Homework 02: Accumulated Reward

Tuesday, March 22, 2022 10:54 AM

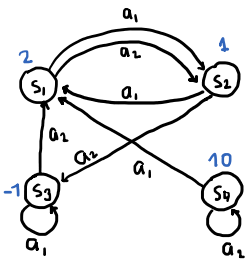
(1) Given the world defined by the following transition function $f_{MT}(s,a)$, the reward function $f_R(s,a,s_f) = f_R(s_f)$ and $\gamma = 0.9$:

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

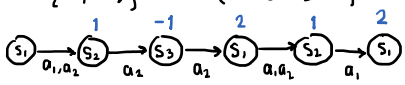
(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :

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

Solution



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

$\tau_1 \rightarrow$ 

$\gamma = 0.9$

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

$= 1 + (0.9)(-1) + (0.9)^2(2) + (0.9)^3(1) + (0.9)^4(2)$

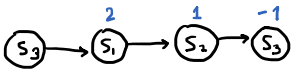
$= 3.76$

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

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

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

Solution



with $\gamma = 0.9$

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

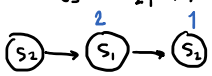
$= 2 + 0.9(1) + (0.9)^2(-1)$

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

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

$\tau_3 = s_2, s_1, s_2$

Solution



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

$= 2 + 0.9(1)$

$= 2 + 0.9$

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

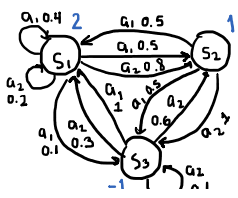
(2) Given the world defined by the transition function $P_{MT}(s_f | s,a)$, the reward function $f_R(s_f, s, a) = f_R(s_f)$ and $\gamma = 0.6$:

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

(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



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

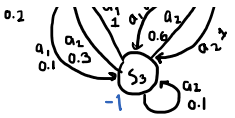
$\tau_1 =$ 

with $\gamma = 0.6$

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

$= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2)$

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

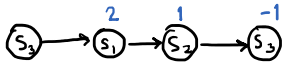


$$f_{AR}(\tau_1) = 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2)$$
$$= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2)$$
$$\boxed{f_{AR}(\tau_1) = 1.59}$$

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

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

Solution



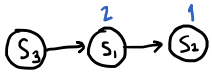
with $\gamma = 0.6$

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

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

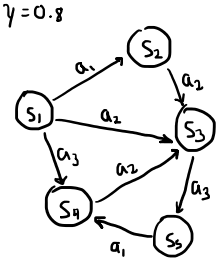
$$\tau_3 = s_3, s_1, s_2$$

Solution



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

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



$$f_R(s, a, s_f) =$$

$s_f = s_1$

	a_1	a_2	a_3
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

$s_f = s_2$

	a_1	a_2	a_3
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

$s_f = s_3$

	a_1	a_2	a_3
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

$s_f = s_4$

	a_1	a_2	a_3
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

$s_f = s_5$

	a_1	a_2	a_3
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

(a) 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

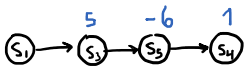
with $\gamma = 0.8$

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

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

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

Solution

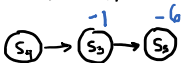


with $\gamma = 0.8$

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

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

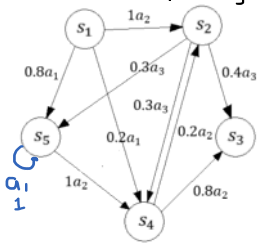
$$\tau_3 = s_4, s_3, s_5$$



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 $\gamma = 0.7$

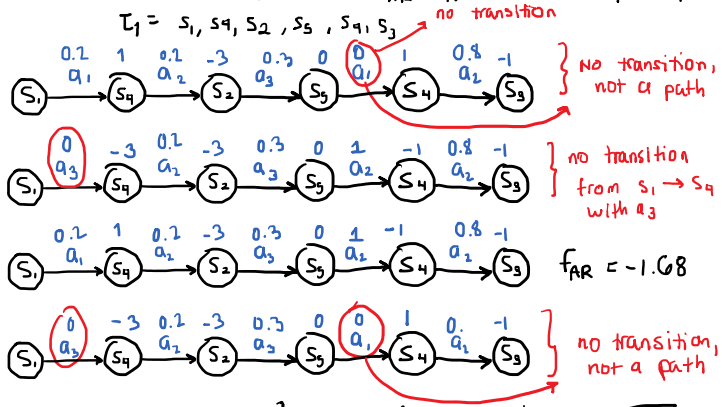


$$f_R(s,a,s_f) = \begin{array}{c} s_f = s_1 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ s_1 & 9 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 1 & 0 \\ s_4 & 0 & 0 & -2 \\ s_5 & 2 & 0 & 0 \end{array} \end{array} \quad \begin{array}{c} s_f = s_2 \\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ s_1 & -2 & 2 & 0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

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

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

Solution

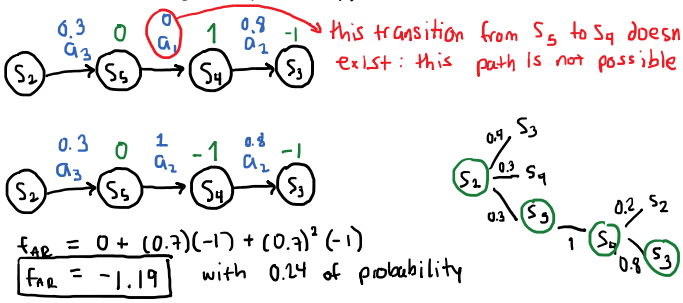


$$f_{AR} = 1 + (0.7)(-3) + (0.7)^2(0) + 0.7^3(-1) + 0.7^4(-1) = \boxed{-1.68}$$

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

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

Solution

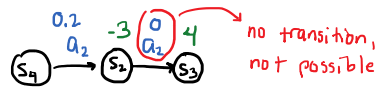


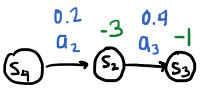
$$f_{AR} = 0 + (0.7)(-1) + (0.7)^2(-1)$$
$$\boxed{f_{AR} = -1.19} \quad \text{with } 0.24 \text{ of probability}$$

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

$\tau_3 = s_4, s_2, s_3$

Solution





$$f_{AR} = -3 + (0.7)(-1) = \boxed{-3.7}$$

with 0.08 probability