Automaty a gramatiky

3

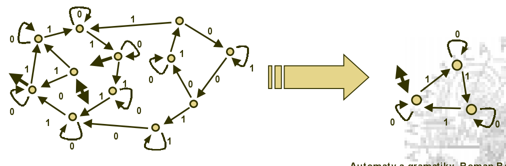
Roman Barták, KTIML

bartak@ktiml.mff.cuni.cz
http://ktiml.mff.cuni.cz/~bartak

Co bylo minule

Hledání jednoznačného minimálního automatu

- dosažitelné stavy
- ekvivalentní stavy
- automatová kongruence a podílový automat
- redukce automatů
  - isomorfismus reduktů (minimální automat)
- normovaný tvar (hledání isomorfismu)

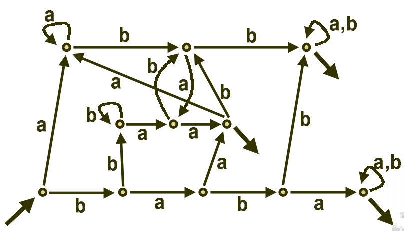


Trochu motivace

Dosud: Stav a písmeno jednoznačně určuje další stav!

Příklad:

L={ w | w=babau ∨ w=uabbv ∨ w=ubaa, u,v∈{a,b}* }

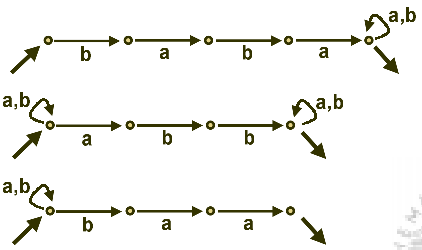


Úvod do nedeterminismu

Stav a písmeno určuje množinu možných dalších stavů!

Příklad:

L={ w | w=babau ∨ w=uabbv ∨ w=ubaa, u,v∈{a,b}* }



Nedeterministický konečný automat

Nedeterministickým konečným automatem nazýváme pětici A = (Q,X,δ,S,F), kde:

Q - konečná neprázdná množina stavů (*stavový prostor*)

X - konečná neprázdná množina symbolů (*vstupní abeceda*)

δ - zobrazení $Q \times X \rightarrow P(Q)$ (*přechodová funkce*)

$S \subseteq Q$ (*množina počátečních stavů*)

$F \subseteq Q$ (*množina koncových stavů*)

Reprezentace:

stavový diagram, tabulka, stavový strom

Jak se počítá s nedeterminismem?

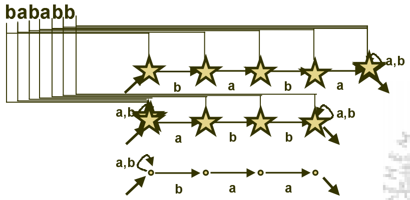
Slovo $w=x_1...x_n$ je ***přijímáno*** nedeterministickým konečným automatem, jestliže existuje posloupnost stavů $q_1,\dots,q_{n+1}$ taková, že:

$q_1 \in S$, $q_{i+1} \in \delta(q_i, x_i)$ pro $i=1\dots n$, $q_{n+1} \in F$.

Prázdné slovo je přijímáno právě když $S \cap F = \emptyset$

Přijímajících výpočtů pro dané slovo může být více!

Př. bababb



Determinismus vs. nedeterminismus

Konečný automat je speciálním případem nedeterministického konečného automatu!

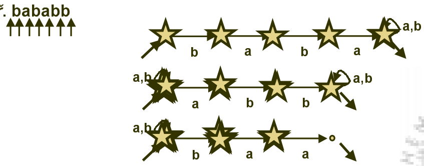
Důsledek: Jazyky rozpoznávané konečnými automaty jsou rozpoznávané nedeterministickými konečnými automaty.

Platí i obrácené tvrzení?

Zkusme to!

- potřebujeme postupovat systematicky a s konečnou pamětí
- pomocí značek na stavech simulujeme všechny možné výpočty tzv. podmnožinová konstrukce

Př. bababb



Převod nedeterminismu na determinismus

Věta: **Je-li A nedeterministický konečný automat, potom lze sestrojit konečný automat B takový, že L(A)=L(B).**

Důkaz: **(podmnožinová konstrukce)**

nechť $A = (Q,X,\delta,S,F)$

potom definujme $B=(P(Q),X,\delta',S,F')$, kde

$F' = \{ K \mid K\in P(Q), K\cap F\neq\emptyset \}$

$\delta'(K,x) = \cup_{q\in K}\, \delta(q,x)$

1) B je definován korektně

2) L(A)=L(B)?

$\lambda\in L(A) \Leftrightarrow S\cap F\neq\emptyset \Leftrightarrow S\in F' \Leftrightarrow \lambda\in L(B)$

$L(A)\subseteq L(B)$

$w= x_1\ldots x_n\in L(A) \Leftrightarrow \exists q_1,\ldots,q_{n+1}\in Q\ \ q_1\in S,\ q_{i+1}\in\delta(q_i, x_i),\ q_{n+1}\in F$

položme $K_1=S\ \ (q_1\in K_1),\ K_{i+1}= \delta'(K_i,x_i)\ \ (q_{i+1}\in K_{i+1})$, potom $K_{n+1}\in F'$

$L(B)\subseteq L(A)$

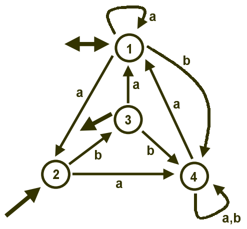
$w= x_1\ldots x_n\in L(B) \Leftrightarrow \exists K_1,\ldots,K_{n+1}\in P(Q)\ \ K_1=S,\ K_{i+1}=\delta'(K_i, x_i),\ K_{n+1}\in F'$

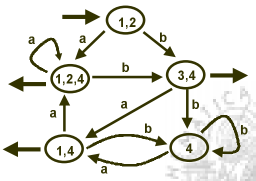
vezměme $q_{n+1}\in F\cap K_{n+1}$, $q_i\in K_i$ tž. $q_{i+1}\in\delta(q_i,x_i),\ldots, q_1\in K_1=S$

Ukázka převodu

		a	b
↔	1	1,2	4
→	2	4	3
←	3	1	4
	4	1,4	4

		a	b
↔	{1,2}	{1,2,4}	{3,4}
←	{1,2,4}	{1,2,4}	{3,4}
←	{3,4}	{1,4}	{4}
←	{1,4}	{1,2,4}	{4}
	{4}	{1,4}	{4}





Poznámky k nedeterminismu

Význam:

- teoretický (např. při převodu gramatik na automaty)
- praktický (zjednodušení návrhu automatu)

U konečných automatů vede nedeterminismus ke stejné třídě jazyků jako determinismus.

- neplatí obecně (zásobníkové automaty)!

Převod na determinismus znamená (až) exponenciální nárůst počtu stavů ($Q\rightarrow P(Q)$).

- obecně je tento nárůst nezbytný!
  - $L_n = \{ w \mid w\in\{0,1\}^*, w=u1v, |v|=n-1 \}$
- není potřeba explicitně převádět.

Existují také *zobecněné nedeterministické automaty*

λ-přechod: změna stavu bez čtení vstupu

Množinové operace nad jazyky

Sjednocení jazyků $L_1 \cup L_2 = \{ w \mid w\in L_1 \vee w\in L_2 \}$

Příklad: jazyk obsahuje slova začínající baba nebo končící baa

Průnik jazyků $L_1 \cap L_2 = \{ w \mid w\in L_1\ \&\ w\in L_2 \}$

Příklad: jazyk obsahuje slova se sudým počtem nul a každý symbol 1 je bezprostředně následován 0

Rozdíl jazyků $L_1 - L_2 = \{ w \mid w\in L_1\ \&\ w\notin L_2 \}$

Příklad: jazyk obsahuje slova začínající baba a neobsahující abb

Doplněk jazyka $-L = \{ w \mid w\notin L \} = X^*- L$

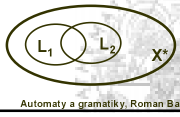
Příklad: slova jazyka neobsahují posloupnost tří symbolů 1

Platí tradiční de Morganova pravidla

$L_1 \cap L_2 = -(-L_1 \cup -L_2)$

$L_1 \cup L_2 = -(-L_1 \cap -L_2)$

$L_1 - L_2 = L_1 \cap -L_2$



Uzavřenost na množinové operace

Nechť L_1 a L_2 jsou jazyky rozpoznávané konečnými automaty. Potom $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 - L_2$ a $-L_1$ jsou také jazyky rozpoznávané konečnými automaty (třída $\mathcal{F}$ je uzavřena na uvedené operace).

Konstruktivní důkaz:

doplněk

- stačí prohodit koncové a nekoncové stavy přijímajícího automatu

sjednocení, průnik a rozdíl

- idea: paralelní běh přijímajících automatů
- $A_1 = (Q_1,X,\delta_1,q_1,F_1)$, $A_2 = (Q_2,X,\delta_2,q_2,F_2)$
- uděláme spojený automat $A = (Q,X,\delta,q,F)$
  - $Q = Q_1 \times Q_2$, $q = (q_1,q_2)$
  - $\delta((p_1,p_2),x) = (\delta_1(p_1,x), \delta_2(p_2,x))$
- sjednocení $F = (F_1 \times Q_2) \cup (Q_1 \times F_2)$
- průnik $F = F_1 \times F_2$
- rozdíl $F = F_1 \times (Q_2 - F_2)$

Množinové operace v příkladě

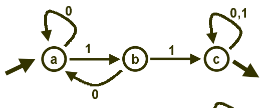
Navrhněte konečný automat přijímající slova, která obsahují 3k+2 symbolů 1 a neobsahují posloupnost 11.

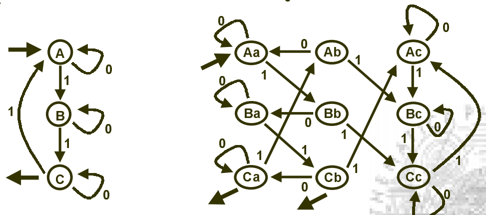
Přímá konstrukce komplikovaná!

$L_1 = \{w \mid w \in \{0,1\}^* \ \& \ |w|_1 = 3k+2\}$

$L_2 = \{w \mid u,v \in \{0,1\}^* \ \& \ w = u11v\}$

$L = L_1 - L_2$





Řetězcové operace nad jazyky

Zřetězení jazyků	$L_1 \,.\, L_2 = \{ uv \mid u \in L_1 \ \& \ v \in L_2 \}$
Mocniny jazyka	$L^0 = \{\lambda\}$
	$L^{i+1} = L^i \,.\, L$
Pozitivní iterace	$L^+ = L^1 \cup L^2 \ldots = \cup_{i \geq 1} L^i$
Obecná iterace	$L^* = L^0 \cup L^1 \ldots = \cup_{i \geq 0} L^i$
zřejmě $L^* = L^+ \cup \{\lambda\}$	
Otočení jazyka	$L^R = \{ u^R \mid u \in L \}$
reverse, zrcadlový obraz	$(x_1 x_2 \ldots x_n)^R = x_n \ldots x_2 x_1$
Levý kvocient L_1 podle L_2	$L_2 \setminus L_1 = \{ v \mid uv \in L_1 \ \& \ u \in L_2 \}$
Levá derivace L podle w	$\partial_w L = \{w\} \setminus L$
Pravý kvocient L_1 podle L_2	$L_1 / L_2 = \{ u \mid uv \in L_1 \ \& \ v \in L_2 \}$
Pravá derivace L podle w	$\partial^R_w L = L / \{w\}$

Uzavřenost zřetězení

$L_1, L_2 \in \mathcal{F} \Rightarrow L_1.L_2 \in \mathcal{F}$

idea:

nejprve počítá automat $A_1=(Q_1,X,\delta_1,q_1,F_1)$ potom $A_2=(Q_2,X,\delta_2,q_2,F_2)$

realizace:

pomocí nedeterministického konečného automatu $B=(Q,X,\delta,S,F)$

nedeterminismus slouží při rozhodování kdy přepnout do A_2

$Q = Q_1 \cup Q_2$	(předpokládáme různá jména stavů, jinak přejmenuj)		
$S = \{q_1\}$	pokud $\lambda \notin L_1$ ($q_1 \notin F_1$)		
$= \{q_1, q_2\}$	pokud $\lambda \in L_1$ ($q_1 \in F_1$), tj. rovnou přejdeme do A_2		
$F = F_2$	končíme až po přečtení slova z L_2		
$\delta(q,x) = \{\delta_1(q,x)\}$	pokud $q \in Q_1 \ \& \ \delta_1(q,x) \notin F_1$	(počítáme v A_1)	
$= \{\delta_1(q,x), q_2\}$	pokud $q \in Q_1 \ \& \ \delta_1(q,x) \in F_1$	(přechod A_1 do A_2)	
$= \{\delta_2(q,x)\}$	pokud $q \in Q_2$	(počítáme v A_2)	

DCV: ověřit $L(B) = L(A_1) \,.\, L(A_2)$

Uzavřenost iterace

$L \in \mathcal{F} \Rightarrow L^* \in \mathcal{F}$

idea: opakovaný výpočet automatu $A=(Q,X,\delta,q_0,F)$

realizace: nedeterministické rozhodnutí, zda pokračovat nebo restart

pozor! $\lambda \in L^*$ i když $\lambda \notin L$, řešíme pomocí speciálního stavu

hledáme nedeterministický automat $B=(Q',X,\delta',S,F')$

$Q' = Q \cup \{s\}$	přidáme nový stav pro příjem λ	
$S = \{q_0, s\}$	nový stav	
$F' = F \cup \{s\}$	končíme po přečtení slova z L nebo v s (pro λ)	
$\delta'(q,x) = \{\delta(q,x)\}$	pokud $q \in Q \ \& \ \delta(q,x) \notin F$	(počítáme uvnitř A)
$= \{\delta(q,x), q_0\}$	pokud $q \in Q \ \& \ \delta(q,x) \in F$	(možný restart)
$\delta'(s,x) = \{\}$	žádné přechody z nového stavu	

$L \in \mathcal{F} \Rightarrow L^+ \in \mathcal{F}$

stejný důkaz, pouze bez použití stavu s

Uzavřenost reverse

$L \in \mathcal{F} \Leftrightarrow L^R \in \mathcal{F}$

zřejmě $(L^R)^R = L$, a tedy stačí ukázat $L \in \mathcal{F} \Rightarrow L^R \in \mathcal{F}$

idea: obrátíme „šipky“ ve stavovém diagramu

realizace: nedeterministický konečný automat

$A=(Q,X,\delta,q_0,F) \quad \rightarrow \quad B=(Q,X,\delta',F,\{q_0\})$

$\delta'(q,x) = \{p \mid \delta(p,x)=q\} \qquad (\delta(p,x)=q \Leftrightarrow p \in \delta'(q,x))$

$w = x_1 x_2 \ldots x_n$

$q_0, q_1, \ldots q_n$ je přijímající výpočet pro w automatu A

tj. $\delta(q_i,x_{i+1})=q_{i+1}$ a $q_n \in F$

$\Leftrightarrow$

$q_n, q_{n-1}, \ldots q_0$ je přijímající výpočet pro w^R automatu B

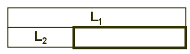
$q_i \in \delta'(q_{i+1},x_{i+1})$

Poznámka:

někdy L nebo L^R má výrazně jednodušší přijímající automat

Uzavřenost kvocientu

$L_1, L_2 \in \mathcal{F} \Rightarrow L_2 \setminus L_1 \in \mathcal{F}$



idea: automat A_1 budeme startovat ve stavech, do kterých se lze dostal slovem z L_2

realizace: nedeterministický automat B „téměř“ totožný s A_1 (rozdíl ve startovních stavech)

$S = \{q \mid q \in Q_1 \ \exists u \in L_2 \ q=\delta_1(q_1,u)\}$ nové startovní stavy

lze nalézt algoritmicky ($A_q=(Q_1,X,\delta_1,q_1,\{q\})$, pak $q \in S \Leftrightarrow L(A_q) \cap L_2 \neq \emptyset$)

$v \in L_2 \setminus L_1$

$\Leftrightarrow \exists u \in L_2 \ uv \in L_1$

$\Leftrightarrow \exists u \in L_2 \ \exists q \in Q_1 \ \delta_1(q_1,u)=q \ \& \ \delta_1(q,v) \in F_1$

$\Leftrightarrow \exists q \in S \ \delta_1(q,v) \in F_1$

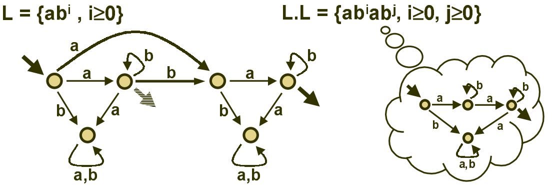
$\Leftrightarrow v \in L(B)$

$L_1, L_2 \in \mathcal{F} \Rightarrow L_1 / L_2 \in \mathcal{F}$

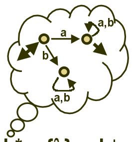
obdobně nebo pomocí $L_1 / L_2 = (L_2^R \setminus L_1^R)^R$

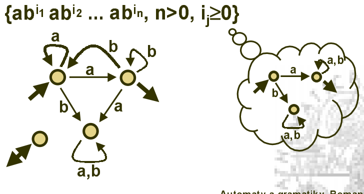
Příklady řetězcových operací

$L = \{ab^i,\ i \geq 0\}$ $L.L = \{ab^i ab^j,\ i \geq 0,\ j \geq 0\}$



$L^+ = \{ab^{i_1} ab^{i_2} \ldots ab^{i_n},\ n>0,\ i_j \geq 0\}$

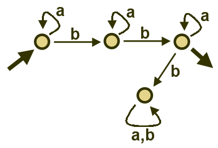


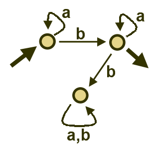


$L^* = \{\lambda\} \cup L^+$

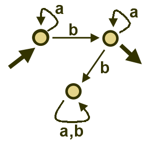
Příklad kvocientu

$L_1 = \{a^{i_1} b a^{i_2} b a^{i_3},\ i_j \geq 0\}$ $L_2 = \{a^{i_1} b a^{i_2},\ i_j \geq 0\}$





$L_2 \setminus L_1 = \{a^{i_1} b a^{i_2},\ i_j \geq 0\} = L_2$



Substituce jazyků

nechť X je konečná abeceda

pro každé $x \in X$ budiž $\sigma(x)$ jazyk v nějaké abecedě Y_x

dále položme:

$\sigma(\lambda) = \{\lambda\}$

$\sigma(u.v) = \sigma(u)\ .\ \sigma(v)$

zobrazení σ: $X^* \to P(Y^*)$, kde $Y = \cup_{x \in X} Y_x$ se nazývá *substituce*

$\sigma(L) = \cup_{w \in L}\ \sigma(w)$

nevypouštějící substituce, žádné $\sigma(x)$ neobsahuje λ

Příklad:

$\sigma(0) = \{a^i b^j,\ i,j \geq 0\}$, $\sigma(1) = \{cd\}$

$\sigma(010) = \{a^i b^j cda^k b^l,\ i,j,k,l \geq 0\}$

homomorfismus h: $h(x) = w_x$ (speciální případ substituce)

nevypouštející homomorfismus: $w_x \neq \lambda$

Věta: $L \in \mathcal{F}$, $\forall x \in X\ \sigma(x) \in \mathcal{F} \Rightarrow \sigma(L), h(L), h^{-1}(L) \in \mathcal{F}$

$h^{-1}(L) = \{w \mid h(w) \in L\}$

Poznámky k uzávěrovým vlastnostem

Zjednodušení návrhu automatů

$L.\emptyset = \emptyset.L = \emptyset$

$\{\lambda\}.L = L.\{\lambda\} = L$

$(L^*)^* = L^*$

$(L_1 \cup L_2)^* = L_1^*(L_2.L_1^*)^* = L_2^*(L_1.L_2^*)^*$

$(L_1.L_2)^R = L_2^R\ .\ L_1^R$

$\partial_w(L_1 \cup L_2) = \partial_w L_1 \cup \partial_w L_2$

$\partial_w(X^* - L) = X^* - \partial_w L$

$h(L_1 \cup L_2) = h(L_1) \cup h(L_2)$

Důkaz regulárnosti

$L = \{w \mid w \in \{0,1\}^*,\ |w|_1 = |w|_0\}$ není regulární

$L \cap \{0^i 1^j,\ i,j \geq 0\} = \{0^i 1^i,\ i \geq 0\}$