

Transformace a normální formy bezkontextových gramatik

ZPRACUJE: Mystik

Warshalův algoritmus?? Má tu být?

Obsah

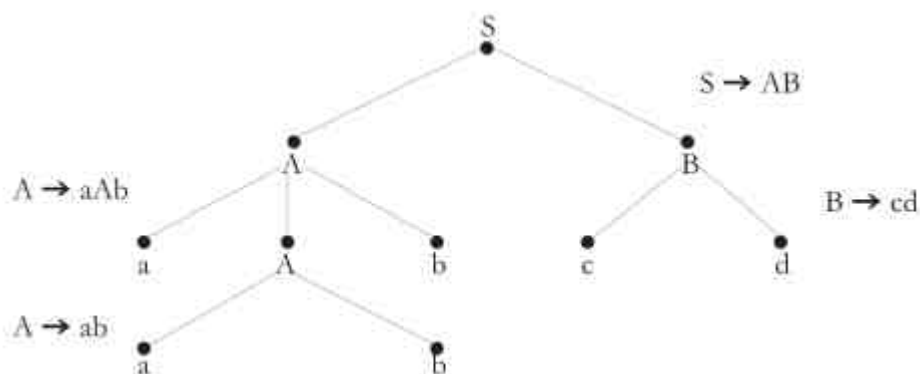
- 1 Derivace a derivační stromy
- 2 Transformace
 - 2.1 Odstranění nedostupných symbolů
 - 2.2 Odstranění zbytečných symbolů
 - 2.3 Odstranění ε -pravidel
 - 2.4 Odstranění jednoduchých pravidel
 - 2.5 Odstranění levé rekurze
- 3 Normální formy
 - 3.1 Chomského normální forma (CNF)
 - 3.1.1 Algoritmus převodu gramatiky do Chomského normální formy
 - 3.2 Greibachova normální forma (GNF)
 - 3.2.1 Algoritmus převodu gramatiky do Greibachovy normální formy

Derivace a derivační stromy

Gramatika (N, Σ, P, S)

Derivační strom

- uzly jsou prvky z $N \cup \Sigma$
- kořen stromu je S
- uzly označené prvky z Σ jsou koncové uzly stromu
- hrany odpovídají přepisovacím pravidlům gramatiky P
- označení koncových uzlů zleva doprava tvoří větnou formu



- Každé derivaci přísluší jeden derivační strom, ale jednomu stromu může příslušet více derivací (více způsobů jak sestavit stejný strom).

Levá/pravá derivace

při sestavování je vždy přepsán nejlevější/nejpravější nonterminál

Fráze větné formy

řada symbolů na koncových uzlech některého podstromu derivačního stromu (řada symbolů, které lze získat derivacemi z jednoho nonterminálu)

β je frází větné formy pokud $S \Rightarrow^* \alpha A \gamma \wedge A \Rightarrow^+ \beta$

Jednoduchá fráze větné formy

řady symbolů na koncových uzlech jednopatrového podstromu derivačního stromu

β je frází větné formy pokud $S \Rightarrow^* \alpha A \gamma \wedge A \Rightarrow \beta$

Nejlevější jednoduchá fráze (l-fráze)

fráze větné formy, která se nachází nejvíce vlevo

Víceznačnost

věta je víceznačná pokud existuje více derivačních stromů tvořících tuto větu

- existují jazyky, které nelze generovat bez víceznačnosti
- gramatika, která obsahuje cykly je vždy víceznačná

ϵ -pravidlo

$A \rightarrow \epsilon$ (přepisují nonterminál na prázdný řetězec)

A-pravidla

všechna pravidla, která mají A na levé straně

Jednoduchá pravidlo

$A \rightarrow B$ (přepisují nonterminál na jiný nonterminál)

Zbytečné symboly

nedostupné symboly (nelze jich dosáhnout žádným přepisem) a symboly negenerující terminální řetězce (nikdy nevytvoří větnou formu)

Cyklus

$A \rightarrow^+ A$

Vlastní gramatika

gramatika bez cyklů a bez ϵ -pravidel

Rekurzivní pravidla

- rekurzivní zleva: $A \rightarrow A\alpha$
- rekurzivní zprava: $A \rightarrow \alpha A$
- rekurzivní: $A \rightarrow \alpha A \beta$
- každý bezkontextový jazyk lze generovat gramatikou bez levé rekurze
- všechna pravidla s levou rekurzí lze nahradit pravidly s pravou rekurzí

Transformace

Odstranění nedostupných symbolů

Převod gramatiky (N, Σ, P, S) na gramatiku bez nedostupných stavů (N', Σ', P', S)

1) Najdeme všechny dostupné terminály a nonterminály

$i := 0$

$V_0 := \{S\}$ (tj. v nultém kroku je dostupný pouze počáteční stav)

repeat

$V_{i+1} := V_i \cup \{X \mid A \rightarrow \alpha X \beta \in P \wedge A \in V_i\}$ (tj. do množiny dostupných přidáme všechny terminály a nonterminály, které lze získat přepisem z některého nonterminálu z množiny dostupných z předchozího kroku)

$i := i + 1$ (tj. přejdeme k dalšímu kroku)

until $V_i = V_{i-1}$ (tj. končíme pokud se nám v posledním kroku množina dostupných nezměnila)

2) Novou gramatiku sestavíme jako:

$N' = V_i \cap N$ (tj. použijeme pouze dosažitelené nonterminály)

$\Sigma' = V_i \cap \Sigma$ (tj. použijeme pouze dosažitelené terminály)

$P' \subset P$ obsahuje pouze pravidla používající pouze symboly z V_i

Odstranění zbytečných symbolů

Převod gramatiky (N, Σ, P, S) na gramatiku bez zbytečných stavů (N^*, Σ, P^*, S)

1) Nalezení nonterminálů, které produkují terminální řetězce

$i := 0$

$N_0 := \emptyset$ (tj. v nultém kroku začínáme s prázdnou množinou nonterminálů generujících terminální řetězce)

repeat

$N_{i+1} := N_i \cup \{A \mid A \rightarrow \alpha \in P \wedge \alpha \in (N_i \cup \Sigma)^*\}$ (tj. do množiny všechny terminály, které lze přepsat na řetězce skládající se pouze z terminálů nebo nonterminálů, které jsou již v množině nonterminálů generujících terminální řetězce)

$i := i + 1$ (tj. přejdeme k dalšímu kroku)

until $N_i = N_{i-1}$ (tj. končíme pokud se nám v posledním kroku množina nezměnila)

2) Vytvoříme gramatiku G^* bez nonterminálů, které neprodukují terminální řetězce

$N^* = N_i$

$P^* \subset P$ obsahuje pouze pravidla používající pouze symboly z N_i

3) Odstraníme nedostupné symboly z gramatiky G^* (algoritmus viz výše)

Odstranění ε -pravidel

Převod gramatiky (N, Σ, P, S) na gramatiku bez ε -pravidel (N', Σ, P', S')

1) Sestavení množiny nonterminálů, které mohou generovat ε (N_ε)

postup je obdobný jako krok 1) algoritmu Odstranění zbytečných symbolů pouze

$N_{i+1} := N_i \cup \{A \mid A \rightarrow \alpha \in P \wedge \alpha \in (N_i \cup \{\varepsilon\})^*\}$

$N_\varepsilon = N_i$

2) Sestavíme novou množinu přepisovacích pravidel P'

Jestliže $A \rightarrow \alpha_0 B_1 \alpha_1 \dots B_k \alpha_k \in P, k \geq 0$ a všechna B_i jsou v N_ε a žádný symbol α_i není v N_ε pak k P' přidej všechna pravidla tvaru:

$A \rightarrow \alpha_0 X_1 \alpha_2 X_2 \dots X_k \alpha_k$

kde X_i je buď B_i nebo ε . Nepřidává se pravidlo $A \rightarrow \varepsilon$

(tj. od každého pravidla sestavíme všechny možné kombinace, kde jsou jednotlivé části přepsatelné na ε vynechány)

3) Zavedeme nový počáteční nonterminál S'

4) Jestliže je S v N_ε pak přidáme do P' pravidla $S' \rightarrow \varepsilon$ a $S' \rightarrow S$

Odstranění jednoduchých pravidel

Převod gramatiky bez ε -pravidel (N, Σ, P, S) na gramatiku bez jednoduchých pravidel (N, Σ, P', S)

1) Pro každé $A \in N$ sestojíme množinu $N_A = \{B \mid A \Rightarrow^* B\}$ (tj. množinu všech samostatných non-terminálů dosažitelných z A)

$i := 0$

$V_0 := \{A\}$ (tj. v nultém kroku začínáme s prázdnou množinou nonterminálů generujících terminální řetězce)

repeat

$N_{i+1} := N_i \cup \{C \mid B \rightarrow C \in P \wedge B \in N_i\}$ (tj. do množiny všechny terminály, které lze přepsat na řetězce skládající se pouze z terminálů nebo nonterminálů, které jsou již v množině nonterminálů generujících terminální řetězce)

$i := i + 1$ (tj. přejdeme k dalšímu kroku)

until $N_i = N_{i-1}$ (tj. končíme pokud se nám v posledním kroku množina nezměnila)

$N_A = N_i$

2) Sestavíme novou množinu P'

Pro všechna nejednoduchá pravidla $B \rightarrow \alpha$ přidáme do P' pravidla $A \rightarrow \alpha$ pro všechna A pro něž platí

$B \in N_A$ (tj. pro každého nejednoduché pravidlo najdeme všechny nonterminály A , které jdou jednoduchými pravidly přepsat na B , a vytvoříme všechny varianty pravidla, kde levou stranu nahradíme za A)

Odstranění levé rekurze

Převod **vlastní gramatiky** (N, Σ, P, S) na gramatiku bez levé rekurze (N', Σ, P', S)

Nonterminály vyjádříme jako $N = \{A_1, A_2, \dots, A_n\}$ tj. zvolíme pořadí jednotlivých nonterminálů.

Gramatiku budeme transformovat tak, že je-li $A_i \rightarrow \alpha$ pravidlo pak α začíná buď terminálem nebo nonterminálem A_j , který je v pořadí až za A_i ($j > i$).

1) $i = 1$ (tj. začneme prvním nonterminálem v pořadí)

2) Vezmeme všechna A_i -pravidla

(která jsou buď ve tvaru $A_i \rightarrow A_i \alpha_x$ nebo $A_i \rightarrow \beta_x$, kde β_x nezačíná nonterminálem, který je v pořadí před A_i)

Přidáme nový stav $A_i' do N'$

Všechna A_i -pravidla nahradíme za:

$A_i \rightarrow \beta_1 | \beta_2 | \dots | \beta_x$ (tj. přepisy na β zachováme)

$A_i \rightarrow \beta_1 A_i' | \beta_2 A_i' | \dots | \beta_x A_i'$ (tj. pro všechny přepisy na β vytvoříme pravidla s vpravo přidaným A_i')

$A_i' \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_x$ (tj. u přepisů na $A_i \alpha$ vytvoříme pravidla pouze s α)

$A_i' \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_x$ (tj. u přepisů na $A_i \alpha$ vytvoříme pravidla s odstraněným A_i a vpravo přidaným A_i')

(tím dosáhneme toho, že všechna A_i -pravidla začínají buď terminálem nebo nonterminálem, který je v pořadí až za A_i)

3) Pokud jsme již prošli všechny nonterminály ($i=n$) končíme

4) $i = i + 1$ (tj. posuneme se na další nonterminál)

5) Pro všechny dosud procházené nonterminály A_j ($1 < j < i$)

Pravidlo $A_i \rightarrow A_j \alpha$ odstraníme

Pro každé A_j -pravidlo ($A_j \rightarrow \beta$) vytvoříme:

$A_i \rightarrow \beta \alpha$

(tj. vezmeme všechna A_j -pravidla, jejichž pravá strana začíná nonterminálem A_j , který je v pořadí před A_i , a nahradíme je za pravidla ve kterých je A_j nahrazeno všemi řetězci na které jej lze přepsat)

6) Jdeme na krok 2)

Normální formy

Chomského normální forma (CNF)

Přepisovací pravidla jsou pouze ve tvaru:

- $A \rightarrow BC$
- $A \rightarrow a$
- případně $S \rightarrow \epsilon$ pokud $\epsilon \in L(G)$ a S nesmí být na pravé straně žádného přepisovacího pravidla

Algoritmus převodu gramatiky do Chomského normální formy

Převod **vlastní gramatiky bez jednoduchých pravidel** (N, Σ, P, S) na gramatiku v CNF (N', Σ, P', S')

- 1) Množina P' obsahuje všechna pravidla z P ve tvaru $A \rightarrow a$
- 2) Množina P' obsahuje všechna pravidla z P ve tvaru $A \rightarrow BC$
- 3) Je-li $S \rightarrow \epsilon$ v P je také v P'
- 4) Každé pravidlo tvaru $A \rightarrow X_1 X_2 \dots X_k$ ($k > 2$) přepíšeme do P' jako sérii pravidel:

$$\begin{aligned} A &\rightarrow X_1 X'_2 \\ X'_2 &\rightarrow X_2 X'_3 \\ &\dots \\ X'_{k-1} &\rightarrow X_{k-1} X_k \end{aligned}$$

Zavedeme nové non-terminální symboly $A \rightarrow X'_2 \dots X'_{k-1}$

(tj. pravidla mající pravou stranu delší než dva prvky rozdělíme na navazující sérii pravidel majících napravo jen dva prvky)

- 5) Ve všech pravidlech v P' které nejsou ve tvaru $A \rightarrow a$, kde se vyskytují na pravé straně terminály, nahradíme tyto terminály a za nové non-terminály A' a přidáme pravidlo:

$$A' \rightarrow a$$

(tj. nepovolená pravidla s terminály na pravé straně upravíme tak, že terminál nahradíme nonterminálem, který se přepisuje na původní terminál)

Greibachova normální forma (GNF)

- nejsou žádná ϵ pravidla
- přepisovací pravidla ve tvaru $A \rightarrow a\alpha$ kde $\alpha \in N^*$
- případně $S \rightarrow \epsilon$ pokud $\epsilon \in L(G)$ a S nesmí být na pravé straně žádného přepisovacího pravidla

Algoritmus převodu gramatiky do Greibachovy normální formy

Převod **vlastní gramatiky bez levé rekurze** (N, Σ, P, S) na gramatiku v CNF (N', Σ, P', S')

- 1) Seřadíme všechny nonterminály tak, aby pravá strana žádného přepisovacího pravidla nezačínala nonterminálem, který je v pořadí před přepisovaným nonterminálem. (v podstatě stejná pořadí v jakém máme nonterminály během algoritmu odstranění levé rekurze)

$$N = \{A_1, A_2, \dots, A_n\}$$

- 2) Postupně pro všechny nonterminály A_i od $A_n - 1$ sestupně po A_1 :

Pro každé pravidlo $A_i \rightarrow A_j \alpha$:

Odstraň toto pravidlo

Pro všechna A_j -pravidla ($A_j \rightarrow \beta$) vytvoř nová pravidla ve tvaru:

$$A_i \rightarrow \beta \alpha$$

(tj. všechna pravidla jejichž pravé strana začíná nonterminálem rozepíšeme tento nonterminál na všechny možnosti na jaké jej lze přepsat)

Po dokončení tohoto kroku všechny pravidla mají pravou stranu začínající terminálem.

3) Ve všech pravidlech v P' ve kterých se na pravé straně vyskytují terminály na jiné než první pozici nahradíme tyto terminály a za nové non-terminály A' a přidáme pravidlo:

$$A' \rightarrow a$$

(tj. nepovolená pravidla s terminály na pravé straně upravíme tak, že terminál nahradíme nonterminálem, který se přepisuje na původní terminál)

Kategorie: Státnice 2011 | Teoretická informatika

Stránka byla naposledy editována 27. 5. 2011 v 14:33.