

IEOR 8100 Lab 3: Pseudocode for DQN

March 2, 2018

Algorithm 1 DQN

```
1: INPUT: target network update period  $\tau$ , total number of episodes  $E$ , initial time steps before update init, learning rate  $\alpha$ ,  
   exploration prob  $\epsilon$ , batchsize for training  $N$   
2: INITIALIZE: DQN principal network  $Q_\theta(s, a)$  with parameters  $\theta$ , target network  $Q_{\theta^-}(s, a)$  with parameters  $\theta^-$ , time steps  
   counter  $counter \leftarrow 0$ , empty buffer  $R \leftarrow \{\}$   
3: for  $e = 1, 2, 3 \dots E$  do  
4:   while episode not terminated do  
5:     Execute actions  
6:      $counter \leftarrow counter + 1$   
7:     Given state  $s_t$ , for prob  $\epsilon$ , take action uniformly random; otherwise, take action by being greedy  $a_t \leftarrow \arg \max_a Q_\theta(s_t, a)$   
8:     Save experience tuple  $\{s_t, a_t, r_t, s_{t+1}\}$  to buffer  $R$   
9:     Training  $\theta$  by gradients  
10:    Sample  $N$  tuples  $\{s_i, a_i, r_i, s'_i\}$  from replay buffer  $R$  (uniformly)  
11:    Compute target  $d_j = r_j + \max_{a'} Q_{\theta^-}(s'_j, a')$  for  $1 \leq j \leq N$   
12:    Compute empirical loss  

$$L = \frac{1}{N} \sum_{j=1}^N (Q_\theta(s_j, a_j) - d_j)^2$$
  
13:    Update  $\theta \leftarrow \theta - \alpha \nabla_\theta L$   
14:    Update target network  $\theta^-$   
15:    if  $counter \bmod \tau = 0$  then  
16:      Update target parameter  $\theta^- \leftarrow \theta$ 
```

There are multiple ways to improve the efficacy of the algorithm.

- **Sampling:** Instead of sampling uniformly from buffer, sample using adaptive distribution (prioritized replay buffer).
- **Double DQN:** DQN overestimates Q values. To mitigate the issue, use Double DQN.
- **Architecture:** Decompose Q value into $Q(s, a) = V(s) + A(s, a)$ and use Dueling DQN architecture.
- **Soft update:** $\theta_{\text{target}} \leftarrow (1 - \beta)\theta_{\text{target}} + \beta\theta$.
- **Exploration constant:** Exploration constant can be made adaptive. Start with large ϵ and gradually decrease ϵ to be small.