

强化学习

state, action, reward, next state.

导弹

当前位置 返回下一个
目标位置 位置

更新
把下一个位置作为
当前位置

next-state, reward, done = step(action)

无人机

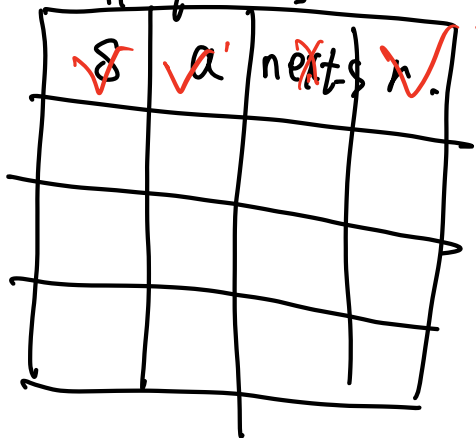
无人机位置 用户位置
用户位置 返回下一个位置

DQN

Q-learning

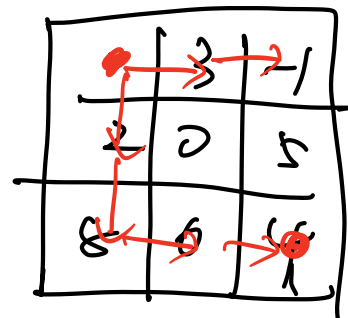
(x, y)
 (x_n, y_n)

Off-policy



格式转换

Value



Reward

$$\text{action} = f(\text{state})$$

x_k, y_k x_k, y_k
 x_k, y_k distance

matlab 数据

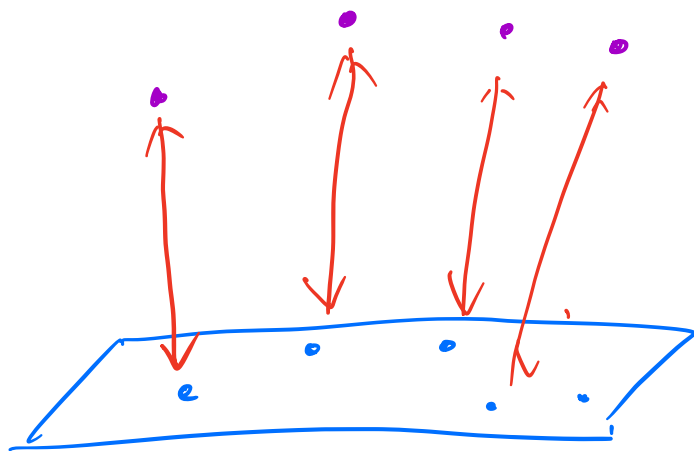
$(x, y, a, \text{orientation} \dots)$

$(\text{state}, \text{action})$
 (x_k, y_k) (x'_k, y'_k)
 (x_k, x_k)

timestep ΔT 100m/s

reward. 奖励函数.

200ms



distance < min value
reward + = 100

> min value
reward - = 100

信道容量

1. matlab 数据再处理.

2. python state, action, next_state (reward)

3. mat → python

$s, a, next-s. = f(\text{matlab 数据.})$

4. DQN ← 历史数据.
replay buffer

5. $action = f(state)$
采取的位置