

بسم الله الرحمن الرحيم

تکلیف ۶ درس هوش مصنوعی
دکتر فدایی و دکتر یعقوب زاده

مهدی وجهی

۸۱۰۱۰۱۵۵۸

MDP

سوال ۱

$$V_{i+1}(s) = \max_{s'} (\sum T(s, a, s') (R(s, a, s') + \gamma V_i(s'))) = \max_{s'} (\sum \gamma T(s, a, s') V_i(s'))$$

$$V_1((1, 2)) = \max_{s'} (\sum \gamma T((1, 2), a, s') V_0(s'))$$

$$\begin{aligned} &= \max_{s'} (0.9 \times T((1, 2), L, s') \times 0, 0.9 \times T((1, 2), R, (1, 3)) V_0((1, 3)), \\ &0.9 \times T((1, 2), U, (1, 3)) V_0((1, 3)), 0.9 \times T((1, 2), D, (1, 3)) V_0((1, 3))) = \\ &= \max(0, 0.9 \times 0.8 \times -5, 0.9 \times 0.1 \times -5, 0.9 \times 0.1 \times -5) = 0 \end{aligned}$$

$$V_1((2, 2)) = \max_{s'} (\sum \gamma T((2, 2), a, s') V_0(s'))$$

$$\begin{aligned} &= \max_{s'} (0.9 \times T((2, 2), L, s') \times 0, 0.9 \times T((2, 2), R, (2, 3)) V_0((2, 3)), \\ &0.9 \times T((2, 2), U, (2, 3)) V_0((2, 3)), 0.9 \times T((2, 2), D, (2, 3)) V_0((2, 3))) = \\ &= \max(0, 0.9 \times 0.8 \times 5, 0.9 \times 0.1 \times 5, 0.9 \times 0.1 \times 5) = 3.6 \end{aligned}$$

بقیه خانه ها چون مقدار مجاورشان همه صفر است، تغییری نمی کنند.

$$V_2((1, 2)) = \max_{s'} (\sum \gamma T((1, 2), a, s') V_1(s'))$$

$$\begin{aligned} &= \max(0.9 \times T((1, 2), L, (2, 2)) V_1((2, 2)) + 0.9 \times T((1, 2), L, (1, 2)) V_1((1, 2)), \\ &0.9 \times T((1, 2), R, (1, 3)) V_1((1, 3)) + 0.9 \times T((1, 2), R, (2, 2)) V_1((2, 2)) + 0.9 \times T((1, 2), R, (1, 2)) V_1((1, 2)), \\ &0.9 \times T((1, 2), D, (1, 2)) V_1((1, 2)) + 0.9 \times T((1, 2), D, (1, 3)) V_1((1, 3)), \\ &0.9 \times T((1, 2), U, (2, 2)) V_1((2, 2)) + 0.9 \times T((1, 2), U, (1, 3)) V_1((1, 3))) \\ &= \max(0.9 \times 0.1 \times 3.6 + 0.9 \times 0.1 \times 0, \\ &0.9 \times 0.8 \times -5 + 0.9 \times 0.1 \times 3.6 + 0.9 \times 0.1 \times 0, 0.9 \times 0.8 \times 0 + 0.9 \times 0.1 \times -5, \\ &0.9 \times 0.8 \times 3.6 + 0.9 \times 0.1 \times -5) \\ &= \max(0.324, -3.276, -0.45, 2.142) = 2.142 \end{aligned}$$

$$V_2((2, 2)) = \max_{s'} (\sum \gamma T((2, 2), a, s') V_1(s'))$$

$$= \max(0.9 \times T((2, 2), L, (2, 2)) V_1((2, 2)) + 0.9 \times T((2, 2), L, (1, 2)) V_1((1, 2)),$$

$$\begin{aligned}
& 0.9 \times T((2,2), R, (2,3))V_1((2,3)) + 0.9 \times T((2,2), R, (2,2))V_1((2,2)) + 0.9 \times T((2,2), R, (1,2))V_1((1,2)), \\
& 0.9 \times T((2,2), D, (1,2))V_1((1,2)) + 0.9 \times T((2,2), D, (2,3))V_1((2,3)), \\
& 0.9 \times T((2,2), U, (2,2))V_1((2,2)) + 0.9 \times T((2,2), U, (2,3))V_1((2,3))) \\
& = \max(0.9 \times 0.1 \times 3.6 + 0.9 \times 0.1 \times 0, \\
& 0.9 \times 0.8 \times 5 + 0.9 \times 0.1 \times 3.6 + 0.9 \times 0.1 \times 0, 0.9 \times 0.8 \times 0 + 0.9 \times 0.1 \times 5, \\
& 0.9 \times 0.8 \times 3.6 + 0.9 \times 0.1 \times 5) \\
& = \max(0.324, 3.924, 0.45, 3.042) = 3.924
\end{aligned}$$

$$\begin{aligned}
V_2((2,1)) &= \max_{s'}(\sum \gamma T((2,1), a, s')V_1(s')) \\
&= \max(0.9 \times T((2,1), U, (2,2))V_1((2,2)), 0.9 \times T((2,1), D, (2,2))V_1((2,2)), \\
&\sum_{s'} 0.9 \times T((2,1), L, s') \times 0, 0.9 \times T((2,1), R, (2,2))V_1((2,2)) \\
&= \max(0.9 \times 0.1 \times 3.6, 0.9 \times 0.1 \times 3.6, 0, 0.9 \times 0.8 \times 3.6) \\
&= \max(0.324, 0.324, 0, 2.592) = 2.592
\end{aligned}$$

S	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
V_0	0	0	-5	0	0	5
V_1	0	0	-5	0	3.6	5
V_2	0	2.142	-5	2.592	3.924	5

مقادیر جدول سوال با شکل متفاوت بود ما شکل و صورت سوال را معیار قرار دادیم.
 بعضی از احتمالاتی که خانه مقصد پاداش صفر داشت به دلیل کوتاه شدن راه حل در برگه نوشته نشده.

سوال ۲

S	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
$\pi * (S)$	U	U	-	R	R	-

سوال ۳

$$V((1,1)) = \frac{-5+5+5}{3} = 1.67$$

$$V((2, 2)) = \frac{5+5}{2} = 5$$

سوال ۴

$$V_{\pi}(s) = V_{\pi}(s) + \alpha(R(s, \pi(s), s') + \gamma V_{\pi}(s') - V_{\pi}(s))$$

$$\begin{aligned} V_{\pi}((1, 1)) &= V_{\pi}((1, 1)) + \alpha(R((1, 1), R, (1, 2)) + \gamma V_{\pi}((1, 2)) - V_{\pi}((1, 1))) \\ &= 0 + 0.1(0 + 0.9 \times 0 - 0) = 0 \end{aligned}$$

$$\begin{aligned} V_{\pi}((1, 2)) &= V_{\pi}((1, 2)) + \alpha(R((1, 2), R, (1, 3)) + \gamma V_{\pi}((1, 3)) - V_{\pi}((1, 2))) \\ &= 0 + 0.1(0 + 0.9 \times (-5) - 0) = -0.45 \end{aligned}$$

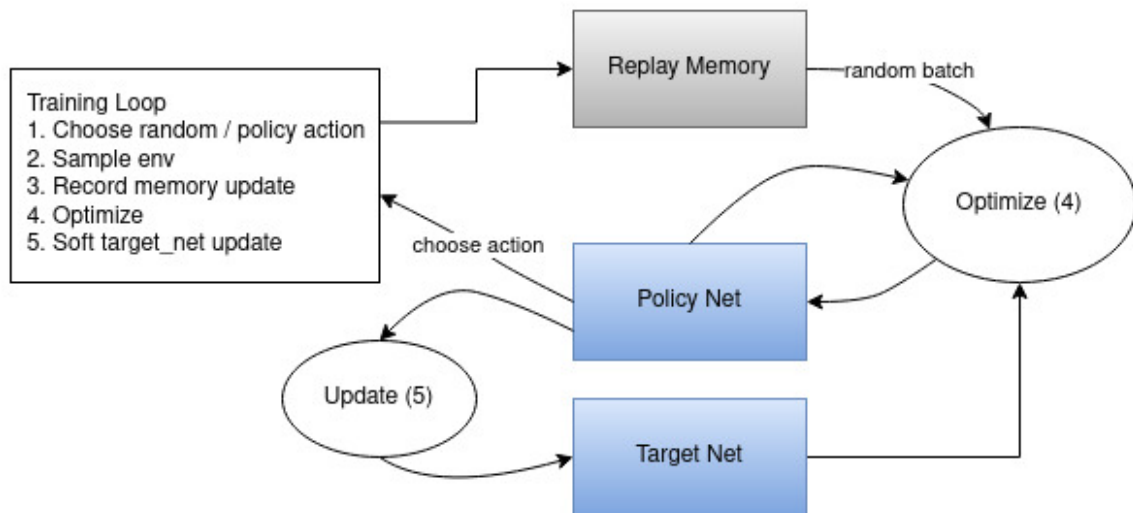
$$\begin{aligned} V_{\pi}((1, 1)) &= V_{\pi}((1, 1)) + \alpha(R((1, 1), R, (1, 2)) + \gamma V_{\pi}((1, 2)) - V_{\pi}((1, 1))) \\ &= 0 + 0.1(0 + 0.9 \times (-0.45) - 0) = -0.0405 \end{aligned}$$

$$\begin{aligned} V_{\pi}((1, 2)) &= V_{\pi}((1, 2)) + \alpha(R((1, 2), R, (2, 2)) + \gamma V_{\pi}((2, 2)) - V_{\pi}((1, 2))) \\ &= -0.45 + 0.1(0 + 0.9 \times 0 + 0.45) = -0.405 \end{aligned}$$

$$\begin{aligned} V_{\pi}((2, 2)) &= V_{\pi}((2, 2)) + \alpha(R((2, 2), R, (2, 3)) + \gamma V_{\pi}((2, 3)) - V_{\pi}((2, 2))) \\ &= 0 + 0.1(0 + 0.9 \times 5 + 0) = 0.45 \end{aligned}$$

S	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
V_0	0	0	-5	0	0	5
V_1	0	-0.45	-5	0	0	5
V_2	-0.0405	-0.405	-5	0	0.45	5

DQN



شبکه های عمیق Q در حالت کلی مانند تصویر بالا عمل می کنند. دلیل استفاده از این مدل ها به این خاطر است که با بزرگ شدن محیط و یا قرار گرفتن در محیط پیوسته حالت های بسیار زیادی را ایجاد می کند که عملاً ذخیره آنها در Q-table ممکن نیست برای حل این موضوع از این ایده استفاده می کنیم که به جای ایجاد جدول یک شبکه عصبی داشته باشیم و با استفاده از تجربیاتی که داشتیم آن را بروزرسانی کنیم. اما همانطور که می دانیم بروزرسانی با یک مشاهده واریانس بالایی دارد و می تواند باعث اختلال در یادگیری مدل شود. به همین علت از ۲ شبکه مشابه استفاده می کنیم مانند شکل. شبکه اول در هر مرحله با استفاده از نتیجه شبکه اول و دوم و با توجه به مشاهده انجام شده بروزرسانی می شود و بعد از تعدادی گام وزن های شبکه اول را به دومی انتقال می دهیم.