正则语言的性质

RL的性质

- **RL**性质
  - 泵引理及其应用
  - 并、乘积、闭包、补、交
  - 正则代换、同态、逆同态的封闭性
- 从RL固有特征寻求表示的一致性
  - Myhill-Nerode定理
  - FA的极小化
- RL的几个判定问题
  - 空否、有穷否、两个DFA等价否、成员关系

RL的泵引理

- **泵引理(pumping lemma)**

  设**L**为一个 **RL** ，则存在仅依赖于**L**的正整数**N**，对于∀**z**∈**L**，如果|**z**|≥**N**，则存在**u**、**v**、**w**，满足

  (1) $z=uvw$；

  (2) $|uv|\leq N$；

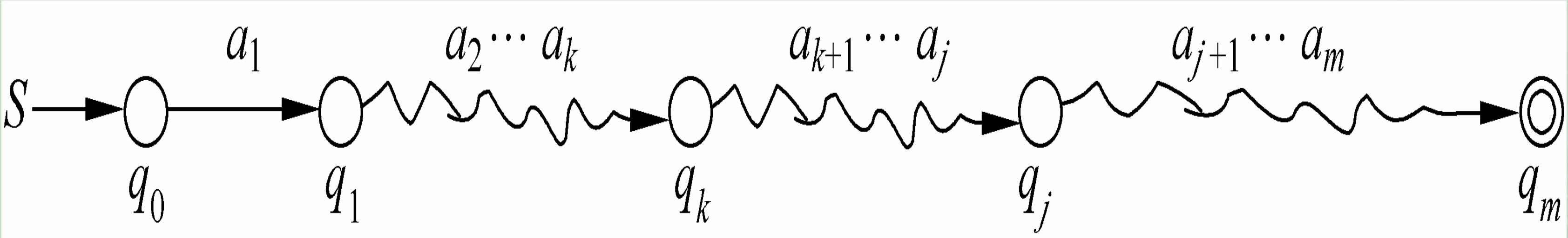
  (3) $|v|\geq 1$；

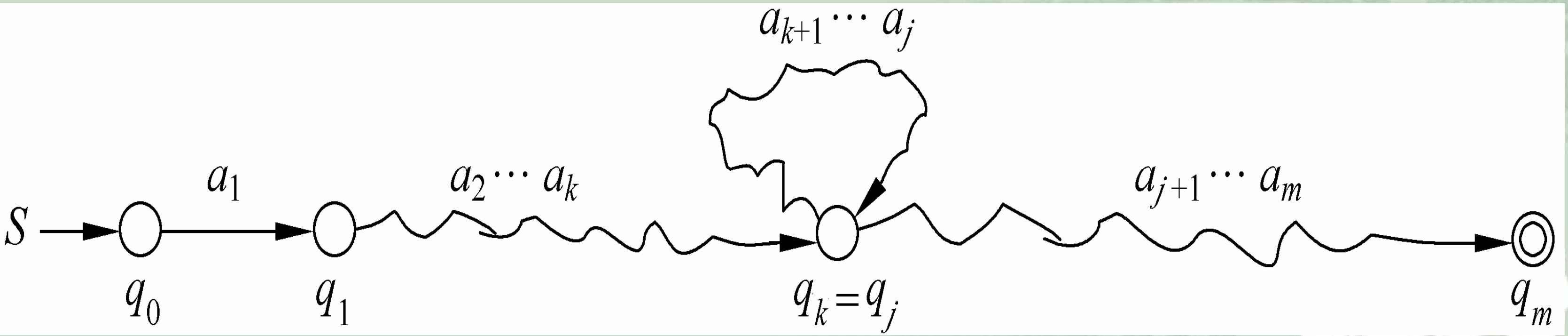
  (4) 对于任意的整数$i\geq 0$，$uv^iw\in L$；

  (5) **N**不大于接受**L**的最小**DFA M**的状态数。

RL的泵引理

- 证明思想


S
a_1
$a_2 \cdots a_k$
$a_{k+1} \cdots a_j$
$a_{j+1} \cdots a_m$
q_0
q_1
q_k
q_j
q_m


$a_{k+1} \cdots a_j$
S
a_1
$a_2 \cdots a_k$
$a_{j+1} \cdots a_m$
q_0
q_1
$q_k = q_j$
q_m

RL的泵引理

证明：

DFA在处理一个足够长的句子的过程中，必定会重复地经过某一个状态。换句话说，在DFA的状态转移图中，必定存在一条含有回路的从启动状态到某个终止状态的路。由于是回路，所以，DFA可以根据实际需要沿着这个回路循环运行，相当于这个回路中弧上的标记构成的非空子串可以重复任意多次。

RL的泵引理

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

$|Q|=N$

$z= a_1a_2\ldots a_m \quad m \geqslant N$

$\delta(q_0, a_1a_2\ldots a_h)=q_h$

状态序列q_0，q_1，…，q_N中，至少有两个状态是相同：$q_k=q_j$

RL的泵引理

$\delta(q_0, a_1a_2\ldots a_k)=q_k$

$\delta(q_k, a_{k+1}\ldots a_j)=q_j=q_k$

$\delta(q_j, a_{j+1}\ldots a_m)=q_m$

对于任意的整数$i\geqslant 0$

$\delta(q_k, (a_{k+1}\ldots a_j)^i)$

$=\delta(q_k, (a_{k+1}\ldots a_j)^{i-1})$

$\ldots$

$=\delta(q_k, a_{k+1}\ldots a_j)=q_k$

RL的泵引理

故，

$\delta(q_0，a_1a_2\ldots a_k(a_{k+1}\ldots a_j)^i a_{j+1}\ldots a_m)=q_m$

也就是说，

$a_1a_2\ldots a_k(a_{k+1}\ldots a_j)^i a_{j+1}\ldots a_m \in L(M)$

$u= a_1a_2\ldots a_k$，

$v=a_{k+1}\ldots a_j$，

$w= a_{j+1}\ldots a_m$

$uv^iw \in L$

RL的泵引理

- **例** 证明$\{0^n1^n|n\geq1\}$不是 RL。

  证明：

  假设$L=\{0^n1^n|n\geq1\}$ 是 RL

  $z=0^N1^N$

  按照泵引理所述

  $v=0^k \qquad k\geq1$

  此时有，

  $u=0^{N-k-j}$

  $w=0^j1^N$

RL的泵引理

从而有，

$uv^iw=0^{N-k-j}(0^k)^i0^j1^N=0^{N+(i-1)k}1^N$

当i=2时，我们有：

$uv^2w=0^{N+(2-1)k}1^N = 0^{N+k}1^N$

注意到k≥1，所以，

N+k>N。

这就是说，

$0^{N+k}1^N \notin L$

这与泵引理矛盾。所以，L不是 RL。

RL的泵引理

- **例** 证明$\{0^n|n$为素数$\}$不是 RL。

证明：假设$L=\{0^n|n$为素数$\}$ 是 RL。

取 $z=0^{N+p} \in L$，

不妨设$v=0^k$，$k \geqslant 1$

从而有，

$$uv^iw=0^{N+p-k-j}(0^k)^i0^j$$

$$=0^{N+p+(i-1)k}$$

RL的泵引理

当$i=N+p+1$时，

$$N+p+(i-1)k=N+p+(N+p+1-1)k$$
$$= N+p+(N+p)k$$
$$= (N+p)(1+k)$$

注意到$k\geq1$，所以

$N+p+(N+p+1-1)k=(N+p)(1+k)$

不是素数。故当$i=N+p+1$时，

$uv^{N+p+1}w=0^{(N+p)(1+k)} \notin L$

这与泵引理矛盾。所以，L不是 RL。

RL的泵引理

- **例** 证明$\{0^n1^m2^{n+m}|m，n\geq 1\}$不是 RL。

证明：假设$L=\{0^n1^m2^{n+m}|m，n\geq 1\}$ 是 RL。

取$z=0^N1^N2^{2N}$

设$v=0^k \quad k\geq 1$

从而有，

$$uv^iw=0^{N-k-j}(0^k)^i0^j1^N2^{2N}$$

$$=0^{N+(i-1)k}1^N2^{2N}$$

RL的泵引理

$uv^0w=0^{N+(0-1)k}1^N2^{2N}$

$= 0^{N-k}1^N2^{2N}$

注意到$k\geqslant 1$，

$N-k+N=2N-k<2N$

$0^{N-k}1^N2^{2N} \notin L$

这个结论与泵引理矛盾。所以，L不是 RL。

RL的泵引理

- **用来证明一个语言不是 RL**
- **不能用泵引理去证明一个语言是 RL。**

(1) 由于泵引理给出的是 **RL** 的必要条件，所以，在用它证明一个语言不是 **RL** 时，我们使用反证法。

(2) 泵引理说的是对 **RL** 都成立的条件，而我们是要用它证明给定语言不是 **RL**，这就是说，相应语言的“仅仅依赖于**L**的正整数**N**”实际上是不存在的。所以，我们一定是无法给出一个具体的数的。因此，人们往往就用符号N来表示这个“假定存在”、而实际并不存在的数。

RL的泵引理

(3) 由于泵引理指出，如果**L**是 **RL** ，则对任意的**z**∈**L**，只要|**z**|≥**N**，一定会存在**u**、**v**、**w**，使**uv**^{i}**w**∈**L**对所有的**i**成立。因此，我们在选择**z**时，就需要注意到论证时的简洁和方便。

(4) 当一个特意被选来用作“发现矛盾”的**z**确定以后，就必须依照|**uv**|≤**N**和|**v**|≥**1**的要求，说明**v**不存在(对“存在**u**、**v**、**w**”的否定) 。

RL的泵引理

(5) 与选z时类似，在寻找i时，我们也仅需要找到一个表明矛盾的“具体”值就可以了(对“所有i”的否定)。

(6) 一般地，在证明一个语言不是 RL 的时候，我们并不使用泵引理的第(5)条。

(7) 事实上，引理所要求的|uv|≤N并不是必须的。这个限制为我们简化相应的证明提供了良好支撑——扩充了的泵引理 。

RL的封闭性

- **封闭性(closure property)**

  如果任意的、属于同一语言类的语言在某一特定运算下所得的结果仍然是该类语言，则称该语言类对此运算是**封闭**的。

- **有效封闭性(valid closure property)**

  给定一个语言类的若干个语言的描述，如果存在一个算法，它可以构造出这些语言在给定运算下所获得的运算结果的相应形式的语言描述，则称此语言类对相应的运算是**有效封闭**的。

RL的封闭性

定理 RL 在并、乘积、闭包运算下是封闭的。

- 根据RE的定义，立即可以得到此定理。

RL的封闭性

定理　RL 在补运算下是封闭的。

证明：

$M=(Q, \Sigma, \delta, q_0, F)$　$L(M)=L$，

取DFA　$M' = (Q, \Sigma, \delta, q_0, Q-F)$

显然，对于任意的$x \in \Sigma^*$，

$\delta(q_0, x)=f \in F \Leftrightarrow \delta(q_0, x)=f \notin Q-F$

即：　$x \in L(M) \Leftrightarrow x \notin L(M')$，

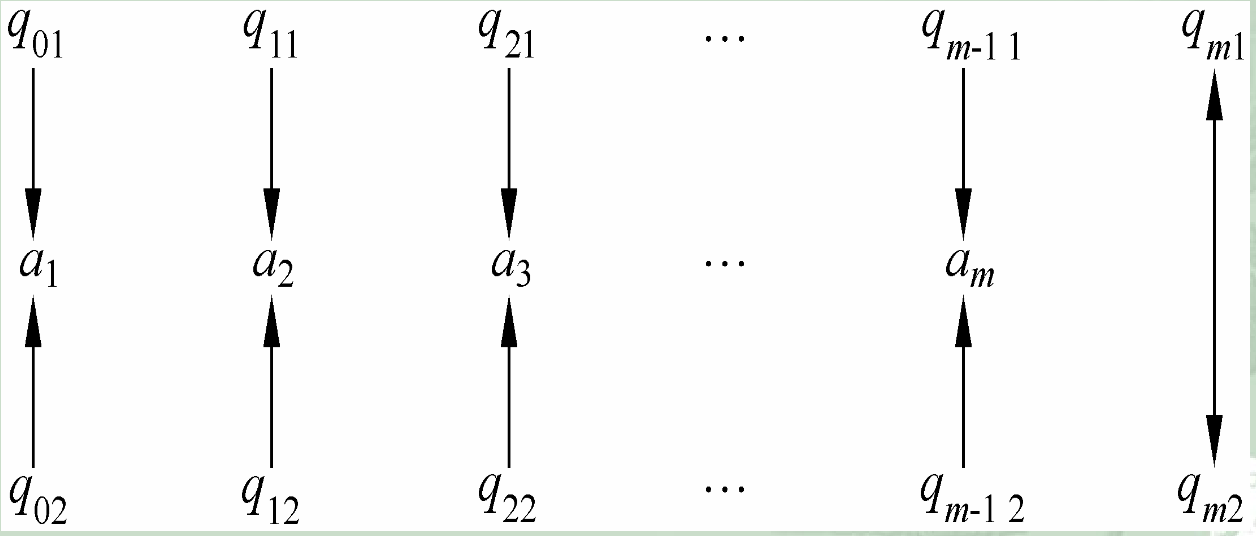
$L(M')= \Sigma^*-L(M)$。

所以，　RL 在补运算下是封闭的。定理得到证明。

RL的封闭性

定理 **RL** 在交运算下封闭。

证明思路



RL的封闭性

- **正则代换(regular substitution)**

设Σ、Δ是两个字母表，映射

$$f:\Sigma \to 2^{\Delta^*}$$

被称为是从Σ到Δ的**代换**。如果对于 ∀**a**∈Σ，**f(a)**是Δ上的 RL ，则称f为**正则代换**。

RL的封闭性

- 先将**f**的定义域扩展到Σ*上：

$$f:\Sigma^* \rightarrow 2^{\Delta^*}$$

(1) f(ε)={ε}；

(2) f(xa)=f(x)f(a)。

RL的封闭性

- 再将**f**的定义域扩展到 2^{Σ^*}

$$f : 2^{\Sigma^*} \to 2^{\Delta^*}$$

对于∀**L**⊆Σ *

$$f(L) = \bigcup_{x \in L} f(x)$$

RL的封闭性

- 例 设Σ={0，1}，Δ={a，b}，f(0)=a，f(1)=b^*，则

  $f(010)=f(0)f(1)f(0)=ab^*a$

  $f(\{11，00\})=f(11)\cup f(00)$

  $=f(1)f(1)\cup f(0)f(0)=b^*b^*+aa=b^*+aa$

RL的封闭性

- f是正则代换, 则

  (1) $f(\Phi)=\Phi$；

  (2) $f(\varepsilon)=\varepsilon$；

  (3) 对于$\forall a\in\Sigma$，$f(a)$是Δ上的RE;

  (4) 如果r，s是Σ上的RE，则

  $f(r+s)=f(r)+f(s)$

  $f(rs)=f(r)f(s)$

  $f(r^*)=f(r)^*$

  是Δ上的RE。

RL的封闭性

定理 设L是Σ上的一个 RL，

$$f : \Sigma \to 2^{\Delta^*}$$

是正则代换，则f(L)也是 RL。

证明：

描述工具：RE。

对r中运算符的个数n施以归纳，证明f(r)是表示f(L)的RE。

RL的封闭性

- 当**n=0**时，结论成立。
- 当**n**≤**k**时定理成立, 即当r中运算符的个数不大于k时：$f(L(r)) = L(f(r))$ 。
- 当**n=k+1**时,

RL的封闭性

(1) $r=r_1+r_2$

$f(L)=f(L(r))$

$=f(L(r_1+r_2))$

$=f(L(r_1)\cup L(r_2))$ RE的定义

$=f(L(r_1))\cup f(L(r_2))$ 正则代换的定义

$=L(f(r_1))\cup L(f(r_2))$ 归纳假设

$=L(f(r_1)+f(r_2))$ RE的定义

$=L(f(r_1+r_2))$ RE的正则代换的定义

$=L(f(r))$

RL的封闭性

（2） $r=r_1r_2$。

$f(L)=f(L(r))$

$=f(L(r_1r_2))$

$=f(L(r_1)\,L(r_2))$ RE的定义

$=f(L(r_1))\,f(L(r_2))$ 正则代换的定义

$=L(f(r_1))\,L\,(f\,(r_2))$ 归纳假设

$=L(f(r_1)\,f\,(r_2))$ RE的定义

$=L(f(r_1r_2))$ RE的正则代换的定义

$=L(f(r))$

RL的封闭性

（3） $r=r_1^*$。

$f(L)=f(L(r))$

$=f(L(r_1^*))$

$=f(L(r_1)^*)$　　RE的定义

$=(f(L(r_1)))^*$　　正则代换的定义

$=(L(f(r_1)))^*$　　归纳假设

$=L(f(r_1)^*)$　　RE的定义

$=L(f(r_1^*))$　　RE的正则代换的定义

$=L(f(r))$

RL的封闭性

例　设 $\Sigma=\{0，1\}$，$\Delta=\{a，b\}$，$f(0)=a$，$f(1)=b^*$，则

$f(L(0^*(0+1)1^*))$

$=L(a^*(a+b^*)(b^*)^*)$

$=L(a^*(a+b^*)b^*)=L(a^*ab^*+a^*b^*b^*)$

$=L(a^*b^*)$

RL的封闭性

- **例** 设Σ={0，1，2}，Δ={a，b}，正则代换f定义为：

  $f(0)=ab$；

  $f(1)=b^*a^*$；

  $f(2)=a^*(a+b)$

则：

(1) $f(00)=abab$；

(2) $f(010)=abb^*a^*ab=ab^+a^+b$；

RL的封闭性

(3) $f((0+1+2)^*)=(ab+b^*a^*+ a^*(a+b))^*$

$=(b^*a^*+a^*(a+b))^*=(a+b)^*$；

(4) $f(0(0+1+2)^*)=ab(ab+b^*a^*+ a^*(a+b))^*$

$=ab(a+b)^*$；

(5) $f(012)=abb^*a^* a^* (a+b)= ab^+a^*(a+b)$;

(6) $f((0+1)^*)=(ab+ b^*a^*)^*$

$=(a+b)^*$ 。

RL的封闭性

- **同态映射(homomorphism)**

  设Σ、Δ是两个字母表，

$$f:\Sigma \rightarrow \Delta^*$$

f为映射，如果对于∀**x**，**y**∈Σ*，

f(xy)=f(x)f(y)，

则称**f**为从Σ到Δ*的**同态映射**。

RL的封闭性

- 对于$\forall L \subseteq \Sigma^*$，L的同态像

$$f(L)=\bigcup_{x\in L} f(x)$$

- 对于$\forall L \subseteq \Delta^*$，L的同态原像

$$f^{-1}(L)=\{x \mid f(x)\in L\}$$

RL的封闭性

- **例** 设Σ={0，1}，Δ={a，b}，同态映射f定义为

  $f(0)=aa$

  $f(1)=aba$

则：

(1) $f(01)=aaaba$；

(2) $f((01)^*)=(aaaba)^*$；

(3) $f^{-1}(aab)=\Phi$；

RL的封闭性

(4) $f^{-1}(aa)=\{0\}$；

(5) $f^{-1}(\{aaa，aba，abaaaaa，abaaaaaa\})$

$=\{1，100\}$；

(6) $f^{-1}((ab+ba)^*a)=\{1\}$；

(7) $f(f^{-1}((ab+ba)^*a))=f(\{1\})=\{aba\}$。

令$L=(ab+ba)^*a$，上述(7)表明，$f(f^{-1}(L)) \neq L$

$f(f^{-1}(L)) \subseteq L$

RL的封闭性

推论 RL 的同态像是 **RL**。

- 证明：

注意到同态映射是正则代换的特例，可以直接得到此结论。

该定理表明，**RL** 在同态映射下是封闭的。

RL的封闭性

定理　RL 的同态原像是 RL 。

证明：

使用**DFA**作为描述工具。

(1) 接受**RL**的同态原像的**FA**的构造思想。

让新构造出的**FA M′**用一个移动去模拟**M**处理**f(a)**所用的一系列移动。

对于∑中的任意字符**a**，如果**M**从状态**q**开始处理**f(a)**，并且当它处理完**f(a)**时到达状态**p**，则让**M′**在状态**q**读入**a**时，将状态变成**p**。

RL的封闭性

M′具有与M相同的状态，并且，在M′对应的状态转移图中，从状态q到状态p有一条标记为a的弧当且仅当在M的状态转移图中，有一条从状态q到状态p的标记为f(a)的路。

(2) 接受RL的同态原像的FA的形式描述。

设DFA M=(Q，Δ，δ，q_0，F)，L(M)=L，

取DFA M′=(Q，Σ，δ′，q_0，F)

δ′(q，a)= δ(q，f(a))

RL的封闭性

(3) 等价证明。

- 施归纳于|**x**|，证明对于∀**x**∈Σ*，

$$\delta'(q_0, x)=\delta(q_0, f(x))$$

当|**x**|=**0**时，结论显然成立。

设当|x|=k是结论成立，往证当|x|=k+1时结论成立。

不妨设x=ya，其中|y|=k

RL的封闭性

$\delta'(q_0, x) = \delta'(q_0, ya)$

$= \delta'(\delta'(q_0, y), a)$

$= \delta'(\delta(q_0, f(y)), a)$ 归纳假设

$= \delta(\delta(q_0, f(y)), f(a))$ δ' 的定义

$= \delta(q_0, f(y)f(a))$ δ 的意义

$= \delta(q_0, f(ya))$ 同态映射性质

$= \delta(q_0, f(x))$

RL的封闭性

这表明，结论对$|x|=k+1$成立。由归纳法原理，结论对$\forall x\in\Sigma^*$成立。

$\forall \mathbf{x}\in\Sigma^*$， $\delta'(\mathbf{q_0}, \mathbf{x})\in \mathbf{F} \Leftrightarrow \delta(\mathbf{q_0}, \mathbf{f(x)})\in \mathbf{F}$。

由于对$\forall x\in\Sigma^*$， $\delta'(q_0, x)=\delta(q_0, f(x))$，

所以，

$\delta'(q_0, x)\in F \Leftrightarrow \delta(q_0, f(x))\in F$。

故

$L(M')=f^{-1}(L(M))$

定理得证。

RL的封闭性

- **商(quotient)**
- 设L_1、$L_2 \subseteq \Sigma^*$， L_1除以L_2的**商**定义为：
  $L_1/L_2=\{x \mid \exists y \in L_2$使得$xy \in L_1\}$。
- 计算语言的商主要是考虑语言句子的后缀。只有当L_1的句子的后缀在L_2中时，其相应的前缀才属于L_1/L_2。所以，当$\varepsilon \in L_2$时，$L_1 \subseteq L_1/L_2$。

RL的封闭性

- 注意以下有意思的情况：

取$L_1=\{000\}$，$L_2=\{\varepsilon\}$，$L_3=\{\varepsilon, 0\}$

$L_4=\{\varepsilon, 0, 00\}$，$L_5=\{\varepsilon, 0, 00, 000\}$

$L_1/L_2=\{000\}=L_1$

$L_1/L_3=\{000, 00\}$

$L_1/L_4=\{000, 00, 0\}$

$L_1/L_5=\{000, 00, 0, \varepsilon\}$

RL的封闭性

定理 L_1、$L_2 \subseteq \Sigma^*$，如果L_1和L_2是 RL，则 L_1/L_2也是 RL 。

证明：设$L_1 \subseteq \Sigma^*$，是 RL ，

DFA $M=(Q, \Sigma, \delta, q_0, F)$，$L(M)=L_1$

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

$F'=\{q \mid \exists y \in L_2, \delta(q, y) \in F\}$

显然，

$L(M')= L_1/L_2$。

定理得证。

Myhill-Nerode 定理与DFA的极小化

- 对给定 **RL L**，**DFA M**接受**L**，**M**不同，由R_M确定的Σ^*上的等价类也可能不同。
- 如果**M**是最小**DFA**，则**M**所给出的等价类的个数应该是最少的。
- 最小**DFA**是不是惟一的？如果是，如何构造？
- 最小**DFA**的状态对应的集合与其他**DFA**的状态对应的集合有什么样的关系，用这种关系是否能从一般的**DFA**出发，求出最小**DFA**？

Myhill-Nerode 定理

- **DFA M确定的等价关系。**

$M=(Q, \Sigma, \delta, q_0, F)$，对于$\forall x, y \in \Sigma^*$

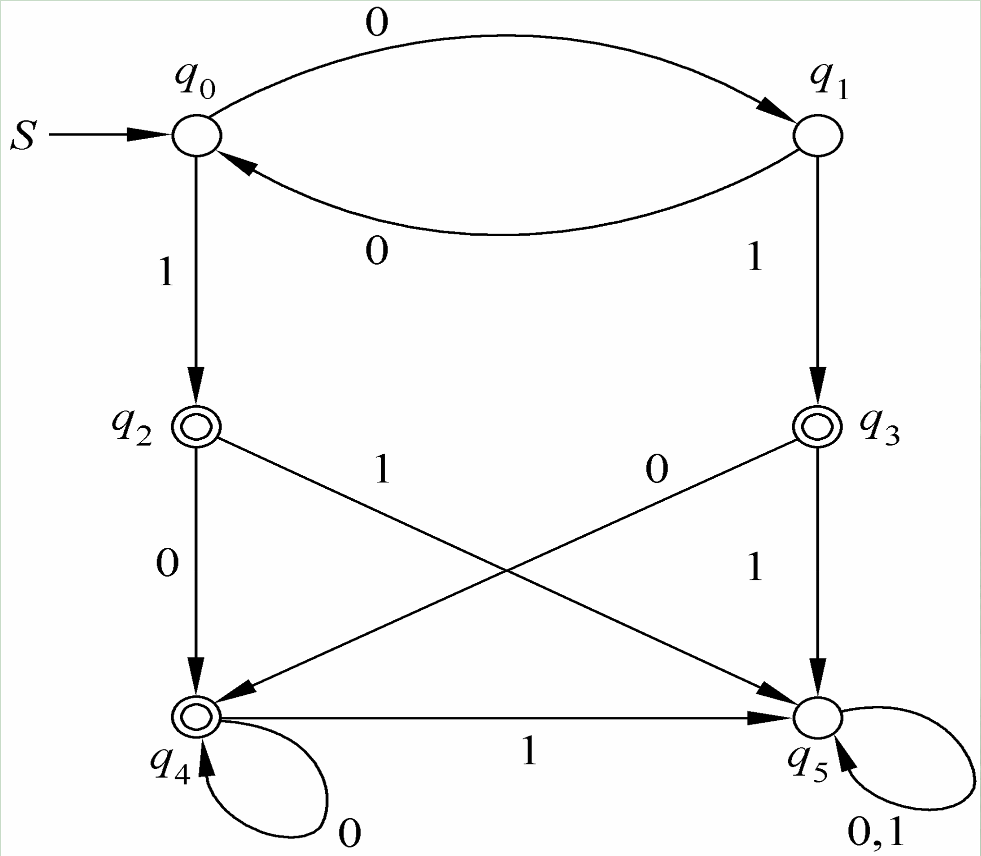
$x\, R_M\, y \Leftrightarrow \delta(q_0, x)=\delta(q_0, y)$。

显然，

$x\, R_M\, y \Leftrightarrow \exists q \in Q, x, y \in set(q)$

Myhill-Nerode 定理

- 例 设 L=0*10*，它对应的DFA M如下图。



Myhill-Nerode 定理

对应于q_0：$(00)^n R_M (00)^m$ n，$m \geqslant 0$；

对应于q_1：$0(00)^n R_M 0(00)^m$ n，$m \geqslant 0$；

对应于q_2：$(00)^n 1 R_M (00)^m 1$ n，$m \geqslant 0$；

对应于q_3：$0(00)^n 1 R_M 0(00)^m 1$ n，$m \geqslant 0$；

对应于q_4：$0(00)^n 10^k R_M 0(00)^m 10^h$ n，$m \geqslant 0$, k, $h \geqslant 1$；

$(00)^n 10^k R_M (00)^m 10^h$ n, $m \geqslant 0$， k, $h \geqslant 1$；

$0(00)^n 10^k R_M (00)^m 10^h$ n, $m \geqslant 0$， k, $h \geqslant 1$；

也就是： $0^n 10^k R_M 0^m 10^h$ n, $m \geqslant 0$， k, $h \geqslant 1$；

对应于q_5：$x R_M y$——x，y为至少含两个1的串。

Myhill-Nerode 定理

- L确定的Σ^*上的关系R_L。

  对于$\forall$x，y$\in\Sigma^*$，

  $x\ R_L\ y \Leftrightarrow$ (对$\forall z\in\Sigma^*$，$xz\in L \Leftrightarrow yz\in L$)

  对于$\forall$x，y$\in\Sigma^*$，如果$x\ R_L\ y$，则在x和y后无论接Σ^*中的任何串z，xz和yz要么都是L的句子，要么都不是L的句子。

Myhill-Nerode 定理

任意x，y∈set(q)，$\delta(q_0, x) = \delta(q_0, y) = q$

对于$\forall z \in \Sigma^*$,

$\delta(q_0, xz) = \delta(\delta(q_0, x), z)$

$= \delta(q, z)$

$= \delta(\delta(q_0, y), z)$

$= \delta(q_0, yz)$

这就是说，

$\delta(q_0, xz) \in F \Leftrightarrow \delta(q_0, yz) \in F$

Myhill-Nerode 定理

即，对于$\forall \mathbf{z} \in \Sigma^*$,

$\mathbf{xz} \in \mathbf{L} \Leftrightarrow \mathbf{yz} \in \mathbf{L}$。

表明,

$\mathbf{x}\ \mathbf{R_L}\ \mathbf{y}$,

也就是

$\mathbf{x}\ \mathbf{R_{L(M)}} \mathbf{y}$。

Myhill-Nerode 定理

- **右不变的(right Invariant)**等价关系

设**R**是Σ^*上的等价关系，对于$\forall$**x**，**y**$\in\Sigma^*$，如果**x R y**，则必有**xz R yz**对于$\forall$**z**$\in\Sigma^*$成立，则称**R**是**右不变的**等价关系。

Myhill-Nerode 定理

命题 对于任意DFA M=(Q，Σ，δ，q_0，F)，M所确定的Σ^*上的关系R_M为右不变的等价关系。

证明：

(1) R_M是等价关系。

自反性显然。

对称性：$\forall x，y \in \Sigma^*$，

$x\ R_M\ y \Leftrightarrow \delta(q_0，x)=\delta(q_0，y)$ 根据R_M的定义；

$\Leftrightarrow \delta(q_0，y)=\delta(q_0，z)$ “=”的对称性；

$\Leftrightarrow y\ R_M\ x$ 根据R_M的定义。

Myhill-Nerode 定理

传递性：设$x\ R_M\ y$，$y\ R_M\ z$。

由于$x\ R_M\ y$，　$\delta(q_0，x)=\delta(q_0，y)$

由于$y\ R_M\ z$，　　$\delta(q_0，y)=\delta(q_0，z)$

由“=”的传递性知，

$\delta(q_0，x)=\delta(q_0，z)$

再由R_M的定义得：

$x\ R_M\ z$

即R_M是等价关系。

Myhill-Nerode 定理

(2) R_M 是右不变的

设$x\ R_M\ y$。则$\delta(q_0, x)=\delta(q_0, y)=q$

所以，对于$\forall z\in\Sigma^*$,

$\delta(q_0, xz)=\delta(\delta(q_0, x), z))$

$=\delta(q, z)$

$=\delta(\delta(q_0, y), z)$

$=\delta(q_0, yz)$

这就是说，$\delta(q_0, xz)=\delta(q_0, yz)$，再由$R_M$的定义，

$xz\ R_M\ yz$

所以，R_M 是右不变的等价关系。

Myhill-Nerode 定理

命题 对于任意$L\subseteq\Sigma^*$，L所确定的Σ^*上的关系R_L为右不变的等价关系 。

证明：

(1) R_L是等价关系。

自反性显然。

对称性：不难看出：$x\ R_L\ y \Leftrightarrow$ (对$\forall z\in\Sigma^*$, $xz\in L \Leftrightarrow yz\in L) \Leftrightarrow y\ R_L\ x$

Myhill-Nerode 定理

传递性：设$x\ R_L\ y$，$y\ R_L\ z$。

$x\ R_L\ y \Leftrightarrow$ (对$\forall w \in \Sigma^*$，$xw \in L \Leftrightarrow yw \in L$)

$y\ R_L\ z \Leftrightarrow$ (对$\forall w \in \Sigma^*$，$yw \in L \Leftrightarrow zw \in L$)

所以，

$(\forall w \in \Sigma^*, xw \in L \Leftrightarrow yw \in L$ 且$yw \in L \Leftrightarrow zw \in L)$

即：

(对$\forall w \in \Sigma^*$，$xw \in L \Leftrightarrow zw \in L$)

故：

$x\ R_L\ z$

即R_L是等价关系。

Myhill-Nerode 定理

(2) R_L 是右不变的。

设 $x \ R_L \ y$。由 R_L 的定义，对 $\forall w$，$v \in \Sigma^*$，$xwv \in L \Leftrightarrow zwv \in L$，注意到 v 的任意性，知，

$xw \ R_L \ yw$。

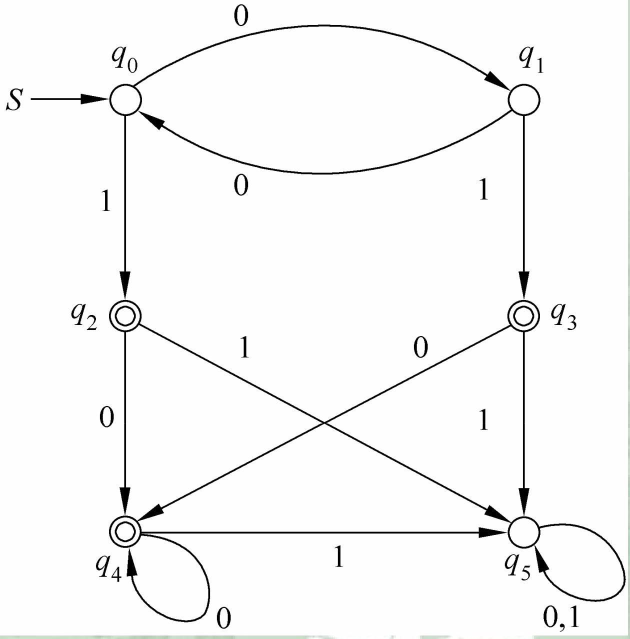
所以，R_L 是右不变的等价关系。

Myhill-Nerode 定理

- **指数(index)**
- 设R是Σ^*上的等价关系，则称$|\Sigma^*/R|$是R关于Σ^*的**指数**，简称为R的指数。简称Σ^*的关于R的一个等价类，也就是Σ^*/R的任意一个元素，为R的一个等价类 。

Myhill-Nerode 定理

- 例 图示所给DFA M所确定的R_M的指数为6。R_M将Σ^*分成6个等价类：



$set(q_0)=\{(00)^n \mid n \geqslant 0\}$；

$set(q_1)=\{0(00)^n \mid n \geqslant 0\}$；

$set(q_2)=\{(00)^n1 \mid n, \ m \geqslant 0\}$；

$set(q_3)=\{0(00)^n 1 \mid n \geqslant 0\}$；

$set(q_4)=\{0^n 10^k \mid n \geqslant 0, \ k \geqslant 1\}$；

$set(q_5)=\{x \mid x$为至少含两个1的串$\}$。

Myhill-Nerode 定理

- R_M是$R_{L(M)}$的“加细”(refinement)
  - $\forall x$，$y\in\Sigma^*$，如果$x\ R_M\ y$，必有$x\ R_{L(M)}\ y$成立。即对于任意DFA $M=(Q, \Sigma, \delta, q_0, F)$。

    $$|\Sigma^*/R_{L(M)}|\leqslant|\Sigma^*/R_M|\leqslant|Q|$$

  - 按照R_M中被分在同一等价类的串，在按照$R_{L(M)}$分类时，一定会被分在同一个等价类。
  - R_M对Σ^*的划分比$R_{L(M)}$对Σ^*的划分更“细”。称R_M是$R_{L(M)}$的“加细”(refinement)。

Myhill-Nerode 定理

- Σ^*/R_M ={set(q_0)，set(q_1)，set(q_2)，set(q_3)，set(q_4)，set(q_5)}

  (1) 取00∈set(q_0)，000∈set(q_1)。

  对于任意的x∈Σ^*，当x含且只含一个1时，00x∈L(M)，000x∈L(M)；当x不含1或者含多个1时，00x$\notin$L(M)，000x$\notin$L(M)。这就是说，对于任意的x∈Σ^*，00x∈L(M) $\Leftrightarrow$ 000x∈L(M)。即按照$R_{L(M)}$，00与000被分在同一个等价类中。从而set(q_0)和set(q_1) 被包含在$R_{L(M)}$的同一个等价类中。

Myhill-Nerode 定理

(2) 取00∈$\mathbf{set(q_0)}$，**001**∈$\mathbf{set(q_2)}$。

取特殊的字符串 **1**∈Σ^*，**001**∈**L(M)**，但 **0011**∉**L(M)**。所以，根据$\mathbf{R_{L(M)}}$，$\mathbf{set(q_0)}$和$\mathbf{set(q_2)}$不能被“合并”到一个等价类中。

类似地，根据$\mathbf{R_{L(M)}}$，$\mathbf{set(q_3)}$、$\mathbf{set(q_4)}$、$\mathbf{set(q_5)}$也都不能被“合并”到$\mathbf{set(q_0)}$的句子所在的等价类中。

Myhill-Nerode 定理

(3) 取001∈**set(q_2)**，**01**∈**set(q_3)**。

对于**任意的x**∈Σ*，**x**要么不含**1**，要么含有**1**。当**x**不含**1**时，**001x**∈**L(M)**，**01x**∈**L(M)**；当**x**含有**1**时，**001x**∉**L(M)**，**01x**∉**L(M)**。这就是说，对于任意的**x**∈Σ*，**001x**∈**L(M)** ⇔ **01x**∈**L(M)**。即按照$R_{L(M)}$，**001**与**01**属于$R_{L(M)}$的同一个等价类中。从而**set(q_2)**和**set(q_3)** 被包含在$R_{L(M)}$的同一个等价类中。

Myhill-Nerode 定理

(4) 取$1 \in set(q_2)$，$10 \in set(q_4)$。

对于任意的$x \in \Sigma^*$，x要么不含1，要么含有1。当x不含1时，$1x \in L(M)$，$10x \in L(M)$；当x含有1时，$1x \notin L(M)$，$10x \notin L(M)$。这就是说，对于任意的$x \in \Sigma^*$，$1x \in L(M) \Leftrightarrow 10x \in L(M)$。即按照$R_{L(M)}$，1与10被分在$R_{L(M)}$的同一个等价类中。从而在$set(q_2)$和$set(q_4)$被包含在$R_{L(M)}$的同一个等价类中。

Myhill-Nerode 定理

(5) 取$1 \in set(q_2)$，$11 \in set(q_5)$。

注意到$1\varepsilon=1$，$11\varepsilon=11$；而$1 \in L(M)$，$11 \notin L(M)$。即1和11不满足关系$R_{L(M)}$，所以，$set(q_2)$和$set(q_5)$不能被“合并”到$R_{L(M)}$的同一个等价类中。在这里，$\varepsilon \in \Sigma^*$是一个特殊的字符串。

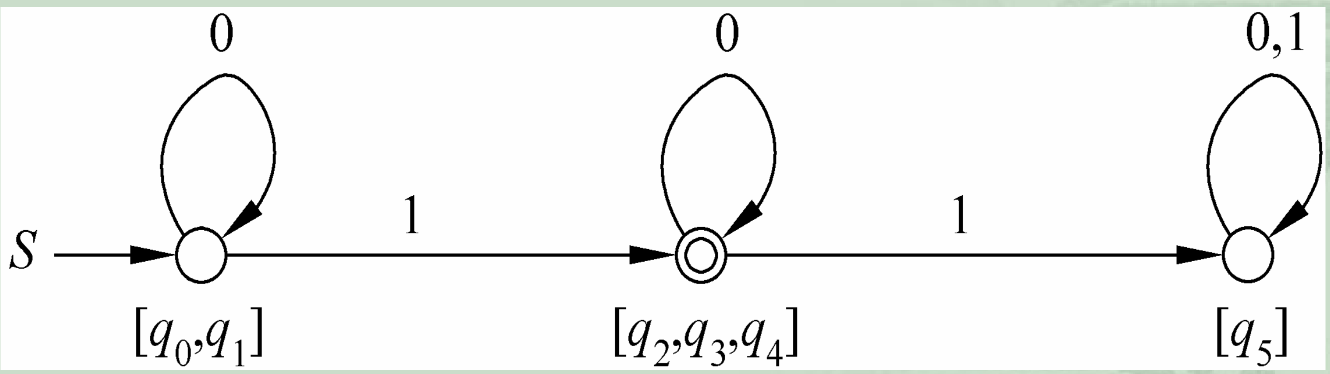
Myhill-Nerode 定理

$\Sigma^*/R_{L(M)}=\{ set(q_0)\cup set(q_1),$
$set(q_2)\cup set(q_3)\cup set(q_4), \quad set(q_5)\}$

不含1：$[0]= set(q_0)\cup set(q_1)=0^*$；

含一个1：$[1]= set(q_2)\cup set(q_3)\cup set(q_4)=0^*10^*$；

含多个1：$[11]= set(q_5)=0^*10^*1(0+1)^*$。



Myhill-Nerode 定理

定理 (Myhill-Nerode定理)如下三个命题等价：

(1) $\mathbf{L} \subseteq \Sigma^*$是 **RL** ；

(2) **L**是Σ^*上的某一个具有有穷指数的右不变等价关系**R**的某些等价类的并；

(3) $\mathbf{R_L}$具有有穷指数。

Myhill-Nerode 定理

证明：

- 由(1)可以推出(2)

设**L**$\subseteq\Sigma^*$是 **RL**，所以，存在**DFA M=(Q**，Σ，δ，**q_0**，**F)**，使得**L(M)=L**。由前述命题，**R_M**是Σ^*上的右不变等价关系，而且$|\Sigma^*/R_M|\leqslant|Q|$，所以，**$R_M$**具有有穷指数。而

$$L=\bigcup_{q\in F} set(q)$$

L是Σ^*上的具有有穷指数的右不变等价关系**R_M**的、对应于**M**的终止状态的等价类的并。

Myhill-Nerode 定理

- 由(2)可以推出(3)。

设x R y，由R的右不变性可知，对于任意$z \in \Sigma^*$，

xz R yz

而L是R的某些等价类的并，所以，

$xz \in L \Leftrightarrow yz \in L$

根据R_L的定义，

$x\ R_L\ y$

故R是R_L的加细。由于R具有有穷指数，所以，R_L具有有穷指数。

Myhill-Nerode 定理

- 由(3)可以推出(1)。

令$M' = (\Sigma^*/R_L, \Sigma, \delta', [\varepsilon], \{[x] | x \in L\})$

$[\varepsilon]$表示ε所在的等价类对应的状态；

$[x]$ 表示x所在的等价类对应的状态。

对于$\forall([x], a) \in (\Sigma^*/R_L) \times \Sigma$，$\delta'([x], a) = [xa]$

- δ' 定义的相容性
- $L(M') = L$

Myhill-Nerode 定理

- **例** 用前述定理证明$\{0^n1^n|n\geqslant 0\}$不是 RL
  - 根据L的句子的特征来寻找R_L的等价类。
  - L的句子的主要特点有两个：

(1) 句子中所含的字符0的个数与所含的字符1的个数相同。

(2) 所有的0都在所有的1的前面

  - 可以得到如下一些等价类。

Myhill-Nerode 定理

[10]={x|x=0^n1^m(m≥n+1)或者x中含子串10}

[ε]—— ε 所在的等价类；

[1]——0所在的等价类；

[2]——00所在的等价类；

[3]——000所在的等价类；

…

[n]——0^n所在的等价类；

…

所以，R_L的指数是无穷的。因此，L不是 RL。

Myhill-Nerode 定理

推论 对于任意的 **RL L**，如果**DFA M=(Q，Σ，δ，q_0，F)**满足**L(M)=L**，则 **|Σ*/R_L|≤|Q|**。

- 表明：对于任意**DFA M=(Q，Σ，δ，q_0，F)**，**|Q|≥|Σ*/$R_{L(M)}$|**。
- 也表明：对任意一个 **RL L**，按照定理中所给的方法构造出来的**DFA M′**是一个接受**L**的最小**DFA**。这个**DFA**是惟一的么？

Myhill-Nerode 定理

推论 对于任意的 **RL L**，在同构意义下，接受**L**的最小**DFA**是惟一的。

证明：

- 接受L的最小DFA M=(Q，Σ，δ，q_0，F)的状态数与R_L的指数相同，也就是说，这个最小DFA的状态数与Myhill-Nerode定理证明中构造的M′=(Σ*/R_L，Σ，δ′，[ε]，{[x]|x∈L})的状态数是相同的。

Myhill-Nerode 定理

- **DFA**同构是指这两个**DFA**的状态之间有一个一一对应，而且这个一一对应还保持状态转换也是相应一一对应的。也就是说，如果**q**与**[w]**对应，**p**与**[z]**对应，当δ**(q，a)=p**时，必定有δ′**([w]，a)=[z]**。
- 这两个**DFA**同构。定义映射**f**

$\mathbf{f(q)=f(}\delta\mathbf{(q_0,\ x))=}\ \delta'\mathbf{([\,\varepsilon\,],\ x)=[x]}$

$\Leftrightarrow \delta\mathbf{(q_0,\ x)=q}$

Myhill-Nerode 定理

- **f为Q与Σ*/R_L之间的一一对应**
  - 如果δ$(q_0$，x)=δ$(q_0$，y)，则x R_M y
  - 由于R_M是R_L的加细，所以，x R_L y
  - 故，[x]=[y]，即，δ′([ε]，x)=δ′([ε]，y)。
  - 如果，δ$(q_0$，x)≠δ$(q_0$，y)
  - 则，δ′([ε]，x)≠δ′([ε]，y)
  - 即，[x]≠[y]
  - 否则，|Σ*/R_M|>|Σ*/R_L|。

Myhill-Nerode 定理

- 如果δ(q，a)=p，f(q)= [x]，必有f(p)=[xa]
  - ∀ q∈Q，如果，f(q)=f(δ(q_0，x))=[x]
  - 所以，∀ a∈Σ，
  - 如果p=δ(q，a)=δ(δ(q_0，x)，a)=δ(q_0，xa)
  - 则 f(p)=f(δ(q，a))= f(δ(δ(q_0，x)，a)) =f(δ(q_0，xa))=[xa]
  - 即，如果M在状态q读入字符a时进入状态p，则M′在q对应的状态f(δ(q_0，x))=[x]读入字符a时，进入p对应的状态f(δ(q_0，xa))=[xa]。所以，f是M和M′之间的同构映射。

DFA的极小化

- **可以区分的(distinguishable) 状态**
- 设**DFA M=(Q，**Σ，δ，**q_0，F)**，如果∃**x**∈Σ*，对**Q**中的两个状态**q**和**p**，使得δ**(q，x)**∈**F** 和δ**(p，x)**∈**F**中，有且仅有一个成立，则称**p**和**q**是**可以区分的**。否则，称**q**和**p**等价。并记作**q≡p** 。

DFA的极小化

算法 DFA的极小化算法

- 算法思想：扫描所有的状态对，找出所有的可区分的状态对，不可区分的状态对一定是等价的。
- 输入：给定的**DFA**。
- 输出：可区分状态表。
- 主要数据结构：状态对的关联链表；可区分状态表。

DFA的极小化

- 主要步骤

⑴ **for ∀(q，p)∈F×(Q-F) do**

标记可区分状态表中的表项**(q，p)**;

⑵ **for∀(q, p)∈F×F∪(Q-F) ×(Q-F)&q≠p do**

⑶ **if ∃a∈∑**，可区分状态表中的表项(δ**(q，a)**，δ**(p，a))**已被标记 **then**

begin

⑷ 标记可区分状态表中的表项**(q，p)**;

⑸ 递归地标记本次被标记的状态对的关联链表上的各个状态对在可区分状态表中的对应表项

end

DFA的极小化

(6)　 **else for** $\forall$ **a**$\in\Sigma$**，do**

(7)　 **if** δ**(q，a)**$\neq$δ**(p，a) &(q，p)**与**(**δ**(q，a)，** δ**(p，a))**不是同一个状态对 **then**

将**(q，p)**放在**(**δ**(q，a)，** δ**(p，a))**的关联链表上。

DFA的极小化

定理 对于任意DFA M=(Q，∑，δ，q_0，F)，Q中的两个状态q和p是可区分的充要条件是(q，p)在DFA的极小化算法中被标记。

证明：

先证必要性。

设q和p是可区分的，x是区分q和p的最短字符串。现施归纳于x的长度，证明(q，p)一定被算法标记。

DFA的极小化

- 当|x|=0时

ε 区分q和p，表明q和p有且仅有一个为M的终止状态，所以，

(q，p)∈F×(Q-F)

因此，它在算法的第(1)行被标记。

- 设当|x|=n时结论成立

x是区分q和p的长度为n的字符串，则(q，p)被算法标记。

DFA的极小化

- 当$|x|=n+1$时

设$x=ay$，其中$|y|=n$。由于x是区分q和p的最短的字符串，所以，$\delta(q，x)\in F$ 和 $\delta(p，x)\in F$中，有且仅有一个成立。不妨假设：

$\delta(q，x)\notin F$，$\delta(p，x)\in F$

即$\delta(\delta(q，a)，y)\notin F$，$\delta(\delta(p，a)，y)\in F$

设$\delta(q，a)=u$，$\delta(p，a)=v$

y是区分u和v的长度为n的字符串。

DFA的极小化

由归纳假设，(u，v)可以被算法标记。

- 如果在考察(q，p)时，(u，v)已经被标记，则(q，p)在算法的第(4)行被标记；
- 如果在考察(q，p)时，(u，v)还没有被标记，则(q，p)在算法的第(7)行被放入到(u，v)的关联链表中，而当(u，v)被标记时，在算法的第(5)行在“递归”过程中(q，p)被标记。
- 结论对|x|=n+1成立。

DFA的极小化

- 充分性。
- 设(q，p)在算法中被标记。对它被标记的顺序n施归纳，证明q和p是可区分的。
- 令$|F\times(Q-F)|=m$，显然，当$1\leqslant n\leqslant m$时，(q，p)是在算法的第(1)行被标记的，此时，ε是区分q和p的字符串：

$\delta(q,\varepsilon)\in F$ 和 $\delta(p,\varepsilon)\in F$

有且仅有一个成立。

DFA的极小化

- 设n≤k(k≥m)时结论成立。即，如果 (q，p)是被算法在第k个或者第k个之前标记的，则存在字符串x，x区分q和p。即：δ(q，x)∈F 和 δ(p，x)∈F有且仅有一个成立。
- 当n=k+1时，如果(q，p)是在算法的第(4)行被标记的，此时，(δ(q，a)，δ(p，a))一定是在第k个之前被标记的。设δ(q，a)=u，δ(p，a)=v，由归纳假设，存在字符串x，x区分u和v:
- δ(u，x)∈F 和 δ(v，x)∈F
- 有且仅有一个成立，从而，
- δ(q，ax)∈F 和 δ(p，ax)∈F有且仅有一个成立。即，ax是区分q和p的字符串。

DFA的极小化

- 如果(q，p)是在算法的第(5)行被标记的，则它必在某个状态对(u，v)的关联链表中，而(u，v)必在(q，p)之前被标记。由归纳假设，存在x区分(u，v)；
- 存在a∈∑，δ(q，a)=u，δ(p，a)=v使得(q，p)被放在(u，v)的关联链表中；ax是区分q和p的字符串。
- 所以，结论对n=k+1成立。由归纳法原理，结论对所有的n成立。

DFA的极小化

定理 由算法构造的**DFA**在去掉不可达状态后是最小**DFA** 。

证明：

设**M=(Q,**Σ，δ**, q$_0$, F)**为上述算法的输入**DFA**，**M′=(Q/≡,** Σ，δ'**, [q$_0$], F′)**是相应的输出**DFA**。**F′={[q]|q**$\in$**F}**。

对于$\forall$ **[q]**$\in$**Q/≡**，$\forall$ a$\in\Sigma$，定义

δ'**([q]，a)=[**δ**(q，a)]**

DFA的极小化

- δ′ 的相容性。
  - 设[q]=[p]，也就是说，q和p等价：q≡p。即根据算法，状态q和p是不可区分的(未被算法标记)。此时，对于∀a∈∑，必须有[δ(q, a)]≡[δ(p，a)]。否则，状态对(δ(q，a), δ(p，a))必定被算法标记，从而最终导致(q，p)被算法标记。此与q≡p矛盾。所以，状态[δ(q，a)]和状态[δ(p，a)] 等价：δ(q, a)=δ(p，a)。所以，δ′ 的定义是相容的。

DFA的极小化

- **L(M′)=L(M)** 。
  - 对∀**x**∈Σ*，现施归纳于|**x**|，证明 δ' **([q_0]，x)=[** δ **(q_0，x)]**
  - **|x|=0**
    - δ' **([q_0]，** ε **)=[q_0]=[** δ **(q_0，** ε **)]**
  - ∀ x∈Σ*并且|**x**|=**n**，
    - δ' **([q_0]，xa)=** δ' **(** δ' **([q_0]，x)，a)**
    - **=** δ' **([** δ **(q_0，x)]，a)**
    - **=[** δ **([** δ **(q_0，x)]，a)]**
    - **=[** δ **(q_0，xa)]**
  - 由归纳法原理，结论对∀**x**∈Σ*成立。

DFA的极小化

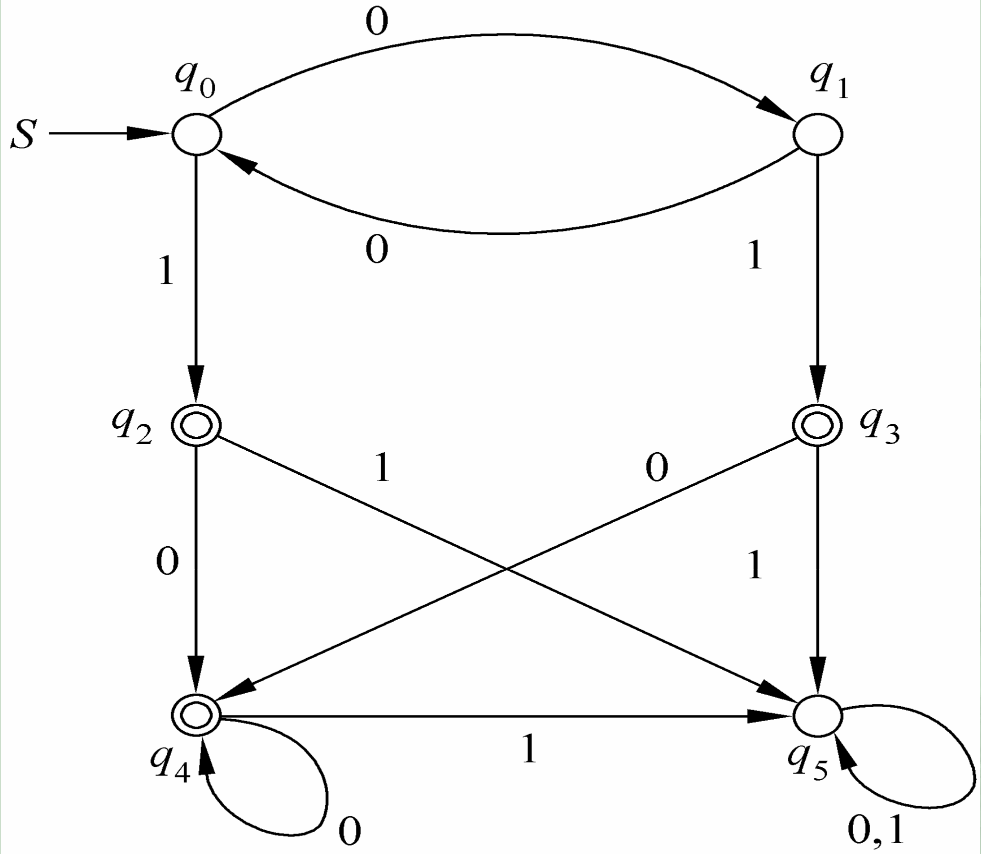
- 再由F'的定义，
- $\delta'([q_0], x)=[\delta(q_0, x)]\in F' \Leftrightarrow \delta(q_0, x)\in F$。
- 所以，
- $x\in L(M') \Leftrightarrow x\in L(M)$。
- 即：
- $L(M')=L(M)$。

DFA的极小化

- 证明所构造的M′去掉不可达状态后是最小DFA。
  - 如果[q]≠[p]，则对于∀x∈set([q])，∀y∈set([p])，x R_L y不成立。事实上，如果[q]≠[p]，则存在z∈Σ*，z区分q和p，有δ(q，z)=q′ 和δ(p，z)= p′有且仅有一个是终止状态，这就是说，xz和yz有且仅有一个是L的句子。所以，x R_L y是不成立的。

DFA的极小化

- 例 用算法对图示DFA进行极小化。

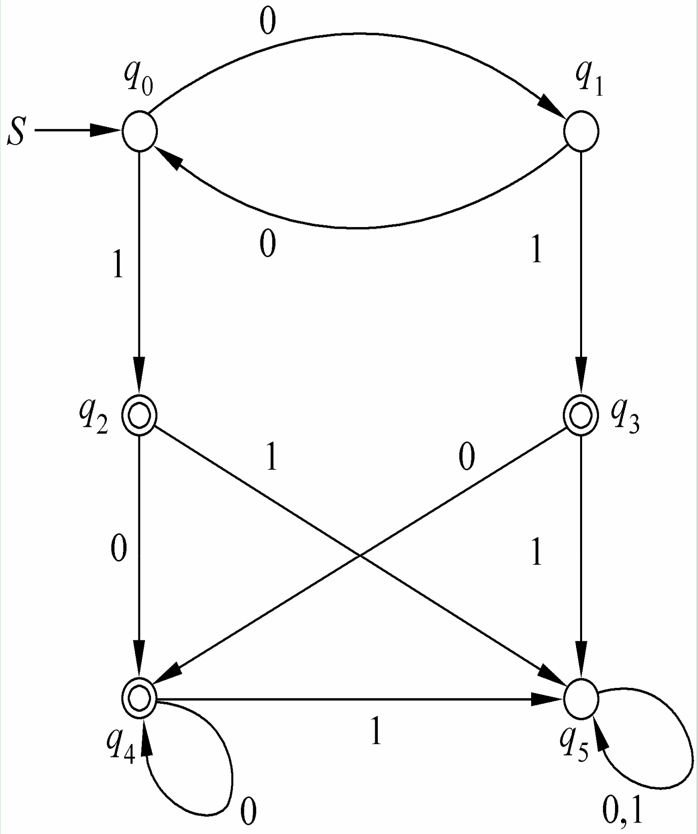


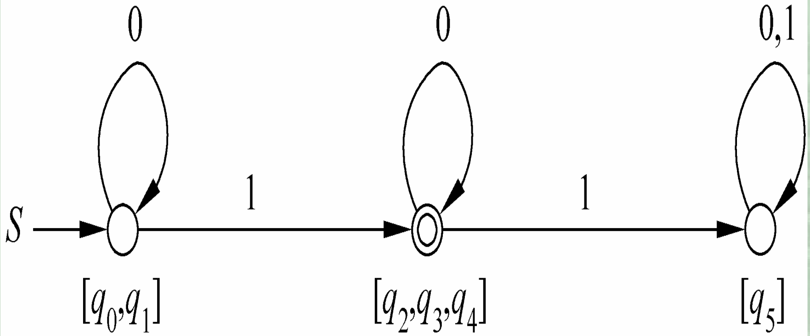
DFA的极小化

- **例** 用算法对图示**DFA**进行极小化。

q_1					
q_2	×	×			
q_3	×	×			
q_4	×	×			
q_5	×	×	×	×	×
	q_0	q_1	q_2	q_3	q_4

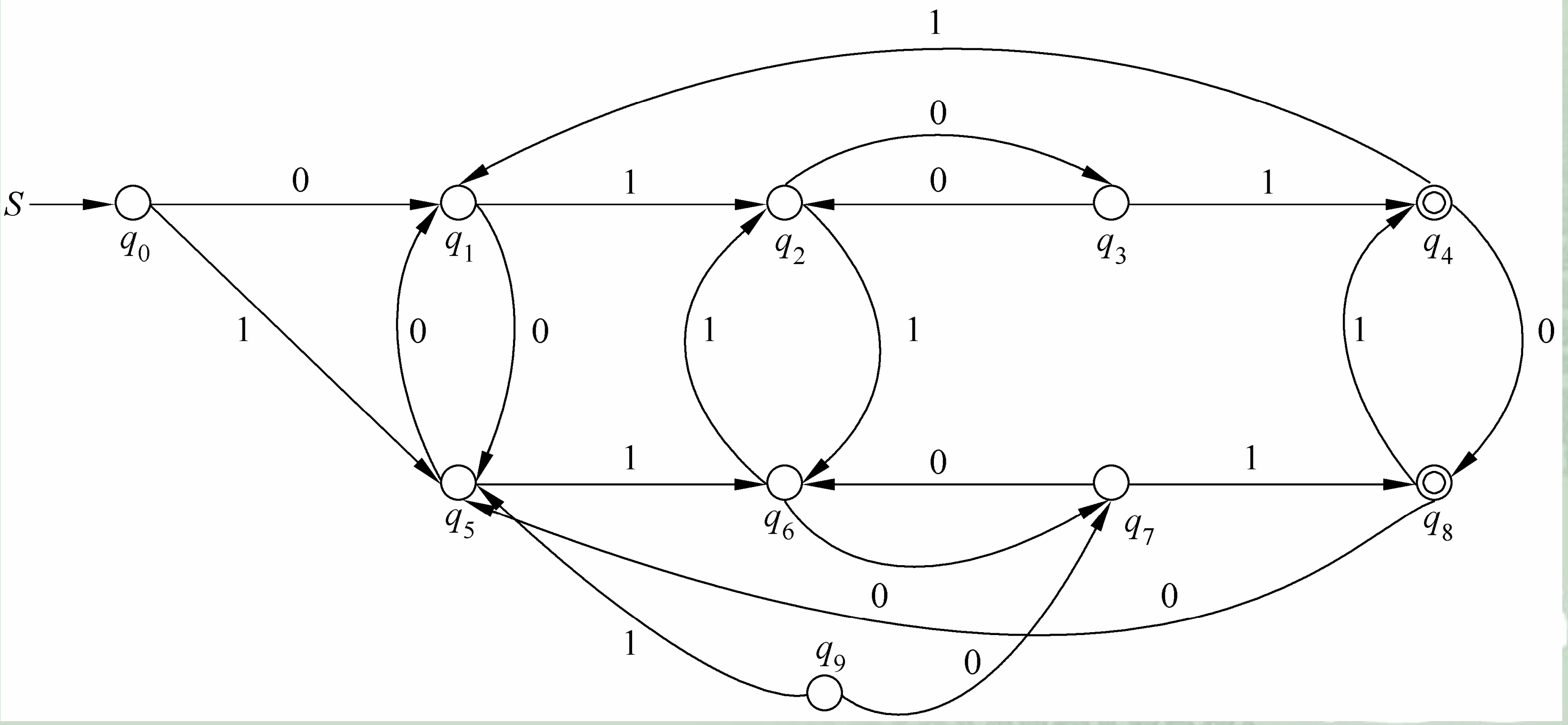
DFA的极小化





DFA的极小化

- 例 用算法对图示DFA进行极小化。



DFA的极小化

q_1	×							
q_2	×	×						
q_3	×	×	×					
q_4	×	×	×	×				
q_5	×	×	×	×	×			
q_6	×	×	×	×	×	×		
q_7	×	×	×	×	×	×	×	
q_8	×	×	×	×	×	×	×	×
	q_0	q_1	q_2	q_3	q_4	q_5	q_6	q_7

关于正则语言的判定算法

定理 设**DFA M=(Q，∑，δ，q_0，F)，L=L(M)**非空的充分必要条件是：存在$x \in \Sigma^*$，$|x|<|Q|$，$\delta(q_0, x) \in F$。

- 证明：充分性显然。
- 必要性：**M**的状态转换图中必存在一条从q_0到某一个终止状态q_f且无重复状态的路，此路中的状态数$n \leq |Q|$。此路的标记**x**满足$|x| \leq n-1$。而$\delta(q_0, x) \in F$。即**x**是**L=L(M)**的长度小于**|Q|**的句子。

关于正则语言的判定算法

定理 设DFA M=(Q，Σ，δ，q_0，F)，L=L(M)为无穷的充分必要条件是：存在$x \in \Sigma^*$，$|Q| \leqslant |x| < 2|Q|$，$\delta(q_0, x) \in F$。

- 算法通过判定是否存在$x \in \Sigma^*$，$|Q| \leqslant |x| < 2|Q|$，$\delta(q_0, x) \in F$即可。

关于正则语言的判定算法

定理 设DFA $M_1=(Q_1, \Sigma, \delta_1, q_{01}, F_1)$，DFA $M_2=(Q_2, \Sigma, \delta_2, q_{02}, F_2)$，则存在判定$M_1$与$M_2$是否等价的算法。

- 通过判定两个**DFA**的极小**DFA**是否同构就可以判定它们是否等价。

关于正则语言的判定算法

定理 设L是字母表Σ上的 **RL**，对任意 **x**∈Σ*，存在判定**x**是不是**L**的句子的算法。

- 从一定的意义上讲，接受**L**的**DFA M**就是判定**x**是否**L**的一个句子的“算法”。

小结

本章讨论了**RL**的性质。包括：**RL** 的泵引理，**RL** 关于并、乘积、闭包、补、交、正则代换、同态、逆同态等运算的封闭性。**Myhill-Nerode**定理与**FA**的极小化。

(1) 泵引理。泵引理是用**RL**的必要条件来用来证明一个语言不是 **RL** 的。它不能用来证明一个语言是 **RL**，而且是采用反证法。

小结

⑵ **RL** 对有关运算的封闭性。**RL** 在并、乘、闭包、补、交、正则代换、同态映射运算下是有效封闭的。**RL** 的同态原像是 **RL** 。

⑶ 设 L_1、$L_2 \subseteq \Sigma^*$，如果L_1是 **RL** ，则L_1/L_2也是 **RL** 。

小结

(4) 如果L是RL，则根据R_L确定的Σ^*的等价类可以构造出接受L的最小DFA。更方便的方法是通过确定给定DFA状态的可区分性构造出等价的最小DFA。

(5) 存在判定L(M)是非空、M_1与 M_2是否等价、L(M)是否无穷、x是不是RL L的句子的算法。

习题

1 下列语言都是字母表$\Sigma=\{0,1,2\}$上的语言,它们哪些是RL，哪些不是RL？如果不是RL，请证明你的结论；如果是RL，请构造其有穷描述（FA、RG或RE）。

(1) $\{0^{2n} \mid n \geq 1\}$。

(2) $\{0^{n^2} \mid n \geq 1\}$。

(3) $\{0^n 1^m 0^n \mid m, n \geq 1\}$。

(4) $\{x \mid x$中0的个数比1的个数恰好多5个$\}$。

(5) $\{xx^T w \mid x, w \in \Sigma^+\}$。

(6) $\{0^m 1^n \mid m \neq n\}$。

习题

2 设字母表{0,1}上的语言L={x|x中1的个数恰好是0的个数的2倍}，求R_L的等价类，判断它是否为RL，用Myhill-Nerode定理证明。

习题

3 设DFA $M_1=(Q_1, \Sigma, \delta_1, q_{01}, F_1)$，DFA $M_2=(Q_2, \Sigma, \delta_2, q_{02}, F_2)$，请分别构造满足如下条件的DFA M_3。

(1) $L(M)= L(M_1)\{a\}L(M_2)$，其中$a \notin \Sigma$。

(2) $L(M)= L(M_1)\cap L(M_2)$。

(3) $L(M)= L(M_1)-L(M_2)$。

(4) $L(M)= L(M_1)\cup L(M_2)$。

习题

4 构造下图所示DFA的最小DFA 。