Learning Diverse Multiagent Behaviors

Doctoral Consortium

Ayhan Alp Aydeniz
Collaborative Robotics and Intelligent Systems Institute
Oregon State University
Corvallis, Oregon, USA
aydeniza@oregonstate.edu

ABSTRACT

Deploying teams of agents for many coordination tasks such as search and rescue missions or deep ocean exploration promises effective solutions. However, these tasks generally require a team of agents to take highly coordinated joint actions, which are difficult to unearth under sparse rewards (returning the same value or zero for many joint actions). The previous works have shown that having a large coverage over behavior space via learning diverse behaviors is an effective method to address reward sparsity. Because multiagent behavior spaces are of higher-dimensions than single-agent spaces, multiagent behavior generation is often intractable. We introduce entropy seeking agents to learn diverse behaviors for multiagent systems to address reward sparsity. Our results show that the promotion of diversity among behaviors within the behavior space effectively, resulting in the discovery of collaborative behaviors.

KEYWORDS

Multiagent Learning, Exploration, Entropy, Reward Shaping

ACM Reference Format:
Ayhan Alp Aydeniz. 2025. Learning Diverse Multiagent Behaviors: Doctoral Consortium. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Applications of multiagent systems in various environments are rapidly evolving [2, 19, 21, 22, 26]. Especially, these systems offer several opportunities in remote exploration missions, such as space exploration [12, 25], and deep ocean exploration [24, 27]. However, determining what these teams should learn, before their deployment in remote missions, is crucial for success. For example, a behavior achieving an objective in a Martian mission (*e.g.*, discovery of river traces on Mars) is difficult to design, because of the lack of knowledge about the environment. Especially considering that tasks often require close coordination among multiple agents, the complexity of behaviors in such problems increases significantly.

In a typical multiagent RL (MARL) framework, agents are trained individually, each with its own reward function that provides feedback after every action. However, their collective goal is to maximize a shared team reward. This dual focus can lead to conflicting objectives, complicating the learning of effective team behaviors. Consider the scenario where two rovers on Mars must coordinate to collect samples from the same rock to achieve a team objective; if each rover collects samples from different rocks to maximize its individual reward, they fail in the team objective. This highlights a crucial challenge where agents need to align their individual actions towards a common objective, particularly when the desired team behavior requires a series of joint actions rewarded sparsely (the reward remains zero for most joint actions). The complexity of aligning these objectives underlines the brittleness of the performance in multiagent tasks, emphasizing the necessity for strategies that effectively integrate individual learning with team goals.

In our work, we propose both evolutionary and purely gradient-based learning frameworks where we introduce entropy seeking agents. We enable agents to explore at agent level, while we exploit the true global objective at team-level. Experiments show that our agents discover highly collaborative behaviors in difficult tasks.

2 BACKGROUND

To learn behaviors achieving close coordination among agents, various approaches have been explored [23]. In multiagent learning, there have been two main trends of algorithms to address this challenge: evolutionary algorithms (EAs), reinforcement learning (RL) algorithms [1, 17, 18, 28]. While these algorithms offer promising solutions, they both have a distinct way of incorporating *feedback mechanisms*. EAs generate a population of behaviors, and execute them in an environment and evaluate them using the accumulated feedback which represents their fitness within the population. On the other hand, RL algorithms, by modeling problems using Markov decision processes (MDPs), leverage immediate feedback—or *reward*—from the environment to refine the behaviors based on the outcomes of specific actions in given states. Ergo, a behavior is modified to obtain a successful mapping of the action given a state.

In multiagent settings, a reward function outputting an agent-specific value after every action often leads the algorithms to narrow areas during the policy search. A fitness value representing the whole team's performance carries noisy information because it includes the effects of all agents' actions. Hence, agents fail to modify their individual behaviors to contribute to the team fitness.

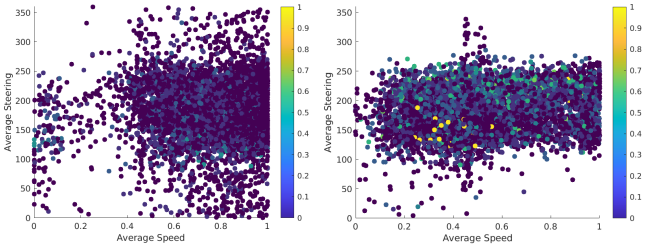
Multiagent evolutionary RL (MERL) [14] addresses the structural problems of both EAs and RL algorithms by implementing a population employing mechanisms used in both families of algorithms. Within MERL, the EA component maintains a population of team behaviors, while the RL module focuses on learning local, agent-specific behaviors. This integrated structure facilitates a hierarchical learning where the EA learns the team behavior and the RL

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

component learns individual behaviors. While MERL has demonstrated effectiveness across various multiagent domains [13, 14], it still suffers from the misalignment between the team and individual objectives. This misalignment, primarily stemming from the difficulty of designing task-specific rewards for individual agents, significantly hinders effective exploration. Ergo, its performance diminishes in complex tasks and, when the team reward is sparse, it tends to get restricted to a narrow portion of the policy space [3].

3 ENTROPY SEEKING AGENTS

In our work, we show that exploring observation space is a powerful technique that enables agents to solve complex multiagent tasks and results in learning diverse behaviors without explicitly searching for diverse behaviors. However, the strategical exploration of state spaces in multiagent systems is of considerable challenges. In our first contribution, we propose Novelty Seeking Multiagent Reinforcement Learning (NS-MERL) [3, 8] where we introduce observation entropy maximizing rewards that enable agents to efficiently explore their observation space. This allows us to balance exploration at agent-level and exploitation at team-level. Agent-level exploration leads to the emergence of team-level cooperative behaviors, even in scenarios where the agents are required to take highly coordinated joint actions. Our experiments show that in rover domain [1], our entropy maximizing rewards result in a much better and efficient coverage of the behavior space (Figure 1).



(a) Exploitative Agents (b) Entropy Seeking Agents

Figure 1: Behavior spaces covered during EA's training. Each behavior is represented by the average speed and steering by a single agent throughout an episode. Covered space by exploitative (task-oriented) agents is on the *Left*, by entropy seeking agents is on the *Right*. The color gradient represents the team fitness of a behavior's episode [3]

An important challenge in the implementation of entropy maximizing rewards is the high dimensionality of the state space we seek to explore. Exploring a joint state space can enable us to discover joint states where agents accomplish collaborative tasks, but typically, increasing number of agents results in a growth in the state space; hence, we need a technique that can handle both the growing dimensions of the joint state space. Therefore, exploring agent-level observation space offers significantly greater feasibility. However, in high dimensional state spaces, regardless of values being discrete or continuous, state vectors can be influenced by minimal actions; hence, each state will likely occur once. Our second work, state entropy maximization for multiagent learning [7], tackles the problem of achieving efficient coverage in high dimensional multiagent state spaces. We build this on NS-MERL. Though NS-MERL discovers is effective in complex tasks, it does not scale to state spaces with high dimensions.

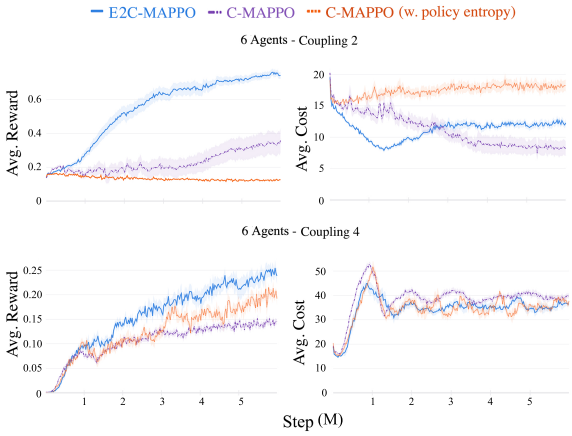


Figure 2: Impact of our observation entropy maximizing reward against employing no entropy (C-MAPPO), and policy entropy (C-MAPPO (w. policy entropy)) [10].

Nevertheless, agents trained to maximize their observation space coverage can result in violations of safety requirements in real-world applications. For instance, behaviors aimed at extensive state space coverage might overlook the risk of collisions with teammates unless it is integrated into the task objective. Recent studies have demonstrated that engineering behaviors through constraints is more effective than optimizing them via reward shaping [10, 15, 20]. However, defining such constraints introduces a new objective to optimize. This often leads to three conflicting objectives in multiagent systems. Our third contribution [4–6] explores the challenge of harmonizing constraints with task objectives in constrained multiagent systems. Our observation entropy seeking agents enable MARL algorithms like MAPPO [29] to discover both high-performing and safe behaviors (Figure 2) compared to the common practices of policy entropy maximization and exploitative agents. The constraints are defined under thresholds and on the top figure has a lower coordination requirement, but due to lower threshold the other methods fail, while E2C agents performs the best. Under higher coordination and relaxed threshold, E2C achieves higher reward, while matching the same cost with the others.

4 PROPOSED RESEARCH

While learning strategies like efficient exploration of the state space of agents, neural networks are capable of maximizing this coverage as an objective; however, they fail to capture the sequential learning of visiting diverse set of states. Recurrent neural networks (RNNs) are able to capture such sequences. Especially, in partially observable settings [16], RNNs have been shown to be effective [9, 11]. In our proposed direction, we will propose episodic multiagent exploration, where we train multiagent systems to effectively cover both joint state space and observation space of the individual agents within the framework of MERL that will allow us to maintain a search in the collaborative portion of the joint state space.

ACKNOWLEDGMENTS

I would like to thank my mentor and advisor, Kagan Tumer; my friends and collaborators, Enrico Marchesini and Robert Loftin; and my labmates at the AADI Lab.

REFERENCES

[1] Adrian K Agogino and Kagan Tumer. 2004. Unifying temporal and structural credit assignment problems. In *Autonomous Agents and Multi-Agent Systems Conference*.

[2] Ebtehal Turki Alotaibi, Shahad Saleh Alqefari, and Anis Koubaa. 2019. Lsar: Multi-uav collaboration for search and rescue missions. *IEEE Access* 7 (2019), 55817–55832.

[3] Ayhan Alp Aydeniz, Robert Loftin, and Kagan Tumer. 2023. Novelty seeking multiagent evolutionary reinforcement learning. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 402–410.

[4] Ayhan Alp Aydeniz, Enrico Marchesini, Christopher Amato, and Kagan Tumer. 2024. Entropy Seeking Multiagent Reinforcement Learning. In *Proceedings of the 2024 International Conference on Autonomous Agents and Multiagent Systems*.

[5] Ayhan Alp Aydeniz, Enrico Marchesini, Robert Loftin, Christopher Amato, and Kagan Tumer. 2024. Safe Multiagent Coordination via Entropic Exploration. *arXiv preprint arXiv:2412.20361* (2024).

[6] Ayhan Alp Aydeniz, Enrico Marchesini, Robert Loftin, Christopher Amato, and Kagan Tumer. 2025. Safe Entropic Agents under Team Constraints. In *Proceedings of the 2025 International Conference on Autonomous Agents and Multiagent Systems*.

[7] Ayhan Alp Aydeniz, Enrico Marchesini, Robert Loftin, and Kagan Tumer. 2023. Entropy Maximization in High Dimensional Multiagent State Spaces. In *2023 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. IEEE, 92–99.

[8] Ayhan Alp Aydeniz, Anna Nickelson, and Kagan Tumer. 2022. Entropy-based local fitnesses for evolutionary multiagent systems. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. 212–215.

[9] Adrià Puigdomènech Badia, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman, Martín Arjovsky, Alexander Pritzel, Andew Bolt, et al. 2020. Never give up: Learning directed exploration strategies. *arXiv preprint arXiv:2002.06038* (2020).

[10] Shangding Gu, Jakub Grudzien Kuba, Munning Wen, Ruiqing Chen, Ziyan Wang, Zheng Tian, Jun Wang, Alois Knoll, and Yaodong Yang. 2021. Multi-Agent Constrained Policy Optimisation. In *arXiv*, Vol. abs/2110.02793.

[11] Steven Kapturowski, Georg Ostrovski, John Quan, Remi Munos, and Will Dabney. 2018. Recurrent experience replay in distributed reinforcement learning. In *International conference on learning representations*.

[12] T Kubota, H Katoh, T Toyokawa, and I Nakatani. 2004. Multi-legged robot system for deep space exploration. In *Proceedings World Automation Congress, 2004.*, Vol. 15. IEEE, 203–208.

[13] Pengyi Li, Jianye Hao, Hongyao Tang, Yan Zheng, and Xian Fu. 2023. Race: improve multi-agent reinforcement learning with representation asymmetry and collaborative evolution. In *International Conference on Machine Learning*. PMLR, 19490–19503.

[14] Somdeb Majumdar, Shauharda Khadka, Santiago Miret, Stephen McAleer, and Kagan Tumer. 2020. Evolutionary reinforcement learning for sample-efficient multiagent coordination. In *International Conference on Machine Learning*. PMLR, 6651–6660.

[15] Enrico Marchesini, Luca Marzari, Alessandro Farinelli, and Christopher Amato. 2023. Safe Deep Reinforcement Learning by Verifying Task-Level Properties. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 1466–1475.

[16] Frans A Oliehoek and Christopher Amato. 2016. *A concise introduction to decentralized POMDPs*. Springer.

[17] Aida Rahmattalabi, Jen Jen Chung, Mitchell Colby, and Kagan Tumer. 2016. D++: Structural credit assignment in tightly coupled multiagent domains. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4424–4429.

[18] Tabish Rashid, Gregory Farquhar, Bei Peng, and Shimon Whiteson. 2020. Weighted qmix: Expanding monotonic value function factorisation for deep multi-agent reinforcement learning. *Advances in neural information processing systems* 33 (2020), 10199–10210.

[19] Luís Paulo Reis, Fernando Almeida, Luís Mota, and Nuno Lau. 2013. Coordination in multi-robot systems: Applications in robotic soccer. In *Agents and Artificial Intelligence: 4th International Conference, ICAART 2012, Vilamoura, Portugal, February 6-8, 2012. Revised Selected Papers 4*. Springer, 3–21.

[20] Julien Roy, Roger Girgis, Joshua Romoff, Pierre-Luc Bacon, and Chris J Pal. 2022. Direct Behavior Specification via Constrained Reinforcement Learning. In *ICML*, Vol. 162. 18828–18843.

[21] Jürgen Scherer and Bernhard Rinner. 2020. Multi-robot persistent surveillance with connectivity constraints. *IEEE Access* 8 (2020), 15093–15109.

[22] Esmaeil Seraj, Andrew Silva, and Matthew Gombolay. 2022. Multi-UAV planning for cooperative wildfire coverage and tracking with quality-of-service guarantees. *Autonomous Agents and Multi-Agent Systems* 36, 2 (2022), 39.

[23] Peng Shi and Bing Yan. 2020. A survey on intelligent control for multiagent systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 51, 1 (2020), 161–175.

[24] Louis Whitcomb, Dana R Yoerger, Hanumant Singh, and Jonathan Howland. 2000. Advances in underwater robot vehicles for deep ocean exploration: Navigation, control, and survey operations. In *Robotics Research*. Springer, 439–448.

[25] E Jay Wyatt, Konstantin Belov, Julie Castillo-Rogez, Steve Chien, Abigail Fraeman, Jay Gao, Sebastian Herzig, TJW Lazio, M Troesch, and T Vaquero. 2018. Autonomous networking for robotic deep space exploration. In *International Symposium on Artificial Intelligence, Robotics, and Automation for Space (ISAIRAS 2018)*.

[26] Logan Yliniemi, Adrian K Agogino, and Kagan Tumer. 2014. Multirobot coordination for space exploration. *AI Magazine* 35, 4 (2014), 61–74.

[27] Dana R Yoerger, Albert M Bradley, Barrie B Walden, M-H Cormier, and William BF Ryan. 2000. Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, Vol. 2. IEEE, 1787–1792.

[28] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems* 35 (2022), 24611–24624.

[29] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre M. Bayen, and Yi Wu. 2022. The Surprising Effectiveness of MAPPO in Cooperative, Multi-Agent Games. In *NeurIPS*.