Rekursivní neterminály a gramatiky

Definice 3.28. Neterminál A v CFG $\mathcal{G} = (N, \Sigma, P, S)$ se nazývá **levorekursivní** jestliže v $\mathcal{G}$ existuje derivace $A \Rightarrow^+ A\beta$.

CFG bez levorekursivních neterminálů se nazývá **nelevorekursivní**.

Praktický význam: některé nástroje pro automatickou tvorbu parserů k zadaným gramatikám vyžadují na vstupu nelevorekursivní gramatiku (např. ANTLR).

Lemma o substituci (pro připomenutí)

Lemma 3.20. (o substituci)
Nechť $\mathcal{G} = (N, \Sigma, P, S)$ je CFG. Nechť $A \to \alpha_1 B \alpha_2 \in P$.
Nechť $B \to \beta_1 \mid \ldots \mid \beta_r$ jsou všechna pravidla v P tvaru $B \to \alpha$.

Definujme $\mathcal{G}' = (N, \Sigma, P', S)$, kde

$$P' = (P \smallsetminus \{A \to \alpha_1 B \alpha_2\}) \;\cup\; \{A \to \alpha_1 \beta_1 \alpha_2 \mid \ldots \mid \alpha_1 \beta_r \alpha_2\}.$$

Pak $L(\mathcal{G}) = L(\mathcal{G}')$.

Algoritmus odstranění přímé levé rekurze

Nechť CFG $\mathcal{G} = (N, \Sigma, P, S)$ je necyklická a bez ε-pravidel, v níž všechna A-pravidla (pravidla mající na levé straně A) jsou tvaru

$$A \to A\alpha_1 \mid \ldots \mid A\alpha_m \mid \beta_1 \mid \ldots \mid \beta_n,$$

kde každý řetěz β_i začíná symbolem různým od A.

Nechť $\mathcal{G}' = (N \cup \{A'\}, \Sigma, P', S)$, kde P' obdržíme z P tak, že všechna výše uvedená pravidla nahradíme pravidly

$$A \to \beta_1 \mid \ldots \mid \beta_n \mid \beta_1 A' \mid \ldots \mid \beta_n A'$$

$$A' \to \alpha_1 \mid \ldots \mid \alpha_m \mid \alpha_1 A' \mid \ldots \mid \alpha_m A'$$

Pak $L(\mathcal{G}) = L(\mathcal{G}')$ a $\mathcal{G}'$ je necyklická a bez ε-pravidel.

Příklad

$$
\begin{aligned}
A &\rightarrow Bd \mid c \\
B &\rightarrow Bdd \mid Ccc \mid aAd \\
C &\rightarrow Aa
\end{aligned}
$$

Algoritmus odstranění levé rekurze

Vstup: Vlastní CFG $\mathcal{G} = (N, \Sigma, P, S)$

Výstup: Ekvivalentní nelevorekursivní gramatika bez ε-pravidel

Uspořádej libovolně N, $N = \{A_1, \ldots, A_n\}$
for $i \leftarrow 1$ **to** n **do**
    for $j \leftarrow 1$ **to** $i - 1$ **do**
        foreach pravidlo tvaru $A_i \to A_j\alpha$ **do**
            přidej pravidla $A_i \to \beta_1\alpha \mid \ldots \mid \beta_k\alpha$
            (kde $A_j \to \beta_1 \mid \ldots \mid \beta_k$ jsou všechna A_j-pravidla)
            vypusť pravidlo $A_i \to A_j\alpha$
        od
    od
    odstraň případnou přímou levou rekurzi na A_i
od

Korektnost algoritmu

Konečnost.

Ekvivalence gramatik: Všechny úpravy jsou dle Lemmatu o substituci nebo odstraňují přímou levou rekurzi.

Výsledná gramatika je nelevorekursivní:

1. po i-té iteraci vnějšího cyklu začíná každé A_i-pravidlo buď terminálem nebo neterminálem A_k, kde $k > i$.
2. po j-té iteraci vnitřního cyklu začíná každé A_i-pravidlo buď terminálem nebo neterminálem A_k, kde $k > j$.

Výsledná gramatika je bez ε-pravidel.

Greibachové normální forma

Definice 3.33. Bezkontextová gramatika $\mathcal{G} = (N, \Sigma, P, S)$ je v **Greibachové normální formě** (GNF) právě když

- $\mathcal{G}$ je bez ε-pravidel a
- každé pravidlo z P je tvaru $A \to a\alpha$, kde $a \in \Sigma$ a $\alpha \in N^*$ (s případnou vyjímkou pravidla $S \to \varepsilon$).

Věta 3.34. Každý bezkontextový jazyk lze generovat bezkontextovou gramatikou v Greibachové normální formě.

Zásobníkové automaty

Definice zásobníkového automatu

Definice 3.36. Nedeterministický zásobníkový automat (PushDown Automaton, PDA) je sedmice $\mathcal{M} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, kde

- Q je konečná množina, jejíž prvky nazýváme **stavy**,
- Σ je konečná množina, tzv. **vstupní abeceda**,
- Γ je konečná množina, tzv. **zásobníková abeceda**,
- $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to \mathcal{P}_{Fin}(Q \times \Gamma^*)$, tzv. (parciální) **přechodová funkce**[1],
- $q_0 \in Q$ je **počáteční stav**,
- $Z_0 \in \Gamma$ je **počáteční symbol v zásobníku**,
- $F \subseteq Q$ je množina **koncových stavů**.

[1] Zápis $\mathcal{P}_{Fin}(Q \times \Gamma^*)$ značí množinu všech **konečných** podmnožin množiny $Q \times \Gamma^*$.

Výpočet zásobníkového automatu

Definice 3.37. Nechť $\mathcal{M} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ je PDA.

Konfigurací nazveme libovolný prvek $(p, w, \alpha) \in Q \times \Sigma^* \times \Gamma^*$.

Na množině všech konfigurací automatu $\mathcal{M}$ definujeme binární relaci **krok výpočtu** $\underset{\mathcal{M}}{\vdash}$ takto:

$$(p, aw, Z\alpha) \underset{\mathcal{M}}{\vdash} (q, w, \gamma\alpha) \overset{def}{\iff} \exists (q, \gamma) \in \delta(p, a, Z) \text{ pro } a \in \Sigma \cup \{\varepsilon\}$$

Reflexivní a tranzitivní uzávěr relace $\underset{\mathcal{M}}{\vdash}$ značíme $\underset{\mathcal{M}}{\overset{*}{\vdash}}$.
Je-li $\mathcal{M}$ zřejmý z kontextu, píšeme pouze $\vdash$ resp. $\overset{*}{\vdash}$.

Akceptující výpočet zásobníkového automatu

Definice 3.37.(pokračování)

Jazyk akceptovaný PDA $\mathcal{M}$ **koncovým stavem** definujeme jako

$$L(\mathcal{M}) = \{w \in \Sigma^* \mid (q_0, w, Z_0) \vdash^* (q_f, \varepsilon, \alpha), \text{ kde } q_f \in F, \alpha \in \Gamma^*\}$$

a jazyk akceptovaný PDA $\mathcal{M}$ **prázdným zásobníkem** definujeme

$$L_e(\mathcal{M}) = \{w \in \Sigma^* \mid (q_0, w, Z_0) \vdash^* (q, \varepsilon, \varepsilon), \text{ kde } q \in Q\}.$$

Příklad

$\mathcal{M} = (\{q_0, p, f\}, \{a, b\}, \{A, B, Z\}, \delta, q_0, Z, \{f\})$, kde

$$
\begin{aligned}
\delta(q_0, a, Z) &= \{(q_0, AZ)\}\\
\delta(q_0, a, A) &= \{(q_0, AA)\}\\
\delta(q_0, a, B) &= \{(q_0, AB)\}\\
\delta(q_0, b, Z) &= \{(q_0, BZ)\}\\
\delta(q_0, b, A) &= \{(q_0, BA)\}\\
\delta(q_0, b, B) &= \{(q_0, BB)\}\\
\delta(q_0, \varepsilon, Z) &= \{(p, Z)\}\\
\delta(q_0, \varepsilon, A) &= \{(p, A)\}\\
\delta(q_0, \varepsilon, B) &= \{(p, B)\}\\
\delta(p, a, A) &= \{(p, \varepsilon)\}\\
\delta(p, b, B) &= \{(p, \varepsilon)\}\\
\delta(p, \varepsilon, Z) &= \{(f, Z)\}
\end{aligned}
$$

Příklad

$\mathcal{M} = (\{q_0\}, \{a, b\}, \{Z, A\}, \delta, q_0, Z, \emptyset)$, kde

$$\begin{aligned}
\delta(q_0, a, Z) &= \{(q_0, A)\} \\
\delta(q_0, a, A) &= \{(q_0, AA)\} \\
\delta(q_0, b, A) &= \{(q_0, \varepsilon)\}
\end{aligned}$$

Ekvivalence dvou způsobů akceptování

Věta 3.39. Pro každý jazyk L platí:
$L = L(\mathcal{N})$ pro nějaký PDA $\mathcal{N} \iff L = L_e(\mathcal{M})$ pro nějaký PDA $\mathcal{M}$.

Důkaz. koncový stav $\Longrightarrow$ prázdný zásobník

Intuice:

K danému $\mathcal{N}$ zkonstruujeme $\mathcal{M}$ simulující jeho činnost.
Vejde-li $\mathcal{N}$ do koncového stavu, $\mathcal{M}$ se nedeterministicky rozhodne

- pokračovat v simulaci automatu $\mathcal{N}$ **nebo**
- přejít do nově přidaného stavu q_ε, v němž vyprázdní zásobník.

Komplikace:

Řešení: před zahájením simulace bude u $\mathcal{M}$ na dně zásobníku nový symbol, který nedovolíme odstranit jinde, než ve stavu q_ε.

Konstrukce: Nechť $\mathcal{N} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$.

Klademe $\mathcal{M} = (Q \cup \{q_0', q_\varepsilon\}\,,\, \Sigma\,,\, \Gamma \cup \{Z'\}\,,\, \delta'\,,\, q_0'\,,\, Z'\,,\, \emptyset)$,
kde $Z' \notin \Gamma$, $q_0', q_\varepsilon \notin Q$ a δ' je definována takto:

- $\delta'(q_0', \varepsilon, Z') = \{(q_0, Z_0Z')\}$
- jestliže $\delta(q, a, Z)$ obsahuje (r, γ), pak $\delta'(q, a, Z)$ obsahuje (r, γ)
- $\delta'(q, \varepsilon, Z)$ obsahuje (q_ε, Z)
  pro všechny $q \in F$ a $Z \in \Gamma \cup \{Z'\}$
- $\delta'(q_\varepsilon, \varepsilon, Z) = \{(q_\varepsilon, \varepsilon)\}$
  pro všechny $Z \in \Gamma \cup \{Z'\}$

Korektnost:

prázdný zásobník $\Longrightarrow$ koncový stav

Intuice:

K danému $\mathcal{M}$ zkonstruujeme $\mathcal{N}$ simulující jeho činnost.

- $\mathcal{N}$ si před simulací přidá na dno zásobníku nový symbol.
- Je-li $\mathcal{N}$ schopen číst tento symbol (tj. zásobník automatu $\mathcal{M}$ je prázdný) tak $\mathcal{N}$ přejde do nově přidaného stavu q_f, který je koncovým stavem.

Konstrukce: Nechť $\mathcal{M} = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, \emptyset)$.

Klademe $\mathcal{N} = (Q \cup \{q_0', q_f\}\,,\, \Sigma\,,\, \Gamma \cup \{Z'\}\,,\, \delta'\,,\, q_0'\,,\, Z'\,,\, \{q_f\})$, kde $Z' \notin \Gamma$, $q_0', q_f \notin Q$ a δ' je definována takto:

- $\delta'(q_0', \varepsilon, Z') = \{(q_0, Z_0 Z')\}$
- jestliže $\delta(q, a, Z)$ obsahuje (r, γ), pak $\delta'(q, a, Z)$ obsahuje (r, γ)
- $\delta'(q, \varepsilon, Z') = \{(q_f, \varepsilon)\}$
  pro všechny $q \in Q$ $\square$

Grafická reprezentace konfigurací