# 多臂老虎机

AAIS in PKU　陈伟杰 1901111420

October 4, 2019

## 1 Problem Setting

设置 K 臂老虎机，其中 $K = 15$，15 台老虎机的平均收益为 [1,2,...,15] 的乱序排列，实际收益是以平均收益为均值，方差为 1 的独立高斯分布。尝试次数设为 $T$=1000.

## 2 Upper Confidential Bounder

简单给出 UCB 的搜索方法，其中 $Q_t(a)$ 与贪婪法相同

$$A_t \doteq \arg\max_a \left[ Q_t(a) + c\sqrt{\frac{\log t}{N_t(a)}} \right] \tag{1}$$

固定随机种子之后对不同的探索方法进行测试，其中 0 表示普通贪婪法，0.1 表示 $\epsilon = 0.1$ 的 $\epsilon$ 贪婪法，2 表示 $c = 2$ 的 UCB 搜索方法

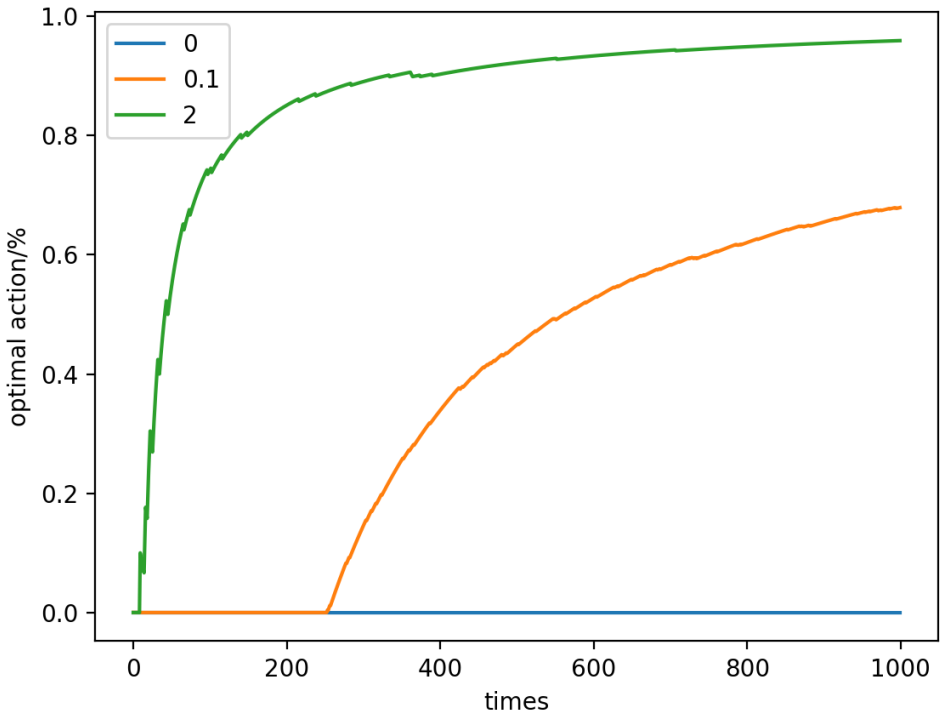


Figure 1: 最优行动率与不同搜索方法的关系

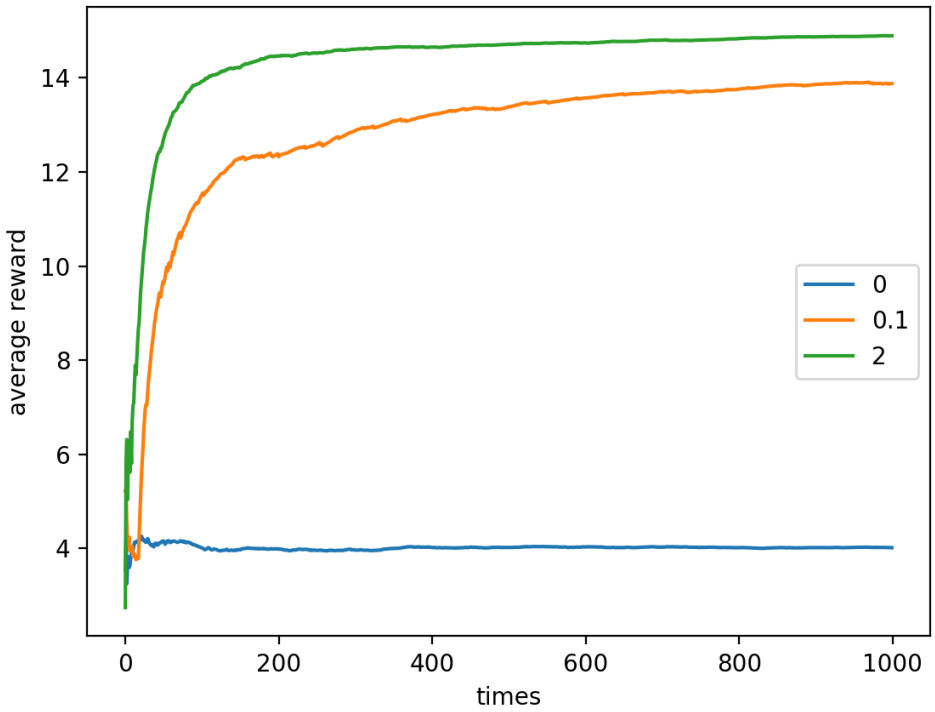


Figure 2: 平均收益与不同搜索方法的关系

可以明显看出，UCB 方法在最优行动率和平均收益上均比 $\epsilon$ 贪婪法效果好，这是因为在 UCB 中会优先遍历所有的老虎机，从而更快找出最优的行动；而 $\epsilon$ 贪婪法则是随机遍历老虎机，其遍历所需的期望步数远大于 UCB 方法。

# 3 README

代码包括 MAB.py 和 test2.py 两个程序，需要 python3 环境。直接运行 test2.py 可以得到上述结果。