Języki, automaty i obliczenia

Wykład 7: Języki bezkontekstowe

Sławomir Lasota

Uniwersytet Warszawski

8 kwietnia 2015

Plan

Gramatyka bezkontekstowa

$$\mathcal{G} \;=\; (A, N, S, \alpha)$$

- A to skończony zbiór symboli końcowych (*terminalnych*)
- N to skończony zbiór symboli niekońcowych (*nieterminalnych*), $A \cap N = \emptyset$
- $S \in N$ to symbol początkowy
- $\alpha \subseteq N \times (A \cup N)^*$ to skończony zbiór reguł przepisywania (*produkcji*)

Notacja

Reguły $(X, v) \in \alpha$ będziemy zapisywać $X \longrightarrow_{\mathcal{G}} v$ albo $X \longrightarrow v$.

Reguły rozszerzamy do relacji $\longrightarrow_{\mathcal{G}} \;\subseteq\; (A \cup N)^* \times (A \cup N)^*$:

$$w \longrightarrow_{\mathcal{G}} w' \quad\Longleftrightarrow\quad \exists (X \longrightarrow v) \in \alpha,\ \exists u, t \in (A \cup N)^*.\ \ w \;=\; u\,X\,t, \quad v \;=\; u\,v\,t$$

i domykamy zwrotno-tranzytywnie: $w \twoheadrightarrow_{\mathcal{G}} w'$ wtw. gdy istnieje ciąg

$$w \;=\; v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_n \;=\; w', \qquad n \geq 0$$

zwany *wyprowadzeniem* (wywodem) w' z w (w gramatyce $\mathcal{G}$).

Języki bezkontekstowe

Język bezkontekstowy to język *generowany* przez gramatykę:

$$L(\mathcal{G}) \;=\; \{w \in A^* \;:\; S \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w\}$$

Przykład

Gramatyka $\mathcal{G}$:

- $A \;=\; \{a, b, +, \cdot\}$
- $N \;=\; \{W, P\}$
- $S \;=\; W$
- $\alpha \;=\; \{\, (W, P),\ (W, W + W),\ (W, W \cdot W),$
  $(P, aPa),\ (P, bPb),\ (P, a),\ (P, b),\ (P, W)\,\}$

$$W \longrightarrow P \mid W + W \mid W \cdot W$$
$$P \longrightarrow a\,P\,a \mid b\,P\,b \mid a \mid b \mid W$$

Słowo $a\,a\,a \cdot b\, + a \in L(\mathcal{G})$, ponieważ ma wyprowadzenie:

$W \longrightarrow_{\mathcal{G}} W \cdot W \longrightarrow_{\mathcal{G}} P \cdot W \longrightarrow_{\mathcal{G}} a\,P\,a \cdot W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot W + W \longrightarrow_{\mathcal{G}}$
$a\,a\,a \cdot P + W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + P \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + a$

$L(\mathcal{G}) \;=\; ?$

Pytanie

Czy każde słowo ma co najwyżej jedno wyprowadzenie w gramatyce?

Pytanie

Czy każde słowo ma co najwyżej jedno wyprowadzenie w gramatyce?

Odpowiedź

$W \longrightarrow W + W \mid \varepsilon$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka spełniająca ten warunek?

Przykłady języków bezkontekstowych

- palindromy nad $\{a, b\}$:

$$P \longrightarrow aPa \mid bPb \mid a \mid b \mid \varepsilon$$

- wyrażenia arytmetyczne:

$$\begin{aligned} S &\longrightarrow S + I \mid I \\ I &\longrightarrow I \cdot C \mid C \\ C &\longrightarrow (S) \mid 0 \mid 1 \end{aligned} \qquad\qquad W \longrightarrow W \cdot W \mid W + W \mid (W) \mid 0 \mid 1$$

- $\{a^n b^n \;:\; n \in \mathbb{N}\}$

$$X \longrightarrow aXb \mid \varepsilon$$

- $\{w \in \{a, b\}^* \;:\; \#_w(a) \;=\; \#_w(b)\}$

$$X \longrightarrow aXb \mid bXa \mid XX \mid \varepsilon$$

- poprawnie zbudowane wyrażenia nawiasowe:

$$W \longrightarrow WW \mid [W] \mid (W) \mid \langle W \rangle \mid \varepsilon$$

- ...

Fakt

Każdy język regularny jest bezkontekstowy.

Dowód:

Niech $\mathcal{A} = (A, Q, \{q_0\}, F, \delta)$. Definiujemy gramatykę $\mathcal{G} = (A, N, S, \alpha)$:

- $N = Q$
- $S = q_0$
- $\alpha = \{(q, a\,q') : (q, a, q') \in \delta\} \cup \{(q, \varepsilon) : q \in F\}$

$$w \in L(\mathcal{G}) \iff q_0 \twoheadrightarrow_{\mathcal{G}} w \iff \exists q \in F.\ \widehat{\delta}(q_0, w, q) \iff w \in L(\mathcal{A})$$

Wyprowadzenia lewostronne

Wyprowadzenie:

$$
\begin{aligned}
w \longrightarrow_{\mathcal{G}} w' &\iff \exists (X \longrightarrow v) \in \alpha,\ \exists u, t \in (A \cup N)^*.\ w \;=\; u\,X\,t,\quad v \;=\; u\,v\,t \\
w \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w' &\iff \exists n \geq 0,\ v_0, \ldots, v_n.\ \ w \;=\; v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_n \;=\; w'
\end{aligned}
$$

Wspólny prefiks: $w \longrightarrow_{\mathcal{G}} w',\ \ w \;\neq\; w'$

trywialne reguły!

$$
\mathrm{wp}(w, w') \;=\; |u|
$$

Wyprowadzenie *lewostronne*:

$$
\begin{aligned}
w \xrightarrow{L}\!\!\!\!\rightarrow_{\mathcal{G}} w' \iff\ & \exists n \geq 0,\ v_0, \ldots, v_n.\ \ w \;=\; v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_n \;=\; w', \\
& \forall 0 < i < n.\ \mathrm{wp}(v_{i-1}, v_i) \leq \mathrm{wp}(v_i, v_{i+1})
\end{aligned}
$$

(Wyprowadzenie *prawostronne*?)

Fakt

Jeśli $w \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w'$ *to* $w \xrightarrow{L}\!\!\!\!\rightarrow_{\mathcal{G}} w'$.

Wyprowadzenia lewostronne (dowód)

Przykład

To wyprowadzenie jest lewostronne:

$W \longrightarrow_{\mathcal{G}} W \cdot W \longrightarrow_{\mathcal{G}} P \cdot W \longrightarrow_{\mathcal{G}} a\,P\,a \cdot W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot W + W \longrightarrow_{\mathcal{G}}$
$a\,a\,a \cdot P + W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + W \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + P \longrightarrow_{\mathcal{G}} a\,a\,a \cdot b + a$

Fakt

Jeśli $w \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w'$ *to* $w \overset{L}{\longrightarrow\!\!\!\!\rightarrow}_{\mathcal{G}} w'$*.*

Dowód:

Indukcja ze względu na długość wyprowadzenia. Krok indukcyjny:

$$(v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_{n-1} \longrightarrow_{\mathcal{G}} v_n) \longrightarrow_{\mathcal{G}} v_{n+1}$$

$$v_{n-1} = uXtYv \qquad v_n = uXtyv \qquad v_{n+1} = uxtyv \qquad (Y \longrightarrow y,\ X \longrightarrow x)$$

$$v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_{n-1} \longrightarrow_{\mathcal{G}} v' \qquad (X \longrightarrow x)$$

$$v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_{n-1} \longrightarrow_{\mathcal{G}} v' \longrightarrow_{\mathcal{G}} v_{n+1} \qquad (Y \longrightarrow y)$$

Język bezkontekstowy jako najmniejszy punkt stały

$L_a = \{a\}$, dla $a \in A$. $L_X \subseteq A^*$, dla $X \in N$.

Zamiast reguły $X \longrightarrow_{\mathcal{G}} X_1 \ldots X_n$, inkluzja $L_X \supseteq L_{X_1} \ldots L_{X_n}$.

Zamiast zbioru wszytkich reguł dla nieterminala X:

$$X \longrightarrow_{\mathcal{G}} X_1 \ldots X_n \mid Y_1 \ldots Y_m \mid Z_1 \ldots Z_k \mid \ldots,$$

inkluzja

$$L_X \supseteq L_{X_1} \ldots L_{X_n} \cup L_{Y_1} \ldots L_{Y_m} \cup L_{Z_1} \ldots L_{Z_k} \cup \ldots$$

Rozważmy N-tuple języków $L = (L_X)_{X \in N} \in (\mathcal{P}(A^*))^N$. Zdefiniujmy

$$f : (\mathcal{P}(A^*))^N \to (\mathcal{P}(A^*))^N$$
$$f(L)_X = L_{X_1} \ldots L_{X_n} \cup L_{Y_1} \ldots L_{Y_m} \cup L_{Z_1} \ldots L_{Z_k} \cup \ldots$$

Gramatyka bezkontekstowa definiuje najmniejszy L t.że

$$\forall X \in N.\ L_X \supseteq f(L)_X \quad \text{czyli} \quad L \supseteq f(L)$$

Plan

Drzewa wyprowadzeń

Drzewo to $T \subseteq \mathbb{N}^*$

- zamknięty na prefiksy: $\forall w, v \in \mathbb{N}^*.\ wv \in T \implies w \in T$
- zamknięty „w lewo”: $\forall w \in \mathbb{N}^*, i, j \in \mathbb{N}.\ wi \in T,\ j < i \implies wj \in T$

Drzewo nad X to para (T, l), gdzie $l : T \to X$.

Plon drzewa (T, l) nad X to słowo $w \in X^*$ otrzymane z etykiet liści $l(w)$, w kojelności leksykograficznej (od lewej do prawej).

Niech $\mathcal{G} = (A, N, S, \alpha)$.

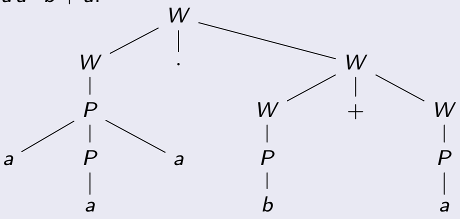
Drzewo wyprowadzenia (wywodu) słowa $w \in (A \cup N)^*$ w gramatyce $\mathcal{G}$ to drzewo nad $A \cup N$ t.że

- S jest etykietą korzenia
- plon drzewa to słowo w
- $$\begin{array}{cccc} & X & & \\ X_1 & X_2 & \dots & X_n \end{array} \implies (X,\ X_1 X_2 \dots X_n) \in \alpha$$

Wyprowadzenia a drzewa wyprowadzeń

Przykład

Drzewo wyprowadzenia dla $a\,a\,a \cdot b + a$:



Fakt

$w \in L(\mathcal{G})$ wtw. gdy istnieje drzewo wyprowadzenia w w gramatyce $\mathcal{G}$.

Dowód:

$\Longleftarrow$: indukcja ze względu na głębokość drzewa

$\Longrightarrow$: indukcja ze względu na długość wyprowadzenia, korzystamy z obserwacji:

jeśli $w\,v \longrightarrow\!\!\!\!\!\rightarrow_{\mathcal{G}} u$ to $u = u_1\,u_2$ dla pewnych $u_1, u_2 \in (A \cup N)^*$ t.że

$$w \longrightarrow\!\!\!\!\!\rightarrow_{\mathcal{G}} u_1 \qquad \text{i} \qquad v \longrightarrow\!\!\!\!\!\rightarrow_{\mathcal{G}} u_2.$$

Jednoznaczność

Pytanie

Czy każde słowo ma co najwyżej jedno drzewo wyprowadzenia w gramatyce?

Pytanie

Czy każde słowo ma co najwyżej jedno drzewo wyprowadzenia w gramatyce?

Odpowiedź

$W \longrightarrow W + W \mid \varepsilon$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka *jednoznaczna*, tzn. taka, w której każde słowo ma co najwyżej jedno drzewo wyprowadzenia?

Jednoznaczność

Pytanie

Czy każde słowo ma co najwyżej jedno drzewo wyprowadzenia w gramatyce?

Odpowiedź

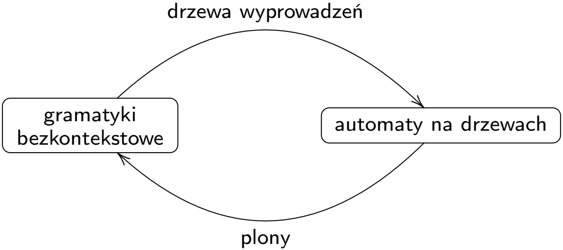
$W \longrightarrow W + W \mid \varepsilon$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka *jednoznaczna*, tzn. taka, w której każde słowo ma co najwyżej jedno drzewo wyprowadzenia?

Odpowiedź

$L = \{a^n b^n c^m d^m : n, m \in \mathbb{N}\} \cup \{a^n b^m c^m d^n : n, m \in \mathbb{N}\}$


drzewa wyprowadzeń
gramatyki bezkontekstowe
automaty na drzewach
plony

Twierdzenie

Niech $\mathcal{A} = (A, Q, I, F, \delta)$ automat na drzewach. Język

$$plony(L(\mathcal{A})) = \{plon(t) \ : \ t \in L(\mathcal{A})\}$$

jest językiem bezkontekstowym.

Dowód:

Definiujemy gramatykę bezkontekstową $\mathcal{G} = (A, N, S, \alpha)$:

- $N = Q \times A \cup \{S\}$
- zbiór reguł α:

$$\begin{aligned}
S &\longrightarrow_{\mathcal{G}} \varepsilon &&\iff I \cap F \neq \emptyset \\
S &\longrightarrow_{\mathcal{G}} (q, a) &&\iff q \in I \\
(q, a) &\longrightarrow_{\mathcal{G}} (q_1, a_1) \dots (q_n, a_n) &&\iff (q, a, q_1 \dots q_n) \in \delta \qquad (\text{arność}(a) = n) \\
(q, a) &\longrightarrow_{\mathcal{G}} a &&\iff \exists w \in F^*.\ (q, a, w) \in \delta
\end{aligned}$$

$w \in L(\mathcal{G}) \iff$ istnieje drzewo wyprowadzenia słowa w w gramatyce $\mathcal{G} \iff$
$\exists t.\ \mathcal{A}$ ma bieg akceptujący nad $t \wedge \text{plon}(t) = w \iff w \in \text{plony}(L(\mathcal{A}))$

Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ *gramatyka bezkontekstowa. Zbiór*

$$dw(\mathcal{G}) = \{\text{drzewa wyprowadzeń gramatyki } \mathcal{G} \text{ wszystkich słów } w \in A^*\}$$

jest regularnym językiem drzew.

Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ *gramatyka bezkontekstowa. Zbiór*

$$dw(\mathcal{G}) = \{\textit{drzewa wyprowadzeń gramatyki } \mathcal{G} \textit{ wszystkich słów } w \in A^*\}$$

jest regularnym językiem drzew.

Konieczna poprawka!

Gramatyka bezkontekstowa $\mapsto$ automat na drzewach

Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ *gramatyka bezkontekstowa. Zbiór*

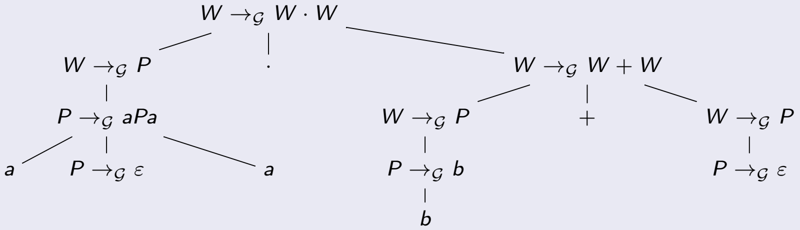
$$dw(\mathcal{G}) = \{\textit{drzewa wyprowadzeń gramatyki } \mathcal{G} \textit{ wszystkich słów } w \in A^*\}$$

jest regularnym językiem drzew.

Konieczna poprawka!

W drzewie wyprowadzenia zastąp każdą etykietę $X \in N$ przez odpowiednią regułę:

Przykład



Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ gramatyka bezkontekstowa. Zbiór

$$dw(\mathcal{G}) = \{\textit{drzewa wyprowadzeń gramatyki } \mathcal{G} \textit{ wszystkich słów } w \in A^*\}$$

jest regularnym językiem drzew.

Dowód:

Definiujemy automat deterministyczny na drzewach $\mathcal{A} = (B, Q, I, F, \delta)$:

- $B = A \cup \alpha$ $\quad$ $\text{arność}(a) = 1, \ \text{arność}((X, w)) = \begin{cases} |w|, & |w| > 0 \\ 1, & \text{w p.p.} \end{cases}$
- $Q = A \cup N \cup \{q_f\}$
- $I = \{S\}$
- $F = \{q_f\}$
- $\delta = \begin{array}{l} \{(X, X, q_f) : X \in A\} \cup \\ \{(X, (X, w), w) : (X, w) \in \alpha, |w| > 0\} \cup \\ \{(X, (X, w), q_f) : (X, w) \in \alpha, |w| = 0\} \end{array}$

$t \in L(\mathcal{A}) \iff t \in \mathrm{dw}(\mathcal{G})$

Plan

Postaci normalne

Załóżmy, że $\varepsilon \notin L(\mathcal{G})$.

Postać normalna Chomskiego

$$X \longrightarrow a \quad (a \in A) \qquad\qquad X \longrightarrow YZ \quad (Y, Z \in N)$$

Dowód:

$$X \longrightarrow a \quad (a \in A) \qquad\qquad X \longrightarrow w \quad (w \in N^*)$$

Postać normalna Greibach

$$X \longrightarrow a\,w \quad (a \in A,\ w \in N^*)$$

Problemy decyzyjne

Pustość

Dane: język bezkontekstowy L

Wynik: czy $L = \emptyset$?

Uniwersalność

Dane: język bezkontekstowy L

Wynik: czy $L = A^*$?

Równość

Dane: języki bezkontekstowe L, M

Wynik: czy $L = M$?

Problem decyzyjny

Dane:	Gramatyka bezkontekstowa $\mathcal{G} = (A, N, S, \alpha)$ w postaci Chomskiego i słowo $a_1 \dots a_n$
Wynik:	czy $a_1 \dots a_n \in L$?

Problem przynależności

Problem decyzyjny

Dane: Gramatyka bezkontekstowa $\mathcal{G} = (A, N, S, \alpha)$ w postaci Chomskiego i słowo $a_1 \ldots a_n$

Wynik: czy $a_1 \ldots a_n \in L$?

$$V_{i,j} = \{X \in N : X \twoheadrightarrow_{\mathcal{G}} a_i \ldots a_j\} \qquad (i \leq j)$$

Algorytm

`dla` $i \in \{1 \ldots n\}$, $V_{i,i} := \{X \in N : X \longrightarrow_{\mathcal{G}} a_i\}$

`dla` $i, j \in \{1 \ldots n\}$, $i < j$, $V_{i,j} := \emptyset$

`powtarzaj dla` $1 \leq i \leq j < k \leq n$

$V_{i,k} := V_{i,k} \cup \{X \in N : \exists Y, Z \in N.\ (X, YZ) \in \alpha, Y \in V_{i,j}, Z \in V_{j+1,k}\}$

`aż do stabilizacji`

`wynik` $:= (S \in V_{1,n})$

W następnym odcinku:

automaty dla języków bezkontekstowych