To generate teams, we first create a larger number of learning agents than the number of physical agents in the system. We generate subsets of these agents to form teams, where the size of the team

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

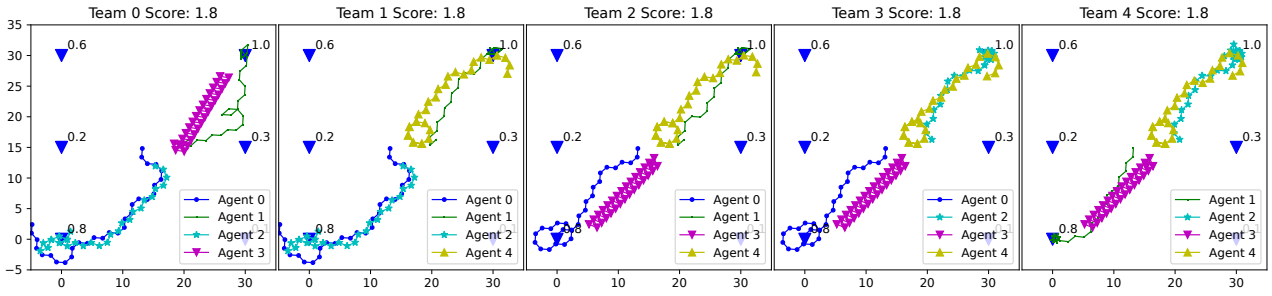

Figure 1: Using MTFC we trained five agents in teams of four. An example run is shown where MTFC was able to train agents to move towards the two highest valued POIs in every team, exhibiting effective teamwork across multiple teams. Agents 0, 2, and 3 demonstrate robust teamwork as they move toward the different POIs needed by each team.

is the number of robots in the system. Each team is one combination of the superset of agents. We iterate through each combination to form the full set of training teams.

3 RESULTS

We evaluate MTFC on the Rover Domain, a two-dimensional exploration domain [3]. Here agents must move around an area and observe Points Of Interest (POI). These POI contain different values and the agents must observe the higher-valued POI to achieve the maximal reward. The difficulty of the task is increased by requiring two agents to view the POI simultaneously.

In our experiments, we trained five agents in teams of four for a total of five teams. Here the agents start in the center and must form two sub-teams at the two highest valued POIs, located in the top right and the bottom left. Figure 1 shows the behaviors of agents trained using MTFC in each of the five teams. Here we can see the agents learn to effectively cooperate across a variety of teams and even change their behavior to meet the needs of the team. For example, in team one, agent two moves to the bottom left POI, but in team three, the agent two moves to the top right POI.

We compare MTFC to the previous state-of-the-art learning signal for robust teaming, D^{Σ} [7]. This signal uses counterfactual state evaluations to form an agent's approximate contribution to multiple teams. We present the findings of this comparison in figure 2. These results represent 12 statistical trials with the shaded regions indicating standard error.

In figure 2, agents trained using MTFC outperforms those trained from D^{Σ}. Unlike D^{Σ}, which uses counterfactual global evaluations of the entire joint-state, MTFC only uses local state information to reward agents.

One attribute of MTFC that allows it to perform this well despite using local information is that MTFC learns a dense approximate of the sparse global objective. In this domain, agents only receive a reward if multiple agents are within the observation radius of a POI. However, the critic learns that states closer to the POIs lead to higher rewards. As a result, the fitness critic is able to reward agents that become close to a POI, providing a gradient-like learning signal. Another factor that contributes to MTFC's success is that it

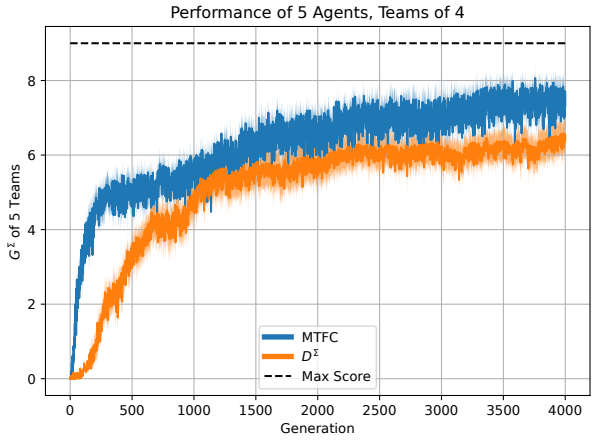

Figure 2: We compare agents trained using MTFC, to another shaped reward, D^{Σ}, across five teams. Despite using local information, MTFC trains the best teaming agents.

evaluates only local states. As a result, this function is less sensitive to the actions of the other agents in the system, alleviating a large portion of the credit assignment problem.

Overall, MTFC is able to more effectively train agents to cooperate across a variety of teams. Unlike D^{Σ}, MTFC does not require access to the global evaluation function or additional team evaluation, instead leveraging local information carried by the local states.

ACKNOWLEDGMENTS

This work was partially supported by the Air Force Office of Scientific Research (grant FA9550-19-1 0195) and the National Science Foundation (grants IIS-1815886 and IS-2112633).

REFERENCES

[1] Arvin Agah and Kazuo Tanie. 1997. Robots Playing to Win: Evolutionary Soccer Strategies. In *Proceedings of International Conference on Robotics and Automation*, Vol. 1. IEEE, 632–637.

[2] A. Agogino, C. Holmes Parker, and K. Tumer. 2012. Evolving Large Scale UAV Communication Systems. In *Proceedings of the Genetic and Evolutionary Computation Conference*. Philadelphia, PA.
[3] Adrian K Agogino and Kagan Tumer. 2008. Analyzing and Visualizing Multiagent Rewards in Dynamic and Stochastic Domains. *Autonomous Agents and Multi-Agent Systems* 17, 2 (2008), 320–338.
[4] Nolan Bard, Michael Johanson, Neil Burch, and Michael Bowling. 2013. Online Implicit Agent Modelling. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 255–262.
[5] Darse Billings, Denis Papp, Jonathan Schaeffer, and Duane Szafron. 1998. Opponent Modeling In Poker. *Aaai/iaai* 493 (1998), 499.
[6] André LV Coelho and Daniel Weingaertner. 2001. Evolving Coordination Strategies in Simulated Robot Soccer. In *Proceedings of the fifth international conference on Autonomous agents*. 147–148.
[7] Joshua Cook and Kagan Tumer. 2022. Fitness Shaping For Multiple Teams. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 332–340.
[8] Cai Luo, Andre Possani Espinosa, Danu Pranantha, and Alessandro De Gloria. 2011. Multi-robot Search and Rescue Ream. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*. IEEE, 296–301.
[9] Esben H Østergaard, Henrik H Lund, and Reality Gap. 2002. Co-Evolving Robot Soccer Behavior. In *From Animals to Animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, Vol. 7. MIT Press, 351.
[10] Charles E Pippin, Henrik Christensen, and Lora Weiss. 2013. Dynamic, Cooperative Multi-robot Patrolling With a Team of UAVs. In *Unmanned Systems Technology XV*, Vol. 8741. International Society for Optics and Photonics, 874103.