

形式语言与自动机理论

正则语言的性质

王春宇

计算机科学与技术学院
哈尔滨工业大学

正则语言的性质

- 正则语言的泵引理
- 正则语言的封闭性
- 正则语言的判定性质
- 自动机的最小化

例 1. $L = \{0^m 1^n \mid m, n \geq 0\}$ 是否是正则语言?

例 2. $L = \{0^m 1^n \mid m \geq 2, n \geq 4\}$ 是否是正则语言?

例 3. $L_{01} = \{0^n 1^n \mid n \geq 0\}$ 是否是正则语言?

正则语言的泵引理

定理 5 (正则语言的泵引理)

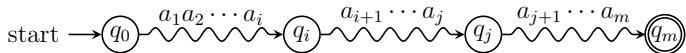
如果语言 L 是正则的, 那么存在正整数 N , 对 $\forall w \in L$, 只要 $|w| \geq N$, 就可以将 w 分为三部分 $w = xyz$ 满足:

- ① $y \neq \varepsilon$ ($|y| > 0$);
- ② $|xy| \leq N$;
- ③ $\forall k \geq 0, xy^kz \in L$.

证明:

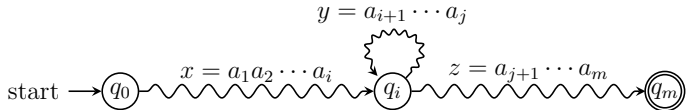
① 如果 L 正则, 那么存在有 n 个状态 DFA A 使 $\mathbf{L}(A) = L$;

② 取 $w = a_1 \dots a_m \in L$ ($m \geq n$), 定义 $q_i = \hat{\delta}(q_0, a_1 \dots a_i)$;



③ 由鸽巢原理, 必有两状态相同 $q_i = q_j$ ($0 \leq i < j \leq n$);

④ 那么 $w = xyz$ 如图, 且有 $\forall k > 0, xy^kz \in L$;



⑤ 而因为 $i < j$ 所以 $y \neq \varepsilon$ (即 $|y| > 0$), 因为 $j \leq n$ 所以 $|xy| \leq n$. □

泵引理的应用

续例3. 证明 $L_{01} = \{0^n 1^n | n \geq 0\}$ 不是正则语言.

证明:

- ① 假设 L_{01} 是正则的.
- ② 那么, 存在 $N \in \mathbb{Z}^+$, 对 $\forall w \in L_{01} (|w| \geq N)$ 满足泵引理.
- ③ 从 L_{01} 中取 $w = 0^N 1^N$, 显然 $w \in L_{01}$ 且 $|w| = 2N \geq N$.
- ④ 那么, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.
- ⑤ 因此 y 只能是 0^m 且 $m > 0$.
- ⑥ 那么 $xy^2z = 0^{N+m} 1^N \notin L_{01}$, 而由泵引理 $xy^2z \in L_{01}$, 矛盾.
- ⑦ 所以假设不成立, L_{01} 不是正则的. □

例 4. 证明 $L_{eq} = \{w \mid w \text{ 由数量相等的 } 0 \text{ 和 } 1 \text{ 构成}\}$ 不是正则的.

思考题

刚刚已经证明了

$$L_{01} = \{0^n 1^n \mid n \geq 0\}$$

不是正则语言, 那么能否使用

$$L_{01} \subseteq L_{eq}$$

来说明 L_{eq} 也不是正则的呢?

续例 4. 证明 $L_{eq} = \{w \mid w \text{ 由数量相等的 } 0 \text{ 和 } 1 \text{ 构成}\}$ 不是正则的.
证明:

- ① 假设 L_{eq} 是正则的.
- ② 那么, 存在 $N \in \mathbb{Z}^+$, 对 $\forall w \in L_{eq} (|w| \geq N)$ 满足泵引理.
- ③ 从 L_{eq} 中取 $w = 0^N 1^N$, 显然 $w \in L_{eq}$ 且 $|w| = 2N \geq N$.
- ④ 那么, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.
- ⑤ 因此 y 只能是 0^m 且 $m > 0$.
- ⑥ 那么 $xy^2z = 0^{N+m}1^N \notin L_{eq}$, 而由泵引理 $xy^2z \in L_{eq}$, 矛盾.
- ⑦ 所以假设不成立, L_{eq} 不是正则的. □

例 5. 证明 $L = \{0^i 1^j \mid i > j\}$ 不是正则的.

证明:

- ① 假设 L 是正则的.
- ② 那么, 存在 $N \in \mathbb{Z}^+$, 对 $\forall w \in L (|w| \geq N)$ 满足泵引理.
- ③ 从 L 中取 $w = 0^{N+1}1^N$, 则 $w \in L$ 且 $|w| = 2N + 1 \geq N$.
- ④ 由泵引理, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.
- ⑤ 那么, y 只能是 0^m 且 $m \geq 1$.
- ⑥ 那么, $xz = xy^0z = 0^{N+1-m}1^N \notin L$, 因为 $N + 1 - m \leq N$, 而由泵引理 $xy^0z \in L$, 矛盾.
- ⑦ 所以假设不成立, L 不是正则的. □

例 6. Prove $L = \{a^3b^nc^{n-3} \mid n \geq 3\}$ is not regular.

证明:

- ① 假设 L 是正则的.
- ② 那么, 存在 $N \in \mathbb{Z}^+$, 对 $\forall w \in L (|w| \geq N)$ 满足泵引理.
- ③ 从 L 中取 $w = a^3b^Nc^{N-3}$, 则 $w \in L$ 且 $|w| = 2N > N$.
- ④ 由泵引理, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.
- ⑤ 那么, 则 y 只可能有 3 种情况 ($m > 0, r > 0, s > 0$):
 - ① $y = a^m$, 则 $xy^2z = a^{3+m}b^Nc^{N-3} \notin L$;
 - ② $y = b^m$, 则 $xy^2z = a^3b^{N+m}c^{N-3} \notin L$;
 - ③ $y = a^rb^s$, 则 $xy^2z = a^3b^sa^rb^Nc^{N-3} \notin L$.
- ⑥ 无论 y 为何种情况, xy^2z 都不可能在 L 中, 与泵引理矛盾.
- ⑦ 所以假设不成立, L 不是正则的. □

思考题

- $L = \{0^n 1^n \mid 0 \leq n \leq 100\}$ 是否是正则语言?
- 有限的语言, 是否符合泵引理呢?
 - \emptyset
 - $\{\varepsilon\}$
 - $\{0, 00\}$

泵引理只是必要条件

- 泵引理只是正则语言的必要条件
- 只能用来证明某个语言不是正则的
- 与正则语言等价的定理 — Myhill-Nerode Theorem

泵引理只是必要条件

例7. 语言 L 不是正则的, 但每个串都可以应用泵引理

$$L = \{ca^n b^n \mid n \geq 1\} \cup \{c^k w \mid k \neq 1, w \in \{a, b\}^*\}$$

- 其中 $A = \{ca^n b^n \mid n \geq 1\}$ 部分不是正则的
- 而 $B = \{c^k w \mid k \neq 1, w \in \{a, b\}^*\}$ 部分是正则的
- 而 A 的任何串 $w = ca^i b^i$, 都可应用泵引理, 因为

$$w = (\varepsilon)(c)(a^i b^i)$$

重复字符 c 生成的新串都会落入 B 中

思考题

泵引理中的正整数 N 与 DFA 的状态数 n 之间有何关系？与 NFA 的状态数之间呢？

思考题

语言

$$L = \{0^n x 1^n \mid n \geq 1, x \in \{0, 1\}^*\}$$

是否是正则语言？