蒙特卡洛算法实验报告

姓名 周雨扬 学号 200013061

## 一、实验要求

（1）阅读《基于蒙特卡洛的路径规划-实验指导书》，尝试运行并理解蒙特卡洛算法在冰湖路径规划问题上的示例代码。

（2）在示例代码的基础上，尝试在策略评估中使用不同数值的迭代次数，比较策略收敛所需的迭代次数（由于随机性建议多次实验），分析策略评估的迭代次数对总的算法效率的影响。

（3）在示例代码的基础上，尝试实现不同的epsilon衰减策略，比较策略收敛所需的迭代次数（由于随机性建议多次实验），分析不同epsilon衰减策略对算法效率的影响。

## 二、代码修改思路

仅修改了 policy\_iteration\_MC 函数中 epsilon 参数的计算方式，用于后续epsilon衰减策略的实验

## 三、实验结果和分析

### 策略评估迭代次数实验

测试采用原始代码以及原始的epsilon衰减策略，设置gamma = 0.95，重复运行 5 次取结果的平均值。结果如下：

观察发现，成功率与策略迭代次数总体会随着策略评估迭代次数的增加而增加。这是因为

策略评估迭代次数越多，得到的策略估值也就越稳定，因此能够以更少的轮数收敛到更好的结果。

但是我们也观察到 1000~5000 内，成功率增长其实较为缓慢。这可能与 epsilon 衰减策略有关。观察发现在 epsilon 较大的时候策略已经完成了收敛，因此导致收敛得到的策略实际在评估时候还没有足够的准确性，仍然有较大的噪声。

|  |  |  |  |
| --- | --- | --- | --- |
| 策略评估迭代次数 | 策略迭代次数 | 运行时间 | 成功率 |
| 500 | 48.2 | 7.02 | 0.424 |
| 1000 | 21.4 | 7.15 | 0.616 |
| 2000 | 11.8 | 6.77 | 0.571 |
| 5000 | 6.0 | 7.23 | 0.554 |
| 10000 | 5.2 | 11.23 | 0.666 |

### Epsilon 衰减策略实验

测试采用原始代码，设置系数 gamma = 0.95，策略评估迭代次数 5000，重复运行 5 次取结果的平均值。

除了原始策略外，其余衰减策略如下：

策略A： 将原始策略中 epsilon 的计算替换为如下计算方式：

Epsilon = min(eps0, eps0 / (1 + decay \* epsilon) / (threshold \*\* 0.5))

这里 threshold = num\_episodes / (nS \* nA \* log(nS) \* 25)，除数处大致为保证每一个状态至少能够被访问到 25 次的期望访问次数。如果仿真数较多，在保证每一个 (s,a) 对访问次数充足的前提下，我们便可以降低 epsilon 来提升策略评估采样的准确性。

观察发现，修改策略后我们在没有显著增加策略迭代次数的同时，能够取得较为明显更好的结果。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Epsilon 衰减策略 | 策略评估迭代次数 | 策略迭代次数 | 运行时间 | 成功率 |
| 原始策略 | 2000 | 11.8 | 6.77 | 0.571 |
| 原始策略 | 5000 | 6.0 | 7.23 | 0.554 |
| 策略A | 2000 | 14.2 | 8.01 | 0.654 |
| 策略A | 5000 | 7.8 | 9.54 | 0.627 |