Formális nyelvek - 5. előadás

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

Lineáris grammatikák és reguláris nyelvek

Definíció

Egy $G = (N, T, P, S)$ környezetfüggetlen grammatikát lineárisnak nevezünk, ha minden szabálya vagy

1. $A \to u$, $A \in N$, $u \in T^*$ vagy
2. $A \to u_1 B u_2$ alakú, ahol $A, B \in N$ és $u_1, u_2 \in T^*$.

Továbbá G-t bal-lineárisnak, illetve jobb-lineárisnak mondjuk, ha $u_1 = \varepsilon$, illetve $u_2 = \varepsilon$ minden 2. alakú szabályra.

Megjegyzés: A jobb-lineáris grammatikák azonosak a reguláris grammatikákkal (3-típusúak).

Lineáris grammatikák és nyelvek

Definíció

Egy L nyelvet lineárisnak, bal-lineárisnak, illetve jobb-lineárisnak mondunk, ha van olyan G lineáris, bal-lineáris, illetve jobb-lineáris grammatika, amelyre $L = L(G)$ teljesül.

Tétel

Minden bal-lineáris grammatika reguláris (3-típusú) nyelvet generál.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ bal-lineáris grammatika és legyen $N = \{S, A_1, \ldots, A_n\}$. Az általánosság megszorítása nélkül feltehetjük, hogy S nem fordul elő egyetlen szabály jobboldalán sem.

Megkonstruálunk egy $G' = (N, T, P', S)$ jobb-lineáris grammatikát, amelyre $L(G) = L(G')$ teljesül.

Legyen

1. $S \to u \in P'$ akkor és csak akkor, ha $S \to u \in P$, $u \in T^*$,
2. $S \to uA_k \in P'$ akkor és csak akkor, ha $A_k \to u \in P$, $u \in T^*$,
3. $A_j \to uA_k \in P'$ akkor és csak akkor, ha $A_k \to A_j u \in P$, $u \in T^*$,
4. $A_j \to u \in P'$ akkor és csak akkor, ha $S \to A_j u \in P$, $u \in T^*$.

Bizonyításvázlat - folytatás

Megmutatjuk, hogy $L(G) = L(G')$.

Legyen $w \in L(G)$. Ha $S \to w \in P$, akkor $S \to w \in P'$, így $w \in L(G')$.

Egyébként w-hez van G-ben egy

$$S \Longrightarrow A_{i_1} w_1 \Longrightarrow \ldots \Longrightarrow A_{i_{m-1}} w_{m-1} \ldots w_1 \Longrightarrow w_m \ldots w_1 = w$$

levezetés. Azonban ekkor G'-ben létezik egy

$$S \Longrightarrow w_m A_{i_{m-1}} \Longrightarrow w_m w_{m-1} A_{i_{m-2}} \Longrightarrow \ldots \Longrightarrow w_m \ldots w_2 A_{i_1} \Longrightarrow w_m \ldots w_1 = w$$

levezetés, azaz $w \in L(G')$. Így $L(G) \subseteq L(G')$. A fordított állítás igaz volta a szimmetria következménye.

Következmény

Korollárium

$\mathcal{L}_3$ zárt a tükrözés (megfordítás) műveletére nézve és minden reguláris nyelv generálható bal-lineáris grammatikával.

A bizonyítás azonnal adódik. Minden jobb-lineáris G grammatikához meg tudunk konstruálni egy G' bal-lineáris grammatikát, amely $L(G)$ tükörképét generálja. Ha $L \in \mathcal{L}_3$, akkor L^{-1} generálható a bal-lineáris G grammatikával és a megelőző tétel alapján L^{-1} reguláris. Akkor $(L^{-1})^{-1}$ szintén reguláris.

Példa

Legyen $L_1 = \{a^n b^n c^k \mid n \geq 1, k \geq 1\}$ és $L_2 = \{a^k b^n c^n \mid n \geq 1, k \geq 1\}$.

L_1 és L_2 lineáris nyelvek és

$$L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 1\}$$

környezetfüggő, de nem környezetfüggetlen nyelv.

L_1 generálható a $G_1 = (\{S, A\}, \{a, b, c\}, P_1, S)$ grammatikával, ahol
$P_1 = \{S \to Sc, S \to Ac, A \to ab, A \to aAb\}$ és
L_2 generálható a $G_2 = (\{S, B\}, \{a, b, c\}, P_2, S)$ grammatikával, ahol
$P_2 = \{S \to aS, S \to aB, B \to bc, B \to bBc\}$.

3-típusú grammatikák normálformája

Tétel

Minden 3-típusú, azaz reguláris nyelv generálható egy olyan grammatikával, amelynek szabályai vagy

1. $X \to aY$, ahol $X, Y \in N$ és $a \in T$ vagy

2. $X \to \varepsilon$ alakúak, ahol $X \in N$.

Bizonyításvázlat

Legyen $G = (N, T, P, S)$ 3-típusú grammatika. Ismeretes, hogy G szabályai vagy:

1. $A \to uB$ vagy

2. $A \to u$ alakúak, ahol $A, B \in N$ és $u \in T^*$.

Bizonyításvázlat - folytatás

Tegyük fel, hogy $|u| > 1$.

Legyen $u = a_1 \dots a_n$, $n \geq 2$. Helyettesítsünk minden $A \to a_1 \dots a_n B$ szabályt az $\{A \to a_1 Z_1, Z_1 \to a_2 Z_2, \dots, Z_{n-1} \to a_n B\}$ szabályhalmazzal, ahol $Z_1, \dots, Z_{n-1}$ új, a szabályhoz bevezetett nemterminálisok.

Hasonlóan, az $A \to a_1 \dots a_m$, $m \geq 1$ alakú szabályokat helyettesítsük a $\{A \to a_1 Y_1, Y_1 \to a_2 Y_2, \dots, Y_{m-1} \to a_m Y_m, Y_m \to \varepsilon\}$ szabályhalmazokkal, ahol $Y_1, \dots, Y_m$ a szabályhoz bevezetett új nemterminálisok.

Az így kapott új P_1 szabályhalmaz elemei $X \to aY$, $X \to Y$, $X \to \varepsilon$ alakúak, ahol $X, Y \in N$ és $a \in T$. Ezután elimináljuk a láncszabályokat.

Bizonyításvázlat - folytatás

Legyen N' a P_1 szabályhalmazban előforduló nemterminálisok halmaza (az új nemterminálisokat is beszámítva). Legyen bármely $X \in N'$ nemterminálisra $U(X) = \{Y \mid Y \Longrightarrow^* X\}$.

Definiáljuk P'-t a következőképpen:

1. $X \rightarrow aY$, akkor és csak akkor, ha létezik olyan $Z \in N'$, amelyre $X \in U(Z)$ és $Z \rightarrow aY \in P_1$,

2. $X \rightarrow \varepsilon$, akkor és csak akkor, ha létezik olyan $Z \in N'$, amelyre $X \in U(Z)$ és $Z \rightarrow \varepsilon \in P_1$.

Könnyen látható, hogy $L(G) = L(G')$, ahol $G' = (N', T, P', S)$.

Korollárium

Minden ε-mentes reguláris (3-típusú) grammatikához konstruálható egy vele ekvivalens reguláris grammatika, amelynek szabályai

1. $X \to aY$ vagy

2. $X \to a$ alakúak, ahol $X, Y \in N$ és $a \in T$.

Környezetfüggetlen nyelvek egy zártsági tulajdonsága

Tétel

A környezetfüggetlen nyelvek osztálya, azaz $\mathcal{L}_2$ zárt a reguláris (3-típusú) nyelvekkel való metszetre nézve.

Bizonyításvázlat

Legyen L egy tetszőleges környezetfüggetlen nyelv ($L \in \mathcal{L}_2$) és legyen let L' ($L' \in \mathcal{L}_3$) reguláris nyelv. Megmutatjuk, hogy $L \cap L'$ környezetfüggetlen.

Először is, tegyük fel, hogy $\varepsilon \notin L$ és az L nyelvet a $G = (N, T, P, S)$ Chomsky-normálformájú grammatika, az L' nyelvet pedig a $G' = (N', T', P', S')$, a megelőzőekben ismertetett normálformájú grammatika generálja.

Legyen $\{X_1, \ldots, X_k\}$ a G' azon nemterminálisainak halmaza, amelyekre $X_i \to \varepsilon$, $1 \leq i \leq k$ teljesül.

Bizonyításvázlat - folytatás

Definiáljuk a következő grammatikákat:

$$G_i = (N' \times (N \cup T) \times N', T \cup T', P'', [S', S, X_i]),$$

$i \in \{1, \ldots, k\}$ ahol

1. $[X, A, Y] \to [X, B, Z][Z, C, Y] \in P''$ minden $X, Y, Z \in N'$ nemterminálisra akkor és csak akkor, ha $A \to BC \in P$,

2. $[X, A, Y] \to [X, a, Y] \in P''$ minden $X, Y \in N'$ nemterminálisra, akkor és csak akkor, ha $A \to a \in P$,

3. $[X, a, Y] \to a \in P''$, akkor és csak akkor, ha $X \to aY \in P'$.

Megmutatjuk, hogy $L \cap L' = \bigcup_{i=1}^{k} L(G_i)$.

Bizonyításvázlat - folytatás

Legyen $w = a_1 \dots a_n$, $a_i \in T$, $1 \leq i \leq n$.

1. w akkor és csak akkor vezethető le G-ben, ha minden i-re és N'-beli nemterminálisok minden $Z_1, \dots, Z_{n-1}$ sorozatára létezik

$[S', S, X_i] \Longrightarrow_{G_i}^* [S', a_1, Z_1][Z_1, a_2, Z_2] \dots [Z_{n-1}, a_n, X_i]$ alakú levezetés G_i-ben. Továbbá,

2. w akkor és csak akkor vezethető le G'-ben, ha létezik nemterminálisok $Z_1, \dots, Z_{n-1}$ sorozata és $X_i \in N'$, ahol $X_i \to \varepsilon \in P'$ úgy, hogy

$$S' \Longrightarrow_{G'} a_1 Z_1 \Longrightarrow_{G'} a_1 a_2 Z_2 \Longrightarrow_{G'} \dots \Longrightarrow_{G'} a_1 \dots a_n X_i \Longrightarrow_{G'} a_1 a_2 \dots a_n.$$

Bizonyításvázlat - folytatás

Ez azt jelenti, hogy w akkor és csak akkor vezethető le G'-ben, ha vannak olyan $Z_1, \ldots, Z_{n-1}$ nemterminálisok N'-ben és van olyan levezetés G_i-ben, ahol

$[S', a_1, Z_1][Z_1, a_2, Z_2] \ldots [Z_{n-1}, a_n, X_i] \Longrightarrow_{G_i}^{*} a_1 a_2 \ldots a_n$ fennáll.

Ebből az következik, hogy $w \in L \cap L'$ akkor és csak akkor, ha $w \in L(G_i)$ valamely i-re. Ha $\varepsilon \in L$, akkor megkonstruálunk egy $(L - \{\varepsilon\}) \cap L'$-t generáló környezetfüggetlen grammatikát és hozzáadjuk az $S_0 \to \varepsilon, S_0 \to S$ szabályokat.

Önbeágyazó környezetfüggetlen grammatika

Definíció

A $G = (N, T, P, S)$ környezetfüggetlen grammatika önbeágyazó, ha van olyan A nemterminálisa, amelyre $A \Longrightarrow_G^* xAy$ teljesül, ahol $x, y \in (N \cup T)^+$.

Tétel

Ha a G környezetfüggetlen grammatika nem önbeágyazó, akkor reguláris.

Bizonyításvázlat

Az általánosság megszorítása nélkül feltehetjük, hogy $G = (N, T, P, S)$ redukált. Két esetet különböztetünk meg.

1. eset

Minden egyes A nemterminálisra létezik $A \Longrightarrow uSz$ alakú levezetés, ahol $u, z \in (N \cup T)^*$.

Minden egyes $A \to v \in P$ szabály, ahol v-ben legalább egy nemterminális van, vagy

1. $A \to xBy$, vagy
2. $A \to xB$, vagy
3. $A \to By$, vagy
4. $A \to B$ alakú, ahol

$A, B \in N$ és $x, y \in (N \cup T)^+$.

Bizonyításvázlat - folytatás

1. eset, 1. alakú szabály

Ha van 1. alakú szabály ($A \to xBy$) a P szabályhalmazban, akkor van

$$A \Longrightarrow xBy \Longrightarrow^* xuSzy \Longrightarrow^* xuvAqzy$$

alakú levezetés, és ez ellentmond annak, hogy G nem önbeágyazó.

1. eset, 2. és 3. alakú szabályok.

Az $A \to xB$ és $C \to Dy$ alakú szabályokra ismét ellentmondást kapunk, hiszen $A \Longrightarrow xB \Longrightarrow^* xu_1Sz_1 \Longrightarrow^* xu_1v_1Cq_1z_1 \Longrightarrow xu_1v_1Dyq_1z_1 \Longrightarrow^* xu_1v_1u_2Sz_2yq_1z_1 \Longrightarrow^* xu_1v_1u_2v_2Aq_2z_2yq_1z_1$.

A 2. alakú szabályok, ahol $X \in T^*$ vagy a 3. alakú szabályok esetében, ahol $Y \in T^*$, a bizonyítás hasonlóan történik. Ebből az következik, hogy a P szabályhalmazban csak $A \to xB$ és $A \to x$ alakú szabályok lehetnek, ahol A, B nemterminálisok és x terminális szó, vagy $A \to By$ és $A \to y$, ahol A, B nemterminálisok és y terminális szó. Azaz, G reguláris.

Bizonyításvázlat - folytatás

2. eset

Van egy olyan A_1 nemterminális, amelyre nem teljesül $A_1 \Longrightarrow^* uSz$ semmilyen $u, z \in (N \cup T)^*$ esetén sem.

Az állítást indukcióval bizonyítjuk. Ha $N = \{S\}$, ez az eset nem állhat fenn, mert $S \Longrightarrow^* S$. Tegyük fel, hogy az állítás teljesül minden, legfeljebb n nemterminálissal rendelkező grammatikára. Megmutatjuk, hogy akkor fenáll minden G grammatikára is, amelynek legfeljebb $n+1$ nemterminálisa van.

Legyen $G_1 = (N - \{S\}, T, P_1, A_1)$, ahol P_1-et a P szabályhalmazból úgy kapjuk meg, hogy minden olyan szabályt törlünk, amely S-et tartalmazza. Hasonlóan, legyen $G_2 = (N - \{A_1\}, T \cup \{A_1\}, P_2, S)$, ahol P_2-t P-ből az A_1 nemterminálist a baloldalon tartalmazó szabályok törlésével kapjuk meg. Sem G_1, sem G_2 nem önbeágyazó, mivel $P_1 \subseteq P$ és $P_2 \subseteq P$. Így $L(G_1)$ és $L(G_2)$ reguláris nyelvek. Látható, hogy az a nyelv, amelyet úgy kapunk, hogy $L(G_1)$-gyel helyettesítjük A_1-et $L(G_2)$-ben, ahol $a \in T$, megegyezik $L(G)$-vel, mivel $\mathcal{L}_3$ zárt a helyettesítés műveletére nézve.

Következmények

A reguláris nyelvek osztálya ($\mathcal{L}_3$) valódi részosztálya a környezetfüggetlen nyelvek osztályának ($\mathcal{L}_2$-nek).

A reguláris nyelvek osztálya ($\mathcal{L}_3$) valódi részosztálya a lineáris nyelvek osztályának.

Reference:

György E. Révész, Introduction to Formal Languages, McGraw-Hill Book Company, 1983, Chapter 3.3