Języki, automaty i obliczenia

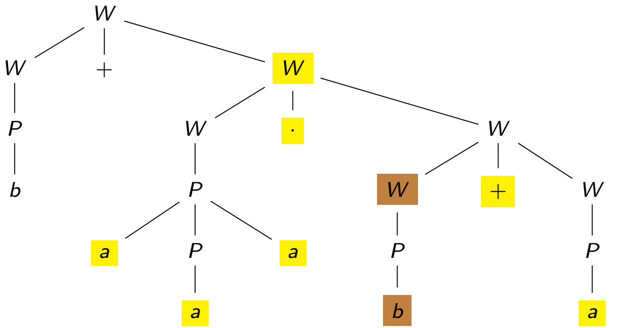
Wykład 7: Własności języków bezkontekstowych

Sławomir Lasota

Uniwersytet Warszawski

17 kwietnia 2024

Plan



$$b + \boxed{a\,a\,a\,\cdot}\ \boxed{b}\ \boxed{+\,a} \quad \in L(\mathcal{G})$$

$$b + \boxed{a\,a\,a\,\cdot}\ \boxed{a\,a\,a\,\cdot}\ \boxed{b}\ \boxed{+\,a}\ \boxed{+\,a} \quad \in L(\mathcal{G})$$

$$b + \boxed{b} \quad \in L(\mathcal{G})$$

Pompowanie

Lemat o pompowaniu

Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w\,y\,u\,z\,v, \text{ gdzie } |y\,z| > 0 \text{ i } |y\,u\,z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w\,y^m\,u\,z^m\,v \in L$.

$$L \text{ bezkontekstowy} \implies \exists n\ \forall x\ \exists x{=}wyuzv\ \forall m.\ w\,y^m\,u\,z^m\,v \in L$$

Wniosek

L jest niebezkontekstowy, jeśli dla każdego n, istnieje słowo $x \in L$, $|x| \geq n$ takie, że dla każdego przedstawienia x jako

$$x = w\,y\,u\,z\,v, \text{ gdzie } |y\,z| > 0 \text{ i } |y\,u\,z| \leq n,$$

istnieje $m \in \mathbb{N}$ takie, że $w\,y^m\,u\,z^m\,v \notin L$.

$$L \text{ niebezkontekstowy} \impliedby \underbrace{\forall n\ \exists x}_{x_n}\ \forall x{=}wyuzv\ \exists m.\ w\,y^m\,u\,z^m\,v \notin L$$

Pompowanie – dowód

Lemat o pompowaniu

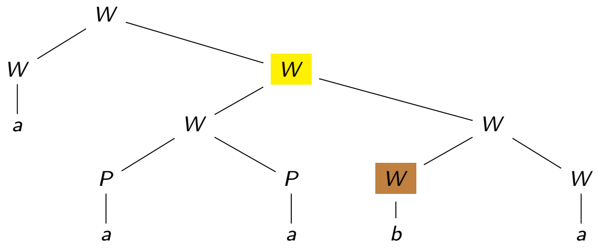
Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w\,y\,u\,z\,v, \text{ gdzie } |y\,z| > 0 \text{ i } |y\,u\,z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w\,y^m\,u\,z^m\,v \in L$.

Dowód:

Niech $\mathcal{G} = (A, N, S, \alpha)$ gramatyka w postaci Chomsky'ego dla $L - \{\varepsilon\}$. Niech n takie, że każde drzewo wyprowadzenia dla słowa długości $\geq n$ ma ścieżkę dłuższą niż $|N|$.



$$L \text{ niebezkontekstowy} \iff \underbrace{\forall n\ \exists x}_{x_n}\ \forall x{=}wyuzv\ \exists m.\ w\,y^m\,u\,z^m\,v \notin L$$

Przykład

Pokażemy, że język

$$L \;=\; \{\, w\, w \;:\; w \in \{a, b\}^* \,\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo

$$x_n \;=\; a^n\, b^n\, a^n\, b^n \in L.$$

Rozważmy dowolne słowa w, y, u, z, v takie, że

$$x_n = w\, y\, u\, z\, v, \text{ gdzie } |y\, z| > 0 \text{ i } |y\, u\, z| \le n.$$

Dwa przypadki: albo obydwie litery a, b występują w którymś ze słów y, z, albo nie.

Przykład

Język

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

nie jest bezkontekstowy.

$$L \text{ niebezkontekstowy} \quad \Longleftarrow \quad \underbrace{\forall n\ \exists x}_{x_n}\ \forall x{=}wyuzv\ \exists m.\ w\, y^m\, u\, z^m\, v \notin L$$

Przykład

Spróbujmy pokazać, że język

$$L \;=\; \{\, a^i\, b^j\, c^k \;:\; i, j, k \text{ parami różne} \,\}$$

nie jest bezkontekstowy.

Niech $n = 4$ i $x = a^i b^j c^k \in L$ dowolne, $|x| \geq n$.

B.u.o. $k > i, j$. Wtedy $k \geq 3$. Niech

$$w \in a^i\, b^j\, c^*, \qquad y\, u\, z\, v \in c^+, \qquad |yz| \neq k - i,\; k - j.$$

Dla każdego $m \in \mathbb{N}$, $w\, y^m\, u\, z^m\, v \in L$. Pompowanie nie działa...

Podobny kontrprzykład można podać dla języków regularnych.

Pompowanie dla zaawansowanych :)

Rozważmy słowa z *zaznaczonymi* pozycjami. Np. słowo

$$a\,\underline{a}\,\underline{a}\,b\,a\,a\,a\,\underline{b}$$

ma zaznaczone 3 pozycje.

Niech $|w|$ oznacza liczbę zaznaczonych pozycji w słowie w.

Lemat o pompowaniu (Ogden 1968)

Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w\,y\,u\,z\,v, \text{ gdzie } |y\,z| > 0 \text{ i } |y\,u\,z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w\,y^m\,u\,z^m\,v \in L$.

$$L \text{ bezkontekstowy} \implies \exists n\ \forall x\ \exists x{=}wyuzv\ \forall m.\ w\,y^m\,u\,z^m\,v \in L$$

Dowód:

Tak samo.

$$L \text{ niebezkontekstowy} \quad \Longleftarrow \quad \underbrace{\forall n\ \exists x}_{x_n}\ \forall x{=}wyuzv\ \exists m.\ w\, y^m\, u\, z^m\, v \notin L$$

Przykład

Pokażemy, że język

$$L \ = \ \{\, a^i\, b^j\, c^k \ : \ i, j, k \text{ parami różne}\,\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo

$$x_n \ = \ \underline{a}^n\, b^{n+n!}\, c^{n+2n!} \in L.$$

Rozważmy dowolne słowa w, y, u, z, v takie, że

$$x_n = w\, y\, u\, z\, v, \ \text{gdzie } |y\, z| > 0 \text{ i } |y\, u\, z| \leq n.$$

Wtedy któreś ze słów y, z musi zawierać przynajmniej jedno a.

Trzy istotne przypadki: albo $yz \in a^+$, albo $y \in a^+, z \in b^*$, albo $y \in a^+, z \in c^*$.

Plan

Deterministyczne automaty ze stosem

Automat ze stosem $\mathcal{A} = (A, Q, q_0, F, S, s_0, \delta)$.

Niech $\delta(q, s, a) = \{ (p, w) \in Q \times S^* \ : \ q \xrightarrow{\mathbf{pop}(s), a, \mathbf{push}(w)} p \}$.

Automat $\mathcal{A}$ jest *deterministyczny*, jeśli dla każdej pary $(q, s) \in Q \times S$:

- albo $|\delta(q, s, a)| = 1$ dla każdego $a \in A$, oraz $\delta(q, s, \varepsilon) = \emptyset$;
- albo $\delta(q, s, a) = \emptyset$ dla każdego $a \in A$, oraz $|\delta(q, s, \varepsilon)| = 1$.

Pytanie

Czy ten automat jest deterministyczny?

- $A = \{a, b\}$
- $Q = \{q_0, q, q_f\}$
- $F = \{q_f\}$
- $S = \{s_0, s\}$

$$q_0 \xrightarrow{\mathbf{pop}(s_0), a, \mathbf{push}(ss_0)} q_0 \qquad q_0 \xrightarrow{\mathbf{pop}(s_0), \varepsilon, \mathbf{push}(s_0)} q \qquad q \xrightarrow{\mathbf{pop}(s_0), \varepsilon, \mathbf{push}(\varepsilon)} q_f$$

$$q_0 \xrightarrow{\mathbf{pop}(s), a, \mathbf{push}(ss)} q_0 \qquad q_0 \xrightarrow{\mathbf{pop}(s), \varepsilon, \mathbf{push}(s)} q \qquad q \xrightarrow{\mathbf{pop}(s), b, \mathbf{push}(\varepsilon)} q$$

Jak go zdeterminizować?

- $Q \;=\; \{q_0, \overline{q}, q, q_f\}$
- $F \;=\; \{q_0, q_f\}$

$$q_0 \xrightarrow{\mathbf{pop}(s_0),a,\mathbf{push}(ss_0)} \overline{q} \qquad \overline{q} \xrightarrow{\mathbf{pop}(s),a,\mathbf{push}(ss)} \overline{q} \qquad q \xrightarrow{\mathbf{pop}(s_0),\varepsilon,\mathbf{push}(\varepsilon)} q_f$$

$$\overline{q} \xrightarrow{\mathbf{pop}(s),b,\mathbf{push}(\varepsilon)} q \qquad q \xrightarrow{\mathbf{pop}(s),b,\mathbf{push}(\varepsilon)} q$$

Pytanie

Czy na każdym słowie automat deterministyczny ma dokładnie jeden bieg?

Pytanie

Czy (deterministyczne języki bezkontekstowe) $\equiv$ (języki bezkontekstowe)?

Język

$$\{\, a^n\, b^n \;:\; n \in \mathbb{N} \,\} \;\cup\; \{\, a^n\, b^{2n} \;:\; n \in \mathbb{N} \,\}$$

jest rozpoznawany przez automat ze stosem, ale nie przez automat deterministyczny.

Warianty

Pytanie

Czy (deterministyczne automaty ze stosem) $\equiv$ (deterministyczne automaty ze stosem bez pustych przejść) ?

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

Pytanie

Czy (deterministyczne automaty ze stosem) $\equiv$ (deterministyczne automaty ze stosem bez stanów) ?

Pytanie

Czy (języki regularne) $\subseteq$ (deterministyczne automaty ze stosem bez stanów)?

Równość deterministycznych języków bezkontekstowych

Problem równości

Dane: języki bezkontekstowe L, M

Wynik: czy $L = M$?

Twierdzenie (Sénizergues 1997)

Problem równości deterministycznych języków bezkontekstowych jest rozstrzygalny.

nagroda Gödla '2002 !

Jednoznaczne języki bezkontekstowe

Automat $\mathcal{A}$ jest *jednoznaczny*, jeśli $\mathcal{A}$ ma dokładnie jeden bieg akceptujący dla każdego akceptowanego słowa.

Przykład

Języki

$$\{\, w\, w^R \ :\ w \in \{a,b\}^* \,\} \qquad \{\, a^n\, b^n \ :\ n \in \mathbb{N} \,\} \cup \{\, a^n\, b^{2n} \ :\ n \in \mathbb{N} \,\}$$

są jednoznaczne, ale nie deterministyczne.

Pytanie

Czy automat deterministyczny jest jednoznaczny?

Twierdzenie

(jednoznaczne automaty ze stosem) $\equiv$ *(jednoznaczne gramatyki bezkontekstowe).*

Plan

Operacje boolowskie, konkatenacja, iteracja

Fakt

Języki bezkontekstowe są zamknięte na sumy, konkatenacje, iteracje (gwiazdkę).

Dowód:

Gramatyki.

Fakt

Języki bezkontekstowe nie są zamknięte na przecięcia.

Dowód:

Język $\{\, a^n\, b^n \;:\; n \in \mathbb{N} \,\}\, c^* \;\cap\; a^* \,\{\, b^n\, c^n \;:\; n \in \mathbb{N} \,\}$ nie jest bezkontekstowy.

Wniosek

Języki bezkontekstowe nie są zamknięte na dopełnienia.

Na przykład języki

$$\{\, w\, w \;:\; w \in \{a, b\}^* \,\} \qquad\qquad \{\, a^i\, b^j\, c^k \;:\; i, j, k \text{ parami różne} \,\}$$

nie są bezkontekstowe, a ich dopełnienia są (dlaczego?).

Fakt (języki bezkontekstowe są zamknięte na podstawienia)

$$\begin{array}{l} L \subseteq A^* \text{ bezkontekstowy} \\ h : A \to \mathcal{P}(B^*),\ \forall a \in A.\ h(a) \text{ bezkontekstowy} \end{array} \quad \Longrightarrow \quad \widehat{h}(L) \text{ bezkontekstowy.}$$

W szczególności języki bezkontekstowe są zamknięte na obrazy homomorficzne:

$$\begin{array}{l} L \subseteq A^* \text{ bezkontekstowy} \\ h : A \to B^* \text{ homomorfizm} \end{array} \quad \Longrightarrow \quad \vec{h}(L) \text{ bezkontekstowy.}$$

Fakt (języki bezkontekstowe są zamknięte na przeciwobrazy homomorficzne)

$$\begin{array}{l} L \subseteq B^* \text{ bezkontekstowy} \\ h : A \to B^* \text{ homomorfizm} \end{array} \quad \Longrightarrow \quad \vec{h}^{-1}(L) \text{ bezkontekstowy.}$$

Twierdzenie

Deterministyczne języki bezkontekstowe są zamknięte na dopełnienia.

Na przykład dopełnienia języków

$$\{\, a^n\, b^n \ :\ n \in \mathbb{N} \,\} \qquad\qquad \{\, w \,\$\, w^R \ :\ w \in A^* \,\}$$

są deterministycznymi językami bezkontekstowymi, a następujący język nie (za chwilę zobaczymy dlaczego):

$$L \;=\; \{\, a^i\, b^j\, c^k \ :\ i = j \ \vee\ j = k \ \vee\ i = k \,\}.$$

Fakt

Deterministyczne języki bezkontekstowe nie są zamknięte na przecięcia.

Dowód:

j.w.

Wniosek

Deterministyczne języki bezkontekstowe nie są zamknięte na sumy.

Zestawienie własności domknięcia

operacja	języki bezkontekstowe	deterministyczne języki bezkontekstowe
konkatenacja	$\surd$	
iteracja (gwiazdka)	$\surd$	
suma	$\surd$	
przecięcie		
dopełnienie		$\surd$
podstawienie/obraz	$\surd$	
przeciwobraz	$\surd$	$\surd$

Dopełnienie języków deterministycznych (idea dowodu)

Niech $\mathcal{A} = (A, Q, q_0, F, S, s_0, \delta)$ deterministyczny automat ze stosem.
Skonstruujemy automat deterministyczny $\mathcal{A}''$ t.że $L(\mathcal{A}'') = A^* - L(\mathcal{A})$.

Bez u.o. załóżmy, że symbol początkowy s_0 nie jest nigdy zdejmowany ze stosu.

Pytanie

Niech $\bar{\mathcal{A}}$ będzie identyczny z $\mathcal{A}$, ale stany akceptujące to $Q - F$. Czy $L(\bar{\mathcal{A}})$ to dopełnienie $L(\mathcal{A})$?

Nie możemy pozbyć się pustych przejść!

Konfigurację (q, w) nazywamy *niestabilną* jeśli $(q, w) \xrightarrow{\varepsilon}$, a w p.p. *stabilną*. Niech

$$L_s(\mathcal{A}) = \{\, w \in A^* \;:\; c_0 \xrightarrow{w} c \text{ dla jakiejś } \textit{stabilnej} \text{ konfiguracji akceptującej } c \,\}$$

Mówimy, że automat $\mathcal{A}$ *nie pętli się*, jeśli $L_s(\mathcal{A}) \cup L_s(\bar{\mathcal{A}}) = A^*$.

Idea dowodu:

$$\mathcal{A} \longmapsto \text{niepętlący się } \mathcal{A}' \longmapsto \mathcal{A}'' \qquad (L(\mathcal{A}'') = L_s(\mathcal{A}''))$$

Język bezkontekstowy z regularnym

Fakt

Jeśli L jest (deterministycznym) językiem bezkontekstowym a R jest regularny, to

$$\left.\begin{array}{l} L \cap R \\ L \cup R \\ L^{-1} R \\ R^{-1} L \\ L \otimes R \end{array}\right. \quad \text{\textit{jest (deterministycznym) językiem bezkontekstowym.}}$$

Dowód: Produkt automatu ze stosem i automatu bez.

Przykład

Gdyby język

$$L \;=\; \{\, a^i\, b^j\, c^k \;:\; i = j \;\vee\; j = k \;\vee\; i = k \,\}.$$

był deterministyczny bezkontekstowy, to język

$$(\{a, b, c\}^* - L) \cap a^*\, b^*\, c^*$$

musiałby być (deterministyczny) bezkontekstowy.

W następnym odcinku:

- W pewnym sensie,

$$(\text{języki bezkontekstowe}) \;=\; (\text{języki regularne})$$

- w kierunku maszyn Turinga: