Udacity Collaboration and Competition

Introduction

This Project is given by Udacity Deep Reinforcement course. The aim is train two tennis playing agent that they can play tennis each other's. The aim is hitting ball ant passing it to other side over net. A reward 0.1 point will be given for each hit that pass over net and will be given a negative reward -0.01 if the ball hits ground.

The aim is obtaining an average score of 0.5 over 100 consecutive episodes.

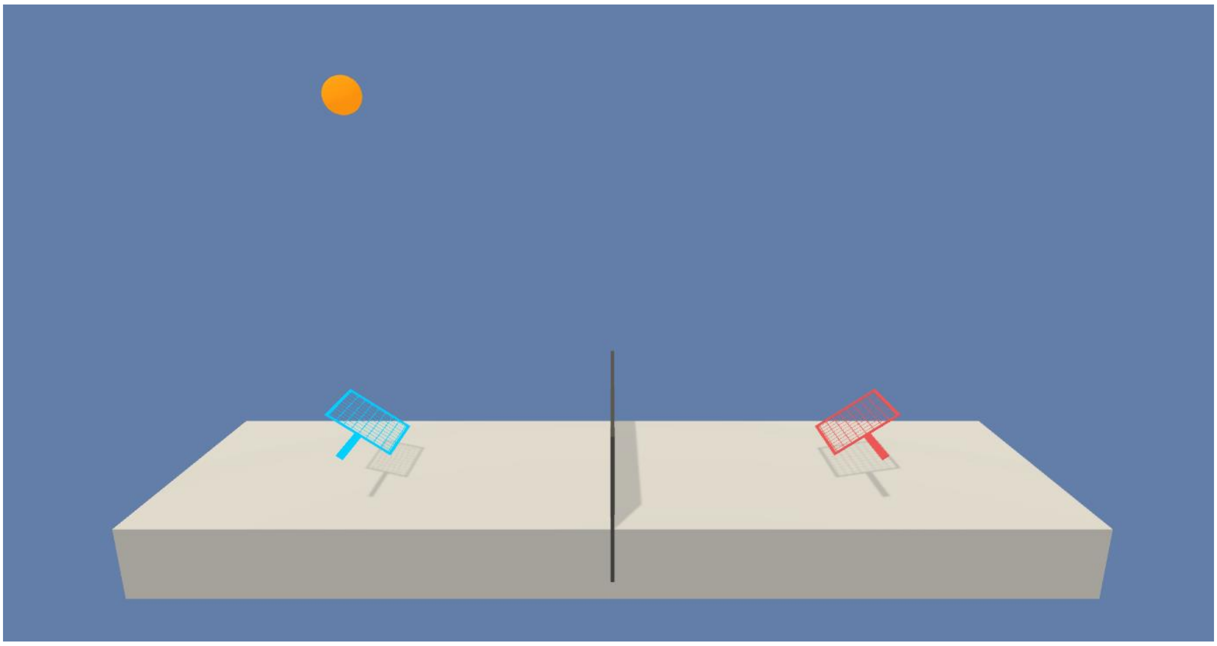

Figure1: An image of environment of tennis.

The environment has continuous action space with size of 2 and these corresponding position and velocity. The state size is 8 and consist of position and velocity of racket and ball.

The agents are trained for score of 0.5 and the result is as following and it run in GPU mode in Udacity environment.

Environment solved in 1082 episodes!   Average Score: 0.51 in 905 seconds.

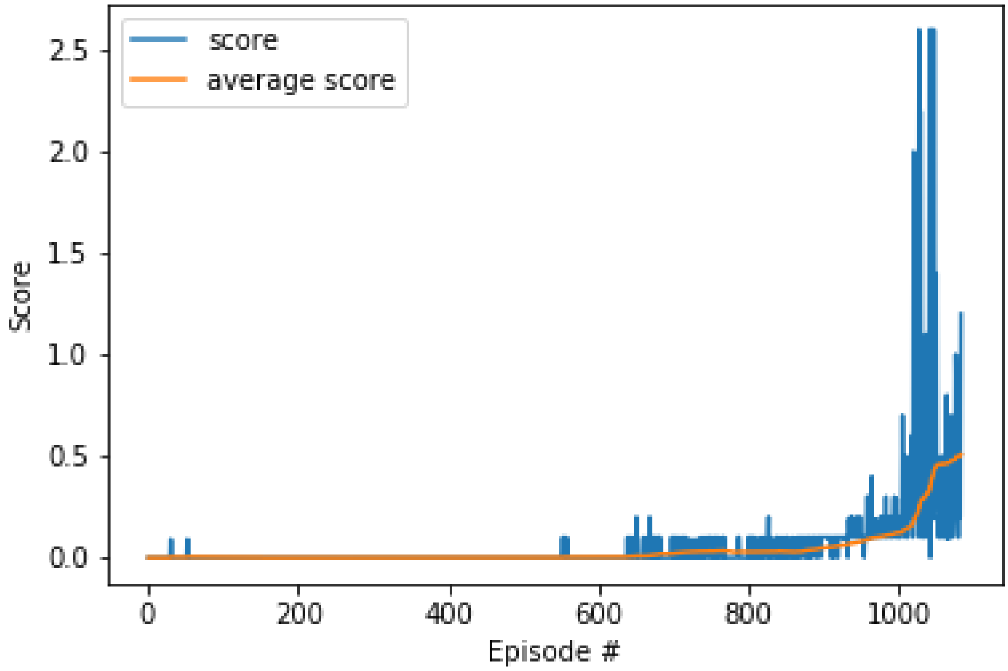

Figure 2: Graph of score 0.5

Learning Algorithm & Model Architecture:

The DDPG algorithm is used in the project. It has 3 layers. 8 inputs, 256 nodes and 128 nodes hidden layer and 2 action node for continious. The same architecture is used both actor and critic side of algorithm.

The architecture is selected with some experiment.

The experiment was run the agent for score 10 and compare the results. Three experiments are done and the same architecture is used both critic and actor side, one for architecture 128*128 nodes, one for 256*128 nodes and one for 128*56 nodes.

The other results are not listed here since the process is cancelled during run time since the process takes too long time. (tried architectures 128*128,128*256,256*128)

Therefore, for this project is chosen an architecture 256*128 due to time concern.

BUFFER_SIZE=int(1e6)

BATCH_SIZE=128

GAMMA=0.99

TAU=1e-3

LR_ACTOR=1e-4

LR_CRITIC=1e-3

WEIGHT_DECAY=0

RANDOM_SEED=0

QUNoise_theta = 0.15

QUNoise_sigma = 0.2

Problems During Project:

During the project, there were two different difficulty. First one was network size problem. It is required to find an optimum architecture since it takes too long time without converging within maximum number of episodes. For this purpose, I ran GPU power in Udacity workspace to make the things quicker to find optimum point. The second one was understanding the process between two agents since during course, we have trained only single agent.

Future Works:

- Parameter optimization
  - The parameters are set by default but there could be better parameters for this environment and architecture.
- Implementing soccer environment solution.
  - A new environment could give better understanding of the course.
- Implementing different agent for the environment.
  - A new algorithm could be used for inspiration.