自动机理论基础

李建文

华东师范大学

一个闸道的小例子

闸道是如何工作的?

一个闸道的小例子

闸道是如何工作的?

1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行

一个闸道的小例子

闸道是如何工作的?

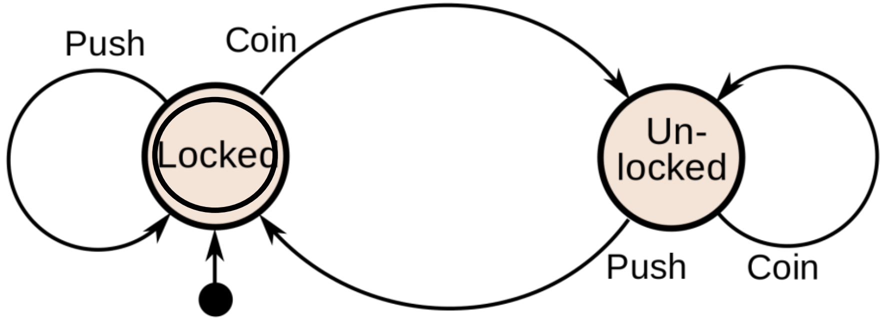
1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行

Current State	Input	Next State
Locked	coin	UnLocked
	push	Locked
UnLocked	coin	UnLocked
	push	Locked

一个闸道的小例子

闸道是如何工作的?

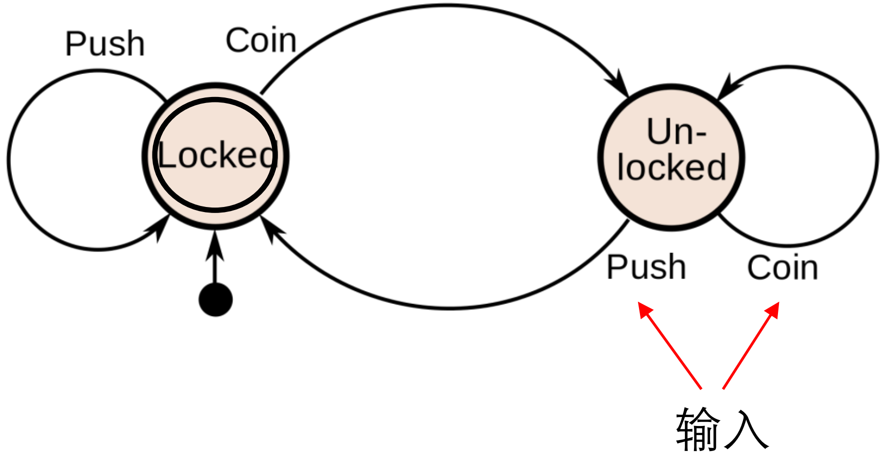
1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行



一个闸道的小例子

闸道是如何工作的?

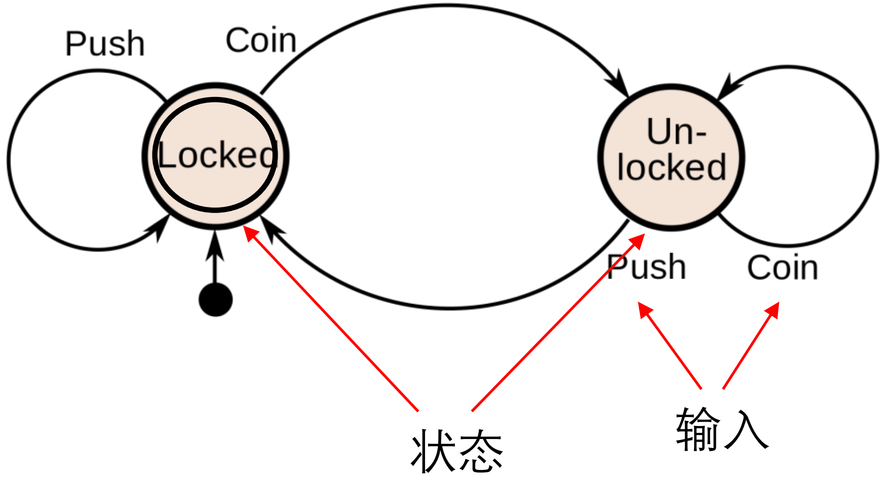
1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行



一个闸道的小例子

闸道是如何工作的?

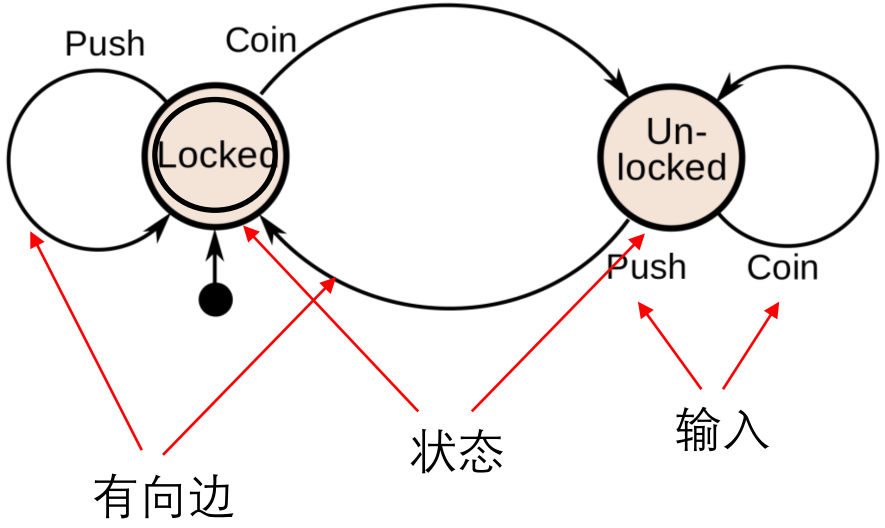
1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行



一个闸道的小例子

闸道是如何工作的?

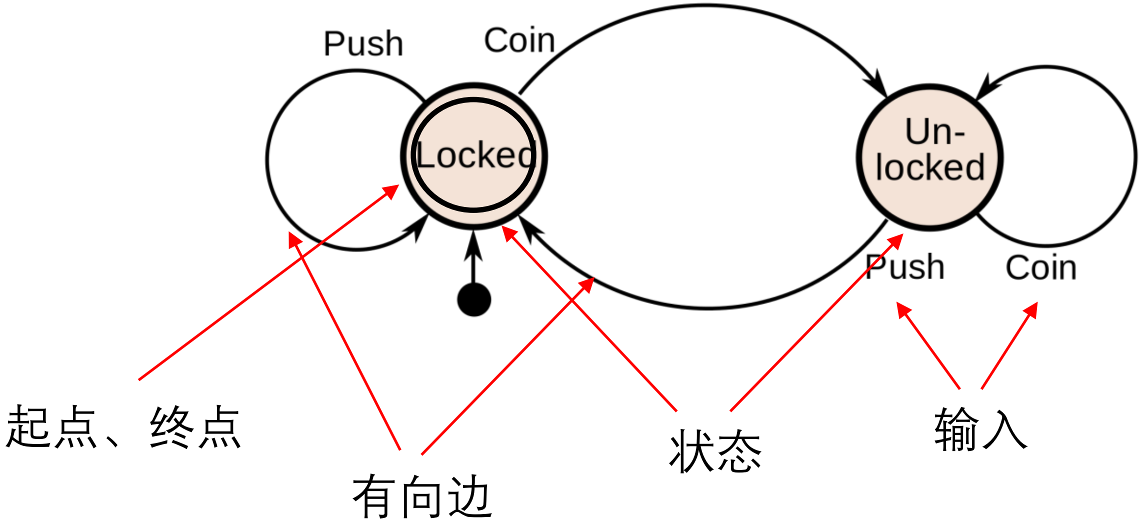
1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行



一个闸道的小例子

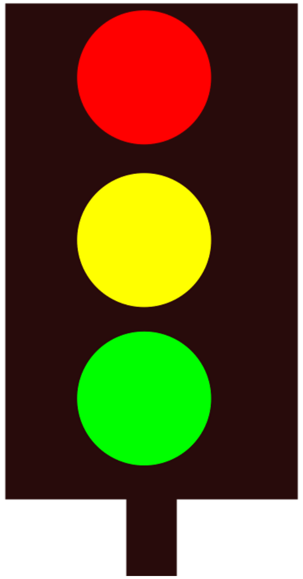
闸道是如何工作的?

1. 不投硬币之前不能通行
2. 投入一个硬币就可以通行
3. 通行一次之后立马变得不能通行



自动机!

课堂练习



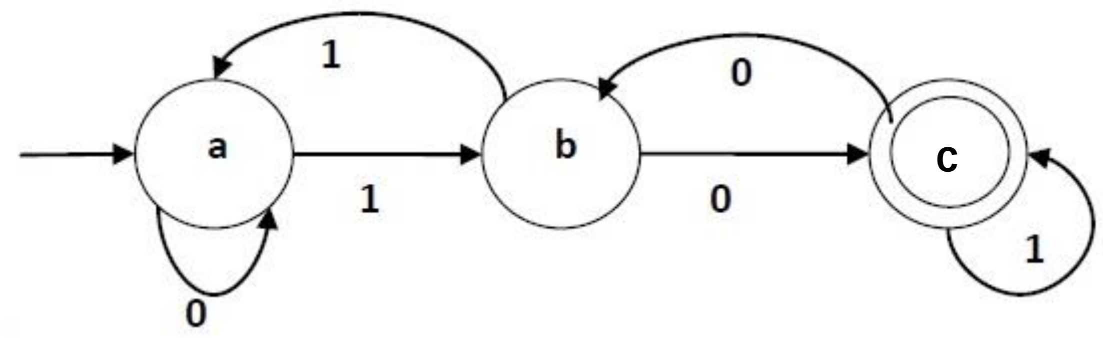
类比上面闸道的例子，画出交通灯对应的状态转换图。

有限状态自动机（Finite Automata）

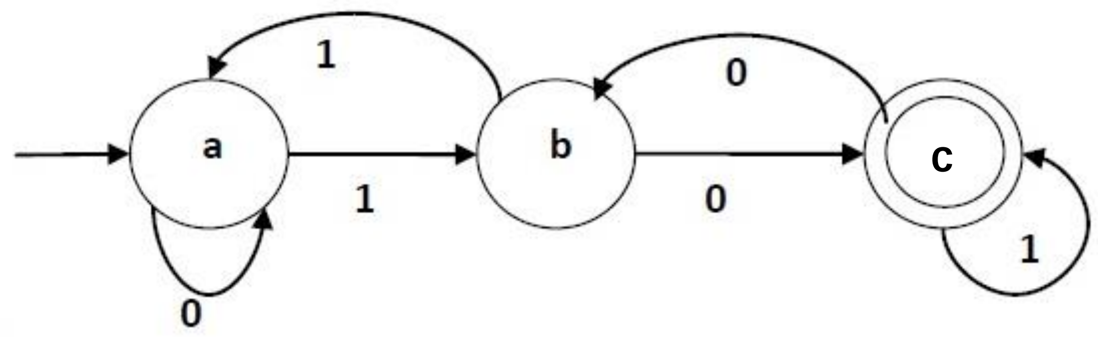
一个有限自动机可以表示成五元组$A = (\Sigma, S, T, I, F)$，其中

- Σ表示字母表的集合
- S表示状态的集合
- $T: S \times \Sigma \times S$表示边的集合
- $I \subseteq S$表示初始状态的集合
- $F \subseteq S$表示终止（接收）状态的集合

有限状态自动机（Finite Automata）



有限状态自动机（Finite Automata）



Σ={0,1}

S={a, b, c}

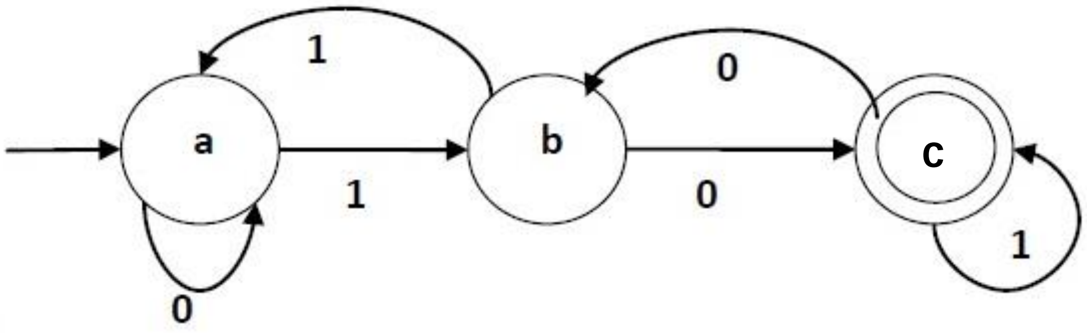
T={(a,0,a), (a, 1, b), (b, 0, c), (b, 1, a),
(c, 0, b), (c, 1, c)}

I = {a}

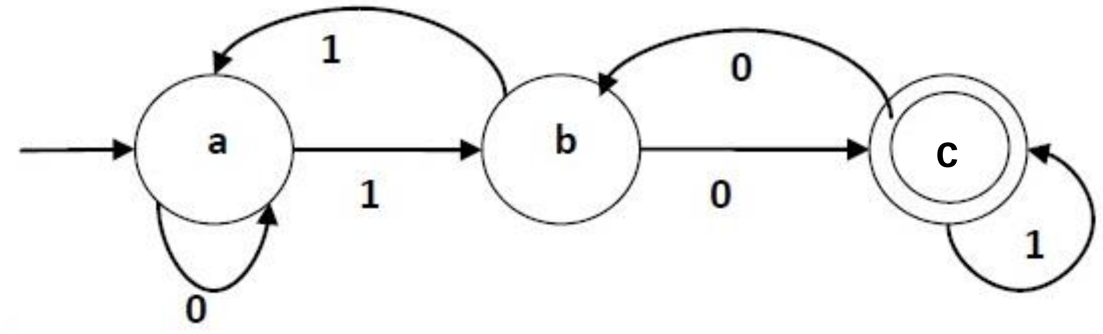
F = {c}

自动机有什么用?

有限自动机接收字符串（正则语言）



有限自动机接收字符串（正则语言）

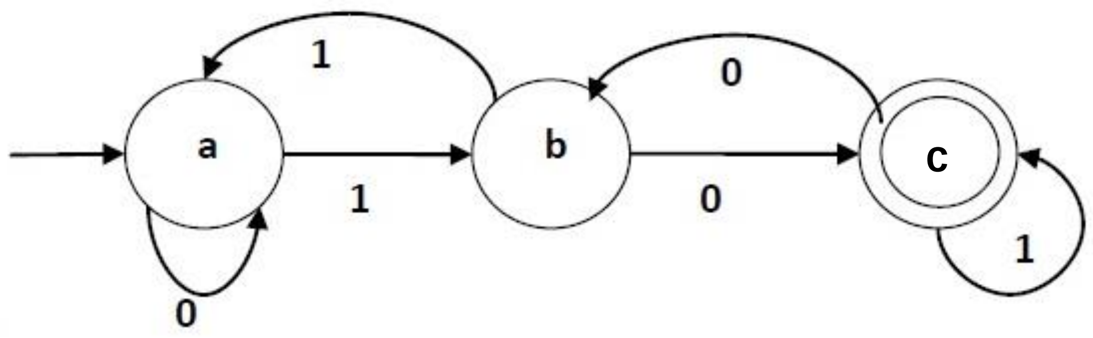


10可以被接收

10010010可以被接收

100100100不可以被接收

有限自动机接收字符串（正则语言）



10可以被接收

10010010可以被接收

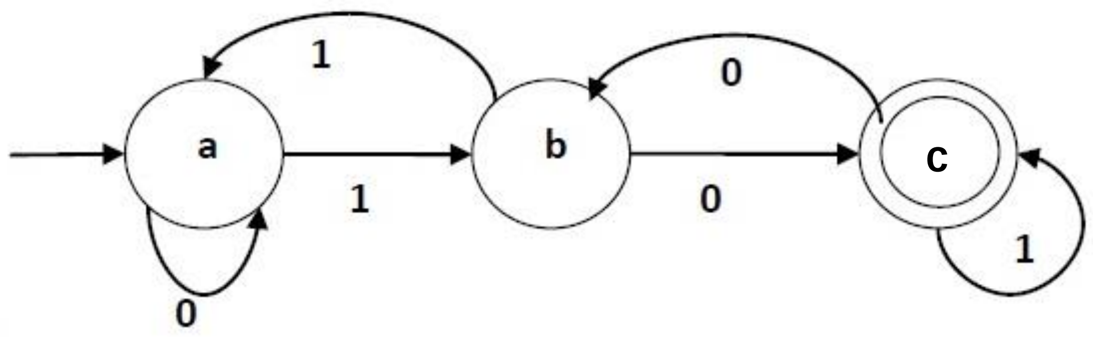
100100100不可以被接收

Q: 怎么样的字符串才可以被自动机接收?

有限自动机的语义

- 有限自动机$A = (\Sigma, S, T, I, F)$接收一组字符串
- 给定一个字符串$\eta = a_0 a_1 \cdots a_n$，$\eta$在$A$上的运行轨迹是一条有限长度的状态序列$s_0 s_1 \cdots s_n\ s_{n+1}$，使得$s_0$是一个初始状态并且$(s_i, a_i, s_{i+1})$是$A$上的一条边。
- 一个字符串$\eta = a_0 a_1 \cdots a_n$可以被$A$接收当且仅当存在$\eta$在$A$上的一条运行轨迹是以$A$中的某个终止（接收）状态结束。
- $L(A)$用来表示A可以接收的所有字符串的集合。

课堂练习： 自动机的轨迹



求下列字符串在自动机上对应的轨迹。

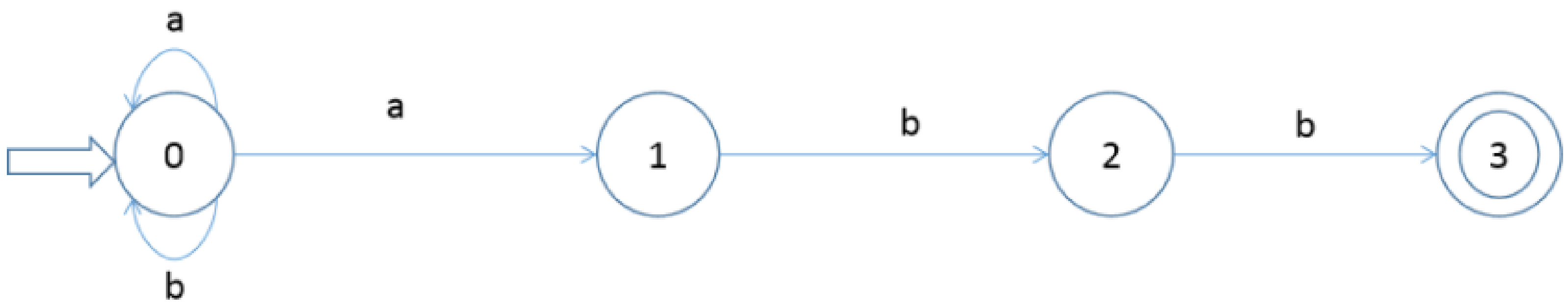
（1）10

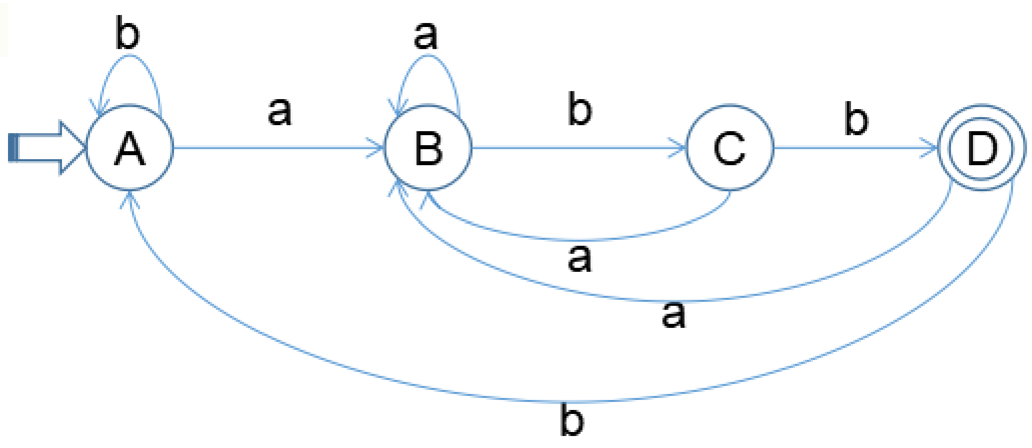
（2）10010010

（3）100100100

引申阅读：关于正则语言和自动机的历史

确定 VS. 非确定自动机





考虑字符串ab，在这两个自动机上的运行轨迹有什么差别？

非确定有限状态自动机

一个非确定有限自动机可以表示成五元组$A = (\Sigma, S, T, I, F)$，其中

- Σ表示字母表的集合
- S表示状态的集合
- $T: S \times \Sigma \rightarrow 2^S$表示迁移函数
- $I \subseteq S$表示初始状态的集合
- $F \subseteq S$表示终止（接收）状态的集合

确定有限状态自动机

一个确定有限自动机可以表示成五元组$A = (\Sigma, S, T, I, F)$，其中

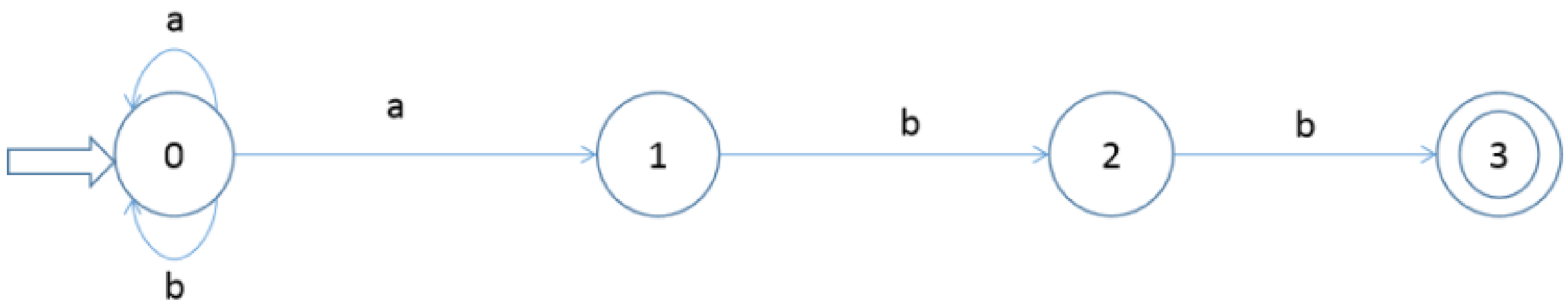
- Σ表示字母表的集合
- S表示状态的集合
- $T: S \times \Sigma \rightarrow S$表示迁移函数
- $I \subseteq S$表示初始状态的集合
- $F \subseteq S$表示终止（接收）状态的集合

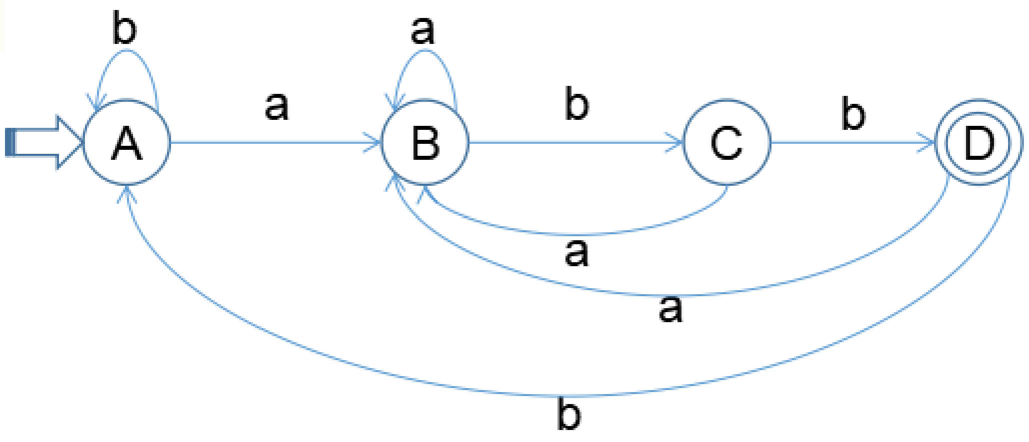
非确定自动机和确定自动机接收的语言有没有不同?

非确定自动机和确定自动机接收的语言有没有不同?

没有不同。

确定 VS. 非确定自动机





子集构造（Subset Construction）

	a	b
{0}	{0,1}	{0}
{0,1}	{0,1}	{0,2}
{0,2}	{0,1}	{0,3}
{0,3}	{0,1}	{0}

课堂练习：子集构造

3. 将图 3.16 中的 NFA 确定化。

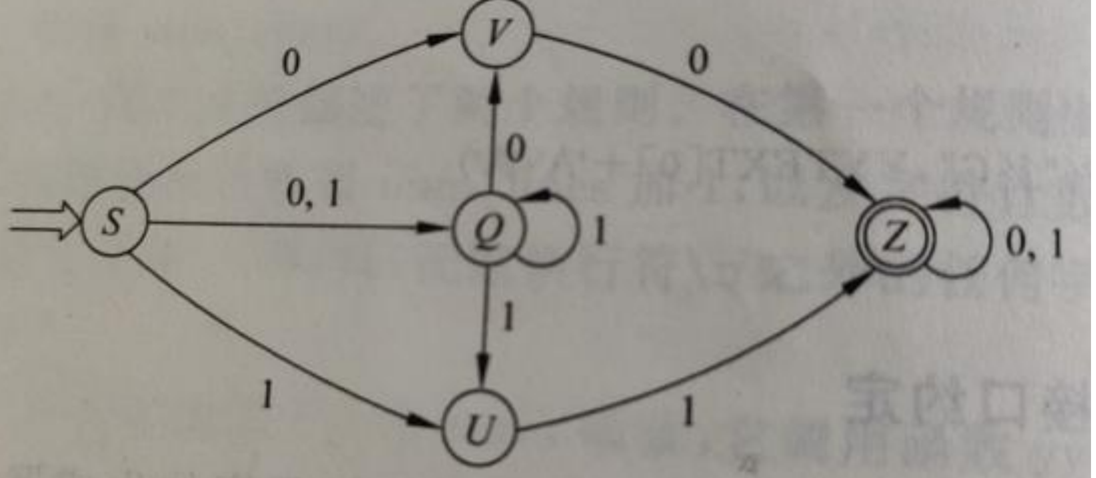
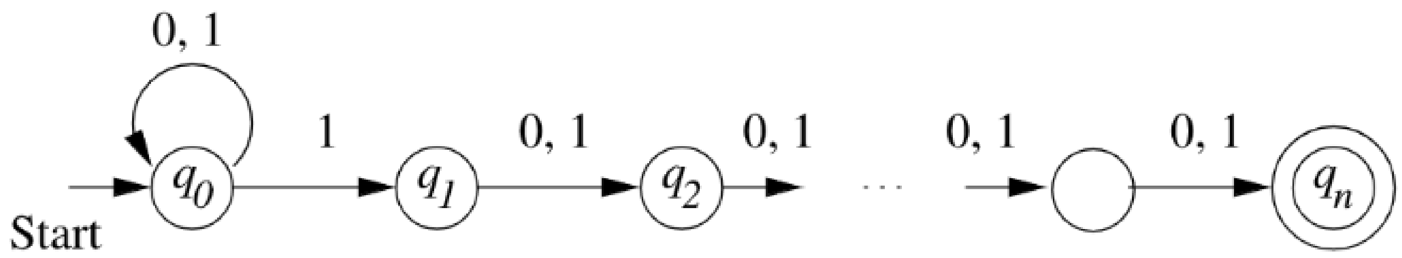


图 3.16 NFA

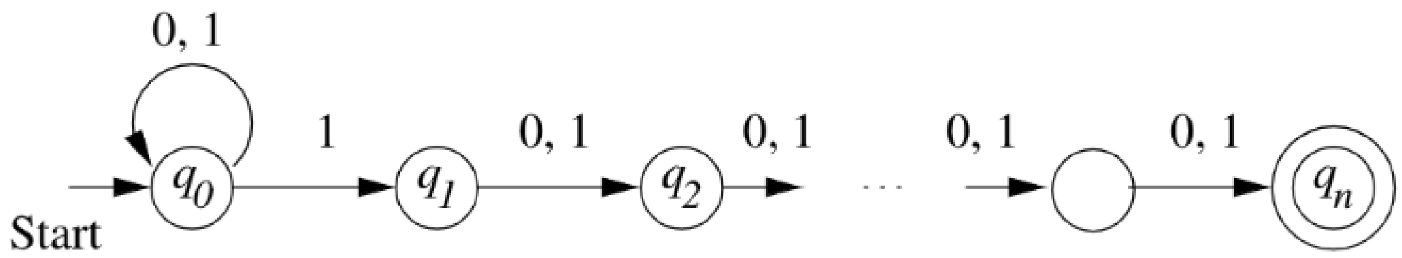
子集构造的代价

用子集构造求下面NFA对应的DFA：



子集构造的代价

用子集构造求下面NFA对应的DFA：



这个NFA对应的DFA状态数不少2^n个。

子集构造的形式化写法

给定一个非确定自动机$A = (\Sigma, S, T, s_0, F)$，与它等价的确定化自动机为$A^d = (\Sigma, Q, \rho, q_0, \Omega)$：

- $Q \subseteq 2^S$为状态的集合；
- $\rho: Q \to Q$为迁移函数，并且$\rho(q) = \{s' | s' \in T(s)$，其中$s \in q\}$成立；
- $q_0 = \{s_0\} \in Q$为初始状态；
- $\Omega \subseteq Q$为接收状态的集合，并且$q \in \Omega$当且仅当$q \cap F \neq \emptyset$。

自动机的取反

给定一个自动机A，它能接收的所有字符串集合记为$L(A)$。$L(A)$也称为A能接收的语言。

自动机的取反

给定一个自动机A，它能接收的所有字符串集合记为$L(A)$。$L(A)$也称为A能接收的语言。

自动机取反是给定一个自动机A_1，求另外一个自动机A_2使得它们接收的语言互补，即$L(A_1) \cap L(A_2) = \emptyset$，并且$L(A_1) \cup L(A_2) = \Sigma^*$。

自动机的取反

流程:

1. 给定一个自动机A，先求解对应的确定化自动机A^d
2. 将A^d中接收状态和非接收状态调换，其他保持不变，得到自动机A';
3. A'即为A的取反自动机。

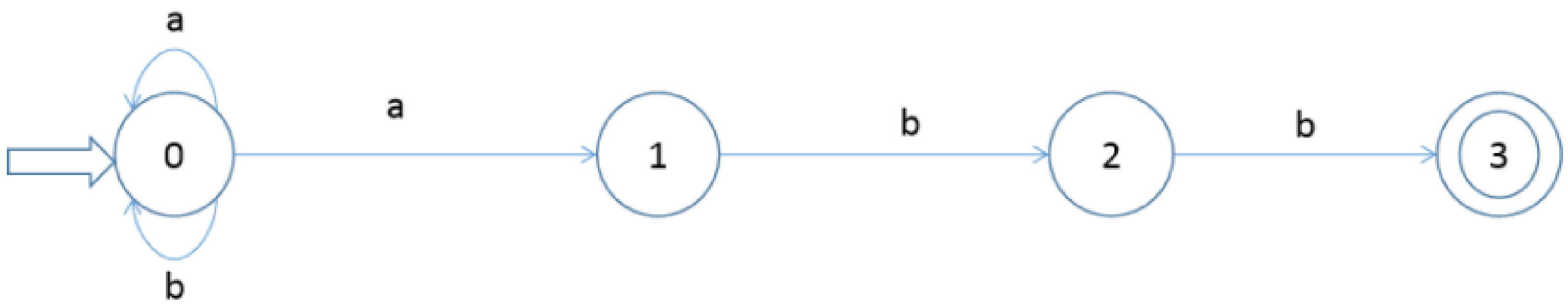
自动机的取反（形式化的写法）

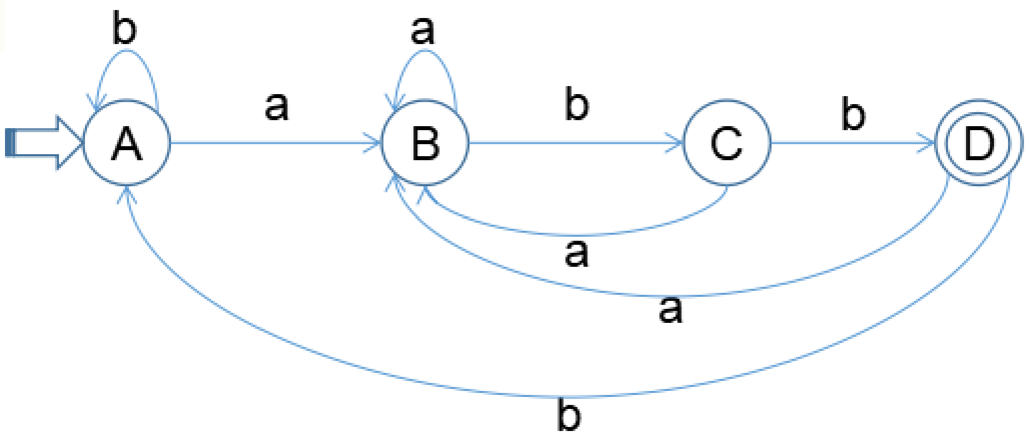
课堂练习（5分钟）

自动机的取反（形式化的写法）

给定一个有限自动机$A=(\Sigma,S,T,I,F)$，其对应的确定化自动机为$A^d=(\Sigma,Q,\rho,q_0,\Omega)$。则与$A$互补的自动机为$\hat{A}=\left(\Sigma,Q,\rho,q_0,\hat{\Omega}\right)$并且$q\in\hat{\Omega}$当且仅当$q\notin\Omega$。

自动机的取反





?

自动机的并（Union）

给定一个自动机A，它能接收的所有字符串集合记为$L(A)$。$L(A)$也称为A能接收的语言。

自动机求并是给定两个自动机A_1和A_2，求另外一个自动机A_3使得$L(A_3) = L(A_1) \cup L(A_2)$。

自动机的并（Union）

流程:

给定两个自动机A_1、 A_2

1. 将A_1和 A_2的 初始状态合并，其他保持不变，得到自动机 A_3 ；
2. A_3即为所求自动机。

自动机的并（形式化写法）

给定两个自动机 $A_1 = (\Sigma_1, S_1, T_1, I_1, F_1)$，$A_2 = (\Sigma_2, S_2, T_2, I_2, F_2)$，它们的并为自动机$A = (\Sigma, S, T, I, F)$使得

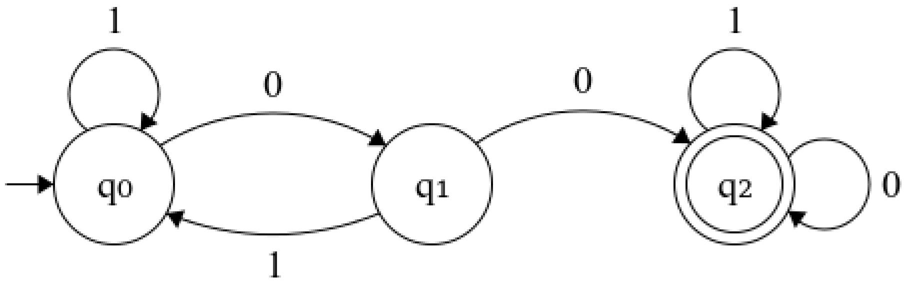
1. $\Sigma = \Sigma_1 = \Sigma_2$；
2. $S = S_1 \cup S_2$；
3. $T = T_1 \cup T_2$；
4. $I = I_1 \cup I_2$；
5. $F = F_1 \cup F_2$。

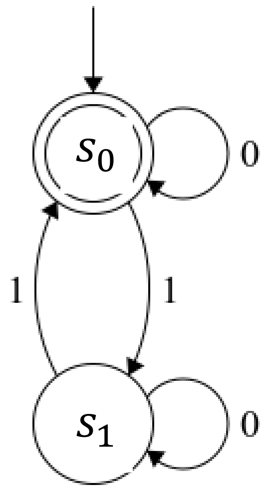
自动机的交（Intersection）

给定一个自动机A，它能接收的所有字符串集合记为$L(A)$。$L(A)$也称为A能接收的语言。

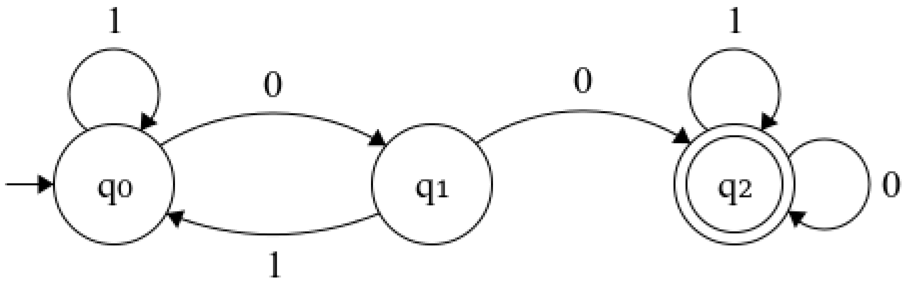
自动机求交是给定两个自动机A_1和A_2，求另外一个自动机A_3使得$L(A_3) = L(A_1) \cap L(A_2)$。

自动机的交--一个例子

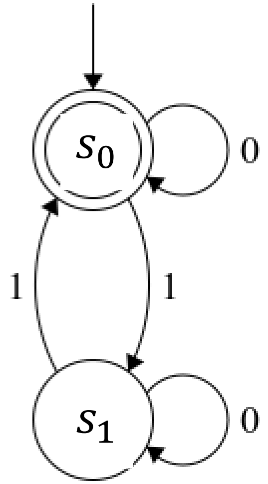




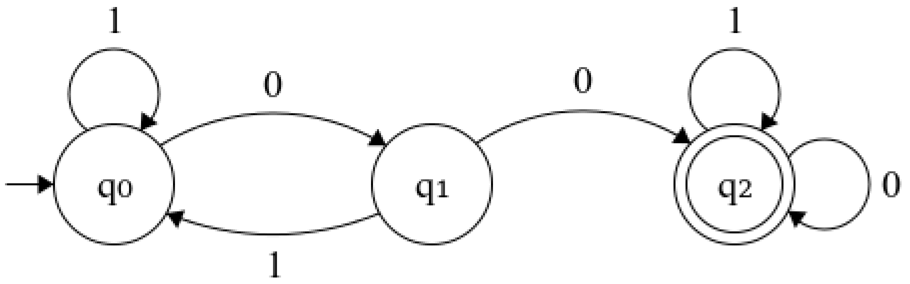
自动机的交--一个例子

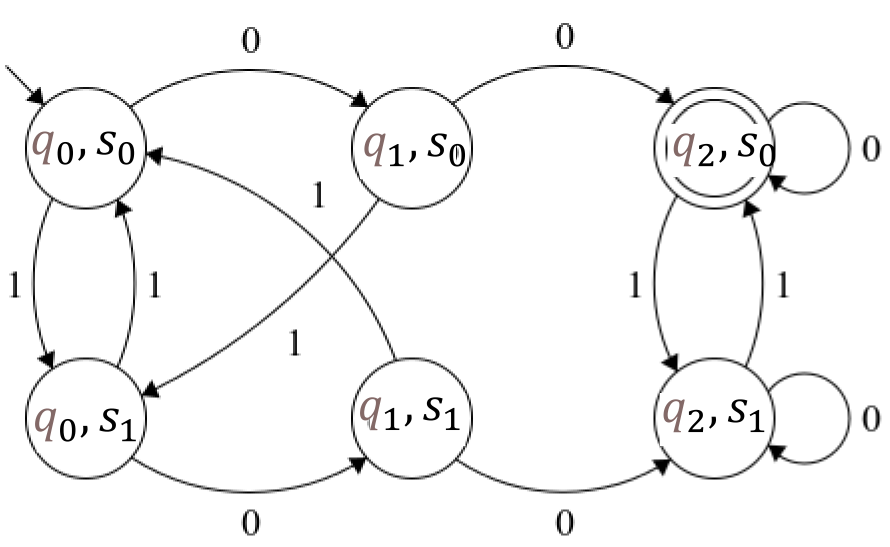


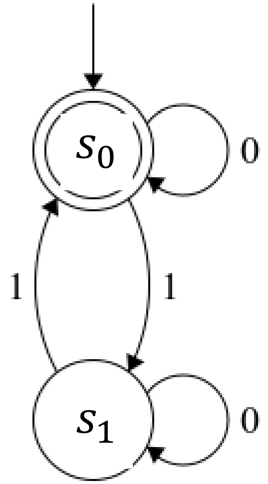
	0	1
(q0, s0)	(q1, s0)	(q0, s1)
(q1, s0)	(q2, s0)	(q0, s1)
(q0, s1)	(q1, s1)	(q0, s0)
(q2, s0)	(q2, s0)	(q2, s1)
(q1, s1)	(q2, s1)	(q0, s0)
(q2, s1)	(q2, s1)	(q2, s0)



自动机的交--一个例子


q0
q1
q2
0
0
1
1
1
0


q_0, s_0
q_1, s_0
q_2, s_0
q_0, s_1
q_1, s_1
q_2, s_1
0
0
0
0
0
0
1
1
1
1
1
1

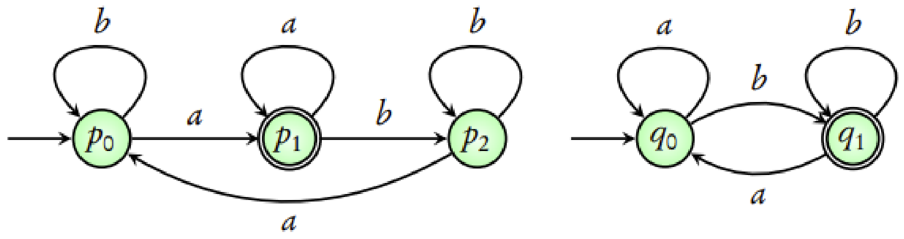

s_0
s_1
0
0
1
1

自动机的交（形式化写法）

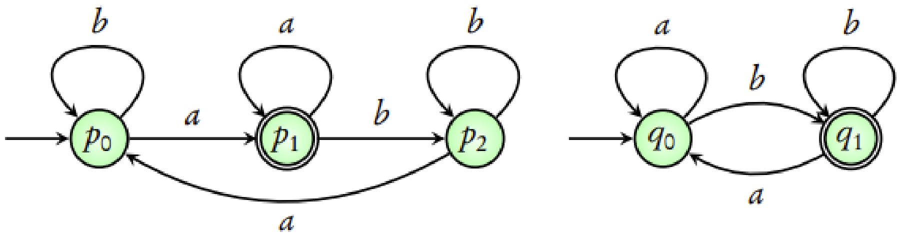
给定两个自动机 $A_1 = (\Sigma_1, S_1, T_1, I_1, F_1)$，$A_2 = (\Sigma_2, S_2, T_2, I_2, F_2)$，它们的交为自动机$A = (\Sigma, S, T, I, F)$使得

1. $\Sigma = \Sigma_1 = \Sigma_2$；
2. $S \subseteq S_1 \times S_2$；
3. $T = T_1 \times T_2$；
   $T = \{(s_1, s_2), a, (s_1', s_2') \mid (s_1, a, s_1') \epsilon T_1$且$(s_2, a, s_2') \epsilon T_2\}$
4. $I = I_1 \times I_2$；
5. $F = F_1 \times F_2$。

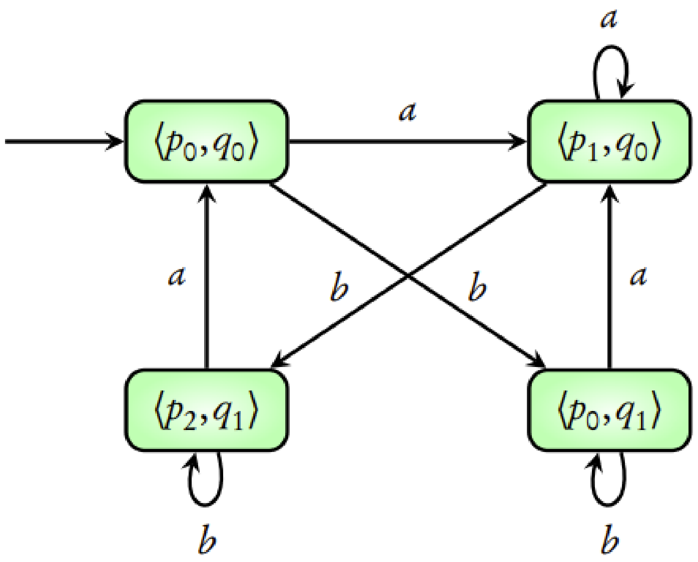
课堂练习-自动机的交



课堂练习-自动机的交



b
a
b
a
b
p_0
p_1
p_2
a
a
b
b
a
q_0
q_1
a

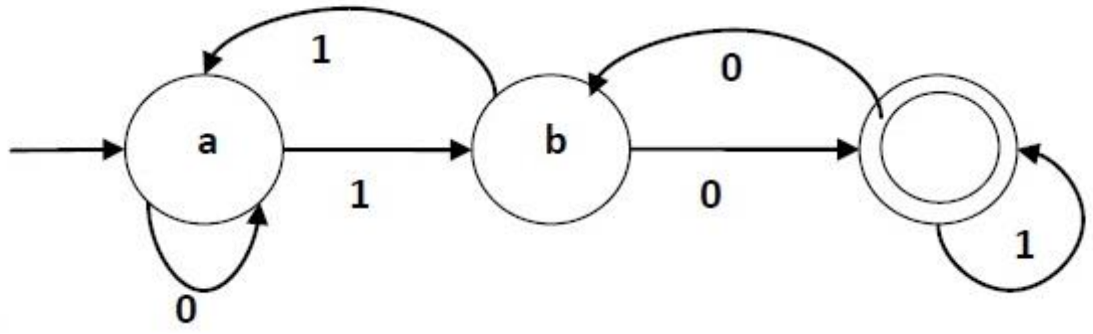


a
a
$\langle p_0,q_0\rangle$
$\langle p_1,q_0\rangle$
a
b
b
a
$\langle p_2,q_1\rangle$
$\langle p_0,q_1\rangle$
b
b

无限自动机（Infinite Automata）

- 状态是有限的
- 接收的字符串长度是无限的

有限 VS. 无限自动机



有限：10， 101， 1000
无限：$10(1)^{\omega}$， $1(0)^{\omega}$

$B\ddot{u}$ ch i自动机

一个$B\ddot{u}$ ch i自动机可以表示成五元组$A = (\Sigma, S, T, I, F)$，其中

- Σ表示字母表的集合
- S表示状态的集合
- $T: S \times \Sigma \times S$表示边的集合
- $I \subseteq S$表示初始状态的集合
- $F \subseteq S$表示终止（接收）状态的集合

$B\ddot{u}$ch i自动机的语义

- 有限自动机$A = (\Sigma, S, T, I, F)$接收一组**无限长度**的字符串

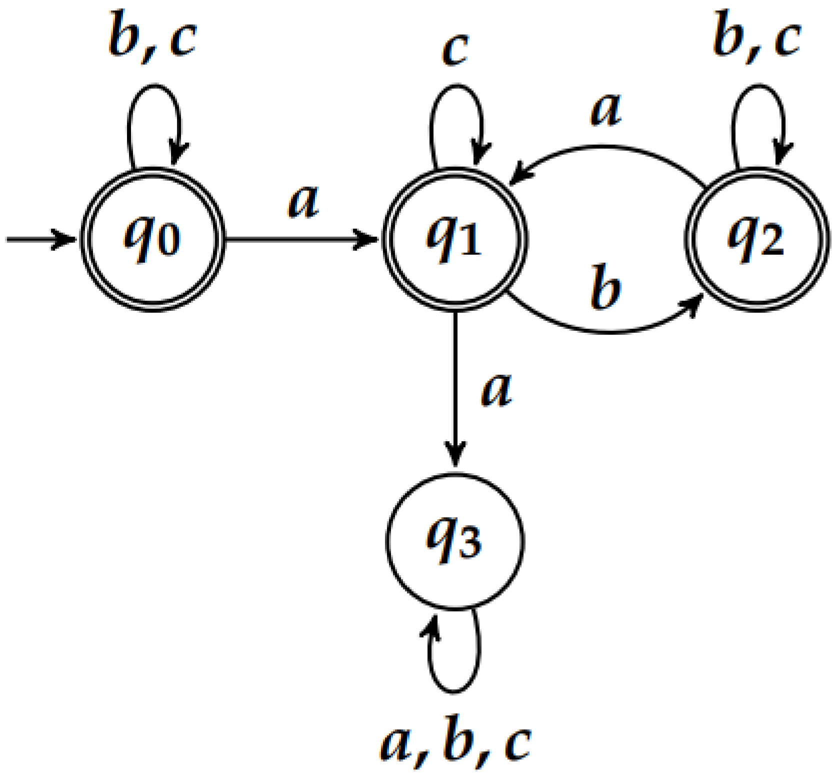
- 给定一个字符串$\eta = a_0 a_1 \cdots$，η在A上的运行轨迹是一条无限长度的状态序列$s_0 s_1 \cdots$，使得s_0是一个初始状态并且(s_i, a_i, s_{i+1})是A上的一条边。

- 一个字符串$\eta = a_0 a_1 \cdots$可以被A接收当且仅当存在η在A上的一条运行轨迹，并且该轨迹**无限次的经过**F中的某个接收状态。

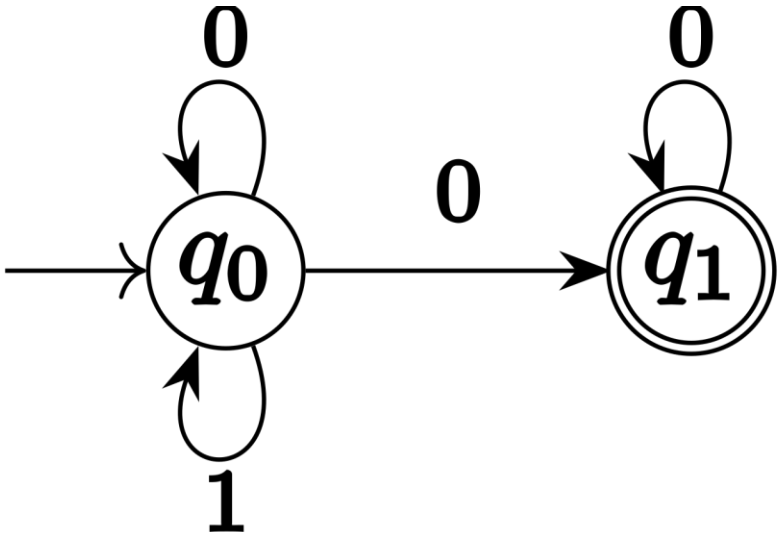
- $L(A)$用来表示A可以接收的所有字符串的集合。

课堂练习 – $B\ddot{u}$ ch i自动机

该$B\ddot{u}$ ch i自动机接收的语言有哪些?



确定 VS. 非确定 Büchi自动机



不存在等价的确定 Büchi自动机。

确定 Büchi自动机要比非确定 Büchi自动机的表达能力弱

$B\ddot{u}$ ch i自动机的确定化

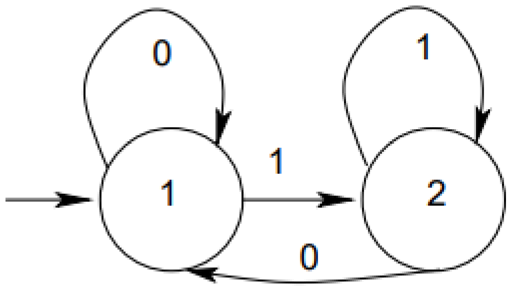
- 一个可以挖掘的research topic
- 基本思路：把非确定的$B\ddot{u}$ ch i自动机转化成等价的其他类型的无限自动机

Rabin自动机

- 接收条件：$\{(\mathrm{B}_1, G_1), (\mathrm{B}_2, G_2), \cdots, (\mathrm{B}_n, G_n)\}$，每个$\mathrm{B}_i$和$G_i$都是S的一个子集

- 一个字符串$\eta = a_0 a_1 \cdots$可以被Rabin自动机A接收当且仅当存在η在A上的一条运行轨迹和一个(B_i, G_i)，使得该轨迹有限次的经过B_i但是无限次的经过G_i中的某个接收状态。

Rabin自动机



若接收条件为：{({1}, {2}), ({}, {1,2})}
那么该自动机接收的语言有哪些？

Safra Construction

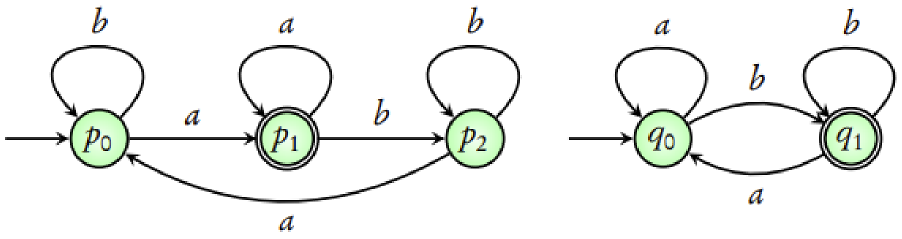
- 将非确定的$B\ddot{u}$ ch i自动机转换为等价确定化Rabin自动机
- 理论上是最优的
- 实际实现时数据结构比较复杂，比较难实现
- 有没有更好的转换方案？

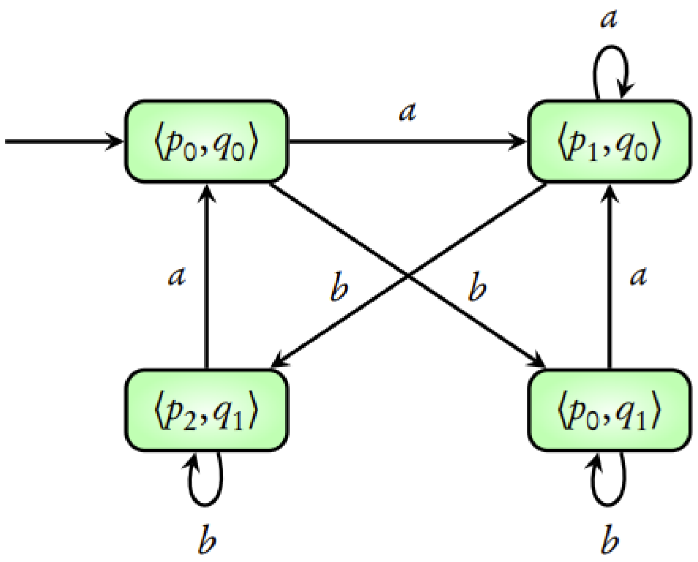
$B\ddot{u}$ch i自动机的交

给定一个自动机A，它能接收的所有字符串集合记为$L(A)$。$L(A)$也称为A能接收的语言。

自动机求交是给定两个自动机A_1和A_2，求另外一个自动机A_3使得$L(A_3) = L(A_1) \cap L(A_2)$。

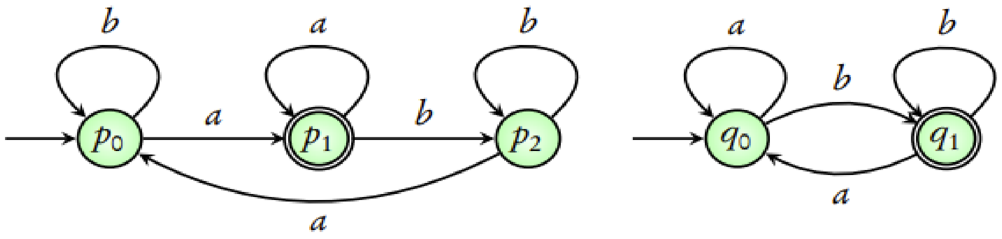
Büchi自动机的交

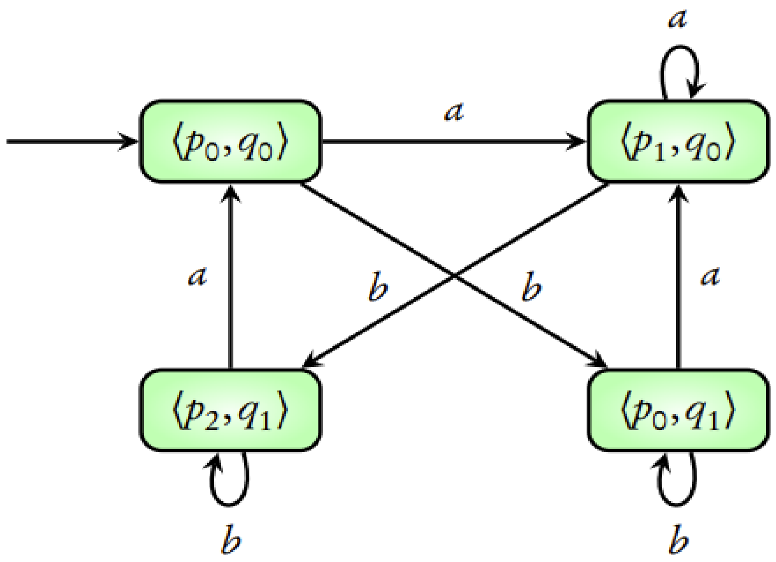




没有接收状态?

$B\ddot{u}$ch i自动机的交

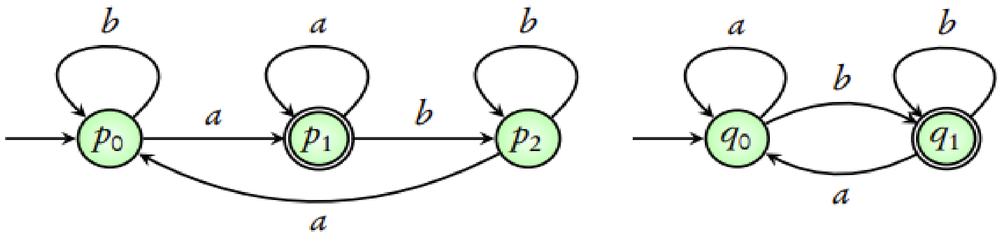




没有接收状态?

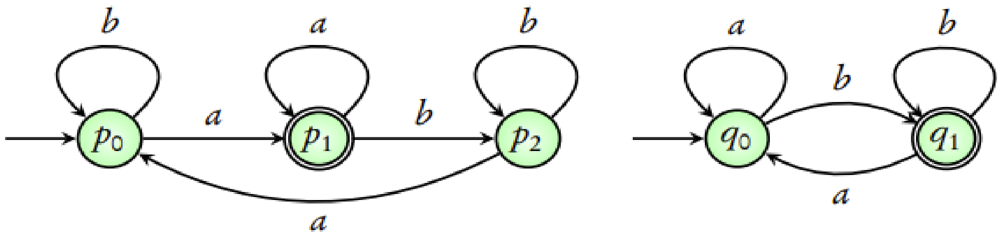
$a, b, (\mathrm{a}, \mathrm{a}, \mathrm{b})^{\omega}$是被两个自动机都接收的语言。

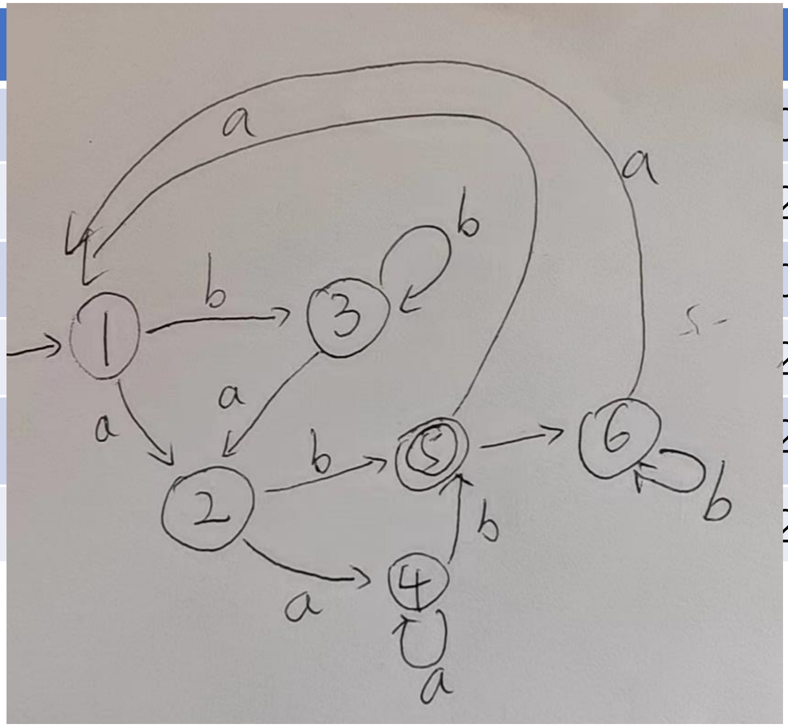
Bü ch i自动机的交



	a	b
(p0, q0, 1)	(p1, q0, 1)	(p0, q1, 1)
(p1, q0, 1)	(p1, q0, 2)	(p2, q1, 2)
(p0, q1, 1)	(p1, q0, 1)	(p0, q1, 1)
(p1, q0, 2)	(p1, q0, 2)	(p2, q1, 2)
(p2, q1, 2)	(p0, q0, 1)	(p2, q1, 1)
(p2, q1, 1)	(p0, q0, 1)	(p2, q1, 1)

Bü ch i自动机的交





(p0, q0, 1)	0, q1, 1)
(p1, q0, 1)	2, q1, 2)
(p0, q1, 1)	0, q1, 1)
(p1, q0, 2)	2, q1, 2)
(p2, q1, 2)	2, q1, 1)
(p2, q1, 1)	2, q1, 1)

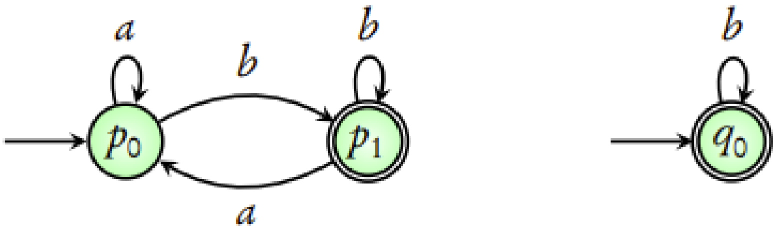
$B\ddot{u}$ch i自动机的交（形式化写法）

给定两个自动机$A_1=(\Sigma_1,S_1,T_1,I_1,F_1)$，$A_2=(\Sigma_2,S_2,T_2,I_2,F_2)$，它们的交为自动机$A=(\Sigma,s,T,I,F)$使得

1. $\Sigma=\Sigma_1=\Sigma_2$；
2. $S\subseteq S_1\times S_2\times\{1,2\}$；
3. $T=\Delta_1\cup\Delta_2$，
   Δ_1={((s_1,s_2,1), a, (s_1',s_2',i) | (s_1, a, s_1')ϵT_1且(s_2, a, s_2')ϵT_2且若$s_1\epsilon F_1$则i=2，否则i=1}；
   Δ_1={((s_1,s_2,2), a, (s_1',s_2',i) | (s_1, a, s_1')ϵT_1且(s_2, a, s_2')ϵT_2且若$s_2\ \epsilon F_2$则i=1，否则i=2}。

4. $I=I_1\times I_2\times\{1\}$；
5. $F=\{(s_1,s_2,2)\ |s_2\epsilon F_2\}$。

课堂练习



本章小结

- 简要接收有限和无限自动机的基本概念

- 介绍自动机的确定化、取反、交等操作