Deep Reinforcement Nano Degree Project 2: Continuous Control

Peng Zhao

December 19, 2019

Abstract

This report briefly summarizes the deep reinforcement nano degree project 2, which is training the a double-jointed arm to move to target locations using Deep Deterministic Policy Gradient (DDPG) method. The problem is solved after episodes in this experiment.

1 Deep Deterministic Policy Gradient (DDPG) Algorithm

Deep Deterministic Policy Gradient (DDPG) is a kind of actor-critic method. The actor neura network is used to approximate the optimal deterministic policy. The critic in DDPG is used to learn the optimal action value function by using the actors best belived action. The facts of DDPG are:

- DDPG is an off-policy algorithm

- DDPG can only be used for environments with continuous action space

- DDPG can be thought of as being deep Q-learning for continuous action spaces

- The Spinning Up implementation of DDPG does not support parallelization

2 Architecture

The actor neural network comprise 3 layers with ReLu activation. The input size of the first layer is the state size. The output and the input size of the second layer is 400. The output size and the input size of the third layer is 300.

The critic neural network has the same architecture with the actor.

3 Result

The environment has been solved in this experiment, which is shown in the following figure. As shown, after an initial training, the score converges to about 40 after around 60 episodes. The required average score (30) is finally reached after 10 episodes.
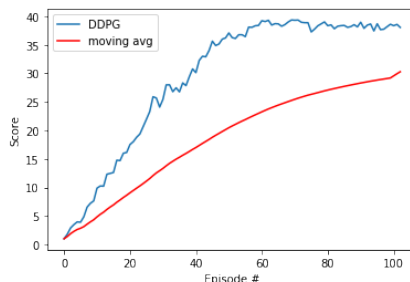

Figure 1: Result