第四讲

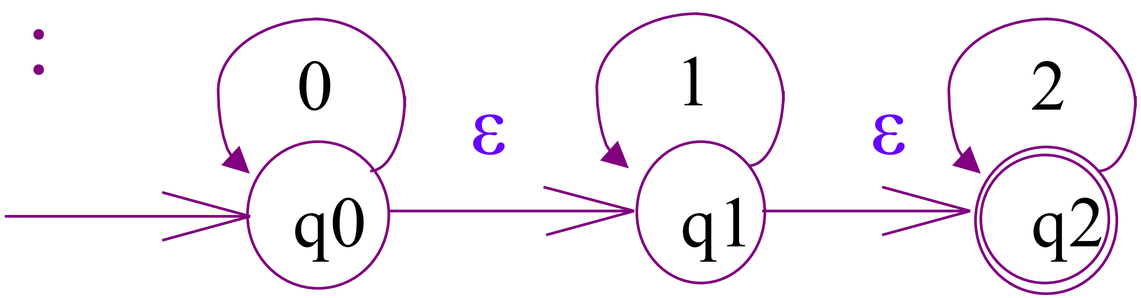
- 带 ε-转移的有限自动机
- 正则表达式
- 右线性文法与正则集

第四节　有ε转换的NFA

一、定义

概念: 当输入空串ε (无输入) 时，也能引起状态的转移.

例：



输入“002”时的转移格局:

$$q_0 \xrightarrow{0} q_0 \xrightarrow{0} q_0 \xrightarrow{\varepsilon} q_1 \xrightarrow{\varepsilon} q_2 \xrightarrow{2} q_2$$

ε - NFA 的形式定义

一个 ε - NFA 是一个五元组 $A = (Q, T, \delta, q_0, F)$.

- ✧ 有限状态集
- ✧ 有限输入符号集
- ✧ 转移函数
- ✧ 一个开始状态
- ✧ 一个终态集合

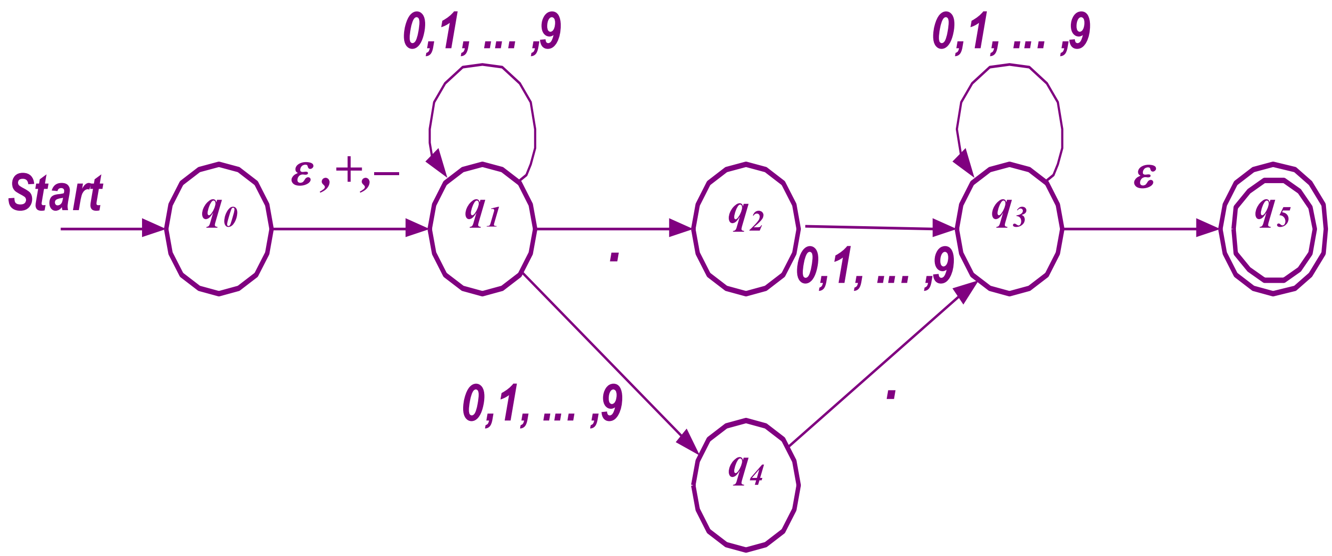
$q_0 \in Q$

$F \subseteq Q$

- ✧ 与 NFA 的不同之处

$$\delta: Q \times (T \cup \{\varepsilon\}) \rightarrow 2^Q$$

ε - NFA 如何接受输入符号串



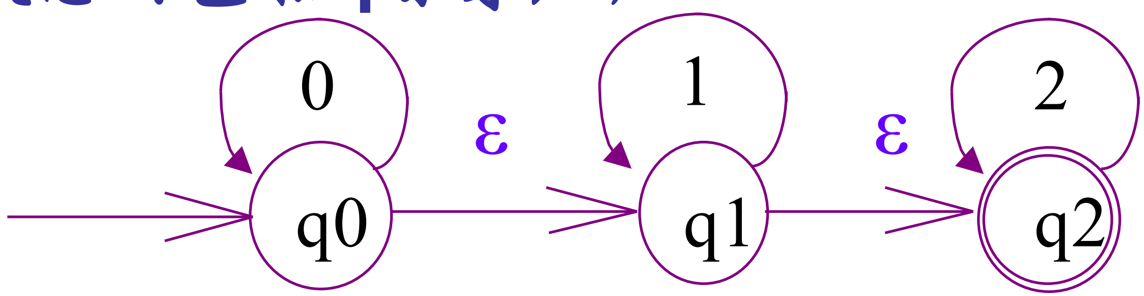
✧ 该 ε - NFA 可以接受的字符串如：

- 3.14
- +.314
- – 314.

二、 ε - 闭包（closure）概念

✧ 状态 q 的ε - 闭包，记为 ε - CLOSURE 或 ECLOSE，定义为从 q 经所有的ε 路径可以到达的状态（包括q自身），

如：



- ε - CLOSURE (q0) = {q0 , q1， q2 }
- ε - CLOSURE (q1) = { q1， q2 }
- ε - CLOSURE (q2) = {q2}

- **状态子集I 的ε-闭包：**

  $$\varepsilon\text{-CLOSURE}(I) = \bigcup_{q \in I} \varepsilon\text{-CLOSURE}(q)$$

例：

$$\varepsilon\text{-CLOSURE}(\{q1, q2\}) = \varepsilon\text{-CLOSURE}(q1) \cup \varepsilon\text{-CLOSURE}(q2) = \{q1, q2\}$$

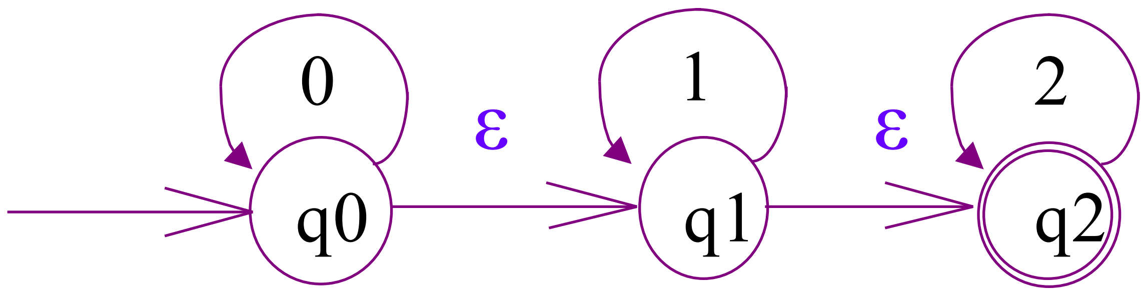
- **Ia 概念：**

  对于状态子集$I \subseteq Q$，任意$a \in T$，定义Ia如下：

  Ia= ε -Closure(P)

  其中P =δ（I，a）. 即P是从I中的状态经过一条标a的边可以到达的状态集合

例：$I=\{q0, q1\}$，求I_1



$$I_1 = \varepsilon\text{-CLOSURE}(\delta(I, 1))$$
$$= \varepsilon\text{-CLOSURE}(\delta(\{q0, q1\}, 1))$$
$$= \varepsilon\text{-CLOSURE}(\Phi \cup \{q1\})$$
$$= \{q1, q2\}$$

扩展转移函数适合于输入字符串

✧ 设一个 ε-*NFA*, δ： $Q \times T \cup \{\varepsilon\} \to 2^Q$

✧ 扩充定义 δ': $Q \times T^* \to 2^Q$

✧ 对任何 $q \in Q$，定义：

1 $\delta'(q, \varepsilon) = \varepsilon\text{-CLOSURE}(q)$

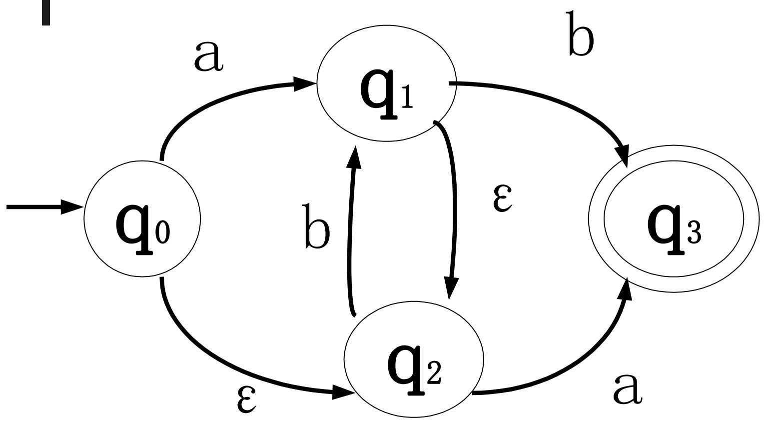
2 $\delta'(q，\omega a) = \varepsilon\text{-CLOSURE}(P)$

其中 $P = \{ p \mid$ 存在 $r \in \delta'(q，\omega) \wedge p \in \delta(r，a)\}$

注意： 此时 $\delta(q，a) \neq \delta'(q，a)$，

因为 $\delta(q，a)$ 表示由 q 出发，只沿着标 a 的路径所能到达的状态，而 $\delta'(q，a)$ 表示由 q 出发，沿着标 a（包括 ε 转换在内）的路径所能到达的状态.

ε-NFA中，δ与δ’ 函数的不同



ε-CLOSURE(q0)＝{q0，q2}
ε-CLOSURE(q1)＝{q1，q2}
ε-CLOSURE(q2)＝{q2}
ε-CLOSURE(q3)＝{q3}

✧ **举例 计算** δ′ (q_0, a)

δ'(q0，ε)＝ε-CLOSURE(q0) ＝{q0，q2}

δ'(q0，a)＝ε-CLOSURE(δ(δ'(q0，ε)，a))

＝ε-CLOSURE(δ({q0，q2}，a))

＝ε-CLOSURE(δ(q0，a) ∪δ(q2，a))

＝ε-CLOSURE({q1}∪{q3})

＝{q1，q2} ∪{q3}

＝{q1，q2，q3}

同理：δ‘(q0，aa)＝ ?

同理：δ‘(q0，aa)＝ {q3}

三、ε - NFA 的语言

- 设一个 ε - NFA $M = (Q, T, \delta, q_0, F)$
- 定义 M 的语言：

$$L(M) = \{ \omega \mid \delta'(q_0, \omega) \cap F \neq \phi \}$$

即 满足 δ'(q0， ω) 含有F的一个状态

四、有 ε 转换的NFA与无 ε 转换的NFA的等价

1. ε -NFA<==>NFA

具有ε转移的NFA是不具ε转移的NFA的一般情况, 所以只要证明下面的定理即可:

定理: 如果L被一个具有 ε 转移的NFA接收, 那么L可被一个不具 ε 转移的NFA 接收.

证明思路: 构造一个不具 ε 转移的NFA, 证明其接收具有 ε 转移的NFA所接受的语言.

从 ε - NFA 构造等价的 无ε NFA

✧ 设 $M = (Q, T, \delta, q_0, F)$ 是一个 ε - NFA，可构造一个无ε的 NFA $M_1 = (Q, T, \delta_1, q_0, F_1)$,

- 对任何 $a \in T$， $\delta_1(q, a) = \delta'(q, a)$.

$$F_1 = \begin{cases} F \cup \{q0\} & \text{若}\varepsilon\text{-CLOSURE}(q0) \cap F \neq \phi \\ F & \text{否则} \end{cases}$$

从ε-NFA 构造等价的 无ε NFA

证明: 采用归纳法证明 δ_1（q0, ω）＝δ’（q0, ω），|ω|>=1。

当| w | = 0, 即 w = ε时，不一定相等.

∵ $\delta_1(q0，\varepsilon)＝\{q0\}$,

而$\delta'(q0，\varepsilon)＝\varepsilon\text{-}CLOSURE(q0)$

因此，从|ω|＝1开始证明

1. 当|ω|=1时，定义相等。

由δ_1定义

δ_1（q0，a）＝δ’（q0，a）

设当$|\omega|<=n$时，δ_1（q0,ω）=δ'（q0, ω），则当$|\omega|=n+1$时，
左侧＝ δ1（q0,ωa）

＝δ_1(δ1（q0，ω），a)

＝δ1 (δ'（q0，ω），a)　　　　由归纳假设

＝δ1 (R，a)　　　　设R＝δ'（q0，ω）

＝∪δ1 (q，a)　　　　q ∈R

＝∪δ' (q，a)　　　　q ∈R. 由δ1定义

＝δ' (R，a)

＝δ'(δ'（q0，ω），a)　　　　∵ R＝δ'（q0，ω）

＝δ' (q0，ωa)

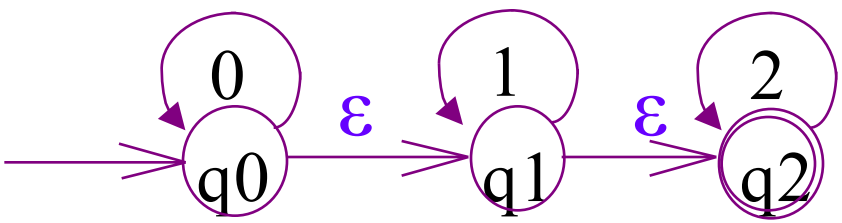
＝ 右侧

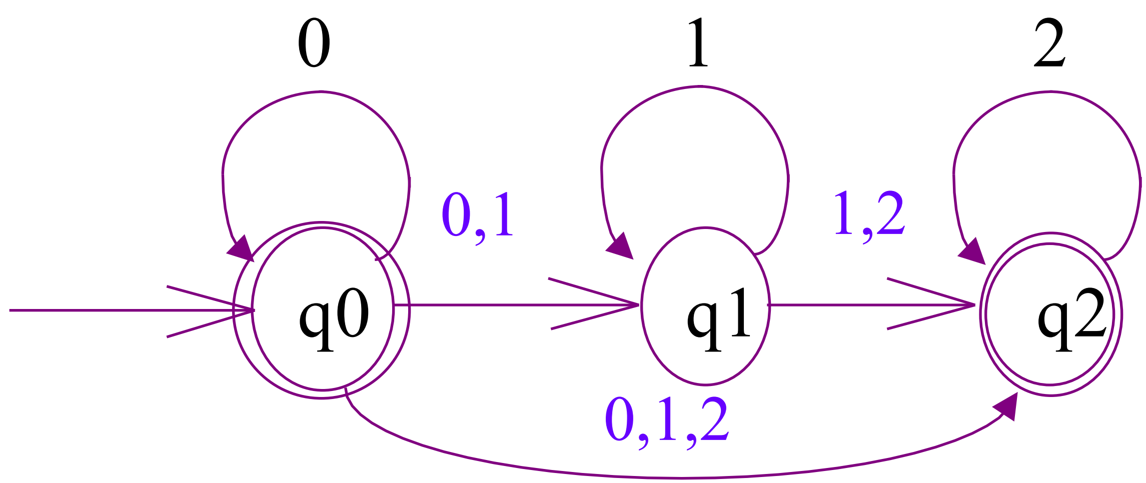
再证：δ_1（q0，ω）含F1的一个状态当且仅当δ'（q0，ω）含F的一个状态　　　　（略）

证明同时展示了一种将 ε -NFA转化为NFA的方法.

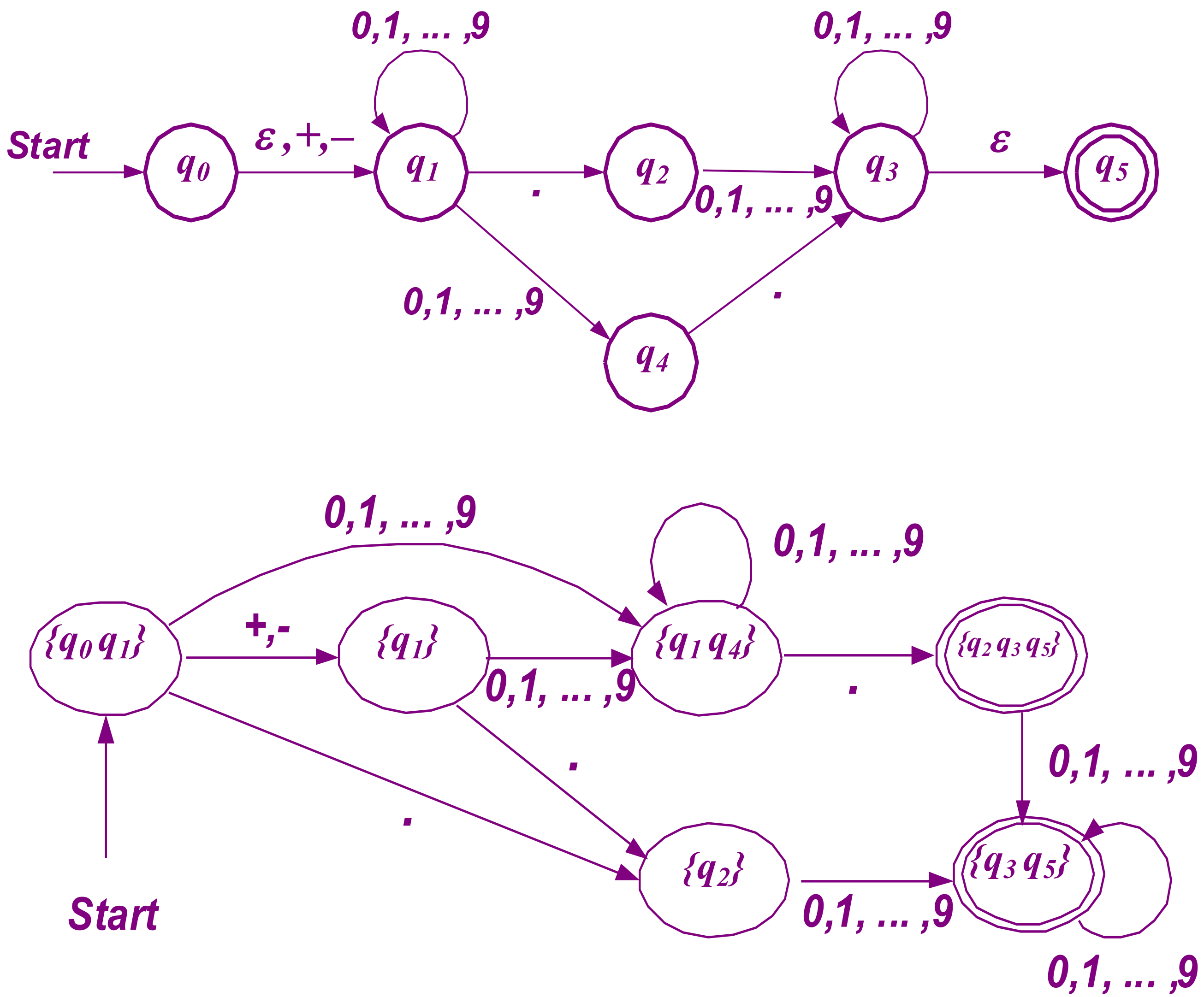
ε -NFA <==> NFA <==> DFA

例：将 ε -NFA转换为NFA.





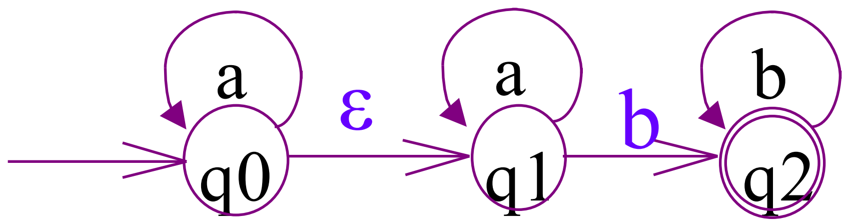
举例



Start
q0
ε,+,–
q1
0,1, … ,9
.
q2
0,1, … ,9
q3
0,1, … ,9
ε
q5
0,1, … ,9
q4
.
0,1, … ,9
{q0q1}
+,-
{q1}
0,1, … ,9
{q1q4}
0,1, … ,9
.
{q2q3q5}
0,1, … ,9
.
.
{q2}
0,1, … ,9
{q3q5}
0,1, … ,9
Start

课堂作业

将 ε -NFA转换为NFA和DFA.



第五节 正则集和正则式

- 正则集：字母表上一些特殊形式的字符串的集合，是正则式所表示的集合.
- 正则式：用类似代数表达式的方法表示正则语言。
- 运算：作用于语言上的三种代数运算
  - 联合（union）
  - 连接（concatenation）
  - （星）闭包（closure）

正则表达式（regular expression）

✧ 归纳定义正则表达式如下：

基础 ：ε, φ, a （a∈T）都是正则式 (原子正则式) ,

相应的正则集为{ε}, φ, {a}

归纳：如果A和B是正则式，且分别代表集合L(A)和L(B),

则(A+B), (A.B), A* 也是正则式，分别表示以下正则集.

L(A) ∪L(B)	(语言A / 语言B的串)
L(A).L(B)	(两个语言中的串的连接)
L(A) *	(语言A中的串的多次连接)

✧ 仅通过有限次使用以上两步定义的表达式，才是字母表T上的正则式。这些正则式所表示的字符串集合是T上的正则集。

正则表达式算符优先级

- 算符优先级（*precedence*）依次为
  - ＊（ *closure* ）
  - • 连接（*concatenation*）
  - ＋（ *union* ）

定义：若两个正则式表示相同的正则集，则称这两个正则式相等。 即 R1＝R2 <==> L(R1)=L(R2)

注1：正则集是T* 的子集。

注2：L^+包含ε当且仅当L包含ε。

注3：每个正则集至少对应一个正则式（可有无穷多个正则式）

正则表达式举例

- 书p55 例1

正则集	正则式
T 上所有 a 和 b 组成的字符串集合	(1) $(a+b)^*$
T 上所有以 a 为首后跟任意个 b 的字符串集合	(2) ab^*
T 上所有以 b 为首后跟由 a 和 b 组成的字符串集合	(3) $b(a+b)^*$
T 上所有含有两个连续 a 或两个连续 b 的字符串集合	(4) $(a+b)^*(aa+bb)(a+b)^*$

- 表示如下语言的正则表达式：语言中的每个字符串由交替的 0's 和 1's 构成
  - $(01)^* + (10)^* + 0(10)^* + 1(01)^*$
  - $(\varepsilon + 1)(01)^*(\varepsilon + 0)$
  - $(\varepsilon + 0)(10)^*(\varepsilon + 1)$

语言的联合（union）运算

- ✧ 两个语言 L 和 M 的联合

$$L \cup M = \{ w \mid w \in L \vee w \in M \}$$

- ✧ 举例

设 $L = \{ 001,10,111 \}$, $M = \{\varepsilon , 001\}$,
则

$$L \cup M = \{ \varepsilon , 10, 001, 111 \}$$

语言的连接（concatenation）运算

- 两个语言 L 和 M 的连接

  $$L \cdot M = \{ w_1 w_2 \mid w_1 \in L \wedge w_2 \in M \}$$

- 有时记 $L \cdot M$ 为 LM
- 举例

  设 $L = \{ 001, 10, 111 \}$, $M = \{\varepsilon, 001\}$, 则

  $$LM = \{ 001, 10, 111, 001001, 10001, 111001 \}$$

语言的闭包（closure）运算

✧ 语言 L 的闭包

$L^* = \{ w^n \mid w \in L \wedge n \geq 0 \}$，其中 w^n 为 w 的 n 次连接

或 $L^* = L^0 \cup L^1 \cup L^2 \cup \ldots = \cup_{i \geq 0} L^i$， 其中

$L^0 = \{\varepsilon\}$， $L^1 = L$, $L^2 = LL$, …

✧ 举例

设 $L = \{ 0, 11 \}$，则

$L^* = \{ \varepsilon, 0, 11, 00, 011, 110, 1111, 000, 0011, 0110, 01111, 1100, 11011, 11110, 111111, \ldots \}$

正则式的性质

交换律（commutativity）和结合律 （associativeity）

(1) $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$

(2) $(\alpha\ \beta)\gamma = \alpha(\beta\ \gamma)$

(3) $\alpha+\beta = \beta+\alpha$

等幂律 （idempotent law）

(4) $\alpha+\alpha = \alpha$

分配律 （distributive law）

(5) $\alpha(\beta+\gamma) = \alpha\ \beta+\alpha\ \gamma$

(6) $(\beta+\gamma)\alpha = \beta\ \alpha+\gamma\ \alpha$

正则式的性质

幺元（identities）和零元（annihilators）

(7) $\alpha + \phi = \phi + \alpha = \alpha$

(8) $\alpha\phi = \phi\alpha = \phi$

(9) $\alpha\varepsilon = \varepsilon\alpha = \alpha$

与闭包相关的定律

(10) $(\alpha^*)^* = \alpha^*$

(11) $\alpha^* = \alpha + \alpha^*$

$\phi^* = \varepsilon$

$\varepsilon^* = \varepsilon$

$L^+ = LL^* = L^*L$ （ L^+的定义）

$L^* = L^+ + \varepsilon$

正则式的性质

(11) $\alpha^* = \alpha + \alpha^*$

证明:

$\because\ \alpha^* = \varepsilon + \alpha + \alpha^2 + \ldots + \alpha^n$

$L(\alpha^*) = \{\varepsilon\} \cup L(\alpha) \cup L(\alpha^2) \cup \ldots \cup L(\alpha^n)$

$L(\alpha + \alpha^*) = L(\alpha) \cup (\{\varepsilon\} \cup L(\alpha) \cup L(\alpha^2) \cup \ldots \cup L(\alpha^n))$

$= L(\alpha) \cup L(\alpha^*)$

而 $L(\alpha) \subseteq L(\alpha^*)$

$\therefore\ \alpha + \alpha^* = \alpha^*$

右线性文法与正则式

右(左)线性文法又称为正则文法，右线性文法与正则式可以用来代表同一正则语言。二者具有等效性。

例： 文法 $S \rightarrow aS$，$S \rightarrow a$

对应正则式：a^+，或者a^*a

从右线性文法导出正则式

求解规则:

设$x \to \alpha x+\beta$， $\alpha \in T^*$， $\beta \in (N+T)^*$, $x \in N$

则x的解为 $x = \alpha^* \beta$

证明:

$x \to \alpha x+\beta$ 表示x有两个生成式: $x \to \alpha x$ 和 $x \to \beta$，生成的语言为（β，αβ，ααβ，αααβ, …），显然该语言可用正则式$\alpha^*\beta$表示。

书p56， 例2

书p57， 例3

例 2 设右线性文法 $G=(\{S,A,B\},\{a,b\},P,S)$

生成式 P 如下：

$$S\rightarrow aA, S\rightarrow bB, S\rightarrow b$$

$$A\rightarrow bA, A\rightarrow \varepsilon$$

$$B\rightarrow bS$$

以上生成式写成联立方程为

$$S=aA+bB+b$$

$$A=bA+\varepsilon$$

$$B=bS$$

对式(2)利用规则 R 和性质 $\alpha\cdot\varepsilon=\alpha$，得

$$A=b^*$$

将式(4)和式(3)代入式(1)中的 A、B，得

$$S=ab^*+bbS+b=bbS+ab^*+b$$

$$=(bb)^*(ab^*+b)$$

所以由 G 产生的语言，用正则式表示为 $(bb)^*(ab^*+b)$。

例 3 设右线性文法 $G=(\{A,S\},\{a\},P,S)$ 生成式 P 如下：

$$S\rightarrow a, S\rightarrow aA, A\rightarrow aS$$

以上生成式写成联立方程为

$$S=aA+a$$

$$A=aS$$

将式(2)代入式(1)中的 A，得

$$S=aaS+a=(aa)^*a$$

表示的正则集是任意一对(包括 0 对)aa 后跟一个 a，即 $\{a^{2n+1}\mid n\geqslant 0\}$。

第六节 正则集和右线性文法

- 正则集是由右线性文法产生的语言
- 由右线性文法产生的语言都是正则集

(一). 求证正则集是由右线性文法产生的语言

证明方法：

找出相应的右线性文法，使之产生的语言是这些正则集。

1. 首先从最简单的正则式出发，求证其正则集Φ，{ε}，{a}是右线性语言。

证明:

对正则集Φ,

有右线性文法G＝{{S}，T， Φ，S}， 使L(G)= Φ

对正则集{ε},

有右线性文法G＝{{S}，T，{S->ε}，S}，使L(G)= {ε}

对正则集{a},

有右线性文法G＝{{S}，T，{S->a}，S}，使L(G)= {a}

2. 将对由并，积，闭包形成的正则集的证明，改为对右线性语言的证明。

设在字母表T上，有语言L_1和L_2，则$L_1 \cup L_2$，$L_1.L_2$，L_1^*都是右线性语言。

证明方法：分别找出相应的右线性文法。

设 G1＝（N_1，T_1，P_1，S_1）产生L_1

G_2＝（N_2，T_2，P_2，S_2）产生L_2

$N_1 \cap N_2 = \Phi$ (若不为空, 则可对N中符号换名)

(a). 求证$L_1 \cup L_2$是右线性语言

构造G＝（N，T，P，S）产生L，使L= $L_1 \cup L_2$

其中　　$N = N_1 \cup N_2 \cup \{S\}$，S为新的非终结符；

$P = P_1 \cup P_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$

$T = T_1 \cup T_2$

先证$L \subseteq L_1 \cup L_2$：

在G中，由G的定义，对于任意，意味着或者（按G1的产生式），或者（按G_2的产生式）

即文法G的每个句子或由G_1产生，或由G_2产生。

$\therefore L(G) \subseteq L(G_1) \cup L(G_2)$

再证 $L_1 \cup L_2 \subseteq L$：

设有$\omega \in L1 \cup L2$，则存在推导

$S_1 \underset{G1}{\Rightarrow^+} \omega$ 或 $S_2 \underset{G2}{\Rightarrow^+} \omega$

$\therefore$ 必有$S \underset{G}{\Rightarrow^+} \omega$。即$L_1 \cup L_2 \subseteq L$。　　命题得证#

(b). 求证L_1L_2是右线性语言

构造G＝（N，T，P，S）

其中$N＝N_1 \cup N_2$，$S＝S_1$，

生成式P为: 若$A \to \alpha B \in P_1$，则它也属于P

若$A \to \alpha \in P_1$，则$A \to \alpha S_2 \in P$

$P_2 \subseteq P$

先证 $L(G_1).L(G_2) \subseteq L(G)$

若有任意$S_1 \Rightarrow^* \omega_1$ 与 $S_2 \Rightarrow^* \omega_2$ 分别属于G_1和G_2定义的语言中，

那么有 $S_1 \Rightarrow \alpha_1 A \Rightarrow \alpha_2 B \Rightarrow \alpha_3 C \Rightarrow \ldots \Rightarrow \omega_1$ 和

$S_2 \Rightarrow \beta_1 A_1 \Rightarrow \beta_2 B_2 \Rightarrow \beta_3 C_3 \Rightarrow \ldots \Rightarrow \omega_2$，

$\therefore$ $S \Rightarrow S_1 \Rightarrow \alpha_1 A \Rightarrow \alpha_2 B \Rightarrow \alpha_3 C \Rightarrow \ldots \Rightarrow \omega_1.S_2 \Rightarrow^* \omega_1.\omega_2$

$\therefore$ $L(G_1).L(G_2) \subseteq L(G)$

(b). 求证$L_1.L_2$是右线性语言

再证 $L(G) \subseteq L(G_1).L(G_2)$

若有$S \Rightarrow S_1 \Rightarrow \alpha_1 A \Rightarrow \alpha_2 B \Rightarrow \alpha_3 C \Rightarrow \ldots$

$$\Rightarrow \omega_1.S_2 \Rightarrow^* \omega_1.\omega_2$$

那么则必然在G_1和G_2中同时有

$S_1 \Rightarrow \alpha_1 A \Rightarrow \alpha_2 B \Rightarrow \alpha_3 C \Rightarrow \ldots \Rightarrow \omega_1$ 和

$S_2 \Rightarrow \beta_1 A_1 \Rightarrow \beta_2 B_2 \Rightarrow \beta_3 C_3 \Rightarrow \ldots \Rightarrow \omega_2$

$\therefore\ L(G) \subseteq L(G_1).L(G_2)$

命题得证#

(c). 求证L_1*是右线性语言

证明：构造G＝（N，T，P，S）

其中；$N = N_1 \cup S$，$S \notin N_1$，S是一个新的非终结符，

生成式P为：如果$A \to \alpha B \in P_1$，则$A \to \alpha B \in P$。

如果$A \to \alpha \in P_1$，则$A \to \alpha S \in P$

$S \to S_1$，$S \to \varepsilon \in P$。

先证 $L(G) \subseteq L(G1)^*$

若有$S \Rightarrow S_1 \Rightarrow \omega_1 S \Rightarrow^* \omega_1 \omega_2 S \Rightarrow \ldots$

$\Rightarrow^* \omega_1 \omega_2 \ldots \omega_i.S \Rightarrow \omega_1 \omega_2 \ldots \omega_i$

则在G_1中必然有

$S_1 \Rightarrow^* \omega_1$；$S_1 \Rightarrow^* \omega_2$；。。。；$S_1 \Rightarrow^* \omega_i$

$\therefore L(G) \subseteq L(G_1)^*$

(c). 求证L_1*是右线性语言(续)

再证 $L(G_1)^* \subseteq L(G)$

若G_1中有

$S_1 \Rightarrow^* \omega_1$ ； $S_1 \Rightarrow^* \omega_2$ ； 。。。；$S_1 \Rightarrow^* \omega_i$

则在G中必然有

$S \Rightarrow S_1 \Rightarrow \omega_1 S \Rightarrow^* \omega_1 \omega_2 S \Rightarrow \ldots$

$\Rightarrow^* \omega_1 \omega_2 \ldots \omega_i.S \Rightarrow \omega_1 \omega_2 \ldots \omega_i$

$\therefore$ $L(G_1)^* \subseteq L(G)$

因此 L* 可以用右线性文法描述。

证毕#

(二). 证明由右线性文法产生的语言是正则集

思路：

由给定的右线性文法可找出正则式，而正则式表示的语言即为正则集。

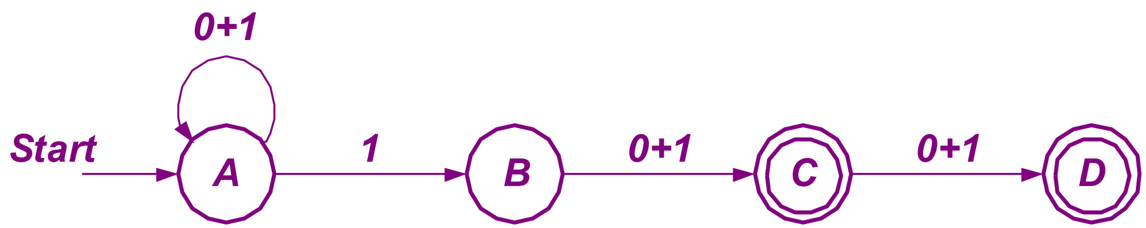
方法：

对右线性文法构造标准形式的正规表达式方程组，应用规则 $x=\alpha x+\beta \Rightarrow x=\alpha^*\beta$ 进行消元，求解方程组，即可得出正规表达式。

由 (一) 和 (二)即可得出下定理:

定理: 一个语言是正则集，当且仅当该语言为右线性语言。

课堂练习



- Exercise 对于以上自动机，已经求出其语言的一个正规表达式如下

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

使用分配律求出两种不同的更简单的与之等价的表达式.

- 参考结果

（1） $(0+1)^*1(\varepsilon+0+1)(0+1)$

（2） $(0+1)^*1(0+1)(\varepsilon+0+1)$

课后练习

chap3 习题

习题15，1，4，5