Heterogeneous Agent Coordination via Adaptive Quality Diversity and Specialization

Gaurav Dixit
Collaborative Robotics and Intelligent Systems Institute
Oregon State University
Corvallis, Oregon, USA
dixitg@oregonstate.edu

Charles Koll
Oregon State University
Corvallis, Oregon, USA
kollch@oregonstate.edu

Kagan Tumer
Collaborative Robotics and Intelligent Systems Institute
Oregon State University
Corvallis, Oregon, USA
kagan.tumer@oregonstate.edu

ABSTRACT

In many real-world multiagent systems, agents must learn diverse tasks and coordinate with other agents. This paper introduces a method to allow heterogeneous agents to specialize and only learn complementary divergent behaviors needed for coordination in a shared environment. We use a hierarchical decomposition of diversity search and fitness optimization to allow agents to speciate and learn diverse temporally extended actions. Within an agent population, diversity in niches is favored. Agents within a niche compete for optimizing the higher level coordination task. Experimental results in a multiagent rover exploration task demonstrate the diversity of acquired agent behavior that promotes coordination.

CCS CONCEPTS

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

KEYWORDS

Heterogeneous Multiagent Coordination, Quality Diversity

ACM Reference Format:
Gaurav Dixit, Charles Koll, and Kagan Tumer. 2021. Heterogeneous Agent Coordination via Adaptive Quality Diversity and Specialization. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459564

1 INTRODUCTION

An important feature of agents in the multiagent setting is the ability to reason about the behavior of other agents and perform tasks at different time scales with a variety of potential partners. This presents a need to learn and plan using skills that abstract temporally extended actions. In the presence of diverging goals, agents must specialize their skills in response to the capabilities of other agents to successfully solve the task as a team.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07…$15.00
https://doi.org/10.1145/3449726.3459564

Quality Diversity (QD) algorithms learn a repertoire of behaviors that favor functional diversity over an objective measure of performance [3, 4]. In cooperative multiagent settings, exploring the entire behavior space is often intractable due to the combinatorial nature of multiagent interaction. In addition, in problems involving diverse goals, agents cannot be expected to be able to perform the entire spectrum of behaviors. Furthermore, hardware and morphology restrictions both encourage specialization of behaviors and require cooperation in complex problems [2].

In this work we introduce an iterative hierarchical learning mechanism that decouples the search for behavioral diversity and the pressure to specialize and coordinate. At the top level, agents operate in a shared environment using skills, temporally extended actions, to achieve a shared goal. At the bottom level, a population of neural networks is evolved to favor high performing and diverse skills using a QD method. This hierarchical combination of diversity and fitness-favoring methods drives the search for diversity that allows for agent specialization, which might be necessary to coordinate and complete diverse tasks. We demonstrate the strength of our method on a multiagent space exploration problem with sparse feedback and a diverse set of goals that require tight agent coupling between heterogeneous agents. Our hierarchical approach shows significant performance improvement over MAP-Elites [4], a state-of-the-art Quality Diversity method.

2 MULTIAGENT ADAPTIVE QUALITY DIVERSITY (MAQD)

Multiagent Adaptive Quality Diversity (MAQD) is an iterative process with three primary operations: niche organization, evaluation and refinement.

(1) Organization: For each sub-goal a skill is learned which is a population of behaviors evolved to both maximize fitness and encourage diversity. Trained behaviors are projected in the behavior space and are clustered into niches. Every behavioral niche has a membership limit that encourages local competition amongst the behaviors within the same niche. Outlier behaviors are allowed to form new niches.
(2) Evaluation: A population of teams of top-level learners is created. Each top-level learner is randomly assigned behaviors from each skill. A top-level learner keeps track of its assigned behaviors as part of its genetic information. This helps top-level learners associate with behaviors across generations after both of them evolve independently. Teams of top-level learners are evolved and evaluated based on the

cumulative team fitness. A random selection of the best performing teams are used for crossover and mutation to create the next generation of teams.

(3) Refinement: The behaviors corresponding to the successful top-level learners are retained. A crossover and mutation operation is applied to the retained behaviors and new behaviors are projected in the behavior space. Niche membership is regulated by imposing a maximum population limit to encourage intra-niche competition. Initially, behavior refinement is performed after every top-level evolutionary step. Over the course of learning, increased niche stability allows completion of several evolutionary steps before reapplying behavior refinement.

3 EXPERIMENT

We use a variant of the rover exploration problem that requires a team of rovers to simultaneously observe multiple point of interests (POIs). Instead of using symbolic POI types [1], we define four types of POIs that require agents to specialize their capabilities: 1) A Rewarding (R) POI that rewards agents with a constant value when observed, 2) a Time-Sensitive (TS) POI - An R POI with a decaying reward, 3) a Low-Energy (LE) POI - An R POI that is observable stochastically, and 4) A Fuel (F) POI that rewards an agent constant fuel units. The average speed and observation radius used by the rover is used as the behavior characterization for niche placement. Agents in a team are assigned 10 fuel units each which can be replenished by visiting an F POI.

Figure 1 shows the cumulative team fitness. In our method, the diversity of the environment puts a selection pressure on finding behavior niches that can work together to satisfy all POI constraints. After several iterations of behavior refinement, the team learns to observe more than 90% of the POIs. Figure 2 shows the discovered niches for the coordination task using all four POI types. The relative sizes of the niches show the subtle balance between fitness of a niche and its utility in the environment. For instance, the niche with the highest speeds and average observation radius has the lowest population of the team. Nevertheless, it is providing sufficient utility in observing LE-POIs and TS-POIs, so it is critical to good team performance.

4 DISCUSSION

We introduced a hierarchical iterative method for diversity coverage in multiagent settings that require tight coordination. Our method decouples the search for behavioral diversity and fitness maximization which allows us to adaptively discover niches in response to higher-level agent goals. MAQD uses a user-defined behavior characterization for the projection and classification of behaviors in the behavior space. The characterization of behaviors is crucial since it drives the exploration and eventual diversity. This requires the user to have some prior domain knowledge. Moreover, in many complex multiagent settings it is difficult to specify subgoals and characterization metrics to measure the fitness required to achieve them. In future work, we will explore behavior space discovery to allow behavior characterization in cooperative tasks without the need for prior knowledge.

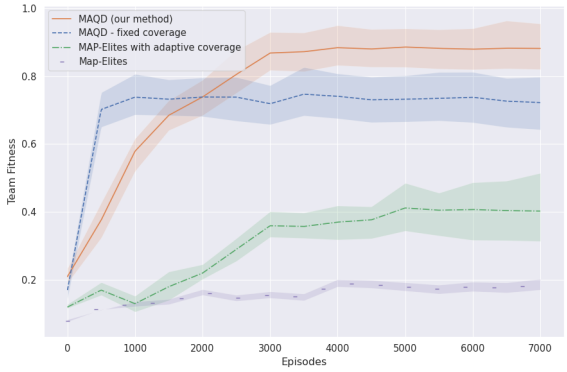

Figure 1: Fitness with uniformly distributed POIs. Our method promotes specialization of capabilities enabling agents to cooperate on diverse POIs. MAP-Elites struggles due to the need to learn the entire behavior repertoire.

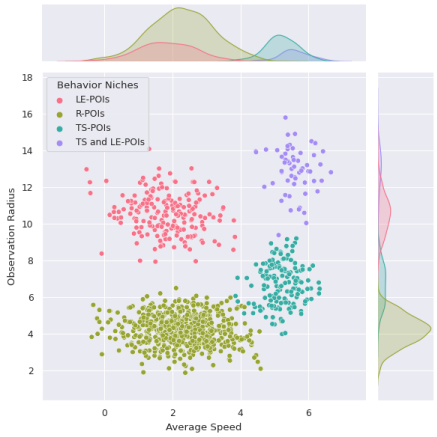

Figure 2: Discovered behavior niches in an environment with uniformly distributed R, TS and LE POIs. Niches with higher observation radii and lower speeds are optimal for LE POIs, whereas niches with higher average speeds are better suited for TS POIs.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under grant No. IIS-1815886 and the Air Force Office of Scientific Research under grant No. FA9550-19-1-0195.

REFERENCES

[1] Gaurav Dixit, Nicholas Zerbel, and Kagan Tumer. 2019. Dirichlet-Multinomial Counterfactual Rewards for Heterogeneous Multiagent Systems. In *2019 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. IEEE, 209–215.

[2] Atil Iscen, Adrian Agogino, Vytas SunSpiral, and Kagan Tumer. 2014. Flop and roll: Learning robust goal-directed locomotion for a tensegrity robot. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2236–2243.

[3] Joel Lehman and Kenneth O Stanley. 2011. Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*. 211–218.

[4] Jean-Baptiste Mouret and Jeff Clune. 2015. Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909* (2015).