人工智能原理

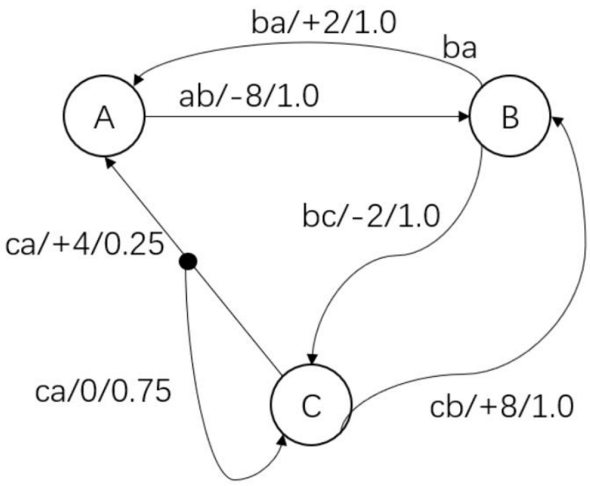
作业 8

注意：

1）请在网络学堂提交电子版；

2）请在 6 月 16 日晚 23:59:59 前提交作业，不接受补交；

1. **价值迭代**：考虑如下图所示的马尔可夫决策过程，折现因子$\gamma = 0.3$。图中大写字母表示状态；状态之间的有向边表示转移；边上的三元组“actions/rewards/probability”给出了动作、回报及转移概率。



现有均匀随机策略$\pi_1(a|s)$，即从一个状态s出发，等概率地选择下一个动作。假设有初始状态值$V_1(a) = V_1(b) = V_1(c) = 2$，请给出：

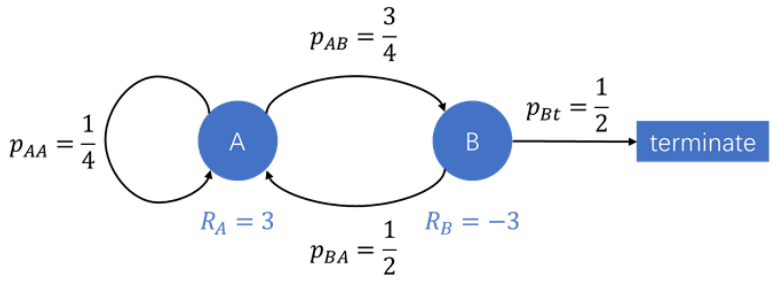
a) 通过同步迭代和确定性贪心策略得到的策略$\pi_2(a|s)$。

b) 通过异步迭代和确定性贪心策略得到的策略$\pi_2'(a|s)$

说明：在下图所有的 action 中，ca 较为特殊，它以 1/4 的概率从状态 C 转移到 A，以 3/4 的概率保持状态 C 不变，保持不变时回报为 0。

2. **蒙特卡洛**：一个无折现($\gamma = 1$)的马尔可夫回报过程，具有 A 和 B 两个状态以及一个终止状态

a) 若状态转移图和回报函数如下图所示，请写出该马尔可夫回报过程的状态价值贝尔曼期望方程，并求解该方程得出状态价值函数$v(A), v(B)$。



b) 若状态转移图及回报函数未知，但已知以下两个观测片段

$$\mathrm{A} \xrightarrow{+2} \mathrm{A} \xrightarrow{+3} \mathrm{B} \xrightarrow{-5} \mathrm{A} \xrightarrow{+5} \mathrm{B} \xrightarrow{-2} \text{terminate}$$
$$\mathrm{B} \xrightarrow{-2} \mathrm{A} \xrightarrow{+3} \mathrm{B} \xrightarrow{-3} \text{terminate}$$

其中$\mathrm{A} \xrightarrow{+2} \mathrm{A}$表示以回报值+2 从 A 状态转移到 A 状态。请分别使用首次访问和每次访问的蒙特卡洛预测，估计状态价值函数$v(A), v(B)$。