אוטומטים 2 - הרצאה 8

אורן דנון

1 פישוטים

- סילוק מסעי ε
- סילוק כללי יחידה

בשיעור הקודם דיברנו על פישוטים שונים של דקדוקים חסרי הקשר. לצורך פשטות נניח היום שמדובר בדקדוק עבור שפה L שאינה מכילה את ε. ואם נבצע סילוק מסעי אפסילון, נקבל דקדוק ללא כללים הגוזרים את אפסילון, בפרט ($s \to \varepsilon$ לא נמצא) וכן נבצע סילוק כללי יחידה, נתבונן בדקדוק שמתקבל G באופן כללי, הדקדוק G מכיל כללים מהצורה $A \to X_1, X_2, ..., X_t$, כאשר $t \geq 2$ וכל $X_i \in V \cup T$. בנוסף, עבור $t = 1$ ייתכנו כללים מהצוקה $A \to X_1$ כאשר $X_i \in T$. נרצה לבנות דקדוק שקול מהצורה הנורמלית של חומסקי.

הגדרה:

דקדוק חסר הקשר הוא מהצורה הנורמלית של חמסקי אם כל כלל ב P הוא מהצורה $A \to BC \;\; a \in T, \;\; A \in V \;\; A, B, C \in V$ או ש $A \to a$

1.0.1 טענה:

לכל דקדוק חסר הקשר עם $\varepsilon \in L(G)$ קיים דקדוק חסר הקשר שקול G' מהצורה הנורמלית של חומסקי.

הוכחה

נתחיל מדקדוק מהצורה 1 נבנה את G' בשני שלבים:

1. נבנה G' שקול, כך שבצד ימין של כל כלל $A \to X_1X_2, ...X_n$ תהיה מחרוזת באורך $t = 1$ או $t = 2$. עבור כלל מהצורה $A \to X_1 \in T$ נקח את הכלל כמו שהוא ונוסיף ל P' כנ"ל עבור $A \to X_1, X_2$ עבור $A \to X_1, X_2, ..., X_t$ כאשר $t \geq 2$ נחליף את הכלל במספר כללים חדשים: נכניס משתנים חדשים $B_{2,t}$ ונוסיף את הכללים $A \to X_1B_{2,t}$, $B_{2,t} \to X_2, ..., X_t$. עכשיו בצורה רקורסיבית נחליף את הכלל $B_{2,t} \to X_2, ..., X_t$ ב $B_{2,t} \to X_2B_{3,t}$ כאשר $B_{3,t} \to X_3, ..., X_t$. נעצור את התהליך ב $B_{t-1,t} \to X_{t-1}X_t$ לא קשה לראות ש G'' שהתקבל שקול ל G המקורי. נשאר לטפל בכללים מהצורה $A \to X_1, X_2$ כאשר אחד ה X_i־ים הוא טרמינל. זה לא מהצורה הנורמלית של חומסקי. לכל טרמינל X_i כזה נכניס משתנה חדש B_i ונחליף את הכלל $A \to X_1, X_2$ בכללים: $B_i \to X_i$ (זה בסדר כי X_i טרמינל) ואת $A \to X_1, X_2$ בכללים $A \to C_1, C_2$ כאשר $C_1 = B_i$ אם X_i טרמינל ו־ $C_i = B_i$ אם X_i טרמינל ו־ $C_i = X_i$ אחרת. לא קשה לראות ש G' שהתקבל דח"ה בצורה הנורמלית של חומסקי.

2 למת הניפוח לשפות חסרות הקשר

נפתח כלי להוכחה ששפה מסוימת אינה ח"ה. שימו לב שמשוקלי ספירה ניתן להראות שקיימות שפות שאינן חסרות הקשר, אבל זה לא עוזר לנו להחליט האם שפה מסוימת היא חסרת הקשר או לא (אם כן אז בדר"כ קל להראות זאת) אחרת לא ברור כיצד להוכיח זאת רק מההגדרה של שפות חסרות הקשר. למת הניפוח מספקת תכונה של שפות ח"ה שיחסית קל לבדוק, אם שפה לא מקיימת אותה אז היא בהכרח לא חסרת הקשר. אבל קיימות גם שפות לא חסרות הקשר המקיימות את התכונה, כלומר שיטת ההוכחה לא תעבוד תמיד.

2.1 למת הניפוח לשפות חסרות הקשר:

תהי L שפה חסרת הקשר אז קיים $n \in \mathbb{N}$ כך שלכל $z \in L$ המקיימת $n \le |Z|$ קיים פירוק מהצורה $z = uvwxy$ כאשר

1. $|vwx| \le n$
2. $|vx| \ge 1$
3. $\forall i \ge 0 \ z^i = uv^iwx^iy \in L$

כיצד נוכיח ש L נתונה אינה חסרת הקשר ע"פ הלמה? נראה שלכל n קיים $z \in L, \ |Z| \ge 0$ כך שלכל פירוק $z = uvwxy$ המקיים את $1, 2$ תכונה שלישית לא מתקיימת כלומר קיים i כך ש $z^i \notin L$.

2.1.1 הוכחה

תהי L שפה חסרת הקשר ויהי G דקדוק חסר הקשר עבור $L - \{0\}$ מהצורה הנורמלית של חומסקי יהיה k מספר המשתנים ב־G.
נראה ש־ L מקיימת את הלמה.

- עבור $n = 2^k$ נקח את $z \in L$ כלשהו נתבונן בעץ הגזירה של z.

ידוע גם שדרגת כל צומת פנימי היא אחד או שתיים. מה גובה העץ בקשתות? בהכרח מתקיים $h \ge lg|z| + 1$ בקשתות. עבור $|z| \ge 2^h$ ונקבל ש $h \ge k+1$. בצמתים גובה העץ הוא $h \ge k+2$ נתבונן בעץ T' בו קצצנו את העלים. נתבונן במסלול הארוך ביותר מהשורש לעלה: מסלול זה מכיל לפחות $k+1$ צמתים, נבצע טיול מהעלה במסלול הזה כלפי מעלה, כלומר נלך מהעלה במסלול כלפי מעלה, ונעצור כאשר נתקלנו לראשונה באותו משתנה פעמיים נסמן A. נסמן את הצמתים ב V_1, V_2 בהתאמה. נתבונן בתתי העצים עם שורש V_1, V_2 בהתאמה. נשים לב שהגובה של V_1 הוא לכל היותר $k+1$ באורך הטיול מהעלה ל V_1 המסלול שבחרנו אורכו $\ge k+1$ כי יש ב k משתנים שונים סה"כ(ע"פ שובך היונים), נשים לב שלא ייתכן מסלול יותר ארוך מהעלה של V_1 ל־ V_1 כי אחרת היינו מקבלים מסלול ארוך יותר בכל העץ. נסמן את החזית של T_1 ב־ vwx נסמן את החזית של T' ב־ $uvwxy$ (כאן החזרנו את העלים ל־T) נראה ש $z = uvwxy$ אכן מקיים את $1, 2, 3$ עבור $n = 2^k$.

1. vwx הוא החזית של T_1 הראינו שהגובה של T_1(קצוץ) הוא לכל היותר $k+1$ קודקודים $\Leftarrow k$ קשתות $\underbrace{\Leftarrow}_{\text{במקרה ש־}T_1\text{ עץ בינארי מלא, מקבלים חזית מקסימלית}}$ בחזית יש פחות מ 2^k עלים כיוון שכל עלה ב T הוא בן יחיד נסיק כי $|vwx| \le 2^k$
2. נראה ש $vx \ge 1$ מדוע זה נכון? נתבונן ב T_1:
   בהכרח ל V_1 יש שני בנים בדיוק המסומנים במשתנה כי יש לו צאצא V_2 שגם מסומן במשתנה ו־ G בצורה הנורמלית של חומסקי $\Leftarrow V_2$ צאצא של אחד הבנים של V_1 או שהוא עצמו בן אם V_2 צאצא של הבן הימני אז לבן השמאלי ישנו צאצא טרמינלי אחד לפחות אז $V \neq \varepsilon$ אחרת (V_2 צאצא של הבן השמאלי) אז בהכרח x לא ריק.
3. נוכיח שלכל $uv^iwx^iy \in L$