

פתרון מטלה 6 – חישוביות וקוגניציה, 6119

15 בדצמבר 2025



שאלה 1

צורך כי מעל הגרפּ ($\{A, B, C, D, E\}, \{(A, B), (B, C), (C, D), (D, E)\}$) ייחד עם פעולה בינהarity. הפעולה מזיהה קדימה או אחורית ביחס סדר הציון של הצמתים והגמול הוא אף אלא אם בוצעה הפעולה a_1 ב- E , ערך ההנחה הוא $1 = \gamma$. נניח כי הייצור משתמש בהסתברות אחידה כרגע.

סעיף א'

נתאר את העולם בהגדרות MDP.
 $r(s, a) = \mathbb{1}_{(E, a_1)}$ הסיבכה היא צמתי הגרפּ, הפעולות $\{s' | a, s\} = \frac{1}{2} \mathcal{A} = \{a_0, a_1\}$, סיכוי המעבר $V_\pi(s)$ פתרון נוכור כי מתקיים $V_\pi(C) = \mathbb{E}(\sum_{i=0}^{\infty} \gamma^i r_i | s_0 = C) = \mathbb{E}(\sum_{i=0}^{\infty} r_i | s_0 = C)$ אבל מטעמי סימטריה נוכל להסיק שכל צעד הסיכוי שהסוכן יהיה באותו מקום $C - ca_i$, ובהתאם בהכרח $V_\pi(C) = \frac{1}{2}(0 + 1)$ בלבד.

סעיף ב'

נמצא את ערכי המצבים בהינתן המדיניות של הסוכן $V_\pi(s)$.
 $V_\pi(A) = \sum_{i=0}^1 \mathbb{P}(a_i | B) \left(r(B, a_i) + \gamma \sum_{s \in \{A, C\}} \mathbb{P}(s | B, a_i) V_\pi(s) \right) = \frac{1}{2}V_\pi(A) + \frac{1}{2}V_\pi(C) = \frac{1}{2}V_\pi(A) + \frac{1}{4}$
 נשתמש במשוואת בלמן כדי לחשב את שאר המצבים,

$$V_\pi(B) = \sum_{i=0}^1 \mathbb{P}(a_i | B) \left(r(B, a_i) + \gamma \sum_{s \in \{A, C\}} \mathbb{P}(s | B, a_i) V_\pi(s) \right) = \frac{1}{2}V_\pi(A) + \frac{1}{2}V_\pi(C) = \frac{1}{2}V_\pi(A) + \frac{1}{4}$$

באופן דומה נקבל שגם,

$$V_\pi(A) = \frac{1}{2}V_\pi(B) + 0.$$

ולכן,

$$V_\pi(B) = \frac{1}{4}V_\pi(A) + \frac{1}{4} \iff V_\pi(B) = \frac{1}{3}.$$

נבחן כי גם נובע,

$$V_\pi(C) = \frac{1}{2}V_\pi(B) + \frac{1}{2}V_\pi(D) \implies \frac{1}{2} = \frac{1}{6} + \frac{1}{2}V_\pi(D) \implies V_\pi(D) = \frac{2}{3}$$

ולבסוף,

$$V_\pi(E) = \frac{1}{2}V_\pi(D) + \frac{1}{2} = \frac{5}{6}$$

וקיבלנו את מפת הערך,

$$\left[\frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6} \right]$$

קיבלנו תוצאה סימטרית באופן שמתכלה עם הסימטריה באסטרטגיה.

משמעות התוצאה היא שככל שהייצור עומד רחוק יותר מהריבוע השמאלי, כך הסיכוי שלבסוף הוא יגיע אליו הוא נמוך יותר, ולכן הסיכוי שהוא קיבל גמול חיובי נמוך אף הוא.

סעיף ג'

נחשב את הליך הלימוד של הייצור בשיטת TD כאשר קצב הלימוד שלו הוא $\eta = 0.1$ וכאשר הייצור עשה שני ניסויים שבהם פעל ב- a_1, a_1, a_1 .

פתרון נחשב בעזרת טבלת מעקב. נשים לב שלאורך כל הניסוי הראשון הייצור לא לומד עד השלב האחרון שכן רק בשלב האחרון מקבל גם גמול חיובי וגם V חיובי.

trial	step	location	$V_\pi(A)$	$V_\pi(B)$	$V_\pi(C)$	$V_\pi(D)$	$V_\pi(E)$
1	0	C	0	0	0	0	0
1	1	D	0	0	0	0	0
1	2	E	0	0	0	0	0
1	3	end	0	0	0	0	0.2
2	0	C	0	0	0	0	0.2
2	1	D	0	0	0	0	0.2
2	2	E	0	0	0	0.02	0.2
2	3	end	0	0	0.02	0.02	0.38

סעיף 4'

מצורף לשאלת גרפ' המתאר את הליך הלימוד של היזכר בשיטת TD עבור 100 הרצות, נבין מה היו הערכיהם ההתחלתיים של כל אחד מהמצבים וنبין איפה הסטיים הניסויי הראשון.

פתורן בגרף נראה שהניסוי האפס מתואר על ידי מוגה קבועה ב- $\frac{1}{2}$, כלומר $V(s) = \frac{1}{2}$ לכל $s \in \{A, B, C, D, E\}$. לפיק הגרף לאחר הניסוי הראשון (ולפני השני) הערך של $\{B, C, D, E\}$ נשאר זהה, ולכן נסיק שלא השתנה, ובהתאם בעזרת האפקט שהוא רואים בסוף הניסוי הראשון בטבלה שהישבנו נוכל להסיק שהיזכר לא הגיע ימין מזמן ניסוי הראשון. נבחן כי גם מהגרף נתון ש- $V(A) < \frac{1}{2} < V(E)$, ולכן נוכל להסיק שגם היה שונה רק במצב A , כלומר היזכר הגיע אליו והמשיך הלאה ובכך למד שאין גבול בפיתוח מסלול זה.

שאלה 2

נthon MPD בעל שני המצבים Home, Out ושתי הפעולות Stay, Switch כמשמעותם Stay, Switch, Switch עלי ידי $r(H, \text{Stay}) = 0, r(H, \text{Switch}) = 1, r(O, \text{Stay}) = 2, r(O, \text{Switch}) = 0$. פרמטר $\gamma = \frac{1}{2}$.

סעיף א'

היא סוכן אקראי מתפלג אחיד, נכתבות את משווהות בלמן ונפתרו אותן במטרה לחשב את V_π .

פתרונות המשוואת בלמן הכללית במקורה שלנו היא,

$$\begin{aligned} V_\pi(s) &= \sum_{a \in \{\text{Stay, Switch}\}} \mathbb{P}(a | X) \left(r(s, a) + \gamma \sum_{s' \in \{O, H\}} \mathbb{P}(s' | s, a) V_\pi(s') \right) \\ &= \frac{1}{2} \sum_{a \in \{\text{Stay, Switch}\}} r(s, a) + \frac{1}{2} \sum_{s' \in \{O, H\}} \mathbb{P}(s' | s, a) V_\pi(s') \end{aligned}$$

נציב ערכיהם בהתאם,

$$V_\pi(H) = \frac{1}{2}(0 + \frac{1}{2}(0 + 1 \cdot V_\pi(H))) + 1 + \frac{1}{2}(0.2 \cdot V_\pi(H) + 0.8V_\pi(O)) = 0.5 + 0.3V_\pi(H) + 0.2V_\pi(O)$$

וכן מחישוב דומה,

$$V_\pi(O) = \frac{1}{2}(2 + \frac{1}{2}(1 \cdot V_\pi(O) + 0 \cdot V_\pi(H))) + 0 + \frac{1}{2}(1 \cdot V_\pi(H) + 0) = 1 + 0.25V_\pi(O) + 0.25V_\pi(H)$$

מהעברת אגפים נסיק,

$$0.7V_\pi(H) = 0.5 + 0.2V_\pi(O), \quad 0.75V_\pi(O) = 1 + 0.25V_\pi(H)$$

נציב את המשווהה השנייה בראשונה,

$$\frac{7}{10}V_\pi(H) = \frac{1}{2} + \frac{4}{15}(1 + \frac{1}{4}V_\pi(H)) = \frac{23}{30} + \frac{1}{15}V_\pi(H) \implies \frac{19}{30}V_\pi(H) = \frac{23}{30} \implies V_\pi(H) = \frac{23}{19} \approx 1.21$$

בהתאם,

$$V_\pi(O) = \frac{4}{3}(1 + \frac{1}{4} \cdot \frac{23}{19}) = \frac{33}{19} \approx 1.736$$

סעיף ב'

ונסה לנחש את המדיניות האופטימלית.

פתרונות ננחש שהמדיניות היא לנסות להחליף במצב H ולהישאר במצב O.

סעיף ג'

نبזוק אם המדיניות שניחשנו מקייםת את משווהת האופטימליות של בלמן.

פתרונות המשווהה היא,

$$V^*(s) = \max \left\{ r(s, a) + \gamma \sum_{s'} \mathbb{P}(s' | s, a) V^*(s') \mid a \right\}$$

נציב ונקבל,

$$V^*(O) = \max \left\{ 2 + \frac{1}{2}(1 \cdot V^*(O)), 0 + \frac{1}{2}(1 \cdot V^*(H)) \right\} = 2 + \frac{1}{2}V^*(O) \implies V^*(O) = 4$$

וכן,

$$\begin{aligned} V^*(H) &= \max \left\{ 0 + \frac{1}{2}(1 \cdot V^*(H) + 0 \cdot V^*(O)), 1 + \frac{1}{2}(0.2V^*(H) + 0.8V^*(O)) \right\} \\ &= \max \left\{ \frac{1}{2}V^*(H), 1 + 0.1V^*(H) + 0.4 \cdot 4 \right\} \\ &= 1 + 0.1V^*(H) + 0.4 \cdot 4 \end{aligned}$$

ולכן $.0.9V^*(H) = 2.6 \Rightarrow V^*(H) = \frac{26}{9} \approx 2.888$
נובע אם כך שהסטרטגיה שבחרנו אכן מקבלת ערכים מקסימליים.

סעיף ד'

נממש את האלגוריתם Value Iteration עבור הבעה ונמצא את ערכי V^* בהתאם לאלגוריתם.
פתרון הקובד נכתוב והורץ בקובץ המצורף, ותוצאתו הינו,

$$V^*(O) = 4, \quad V^*(H) = 3.333$$

נבחן כי הערכים שהתקבלו קרובים מאוד לערכים שנמצאו בסעיף הקודם, למעשה הערך $V^*(O)$ זהה לגורדי, בעוד יש פער ב- $V^*(H)$ שכנראה נובע מטעות חישוב.

סעיף ה'