Generalized Multi-Agent Reinforcement Learning in Cooperative and Competitive Environments

Diana Huang * [1] **Nitin Viswanathan** * [1 2] **Shalini Keshavamurthy** [2]

Abstract

This document provides the current progress on our project. The objective of the

1. Introduction

There are many applications where multiple agents need to learn how to act together, such as communication scenarios and multiplayer games. Traditional reinforcement learning approaches do not work well on these problems because the environment for any single agent changes over time as other agents change their policies, leading to instability during training and high variance in results (Lowe et al., 2017).

Existing efforts in Multi-Agent Reinforcement Learning (MARL) propose new techniques to get around these issues, but they do not generalize well to larger numbers of agents due to the growing computational complexity. We will explore generalizable training approaches for MARL to find policies that allow training on a smaller number of agents but scale well when applied in environments with more agents.

We will use an existing OpenAI Gym environment created for MARL [1]. This environment supports a variety of cooperative and competitive scenarios, as well as environments that allow communication between agents (beyond just their locations in the environment).

We have reviewed many papers in the MARL space as preparation for our project. (Lowe et al., 2017) is the most relevant paper as it uses the same environment that we plan to use, and modifies Deep Deterministic Policy Gradients (DDPG) to work in a multi-agent setting. [3] and [4] focus specifically on the task of communication amongst cooperative agents.

As a starting baseline for the project we have implemented the policy gradient method with OpenAI multi-particle-envs. We have chosen 'simple-spread' environment to evaluate the performance of policy gradient method.

2. Approach

This section will describe in detail the environment setup, evaluation metric used and results.

2.1. Environment setup

Write about open ai gym environments. Write more in detail about environment we are using for the experiments. Describe agents, rewards, observation dimension, etc.

2.2. Experiments

Write about
1. environment
2. configuration parameters
3. algorithm used 4. expected results

2.3. Evaluation

Write about evaluation metrics for multi-agent environments. which metric is chosen and why?

2.4. Results

write about the results we get from the experiments. Include video snapshots , tables.

3. Future work

As a next baseline to gauge performance, we will implement the approach used in (Lowe et al., 2017) and test it on a variety of competitive and cooperative environments. In order to generalize this approach, we will train agents using information only on nearby agents, not on every agent in the environment, as this limits the complexity and specificity of any learned policies.

We will evaluate our model performance both during train and test time. During train time, we will see how quickly we are able to reach convergence and how stable the training process is across the different implemented methods. During test time, we will compare our different methods with the average reward over many trials as well as the variance of the average, and also see how performance changes as we change the number of agents present to test generalization. As a stretch goal, we will experiment with on-policy approaches (e.g. SARSA) and see if they can be used to further improve an optimal policy produced by an off-policy approach.

References

Lowe, Ryan, Wu, Yi, Tamar, Aviv, Harb, Jean, Abbeel, Pieter, and Mordatch, Igor. Multi-agent actor-critic for mixed cooperative-competitive environments. *CoRR*, abs/1706.02275, 2017. URL http://arxiv.org/abs/1706.02275.