



华中农业大学

HUAZHONG AGRICULTURAL UNIVERSITY
Wuhan 430070, Hubei, P.R. China 中国·武汉 http://www.hzau.edu.cn 传真: 027-87384670

基础公式 $V_{\pi}(s) = \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r|s, a) [r + \gamma V_{\pi}(s')]$ $s \in S$

$$V_{\pi}(s_A) = 0.25(5 + 0.7V_{\pi}(s_B)) + 0.25(0 + 0.7V_{\pi}(s_C)) + 0.5(0 + 0.7V_{\pi}(s_A))$$

$$V_{\pi}(s_B) = 0.5(5 + 0.7V_{\pi}(s_B)) + 0.25(0 + 0.7V_{\pi}(s_A)) + 0.25(0 + 0.7V_{\pi}(s_D))$$

$$V_{\pi}(s_C) = 0.25(0 + 0.7V_{\pi}(s_A)) + 0.25(0 + 0.7V_{\pi}(s_D)) + 0.5(0 + 0.7V_{\pi}(s_C))$$

$$V_{\pi}(s_D) = 0.25(5 + 0.7V_{\pi}(s_B)) + 0.25(0 + 0.7V_{\pi}(s_C)) + 0.5(0 + 0.7V_{\pi}(s_D))$$

简化计算令 $V_{\pi}(s_A) = A, V_{\pi}(s_B) = B, V_{\pi}(s_C) = C, V_{\pi}(s_D) = D$

$$\begin{cases} A = 0.25(5 + 0.7B) + 0.25(0 + 0.7C) + 0.5(0 + 0.7A) & 1 - 0.35 = -0.65 \\ B = 0.5(5 + 0.7B) + 0.25(0 + 0.7A) + 0.25(0 + 0.7D) & 0.25 \times 0.7 = 0.175 \\ C = 0.25(0 + 0.7A) + 0.25(0 + 0.7D) + 0.5(0 + 0.7C) & 0.25 \times 0.5 = 0.125 \\ D = 0.25(5 + 0.7B) + 0.25(0 + 0.7C) + 0.5(0 + 0.7D) \end{cases}$$

$$\Rightarrow \begin{cases} -0.65A + 0.175B + 0.125C + 0D + 0.125 = 0 \\ 0.175A + (-0.65)B + 0C + 0.175D + 0.25 = 0 \\ 0.125A + 0B + (-0.65)C + 0.125D + 0 = 0 \\ 0A + 0.175B + 0.125C + (-0.65)D + 1.25 = 0 \end{cases} \Rightarrow Ax = b$$

$$\begin{pmatrix} -0.65 & 0.175 & 0.125 & 0 \\ 0.175 & -0.65 & 0 & 0.175 \\ 0.125 & 0 & -0.65 & 0.125 \\ 0 & 0.175 & 0.125 & -0.65 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -1.25 \\ -2.5 \\ 0 \\ -1.25 \end{pmatrix}_{4 \times 1}$$

$$\Rightarrow \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} -0.65 & 0.175 & 0.125 & 0 \\ 0.175 & -0.65 & 0 & 0.175 \\ 0.125 & 0 & -0.65 & 0.125 \\ 0 & 0.175 & 0.125 & -0.65 \end{pmatrix}^{-1} \begin{pmatrix} -1.25 \\ -2.5 \\ 0 \\ -1.25 \end{pmatrix}$$

$$= \begin{pmatrix} -1.85 & -0.58 & -0.58 & -0.31 \\ -0.58 & -1.85 & -0.31 & -0.58 \\ -0.58 & -0.31 & -1.85 & -0.58 \\ -0.31 & -0.58 & -0.58 & -1.85 \end{pmatrix} \begin{pmatrix} -1.25 \\ -2.5 \\ 0 \\ -1.25 \end{pmatrix}$$

$$= \begin{pmatrix} 4.15 \\ 6.075 \\ 2.225 \\ 4.15 \end{pmatrix}$$

$$\therefore V_{\pi}(s_A) = 4.15$$

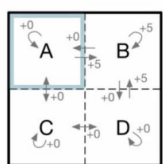
$$V_{\pi}(s_B) = 6.075$$

$$V_{\pi}(s_C) = 2.225$$

$$V_{\pi}(s_D) = 4.15$$

例题：一个简单有限马尔可夫决策过程

$$v_{\pi}(s) = \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r|s, a) [r + \gamma v_{\pi}(s')] \text{ for all } s \in S$$



π 25% 25% 25% 25%

- 状态空间: $S = \{s_A, s_B, s_C, s_D\}$
- $\gamma = 0.7$, 奖励如图所示
- 问题: 求解状态价值函数 (即计算每个状态对应的价值)
- 解法: 解决线性方程组即可 (n个未知量, n个线性方程组)

$$\begin{cases} v_{\pi}(s_A) = 0.25[5 + 0.7v_{\pi}(s_B)] + 0.25[0 + 0.7v_{\pi}(s_C)] + 0.5[0 + 0.7v_{\pi}(s_A)] \\ v_{\pi}(s_B) = 0.5[5 + 0.7v_{\pi}(s_B)] + 0.25[0 + 0.7v_{\pi}(s_A)] + 0.25[0 + 0.7v_{\pi}(s_D)] \\ v_{\pi}(s_C) = 0.25[0 + 0.7v_{\pi}(s_A)] + 0.25[0 + 0.7v_{\pi}(s_D)] + 0.5[0 + 0.7v_{\pi}(s_C)] \\ v_{\pi}(s) = \sum_a \pi(a|s) \sum_{s'} \sum_r p(s', r|s, a) [r + \gamma v_{\pi}(s')] \end{cases}$$

$$\pi, p, \gamma \xrightarrow{\text{Linear System Solver}} v_{\pi}$$
$$\begin{cases} v_{\pi}(s_C) = 2.2 \\ v_{\pi}(s_D) = 4.2 \end{cases}$$

卢东黎 2021317220603

作业3

□ 题目: 智能体与环境进行一个回合, 获得如下奖励序列 ($H = 5$),

$$R_1 = -1, R_2 = 2, R_3 = 6, R_4 = 3, R_5 = 2$$

□ 问题: $\gamma = 0.5$, 计算 $G_0, G_1, G_2, \dots, G_5$ (提示: 从后往前计算)

$$G_t = R_{t+1} + \gamma G_{t+1}$$

$$G_5 = 0 \text{ 代入 } G_4 = R_5 + \gamma G_5 = 2$$

$$G_4 = 2 \text{ 代入 } G_3 = R_4 + \gamma G_4 = 4$$

$$G_3 = 4 \text{ 代入 } G_2 = R_3 + \gamma G_3 = 8$$

$$G_2 = 8 \text{ 代入 } G_1 = R_2 + \gamma G_2 = 6$$

$$G_1 = 6 \text{ 代入 } G_0 = R_1 + \gamma G_1 = 2$$