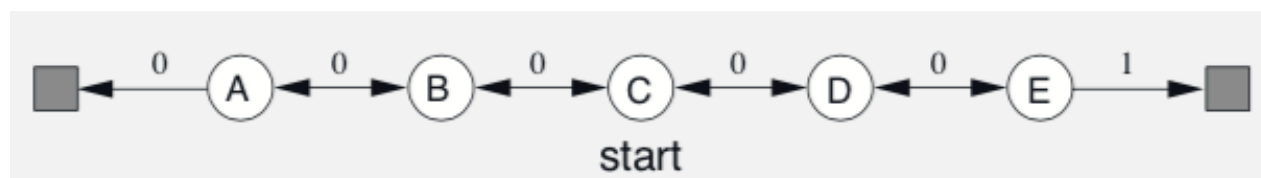


作业报告——Random-Walk

方鸿宇 2001213098

问题描述

本实验使用蒙特卡洛算法和TD(0)算法实现了对Random Walk过程中各状态价值的评估。该过程的状态转移如下图所示，每个episode从状态C出发，可向左或向右转移，概率均为50%。最左及最右的深色方块表示终止状态，到达时过程结束。当进入最右的状态时，获得reward为1，其他情况reward为0。设定折扣因子为1，求A、B、C、D、E各状态的价值。



算法说明

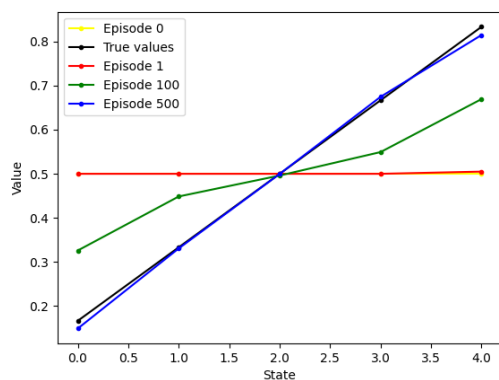
本实验分别使用常数 α 蒙特卡洛(constant- α MC)算法（以下记为MC）和TD(0)算法（以下记为TD）实现了对上述问题中价值表的评估，其中每步价值更新的步长为 α 。

实验设置

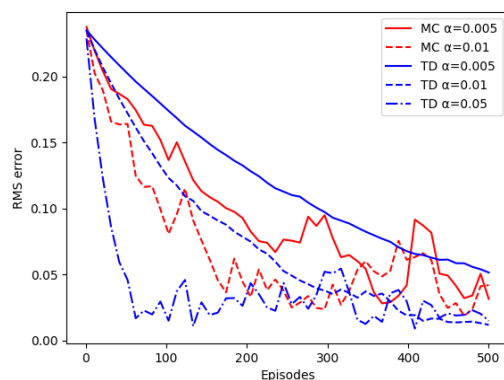
对MC算法，实验中设置 α 为0.005、0.01。对TD算法，实验中设置 α 为0.005、0.01、0.05。每组实验进行500个episode。

实验结果

下图展示了本实验的实验结果。(a)图展示了TD实验中各episode的价值表，该实验中 α 设为0.01，黑线为价值表的理论数值。可看到随着episode的增加，算法逐渐收敛到理论值。(b)为各实验的RMS误差，可见相比MC算法，TD算法的收敛过程更加稳定，且最终收敛时的RMS误差更低。



(a)TD实验($\alpha=0.01$)中各episode价值表



(b)RMS误差变化曲线

代码说明

本实验程序基于python3实现，代码文件包括 `random_walk.py` 和 `visualize.py`，前者为实验代码，后者为可视化代码，实验结果保存在 `log` 文件夹。