מודלים חישוביים
תרגול מס' 5

6 במאי 2017

נושאי התרגול:

- דקדוקים חסרי 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 $uAv \to uwv$, ובאופן כללי נסמן $u \to^\star v$ אם ניתן לעבור מ־ u ל־ v ע"י מספר סופי של הפעלות כללים מ־ R. השפה הנוצרת ע"י G מוגדרת ע"י $L(G) = \{w \in \Sigma^\star \mid S \to^\star w\}$.

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.

דוגמא 1

נראה דקדוק היוצר 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_1, 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:

$$S \to SaSb \to SaSbaSbb \to abaSbbb \to ababbb$$

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 $L\left(G\right) = L_2$ ע"י הכלה דו כיוונית. באופן כללי, את הכיוון $L \subseteq L\left(G\right)$ בד"כ נוכיח באינדוקציה על אורך המילה ואת הכיוון $L\left(G\right) \subseteq L$ נוכיח באינדוקציה על אורך הגזירה. הדקדוק עבור L_2 יהיה:

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

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

הכיוון הראשון שנוכיח יהיה $L_2 \subseteq L\left(G\right)$. 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:

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

ולכן $S \rightarrow^\star w$ (שימו לב כי היתה כאן אינדוקציה "מובלעת" שחסכנו אותה. איזו?).
בכיוון השני, תהא $w \in L\left(G\right)$. 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 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 2

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

פתרון

תהא L שפה ח"ה מעל Σ, אזי קיים דקדוק ח"ה $G=\left(V,\Sigma,R,S\right)$ כך ש־ $L=L\left(G\right)$. 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:

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

כך שהמעבר האחרון, דהיינו $S\to^{\star}w$ ב־ G אס"ם $S\to^{\star}w^{R}$ ב־ G^{R}, דורש הוכחה נפרדת. בכיוון הראשון, נוכיח טענה חזקה יותר: לכל k טבעי ו־ $A\in V$, אם $A\to^{(k)}w$ ב־ G אז $A\to^{(k)}w^{R}$ 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

$$A\to^{(k-1)}w_{1}A_{1}w_{2}\cdots w_{n}A_{n}w_{n+1}\to w_{1}x_{1}w_{2}\cdots w_{n}x_{n}w_{n+1}=w$$

עבור $n\geq1$, $A_{1},\dots,A_{n}\in V$ ו־ $x_{1},\dots,x_{n}\in\Sigma^{\star}$. לפי הבנייה, $A_{i}\to x_{i}^{R}$ ב־ G^{R}. מהנחת האינדוקציה, מתקיים ב־ G^{R} כי:

$$A\to^{(k-1)}w_{n+1}^{R}A_{n}w_{n}^{R}\cdots w_{2}^{R}A_{1}w_{1}^{R}\to w_{n+1}^{R}x_{n}^{R}w_{n}^{R}\cdots w_{2}^{R}x_{1}^{R}w_{1}^{R}=w^{R}$$

כנדרש.

הכיוון השני, ש־ $S\to^{\star}w^{R}$ ב־ G^{R} גורר $S\to^{\star}w$ 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.

3 הצורה הנורמלית של חומסקי

בהרצאות ראינו את הצורה הנורמלית של חומסקי (CNF) לדקדוקים חסרי הקשר, שלמשל עזרה לנו בלמצוא אלגוריתם לבעית השייכות (בהנתן דקדוק G, משתנה A ומילה w, האם $A\Rightarrow^{\star}w$ ב־ G). נזכיר שדקדוק הוא ב־ CNF אם כל כלל גזירה הוא מאחת מהצורות הבאות: $S\to\epsilon$ (כך ש־ S משתנה התחלתי), $A\to a$ (עבור משתנה A כלשהו וליטרל a כלשהו) או $A\to BC$ (עבור משתנים כלשהם A,B,C).

כמו כן, ראינו כי כל דקדוק חסר הקשר ניתן להמרה ל־ CNF ע"י הצעדים הבאים:

1. הוספת משתנה התחלתי חדש.
2. המרת "כללים ארוכים" (תחילה טיפול בליטרלים ואז המרת כללים עם משתנים).
3. ביטול כללי ϵ מהצורה $A\to\epsilon$.
4. ביטול כללי זהות מהצורה $A\to B$.

נדגים על הדקדוק $G=(V=\{S,X,Y\},\Sigma=\{a,b,c\},R,S)$ כאשר R מכיל את:

$$\begin{aligned} S &\to aXbX \\ X &\to aY \mid bY \mid \epsilon \\ Y &\to X \mid c \end{aligned}$$

לאחר הוספת משתנה התחלתי חדש S_0 וביצוע שלב 2 לטרמינלים, נקבל את הכללים הבאים:

$$\begin{aligned} S_0 &\to S \\ S &\to AXBX \\ X &\to AY \mid BY \mid \epsilon \\ Y &\to X \mid C \\ A &\to a \\ B &\to b \\ C &\to c \end{aligned}$$

לאחר ביצוע שלב 2 למשתנים, נקבל:

$$\begin{aligned} S_0 &\to S \\ S &\to AN_1 \\ N_1 &\to XN_2 \\ N_2 &\to BX \\ X &\to AY \mid BY \mid \epsilon \\ Y &\to X \mid C \\ A &\to a \\ B &\to b \\ C &\to c \end{aligned}$$

לאחר ביצוע שלב 3, נקבל:

$$\begin{aligned} S_0 &\to S \\ S &\to AN_1 \\ N_1 &\to XN_2 \mid N_2 \\ N_2 &\to BX \mid B \\ X &\to AY \mid BY \mid B \mid A \\ Y &\to X \mid C \\ A &\to a \\ B &\to b \\ C &\to c \end{aligned}$$

לבסוף (אחרי שלב 4), הכללים שנקבל הם:

$$\begin{aligned} S_0 &\to AN_1 \\ N_1 &\to XN_2 \mid BX \mid b \\ N_2 &\to BX \mid b \\ X &\to AY \mid BY \mid b \mid a \\ Y &\to AY \mid BY \mid b \mid a \mid c \\ A &\to a \\ B &\to b \\ C &\to c \end{aligned}$$