Expressing Equations

Sunday, May 29, 2022 3:43 PM

1. Value Iteration

a) Deterministic : do one iteration for $V(s)$ and $Q(s,a)$

- $V(s)$

$S = \{SF_1, s_1, s_2, s_3, SF_2\}$

$$f_{MT} = \begin{array}{c} SF_1 \\ s_1 \\ s_2 \\ s_3 \\ SF_2 \end{array} \begin{bmatrix} \rightarrow & \leftarrow \\ 1 & 0 \\ 2 & 0 \\ 3 & 1 \\ 4 & 2 \\ 4 & 3 \end{bmatrix} \qquad f_R(s_f) = \begin{array}{c} SF_1 \\ s_1 \\ s_2 \\ s_3 \\ SF_2 \end{array} \begin{bmatrix} -10 \\ 0 \\ -0.4 \\ -0.4 \\ 10 \end{bmatrix}$$

$\gamma = 0.8$

$$V(s) = \max_a \left[\sum_{s_f \in S} P_{MT}(s_f \mid s, a) \left[f_R(s, a, s_f) + \gamma V(s_f) \right] \right]$$

(INIT WITH Ø → $V(s_1)$)

$$V(SF_1) = \max\left(f_R(SF_1, \rightarrow, s_1) + 0.8(V(s_1)),\ f_R(SF_1, \leftarrow, SF_1) + 0.8(V(s_1)) \right)$$

$$= \max(\emptyset + 0.8 * \emptyset,\ -10 + 0.8 * \emptyset)$$

$V(SF_1) = \emptyset$

- $Q(s,a) = \sum_{s_f \in S} \left[f_R(s, a, s_f) + \gamma \max_{a_f} (Q(s_f, a_f)) \right]$

$$Q(SF_1, \rightarrow) = f_R(SF_1, \rightarrow, s_1) + 0.8 \max(Q(s_1, \rightarrow), Q(s_1, \leftarrow))$$

$$= \emptyset + 0.8 \max(\emptyset, \emptyset)$$

$Q(SF_1, \rightarrow) = \emptyset$

b) Non-deterministic

- $V(s) = \max_a \left[\sum_{s_f \in S} P_{MT}(s_f \mid s, a) \left[f_R(s, a, s_f) + \gamma V(s_f) \right] \right]$

$P_{MT}(s_f \mid s, a)$:

$s_f = SF_1$

	$\rightarrow$	$\leftarrow$
SF_1	0.2	0.8
s_1	0.2	0.8
s_2	0	0
s_3	0	0
SF_2	0	0

$s_f = s_1$

	$\rightarrow$	$\leftarrow$
SF_1	0.8	0.2
s_1	0	0
s_2	0.2	0.8
s_3	0	0
SF_2	0	0

$s_f = s_2$

	$\rightarrow$	$\leftarrow$
SF_1	0	0
s_1	0.8	0.2
s_2	0	0
s_3	0.2	0.8
SF_2	0	0

$s_f = s_3$

	$\rightarrow$	$\leftarrow$
SF_1	0	0
s_1	0	0
s_2	0.8	0.2
s_3	0	0
SF_2	0.2	0.8

$s_f = SF_2$

	$\rightarrow$	$\leftarrow$
SF_1	0	0
s_1	0	0
s_2	0	0
s_3	0.8	0.2
SF_2	0.8	0.2

$$V(SF_1) = \max_a \Big[(P_{MT}(SF_1 \mid SF_1, \rightarrow))(f_R(SF_1, \rightarrow, SF_1) + \gamma V(SF_1)) +$$
$$(P_{MT}(s_1 \mid SF_1, \rightarrow))(f_R(SF_1, \rightarrow, s_1) + \gamma V(s_1)) +$$
$$(P_{MT}(s_2 \mid SF_1, \rightarrow))(f_R(SF_1, \rightarrow, s_2) + \gamma V(s_2)) +$$
$$(P_{MT}(s_3 \mid SF_1, \rightarrow))(f_R(SF_1, \rightarrow, s_3) + \gamma V(s_3)) +$$
$$(P_{MT}(SF_2 \mid SF_1, \rightarrow))(f_R(SF_1, \rightarrow, SF_2) + \gamma V(SF_2)), \quad \text{— term 1}$$
$$(P_{MT}(SF_1 \mid SF_1, \leftarrow))(f_R(SF_1, \leftarrow, SF_1) + \gamma V(SF_1)) +$$
$$(P_{MT}(s_1 \mid SF_1, \leftarrow))(f_R(SF_1, \leftarrow, s_1) + \gamma V(s_1)) +$$
$$(P_{MT}(s_2 \mid SF_1, \leftarrow))(f_R(SF_1, \leftarrow, s_2) + \gamma V(s_2)) +$$
$$(P_{MT}(s_3 \mid SF_1, \leftarrow))(f_R(SF_1, \leftarrow, s_3) + \gamma V(s_3)) +$$
$$(P_{MT}(SF_2 \mid SF_1, \leftarrow))(f_R(SF_1, \leftarrow, SF_2) + \gamma V(SF_2)) \Big] \quad \text{— term 2}$$

max (term 1, term 2)

$$= \max\left[(0.2)(-10 + 0.8(\emptyset)) + (0.8)(0 + 0.8(\emptyset)),\ (0.8)(-10 + 0.8(\emptyset)) + (0.2)(0 + 0.8(\emptyset)) \right]$$

$$= \max\left[(0.2)(-10) + (0.8)(\emptyset),\ (0.8)(-10) + (0.2)(\emptyset) \right]$$

$$= \max[-2, -8]$$

$V(SF_1) = -2$

- $Q(s,a) = \sum_{s_f \in S} P_{MT}(s_f \mid s, a) \left[f_R(s, a, s_f) + \gamma \max_{a_f} (Q(s_f, a_f)) \right]$

$$Q(SF_1, \rightarrow) = P_{MT}(SF_1 \mid SF_1, \rightarrow) \left[f_R(SF_1, \rightarrow, SF_1) + \gamma (Q(SF_1, \rightarrow), Q(SF_1, \leftarrow)) \right] +$$

$P_{MT}(s_1|SF_1,\rightarrow)(f_R(SF_1\rightarrow,s_1)+\gamma(Q(s_1,\rightarrow),Q(s_1,\leftarrow))]$ +
$P_{MT}(s_2|SF_1,\rightarrow)(f_R(SF_1\rightarrow,s_2)+\gamma(Q(s_2,\rightarrow),Q(s_2,\leftarrow))]$ +
$P_{MT}(s_3|SF_1,\rightarrow)(f_R(SF_1,\rightarrow,s_3)+\gamma(Q(s_3,\rightarrow),Q(s_3,\leftarrow))]$ +
$P_{MT}(SF|SF_1,\rightarrow)(f_R(SF_1,\rightarrow,SF_2)+\gamma(Q(SF_2,\rightarrow),Q(SF_2,\leftarrow))]$

$= (0.2)[-10 + 0.8\max[0,0)] +$
$(0.8)[\emptyset + 0.9\max(\emptyset,\emptyset)] +$
$\emptyset$ +
$\emptyset$ +
$\emptyset$

$= (0.2)(-10) + (0.8)(\emptyset)$

$Q(SF_1,\rightarrow) = -2$

2. Temporary Difference, $\alpha = 0.9$ $\gamma = 0.8$ $V(s_1) = 0.83$ $V(SF_1) = 0.15$ (random)

- $V(s) \leftarrow V(s) + \alpha\Delta V(s)$ $\Delta V(s) = [f_R(s,a,s_f) + \gamma V(s_f)] - V(s)$

$V(SF_1) = V(SF_1) + 0.9\,\Delta V(SF_1)$ $\Delta V(SF_1) = [f_R(SF_1,\rightarrow,s_1) + 0.8\,V(s_1)] - V(SF_1)$

(random) $V(s_1) \rightarrow V(s_f)$ $V(SF_1)$

$a = \max(f_R(SF_1\rightarrow,s_1) + 0.8(0.83), f_R(SF_1\leftarrow,SF_1) + 0.8(0.15))$

$a = \max(\emptyset + 0.664, -10 + 0.12)$

$a = \max(0.664, -9.88)$

$a = \rightarrow$

$V(SF_1) = 0.15 + 0.9((\emptyset + 0.8(0.83)) - 0.15)$

$= 0.15 + 0.9(0.664 - 0.15)$

$= 0.15 + 0.9(0.514)$

$= 0.15 + 0.4626$

$V(SF_1) = 0.6126$

- $Q(s,a) \leftarrow Q(s,a) + \alpha\Delta Q(s,a)$ $\Delta Q(s,a) = [f_R(s,a,s_f) + \gamma\max_{a_f}(Q(s_f,a_f))] - Q(s,a)$

$Q(SF_1,\rightarrow) = 0.67$ $Q(s_1,\rightarrow) = 0.54$ $Q(s_1,\leftarrow) = 0.15$ (random)

$Q(SF_1,\rightarrow) = 0.67 + 0.9\,\Delta Q(SF_1,\rightarrow)$

$\Delta Q(SF_1,\rightarrow) = (f_R(SF_1,\rightarrow,s_1) + 0.8\max[Q(s_1,\rightarrow),Q(s_1,\leftarrow)]) - 0.67$

$= [\emptyset + 0.8\max(0.54, 0.15)) - 0.67$

$= (0.8)(0.54) - 0.67$

$\Delta Q(SF_1,\rightarrow) = -0.238$

$Q(SF_1,\rightarrow) = 0.67 + 0.9(-0.238)$

$= 0.67 - 0.2142$

$Q(SF_1,\rightarrow) = 0.4558$