Learning Under the Influence of Others

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

KEYWORDS

Multiagent Learning, Communication

ACM Reference Format:
Andrew Festa, Gaurav Dixit, and Kagan Tumer. 2023. Learning Under the Influence of Others. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 2 pages.

1 INTRODUCTION

Multiagent learning has shown to be a highly effective tool when designing large-scale distributed systems. Their inherent distributed nature allows designers to build on each component individually as parts of a whole, leading to a simplification of designing the individual components.

However, the strength of these systems also belies their underlying weakness. Carving out the individual components is difficult, and even more so, it is difficult to train complex cooperation between these separate components with different capabilities that all must interact with each other in different ways in order to achieve a team objective. Learning in such a scenario thus gives rise to two particular challenges.

The first comes from the singular actions and sequences of actions in the environment. As there are numerous entities interacting simultaneously, the immediate and long-term effects of any singular action (of an agent) or interactions (between agents) can be difficult to precisely dictate to lead to a particular desired behavior. In traditional reinforcement learning, this can be seen as the reward hacking problem, where an agent may learn to exploit a nuance of the environment to maximize its long-term returns. This is magnified in a multi-agent setting where not only must the individual rewards be designed such as to lead a singular agent to maximize its reward, but also such that the interactions between agents aid in furthering the team objective of the agents in the system.

The second issue comes from the multiples agents in the system all learning simultaneously. From the perspective of any single agent in the system, this causes the environment to appear non-stationary. Given the same state-action pair of an agent, the system may provide a different next state as the other agents in the system as likewise making decisions based on their own state perceptions. This non-stationarity issue, inherent in multi-agent learning systems, introduces large levels of noise into the learning signal provided to an agent, making it more difficult for an agent to reinforce effective policies.

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. © 2023 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

The first challenge is particularly present in reinforcement learning, where an agent learns to map a state, or state-action pair, to a value function. This means that the system designer must consider all the low-level interactions in the environment when designing the reward function. Evolutionary learning, over reinforcement learning, alleviates this difficulty by changing the method by which feedback is provided to an agent. Instead of the reward being provided based on state, or state-action pairs, an agent's performance is evaluated based on a fitness function which scores the entire performance of the agent's policy. This can often be easier to design conceptually as it is at a high-level, although it can be much harder and less sample efficient for an agent to learn what constitutes a good behavior, as there is less feedback it can leverage to reinforce good behaviors, or good sub-behaviors.

Approaches leveraging temporal abstractions, such as multi-fitness learning or multi-agent options, help primarily with the second challenge by providing the agents with meaningful behaviors, rather than single actions, to plan over. This effectively reduces the temporal scale over which the agent must plan, making it more likely that the agents will discover sequences of joint-actions that lead to a positive system reward. These approaches, however, still suffer from the first challenge described above, albeit in a slightly different way. As MFL or multi-agent options does not allow for changing the behaviors while the agents are all learning simultaneously in the system, the designer must carefully consider which *behaviors* are appropriate for the agent to use to reach its desired behavior.

In this work, we present an evolutionary method for reinforcing complex interactions between agents with different capabilities where the success of an agent on the team is both dependent on its own behaviors and support from other agents on the team, both of the same type and of a different type. By leveraging an asymmetric island model, we are able to design and train agents of the team, and by separating out the agents, each type of agent is learning in a more stationary version of the environment, from the perspective of the learning agents. We then allow for intermittent migrations of learned policies not only to a mainland for joint-learning of all agent types, but between the sub-islands as well.

The agent islands are learning how to maximize their own rewards in the presence of other types of agents, and the mainland provides selective pressure to push all the agents towards collaborative policies. In doing this, the agents are able to learn more collaborative and effective policies to maximize the team objective. Additionally, not only is this island-based learning framework inherently both horizontally and vertically scalable, we show that this approach is more readily transferable to other variations of the environment, suggesting the agents learned a policy that is more closely aligned to the desired behaviors and does not exploit any particularities of the training environment.

The contributions of this work are a multi-agent evolutionary island-model:

- excels in reinforcing complex and extended-time inter-agent coordination that learns temporally abstract behaviors
- robust and transferable to different environmental dynamics
- inherently scalable to differing resource limitations of training machines

REFERENCES