第5章 正则语言的性质

主要内容：

5.1 FA与RG的等价性

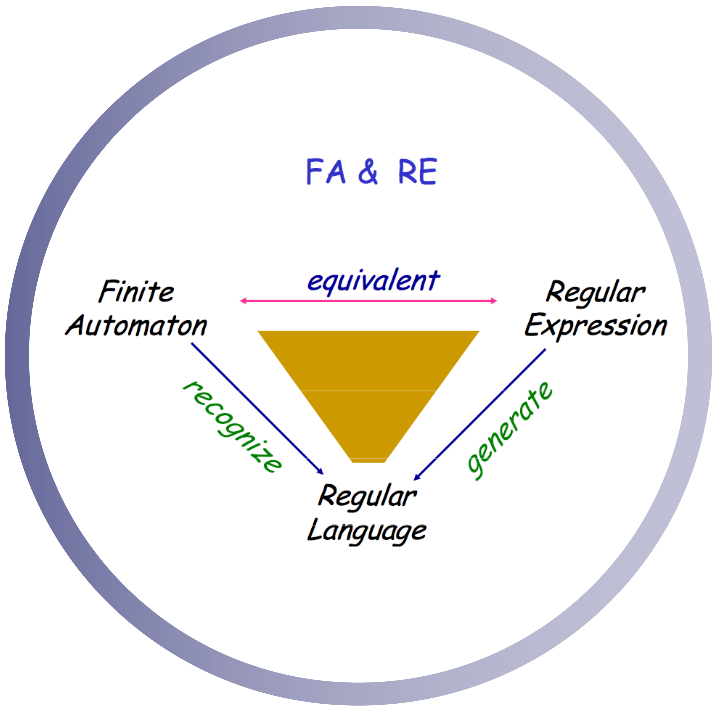
5.2 正则语言的泵引理

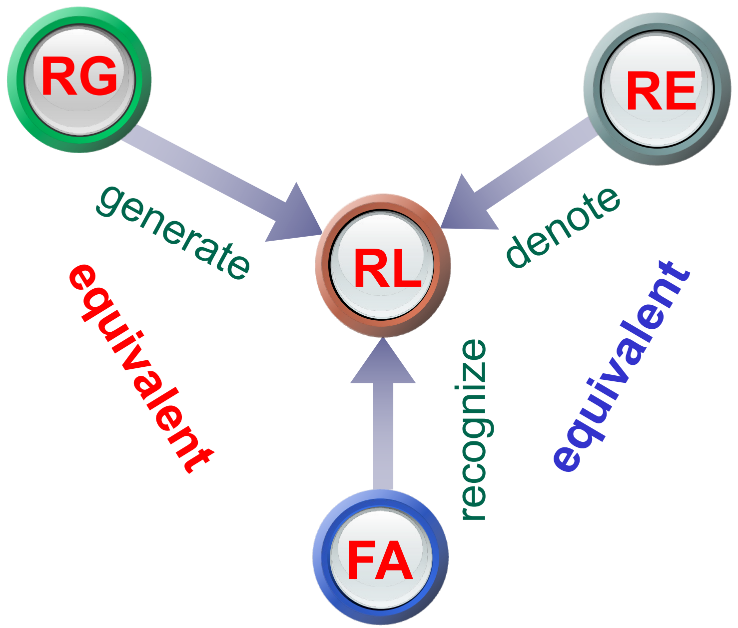
5.3 正则语言的封闭性

5.4 正则语言的判定算法

5.5 自动机的等价性与最小化

5.1 FA与RG的等价性





5.1 FA与RG的等价性

RG的工作机制

$A_0 \Rightarrow a_1A_1$	对应产生式$A_0 \to a_1A_1$
$\Rightarrow a_1a_2A_2$	对应产生式$A_1 \to a_2A_2$
…	
$\Rightarrow a_1a_2\ldots a_{n-1}A_{n-1}$	对应产生式$A_{n-2} \to a_{n-1}A_{n-1}$
$\Rightarrow a_1a_2\ldots a_{n-1}a_n$	对应产生式$A_{n-1} \to a_n$

FA的工作机制

$q_0\ a_1a_2\ldots a_{n-1}a_n$	
$\vdash a_1q_1\ a_2\ldots a_{n-1}a_n$	对应$\delta(q_0, a_1)=q_1$
$\vdash a_1a_2q_2\ldots a_{n-1}a_n$	对应$\delta(q_1, a_2)=q_2$
……	
$\vdash a_1a_2\ldots a_{n-1}q_{n-1}a_n$	对应$\delta(q_{n-2}, a_{n-1})=q_{n-1}$
$\vdash a_1a_2\ldots a_{n-1}a_nq_n$	对应$\delta(q_{n-1}, a_n)=q_n$

$\delta(A_0, a_1)=A_1$
$\delta(A_1, a_2)=A_2$

……

$\delta(A_{n-2}, a_{n-1})=A_{n-1}$
$\delta(A_{n-1}, a_n)=A_n=f$，
其中，$f \in F$。

可见：正则文法的推导与 FA的状态转移可 **相互模拟.**

定理 5.1 FA接受的语言是正则语言。

构造出来的文法G与M等价吗？

证明：

(1) 根据FA M，构造 RG G。

基本思想是让RG的派生与DFA的状态转移相对应。

设DFA $M=(Q，\sum，\delta，q_0，F)$，

取右线性文法 $G=(Q，\sum，P，q_0)$，

$P=\{ q\rightarrow ap|\delta(q，a)=p \}\cup\{ q\rightarrow a|\delta(q，a)=p，p\in F \}$

（2）证明 L(G)=L(M)-{ε}。

对于$a_1a_2\ldots a_{n-1}a_n \in L(G)$，其中 $a_1a_2\ldots a_{n-1}a_n\in\Sigma^+$

有：

$q_0\Rightarrow^+ a_1a_2\ldots a_{n-1}a_n$

$\Leftrightarrow q_0\rightarrow a_1q_1$，$q_1\rightarrow a_2q_2$，…，

$q_{n-2}\rightarrow a_{n-1}q_{n-1}$，$q_{n-1}\rightarrow a_n\in P$

正则文法领域

$\Leftrightarrow\delta(q_0, a_1)=q_1$，$\delta(q_1, a_2)=q_2$，…，$\delta(q_{n-2}, a_{n-1})=q_{n-1}$，$\delta(q_{n-1}, a_n)=q_n$，且$q_n\in F$

$\Leftrightarrow\delta(q_0, a_1a_2\ldots a_{n-1}a_n)= q_n\in F$

$\Leftrightarrow a_1a_2\ldots a_{n-1}a_n\in L(M)$

自动机领域

（3）关于ε句子。

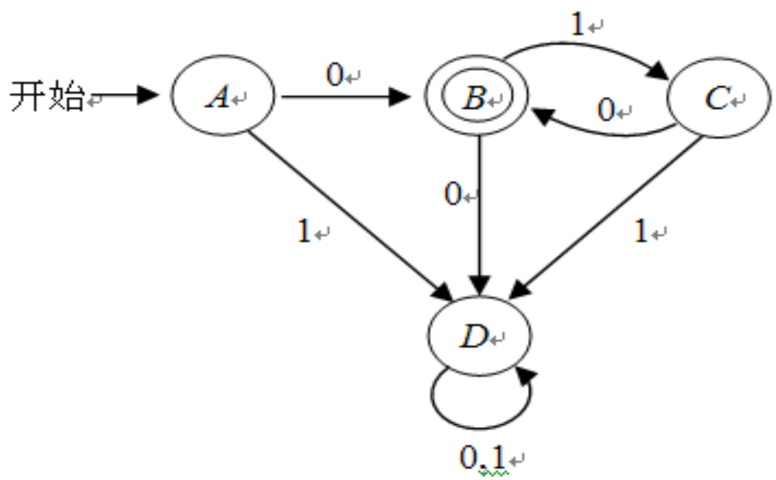
- 如果$q_0 \notin F$，则$\varepsilon \notin L(M)$，$L(G)=L(M)$。
- 如果$q_0 \in F$，则扩充G为G′，

  $L(G')=L(G)\cup\{\varepsilon\}=L(M)$。其中，G′比G增加一个开始符号S和两个产生式$S \to q_0 \mid \varepsilon$。

综上所述，对于任意DFA M，存在正则文法G，使得$L(G)=L(M)$。

定理得证。

EXP5.1 将下列DFA转化为等价的正则文法。



按照定理5 .1的构造方法，得出对应的正则文法是（A为开始符号）：

A→0B|0 , A→1D ,
B→0D ,　B→1C ,
C→0B|0 , C→1D ,
D→0D ,　D→1D 。

定理5.2 正则语言可以由FA接受。

证明：

（1）根据RG，构造FA。

基本思想：让FA**模拟**RG的派生过程。

设G=(V，T，P，S)，且$\varepsilon \notin L(G)$，

取FA M=(V∪{f}，T，δ，S，{f})，$f \notin V$。

5.1 FA与RG的等价性

定义转移函数如下：

$$\delta(A，a)=\begin{cases} \{B \mid A\to aB \in P\}\cup\{f\}， & 如果A\to a\in P \\ \{B \mid A\to aB \in P\}， & 如果A\to a\notin P \end{cases}$$

其中 $a \in T$

也就是:

- 用$B\in\delta(A，a)$与产生式 $A\to aB$ 对应;
- 用 $f\in\delta(A，a)$与产生式 $A\to a$ 对应。

思考：为什么要定义接受状态{ f }呢？

（2）证明L(M)=L(G)

对于$a_1a_2\ldots a_{n-1}a_n \in T^+$，

$a_1a_2\ldots a_{n-1}a_n \in L(G) \Leftrightarrow S \Rightarrow^+ a_1a_2\ldots a_{n-1}a_n$

$\Leftrightarrow S \Rightarrow a_1A_1 \Rightarrow a_1a_2A_2 \Rightarrow \ldots$

$\Rightarrow a_1a_2\ldots a_{n-1}A_{n-1} \Rightarrow a_1a_2\ldots a_{n-1}a_n$

$\Leftrightarrow S \to a_1A_1$，$A_1 \to a_2A_2$，…，

$A_{n-2} \to a_{n-1}A_{n-1}$，$A_{n-1} \to a_n \in P$

5.1 FA与RG的等价性

$\Leftrightarrow A_1 \in \delta(S，a_1)，A_2 \in \delta(A_1，a_2)，\ldots，$

$A_{n-1} \in \delta(A_{n-2}，a_{n-1})，f \in \delta(A_{n-1}，a_n)$

$\Leftrightarrow f \in \delta(S，a_1 a_2 \ldots a_{n-1} a_n)$

$\Leftrightarrow a_1 a_2 \ldots a_{n-1} a_n \in L(M)$

5.1 FA与RG的等价性

例5.2 给出正则文法G_1如下：

$$S \to 0B, \quad B \to 0B,$$

$$B \to 1S, \quad B \to 0。$$

根据定理5.2给出的方法，我们构造对应的有穷自动机 $M=(\{S,B,f\},\{0,1\},\delta,S,\{f\})$,其中：

$$\delta(S,0)=\{B\}, \quad \delta(B,0)=\{B,f\},$$

$$\delta(B,1)=\{S\}。$$

例5.3 给出正则文法G_2如下：

$S \to 0A$， $A \to 1A$，

$A \to B$, $B \to 0$, $B \to \varepsilon$。

我们构造对应的有穷自动机 $M = (\{S,A,B,f\},\{0,1\},\delta,S,\{f\})$,其中：

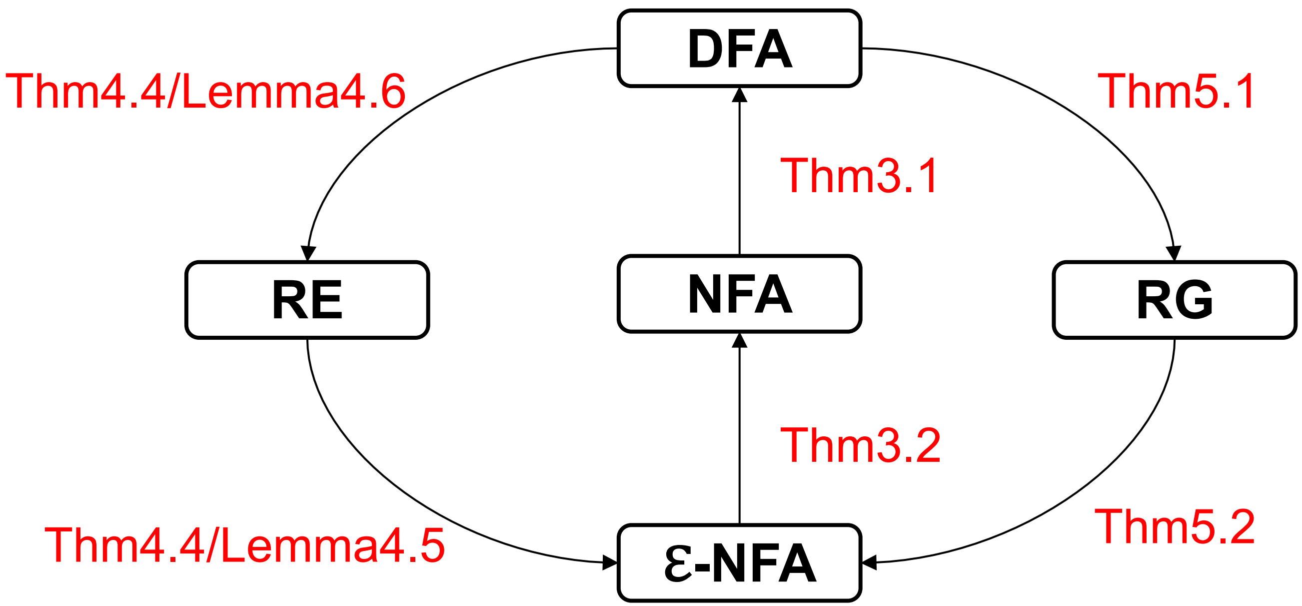
$\delta(S,0)=\{A\}$， $\delta(A,1)=\{A\}$，

$\delta(A,\varepsilon)=\{B\}$， $\delta(B,0)=\{f\}$， $\delta(B,\varepsilon)=\{f\}$。

注意：这个有穷自动机 M是具有ε-转移功能的。由于要考虑到一般情况，所以定理5.2 中必须要构造一个具有ε-转移的NFA M，才能接受一切由正则文法产生的语言。

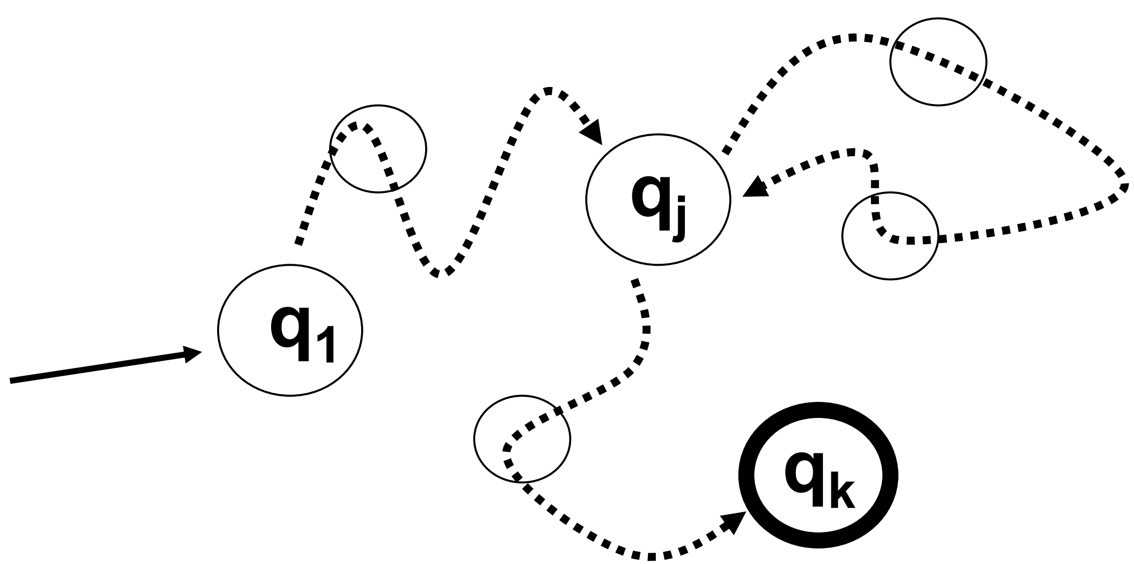
正则语言各种表达形式之间的关系

- 有向边代表两种表达式之间的**构造关系**



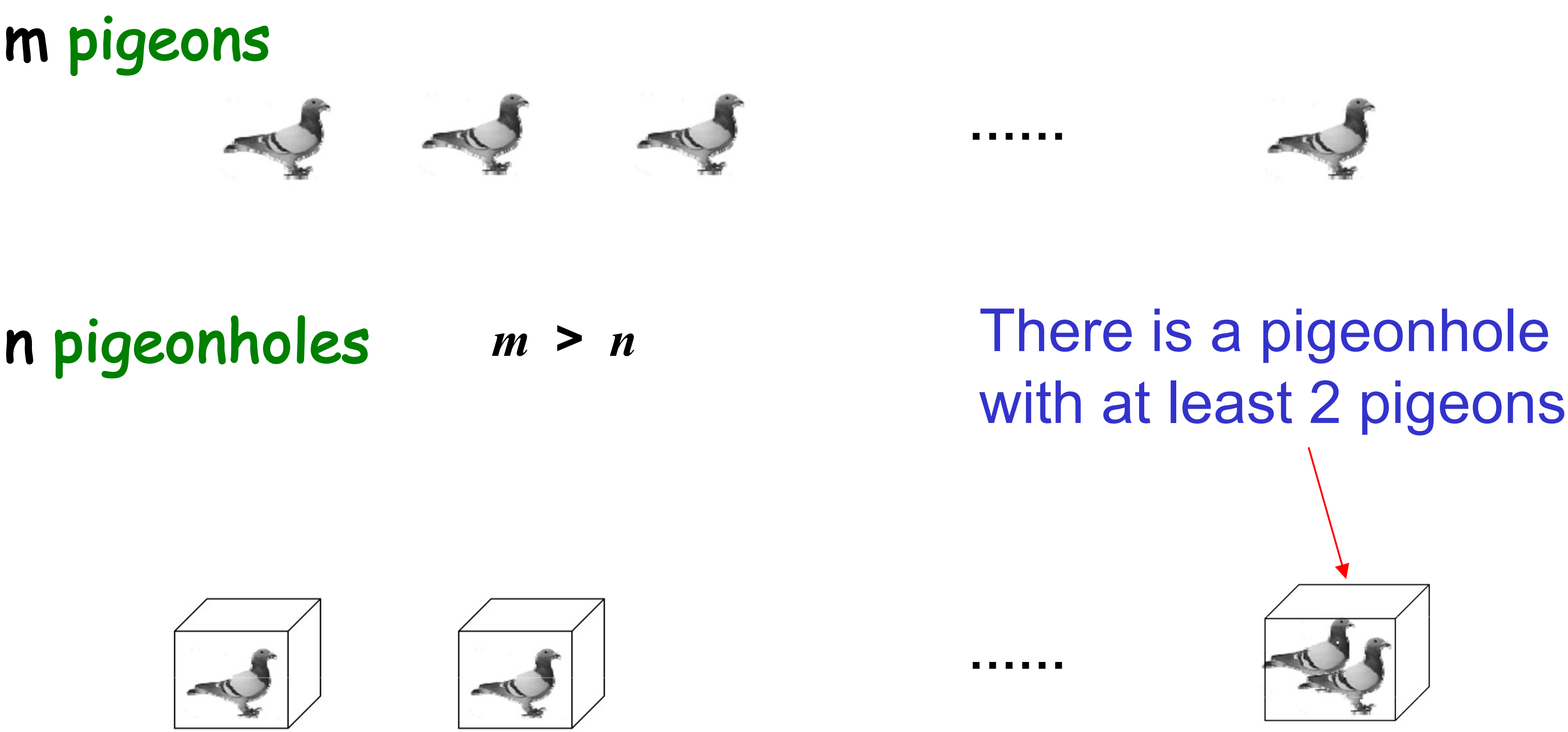
5.2 正则语言的泵引理

1. 有穷语言 *一定*是正则的，无穷语言 *可能*是正则的。
2. **如何判断一个无穷语言是否正则呢？**
3. 正则语言是靠***打圈***，来描述（有某种规律的）无限集合；



5.2 正则语言的泵引理

Pigeonhole principle(鸽巢原理): If m pigeons are placed into fewer than m holes, some hole has to have more than one pigeon in it.



5.2 正则语言的泵引理

The DFA Principle

m symbols

$$w = a_1 a_2 \cdots\cdots a_m$$

n states

$$a_n \cdots\cdots a_m \,?$$

$$m \geq n$$

$$\rightarrow (q_1) \xrightarrow{a_1} (q_2) \xrightarrow{a_2} (q_3) \cdots \xrightarrow{a_{n-2}} (q_{n-1}) \xrightarrow{a_{n-1}} (q_n) \xrightarrow{a_n} (?)$$

Property of regular languages

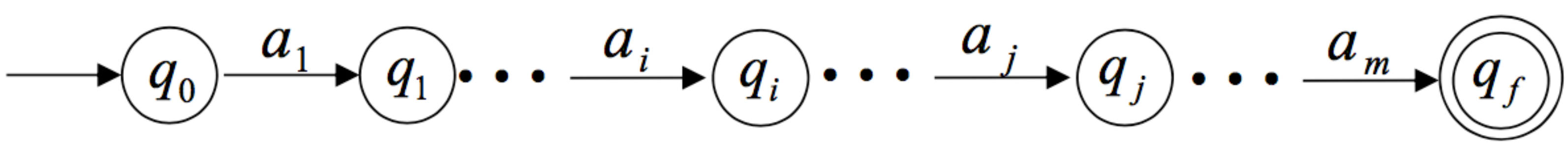
L is a regular language $\Rightarrow \exists$DFA $A : L(A) = L$

Let $A = (\mathrm{Q}, \Sigma, \delta, \mathrm{q}_0, F)$, and $n = |Q|$

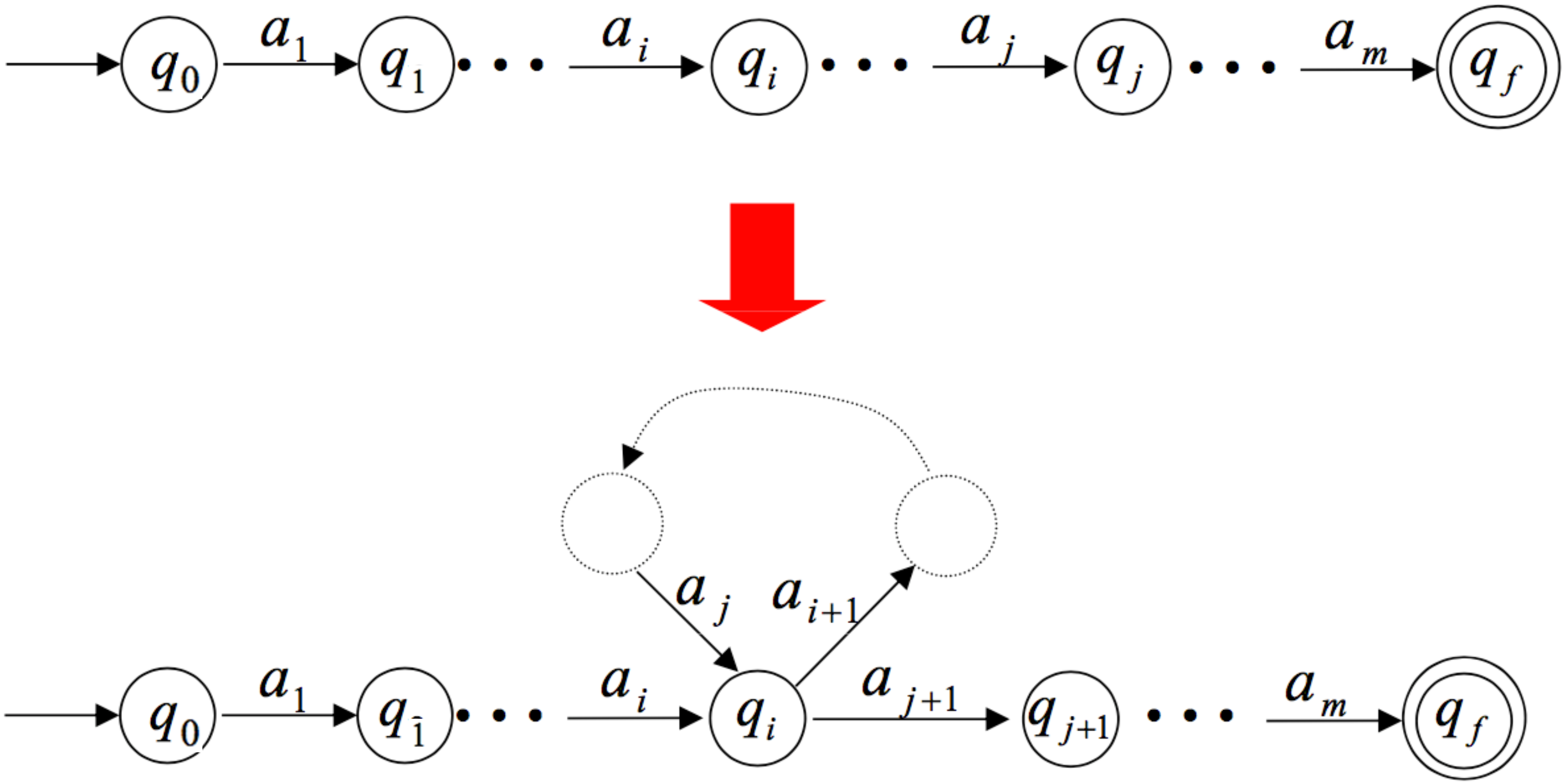
Get $w \in L$, and suppose $w = a_1 a_2 \cdots\cdots a_m, m \geq n$

Let $q_{\mathrm{i}} = \overline{\delta}(q_0, a_1 a_2 \cdots a_i)$

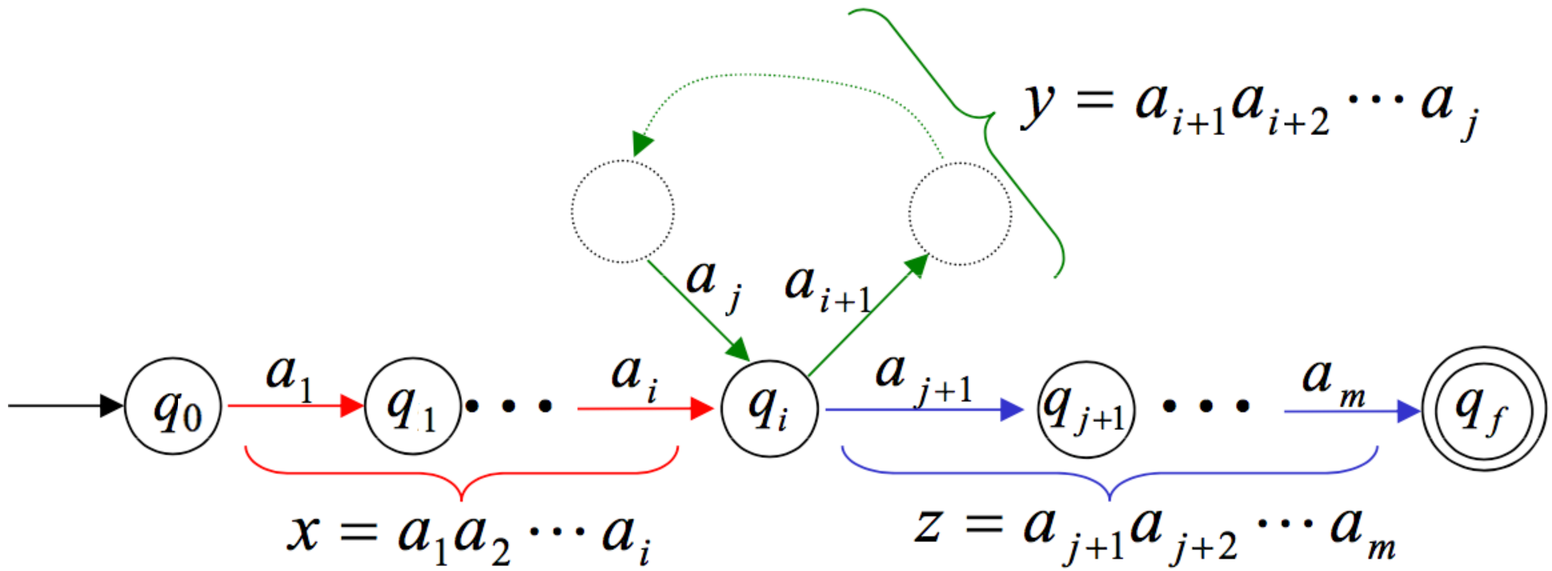
$$\Rightarrow \exists 0 < i < j \leq n : q_i = q_j$$



Property of regular languages



Property of regular languages



$$\Rightarrow w = xyz \begin{cases} |xy| \le n \\ |y| \ge 1 \text{ or } y \ne \varepsilon \\ xy^kz \in L, \text{ for any } k \ge 0 \end{cases}$$

同濟大學 TONGJI UNIVERSITY

5.2 正则语言的泵引理

Pumping lemma: For every regular language L, there is a *pumping length* **p**, such that for any string $s \in L$ and $|s| \geq p$, we can write $s = xyz$ with

1) $x\,y^i z \in L$ for every $i \in \{0,1,2,\ldots\}$ 为什么打圈？
2) $|y| > 0$
3) $|xy| \leq p$ 什么时候打圈？

Note that

1) implies that $xz \in L$
2) says that y cannot be the empty string ε
3) is not always used

■ 经得起泵测试是**RL**的必要条件（不充分）。

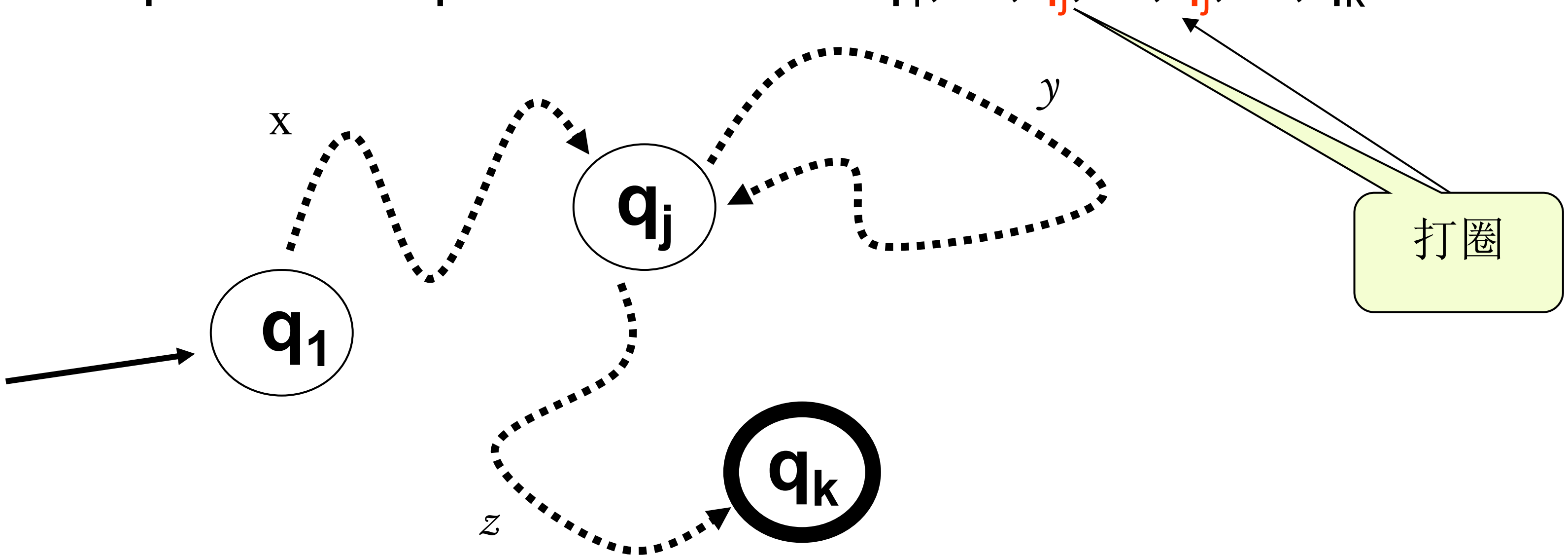
5.2 正则语言的泵引理

Proof Idea:

① Consider an accepting DFA M with size $|Q|$ 机器状态数

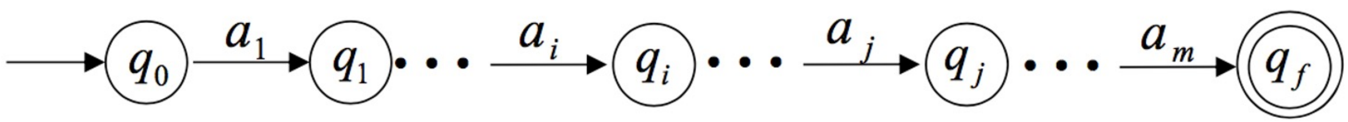
② On a string of length p, **p+1** states（识别某词的状态路径长度）

③ get visited for $p \geq |Q|$, there must be q_j ,such that the computational path looks like: $q_1,\dots,q_j,\dots,q_j,\dots,q_k$

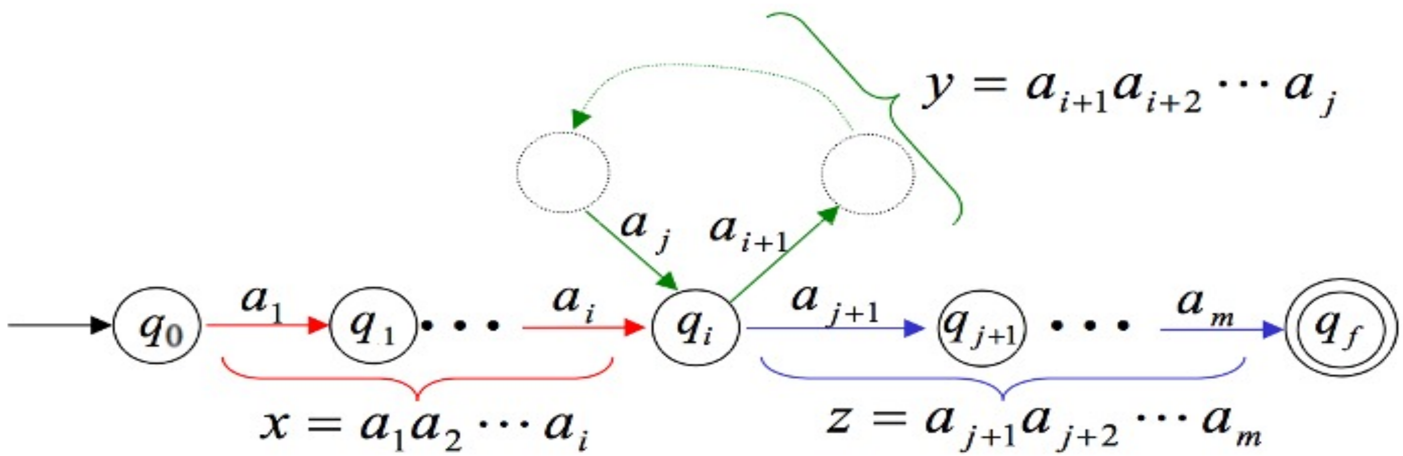


5.2 正则语言的泵引理

设p=｜Q｜=n； $s=a_1a_2\ldots a_m$, 其中 m>n； q_f 是可接受状态。



根据鸽巢原理，上图中一定有两个状态相同：$q_i = q_j$, 其中 $0 < i < j \le n$.



由上图可知：

1. $s = xy^kz,\ \forall k \ge 0 \Rightarrow\ s\in L$;
2. $\because i \ne j,\ \therefore |y| > 0$;
3. $\because j \le n,\ \therefore |xy| \le n$.

Proof Let M = (Q, Σ, δ, q_1, F) be a DFA recognizing Language A and p be the number of states of M. (p = |Q|)

Let s=$s_1s_2\cdots s_n$ be a string in A of length n, where $n \geq p$. Let $r_1,\ldots,r_{n+1}$ be the sequence of states that M enters while processing s, so $r_{i+1} = \delta(r_i, s_i)$ for $1 \leq i \leq n$. This sequence has length n+1, which is at least p+1. Among the first p + 1 elements in the sequence, two must be the same state, by the pigeonhole principle. We call the first of these r_j and the second r_k. Because r_k occurs among the first p+1 places in a sequence starting at r_1, we have $k \leq p+1$. Now let $x = s_1 \cdots s_{j-1}$, $y = s_j \cdots s_{k-1}$, and $z = s_k \cdots s_n$.

As x takes M from r_1 to r_j, y takes M from r_j to r_j, and z takes M from r_j to r_{n+1}, which is an accept state, M must accept xy^iz for $i \geq 0$. We know that $j \neq k$, so $|y| > 0$; and $k \leq p+1$, so $|xy| \leq p$. Thus we have satisfied all conditions of the pumping lemma.

EXP1:Prove B = $\{0^n1^n \mid n\geq 0\}$ is not regular.

1. Assume that B is regular 反证法
2. Let p be the pumping length, and s = $0^p1^p \in B$
   s = xyz = 0^p1^p, with $xy^iz \in B$ for all $i\geq 0$
   Three options for y:
   1) $y=0^k$, hence $xyyz = 0^{p+k}1^p \notin B$
   2) $y=1^k$, hence $xyyz = 0^p1^{k+p} \notin B$
   3) $y=0^k1^l$, hence $xyyz = 0^p1^l0^k1^p \notin B$
3. Conclusion: The pumping result does not hold, the language B is not regular.

5.2 正则语言的泵引理

EXP2, Show F = { ww | w$\in$\{0,1\}* } is not RL.

反证法

Let p be the pumping length, and take word s = 0^p10^p1 , $w=0^p1$

Let s = xyz = 0^p10^p1,

with condition 3) $|xy|\leq p$

Only one option: $x=0^{p-k}$,$y=0^k$, $z=10^{p-k}1$, （保证xz in L)

with $xyyz = 0^{p+k}10^{p-k}1 \notin F$

Without 3) this would have been a pain.

思考题：

1. $L = \{ 0^n1^n \mid 0 \le n \le 100 \}$ 是正则语言吗？

2. 有限语言是否符合泵引理？

5.2 正则语言的泵引理

EXP3, Show $E=\{0^i1^j \mid i > j\}$ is not RL. *证明思路回顾*

Step 1: 选择反证法；

Step 2：构造 string $s = 0^{p+1}1^p$; 利用泵长度p

Step 3：发现矛盾

$s=xyz$，由引理3）$|xy|\le p$知：$y=0s$，令 $y=0^k$, $k>0$；

y不可能包含1

$xy^iz = 0^{p-k}0^{k*i+1}1^p = 0^{p+k(i-1)+1}1^p$，当i=2时，显然 $xyyz \in E$;

可惜没矛盾，只好换一条路走

Pumping Down：The pumping lemma states that $xy^iz \in E$ even if when i=0, so lets consider the string $xy^0z=xz$.

结果怎么样呢？

$xz=0^n1^p$, $\because y=0^k$, $k>0$, $\therefore n\le p$, 即0的个数不比1的个数多，显然，这与$s = 0^{p+1}1^p$相互矛盾 $\therefore xz \notin E$。

Step 4：得出结论。

同济大学 TONGJI UNIVERSITY

再论泵引理

Pumping lemma: For every regular language L, there is a *pumping length* **p**, such that for any string $s \in L$ and $|s| \geq p$, we can write $s=xyz$ with

1) $x y^i z \in L$ for every $i \in \{0,1,2,\ldots\}$ 为什么打圈？鸽子比笼子多。
2) $|y| > 0$
3) $|xy| \leq p$ 什么时候打圈？字符数 ≥ 状态数。

Note that

1) implies that $xz \in L$，正则语言靠打圈
2) （1）$y \neq \varepsilon$，但x,z可以为空；（2）如果$y = \varepsilon$,引理也成立，只是毫无意义；（3）y不能为空，因为打的不是空圈（此q_i非彼q_j）。
3) $|xy| = p$ 时，至少打圈一次。

泵引理是正则语言必须满足的条件（必要条件，不是充分条件）。

再论泵引理

EXP4: 证明 L = {1^n|n 是素数 } 不是正则语言。

证明：假设 L 是正则语言，则存在 p 满足泵引理的性质。那么由于串 w = xyz = 1^{p_0} （其中 p_0 为大于 p 的素数）属于 L, 串 w′ = xy^kz = $1^{p_0+(k-1)|y|}$ 也属于 L。取 k = p_0 + 1， w′ = $1^{p_0(1+|y|)}$ 显然不属 于 L，产生矛盾，因此 L 不是正则语言。

注意：这里使用了素数有无穷多个的引理。

课堂作业：试证 L = {1^n|n 是合数 } 不是正则语言。

狄利克雷定理 对于任意互质的正整数 a, d，形式如 a + nd 的素数有无限多个，其中 n 为正整数。

再论泵引理

课堂作业：试证 $L = \{1^n | n$ 是合数 $\}$ 不是正则语言。

> **狄利克雷定理** 对于任意互质的正整数 a，d，形式如 a + nd 的素数有无限多个，其中 n 为正整数。

证明： 若 L 是正则语言，则存在 p 满足泵引理的性质。

那么由于串 $w = xyz = 1^{p_0 \cdot p_0}$ (其中 p_0 为大于 p 的素数)属于 L，串 $w' = xy^kz = 1^{p_0 \cdot p_0 + (k-1)|y|}$ 也属于 L。

由于 $|y| \le p$，p_0^2 和 $|y|$ 互素，由狄利克雷定理，存在正整数 k 使得 $p_0^2 + (k - 1)|y|$ 为素数，有 w' 不属于 L，产生矛盾，因此 L不是正则语言。

5.3 正则语言的封闭性

定义 5.1 如果属于某个语言类的任何语言在某个特定运算下所得的结果仍然属于该语言类，则称该语言类对这个运算是封闭的，并称该语言类对这个运算具有**封闭性**。

可以证明：

①两个正则语言的**并**是正则语言。
②两个正则语言的**连接**是正则语言。
③正则语言的**闭包**是正则语言。
④两个正则语言的**交**是正则语言。
⑤正则语言（对全集）的**补**是正则语言。
⑥两个正则语言的**差**是正则语言。
⑦正则语言的**逆转**是正则语言。
⑧正则语言的**同态**是正则语言。
⑨正则语言的**逆同态**是正则语言。

5.4 正则语言的判定

- "给定的DFA接受的集合是空集吗？"
- "给定的DFA接受的集合是有穷集（或无穷集）吗？"
- "给定两个DFA是否接受同一个集合（是否等价）？"
- "给定一个DFA M和一个字符串x，M能接受x吗？"

对于这些问题，是否存在一个算法能回答"是"或者"否"，如存在，则称该算法为判定算法，相应的问题称为可判定的，否则称为不可判定的。

5.4 正则语言的判定

1. 判定（decide）和识别（recognize）的区别？

2. 可判定（decidable）与可计算是等价的；

3. 一个问题能否被一个算法解决，也就是这个问题是否可计算；

4. 通常，我们将计算问题转化成语言的归属问题；

EXP：问题1: 检测DFA B是否接受输入字符串ω？

等价于 问题2: 检测<B,ω>是否属于语言A_{DFA}？

$A_{DFA}=\{\langle B,\omega\rangle \mid B \text{ is a DFA that accepts input string } \omega\}$

5.4 正则语言的判定

定理 5.10 设DFA M=(Q，Σ，δ，q_0;，F)，L=L(M)非空的充分必要条件是：存在x∈Σ*，|x|<|Q|，δ(q_0，x)∈F。

定理5.11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∈Σ*，|Q|≤|x|<2|Q|，δ(q_0，x)∈F。

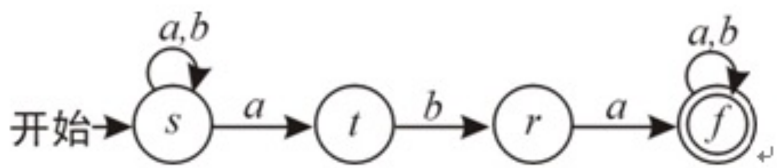
定理 5.12 设DFA M_1=(Q_1，Σ，δ_1，q_{01}，F_1)，DFA M_2=(Q_2，Σ，δ_2，q_{02}，F_2)，则存在判定M_1与 M_2是否等价的算法。

定理 5.13 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。

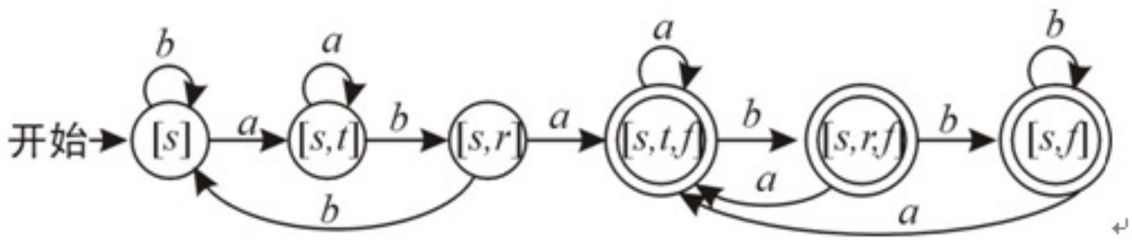
关于判定的详细内容，参见《Introduction to the Theory of Computation 》Part Two: Computability Theory.

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。

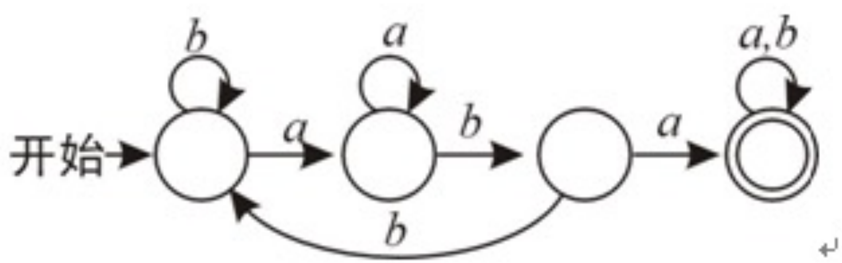


构造一个等价的DFA，需要设2^4=16个状态，除去不可到达的状态外，还剩下6个状态。这个DFA如下图所示。



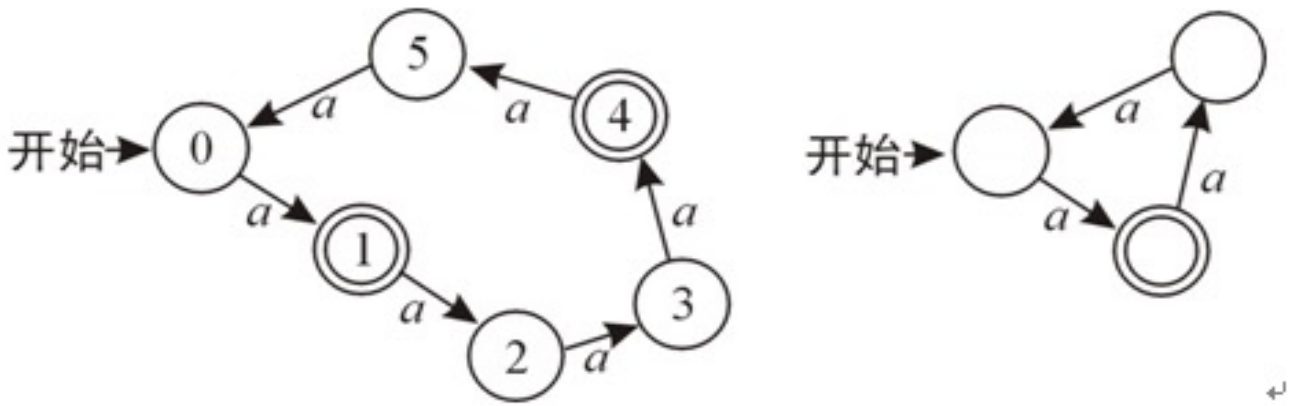
5.5 自动机的等价性与最小化

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：



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$\{a,a^4,a^7,a^{10},\ldots\}$。

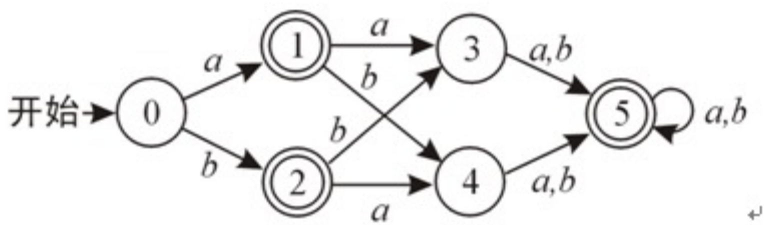


左边的图所示的DFA共用6个状态，显然是太多了。我们可以将状态1，4合并，状态2，5合并，状态0，3合并，得出的DFA只用3个状态，也能接受同样的集合。这个化简后的DFA示于原图的右边。

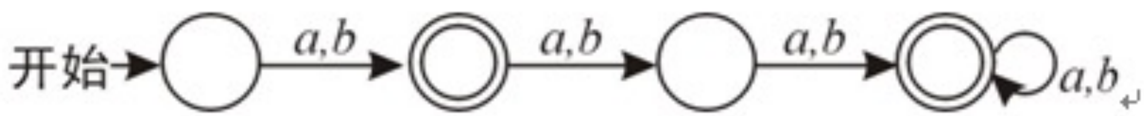
同濟大學 TONGJI UNIVERSITY

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：

是否存在最小化的**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**

  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。

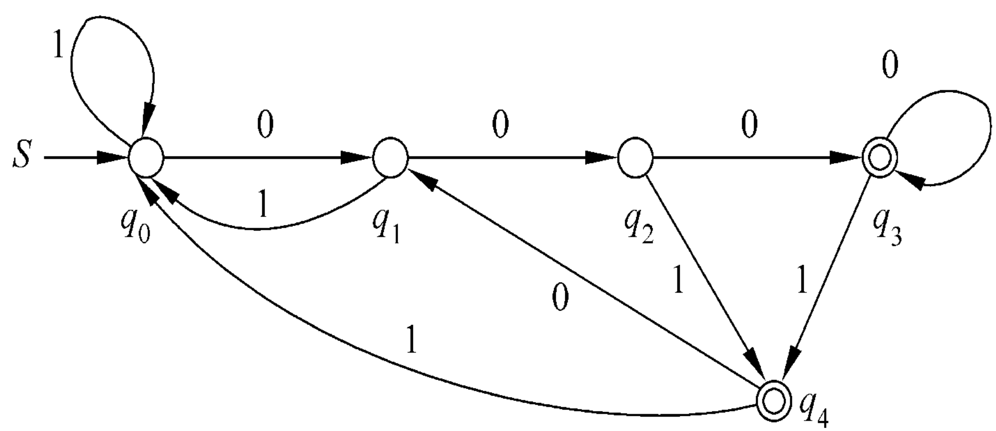
➢ 等价关系R将S分成的等价类的个数称为R在S上的**指数**。

5.5 自动机的等价性与最小化

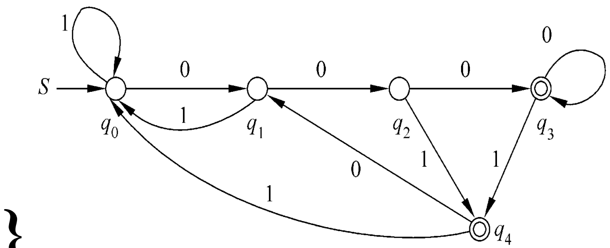
定义5.6-1 能引导FA从开始状态q_0到达状态q的所有字符串的集合为：

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

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 | x∈Σ^*，x= ε 或者x以1结尾}
set(q_1)={x | x∈Σ^*，x=0或者x以10结尾}
set(q_2)={x | x∈Σ^*，x=00或者x以100结尾}
set(q_3)={x | x∈Σ^*，x以000结尾}
set(q_4)={x | x∈Σ^*，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；

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，

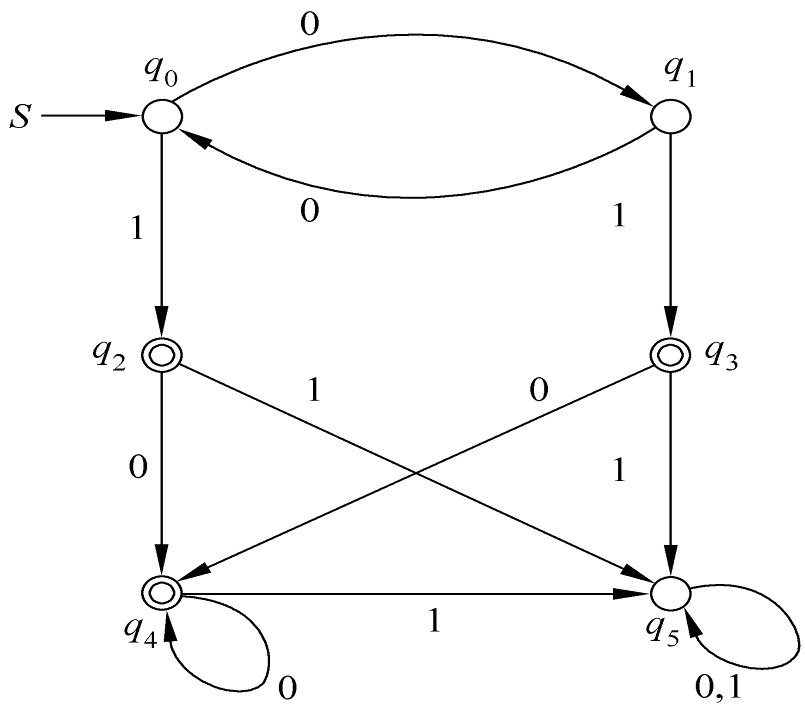
根据定义5.6-1而得

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

$x\ R_M\ y$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， $q \in Q$ 。

同濟大學 TONGJI UNIVERSITY

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，$m \geqslant 0$；

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。

5.5 自动机的等价性与最小化

定义5-7 语言L确定的Σ*上的关系R_L。

对于∀x，y∈Σ*，

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

即：对于∀x，y∈Σ*，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。

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？

同濟大學 TONGJI UNIVERSITY

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。

5. 5 自动机的等价性与最小化

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. $S_4 = \Sigma^*11 \cup 1 \cup \varepsilon$

思考:

L是正则语言吗？为什么？

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. $S5 =\{0^n1^{n-3} : n\geq 3\}$

……

思考：

B是正则语言吗？为什么？

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。

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$x\ R_L\ y$，$y\ R_L\ z$。

$x\ R_L\ y \Leftrightarrow$ (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。

(2) R_L 是右不变的。

设 $x\ R_L\ y$。由 R_L 的定义，对 $\forall w$，$v \in \Sigma^*$，$xwv \in L$ $\Leftrightarrow$ $ywv \in L$，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。

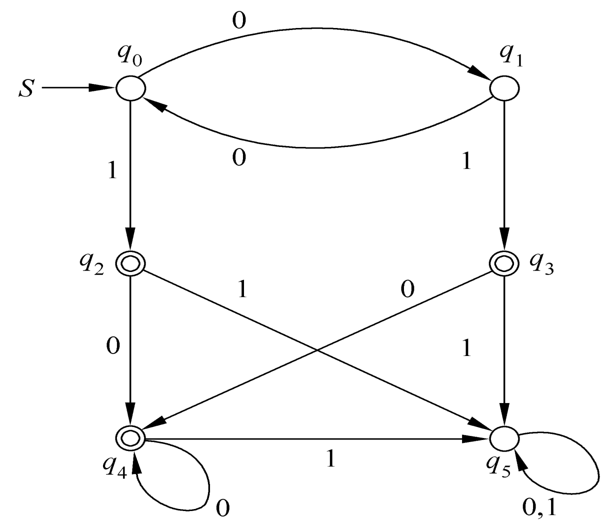
定义5-9

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）*



$set(q_0)=\{(00)^n \mid n\geq 0\}$；

$set(q_1)=\{0(00)^n \mid n\geq 0\}$；

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

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

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

$set(q_5)=\{x \mid x$为至少含两个1的串$\}$。

同濟大學 TONGJI UNIVERSITY

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？）；

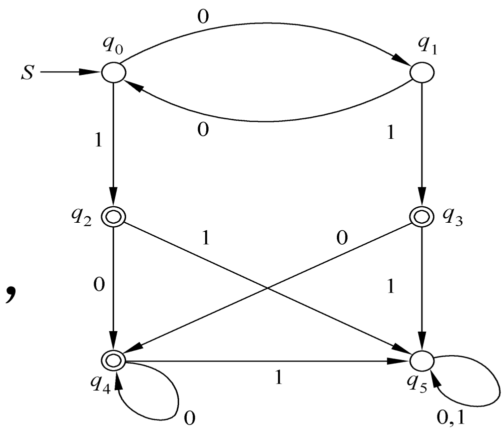
即对于任意DFA $M=(Q, \Sigma, \delta, q_0, F)$。必有：

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

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z∈Σ*，当z含且只含一个1时，00z∈L(M)，000z∈L(M)；当z不含1或者含多个1时，00z∉L(M)，000z∉L(M)。这就是说，对于任意的z∈Σ*，**00z∈L(M) ⇔ 000z∈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。

同济大学 TONGJI UNIVERSITY

(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。

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

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

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

对于任意的$z \in \Sigma^*$，z要么不含1，要么含有1。当z不含1时，$1z \in L(M)$，$10z \in L(M)$；当z含有1时，$1z \notin L(M)$，$10z \notin L(M)$。这就是说，对于任意的$z \in \Sigma^*$，$1z \in L(M) \Leftrightarrow 10z \in L(M)$。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。

5.5 自动机的等价性与最小化

(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。

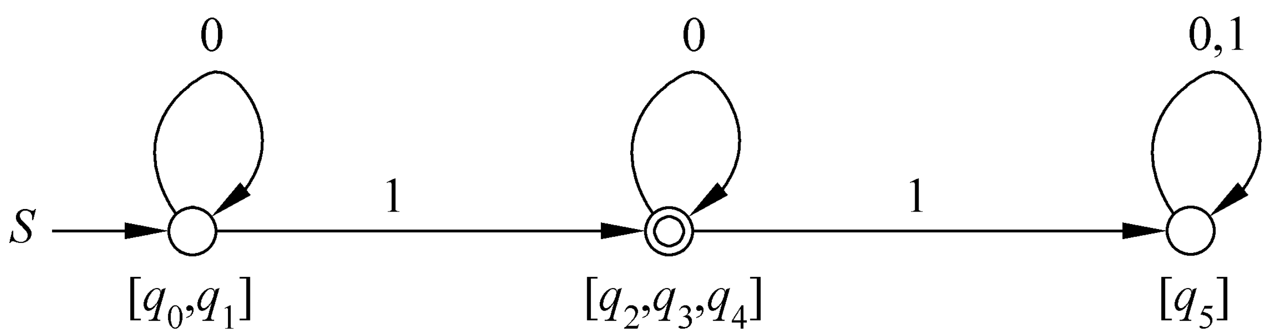
5. 5 自动机的等价性与最小化

综合上述分析，得：

$\Sigma^*/R_{L(M)}$={ set(q_0)∪set(q_1)， set(q_2)∪set(q_3)∪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： [ε] = set(q_0)∪set(q_1)=0^*；
2. 含一个1：[1] = set(q_2)∪set(q_3)∪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

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。

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的对应于M的接受状态的等价类的并。

$\Updownarrow$

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？

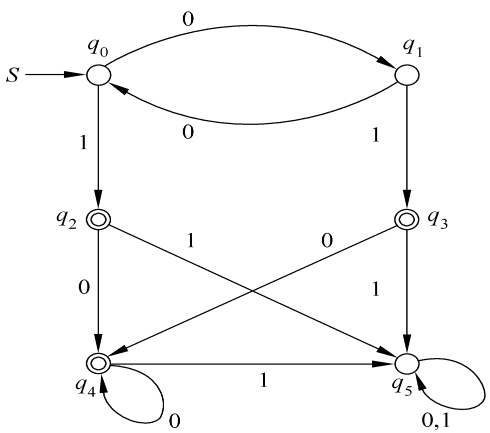
根据R_L的定义，

$x\ R_L\ y$

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。

由(3)可以推出(1)

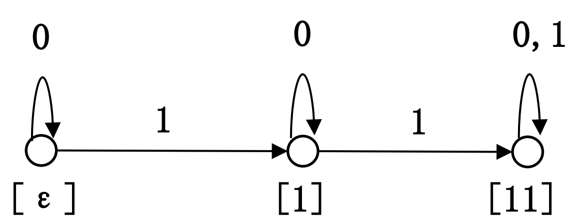
思路：已知R_L具有有穷指数，证存在DFA M，使得L(M) = L。



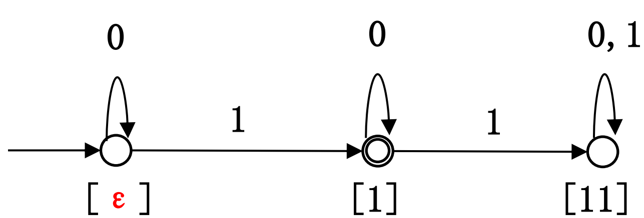
$\Sigma^*/R_{L(M)}$={ set(q_0)∪set(q_1)，set(q_2)∪set(q_3)∪set(q_4)，set(q_5)}

不妨采用新的符号标记这3个等价类：

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



等价类的转换



DFA的状态转移

关键：①等价类对应状态，②含有 ε 的等价类对应M的起始状态。

同濟大學 TONGJI UNIVERSITY

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，使得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.

可以得到如下一些等价类

[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；

[ε]——ε所在的等价类；

[10]= $\{x \mid x=0^n1^m\ (m>n)$或者x中含子串10$\}$

…

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$p\in Q_m$，$a\in\Sigma$,

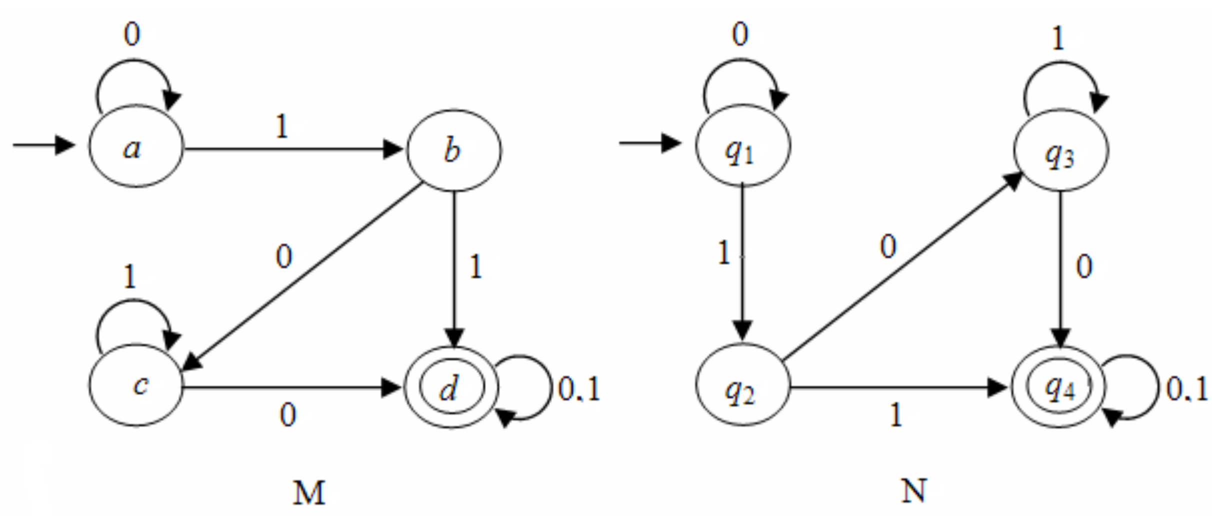
（3）$p\in F_m$ 当且仅当 $f(p)\in F_n$ 。

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, 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 。

极小化算法

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×），表示(p,q)是可以区分的。

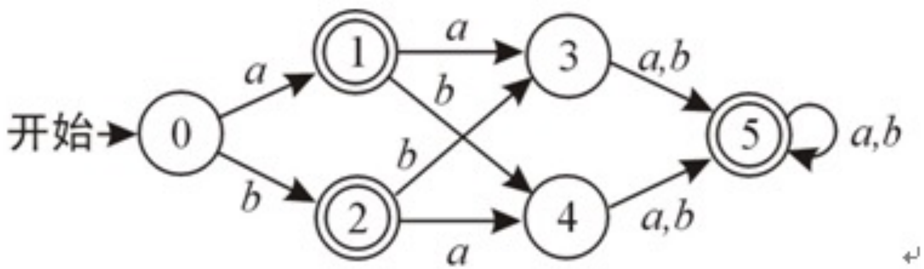
③重复下述过程，直到表中内容不再改变为止：

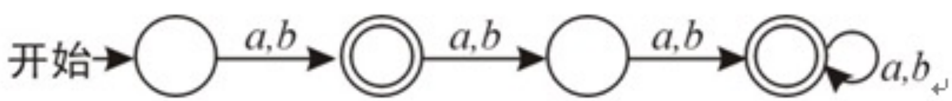
如果存在一个未被标记的状态对(p,q)，且对于某个a∈∑，如果（r=δ(p,a)，s=δ(q,a)）已做了标记，则在(p,q) 相应的格子内做标记。

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，状态p和状态q可以合并。

3. 5 自动机的等价性与最小化

例 5.15 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 5.15 对下图 给出的DFA用极小化算法进行化简。

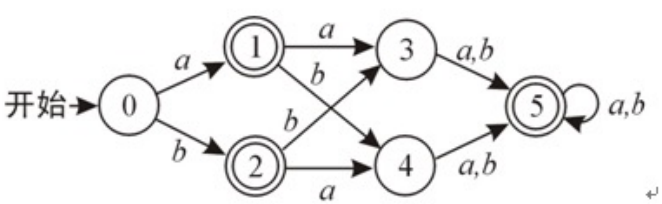
- 在算法的第1步，我们对该DFA中的6个状态建立一个空白表，表中所有格子皆为空。因为p和q等价是对称的，所以只用表的下三角部分即可（阶梯形的）。这张表如左下图所示。
- 在算法的第2步之后，对接受状态和非接受状态的状态对的格子内做了标记，结果如右下图所示。

1					
2					
3					
4					
5					
	0	1	2	3	4

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

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）。

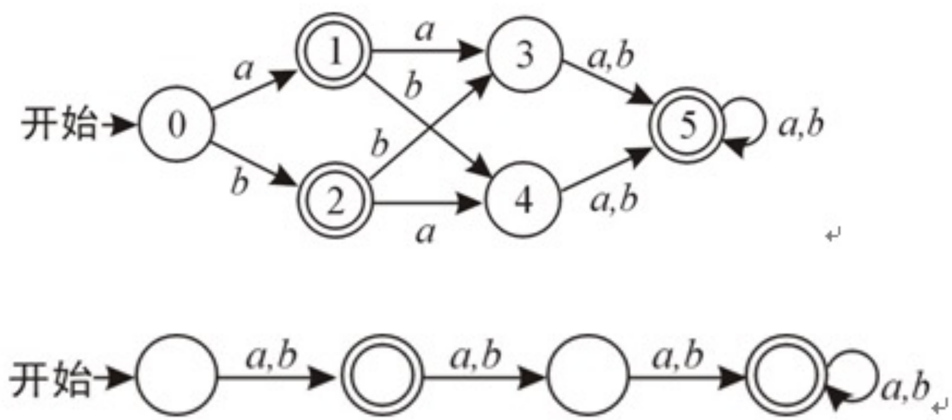


1	×				
2	×				
3		×	×		
4		×	×		
5	×	×	×	×	×
	0	1	2	3	4

同濟大學 TONGJI UNIVERSITY

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。

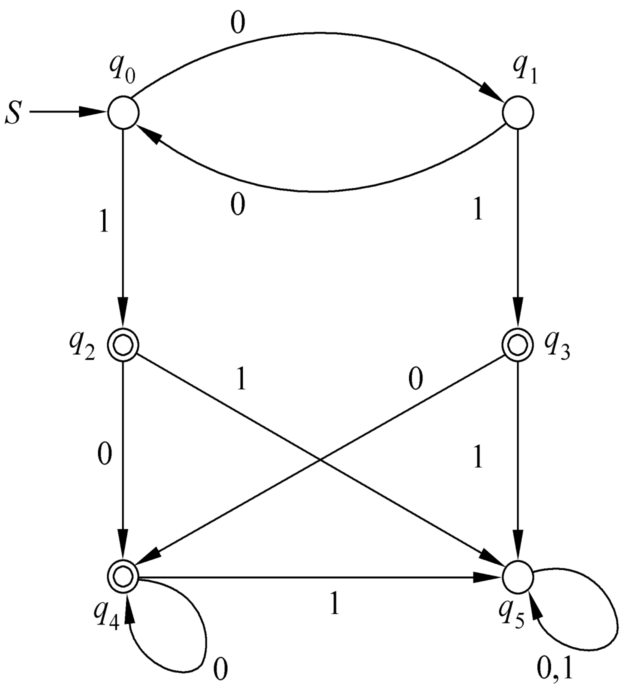


1	×				
2	×				
3	×	×	×		
4	×	×	×		
5	×	×	×	×	×
	0	1	2	3	4

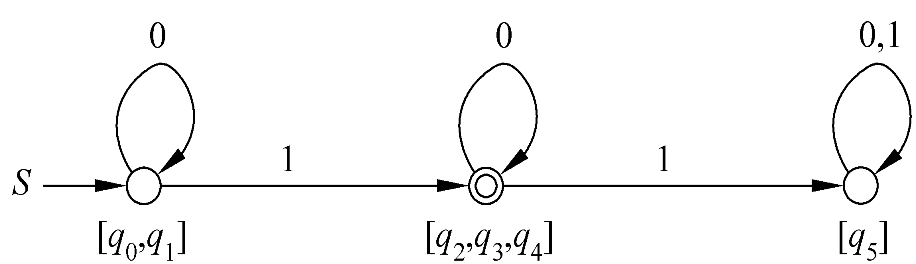
同濟大學 TONGJI UNIVERSITY

3. 5 自动机的等价性与最小化

例5. 16 对下图所示的DFA进行极小化。



1					
2					
3					
4					
5					
	0	1	2	3	4



Myhill–Nerode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

总　结

如何求R_L的等价类

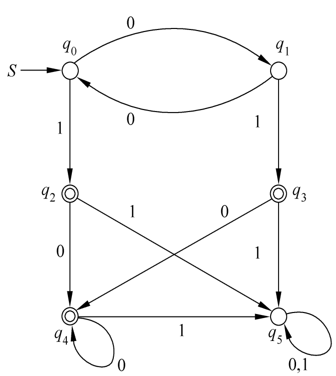
方法一：间接法（适用于L是正则语言）

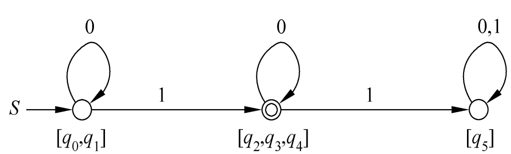
　如果L是正则语言，DFA M识别语言L。先根据DFA M求R_M的等价类，再对R_M的等价类进行合并，得到R_L的等价类。

说明：

1. $\{set(q)|q\in Q\}=\Sigma^*/R_M$，即$\{set(q)|q\in Q\}$就是$R_M$的等价划分，见定义**5.6-1**。

2. DFA的极小化算法（主要是合并不可区分的状态部分）。





总　结

方法二：直接法

根据语言的**结构特征**，对$\sum^*$进行划分。举例如下：

例17. L1={w∈{0,1}*|w包含的数字之和能被3整除}

首先，分析句子的结构特征---和数模3取0，这是关键步骤。

[0]：模3取0，比如00，011100
[1]：模3取1，比如001，1111
[2]：模3取2，比如11，011

特别提醒：

1、**必须考虑[ε]**（表示ε所在的等价类），因为自动机起始状态的输入总是ε。例17中的[ε]= [0]，这是比较特殊的情况。

2、求得的等价类是必需的，也就是不能再合并了。见例18的分析。

总 结

例18. $L2 = \{0w \mid w \in \Sigma^*, \Sigma=\{0,1\}\}$，求$R_L$的等价类。

分析句子的特征，都是以0开头的，这是关键步骤。

$[\varepsilon]=\{\varepsilon\}$

$[0]=\{x \mid x$以0开头的所有字符串$\}$

$[1]=\{x \mid x$以1开头的所有字符串$\}$

注意：

1. 这三个等价类并，正好是Σ^*。
2. 在R_L的约束下，这三个等价类不能再合并了，举例如下：

   令$z=\varepsilon$，因为$0z\in L2$，而$1z\notin L2$，根据R_L的定义，$0R_{L2}1$是不成立的；同理，$1\ R_{L2}\ \varepsilon$，$\varepsilon\ R_{L2}0$ 也是不成立的。所以，$[\varepsilon]$、$[0]$、$[1]$是不能再合并了。

同濟大學 TONGJI UNIVERSITY

总　结

例19. L3={Σ*0Σ｜ Σ={0, 1}}，求R_L的等价类。

$A1 = \Sigma^{*}00$,
$A2 = \Sigma^{*}01$,
$A3 = \Sigma^{*}10 \cup 0$,
$A4 = \Sigma^{*}11 \cup 1 \cup \varepsilon$

总　结

例20. 语言$L=\{xwx^R \mid x, w \in\{0,1\}^+\}$ 是正则语言吗？

分析：语言L的特征有：①句子长度≥3，w非空；②头尾相同；③重点关注|x|=1的情况。

解法1: RE：$0(0+1)^+0 +1(0+1)^+1$

解法2: RG：

$S \longrightarrow 0A|0B$

$A \longrightarrow 0A|1A|00|10$

$B \longrightarrow 0B|1B|01|11$

解法3:M-N定理---句子长度≥3

[ε]={ε}

[0]={0}

[1]={1}

[01]={x|x以0开头，1结尾，长度≥2}∪{00}

[10]={x|x以1开头，0结尾，长度≥2}∪{11}

[00]={x|x以0开头，0结尾，长度≥3}

[11]={x|x以1开头，1结尾，长度≥3}

总　结

例20. 语言L={xwx^R | x, w ∈{0,1}$^+$} 是正则语言吗?

{xwx^R|x,w∈{0,1}+}不是正则语言

取p为泵长，不妨设$x=01^{\frac{p}{2}-1}, w=1$

那么$xwx \wedge R=01^{p-1}0$

使$s=xyz=01^{p-1}0, \because |xy| \leq p \therefore\ x=0, y=1^{p-1}, z=0$

此时$xz=00 \notin L$

因此该语言不是正则语言

问题出在哪里?

再论泵引理

Theorem 1.70 Pumping lemma If A is a regular language, then there is a number p (the pumping length) where if s is any string in A of length at least p, then s may be divided into three pieces, s=xyz, satisfying the following conditions:

1. for each $i \geq 0$, $xy^iz \in A$,
2. $|y|>0$, and
3. $|xy| \leq p$.

Sipser M . Introduction to the Theory of Computation[J]. Acm Sigact News, 2008, 27(1):27-29.

再论泵引理

Theorem 4.1 (The pumping lemma for regular languages)
Let L be a regular languages. Then there exists a constant n (which depends on L) such that for every string w in L such that $|w| \geq n$, we can break w into three strings, $w = xyz$, such that:

1. $y \neq \varepsilon$
2. $|xy| \leq n$
3. For all $k \geq 0$, the string xy^kz is also in L.

Hopcroft J E ， Ullman J D ， Hopcroft J E . Introduction to automata theory, languages, and computation /[M]. Addison-Wesley, 2001.

再论泵引理

Theorem 4.8

Let L be an infinite regular language. Then there exists some positive integer m such that any $w \in L$ $|w| \geq m$ can be decomposed as $w = xyz$

with $|xy| \leq m$,

and $|y| \geq 1$,

such that $xy^i z \in L$

"We have given the pumping lemma only for infinite languages. <u>Finite languages, although always regular</u>, cannot be pumped since pumping automatically creates an infinite set. The theorem does hold for finite languages, but it is vacuous. The m in the pumping lemma is to be taken larger than the longest string, so that no string can be pumped."

Linz P. An introduction to formal languages and automata[M]. Jones & Bartlett Learning, 2012.

再论泵引理

引理4-1 设L为一个RL，则存在仅依赖于L的正整数N，对于$\forall z \in L$，如果$|z| \geqslant N$，则存在u，v，w，满足：

1. $z = uvw$
2. $|uv| \leqslant N$
3. $|v| \geqslant 1$
4. 对于任意的整数$i \geqslant 0$，$uv^iw \in L$
5. N不大于接受L的最小DFA M的状态数。

蒋宗礼，姜守恒. 形式语言与自动机理论(第3版)[M]. 清华大学出版社，2007.

扩充泵引理

引理5.1（扩充泵引理）设L是一个正则语言，则对于L有下列性质：
存在只依赖于L的正整数k，对于任何串x，y，z（这里xyz∈L），只要$|y| \geq k$，就可以将y写成y=uvw（这里$v \neq \varepsilon$，$|uv| \leq k$），使得对于任何$i \geq 0$，都有$xuv^iwz \in L$。

陈有祺. 形式语言与自动机 : Formal languages and automata[M]. 机械工业出版社, 2008.

扩充泵引理

问题：有限语言是正则语言吗？它符合泵引理？

1. 有限语言都可以通过有限次的正则运算得到，所以它一定是正则语言。
2. 有限语言符合泵引理，有两种说法：
   a) 有限语言没有泵，泵长度p=0；
   b) 有限语言存在泵，泵长度p≥|Q|；

扩充泵引理

例21. 证明语言$L_{21} = \{ xx^R w \mid x, w \in \{0,1\}^+ \}$不是正则语言。

分析：句子的前面两个字符不能相同。

解法一（泵引理）：令$s = xyz = (01)^p(10)^p w$，其中p是泵的长度。

$\because |xy| \le p$，$\therefore$ 可以令 $x = \varepsilon$, $y = (01)^k$，$z = (10)^p w$ 。

$xy^2z = = (01)^{p-k}(01)^{2k}(10)^p w = = (01)^p (01)^k (10)^p w \notin L_{21}$

，与泵引理矛盾。

解法二（扩充泵引理）：

解法三（M-N定理）：证明RL的指数是无穷的。

扩充泵引理

例22. 证明语言$L_{22} = \{0^n1^m0^m|n,m\geq1\}$不是正则语言。

证明：反证法，利用扩充泵引理

设 $s = xyz=0^p1^p0^p$， $x = 0^p$, $y = 1^p$, $z = 0^p$, 其中p是泵长度，则：

$y= uvw = 1^p$，可令 $v=1^k,k\geq1$。

从而

$xuv^iw = 0^p\ 1^{p-k}\ 1^{2k}0^p = 0^p1^{p+k}0^p \notin L_{22}$.