

# 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

# 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,  
称正则语言在这些运算下封闭.

正则语言  $L$  和  $M$ , 在这些运算下封闭

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

### 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  $\Sigma$  上的正则语言, 那么  $\bar{L} = \Sigma^* - L$  也是正则的.

证明:

设接受语言  $L$  的 DFA

$$A = (Q, \Sigma, \delta, q_0, F) \quad \text{被 A 接收就不会被 B 接收}$$

即  $L(A) = L$ . 构造 DFA

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

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

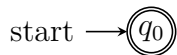
$$w \in \bar{L} \iff \hat{\delta}(q_0, w) \notin F \iff \hat{\delta}(q_0, w) \in Q - F \iff w \in L(B). \quad \square$$

## 注意

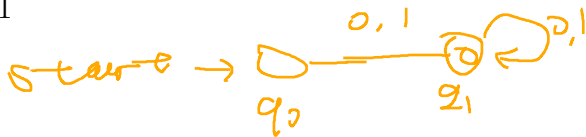
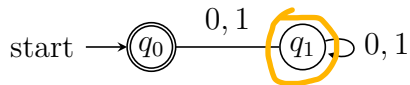
使用这种方法求正则语言的补时, DFA 不能有缺失状态.

例 8.

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



应使用完整的 DFA 去求补:



$$\overline{\{\varepsilon\}} = \Sigma^+$$

## 思考题

如何求正则表达式的补?

1.  $R \rightarrow \text{DFA}$
2. complete DFA
3. find DFA's complement
4. convert to regular expression

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

- 由泵引理不易直接证明  $L_{neq}$  不是正则的
- 因为无论如何取  $w$ , 将其分为三部分  $w = xyz$  时, 都不易产生  $L_{neq}$  之外的串
- 而证明  $L_{eq} = \overline{L_{neq}}$  非正则很容易
- 由补运算的封闭性, 所以  $L_{neq}$  也不是正则的

### 定理 8

若 DFA  $A_L$ ,  $A_M$  和  $A$  的定义如下

$$A_L = (Q_L, \Sigma, \delta_L, q_L, F_L)$$

$$A_M = (Q_M, \Sigma, \delta_M, q_M, F_M)$$

$$A = (\underbrace{Q_L \times Q_M}, \Sigma, \underbrace{\delta, (q_L, q_M)}, \underbrace{F_L \times F_M})$$

其中

$$\delta : (Q_L \times Q_M) \times \Sigma \rightarrow Q_L \times Q_M$$

$$\underbrace{\delta((p, q), a) = (\delta_L(p, a), \delta_M(q, a))}.$$

则对任意  $w \in \Sigma^*$ ,

$$\hat{\delta}((q_L, q_M), w) = (\hat{\delta}(q_L, w), \hat{\delta}(q_M, w)).$$



证明: 通过对  $w$  的归纳来证明.

归纳基础: 当  $w = \varepsilon$  时

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

归纳递推: 当  $w = xa$  时

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

### 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.  $\square$

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

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

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

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

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

也不是正则的.

证明:

- ① 首先, 因为  $0^*1^*$  是正则语言;
- ② 而  $L_{01} = L(0^*1^*) \cap L_{eq}$ ;
- 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.  $\square$

### 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

$$L_1 = \{0^n 1^n \mid n \geq 0\}$$

$$L_2 = \{a^n b^n \mid n \geq 0\}$$

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:  $L - M = L \cap \overline{M}$ .  $\square$

---

## 定义

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

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

语言  $L$  的反转, 记为  $L^R$ , 定义为

$$L^R = \{w^R \in \Sigma^* \mid w \in L\}.$$

### 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

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

- 由识别  $L$  的 DFA  $A = (Q, \Sigma, \delta_A, q_0, F)$ , 构造识别  $L^R$  的  $\varepsilon$ -NFA

$$B = (Q, \Sigma, \delta_B, q_s, \{q_0\})$$

- ① 将  $A$  的初始状态  $q_0$ , 改为唯一的接受状态;
- ② 将  $A$  的边调转方向: 如果  $\delta_A(q, a) = p$ , 那么  $\delta_B(p, a) = q$ ;
- ③ 新增初始状态  $q_s$ , 且令  $\delta_B(q_s, \varepsilon) = F$ ;
- ④ 往证  $\mathbf{L}(B) = L^R$ .



例 12. 语言  $L$  及其反转  $L^R$  分别为

$$L = \{w \in \{0, 1\}^* \mid w \text{ ends in } 01.\}$$

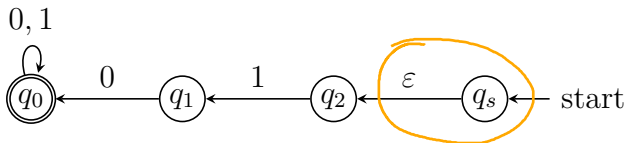
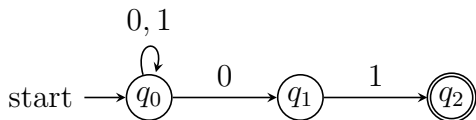
$$L^R = \{w \in \{0, 1\}^* \mid w \text{ starts with } 10.\}$$

正则表达式分别为

$$L = (0 + 1)^* 01$$

$$L^R = 10(0 + 1)^*.$$

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

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

归纳基础:

- ① 当  $E = \emptyset$  时, 有  $\emptyset^R = \emptyset$ ;
- ② 当  $E = \epsilon$  时, 有  $\epsilon^R = \epsilon$ ;
- 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  $(E_1 + E_2)^R = E_1^R + E_2^R$

$$\underline{(\mathbf{L}(E_1 + E_2))^R}$$

$$= (\mathbf{L}(E_1) \cup \mathbf{L}(E_2))^R$$

$$= \{w^R \mid w \in \mathbf{L}(E_1) \cup w \in \mathbf{L}(E_2)\}$$

$$= (\mathbf{L}(E_1))^R \cup (\mathbf{L}(E_2))^R$$

$$= \mathbf{L}(E_1^R) \cup \mathbf{L}(E_2^R)$$

$$= \underline{\mathbf{L}(E_1^R + E_2^R)}$$

正则表达式的加

语言的反转

同上

归纳假设

正则表达式的加

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  $(E_1 E_2)^R = E_2^R E_1^R$

$$(\mathbf{L}(E_1 E_2))^R = (\mathbf{L}(E_1) \mathbf{L}(E_2))^R$$

$$= \{w_1 w_2 \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2)\}^R$$

$$= \{(w_1 w_2)^R \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2)\}$$

$$= \{w_2^R w_1^R \mid w_1 \in \mathbf{L}(E_1), w_2 \in \mathbf{L}(E_2)\}$$

$$= \{w_2^R \mid w_2 \in \mathbf{L}(E_2)\} \{w_1^R \mid w_1 \in \mathbf{L}(E_1)\}$$

$$= (\mathbf{L}(E_2))^R (\mathbf{L}(E_1))^R$$

$$= \mathbf{L}(E_2^R) \mathbf{L}(E_1^R) = \mathbf{L}(E_2^R E_1^R)$$

正则表达式的连接

语言的连接

语言的反转

字符串的反转

语言的连接

语言的反转

正则表达式的连接

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  $(E_1 E_2)^R = E_2^R E_1^R$
- ③ 当  $E = E_1^*$  时, 有  $(E_1^*)^R = (E_1^R)^*$

$$(\mathbf{L}(E_1^*))^R$$

$$= \{w_1 w_2 \dots w_n \mid n \geq 0, w_i \in \mathbf{L}(E_1)\}^R$$

正则表达式的闭包

$$= \{(w_1 w_2 \dots w_n)^R \mid n \geq 0, w_i \in \mathbf{L}(E_1)\}$$

语言的反转

$$= \{w_n^R w_{n-1}^R \dots w_1^R \mid n \geq 0, w_i \in \mathbf{L}(E_1)\}$$

字符串的反转

$$= \{w_n^R w_{n-1}^R \dots w_1^R \mid n \geq 0, w_i^R \in \mathbf{L}(E_1^R)\}$$

归纳假设

$$= \{w_1 w_2 \dots w_n \mid n \geq 0, w_i \in \mathbf{L}(E_1^R)\}$$

变量重命名

$$= \mathbf{L}((E_1^R)^*)$$

正则表达式的闭包

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.  $\square$

# 同态

## 定义

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

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

扩展  $h$  的定义到字符串,

$$(1) \quad h(\varepsilon) = \varepsilon$$

$$(2) \quad h(xa) = \underline{h(x)h(a)}$$

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

$$\boxed{h(L) = \{h(w) \mid w \in L\}.}$$

例 13. 若由  $\Sigma = \{0, 1\}$  到  $\Gamma = \{a, b\}$  的同态函数  $h$  为

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

则  $\Sigma$  上的字符串 0011, 在  $h$  的作用下

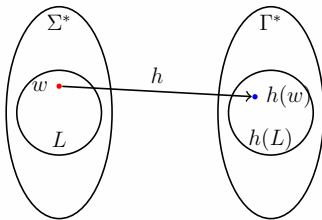
$$\begin{aligned} h(0011) &= h(0)h(0)h(1)h(1) \\ &= abab\varepsilon\varepsilon = abab. \end{aligned}$$

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

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

### 定理 12 (同态的封闭性)

若  $L$  是字母表  $\Sigma$  上的正则语言,  $h$  是  $\Sigma$  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  $h(E)$

$$h(\emptyset) = \emptyset$$

$$h(\varepsilon) = \varepsilon$$

$$\forall a \in \Sigma, h(\mathbf{a}) = h(a)$$

$$h(\mathbf{r} + \mathbf{s}) = h(\mathbf{r}) + h(\mathbf{s})$$

$$h(\mathbf{rs}) = h(\mathbf{r})h(\mathbf{s})$$

$$h(\mathbf{r}^*) = (h(\mathbf{r}))^*$$

- 往证  $\mathbf{L}(h(E)) = h(\mathbf{L}(E))$ , 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  $\mathbf{L}(h(E)) = h(\mathbf{L}(E))$ .

归纳基础:

- 当  $E = \varepsilon$  时

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

- 当  $E = \emptyset$  时

$$h(\mathbf{L}(\emptyset)) = h(\emptyset) = \emptyset = \mathbf{L}(\emptyset) = \mathbf{L}(h(\emptyset))$$

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

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

所以命题成立.

归纳递推: 假设对正则表达式  $F, G$  分别有

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

- 当  $E = F + G$  时:

$$\begin{aligned} h(\mathbf{L}(F + G)) &= h(\mathbf{L}(F) \cup \mathbf{L}(G)) \\ &= h(\mathbf{L}(F)) \cup h(\mathbf{L}(G)) \\ &= \mathbf{L}(h(F)) \cup \mathbf{L}(h(G)) \\ &= \mathbf{L}(h(F) + h(G)) \\ &= \mathbf{L}(h(F + G)) \end{aligned}$$

正则表达式的加

$h$  作用在每个集合的串上

归纳假设

正则表达式的加

$h(F + G)$  的定义

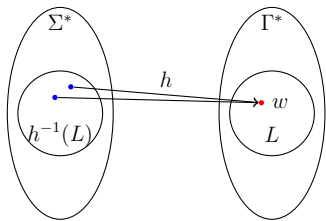
- 当  $E = FG$  时: 略
- 当  $E = F^*$  时: 略  $\square$

## 逆同态

### 定义

若  $h$  是字母表  $\Sigma$  到  $\Gamma$  的同态, 且  $L$  是  $\Gamma$  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

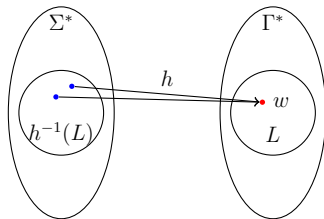
$$h^{-1}(L) = \{w \in \Sigma^* \mid h(w) \in L\}.$$



$$h^{-1}(L) = \{h(w) \mid w \in L\}$$

### 定理 13 (逆同态的封闭性)

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

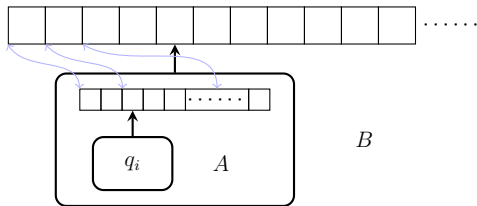


证明: 由  $L$  的 DFA  $A = (Q, \Gamma, \delta, q_0, F)$ , 构造识别  $h^{-1}(L)$  的 DFA

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

其中

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



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

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

① 归纳基础: 若  $w = \varepsilon$

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

② 归纳递推: 若  $w = xa$

$$\hat{\delta}'(q, xa) = \delta'(\hat{\delta}'(q, x), a) \quad \delta' \text{ 定义}$$

$$= \delta'(\hat{\delta}(q, h(x)), a) \quad \text{归纳假设}$$

$$= \hat{\delta}(\hat{\delta}(q, h(x)), h(a)) \quad \delta' \text{ 构造}$$

$$= \hat{\delta}(q, h(x)h(a)) \quad \text{DFA 节例 5}$$

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

所以  $\forall w \in \Sigma^*, \hat{\delta}'(q_0, w) = \hat{\delta}(q_0, h(w)) \in F$ , 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.  $\square$

例 14. Prove that  $L = \{0^n 1^{2n} \mid n \geq 0\}$  is a language not regular.

证明: 设同态  $h: \{0, 1\}^* \rightarrow \{0, 1\}^*$  为

$$\begin{cases} w = h^{-1}(L) \\ L = h(w) \end{cases} \quad (\psi) \quad h(w) =$$

~~$$h(w) =$$~~

$$\begin{cases} h(0) = 0, \\ h(1) = 11, \end{cases}$$

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.  $\square$

$$w = 0^a 1^b, \Rightarrow h(w) = 0^a 1^{2b}$$

if  $h(w) \in L$

$$a = b = N, \text{ so}$$

$$w = 0^n 1^n$$

$$h^{-1}(L) = 0^n 1^n = L_{01}$$

- ①  $h(w) \in L \Rightarrow h^{-1}(L)$
- ② strings in  $L$   $0 \dots 01 \dots 1$
- ③  $h(0) = 0, h(1) = 11$

例 15. 若语言  $L = (00 + 1)^*$ , 同态  $h : \{a, b\} \rightarrow \{0, 1\}^*$  为

$$w = (ba)^n \quad h(a) = 01, \quad h(b) = 10,$$

请证明  $h^{-1}(L) = (ba)^*$ .

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  $w \notin (ba)^*$ , 则只能有四种情况:

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

因此  $w$  只能是  $(ba)^n, n \geq 0$  的形式.  $\square$