Homework 02: Accumulated Reward

Tuesday, March 22, 2022 10:54 AM

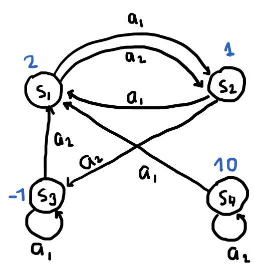
1. Given the world defined by the following transition function $f_{MT}(s,a)$, the reward function $f_R(s, a, s_f) = f_R(s_f)$ and $\gamma = 0.9$:

$$f_{M_T}(s,a) = \begin{array}{c c} & \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} & \begin{bmatrix} s_2 & s_2 \\ s_1 & s_3 \\ s_3 & s_1 \\ s_1 & s_4 \end{bmatrix} \end{array} \qquad f_R(s_f) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{bmatrix} 2 \\ 1 \\ -1 \\ 10 \end{bmatrix}$$

(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

$\tau_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution



$A = \{a_1, a_2\}$ $S = \{s_1\ s_2\ s_3\ s_4\}$

$$\tau_1 \rightarrow s_1 \xrightarrow[a_1, a_2]{} \overset{1}{s_2} \xrightarrow[a_2]{} \overset{-1}{s_3} \xrightarrow[a_2]{} \overset{2}{s_1} \xrightarrow[a_1, a_2]{} \overset{1}{s_2} \xrightarrow[a_1]{} \overset{2}{s_1}$$

$\gamma = 0.9$

$$\begin{aligned} f_{AR}(\tau_1) &= 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2) \\ &= 1 + (0.9)(-1) + (0.9)^2(2) + (0.9)^3(1) + (0.9)^4(2) \\ &= 3.76 \end{aligned}$$

$$\boxed{f_{AR}(\tau_1) = 3.76}$$

(b) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

$\tau_2 = s_3, s_1, s_2, s_3$

Solution

$$s_3 \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2} \rightarrow \overset{-1}{s_3}$$

with $\gamma = 0.9$

$$\begin{aligned} f_{AR}(\tau_2) &= 2 + \gamma(1) + \gamma^2(-1) \\ &= 2 + 0.9(1) + (0.9)^2(-1) \end{aligned}$$

$$\boxed{f_{AR}(\tau_2) = 2.09}$$

(c) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

$\tau_3 = s_2, s_1, s_2$

Solution

$$s_2 \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2}$$

$$\begin{aligned} f_{AR}(\tau_3) &= 2 + \gamma(1) \\ &= 2 + 0.9(1) \\ &= 2 + 0.9 \end{aligned}$$

$$\boxed{f_{AR}(\tau_3) = 2.9}$$

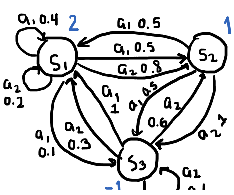
2. Given the world defined by the transition function $P_{MT}(s_f | s, a)$, the reward function $f_R(s_f, s, a) = f_R(s_f)$ and $\gamma = 0.6$:

$$f_{M_T}(s,a) = \begin{array}{c c} & \begin{array}{cc} s_f = s_1 & \\ a_1 & a_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & 0 \\ 1 & 0.3 \end{bmatrix} \end{array} \quad \begin{array}{c c} & \begin{array}{cc} s_f = s_2 & \\ a_1 & a_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \\ 0 & 0.6 \end{bmatrix} \end{array} \quad \begin{array}{c c} & \begin{array}{cc} s_f = s_3 & \\ a_1 & a_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.1 & 0 \\ 0.5 & 1 \\ 0 & 0.1 \end{bmatrix} \end{array} \qquad f_R(s_f) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

$\tau_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution



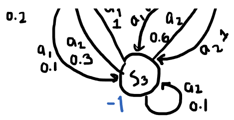
$S = \{s_1\ s_2\ s_3\}$ $A = \{a_1, a_2\}$

$$\tau_1 = s_1 \rightarrow \overset{1}{s_2} \rightarrow \overset{-1}{s_3} \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2} \rightarrow \overset{2}{s_1}$$

with $\gamma = 0.6$

$$\begin{aligned} f_{AR}(\tau_1) &= 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2) \\ &= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2) \end{aligned}$$

$$\boxed{f_{AR}(\tau_1) = 1.59}$$



$$f_{AR}(\tau_1) = 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2)$$
$$= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2)$$
$$\boxed{f_{AR}(\tau_1) = 1.59}$$

ⓑ Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory :

$\tau_2 = s_3, s_1, s_2, s_3$

Solution

$$S_3 \xrightarrow{2} S_1 \xrightarrow{1} S_2 \xrightarrow{-1} S_3$$

with $\gamma = 0.6$

$$f_{AR}(\tau_2) = 2 + \gamma(1) + \gamma^2(-1)$$
$$= 2 + (0.6)(1) + (0.6)^2(-1)$$
$$\boxed{f_{AR}(\tau_2) = 2.24}$$

ⓒ Calculate the accumulated reward function $f_{AR}(\tau_3)$ for the trajectory :

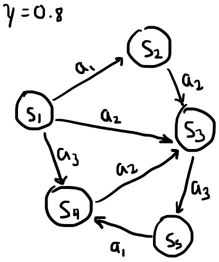
$\tau_3 = s_3, s_1, s_2$

Solution

$$S_3 \xrightarrow{2} S_1 \xrightarrow{1} S_2$$

$$f_{AR}(\tau_3) = 2 + \gamma(1)$$
$$= 2 + 0.6(1)$$
$$\boxed{f_{AR}(\tau_3) = 2.6}$$

③ Given the world defined by the following graph, the reward function $f_R(s_f, s, a)$ and $\gamma = 0.8$



$$f_R(s,a,s_f) = \begin{array}{c} s_f = s_1 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array} \end{array} \quad \begin{array}{c} s_f = s_2 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & -2 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array} \end{array} \quad \begin{array}{c} s_f = s_3 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 5 & 0 \\ s_2 & 0 & 4 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & -1 & 0 \\ s_5 & 0 & 0 & 0 \end{array} \end{array} \quad \begin{array}{c} s_f = s_4 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & -3 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 1 & 0 & 0 \end{array} \end{array} \quad \begin{array}{c} s_f = s_5 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & -6 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array} \end{array}$$

ⓐ Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

$\tau_1 = s_1, s_2, s_3, s_5, s_4, s_3, s_5$

Solution

$$S_1 \xrightarrow{-2} S_2 \xrightarrow{4} S_3 \xrightarrow{-6} S_5 \xrightarrow{1} S_4 \xrightarrow{-1} S_3 \xrightarrow{-6} S_5$$

with $\gamma = 0.8$

$$f_{AR}(\tau_1) = -2 + \gamma(4) + \gamma^2(-6) + \gamma^3(1) + \gamma^4(-1) + \gamma^5(-6)$$
$$= -2 + (0.8)(4) + (0.8)^2(-6) + (0.8)^3(1) + (0.8)^4(-1) + (0.8)^5(6)$$
$$\boxed{f_{AR}(\tau_1) = -0.57}$$

ⓑ Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory :

$\tau_2 = s_1, s_3, s_5, s_4$

Solution

$$S_1 \xrightarrow{5} S_3 \xrightarrow{-6} S_5 \xrightarrow{1} S_4$$

with $\gamma = 0.8$

$$f_{AR}(\tau_2) = 5 + \gamma(-6) + \gamma^2(1)$$
$$= 5 + (0.8)(-6) + (0.8)^2(1)$$
$$\boxed{f_{AR}(\tau_2) = 0.84}$$

ⓒ Calculate the accumulated reward function $f_{AR}(\tau_3)$ for the trajectory :

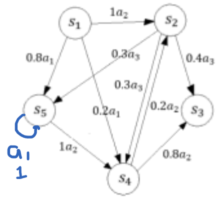
$\tau_3 = s_4, s_3, s_5$

$$S_4 \xrightarrow{-1} S_3 \xrightarrow{-6} S_5$$

with $\gamma = 0.8$

$$f_{AR}(\tau_3) = -1 + \gamma(6)$$
$$= -1 + (0.8)(6)$$
$$= -1 + 4.8$$
$$\boxed{f_{AR}(\tau_3) = 3.8}$$

(4) Given the world defined by the following graph where $0.8a_1$ means a_1 with probability 0.8, and so on, and with the following reward function and $\gamma = 0.7$



$$f_R(s,a,s_f) = \begin{array}{c}s_f = s_1\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3\\ \hline s_1 & 9 & 0 & 0\\ s_2 & 0 & 0 & 0\\ s_3 & 0 & 1 & 0\\ s_4 & 0 & 0 & -2\\ s_5 & 2 & 0 & 0\end{array}\end{array} \quad \begin{array}{c}s_f = s_2\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3\\ \hline s_1 & -2 & 2 & 0\\ s_2 & 0 & 0 & 0\\ s_3 & 0 & 0 & 0\\ s_4 & 0 & -3 & 0\\ s_5 & 0 & 0 & 0\end{array}\end{array} \quad \begin{array}{c}s_f = s_3\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3\\ \hline s_1 & 0 & 5 & 0\\ s_2 & 0 & 4 & -1\\ s_3 & 0 & 0 & 0\\ s_4 & 0 & -1 & 0\\ s_5 & 0 & 0 & 0\end{array}\end{array} \quad \begin{array}{c}s_f = s_4\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3\\ \hline s_1 & 1 & 0 & -3\\ s_2 & 0 & 0 & 2\\ s_3 & 0 & 0 & 0\\ s_4 & 0 & 0 & 0\\ s_5 & 1 & -1 & 0\end{array}\end{array} \quad \begin{array}{c}s_f = s_5\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3\\ \hline s_1 & 3 & 0 & 0\\ s_2 & 0 & 0 & 0\\ s_3 & 0 & 0 & -6\\ s_4 & 0 & 0 & 0\\ s_5 & 0 & 0 & 0\end{array}\end{array}$$

(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory:

$\tau_1 = s_1, s_4, s_2, s_5, s_4, s_3$

no transition

Solution

$$S_1 \xrightarrow[0.2\,a_1]{1} S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[0\,a_1]{1} S_4 \xrightarrow[0.8\,a_2]{-1} S_3$$

no transition, not a path

$$S_1 \xrightarrow[0\,a_3]{-3} S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[1\,a_2]{-1} S_4 \xrightarrow[0.8\,a_2]{-1} S_3$$

no transition from $s_1 \to s_4$ with a_3

$$S_1 \xrightarrow[0.2\,a_1]{1} S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[1\,a_2]{-1} S_4 \xrightarrow[0.8\,a_2]{-1} S_3 \qquad f_{AR} = -1.68$$

$$S_1 \xrightarrow[0\,a_3]{-3} S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[0\,a_1]{1} S_4 \xrightarrow[0.\,a_2]{-1} S_3$$

no transition, not a path

$$f_{AR} = 1 + (0.7)(-3) + (0.7)^2(0) + 0.7^3(-1) + 0.7^4(-1) = \boxed{-1.68}$$

(b) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory:

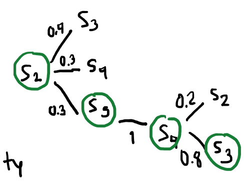
$\tau_2 = s_2, s_5, s_4, s_3$

Solution

$$S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[0\,a_1]{1} S_4 \xrightarrow[0.9\,a_2]{-3} S_3$$

this transition from S_5 to S_4 doesn't exist: this path is not possible

$$S_2 \xrightarrow[0.3\,a_3]{0} S_5 \xrightarrow[1\,a_2]{-1} S_4 \xrightarrow[0.8\,a_2]{-3} S_3$$



$$f_{AR} = 0 + (0.7)(-1) + (0.7)^2(-3)$$
$\boxed{f_{AR} = -2.17}$ with 0.24 of probability

(c) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory:

$\tau_3 = s_4, s_2, s_3$

Solution

$$S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0\,a_2]{4} S_3$$

no transition, not possible

$$S_4 \xrightarrow[0.2\,a_2]{-3} S_2 \xrightarrow[0.4\,a_3]{-1} S_3 \qquad f_{AR} = -3 + (0.7)(-1) = \boxed{-3.7}$$

with 0.08 probability