Indirect Credit Without A Heuristic For Coevolving Agents

Everardo Gonzalez
gonzaeve@oregonstate.edu
Oregon State University
Corvallis, Oregon, USA

Raghav Thakar
thakarr@oregonstate.edu
Oregon State University
Corvallis, Oregon, USA

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

Abstract

Cooperative Coevolutionary Algorithms have discovered coordinated behaviors in a variety of multiagent systems. Even though the team's performance is captured in team fitness, efficient learning requires shaped, local fitnesses for each agent. While traditional fitness shaping approaches capture each agent's direct contribution to team fitness, they are unable to capture agents' indirect contributions, such as supporting other teammates' actions. Recent work has incorporated a domain-specific heuristic to measure the influence each agent has on its teammates, and provide credit for those teammates' actions to the influencing agent. However, such a heuristic may be non-trivial to define for every domain, and may not capture all influential interactions between agents. We propose randomly choosing which teammates an agent receives credit for, and combining that with a mixed elites selection mechanism to retain high performing teams. This requires no domain expertise, and shows performance gains over traditional shaping techniques. Lastly, we propose bootstrapping as future work to gradually improve this random credit assignment through training to bias the coevolutionary search towards beneficial influence sets.

CCS Concepts

• **Computing methodologies** → **Cooperation and coordination**; **Multi-agent systems**; *Intelligent agents.*

Keywords

Multiagent Learning, Multiagent Credit Assignment, Cooperative Coevolutionary Algorithms

1 Introduction

Cooperative coevolution in multiagent systems has been applied to many real world problems including underwater monitoring, satellite management, and search and rescue [3]. A key challenge in these applications is deriving a fitness value for each agent from a fitness that evaluates the entire team.

One aspect involves agents that *influence* one another. Here, agent-specific fitnesses must meld two key effects: 1) the agent's direct contribution to team performance, and 2) its indirect impacts on other agents. An example of an indirect impact is a high-altitude agent whose flying patterns guide ground-level agents towards desirable states. Fitness shaping for such flying agents requires both prior knowledge that influential behaviors are beneficial, and domain expertise to measure them.

Recent work has used a domain-specific heuristic to measure the influence agents have on their teammates, and used this measurement to shape each agent's fitness. However, such heuristics are difficult to produce for every possible environment, and imperfect heuristics may not accurately capture all possible influential interactions between agents. Thus, it is crucial to develop fitness shaping techniques that can attribute indirect contributions to agents without relying on domain-specific knowledge.

In this paper, we propose deriving agent-specific local fitness values by attributing an agent's indirect impacts on other agents randomly, making it possible to incorporate this information into the coevolutionary search without any prior domain expertise. We compare this approach to other heuristic-free shaping techniques, and find that our approach shows modest performance gains and compelling potential in influence-based settings. Lastly, we outline a future extension that would enable agents to bootstrap an estimate of the influence they have on their teammates, and learn over time which teammates should be included in their estimate.

2 Background

Existing methods partially address the problem of how to incorporate an agent's indirect impacts into its shaped fitness. The Indirect Difference Evaluation in particular uses a domain-specific influence heuristic to determine which teammates an agent influenced, and attributes the direct contributions of those teammates to that agent [3]. The set of teammates influenced by an agent is its influence set. The Indirect Difference Evaluation compares system performance with and without an agent's direct and indirect contributions:

$$D_i^{IND} = G(T) - G(T_{-F_i}) \tag{1}$$

The original (direct) Difference Evaluation D_i uses the same structure, but there are not multiple agents in the influence set F_i. Agent i only includes itself [1, 2]. Thus on the one hand, the Indirect Difference Evaluation captures indirect contributions of agents, but it requires a domain-specific heuristic to do so. On the other hand, the original Difference Evaluation does not require a heuristic, but only captures an agent's direct contribution.

3 Random Influence Sets

The influence set F_i determines which teammates' actions agent i gets credit for. Prior work builds this set using a distance-based heuristic to determine which teammates an agent influenced [3]. The use of a heuristic to determine influence limits the application of indirect credit to domains where the nature of agent interactions are known apriori. This heuristic could be misleading if agents discover unexpected interactions that are beneficial to team performance.

Instead of a heuristic, we use random chance to determine which teammates are included in an influence set. An agent always includes itself. For each teammate, we generate a random number from 0 to 1. If it is greater than an ϵ threshold, that teammate is included in F_i. Otherwise, that teammate is excluded. Even though this builds a random subset, the shaped fitness is still aligned with the global team fitness G because agent i always includes itself [2].

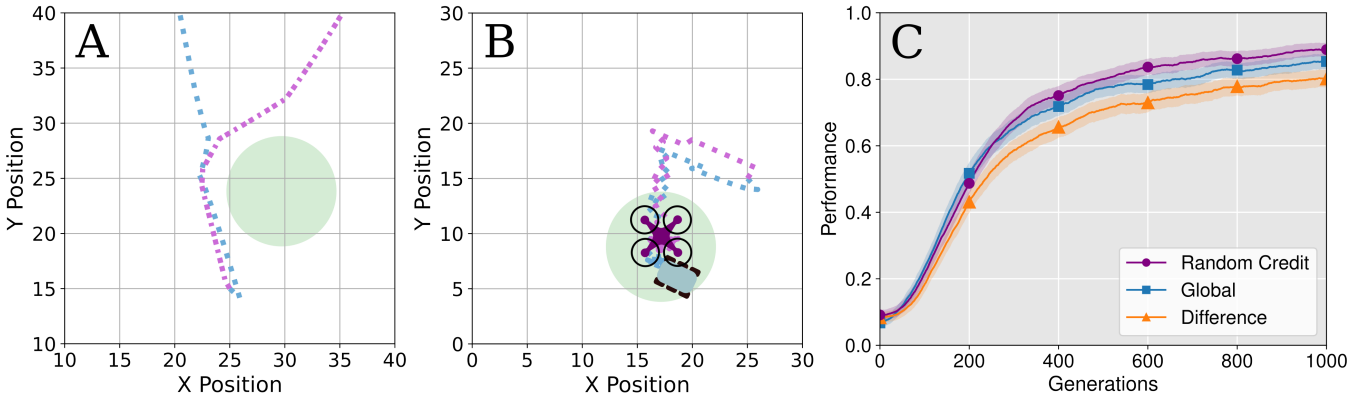

Figure 1: A drone (purple) must guide a rover (blue) to a POI (green, shading represents capture radius). Dotted lines indicate the paths taken by the drone and rover. The team explores various joint behaviors with random credit, including a separation behavior (A), and aggressive turning (B). Random credit offers improved joint policy exploration, leading to modestly better performance than deterministic baselines (C).

These random influence sets make it possible for agents to receive indirect credit for their teammates' actions without a heuristic. However, the randomness in set-building means an agent might incorrectly receive credit for a teammate it had no interactions with, or vice versa. This approach requires a mechanism to protect high performing joint policies from being lost in exploration due to incorrectly assigned credit. This is why we combine random influence sets with a mixed elites selection strategy during coevolution. We select both elite individuals (based on random credit), and elite teams (based on that team's performance) to move onto the next generation during training.

4 Preliminary Results

We test random credit in a 2D POI (point of interest) capture problem [3]. A rover-drone team must capture a randomly spawning POI. The challenge is that only the rover can capture the POI, but only the drone can sense the POI. The rover must rely on the influence of the drone to guide it to the POI; otherwise the rover wanders blindly through the environment.

The rover and drone start at (25, 15), and a POI spawns randomly around them. The team fitness G is 0 unless the rover captures the POI. If the rover captures the POI, G is 1. We compare the global team evaluation G, difference evaluation D, and random credit (ϵ=0.5). Random credit discovers a separation behavior that might be useful if the drone must accomplish other tasks, shown in Figure 1(A). Another behavior is a sharp turn that might be useful if the team must navigate around obstacles, shown in Figure 1(B). In this experiment, the team does not encounter either of these situations, so it is important to guide this random exploration towards beneficial team behaviors. Mixed elites selection ensures we are selecting strong joint policies based on G and individual policies based on random credit for influence-based exploration. Each shaping method uses a mixed elites selection strategy, but the exploratory noise in random credit best leverages this selection mechanism. The combination of random credit with mixed elites yields a moderate performance improvement against baseline heuristic-free methods. Figure 1(C) shows the mean and standard error in training performance across 200 stat runs.

5 Future Work: Bootstrapping Influence Sets

Random credit struggles scaling up to many agents. Each additional agent doubles the possible combinations of teammates that could be included in an influence set. To address this, we can bootstrap influence sets randomly, and learn a distribution of which teammates belong in each agent's influence set. We can initialize an ϵ_j for each teammate j. As we train, we update ϵ_j based on how beneficial including or excluding this teammate was. This could be tracked throughout generations so it is updated on a slower timescale than individual policies, meaning we have time to collect information on which subsets work well together and which do not.

References

[1] Adrian Agogino and Kagan Tumer. 2004. Efficient Evaluation Functions for Multi-rover Systems. In *Genetic and Evolutionary Computation – GECCO 2004*, Kalyanmoy Deb (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–11.

[2] Adrian Agogino and Kagan Tumer. 2008. Analyzing and visualizing multiagent rewards in dynamic and stochastic domains. *Autonomous Agents and Multi-Agent Systems* 17 (10 2008), 320–338. https://doi.org/10.1007/s10458-008-9046-9

[3] Everardo Gonzalez, Siddarth Viswanathan, and Kagan Tumer. 2024. Influence Based Fitness Shaping for Coevolutionary Agents. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Melbourne, VIC, Australia) *(GECCO '24)*. Association for Computing Machinery, New York, NY, USA, 322–330. https://doi.org/10.1145/3638529.3654175