
Algorithm 1 Deep Q-learning with experience replay

Input: $n_{\text{batch}}, n_{\text{rmem}}, l_{\text{history}}, f_{\text{tupdate}}, \gamma, r_{\text{learn}}, m_{\text{gradient}}, m_{\text{sgradient}}, g_{\text{min}}, \epsilon_{\text{first}}, \epsilon_{\text{final}}, l_{\text{expl}}, n_{\text{rpstart}}, l_{\text{nomax}}$

```
1: Replay Memory  $D$  を初期化
2: Q-Network  $Q$  をランダムな重み  $\theta$  で初期化
3: Target network  $Q^-$  を重み  $\theta^- = \theta$  で初期化
4: for episode = 1,  $\dots$ ,  $M$  do
5:    $T \sim U(1, l_{\text{nomax}})$ 
6:   for  $t' = 1, \dots, T$  do
7:      $a_t$  = (do nothing) の実行  $\rightarrow$  初期状態の生成
8:   end for
9:    $t = 1$ 
10:  while not done do
11:    if  $t \leq n_{\text{rpstart}}$  then
12:       $a_t$  をランダムに決定  $\rightarrow$  Replay Memory の確保
13:    else
14:      if  $(t-1)\%l_{\text{history}} = 0$  then
15:         $\epsilon$ -greedy に従って行動  $a_t$  を選択
16:      else
17:         $a_t = a_{t-1}$ 
18:      end if
19:       $\epsilon = \max(\epsilon_{\text{final}}, \epsilon - \frac{\epsilon_{\text{first}} - \epsilon_{\text{final}}}{l_{\text{expl}}}) \rightarrow \epsilon$  を線形減少
20:    end if
21:    行動  $a_t$  を実行し, 報酬  $r_t$  と次の画面  $x_{t+1}$  と done を観測
22:    前処理して次の状態  $s_{t+1}$  を生成
23:     $D$  に  $(s_t, a_t, r_t, s_{t+1}, \text{done})$  を追加,  $|D| > n_{\text{mem}}$  なら古いものを削除する.
24:    if  $t > n_{\text{rpstart}}$  then
25:      if  $(t-1)\%l_{\text{history}} = 0$  then
26:         $D$  からランダムに  $(s_j, a_j, r_j, s_{j+1}, \text{done})$  を  $n_{\text{batch}}$  個の履歴をサンプル
27:        
$$y_j = \begin{cases} r_j & (\text{done}) \\ r_j + \gamma \max_{a'} Q^-(s_{j+1}, a'; \theta^-) & (\text{otherwise}) \end{cases}$$

28:         $\theta$  を  $(y_j - Q(s_j, a_j; \theta))^2$  の勾配方向に RMSProp( $r_{\text{learn}}, m_{\text{gradient}}, m_{\text{sgradient}}, g_{\text{min}}$ ) を用いて更新 (勾配計算の後の更新の際,  $r_j$  は  $[-1, 1]$  にクリップされる)
29:      end if
30:      if  $(t-1)\%(f_{\text{tupdate}} \times l_{\text{history}}) = 0$  then
31:         $Q^- = Q$ 
32:      end if
33:    end if
34:     $t = t + 1$ 
35:  end while
36: end for
```
