インテリジェントシステム レポート課題 4

21T2166D 渡辺大樹

2024/07/08

1

(a) 状態価値関数 U(s) はここでは次の式で表すことができる。

$$U(s) = \max(R(s, stop), \sum_{s'} P(s'|s, spin)[R(s, spin, s') + \gamma U(s')])$$

状態 s=0 のときは

$$U(0) = \max(0, \sum_{s'} P(s'|0, spin)[R(0, spin, s') + \gamma U(s')])$$

となる。ここで

$$\sum_{s'} P(s'|0, spin)[R(0, spin, s') + \gamma U(s')]$$

を計算すると

$$\begin{aligned}\sum_{s'} P(s'|0, spin)[R(0, spin, s') + \gamma U(s')] &= \frac{1}{3}(U_0(2) + U_0(3) + U_0(4)) \\ &= \frac{1}{3}(0 + 0 + 0) \\ &= 0\end{aligned}$$

となる。よって

$$U(0) = \max(0, 0) = 0$$

となる。

状態 s=2 のときは

$$U(2) = \max(2, \sum_{s'} P(s'|2, spin)[R(2, spin, s') + \gamma U(s')])$$

となる。ここで

$$\sum_{s'} P(s'|2, spin)[R(2, spin, s') + \gamma U(s')]$$

を計算すると

$$\begin{aligned}\sum_{s'} P(s'|2, spin)[R(2, spin, s') + \gamma U(s')] &= \frac{1}{3}(U_0(4) + U_0(5) + 0) \\ &= \frac{1}{3}(0 + 0 + 0) \\ &= 0\end{aligned}$$

となる。よって

$$U(2) = \max(2, 0) = 2$$

状態 s=3 のときは

$$U(3) = \max(3, \sum_{s'} P(s'|3, spin)[R(3, spin, s') + \gamma U(s')])$$

となる。ここで

$$\sum_{s'} P(s'|3, spin)[R(3, spin, s') + \gamma U(s')]$$

を計算すると

$$\begin{aligned}\sum_{s'} P(s'|3, spin)[R(3, spin, s') + \gamma U(s')] &= \frac{1}{3}(U_0(5) + 0 + 0) \\ &= \frac{1}{3}(0 + 0 + 0) \\ &= 0\end{aligned}$$

となる。よって

$$U(3) = \max(3, 0) = 3$$

状態 s=4 のときは

(b)

(c)

(d)

(e)

2

(a)

状態 s_2, s_4, s_5 はそれぞれ s_4, s_5, s_0 にしか遷移しないため価値関数 U はそれぞれ

$$U(s_2) = 80, U(s_4) = 90, U(s_5) = 100$$

となる。

(b)

状態 s_3 での状態価値関数 $U(s_3)$ は

$$\begin{aligned} U(s_3) &= \max_{\mathbf{a}\in\{a,b\}} \sum_{s'} P(s'|s_3,\mathbf{a})[R(s_3,\mathbf{a},s') + \gamma U(s')] \\ &= \sum_{s'} P(s'|s_3,b)[R(s_3,b,s') + \gamma U(s')] \\ &= P(s_4|s_3,b)[R(s_3,b,s_4) + \gamma U(s_4)] \\ &\quad + P(s_5|s_3,b)[R(s_3,b,s_5) + \gamma U(s_5)] \end{aligned}$$

となる。$P(s_4|s_3,b) = p, P(s_5|s_3,b) = q$ を代入し、ほかの値も課題資料中の表の値を用いると、

$$U(s_3) = 80p + 96q$$

$p + q = 1$ より、p または q で整理すると

$$\begin{aligned} &= 96 - 16p \\ &= 16q + 80 \end{aligned}$$

となる。

(c)

状態 s_1 での状態価値関数 $U(s_1)$ は

$$\begin{aligned} U(s_1) &= \max_{\mathbf{a}\in\{a,b\}} \sum_{s'} P(s'|s_1,\mathbf{a})[R(s_1,\mathbf{a},s') + \gamma U(s')] \\ &= \max(P(s_3|s_1,a)[R(s_1,a,s_3) + \gamma U(s_3)], P(s_2|s_1,b)[R(s_1,b,s_2) + \gamma U(s_2)]) \end{aligned}$$

である。ここに資料中の値と前問で出した答えを代入すると

$$U(s_1) = \max(86 - 16p, 85)$$

となる。この関数の解は

$$U(s_1) = \begin{cases} 86 - 16p & (p <= \frac{1}{16}) \\ 85 & \text{otherwise} \end{cases}$$

となる。