

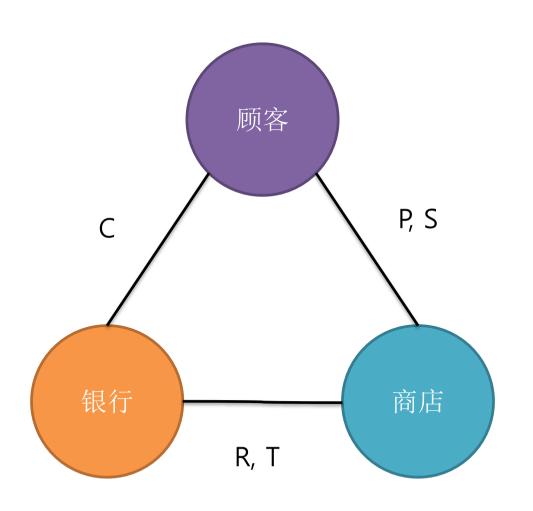


>> 本章内容

- 在对于DFA有感性认识的前提下,本章采用形式化方式介绍DFA的完整内容,并 过渡到非确定型有穷自动机NFA以及有ε-转移的非确定有穷自动机ε-NFA,证明 它们三者在枚举语言的能力上等价。
- 先从DFA开始,包括DFA定义及表示、DFA如何判定输入串为接受还是拒绝、扩展DFA转移函数以表示DFA的连续转移,并用于表示DFA的语言。
- 类似的叙述方式用于NFA和ε-NFA的介绍,并引入ε-闭包和ε-闭集概念,以及将NFA转换为DFA的子集构造法。
- 概括粗粒度知识点:语言识别器;判定性质;等价性质。



→ 一个简单购物系统的DFA建模示例 (F2.1)



P付款: 顾客把电子货币发给商店, 商店收到。

S送货:商店送货给顾客,顾客收到。

C取消: 顾客让银行取消电子货币,银行取消。

R兑换: 商店将电子货币发给银行,银行收到

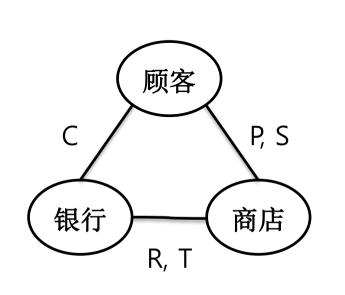
并予以认证。

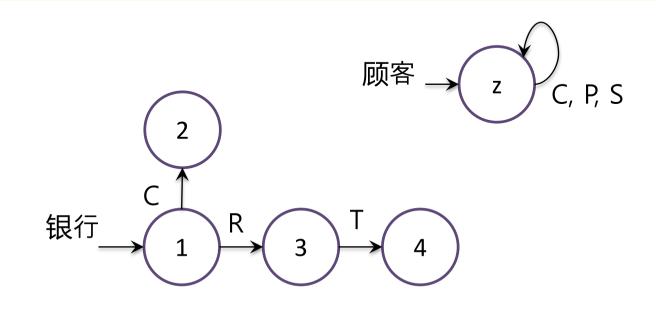
T转帐:银行签发给商店电子货币,商店收到。





>>> 顾客、商店、银行的自动机模型





P付款:顾客把电子货币发给商店,商店收到。

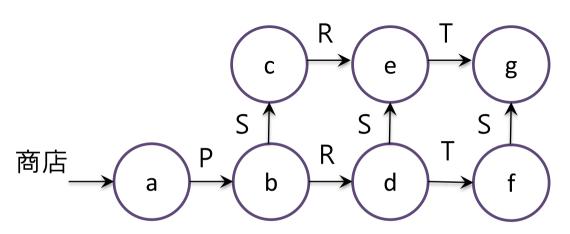
S送货:商店送货给顾客,顾客收到。

C取消:顾客让银行取消电子货币,银行取消。

R兑换:商店将电子货币发给银行,银行收到

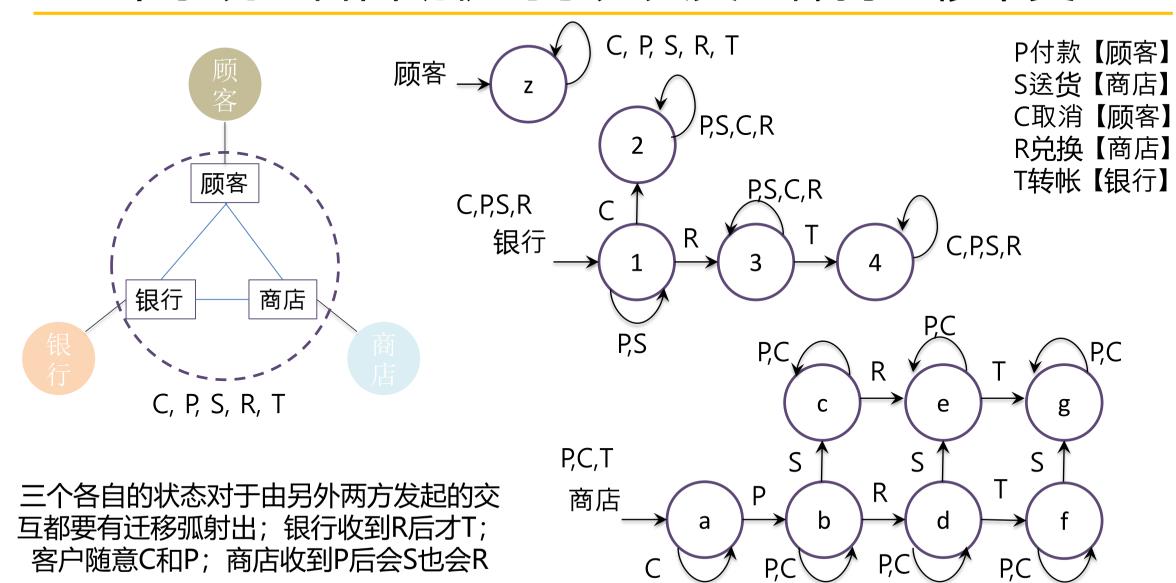
并予以认证。

T转帐:银行签发给商店电子货币,商店收到。



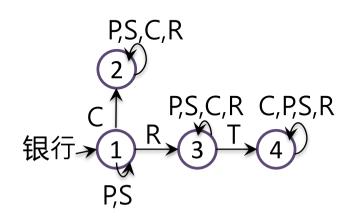


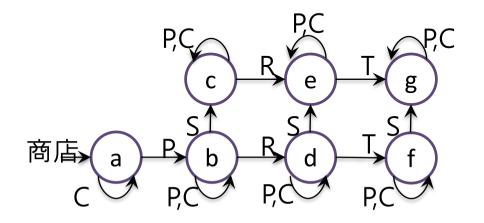
>> 一个系统: 各自动机对于无关交互保持迁移不变

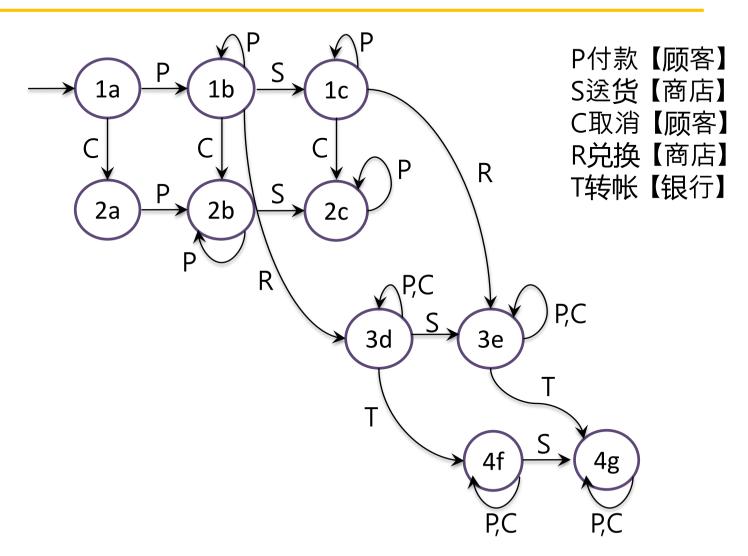




>> 乘积自动机



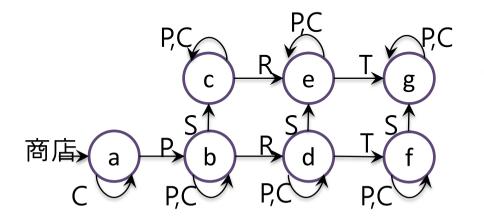






>> 乘积自动机

P付款【顾客】 S送货【商店】 C取消【顾客】 P,S,C,R R兑换【商店】 2 T转帐【银行】 C↑ P,S,C,R C,P,S,R 银行→1 R 3 T 4 P,S



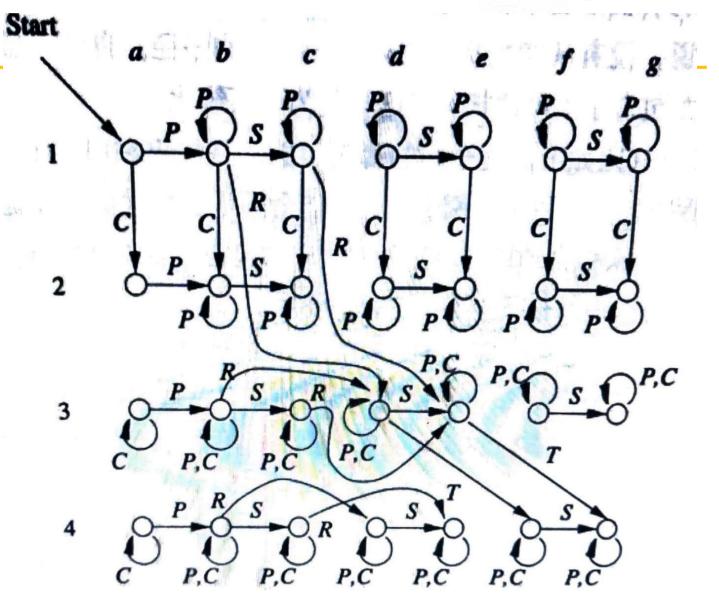
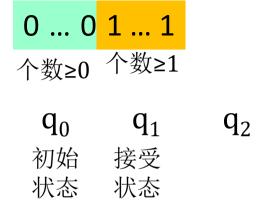


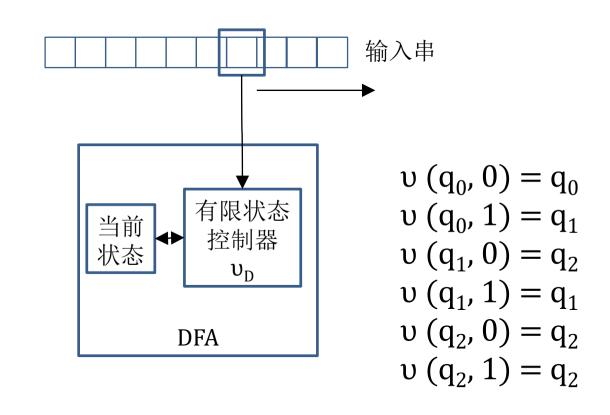
图2-3 商店和银行的乘积自动机



>>> 2.1 确定型有穷自动机DFA

- DFA A = $(Q, \Sigma, \upsilon, q_0, F)$
- 有穷个状态的集合Q
- 字母表Σ,有穷个符号的集合
- 状态转移函数υ: Q×Σ→Q
- q₀ ∈Q 初始状态
- F⊆Q接受状态





DFA (
$$\{q_0, q_1, q_2\}, \{0, 1\}, \upsilon, q_0, \{q_1\}$$
)

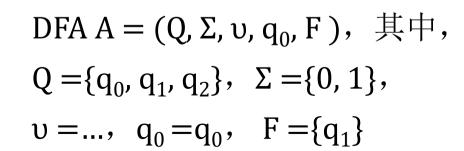


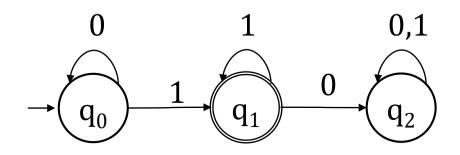
>> DFA表示:代数、转移图、转移表

DFA A = ({q₀, q₁, q₂}, {0, 1}, v, q₀, {q₁})

$$v = ((q_0, 0), q_0), ((q_0, 1), q_1), ((q_1, 0), q_2),$$

 $((q_1, 1), q_1), ((q_2, 0), q_2), ((q_2, 0), q_2))$





$$G(A) = (Q, E, \Sigma)$$

 $E = \{(p, q, a) \mid p, q \in Q, a \in \Sigma, q = \upsilon(p, a)\}$

	0	1
$\rightarrow q_0$	q_0	q_1
*q ₁	q_2	q_1
q_2	q_2	q_2

$$T[ind(q), 0]$$

 $T[0, ind(a)]$
 $T[ind(q), ind(a)]$



>>> 三种DFA表示都是等价的

代数A	转移图 $\mathcal{G}(A)$	转移表℃
Q元素q	顶点(圆圈)	$\mathbb{T}[ind(q), 0]$
Σ元素a	权(弧标记)	$\mathbb{T}[0, ind(a)]$
υ元素	边(转移弧)	T[1 Q , 1 Σ]元素
q_0	箭头标记顶点	箭头标记T[1,0]
F元素	双圈顶点	*标记T[1 Q ,0]元素

代数系统:集合Q、 Σ 、运算 υ 、封闭性

2024/3/5

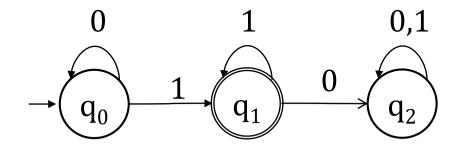


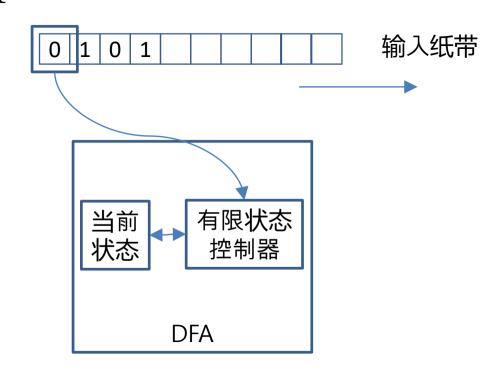


>>> DFA的判定性质: DFA接受、拒绝输入串

0, 1, 00, 01, 000, 001, 010, 0101

	0	1
$\rightarrow q_0$	q_0	q_1
q_1	q_2	q_1
q_2	q_2	q_2





(当前状态,当前输入符号)->当前状态



>> 用算法描述DFA的判定性质(运行过程)

输入: DFA A=(Q, Σ, υ, q_0 , F)

输入: 给定输入串w \in Σ*

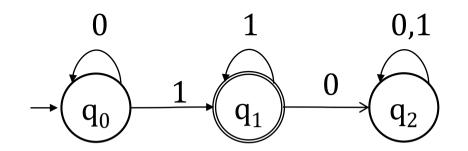
输出: dfa(w)为真接受否则拒绝

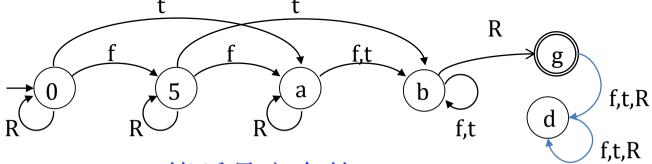
```
int dfa(w) {
                //当前状态
q=q_0;
                //剩余串
x=w:
while (x != \varepsilon) {
                //当前输入符号a
     x=ay;
     q=υ(q, a); //状态转移
                //修改剩余串
     x=y;
                //为真表示接受
return q∈F;
```

习惯约定: w, x, y, ... 是串, a, b, c, ... 是符号。

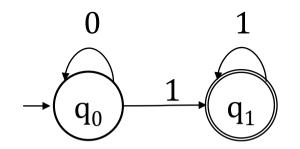


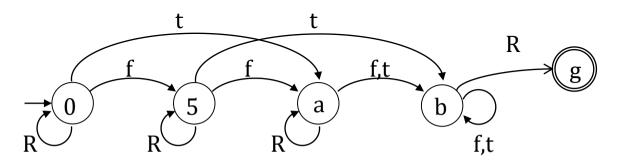
➤ 简化的DFA (DFA的一般形式)





箭弧是完全的 即, v是映射





箭弧允许有缺失 即,v是关系



>>> DFA的通用判定算法

• 算法: DFA的判定性质

输入: DFA A=(Q, Σ, υ, q_0 , F)

输入: 给定输入串w ∈ Σ*

输出: dfa(w)为真接受否则拒绝

```
DFA A = (Q, \Sigma, \upsilon, q_0, F),
```

完全的: $\forall q \in Q \cdot \forall a \in \Sigma \cdot \upsilon(q, a) \in Q$,

简化的: $\forall q \in Q \cdot \forall a \in \Sigma \cdot \upsilon(q, a) \in Q \cup \{\bot\}$,

为了描述方便, υ的值为_上表示无定义。

```
int dfa(w) {
q=q_0;
x=w:
while (x != \varepsilon) {
       x=ay;
       //自动机突然死亡以示拒绝
       if((q=v(q, a))\notin Q) return 0;
       x=y;
return q∈F;
```



>>> 形式描述DFA接受串w

- DFA A=(Q, Σ, υ, q₀, F) 接受 w 如果对于 r_0 , r_1 , ..., $r_n \in Q$ 满足:
 - $r_0 = q_0$
 - $\upsilon(r_i, w_{i+1}) = r_{i+1}$ 其中 i = 0, ..., n-1
 - $r_n \in F$
- 为方便描述,使用路径术语:
 - 路径起点(始端) r_0 ,终点(末端) r_n ,路径标记w
 - 路径隐含 $\upsilon(r_i, w_{i+1}) = r_{i+1}$ 其中 i = 0, ..., n-1
- 何时不接受(拒绝)?
 - 不存在末端为接受状态的路径(路径长度为|w|)
 - 或自动机突然死亡(路径长度<|w|且以w的一个前缀为标记)



>> 扩展转移函数

- · 为了方便表示连续的转移,扩展转移函数v为 ũ
- 归纳定义ῦ:

基础: $\tilde{v}(q, \varepsilon) = q$

归纳: $\tilde{\upsilon}(q, wa) = \upsilon(\tilde{\upsilon}(q, w), a)$

- ũ(q,w)表示DFA中存在唯一一条始端为q末端为ũ(q,w)的标记为w的路径。
- 对于简化DFA, ῦ(q, w)的值又可能为

对一个 DFA,

给定状态q,

给定输入串 $w=a_1a_2...a_n$,

ũ 是这样计算的:

从q开始依次选择标记为 a_1 , a_2 , ..., a_n 的弧走过一条路径,路径末端就是 $\tilde{\upsilon}(q,w)$ 。



>> 扩展转移函数举例

 $\tilde{v}(q_0, 011)$ = $v(\tilde{v}(q_0, 01), 1)$ = $v(v(\tilde{v}(q_0, 0), 1), 1)$ = $v(v(v(\tilde{v}(q_0, \epsilon), 0), 1), 1)$ = $v(v(v(q_0, \epsilon), 0), 1), 1)$ = $v(v(q_0, 1), 1)$ = $v(q_1, 1)$ = $v(q_1, 1)$

```
\begin{array}{cccc} & 0 & 1 \\ \rightarrow^* q_0 & q_0 & q_1 \\ ^* q_1 & q_2 & q_1 \\ q_2 & q_2 & q_2 \end{array}
```

$$\tilde{v}(q_0, 101)$$
= $v(\tilde{v}(q_0, 10), 1)$
= $v(v(\tilde{v}(q_0, 1), 0), 1)$
= $v(v(q_1, 0), 1)$
= $v(q_2, 1)$
= q_2

$$\begin{array}{cccc} & 0 & 1 \\ \rightarrow^* q_0 & q_0 & q_1 \\ ^* q_1 & \bot & q_1 \end{array}$$

$$\tilde{v}(q_0, 101) =$$
= $v(\tilde{v}(q_0, 10), 1)$
= $v(v(\tilde{v}(q_0, 1), 0), 1)$
= $v(v(q_1, 0), 1)$
= $v(\bot, 1)$
= \bot

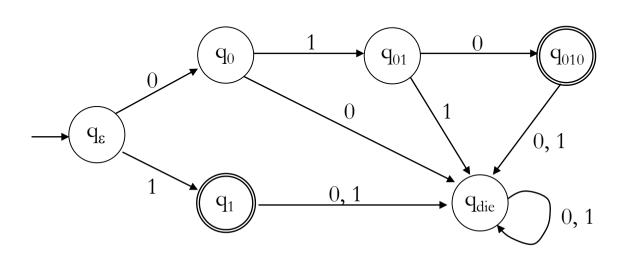
17



>>> DFA状态的可达性

- 对于 $q \in \mathbb{Q}$,不存在始端为 q_0 末端为q的路径,则q是从初始状态不可达的。如果 DFA有这样的状态q,则称其具有可达性情形①。
- 对于 $q \in Q$, $q_1 \in F$,不存在始端为q末端为 q_1 的路径,则q是不能到达接受状态的。如果DFA有这样的状态q,则称其具有可达性情形(2)。
- 对于 $q \in Q$, $q_1 \in F$,同时存在始端为 q_0 末端为q的路径、始端为q末端为 q_1 的路径。如果DFA有这样的状态q,则称其具有可达性情形(3)。

- $\bigcirc \forall w \in \Sigma^* \cdot q \neq \tilde{\upsilon}(q0, w)$
- ② $\forall p \in F, w \in \Sigma^* \cdot p \neq \tilde{\upsilon}(q, w)$
- $\exists p \in F, \exists x,y \in \Sigma^* \cdot q = \tilde{\upsilon}(q0, x) \land p = \tilde{\upsilon}(q, y)$





>>> DFA定义语言

- DFA A = $(Q, \Sigma, \upsilon, q_0, F)$
- $L(A) = \{ w \in \Sigma^* \mid \tilde{v}(q_0, w) \in F \}$
- DFA定义的语言是正则语言
- 语言是正则的如果它是某个DFA的语言

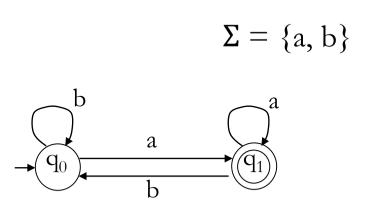
2024/3/5

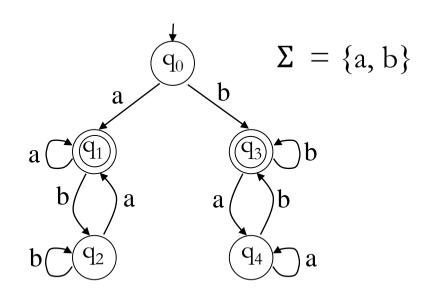


>>> DFA的语言示例



- DFA M = $(Q, \Sigma, \upsilon, q_0, F)$
- $L(M) = \{w \in \Sigma^* \mid \tilde{v}(q_0, w) \in F\}$







>>> {0,1}上的以11为前缀的串

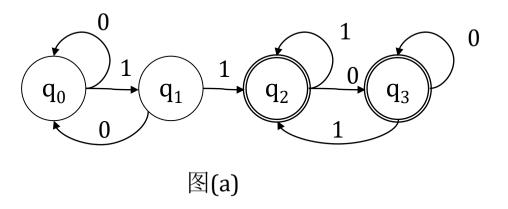
	0	1
→q0	丄	q1
q1 *q2	\perp	q2 q2
*q2	q2	q2



>>> {0,1}上的含有子串11的串

	0	1
$\rightarrow q_0$	q0	q_1
q_1	q_0	q_2
q ₁ *q ₂ *q ₃	q_3	q_2
$ *q_3 $	q_3	q_2

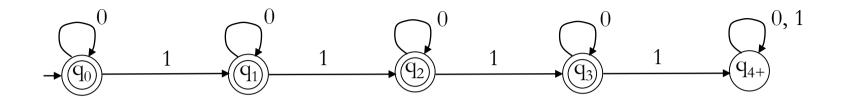
	0	1
$\rightarrow q_0$	q_0	q_1
$\begin{vmatrix} q_1 \\ *q_2 \end{vmatrix}$	q_0	q_2
*q_2	q_2	q_2







• 构造字母表{0,1}上的DFA 接受所有最多含有三个1的串

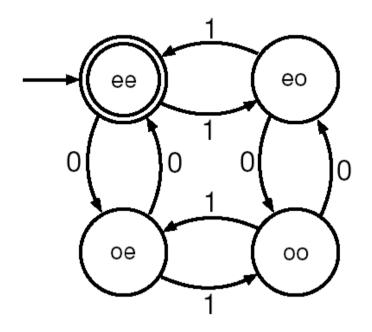




>> 正则语言的例子

• {0,1}上的,含有偶数个0和偶数个1的串的全集,是正则语言。

思路:输入串已消耗掉的部分只有四种情况,分别用状态来记住,继续读入符号将继续发生状态转移,当输入串消耗完以后,所达到的状态就决定了原输入串是什么情况,比如偶数个0和偶数个1



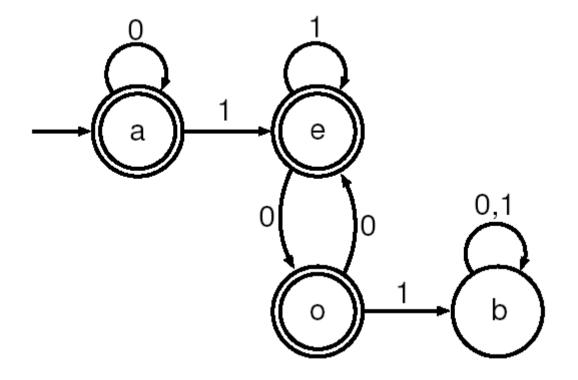
2024/3/5



>> 正则语言的例子

• {0,1}上不包含一对1且中间被奇数个0隔开的任何串。

思路:排除包含模式10...01的串,该模式中的0为奇数个

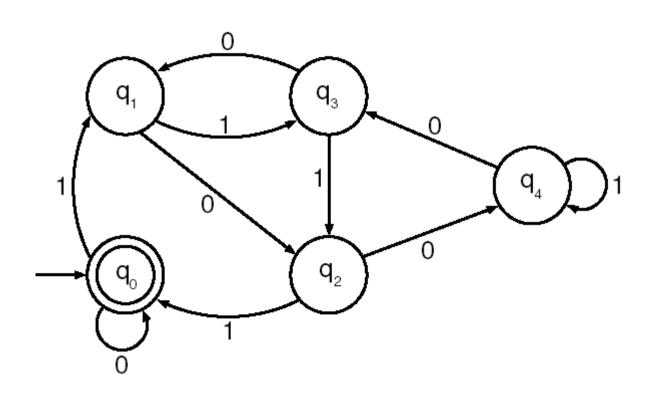


2024/3/5



>> 正则语言的例子

• $\diamondsuit L = \{ w \in \{0,1\}^* | w 看作二进制数的话能被5整除 \}$



分析:一个数除以5,其余数只能是 0, …, 4五种, 因此我们以q₀, …, q₄分别表示这五种状态。因为接受能 被5整除的数,故状态0既为初始状态, 又为终结状态。接着,考虑二进制数 在其串后增添0或1时,状态的转化情 况。在二进制串后添1位,即可理解 为将先前的串值乘以二再加上所添的 数值。那么, 串尾添数后新的数值模 5的余数便可以计算出来。即可以得 到添0或1后的新的状态。



>>> 归纳{0,1}上语言的DFA设计题目

- 前缀有这个、没这个、同时有这两个、或者有这个或者那个;
- 子串同上;
- 后缀同上;
- 汇总性:有n个1、没有n个1、有n个1和m个0、没有n个1同时也没有m个0;
- 计算性:被5整除、是素数、是奇数、是偶数
- 整数、定点数
- 对题目要求: 典型且简单



>> 用状态作记忆

• {0,1}上以01结尾的串

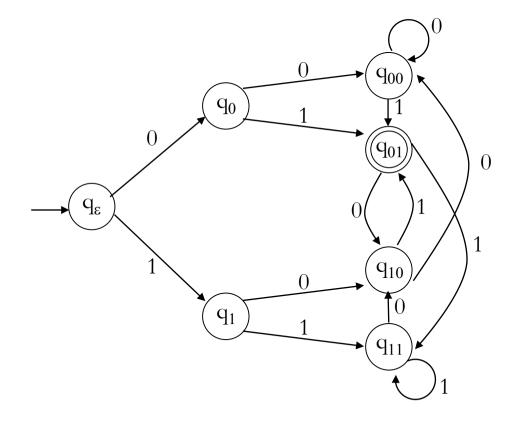
用四个状态记住输入串的前缀的最后两位(已消耗串的最后两位)。

 q_{00}

 q_{01}

 q_{10}

 q_{11}





>>> 用状态作记忆局限性

• {0,1}上以101结尾的串

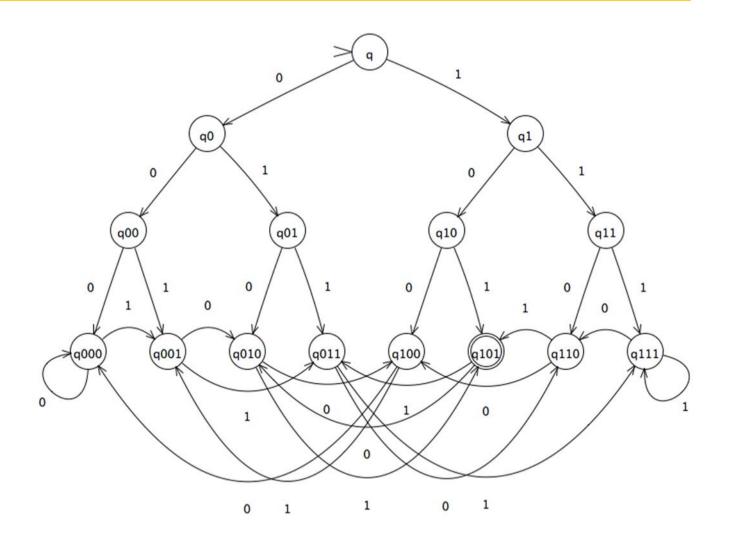
用八个状态记住已消耗串的最后三位。

 q_{000}

 q_{001}

...

 q_{111}

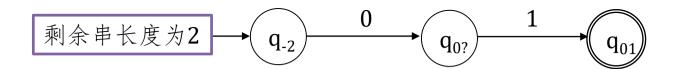


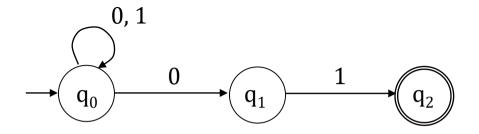




>>> 引入不确定性来避免状态爆炸

猜测

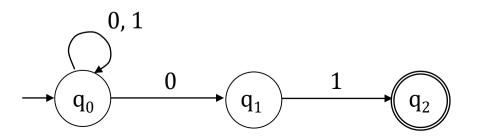


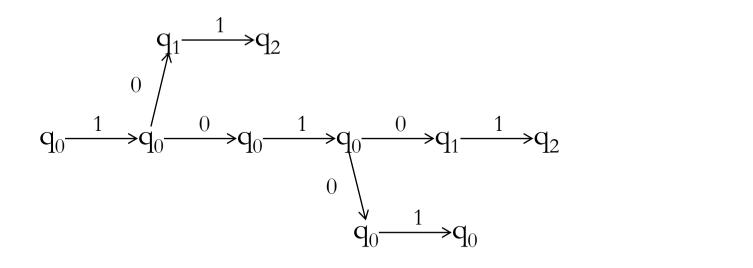




>> 穷举与猜测

• 输入串10101





死亡

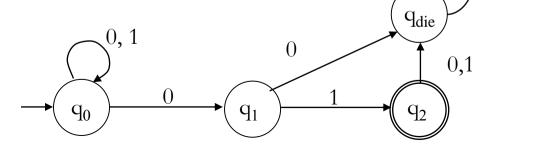
接受

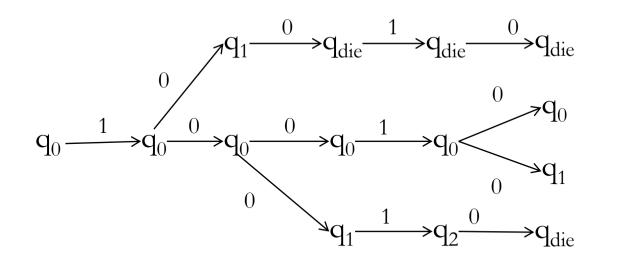
拒绝



>> 穷举与猜测

• 输入串10010





陷入

0,1

拒绝

拒绝



>> 线索穷举

- 对于输入串,线索是自动机的一条以初始状态为起点的状态转移路径,该路径正 好消耗掉输入串的某个前缀。线索穷举是有穷自动机的运行方式
 - DFA运行时只有一条线索(路径)。
 - 非确定有穷自动机(NFA)运行时有有穷个线索。
- 对于给定的输入串w,只要有一个线索成功,则w被自动机接受,否则不被接受
 - 成功的线索: 始端为初始状态,标记为w,末端为接受状态。
- 在前例中,
 - 以101结尾的简化型NFA对于输入串110101有5个线索,其中有一个成功。
 - 以101结尾的完全型NFA对于输入串110010有4个线索,均不成功。
 - 以101结尾的完全型NFA对于输入串110101有几个个线索?成功者为何?

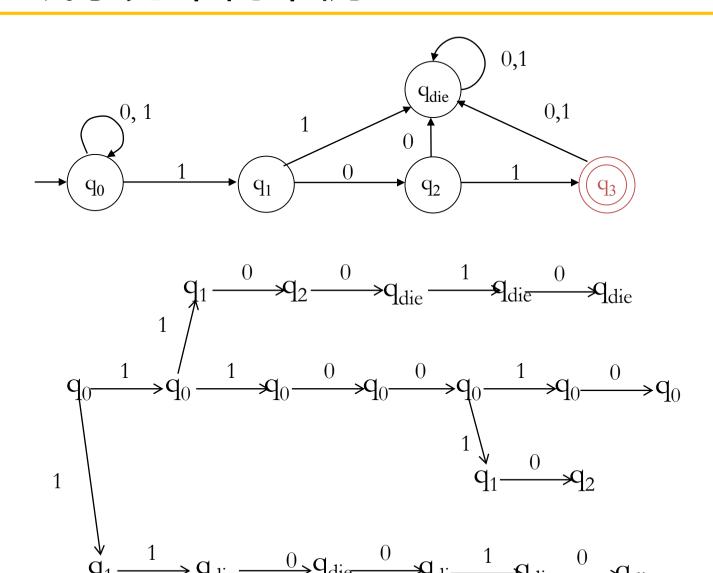


>> NFA的线索穷举成为一个树

- w为输入串,w的第i(i≥0)个前缀就是长度为i的前缀,记为i-前缀
- w=hello-world
- w前缀: ε, h, he, hel, hell, hello, hello-, hello-w, ..., hello-world
- w有|w|+1个前缀
- 给定输入串的线索穷举树▶
- 根为初始状态,层数为0
- 从根到每一个叶子的路径都是一条线索(每个线索都在树中恰有一条从根到叶子的路径对应)
- 第i层(i>0)的状态集合=Ti
- 以根为起点,无论沿着那条线索,都消耗掉输入串的i-前缀并转移到Ti中的状态
- Ti就是活动状态集,也是w的i-前缀的当前状态集。



>> 线索穷举树举例

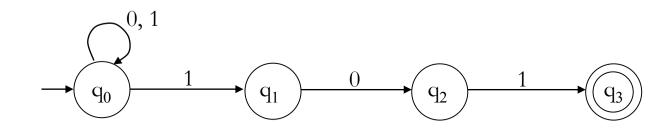


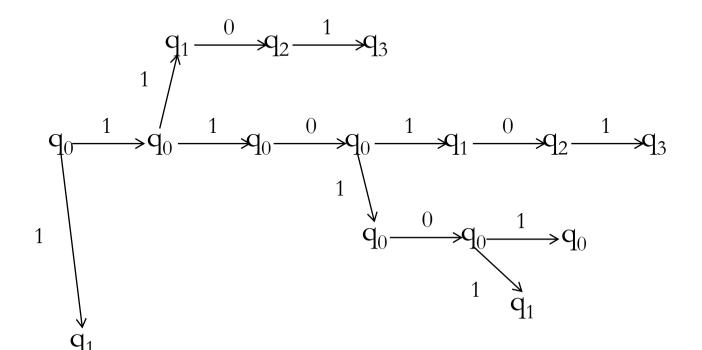
输入串: 110010

i	Т	i-前缀
0	$\{q_0\}$	ε
1	$\{q_0,q_1\}$	1
2	$\{q_0,q_1,q_{die}\}$	11
3	$\{q_0,q_2,q_{die}\}$	110
4	$\{q_0, q_{die}\}$	1100
5	$\{q_0,q_1,q_{die}\}$	11001
6	$\{q_0,q_2,q_{die}\}$	110010



>> 线索穷举树举例





输入串: 110101

i	T	i-前缀
0	$\{q_0\}$	3
1	$\{q_0,q_1\}$	1
2	$\{q_0,q_1\}$	11
3	$\{q_0,q_2\}$	110
4	${q_0, q_1, q3}$	1101
5	{q ₀ ,q2}	11010
6	${q_0, q_1, q3}$	110101



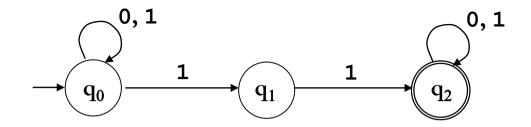


• 构造字母表{0,1}上的NFA接受含有子串11的串。

10011010, 11001, 11101 接受

ε, 000, 010101

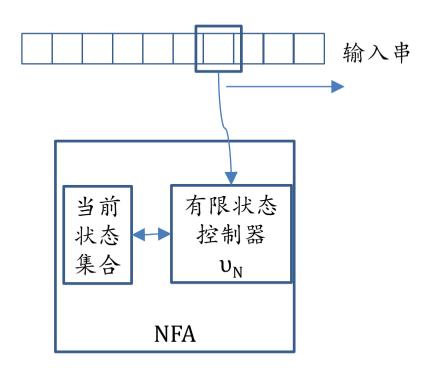
拒绝





>>> NFA与DFA有什么不同?

- DFA: 从同一个状态射出的带有相同标记的箭弧最多有一条。
- NFA: 从同一个状态射出的带有相同标记的箭弧可以有多条(NFA)。
- ε-NFA: 有ε-转移的NFA。
- ε-转移: 状态间的ε-转移不消耗输入符号。
- 本章2.3小节介绍NFA
- 本章2.4小节介绍ε-NFA
- 之后,二者不再区分,都是NFA



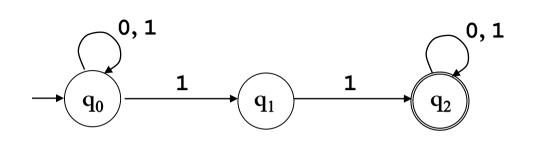




- NFA A=(Q, Σ , υ , q₀, F)
 - 状态集 Q={q₀, q₁, q₂}
 - 字母表 Σ={0,1}
 - 转移函数υ: Q×Σ→2^Q
 - 初始状态q₀∈Q
 - 接受状态集合 F={q₂}⊆Q

- $\upsilon = \{((q_0,0), \{q_0\}), ((q_0,1), \{q_0,q_1\}), ((q_1,0), \{\}), ((q_1,1), \{q_2\}), ((q_2,1), \{q_2\})\}$
- NFA A=($\{q_0,q_1,q_2\},\{0,1\}, \nu, q_0,\{q_2\}$)

两种定义:转移函数 υ 是映射; υ 允许 \bot 值用 υ (q, a)= \bot 表示 υ (q, a)无定义。



	0	1
$\rightarrow q_0$	$\{q_0\}$	$\{q_0, q_1\}$
q_1	φ	$\{q_2\}$
$*q_2$	$\{q_2\}$	$\{q_2\}$



>>> NFA的判定性质

- 输入: NFA A=(Q, Σ, υ, q₀, F), w∈Σ*
- 输出:接受、拒绝w

2024/3/5



>> NFA的扩展转移函数

为了方便表示读入一个串所发生的 连续转移,扩展转移函数v为v,v被称为扩展转移函数。

- 归纳方式定义为:
- 基础: ῦ(q, ε) = {q}
- 归纳: $\tilde{\upsilon}(q, wa) = Up \in \tilde{\upsilon}(q, w) \cdot \upsilon(p, a)$

q是NFA的状态, 对于串 $w=a_1a_2...a_n$, $T_0=\{q\}$, $T_i=Up\in T_{i-1}\cdot \upsilon(p,a_i), 1\leq i\leq n$, 那么 $\tilde{\upsilon}(q,w)=T_n$

• $\tilde{v}(q,w)$: 对于始端为q,标记为w的所有路径,由它们的末端组成的集合。





>> 例2-12的示例

```
\tilde{v}(q_0, 01011) = \mathsf{Up} \in \tilde{v}(q_0, 0101) \cdot v(p, 1)
= Up\in[ Up\in\tilde{\upsilon}(q<sub>0</sub>, 010) \cdot \upsilon(p, 1) ] \cdot \upsilon(p, 1)
= \mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in \tilde{\upsilon}(q_0, 01) \cdot \upsilon(p, 0)] \cdot \upsilon(p, 1)] \cdot \upsilon(p, 1)
= \mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in \tilde{\upsilon}(\mathsf{q}_0, 0) \cdot \upsilon(\mathsf{p}, 1)] \cdot \upsilon(\mathsf{p}, 0)] \cdot \upsilon
(p, 1) ] \cdot v(p, 1)
= Upe[ Upe[ Upe[ Upe[ Upe\tilde{v}(q_0, \epsilon) \cdot v(p, 0) ] \cdot v(p, 0)
1) ] \cdot \upsilon(p, 0) ] \cdot \upsilon(p, 1) ] \cdot \upsilon(p, 1)
= \mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in \{\mathsf{q}_0\} \cdot \upsilon(\mathsf{p}, 0)] \cdot \upsilon(\mathsf{p}, 1)] \cdot
\upsilon(p, 0) \cdot \upsilon(p, 1) \cdot \upsilon(p, 1)
= \mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in \{\mathsf{q}_0, \mathsf{q}_1\} \cdot \upsilon(\mathsf{p}, 1)] \cdot \upsilon(\mathsf{p}, 0)] \cdot \upsilon
(p, 1) ] \cdot v(p, 1)
= \mathsf{Up} \in [\mathsf{Up} \in [\mathsf{Up} \in \{\mathsf{q}_0, \mathsf{q}_2\} \cdot \upsilon(\mathsf{p}, 0)] \cdot \upsilon(\mathsf{p}, 1)] \cdot \upsilon(\mathsf{p}, 1)
= \mathsf{Up} \in [\; \mathsf{Up} \in \{\mathsf{q}_0, \mathsf{q}_1\} \cdot \upsilon(\mathsf{p}, 1) \;] \cdot \upsilon(\mathsf{p}, 1)
= \mathsf{Up} \in \{\mathsf{q}_0, \mathsf{q}_2\} \cdot \mathsf{v}(\mathsf{p}, 1)
= \{q_0\}
```

$$\begin{split} \tilde{\upsilon}(q_0,\epsilon) &= \{q_0\} \\ \tilde{\upsilon}(q_0,0) &= \mathsf{U} p \in \tilde{\upsilon}(q_0,\epsilon) \cdot \upsilon(p,0) = \{q_0,q_1\} \\ \tilde{\upsilon}(q_0,01) &= \mathsf{U} p \in \tilde{\upsilon}(q_0,0) \cdot \upsilon(p,1) = \upsilon(q_0,1) \; \cup \\ \upsilon(q_1,1) &= \{q_0,q_2\} \\ \tilde{\upsilon}(q_0,010) &= \mathsf{U} p \in \tilde{\upsilon}(q_0,01) \cdot \upsilon(p,0) = \upsilon(q_0,0) \\ \mathsf{U}\,\upsilon(q_2,0) &= \{q_0,q_1\} \\ \tilde{\upsilon}(q_0,0101) &= \mathsf{U} p \in \tilde{\upsilon}(q_0,010) \cdot \upsilon(p,1) = \upsilon(q_0,1) \\ \mathsf{U}\,\upsilon(q_1,1) &= \{q_0,q_2\} \\ \tilde{\upsilon}(q_0,01011) &= \mathsf{U} p \in \tilde{\upsilon}(q_0,0101) \cdot \upsilon(p,1) = \upsilon(q_0,1) \; \cup \upsilon(q_2,1) = \{q_0\} \end{split}$$



>> NFA的语言

- NFA A=(Q, Σ, υ, q₀, F)的语言,
- $L(A) = \{ w \in \Sigma^* \mid \tilde{\upsilon}(q_0, w) \cap F \neq \phi \}_{\circ}$
- 这个定义也显示,只要有一个成功线索,就表示接受该输入串。
- 证明例2.12接受{0,1}上后缀为01的串组成的语言
 - 命题一: w∈Σ*, q₀∈ῦ(q₀, w);
 - 命题二: $q_1 \in \tilde{v}(q_0, w)$, 当且仅当w以0结尾;
 - 命题三: q_2 ∈ $\tilde{v}(q_0, w)$,当且仅当w以01结尾。



>> 小结

• DFA: 定义及表示、判定性算法、扩展转移函数、语言与DFA构建。

• NFA: 定义及表示,判定性算法、扩展转移函数、线索穷举,活动状态集

• 习题

• p50: 2.1~2.4