Automaty a regulární jazyky

Anša Lauschmannová

Definice. *Abecedou* rozumíme libovolnou konečnou neprázdnou množinu znaků. Obvykle ji budeme značit Σ. *Slovo* nad abecedou Σ je nějaká (i prázdná) posloupnost znaků z dané abecedy. Množinu všech slov nad danou abecedou značíme Σ^*, množinu neprázdných slov nad danou abecedou Σ^+, prázdné slovo obvykle označujeme λ. *Jazykem* rozumíme libovolnou množinu L slov; $L \subseteq \Sigma^*$.

Příklady.

(i) Nechť $\Sigma = \{1, 2, 5, 10, 20\}$ je množina českých platných mincí (přijímaných nápojovými automaty). Pořadí, v jakém naházím několik mincí do automatu na nápoje, určuje slovo nad touto abecedou. Jazykem L mohou být například všechny posloupnosti, jejichž součet je roven alespoň ceně nabízeného nápoje, nebo všechny posloupnosti obsahující alespoň jednu dvacetikorunu.

(ii) Budeme-li každé slovo (v obvyklém významu) českého jazyka považovat za jediný znak, bude každá posloupnost slov (třeba „venku prší“, ale také „koloběžka stínili domácí kam“) slovem ve formálním významu. Jazyk L může být tvořen například všemi správně utvořenými větami českého jazyka.

Definice. *Konečný automat* je zařízení popsané těmito parametry:

- Q je konečná neprázdná množina stavů,
- Σ je abeceda,
- $\delta : Q \times \Sigma \to Q$ je přechodová funkce,
- q_0 je počáteční stav,
- $F \subseteq Q$ je množina přijímacích (koncových) stavů.

Výpočet pro slovo w je posloupnost stavů $q_0, \ldots, q_n$ taková, že $n = |(w)|$ a platí

$$q_i = \delta(q_{i-1}, a_i), \quad i = 1, \ldots, n,$$

kde $w = a_1 \ldots a_n$.

Slovo w je *přijato*, pokud příslušný výpočet skončí v přijímacím stavu. Je-li M konečný automat, pak symbolem $L(M)$ označujeme jazyk obsahující právě ta slova, která automat M přijímá. Jazyk L, pro který existuje automat M, který ho rozpoznává (tedy $L = L(M)$), nazýváme *regulární*.

Rozšířená přechodová funkce je definována následovně:

- $\delta^* : Q \times \Sigma^* \to Q$,
- $\delta^*(q, \lambda) = q$,
- $\delta^*(q, wx) = \delta(\delta^*(q, w), x)$.

M přijímá $w \equiv \delta^*(q_0, w) \in F$.

Definice. *Pravá kongruence* je každá relace $\sim$, která pro všechny prvky nosné množiny splňuje následují podmínky:

- $u \sim u$ (reflexivita),
- $u \sim v \equiv v \sim u$ (symetrie),
- $(u \sim v \ \ \& \ \ v \sim w) \rightarrow u \sim w$ (tranzitivita),
- $u \sim v \rightarrow uw \sim vw$ (kongruence).

Možná jsi poznal, že první tři vlastnosti říkají, že $\sim$ je ekvivalence; má-li konečný počet tříd rozkladu, řekneme o ní, že je *konečného indexu.*

Věta. (Nerodova) *Nechť L je jazyk nad konečnou abecedou Σ. Potom následující dvě tvrzení jsou ekvivalentní:*

(i) L je regulární jazyk.

(ii) Existuje pravá kongruence konečného indexu na Σ^* tak, že L je sjednocením jistých tříd rozkladu této kongruence.

Cvičení.

(1) Rozhodněte, zda jsou následující jazyky regulární:
- $L_1 = \{0^k 1^n, k \leq n\}$,
- $L_2 = \{0^{2n}\}$,
- $L_3 = \{0^{2^n}\}$.

(2) Nechť L je regulární jazyk. Je také jazyk $\{u; *u* \in L\}$ regulární?

Přechodová funkce v definici konečného automatu na základě stavu automatu a vstupního symbolu určuje jediný stav, do kterého automat přejde – je jednoznačná. Při řešení problémů se často ukazuje jako výhodné pracovat s automatem, ve kterém pro jeden vstupní symbol existuje více možných přechodů, a až v okamžiku výpočtu se rozhodnout, který z nich zvolit. Takový *nedeterministický automat* přijímá vstupní slovo tehdy, když existuje výpočet, který končí v přijímajícím stavu. Lze ukázat, že každý nedeterministický automat lze převést na obyčejný konečný automat přijímající stejný jazyk.

V praxi se často stává, že chceme nejenom poznat, jestli vstupní slovo patří do nějakého jazyka, ale zároveň s ním chceme provést nějakou operaci – očekáváme, že automat nám dá obsáhlejší odpověď než pouhé „ano“, „ne“. K tomu slouží následující dvě modifikace konečného automatu:

Definice. *Mealyho automat* je konečný automat, jehož výstup je spojen s přechody mezi stavy; formálně je popsán jako konečný automat rozšířený o následující parametry:

- Σ^* – výstupní abeceda
- $\epsilon : Q \times \Sigma \rightarrow \Sigma^*$ – výstupní funkce.

Definice. *Mooreův automat* je konečný automat, jehož výstup je určen aktuálním stavem; formálně je popsán jako konečný automat rozšířený o následující parametry:

- Σ^* – výstupní abeceda
- $\epsilon : Q \to \Sigma^*$ – výstupní funkce.

Jediný rozdíl mezi Mealyho a Mooreovým automatem je ve tvaru výstupní funkce. U obou typů automatů se většinou neuvádí množina koncových stavů – výstup automatu nás zajímá víc než stav, ve kterém výpočet skončí. Na příkladech si ukážeme, že ke každému Mealyho automatu lze sestrojit Mooreův automat, který dává stejné výstupy, a naopak (až na jednu maličkost).

Cvičení. Navrhněte Mealyho stroj nad abecedou $\{0, 1\}$, který

(1) invertuje vstupní slovo ($0 \to 1, 1 \to 0$). Převeďte jej na Mooreův stroj.
(2) dává výstup 1, je-li na vstupu člen skupiny symbolů 1, která je přímo předcházena skupinou symbolů 00, a jinak dává výstup 0.
(3) realizuje sčítání binárních čísel – součástí zadání je i definice vstupu takového stroje.

Převeďte nalezené automaty na Mooreovy.