

DDPG算法

一个可以处理动作空间无限的环境并且是离散策略的算法。

由于DQN中有一步取max的操作，即在动作价值网络中输入动作输出每个动作的价值，然后取最高价值对应的动作，因此它无法处理连续情况，只能通过离散处理。DDPG则是将状态和动作拼接成一个状态动作对一起输入到网络中，然后输出这个指定的状态动作对对应的Q值，作为action网络的指导信息，即让action网络找到能够最大化这个Q值的状态动作对。这样action网络的输出就可以直接当作最优输出了，不需要再取max了。