Report

July 26, 2019

Deep Reinforcement Learning Project 3 - Collaboration and Competition

1   Learning Algorithm

The learning algorithm used to train this agent was MADDPG. This is an ActorCritic Algorithm. ActorCritic algorithms takes the best of Value-Based and Policy-Based algorithms into consideration.

2   MADDPG Agent with batch normalization

- Agents Hyper Parameter:

```
BUFFER_SIZE = int(1e6)  # replay buffer size
BATCH_SIZE = 512        # minibatch size
GAMMA = 0.99            # discount factor
TAU = 5e-2              # for soft update of target parameters
LR_ACTOR = 5e-4         # learning rate of the actor
LR_CRITIC = 5e-4        # learning rate of the critic
WEIGHT_DECAY = 0.0      # L2 weight decay
EPSILON = 1.0           # noise decay
```
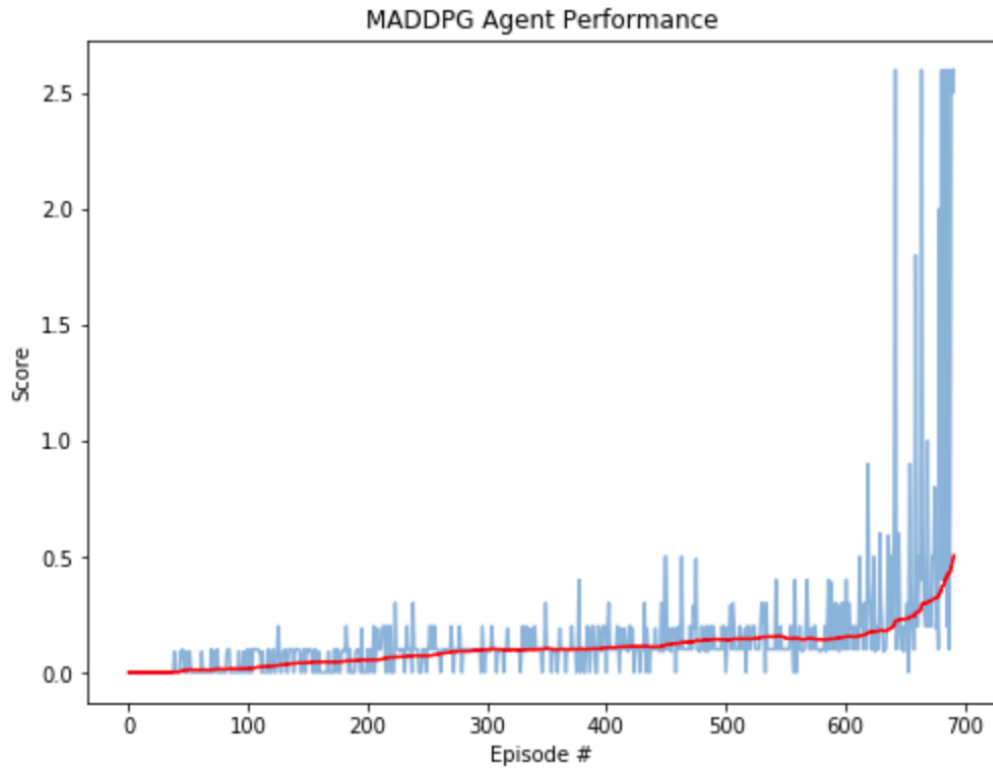
- ModelStructureActor:

  Actor:

```
state_size--> BN0--> fc1--> BN1--> ReLU--> fc2--> BN2--> ReLU--> fc3--> tanh
                                                      --> action_size

#fc1_units = 256
# fc2_units = 128
```

  Critic:

```
state_size--> BN0--> fc1--> BN1--> ReLU--> fc2--> BN2--> ReLU--> fc3
                                                      --> state-valuefunction
```

MADDPG Agent Performance

ScoresPlot.png

3   Plot of Rewards:

```
Episode 100 Average Score: 0.02
Episode 200 Average Score: 0.05
Episode 300 Average Score: 0.10
Episode 400 Average Score: 0.11
Episode 500 Average Score: 0.14
Episode 600 Average Score: 0.15
Episode 692 Average Score: 0.50
Environment solved in 592 episodes! Average Score: 0.50
CPU times: user 1h 56min 13s, sys: 25.8 s, total: 1h 56min 39s
Wall time: 40min 28s
```

4   Ideas for Future Work

It would be interesting to try the following modidifications:

```
Training the agent with PPO2
```