Learning to Follow: Coordination Through Time

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

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

ABSTRACT

KEYWORDS

Multiagent Learning, Communication

ACM Reference Format:
Andrew Festa and Kagan Tumer. 2023. Learning to Follow: Coordination Through Time. 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

(1) What is the problem and why do I care?
Many complex tasks require not only that agents establish a point of coordination, but they maintain this coordination throughout a period of time. Picking up a heavy object may require multiple agents, but carrying it to another location would require the agents to move and act in tandem.

(2) Why is it important/difficult?
This is made challenging due to the high likelihood that any errant action may cause the group of agents to fail the task, or even be unable to make effective progress. If any agent makes a misstep, then the entire effort of the group of agents may be negated.

(3) What has been done already in this problem area?
- DPP
- Long Term Temporal Credit Assignment
- MFL
- MERL

Efforts towards problems of these types generally approach one aspect of the problem. Credit assignment techniques, such as difference rewards and DPP, seek to provide information to agents about how their actions contributed to the task. However, these solutions tend to break down when considering the temporal aspect of the problem largely due to how much information is lost when trying to condense the contribution of a whole sequence of actions into a single value.
Long Term Temporal Credit assignment addresses the temporal aspect of agent interactions, but it is more focused towards single agent learning. A single agent learns to effectively act over a sequence in order to allow for another agent to perform its task. MFL is similar to this in that each agent learns its own policy about what matters when. The temporal coordination is achieved by allowing agents to select from complementary behaviors.

(4) What particular problem remains unsolved?
While individual parts of the problem have been addressed, these techniques do not address the entire problem as a whole, particularly as the coupling or temporal requirement increases. DPP fails particularly with respect to the temporal aspect of the problem, long term credit assignment does not scale with the coupling requirement of the task, and MFL requires providing a set of unchanging behaviors to select from during the episode. If there is not a behavior (or set of joint behaviors) that is capable to solving the task, then it is not possible for the agents to learn to coordinate over time to achieve the task.
Differentiating from MFL is the hardest, but it is meant to get at that MFL's coordination is more rigid (unchanging) and requires the behaviors to select from prior to learning to coordinate.

(5) How did you solve it?
At the core, all of the agents are capable of learning their own policies. But they also have a simplistic hard-coded policy (magnetic policy) that guides their actions in the presence of other agents. The agents essentially learn to update both of these policies during the learning process. During an episode, an agent can chose to follow the magnetic policy, and during this time, it will update the parameters of its magnetic force to align more closely with the average action of the agents around itself.
The agents thus are learning to rely on other agents to help guide their decision making process during the course of an episode, but they do not throw away potentially useful knowledge about *how* they should be leveraging the learning of other agents between episodes.
Possibly another policy where an agent selects another nearby agent to dictate the action it should take for that step.

(6) What is cool about your approach?
In many learning problems, agents have either fully their own policies, a shared policy amongst all agents, or between subsets of agents. This method for sharing knowledge is overly strict in that they learn the same thing for a state representation, or they all learn their own individualistic behaviors. Our approach blends these two extremes and allows agents to learn when to leverage the learning of others to guide their own actions. So they still learn how to behave individually, but they do not waste the discard the learning of others.
Note: Make note of what happens when all agents decide to follow the magnetic policy for a sequence of actions. Emergence of a leader versus emergence of a swarm.
Sometimes it is better to rely on others, or to act cohesively, rather than operating individualistically optimal.

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.

(7) What were your key results?

The experiments are based on an OpenAI paper where an agent has to learn to collect a key, allowing another agent to capture a reward [1]. We extend this problem by requiring each of these parts of the task to require multiple agents. We compare our results against MFL in particular and show that, especially as the coupling requirement increases, our approach is able to consistently perform the task by leveraging the simplification of the learning process. This can be seen not only in the learning curves of each agent, but also through examining how much each agents had to explore before finding the valid solutions.

Note: The coordination aspect is achieved through relying on the actions of other agents to guide ones own actions. Thus, the analysis should include some aspect about how much the agents leverage the learning of others to guide their decisions.

(8) What are the contributions of this paper?

This work provides a method for agents to effectively learn to coordinate throughout a sequence of actions for scaling, tightly coupled tasks. Additionally, it provides a framework for leveraging the learning of nearby agents to reduce reduce the difficulty for any given agent helping to achieve the goal.

Note: I want to capture the two points that it allows an easier method for agents to coordinate through time. And it relies on the actions of others to guide how it makes its decisions.

REFERENCES

[1] Cathy Yeh. [n. d.]. Long Term Credit Assignment with Temporal Reward Transport. https://openai.com/blog/openai-scholars-2020-final-projects/#cathy