

GridWorld 网格世界问题

0.0	A	0.0	B	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	B _{TO}	0.0
0.0	0.0	0.0	0.0	0.0
0.0	A _{TO}	0.0	0.0	0.0

【规则】 网格中的每一个小格都对应于环境中的状态. 在一个小格上, 有 4 种可能的动作: 上移/下移/左移/右移, 其中各个动作都**确定性地**使智能体在网格上沿对应的方向移动一格. 如果所采取的动作将令智能体超出网格范围, 那么该动作的结果为智能体的位置保持不变, 且造成 -1 的回报. 除了上述动作与将智能体移出特殊状态 A 与 B 的动作外, 其他的动作只会造成 0 的回报. 在状态 A 上, 所有的 4 个动作会产生 A_Reward 的回报, 并将智能体带至 A_TO. 在状态 B 上, 所有的 4 个动作会产生 B_Reward 的回报, 并将智能体带至 B_TO.

【输入】

A 点坐标 A_TO 点坐标 A_Reward

B 点坐标 B_TO 点坐标 B_Reward

示例:

[0,1] [3,1] 10.0

[0,2] [2,3] 5.0

【输出】

值迭代和策略迭代的值函数估计, 以矩阵形式表示

示例:

Value Iteration

[26.2, 29.1, 22.2, 20.0, 18.0]

[23.6, 26.2, 23.6, 21.2, 19.1]

[21.2, 23.6, 21.2, 19.1, 17.2]

[19.1, 21.2, 19.1, 17.2, 15.5]

[17.2, 19.1, 17.2, 15.5, 13.9]

Policy Iteration

[26.2, 29.1, 22.2, 20.0, 18.0]

[23.6, 26.2, 23.6, 21.2, 19.1]

[21.2, 23.6, 21.2, 19.1, 17.2]

[19.1, 21.2, 19.1, 17.2, 15.5]

[17.2, 19.1, 17.2, 15.5, 13.9]