מודלים חישוביים - הרצאה 3

שרון מלטר, אתגר 17

20 במרץ 2024

1 המשך הוכחה

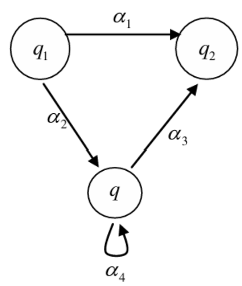
נמשיך בהוכחת הטענה מההרצאה הקודמת, כלומר נוכיח שכל שפה שמתקבלת על ידי אס"ד היא רגולרית.
בהנתן אס"ד M, נרצה לבנות ביטוי רגולרי α כך ש- $L(\alpha) < L(M)$ וכך נוכל לחפש ביטוי רגולרי מתאים (בסגנון של הפרד ומשול).
מדובר בהוכחה מעט מוזרה, אך היא מייצגת קונספט שמשתמשים בו רבות במדעי המחשב.
אסטרטגיה:
1. נגדיר מודל חדש של אוטומט סופי, 'אוטומט מוכלל'
2. נוכיח שאוטומט מוכלל מכיל את שפות האס"דים (הוא הרבה יותר חזק מאסל"ד)
3. נראה שקל להפוך אוטומט מוכלל עם שני מצבים לביטוי רגולרי.
4. נראה שלא קשה להפוך אוטומט מוכלל עם $t > 2$ מצבים לאוטומט מוכלל עם 2 מצבים.

לא נלמד על אוטומטים מוחלטים מעבר, אך נוכיח בהמשך שהם שקולים לביטויים רגולריים.

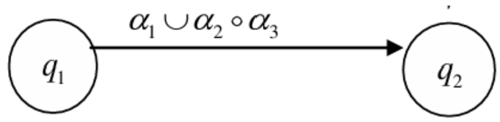
1.1 אוטומט מוכלל

אוטומט מוכלל הינו הכללה של אסל"ד.
באסל"ד מוגדרים מעברים אפשריים לקריאה של תווים יחידים, אך באוטומט מוכלל ניתן להגדיר מעברים אפשריים למספר תווים.
למשל נגדיר אוטומט מוכלל שיש לו מצב התחלתי ומעבר למצב מקבל, כך שניתן לעבור בו רק אם קבוצת התווים הבאה היא $(o \cup 1)^* \bigcirc 1$ כלומר זוהי מילה עם אפסים ואחדות שנגמרת ב-1.
באופן יותר פורמלי, נגדיר שניתן לעבור ממצב q_1 ל- q_2 אם"ם קוראים מילה $u \in L(\alpha)$. כמו כן נקבע שבאוטומט מוכלל יש בדיוק חץ אחד בין כל זוג מצבים, (כלומר חץ מכל מצב לאחר וחץ מהאחר אליו) , כולל בין מצב לעצמו, פרט לכך שאין חצים למצב ההתחלתי ואין חצים שיוצאים ממצב מקבל ושהמצב ההתחלתי איננו מקבל.
נראה שקל להפוך כל ביטוי רגולרי לאוטומט מוכלל עם שני מצבים:
נעביר חץ מהמצב ההתחלתי למצב מקבל שניתן לעבור בו רק אם קוראים מילה מ- $L(\alpha)$.
באופן הפוך ניתן להפוך כל אוטומט מוכלל לביטוי רגולרי.

<u>טענה:</u>
לכל אוטומט מוכלל עם $t > 2$ מצבים ישנו אוטומט מוכלל עם $t - 1$ מצבים לאותה השפה.
בסיס: נתחיל במקרה $t = 3$.
אם אף מצב איננו מקבל אז סיימנו.
אחרת, נקבל אוטומט מוכלל כזה;



נרצה לבנות אוטומט מוכלל חדש שמקבל את אותה השפה, שבה יש מצב מקבל מתוך המקורי.
נוכל פשוט לבנות את האוטומט הבא:



ברור כי ניתן להשתמש באותו העיקרון כדי לבנות אוטומט מוכלל עם פחות מצבים בשביל אוטומט שבו ישנם שני מצבים מקבלים.

צעד האינדוקציה: נניח את נכונות הטענה עבור $t-1$ ונבנה אלגוריתם היוצר אוטומט מוכלל כנדרש עבור t.
האלגוריתם־ נעבור בלולאות על כל זוגות המצבים q_1, q_2 כך שקיים חץ מ־ q_1 ל־ q_2. נסתכל בציור ונחליף את הביטוי הרגולרי α_1 שאותו צריכים לקרוא כדי לעבור מ־ q_1 ל־ q_2. ב־ $\alpha_1 \cap (\alpha_2 \bigcirc (\alpha_4^*) \bigcirc \alpha_3)$ כאשר α_2 על החץ מ־ q_1 ל־ q, α_3 על החץ מ־ q ל־ q_2 ו־ α_3 על החץ מ־ q לעצמו. (שימי לב ש־ q הוא כל מצב שאיננו מקבל ואיננו התחלתי)
מדוע עושים זאת? מכיוון שכעת כל מילה שהייתה צריכה לעבור דרך q כדי להגיע מ־ q_1 ל־ q_2 יכולה לעשות זאת ישירות מ־ q_1. ניתן לחשוב על זה כעל הוספת קשת ישירה בין שני צמתים שיש ביניהם מסלול. לא הוספנו או מחקנו קשירות, אלא רק קיצרנו מרחק.

וזהו, סיימנו את כל הצעדים שבאסטרטגיה, כך שהטענה נכונה :)
מאחר שרגולריות הינו מושג שנקבע לפני אס"דים ואסל"דים, מעתה לא נאמר ששפה מתקבלת על ידי אס"ד, אלא שהיא שפה רגולרית, כלומר שפה הנוצרת ע"י ביטוי רגולרי.
אבל כל החלק הזה מעלה את השאלה הבאה; מה הן שפות לא רגולריות?

2 שפות לא רגולריות

כפי שנאמר, הדבר היחיד שמכונת מצבים זוכרת הוא מה המצב האחרון שהיא הייתה בו. כלומר, אם מילה עברה במעגל של מצבים, באף שלב לא נוכל לדעת כמה פעמים עברנו בו. בגלל העיקרון הזה, אם חשוב לנו לדעת כמה פעמים עברנו במעגל כלשהו, ניתקל בבעיה.
בדיחה של רונן: אוטומטית סופית דטרמיניסטית יושבת בבר, ואז נכנס אליו אוטומט סופי דטרמיניסטי. הוא נשען על הבר ושואל "את באה לפה הרבה?" והאוטומטית הסופית הדטרמיניסטית עונה "אני לא יודעת."
(ניתן לקנות את זכויות היוצרים מרונן)

נראה דוגמה לסיטואציה בה חוסר הזיכרון מפריע לנו. נסתכל על הא"ב והשפה הבאים;

$$\Sigma = \Big\{a, b\Big\}, \quad L = \Big\{a^n b^n : n \geqslant 0\Big\}$$

טענה:
השפה L איננה רגולרית. ניעזר בכך שאס"ד איננו יכול לספור ובלמת הניפוח:
<u>למת הניפוח</u>
לכל שפה רגולרית L קיים מספר n_0 כך שלכל $w \in L$ עבורה $|w| \geqslant n_0$ קיימים $x, y, z \in \Sigma^*$ כך ש־ $w = xyz$ וגם:
1. $y \neq \epsilon$
2. $\forall k \geqslant 0 : \quad w' = xy^k z \in L$
3. $|xy| \leqslant n_0$

חשוב: למה זו היא לא אם"ם. כלומר, קיימות שפות לא רגולריות שמקיימות את התנאים בה.
נניח בשלילה ש־ L רגולרית.
אזי מהלמה קיים מספר n_0 כך שאם $|w| \geqslant n_0$ קיימים $x, y, z \in \Sigma^*$ שעבורם מתקיימים התנאים.
נסתכל על $w = a^{n_0} b^{n_0}$. מילה זו מקיימת $w \in L$ וגם $|w| = 2n_0 \geqslant n_0$ ולכן קיימות $x, y, z \in \Sigma^*$ כך ש־ $w = xyz$ ומתקיימים שלושת התנאים.
עקב תנאי 1, נקבל ש־ $y \neq \epsilon$ ומתנאי 3 נקבל ש־ $|xy| \leqslant n_0$ כך שב־ xy אין b ־ים.
נסתכל על תנאי 2 עם $k = 2$ ואז $w' = xyyz \in L$ וזו סתירה, מכיוון שנוספו a ־ים אבל לא b ־ים. לכן הטענה נכונה.

2.1 למת הניפוח ־ הוכחה

תהי L שפה רגולרית, כך שהיא מתקבלת ע"י אס"ד $M=(Q,\Sigma,\delta,q_0,F)$. נבחר $n_0=|Q|+2$ ותהי $w\in L$ כך ש־ $|w|\geqslant n_0$.
נסמן $|w|=n$ ונתחיל את מסלול החישוב של M על w. זוהי סדרה של $n+1$ מצבים.
כיוון שמדובר במצבים בסדרה שאורכה גדול־שווה ל־ n_0, כלומר גדול ממש מ־ $|Q|$, לפי עיקרון שובך היונים קיימים $i<j$ כך ש־ $r_i=r_j$.
נסתכל על הזוג הראשון $i<j$ שבו קורה מקרה זה ואז $i<i\leqslant n_0$.
נסמן $w=a_0,...,a_n$ כאשר $\forall i: a_i\in\Sigma$. נסמן

$$x=a_0,...,a_i \quad y=a_{i+1},...,a_j \quad z=a_{j+1},...,a_n$$

תחילה, נשים לב ש־ $i<j$ ולכן לא ייתכן ש־ $y\neq\epsilon$.
נסמן ב־ q את המצב שאליו מגיעים לאחר קריאת המילה x ב־M. לפי איך שהגדרנו את x,y,z, זהו גם המצב אליו מגיעים לאחר קריאת המילה y. מאחר שקריאת המילה y מהמצב q מחזירה לאותו המצב, לכל $k\geqslant 0$ נקבל ש־ k קריאות של y מהמצב q יחזירו אותנו למצב q ולכן עבור כל $k\geqslant 0$ כל מילה xy^kz תעבור בסוף במצבים ש־ w עוברת בהם כאשר z נקראת.
אז בסך הכל, מאחר ש־ $w\in L$ גם $\forall k\geqslant 0: \quad xy^kz\in L$
מאחר ש־ a_i,a_j הן שתי המילים הראשונות ב־ w שבעת קריאתן נמצאים באותו המצב, לא ייתכן ש־ $|xy|>|Q|+1$ ולכן $|xy|\leqslant n_0$ כנדרש.
כל תנאי למת הניפוח מתקיימים.

2.2 בעיית המספרים הראשוניים

נרצה לזהות שפה של מספרים ראשוניים;

$$\Sigma=\Big\{1\Big\}, \quad L=\Big\{1^n: \quad n \quad is \quad primary\Big\}$$

<u>טענה:</u>
השפה L איננה רגולרית.
("אס"ד לא יכול לספור")

<u>הוכחה:</u> נשתמש בלמת הניפוח ונוכיח ש־ L לא רגולרית.
מתנאי 1 אנו מקבלים ש־ $y\neq\epsilon$ ולכן $y=1^l: l>0$.
מתנאי 2;

$$w=xy^kz=1^{n+(k-1)l}$$

$$|w|=n+(k-1)l$$

כמו כן מתנאי 2 גם נובע שהמספרים הבאים ראשוניים; $(n+l)$ $(n+2l)$ $(n+3l)$ ...
אך עבור $k=n+1$ לא מתקבל מספר ראשוני, $(n+nl)=n(1+l)$. (ניתן לפרק אותו לשני מספרים שלמים שגדולים מ־1)
אזי L לא יכולה להיות רגולרית.

3 משפט מייהיל־נרוד

3.1 הגדרות

בהינתן שפה L מעל א"ב Σ נגדיר שני יחסים על מילים ב־ Σ^*:
יחס ההסכמה של L־

$$w_1\sim w_2 \Leftrightarrow (w_1,w_2\in L)\cup(w_1,w_2\notin L)$$

יחס השקילות של L־

$$w_1\equiv w_2 \Leftrightarrow \forall z\in\Sigma^*: \quad w_1z\sim_L w_2z$$

__טענה:__
M אס"ד ויהיו $w_1, w_2 \in \Sigma^*$. אם $w_1 \not\equiv_L w_2$ אזי $S_M(w_1) \neq S_M(w_2)$

__הוכחה:__
נניח בשלילה ש־ $w_1 \not\equiv_L w_2$ וגם $S_M(w_1) = S_M(w_2)$.
כלומר קיים $z \in \Sigma^*$ כך ש־ $w_1 z \not\sim_L w_2 z$. בה"כ נניח ש־ $w_2 z \notin L$, $w_1 z \in L$.
אבל $S_M(w_1 z) = S_M(w_2 z)$ כיוון ש־ $S_M(w_1) = S_M(w_2)$ וזו סתירה כי $S_M(w_1 z) \in F$, $w_1 z \in L$ ו־ $S_M(w_2 z) \notin F$ $w_2 z \notin L$.

3.2 מסקנות

מסקנה: תהי שפה L ויהיו $w_1, ..., w_t \in \Sigma^*$ כך שלכל $i \neq j$ מתקיים $w_i \not\equiv_L w_j$, כך שאין ל־ L אס"ד עם פחות מ־ t מצבים.
לדוגמה בשפת המילים

$$L = \left\{ 1^n | 0 \leqslant n \leqslant 1000 \right\}$$

בהכרח ישנם לפחות 1001 מצבים בכל אס"ד שמקבל את השפה.

מסקנה נוספת: תהי L שפה ויהיו $\{w_i\}_{i=1}^{\infty}$ מילים מתוכה, כך שלכל $i \neq j$ מתקיים $w_i \not\equiv w_j$. אז L לא רגולרית (היא צריכה יותר ממספר סופי של מצבים)

נשתמש במסקנה זו כדי למצוא הוכחה נוספת לכך שהשפה

$$L = \left\{ a^n b^n : n \geqslant 0 \right\}$$

היא לא רגולרית.
__הוכחה:__
מספיק להראות שישנם אינסוף מילים מ־ L שכל שתיים מהן לא שקולות.
נבחר קבוצת מילים מ־ L $\{w_i\}_{i=1}^{\infty}$ כך ש־ $w_i = a^i$.
נוכיח שלכל $i \neq j$ מתקיים $w_i \not\equiv_L w_j$.
לשם כך מספיק להראות שקיים $z \in \Sigma^*$ כך ש־ $w_i z \not\sim w_j z$. נבחר $z = b^i$;

$$a^j b^i (\notin L) = w_j z$$

$$w_i z = a^i b^i \in L$$

$$\Downarrow$$

$$w_i z \not\sim w_j z$$

כנדרש.

3.3 המשפט

נסמן ב־ $\#_L$ את מספר מחלקות השקילות של $\equiv_L$ כאשר $\#_L$ יכול להיות אינסוף.
L לא רגולרית אס"ם $\#_L = \infty$
ל־L קיים אס"ד עם t מצבים ואין אס"ד עם $t-1$ מצבים אס"ם $\#_L = t$.
שימי לב; משפט זה הוא אפיון של שפות רגולריות. כלומר, שלא כמו למת הניפוח, הוא מתקיים רק עבור שפות רגולריות.

3.4 הוכחה

כבר הוכחנו שאם לשפה יש אינסוף מחלקות שקילות היא לא רגולרית, כך שאם היא רגולרית אז יש לה מספר סופי של מחלקות שקילות. הוכחנו גם שאין לשפה L אס"ד עם פחות מ־ $\#_L$ מצבים.
כלומר כבר הוכחנו חלק משני חלקי המשפט. מספיק לנו להוכיח את הטענה הבא;
טענה:
אם $\#_L = t$ ו־ t סופי, אז ל־ L קיים אס"ד עם t מצבים שמקבל את השפה.
הוכחה:
תהי שפה L עבורה $\#_L = t$ כך ש־ t סופי. נוכיח שקיים אס"ד עם t מצבים שמקבל את L.
הרעיון הוא שקבוצת המצבים תכיל (ותוכל) במחלקות השקילות של $\equiv_L$.
בנייה: הקבוצה Q הינה קבוצה של מחלקות השקילות (לכל מחלקת שקילות יש מצב)

$$q_0 = [\epsilon]_L \quad F = \Big\{[w]_L : w \in L\Big\}$$

מצבי האס"ד מוגדרים היטב מכיוון שלכל מילה מחלקת שקילות יחידה, ומילים שנמצאות באותה מחלקת שקילות מסכימות, כלומר כולן שייכות ל־ L או לא שייכות ל־ L.
נשאר רק להגדיר את פונקציית המעבר;

$$\delta([u]_L, a) = [u \bigcirc a]_L$$

כאשר u זוהי המילה שנקראה עד כה (היא לא נשמרת בזיכרון, אך המצב $[u]_L$ כן) נוודא ש־ δ מוגדרת היטב;
כלומר יש לבדוק האם $w \equiv_L w'$ גורר $[wa]_L = [w'a]_L$.
גרירה זו נכונה מכיוון שלכל $z \in \Sigma^*$ מתקיים $w(az) \sim_L w'(az)$, כך ש־ $(wa)z \sim_L (w'a)z$.

טענת עזר:
לכל $w \in \Sigma^*$ מתקיים $S_M(w) = [w]_L$.
ניתן להוכיח זאת באינדוקציה (תרגיל לקורא :)

4 סיכום

הוכחנו סופית את השקילות בין אס"דים, אסל"דים ורגולריות. אבל גם הבנו שאוטומטים סופיים הם לא הכי חזקים, ואין להם אפילו את היכולת לספור.
למדנו על שפות שונות שאינן רגולריות ואיך להוכיח את אי־הרגולריות שלהן.