Report - Continuous Control

1. Environment Description

In this environment, a double-jointed arm can move to target locations. A reward of +0.1 is provided for each step that the agent's hand is in the goal location. Thus, the goal of the agent is to maintain its position at the target location for as many time steps as possible.

The observation space consists of 33 variables corresponding to position, rotation, velocity, and angular velocities of the arm. Each action is a vector with four numbers, corresponding to torque applicable to two joints. Every entry in the action vector should be a number between -1 and 1.

The solution uses the environment of version 2, which contains 20 identical agents, each with its own copy of the environment.

2. Learning Algorithm

The solution in this submission adopts DDPG as the learning model. It consists 2 neural networks, one actor model and one critic model. Each has 2 hidden layers with respectively 64 and 128 nodes and 1 output layer. Batch Normalization is followed by the first layer and the number of nodes in output layer in actor is number of actions which has been described in the first section, 4 in this case and 1 in critic. The activation function in the output layer for actor and critic respectively are tanh and relu.

The details of the neural net can be found in “model.py”.

Both actor and critic take two neural networks with identical structures, one local and one target. It also sets a replay memory to replay the experiences. In the submission, soft update is adopted, which updates both target and local network at the same time but uses a parameter tau to determine the parameters of the updated target network

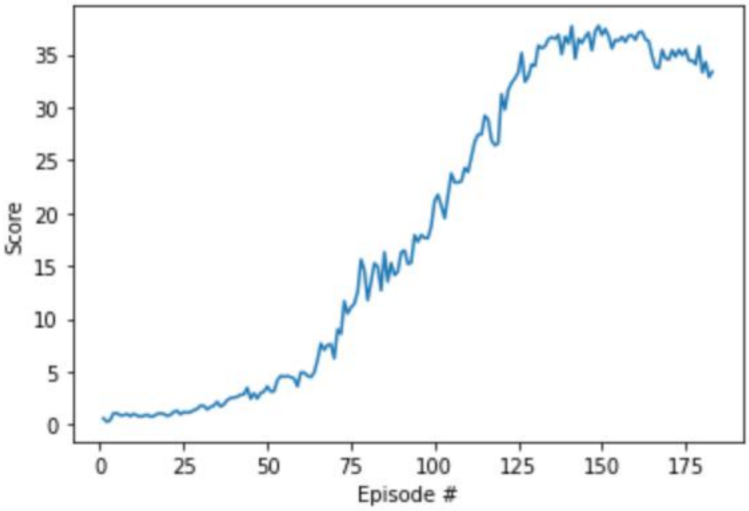
The hyper parameters of the model is set as follows:

Replay Buffer Size: 100000
Sampled number for relay: 128
Discount factor - Gamma: 0.99
Learning rate: 0.0002
Number of steps to update target network: 4

Tau: 0.001

3. Result

The environment is solved in 183 episodes and the plot of rewards is as below:



4. Future Work

Since this work only tries DDPG and there are a few things that can be tried in future work.

1) D4PG
2) PPO