Influence-Focused Asymmetric Island Model

Extended Abstract

Andrew Festa
Oregon State University
Corvallis, United States
festaa@oregonstate.edu

Gaurav Dixit
Oregon State University
Corvallis, United States
dixitg@oregonstate.edu

Kagan Tumer
Oregon State University
Corvallis, United States
ktumer@oregonstate.edu

ABSTRACT

Learning good joint-behaviors is challenging in multiagent settings due to the inherent non-stationarity: agents adapt their policies and act simultaneously. This is aggravated when the agents are asymmetric (agents have distinct capabilities and objectives) and must learn complementary behaviors required to work as a team. The Asymmetric Island Model partially addresses this by independently optimizing class-specific and team-wide behaviors. However, optimizing class-specific behaviors in isolation can produce egocentric behaviors that yield sub-optimal inter-class behaviors. This work introduces the Influence-Focused Asymmetric Island model (IF-AIM), a hierarchical framework that explicitly reinforces inter-class behaviors by optimizing class-specific behaviors conditioned on the expectation of behaviors of the complementary agent classes. An experiment in the harvest environment highlights the effectiveness of our method in optimizing adaptable inter-class behaviors.

KEYWORDS

Multiagent learning; Asymmetric agents; Optimization architecture

ACM Reference Format:
Andrew Festa, Gaurav Dixit, and Kagan Tumer. 2024. Influence-Focused Asymmetric Island Model: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Multiagent leaning is a promising paradigm to address challenging problems such as search and rescue [15] and air traffic control [6, 12]. Coordination in such settings requires agents to learn good joint-actions [1]. This is particularly challenging in problems with asymmetric agents (agents with distinct capabilities and objectives) that must learn diverse generalizable inter-agent relationships [7].

Multi-Fitness Learning (MFL), a hierarchical framework, facilitates the discovery of beneficial joint-actions by injecting pre-trained class-specific behaviors in the team optimization process [16]. However, pre-trained behaviors can be brittle to changes in the task dynamics. The Asymmetric island model (AIM) evolves asymmetric agents to acquire diverse generalizable behaviors by leveraging a combination of Quality-Diversity (QD) and evolutionary optimization [2]. The QD process, known as 'island' for each agent class, allows agents to learn diverse primitive class-specific behaviors [8, 14], whereas the evolutionary optimization, known as 'mainland', evolves populations of teams to maximize team fitness across several tasks. Periodic policy migration from the island to the mainland biases diversity search toward regions of the policy space that produce useful team behaviors [4]. However, the diversity search is performed for each agent class independently, which can produce egocentric behaviors that are sub-optimal in teams.

This work introduces Influence-Focused AIM (IF-AIM), a framework that augments AIM via inter-island migrations to reinforce inter-class behaviors. Each island periodically migrates its highest performing behaviors to other islands facilitating QD optimization to learn class-specific behaviors in the presence of other agent classes. This speeds the flow of information between islands by removing the need for policies to pass through the mainland. Empirical result in the harvest environment highlights the benefits of introducing inter-island migrations to produce agents that can express robust inter-class relationships.

2 BACKGROUND AND RELATED WORK

Hierarchical methods applied to multiagent learning can partially address key challenges such as temporal credit assignment and non-stationarity [3, 5, 9, 11, 13, 17]. Multi-fitness learning (MFL), a two-tier optimization framework, leverages multiple fitness functions to learn *what matters when* [12, 16]. However, defining several fitness functions for learning diverse behaviors requires domain knowledge and can potentially lead misaligned objectives.

Asymmetric Island Model (AIM) is a hierarchical method for evolving asymmetric agents that explicitly maximizes both agent diversity and team objectives [2]. AIM achieves this via a combination of distributed Quality Diversity (QD) and evolutionary algorithms. A QD process for each agent class, called an 'island', evolves a population of policies to maximize a class-specific reward [8]. An evolutionary algorithm samples policies from the islands to create a population of teams that is evolved to optimize the team reward. Periodically, policies from the high-fitness teams are migrated to the islands to bias QD towards regions of the behavior space that yield good team behaviors [2]. The policy migration ensures that the diversity search on the islands is aligned with the team objective.

3 INFLUENCE-FOCUSED ASYMMETRIC ISLAND MODEL

This work introduces the Influence-Focused Asymmetric Island Model (IF-AIM), an extension of the Asymmetric Island Model that explicitly enables agents to learn efficient interactions with other agent classes. An island in IF-AIM optimizes a class-specific behavior for a particular agent class. However, unlike AIM, IF-AIM

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

maintains a small non-learning population of all other agent classes on an island, achieved via periodic inter-island migrations.

$$m_i = m_{i-1} + (i * L) + K \quad (1)$$

We use a migration schedule given by equation 1, where, m_i is the number of generations between each migration, and K and L are hyperparameters that control the migration frequency. For the experiment in section 4, $K = 50$, $L = 25$ and $m_0 = 0$. Maintaining non-learning agents of each class on an island ensures that agents of each class learn in the presence of other agent classes.

An evolutionary optimization process (CCEA [10]), called as the 'mainland', samples policies from the islands at each migration, groups them into teams, and evaluates them on the team task. The sampling ensures that the behavioral diversity acquired on the islands permeates to the mainland teams. Similarly, policies from the highest performing teams on the mainland are migrated to the islands to bias the diversity search process [2].

4 EXPERIMENTAL PARAMETERS

We adopt the problem introduced in [2]. Agents of two classes, harvesters and excavators, must learn to coordinate in order to collect and clear two points of interests (POIs): resources and obstacles. We introduce a penalty in reward if a harvester collides with a resource or if an excavator collides with a resource. IF-AIM is compared with several baselines: a standard CCEA [10], multi-fitness learning [16], and the Asymmetric Island Model [2]. The environment has eight harvesters and excavators each, 16 obstacles and resources. we report the normalized fitness averaged over 10 statistical trials. Each agent has an observation radius of five units and uses the density sensors to capture the density of other agents, resources and obstacles [2]. The POIs use a radius of two units and density sensors described in [2]. The length of each episode is 50 time-steps. An agent's policy is defined as a fully connected feed-forward neural network with 2 hidden layers (adopted from [2]).

4.1 Dynamic Reward Penalties

This experiment inspects the effect of increasing reward penalties for collisions and how that affects the training of IF-AIM compared to several baselines. When examining the effect of increasing reward penalties, we look not only at the team fitness, but also the number of generations it takes to learn robust team behaviors. It is important to note that the learning curves for MFL, in figure 1, do not show the pre-training required for its class-specific behaviors.

The general trend shown in figure 1 highlights the strength and weakness of IF-AIM compared with MFL. As MFL is based on learning from a set of temporally abstracted behaviors, it can plan over much longer sequences much more quickly than approaches such as IF-AIM, or AIM, which train policies bootstrapped from primitive class-specific behaviors. However, MFL relies on intimate domain knowledge to pre-train diverse behaviors. As the reward penalty increases (from (a) to (c) in figure 1), the pre-trained behaviors provided to MFL start to become misaligned with the demands of the problem. Therefore, while MFL learns its policy the fastest, agents trained with MFL fail to adapt their class-specific behaviors to account for changes in the environment. In contrast, IF-AIM takes longer to produce meaningful behaviors, and for agent classes to reinforce how to use those behaviors alongside one another on the mainland, but the learnt behaviors can adapt fluidly via the island to mainland migrations.

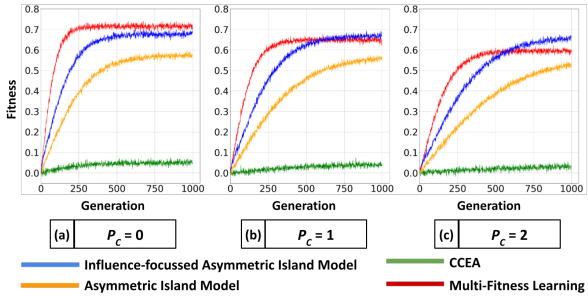

Figure 1: Normalized team fitness for CCEA, MFL, AIM, and IF-AIM, trained using varying reward penalties. MFL is consistently able to learn the fastest, but its performance degrades the most as the reward penalty increases, despite being provided the most amount of information. IF-AIM takes longer to learn, it outperforms the baselines in spite of the change in rewards.

5 DISCUSSION

This work presents the Influenced-Focused Asymmetric Island Model (IF-AIM), a framework for learning inter-class agent dependencies required to coordinate as robust teams in dynamic environments. By periodically incorporating non-learning representation of agent classes, the optimization for each agent class is able to learn behaviors that are conditioned on the actions of other classes. While pre-trained temporal abstractions (such as options [11]) are particularly useful when domain knowledge is available, relying on them is prone to producing teams that fail when the environment undergoes change. The migration of policies between the islands and the mainland ensures that agents trained with IF-AIM are able to discover and learn inter-class dependencies in response to changes in the environment.

IF-AIM achieves a balance between learning over pre-trained abstractions (MFL) and learning via independent optimization (AIM). The islands optimize their corresponding class-specific behaviors. The inter-island migrations allow for rapid sharing of these learned behaviors. Initially, there are more improvements to be made on the class-specific behaviors. Therefore, faster migration (equation 1) ensures that the learning process on each island uses an updated representation of other agent classes. The migration rate is reduced over the course of training to allow each island to fine-tune class-specific behaviors. In future work, we will consider nuanced migration schedules required in problems with potentially conflicting class-specific objectives.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation with grant No. IIS-2112633 and the Air Force Office of Scientific Research with grant No. FA9550-19-1-0195.

REFERENCES

[1] Stefano V Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (2018), 66–95.

[2] Anonymous. 2023. Learning Inter-Agent Synergies in Asymmetric Multiagent Systems. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 1569–1577.

[3] Craig Boutilier. 1996. Planning, Learning and Coordination in Multiagent Decision Processes. In *Proceedings of the 6th Conference on Theoretical Aspects of Rationality and Knowledge* (The Netherlands) *(TARK '96)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 195–210.

[4] Antoine Cully. 2019. Autonomous skill discovery with quality-diversity and unsupervised descriptors. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 81–89.

[5] Hoong Chuin Lau Duc Thien Nguyen, Akshat Kumar. 2018. Credit assignment for collective multiagent RL with global rewards. In *Advances in Neural Informaiton Processing Systems (NIPS 2018): Montreal, Canada, December 2-8*. 8102–8113.

[6] Jared Hill, James Archibald, Wynn Stirling, and Richard Frost. 2005. A multi-agent system architecture for distributed air traffic control. In *AIAA guidance, navigation, and control conference and exhibit*. 6049.

[7] Joel Z Leibo, Edward Hughes, Marc Lanctot, and Thore Graepel. 2019. Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. *arXiv preprint arXiv:1903.00742* (2019).

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

[9] Alexander Politowicz and Bing Liu. 2021. Learning to Dynamically Select Between Reward Shaping Signals. https://openreview.net/forum?id=NrN8XarA2Iz

[10] Mitchell A Potter and Kenneth A De Jong. 1994. A cooperative coevolutionary approach to function optimization. In *International Conference on Parallel Problem Solving from Nature*. Springer, 249–257.

[11] Richard S. Sutton, Doina Precup, and Satinder Singh. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence* 112, 1 (1999), 181–211. https://doi.org/10.1016/S0004-3702(99)00052-1

[12] Kagan Tumer, Zachary T Welch, and Adrian Agogino. 2008. Aligning social welfare and agent preferences to alleviate traffic congestion. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*. Citeseer, 655–662.

[13] Karl Tuyls and Gerhard Weiss. 2012. Multiagent learning: Basics, challenges, and prospects. *Ai Magazine* 33, 3 (2012), 41–41.

[14] Vassilis Vassiliades, Konstantinos Chatzilygeroudis, and Jean-Baptiste Mouret. 2017. A comparison of illumination algorithms in unbounded spaces. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 1578–1581.

[15] Jijun Wang, Michael Lewis, and Paul Scerri. 2006. Cooperating robots for search and rescue. In *Proceedings of AAMAS Workshop on Agent Technology for Disaster Management*.

[16] Connor Yates, Reid Christopher, and Kagan Tumer. 2020. Multi-Fitness Learning for Behavior-Driven Cooperation *(GECCO '20)*. Association for Computing Machinery, New York, NY, USA, 453–461. https://doi.org/10.1145/3377930.3390220

[17] Xiangbin Zhu, Chongjie Zhang, and Victor Lesser. 2013. Combining Dynamic Reward Shaping and Action Shaping for Coordinating Multi-agent Learning. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 2. 321–328. https://doi.org/10.1109/WI-IAT.2013.127