Project report

1. Learning algorithm

In this project, a multi-agent version of ddpg algorithm was implemented and tested for the Tennis environment. According to the paper, Multi-Agent Actor-Critic for Mixed Cooperative Competitive Environments, the **centralized training and decentralized execution** framework was implemented.

Basically, the ddpg algorithm is an Actor Critic Model that consists of 2 neural networks, where the Actor model is used to approximate the policy function and takes the current state as input and outputs the action to take at that step. The Critic Model is used to approximate the value function from that input state. In the Multi Agent algorithm, multiple Agents are instantiated together, where each of them sample experience events from the same shared Replay Buffer. The Environment returns states for multiple agents, and each is used to update and take an action from the individual agent.

During training, a critic for each agent uses the state observed and extra information like states observed and actions taken by other agents through the shared replay buffer. But, each actor has access to only its observations and actions. During execution time, only the actors are present, and hence all observations and actions are used. A learning agent for each agent allows us to use a different reward structure for each. Hence, the algorithm can be used in all cooperative, competitive, and mixed scenarios.

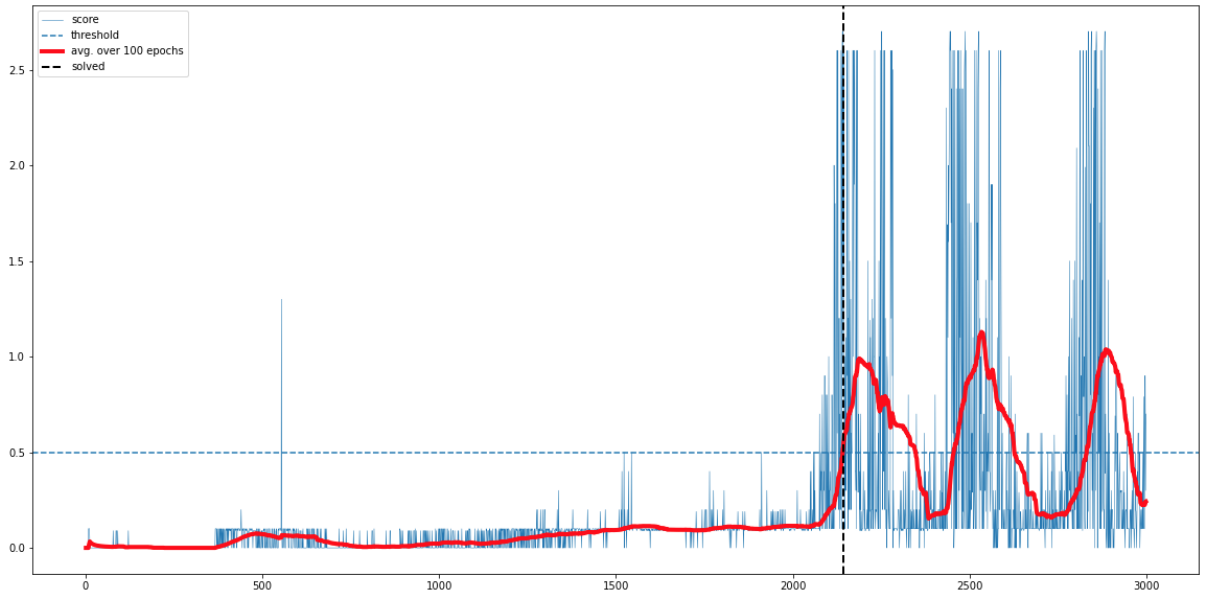
2. multi-agent DDPG Hyper-parameters and models

<table>
<tr><td>Hyper-parameters</td><td colspan="2">BUFFER_SIZE = int(1e5)
BATCH_SIZE = 128
GAMMA = 0.99
TAU = 1e-3
LR_ACTOR = 1e-4
LR_CRITIC = 1e-4
WEIGHT_DECAY = 1e-6</td></tr>
<tr><td rowspan="2">Model</td><td>Actor Network</td><td>Critic Network</td></tr>
<tr><td>24 x 128 input layer
128 batch normalizer
relu
128 x 64 hidden layer
relu
64 x 2 output layer 2
Tanh</td><td>24 x 128 input layer
128 batch normalizer
relu
(128+2) x 64 hidden layer
relu
64 dropout(p=0.2)
64 x 2 output layer</td></tr>
</table>

3. Results

The average test score first reached the target 0.5 at the 2145th episode, successfully completing the task.

In [ ]:



4. Future Work

Even though the implementation completed the task, it looks like more improvements are needed in that the ddpg algorithm with 2 Agents was tried multiple times but never exceeded avg. 1.5 score over 100 episodes. Moreover, in the long-term, the avg. performance became unstable and not converged, oscillating in a certain interval. In the future work, more parameter tunings and modifications focused on this issue should be tested.