Teoria Automatów i Języków Formalnych

Ćwiczenia 5: Języki bezkontekstowe

dr inż. Marcin Luckner
mluckner@mini.pw.edu.pl

Wydział Matematyki i Nauk Informacyjnych

Wersja 1.3
3 marca 2021

Fundusze Europejskie Wiedza Edukacja Rozwój
Rzeczpospolita Polska
Politechnika Warszawska
Unia Europejska Europejski Fundusz Społeczny

Projekt „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca" współfinansowany jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Zadanie 10 pn. „Modyfikacja programów studiów na kierunkach prowadzonych przez Wydział Matematyki i Nauk Informacyjnych", realizowane w ramach projektu „NERW 2 PW. Nauka – Edukacja – Rozwój – Współpraca", współfinansowanego jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Fundusze Europejskie Wiedza Edukacja Rozwój
Rzeczpospolita Polska
Politechnika Warszawska
Unia Europejska Europejski Fundusz Społeczny

Język bezkontekstowy

- Języki bezkontekstowe są generowane przez gramatyki bezkontekstowe.
- Jedynym sposobem, aby wykazać, że język jest bezkontekstowy jest stworzenie jego gramatyki.
- Można wykazać, że język nie jest bezkontekstowy stosując odmianę kontrapozycji lematu o pompowaniu.
- Algorytm Cocke–Younger–Kasami (CYK) pozwala wykazać, że dane słowo należy do języka.

Kontrapozycja lematu o pompowaniu

Kontrapozycja lematu o pompowaniu

Jeżeli dla każdej stałej n_L istnieje słowo $z \in L$ takie, że

$$\begin{aligned}(|z| \geqslant n_L) \wedge \\ [(\forall_{u,v,w,x,y} z = uvwxy \wedge |vwx| \leqslant n_L \wedge |vx| \geqslant 1) \\ \exists_{i=0,1,2\ldots} z_i = uv^i wx^i y \notin L]\end{aligned}$$

to język nie jest bezkontekstowy.

Lemat dla języków regularnych, a dla bezkontekstowych

- Lematy dla języków regularnych i języków bezkontekstowych są podobne, ale różnią się kilkoma istotnymi szczegółami.

Język regularny

- $z = uvw$
- $|uv| \leqslant n_L$
- $|v| \geqslant 1$
- $z_i = uv^i w$

Język bezkontekstowy

- $z = uvwxy$
- $|vwx| \leqslant n_L$
- $|vx| \geqslant 1$
- $z_i = uv^i wx^i y$

Dyskusja lematu dla języków bezkontekstowych

- Dodanie w słowie $z = uvwxy$ części u powoduje, że okno, w którym możemy dokonać modyfikacji może wędrować po całym słowie.
- Okno ma szerokość n_L więc możemy ograniczyć jego zasięg dobierając odpowiednie słowo z.
- Pompowanie dwóch elementów na raz powoduje, że może być trudniej udowodnić dla każdego podziału, że słowo $z_i = uv^i wx^i y$ nie należy do języka.

Lematu o pompowaniu - przykład I

Korzystając z kontrapozycji lematu o pompowaniu wykaż, że język $L = \{a^i b^j : j = i^2\}$ nad alfabetem $\Sigma = \{a, b\}$ nie jest bezkontekstowy.

Lematu o pompowaniu - przykład II

$\forall n \in N$ słowo

$$z = a^n b^{n*n}$$

należy do języka L.

Ponieważ $|vwx| \leqslant n$ to musimy rozpatrzyć następujące przypadki:

1. $vwx = a^k \implies v = a^p \wedge x = a^q$
2. $vwx = b^k \implies v = b^p \wedge x = b^q$
3. $$\begin{aligned} vwx = a^k b^j \implies & (v = a^p \wedge x = a^q b^j) \vee \\ & (v = a^k b^p \wedge x = b^q) \vee \\ & (v = a^p \wedge x = b^q) \end{aligned}$$

gdzie $k, j \in \{1, \ldots, n\} \wedge p, q \in \{0, \ldots, n\}$.

Lematu o pompowaniu - przykład III

- W przypadku 1 ($v = a^p \wedge x = a^q$) wybierzmy $i = 0$.
  - Wtedy słowo $z_0 = a^{n-(p+q)} b^{n*n} \notin L$
- W przypadku 2 ($v = b^p \wedge x = b^q$) wybierzmy $i = 0$.
  - Wtedy słowo $z_0 = a^n b^{n*n-(p+q)} \notin L$

Lematu o pompowaniu - przykład IV

- W przypadku 3 są trzy dalsze przypadki, w tym dwa podobne.
  - W przypadku pierwszym ($v = a^p \wedge x = a^q b^j$) wybierzmy $i = 2$
    - Po zwielokrotnieniu $z_2 = a^{n-k}(a^p)(a^p)a^{k-p-q}(a^q b^j)(a^q b^j)b^{n*n-j}$
    - Czyli $z_2 = a^{n+p} b^j a^q b^{n*n} \notin L$
  - W przypadku drugim ($v = a^k b^p \wedge x = b^q$) wybierzmy $i = 2$
    - Po zwielokrotnieniu $z_2 = a^{n-k}(a^k b^p)(a^k b^p)b^{j-p-q}(b^q)(b^q)b^{n*n-j}$
    - Czyli $z_2 = a^n b^p a^k b^{n*n+q} \notin L$

Lematu o pompowaniu - przykład IV

- W ostatnim przypadku ($v = a^p \wedge x = b^q$) wybierzmy $i = 2$.
- Wtedy $z_2 = a^{n+p} b^{n*n+q}$.
- Jednakże z_2 może należeć do języka jeżeli $(n+p)^2 = n^2 + q$.
- Załóżmy, że $p = 1$. W takim wypadku liczba b wynosi $(n+1)^2 = n^2 + 2n + 1$.
- To jednak oznacza, że $q > n$, a wiemy, że $|vwx| \leqslant n$, więc jest to niemożliwe.
- Dla $p > 1$ wielkość q wzrasta jeszcze bardziej więc $z_2 = a^{n+p} b^{n*n+q} \notin L$
- Zatem język nie jest bezkontekstowy.

Algorytm Cocke–Younger–Kasami

- Algorytm Cocke–Younger–Kasami (CYK) pozwala określić czy słowo w długości n jest generowane przez gramatykę G.
- Gramatyka wejściowa musi być w postaci normalnej Chomskiego.
- Algorytm opiera się na programowaniu dynamicznym.

Algorytm CYK

1. Dzielimy słowo w na n jednoliterowych łańcuchów i odnajdujemy generujące je zmienne. W tym celu przeglądamy produkcje typu $A \to a$, gdzie $a \in T$.
2. Znając nieterminale generujące dla słowa w podciągi długości k znajdujemy nieterminale generujące podciągi długości $k+1$:
   - 2.1 Dzielimy łańcuch o długości $k+1$ na wszystkie możliwe pary podłańcuchów.
   - 2.2 Każdy prefiks i odpowiadający mu sufiks mają już ustalone generujące je nieterminale.
   - 2.3 Znajdujemy zbiór nieterminali A posiadających produkcje $A \to BC$, gdzie $B, C \in V$ generują prefiks i sufiks.
   - 2.4 Nieterminale A z punktu 2.3 generują podciąg o długości $k+1$.
3. Na koniec, otrzymujemy zestaw nieterminali produkujących ciąg długości n tworzący słowo w. Słowo w jest generowane przez gramatykę $\iff$ jest wśród nich symbol startowy.

Algorytm CYK - przykład I

Użyj algorytmu CYK, aby sprawdzić czy słowo *anna* należy do języka generowanego przez gramatykę
$G = \{V = \{A, B, C, D, E, F, N, S\}, T = \{a, n, w\}, P, S\}$

P:

$$
\begin{aligned}
S &\rightarrow NN|NA \\
N &\rightarrow BB|BD|w \\
B &\rightarrow n \\
D &\rightarrow NB \\
A &\rightarrow CC|EC|CF \\
C &\rightarrow a \\
E &\rightarrow CA \\
F &\rightarrow NC
\end{aligned}
$$

Algorytm CYK - przykład II

P:

$$
\begin{aligned}
S &\rightarrow NN|NA \\
N &\rightarrow BB|BD|w \\
B &\rightarrow n \\
D &\rightarrow NB \\
A &\rightarrow CC|EC|CF \\
C &\rightarrow a \\
E &\rightarrow CA \\
F &\rightarrow NC
\end{aligned}
$$

- Znajdź zmienne generujące terminale

C	B	B	C
a	n	n	a

Algorytm CYK - przykład III

P:

$$S \longrightarrow NN|NA$$
$$N \longrightarrow BB|BD|w$$
$$B \longrightarrow n$$
$$D \longrightarrow NB$$
$$A \longrightarrow CC|EC|CF$$
$$C \longrightarrow a$$
$$E \longrightarrow CA$$
$$F \longrightarrow NC$$

- Znajdź zmienne generujące podciągi dwuliterowe

$\emptyset$	N_{BB}	$\emptyset$	
C	B	B	C
a	n	n	a

Algorytm CYK - przykład IV

P:

$$S \rightarrow NN|NA$$
$$N \rightarrow BB|BD|w$$
$$B \rightarrow n$$
$$D \rightarrow NB$$
$$A \rightarrow CC|EC|CF$$
$$C \rightarrow a$$
$$E \rightarrow CA$$
$$F \rightarrow NC$$

- Znajdź zmienne generujące podciągi trzyliterowe

$\emptyset$	F_{NC}		
$\emptyset$	N_{BB}	$\emptyset$	
C	B	B	C
a	n	n	a

Algorytm CYK - przykład V

P:

$$S \rightarrow NN|NA$$
$$N \rightarrow BB|BD|w$$
$$B \rightarrow n$$
$$D \rightarrow NB$$
$$A \rightarrow CC|EC|CF$$
$$C \rightarrow a$$
$$E \rightarrow CA$$
$$F \rightarrow NC$$

- Znajdź zmienne generujące słowo *anna*

A_{CF}			
$\emptyset$	F_{NC}		
$\emptyset$	N_{BB}	$\emptyset$	
C	B	B	C
a	n	n	a

- Zmienna A generuje słowo $w = anna$, ale $w \notin L$ gdyż $S \neq A$.

Algorytm CYK - przykład VI

- Czy słowo *wanna* należy do języka?

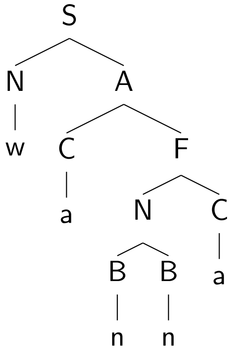
P:

$$S \rightarrow NN|NA$$
$$N \rightarrow BB|BD|w$$
$$B \rightarrow n$$
$$D \rightarrow NB$$
$$A \rightarrow CC|EC|CF$$
$$C \rightarrow a$$
$$E \rightarrow CA$$
$$F \rightarrow NC$$

S_{NA}				
$\emptyset$	A_{CF}			
$\emptyset$	$\emptyset$	F_{NC}		
F_{NC}	$\emptyset$	N_{BB}	$\emptyset$	
N	C	B	B	C
w	a	n	n	a

- Słowo $w = $ *wanna* należy do języka, gdyż da się je wyprowadzić z S

Drzewa wyprowadzenia



- Korzystając z indeksowania nieterminali w taki sposób, aby wskazywały one na użytą produkcję, możemy odtworzyć sposób wyprowadzenia słowa.
- Takie wyprowadzenie nazywamy drzewem wyprowadzenia.
- Słowo może mieć więcej niż jedno drzewo wyprowadzenia.

Zadania I

1. Użyj algorytmu CYK, aby sprawdzić czy
   1.1 *baaba*, *aaaaa*, *aaaaaa* należą do następującej gramatyki
   $G = \{V = \{A, B, C, S\}, T = \{a, b\}, P, S\}$

   $$\begin{aligned} P:\quad S &\rightarrow AB|BC \\ A &\rightarrow BA|a \\ B &\rightarrow CC|b \\ C &\rightarrow AB|a \end{aligned}$$

   1.2 (()(())) należy do następującej gramatyki
   $G = \{V = \{S\}, T = \{C,)\}, P, S\}$

   $$P:\ S \rightarrow SS|(S)|\epsilon$$

Zadania II

2. Korzystając z kontrapozycji lematu o pompowaniu wykaż, że następujące języki nie są bezkontekstowe.
   2.1 Język L nad alfabetem $\Sigma = \{a, b, c\}$, taki, że liczba wystąpień symboli a, b, c jest równa. $abbacc \in L$, $accaabb \notin L$
   2.2 Język $L = \{a^i b^j c^k : i \leqslant j \wedge j \leqslant k\}$ nad alfabetem $\Sigma = \{a, b, c\}$
   2.3 Język $L = \{a^i b^j c^k : i \neq j \wedge j \neq k\}$ nad alfabetem $\Sigma = \{a, b, c\}$