DQN Assignment

Advances in robotics and control

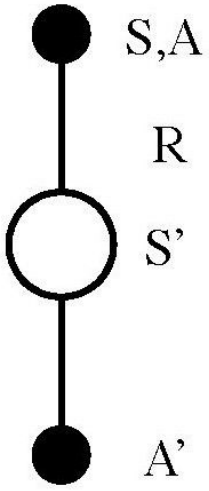
Animesh sahu

20161028

2.

1)

For SARSA

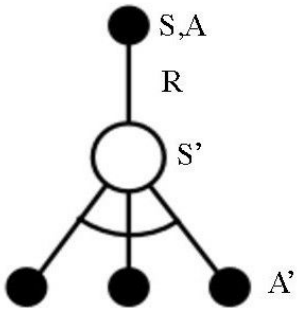


$$Q(S,A) \leftarrow Q(S,A) + \alpha\left(R + \gamma Q(S',A') - Q(S,A)\right)$$

S ➔ current state, A ➔ current action, R ➔ current reward

S` ➔ next state, A` ➔ next action

FOR Q learning



$$Q(S,A) \leftarrow Q(S,A) + \alpha \left(R + \gamma \max_{a'} Q(S',a') - Q(S,A) \right)$$

2). Monte Carlo control methods do not suffer from this bias, as each update is made using a true sample of what Q(s,a) should be. However, Monte Carlo methods can suffer from high variance, which means more samples are required to achieve the same degree of learning compared to TD.

TD exploits the Markov property, i.e. the future states of a process rely only upon the current state, and therefore it's usually more efficient to use TD in Markov environments.MC does not exploit the Markov property as it bases rewards on the entire learning process, which lends itself to non-Markov environments.

MC must wait until the end of the episode before the return is known. But TD can learn online after every step and does not need to wait until the end of episode.

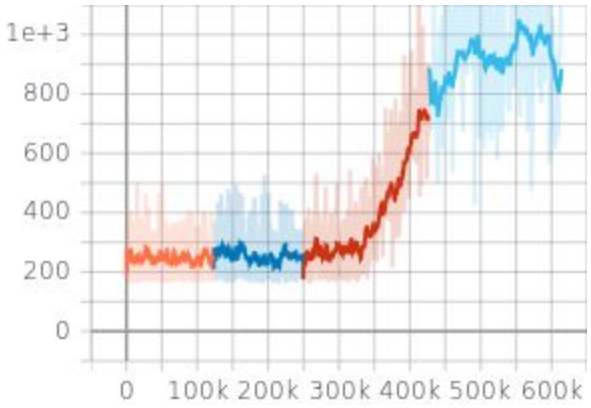
3.) No.SARSA is an on-policy algorithm (it follows the policy that is learning) and Q-learning is an off-policy algorithm (it can follow any policy (that fulfills some convergence requirements).though we know that In both cases the policy followed by the agent is epsilon-greedy which is important for exploration.If we use only the greedy policy then there will be no exploration so the learning will not work. In the limiting case where epsilon goes to 0 (like 1/t for example), then SARSA and Q-Learning would converge to the optimal policy q*. However with epsilon being fixed, SARSA

will converge to the optimal epsilon-greedy policy while Q-Learning will converge to the optimal policy q*.

DQN

Here is the graph of episode length.

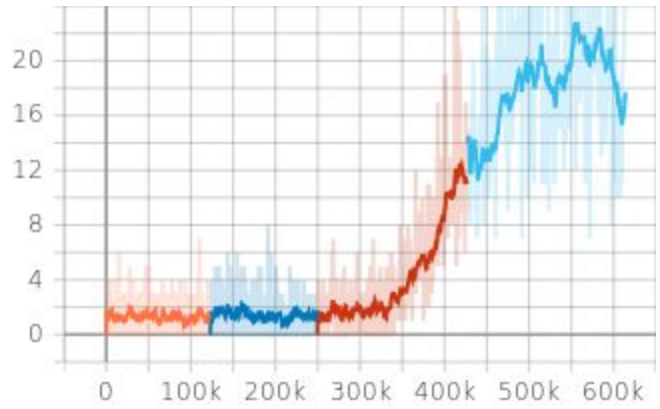
According to our observations,The more we hit the ball , the more is the time spent in an episode.



Episode length

And below is graph for episode reward

Here we can see that as our agent learns more and more our reward increases.



Video is in the submission folder.