

۱. الف) ۲ به discount factor هفت است. (فقداری بین ۰ و ۱). در آینده پاداش
۲. ها ارزش کمتری خواهند داشت نسبت به الان. همچنین به هکلهایی در پاداش
۳. ها کمکی کند که تا بعد از آن دار نشود.
۴. هرچه γ آتیه باشد اهمیت بلند مدت تاثیر بیشتری دارد و بالعکس.
۵. از لحاظ سرعت هکلهایی هر چه γ آتیه باشد سرعت آن است و در
۶. (ب) در ارزشیابی تکلهایی، ابتدا $V(s)$ ها به حسب واکتیزم (s, a) ها محاسبه
۷. می شوند. حال Policy را از روی آن فاکتورهای مشخص می کنیم. **کاربرد: تقوای حالت**
۸. در سیاستگذاری تکلهایی یک Policy ثابت اولیه داریم. هر بار که $V(s)$ ها پیدا
۹. می کنند Policy بهبود یافته از قبلی ای را می گیریم. **کاربرد: تقوای حالت**
۱۰. ج) وقتی دیکه تعویج تقییدی در سیاست فعلی بود و در سیاست
۱۱. داریم.

۱۲. (د) این کار صرفاً Partial evaluation می باشد. نه لزومی ندارد به سیاست بهینه برسیم.

۱۴. الف) $T(s, a) = P(s' | s, a) =$ احتمال این که از وضعیت s به s' برویم با action a قدر است P .

۱۵. وضعیت ها: $\{done, done, done, done\}$ عمل ها: $\{move, stop\}$

۱۷. به ازای دهی state های s به $done$ که هیچ actionی در state های s که $done$ است.

۱۸. این ها به صورت زیر است: $T(s, move, done) = \frac{1}{4}$ $T(s, stop, done) = 1$ $T(s, stop, s) = 0$ $T(s, move, s) = \frac{1}{4}$ $T(s, stop, s) = 0$ $T(s, move, s) = \frac{1}{4}$ $T(s, stop, s) = 0$

۲۰. در غیر این صورت $T(s, move, done) = 1$ $T(s, stop, done) = 1$ $T(s, stop, s) = 0$ $T(s, move, s) = \frac{1}{4}$

۲۱. $V(s, a) =$
 if $a = stop$; $V(s, a) = 1$
 else: $V(s, a) = \frac{1}{4}$
 $T(s, move, done) = 1$
 $T(s, stop, done) = 1$
 $T(s, stop, s) = 0$
 $T(s, move, s) = \frac{1}{4}$

done	W	K	M	T	O	done
0	0	0	0	0	0	0
0	W	K	M	T	0	1
0	W	K	stop	M	M	2
0	W	K	M	M	10/12	3
0	W	K	M	M	10/12	4

$$V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s, a, s') (R(s, a, s') + \gamma V_k(s'))$$

توی تابلو این رو با آله میخوریم
 بهر از 4 حالت که میخوره میور
 done رو به جای تابلو

$\pi^*(0, T) = \text{move}$
 $\pi^*(T, T, W, done) = \text{stop}$

done	W	K	M	T	O	done
0	0	0	0	0	0	0
0	W	K	M	T	0	1
0	W	K	M	T	0, 2	2
0	W	K	M	T	0, 2	3


در وقتی که با تابلو میور
 - میور

$\lambda = 0.1$
 حال میور که تابلو = 0
 در تابلو که تابلو به تابلو

$\pi^*(0) = \text{move}$ $\pi^*(T, T, T, done) = \text{stop}$

$\gamma = 1$ $\pi_0 = \text{move}, \text{stop}, \text{move}, \text{stop}, \text{move}$

$$V(0) = \frac{1}{4} (V(T) + V(W) + V(K) + V(O)) \quad V_T = T \quad V_W = \frac{1}{4} (V(W)) \quad V_K = K$$

$V_W = 0$ $V = \{T, T, 0, K, 0\}$

 $\Rightarrow \pi_1 = \text{move}, \text{stop}, \text{stop}, \text{stop}, \text{stop}$
 $V(0) = \frac{1}{4} (V(T) + V(M) + V(K) + V(O))$

ω	κ	κ	κ	0
ω	κ	κ	κ	κ

=) $M_F = \text{more, more, stop, stop, stop}$

Subject:

Year:

Month:

Date:



Sa Su Mo Tu We Th Fr

$$V(A) = 1 + V(B) = 19$$

$$V(B) = 1 + V(C) = 10$$

$$V(C) = \frac{1}{2}(1 + V(D)) + \frac{1}{2}(1 + V(E)) = 16$$

$$V(D) = 1 + V(E) = 15$$

$$V(E) = 1 + V(F) = 11$$

$$V(F) = 1 + V(G) = 10$$

$$V(G) = 0$$

$$V_r(B) = \kappa \cdot \omega$$

$$Q_r(B, \text{right}) = \kappa \cdot \omega$$

$$Q_r(B, \text{left}) = -1$$

G	F	E	D	C	B	A
0	0	0	0	0	0	0
0	10	11	15	16	19	19
0	10	11	15	16	19	19

م

2

0

1

1

1

1

(5)

$\kappa(\omega + \kappa \cdot \omega)$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

$\kappa \cdot \omega$

