Języki, automaty i obliczenia

Wykład 9: Własności języków bezkontekstowych

Sławomir Lasota

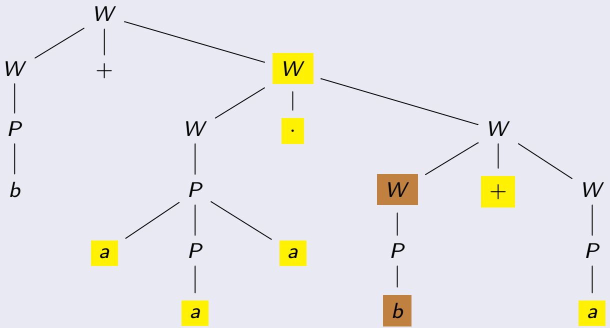
Uniwersytet Warszawski

27 kwietnia 2016

Plan

1. Pompowanie języków bezkontekstowych
2. Własności domknięcia
3. Obrazy przemienne języków bezkontekstowych są semiliniowe

Przykład



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

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

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

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

Pompowanie – dowód

Lemat o pompowaniu

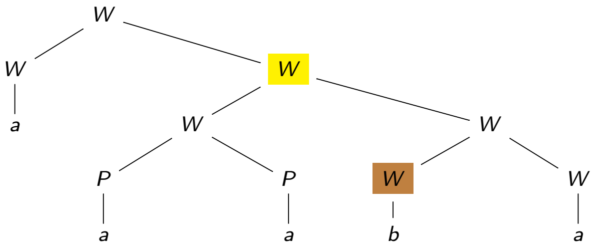
Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

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

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 Chomskiego dla $L - \{\varepsilon\}$. Niech n takie, że każde drzewo wyprowadzenia dla słowa długości $\geq n$ ma ścieżkę dłuższą niż $|N|$.



Pompowanie – przykłady pozytywne

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

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.

Pompowanie – przykład negatywny

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

Przykład

Spróbujmy pokazać, że język

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

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

$$x_n \;=\; a^{f(n)}\, b^{g(n)}\, c^{h(n)} \in L.$$

B.u.o. $h(n) > f(n), g(n)$. Wtedy niech

$$w \in a^{f(n)}\, b^{g(n)}\, c^*, \qquad y\, u\, z\, v \in c^+, \qquad |yz| \neq h(n) - f(n),\ h(n) - g(n).$$

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

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.

Pompowanie dla zaawansowanych – przykład

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

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

Plan

1 Pompowanie języków bezkontekstowych

2 Własności domknięcia

3 Obrazy przemienne języków bezkontekstowych są semiliniowe

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 $\{\, a^n\, b^n\, c^m \;:\; n, m \in \mathbb{N} \,\} \;\cap\; \{\, a^n\, b^m\, c^m \;:\; n, m \in \mathbb{N} \,\}$ 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?).

Podstawienia, (przeciw)obrazy homomorficzne

Fakt (języki bezkontekstowe są zamknięte na podstawienia)

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

W szczególności języki bezkontekstowe są zamknięte na obrazy homomorficzne:

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

Fakt (języki bezkontekstowe są zamknięte na przeciwobrazy homomorficzne)

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

Operacje boolowskie (języki deterministyczne)

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:

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

Fakt

Deterministyczne języki bezkontekstowe nie są zamknięte na przecięcia.

Dowód:

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

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		$\surd$
różnica		
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:

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

Dopełnienie języków deterministycznych (dowód, krok 1)

$$\mathcal{A} \quad \longmapsto \quad \text{niepętlący się } \mathcal{A}' \quad \longmapsto \quad \mathcal{A}''$$

Mówimy, że para $(q, s) \in Q \times S$ *pętli się*, jeśli automat $\mathcal{A}$ ma nieskończony bieg:

$$(q, s) \xrightarrow{\varepsilon} (p_1, w_1) \xrightarrow{\varepsilon} (p_2, w_2) \xrightarrow{\varepsilon} \dots$$

Stany $\{q, p_1, p_2, \dots\}$ nazwijmy osiągalnymi z (q, s).

Automat $\mathcal{A}'$ otrzymujemy przez eliminację pętlących się par (q, s):

$$\delta'(q, s, \varepsilon) \; = \; \begin{cases} (q_{\mathrm{tak}}, s) & \text{jeśli jakiś stan akceptujący jest osiągalny z } (q, s) \\ (q_{\mathrm{nie}}, s) & \text{w p.p.} \end{cases}$$

$L(\mathcal{A}') \; = \; L(\mathcal{A})$. Czy $\mathcal{A}'$ nie pętli się?

Gdyby się pętlił, to

$$(q_0, s_0) \xrightarrow{v} (p_0, w_0) \xrightarrow{\varepsilon} (p_1, w_1) \xrightarrow{\varepsilon} (p_2, w_2) \xrightarrow{\varepsilon} \dots$$

a wtedy para (p_i, w_i), gdzie w_i jest najkrótsze spośród wszystkich w_j, pętliłaby się.

Dopełnienie języków deterministycznych (dowód, krok 2)

$\mathcal{A} \quad \longmapsto \quad$ niepętlący się $\mathcal{A}' \quad \longmapsto \quad \mathcal{A}''$

- $Q'' \; = \; Q' \times \{\mathrm{t},\ \mathrm{n},\ ?\}$
- $q_0'' \; = \; (q_0', ?)$ (bez u.o. zakładamy, że $q_0 \notin F$)
- $F'' \; = \; \{ \; (q, \mathrm{n}) \; : \; q \in Q' \; \}$
- $S'' \; = \; S'$
- $s_0'' \; = \; s_0'$
- jeśli $\delta'(q, s, \varepsilon) \; = \; (p, w)$, to
  - $\delta''((q, \mathrm{t}), s, \varepsilon) \; = \; ((p, \mathrm{t}), w)$
  - $\delta''((q, ?), s, \varepsilon) \; = \; \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, \mathrm{t}), w) & \text{o ile } p \in F' \end{cases}$
- jeśli $\delta'(q, s, a) \; = \; (p, w)$, dla $a \in A$, to
  - $\delta''((q, ?), s, \varepsilon) \; = \; (\; (q, \mathrm{n}) \; , w)$
  - $\delta''((q, \mathrm{t}), s, a) \; = \; \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, \mathrm{t}), w) & \text{o ile } p \in F' \end{cases}$
  - $\delta''((q, \mathrm{n}), s, a) \; = \; \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, \mathrm{t}), w) & \text{o ile } p \in F' \end{cases}$

Język bezkontekstowy z regularnym

Fakt

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

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

Dowód:

Produkt automatu ze stosem i automatu skończonego.

Przykład

Gdyby język

$$L = \{ a^i b^j c^k : i = j \lor j = k \lor i = k \}.$$

był deterministyczny bezkontekstowy, to język

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

musiałby być bezkontekstowy.

Plan

1. Pompowanie języków bezkontekstowych
2. Własności domknięcia
3. Obrazy przemienne języków bezkontekstowych są semiliniowe

Obraz przemienny języka

- Dla $w = a_1 \dots a_n \in A^*$, zdefiniujmy funkcję (multizbiór nad A)

$$\mathcal{P}(w) : A \to \mathbb{N} \qquad\qquad \mathcal{P}(w)(a) = |\{ i \in \{1 \dots n\} : a_i = a \}|.$$

- *Obraz przemienny* języka L nad A to:

$$\mathcal{P}(L) = \{ \mathcal{P}(w) : w \in L \}.$$

- Dla ustalonego liniowego porządku na alfabecie $A = \{b_1, \dots b_k\}$,

$$b_1 < b_2 < \dots < b_k,$$

  obraz przemienny języka L można utożsamiać z podzbiorem $\mathbb{N}^k$:

$$\mathcal{P}(w)(i) = \mathcal{P}(w)(b_i).$$

Przykład

Niech $A = \{a, b\}$, $a < b$.

$$\mathcal{P}(\{ a^n b^n a : n \in \mathbb{N} \}) = \mathcal{P}((a\,b)^*\,a) = \{ (n+1, n) : n \in \mathbb{N} \}$$
$$\mathcal{P}(\text{palindromy}) = \{ (n, m) : \text{przynajmniej jedno spośród } n, m \text{ parzyste} \}$$

Zbiory semiliniowe

Dla $P \subseteq \mathbb{N}^k$, zdefiniujmy $P^\oplus = \{ p_1 + p_2 + \ldots + p_m : p_1, p_2, \ldots, p_m \in P^* \}$.

Zbiór *liniowy* $X \subseteq \mathbb{N}^k$ to dowolny zbiór postaci

$$X = b + P^\oplus = \{ b + p_1 + \ldots + p_m : p_1, \ldots, p_m \in P^* \}, \quad \text{dla } b \in \mathbb{N}^k,\ P \subseteq_{\text{fin}} \mathbb{N}^k.$$

b baza, P okresy

Przykład

Dla $b = (1, 2)$, $P = \{(2, 0), (0, 3)\}$,

$$b + P^\oplus = \{ (n, m) \in \mathbb{N}^2 : n \equiv 1 \,(\text{mod} 2),\ m \equiv 2 \,(\text{mod} 3) \}.$$

Zbiór *semiliniowy* $X \subseteq \mathbb{N}^k$ to dowolna skończona suma zbiorów liniowych:

$$b_1 + {P_1}^\oplus \quad \cup \quad b_2 + {P_2}^\oplus \quad \cup \quad \ldots \cup \quad b_n + {P_n}^\oplus$$

Pytanie

Niech $B, P \subseteq_{\text{fin}} \mathbb{N}^k$. Czy $B + P^\oplus$ jest zbiorem (semi)liniowym?

Dla $k = 1$, zbiory semiliniowe to zbiory prawie okresowe.

Obraz przemienny jest semiliniowy

Twierdzenie (Parikh 1961)

Obrazy przemienny języka bezkontekstowego jest semiliniowy.

Wniosek

Dla każdego języka bezkontekstowego L istnieje język regularny R t.że

$$\mathcal{P}(L) \;=\; \mathcal{P}(R).$$

Dowód:

$$b_1 + {P_1}^{\oplus} \quad \cup \quad \ldots \;\cup \quad b_n + {P_n}^{\oplus}$$

$$v + \{v_1, \ldots, v_m\}^{\oplus} \subseteq \mathbb{N}^k \qquad \longmapsto \qquad w + (w_1 + \ldots + w_m)^* \subseteq A^*$$

Wniosek

Dla alfabetów jednoliterowych, (języki bezkontekstowe) $=$ (języki regularne).

Obraz przemienny jest semiliniowy (dowód)

Niech $L = L(\mathcal{G})$, gdzie $\mathcal{G} = (A, N, S, \alpha)$. Pokażemy, że $\mathcal{P}(L)$ jest semiliniowy.

- Dla $M \subseteq N$, niech $L_M \subseteq L$ zawiera słowa posiadające *M-drzewo wyprowadzenia*, tz. takie, w którym pojawiają się wszystkie nieterminale z M (i żadne inne).
- Ponieważ
$$L = \bigcup_{M \subseteq N} L_M,$$
wystarczy pokazać, że $\mathcal{P}(L_M)$ jest semiliniowy, dla dowolnego $M \subseteq N$.
- Drzewo wyprowadzenia nazwijmy *płytkim*, jeśli na żadnej ścieżce żaden nieterminal nie pojawia się więcej niż $|N|$ razy.
- Pokażemy, że $\boxed{\mathcal{P}(L_M) = B_M + P_M{}^{\oplus}}$, gdzie
$$\begin{aligned} B_M &= \{\mathcal{P}(w) \ : w \text{ ma } \textit{płytkie } M\textit{-drzewo wyprowadzenia}\ \} \\ P_M &= \{\mathcal{P}(wv) \ : wXv \text{ ma } \textit{płytkie } M\textit{-drzewo wyprowadzenia} \text{ o korzeniu } X, \\ &\qquad \text{dla pewnego } X \in N\}. \end{aligned}$$

rysunek

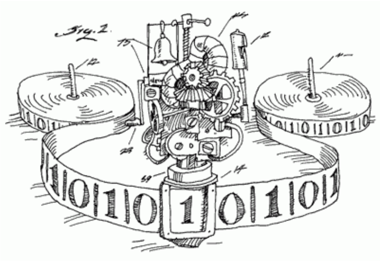


Fig. 2