复习课

- 第二章 语言及文法
- 第三章 有限自动机和右线性文法

第二章 语言及文法

- 语言的定义
  - 字母表 字符串（句子）
  - 连接运算
  - T^*，T^+
  - 语言，语言的积，语言的幂，L^*，L^+
- 文法
  - 文法的形式定义
  - 推导，句型，句子
- 文法的分类
  - 0 型，1 型，2 型和 3 型

文法的分类

- **0型文法(无限制文法)**
- **1型文法(上下文有关文法):**
  - 生成式: $\alpha\rightarrow\beta$，　$|\alpha|\le|\beta|$，

    且 $\beta\in（N\cup T)^{+}$，　$\alpha\in（N\cup T)^{*} N^{+}（N\cup T)^{*}$
- **2型文法(上下文无关文法):**
  - $A\rightarrow\alpha$, $A\in N$, 且 $\alpha\in(N\cup T)^{*}$
- **3型文法(正则文法):**
  - 右线性文法：$A\rightarrow\omega B$ 或 $A\rightarrow\omega$，　A、$B\in N$, $\omega\in T^{*}$
  - 左线性文法：$A\rightarrow B\omega$ 或 $A\rightarrow\omega$，　A、$B\in N$, $\omega\in T^{*}$
  - 对应的语言：　正则语言
  - 对应的自动机：有限自动机

例题

例 3 设 $T=\{0,1\}$，请给出 T 上下列语言的文法：

（1）$L=\{0^{3m}1^{2m} \mid m\geqslant 1\}$；

（2）所有以 1 开头，以 0 结尾的串；

（3）$L=\{\omega \mid \omega=\tilde{\omega}, \omega\in\{0,1\}^{+}\}$。

(1) $G=(N,T,P,S)$，其中 $N=\{S\}$，$T=\{0,1\}$，P：$S\rightarrow 000S11 \mid 00011$。

(2) $G=(N,T,P,S)$，其中 $N=\{S,A\}$，$T=\{0,1\}$，P：$S\rightarrow 1A0$，$A\rightarrow 0A \mid 1A \mid \varepsilon$

或 $G=(N,T,P,S)$，其中 $N=\{S,A\}$，$T=\{0,1\}$，P：$S\rightarrow 1A$，$A\rightarrow 0A \mid 1A \mid 0$。

(3) $G=(N,T,P,S)$，其中 $N=\{S\}$，$T=\{0,1\}$，P：$S\rightarrow 0S0 \mid 1S1 \mid 0 \mid 1 \mid 00 \mid 11$。

例题

例 4 找出由下列各组生成式产生的语言(起始符为 S)：

(1) $S \to SaS, S \to b$；

(2) $S \to aSb, S \to c$；

(3) $S \to a, S \to aE, E \to aS$。

(1) $\{b(ab)^n \mid n \geq 0\}$ 或 $L = \{(ba)^n b \mid n \geq 0\}$；

(2) $L = \{a^n c b^n \mid n \geq 0\}$；

(3) $L = \{a^{2n+1} \mid n \geq 0\}$。

练习1

- 设文法 G ＝ ({S， A}, {a, b}, P, S) 的产生式P为：
  S →ε , S → aS, S → bA, A → bA, A →ε
  说明其产生的语言形式。
- 分析
  - 文法为3型文法
  - 由产生式可得联立方程 $\begin{cases} S = aS + bA + \varepsilon \\ A = bA + \varepsilon \end{cases}$
  - 由此可得A=b*； S=aS+bb*+ ε = a*b* (不唯一)
  - 其产生的语言为$L=\{a^m b^n | m,n \geq 0\}$

练习2

指出由下列各组生成式产生的语言形式

L： $S \rightarrow aB \mid bA$

$A \rightarrow a \mid bAA \mid aS$

$B \rightarrow b \mid aBB \mid bS$

练习3

- 语言L中的每个串由零个或多个0, 紧跟一个或多个1, 然后再紧跟两个或多个2构成。试定义该语言的一个上下文无关文法。

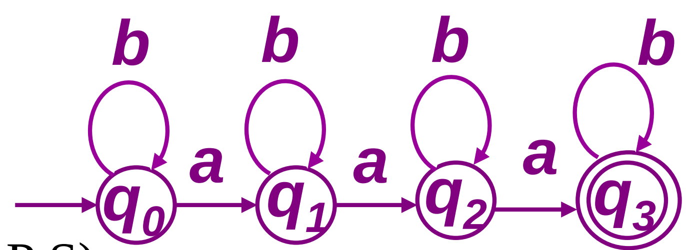
- 分析
  - 语言L对应的正则式为： $0^*11^*222^*$
  - 可得0^*; 11^*; 222^*对应的产生式 $\begin{cases} A \to 0A \mid \varepsilon \\ B \to 1B \mid 1 \\ C \to 2C \mid 22 \end{cases}$
  - 由此可得上下文无关文法 G=({S,A,B,C},{0,1,2},P S)

P: $S \to ABC$; $A \to 0A|\varepsilon$; $B \to 1B|1$; $C \to 2C|22$

练习4

- **给出T={a,b} 上能满足下列条件的语言的文法：**
  - 至少有3个a的所有符号串
  - a的个数不多于3的所有符号串
- 分析
  - 至少有3个a的所有符号串
    - 对应的正则式$(a+b)^*a(a+b)^*a(a+b)^*a\ (a+b)^*$
    - 由此可得上下文无关文法 G=({S,A},{a,b},P S)

    P: $S \rightarrow AaAaAaA$;　$A \rightarrow aA|\ bA|\varepsilon$



其线性文法？

  - a的个数不多于3的所有符号串
    - 对应的正则式 $b^*+b^*ab^* + b^*ab^*ab^* + b^*ab^*ab^*ab^*$
    - 由此可得上下文无关文法 G=({S,A},{a,b},P S)

    P: $S \rightarrow A|AaA|AaAaA\ |\ AaAaAaA$;　$A \rightarrow bA|\varepsilon$

练习5

给出T={a} 上能满足下列条件的语言的文法：

(a) L={w | |w| mod 3 =0 }

G=({S},{a},P S) P: S→aaaS|ε

(b) L={w | |w| mod 3 >0 }

G=({S,A,B},{a},P S)

P: S → A|B; A → aaaA|*a*; B → aaaB|*aa* (可简化)

(c) L={w | |w| mod 3 <> |w| mod 2 }

|w| mod 6 = 2，3，4，5

G=({S},{a},P S) P: S → aaaaaaS|*aa*|*aaa* |*aaaa* |*aaaaa*

练习6

设计下面语言的文法:

（1） $L1=\{a^m b^n \mid 0 < m \leq n \leq 3m \}$

（2） $L2=\{ 0^n 1^{n+k} 0^k \mid n、k \geq 0 \}$

第三章 有限自动机和右线性文法

- 有限自动机定义，识别的语言，格局
- 不确定的有限自动机NFA
- DFA与NFA等效
- 有 ε转换的NFA
- 有ε转换的NFA和无ε转换的NFA等效
- 正则集与正则式
- 右线性文法与正则集
- 正则表达式和有限自动机
- 右线性语言与有限自动机
- 右线性语言的性质　　　DFA的化简，泵浦引理
- 双向和有输出的有限自动机　2 DFA，米兰机，摩尔机

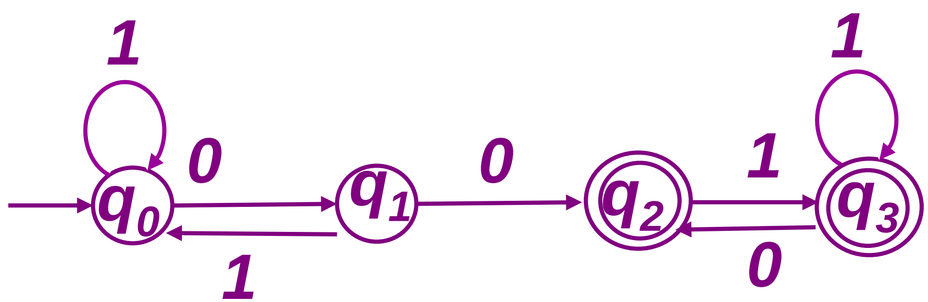
第三章 有限自动机和右线性文法

- **有限状态系统**
  - 可以认为有限状态系统定义了一个无穷的语言，这个语言是由那些能使状态从初始状态经过任意可能的路径到达终止状态的字符串组成。
- **有限自动机(DFA / NFA):**
  - 五元组 M＝（Q，T，δ，q0，F）
  - DFA: δ: Q×T --> Q ; NFA: δ: Q×T --> 2^Q
  - 格局: 表现自动机所处工作状态 (q，ω)
  - 双向自动机的格局: $(\omega_1 q \omega_2)$
  - 自动机接受的语言：
    - DFA L(M)＝{ω|δ(q0,ω) ∈F}; NFA L(M)＝{ω|δ(q0,ω) 含F的一个状态}
  - 接受一个字符的状态转换函数 -- δ
  - 接受一个字符串的状态转换函数 -- δ’
  - δ’ (q,ωa)＝δ（δ’(q,ω), a）
  - NFA 到DFA的转换—(子集构造法)
  - (实践中最好方法是从q0出发)

练习1

构造一个自动机，使它接受下面的字符集

R＝{ω|ω∈{0，1}*，ω恰好包含一个“00”而不含“000”序列}



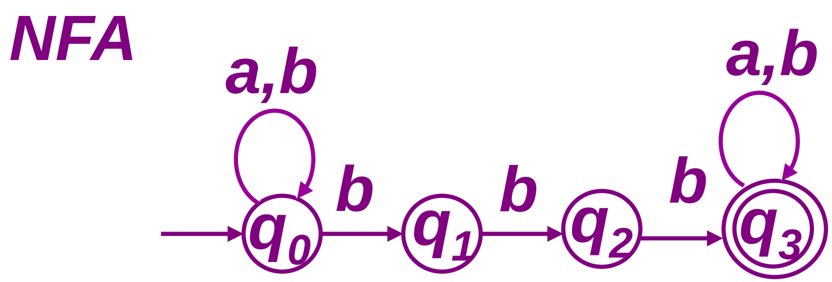
注意图的画法:

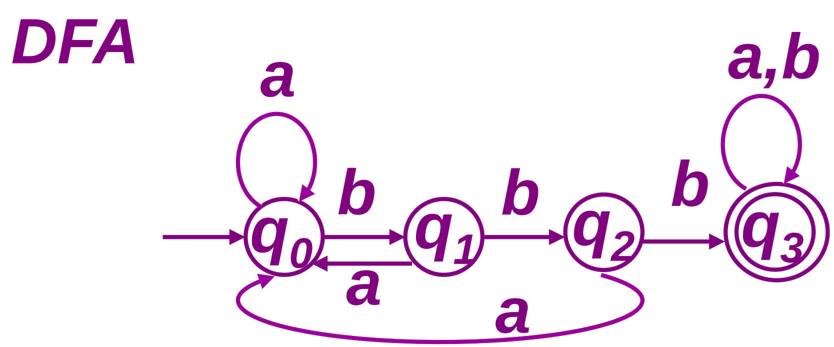
一个起始(必须有起始箭头),必须有终止(双圈). 异常情况可以不画．如 遇到输入”2”之类.

练习2

- **设字母表T={a, b}，对语言**G = 含有3个连续b的所有字符串的集合,分别写出其NFA和 DFA .

G对应的正则式 $(a+b)^*bbb(a+b)^*$





练习3

- **T={a，b}，L中的任意三个连续符号中最多含2个a，设计语言L的自动机.**

- <u>ε转换</u>:当输入空串ε(无输入)时,也能引起状态的转移.
- <u>ε闭包</u>:
  - 状态q的ε-闭包：ε-CLOSURE(q) -- 表示从q出发可用ε转换到达的所有状态的集合
  - 状态子集I 的ε-闭包：$\varepsilon\text{-CLOSURE}(I) = \bigcup_{q \in I} \varepsilon\text{-CLOSURE}(q)$
  - <u>有ε转换的NFA中，δ与δ' 函数的区别:</u>
  - δ' 是接受字符串的状态转移函数：δ' ＝$Q \times T^* \to 2^Q$
  - δ' (q，ω)＝ {p1，p2，… pn} （ω路径中含有标ε的边）
  - （1）δ' (q，ε)＝ε-CLOSURE(q)
  - （2）δ' (q，ωa)＝ε-CLOSURE(P)
  - 其中P＝{ p | 存在r∈δ'(q，ω) ∧ p∈δ(r，a)}
  - 注意：此时δ(q，a) ≠δ'(q，a)，因为δ(q，a)表示由q出发，只沿着标a的路径所能到达的状态，而δ'(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.

- **有ε转换的NFA所能接受的语言:**
  - L（M）＝{ω | ∃p（p∈δ'(q0，ω) ∧ p∈F）}
  - （即　　　满足δ' (q0，ω) 含有F的一个状态）
- **<u>有ε转换的NFA 到无ε转换的NFA 的等效:</u>**
  - 设具有ε转移的NFA： M＝（Q，T，δ，q0，F）
  - 可构造不具有ε转移的NFA：M1＝（Q，T，δ1，q0，F1），
  - 其中
  - F1 ＝ ⎧ F∪{q0}　若ε-CLOSURE(q0) 含F中一个状态
　　　⎩　**F**　　　否则
  - δ1（q，a）＝δ’（q，a）

正则集与正则式

- 字母表上一些特殊的字符串的集合称为正则集。
- 表示正则集可用正则式。正则集是正则式所表示的集合。
- 关键：用类似代数表达式的方法来表示正则集。
- 定义:
  - 字母表T上的一个正则式及其所代表的正规集合可以递归地定义如下:
  - 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}
  - 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，分别表示正则集 (语言)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。

正则式的性质

- 若两个正则式表示相同的正则集，则称两个正则式相等。
- 正则集是T* 的子集。
- 定义 $L^0=\varepsilon$， $L^1=L$， $L^i=L.L^{i-1}$
- L^+包含ε当且仅当L包含ε。
- 每个正则集至少对应一个正则式（可有无穷多个正则式）；每个正则式只对应一个正则集。

由正则文法求解正则式

- 右线性文法与正则式具有等效性。即可以用来代表同一正则语言。
- 求解规则:
  - 设$x=\alpha x+\beta$，$\alpha\in T^*$，$\beta\in(N\cup T)^*$，$x\in N$
    (更严格地说，应为α和β为正则表达式）
  - 则x的解为 $x=\alpha^*\beta$

练习4

设正则文法G的产生式为:

$S \rightarrow 0A \mid 1S \mid \varepsilon;\quad A \rightarrow 0B \mid 1A;\quad B \rightarrow 0S \mid 1B$

求解其正则式.

联立方程:

$$\mathbf{S = 1S + 0A + \varepsilon}$$
$$\mathbf{A = 1A + 0B}$$
$$\mathbf{B = 0S + 1B}$$

$$\mathbf{B = 1^*0S}$$
$$\mathbf{A = 1^*0B = 1^*01^*0S}$$
$$\mathbf{S = 1S + 01^*01^*0S + \varepsilon}$$
$$S = (1 + 01^*01^*0)^*$$

右线性文法与正则集的等价

- **一个右线性文法表示的语言可以用正则式来表示**
  - (求解联立方程)
- **一个用正则式表示的语言可以用右线性文法来表示**
- **右线性文法与正则式两者等价，进而它们与FA等价**

- **FSM的局限性:**
  - 必须是有限个状态 (但其语言可以是无限的).

- 正则表达式和有限自动机
  - 已知正则式，构造 ε -NFA 或 NFA
  - 已知NFA / DFA，构造等价正则式
- 右线性语言与有限自动机
  - 已知右线性文法，构造NFA
  - 已知NFA / DFA，构造右线性文法

- **右线性语言的性质**
  - DFA的化简
    - 最小化算法
    - 填表法
  - 泵浦引理
    - 用于证明某个语言 L 不是正规语言

练习5

- **证明$L=\{\omega\varpi|\omega\in\{a,b\}^*\}$不是3型语言。**

证明 假设L是正则集，取足够大的整数k，取$\omega = a^k b^k b^k a^k \in L$

有$|\omega| = |\omega_1\omega_0\omega_2| = 4k \geq k$，

其中 $0<|\omega_0|\leq k$， $0<|\omega_1\omega_0|\leq k$

$\because$ $0<|\omega_1\omega_0|\leq k$ $\therefore$ 限制了ω_0只能处于最左边的a^k段

设$\omega_0=a^m$,$0<m\leq k$,

当取 $i=0$时， $\omega_1\omega_0^0\omega_2 = a^{k-m}b^k b^k a^k$

$\therefore$ 有 $\omega_1\omega_0^0\omega_2 \notin L$ (a的个数不同，不是回文)

与假设矛盾。

$\therefore$ L不是正则集。

- 双向和有输出的有限自动机
  - 2 DFA
  - 米兰机
  - 摩尔机
- 练习6：构造摩尔机，输入是十进制整数的二进制形式，输出为输入的模5余数。要求给出简单分析并画出状态转换图.

习题难点讲解： ch3 习题1， 5（4）；17（1）

1. 下列集合是否为正则集，若是正则集写出其正则式：

（1）含有奇数个 0 和偶数个 1 的$\{0,1\}^*$上的字符串集合；

（2）含有 a 的个数是 b 的个数的 2 倍的 a 和 b 的字符串集合；

1.（1） 参考书 P87 例3

1.（2） 应用泵浦引理， 可取$w=b^na^{2n}$， 或取$w=a^nb^na^n$ 亦可

5. 为下列正则集构造右线性文法：

（4）含有两个相继 a 或两个相继 b 的由 a 和 b 组成的所有字符串集合。

5.（4） 参考书 P88 例5 注意那里是“and”

17. 使用泵浦引理，证明下列集合不是正则集：

（1）由文法 G 的生成式 $S\rightarrow aSbS \mid c$产生的语言 $L(G)$；

17.（1） 可推导出该句子的一般形式是 $a^n S (bS)^n$

所以可取句子 $w= a^n c (bc)^n$ 应用泵浦引理