Report - Collaboration and Control

1. Environment Description

In this environment, two agents control rackets to bounce a ball over a net. If an agent hits the ball over the net, it receives a reward of +0.1. If an agent lets a ball hit the ground or hits the ball out of bounds, it receives a reward of -0.01. Thus, the goal of each agent is to keep the ball in play.

The observation space consists of 8 variables corresponding to the position and velocity of the ball and racket. Each agent receives its own, local observation. Two continuous actions are available, corresponding to movement toward (or away from) the net, and jumping.

2. Learning Algorithm

The solution in this submission adopts DDPG as the base learning model and adjust it for multiple agents. 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, 2 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:

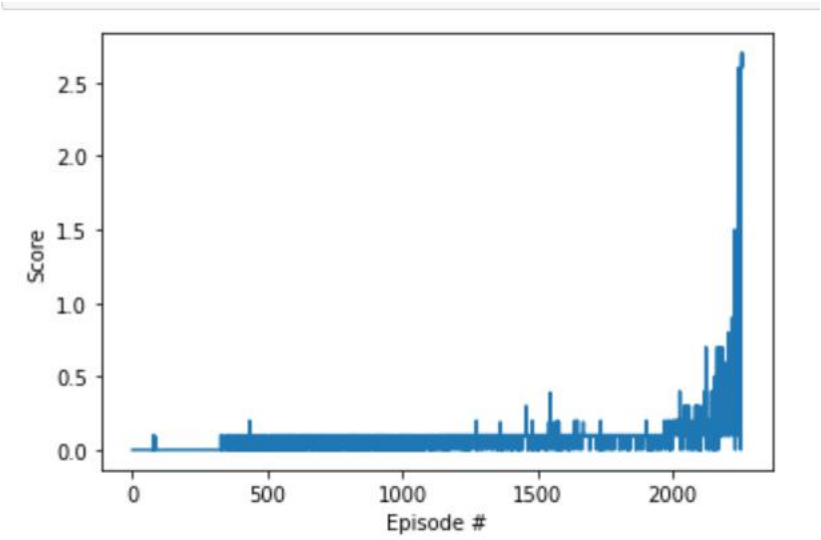
Replay Buffer Size: 100000
Sampled number for relay: 128
Discount factor - Gamma: 0.99
Learning rate: 0.0001
Tau: 0.001

The algorithm was adjusted for multiple agents in two ways:

- Each agent has it's own actor and critic network while sharing the same replay buffer
- States include observations from all the agents instead of from single agent

3. Result

The environment is solved in 2258 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) Prioritized experience replay
2) PPO, TRPO