Automaty a gramatiky (BI-AAG)

8. Vlastnosti regulárních jazyků

Jan Holub

Katedra teoretické informatiky

Fakulta informačních technologií

ČVUT v Praze

Pumping lemma – podstata problému

Předpokládejme, že by jazyk $L = \{0^n1^n : n \geq 1\}$ byl regulární.

V takovém případě by jazyk L byl přijímaný konečným automatem $\mathcal{A}$ s k stavy.

Pumping lemma – podstata problému

Nechť automat $\mathcal{A}$ čte posloupnost 0^k. Při čtení této posloupnosti je průchod jednotlivými stavy následující:

ε	p_0
0	p_1
00	p_2
$\dots$	$\dots$
0^k	p_k

t.j., $\exists i < j : p_i = p_j$. Označme takový stav q.

Pumping lemma – vyřešení problému

Předpokládejme dále, že přečtením vstupní posloupnosti 1^i se automat dostane ze stavu q do stavu r.
Platí, že:

- Pokud by stav r byl koncovým stavem, pak by automat přijímal větu 0^j1^i, což nechceme.
- Pokud by stav r nebyl koncovým stavem, pak by automat nepřijal větu 0^i1^i, což také ale nechceme.

Proto jazyk $L = \{0^n1^n : n \geq 1\}$ nemůže být regulární.

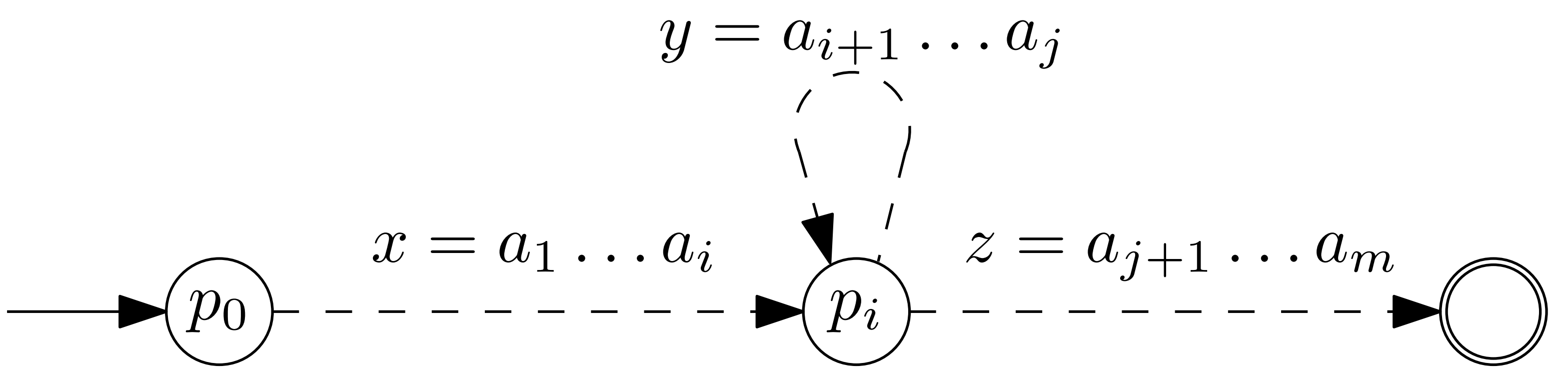
Pumping lemma formálně

Pumping lemma
Nechť L je regulární jazyk. Pak pro jazyk L existuje konstanta $p \geq 1$ taková, že pro každou větu $w \in L$ platí: Jestliže $|w| \geq p$, pak w lze zapsat ve tvaru $w = xyz$ tak, že:

- $y \neq \varepsilon$ (t.j. $|y| \geq 1$),
- $|xy| \leq p$,
- $\forall i \geq 0$ platí, že $xy^i z \in L$.

Pumping lemma neformálně

Pro jazyk L existuje konečný automat a v něm je smyčka. Tato smyčka čte neprázdný podřetěz y. Platí, že tato smyčka se může libovolně-krát „pumpovat“. Pokud je z L, pak i pro každé i-násobné „pumpování“ pro $i \geq 0$ platí, že $xy^iz \in L$.



Použití PL k důkazu, že L není reg.

Důkaz, že jazyk $L = \{0^n 1^n : n \geq 1\}$ není regulární

Předpokládejme, že L je regulární. Pak pomocí pumping lemmatu musí platit, že existuje konstanta $p \geq 1$ taková, že pro každou větu $w \in L$ platí, že:

Jestliže $|w| \geq p$, pak w lze zapsat ve tvaru $w = xyz$ tak, že:

- $y \neq \varepsilon$ (t.j. $|y| \geq 1$),
- $|xy| \leq p$,
- $\forall i \geq 0$ platí, že $xy^i z \in L$.

Věta $w = 0^p 1^p \in L$ je zřejmě delší než p a tudíž musí splňovat požadavky PL. K důkazu, že L není regulární, zkusíme všechna možná rozdělení věty w na xyz. Musíme pak dokázat, že PL neplatí pro žádné z nich.

Použití PL k důkazu, že L není reg.

Důkaz, že jazyk $L = \{0^n 1^n : n \geq 1\}$ **není regulární**

Podle prvních dvou podmínek musí platit:

- xy je neprázdná posloupnost nul,
- y je neprázdná posloupnost nul,
- z obsahuje všechny jedničky.

Ale pak xy^0z (čili odstraníme y z $w = xyz$) nepatří do L (protože počet nul v xy^0z je určitě menší než počet jedniček)!

Proto pumping lemma neplatí pro L a L není regulární jazyk.

Použití PL k důkazu, že L není reg.

Důkaz, že jazyk $L = \{1^m : m \text{ je prvočíslo}\}$ **není regulární**

Předpokládejme, že L je regulární. Pak pomocí pumping lemmatu musí platit, že existuje konstanta $p \geq 1$ taková, … (viz Pumping Lemma):

Předpokládejme větu $w = 1^m$ pro prvočíslo $m \geq p + 2$.

Nyní předpokládejme rozdělení $w = xyz$ a „napumpovanou“ větu $w_1 = xy^{m-|y|}z$. Ukážeme, že w_1 nepatří do L, což rozporuje Pumping lemma.

Použití PL k důkazu, že L není reg.

Důkaz, že jazyk $L = \{1^m : m \text{ je prvočíslo}\}$ **není regulární**

Uvažujme délku věty $w_1 = xy^{m-|y|}z$. Pak platí, že
$|xy^{m-|y|}z| =$
$|xz| + (m - |y|) * |y| =$
$(m - |y|) + (m - |y|) * |y| =$
$(m - |y|) * (1 + |y|)$.

w_1 by bylo prvočíslem, pouze kdyby buďto $(m - |y|)$ nebo $(1 + |y|)$ byly rovné jedné.

Použití PL k důkazu, že L není reg.

Důkaz, že jazyk $L = \{1^m : m \text{ je prvočíslo}\}$ **není regulární**

- $(1 + |y|) \neq 1$, protože $|y| \geq 1$.
- $m \geq p + 2$, $|y| \leq |xy| \leq p$, proto $m - |y| \geq p + 2 - p = 2$.

Pumping Lemma neplatí pro jazyk L, neboť pro libovolné rozdělení $w = xyz \in L$ podle první ze dvou podmínek PL platí, že $xy^{m-|y|}z$ nepatří do L.

Proto, L **není regulární**.

PL pouze k důkazu, že L není reg.

Příklad jazyka, který splňuje podmínky PL, ale není regulární.

- $L = \{u \mid u = a^+b^ic^i \vee u = b^ic^j\}$

Kontrolní otázka

Jak velká je konstanta p v Pumping lemmatu pro *konečný* jazyk?

(Poznámka: Každý konečný jazyk je regulární, a proto pro něj musí platit podmínky Pumping Lemmatu.)

Myhill-Nerodova věta: motivace

- MNv charakterizuje některé zásadní vztahy mezi konečnými automaty nad abecedou Σ a jistými ekvivalenčními relacemi nad řetězci ze Σ^*,
- MNv popisuje některé z nutných a postačujících podmínek pro to, aby daný jazyk byl jazykem regulárním (používá se často k důkazu neregularity jazyka),
- MNv poskytuje formální bázi pro elegantní důkaz existence unikátního (až na isomorfismus) minimálního DKA k danému regulárnímu jazyku.

Ekvivalence

Definice

Ekvivalence $\sim$ je binární relace, která je *reflexivní, symetrická* a *tranzitivní*.

Definice

Třída ekvivalence prvku a na množině X je podmnožina X, obsahující prvky ekvivalentní s a.

Definice

Množina všech tříd ekvivalence v X se nazývá *rozklad množiny* X podle $\sim$ a značí se $X/\sim$.

Definice

Index ekvivalence $\sim$ je počet tříd ekvivalence v $\Sigma/\sim$. Jestliže existuje nekonečně mnoho tříd ekvivalence, definujeme index jako ∞.

Pravá kongruence a prefixová ekv.

Definice

Nechť Σ je abeceda a $\sim$ je ekvivalence na Σ^*. Ekvivalence $\sim$ je *pravou kongruencí*, jestliže pro každé $u, v, w \in \Sigma^*$ platí, že:

$$u \sim v \Rightarrow uw \sim vw$$

Definice

Nechť L je libovolný (ne nutně regulární) jazyk nad abecedou Σ. Definujeme *prefixovou ekvivalenci* pro L, jako relaci $\sim_L$ na množině Σ^* následovně:

$$u \sim_L v \Leftrightarrow \forall w \in \Sigma^* : uw \in L \Leftrightarrow vw \in L$$

Myhill-Nerodova věta

Myhill-Nerodova věta
Nechť L je jazyk nad Σ. Pak následující tvrzení jsou ekvivalentní:

1. L je jazyk přijímaný deterministickým konečným automatem.
2. L je sjednocením některých tříd rozkladu určeného pravou kongruencí na Σ^* s konečným indexem.
3. Relace $\sim_L$ má konečný index.

Důkaz. Dokážeme následující implikace:

- $1 \Rightarrow 2$.
- $2 \Rightarrow 3$.
- $3 \Rightarrow 1$.

Myhill-Nerodova věta: $1 \Rightarrow 2$

Je-li L přijímán DKA, pak L je sjednocením některých tříd rozkladu určeného pravou kongruencí na Σ^* s konečným indexem.

Zaveďme si pro DKA $M = (Q, \Sigma, \delta, q_0, F)$ zobecněnou přechodovou funkci $\hat{\delta}$.

$\hat{\delta} : Q \times \Sigma^* \to Q$ tak, že

$\forall q_1, q_2 \in Q, w \in \Sigma^* : \hat{\delta}(q_1, w) = q_2 \Leftrightarrow (q_1, w) \vdash_M^* (q_2, \varepsilon).$

Myhill-Nerodova věta: $1 \Rightarrow 2$

Důkaz. Pro daný jazyk L přijímaný DKA M zkonstruujeme $\sim$ s potřebnými vlastnostmi:

- Nechť $M = (Q, \Sigma, \delta, q_0, F)$ a δ je úplná.
- Zvolíme $\sim$ jako binární relaci na Σ^* takovou, že $u \sim v \Leftrightarrow \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$.
- Ukážeme, že $\sim$ má potřebné vlastnosti:
  - $\sim$ je ekvivalence: je reflexivní, tranzitivní a symetrická.
  - $\sim$ má konečný index: třídy rozkladu odpovídají stavům DKA.
  - $\sim$ je pravá kongruence: Nechť $u \sim v$ a $a \in \Sigma$. Pak $\hat{\delta}(q_0, ua) = \delta(\hat{\delta}(q_0, u), a) = \delta(\hat{\delta}(q_0, v), a) = \hat{\delta}(q_0, va)$ a tedy $ua \sim va$.
  - L je sjednocením některých tříd ekvivalence $\Sigma^*/\sim$ – těch, které odpovídají F.

Myhill-Nerodova věta: $2 \Rightarrow 3$

Existuje-li relace $\sim$ splňující podmínku 2, pak $\sim_L$ má konečný index.

Důkaz.

- Pro všechny $u, v \in \Sigma^*$ takové, že $u \sim v$, platí $u \sim_L v$:
  - Nechť $u \sim v$. Ukážeme, že také $u \sim_L v$, tj. $\forall w \in \Sigma^* : uw \in L \Leftrightarrow vw \in L$.
  - Víme, že $uw \sim vw$ a protože L je sjednocením některých tříd rozkladu $\Sigma^*/\sim$, platí také $uw \in L \Leftrightarrow vw \in L$.
- Víme tedy, že $\sim \subseteq \sim_L$ (tj. $\sim_L$ je největší pravá kongruence s danými vlastnostmi).
- Každá třída $\sim$ je obsažena v nějaké třídě $\sim_L$.
- Index $\sim_L$ nemůže být větší než index $\sim$.
- $\sim$ má konečný index a tedy i $\sim_L$ má konečný index.

Myhill-Nerodova věta: $3 \Rightarrow 1$

Má-li $\sim_L$ konečný index, pak L je přijímán nějakým konečným automatem.

Důkaz.

- Vytvoříme $M = (Q, \Sigma, \delta, q_0, F)$ přijímající L:
  - $Q = \Sigma^* / \sim_L$ (stavy jsou třídy rozkladu Σ^* relací $\sim_L$),
  - $\forall u \in \Sigma^*, a \in \Sigma : \delta([u], a) = [ua]$,
  - $q_0 = [\varepsilon]$,
  - $F = \{[x] \mid x \in L\}$.
- Uvedená konstrukce je korektní, tj. $L = L(M)$:
  - Indukcí nad délkou slova v ukážeme, že $\forall v \in \Sigma^* : \hat{\delta}([\varepsilon], v) = [v]$.
  - $v \in L \Leftrightarrow [v] \in F \Leftrightarrow \hat{\delta}([\varepsilon], v) \in F$.

Důkaz neregularity pomocí M.-N. věty

Dokažte, že jazyk $L = \{0^n 1^n : n \geq 1\}$ není regulární.

Důkaz.

- Žádné řetězce $\varepsilon, 0, 0^2, 0^3, \ldots$ nejsou $\sim_L$-ekvivalentní, protože $0^i 1^i \in L$, ale $0^i 1^j \notin L$ pro $i \neq j$.
- L má tedy nekonečně mnoho tříd (neboli nekonečný index).
- Podle Myhill-Nerodovy věty tedy nemůže být jazyk L přijímán žádným konečným automatem.

M.-N. věta a minimalita DKA

Věta (2. varianta M.-N. věty)
Počet stavů libovolného minimálního DKA přijímajícího jazyk L je roven indexu $\sim_L$. (Takový DKA existuje právě tehdy, když je index $\sim_L$ konečný.)

Důkaz.

- Každý DKA (bez nedosažitelných stavů) určuje jistou pravou kongruenci s konečným indexem a naopak.
- Je-li L regulární, $\sim_L$ je největší pravou kongruencí s konečným indexem takovou, že L je sjednocením některých tříd příslušného rozkladu.
- Konečný automat, který odpovídá $\sim_L$ (viz důkaz $3 \Rightarrow 1$ M.-N. věty), je tedy minimální konečný automat přijímající L.