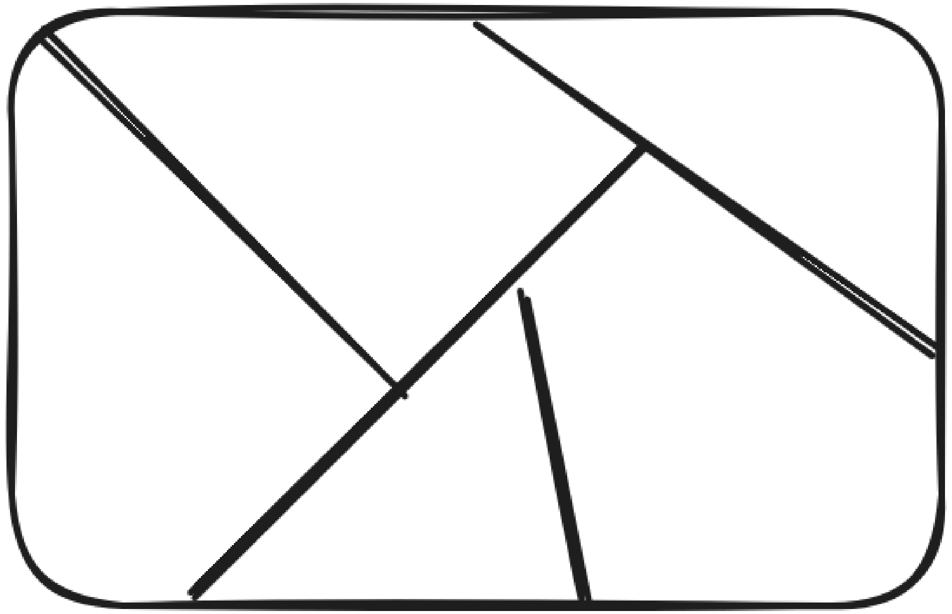
Izrek Myhill-Nerode in minimizacija KA

Uroš Čibej



Pregled

- (Ne)razločljivost nizov
- Izrek Myhill-Nerode
- Uporaba za dokazovanje neregularnosti
- Minimizacija avtomatov

Literatura

- https://www.cs.columbia.edu/~tal/3261/fall24/Handouts/3_Myhill_Nerode.pdf
- https://www.cs.columbia.edu/~tal/3261/fall22/handouts/4_MyhillNerodePlus.pdf
- Hopcroft, Motwani, Ullman - razdelek 4.4.

Razločljivost

Podan imamo poljuben jezik $L \subseteq \Sigma^*$

Dva niza

$$x, y \in \Sigma^*$$

sta razločljiva v L, če

$$\exists z \in \Sigma^* : (xz \in L \land yz \notin L) \lor (yz \in L \land xz \notin L)$$

Primer

Vzemimo jezik

$$L = \{w \in \{0,1\}^* \mid w \text{ se konča na } 001\}$$

$x = 00100$
$y = 101$

Pri $z = 1$

$$xz \in L \land yz \notin L$$

Povezava z DKA

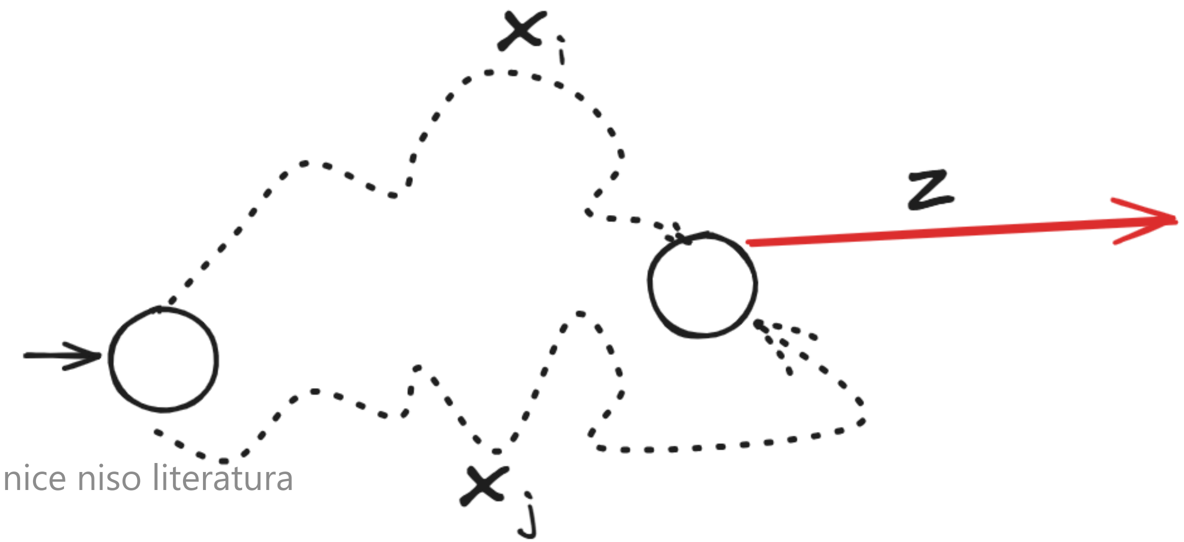
Lema: Naj bo $L \subseteq \Sigma^*$ in množica vzajemno razločljivih nizov $R = \{x_1, \ldots, x_k\}$. Potem ne obstaja DKA z manj kot k stanji, ki bi razpoznaval L.

Dokaz

Predpostavimo obstoj takega avtomata z manj kot k stanji za L. V R obstajata dva niza x_i, x_j za katera velja

$$\hat{\delta}(q_0, x_i) = \hat{\delta}(q_0, x_j)$$

S poljubnim podaljškom z velja $x_i z \in L \iff x_j z \in L$. Torej x_i in x_j nista razločljiva, kar je v protislovju s predpostavko da sta v R.



Relacija nerazločljivosti $\sim_L$

$$x \sim_L y : \forall z \in \Sigma^* : xz \in L \iff yz \in L$$

Primer za prejšnji jezik

$$00101 \sim_L 101$$

Relacija je ekvivalenčna

1. refleksivnost $a \sim_L a$
2. simetričnost $a \sim_L b \iff b \sim_L a$
3. tranzitivnost $a \sim_L b, b \sim_L c \implies a \sim_L c$

Refleksivnost

$$\forall z : xz \in L \iff xz \in L$$

Simetričnost

$$(xz \in L \iff yz \in L) \implies (yz \in L \iff xz \in L)$$

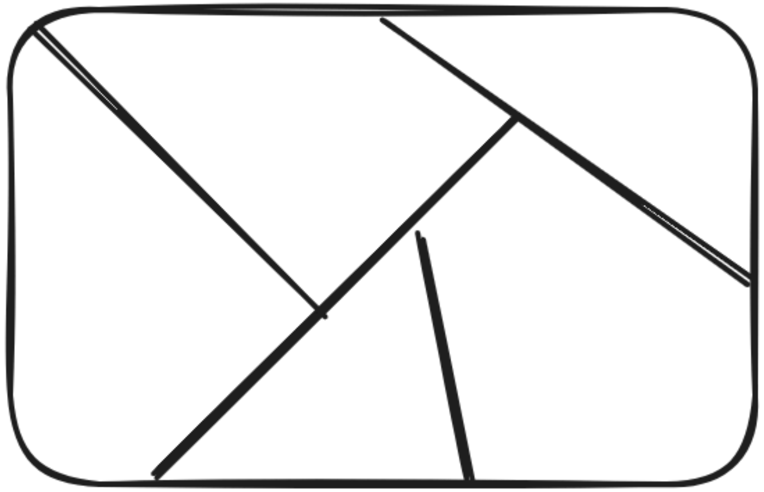
Tranzitivnost

$$[(az \in L \iff bz \in L) \land (bz \in L \iff cz \in L)] \implies az \in L \iff cz \in L$$

Posledica

Množica razpade na ekv. razrede, lahko jih predstavimo z enim predstavnikom:

$$[x] = \{y \mid y \sim_L x\}$$



Primer I

Končen jezik

$$L = \{ab, aa\}$$

Primer I

Končen jezik

$$L = \{ab, aa\}$$

1. $[\varepsilon]$
2. $[a]$
3. $[ab]$
4. vse ostalo

Primer II

Vzemimo jezik

$$L = \{w \in \{0,1\}^* \mid w \text{ se konča na } 001\}$$

Relacija $\sim_L$ ima 4 ekv. razrede

1. $[\varepsilon]$
2. $[0]$
3. $[00]$
4. $[001]$

Izrek Myhill-Nerode (formalno)

Izrek: Jezik $L \subseteq \Sigma^*$ je regularen $\iff$ ima relacija $\sim_L$ ima končno število ekv. razredov. Število ekv. razredov je tudi enako številu stanj najmanjšega determinističnega končnega avtomata.

Dokaz 1. $L \in RJ \implies \sim_L$ ima končno ekv. razredov

Predpostavimo, da ima $\sim_L$ neskončno ekv. razredov. Potem vemo, da obstaja neskončna množica $\{x_1, x_2, \ldots\}$ (iz vsakega razreda en element), ki so paroma razločljivi.

(iz leme) $\nexists DKA$, $|Q|$ je končno.

Dokaz 2. $\sim_L$ ima končno ekv. razredov $\implies L \in RJ$

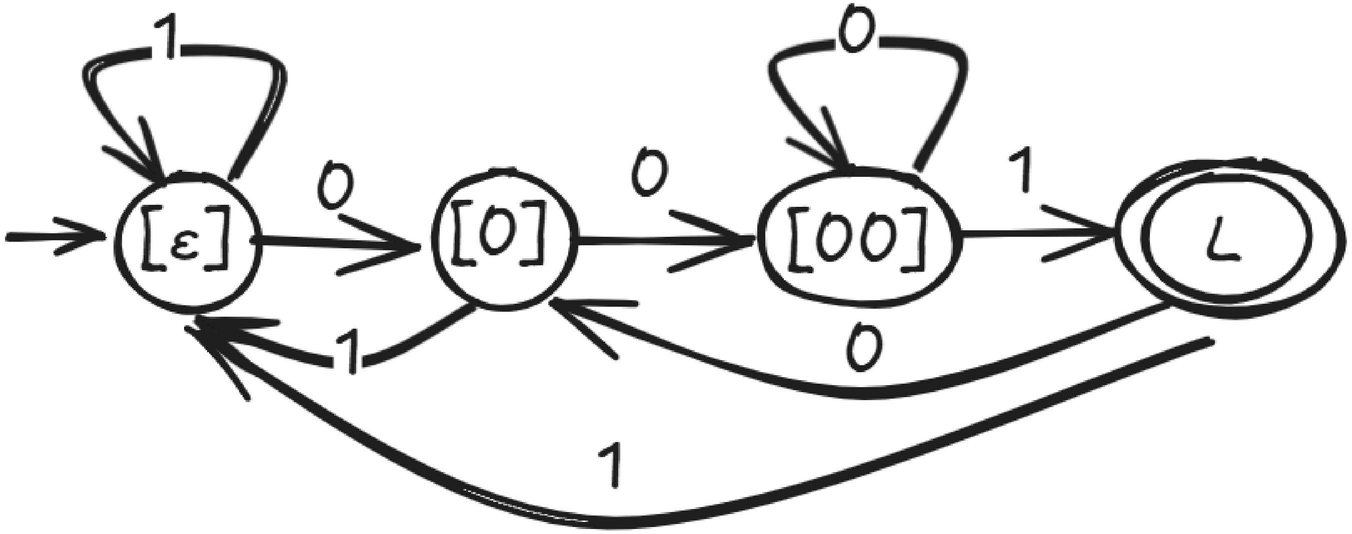
Konstrukcija

$$M = \langle Q, \Sigma, \delta, q_0. F \rangle$$

- $Q = \{[x] \mid x \in \Sigma^*\}$ - vsak ekv. razred predstavlja stanje
- $q_0 = [\varepsilon]$
- $F = \{[x] \mid x \in L\}$
- $\delta([x], a) = [xa]$

Primer konstrukcije

$$L = \{w \in \{0,1\}^* \mid w \text{ se konča na } 001\}$$



Dokaz 3. Skonstruiran avtomat je najmanjši

Sledi neposredno iz leme, skonstruirali smo avtomat, ki ima natanko toliko stanj kot je ekv. razredov $\sim_L$. Po lemi manjši avtomat ne obstaja.

Dokazovanje neregularnosti z M-N

Skonstruiramo neskončno množico:

$$R = \{x_1, x_2, x_3, \ldots\}, i \neq j \implies x_i \nsim_L x_j$$

Po lemi: ne obstaja končni avtomat.

Primer 1.

$$L = \{a^n b^n\}$$

Razločljiva množica:

$$R = \{a^k \mid k \in \mathbb{N}\}$$

$i \neq j \implies a^i \not\sim_L a^j$

Množica R predstavlja neskončno razločljivo množico $\implies L \notin RJ$

Primer 2.

$$L = \{w \in \{0,1\}^* \mid w = w^r\}$$

(^{r} je operacija obrata niza - reverse)

Minimizacija

Začnemo z

$$M = \langle Q, \Sigma, \delta, q_o, F \rangle$$

1. predpostavimo, da so vsa stanja dosegljiva
2. detektiramo nerazločljiva stanja
3. nerazločljiva stanja združimo

Algoritem za določitev nerazločljivih stanj

Polnjenje tabele: eq

Inicializacija: $eq(p, q) = p \in F \iff q \in F$

Ponavljaj dokler se eq spreminja:

- $\forall (p, q), a \mid eq(\delta(p, a), \delta(q, a)) = False$:
  - $eq(p, q) = False$

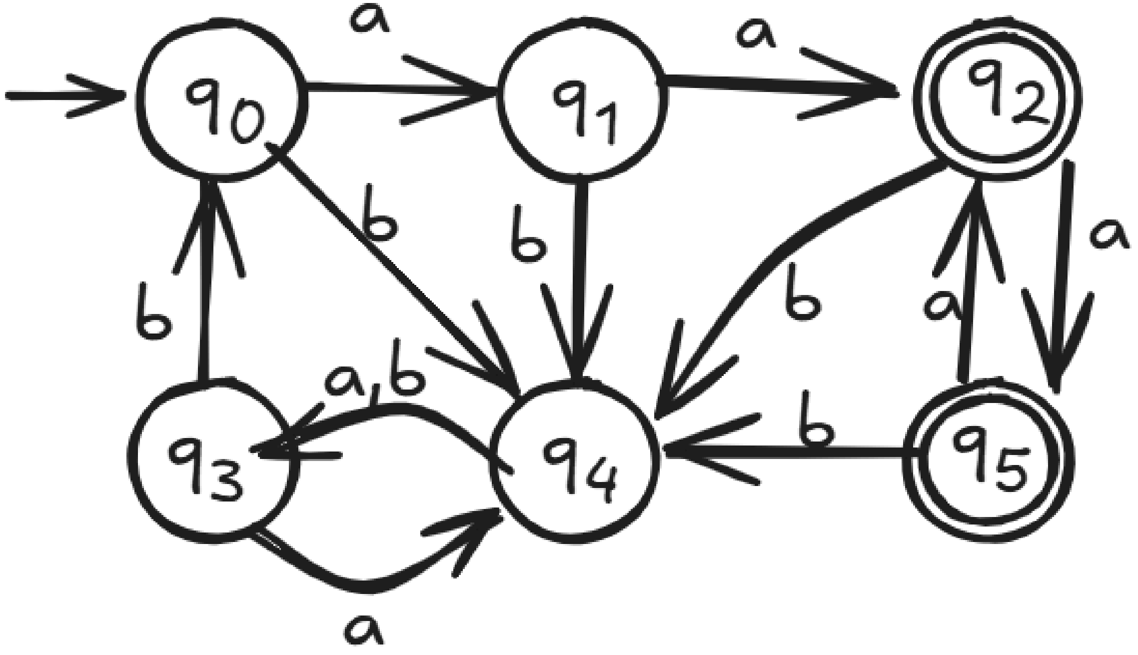
Združitev nerazločljivih stanj

Vsa ekv. stanja zapišemo kot: $[q] = \{p \in Q \mid eq(p, q)\}$

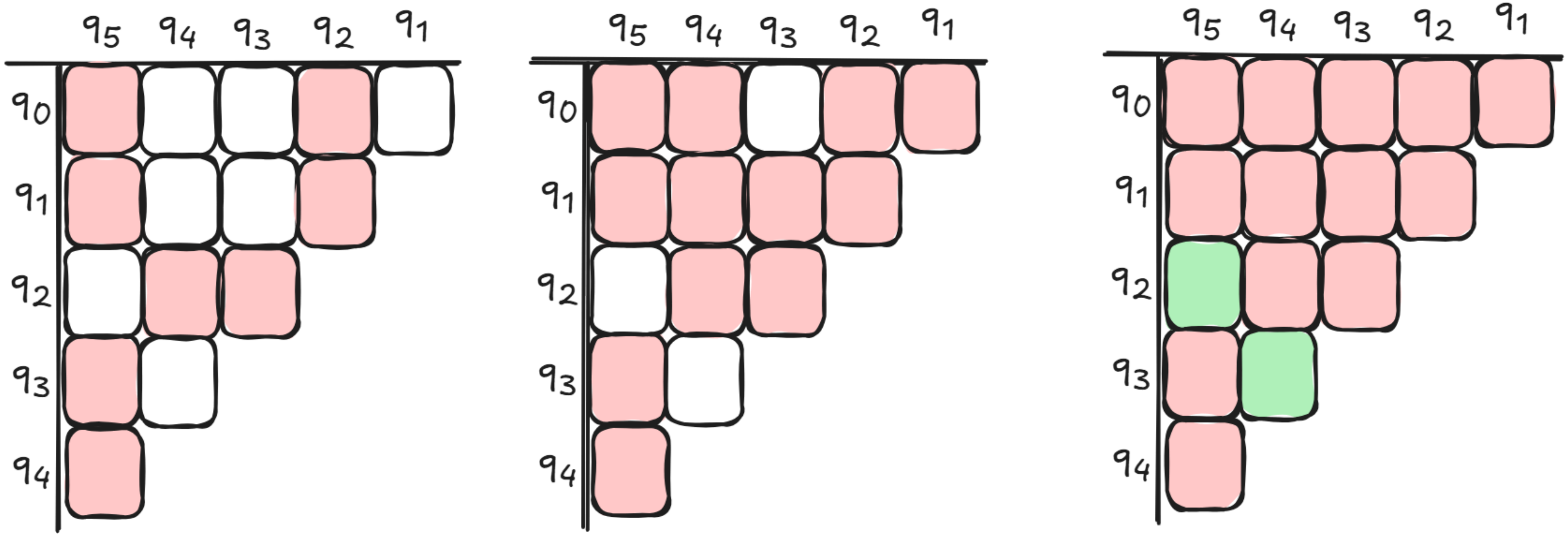
Nov avtomat

- $Q' = \{[q] \mid q \in Q\}$
- $q_0' = [q_0]$
- $\delta'([q], a) = [\delta(q, a)]$
- $F' = \{[q] \mid q \in F\}$

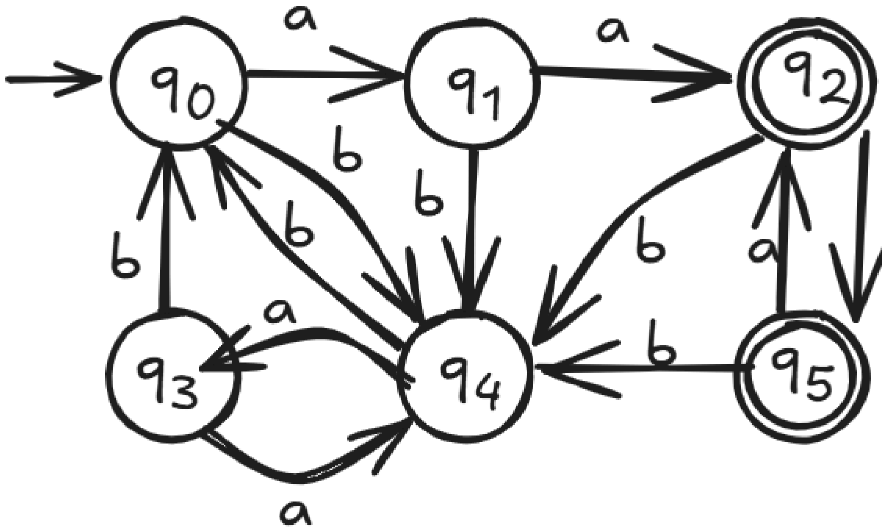
Primer

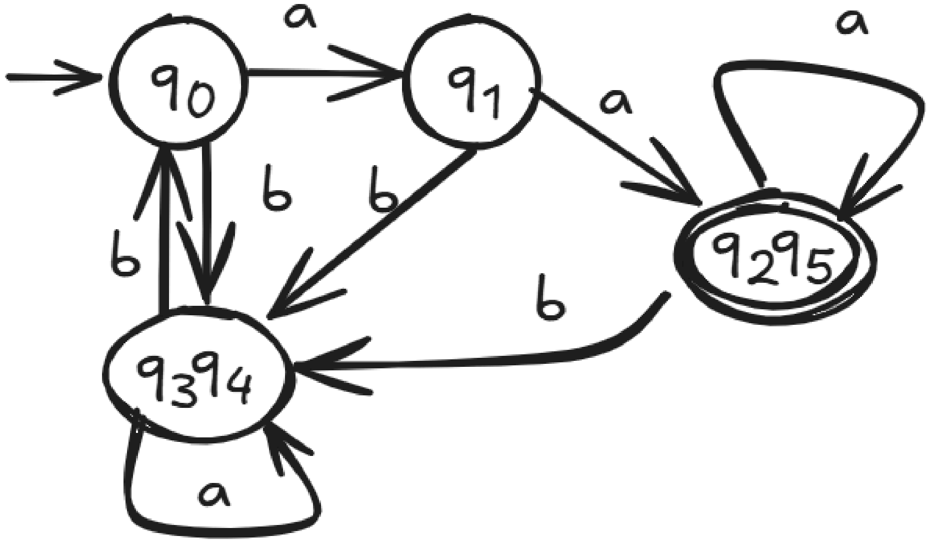


Primer: min. tabela



Minimizirani avtomat





"Pokaz" pravilnosti

1. Vsako stanje predstavlja podmnožico ekv. razreda relacije $\sim_L$
2. algoritem preizkusi vse podaljške z (indukcija) s katerimi bi lahko pokazali, da dve stanji nista v istem ekv. razredu
3. stanja, za katere nismo našli takega podaljška, so ekvivalentna

Domača naloga: napišite bolj rigorozen dokaz

KONEC

(prvega dela)

Povzetek

Znamo:

- opisati problem kot **jezik**
- formalno opisati **računski model**
- pokazati **ekvivalenco** računskih modelov
- pokazati česa model **ni sposoben** rešiti