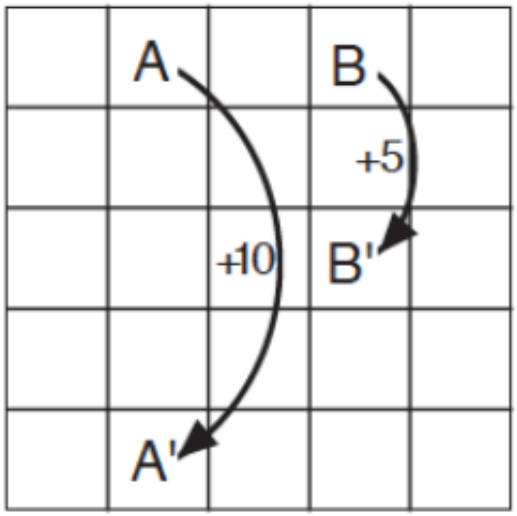
作业报告——GridWorld

方鸿宇 2001213098

问题描述

本算法实现了对GridWorld问题中价值表的学习。该问题设定了一个5x5的网格，在网格的每个位置，智能体以相同的概率向上、下、左、右四个方向跳转。每当跳转到网格外的点时，智能体获得-1的奖励，并停留在原位置；每当智能体位于A点时，无论采取任何动作，智能体均跳转到A'点，并获得+10的奖励；每当智能体位于B点时，无论采取任何动作，智能体均跳转到B'点，并获得+5的奖励；其余情况下智能体执行动作后的奖励均为0。下图展示了本实验中的网格。



算法说明

在每轮迭代中，按照从上到下、从左到右的顺序依次对每个格点的价值进行更新，更新方式依照贝尔曼方程进行：

$$V(s) = \mathbb{E}_a[r(s,a) + \gamma V(s')].$$

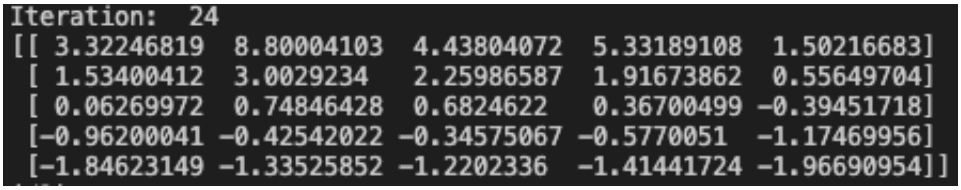
每轮迭代中对每个格点价值变化的平方进行求和，若和小于或等于1e-4，则结束迭代，输出价值表。

实验设置

实验中设置参数γ为0.9。

实验结果

实验中共进行了24次迭代，最终输出的价值表如下。



```
Iteration:  24
[[ 3.32246819  8.80004103  4.43804072  5.33189108  1.50216683]
 [ 1.53400412  3.0029234   2.25986587  1.91673862  0.55649704]
 [ 0.06269972  0.74846428  0.6824622   0.36700499 -0.39451718]
 [-0.96200041 -0.42542022 -0.34575067 -0.5770051  -1.17469956]
 [-1.84623149 -1.33525852 -1.2202336  -1.41441724 -1.96690954]]
```

代码说明

代码见`gridworld.py`，使用python3运行，需安装numpy。