

Home task (written)

1. Efficient Routing MDP

1	2	3	4	5	6
1	7	13	19	25	31
2	8	14	20	26	32
3	9	15	21	27	33
4	10	16	22	28	34
5	11	17	23	29	35
6	12	18	24	30	36

$$\gamma = 0.9$$

$$r_g = +5$$

$$r_2 = -5$$

$$r_s$$

$$G = R$$

$$V_{k+1}(s) = \mathbb{E}_\pi [G_t | s=s]$$

$$4.9$$

$$V_{k+1}(s) = \max_a [R(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V_k(s')]$$

$$\pi_{k+1} = \arg \max_a$$

a) $r_s \in \{-5, -0.5, 0, 2\}$

starting: square 2

$\gamma_s = -5$. Розглянемо з початку: $V_{\Pi}(s) = \mathbb{E}_{\Pi} [G_t^c / S_t = s]$

$\Pi_1(2) = 7$; $\Pi_2(2) = 9$, $\Pi_2(9) = 16, \dots, \Pi_2(26) = 33$

$V_{\Pi_1}(2) = -5$

$$V_{\Pi_2}(2) = -5 + \gamma \cdot (-5) + \gamma^2 \cdot (-5) + \gamma^3 \cdot (-5) + \gamma^4 \cdot 5 = -5 \cdot \sum_{k=0}^3 \gamma^k +$$

$$+ \gamma^4 \cdot 5 = -5 \frac{1-\gamma^5}{1-\gamma} + \gamma^4 \cdot 5 = 5 \cdot \frac{\gamma^4 - \gamma^5 + \gamma^3 \cdot \gamma}{1-\gamma}$$

$\forall \gamma \in (0, 1) : 5 \cdot \frac{\gamma^4 - \gamma^5 + \gamma^3 \cdot \gamma}{1-\gamma} < -5$

отже, з цих двох оптимізації Π_1 не єдина, Π_2 є *big gear* з γ .

Далі з Π_3 , якою не бере до $s=33$, то має

две інших $s \in S$ Π_1 не єдина, наприклад, однією з оптимального буде:

$\Pi_1(28) = 33$ $\Pi_1(21) = 26$ та $\Pi_1(21) = 28$ мають однакові $V_{\Pi_1}(21) = -5 + \gamma \cdot 5$

але це є проблема, адже за умови Π_1 не можна

не дочекати $s = 21$.

$\gamma_s = -0,5$. Оптимальна початок:

$$\Pi(26) = \Pi(28) = 33 \quad V_{\Pi}(26) = V_{\Pi}(28) = 5$$

$$\Pi(21) = 26 / 28 \quad V_{\Pi}(21) = -0,5 + \gamma \cdot 5$$

$$\Pi(16) = 21, \text{ отже} \quad V_{\Pi}(16 \rightarrow 21) = -0,5 - 0,5 \cdot \gamma + \gamma^2 \cdot 5 \quad \forall \gamma \in (0, 1)$$

$$\Pi(g) = 16, \text{ отже} \quad V_{\Pi}(g \rightarrow 16) = -0,5 - 0,5 \cdot \gamma - 0,5 \cdot \gamma^2 + \gamma^3 \cdot 5 \quad \forall \gamma \in (0, 1)$$

$$\Pi(2) = 9, \quad V_{\Pi}(2 \rightarrow g) = -0,5 - 0,5 \cdot \gamma - 0,5 \cdot \gamma^2 - 0,5 \cdot \gamma^3 + \gamma^4 \cdot 5 \quad \forall \gamma \in (0, 1)$$

така не єдина, хоча думка $\Pi(21) = 26$ або $\Pi(21) = 28$ і не дочекати *big* γ

$\gamma_s = 0$. При такому значенні дочеканням можна до $s=33$, проміжні будуть багато недобреагуючих кроків. Але ритмік не зможе викликати *насікоротких* чиєх. Проте, оскільки ми стартуємо з $s=2$, то це не має ніякого значення.

Для кінця $2 \rightarrow 9 \rightarrow 16 \rightarrow 21 \rightarrow 26/28 \rightarrow 33$ початок маємо, як у попередньому.

$\gamma_s = 2$ початок піднімання до довших маршрутив.

Але знову, при стартовій починці 2 є *єдиний* (тобто два симетрических) чиєх, де ми одразу пішли *першого*; другого кінця немає. Початок (оптимізація) не єдина і не дочекати *big* γ .

b) найкоротшіших породити $\gamma_s = -0,5$

$\gamma_s = -5$ уялай не даст такого чимеху, але $\gamma_s = 0$ буде неважливим і-но кінцевих, тому найкоротшіших чимех не гарантується а $\gamma_s = 2$ гарантує чимех що ми дослідимо добре, існіть наступна висноворогу за доказаної вище кінцевки.

Отже, обираємо $\gamma_s = -0,5$

$V_{\pi}(32) = -5$ (будь-які гії будуть впору на одну кінцевку - підміто в 31)

$V_{\pi}(21) = -0,5 + \gamma \cdot 5$

$V_{\pi}(13) = -5$

$V_{\pi}(2) = -0,5 + \gamma \cdot (-0,5) + \gamma^2 \cdot (-0,5) + \gamma^3 \cdot (-0,5) + \gamma^4 \cdot 5$

c) $e = \{\rightarrow, \downarrow\}$; $r_e = \{-5, -0,5; 0; 2\}$

start: state 2.

$\delta \beta = 2$

$\gamma_e = -5$. тут оптимальна поліс $\sqrt{\text{меншіше (більше)}}$ до верхніх кінцевок, оскільки до $s=33$ дадуть добрий чимех. Поганаком з кінцевими 2 можна приступити в більше до 14 і отримати висноворогу $-20, \sqrt{\text{меншіше}}$ β^{10} більше. Така поганаком з 4 стовбура, чи вже можна отримати чимех до 33. Така поганаки не сума. $(26 \rightarrow 32 \rightarrow 33 / 26 \rightarrow 27 \rightarrow 33)$

$\gamma_e = -0,5$. оптимальна поганаки дають найкоротшіших чимех до 33. Добрий чимех дратуємо висноворогу, а кілька разів в верхніх кінцевих дратуємо, але не зменшує висноворогу. Таку оптимальну поганаки не сума, та замінити $\beta \gamma$ з, але врешті може привести до 35.

Така поганаки не сума (можна вже $2 \rightarrow 8 \rightarrow 9$, а можна $2 \rightarrow 3 \rightarrow 9$) і не замінить $\beta \gamma$.

$\gamma_e = 0$. Тут все можна брати дуже добрий чимех, і не не побачивши на висноворогу. Таку оптимальну поганаки не сума, та замінити $\beta \gamma$, але врешті може привести до 35.

d) $\gamma_s = 2$. Альтернативи буде меншіше до добрих чимех, більше, з поганаки; $s=2$ є відносно меншою чимех, і є більшістю найкоротшими. Поганаки оптимальна, чимехи 2, але з однаковою висноворогами). $\beta \gamma$.

d) $\gamma_s = 0$. optimal path from 2 to 33 using only efficient actions (\uparrow, \downarrow) is strictly more rewarding than the optimal path using only inefficient actions ($\{\uparrow, \downarrow\}$)

Optimal path using only γ_s : $2 \rightarrow 9 \rightarrow 16 \rightarrow 21 \rightarrow \frac{26}{28} \rightarrow 33$

reward: 5 (given γ_s)

optimal path using only e : $2 \rightarrow \frac{3}{8} \rightarrow 9 \rightarrow 15 \rightarrow 21 \rightarrow 27 \rightarrow 33$

reward $5r_e + 5$

$$5 < 5r_e + 5 \Rightarrow r_e > 0.$$

e) efficient

32
26
27
20
21
15
8
9
2
3

inefficient

34
26
28
27
20
21
22
15
16
17
8
9
10
2
3
4
5

Одно, efficient & inefficient

некоторые решения -
是最好的, & других
не имеют значения
33 единиц. \rightarrow , 13
больше не имеет
big 33, a броу
параметров не имеет.

f) Assume horizon is infinite (no termination)

$r_{new} = c + \gamma r_{old}$ \rightarrow can change the optimal policy
for mdp?

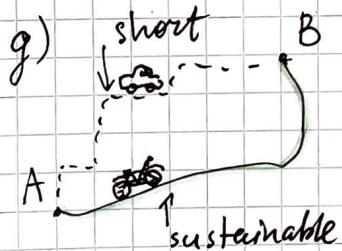
Так, например, якою $r_{old} < 0$, а $c > |r_{old}|$, то

$r_{new} > 0$. Тоги же новых интересов оптимизация
наши будущие действия лучше, в том числе

как r_{old} будущая короткий.

Например, якою $r_g = 5$, $r_s = -5$, $r_e = -5$

важен ли коэффициент $c = 5$: $r_g = 10$, $r_s = 0$, $r_e = 0$



здесь, оправдание учебника
картины из Google Maps

изображено на карте мониторинга
и показывает обработка (или действие)
варианты наименее плохое

"sustainable". Таким образом
доказано в наименее важных моментах

зубами или же "лучшими" случаях, это правило

Например, якою и нее не имеет значения
но и для более высоких значений это опять, это
якою имеет значение, это правило.

2. Value Iteration theorem

$$(BV)(s) = \max_a [R(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) \cdot V(s')]$$

$0 \leq \gamma \leq 1, \|V\| = \|V'\|_\infty$

$$(B_\pi V)(s) = \mathbb{E}_{a \sim \pi} [R(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V(s')]$$

a) Prove B_π is a contraction mapping:

$$\|B_\pi V - B_\pi V'\| \leq \gamma \|V - V'\|$$

$$\begin{aligned} \|B_\pi V - B_\pi V'\| &= \|R^\pi(s) + \gamma \sum_{s' \in S} P^\pi(s'|s) V(s') - \\ &\quad - R^\pi(s) - \gamma \sum_{s' \in S} P^\pi(s'|s) V'(s')\| = \|\gamma \sum_{s' \in S} P^\pi(s'|s) [V(s') - \\ &\quad - V'(s')]\| = \|\gamma \sum_{s' \in S} P^\pi(s'|s) [V(s') - V'(s')]\| = \\ &= \gamma \cdot \max_s \left| \sum_{s' \in S} P^\pi(s'|s) \cdot [V(s') - V'(s')] \right| \leq \\ &\leq \gamma \cdot \max_s \left| \sum_{s' \in S} P^\pi(s'|s) \cdot \|V - V'\| \right| = \gamma \cdot \|V - V'\| \times \\ &\times \max_s \left| \sum_{s' \in S} P^\pi(s'|s) \right| = \gamma \cdot \|V - V'\| \cdot \max_s \left| \sum_{s' \in S} \sum_{a \in A} \pi(a|s) \times \right. \\ &\quad \left. \times p(s'|s, a) \right| \leq \gamma \cdot \|V - V'\| \cdot \sum_{a \in A} \max_s |P^\pi(a|s)| \\ &\times \underbrace{\max_s \left| \sum_{a \in A} \pi(a|s) \sum_{s' \in S} p(s'|s, a) \right|}_{\text{we gaan nu big } s} \leq \gamma \cdot \|V - V'\|. \end{aligned}$$

b) Hexaei V' ma V^* -pizre qikobari π -oneparox B^π

$$\text{mogi } \|V^* - V'\| = \|B_\pi V^* - B_\pi V'\| \leq \gamma \|V^* - V'\|$$

$$\|V^* - V'\| \leq \gamma \|V^* - V'\|, \gamma < 1$$

$$\|V^* - V'\| \leq 0 \Leftrightarrow \|V^* - V'\| = 0 \Rightarrow V^* = V'$$

$$B_\pi V^* = V^*$$

$$\mathbb{E}_{a \sim \pi} [R(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V^*(s')] = V^*(s)$$

$$\begin{pmatrix} V^*(s_1) \\ \vdots \\ V^*(s_N) \end{pmatrix} = \begin{pmatrix} R^\pi(s_1) \\ \vdots \\ R^\pi(s_N) \end{pmatrix} + \gamma \begin{pmatrix} P^\pi(s_1|s_1) \cdot P^\pi(s_2|s_1) & V^*(s_1) \\ \vdots & \vdots \\ P^\pi(s_N|s_1) \cdot P^\pi(s_N|s_N) & V^*(s_N) \end{pmatrix}$$

$$V^* = R^\pi + \gamma P^\pi \cdot V^*$$

$$V^* = (I - \gamma P^\pi)^{-1} R^\pi.$$

c) consider greedy policy $\pi(s) = \operatorname{argmax}_a [\gamma \sum_{s' \in S} p(s'|s, a) V(s')]$

let on step k be generated

$$V_{k+1}(s) = \max_a [\gamma \sum_{s' \in S} p(s'|s, a) V_k(s')]$$

therefore $\pi_{k+1}(s) = \operatorname{argmax}_a [\dots]$

$$V^{\pi_{k+1}}(s) = \mathbb{E}_{a \sim \pi_{k+1}} [\dots] = r(s, \pi_{k+1}(s)) + \gamma \sum_{s' \in S} p(s'|s, \pi_{k+1}(s)) \times V^{\pi_k}(s)$$

π is determined

prove by induction

$$k=0: V_1(s) = \max_a [r(s, a)]$$

$$\pi_1(s) = \operatorname{argmax}_a [r(s, a)]$$

$$V^{\pi_1}(s) = r(s, \pi_1(s)) = V_1(s)$$

↑ maximizes $r(s, \cdot)$

induction step:

$$\text{let } V_k(s) = V^{\pi_k}(s)$$

$$\text{then for } k+1: V_{k+1}(s) = \max_a [r(s, a) + \gamma \sum_{s' \in S} p(s'|s, a) V_k(s')]$$

$$\pi = \operatorname{argmax}_{\pi_{k+1}} (\dots)$$

$$V^{\pi_{k+1}}(s) = r(s, \pi_{k+1}(s)) + \gamma \sum_{s' \in S} p(s'|s, \pi_{k+1}(s)) \cdot V_k(s')$$

or we, $\pi_{k+1}(s)$ maximizes also $r(s, \cdot)$, $V_k(s')$

$$\therefore V_{k+1}(s) = V^{\pi_{k+1}}(s) \quad \forall k > 0.$$

d) π -greedy. $\# V$:

$$BV = \max_a [\gamma \sum_{s' \in S} p(s'|s, a) V(s')]$$

$$B^\pi V = \mathbb{E}_{a \sim \pi} [\gamma \sum_{s' \in S} p(s'|s, a) V(s')] \Leftrightarrow$$

$\pi = \operatorname{argmax}_a [\dots]$ - by definition

newly $p_{\text{new}}(\pi(a|s))$ byge

$$(0, 0, \dots, 1, \dots, 0)$$

to which $\operatorname{argmax}_a [\dots]$

$$\Leftrightarrow r(s, \pi(s)) + \gamma \sum_{s' \in S} p(s'|s, \pi(s)) V(s') = BV.$$

$\downarrow \operatorname{argmax}$ $\downarrow \operatorname{argmax}$

e) let V_n and V_{n+1} - outputs of value iteration at the n^{th} and $(n+1)^{th}$ iter.

$$\varepsilon > 0. \quad \|V_{n+1} - V_n\| < \frac{\varepsilon(1-\gamma)}{2\gamma}$$

π -greedy policy given V_{n+1}

$$\text{prove: } \|V^\pi - V_{n+1}\| \leq \varepsilon/2$$

$$\begin{aligned} \|V^\pi - V_{n+2} + V_{n+2} - V_{n+1}\| &\leq \|V^\pi - V_{n+2}\| + \|V_{n+2} - V_{n+1}\| \leq \\ &\leq \|BV_{n+1} - BV_n\| \leq \gamma \|V_{n+1} - V_n\| \leq \gamma \cdot \frac{\varepsilon(1-\gamma)}{2\gamma} \leq \frac{\varepsilon}{2}. \end{aligned}$$

f) Prove $\|B^K V - B^K V'\| \leq \gamma^k \|V - V'\|$

$$\begin{aligned} \|B^K V - B^K V'\| &\leq \gamma \|B^{k-1} V - B^{k-1} V'\| \leq \gamma^2 \|B^{k-2} V - B^{k-2} V'\| \leq \\ &\leq \dots \leq \gamma^k \|V - V'\| \end{aligned}$$

g) Prove: $\|V^* - V_{n+1}\| \leq \varepsilon/2$ $\xrightarrow{\text{repeat}}$

$$\begin{aligned} \|V^* - V_{n+1}\| &= \|V^* + V_{n+2} - V_{n+2} - V_{n+1}\| \leq \|V^* - V_{n+2}\| + \\ &+ \|V_{n+2} - V_{n+1}\| \leq \left| \begin{array}{l} V^* = \text{fixed point} \\ BV^* = V^* \\ BV_{n+2} = V_{n+2} \\ BV_n = V_{n+1} \end{array} \right| \leq \|BV^* - BV_{n+1}\| + \\ &+ \|BV_{n+1} - V_n\| \leq \gamma \|V^* - V_{n+1}\| + \gamma \|V_{n+1} - V_n\| \leq \underbrace{\gamma \|V^* - V_{n+1}\|}_{\frac{\varepsilon(1-\gamma)}{2\gamma}} + \\ &+ \gamma^k \frac{\varepsilon(1-\gamma)}{2\gamma} \end{aligned}$$

$$\|V^* - V_{n+1}\| \leq \gamma \|V^* - V_{n+1}\| + \varepsilon(1-\gamma) \cdot \frac{1}{2}$$

$$\|V^* - V_{n+1}\| (1-\gamma) \leq \frac{\varepsilon}{2} \cdot (1-\gamma) \quad |1-\gamma>0.$$

$$\|V^* - V_{n+1}\| \leq \frac{\varepsilon}{2}.$$

h) $\|V^\pi - V^*\| \leq \varepsilon \leftarrow \text{prove}$

$$\begin{aligned} \|V^\pi - V^*\| &\leq \|V^\pi - V_{n+1} + V_{n+1} - V^*\| \leq \|V^\pi - V_{n+1}\| + \\ &+ \|V_{n+1} - V^*\| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$