形式语言与自动机理论

正则语言的性质

王春宇
chunyu@hit.edu.cn

计算学部
哈尔滨工业大学

2021 年 4 月

正则语言的性质

- 证明语言的非正则性
  - 正则语言的泵引理
  - 泵引理的应用
  - 泵引理只是必要条件
- 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)

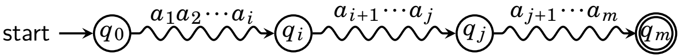
如果语言 L 是正则的, 那么存在正整数 N, 它只依赖于 L, 对 $\forall w \in L$, 只要 $|w| \geq N$, 就可以将 w 分为三部分 $w = xyz$ 满足:

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

证明:

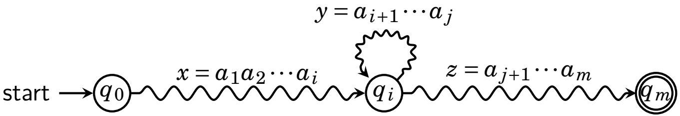
1. 如果 L 正则, 那么存在有 n 个状态 DFA A 使 $\mathbf{L}(A) = L$;

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 $\forall k \geq 0$, $xy^kz \in L$;



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

泵引理的应用

续例 3. 证明 $L_{01} = \{ 0^n 1^n \mid n \geq 0 \}$ 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. $\square$

例 4. 证明 $L_{\text{eq}} = \{\, w \in \{0,1\}^* \mid w \text{ 由数量相等的 0 和 1 构成} \,\}$ 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

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

来说明 L_{eq} 也不是正则的呢?

续例 4. 证明 $L_{\text{eq}} = \{\, w \in \{0,1\}^* \mid w$ 由数量相等的 0 和 1 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 $\forall w \in L_{\text{eq}} \left(|w| \geq N\right)$ 满足泵引理.
3. 从 L_{eq} 中取 $w = 0^N 1^N$, 显然 $w \in L_{\text{eq}}$ 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 $xy^2z = 0^{N+m}1^N \notin L_{\text{eq}}$, 而由泵引理 $xy^2z \in L_{\text{eq}}$, 矛盾.
7. 所以假设不成立, L_{eq} 不是正则的. $\square$

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 $\forall w \in L \left(|w| \geq N\right)$ 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. $\square$

例 6. Prove $L = \{ a^3 b^n c^{n-3} \mid n \geq 3 \}$ 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 $w = a^3 b^N c^{N-3}$, 则 $w \in L$ 且 $|w| = 2N \geq N$.
4. 由泵引理, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.
5. 那么, 则 y 只可能有 3 种情况 $(m > 0, r > 0, s > 0)$:
   1. $y = a^m$, 则 $xy^2z = a^{3+m} b^N c^{N-3} \notin L$;
   2. $y = b^m$, 则 $xy^2z = a^3 b^{N+m} c^{N-3} \notin L$;
   3. $y = a^r b^s$, 则 $xy^2z = a^3 b^s a^r b^N c^{N-3} \notin L$.
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. $\square$

思考题

- $L = \left\{ 0^n 1^n \;\middle|\; 0 \le n \le 100 \right\}$ 是否是正则语言?
- 有限的语言，是否符合泵引理呢?
  - $\varnothing$
  - $\{\varepsilon\}$
  - $\{0, 00\}$

课堂练习. Prove that $L = \{a^i b^j c^k \mid i + j = k\}$ is not regular with pumping lemma.

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

例 7. 语言 L 不是正则的, 但每个串都可以应用泵引理

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

- 其中 $A = \{\, ca^n b^n \mid n \ge 1 \,\}$ 部分不是正则的
- 而 $B = \{\, c^k w \mid k \ne 1, w \in \{a, b\}^* \,\}$ 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

对任何正则语言 L, 在泵引理中, 与 L 相关联的正整数 N

- 与识别 L 的 DFA 的状态数 n 之间有何关系?
- 与识别 L 的 NFA 的状态数之间呢?

思考题

语言

$$L = \left\{ 0^n x 1^n \;\middle|\; n \geq 1, x \in \{0,1\}^* \right\}$$

是否是正则语言?

正则语言的性质

- 证明语言的非正则性
- 正则语言的封闭性
- 正则语言的判定性质
- 自动机的最小化

正则语言的封闭性

定义

正则语言经某些**运算**后得到的新语言仍保持正则, 称为在这些运算下**封闭**.

正则语言 L 和 M, 在这些运算下封闭

- 并: $L \cup M$
- 连接: LM
- 闭包: L^*
- 补: $\overline{L}$
- 差: $L - M$
- 交: $L \cap M$
- 反转: $L^R = \{ w^R \mid w \in L \}$
- 同态: $h(L) = \{ h(w) \mid w \in L, \text{同态}\, h : \Sigma \to \Gamma^* \}$
- 逆同态: $h^{-1}(L) = \{ w \in \Sigma^* \mid h(w) \in L \subseteq \Gamma^*, \text{同态}\, h : \Sigma \to \Gamma^* \}$

定理 6 (并/连接/闭包的封闭性)

正则语言在并, 连接和闭包运算下保持封闭.

证明: 由正则表达式的定义得证. $\square$

定理 7 (补运算封闭性)

如果 L 是 Σ 上的正则语言, 那么 $\overline{L} = \Sigma^* - L$ 也是正则的.

证明: 设接受语言 L 的 DFA

$$A = (Q, \Sigma, \delta, q_0, F)$$

即 $\mathbf{L}(A) = L$. 构造 DFA

$$B = (Q, \Sigma, \delta, q_0, Q - F)$$

则有 $\overline{L} = \mathbf{L}(B)$, 因为 $\forall w \in \Sigma^*$

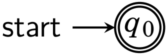
$$w \in \overline{L} \iff \hat{\delta}(q_0, w) \notin F \iff \hat{\delta}(q_0, w) \in Q - F \iff w \in \mathbf{L}(B).$$

$\square$

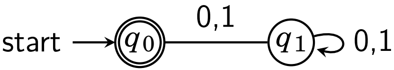
注意

使用这种方法求正则语言的补时, DFA 不能有缺失状态.

例 8. 若 $\Sigma = \{0, 1\}$, $L = \{\varepsilon\}$ 的 DFA 如图, 请给出 $\overline{L}$ 的 DFA.



应使用完整的 DFA 去求补:



思考题

如何求正则表达式的补?

例 9. 证明 $L_{\text{neq}} = \{ w \mid w$ 由数量不相等的 0 和 1 构成 $\}$ 不是正则的.

证明:

- 由泵引理不易直接证明 L_{neq} 不是正则的;
- 因为无论如何取 w, 将其分为 $w = xyz$ 时, 都不易产生 L_{neq} 之外的串;
- 而证明 L_{eq} 非正则很容易;
- 由补运算的封闭性, 所以 $L_{\text{neq}} = \overline{L_{\text{eq}}}$ 也不是正则的. □

定理 8

若 DFA A_L, A_M 和 A 的定义如下

$$\begin{aligned}
A_L &= \left(Q_L, \Sigma, \delta_L, q_L, F_L\right) \\
A_M &= \left(Q_M, \Sigma, \delta_M, q_M, F_M\right) \\
A &= \left(Q_L \times Q_M, \Sigma, \delta, (q_L, q_M), F_L \times F_M\right)
\end{aligned}$$

其中

$$\begin{aligned}
\delta : \left(Q_L \times Q_M\right) \times \Sigma \to Q_L \times Q_M \\
\delta\left((p, q), a\right) = \left(\delta_L(p, a), \delta_M(q, a)\right).
\end{aligned}$$

则对任意 $w \in \Sigma^*$,

$$\hat{\delta}\left((q_L, q_M), w\right) = \left(\hat{\delta}_L\left(q_L, w\right), \hat{\delta}_M\left(q_M, w\right)\right).$$

证明: 对 w 的结构归纳.

归纳基础：当 $w=\varepsilon$ 时

$$\begin{aligned}
\hat{\delta}\big((q_L,q_M),\varepsilon\big) &= (q_L,q_M) && \hat{\delta}\text{的定义} \\
&= \left(\hat{\delta}_L(q_L,\varepsilon),\hat{\delta}_M(q_M,\varepsilon)\right) && \text{同理}
\end{aligned}$$

归纳递推：当 $w=xa$ 时

$$\begin{aligned}
\hat{\delta}\big((q_L,q_M),xa\big) &= \delta\left(\hat{\delta}\big((q_L,q_M),x\big),a\right) && \hat{\delta}\text{的定义} \\
&= \delta\left(\left(\hat{\delta}_L(q_L,x),\hat{\delta}_M(q_M,x)\right),a\right) && \text{归纳假设} \\
&= \left(\delta_L\big(\hat{\delta}_L(q_L,x),a\big),\delta_M\big(\hat{\delta}_M(q_M,x),a\big)\right) && \delta\text{的构造} \\
&= \left(\hat{\delta}_L(q_L,xa),\hat{\delta}_M(q_M,xa)\right) && \hat{\delta}\text{的定义}
\end{aligned}$$

$\square$

定理 9 (交运算封闭性)

如果 L 和 M 是正则语言, 那么 $L \cap M$ 也是正则语言.

证明 1: 由 $L \cap M = \overline{\overline{L} \cup \overline{M}}$ 得证. □

证明 2: 由定理 8 构造识别 $L \cap M$ 的 DFA A, 则 $\forall w \in \Sigma^*$,

$$\begin{aligned} w \in L \cap M &\iff \hat{\delta}_L(q_L, w) \in F_L \wedge \hat{\delta}_M(q_M, w) \in F_M \\ &\iff \left(\hat{\delta}_L(q_L, w), \hat{\delta}_M(q_M, w)\right) \in F_L \times F_M \\ &\iff \hat{\delta}\left((q_L, q_M), w\right) \in F_L \times F_M \\ &\iff w \in \mathbf{L}(A). \end{aligned}$$

因此 $\mathbf{L}(A) = L \cap M$, 所以 $L \cap M$ 也是正则的. □

思考题

有语言 M, N 和 L, 满足

$$M \cap N = L$$

1. 若 M 和 N 都是正则语言, 由定理9, 则 L 是一定正则语言.
2. 若 M 和 N 都不是正则语言, 则 L 一定不是正则语言吗?
3. 若 M 和 L 都是正则语言, 则 N 一定是正则语言吗?
4. 若 M 正则而 L 非正则, 则 N 一定非正则吗?
5. 若 M 非正则而 L 正则, 则 N 呢?

例 10. 如果已知语言

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

不是正则的, 请用封闭性证明语言

$$L_{\text{eq}} = \{\, w \in \{0,1\}^* \mid w \text{ 由数量相等的 0 和 1 构成} \,\}$$

也不是正则的.

证明:

1. 首先, 因为 $\mathbf{0}^*\mathbf{1}^*$ 是正则语言;
2. 而 $L_{01} = \mathbf{L}(\mathbf{0}^*\mathbf{1}^*) \cap L_{\text{eq}}$;
3. 如果 L_{eq} 是正则的, L_{01} 必然也是正则的;
4. 因为已知 L_{01} 不是正则的, 所以 L_{eq} 一定不是正则的. □

思考题

为什么又能用 L_{eq} 的子集 L_{01} 是非正则的，来证明 L_{eq} 是非正则的呢？

例 11. 如果 L_1 和 L_2 都不是正则的, 那么 $L_1 \cap L_2$ 一定不是正则的吗?

不一定. 因为, 如果令

$$L_1 = \left\{ 0^n 1^n \,\middle|\, n \geq 0 \right\}$$

$$L_2 = \left\{ a^n b^n \,\middle|\, n \geq 0 \right\}$$

显然两者都不是正则语言, 但

$$L_1 \cap L_2 = \{\varepsilon\}$$

是正则语言.

定理 10 (差运算封闭性)

如果 L 和 M 都是正则语言, 那么 $L - M$ 也是正则的.

证明: $L - M = L \cap \overline{M}$. □

例 12. 证明正则语言在以下运算下封闭

$$\min(L) = \left\{ w \,\middle|\, w \text{ is in } L, \text{but no proper prefix of } w \text{ is in } L \right\}$$

例 12. 证明正则语言在以下运算下封闭

$$\min(L) = \{ w \mid w \text{ is in } L, \text{but no proper prefix of } w \text{ is in } L \}$$

证明 1:
设接受 L 的 DFA $A = (Q, \Sigma, \delta, q_0, F)$, 构造接受 $\min(L)$ 的 DFA $B = (Q, \Sigma, \delta', q_0, F)$, 其中 δ' 如下, 往证 $L(B) = \min(L)$.

$$\delta'(q, a) = \begin{cases} \delta(q, a) & \text{if } q \notin F \\ \varnothing & \text{if } q \in F \end{cases}$$

1. $\forall w \in L(B)$, 存在状态转移序列 $q_0, q_1, \cdots q_{n-1}, q_n \in F$ 使 B 接受 w, 其中 $q_i \notin F\ (0 \le i \le n-1)$. $\therefore w \in \min(L)$.
2. $\forall w \in \min(L) \subseteq L$, A 接受 w 的状态序列如果为 $q_0, q_1, \cdots, q_n$, 则显然 $q_i \notin F\ (0 \le i \le n-1)$ 且 $q_n \in F$, 否则 w 会有 L 可接受的前缀. $\therefore w \in L(B)$.

$\square$

例 12. 证明正则语言在以下运算下封闭

$$\min(L) = \left\{ w \,\middle|\, w \text{ is in } L, \text{but no proper prefix of } w \text{ is in } L \right\}$$

证明 2:

由封闭性

$$\min(L) = L - L\Sigma^{+}$$

得证. $\square$

定义

字符串 $w = a_1 a_2 \ldots a_n$ 的反转, 记为 w^R, 定义为

$$w^R = a_n a_{n-1} \ldots a_1.$$

定义

语言 L 的反转, 记为 L^R, 定义为

$$L^R = \left\{ w^R \in \Sigma^* \,\middle|\, w \in L \right\}.$$

定理 11 (反转的封闭性)

如果 L 是正则语言, 那么 L^R 也是正则的.

两种证明方法:

- 对正则表达式 E 的结构归纳, 往证
$$\mathbf{L}(E^R) = (\mathbf{L}(E))^R.$$
- 构造识别 L 的 NFA $A = (Q, \Sigma, \delta_A, q_0, F)$, 将其转换为识别 L^R 的 NFA
$$B = (Q \cup \{q_s\}, \Sigma, \delta_B, q_s, \{q_0\})$$
  1. 将 A 的边调转方向;
  2. 将 A 的初始状态 q_0, 改为唯一的接受状态;
  3. 新增初始状态 q_s, 且令 $\delta_B(q_s, \varepsilon) = F$.

例 13. 语言 L 及其反转 L^R 分别为

$$L = \left\{ w \in \{0,1\}^* \;\middle|\; w \text{ ends in } 01. \right\}$$

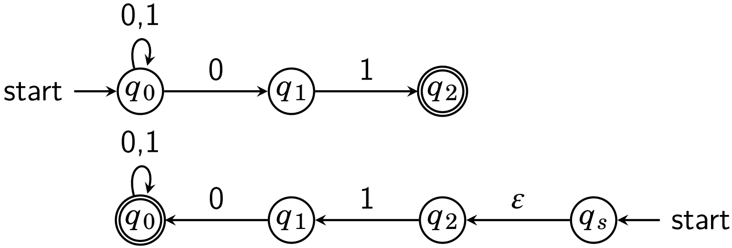
$$L^R = \left\{ w \in \{0,1\}^* \;\middle|\; w \text{ starts with } 10. \right\}$$

正则表达式分别为

$$L = (\mathbf{0}+\mathbf{1})^*\mathbf{01}$$

$$L^R = \mathbf{10}(\mathbf{0}+\mathbf{1})^*.$$

自动机分别为



证明: 往证如果有正则表达式 E, 则存在正则表达式 E^R 使

$$\mathbf{L}(E^R) = (\mathbf{L}(E))^R.$$

归纳基础:

1. 当 $E = \emptyset$ 时, 有 $\emptyset^R = \emptyset$;
2. 当 $E = \boldsymbol{\varepsilon}$ 时, 有 $\boldsymbol{\varepsilon}^R = \boldsymbol{\varepsilon}$;
3. $\forall a \in \Sigma$, 当 $E = \mathbf{a}$ 时, 有 $\mathbf{a}^R = \mathbf{a}$;

都满足 $\mathbf{L}(E^R) = (\mathbf{L}(E))^R$, 因此命题成立.

归纳递推:

1. 当 $E = E_1 + E_2$ 时, 有 $(E_1 + E_2)^R = E_1^R + E_2^R$

$$
\begin{aligned}
& \left(\mathbf{L}(E_1 + E_2)\right)^R \\
&= \left(\mathbf{L}(E_1) \cup \mathbf{L}(E_2)\right)^R && \text{正则表达式的加} \\
&= \left\{ w^R \mid w \in \mathbf{L}(E_1) \ \text{or} \ w \in \mathbf{L}(E_2) \right\} && \text{语言的反转} \\
&= \left(\mathbf{L}(E_1)\right)^R \cup \left(\mathbf{L}(E_2)\right)^R && \text{同上} \\
&= \mathbf{L}\left(E_1^R\right) \cup \mathbf{L}\left(E_2^R\right) && \text{归纳假设} \\
&= \mathbf{L}\left(E_1^R + E_2^R\right) && \text{正则表达式的加}
\end{aligned}
$$

归纳递推:

1. 当 $E = E_1 + E_2$ 时，有 $(E_1 + E_2)^R = E_1^R + E_2^R$
2. 当 $E = E_1 E_2$ 时，有 $(E_1 E_2)^R = E_2^R E_1^R$

$$
\begin{aligned}
\left(\mathbf{L}(E_1E_2)\right)^R &= \left(\mathbf{L}(E_1)\mathbf{L}(E_2)\right)^R && \text{正则表达式的连接} \\
&= \{\, w_1w_2 \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2) \,\}^R && \text{语言的连接} \\
&= \{\, (w_1w_2)^R \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2) \,\} && \text{语言的反转} \\
&= \{\, w_2^R w_1^R \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2) \,\} && \text{字符串的反转} \\
&= \{\, w_2^R \mid w_2 \in \mathbf{L}(E_2) \,\}\{\, w_1^R \mid w_1 \in \mathbf{L}(E_1) \,\} && \text{语言的连接} \\
&= \left(\mathbf{L}(E_2)\right)^R \left(\mathbf{L}(E_1)\right)^R && \text{语言的反转} \\
&= \mathbf{L}(E_2^R)\mathbf{L}(E_1^R) = \mathbf{L}(E_2^R E_1^R) && \text{正则表达式的连接}
\end{aligned}
$$

归纳递推:

❶ 当 $E = E_1 + E_2$ 时, 有 $(E_1 + E_2)^R = E_1^R + E_2^R$

❷ 当 $E = E_1 E_2$ 时, 有 $(E_1 E_2)^R = E_2^R E_1^R$

❸ 当 $E = E_1^*$ 时, 有 $(E_1^*)^R = (E_1^R)^*$

$$
\begin{aligned}
&\left(\mathbf{L}(E_1^*)\right)^R \\
&= \left\{ w_1 w_2 \dots w_n \mid n \ge 0, w_i \in \mathbf{L}(E_1) \right\}^R && \text{正则表达式的闭包} \\
&= \left\{ (w_1 w_2 \dots w_n)^R \mid n \ge 0, w_i \in \mathbf{L}(E_1) \right\} && \text{语言的反转} \\
&= \left\{ w_n^R w_{n-1}^R \dots w_1^R \mid n \ge 0, w_i \in \mathbf{L}(E_1) \right\} && \text{字符串的反转} \\
&= \left\{ w_n^R w_{n-1}^R \dots w_1^R \mid n \ge 0, w_i^R \in \mathbf{L}(E_1^R) \right\} && \text{归纳假设} \\
&= \left\{ w_1 w_2 \dots w_n \mid n \ge 0, w_i \in \mathbf{L}(E_1^R) \right\} && \text{变量重命名} \\
&= \mathbf{L}\left((E_1^R)^*\right) && \text{正则表达式的闭包}
\end{aligned}
$$

都满足 $(\mathbf{L}(E))^R = \mathbf{L}(E^R)$, 因此命题成立, 所以 L^R 也是正则语言. □

同态

定义

若 Σ 和 Γ 是两个字母表, 同态定义为函数 $h : \Sigma \to \Gamma^*$

$$\forall a \in \Sigma,\ h(a) \in \Gamma^*.$$

扩展 h 的定义到字符串,

$$\begin{aligned} &(1) \quad h(\varepsilon) = \varepsilon \\ &(2) \quad h(xa) = h(x)h(a) \end{aligned}$$

再扩展 h 到语言, 对 $\forall L \subseteq \Sigma^*$,

$$h(L) = \left\{ h(w) \,\middle|\, w \in L \right\}.$$

例 14. 若由 $\Sigma=\{0,1\}$ 到 $\Gamma=\{a,b\}$ 的同态函数 h 为

$$h(0)=ab,\quad h(1)=\varepsilon.$$

则 Σ 上的字符串 0011, 在 h 的作用下

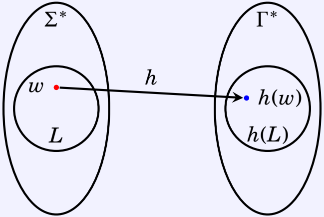
$$\begin{aligned}h(0011)&=h(\varepsilon)h(0)h(0)h(1)h(1)\\&=\varepsilon\cdot ab\cdot ab\cdot\varepsilon\cdot\varepsilon\\&=abab.\end{aligned}$$

语言 $L=\mathbf{1}^*\mathbf{0}+\mathbf{0}^*\mathbf{1}$, 在 h 的作用下, $h(L)$ 为:

$$\begin{aligned}h(\mathbf{1}^*\mathbf{0}+\mathbf{0}^*\mathbf{1})&=\big(h(\mathbf{1})\big)^*h(\mathbf{0})+\big(h(\mathbf{0})\big)^*h(\mathbf{1})\\&=(\boldsymbol{\varepsilon})^*(\mathbf{ab})+(\mathbf{ab})^*(\boldsymbol{\varepsilon})\\&=(\mathbf{ab})^*\end{aligned}$$

定理 12 (同态的封闭性)

若 L 是字母表 Σ 上的正则语言, h 是 Σ 上的同态, 则 $h(L)$ 也是正则的.



- 若 L 的正则表达式为 E, 即 $L = \mathbf{L}(E)$, 按如下规则构造表达式 $h(E)$

$$
\begin{aligned}
h(\emptyset) &= \emptyset & \qquad h(\mathbf{r}+\mathbf{s}) &= h(\mathbf{r}) + h(\mathbf{s}) \\
h(\boldsymbol{\varepsilon}) &= \boldsymbol{\varepsilon} & h(\mathbf{rs}) &= h(\mathbf{r})h(\mathbf{s}) \\
\forall a \in \Sigma,\ h(\mathbf{a}) &= h(a) & h(\mathbf{r}^*) &= \big(h(\mathbf{r})\big)^*
\end{aligned}
$$

- 往证 $\mathbf{L}\big(h(E)\big) = h\big(\mathbf{L}(E)\big)$, 而 $h(E)$ 显然也是正则表达式, 因此 $h(L)$ 正则

证明: 对 E 的结构归纳, 往证 $\mathbf{L}\big(h(E)\big) = h\big(\mathbf{L}(E)\big)$.

归纳基础:

- 当 $E = \boldsymbol{\varepsilon}$ 时

$$h\big(\mathbf{L}(\boldsymbol{\varepsilon})\big) = h\big(\{\varepsilon\}\big) = \{\varepsilon\} = \mathbf{L}(\boldsymbol{\varepsilon}) = \mathbf{L}\big(h(\boldsymbol{\varepsilon})\big)$$

- 当 $E = \varnothing$ 时

$$h\big(\mathbf{L}(\varnothing)\big) = h(\varnothing) = \varnothing = \mathbf{L}(\varnothing) = \mathbf{L}\big(h(\varnothing)\big)$$

- $\forall a \in \Sigma$, 当 $E = \mathbf{a}$ 时

$$h\big(\mathbf{L}(\mathbf{a})\big) = h\big(\{a\}\big) = \big\{h(a)\big\} = \mathbf{L}\big(h(a)\big) = \mathbf{L}\big(h(\mathbf{a})\big)$$

所以命题成立.

归纳递推: 假设对正则表达式 F, G 分别有

$$\mathbf{L}\big(h(F)\big) = h\big(\mathbf{L}(F)\big), \quad \mathbf{L}\big(h(G)\big) = h\big(\mathbf{L}(G)\big)$$

- 当 $E = F + G$ 时:

$$\begin{aligned}
h\big(\mathbf{L}(F+G)\big) &= h\big(\mathbf{L}(F) \cup \mathbf{L}(G)\big) && \text{正则表达式的加} \\
&= h\big(\mathbf{L}(F)\big) \cup h\big(\mathbf{L}(G)\big) && h\,\text{作用在每个集合的串上} \\
&= \mathbf{L}\big(h(F)\big) \cup \mathbf{L}\big(h(G)\big) && \text{归纳假设} \\
&= \mathbf{L}\big(h(F) + h(G)\big) && \text{正则表达式的加} \\
&= \mathbf{L}\big(h(F+G)\big) && h(F+G)\,\text{的定义}
\end{aligned}$$

- 当 $E = FG$ 时: 略
- 当 $E = F^*$ 时: 略

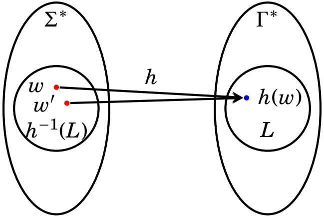
$\square$

逆同态

定义

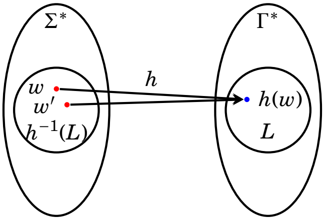
若 h 是字母表 Σ 到 Γ 的同态, 且 L 是 Γ 上的语言, 那么使 $h(w) \in L$ 的 w $(w \in \Sigma^*)$ 的集合, 称为语言 L 的 h 逆, 记为 $h^{-1}(L)$, 即

$$h^{-1}(L) = \{\, w \in \Sigma^* \mid h(w) \in L \,\}.$$



定理 13 (逆同态的封闭性)

如果 h 是字母表 Σ 到 Γ 的同态, L 是 Γ 上的正则语言,
那么 $h^{-1}(L)$ 也是正则语言.



证明: 由 L 的 DFA $A = \big(Q, \Gamma, \delta, q_0, F\big)$, 构造识别 $h^{-1}(L)$ 的 DFA

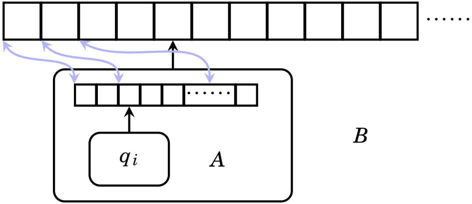
$$B = \big(Q, \Sigma, \delta', q_0, F\big),$$

证明: 由 L 的 DFA $A=(Q,\Gamma,\delta,q_0,F)$, 构造识别 $h^{-1}(L)$ 的 DFA

$$B=(Q,\Sigma,\delta',q_0,F),$$

其中

$$\delta'(q,a)=\hat{\delta}(q,h(a)).$$



为证明 $\mathbf{L}(B)=h^{-1}(L)$, 先证明 $\widehat{\delta'}(q,w)=\hat{\delta}(q,h(w))$.

对 $|w|$ 归纳, 往证 $\widehat{\delta'}(q,w) = \hat{\delta}(q,h(w))$.

1. 归纳基础: 若 $w = \varepsilon$

$$\hat{\delta}(q,h(\varepsilon)) = \hat{\delta}(q,\varepsilon) = q = \widehat{\delta'}(q,\varepsilon),$$

2. 归纳递推: 若 $w = xa$

$$\begin{aligned}
\widehat{\delta'}(q,xa) &= \delta'(\widehat{\delta'}(q,x),a) && \widehat{\delta'}\text{ 定义} \\
&= \delta'(\hat{\delta}(q,h(x)),a) && \text{归纳假设} \\
&= \hat{\delta}(\hat{\delta}(q,h(x)),h(a)) && \delta'\text{ 构造} \\
&= \hat{\delta}(q,h(x)h(a)) && \text{DFA 节例 5} \\
&= \hat{\delta}(q,h(xa)).
\end{aligned}$$

所以 $\forall w \in \Sigma^*$, $\widehat{\delta'}(q_0,w) = \hat{\delta}(q_0,h(w)) \in F$, 即 w 被 B 接受当且仅当 $h(w)$ 被 A 接受, B 是识别 $h^{-1}(L)$ 的 DFA, 因此 $h^{-1}(L)$ 是正则的. □

例 15. Prove that $L = \{ 0^n 1^{2n} \mid n \geq 0 \}$ is a language not regular.

证明: 设同态 $h : \{0,1\} \to \{0,1\}^*$ 为

$$h(0) = 0,$$

$$h(1) = 11,$$

那么

$$h^{-1}(L) = \{ 0^n 1^n \mid n \geq 0 \} = L_{01},$$

我们已知 L_{01} 非正则, 由封闭性, L 不是正则的. □

例 16. 若语言 $L = (\mathbf{00} + \mathbf{1})^*$, 同态 $h : \{a, b\} \to \{0, 1\}^*$ 为

$$h(a) = 01, \quad h(b) = 10,$$

请证明 $h^{-1}(L) = (\mathbf{ba})^*$.

证明: 往证 $h(w) \in L \iff w = (ba)^n$.

($\Leftarrow$) 若 $w = (ba)^n$, 而 $h(ba) = 1001$, 因此 $h(w) = (1001)^n \in L$.

($\Rightarrow$) 若 $h(w) \in L$, 假设 $w \notin (\mathbf{ba})^*$, 则只能有四种情况:

1. w 以 a 开头, 则 $h(w)$ 以 01 开头, 显然 $h(w) \notin (\mathbf{00} + \mathbf{1})^*$;
2. w 以 b 结尾, 则 $h(w)$ 以 10 结尾, 显然 $h(w) \notin (\mathbf{00} + \mathbf{1})^*$;
3. w 有连续的 a, 即 $w = xaay$, 则 $h(w) = z1010v$, 显然 $h(w) \notin (\mathbf{00} + \mathbf{1})^*$;
4. w 有连续的 b, 即 $w = xbby$, 则 $h(w) = z0101v$, 显然 $h(w) \notin (\mathbf{00} + \mathbf{1})^*$;

因此 w 只能是 $(ba)^n, n \geq 0$ 的形式. □

例 17. 证明正则语言在以下运算下封闭

$$\max(L) = \left\{ w \,\middle|\, w \text{ is in } L \text{ and for no } x \text{ other than } \varepsilon \text{ is } wx \text{ in } L \right\}.$$

证明 1:

由接受 L 的 DFA $M = (Q, \Sigma, \delta, q_0, F)$, 构造接受 $\max(L)$ 的 DFA $M' = (Q, \Sigma, \delta, q_0, F')$, 其中

$$F' = F - \{q \mid q \in F,\ \exists x \in \Sigma^+,\ \hat{\delta}(q, x) \in F\}$$

则 $L(M') = \max(L)$. $\square$

注意: $\min(L)$ 和 $\max(L)$ 不同, 设 $L = \mathbf{10}^+$, 则 $\min(L) = \{10\}$ 而 $\max(L) = \varnothing$.

例 17. 证明正则语言在以下运算下封闭

$$\max(L) = \left\{ w \,\middle|\, w \text{ is in } L \text{ and for no } x \text{ other than } \varepsilon \text{ is } wx \text{ in } L \right\}.$$

证明 2: 利用封闭性.

如果 $\Sigma = \{a, b\}$, 设 $\Gamma = \{a, \hat{a}, b, \hat{b}\}$, 定义同态 h $(\Gamma \to \Sigma^*)$ 和 g $(\Gamma \to \Sigma^*)$:

$$\begin{array}{llll} h(a) = a & h(\hat{a}) = a & g(a) = a & g(\hat{a}) = \varepsilon \\ h(b) = b & h(\hat{b}) = b & g(b) = b & g(\hat{b}) = \varepsilon \end{array}$$

那么

$$\max(L) = L - g\left(h^{-1}(L) \cap (\mathbf{a} + \mathbf{b})^* (\hat{\mathbf{a}} + \hat{\mathbf{b}})^+\right).$$

$\square$

注意: $\min(L)$ 和 $\max(L)$ 不同, 设 $L = \mathbf{10}^+$, 则 $\min(L) = \{10\}$ 而 $\max(L) = \varnothing$.

例 18. For a language L, define head(L) to be the set of all prefixes of strings in L. Prove that if L is regular, so is head(L).

证明. 设 L 是 Σ 上的正则语言且 $\Sigma = \{a, b\}$, $\Gamma = \{a, b, \hat{a}, \hat{b}\}$.
定义同态 $h : \Gamma \to \Sigma^*$ 和 $g : \Gamma \to \Sigma^*$ 分别为:

$$\begin{aligned} h(a) &= a & h(\hat{a}) &= a & g(a) &= a & g(\hat{a}) &= \varepsilon \\ h(b) &= b & h(\hat{b}) &= b & g(b) &= b & g(\hat{b}) &= \varepsilon \end{aligned}$$

则因为 $(\mathbf{a}+\mathbf{b})^*(\hat{\mathbf{a}}+\hat{\mathbf{b}})^*$ 是 Γ 上的正则语言, 所以

$$\mathrm{head}(L) = g\big((\mathbf{a}+\mathbf{b})^*(\hat{\mathbf{a}}+\hat{\mathbf{b}})^* \cap h^{-1}(L)\big)$$

是 Σ 上的正则语言, 因此 $\mathrm{head}(L)$ 是正则的. □

正则语言的性质

- 证明语言的非正则性
- 正则语言的封闭性
- 正则语言的判定性质
  - 空性，有穷性和无穷性
  - 等价性
- 自动机的最小化

正则语言的判定性质

正则语言, 或任何语言, 典型的 3 个判定问题:

1. 以某种形式化模型描述的语言是否为空? 是否无穷?
2. 某个特定的串 w 是否属于所描述的语言?
3. 以两种方式描述的语言, 是否是相同的? — 语言的等价性

我们想知道, 要回答这类问题的具体算法, 是否存在.

空性, 有穷性和无穷性

定理 14

具有 n 个状态的有穷自动机 M 接受的集合 S:

1. S 是非空的, 当且仅当 M 接受某个长度小于 n 的串;
2. S 是无穷的, 当且仅当 M 接受某个长度为 m 的串, $n \le m < 2n$.

所以, 对于正则语言:

- 存在算法, 判断其是否为空, 只需检查全部长度小于 n 的串;
- 存在算法, 判断其是否无穷, 只需检查全部长度由 n 到 $2n-1$ 的串.

证明: 设接受正则语言 S 的 DFA 为 A.

1. 必要性: 显然成立. 充分性:
   1. 如果 S 非空, 设 w 是 A 接受的串中长度最小者之一;
   2. 必然 $|w| < n$, 否则由泵引理 $w = xyz$, 接受 xz 更短.
2. 必要性: 由泵引理, 显然成立. 充分性:
   1. 如果 S 无穷, 假设没有长度 n 到 $2n-1$ 之间的串;
   2. 那么取 $w \in \mathbf{L}(A)$ 是长度 $\geq 2n$ 中最小者之一;
   3. 由泵引理 $w = xyz$, 且 A 会接受更短的串 xz;
   4. 于是, 或者 w 不是长度最小的, 或者长度 n 到 $2n-1$ 之间有被接受的串, 因此假设不成立. $\square$

正则语言的等价性

定理 15

存在算法, 判定两个有穷自动机是否等价(接受语言相同).

证明:

1. 设 M_1 和 M_2 是分别接受 L_1 和 L_2 的有穷自动机;
2. 则 $(L_1 \cap \overline{L_2}) \cup (\overline{L_1} \cap L_2)$ 是正则的, 所以可被某个有穷自动机 M_3 接受;
3. 而 M_3 接受某个串, 当且仅当 $L_1 \neq L_2$;
4. 由于存在算法判断 $\mathbf{L}(M_3)$ 是否为空, 因此得证. □

正则语言的性质

- 证明语言的非正则性
- 正则语言的封闭性
- 正则语言的判定性质
- 自动机的最小化
  - DFA 状态的等价性
  - 填表算法与 DFA 最小化

DFA 状态的等价性

定义

DFA $A = (Q, \Sigma, \delta, q_0, F)$ 中两个状态 p 和 q, 对 $\forall w \in \Sigma^*$:

$$\hat{\delta}(p, w) \in F \Leftrightarrow \hat{\delta}(q, w) \in F,$$

则称这两个状态是等价的, 否则称为可区分的.

- 等价性只要求 $\hat{\delta}(p, w)$ 和 $\hat{\delta}(q, w)$ 同时在或不在 F 中, 而不必相同.

填表算法

递归寻找 DFA 中全部的可区分状态对:

1. 如果 $p \in F$ 而 $q \notin F$, 则 $[p,q]$ 是可区分的;
2. $\exists a \in \Sigma$, 如果
$$\left[r = \delta(p,a), s = \delta(q,a)\right]$$
是可区分的, 则 $[p,q]$ 是可区分的.

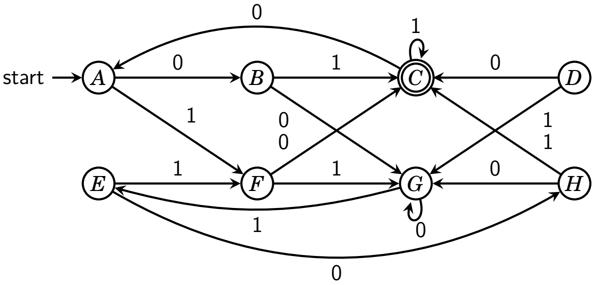
定理 16

如果填表算法不能区分两个状态, 则这两个状态是等价的.

Algorithm 1 MinimizeDFA$(Q, \Sigma, \delta, q_0, F)$

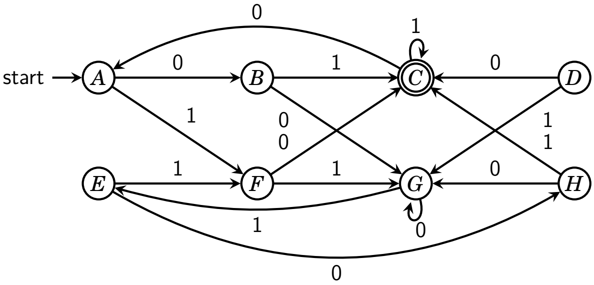
```
for all (p, q) ∈ Q × Q do
    if (p ∈ F and q ∉ F) or (p ∉ F and q ∈ F) then
        T[p, q] ← ★
repeat
    done ← True
    for all (p, q) ∈ Q × Q do
        if T[p, q] ≠ ★ then
            for all a ∈ Σ do
                if T[δ(p, a), δ(q, a)] = ★ then
                    T[p, q] ← ★
                    done ← False
until done
return T
```

例 19. 用填表算法找到如图 DFA 中全部可区分状态对.



B							
C							
D							
E							
F							
G							
H							
	A	B	C	D	E	F	G

例 19. 用填表算法找到如图 DFA 中全部可区分状态对.

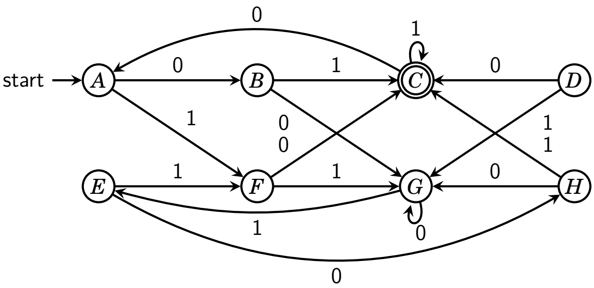


	A	B	C	D	E	F	G
B							
C	×	×					
D			×				
E			×				
F			×				
G			×				
H			×				

❶ 直接标记终态和非终态之间的状态对:

$$\{C\} \times \{A, B, D, E, F, G, H\}.$$

例 19. 用填表算法找到如图 DFA 中全部可区分状态对.

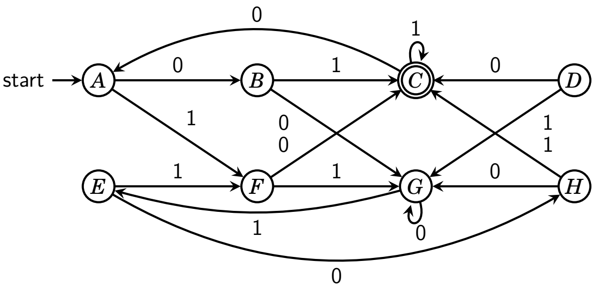


B							
C	×	×					
D	×	×	×				
E			×	×			
F	×	×	×		×		
G			×	×		×	
H			×	×		×	
	A	B	C	D	E	F	G

2. 标记所有经过字符 0 到达终态和非终态的状态对:

$$\{D, F\} \times \{A, B, C, E, G, H\}.$$

例 19. 用填表算法找到如图 DFA 中全部可区分状态对.

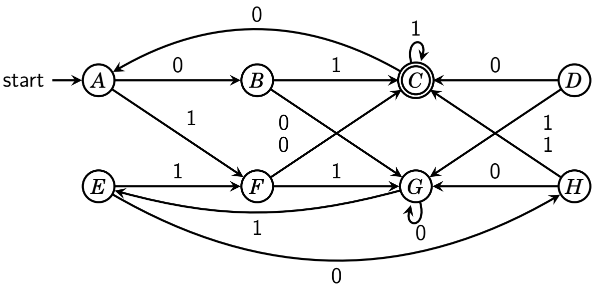


	A	B	C	D	E	F	G
B	×						
C	×	×					
D	×	×	×				
E		×	×	×			
F	×	×	×		×		
G		×	×	×		×	
H	×		×	×	×	×	×

❸ 标记所有经过字符 1 到达终态和非终态的状态对:

$$\{B, H\} \times \{A, C, D, E, F, G\}.$$

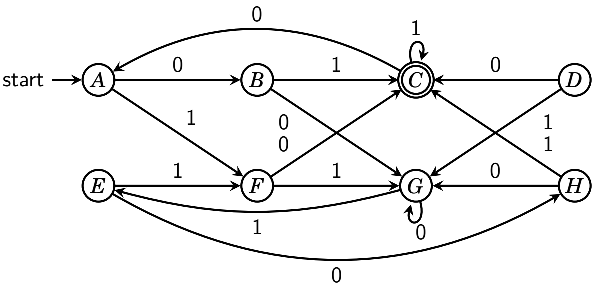
例 19. 用填表算法找到如图 DFA 中全部可区分状态对.



	A	B	C	D	E	F	G
B	×						
C	×	×					
D	×	×	×				
E		×	×	×			
F	×	×	×		×		
G		×	×	×		×	
H	×		×	×	×	×	×

4. 此时还有 $[A,E]$, $[A,G]$, $[B,H]$, $[D,F]$, $[E,G]$ 未标记, 只需逐个检查.

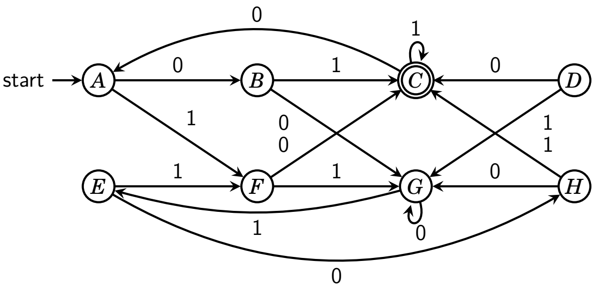
例 19. 用填表算法找到如图 DFA 中全部可区分状态对.



B	×						
C	×	×					
D	×	×	×				
E		×	×	×			
F	×	×	×		×		
G	×	×	×	×	×	×	
H	×		×	×	×	×	×
	A	B	C	D	E	F	G

4. 此时还有 $[A,E]$, $[A,G]$, $[B,H]$, $[D,F]$, $[E,G]$ 未标记, 只需逐个检查.
   - × $[A,G]$ 是可区分的, 因为字符 0 到可区分的 $[B,G]$;
   - × $[E,G]$ 是可区分的, 因为字符 1 到可区分的 $[E,F]$.

例 19. 用填表算法找到如图 DFA 中全部可区分状态对.



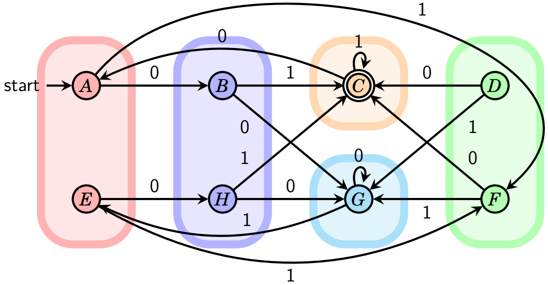
B	×						
C	×	×					
D	×	×	×				
E		×	×	×			
F	×	×	×		×		
G	×	×	×	×	×	×	
H	×		×	×	×	×	×
	A	B	C	D	E	F	G

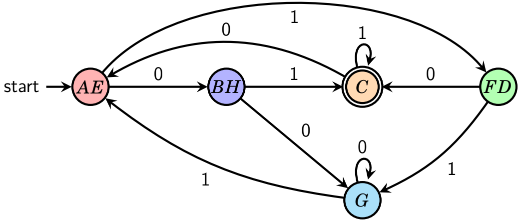
5 而 $[A,E]$, $[B,H]$ 和 $[D,F]$ 在经过很短的字符串后, 都会到达相同状态, 因此都是等价的.

DFA 最小化

根据等价状态, 将状态集划分成块, 构造等价的最小化 DFA.

续例 19. 构造其最小化的 DFA.

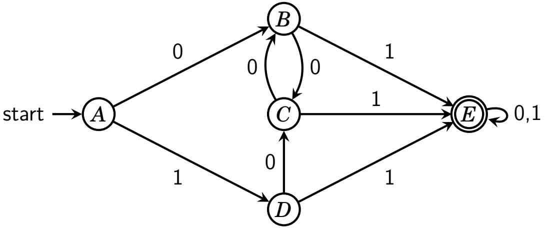




思考题

NFA 能否最小化?

课堂练习. Minimize the given DFA.



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

chunyu@hit.edu.cn
http://nclab.net/~chunyu