Prezentácia úlohy 2.1 “ekvivalencia Chomského a štandartnej definícií kontextových gramatík”

Alexander Muchovikov, 2i3

6. marca 2001

Našou úlohou je dokázať ekvivalenciu Chomského definicie a štandartnej definície regulárnych gramatik (CSG).
Pre CSG s množinou terminálov T a množinou neterminálov N:

* Chomského definícia je, že pravidlá CSG majú tvar

$$u\alpha v \rightarrow u\beta v, \text{ kde}$$

$u, v \in (N \bigcup T)^*$, $\alpha \in N, \beta \in (N \bigcup T)^+$

* Štandartná definícia je, že pravidlá CSG majú tvar

$$u \rightarrow v, \text{ kde}$$

$u \in (N \bigcup T)^* N (N \bigcup T)^*$, $v \in (N \bigcup T)^+$ a zároveň $|u| \leq |v|$

Dôkaz
Majme dve gramatiky $G_1 = (N_1, T_1, P_1, \sigma_1)$ a $G_2 = (N_2, T_2, P_2, \sigma_2)$, ktoré generujú ten istý jazyk, ale majú rôzne množiny pravidiel: pravidlá P_1 sú v Chomského tvare, pravidlá P_2 sú v štandardnom tvare.
Keďže máme dokázať ekvivalenciu gramatík, tak to je to isté, ako dokázať, že každá z daných gramatík je normálnym tvarom pre tú druhú (dá sa prepísať všetky pravidlá tak, že ku každému odvodeniu v jednej gramatike existuje odvodenie aj v druhej gramatike, pričom prislúchajúce jazyky sú rovnaké).

I **Všetky pravidlá CSG v Chomského tvare sa dajú prepísať do štandardného tvaru.**

Táto časť dôkazu bude pomerne jednoduchá, pretože my ani nepotrebujeme meniť množinu pravidiel P_1 aby sme získali P_2 v štandardnom tvare. Keď si uvedomíme, že štandardný tvar vyžaduje len prítomnosť aspoň jedného neterminála na ľavej strane a aby bola splnená nerovnosť $|u| \leq |v|$ (pozri definíciu pravidiel CSG v štandardnom tvare), tak obe tieto podmienky platia.
Inými slovami, položíme $N_2 = N_1, T_2 = T_1, P_2 = P_1, \sigma_2 = \sigma_1$. Je zrejme, že G_2 generuje ten istý jazyk ako G_1. Preto nám stačí skontrolovať, či $\forall (u \Rightarrow v) \in P_2$ platí $u \in (N \bigcup T)^* N (N \bigcup T)^*$, $v \in (N \bigcup T)^+$ a zároveň $|u| \leq |v|$
Z definície Chomského tvaru CSG pravidiel vyplýva, že na ľavej strane

pravidla musí byť jeden neterminál, ktorý podľa jeho kontextu prepíšeme na neprázdnu vetnú formu. Je zrejme, že každé také pravidlo spĺňa aj podmienky štandardnej definície: množina $(N\bigcup T)^+$ (vetné formy dĺžky aspoň 1) - ako podmnožinu má množinu $(N\bigcup T)^*N(N\bigcup T)^*$ - vetné formy obsahujúce aspoň jeden neterminál. Pričom z toho, že $(\forall\alpha\in N)$: $|\alpha|=1$, $w_1\alpha w_2\to w_1\beta w_2$, $\beta\in(N\bigcup T)^+$, $|\beta|\geq 1$ vyplýva aj $|w_1\alpha w_2|\leq|w_1\beta w_2|$ a naše pravidlo vyhovuje všetkým podmienkam štandardného tvaru CSG pravidiel.

II **Všetky pravidlá CSG v štandardnom tvare sa dajú prepísať do Chomského tvaru.**

V danej časti dôkazu narážame na problém prevodu pravidiel zo štandardného tvaru do Chomského. Ide o to, že v štandardnom tvare môžeme prepísať aj niekoľko znakov z ľavej strany, pričom nekontrolujeme rovnosť kontextu (prefixov a sufixov) prepísaného podslova na ľavej a pravej strane.
Majme gramatiku G_2 s množinou pravidiel v štandardnom tvare. Potom $G_1=(N_1,T_1,P_1,\sigma_1)$ zostrojíme takto:

- $N_{ter}=\{N_\alpha|\alpha\in T_2\}$
- $T_1=T_2$
- Zavedieme homomorfizmus h : $h(a)=a$, ak $a\in N_2$; a $h(a)=N_a$, ak $a\in T_2$ (a $N_a\in N_{ter}$). Potom $P_{term}=\{h(u_1\dots u_n)\to h(v_1\dots v_m)|u_1\dots u_n\to v_1\dots v_m\in P_2\}$
- pre $\forall(u_1\dots u_S\to v_1\dots v_T)\in P_{term}$ položme $k=$ porádové číslo tohto pravidla v pôvodnej gramatike G_2 $k\leq|P_2|$ a nasledovne pridáme
  1) $P_1=P_1\bigcup\{N_\alpha\to\alpha|\alpha\in T_1\&N_\alpha\in N_{ter}\}$
  2) $P_1=P_1\bigcup\{u_1\dots u_S\to N_1^ku_2\dots u_S\}\bigcup\{N_1^ku_2\dots u_S\to N_1^kN_2^ku_3\dots u_S\}$
     $\bigcup\dots$ atď. $\dots\bigcup\{N_1^k\dots N_{S-1}^ku_S\to N_1^k\dots N_S^k\}$
  3) $N_1=N_{ter}\bigcup N_2\bigcup\{N_i^k|i=1\dots S\}$
  4) Vieme, že $S\leq T$, potom $P_1=P_1\bigcup\{N_1^k\dots N_S^k\to v_1\dots v_{T-S+1}N_2^k\dots N_S^k\}\bigcup\{v_1\dots v_{T-S+1}N_2^k\dots N_S^k\to v_1\dots v_{T-S+1}v_{T-S+2}N_3^k\dots N_S^k\}$
     $\bigcup\dots$ atď. $\dots\bigcup\{v_1\dots v_{T-1}N_S^k\to v_1\dots v_{T-1}v_T\}$

Táto konštrukcia funguje nasledovne: najrpv prepíšeme ľavú časť pravidla na S krokov na postupnosť neterminálov s dvoma indexami, potom ju na S krokov prepíšeme na $v_1 \dots v_T$. Pričom v prvom kroku neterminál N_1^k prepíšeme hneď na $T - S + 1$ neterminálov (toto si môžeme dovoliť, pretože platí $S \leq T$) a potom pokračujeme v prepisovaní po jednom neterminále. (Na konci dôkazu bude príklad) Treba dokázať, že pre každé odvodenie v G_2 existuje odvodenie v G_1, pričom $L(G_1) = G(L_2)$.

Indukcia vzhľadom na dĺžku odvodenia i.

Nech máme odvodenie dĺžky i. Dokážeme, že vieme prepísať aj odvodenie dĺžky $i + 1$

- Ak $i = 0$, tak je to zrejmé
- Indukčný predpoklad: vieme prepísať i ($i > 0$) krokov odvodenia v G_2 na odvodenie v G_1, čiže ak $\sigma_2 \Rightarrow_{G_2}^{i} u_1 \dots u_S$ potom $\sigma_1 \Rightarrow_{G_1}^{*} u_1 \dots u_S$
- Indukčný krok: ak

$$u_1 \dots u_S \Rightarrow_{G_2} v_1 \dots v_T \ \mathbf{(1)}$$

potom

$$u_1 \dots u_S \Rightarrow_{G_1}^{+} v_1 \dots v_T \ \mathbf{(2)}$$

V odvodení **(1)** sme použili pravidlo $u_1 \dots u_S \to v_1 \dots v_T \in P_2$. Nech je to pravidlo číslo k v množine P_2. V G_1 také odvodenie získame nasledovne: najprv prepíšeme na S krokov $u_1 \dots u_S \Rightarrow^S N_1^k \dots N_S^k$, ďalej potrebujeme už len $(|T| - |S| + 1)$ krokov:

$$u_1 \dots u_S \Rightarrow_{G_1} N_1^k u_2 \dots u_S \Rightarrow_{G_1}^{*} N_1^k \dots N_S^k \Rightarrow_{G_1}$$

$$v_1 \dots v_{T-S+1} N_2^k \dots N_S^k \Rightarrow_{G_1}^{T-S} v_1 \dots v_T$$

Prepisovanie prebieha zľava doprava od N_1^k až po N_S^k. Horný index k nám zabezpečuje, aby sme nepreskočili do inej sady pravidiel podľa našej konštrukcie a dokončili prepisovanie vetnej formy až po posledný neterminál N_S^k. Aj keby sme v procese odvodzovania začali prepisovať prefix $v_1 \dots v_{T-S+i}$ vetnej formy $v_1 \dots v_{T-S+i} N_{T-S+i+1}^k \dots N_S$ tak:

- buď sa nám aj po niekoľkých krokoch nič nezmení a prefix (tým pádom aj celý kontext) ostane ten istý. Toto môže nastať jedine v tom prípade, že by sme začali prepísovať niektoré neterminály a nakoniec by sme sa vrátili ku pôvodnej vetnej forme, čo ovplyvni iba dlžku odvodenia;
- alebo prepišeme kontext a ďalej sa už nepohneme, pretože to nebude vyhovovať žiadnemu pravidlu s neprepísanými neterminálmi N_i^k na konci pravidla.(Prepisujeme zľava doprava)

Dôkaz korektnosti prepisu $u_1 \dots u_S$ na $N_1^k \dots N_S^k$ v (2) kroku nášho algoritmu na S krokov je analogický.

Príklad
Majme CSG gramatiku s pravidlami v štandardnom tvare. G
$G = (N, T, P, \sigma)$, kde $N = \{\sigma, A\}$, $T = \{a, b\}$,

$$\begin{aligned} P = \{\sigma \to a\sigma b\ |A, \\ Ab \to Abab, \\ aA \to aa\}. \end{aligned}$$

Potom G_1 =($N = \{\sigma, A,\ N_1^1,\ N_1^2,\ N_1^3,\ N_2^3,\ N_1^4,\ N_2^4\}$, $T = \{a, b\}$, P_1, σ))}

$P_{ter} = \{N_a \to a, N_b \to b\}$.

Homomorfizmus $h : N \to \{N \bigcup T\}$ taký, že $h(\sigma) = \sigma$, $h(A) = A$, $h(a) = N_a$, $h(b) = N_b$.

Najprv pridáme množinu P_{ter} s pravidlami na prepisovanie neterminálov na terminály a potom (prvé tri riadky) prevedieme ľavé strany všetkých pravidiel $h(u_1 \dots u_S) \to h(v_1 ... hv_T)$ do potrebného tvaru $N_1^k \dots N_S^k$
$P_1 = P_{ter} \bigcup \{\ \sigma \to N_1^1\ ,\ \sigma \to N_1^2\ ,$
$AN_b \to N_1^3 N_b\ ,\ N_1^3 N_b \to N_1^3 N_2^3,$
$N_a A \to N_1^4 A\ ,\ N_1^4 A \to N_1^4 N_2^4,$
A teraz už môžeme začať prepisovať jednotlivé takýmto spôsobom získané vetné formy na homomorfné obrazy zodpovedajúcich pravých častí pôvodných pravidiel:
$N_1^1 \to N_a \sigma N_b,\ N_1^2 \to A,$
$N_1^3 N_2^3 \to AN_b N_a N_2^3\ ,\ AN_b N_a N_2^3 \to AN_b N_a N_b,$
$N_1^4 N_2^4 \to N_a N_2^4\ ,\ N_a N_2^4 \to N_a N_a\}.$