

$$1.(1) V_{k+1}(s) = \max_{a \in A} \{ r_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_k(s') \}$$

$$A: V_2(A) = -8 + 0.5 \times V_1(B) = -6$$

$$B: V_2(B) = \max \{ 2 + 0.5 V_1(A), -2 + 0.5 V_1(C) \} = 4$$

$$C: V_2(C) = \max \{ 0.75 \times 0 + 0.25 \times 4 + 0.5 \times (0.75 \times V_1(C) + 0.25 \times V_1(A)), 8 + 0.5 \times V_1(B) \} = \max \{ 3, 10 \} = 10$$

$$\pi_2(a|s) = \operatorname{argmax} \{ r_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_2(s') \}$$

$$\pi_2(a|A) = ab$$

$$\pi_2(a|B) = \operatorname{argmax} \{ 2 + 0.5 V_2(A), -2 + 0.5 V_2(C) \} = \operatorname{argmax} \{ -1, 3 \} = bc$$

$$\begin{aligned} \pi_2(a|C) &= \operatorname{argmax} \{ 0.75 \times 0 + 0.25 \times 4 + 0.5 \times (0.75 \times V_2(C) + 0.25 \times V_2(A)), 8 + 0.5 \times V_2(B) \} \\ &= \operatorname{argmax} \{ 4, 10 \} = cb \end{aligned}$$

$$(2) A: V_2(A) = -8 + 0.5 V_1(B) = -6$$

$$B: V_2(B) = \max \{ 2 + 0.5 V_2(A), -2 + 0.5 V_1(C) \} = \max \{ -1, 0 \} = 0$$

$$\begin{aligned} C: V_2(C) &= \max \{ 0.75 \times 0 + 0.25 \times 4 + 0.5 \times (0.75 \times V_1(C) + 0.25 \times V_2(A)), 8 + 0.5 \times V_2(B) \} \\ &= \max \{ 1.75, 8 \} = 8 \end{aligned}$$

$$\pi'_2(a|A) = ab$$

$$\pi'_2(a|B) = \operatorname{argmax} \{ 2 + 0.5 V_2(A), -2 + 0.5 V_2(C) \} = \operatorname{argmax} \{ -1, 2 \} = bc$$

$$\begin{aligned} \pi'_2(a|C) &= \operatorname{argmax} \{ 0.75 \times 0 + 0.25 \times 4 + 0.5 \times (0.75 \times V_2(C) + 0.25 \times V_2(A)), 8 + 0.5 \times V_2(B) \} \\ &= \operatorname{argmax} \{ 3.25, 8 \} = cb \end{aligned}$$

$$2.(1) V(A) = r(A) + \gamma (P_{AA} V(A) + P_{AB} V(B))$$

$$V(B) = r(B) + \gamma (P_{BA} V(A) + P_{Bt} V(t))$$

$$V(t) = 0 \quad \gamma = 1$$

$$\Rightarrow \begin{cases} V(A) = 3 + \frac{1}{4} V(A) + \frac{3}{4} V(B) \\ V(B) = -3 + \frac{1}{2} V(A) + \frac{1}{2} \times 0 \end{cases} \Rightarrow \begin{cases} V(A) = 2 \\ V(B) = -2 \end{cases}$$

$$(2) \text{首次访问: 片段1: } A_0: \text{return} = 3 + 2 - 4 + 4 - 3 = 2$$

$$B_2: \text{return} = -4 + 4 - 3 = -3$$

$$\text{片段2: } B_0: \text{return} = -2 + 3 - 3 = -2$$

$$A_1: \text{return} = 3 - 3 = 0$$

$$V(A) = \frac{2+0}{2} = 1$$

$$V(B) = \frac{-3-2}{2} = -2.5$$

每次访问：片段1： $A_0 = 3+2-4+4-3=2$

$$A_1 = 2-4+4-3=-1$$

$$A_3 = 4-3=1$$

$$B_2 = -4+4-3=-3$$

$$B_4 = -3$$

片段2： $B_0 = -2+3-3=-2$

$$B_2 = -3$$

$$A_1 = 3-3=0$$

$$V(A) = \frac{2-1+1+0}{4} = 0.5$$

$$V(B) = \frac{-3-3-2-3}{4} = -2.75$$