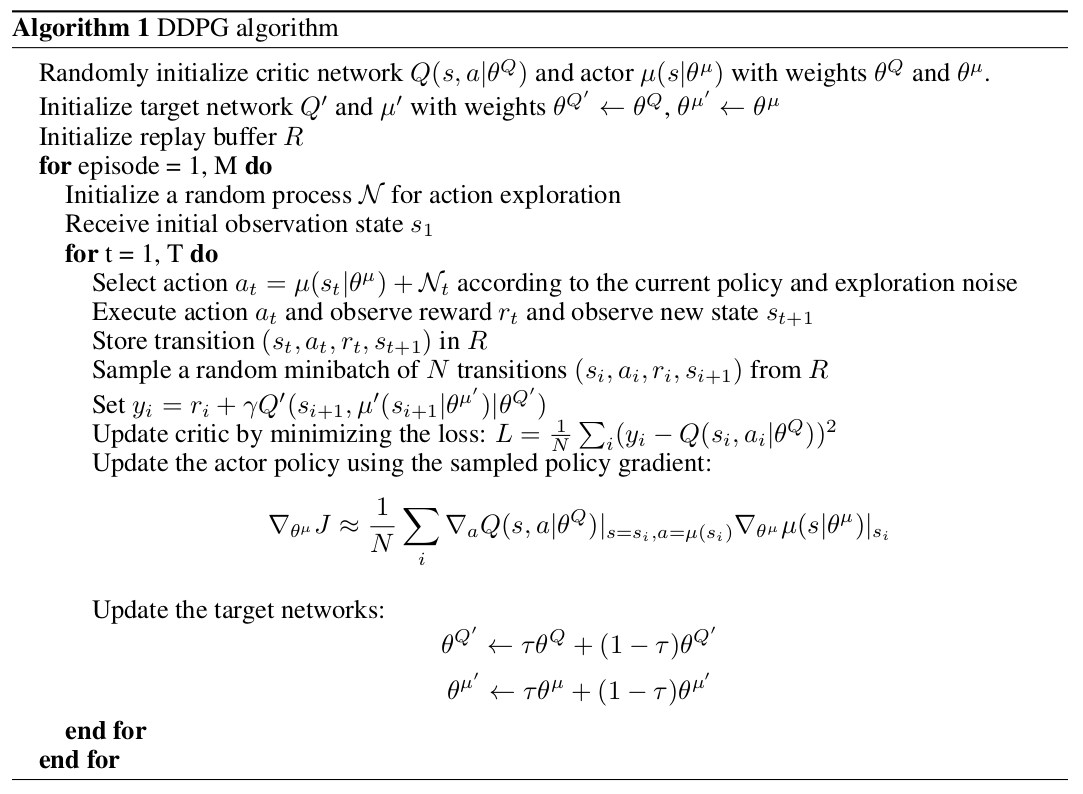
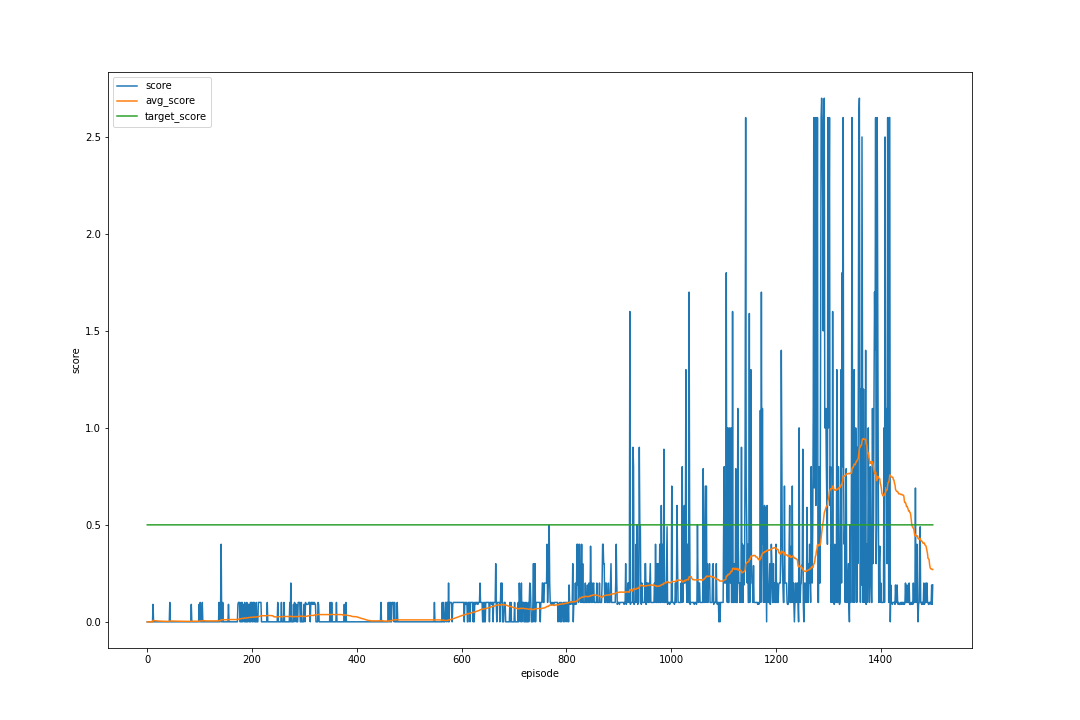
1. Learning Architecture
   1. Algorithm

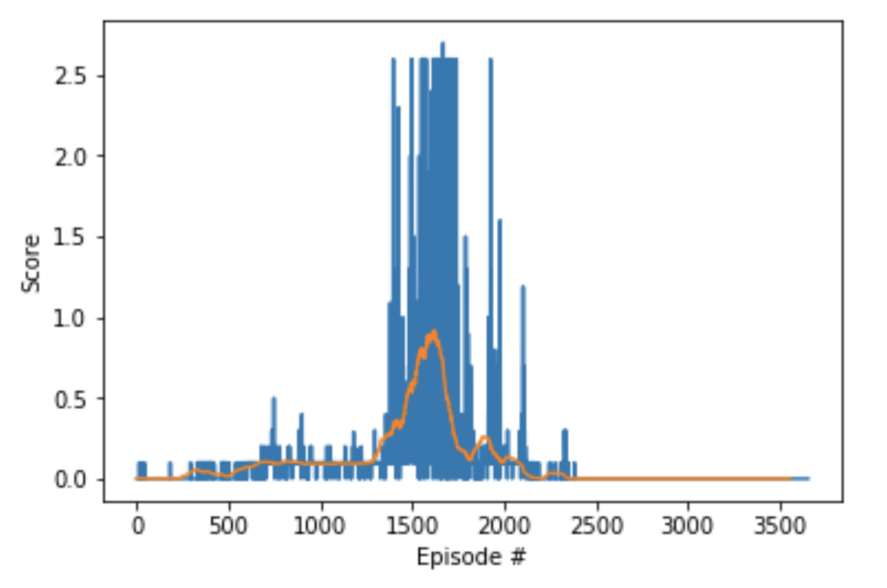


* 1. Model Structure
     1. Actor:
        1. Input size = State size = 24
        2. Hidden layers(2)
           1. Fully connected with 128 batch-normalized rectifiers
           2. Fully connected with 256 rectifiers
        3. Output size = Action size = 2
     2. Critic:
        1. Input 1 size = State size = 24 at 1st layer
        2. Input 2 size = Action size = 2 concat to 2nd layer
        3. Hidden layers(2)
           1. Fully connected with 128 batch-normalized rectifiers
           2. Fully connected with 256+4 rectifiers
        4. Output size = Value function = 1
  2. Hyperparameters(Final)
     1. Batch size 40
     2. Memory buffer size 1e6
     3. Number of episodes 1500
     4. Target score 30.0
     5. Discount factor gamma 1e-3
     6. Learning rate for Actor 1e-4
     7. Learning rate for Critic 1e-3
     8. Update Period 5
     9. Update Times per update 10
     10. Weight Decay 0
     11. Agent number 20
     12. Alpha(prioritized exp replay) 0.7
     13. Beta(prioritized exp replay) 0.8

1. Results
   1. DDPG with prioritized experience replay
      1. Folder: ddpg\_result\_2023\_05\_21\_22\_08\_57
      2. Result:



1. Conclusions
   1. The prioritized exp replay method do have accelerations on training(<1300 episodes), comparing to the original DDPG baseline provided by course instructions below(>1500 episodes).



1. Future Improvements
   1. Will try to implement n-step bootstrapping
   2. Will try to use array to represent replay buffer’s binary tree
   3. Will try to implement other algorithms like Reinforce and TRPO