מודלים חישוביים
תרגול מס' 5

29 במאי 2015

נושאי התרגול:

- דקדוקים חסרי הקשר.
- למת הניפוח לשפות חסרות הקשר.
- פעולות סגור לשפות חסרות הקשר.

1 דקדוקים חסרי הקשר

נזכיר כי דקדוק חסר הקשר הוא רביעיה $G = (V, \Sigma, R, S)$, כך ש:

- V היא קבוצת סופית של משתנים (בד"כ נסמנם באותיות אנגליות גדולות).
- Σ היא קבוצה סופית של ליטרלים, זרה ל־ V (בד"כ נסמנם באותיות אנגליות קטנות).
- R היא קבוצה של כללי גזירה כך שכל כלל הוא מהצורה $A \to \alpha$ עבור $A \in V$ ו־ $\alpha \in (V \cup \Sigma)^\star$.
- $S \in V$ הוא המשתנה ההתחלתי.

הגדרה 1.1 יהיו $u, v, w \in (V \cup \Sigma)^\star$ כך ש־ $A \to w$ הוא כלל ב־ R. נוכל נסמן במקרה זה $uAv \Rightarrow uwv$, ובאופן כללי נסמן $u \Rightarrow^\star v$ אם ניתן לעבור מ־ u ל־ v ע"י מספר סופי של הפעלות כללים מ־ R. השפה הנוצרת ע"י G מוגדרת ע"י $L(G) = \{w \in \Sigma^\star \mid S \Rightarrow^\star w\}$.

לשפות המתקבלות ע"י דקדוקים חסרי הקשר כנ"ל אנו קוראים שפות חסרות הקשר. ניתן להגביל את כללי הגזירה באופן כזה שקבוצת השפות המתקבלות יהיו השפות הרגולריות. לדקדוקים כאלו אנו קוראים דקדוקים רגולריים. פירוט נוסף על כך ניתן בהרצאות.

דוגמא 1

נראה דקדוק המקבל את השפה:

$$L_1 = \left\{w \in \{a, b\}^\star \mid \forall u \in \mathrm{Suffix}(w) . \#_a(u) \le \#_b(u)\right\}$$

כלומר, בכל סיפא של w ב־ L_3, מספר ה־ b־ים הוא לפחות כמו מספר ה־ a־ים. הדקדוק $G = (V, \Sigma, R, S)$ יהיה:

- $V = \{S\}$.
- $\Sigma = \{a, b\}$.
- R יכיל את הכלל: $S \to SaSb \mid Sb \mid \epsilon$.

דוגמא לגזירה של מילה בשפה:

$$S \Rightarrow SaSb \Rightarrow SaSbaSbb \Rightarrow abaSbbb \Rightarrow ababbb$$

תרגיל 1

הוכיחו כי השפה $L_2 = \left\{a^i b^j c^{i+j} \mid i,j \geq 0\right\}$ היא חסרת הקשר.

פתרון

נבנה עבורה דקדוק חסר הקשר $G = (V, \Sigma, R, S)$ ונראה כי $L(G) = L_2$ ע"י הכלה דו כיוונית. באופן כללי, את הכיוון $L \subseteq L(G)$ בד"כ נוכיח באינדוקציה על אורך המילה ואת הכיוון $L(G) \subseteq L$ נוכיח באינדוקציה על אורך הגזירה. הדקדוק עבור L_2 יהיה:

- $V = \{S, T\}$.
- $\Sigma = \{a, b, c\}$.
- R יכיל את הכללים:

$$\begin{array}{rcl} S & \rightarrow & aSc \mid T \\ T & \rightarrow & bTc \mid \epsilon \end{array}$$

הכיוון הראשון שנוכיח יהיה $L_2 \subseteq L(G)$. תהא $w = a^i b^j c^{i+j} \in L_2$. נראה כי $S \Rightarrow^\star w$ ב־ G. סדר הגזירה יהיה i הפעלות של הכלל aSc, מעבר ל־ T, j הפעלות של הכלל bTc ולבסוף גזירת ϵ. כלומר:

$$\begin{array}{rcl} S & \Rightarrow & aSc \Rightarrow a^2Sc^2 \Rightarrow \ldots \Rightarrow a^iSc^i \\ & \Rightarrow & a^iTc^i \\ & \Rightarrow & a^ibTc^{i+1} \Rightarrow a^ib^2Tc^{i+2} \Rightarrow \ldots \Rightarrow a^ib^jTc^{i+j} \\ & \Rightarrow & a^ib^jc^{i+j} = w \end{array}$$

ולכן $S \Rightarrow^\star w$ (שימו לב כי היתה כאן אינדוקציה "מובלעת" שחסכנו אותה. איזו?).

בכיוון השני, תהא $w \in L(G)$. כלומר, קיימת סדרה $\alpha_0, \ldots, \alpha_n$ כך ש־

$$S = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \ldots \Rightarrow \alpha_n = w$$

הלמה הבאה תשלים את הוכחת הכיוון השני:

למה 1.2 לכל $0 \leq k \leq n$, α_k הוא באחת מן הצורות הבאות:

1. a^iSc^i עבור $i \geq 0$.
2. $a^ib^jTc^{i+j}$ עבור $i,j \geq 0$.
3. $a^ib^jc^{i+j}$ עבור $i,j \geq 0$.

הוכחת הלמה תשלים את הוכחת הכיוון השני מכיוון שהצורה היחידה שבה המחרוזת מורכבת מליטרלים בלבד היא 3, וכל מחרוזת ב־ 3 שייכת ל־ L_2. ההוכחה היא באינדוקציה על k, אורך הגזירה:

- עבור $k = 0$ מתקיים ש־ $\alpha_0 = S$ הוא מצורה 1.
- נניח את נכונות הלמה עבור $k - 1$ ונפריד למקרים לפי הצורה של α_{k-1}.
  - אם $\alpha_{k-1} = a^iSc^i$, ניתן להשתמש בכלל $S \rightarrow aSc$ כך ש־ $\alpha_k = a^{i+1}Sc^{i+1}$ ולהיות מצורה 1 או להשתמש בכלל $S \rightarrow T$ כך ש־ $\alpha_k = a^iTc^i$ ולהיות מצורה 2.
  - אם $\alpha_{k-1} = a^ib^jTc^{i+j}$, ניתן להשתמש בכלל $T \rightarrow bTc$ כך ש־ $\alpha_k = a^ib^{j+1}Tc^{i+j+1}$ ולהיות מצורה 2 או להשתמש בכלל $T \rightarrow \epsilon$ כך ש־ $\alpha_k = a^ib^jc^{i+j}$ ולהיות מצורה 3.
  - המקרה $\alpha_{k-1} = a^ib^jc^{i+j}$ אינו אפשרי, שכן אין משתנים שיכולים להוביל לגזירת α_k.

2 למת הניפוח לשפות חסרות הקשר

נזכר בלמת הניפוח לשפות ח"ה, שגם כאן תהווה עבורנו טכניקה להוכחה כי שפה אינה ח"ה:

למה 2.1 לכל שפה ח"ה L קיים $\ell > 0$ כך שלכל $s \in L$ המקיימת $|s| \geq \ell$ קיים פירוק מהצורה $s = uvxyz$ כך ש:

1. לכל $i \geq 0$, $uv^ixy^iz \in L$.
2. $|vy| > 0$.
3. $|vxy| \leq \ell$.

איך נשתמש בלמת הניפוח כדי להוכיח ששפה L כלשהי האינה ח"ה? כמו שעשינו עבור למת הניפוח לשפות רגולריות: נניח בשלילה ש־ L ח"ה ויהא ℓ המובטח לנו. נבחר מילה $s \in L$ באורך גדול מ־ ℓ ונראה ש<u>לכל</u> חלוקה $uvxyz$ של s כך ש־ $|vy| > 0$ ו־ $|vxy| \leq \ell$ קיים $i \geq 0$ כך ש־ $uv^ixy^iz \notin L$, בסתירה ללמת הניפוח.

תרגיל 2

הוכיחו כי השפה הבאה אינה ח"ה:

$$L_3 = \left\{a^ib^jc^k \mid j = \max(i,k)\right\}$$

פתרון

ניח בשלילה ש־ L_3 ח"ה ויהא ℓ המובטח לנו. נבחר:

$$s = a^\ell b^\ell c^\ell \in L_3$$

ונשים לב ש־ $|s| \geq \ell$. נסתכל על כל הפירוקים האפשריים של s ל־ $uvxyz$ כך ש־ $|vy| > 0$ ו־ $|vxy| \leq \ell$. מכך ש־ $|vxy| \leq \ell$, גם v וגם y לא יכולות להכיל יותר משני סוגים שונים של תווים. נפריד למקרים:

- אם vy מכילה את האות b, נסתכל על

$$w = uv^0xy^0z = uxz$$

וקיבלנו שמספר ה־ b־ים בהכרח קטן אך או שמספר ה־ a־ים לא השתנה או שמספר ה־ c־ים לא השתנה (מהאבחנה הקודמת שלנו). מכאן, ש־ $j < \max(i,k) = \ell$ ולכן $w \notin L_3$.

- אם vy אינה מכילה את האות b אז vxy כולה a־ים או כולה c־ים. נסתכל על

$$w = uv^2xy^2z$$

מספר ה־ b־ים נשאר ℓ, אך בהכרח מספר ה־ a־ים גדלו או מספר ה־ c־ים גדלו. מכאן, ש־ $\ell = j < \max(i,k)$ ולכן $w \notin L_3$.

בכל חלוקה אפשרית קיים i כך ש־ $uv^ixy^iz \notin L_3$, בסתירה ללמת הניפוח. לכן, L_3 אינה ח"ה.

תרגיל 3

הוכיחו כי השפה הבאה אינה ח"ה:

$$L_4 = \{1^n \mid n \text{ is prime}\}$$

פתרון

ניח בשלילה ש־ L_3 ח"ה ויהא ℓ המובטח לנו. נבחר $s = 1^m$ כך ש־ m הוא הראשוני הקטן ביותר שהוא לפחות ℓ. מהגדרתינו, $s \in L_4$ ו־ $|s| \geq \ell$. נסתכל על כל הפירוקים האפשריים של s ל־ $uvxyz$ כך ש־ $|vy| > 0$ ו־ $|vxy| \leq \ell$. נסמן $v = 1^k$ ו־ $y = 1^t$. מתקיים אם כך ש־ $t + k > 0$. נבחר $i = m + 1$, ואז:

$$w = uv^ixy^iz = 1^{m+m(k+t)}$$

ומכיוון ש־ $m + m(k+t) = m(1+k+t)$, ברור כי הוא אינו ראשוני ולכן $w \notin L_4$.

תרגיל 4

הוכיחו כי השפה $L_5 = \left\{a^{3j}b^{2^k} \mid j,k > 0\right\} \cup \{b\}^\star$ מקיימת את תנאי למת הניפוח לשפות ח"ה.

פתרון

נראה כי L_5 מקיימת את הלמה עבור קבוע הניפוח $\ell = 3$. תהא $s \in L_5$ כך ש־ $|s| \geq 3$. נפריד למקרים:

- אם $s \in \{b\}^\star$ אז $s = b^m$ עבור $m \geq 3$. נבחר פירוק $u = \epsilon$, $v = b$, $x = y = \epsilon$ ו־ $z = b^{m-1}$. נראה כי הפירוק מקיים את תנאי הלמה:
  - $|vxy| = 1 \leq \ell$.
  - $|vy| = 1 > 0$.
  - לכל $i \geq 0$, $uv^ixy^iz = b^{i+m-1} \in \{b\}^\star \subseteq L$.
- אם $s \notin \{b\}^\star$ הרי ש־ $s \in \left\{a^{3j}b^{2^k} \mid j,k > 0\right\}$. כלומר, קיימים $j,k > 0$ כך ש־ $s = a^{3j}b^{2^k}$. נבחר פירוק $u = \epsilon$, $v = a^3$, $x = y = \epsilon$ ו־ $z = a^{3(j-1)}b^{2^k}$. נראה כי הפירוק מקיים את תנאי הלמה:
  - $|vxy| = 3 \leq \ell$.
  - $|vy| = 3 > 0$.
  - לכל $i \geq 0$,

$$uv^ixy^iz = a^{3i}a^{3(j-1)}b^{2^k} = a^{3(i+j-1)}b^{2^k} \in \left\{a^{3j}b^{2^k} \mid j,k > 0\right\} \subseteq L$$

3 פעולות סגור לשפות חסרות הקשר

כפי שדנו בפעולות אשר השפות הרגולריות סגורות עבורן, ניתן לשאול את אותן השאלות על שפות חסרות הקשר עם טכניקות הוכחה (או הפרכה) דומות.

תרגיל 5

הוכיחו כי השפות חסרות ההקשר סגורות תחת היפוך. כלומר, עבור שפה ח"ה L מעל Σ, הוכיחו כי $rev(L) = \left\{w^R \mid w \in L\right\}$ היא ח"ה.

פתרון

תהא L שפה ח"ה מעל Σ, אזי קיים דקדוק ח"ה $G = (V, \Sigma, R, S)$ כך ש־ $L = L(G)$. נבנה $G^R = \left(V, \Sigma, R^R, S\right)$ כך שעבור כל כלל מהצורה $A \to \alpha$ ב־ R, נוסיף את הכלל $A \to \alpha^R$ ל־ R^R. ואז:

$$\begin{aligned} w^R \in rev(L) \quad &\Leftrightarrow \quad w \in L \\ &\Leftrightarrow \quad w \in L(G) \\ &\Leftrightarrow \quad w^R \in L\left(G^R\right) \end{aligned}$$

כך שהמעבר האחרון, דהיינו $S \Rightarrow^\star w$ ב־ G אם"ם $S \Rightarrow^\star w^R$ ב־ G^R, דורש הוכחה נפרדת. בכיוון הראשון, נוכיח טענה חזקה יותר: לכל $A \in V$, אם $A \Rightarrow^\star w$ ב־ G אז $A \Rightarrow^\star w^R$ ב־ G^R. נוכיח באינדוקציה על k, אורך הגזירה.

- עבור $k=1$ - טריוויאלי.
- נניח את הנכונות לגזירות באורך לכל היותר $k-1$ ונניח ב־ G כי

$$A \Rightarrow^{(k-1)} w_1A_1w_2\cdots w_nA_nw_{n+1} \Rightarrow w_1x_1w_2\cdots w_nx_nw_{n+1} = w$$

עבור $n \geq 1$, $A_1,\ldots,A_n \in V$ ו־ $x_1,\ldots,x_n \in \Sigma^\star$. לפי הנחת האינדוקציה, $A_i \Rightarrow^{\leq(k-1)} x_i^R$ ב־ G^R. מהגדרת G^R, מתקיים כי:

$$A \Rightarrow w_{n+1}^RA_nw_n^R\cdots w_2^RA_1w_1^R \Rightarrow^{(k-1)} w_{n+1}^Rx_n^Rw_n^R\cdots w_2^Rx_1^Rw_1^R = w^R$$

כנדרש.

הכיוון השני, ש־ $S \Rightarrow^\star w^R$ ב־ G^R גורר $S \Rightarrow^\star w$ ב־ G, נובע ישירות מהכיוון הראשון עבור הדקדוק G^R במקום G, כי $\left(G^R\right)^R = G$. שימו לב כי היינו יכולים גם להניח כי G הוא בצורה הנורמלית של חומסקי (CNF) ובכך לפשט מעט את ההוכחה.