INPUT: Actor current network, Actor goal network, Critic current network, respectively the Actor-network parameter $\theta, \theta^{'}$, Critic-network parameter ω, $\omega^{'}$, Attenuation factor γ, Number of samples with batch gradient descent m, Target Q network update frequency C, maximum number of iterations T, random noise \ mathcal{N}.

OUTPUT: Optimal Actor current network parameter, critical current network parameter ω.

1. Randomly initialize θ, ω, $\omega^{'} = \omega, \theta^{'} = \theta$, and clear the experience playback set.

2. Iterate from 1 to T：

2.1 Initialize S as the first state of the current state sequence and get its eigenvector $\phi(S)$

2.2 In the Actor, the current network obtains the action $A = \pi\theta\left(\phi(S)\right) + N$ based on the state S.

2.3 Execute action A to get a new status S', reward R, and judge terminate status or not.

2.4 Store the quintuple $\{\phi(S), A, R, \phi(S'), is\_end\}$ into the experience playback set D.

2.5 Let S=S':

2.6 Sample m samples $\{\phi(S), A, R, \phi(S'), is\_end_j\}$ (j = 1,2,..., m) from the empirical playback set D, and calculate the current target Q value y_j:

$$y_j = \begin{cases} R_j & is\_end_j\ is\ true \\ R_j + \gamma Q^{'}\left(\phi(S_j^{'}), \pi_{\theta^{'}}\left(\phi(s_j^{'})\right), \omega^{'}\right) & is\_end_j\ is\ false \end{cases}$$

2.7 Use mean square loss function $\frac{1}{m}\sum_{j=1}^{m}\left(y_j - Q\left(\phi(S_j), A_j, \omega\right)\right)^2$, all parameters ω of critical current network are updated by gradient back propagation of neural network.

2.8 Use $J(\theta) = -\frac{1}{m}\sum_{j=1}^{m} Q(s_i, a_i, \theta)$, all parameters θ of actor's current network are updated by gradient back propagation of neural network .

2.9 If T%C = 1, then update the critical target network and actor target network parameters:

$$\omega^{'} \leftarrow \tau\omega + (1-\tau)\omega^{'}$$
$$\theta^{'} \leftarrow \tau\theta + (1-\tau)\theta^{'}$$

If S' is terminated, the current round of iteration is completed, otherwise go to step 2.2.

The above is the main process of DDPG algorithm.

Please pay attention: in the step 2.6, $\pi_{\theta^{'}}\left(\phi(S_j^{'})\right)$ is acquired by the Actor goal network, and $Q^{'}\left(\phi(S_j^{'}), \pi_{\theta^{'}}\left(\phi(s_j^{'})\right), \omega^{'}\right)$ is acquired by the Critic goal network.