دانشکده مهندسی کامپیوتر

هوش مصنوعی و سیستم‌های خبره

آزمون میانترم

نام و نام خانوادگی . . . . . .

شماره دانشجویی . . . . . .

مدرس . . . . . محمدطاهر پیله‌ور - سید صالح اعتمادی

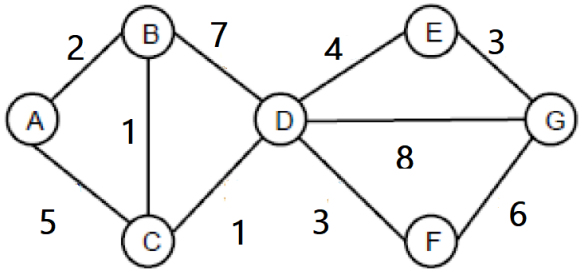
طراحی و تدوین . . . . . اساتید حل تمرین

تاریخ آزمون . . . . . ۱۱ آذر ۱۳۹۹

مدت زمان آزمون . . . . . ۱ ساعت و ۳۰ دقیقه

۱ جستجو (۱۷ نمره)

گراف زیر را در نظر بگیرید. گره A گره شروع و G گره پایانی است. یال‌ها بدون جهت و وزن‌دار هستند.



Node	h1	h2
A	9.5	10
B	9	12
C	8	10
D	7	8
E	1.5	1
F	4	4.5
G	0	0

۱.۱) مشخص کنید هر کدام از توابع heuristic آیا consistent هستند یا خیر.

Solution:

h_1 is not consistent since $h_1(D) > c(D,E) + h_2(E), 7 > 4 + 1.5$.

h_2 is not consistent since $h_2(D) > c(D,E) + h_2(E), 8 > 4 + 1$

۲.۱) ترتیب گره‌های expand شده و مسیر پیدا شده در هر کدام از روش‌های جستجوی زیر توسط graph search را مشخص کنید.

Solution:

1. DFS. Expanded: A, B, C, D, E, G, Path: $A - B - C - D - E - G$
2. BFS. Expanded: A, B, C, D, E, F, G, Path: $A - B - D - G$
3. UCS, Expanded: A, B, C, D, F, E, G, Path: $A - B - C - D - E - G$
4. A* with h1, Expanded: A, B, C, D, E, F, G, Path: $A - B - C - D - E - G$
5. A* with h2, Expanded: A, B, C, D, E, G, Path: $A - B - C - D - E - G$

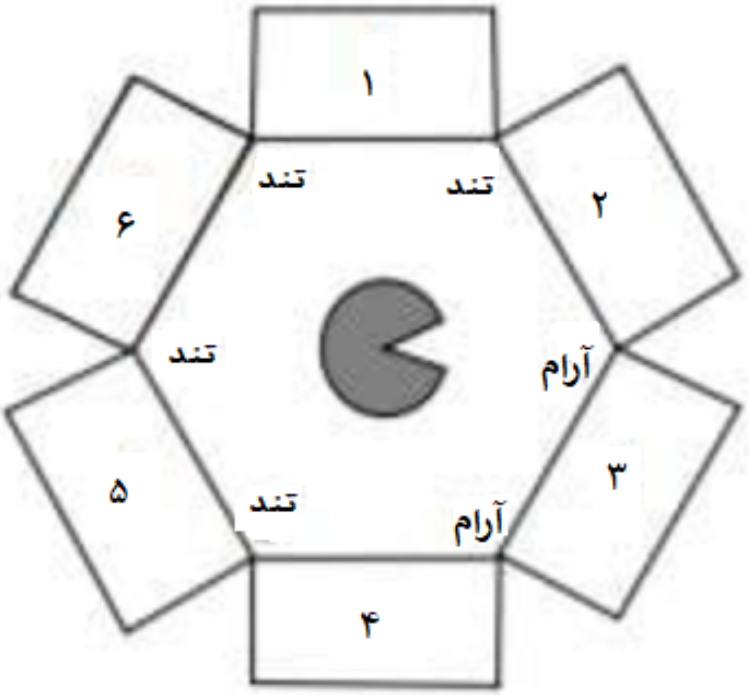
۲ CSP (۲۳ نمره)

شما در یک اتاق مثل شکل زیر گرفتار شده‌اید که ۶ راهرو دارد. راهروها یا به خروجی منجر می‌شوند یا به حیوان درنده و یا به پرتگاه. وقتی یک راهرو را انتخاب کنید دیگر راه برگشت ندارید. هدف شما انتخاب راهرو خروجی است.

نشانه‌هایی که دارید این است که از راهروی پرتگاه باد تند می‌آید. از راهروی حیوان درنده بادی نمی‌آید و از راهروی خروجی نسیم آرامی می‌وزد.

چون راهروها به هم چسبیده هستند، شما باد هر کدام را جداگانه نمی‌توانید بسنجید. در عوض وقتی که بین دو راهرو می‌ایستید باد غالب (بیشتر) را احساس خواهید کرد. مثلا اگر بین راهرو حیوان درنده و پرتگاه بایستید باد تند احساس می‌کنید ولی اگر بین راهرو حیوان درنده و خروجی بایستید نسیم آرامی احساس می‌کنید.

فرض کنید بین هر دو راهرو ایستاده‌اید و بادهای زیر را احساس کرده‌اید.



۱.۲) برای هر راهرو یک متغیر X_i در نظر گرفته که می‌تواند مقادیر Exit، Animal یا Pit را داشته باشد (A, E, P). با توجه به شکل بالا محدودیت‌های unary و/یا binary این متیغرها را بنویسید.

Solution:

Binary:

1. $X_1 = P | X_2 = P$
2. $X_2 = E | X_3 = E$
3. $X_3 = E | X_4 = E$
4. $X_4 = P | X_5 = P$
5. $X_5 = P | X_6 = P$
6. $X_1 = P | X_6 = P$

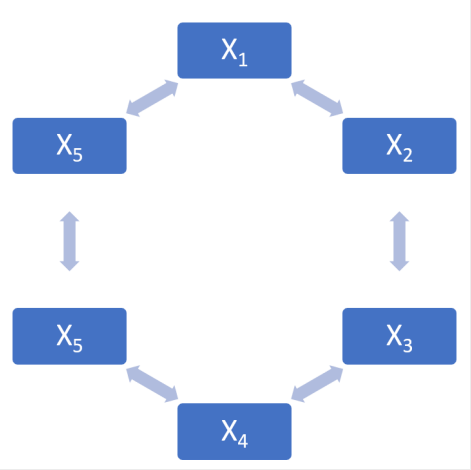
Unary: $X_2 \neq P$, $X_3 \neq P$, $X_4 \neq P$

If you copied the results from the Internet you would have the following additional condition:

$$\forall i, j, Adj(i, j) and \neg(X_i = E and X_j = E)$$

۲.۲) گراف محدویت متناظر را رسم کنید.

Solution:



۳.۲) مقادیر معتبر برای هر متغیر پس از اعمال محدودیت های یگانه را بنویسید.

Solution:

$X_1 = A, E, P$

$X_2 = A, E$

$X_3 = A, E$

$X_4 = A, E$

$X_5 = A, E, P$

$X_6 = A, E, P$

۴.۲) مقادیر معتبر برای هر متغیر پس از اعمال arc-consistency را بنویسید.

Solution:

$X_1 = P$

$X_2 = A, E$

$X_3 = A, E$

$X_4 = A, E$

$X_5 = P$

$X_6 = A, E, P$

This would be the result if you copied from the Internet:

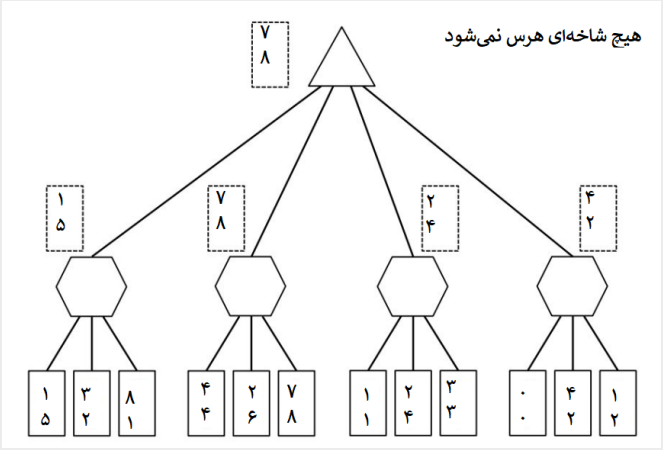
X_1	P		
X_2		G	E
X_3		G	E
X_4		G	E
X_5	P		
X_6	P	G	E

۳ Game (۲۰ نمره)

درخت بازی زیر را در نظر بگیرید. در این بازی شما در مقابل کامپیوتر بازی می‌کنید. هر بازیکن (شما و کامپیوتر) جداگانه امتیاز میگیرید و فقط بیشینه کردن امتیاز خود برای هر بازیکن اهمیت دارد. در این درخت امتیاز شما در بالا و

امتیاز کامپیوتر در پایین قرار دارد. با توجه به این نکته، جاهای خالی را پر کنید.

پاسخ:



با فرض اینکه شاخه‌ها از چپ به راست بررسی می‌شوند، شاخه‌هایی که نیاز به بررسی ندارند را با علامت ضربدر مشخص کنید.

۴ سوالات بخش MDP (۴۰ نمره)

۱.۴ سوال ۱

یک MDP را فرض کنید که دارای تابع انتقال و تابع پاداش زیر می‌باشند.
توجه فرمایید که ردیف‌هایی که در آن‌ها $T(s,a,s')=0$ می‌باشد، از جدول حذف شده‌اند.

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
A	۱	A	۰.۲	۰
A	۱	B	۰.۶	-۱
A	۱	C	۰.۲	۱
A	۲	A	۰.۱	۰
A	۲	B	۰.۱	-۱
A	۲	C	x	۱

جدول ۱

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
B	۱	A	۰.۸	۰
B	۱	B	y	-۱
B	۲	B	۰.۳	-۱
B	۲	C	۰.۷	۱
B	۳	A	۰.۵	۰
B	۳	C	۰.۵	۱

جدول ۲

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
C	۱	B	۱	-۱
C	۲	A	۱	۰
C	۳	C	۱	۱

جدول ۳

با فرض ضریب تخفیف ۱ به سوالات زیر پاسخ دهید:

۱. مقادیر مجهول موجود در جداول فوق را بیابید.

۲. یک لایه از درخت این MDP را رسم کنید.

۳. با فرض آنکه ارزش ابتدایی تمام حالات برابر صفر باشد، ارزش هر حالت را پس از دو مرحله Value Iteration بیابید.

۴. با توجه به مقادیر ارزش‌های هر حالت در هر مرحله، راهبرد هر حالت را در هر مرحله بیابید.

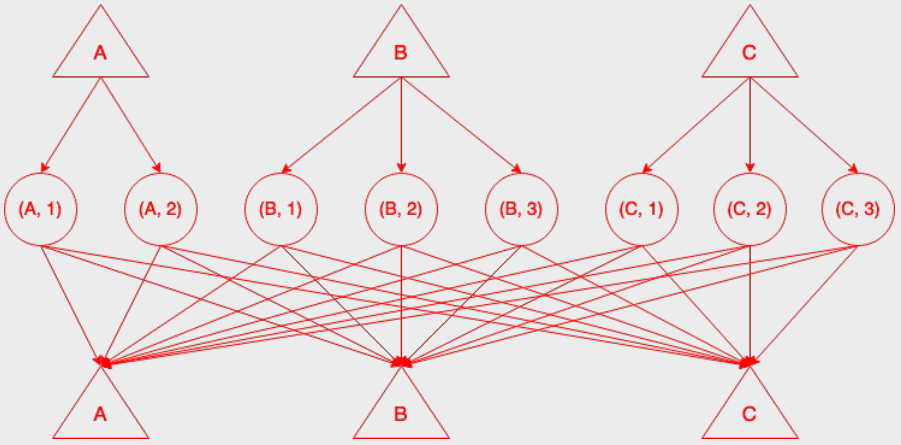
پاسخ:

۱. از آنجایی که جمع احتمالات هر action بایستی برابر یک شود، برای دو مقدار مجهول x و y داریم:

$$x + 0.1 + 0.1 = 1 \rightarrow x = 1 - (0.1 + 0.1) = x = 1 - 0.2 = 0.8$$

$$y + 0.8 = 1 \rightarrow y = 1 - 0.8 = 0.2$$

۲. درخت این مسئله برای یک لایه به شکل زیر می‌باشد:



شکل ۱: تصویر درخت مسئله MDP

۳. ابتدا روابط مورد نیاز برای حل این قسمت را بازنویسی می‌کنیم:

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')] \quad (۱)$$

$$V^*(s) = \max_a Q^*(s,a) \quad (۲)$$

از انجایی که در این مسئله $\gamma = 1$ در نظر گرفته شده است، روابط فوق به شکل زیر تغییر می‌کنند:

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + V^*(s')] \quad (۳)$$

$$V^*(s) = \max_a Q^*(s,a) \quad (۴)$$

حال به کمک روابط فوق به یافتن ارزش حالات Q و V در گام اول می‌پردازیم:

$$Q_1^*(A,1) = 0.2 \times (0+0) + 0.6 \times (-1+0) + 0.2 \times (1+0) = -0.4$$
$$Q_1^*(A,2) = 0.1 \times (0+0) + 0.1 \times (-1+0) + 0.8 \times (1+0) = 0.7$$
$$\rightarrow V_1^*(A) = \max(-0.4, 0.7) = 0.7$$

$$Q_1^*(B,1) = 0.8 \times (0+0) + 0.2 \times (-1+0) = -0.2$$
$$Q_1^*(B,2) = 0.3 \times (-1+0) + 0.7 \times (1+0) = 0.4$$
$$Q_1^*(B,3) = 0.5 \times (0+0) + 0.5 \times (1+0) = 0.5$$
$$\rightarrow V_1^*(B) = \max(-0.2, 0.4, 0.5) = 0.5$$

$$Q_1^*(C,1) = 1 \times (-1+0) = -1$$
$$Q_1^*(C,2) = 1 \times (0+0) = 0$$
$$Q_1^*(C,3) = 1 \times (1+0) = 1$$
$$\rightarrow V_1^*(C) = \max(-1, 0, 1) = 1$$

حال با استفاده از ارزش حالات V در گام اول، به محاسبه ارزش حالات Q و V در گام دوم می‌پردازیم:

$$Q_2^*(A,1) = 0.2 \times (0+0.7) + 0.6 \times (-1+0.5) + 0.2 \times (1+1) = 0.24$$
$$Q_2^*(A,2) = 0.1 \times (0+0.7) + 0.1 \times (-1+0.5) + 0.8 \times (1+1) = 1.62$$
$$\rightarrow V_2^*(A) = \max(0.24, 1.62) = 1.62$$

$$Q_2^*(B,1) = 0.8 \times (0+0.7) + 0.2 \times (-1+0.5) = -0.46$$
$$Q_2^*(B,2) = 0.3 \times (-1+0.5) + 0.7 \times (1+1) = 1.25$$
$$Q_2^*(B,3) = 0.5 \times (0+0.7) + 0.5 \times (1+1) = 1.35$$
$$\rightarrow V_2^*(B) = \max(-0.46, 1.25, 1.35) = 1.35$$

$$Q_2^*(C,1) = 1 \times (-1+0.5) = -0.5$$
$$Q_2^*(C,2) = 1 \times (0+0.7) = 0.7$$
$$Q_2^*(C,3) = 1 \times (1+1) = 2$$
$$\rightarrow V_2^*(C) = \max(-0.5, 0.7, 2) = 2$$

در نتیجه مقادیر نهایی ارزش حالات V پس از دو گام Value Iteration برابر است با:

$$V_2^*(A) = 1.62 \qquad V_2^*(B) = 1.35 \qquad V_2^*(C) = 2$$

۴. حال که مقادیر ارزش‌های نهایی هر حالت V را پس از دو گام Value Iteration محاسبه کردیم، با استفاده از رابطه زیر به محاسبه راهبرد (policy) مناسب می‌پردازیم:

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\, Q^*(s,a) \qquad (۵)$$

$\pi^*(A) = \operatorname{argmax}(0.24, 1.62) = 2$

$\pi^*(B) = \operatorname{argmax}(-0.46, 1.25, 1.35) = 3$

$\pi^*(C) = \operatorname{argmax}(-0.5, 0.7, 2) = 3$

۵ RL ً (۲۰ نمره)

Imagine an unknown game which has only two states {A,B} and in each state the agent has two actions to choose from: {Up,Down}.Suppose a game agent chooses actions according to some policy π and generates the following sequence of actions and rewards in the unknown game:

r_t	s_{t+1}	a_t	s_t	t
-1	B	Down	B	0
2	A	Down	B	1
-2	B	Up	A	2
+4	A	Up	B	3
-1	A	Down	A	4
+1	B	Down	A	5

Assume a discount factor $\Upsilon = 0.6$ *and a learning rate* $\alpha = 0.4$. *You can use calculator and calculations to two digits after decimal point is enough*

۱.۵

Assume that all Q-values initialized as 0. What are the following Q-values learned by running Q-learning with the above experience sequence?

Solution:

$Q(B, Down) = 0.6(0) + 0.4(-1 + 0.6(0)) = -0.4$

$Q(B, Down) = 0.6(-0.4) + 0.4(2 + 0.6(0)) = -0.24 + 0.8 = 0.56$

$Q(A, Up) = 0.6(0) + 0.4(-2 + 0.6(0.56)) = -0.66$

$Q(B, Up) = 0.6(0) + 0.4(4 + 0.6(0)) = 1.6$

$Q(A, Down) = 0.6(0) + 0.4(-1 + 0.6(0)) = -0.4$

$Q(A, Down) = 0.6(-0.4) + 0.4(1 + 0.6(1.6)) = 0.54$

So: $Q(A, Down) = 0.54$ $Q(B, Up) = 1.6$

5	4	3	2	1	0	initial	t
-0.66	-0.66	-0.66	-0.66	0	0	0	Q(A,Up)
0.54	-0.4	0	0	0	0	0	Q(A,Down)
1.6	1.6	1.6	0	0	0	0	Q(B,Up)
0.56	0.56	0.56	0.56	0.56	-0.4	0	Q(B,Down)

۲.۵

In model-based reinforcement learning, we first estimate the transition function T(s,a,s') and the reward function R(s,a,s').What are the following estimates of T and R, estimated from the experience above? Write "n/a" if not applicable or undefined.

Solution:

$\hat{T}(B, Down, A) = 1/2 \quad \hat{T}(B, Up, A) = 1$

$\hat{T}(A, Down, A) = 1/2 \quad \hat{T}(A, Up, A) = 0$

$\hat{R}(B, Down, A) = 2 \quad \hat{R}(B, Up, A) = 4$

$\hat{R}(A, Down, A) = -1 \quad \hat{R}(A, Up, A) = n/a$