Limity konečných automatů

$$L = \{a^n b^n \mid n \geq 0\} = \{\varepsilon, ab, aabb, aaabbb, aaaabbbb \ldots\}$$

$$a \ a \ a \ a \ a \ b \ b \ b \ b \ b$$

Předpokládejme, že existuje automat $\mathcal{M}$ příjímající jazyk L.
Nechť $\mathcal{M}$ má k stavů.

Uvažme výpočet $\mathcal{M}$ na slově $a^n b^n$ kde $n > k$.

$$aaaaaaaaaaaaaaaaaaa\ bbbbbbbbbbbbbbbbbbb$$

Protože $n > k$, musí existovat (z Dirichletova principu) stav p takový, že při čtení iniciální posloupnosti symbolů a projde automat stavem p (alespoň) dvakrát.

$$aaaaaaaa\ aaaaaaa\ aaaabbbbbbbbbbbbbbbbbbb$$

Platí

$$\hat{\delta}(q_0, x) = p \qquad \hat{\delta}(p, y) = p \qquad \hat{\delta}(p, z) = r \in F$$

Pak ale

$$\hat{\delta}(q_0, xz) \;=\; \hat{\delta}(\hat{\delta}(q_0, x), z) \;=\; \hat{\delta}(p, z) \;=\; r \in F$$

$$aaaaaaaa\ \ aaaabbbbbbbbbbbbbbbbbbb$$

Analogicky můžeme "vsunout" slovo y

$$\begin{aligned}
\hat{\delta}(q_0, xyyyz) &= \hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(q_0, x), y), y), y), z) \\
&= \hat{\delta}(\hat{\delta}(\hat{\delta}(\hat{\delta}(p, y), y), y), z) \\
&= \hat{\delta}(\hat{\delta}(\hat{\delta}(p, y), y), z) \\
&= \hat{\delta}(\hat{\delta}(p, y), z) \\
&= \hat{\delta}(p, z) \\
&= r \in F
\end{aligned}$$

Lemma 1. [o vkládání, pumping lemma] Nechť L je regulární jazyk. Pak existuje $n \in \mathbb{N}$ takové,

že libovolné slovo $w \in L$ délky alespoň n lze psát ve tvaru
$w = xyz$, kde $|xy| \leq n$, $y \neq \varepsilon$ a $xy^iz \in L$ pro každé $i \in \mathbb{N}_0$.

Důkaz. Nechť FA $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ rozpoznává jazyk L.
Položme $n = card(Q)$.
Pro libovolné slovo $w \in L$ délky alespoň n platí, že automat $\mathcal{M}$ projde při akceptování slova w (alespoň) dvakrát stejným stavem.

Slovo w se tedy můžeme rozdělit na tři části: $w = xyz$, kde $y \neq \varepsilon$ a $\hat{\delta}(q_0, x) = p$, $\hat{\delta}(p, y) = p$ a $\hat{\delta}(p, z) = r \in F$. Je zřejmé, že ke zopakování nějakého stavu dojde nejpozději po zpracování prvních n znaků a tedy dostáváme $|xy| \leq n$.

Dále $\hat{\delta}(p, y^i) = p$ pro libovolné $i \in \mathbb{N}_0$, proto také $\hat{\delta}(q_0, xy^iz) = r$, tj. $xy^iz \in L(\mathcal{M})$ pro každé $i \in \mathbb{N}_0$. $\square$

$$L \text{ je regulární} \implies \exists n \in \mathbb{N} \;.$$
$$\forall w \in L \;.\; \Big(|w| \geq n \implies$$
$$\exists\, x, y, z \;.\; (w = xyz \;\wedge\; y \neq \varepsilon \;\wedge\; |xy| \leq n \;\wedge$$
$$\forall i \geq 0 \;.\; xy^iz \in L \,) \Big)$$

Pomocí Lemmatu lze dokázat, že nějaký jazyk není regulární

Nechť pro jazyk L platí:

- pro *libovolné* $n \in \mathbb{N}$
- *existuje* takové slovo $w \in L$ délky alespoň n, pro které platí, že
- při *libovolném* rozdělení slova w na takové tři části x, y, z, že $|xy| \leq n$ a $y \neq \varepsilon$
- *existuje alespoň jedno* $i \in \mathbb{N}_0$ takové, že $xy^iz \notin L$.

Pak z Lemma o vkládání plyne, že L není regulární.

Příklad důkazu ne-regularity pomocí Lemmatu o vkládání

$L = \{uc^m u^R \mid u \in \{a,b\}^*, m > 0\}$

Myhill-Nerodova věta
Motivace I

$L = \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\}$

Pravá kongruence

Definice 2.20. Nechť Σ je abeceda a $\sim$ je ekvivalence na Σ^*.

Ekvivalence $\sim$ je **pravá kongruence (zprava invariantní)**, pokud pro každé $u, v, w \in \Sigma^*$ platí

$$u \sim v \Longrightarrow uw \sim vw$$

Index ekvivalence $\sim$ je počet tříd rozkladu $\Sigma^*/_{\sim}$.
Je-li těchto tříd nekonečně mnoho, klademe index $\sim$ roven ∞.

Tvrzení 2.21. Ekvivalence $\sim$ na Σ^* je pravá kongruence právě když pro každé $u, v \in \Sigma^*$, $a \in \Sigma$ platí $u \sim v \Longrightarrow ua \sim va$.
(Implikace $\Longrightarrow$ je triviální, implikace $\Longleftarrow$ se snadno ukáže indukcí k délce zprava přiřetězeného slova w.)

Příklad

$\Sigma = \{a, b\}$

$\sim$: $u \sim v \iff u$ a v začínají stejným symbolem
(počet tříd ekvivalence 3)

$\sim$: $u \sim v \iff \#_a(u) = \#_a(v)$
(počet tříd ekvivalence nekonečný)

$\sim$: $u \sim v \iff u$ a v mají stejné předposlední písmeno

Myhill-Nerodova věta
Motivace II

Prefixová ekvivalence

Definice 2.25. Nechť L je libovolný (ne nutně regulární) jazyk nad abecedou Σ. Na množině Σ^* definujeme relaci $\sim_L$ zvanou **prefixová ekvivalence pro** L takto:

$$u \sim_L v \stackrel{def}{\iff} \forall w \in \Sigma^* : uw \in L \iff vw \in L$$

Příklad

$L = \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\}$
$\sim_L$ má . . . třídy ekvivalence

$L = \{a^n b^n \mid n \geq 0\}$
$\sim_L$ má . . . tříd ekvivalence

Myhill-Nerodova věta

Věta 2.28. Nechť L je jazyk nad Σ.
Pak tato tvrzení jsou ekvivalentní:

1. L je rozpoznatelný 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.

$\mathbf{1} \Longrightarrow \mathbf{2}$

$\mathbf{2} \Longrightarrow \mathbf{3}$

$\mathbf{3} \Longrightarrow \mathbf{1}$

$\square$

$1 \Longrightarrow 2$

Jestliže L je rozpoznatelný konečným automatem
pak L je sjednocením některých tříd rozkladu určeného pravou kongruencí na Σ^* s konečným indexem.

- pro daný L rozpoznávaný automatem $\mathcal{M}$ zkonstruujeme relaci požadovaných vlastností
- $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$, δ je totální
- na Σ^* definujme binární relaci $\sim$ předpisem

$$u \sim v \overset{def}{\Longleftrightarrow} \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$$

- ukážeme, že $\sim$ má požadované vlastnosti

$$u \sim v \overset{def}{\iff} \hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$$

– $\sim$ je ekvivalence (je reflexivní, symetrická, tranzitivní)

– $\sim$ má konečný index
třídy rozkladu odpovídají stavům automatu

– $\sim$ 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 těch tříd rozkladu určeného relací $\sim$, které odpovídají koncovým stavům automatu $\mathcal{M}$ ■

$2 \Longrightarrow 3$

Nechť L je sjednocením některých tříd rozkladu určeného pravou kongruencí R na Σ^* s konečným indexem.
Pak prefixová ekvivalence $\sim_L$ má konečný index.

- $uRv \Longrightarrow u \sim_L v$ pro všechna $u, v \in \Sigma^*$ (tj. $R \subseteq \sim_L$)

- každá třída ekvivalence relace R je **celá** obsažena v nějaké třídě ekvivalence $\sim_L$
- index ekvivalence $\sim_L$ je menší nebo roven indexu ekvivalence R
- R má konečný index $\Longrightarrow \sim_L$ má konečný index ■

$3 \Longrightarrow 1$

Nechť prefixová ekvivalence $\sim_L$ má konečný index.
Pak jazyk L je rozpoznatelný konečným automatem.

Zkonstruujeme automat $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ přijímající L:

- $Q = \Sigma^*/\sim_L$
  Stavy jsou třídy rozkladu Σ^* určeného ekvivalencí $\sim_L$. (Konečnost!)

- $F = \{[v] \mid v \in L\}$

- $q_0 = [\varepsilon]$

- δ je definována pomocí reprezentantů: $\delta([u], a) = [ua]$
  Definice δ je korektní, protože nezávisí na volbě reprezentanta.

Důkaz korektnosti, tj. $L = L(\mathcal{M})$

- $\hat{\delta}([\varepsilon], v) = [v]$ pro každé $v \in \Sigma^*$ (indukcí k délce slova v)
- $v \in L(\mathcal{M}) \iff \hat{\delta}([\varepsilon], v) \in F \iff [v] \in F \iff v \in L$ ■

Použití Myhill-Nerodovy věty I

$$L = \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\}$$

Třídy ekvivalence relace $\sim_L$:
$$\begin{aligned} T_1 &= \{w \in \{a,b\}^* \mid \#_a(w) = 0\} \\ T_2 &= \{w \in \{a,b\}^* \mid \#_a(w) = 1\} \\ T_3 &= \{w \in \{a,b\}^* \mid \#_a(w) = 2\} \\ T_4 &= \{w \in \{a,b\}^* \mid \#_a(w) \geq 3\} \end{aligned}$$

$\sim_L$ **má** konečný index $\implies$
L **je** rozpoznatelný konečným automatem, tj. **regulární** $(3 \implies 1)$

Použití Myhill-Nerodovy věty II

$$L = \{a^n b^n \mid n \geq 0\}$$

Nechť $i \neq j$. Pak $a^i \not\sim_L a^j$, protože $a^i b^i \in L$ ale $a^j b^i \notin L$.

Žádné ze slov $a, a^2, a^3, a^4, a^5, a^6, \ldots$ nepadnou do stejné třídy ekvivalence relace $\sim_L$.

$\sim_L$ **nemá** konečný index $\implies$
L **není regulární** $(\neg 3 \implies \neg 1)$

Použití Myhill-Nerodovy věty III

Věta 2.29. a 2.31. Minimální konečný automat akceptující jazyk L je určen jednoznačně až na isomorfismus (tj. přejmenování stavů). Počet stavů minimálního automatu rozpoznávajícího jazyk L je roven indexu prefixové ekvivalence $\sim_L$.