

# 强化学习：作业二

黄彦骁 502023370016

October 26, 2023

## 1 作业内容

在gridworld环境中实现Q-learning算法。

## 2 实现过程

Q-learning算法伪代码如图1所示。

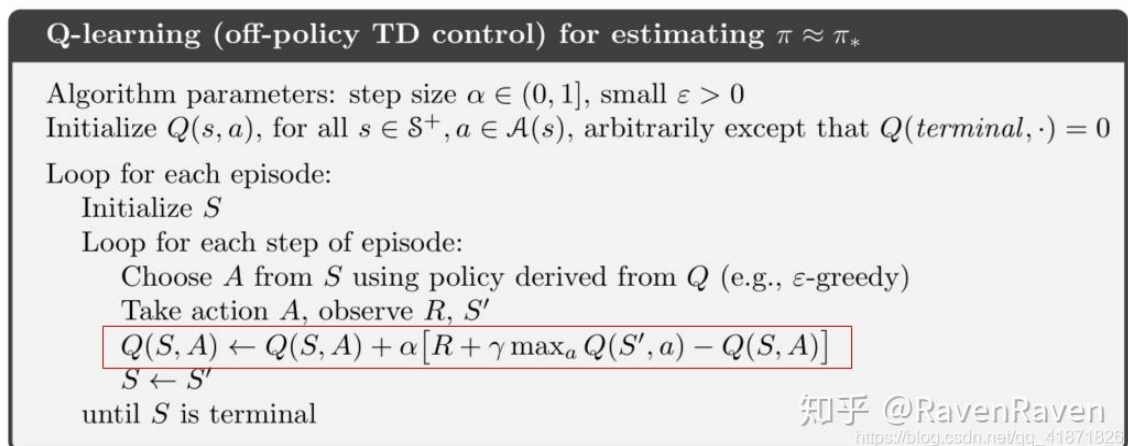


Figure 1: Q-learning算法伪代码

算法根据每次探索之后新的状态的最大可能Q值来更新当前状态的Q值，这也是Q-learning算法离线学习的本质，因此在algo.py中设计自己的Agent，初始化一个Q矩阵，根据伪代码中标红的部分不断更新Q矩阵，最后得到的Q矩阵可以通过不断最大化Q值选出一条最佳路径。同时在main.py中将新的状态赋值给当前状态之前调用Myagent.update进行Q矩阵更新。

### 3 复现方式

在code文件夹下运行 `python main.py`.

### 4 实验效果

实验的默认参数为 $\epsilon=0.1$ ,  $lr=0.01$ ,  $\gamma=0.9$ , 实验效果如图2所示

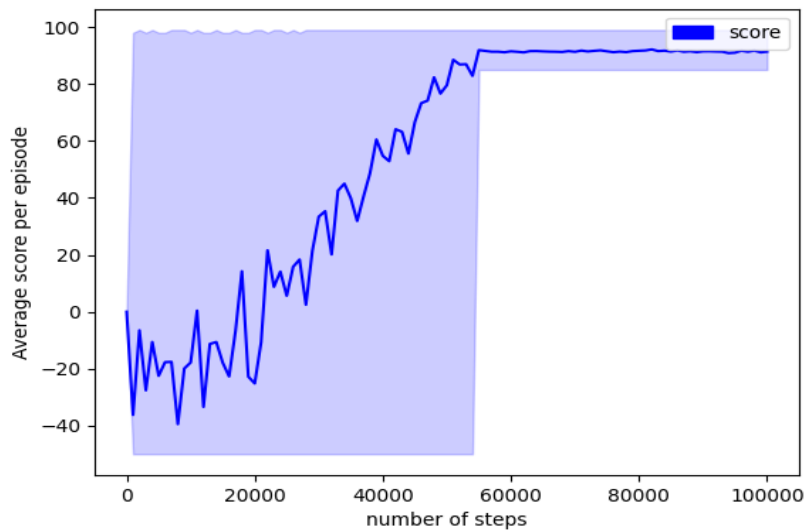


Figure 2: 默认参数效果

接下来针对不同参数对实验效果的影响做了探究，首先是学习率 ( $\epsilon = 0.1$ ,  $\gamma = 0.9$ ) 对效果的影响，如图3所示：

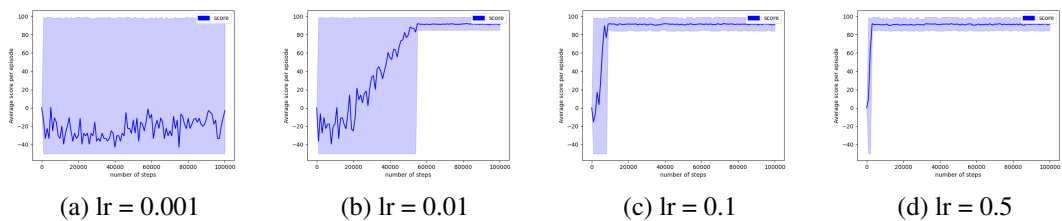


Figure 3: 学习率对实验效果影响

可以看出学习率过低会导致算法在固定步长无法收敛，而在一定范围内学习率越高算法收敛越快。

其次是Gamma的影响（ $\epsilon = 0.1$ ,  $lr = 0.01$ ），如图4所示

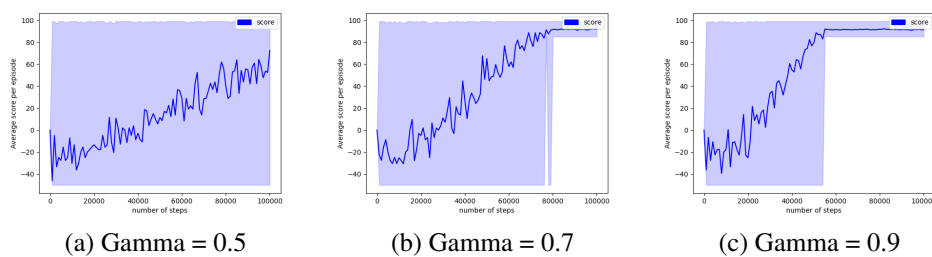


Figure 4: Gamma对实验效果影响

可以看出Gamma最好是维持在接近1的水准，这样可以使得算法更好收敛。

## 5 小结

本次实验采用Q-learning算法，实现较为简单，主要是通过该次作业能够对这种无模型的强化学习方法有一个简单的认识，同时认识到参数对于强化学习过程的影响。