第5章　正则语言的性质

主要内容：

5.1 FA与RG的等价性

5.2 正则语言的泵引理

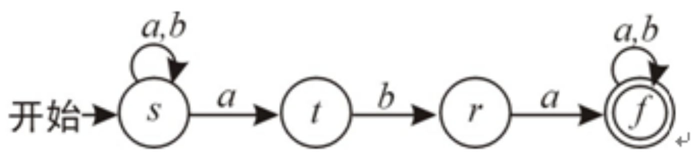
5.3 正则语言的封闭性

5.4 正则语言的判定算法

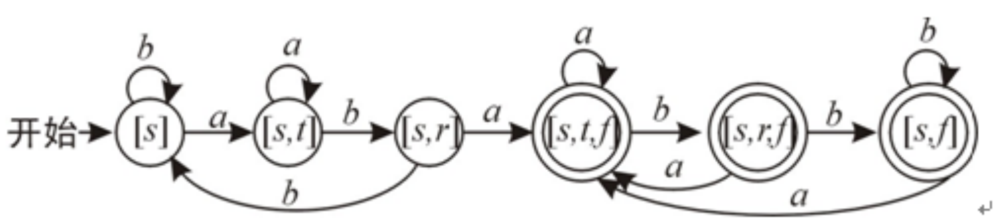
5.5 自动机的等价性与最小化

5.5 自动机的等价性与最小化

例5.6　由下图给出的NFA共有4个状态，它接受字母表{a,b}上所有包含aba的字符串。

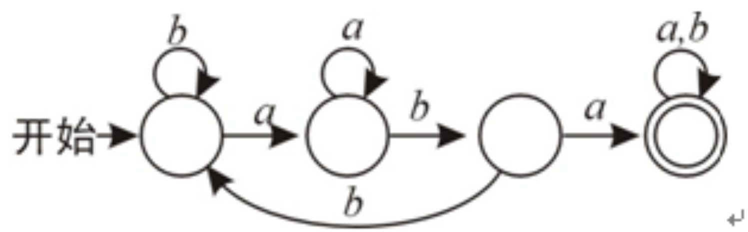


若按照第三章定理给出的方法，构造一个等价的DFA，需要设2^4=16个状态，除去不可到达的状态外，还剩下6个状态。这个DFA如下图所示。



5.5 自动机的等价性与最小化

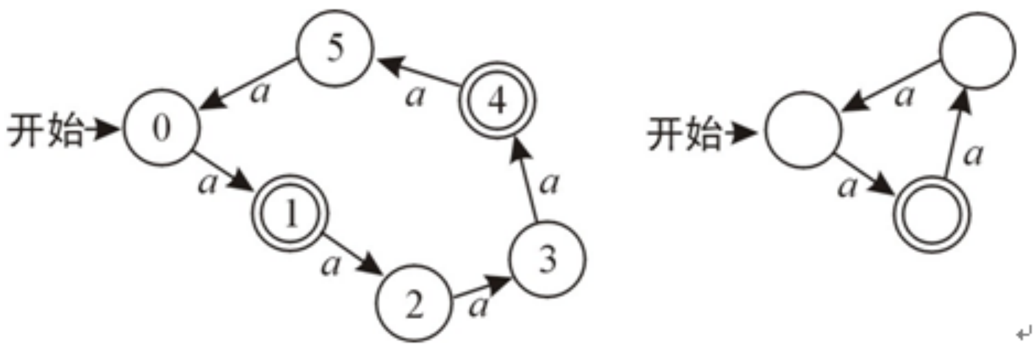
- 注意，在上图中最右边的三个终结状态可以合并为一个，这样一来，接受同一集合的DFA只用4个状态就够了。这个DFA如下：



- 后面将要证明，接受上述集合的任何DFA都不能少于4个状态。换句话说，最后这个DFA已经是最简单的了。

5.5 自动机的等价性与最小化

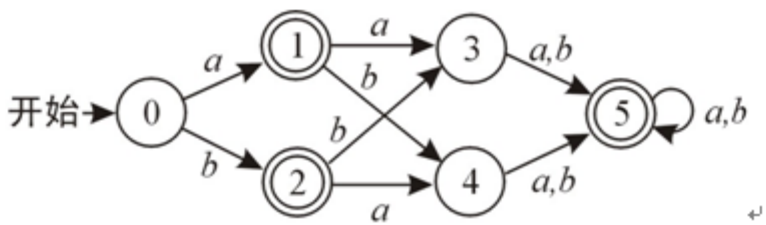
- 例5.7 由下图给出的DFA接受集合$\{a^m | m \bmod 3 = 1\}$，即$\{a,a^4,a^7,a^{10},\ldots\}$。



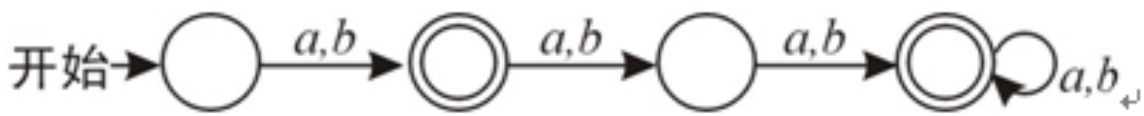
- 左边的图所示的DFA共用6个状态，显然是太多了。我们可以将状态1，4合并，状态2，5合并，状态0，3合并，得出的DFA只用3个状态，也能接受同样的集合。这个化简后的DFA示于原图的右边。

5.5 自动机的等价性与最小化

例5.8 给出一个DFA接受集合 $\{x|x\in\{a,b\}^+$，并且$|x|\neq 2\}$，图示如下：



在图中，状态1、2的功能相同，它们都表明读入的串的长度为1，应当接受。状态3、4的功能也相同，它们都表明读入的串的长度为2，不能接受。因此，状态1、2可以合并，状态3、4也可以合并。上图可以化简为的等价的DFA（见下图）：



总结：从上面的四个例子发现，在保持等价的条件下，有穷自动机确实存在化简的问题。对于简单的有穷自动机，可以根据它所接受的字符串的集合，分析每个状态的作用，决定哪些可以合并。但是对于复杂的有穷自动机，这种直观的方法就不行了，必须寻求形式化的方法，给出一个化简的算法。

5.5 自动机的等价性与最小化

问题提出：

存在最小化的**DFA**吗？如果存在，唯一吗？

Myhill–Nerode theorem provides a necessary and sufficient condition for a language to be regular.

The theorem is named for John Myhill and Anil Nerode, who proved it at the University of Chicago in 1958 (Nerode 1958).

5.5 自动机的等价性与最小化

- 二元关系 ---是一个集合
  - 任意的R⊆A×B，R是A到B的二元关系。
  - （a，b）∈R，也可表示为：aRb。
  - A称为定义域(domain)，B称为值域(range)。
  - 当A=B时，则称R是A上的二元关系。
- 二元关系的性质
  - 自反(reflexive)性、反自反(irreflexive)性、对称(symmetric)性、反对称(asymmetric)性、传递(transitive)性。
- 等价关系(equivalence relation)
  - 具有自反性、对称性、传递性的二元关系称为等价关系。
  - 如：“=”关系是等价关系。

5.5 自动机的等价性与最小化

- 等价类（equivalence class)---由等价关系R决定的

S的满足如下要求的划分：S_1、S_2、S_3、…、S_n…称为S关于R的等价划分，S_i称为等价类。

(1) $S = S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_n \cup \cdots$;

(2) 如果$i \neq j$，则$S_i \cap S_j = \Phi$;

(3) 对任意的i，S_i中的任意两个元素a、b，aRb恒成立；

(4) 对任意的i，j，$i \neq j$，S_i中的任意元素a和S_j中的任意元素b，aRb恒不成立 。

➢ 等价类：指的是该类中的元素之间存在等价关系。

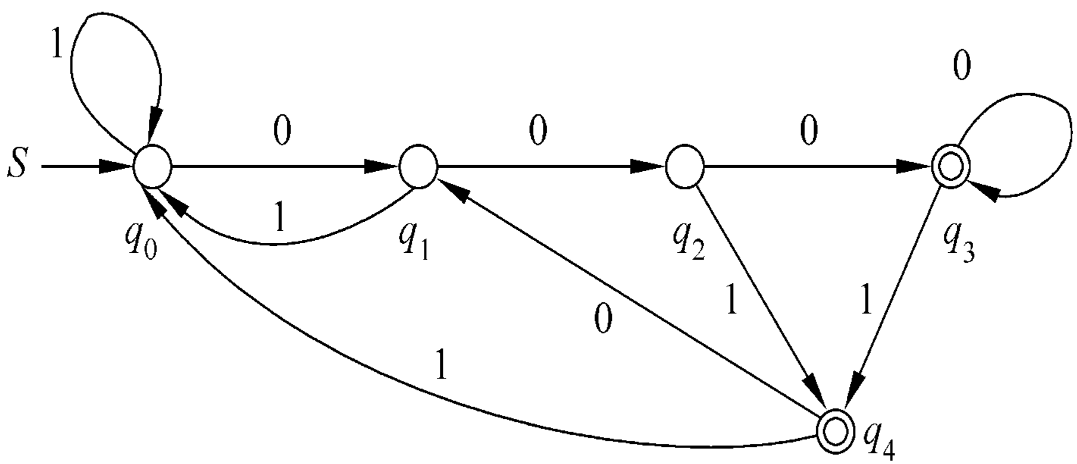
➢ 等价关系R将S分成的等价类的个数称为是R在S上的**指数**。

5.5 自动机的等价性与最小化

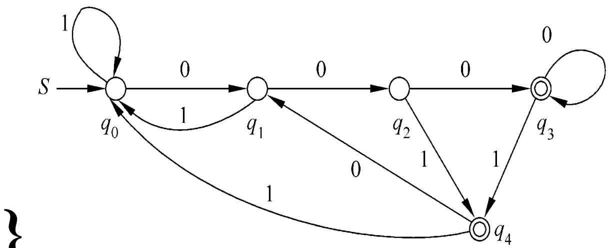
定义5.6-1 能引导FA从开始状态q_0到达状态q的字符串的集合为：

$$\text{set}(q)=\{x \mid x\in \Sigma^*,\ \delta(q_0, x) = q\}$$

对下图所给的DFA 中的所有q，求set(q)。



5.5 自动机的等价性与最小化



$set(q_0)=\{x \mid x\in\Sigma^*, x=\varepsilon$ 或者x以1结尾$\}$

$set(q_1)=\{x \mid x\in\Sigma^*, x=0$或者$x$以10结尾$\}$

$set(q_2)=\{x \mid x\in\Sigma^*, x=00$或者$x$以100结尾$\}$

$set(q_3)=\{x \mid x\in\Sigma^*, x$以000结尾$\}$

$set(q_4)=\{x \mid x\in\Sigma^*, x$以001结尾$\}$

这5个集合具有如下性质（涉及概念：等价关系、等价类、划分、指数）：

1. 是两两互不相交；
2. 5个集合的并，构成了该DFA的输入字母表｛0，1｝的克林闭包；
3. 这5个集合是｛0，1｝*的一个划分；
4. 按照这个划分，可以定义一个等价关系，每个集合中的字符串满足该等价关系；

5. 5 自动机的等价性与最小化

定义5.6 DFA M确定的Σ*上的等价关系 R_M。

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

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

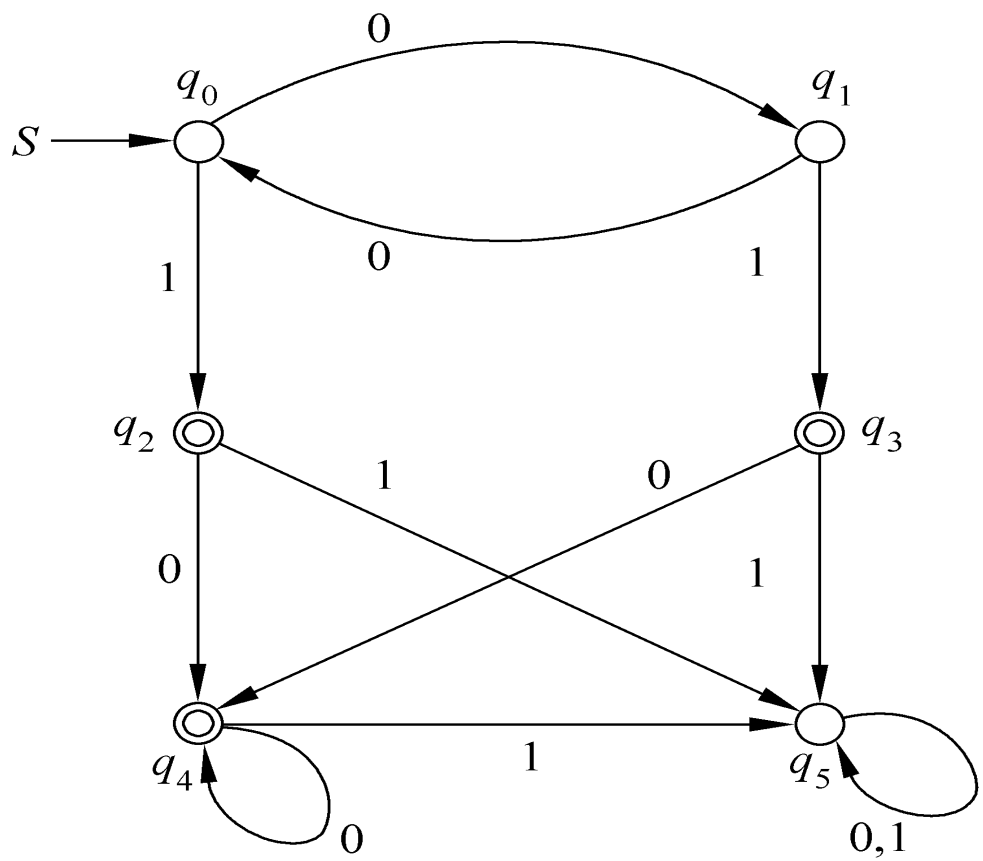
显然，

根据定义5.6-1而得

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

x R_M y的直观含义：从初始状态q_0出发，x和y都能把自动机M引导到相同的状态q，$q \in Q$ 。

例 5.9 设 $L=0^*10^*$，它对应的DFA M如下图。



5.5 自动机的等价性与最小化

对应于q_0： $(00)^n R_M (00)^m$ n， m≥0；

对应于q_1： $0(00)^n R_M 0(00)^m$ n， m≥0；

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

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

对应于q_4：

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

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

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

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

对应于q_5： $x R_M y$——x， y为至少含两个1的串。

定义5-7 语言L确定的Σ^*上的关系R_L。

对于$\forall x, y \in \Sigma^*$,

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

即：对于$\forall x, y \in \Sigma^*$，如果$x \; R_L \; y$，则在x和y后无论接$\Sigma^*$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。

注意：这里的语言L不一定是正则的。但是，如果L是正则的，又会怎么样呢？

5.5 自动机的等价性与最小化

试证：如果xR_My，则一定有xR_Ly。

证明：因为L是正则的，所以一定存在DFA M 识别语言L。
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,

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

接下页

5.5 自动机的等价性与最小化

即，对于$\forall z \in \Sigma^*$，

$xz \in L \Leftrightarrow yz \in L$。

表明，

$x\ R_L\ y$，

也就是

$x\ R_{L(M)}\ y$。

讨论：R_L、$R_{L(M)}$，R_M 三者之间的关系？

1. 如果xR_My，则一定有xR_Ly 。反之不一定成立。
2. $R_{L(M)}$中的M是DFA，L(M)是正则语言，但R_L中的L不一定是正则的。

例5.10 if $\Sigma=\{0,1\}$ and $L=\Sigma^*0\Sigma$, then R_L has four equivalence classes:

1. $S_1 = \Sigma^*00$
2. $S_2 = \Sigma^*01$
3. $S_3 = \Sigma^*10 \cup 0$
4. $S_4 = \Sigma^*11 \cup 1 \cup \varepsilon$

5.5 自动机的等价性与最小化

R_M与$R_{L(M)}$的关系

- $R_{L(M)}$---按区域划分
- R_M ---按省份划分

例5.11 If $\Sigma = \{0, 1\}$ and $B = \{0^n1^n : n \geq 0\}$, then R_L has infinitely many equivalence classes:

1. $S1=\{0^n1^m : m>n\geq 0\} \cup \Sigma^*1\Sigma^*0\Sigma^*$
2. $S2 =\{0^n1^n : n\geq 0\}$
3. $S3 =\{0^n1^{n-1} : n\geq 1\}$
4. $S4 =\{0^n1^{n-2} : n\geq 2\}$
5. $S5 =\{0^n1^{n-3} : n\geq 3\}$

……

定义5-8 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。

注意：这里的R不一定是R_M，也不一定是R_L。

命题 5-1 对于任意DFA $M=(Q, \Sigma, \delta, q_0, F)$，M所确定的$\Sigma^*$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；

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

$\Leftrightarrow y\ R_M\ x$　　根据R_M的定义。

5.5 自动机的等价性与最小化

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。

(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。

命题 5-2 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$\forall z \in \Sigma^*$，$xz \in L \Leftrightarrow yz \in L$）$\Leftrightarrow y\ R_L\ x$

5.5 自动机的等价性与最小化

传递性：设$x\ R_L\ y$，$y\ R_L\ z$。

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

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

所以，

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

即：

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

故：

$$x\ R_L\ z$$

即R_L是等价关系。

(2) R_L 是右不变的。

设 x R_L y。由 R_L 的定义，对 $\forall$ x，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。

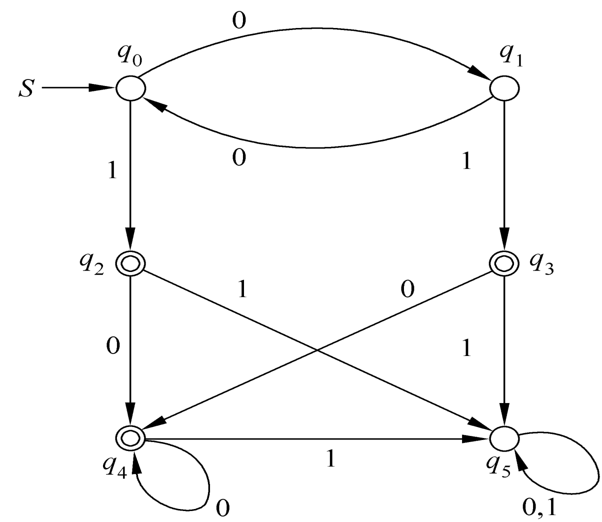
定义5-9

关系R的指数（index）———设R是Σ^*上的等价关系，则称$|\Sigma^*/R|$是R关于Σ^*的指数（index），简称为R的指数。注意：Σ^*/R是指关系对Σ^*的划分。

R的一个等价类——— Σ^*的关于R的一个等价类，也就是Σ^*/R的任意一个元素（这里指一个集合，或一个划分），简称为R的一个等价类 。

5.5 自动机的等价性与最小化

例 5.12 下图所示DFA M所确定的R_M的指数为6。R_M将Σ^*分成6个等价类： *（见例5.9）*



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

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

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

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

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

$set(q_5)=\{x \mid x$为至少含两个1的串$\}$。

5.5 自动机的等价性与最小化

R_M与$R_{L(M)}$的关系讨论：

1. $\forall x, y \in \Sigma^*$，如果$x\ R_M\ y$，必有$x\ R_{L(M)}\ y$成立；如果$x\ R_{L(M)}\ y$成立，$x\ R_M\ y$不一定成立；

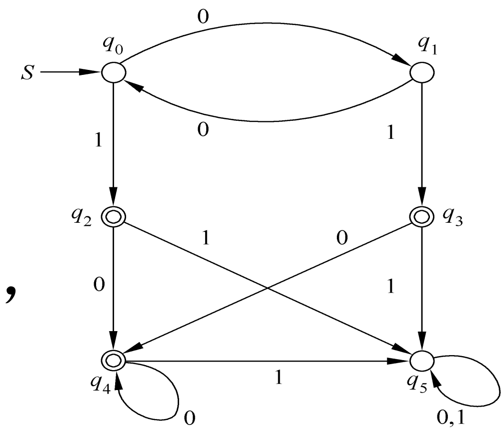
如：例5.9中，$0R_M00$不成立，但$0R_{L(M)}00$成立（为什么？）；即对于任意DFA $M=(Q, \Sigma, \delta, q_0, F)$。必有：

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

2. R_M是$R_{L(M)}$的“加细”(refinement)

- 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)。

5. 5 自动机的等价性与最小化



以例5. 9为例，解释R_M和$R_{L(M)}$之间的区别

第一步：以R_M进行等价划分（等价分类）

Σ^*/R_M ={set(q_0)，set(q_1)，set(q_2)，set(q_3)，set(q_4)，set(q_5)} ————分类依据set(q_i)

第二步：以$R_{L(M)}$进行等价分类

(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∉L(M)，000x∉L(M)。这就是说，对于任意的x∈Σ*，**00x∈L(M) ⇔ 000x∈L(M)**。即按照$R_{L(M)}$的定义，00与000被分在同一个等价类中。所以，set(q_0)和set(q_1) 被包含在$R_{L(M)}$的同一个等价类中。

(2) 取$00 \in set(q_0)$，$001 \in set(q_2)$。

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

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

5.5 自动机的等价性与最小化

(3) 取$001 \in set(q_2)$，$01 \in set(q_3)$。

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，$001x \in L(M)$，$01x \in L(M)$；当x含有1时，$001x \notin L(M)$，$01x \notin L(M)$。这就是说，对于任意的$x \in \Sigma^*$，$001x \in L(M) \Leftrightarrow 01x \in L(M)$。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。

5.5 自动机的等价性与最小化

(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。

(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。

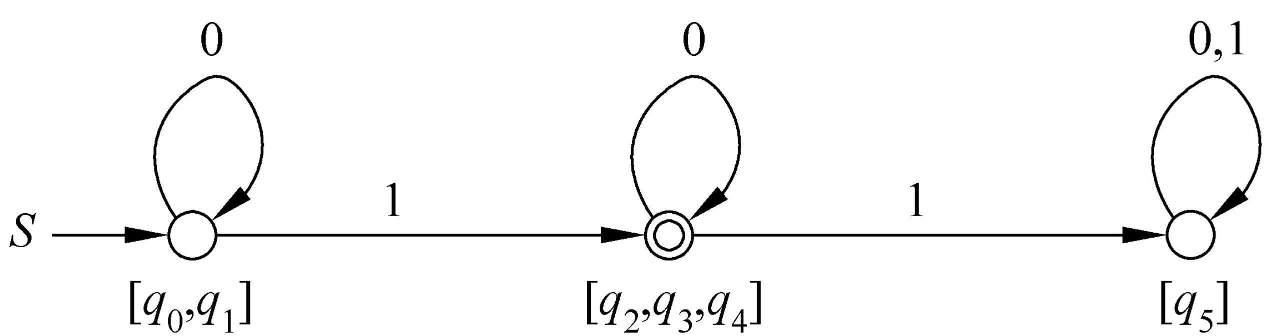
5.5 自动机的等价性与最小化

综合上述分析，得：

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

不妨采用新的符号标记这3个等价类：

1. 不含1： $[\varepsilon] = set(q_0)\cup set(q_1)=0^*$；
2. 含一个1：$[1] = set(q_2)\cup set(q_3)\cup set(q_4)=0^*10^*$；
3. 含多个1：$[11] = set(q_5)=0^*10^*1(0+1)^*$。



根据$R_{L(M)}$构造的DFA

5.5 自动机的等价性与最小化

定理5-1 (Myhill-Nerode定理)下列三个命题等价：

(1) $L \subseteq \Sigma^*$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) R_L具有有穷指数。

5.5 自动机的等价性与最小化

证明：

由(1)可以推出(2)

设L⊆Σ*是 RL ，所以，存在DFA M=(Q，Σ，δ，q_0，F)，使得L(M)=L。由命题5-3-1，R_M是Σ*上的右不变等价关系，而且|Σ*/R_M|≤|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<u>对应于M的接受状态的等价类</u>的并。

某些等价类

由(2)可以推出(3)

思路：已知R的指数是有穷，如果R是R_L的加细（即$xRy \Rightarrow xR_Ly$），则R_L的指数也是有穷的。

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，

$xz \in L \Leftrightarrow yz \in L$　　何时L才是R的所有等价类的并？

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。

5.5 自动机的等价性与最小化

由(3)可以推出(1)

思路：由R_L对Σ^*的分类构造DFA M，且L(M)=L。

一、根据R_L，构造 $M=(Q, \Sigma, \delta, q_0, F)$

其中：$Q=\Sigma^*/R_L$，$q_0=[\varepsilon]$，$F=\{[x] \mid x\in L\}$

$[\varepsilon]$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具有相容性(无论在等价类[x]中取哪个元素为代表，得到的函数值都是相同的，又称一致性)

二、证明L(M)=L? 注意：L是R_L中的L，L(M)是M识别的语言

1. 若$x\in L(M)$，则$\delta([\varepsilon],x)\in F$，即$[x]\in F$。根据F的定义，$x\in L$;
2. 若$x\in L$，则$[x]\in F$。 $\because\ \delta([\varepsilon],x)=[x]$，$\therefore x\in L(M)$。

例5.13 证明$\{0^n1^n \mid n \geq 0\}$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.

5.5 自动机的等价性与最小化

可以得到如下一些等价类

[10]={x ｜ $x=0^n1^m$ (m>n)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。

5.5 自动机的等价性与最小化

推论 5-2 对于任意的 RL L，如果DFA $M=(Q, \Sigma, \delta, q_0, F)$满足$L(M)=L$，则$|\Sigma^*/R_L| \leqslant |Q|$。

- 表明，对于任意DFA $M=(Q, \Sigma, \delta, q_0, F)$，$|Q| \geqslant |\Sigma^*/R_{L(M)}|$。
- 也表明，对任意一个 RL L，按照定理证明中（即，由(3) 推(1)的证明过程）所给的方法构造出来的DFA M是一个接受L的状态最少的DFA。这个DFA是惟一的么？

5.5 自动机的等价性与最小化

定义5.5　给出两个DFA

$M=(Q_m,\Sigma,\delta_m,q_m,F_m)$，$N=(Q_n,\Sigma,\delta_n,q_n,F_n)$。

如果在它们的状态集之间存在一个一对一的映射f: $Q_m\to Q_n$，满足：

（1）$f(q_m)=q_n$，

（2）$f(\delta_m(p,a))=\delta_n(f(p),a)$，对一切$p\in Q_m$，$a\in\Sigma$,

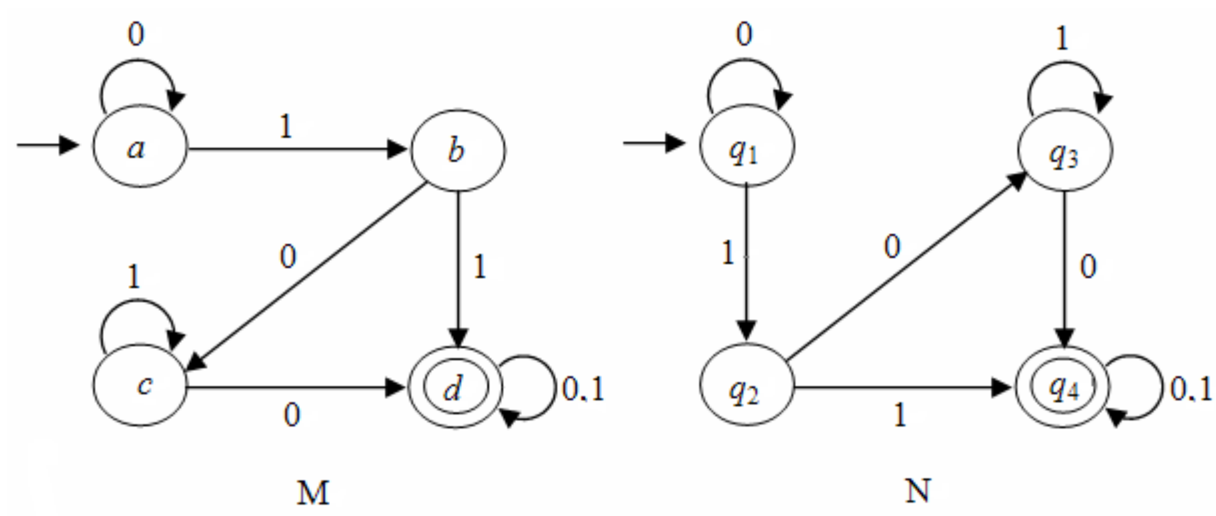
（3）$p\in F_m$ 当且仅当 $f(p)\in F_n$。

则称M 和N是**同构**的。

1. M、N的初始状态互相对应，终结状态（可能不只一个）一一对应。
2. M和N中两个对应的状态（对任何符号a∈∑）经过一次转移后，所得的状态仍然是对应的。
3. 实际上，两个同构的DFA，除了状态名字可以不同以外，本质上是同一个DFA。显然，两个同构的DFA是等价的。

5.5 自动机的等价性与最小化

例5.14 在下图中给出两个DFA M和N，它们各有4个状态，Q_m ={a，b，c，d}，Q_n ={q_1，q_2 ,q_3 ,q_4 }。一对一的映射f为：f(a)=q_1，f(b)=q_2，f(c)=q_3 , f(d)=q_4。其中a、q_1 为各自的初始状态，d、q_4 为各自的终结状态，满足互相对应的要求。另外，$f(\delta_m(a,0))=f(a)= q_1$，$\delta_n(f(a),0)=\delta_n(q_1 ,0)=q_1$；$f(\delta_m(a,1))=f(b)=q_2$，$\delta_n(f(a),1)=\delta_n(q_1 ,1)=q_2$；$f(\delta_m(b,0))=f(c)= q_3$，$\delta_n(f(b),0)=\delta_n(q_2 ,0)=q_3$ 等等，均满足定义5.5第（2）条的要求，因此M和N是两个同构的DFA。



推论5-3 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 $M=(Q, \Sigma, \delta, q_0, F)$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 $M'=(\Sigma^*/R_L, \Sigma, \delta', [\varepsilon], \{[x] \mid x\in L\})$的状态数是相同的。

5.5 自动机的等价性与最小化

- 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移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$\delta(q, a)=p$时，必定有$\delta([w], a)=[z]$。
- 这两个DFA是同构。定义映射f

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

$$\Leftrightarrow \delta(q_0, x)=q$$

5.5 自动机的等价性与最小化

- f为Q与Σ^*/R_L之间的一一对应
  - 如果$\delta(q_0, x)=\delta(q_0, y)$，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，$\delta'([\varepsilon], x)=\delta'([\varepsilon], y)$。
  - 如果，$\delta(q_0, x)\neq\delta(q_0, y)$
  - 则，$\delta'([\varepsilon], x)\neq\delta'([\varepsilon], y)$
  - 即，$[x]\neq[y]$
  - 否则，$|\Sigma^*/R_M|>|\Sigma^*/R_L|$。

5.5 自动机的等价性与最小化

- 如果 $\delta(q, a)=p$，$f(q)=[x]$，必有 $f(p)=[xa]$
  - $\forall q\in Q$，如果，$f(q)=f(\delta(q_0, x))=[x]$
  - 所以，$\forall a\in\Sigma$，如果，
  - $p=\delta(q, a)=\delta(\delta(q_0, x), a)=\delta(q_0, xa)$
  - 则 $f(p)=f(\delta(q, a))= f(\delta(\delta(q_0, x), a))=f(\delta(q_0, xa))=[xa]$
  - 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 $f(\delta(q_0, x))=[x]$ 读入字符a时，进入p对应的状态 $f(\delta(q_0, xa))=[xa]$。所以，f是M和M′之间的同构映射。

定义5.10 可以区分的(distinguishable) 状态对

设DFA $M=(Q, \Sigma, \delta, q_0, F)$，如果$\exists x \in \Sigma^*$， 对Q中的两个状态q和p，使得$\delta(q, x) \in F$ 和$\delta(p, x) \in F$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$q \equiv p$ 。

算法5-1 DFA的极小化算法一

- 算法思想：扫描所有的状态对，找出所有的可区分的状态对，不可取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。

5.5 自动机的等价性与最小化

主要步骤

(1) for $\forall(q, p)\in F\times(Q-F)$ do

标记可区分状态表中的表项(q，p)；

(2) for $\forall(q,p)\in F\times F\cup(Q-F)\times(Q-F)$ & $q\neq p$ do

(3) if $\exists a\in\Sigma$，可区分状态表中的表项($\delta(q, a)$，$\delta(p, a)$)已被标记 then

begin

(4) 标记可区分状态表中的表项(q，p)；

(5) 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

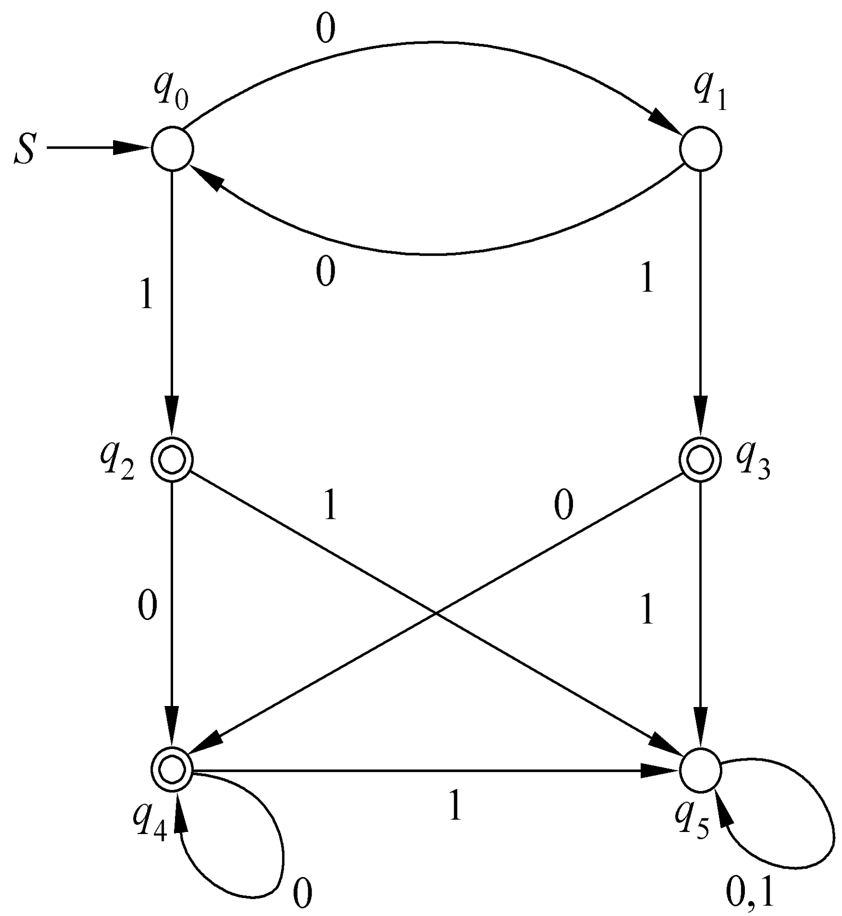
end

(6) else for $\forall a\in\Sigma$，do

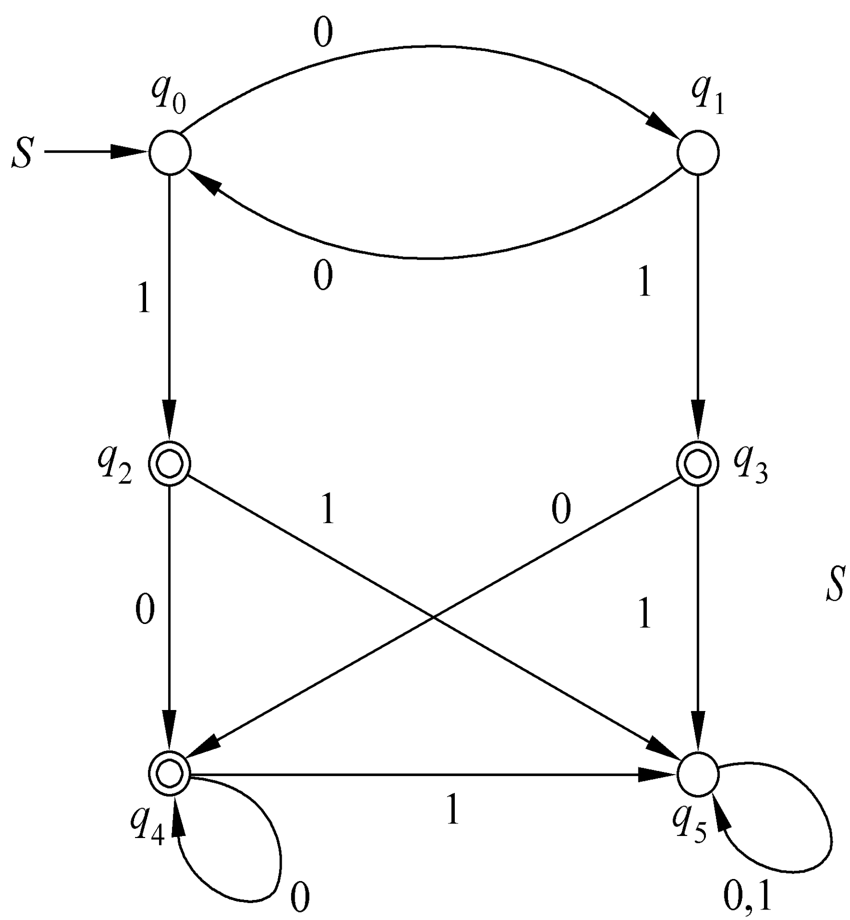
(7) if $\delta(q, a)\neq\delta(p, a)$ &(q，p)与($\delta(q, a)$，$\delta(p, a)$)不是同一个状态对 then

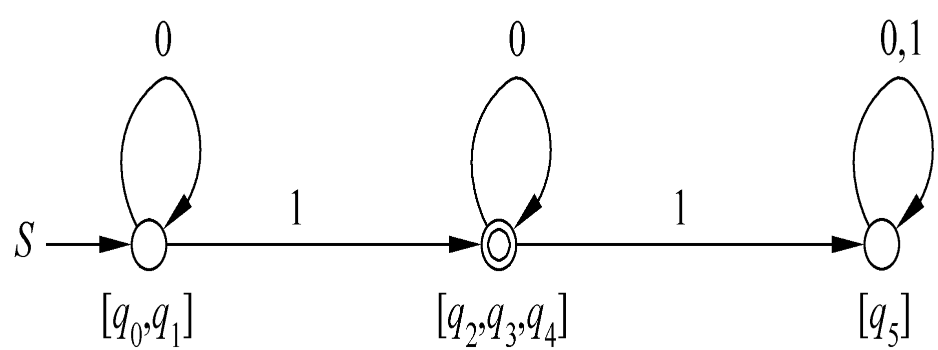
将(q，p)放在($\delta(q, a)$，$\delta(p, a)$)的关联链表上。

- 例5.15 用算法5-1对下图所示的DFA进行极小化。



3.5 自动机的等价性与最小化

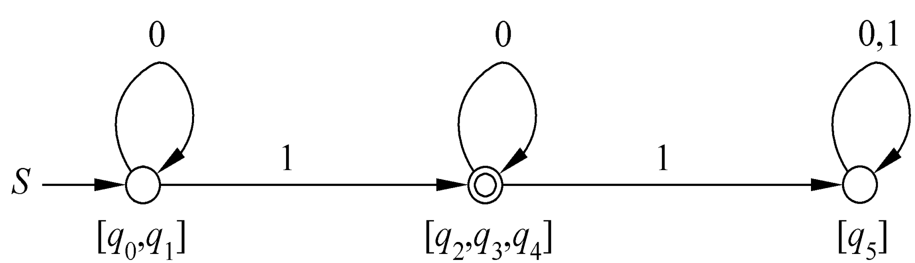




3.5 自动机的等价性与最小化

可区分状态图

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



极小化算法二

①为所有状态对(p,q)(p,q∈Q)画一张表，开始时表中每个格子内均为空白（未做任何标记）。

②对p∈F, q∉ F的一切状态对(p,q)，在相应的格子内做标记（例如画一个×）。

③重复下述过程，直到表中内容不再改变为止：

如果存在一个未被标记的状态对(p,q)，且对于某个a∈∑，（δ(p,a)，δ(q,a)）已做了标记，则在(p,q) 相应的格子内做标记。

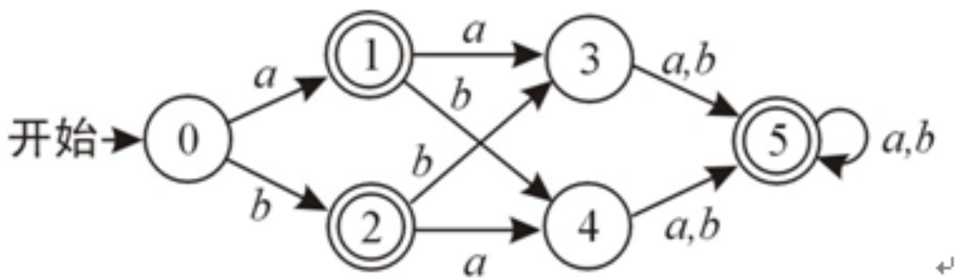
④在完成1，2，3之后，所有未被标记的状态对(p,q)都是等价的，即p≡q。

- 对上述算法有以下几点说明：

1. 在第2步，因为终结状态和非终结状态肯定是不等价的，所以对这些状态对首先做了标记。
2. 第3步是算法的主体，它由三重循环构成。中循环是扫视代表状态对的所有格子，如果存在一个尚未被标记的状态对(p,q)，则对于所有的a∈∑，检查（δ(p,a)，δ(q,a)）是否已做了标记，如果已做了标记，则在(p,q) 相应的格子内做标记；否则，对下一个a再检查（这是内循环）。最外层循环是不断重复中循环的工作，直到表中内容不再改变为止。在这一过程中，对某些状态对(p,q)可能查看多次，因为这一轮扫视时未被标记，到下一轮扫视时就有可能被标记。
3. 第4步实际上是一个结论，它指明算法的正确性，这是需要证明的。后面我们将用定理形式对算法的正确性加以证明。

3.5 自动机的等价性与最小化

例 5.16 对下图 给出的DFA用极小化算法进行化简，该DFA接受$\{x|x\in\{a,b\}^+$，并且$|x|\neq 2\}$。



3. 5 自动机的等价性与最小化

例 5.16 对下图 给出的DFA用极小化算法进行化简。

- 在算法的第1步，我们对该DFA中的6个状态建立一个空白表，表中所有格子皆为空。因为p和q等价是对称的，所以只用表的下三角部分即可（阶梯形的）。这张表如左下图所示。
- 在算法的第2步之后，对终结状态和非终结状态的状态对的格子内做了标记。这个结果如右下图所示。

	0	1	2	3	4
1	—				
2	—	—			
3	—	—	—		
4	—	—	—	—	
5	—	—	—	—	—

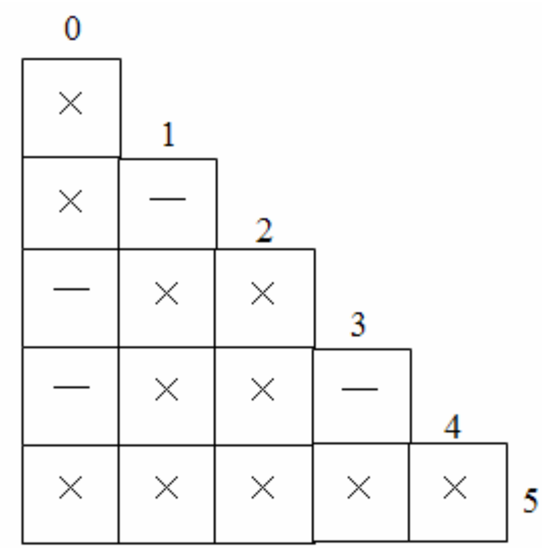
(*a*)

	0	1	2	3	4
1	×				
2	×	—			
3	—	×	×		
4	—	×	×	—	
5	×	—	—	×	×

(*b*)

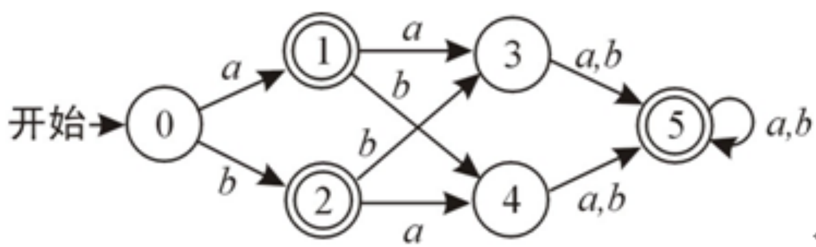
3.5 自动机的等价性与最小化

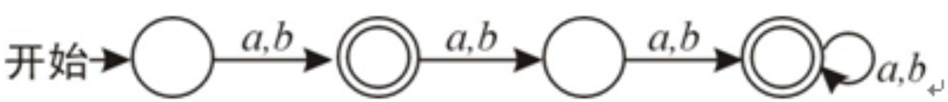
- 在第3步，我们找出尚未标记的状态对，例如（0，3），对于a∈∑，有δ(0,a)=1，δ(3,a)=5，因为（1，5）未被标记，所以现在也不能标记（0，3）。对于b∈∑，有δ(0,b)=2, δ(3,b)=5，因为（2，5）未被标记，所以现在仍不能标记（0，3）。由于∑中只有a、b两个符号，故对（0，3）的考察暂时停止。再看（0，4）和（1，2），基于同样的理由也不能被标记。但是对于（1，5），对a∈∑，有δ(1,a)=3，δ(5,a)=5，而此时（3，5）已被标记，所以（1，5）也应被标记，对b就不用再看了。类似地，可以标记（2，5）。



3.5 自动机的等价性与最小化

- 现在开始下一遍考察。对于（0，3），仍和上次一样，有δ(0,a)=1，δ(3,a)=5。但此时（1，5）已于上一遍被标记，所以这一遍我们标记（0，3）。类似地，可以标记（0，4）。
- 最后得出的结论是：1≡2和3≡4。
- 依照这个算法，得出等价状态后构造的自动机，就是具有4个状态的那个DFA。





	0	1	2	3	4
1	×				
2	×	—			
3	×	×	×		
4	×	×	×	—	
5	×	×	×	×	×

5.5 自动机的等价性与最小化

极小化算法三：P问题---在多项式时间内完成。

Suppose we have a DFA – how to construct a corresponding minimal DFA ?

1. Remove unreachable states.
2. Identify equivalent states (and merge them):
    - construct graph with vertices=states
    - place edges between every accepted and nonaccept state
    - continue placing edges as follows while can:
      for $q,r \in Q$, $q \neq r$, place edge (q,r) if there exists $a \in \Sigma$ s.t. $(\delta(q,a),\delta(r,a))$ is an edge.
    - Merge all states that do not have edges between them into a single state

参见 ep327 7.42

5. 5 自动机的等价性与最小化

The Myhill–Nerode theorem states that：

1. L is regular if and only if R_L has a finite number of equivalence classes.
2. and that the number of states in the smallest deterministic finite automaton (DFA) recognizing L is equal to the number of equivalence classes in R_L.
3. In particular, this implies that there is a unique minimal DFA with minimum number of states (Hopcroft & Ullman 1979).

5. 5 自动机的等价性与最小化

定理的应用：

1. The Myhill–Nerode theorem may be used to show that a language *L* is regular by proving that the number of equivalence classes of R_L is finite.（证明一个语言是正则的）
2. Another immediate corollary of the theorem is that if a language defines an infinite set of equivalence classes, it is *not* regular. It is this corollary that is frequently used to prove that a language is not regular.（证明一个语言是非正则的）
3. 极小化DFA