JAO - lematy o “pompowaniu” dla jezykow bezkontekstowych

Gramatyka G ze zbiorem nieterminali N i zbiorem terminali T jest w postaci normalnej Chomsky'ego wtw gdy każda produkcja (reguła) gramatyki jest jednej z dwóch postaci

$A \rightarrow B\,C$, gdzie $A, B, C \in N$;

$A \rightarrow a$, gdzie $A \in N$, $a \in T$.

(Postać normalna jest wygodna w dowodach wykorzystujących drzewa wyprowadzenia, gdyż drzewa stają się prawie binarne).

Twierdzenie

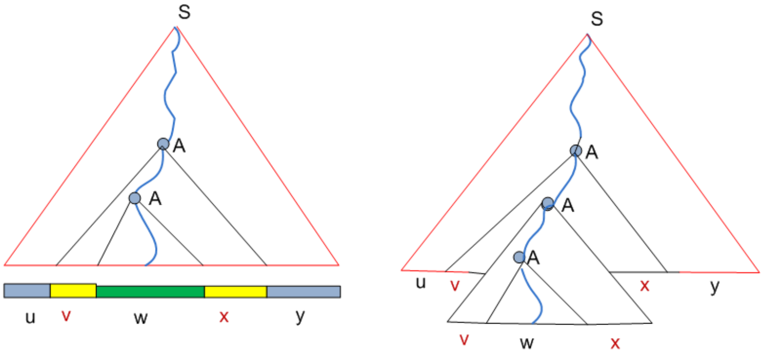
Jeśli $\epsilon \notin L(G)$ to można przekształcić G do postaci normalnej Chomsky'ego.

Transformacja polega na usunięciu symboli pustych z produkcji gramatyki oraz produkcji postaci $A \rightarrow B$. Produkcję postaci $A \rightarrow A_1 A_2 A_3 \ldots A_k$, gdzie $A_i \in N$ zastępujemy przez zbiór:

$$A \rightarrow A_1 D_1,\; D_1 \rightarrow A_2 D_2,\; D_2 \rightarrow A_3 D_3,\; \ldots D_{k-1} \rightarrow A_{k-1} A_k.$$

Lemat

Jeśli L *bezkontekstowy, to dla* $z \in L$ *dostatecznie dużych (większych od pewnej stałej* c*) istnieje rozkład* $z = uvwxy, vx \neq \epsilon$ *spełniający* $[(\forall\, i)\; uv^i wx^i y \in L]\ \&\ |vwx| \leq c$.



Z lematu o pompowaniu natychmiast wynika, że następujące języki nie są bezkontekstowe:

$$L1 \;=\; \{\, a^n b^n c^n \;:\; n \geq 1 \,\}, \quad L2 \;=\; \{\, a^i b^j a^i b^j \;:\; i,j \geq 1 \,\}$$

Nieskończony język nie zawierający słów postaci x^3, nie jest bezkontekstowy (nie ma "pompowania"). Tak więc zbiory prefiksów/podsłów słowa Thue-Morse'a i Fibonacciego nie są bezkontekstowe ponieważ słowo $\mathcal{T}$ nie zawiera x^3, a słowo $\mathcal{F}$ nie zawiera x^4. Język $L3$ słów binarnych typu xx nie jest bezkontekstowy, gdyż

$$L3 \cap a^+ b^+ a^+ b^+ \;=\; L2$$

Dla $L \;=\; \{a^i b^j a^i b^k : k \geq 1\} \cdot \{a^k b^j a^i b^j : k \geq 1\}$ mamy $\sqrt{L} \;=\; L2$. Zatem operacja $\sqrt{\cdot}$ wyprowadza poza klasę języków bezkontekstowych.

Twierdzenie

Klasa języków bezkontekstowych nie jest zamknięta na dopełnienie i przecięcie teoriomnogościowe.

- Dopełnienie $L1$ jest bezkontekstowe, a $L1$ nie jest bezkontekstowy, dowodzi to niezamkniętości na dopełnienie.
- Języki $\{a^i b^i c^j : i,j \geq 1\}$, $\{a^j b^i c^i : i,j \geq 1\}$ są bezkontekstowe, ich przecięcie, równe $L1$, nie jest językiem bezkontekstowym:

$$L1 = \{a^i b^i c^j : i,j \geq 1\}, \cap \{a^j b^i c^i : i,j \geq 1\}$$

Twierdzenie

Klasa języków bezkontekstowych jest zamknięta na operacje $\cup$, $*$, , *konkatencję, odwracanie słów, oraz branie wszystkich przesunięć cyklicznych.*

Dla języka $a^*bc \cup \{a^p ba^n ca^n \ : \ p$ liczba pierwsza, $n \geq 1\}$ zwyczajny lemat o pompowaniu "nie chwyta", wzmocnijmy go.

Lemat (Lemat "o pompowaniu kontrolowanym", uproszczony lemat Ogdena)

Istnieje taka stała c, że jeśli zaznaczymy co najmniej c pozycji w słowie $z \in L$ to jedna z części pompujących (v, x w dekompozycji $uvwxy$) zawiera wyróżnioną pozycję

Uzasadnienie. Tym razem rozważamy tylko **istotne** wierzchołki drzewa wyprowadzenia, tzn. te z których wychodzą co najmniej dwie różne krawędzie będące początkami ścieżek do liścia (pozycji w słowie terminalnym) wyróżnionego. Istnieje ścieżka na której są dwa *istotne* wierzchołki o tej samej etykiecie (nieterminalu). Teraz drzewo "pączkuje" tak jak poprzednio.

Lemat o pompowaniu kontrolowanym zastosujemy do języka

$$\{a^i\, b^j\, a^k \ :\ i \neq j \ i \neq k \ j \neq k\}$$

Niech c będzie stałą z lematu. Weźmy słowo $a^c b^{c!} a^{2c!}$.
Zaznaczamy prefiks długości c, wymuszamy by jedna z części pompujących była w pierwszej grupie liter a.
Podobnie dla języka

$$\{a^i\, b^j\, c^k \ :\ (\, i = j \text{ lub } j = k\,)\ \&\ i \neq k\,\}$$

Korzystając z Lematu Ogdena można udowodnić, że zbiór słów bez nietrywialnego okresu oraz jego dopełnienie nie są bezkontekstowe.
Natomiast język słów zawierających pewne podsłowo typu xx, $x \neq \epsilon$, dla alfabetu ternarnego, nie jest bezkontekstowy, ale nie widać jak wykorzystać tutaj lemat Ogdena.

Jednoznaczność języków bezkontekstowych

Gramatyka bezkontekstowa G jest jednoznaczna jeśli dla każdego słowa $w \in L(G)$ istnieje tylko jedno drzewo wyprowadzenia w tej gramatyce.
Na przykład gramatyka dla D_1:

$$S \rightarrow (\, S\,) \mid S\, S \mid (\,)$$

jest niejednoznaczna, natomiast jest jednoznazna gramatyka:

$$S \rightarrow (\, S\,) \mid (\, S\,)\, S \mid (\,)$$

Język bezkontekstowy nazywamy jednoznacznym jeśli jest generowany przez pewną gramatykę jednoznaczną.

Tak więc język D_1 jest językiem jednoznacznym (jak również języki nawiasowe D_k z k typami nawiasów).

Twierdzenie

Język $\{ a^i b^i c^j \ : \ i,j \geq 1 \} \cup \{ a^i b^j c^j \ : \ i,j \geq 1 \}$ *jest niejednoznaczny.*

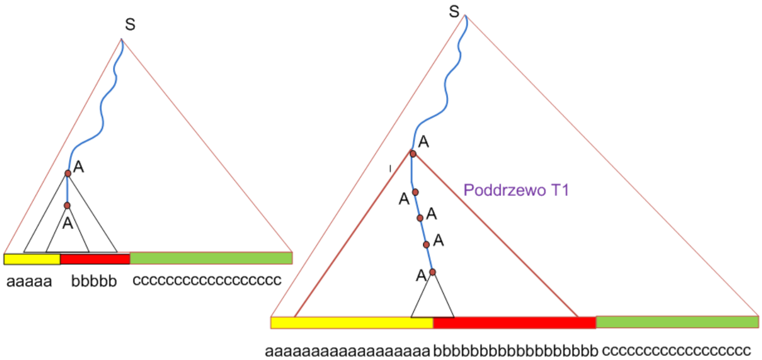
Stosujemy takie rozumowanie jak w dowodzie lematu o pompowaniu kontrolowanym. Przypuśćmy, że gramatyka jednoznaczna G generuje nasz język, niech p będzie stałą z lematu. Weźmy słowo $a^p b^p c^{p+p!}$ i wyróżnijmy wszystkie a. Wtedy pierwsza częśc pompująca musi bć w bloku a (na rysunku żółta), a druga w bloku symboli b (czerwony na rysunku). Po *podpompowaniu* otrzymujemy wyyprowadzenie słowa $a^{p+p!} b^{p+p!} c^{p+p!}$ w którym pewne poddrzewo $T1$ zaczyna się w bloku liter a i w bloku liter b, nie dalej niż p symboli od prawego końca tego bloku (kolor czerwony).

Dowod twierdzenia o niejednoznacznosci

Stosując symetryczne rozumowanie dla słowa $a^{p+p!}b^p c^p$, w którym wyróżniamy ostatni blok, otrzymamy drzewo wyprowadzenia słowa

$$a^{p+p!}b^{p+p!}c^{p+p!}$$

w którym pewne poddrzewo $T2$ zaczyna się w bloku liter b, nie dalej niż p pozycji od lewego końca tego bloku, a kończy w bloku liter c.



Niejednoznaczność

Drzewa $T1$, $T2$ nie mogą być poddrzewami tego samego drzewa, ponieważ się *przecinają* (patrz rysunek). Zatem mamy więcej niż jedno drzewo wyprowadzenia (niejednoznaczność).

Twierdzenie Parikha dla języków bezkontekstowych

Przypomnienie: Ψ jest funkcją Parikha. Następujące twierdzenie jest również konsekwencją własności *pompujących* dla języków bezkontekstowych.

Twierdzenie

Jeśli L jest językiem bezkontekstowym to istnieje język R regularny taki, że $\Psi(L) = \Psi(R)$.

(Szkic dowodu)

Załóżmy, że gramatyka bezk. G jest w postaci normalnej Chomsky'ego i ma k nieterminali.

Niech N będzie taką stałą, że w każdym drzewie wyprowadzenia słowa długości co najmniej N istnieje ścieżka na której pewien nieterminal pojawia się co najmniej $k+2$ razy.

Szkic dowodu twierdzenia Parikha

Niech Q będzie podzbioremz zbioru nieterminali i niech L_Q będzie zbiorem słów mających drzewo wyprowadzenia zawierające dokładnie wszystkie nieterminale z Q i żadnych innych.

Niech F_Q będzie zbiorem słów z L_Q o długości co najwyżej N, a T_Q będzie zbiorem słow xy takich, że $|xy| \leq N$ oraz istnieje wyprowadzenie $A \to^* xAy$ w którym występują jedynie nieterminale z Q. Wtedy

$$\Psi(L_q) \;=\; \Psi(F_Q\ T_Q^*)$$

Zatem $L_Q \;=\; \bigcup_Q\ L_Q$ jest w sensie Parikha równoważny językowi który jest regularny, gdyż zbiory F_Q i T_q są skończone.
Mamy

$$\Psi(L) \;=\; \bigcup_Q \Psi(F_Q\ T_Q^*)$$

Lemat

Jeśli L jest liniowym językiem bezkontekstowym, to dla $z \in L$ dostatecznie dużych (większych od pewnej stałej c) istnieje rozkład $z = uvwxy, vx \neq \epsilon$ spełniający

$$[\,(\forall\, i\,)\ uv^i wx^i y \in L\,]\ \&\ |u\,v\,x\,y| \leq c$$

Części pompujące są blisko brzegów słowa. Wynika stąd

Fakt

Istnieją języki bezkontekstowe, które nie są liniowe, na przykład język

$$\{\, a^i b^i a^j b^j \ :\ i,j \geq 1\,\}$$