形式语言与自动机期末速通

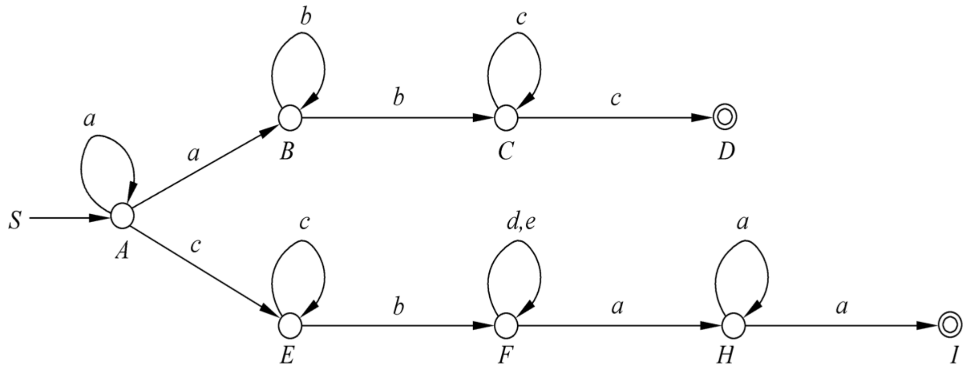
5. RE

5.1 正则表达式

[**例5.1.1**] 考察语言$L=\{a^nb^mc^k \mid n,m,k\geq 1\}\bigcup\{a^ic^nbxa^m \mid i\geq 0,n\geq 1,m\geq 2,x\in\{d,e\}^*\}$.

(1)产生L的RL之一为:$A\rightarrow aA|aB|cE,B\rightarrow bB|bC,C\rightarrow cC|c,E\rightarrow cE|bF,F\rightarrow dF|eF|aH,H\rightarrow aH|a$.

(2)接受L的NFA M之一如下图所示:



各个状态的$set(q)$:

①$set(A)=\{a^n \mid n\geq 0\}=\{a\}^*$.

②$set(B)=set(A)\{a\}\{b^n \mid n\geq 0\}=\{a^nab^m \mid m,n\geq 0\}=\{a\}^*\{a\}\{b\}^*=\{a\}^+\{b\}^*$.

③$set(C)=set(B)\{b\}\{c\}^*=\{a\}^*\{a\}\{b\}^*\{b\}\{c\}^*=\{a\}^+\{b\}^+\{c\}^*$.

④$set(D)=set(C)\{c\}=\{a\}^+\{b\}^+\{c\}^*\{c\}=\{a\}^+\{b\}^+\{c\}^+$.

⑤$set(E)=set(A)\{c\}\{c\}^*=\{a\}^*\{c\}\{c\}^*=\{a\}^*\{c\}^+$.

⑥$set(F)=set(E)\{b\}\{d,e\}^*=\{a\}^*\{c\}^+\{b\}\{d,e\}^*$.

⑦$set(H)=set(F)\{a\}\{a\}^*=\{a\}^*\{c\}^+\{b\}\{d,e\}^*\{a\}\{a\}^*=\{a\}^*\{c\}^+\{b\}\{d,e\}^*\{a\}^+$.

⑧$set(I)=set(H)\{a\}=\{a\}^*\{c\}^+\{b\}\{d,e\}^*\{a\}^+\{a\}$.

⑨$L(M)=set(D)\bigcup set(I)=\{a\}^+\{b\}^+\{c\}^+\bigcup\{a\}^*\{c\}^+\{b\}\{d,e\}^*\{a\}^+\{a\}$.

(3)注意到$\{d,e\}=\{d\}\bigcup\{e\}$,则$\{d,e\}^*=(\{d\}\bigcup\{e\})^*$,

进而$L(M)=\{a\}^+\{b\}^+\{c\}^+\bigcup\{a\}^*\{c\}^+(\{d\}\bigcup\{e\})^*\{a\}^+\{a\}$,

记作$a^+b^+c^++a^*c^+(d+e)^*a^+a=aa^*bb^*cc^*+a^*cc^*(d+e)^*aaa^*$.

[**定义5.1.1**] 字母表Σ上的正则表达式(Regular Expression, RE)定义为:

(1)Φ是Σ上RE,它表示语言Φ.

(2)ε是Σ上的RE,它表示语言$\{\varepsilon\}$.

(3)$\forall a\in\Sigma$是Σ上的RE,它表示语言$\{a\}$.

(4)设r和s分别是Σ上表示语言R和语言S的RE.

①r与s之和$(r+s)$是Σ上表达语言$R\bigcup S$的RE.

②r与s之积rs是Σ上表达语言RS的RE.

③r的**Kleene闭包**r^*是Σ上表达语言R^*的RE.

(5)只有满足(1)(2)(3)(4)的表达式才是Σ上的RE.

意义明确时,RE r表示的语言记作$L(r)$或r.

[**例5.1.2**] 设字母表$\Sigma=\{0,1\}$.

(1)RE 0表示语言$\{0\}$.

(2)RE 1表示语言$\{1\}$.

(3)RE $(0+1)$表示语言$\{0,1\}$.

(4)RE (01)表示语言$\{01\}$.

(5)RE $((0+1)^*)$表示语言$\{0,1\}^*$.

(6)RE $((00)((00)^*)$表示语言$\{00\}\{00\}^*$.

(7)RE $((((0+1)^*)(0+1))((0+1)^*)\ )$表示语言$\{0,1\}^+$.

(8)RE $((((0+1)^*)000)((0+1)^*))$表示$\{0,1\}$上至少含有$3$个连续$0$的串组成的语言.

(9)RE $((((0+1)^*)0)1)$表示$\{0,1\}$上以01结尾的的串组成的语言.

(10)RE $(1((0+1)^*)0))$表示$\{0,1\}$上以1开头、以0结尾的串组成的语言.

[**定义5.1.2**] RE r的**正闭包**r^+定义为RE r与RE (r^*)的乘积及(r^*)与r的乘积,即$r^+=(r(r^*))=((r^*)r)$.加、乘、闭包运算都执行左结合规则.

[**注**] 闭包运算优先级最高,乘运算其次,加运算优先级最低.故意义明确时,可省略某些括号.

如$((((0+1)^*)000)((0+1)^*))=(0+1)^*000(0+1)^*$,

$((((0+1)^*)(0+1))((0+1)^*))=(0+1)^*(0+1)(0+1)^*$.

[**例5.1.3**] 设字母表$\Sigma=\{0,1\}$.

(1)RE $(0+1)^*00(0+1)^*$表示Σ上所有至少包含2个连续0的串组成的语言.

(2)RE $(0+1)^*1(0+1)^9$表示Σ上所有倒数第100个字符为1的串组成的语言.

(3)RE $L((0+1)^*011)=\{x\mid x\in\Sigma^*\wedge x\text{以}001\text{结尾}\}$.

(4)RE $L(0^+1^+2^+)=\{0^n1^m2^k\mid n,m,k\geq 1\}$.

(5)RE $L(0^*1^*2^*)=\{0^n1^m2^k\mid n,m,k\geq 0\}$.

(6)RE $L(1(0+1)^*1+0(0+1)^*0)=\{x\mid x\in\Sigma^*\wedge x\text{的首尾字符相同}\}$.

[**定义5.1.3**] 对字母表Σ上的两个RE r和s,若$L(r)=L(s)$,则称r和s**相等**或**等价**,记作$r=s$.

[**定义5.1.4**] 对字母表Σ上的RE r,其n次**幂**定义为:

(1)$r^0=\varepsilon$.

(2)$r^n=r^{n-1}r$.

[**注**] 注意(2)中r^{n-1}在前.

[**定理5.1.1**] [**RE的常用结论**]

(1)结合律:

①$(rs)t=r(st)$.

②$(r+s)+t=r+(s+t)$.

(2)分配律:

①$r(s+t)=rs+rt$.

②$(s+t)r=sr+tr$.

(3)交换律:$r+s=s+r$.

(4)幂等律:$r+r=r$.

(5)加法零元:$r+\varPhi=r$.

(6)乘法幺元:$r\varepsilon=\varepsilon r=r$.

(7)乘法零元:$r\varPhi=\varPhi r=\varPhi$.

(8)$L(\varPhi)=\varPhi$.

(9)$L(\varepsilon)=\{\varepsilon\}$.

(10)$L(a)=\{a\}$.

(11)$L(rs)=L(r)L(s)$.

(12)$L(r+s)=L(r)\bigcup L(s)$.

(13)$L(r^*)=(L(r))^*$.

(14)$L(\varPhi^*)=\{\varepsilon\}$.

(15)$L((r+\varepsilon)^*)=L(r^*)$.

(16)$L((r^*)^*)=L(r^*)$.

(17)$L((r^*s^*)^*)=L((r+s)^*)$.

(18)若$L(r)\subseteq L(s)$,则$r+s=s$.

(19)$L(r^n)=(L(r))^n$.

(20)$r^nr^m=r^{n+m}$.

[**证**] (17)设$L(r)=\{r_1,\cdots,r_n\},L(s)=\{s_1,\cdots,s_m\}$,

则$r^*=\{\varepsilon,r_1,\cdots,r_n,\cdots\},s^*=\{\varepsilon,s_1,\cdots,s_m,\cdots\}$,

进而$r^*s^*=\{\varepsilon,r_1,\cdots,r_n,s_1,\cdots,s_m,\cdots\},r+s=\{r_1,\cdots,r_n,s_1,\cdots,s_m\}$.

故$L((r^*s^*)^*)=L((r+s)^*)$.

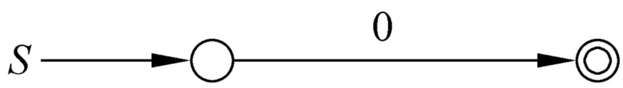
5.2 RE与FA等价

5.2.1 RE转FA

[**定义5.2.1**] 称RE r与FA M等价,如果$L(r)=L(M)$.

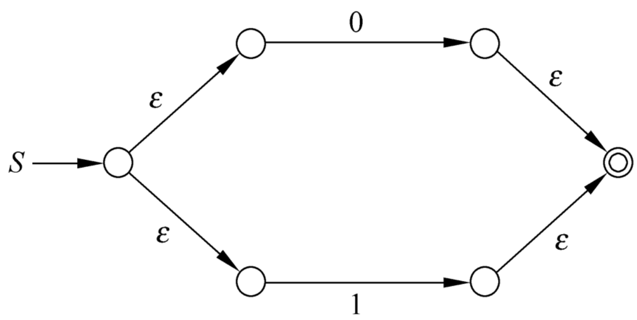
[**例5.2.1**]

(1)RE 0对应的FA:

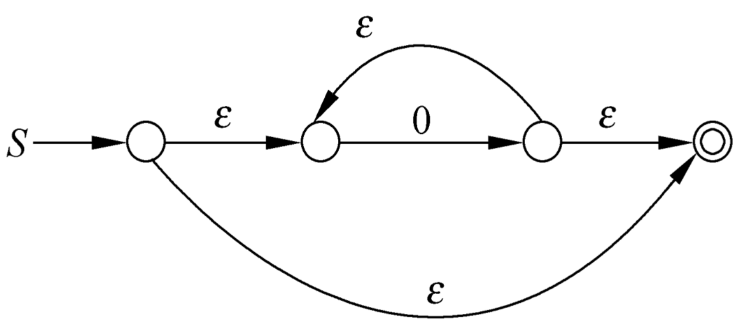


(2)RE 01对应的FA:

(3)RE $(0+1)$对应的FA:



(4)RE 0^*对应的FA:



[**定理5.2.1**] RE表示的语言是RL.

[**证**] 对RE包含的运算符的个数归纳.

下证:对字母表Σ上的$\forall$RE r,$\exists$FA $M\ s.t.\ L(M)=L(r)$,且M恰有一个终止状态,且M在终止状态下不作移动.

(1)$n=0$时,(a)$r=\varepsilon$,(b)$r=\varPhi$,(c)$r=a$对应的FA分别如下图所示:

(a) (b) (c)

(2)设$n=k$时结论成立.

此时对两个RE r_1,r_2,$\exists$两个FA $M_1=(Q_1,\Sigma,\delta_1,q_{01},\{f_1\})$和$M_2=(Q_2,\Sigma,\delta_2,q_{02},\{f_2\})$,

$s.t.\ L(M_1)=L(r_1),L(M_2)=L(r_2),Q_1\bigcap Q_2=\Phi.$

(3)$n=k+1$时,

①取$q_0,f\notin Q_1\bigcup Q_2$,FA $M=(Q_1\bigcup Q_2\bigcup\{q_0,f\},\Sigma,\delta,q_0,\{f\})$,其中$\delta$定义为:

(i)$\delta(q_0,\varepsilon)=\{q_{01},q_{02}\}.$

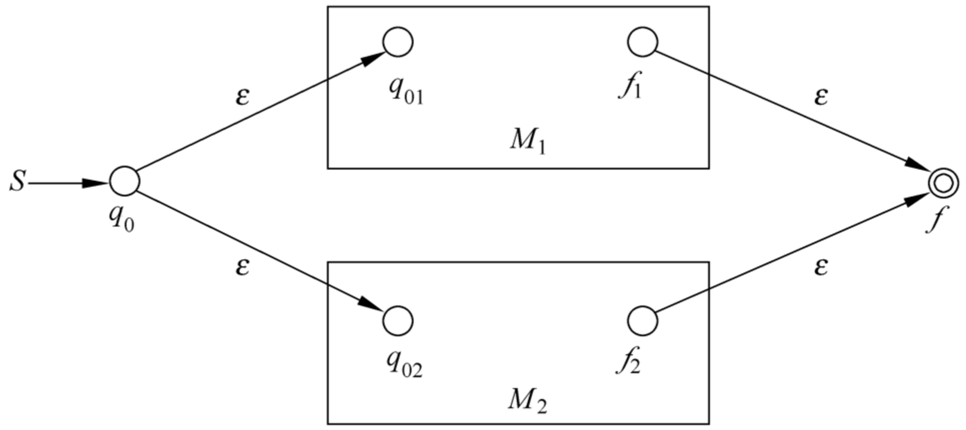
(ii)对$\forall q\in Q_1,\forall a\in\Sigma\bigcup\{\varepsilon\}$,有$\delta(q,a)=\delta_1(q,a).$

对$\forall q\in Q_2,\forall a\in\Sigma\bigcup\{\varepsilon\}$,有$\delta(q,a)=\delta_2(q,a).$

(iii)$\delta(f_1,\varepsilon)=\{f\}.$

(iv)$\delta(f_2,\varepsilon)=\{f\}.$

则RE $r=r_1+r_2$对应的FA如下图所示:



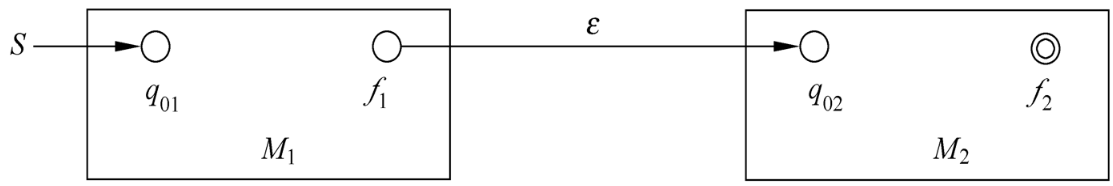
②取FA $M=(Q_1\bigcup Q_2,\Sigma,\delta,q_{01},\{f_2\})$,其中$\delta$定义为:

(i)对$\forall q\in Q_1\backslash\{f_1\},\forall a\in\Sigma\bigcup\{\varepsilon\}$,有$\delta(q,a)=\delta_1(q,a).$

(ii)对$\forall q\in Q_2\backslash\{f_2\},\forall a\in\Sigma\bigcup\{\varepsilon\}$,有$\delta(q,a)=\delta_2(q,a).$

(iii)$\delta(f_1,\varepsilon)=\{q_{02}\}.$

则RE $r=r_1r_2$对应的FA如下图所示:



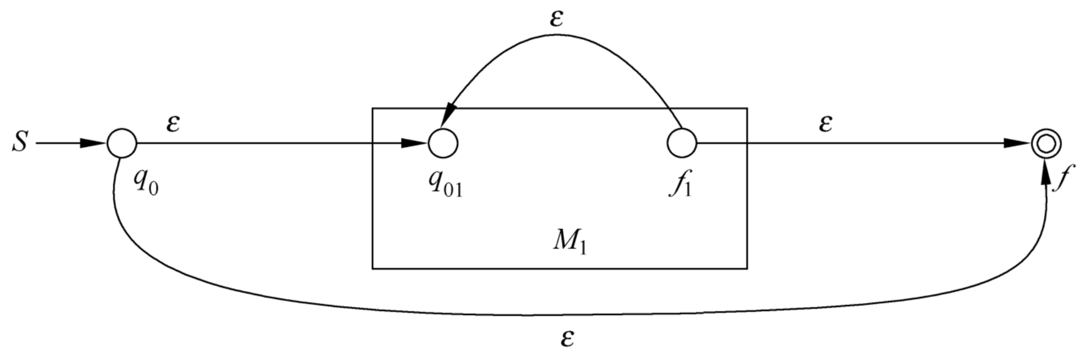
③取$q_0,f\notin Q_1$,FA $M=(Q_1\bigcup\{q_0,f\},\Sigma,\delta,q_0,\{f\})$,其中$\delta$定义为:

(i)对$\forall q\in Q_1\backslash\{f_1\},\forall a\in\Sigma$,有$\delta(q,a)=\delta_1(q,a).$

(ii)$\delta(f_1,\varepsilon)=\{q_{01},f\}.$

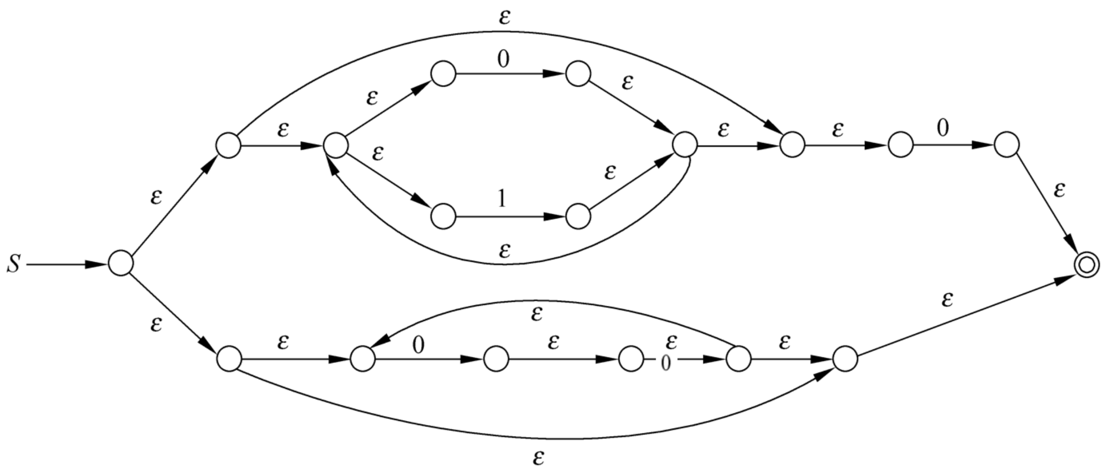
(iii)$\delta(q_0,\varepsilon)=\{q_{01},f\}.$

则RE $r=r_1^*$对应的FA如下图所示:

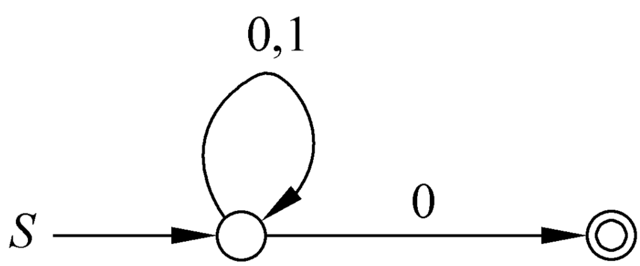


[**例5.2.2**] 构造与RE $(0+1)^*0+(00)^*$等价的FA.

[**解1**]

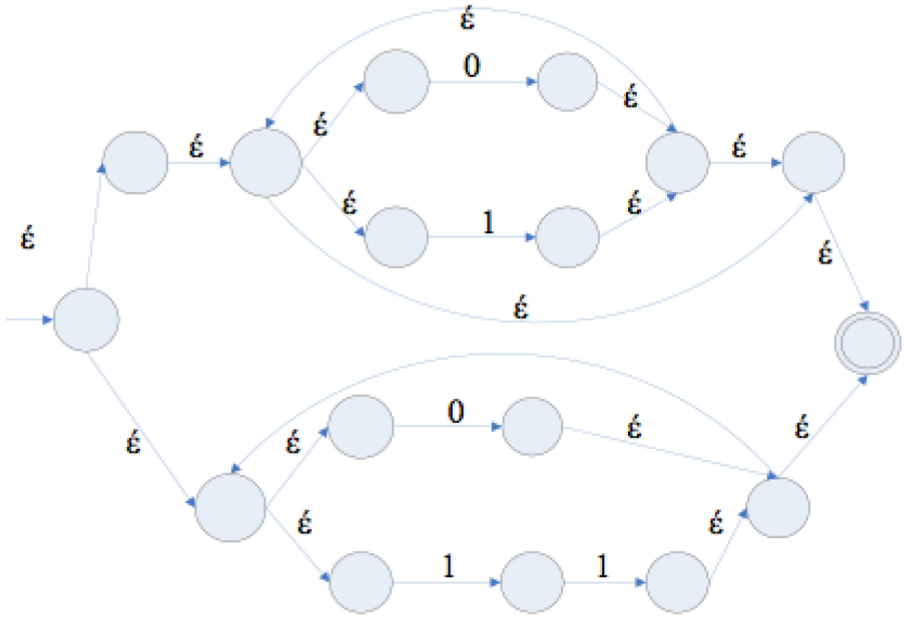


[**解2**]



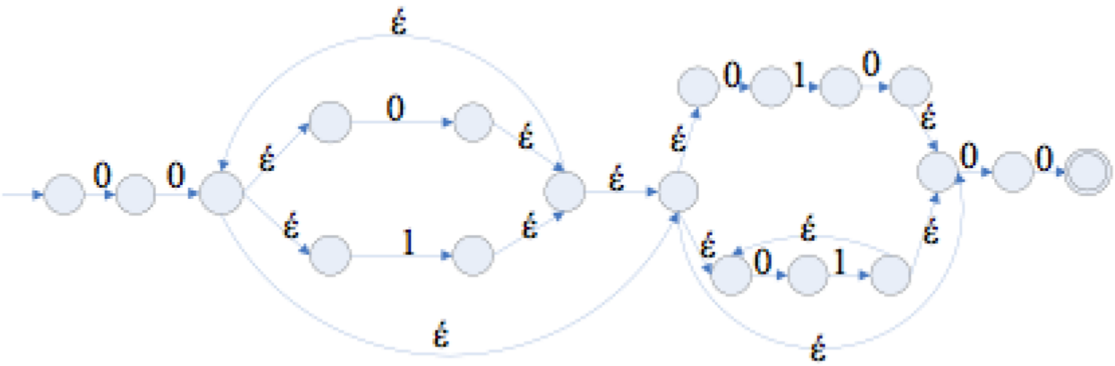
[**例5.2.3**] 构造与RE $(0+1)^*+(0+11)^*$等价的FA.

[**解**]



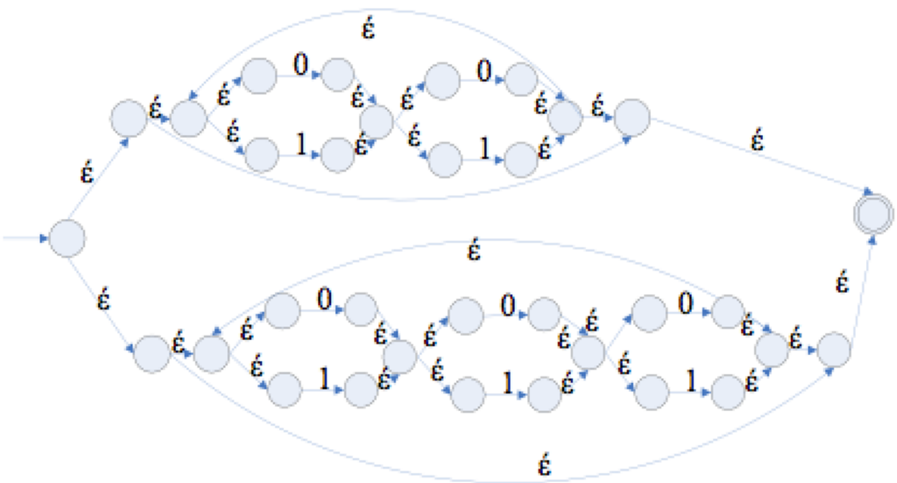
[例5.2.4] 构造与RE $00(0+1)^*((01)^*+010)00$等价的FA.

[解]



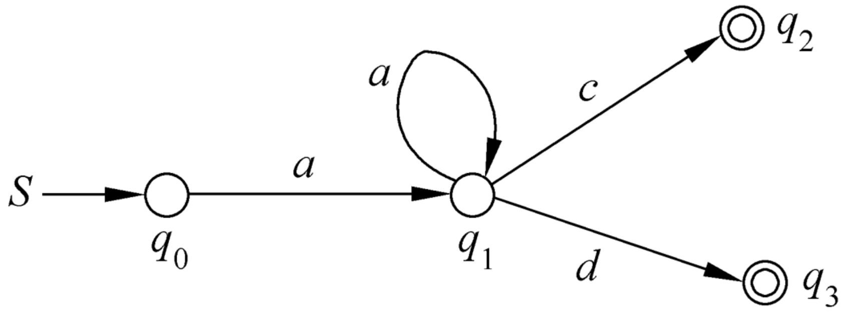
[例5.2.5] 构造与$((0+1)(0+1))^*+((0+1)(0+1)(0+1))^*$.

[解]

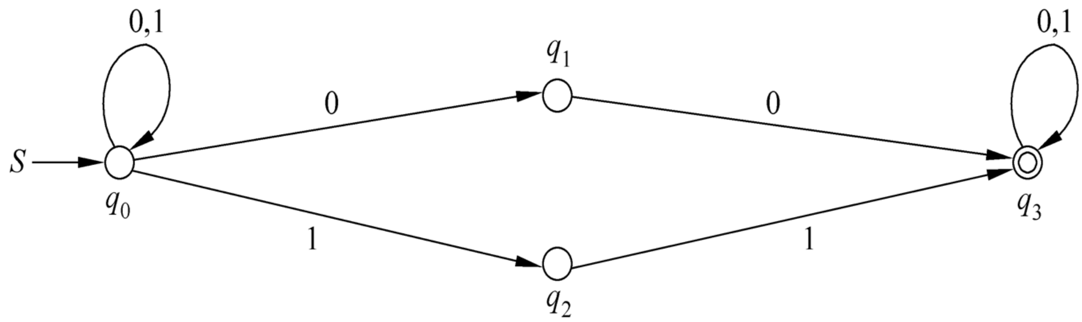


5.2.2 FA转RE

[**例5.2.6**] 下图所示的FA对应的RE为$aa^*(c+d)$.



[**例5.2.7**] 下图所示的FA对应的RE为$(0+1)^*(00+11)(0+1)^*$.



[**定理5.2.2**] RL可用RE表示.

[**证**] 设DFA $M=(\{q_1,\cdots,q_n\},\Sigma,\delta,q_1,F)$.

设集合$R_{ij}^k=\{x\mid x\in\Sigma^*,\delta(q_i,x)=q_j,\text{且对}x\text{的任一非空真前缀}y,\text{若}\delta(q_i,y)=q_l,\text{则}l\leq k\}$,其中:

(1)$R_{ij}^0=\begin{cases}\{a\mid a\in\Sigma^*\wedge\delta(q_i,a)=q_j\},i\neq j\\ \{a\mid a\in\Sigma^*\wedge\delta(q_i,a)=q_j\}\bigcup\{\varepsilon\},i=j\end{cases}$

(2)$R_{ij}^k=R_{ik}^{k-1}(R_{kk}^{k-1})^*R_{kj}^{k-1}\bigcup R_{ij}^{k-1}$.

(3)$L(M)=\bigcup\limits_{q_f\in F}R_{1f}^n$.

[**推论**] RE与FA、RG等价,是RL的表示模型.

[**图上作业法**] 用图上作业法将FA转化为RE的步骤:

(1)预处理:

①用标记为X和Y将M"括起来":在状态转移图中添加标记X和Y的状态,从状态X向状态q_0连一条标记为ε的弧,从状态$q\ \ (q\in F)$向状态Y连一条标记为ε的弧.

②去除不可达状态.

(2)重复如下操作,直至状态转移图中只包含状态X和状态Y,且这两个状态间至多有一条弧:

①并弧:将从状态q到状态p的标记为$r_1,\cdots,r_g$的并行弧替换为从q到p的标记为$(r_1+\cdots+r_g)$的弧.

②去除状态1:若从状态q到状态p有一条标记为r_1的弧,从p到状态t有一条标记为r_2的弧,但不存在从p到p的弧,则去除状态p及与其关联的两条弧,替换为一条从q到t的标记为r_1r_2的弧.

③去除状态2:若从状态q到状态p有一条标记为r_1的弧,从p到状态t有一条标记为r_2的弧,从p到p有一条标记为r_3的弧,则去除状态p及与其关联的三条弧,替换为一条从q到t的标记为$r_1r_3^*r_2$的弧.

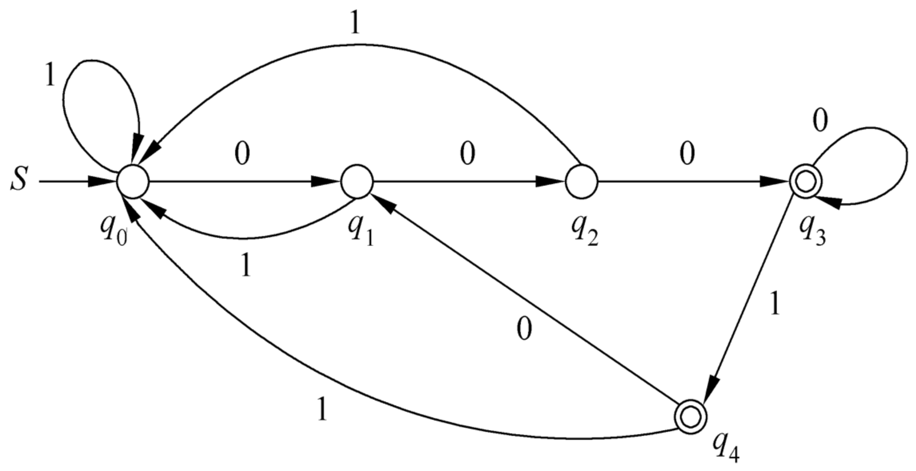
③去除状态3:若状态转移图中只有三个状态,且不存在从状态X到状态Y的路径,则删除X和Y以外的其他状态及与其相关的弧.

(3)从状态X到状态Y的弧的标记即所求的RE.若该弧不存在,则所求的RE为Φ,即原FA中的终止状态都不可达.

[**注1**] 去除状态的顺序不同,得到的RE的形式可能不同,但它们等价.

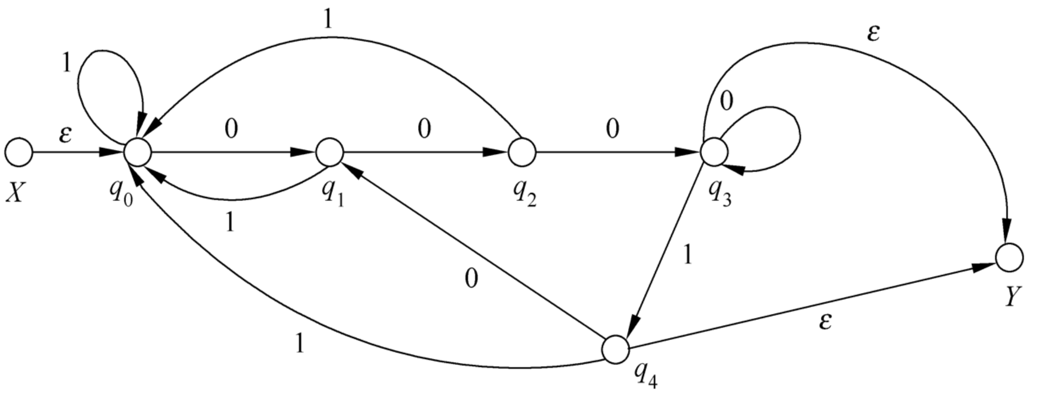
[**注2**] 不考虑节点处的自环,若状态转移图中状态q对应的节点的入度为n,出度为m,则将其去除后需添加nm条弧.

[**例5.2.8**] 求与如下图所示的FA等价的RE.

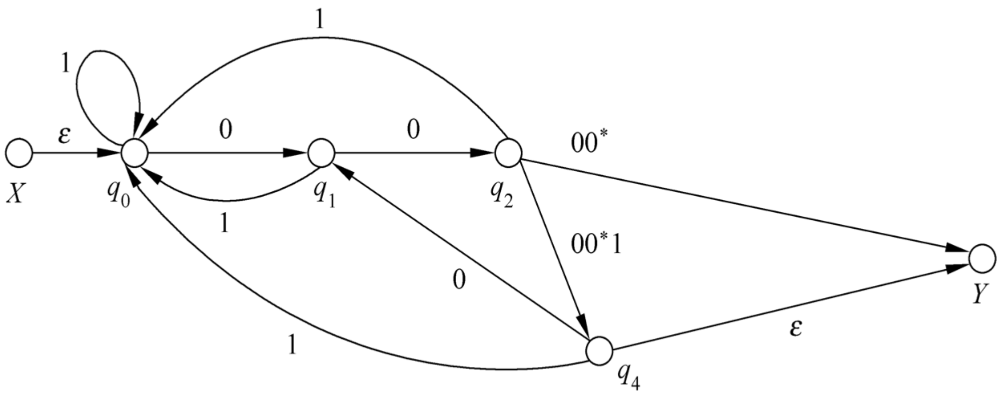


[**解**]

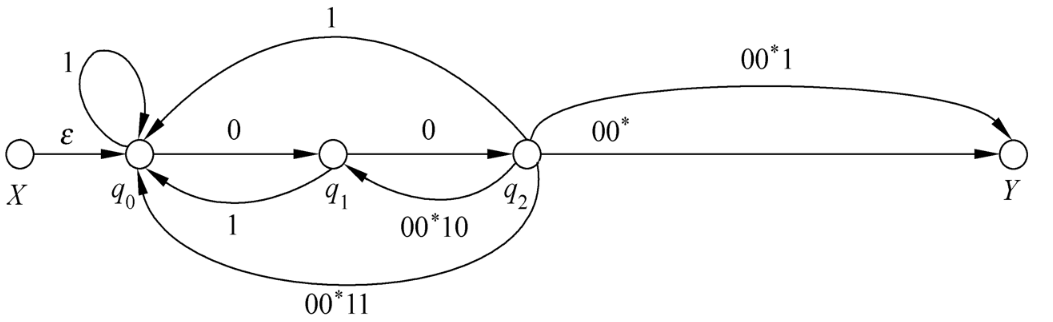
(1)预处理:



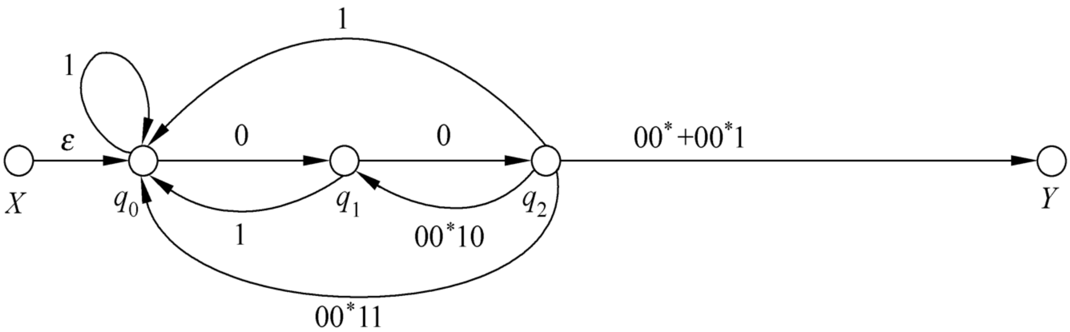
(2)去除状态q_3:



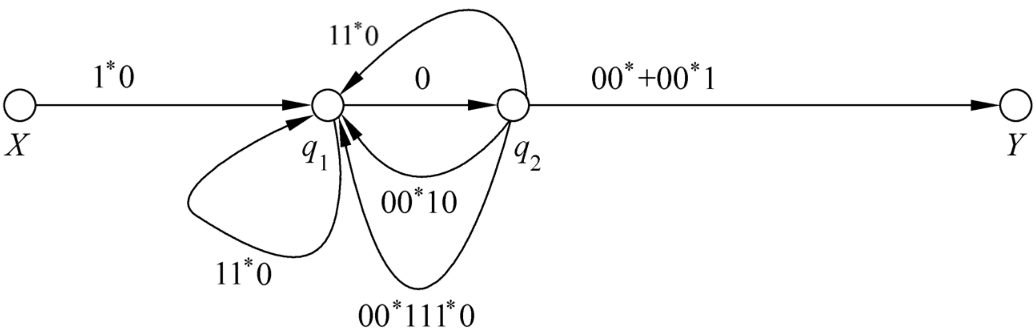
(3)去除状态q_4:



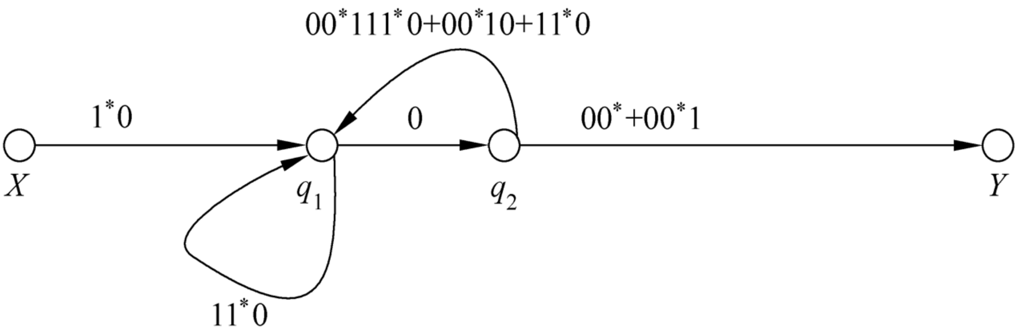
(4)合并从状态q_2到状态Y的两条并行弧:



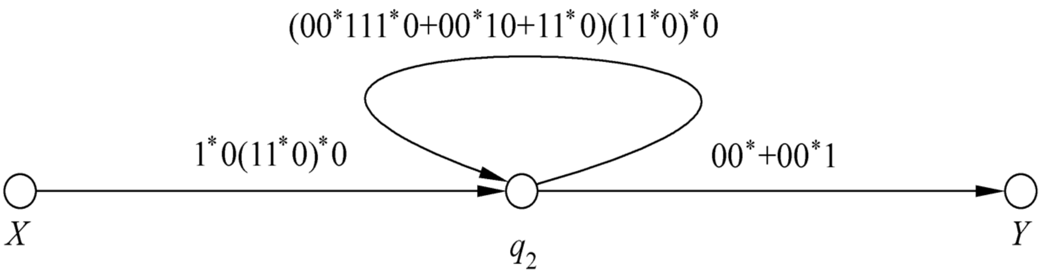
(5)去除状态q_0:



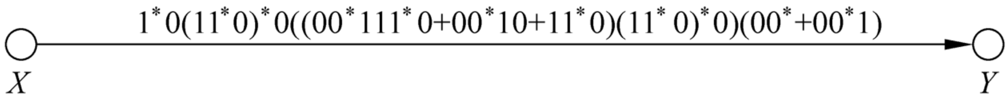
(6)合并从状态q_2到状态q_2的三条并行弧:



(7)去除状态q_1:

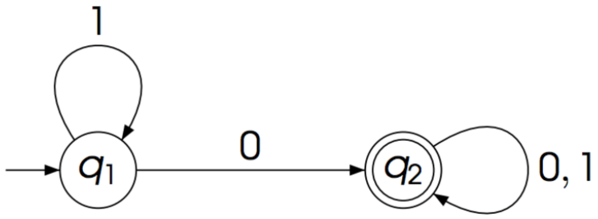


(8)去除状态q_2:



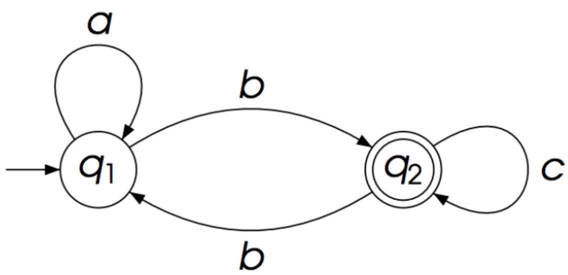
故RE $1^*0(11^*0)^*0((00^*111^*0+00^*10+11^*0)(11^*0)^*0)^*(00^*+00^*1)$.

[**例5.2.9**] 将如下图所示的FA转化为RE(用图上作业法,先消状态q_1,再消状态q_2).



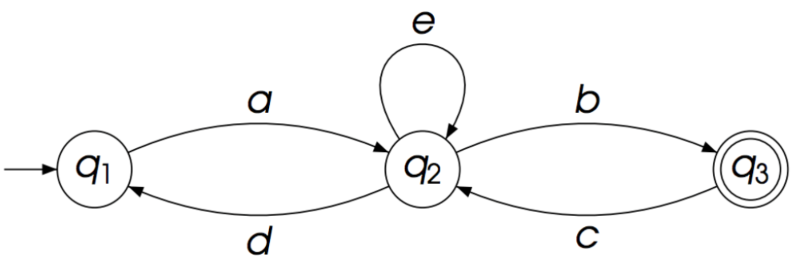
[**解**] $1^*0(0+1)^*$.

[**例5.2.10**] 将如下图所示的FA转化为RE(用图上作业法,先消状态q_2,再消状态q_1).



[**解**] $(a+bc^*b)^*bc^*$.

[**例5.2.11**] 将如下图所示的FA转化为RE(用图上作业法,先消状态q_2,再消状态q_1,最后消状态q_3).



[**解**] $(ae^*d)^*ae^*b(ce^*b+ce^*d(ae^*d)^*ae^*b)^*$.