

林柏均 107062240

May 30, 2021

Deep Reinforcement Learning

Homework 4

What have I done in this assignment?

Instead of using traditional Deep Q-Learning, this time I used TD3 to train my agent. The reason why I abandoned traditional way to train my agent because of the performance. After 10000 episodes of training, we can notice that the runner can do something different.

Description of my model

TD3:

- state_dim (float): max priority
- action_dim (int): next index of tree
- max_action (float): alpha parameter for prioritized replay buffer

While DDPG can achieve great performance sometimes, it is frequently brittle with respect to hyper-parameters and other kinds of tuning. A common failure mode for DDPG is that the learned Q-function begins to dramatically overestimate Q-values, which then leads to the policy breaking, because it exploits the errors in the Q-function. Twin Delayed DDPG (TD3) is an algorithm that addresses this issue by introducing three critical tricks:

Conclusion

After training so long and times, we can notice that the skeleton is not easy to move forward. I cannot understand what reason will cause it so hard to train. In the evaluation phase, I notice the return value of select_action function always return the value approaching to 0 or 1, it quite weird since I printed out the action array during the training phase, the distribution of the action value is quite normal.

In this assignment, I think that I should try different exploration method, so that the skeleton can try more different strategy. Below figure is part of the training duration, the reward value returned from the environment cannot increase dramatically. It always retains between 10 to 20 !

```
Total T: 50596 Episode Num: 400 Episode T: 111 Reward: 6.903
Total T: 50694 Episode Num: 401 Episode T: 98 Reward: 3.211
Total T: 50834 Episode Num: 402 Episode T: 140 Reward: 10.663
Total T: 50965 Episode Num: 403 Episode T: 131 Reward: 9.937
Total T: 51065 Episode Num: 404 Episode T: 100 Reward: 2.614
Total T: 51194 Episode Num: 405 Episode T: 129 Reward: 11.604
Total T: 51343 Episode Num: 406 Episode T: 149 Reward: 12.949
Total T: 51465 Episode Num: 407 Episode T: 122 Reward: 9.400
Total T: 51581 Episode Num: 408 Episode T: 116 Reward: 9.105
Total T: 51704 Episode Num: 409 Episode T: 123 Reward: 11.700
Total T: 51824 Episode Num: 410 Episode T: 120 Reward: 9.940
Total T: 51920 Episode Num: 411 Episode T: 96 Reward: 3.582
Total T: 52018 Episode Num: 412 Episode T: 98 Reward: 5.315

simbody-visualizer: received Shutdown message. Goodbye.
Total T: 52112 Episode Num: 413 Episode T: 94 Reward: 9.400
Total T: 52274 Episode Num: 414 Episode T: 162 Reward: 8.292
Total T: 52427 Episode Num: 415 Episode T: 153 Reward: 8.600
Total T: 52518 Episode Num: 416 Episode T: 91 Reward: 9.100
Total T: 52663 Episode Num: 417 Episode T: 145 Reward: 9.645
Total T: 52780 Episode Num: 418 Episode T: 117 Reward: 7.667
Total T: 52901 Episode Num: 419 Episode T: 121 Reward: 10.733
Total T: 53037 Episode Num: 420 Episode T: 136 Reward: 10.469
Total T: 53162 Episode Num: 421 Episode T: 125 Reward: 6.809
```