Myhill-Nerodova věta
Motivace I

Každý DFA $(Q, \Sigma, \delta, q_0, F)$ s totální přechodovou funkcí definuje ekvivalenci $\sim$ na Σ^*:

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

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

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

$$u \sim v \implies uw \sim vw$$

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 $u \sim v \implies ua \sim va$.
(Implikace $\Longrightarrow$ je triviální, implikace $\Longleftarrow$ se snadno ukáže indukcí k délce zprava přiřetězeného slova w.)

Příklad

Jsou tyto relace na slovech nad abecedou $\Sigma = \{a, b\}$ pravé kongruence?

1. $u \sim v \iff u$ a v začínají stejným symbolem nebo $u = v = \varepsilon$
   index $\sim$ je ...

2. $u \sim v \iff \#_a(u) = \#_a(v)$
   index $\sim$ je ...

3. $u \sim v \iff u$ a v mají stejné předposlední písmeno nebo $|u|, |v| \leq 1$
   index $\sim$ je ...

Myhill-Nerodova věta
Motivace II

Pravá kongruence $\sim$ na Σ^* s konečným indexem jednoznačně (až na pojmenování stavů) určuje DFA $(Q, \Sigma, \delta, q_0, \emptyset)$ s totální přechodovu funkcí a bez nedosažitelných stavů slpňující:

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

$u \sim v \iff u$ a v začínají stejným symbolem nebo $u = v = \varepsilon$

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:

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

Věta. $\sim_L$ je pravá kongruence.

Příklad

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

2. $L = \{a^n b^n \mid n \geq 0\}$
   index $\sim_L$ je ...

Myhill-Nerodova věta

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

1. L je rozpoznatelný 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.

$1 \implies 2$

$2 \implies 3$

$3 \implies 1$

□

$1 \implies 2$

Jestliže L je rozpoznatelný deterministickým 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

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

- 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 $\mathcal{M}$ $\square$

$2 \implies 3$

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

- $uRv \implies u \sim_L v$ 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 $\implies$ $\sim_L$ má konečný index $\square$

$3 \implies 1$

Nechť prefixová ekvivalence $\sim_L$ má konečný index.
Pak jazyk L je rozpoznatelný deterministickým 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!)
- $q_0 = [\varepsilon]$
- δ je definována pomocí reprezentantů: $\delta([u], a) = [ua]$
  Definice δ je korektní, protože nezávisí na volbě reprezentanta.
- $F = \{[v] \mid v \in L\}$

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

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

Použití Myhill-Nerodovy věty k důkazu neregularity

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

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 $a^1, a^2, a^3, a^4, a^5, a^6, \ldots$ 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 k důkazu regularity

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

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 deterministickým konečným automatem, tj. **regulární** $(3 \implies 1)$

Další použití Myhill-Nerodovy věty

Věta 2.29. a 2.31. Minimální deterministický konečný automat s totální přechodovou funkcí akceptující jazyk L je určen jednoznačně až na isomorfismus (tj. přejmenování stavů). Počet stavů tohoto automatu je roven indexu prefixové ekvivalence $\sim_L$.

Příklad

Dokažte, že jazyk $L = \{a^i b^j \mid i \neq j,\ i, j \geq 0\}$ není regulární.

Příklad 3.8 a)

Dokažte, že neexistuje DFA se 4 stavy (a s totální přechodovou funkcí) akceptující jazyk $L = \{w \in \{a, b\}^* \mid |w| \geq 4\}$.

Příklad 3.12

Určete relaci $\sim_L$ a její index pro tyto jazyky:

a) $L = \{a\}^*.\{b\}^*.\{c\}^*$

b) $L = \{a^n b^n c^n \mid n > 0\}$

Příklad 3.13

Nechť $\Sigma = \{a, b\}$. Uvažte následující relace na množině Σ^*:

a) $u \sim v \iff \#_a(u) \bmod 4 = \#_a(v) \bmod 4$

b) $u \sim v \iff \#_a(u) \bmod 4 = \#_a(v) \bmod 4$
nebo u i v končí na stejné písmeno

c) $u \sim v \iff \#_a(u) \bmod 4 = \#_a(v) \bmod 4$
a u i v končí na stejné písmeno (nebo $u = v = \varepsilon$)

U každé relace určete, zda je to ekvivalence. Pokud ano, určete její index a zda je pravou kongruencí. Pokud ano, nalezněte jazyk L takový, že $\sim_L = \sim$. Nakonec nalezněte jazyk L', který je sjednocením některých tříd rozkladu $\Sigma^*/\sim$, ale přitom $\sim_{L'} \neq \sim$.

Eliminace ekvivalentních stavů

Nechť $\mathcal{M} = (Q, \Sigma, \delta, q_0, F)$ je DFA bez nedosažitelných stavů, jehož přechodová funkce je totální.

Definice 2.32. Stavy p, q nazveme **jazykově ekvivalentní**, psáno $p \equiv q$, pokud

$$p \equiv q \iff \forall w \in \Sigma^* : (\hat{\delta}(p, w) \in F \iff \hat{\delta}(q, w) \in F).$$