Formális nyelvek - 6.

Csuhaj Varjú Erzsébet

Algoritmusok és Alkalmazásaik Tanszék
Informatikai Kar
Eötvös Loránd Tudományegyetem
H-1117 Budapest
Pázmány Péter sétány 1/c

E-mail: csuhaj@inf.elte.hu

Hossz-nemcsökkentő grammatikák

Definíció

Egy $G = (N, T, P, S)$ 0-típusú grammatikát hossz-nemcsökkentőnek mondunk, ha bármely $u \to v \in P$ szabályra $|u| \leq |v|$ teljesül.

Tétel

Minden hossz-nemcsökkentő grammatika környezetfüggő nyelvet generál.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ egy tetszőleges hossz-nemcsökkentő grammatika. Az általánosság megszorítása nélkül feltehetjük, hogy terminális szimbólum csak $A \to a$ alakú szabályban szerepel a P szabályhalmazban, ahol A nemterminális és a terminális.

Legyen $u \to v$ olyan szabály P-ben, amelyre $|u| \geq 2$ és $u \to v$ alakja legyen

$$X_1X_2 \ldots X_m \to Y_1Y_2 \ldots Y_n,$$

$n \geq m \geq 2$. Akkor az előbbi megjegyzés alapján $X_i, Y_j \in N$, $1 \leq i \leq m$, $1 \leq j \leq n$.

Legyen $n > m$. Helyettesítsünk minden

$$X_1X_2 \ldots X_m \to Y_1Y_2 \ldots Y_n$$

alakú szabályt a következő szabályhalmazzal, ahol $Z_i \notin N$, $1 \leq i \leq m$, új, a szabályhoz bevezetett nemterminálisok. (Különböző szabályokhoz páronként diszjunkt új nemterminálishalmazokat vezetünk be.)

Bizonyításvázlat - folytatás:

$$X_1X_2\dots X_m \to Y_1Y_2\dots Y_n$$

Helyettesítve:

$$\begin{aligned}
X_1X_2\dots X_m &\to Z_1X_2\dots X_m \\
Z_1X_2\dots X_m &\to Z_1Z_2\dots X_m \\
&\vdots \\
Z_1\dots Z_{m-1}X_m &\to Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n \\
Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n \\
&\vdots \\
Y_1\dots Y_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Y_{m-1}Y_mY_{m+1}\dots Y_n \\
&.
\end{aligned}$$

Az $n = m$ eset értelemszerű módosítással adódik.

Könnyen látható, hogy így minden egyes hossz-nemcsökkentő szabályt helyettesíthetünk környezetfüggő szabályok halmazával úgy, hogy a generált nyelv nem változik.

Bizonyításvázlat - szemléltetés:

$X_1X_2\dots X_m \Longrightarrow Z_1X_2\dots X_m \Longrightarrow Z_1Z_2\dots X_m$

$\Longrightarrow^* Z_1Z_2\dots Z_{m-1}X_m \Longrightarrow Z_1Z_2\dots Z_{m-1}Z_mY_{m+1}\dots Y_n$

$\Longrightarrow^* Y_1Y_2\dots Y_{m-1}Z_mY_{m+1}\dots Y_n \Longrightarrow Y_1Y_2\dots Y_{m-1}Y_mY_{m+1}\dots Y_n.$

$$\begin{aligned}
X_1X_2\dots X_m &\to Z_1X_2\dots X_m\\
Z_1X_2\dots X_m &\to Z_1Z_2\dots X_m\\
&\vdots\\
Z_1\dots Z_{m-1}X_m &\to Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n\\
Z_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Z_{m-1}Z_mY_{m+1}\dots Y_n\\
&\vdots\\
Y_1\dots Y_{m-1}Z_mY_{m+1}\dots Y_n &\to Y_1\dots Y_{m-1}Y_mY_{m+1}\dots Y_n\\
&.
\end{aligned}$$

Korollárium

$$\mathcal{L}_2 \subset \mathcal{L}_1.$$

Bizonyítás: Az $L = \{a^n b^n c^n \mid n \geq 1\}$ nyelv hossz-nemcsökkentő grammatikával generálható, így környezetfüggő, viszont nem generálható környezetfüggetlen grammatikával.

Megjegyzés: A hossz-nemcsökkentő tulajdonság íly módon ekvivalens a környezetfüggőséggel, azzal a kivétellel, hogy környezetfüggő grammatikák esetében az $S \to \varepsilon$ szabály megléte megengedett.

Megjegyzések a korolláriumhoz

Legyen $G = (\{S, A, B\}, \{a, b, c\}, P, S)$, ahol

$$\begin{aligned}\{S \to abc, &\quad S \to aAbc, \quad Ab \to bA,\\ Ac \to Bbcc, &\quad bB \to Bb, \quad aB \to aaA,\\ aB \to aa\}\end{aligned}$$

Akkor $L = \{a^n b^n c^n \mid n \geq 1\}$ és G hossz-nemcsökkentő. Legyen $S \Longrightarrow^* a^i A b^i c^i$. Akkor alkalmazzuk a 3. szabályt, majd a 4. szabályt és az $a^i b^i B b c^{i+1}$ szót kapjuk. Ezután csak az 5. szabályt alkalmazhatjuk és $a^i B b^{i+1} c^{i+1}$-hez jutunk. Ebből a szóból vagy az $a^{i+1} A b^{i+1} c^{i+1}$ szót vagy az $a^{i+1} b^{i+1} c^{i+1}$ szót nyerhetjük.

Megjegyzések a korolláriumhoz

Megmutatjuk, hogy L nem környezetfüggetlen. Tegyük fel az ellenkezőjét. Legyen $w = a^q b^q c^q \in L$, ahol p, q a Bar-Hillel lemmának megfelelő konstansok. Nyilvánvaló, hogy $|w| > q > p$. (Lásd a lemma bizonyítását.) Ekkor bármely $u, v, x, y, z \in \{a, b, c\}^*$-ra, amelyre $w = uxvyz$, $|xvy| \leq q$, $|xy| > 0$, a lemma alapján fennáll, hogy $uvz \in L$. Viszont, mivel xy az $\{a, b, c\}$-ből legalább egy betűt nem tartalmaz, így uvz nem lehet L eleme, és így a nyelv nem teljesíti a Bar-Hillel lemma feltételeit, vagyis a nyelv nem környezetfüggetlen.

Bar-Hillel lemma: Minden L környezetfüggetlen nyelvhez meg tudunk adni két p és q természetes számot úgy, hogy minden olyan w szó L-ben, amely hosszabb, mint p $w = uxvyz$ alakú, ahol $|xvy| \leq q$, $xy \neq \varepsilon$, és minden ux^ivy^iz szó is benne van az L nyelvben minden $i \geq 0$ egész számra ($u, x, v, y, z \in T^*$).

Tartalmazás (szóprobléma)

Tétel:

Minden környezetfüggő $G = (N, T, P, S)$ grammatika és minden $u \in T^*$ szó esetén eldönthető, hogy G-ben u levezethető-e vagy sem.

Bizonyításvázlat:

Ha $u = \varepsilon$, akkor a válasz triviális. Ha $u \in T^+$, akkor tekintsük az összes olyan véges

$$S = u_0, u_1, u_2, \ldots, u_n = u$$

sorozatot, ahol $u_i \in (N \cup T)^+$ és $|u_i| \leq |u_{i+1}|$ teljesül $0 \leq i \leq n-1$-re , valamint $u_i \neq u_j$, hacsak $i \neq j$.

Azon $v \in (N \cup T)^+$ szavak száma, amelyekre $|v| \leq |u|$ fennáll, véges, ezért a fenti sorozatok száma is véges.

Ezért szisztematikusan elő tudjuk az összes ilyen sorozatot állítani és le tudjuk ellenőrizni, hogy $u_i \Longrightarrow u_{i+1}$, $0 \leq i \leq n-1$ teljesül-e vagy sem.

Ha van ilyen sorozat, akkor $u \in L(G)$. (A legrövidebb levezetésben nincs szóismétlés.)

Kuroda normálforma

Definíció:

Egy hossz-nemcsökkentő $G = (N, T, P, S)$ grammatikát Kuroda normálformájúnak mondunk, ha minden egyes szabálya

1. vagy $A \to a$,

2. vagy $A \to B$,

3. vagy $A \to BC$

4. vagy $AB \to CD$

alakú, ahol $a \in T$ és $A, B, C, D \in N$.

Kuroda normálforma - folytatás

Tétel:

Minden hossz-nemcsökkentő grammatikához meg tudunk konstruálni egy vele ekvivalens Kuroda normálformájú grammatikát.

Bizonyításvázlat:

Legyen $G = (N, T, P, S)$ egy hossz-nemcsökkentő grammatika, és tegyük fel, hogy terminális szimbólumok csak $A \to a$ alakú szabályokban fordulnak elő, ahol $A \in N, a \in T$.

Ha $u \to v \in P$, valamint $|u| = 1$ és $|v| > 2$, akkor a $u \to v$-t 3. alakú (azaz, $A \to BC$ alakú) szabályokkal tudjuk helyettesíteni a Chomsky normálformára hozás algoritmusának megfelelően.

Ha $|u| = |v| = 2$, akkor az $u \to v$ szabály 4. alakú (azaz, $AB \to CD$ alakú).

Vagyis elegendő azzal az esettel foglalkoznunk, ha $|u| \geq 2$ és $|v| > 2$.

Bizonyításvázlat - folytatás

Legyen $u = X_1X_2\ldots X_m$ és $v = Y_1Y_2\ldots Y_n$, $2 \leq m$, $2 < n$, $X_i \in N, 1 \leq i \leq m$ és $Y_j \in N, 1 \leq j \leq n$. Tegyük fel, hogy $n > m$, az $n = m$ eset értelemszerű módosítással adódik. Akkor az

$$X_1X_2\ldots X_m \to Y_1Y_2\ldots Y_n$$

szabályt helyettesítjük a

$$\begin{array}{rr} X_1X_2 \to Y_1Z_2 & Z_m \to Y_mZ_{m+1} \\ Z_2X_3 \to Y_2Z_3 & Z_{m+1} \to Y_{m+1}Z_{m+2} \\ \vdots & \\ Z_{m-1}X_m \to Y_{m-1}Z_m & Z_{n-1} \to Y_{n-1}Y_n \end{array}$$

szabályokkal, ahol $Z_2, \ldots, Z_{n-1}$ új, a szabályhoz bevezetett nemterminálisok. (Minden fenti alakú szabályhoz páronként diszjunkt új nemterminálishalmazt vezetünk be.)

Bizonyításvázlat - szemléltetés

A $X_1X_2\ldots X_m \to Y_1Y_2\ldots Y_n$ szabály végrehajtásának szimulálása:

$X_1X_2X_3\ldots X_{m-1}X_m \Longrightarrow Y_1Z_2X_3\ldots X_{m-1}X_m \Longrightarrow Y_1Y_2Z_3\ldots X_{m-1}X_m$

$\Longrightarrow Y_1Y_2Y_3\ldots Z_{m-1}X_m \Longrightarrow Y_1Y_2Y_3\ldots Y_{m-1}Z_m \Longrightarrow Y_1Y_2Y_3\ldots Y_{m-1}Y_mZ_{m+1}$

$\Longrightarrow Y_1Y_2Y_3\ldots Y_mY_{m+1}Z_{m+2} \Longrightarrow^* Y_1Y_2Y_3\ldots Y_{m-1}Y_{m+1}Y_{m+2}\ldots Y_{n-1}Y_n$

Az alkalmazott szabályok:

$$\begin{array}{rr}
X_1X_2 \to Y_1Z_2 & Z_m \to Y_mZ_{m+1} \\
Z_2X_3 \to Y_2Z_3 & Z_{m+1} \to Y_{m+1}Z_{m+2} \\
\vdots & \\
Z_{m-1}X_m \to Y_{m-1}Z_m & Z_{n-1} \to Y_{n-1}Y_n
\end{array}$$

Bizonyításvázlat - folytatás

Legyen N' az új és az eredeti nemterminálisok halmaza, P' pedig a P halmazból megmaradó és a fenti átalakítással kapott szabályok halmaza. Akkor könnyen látható, hogy a $G' = (N', T, P', S)$ grammatikára $L(G) \subseteq L(G')$ teljesül. A fordított irányú tartalmazás is igazolható.

Kuroda normálforma - folytatás

Korollárium

Minden ε-mentes környezetfüggő nyelv generálható Kuroda normálformájú grammatikával.

0-**típusú grammatikák egy normálformája**

Tétel

Minden $G = (N, T, P, S)$ 0-típusú grammatikához létezik egy vele ekvivalens G' generatív grammatika, amelynek minden egyes szabályának alakja egyike az alábbiaknak:

$$\begin{array}{rl} S \to \varepsilon, & A \to a, \\ A \to B, & A \to BC, \\ AB \to AC, & AB \to CB, \\ AB \to B, & \end{array}$$

ahol $a \in T$, $S, A, B, C, \in N$ és az S kezdőszimbólum csak a szabályok baloldalán fordul elő.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ egy tetszőleges 0-típusú grammatika. Tegyük fel, hogy P tartalmaz $u \to \varepsilon$ alakú szabályt, ahol $u \in (N \cup T)^+$. Ekkor minden $u \to \varepsilon$ alakú szabályt, ahol $u \in (N \cup T)^+$, helyettesítsünk $uX \to X$ és $Xu \to X$ alakú szabályokkal minden egyes $X \in (N \cup T)$-re. Jelöljük az újonnan kapott szabályhalmazt P'-vel. Könnyen látható, hogy a $G' = (N, T, P', S)$ grammatika az $L(G') = L(G) - \{\varepsilon\}$ nyelvet generálja.

Ha $\varepsilon \in L(G)$, akkor adjuk hozzá P-hez az $S' \to \varepsilon$ szabályt, ahol $S' \notin N$.

Ezután P'-t átalakítjuk úgy, hogy terminális szimbólumok csak 2. alakú, azaz $A \to a$, $(A \in N, a \in T)$ alakú szabályokban fordulhassanak elő. Ekkor minden további szabály $u \to v$ alakú lesz, ahol $u, v \in N^+$.

A hossz-nemcsökkentő szabályok (azaz, ahol $|u| \leq |v|$) esetében a Kuroda normálformára hozás során alkalmazott módszert alkalmazzuk.

Bizonyításvázlat - folytatás

Ezután a hossz-csökkentő szabályokkal foglalkozunk. Minden $X_1 \dots X_m \to Y_1 \dots Y_n$, ahol $m > n \geq 1$, és $X_i, Y_j \in N$, alakú szabályt helyettesítünk egy szabályhalmazzal, ahol $U_1, \dots U_m$, és $Z_1, \dots Z_n$ új, a szabályhoz bevezetett nemterminálisok:

$$
\begin{array}{rl}
X_{m-1}X_m \to Z_m U_m & Z_m U_m \to U_m \\
X_{m-2}U_m \to Z_{m-1}U_{m-1} & Z_{m-1}U_{m-1} \to U_{m-1} \\
\vdots & \\
X_n U_{n+2} \to Z_{n+1}U_{n+1} & Z_{n+1}U_{n+1} \to U_n Y_n \\
X_{n-1}U_n \to U_{n-1}Y_{n-1} & \\
\vdots & \\
X_1 U_2 \to U_1 Y_1 & \\
U_1 Y_1 \to Y_1, &
\end{array}
$$

Itt $AB \to CD$ típusú szabályokat váltunk ki, négy 5. vagy 6. alakú szabállyal ($AB \to AC$, $AB \to CB$). Így egy olyan grammatikát kapunk, amely ugyanazt a nyelvet generálja, mint az eredeti G grammatika.

Bizonyításvázlat - szemléltetés

A szabály alkalmazása:

$$X_1X_2\dots X_{m-2}X_{m-1}X_m \Longrightarrow X_1X_2\dots X_{m-2}Z_mU_m \Longrightarrow X_1X_2\dots X_{m-2}U_m \Longrightarrow X_1X_2\dots Z_{m-1}U_{m-1}$$
$$\Longrightarrow X_1X_2\dots U_{m-1} \Longrightarrow^* X_1X_2\dots X_nU_{n+2} \Longrightarrow X_1X_2\dots Z_{n+1}U_{n+1} \Longrightarrow X_1X_2\dots U_nY_n$$
$$\Longrightarrow X_1X_2\dots U_nY_n \Longrightarrow X_1U_2Y_2\dots Y_n \Longrightarrow^* U_1Y_1Y_2\dots Y_n \Longrightarrow Y_1Y_2\dots Y_n$$

$$\begin{array}{rl}
X_{m-1}X_m \to Z_mU_m & Z_mU_m \to U_m \\
X_{m-2}U_m \to Z_{m-1}U_{m-1} & Z_{m-1}U_{m-1} \to U_{m-1} \\
\vdots & \\
X_nU_{n+2} \to Z_{n+1}U_{n+1} & Z_{n+1}U_{n+1} \to U_nY_n \\
X_{n-1}U_n \to U_{n-1}Y_{n-1} & \\
\vdots & \\
X_1U_2 \to U_1Y_1 & \\
U_1Y_1 \to Y_1, &
\end{array}$$