Języki, automaty i obliczenia

Wykład 3: Języki regularne a automaty

Sławomir Lasota

Uniwersytet Warszawski

13 marca 2024

Plan

Automaty z pustymi przejściami

- $\delta \subseteq Q \times (A \cup \{\varepsilon\}) \times Q$
- Puste przejścia (ε-przejścia)

$$\delta(q, \varepsilon, q') \qquad\qquad q \xrightarrow{\varepsilon} q'$$

- Rozszerzamy puste przejścia:

$$\widehat{\delta}(q, \varepsilon, q') \qquad\qquad q \xrightarrow{\varepsilon} \xrightarrow{\varepsilon} \dots \xrightarrow{\varepsilon} q'$$

- Usuwanie pustych przejść:

$$(q, a, q') \in \rho \iff \exists \bar{q}, \bar{q}'. \begin{cases} \widehat{\delta}(q, \varepsilon, \bar{q}) \wedge \\ \delta(\bar{q}, a, \bar{q}') \wedge \\ \widehat{\delta}(\bar{q}', \varepsilon, q') \end{cases}$$

$$\begin{tikzcd} q \arrow[r, "\varepsilon"] \arrow[rrrrrr, bend right, dotted, "a"'] & \dots \arrow[r, "\varepsilon"] & \bar{q} \arrow[r, "a"] & \bar{q}' \arrow[r, "\varepsilon"] & \dots \arrow[r, "\varepsilon"] & q' \end{tikzcd}$$

Wyrażenie regularne $\mapsto$ automat

Lemat

Dla każdego wyrażenia regularnego istnieje równoważny automat.

(Każdy język regularny jest rozpoznawany przez automat.)

Dowód:

Konstruujemy indukcyjnie automat $\mathcal{A}_L$ z pustymi przejściami t. że $L(\mathcal{A}_L) = L$.

- początek:

$$\mathcal{A}_a \ (a \in A) \qquad\qquad \mathcal{A}_\emptyset$$

- krok indukcyjny:
  - $\mathcal{A}_L, \mathcal{A}_M \mapsto \mathcal{A}_{LM}$
  - $\mathcal{A}_L, \mathcal{A}_M \mapsto \mathcal{A}_{L+M}$
  - $\mathcal{A}_L \mapsto \mathcal{A}_{L^*}$

Automat $\mapsto$ wyrażenie regularne

Lemat

Dla każdego automatu istnieje równoważne wyrażenie regularne.

(Każdy język ropoznawany przez automat jest regularny.)

Dowód:

Bez utraty ogólności załóżmy, że automat ma dokładnie jeden stan początkowy i jeden stan akceptujący:

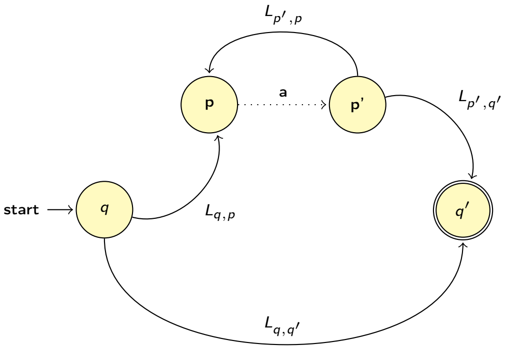
$$L(\mathcal{A}) \;=\; \bigcup_{q \in I, q' \in F} L_{q,q'}(\mathcal{A})$$

Indukcja ze względu na liczbę przejść $|\delta|$ automatu.

- początek: brak przejść

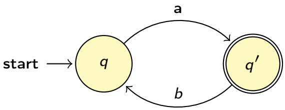
$$L \;=\; \emptyset \qquad \text{albo} \qquad L \;=\; \varepsilon$$

- krok indukcyjny: ...

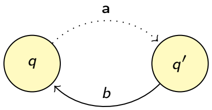


$$L \quad = \quad L_{q,q'} \quad + \quad L_{q,p}\ a\ (L_{p',p}\ a)^*\ L_{p',q'}$$

Przykład



$$L \;=\; ?$$

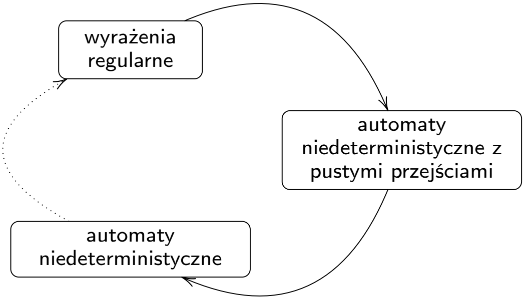


$$L \;=\; L_{q,q'} + L_{q,q}\, a (L_{q',q}\, a)^* L_{q',q'} \;=\; \emptyset + \varepsilon\, a (b\, a)^* \varepsilon \;=\; a(b\, a)^*$$

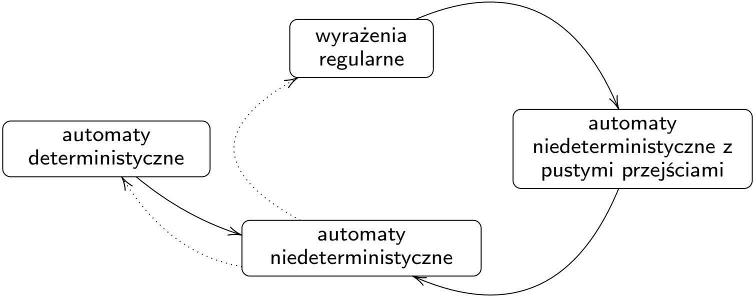
Równoważność

Twierdzenie (Kleene 1956)

Automaty rozpoznają dokładnie języki regularne.



Wykładniczy wzrost rozmiaru *



Wykładniczy wzrost rozmiaru *

Automat deterministyczny $\mathcal{A}_n$:

- $A \;=\; \{\,(j,k) \in \{1\ldots n\}^2 \;:\; j \neq k\,\}$
- $Q \;=\; \{1\ldots n\} \cup \{\text{śmietnik}\}$
- $I \;=\; \{1\}$
- $F \;=\; \{n\}$
- $\delta(i,(j,k)) \;=\; \begin{cases} k & \text{jeśli } i = j \\ \text{śmietnik} & \text{wpp.} \end{cases}$

  $\delta(\text{śmietnik},(j,k)) = \text{śmietnik}$

Wyrażenie regularne równoważne automatowi $\mathcal{A}_n$ ma rozmiar $\Omega(2^{n-1})$.

L_n zdefiniowany przez wyrażenie regularne: $(a+b)^* a\,(a+b)^{n-2}$.

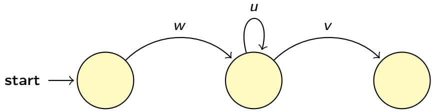
Automat deterministyczny rozpoznający L_n ma $\Omega(2^{n-1})$ stanów.

Plan

Pompowanie

Obserwacja

Bieg automatu o n stanach, o długości $\geq n$, odwiedza dwukrotnie jakiś stan.



Obserwacja

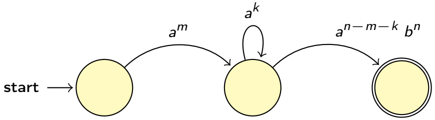
Jeśli automat o n stanach akceptuje słowo o długości $\geq n$, to słowo to można przedstawić jako

$$w\,u\,v, \qquad \text{gdzie } |u| > 0 \text{ i } |w\,u| \leq n$$

i automat akceptuje słowo $w\,u^m\,v$, dla każdego $m \in \mathbb{N}$.

Pytanie

$A = \{a, b\}$. Czy język $L = \{\, a^n\, b^n \;:\; n \in \mathbb{N} \,\}$ jest regularny?



Odpowiedź

Nie: Dowód nie-wprost.

Załóżmy, że L jest rozpoznawany przez automat $\mathcal{A}$ o n stanach. $\mathcal{A}$ akceptuje $a^n\, b^n \in L$, więc $a^{n+k}\, b^n \in L$ dla pewnego $k > 0$. Sprzeczność.

Lemat o pompowaniu

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

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

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

$$L \text{ regularny} \quad \Longrightarrow \quad \exists n.\ \forall x.\ \exists x{=}w\,u\,v.\ \forall m.\ \ w\,u^m\,v \in L$$

Wniosek

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

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

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

$$L \text{ nieregularny} \quad \Longleftarrow \quad \forall n.\ \exists x.\ \forall x{=}w\,u\,v.\ \exists m.\ \ w\,u^m\,v \notin L$$

Pompowanie – przykład

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

Pytanie

$A = \{a, b\}$. Czy język $L = \{\, a^m\, b^n \;:\; m \leq n \,\}$ jest regularny?

Odpowiedź

Nie. Dla *dowolnego* $n \in \mathbb{N}$, rozważmy słowo $x_n = a^n\, b^n \in L$.
Rozważmy *dowolne* słowa w, u, v takie, że

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

Czyli $w \in a^*$, $u \in a^+$ i $v \in a^* b^n$. Wtedy $w\,u^2\,v \notin L$. Zatem L nieregularny.

Pompowanie – przykład 2

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

Pytanie

Czy język palindromów $L \;=\; \{\, w \in \{a, b\}^* \;:\; w \;=\; w^{\mathrm{R}}\,\}$ jest regularny?

Odpowiedź

Nie. Dla *dowolnego* $n \in \mathbb{N}$, rozważmy słowo $x_n \;=\; a^n\,b\,a^n \in L$.
Rozważmy *dowolne* słowa w, u, v takie, że

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

Czyli $w \in a^*$, $u \in a^+$ i $v \in a^*\,b\,a^n$. Wtedy $w\,v \notin L$. Zatem L nieregularny.

Pytanie

Czy język $L = \{\, (c^* \, a)^n \, c^* \, (b \, c^*)^n \;:\; n \in \mathbb{N} \,\}$ jest regularny?

Odpowiedź 1

Nie. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo $x_n = a^n \, b^n \in L \ldots$

Odpowiedź 2

Nie. Dowód nie-wprost.

Przypuśćmy, że L jest regularny. Rozważmy homomorfizm $h : \{a, b, c\}^* \to \{a, b\}^*$ wyznaczony przez

$$a \mapsto a, \;\; b \mapsto b, \;\; c \mapsto \varepsilon.$$

Skoro L jest regularny, to $\vec{h}(L) = \{\, a^n \, b^n \;:\; n \in \mathbb{N} \,\}$ też. Sprzeczność.

Pytanie

Niech L = (zbiór wyrażeń regularnych nad A). Czy L jest językiem regularnym?

Alfabet to $A' = A \cup \{+, ^*, \emptyset, (,)\}$.

Odpowiedź

Nie. Dowód nie-wprost.

Przypuśćmy, że L jest regularny. Rozważmy homomorfizm

$$h : (A')^* \to \{(,)\}^*,$$

który zachowuje symbole (i) a „wymazuje” pozostałe. Zatem $\vec{h}(L)$ to poprawnie zbudowane wyrażenia nawiasowe. Skoro L jest regularny, to $\vec{h}(L)$ też.

Dla dowolnego $n \in \mathbb{N}$, rozważmy słowo $x_n = \underbrace{(\ldots(}_{n}\ \underbrace{)\ldots)}_{n} \in L \ldots$

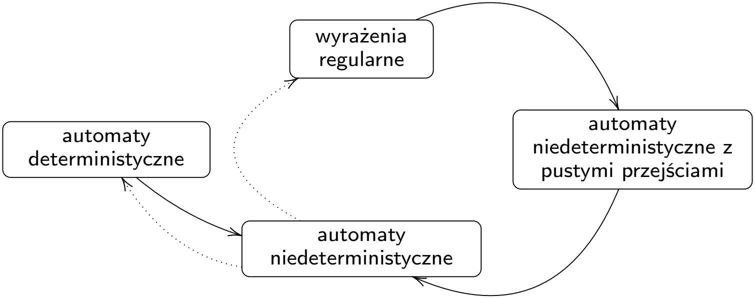
Plan

Problem przynależności

Problem decyzyjny

Dane: język regularny L i słowo w

Wynik: czy $w \in L$?



Problem przynależności

Problem decyzyjny

Dane: automat niedeterministyczny $\mathcal{A} = (A, Q, I, F, \delta)$ i słowo $w = a_1 \ldots a_n$

Wynik: czy $w \in L(\mathcal{A})$?

Algorytm

$X \;:=\; I$

`powtarzaj dla i = 1...n`

$\quad X \;:=\; \{\, q' \in Q \;:\; \exists q \in X.\; q \xrightarrow{a_i}_{\mathcal{A}} q' \,\}$

`wynik` $:= \; (X \cap F \;\neq\; \emptyset)$

Problem niepustości

Problem decyzyjny

Dane: automat niedeterministyczny $\mathcal{A} = (A, Q, I, F, \delta)$

Wynik: czy $L(\mathcal{A}) \neq \emptyset$?

Algorytm

$X := I$

`powtarzaj`

$\quad X' := X$

$\quad X := X \cup \{ q' \in Q : \exists q \in X, a \in A.\ q \xrightarrow{a}_{\mathcal{A}} q' \}$

`aż` $X = X'$

`wynik` $:= (X \cap F \neq \emptyset)$

Problem nieskończoności

Problem decyzyjny

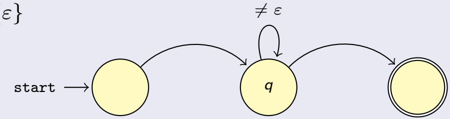
Dane: automat niedeterministyczny $\mathcal{A} = (A, Q, I, F, \delta)$

Wynik: czy $L(\mathcal{A})$ nieskończony?

Algorytm

`sprawdź, czy istnieje stan` q `taki, że`

- $L_{I,q}(\mathcal{A}) \neq \emptyset$
- $L_{q,F}(\mathcal{A}) \neq \emptyset$
- $L_{q,q}(\mathcal{A}) \neq \{\varepsilon\}$



Problem równości

Problem decyzyjny

Dane: automaty niedeterministyczne $\mathcal{A}, \mathcal{B}$

Wynik: czy $L(\mathcal{A}) \subseteq L(\mathcal{B})$?

Algorytm

- oblicz automat $\bar{\mathcal{B}}$ dla dopełnienia języka $L(\mathcal{B})$
- sprawdź, czy $L(\mathcal{A}) \cap L(\bar{\mathcal{B}}) \neq \emptyset$

Problem decyzyjny

Dane: automaty niedeterministyczne $\mathcal{A}, \mathcal{B}$

Wynik: czy $L(\mathcal{A}) = L(\mathcal{B})$?

W następnym odcinku: minimalizacja automatu