LFA CURS 9

Simplificarea gramaticilor independente de context – continuare

Def Gramatica independentă de context G' obținută prin eliminarea neterminalelor și a producțiilor nefolositoare din gramatica G se numește gramatică redusă.

Obs În cazul în care prin aplicarea algoritmului de reducere pentru gramatica independentă de context $G = (N, \Sigma, S, P)$ simbolurile terminale ce apar în producțiile gramaticii reduse, G', constituie o submulțime proprie, Σ', a lui Σ, $\Sigma' \subsetneq \Sigma$, atunci în gramatica redusă G' vom considera ca alfabet terminal pe Σ' și nu pe Σ.

Pentru simplificarea unei gramatici independente de context vom proceda în felul următor

1. Se aplică algoritmul de reducere (eliminarea simbolurilor și a producțiilor nefolositoare
2. Eliminarea λ-producțiilor
3. Eliminarea redenumirilor
4. Se aplică încă o dată algoritmul de reducere

Observații 1) Este posibil ca prin aplicarea algoritmului de eliminare a λ-producțiilor să se obțină neterminali / producții nefolositoare.

2) Dacă λ este în limbajul gramaticii inițiale, atunci în gramatica obținută după eliminarea λ-producțiilor vom avea $S \to \lambda$ producție pentru simbolul de start S (pentru a permite ca λ să fie derivat în gramatica respectivă), iar S nu apare în membrul drept al niciunei producții

3) Este posibil ca după aplicarea algoritmului de eliminare a redenumirilor să se obțină neterminali / producții nefolositoare

4) Este necesar ca după aplicarea pașilor 1-3 să se aplice din nou reducerea (pasul 4)

5) Numărul t al terminalilor ce apar în șirul curent dintr-o derivare în G simplificată nu poate descrește. Deoarece G nu are λ-producții, (exceptând $S \to \lambda$ când $\lambda \in L(G)$, S simbolul de start ce nu apare în membrul drept al niciunei producții), atunci lungimea l a șirului curent din derivare nu poate, de asemenea, să descrească, ceea ce înseamnă că suma $t + l$ nu descrește. Mai mult decât atât, deoarece în G nu există redenumiri

(producții de forma $A \to B$), înseamnă că suma $t+l$ crește la fiecare pas al derivării.

O derivare a unui șir $x \in L(G)$ începe cu S, pentru care $l+t=1$, și se termină cu x, pentru care $l+t = 2|x|$. Rezultă că orice derivare a lui x nu are mai mult de $2|x|-1$ pași.

Dacă vom încerca toate derivările cu cel mult $2|x|-1$ pași și nu găsim printre acestea niciuna care să îl genereze pe x, atunci $x \notin L(G)$.

<u>ALGORITM</u> Verificarea apartenenței unui șir la limbajul generat de o gramatică independentă de context

<u>Input</u> • $G=(N, \Sigma, S, P)$ redusă, fără redenumiri, fără λ-producții (exceptând $S \to \lambda$ dacă $\lambda \in L(G)$; în acest caz S nu apare în membrul drept al niciunei producții).

• $w \in \Sigma^*$, șirul analizat

<u>Output</u> "DA", dacă $w \in L(G)$

"NU", dacă $w \notin L(G)$

```
if (w == λ)
    if (S → λ ∈ P) { printf("DA"); return }
    else { printf("NU"); return }
M = {S}; flag = true;   // |w| ≥ 1 ; M coada
while (flag) {
    flag = false;
    for (fiecare α ∈ M)
        for (fiecare α ⇒s β, |β| ≤ |w|)
            if (w == β) { printf("DA"); return }
            else if (β ∉ M && |β|N ≥ 1)
                { M = M ∪ {β}; flag = true }
} // end while
printf("NU"); return
```

Forma normala Chomsky (FNC)

Teoremă Fie $G = (N, \Sigma, S, P)$ gramatică independentă de context redusă, fără λ-producții, fără redenumiri. Atunci există $G' = (N', \Sigma, S', P')$ cu $L(G') = L(G)$ și toate producțiile lui G' au una dintre formele:

$A \to BC \qquad B, C \in N$

$A \to a \qquad a \in \Sigma$

Dem Vom construi G' în 2 pași.

I. • Inițializăm $N''=N$, $S''=S$, $P''=P$

• Pentru fiecare $x\in\Sigma$ adăugăm la N'' neterminalul nou A_x.

• Fiecare producție din P de forma:

$A\to\alpha_0A_1\alpha_1\cdots A_n\alpha_n$, $n\geq 1$, $A_1,\dots,A_n\in N$,

$\alpha_0,\alpha_1,\dots,\alpha_n\in\Sigma^*$, $|\alpha_0\alpha_1\cdots\alpha_n|\geq 1$,

o vom înlocui cu producțiile:

$A\to h(\alpha_0)A_1h(\alpha_1)\cdots A_nh(\alpha_n)$,

unde $h:\Sigma\to\{A_x\mid x\in\Sigma\}$, $h(x)=A_x$

la care adăugăm toate producțiile de forma $A_y\to y$, unde $y\in\Sigma$ și apare într-unul dintre șirurile $\alpha_0,\alpha_1,\dots,\alpha_n$.

Obținem astfel $G''=(N'',\Sigma,S,P'')$.

Este evident că $A\to\alpha_0A_1\alpha_1\cdots A_n\alpha_n\in P\iff$ $A\underset{G''}{\overset{*}{\Rightarrow}}\alpha_0A_1\alpha_1\cdots A_n\alpha_n$ și derivarea

$A\underset{G''}{\overset{*}{\Rightarrow}}\alpha_0A_1\alpha_1\cdots A_n\alpha_n$ este unic determinată

de neterminalii noi ce apar în $h(\alpha_0),\dots,h(\alpha_n)$,

deci $L(G'')=L(G)$.

În G'' avem acum producții de forma

$A\to x$, $x\in\Sigma$

$A\to Y_1\cdots Y_m$, $m\geq 2$, $Y_1,\dots,Y_m\in N$.

II. • Inițializăm $N' = N''$, $P' = P''$

• Fiecare producție din P'' de forma

$A \to Y_1 \cdots Y_m$, $m \geq 3$, $Y_1, \dots, Y_m \in N$

va fi înlocuită de producțiile

$A \to Y_1 Z_1$, $Z_1 \to Y_2 Z_2, \dots, Z_{m-2} \to Y_{m-1} Y_m \in P'$,

unde $Z_1, \dots, Z_{m-2}$ sunt neterminali ~~noi~~ $\in N'$, unic asociați cu $A \to Y_1 \cdots Y_m$.

Este evident că în G' avem $A \overset{*}{\underset{G'}{\Rightarrow}} Y_1 \cdots Y_m$, iar această derivare este unic determinată de neterminalii noi $Z_1, \dots, Z_{m-2} \in N'$, deci $L(G') = L(G'') = L(G)$.

În felul acesta am obținut în G' producții de forma:

$A \to x$, $x \in \Sigma$,

$A \to BC$, $A, B, C \in N'$

<u>Observație</u> Dacă vrem ca $\lambda \in L(G')$, atunci introducem în N' neterminalul nou S', care devine noul simbol de start al lui G', iar în P' adăugăm producția $S' \to \lambda$. În plus, pentru orice producție $S \to \alpha$ din P' adăugăm $S' \to \alpha$ la P'. Obținem $L(G') = L(G) \cup \{\lambda\}$

Dacă S nu apare în membrul drept al niciunei producții din P', atunci din P' putem elimina toate producțiile $S \to \alpha$

APLICAȚII

1). Se dă gramatica G_1 cu producțiile

$$S \to aB \mid bA$$
$$A \to a \mid aS \mid bAA$$
$$B \to b \mid bS \mid aBB$$

a) Să se verifice dacă abba este în $L(G_1)$

b) Să se arate că $L(G_1) = \{ w \in \{a,b\}^+ \mid |w|_a = |w|_b \}$

Dem. a) Observăm că G_1 este redusă, nu are λ-producții și nici redenumiri.

Aplicăm algoritmul de verificare a apartenenței unui șir la limbajul generat de G_1:

1. $M = \{S\}$
2. $M = \{aB, bA\}$
3. $M = \{abS, aaBB, baS, bbAA\}$
4. $M = \{abaB, abbA, aabB, baaB, babA, bbaA\}$
5. $M = \{abba, \dots \}$

$\downarrow$

"DA"

b) Exercițiu

2) Se dă gramatica G_2 cu producțiile

$S \to abAB$

$A \to bAB \mid \lambda$

$B \to BAa \mid A \mid \lambda$

Să se aducă G_2 la forma normală Chomsky

i) G_2 este redusă

ii) eliminăm λ-producțiile:

- neterminalii anulabili: $\{A, B\}$

$S \to abAB \mid abA \mid abB \mid ab$

$A \to bAB \mid bA \mid bB \mid b$

$B \to BAa \mid Ba \mid Aa \mid a \mid A$

iii) eliminăm redenumirile $(B \to A)$

$S \to abAB \mid abA \mid abB \mid ab$

$A \to bAB \mid bA \mid bB \mid b$

$B \to BAa \mid Ba \mid Aa \mid a \mid bAB \mid bA \mid bB \mid b$

iv) gramatica de la iii) este redusă

v) aducem gramatica la FNC

I: $S \to A_a A_b AB \mid A_a A_b A \mid A_a A_b B \mid A_a A_b$

$A \to A_b AB \mid A_b A \mid A_b B \mid b$

$B \to BAA_a \mid BA_a \mid AA_a \mid a \mid A_b AB \mid A_b A \mid A_b B \mid b$

$A_a \to a$

$A_b \to b$

II $S \to A_a A_1 \mid A_a B_1 \mid A_a C_1 \mid A_a A_b$

$A_1 \to A_b A_2$

$A_2 \to AB$

$B_1 \to A_b A$

$C_1 \to A_b B$

$A \to A_b D_1 \mid A_b A \mid A_b B \mid b$

$D_1 \to AB$

$B \to BE_1 \mid BA_a \mid AA_a \mid a \mid A_b F_1 \mid A_b A \mid A_b B \mid b$

$E_1 \to AA_a$

$F_1 \to AB$

$A_a \to a$

$A_b \to b$