

Greibachovej normálny tvar

Peter Kostolányi

5. marca 2024

1 Definícia a štandardná konštrukcia

Ku každej bezkontextovej gramatike existuje až na ε ekvivalentná bezkontextová gramatika, v ktorej sa všetky pravé strany pravidiel začínajú terminálom – o takýchto gramatikách budeme hovoriť, že sú v *Greibachovej normálnom tvare*. V každom kroku odvodenia gramatiky v Greibachovej normálnom tvare sa tak vygeneruje aspoň jeden terminálny symbol výsledného slova. Keďže sa navyše terminály generujú zľava, možno Greibachovej normálny tvar použiť na dôkaz, že ku každému zásobníkovému automatu akceptujúcemu prázdny zásobník resp. stavom existuje až na ε resp. plne ekvivalentný automat rovnakého druhu neobsahujúci prechody na prázdne slovo ε .

Definícia 1. Hovoríme, že bezkontextová gramatika $G = (N, T, P, \sigma)$ je v *Greibachovej normálnom tvare*, ak $P \subseteq N \times T(N \cup T)^*$.

Pomenovanie „normálny tvar“ je odôvodnené nasledujúcou vetou, ktorej detailný dôkaz možno nájsť v knihe Hopcrofta a Ullmana [1].

Veta 1. *Nech G je bezkontextová gramatika. Potom existuje bezkontextová gramatika G' v Greibachovej normálnom tvare taká, že $L(G') = L(G) - \{\varepsilon\}$.*

Poznámka 1. Greibachovej normálny tvar sa často definuje prostredníctvom o niečo silnejšej požiadavky $P \subseteq N \times TN^*$. Každú gramatiku v tvare z definície 1 možno do tejto podoby previesť zavedením nového neterminálu ξ_c s pravidlom $\xi_c \rightarrow c$ pre všetky terminály c a nahradením iných ako začiatočných terminálov na pravých stranách pôvodných pravidiel neterminálom ξ_c .

Niekde tiež možno naraziť na definíciu, pri ktorej sa v Greibachovej normálnom tvare povoľuje pravidlo $\sigma \rightarrow \varepsilon$ za predpokladu, že sa σ nevyskytuje na pravej strane žiadneho pravidla; takéto gramatiky dokážu vygenerovať aj prázdne slovo.

Vstupom nasledujúceho algoritmu na prevod bezkontextovej gramatiky do Greibachovej normálneho tvaru je „bezepsilonová“ bezkontextová gramatika $G = (N, T, P, \sigma)$ bez *reťazových pravidiel* (angl. *chain rules*) $\alpha \rightarrow \beta$ pre $\alpha, \beta \in N$. Predpokladajme, že je dané očíslovanie neterminálov gramatiky G ; čiže $N = \{\xi_1, \dots, \xi_n\}$ pre nejaké $n \in \mathbb{N} - \{0\}$.

1. Opakuj pre $k = 1, \dots, n$:
 - 1.1. Opakuj pre $j = 1, \dots, k - 1$:
 - 1.1.1. Opakuj pre všetky pravidlá $\xi_k \rightarrow \xi_j w$, kde w je slovo:
 - 1.1.1a. Pre každé pravidlo $\xi_j \rightarrow u$, kde u je slovo, pridaj pravidlo $\xi_k \rightarrow uw$.
 - 1.1.1b. Odober pravidlo $\xi_k \rightarrow \xi_j w$.
 - 1.2. Pridaj nový neterminál ξ_{k-n} .
 - 1.3. Opakuj pre všetky pravidlá $\xi_k \rightarrow \xi_k w$, kde w je slovo:
 - 1.3a. Pridaj pravidlá $\xi_{k-n} \rightarrow w \xi_{k-n}$ a $\xi_{k-n} \rightarrow w$.
 - 1.3b. Odober pravidlo $\xi_k \rightarrow \xi_k w$.
 - 1.4. Pre všetky pravidlá $\xi_k \rightarrow u$, kde u je slovo, pridaj pravidlo $\xi_k \rightarrow u \xi_{k-n}$.
2. Opakuj pre $k = n - 1, \dots, -(n - 1)$:
 - 2.1. Opakuj pre $j = n, \dots, k + 1$:
 - 2.1.1. Opakuj pre všetky pravidlá $\xi_k \rightarrow \xi_j w$, kde w je slovo:
 - 2.1.1a. Pre každé pravidlo $\xi_j \rightarrow u$, kde u je slovo, pridaj pravidlo $\xi_k \rightarrow uw$.
 - 2.1.1b. Odober pravidlo $\xi_k \rightarrow \xi_j w$.

Korektnosť uvedeného algoritmu je založená na nasledujúcich pozorovaniach.

- Po vykonaní kroku 1 pre $k = s$ je gramatika ekvivalentná pôvodnej a všetky pravé strany pravidiel z neterminálov $\xi_1, \dots, \xi_s$ sa začínajú buď terminálom, alebo neterminálom s väčším indexom.
- Všetky pravidlá vedúce z neterminálov $\xi_{-(n-1)}, \dots, \xi_0$ pridaných v rámci kroku 1 majú na pravej strane slovo začínajúce buď terminálom, alebo neterminálom s kladným indexom.
- Po vykonaní celého kroku 1 je tak gramatika ekvivalentná pôvodnej a pre všetky pravidlá $\xi_i \rightarrow \xi_j w$, kde $i, j \in \{-(n-1), \dots, n\}$, je $i < j$. Špeciálne sa už teda všetky pravé strany pravidiel z ξ_n musia začínať terminálom.
- Po vykonaní kroku 2 pre $k = s$ je gramatika ekvivalentná pôvodnej a terminálom sa začínajú všetky pravé strany pravidiel z neterminálov $\xi_s, \dots, \xi_n$.
- Po vykonaní celého kroku 2 je tak gramatika ekvivalentná pôvodnej a terminálom sa musia začínať všetky pravé strany pravidiel – gramatika je v Greibachovej normálnom tvare.

1.1 Riešená úloha

Úloha 1. Nech $G = (N, T, P, \sigma)$ je bezkontextová gramatika s $N = \{\sigma, \alpha\}$, $T = \{a, b\}$ a

$$P = \{\sigma \rightarrow \sigma a \mid \alpha \alpha \\ \alpha \rightarrow \sigma \sigma \mid \alpha \alpha \mid b\}.$$

Štandardnou konštrukciou prevedte gramatiku G do Greibachovej normálneho tvaru.

Riešenie. Po zavedení očíslovania $\sigma =: \xi_1$, $\alpha =: \xi_2$ dostávame gramatiku $G'' = (N'', T, P'', \xi_1)$, kde $N'' = \{\xi_1, \xi_2\}$ a

$$P'' = \{\xi_1 \rightarrow \xi_1 a \mid \xi_2 \xi_2 \\ \xi_2 \rightarrow \xi_1 \xi_1 \mid \xi_2 \xi_2 \mid b\}.$$

Táto gramatika zrejme spĺňa všetky predpoklady na vstup štandardného algoritmu opísaného vyššie. Výpočet algoritmu bude prebiehať nasledujúcim spôsobom.

- Krok 1, $k = 1$. V rámci kroku 1.1 iterujeme cez prázdnu množinu, preto ostane gramatika nezmenená. V kroku 1.2 pridáme neterminál ξ_{-1} . V kroku 1.3 pridáme pravidlá $\xi_{-1} \rightarrow a\xi_{-1}$, $\xi_{-1} \rightarrow a$ a odoberieme pravidlo $\xi_1 \rightarrow \xi_1 a$. V kroku 1.4 následne pridáme pravidlo $\xi_1 \rightarrow \xi_2 \xi_2 \xi_{-1}$. Výsledná sada pravidiel je

$$\xi_{-1} \rightarrow a\xi_{-1} \mid a \\ \xi_1 \rightarrow \xi_2 \xi_2 \mid \xi_2 \xi_2 \xi_{-1} \\ \xi_2 \rightarrow \xi_1 \xi_1 \mid \xi_2 \xi_2 \mid b.$$

- Krok 1, $k = 2$. V kroku 1.1 pridáme pravidlá $\xi_2 \rightarrow \xi_2 \xi_2 \xi_1$, $\xi_2 \rightarrow \xi_2 \xi_2 \xi_{-1} \xi_1$ a odoberieme pravidlo $\xi_2 \rightarrow \xi_1 \xi_1$. Výsledná sada pravidiel je

$$\xi_{-1} \rightarrow a\xi_{-1} \mid a \\ \xi_1 \rightarrow \xi_2 \xi_2 \mid \xi_2 \xi_2 \xi_{-1} \\ \xi_2 \rightarrow \xi_2 \xi_2 \mid b \mid \xi_2 \xi_2 \xi_1 \mid \xi_2 \xi_2 \xi_{-1} \xi_1.$$

V rámci kroku 1.2 pridáme neterminál ξ_0 . V kroku 1.3 pridáme pravidlá $\xi_0 \rightarrow \xi_2 \xi_0$, $\xi_0 \rightarrow \xi_2$, $\xi_0 \rightarrow \xi_2 \xi_1 \xi_0$, $\xi_0 \rightarrow \xi_2 \xi_1$, $\xi_0 \rightarrow \xi_2 \xi_{-1} \xi_1 \xi_0$, $\xi_0 \rightarrow \xi_2 \xi_{-1} \xi_1$ a odoberieme pravidlá $\xi_2 \rightarrow \xi_2 \xi_2$,

$\xi_2 \rightarrow \xi_2 \xi_2 \xi_1$ a $\xi_2 \rightarrow \xi_2 \xi_2 \xi_{-1} \xi_1$. V kroku 1.4 pridáme pravidlo $\xi_2 \rightarrow b \xi_0$. Výsledná sada pravidiel je

$$\begin{aligned} \xi_{-1} &\rightarrow a \xi_{-1} \mid a \\ \xi_0 &\rightarrow \xi_2 \xi_0 \mid \xi_2 \mid \xi_2 \xi_1 \xi_0 \mid \xi_2 \xi_1 \mid \xi_2 \xi_{-1} \xi_1 \xi_0 \mid \xi_2 \xi_{-1} \xi_1 \\ \xi_1 &\rightarrow \xi_2 \xi_2 \mid \xi_2 \xi_2 \xi_{-1} \\ \xi_2 &\rightarrow b \mid b \xi_0. \end{aligned}$$

Pravidlá z neterminálu ξ_2 sú už teraz požadovaného tvaru. Vykonávanie algoritmu pokračuje krokom 2.

- Krok 2, $k = 1$. Pridáme pravidlá $\xi_1 \rightarrow b \xi_2$, $\xi_1 \rightarrow b \xi_0 \xi_2$, $\xi_1 \rightarrow b \xi_2 \xi_{-1}$ a $\xi_1 \rightarrow b \xi_0 \xi_2 \xi_{-1}$. Odoberieme pravidlá $\xi_1 \rightarrow \xi_2 \xi_2$ a $\xi_1 \rightarrow \xi_2 \xi_2 \xi_{-1}$. Výsledná sada pravidiel je

$$\begin{aligned} \xi_{-1} &\rightarrow a \xi_{-1} \mid a \\ \xi_0 &\rightarrow \xi_2 \xi_0 \mid \xi_2 \mid \xi_2 \xi_1 \xi_0 \mid \xi_2 \xi_1 \mid \xi_2 \xi_{-1} \xi_1 \xi_0 \mid \xi_2 \xi_{-1} \xi_1 \\ \xi_1 &\rightarrow b \xi_2 \mid b \xi_0 \xi_2 \mid b \xi_2 \xi_{-1} \mid b \xi_0 \xi_2 \xi_{-1} \\ \xi_2 &\rightarrow b \mid b \xi_0. \end{aligned}$$

- Krok 2, $k = 0$. Pridáme prepisovacie pravidlá $\xi_0 \rightarrow b \xi_0$, $\xi_0 \rightarrow b \xi_0 \xi_0$, $\xi_0 \rightarrow b$, $\xi_0 \rightarrow b \xi_0$, $\xi_0 \rightarrow b \xi_1 \xi_0$, $\xi_0 \rightarrow b \xi_0 \xi_1 \xi_0$, $\xi_0 \rightarrow b \xi_1$, $\xi_0 \rightarrow b \xi_0 \xi_1$, $\xi_0 \rightarrow b \xi_{-1} \xi_1 \xi_0$, $\xi_0 \rightarrow b \xi_0 \xi_{-1} \xi_1 \xi_0$, $\xi_0 \rightarrow b \xi_{-1} \xi_1$ a $\xi_0 \rightarrow b \xi_0 \xi_{-1} \xi_1$. Odoberieme všetkých šesť pôvodných pravidiel z neterminálu ξ_0 . Výsledná sada pravidiel (po odstránení duplikátov) je

$$\begin{aligned} \xi_{-1} &\rightarrow a \xi_{-1} \mid a \\ \xi_0 &\rightarrow b \xi_0 \mid b \xi_0 \xi_0 \mid b \mid b \xi_1 \xi_0 \mid b \xi_0 \xi_1 \xi_0 \mid b \xi_1 \mid b \xi_0 \xi_1 \mid b \xi_{-1} \xi_1 \xi_0 \mid b \xi_0 \xi_{-1} \xi_1 \xi_0 \mid b \xi_{-1} \xi_1 \mid b \xi_0 \xi_{-1} \xi_1 \\ \xi_1 &\rightarrow b \xi_2 \mid b \xi_0 \xi_2 \mid b \xi_2 \xi_{-1} \mid b \xi_0 \xi_2 \xi_{-1} \\ \xi_2 &\rightarrow b \mid b \xi_0. \end{aligned}$$

- Krok 2, $k = -1$. Keďže iterujeme cez prázdnu množinu pravidiel, gramatika ostane nezmenená.

Môžeme teda uzavrieť, že gramatika G je ekvivalentná gramatike $G' = (N', T, P', \xi_1)$ v Greibachovej normálnom tvare, kde $N' = \{\xi_{-1}, \xi_0, \xi_1, \xi_2\}$ a

$$\begin{aligned} P' &= \{ \xi_{-1} \rightarrow a \xi_{-1} \mid a \\ &\quad \xi_0 \rightarrow b \xi_0 \mid b \xi_0 \xi_0 \mid b \mid b \xi_1 \xi_0 \mid b \xi_0 \xi_1 \xi_0 \mid b \xi_1 \mid b \xi_0 \xi_1 \mid b \xi_{-1} \xi_1 \xi_0 \mid b \xi_0 \xi_{-1} \xi_1 \xi_0 \mid b \xi_{-1} \xi_1 \mid b \xi_0 \xi_{-1} \xi_1 \\ &\quad \xi_1 \rightarrow b \xi_2 \mid b \xi_0 \xi_2 \mid b \xi_2 \xi_{-1} \mid b \xi_0 \xi_2 \xi_{-1} \\ &\quad \xi_2 \rightarrow b \mid b \xi_0 \}. \end{aligned}$$

□

2 „Bezepsilonové“ zásobníkové automaty

Lahko vidieť, že ak ku gramatike v Greibachovej normálnom tvare zostrojíme pomocou štandardnej konštrukcie ekvivalentný zásobníkový automat, tak ľubovoľná jeho konfigurácia s neterminálom na vrchu zásobníka môže byť nasledovaná iba konfiguráciou, v ktorej je na vrchu zásobníka terminálny symbol. Ak upravíme prechodovú funkciu tohto automatu tak, aby tieto terminály čítal zo vstupu vždy už v predchádzajúcom kroku výpočtu, získame ekvivalentný automat neobsahujúci prechody na prázdne slovo.

Veta 2. *Nech $G = (N, T, P, \sigma)$ je bezkontextová gramatika v Greibachovej normálnom tvare. Potom existuje zásobníkový automat $A = (K, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ taký, že $N(A) = L(G)$ a pre všetky $q \in K$ a $Z \in \Gamma$ je $\delta(q, \varepsilon, Z) = \emptyset$.*

Dôkaz. Automat A zostrojíme nasledovne: $K = \{q_0\}$, $\Sigma = T$, $\Gamma = N \cup T$, $Z_0 = \sigma$, $F = \emptyset$, pričom pre všetky $\xi \in N$, $c \in T$ a $x \in (N \cup T)^*$ také, že $\xi \rightarrow cx \in P$ je $(q_0, x^R) \in \delta(q_0, c, \xi)$ a pre všetky $c \in T$ je $(q_0, \varepsilon) \in \delta(q_0, c, c)$. Automat A nemá definované žiadne ďalšie prechody. Dôkaz tvrdenia $N(A) = L(G)$ prenechávame čitateľovi. $\square$

Zásobníkové automaty $A = (K, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ spĺňajúce pre všetky $q \in K$ a $Z \in \Gamma$ podmienku $\delta(q, \varepsilon, Z) = \emptyset$ môžeme nazvať „bezepsilonovými“. Nasledujúci dôsledok hovorí, že pri akceptácii prázdny zásobník ide až na ε o normálny tvar zásobníkových automatov.

Dôsledok 1. *Ku každému bezkontextovému jazyku L existuje „bezepsilonový“ zásobníkový automat taký, že $N(A) = L - \{\varepsilon\}$.*

V prípade akceptácie stavom ide dokonca o plnohodnotný normálny tvar zásobníkových automatov – pre všetky bezkontextové jazyky L existuje „bezepsilonový“ zásobníkový automat A taký, že $L(A) = L$. Dôkaz tohto tvrdenia je náplňou jednej z úloh na nasledujúce cvičenie.

Literatúra

- [1] Hopcroft, J. E., Ullman, J. D.: *Introduction to Automata Theory, Languages, and Computation*. Reading : Addison-Wesley, 1979. ISBN 0-201-02988-X.