

第五讲

正则文法与正则语言

2022/3/22

School of Software

1

正则文法和正则语言

- 文法
- 线性文法
- 正则文法和正则语言
- 自动机的积

2022/3/22

School of Software

2

正则文法和正则语言

- 文法
- 线性文法
- 正则文法和正则语言
- 自动机的积

2022/3/22

School of Software

启示

- 文法的概念最早是由语言学家们在研究自然语言中形成。
- 归纳如下句子的描述：
 - (1) 形式语言是字符串的集合。
 - (2) 学生查阅资料。
 - (3) 软件的测试是必要的。
 - (4) 科学推动社会的发展。

2022/3/22

School of Software

3

启示

- 文法的概念最早是由语言学家们在研究自然语言中形成。
- 归纳如下句子的描述：
 - (1) <形式语言><是字符串的集合><。>
 - (2) <学生><查阅资料><。>
 - (3) <软件的测试><是必要的><。>
 - (4) <科学><推动社会的发展><。>

2022/3/22

School of Software

4

启示

4个句子的主体结构:

<名词短语> <动词短语> <句号>
 <名词短语> = {形式语言, 学生, 软件的测试, 科学, 字符串的集合, 资料, 社会的发展}
 <动词短语> = {是字符串的集合, 查阅资料, 是必要的, 推动社会的发展}
 <句号> = {。}

2022/3/22

School of Software

5

启示

<动词短语>可以是<动词><形容词短语>
或者<动词><名词短语>。

<动词>={是, 查阅, 推动}。

<形容词短语>={必要的}。

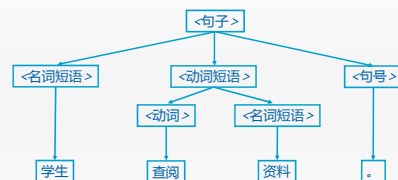
<名词短语>={形式语言, 学生, 软件的测试, 科学, 字符串的集合, 资料, 社会的发展}。

<名词短语><动词短语><句号> 取名为<句子>。

2022/3/22

School of Software

启示



2022/3/22

School of Software

启示

表示成 $\alpha \rightarrow \beta$ 形式

<句子> \rightarrow <名词短语><动词短语><句号>

<动词短语> \rightarrow <动词><形容词短语>

<动词短语> \rightarrow <动词><名词短语>

<名词短语> \rightarrow 学生

<动词> \rightarrow 查阅

<名词短语> \rightarrow 资料

<句号> \rightarrow 。

2022/3/22

School of Software

英文文法

英语语言文法：

$\langle sentence \rangle \rightarrow \langle noun_phrase \rangle \langle predicate \rangle$

$\langle noun_phrase \rangle \rightarrow \langle article \rangle \langle noun \rangle$

$\langle predicate \rangle \rightarrow \langle verb \rangle$

2022/3/22

School of Software

英文文法

英语语言文法：

$\langle article \rangle \rightarrow a$

$\langle article \rangle \rightarrow the$

$\langle noun \rangle \rightarrow cat$

$\langle noun \rangle \rightarrow dog$

$\langle verb \rangle \rightarrow runs$

$\langle verb \rangle \rightarrow walks$

2022/3/22

School of Software

英文文法

“the dog walks” 的推导

$\langle sentence \rangle \Rightarrow \langle noun_phrase \rangle \langle predicate \rangle$

$\Rightarrow \langle noun_phrase \rangle \langle verb \rangle$

$\Rightarrow \langle article \rangle \langle noun \rangle \langle verb \rangle$

$\Rightarrow the \langle noun \rangle \langle verb \rangle$

$\Rightarrow the \text{ dog } \langle verb \rangle$

$\Rightarrow the \text{ dog walks}$

2022/3/22

School of Software

英文文法

“a cat runs” 的推导

$$\begin{aligned}
 \langle sentence \rangle &\Rightarrow \langle noun_phrase \rangle \langle predicate \rangle \\
 &\Rightarrow \langle noun_phrase \rangle \langle verb \rangle \\
 &\Rightarrow \langle article \rangle \langle noun \rangle \langle verb \rangle \\
 &\Rightarrow a \langle noun \rangle \langle verb \rangle \\
 &\Rightarrow a \text{ cat } \langle verb \rangle \\
 &\Rightarrow a \text{ cat runs}
 \end{aligned}$$

2022/3/22

School of Software

15

英文文法

英语语言：

$$L = \{ \begin{aligned} &\text{"a cat runs"}, \\ &\text{"a cat walks"}, \\ &\text{"the cat runs"}, \\ &\text{"the cat walks"}, \\ &\text{"a dog runs"}, \\ &\text{"a dog walks"}, \\ &\text{"the dog runs"}, \\ &\text{"the dog walks"} \end{aligned} \}$$

2022/3/22

School of Software

16

文法

产生规则

$$\left\{ \begin{aligned} &\langle noun \rangle \rightarrow cat \\ &\langle noun \rangle \rightarrow dog \end{aligned} \right.$$

\uparrow \uparrow
 变量 终结符

2022/3/22

School of Software

17

推导

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

句子 ab 的推导
$$\begin{aligned}
 S &\Rightarrow aSb \Rightarrow ab \\
 S &\rightarrow aSb \quad S \rightarrow \varepsilon
 \end{aligned}$$

2022/3/22

School of Software

18

另一推导

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

句子 $aabb$ 的推导
$$\begin{aligned}
 S &\Rightarrow aSb \Rightarrow aaSbb \Rightarrow aabb \\
 S &\rightarrow aSb \quad S \rightarrow \varepsilon
 \end{aligned}$$

2022/3/22

School of Software

19

其它推导

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

$$\begin{aligned}
 S &\Rightarrow aSb \Rightarrow aaSbb \\
 &\Rightarrow aaaSbbb \Rightarrow aaabbbb \\
 S &\Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \\
 &\Rightarrow aaaaSbbbb \Rightarrow aaabbbbbb
 \end{aligned}$$

2022/3/22

School of Software

20

文法的语言

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

$$L = \{a^n b^n | n \geq 0\}$$

2022/3/22

School of Software

15

文法的定义

文法 $G = (V, T, S, P)$

V : 变量的集合

T : 终结符的集合

S : 开始变量

P : 产生式的集合

满足
 $V \cap T = \emptyset$
 $S \in V$

产生式 P :

$A \rightarrow \alpha$
 $A \in V, \alpha \in (V \cup T)^*$

2022/3/22

School of Software

16

例

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

$$G = (V, T, S, P)$$

$$V = \{S\} \quad T = \{a, b\} \quad P = \{S \rightarrow aSb, S \rightarrow \varepsilon\}$$

2022/3/22

School of Software

17

句型

由变量和终结符构成的字符串称为句型；

由终结符构成的字符串称为句子。

例: $S \rightarrow aSb$

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaSbbb \Rightarrow aaabbbb$$

句型

句子

2022/3/22

School of Software

18

推导闭包

文法: $S \rightarrow aSb$
 $S \rightarrow \varepsilon$

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaSbbb \Rightarrow aaabbbb$$

只强调最后结果, 不关注中间推导, 记为:

$$S \stackrel{*}{\Rightarrow} aaabbbb$$

2022/3/22

School of Software

19

推导闭包

一般的情形, 如果有:

$$w_1 \Rightarrow w_2 \Rightarrow w_3 \Rightarrow \cdots \Rightarrow w_n$$

则:

$$w_1 \stackrel{*}{\Rightarrow} w_n$$

规定:

$$w \stackrel{*}{\Rightarrow} w$$

可以通过
归纳法定义

2022/3/22

School of Software

20

例

文法G:

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

推导:

$$S \Rightarrow^* \varepsilon$$

$$S \Rightarrow^* ab$$

$$S \Rightarrow^* aabb$$

$$S \Rightarrow^* aaabbb$$

例

文法G:

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

推导:

$$S \Rightarrow^* aaSbb$$

$$aaSbb \Rightarrow^* aaaaaaSbbbbbb$$

符号

$$A \rightarrow aAb$$

$$A \rightarrow \varepsilon$$

$$\longrightarrow A \rightarrow aAb \mid \varepsilon$$

$$\alpha \rightarrow \beta$$

$$\alpha \rightarrow \gamma$$

$$\longrightarrow \alpha \rightarrow \beta \mid \gamma$$

另一例

文法G:

$$S \rightarrow Ab$$

$$A \rightarrow aAb \mid \varepsilon$$

推导:

$$S \Rightarrow Ab \Rightarrow b$$

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow abb$$

$$S \Rightarrow Ab \Rightarrow aAbb$$

$$\Rightarrow aaAbbb \Rightarrow aabbb$$

另一例

文法G:

$$S \rightarrow Ab$$

$$A \rightarrow aAb \mid \varepsilon$$

推导:

$$S \Rightarrow Ab \Rightarrow aAbb \Rightarrow aaAbbb \Rightarrow aaaAbbbb$$

$$\Rightarrow aaaaAbbbbbb \Rightarrow aaaabbbbbb$$

$$S \Rightarrow^* aaaabbbbbb$$

$$S \Rightarrow^* aaaaaabbbbbbbb$$

$$S \Rightarrow^* a^n b^n b$$

文法的语言

对于文法 $G=(V, T, S, P)$, 其语言为

$$L(G) = \{w \mid S \xRightarrow{*} w, w \in T^*\}$$

终结符构成的字符串

若文法G1与G2的语言相同, 则称G1和G2等价

例

文法G:

因为:

$$S \rightarrow Ab$$

$$A \rightarrow aAb \mid \varepsilon$$

语言:

$$S \Rightarrow a^n b^n b$$

$$L(G) = \{a^n b^n b \mid n \geq 0\}$$

正则文法和正则语言

- 文法
- 线性文法
- 正则文法和正则语言
- 自动机的积

线性文法

线性文法: 产生式右侧至多有一个变量

例:

$$S \rightarrow aSb$$

$$S \rightarrow Ab$$

$$S \rightarrow \varepsilon$$

$$A \rightarrow aAb$$

$$A \rightarrow \varepsilon$$

线性文法

例文法G:

$$S \rightarrow A$$

$$A \rightarrow aB \mid \varepsilon$$

$$B \rightarrow Ab$$

$$L(G) = \{a^n b^n \mid n \geq 0\}$$

非线性文法例

文法G:

$$S \rightarrow SS$$

$$S \rightarrow \varepsilon$$

$$S \rightarrow aSb$$

$$S \rightarrow bSa$$

$$L(G) = \{w \mid n_a(w) = n_b(w)\}$$

串w中a的个数

右线性文法

右线性文法G: 所有产生式只有两类

$$A \rightarrow xB \text{ 或 } A \rightarrow x$$

终结符串

例:

$$S \rightarrow abS$$

$$S \rightarrow a$$

产生式右侧: 1) 变量最多一个; 2) 变量要有, 则一定在最右边

左线性文法

左线性文法 G : 所有产生式只有两类

$$A \rightarrow Bx \text{ 或 } A \rightarrow x$$

例:
$$\begin{aligned} S &\rightarrow Aab \\ A &\rightarrow Aab \mid B \\ B &\rightarrow a \end{aligned}$$

正则文法

右线性文法或左线性文法称为正则文法 (3型文法)

例: 只能有一个, 不能既有右线性文法又有左

$$\begin{aligned} G_1 : S &\rightarrow abA & G_2 : S &\rightarrow Aa \\ A &\rightarrow baA \mid B & A &\rightarrow Aba \mid B \\ B &\rightarrow a & B &\rightarrow ab \\ L(G_1) &= L(G_2) = L((ab)(ba)^*a) \end{aligned}$$

正则文法和正则语言

- 文法
- 线性文法
- 正则文法和正则语言
- 自动机的积

定理

$$\left\{ \begin{array}{l} \text{正则文法} \\ \text{产生的语言} \end{array} \right\} = \left\{ \begin{array}{l} \text{正则语言} \end{array} \right\}$$

定理

定理第一部分

$$\left\{ \begin{array}{l} \text{正则文法} \\ \text{产生的语言} \end{array} \right\} \subseteq \left\{ \begin{array}{l} \text{正则语言} \end{array} \right\}$$

正则文法产生的语言为正则语言

定理

定理第二部分

$$\left\{ \begin{array}{l} \text{正则文法} \\ \text{产生的语言} \end{array} \right\} \supseteq \left\{ \begin{array}{l} \text{正则语言} \end{array} \right\}$$

正则语言为一正则文法产生的语言

定理证明

证明定理第一部分

$$\left\{ \begin{array}{l} \text{正则文法} \\ \text{产生的语言} \end{array} \right\} \subseteq \left\{ \text{正则语言} \right\}$$

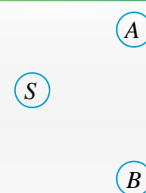
即证正则文法G产生的语言 $L(G)$ 为正则语言

2022/3/22

School of Software

15

举例

 $G:$ $S \rightarrow aA \mid B$ $A \rightarrow aa \mid B$ $B \rightarrow bB \mid a$ 

增加终态

构造NFA M , 其每一个状态都对应一个变量

2022/3/22

School of Software

16

举例

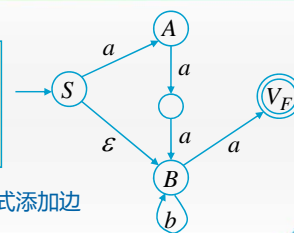
 $G:$ $S \rightarrow aA \mid B$ $A \rightarrow aa \mid B$ $B \rightarrow bB \mid a$

2022/3/22

School of Software

17

举例

 $G:$ $S \rightarrow aA \mid B$ $A \rightarrow aa \mid B$ $B \rightarrow bB \mid a$ 

为每一个产生式添加边

2022/3/22

School of Software

18

定理证明

假设 G 是右线性文法, 证明其对应的语言 $L(G)$ 是正则语言

构造一个NFAM, 使得

$$L(M) = L(G)$$

2022/3/22

School of Software

19

定理证明

设右线性文法 G 的变量为: V_0, V_1, V_2, \dots

产生式:

 $V_i \rightarrow a_1 a_2 \dots a_m V_j$

或

 $V_i \rightarrow a_1 a_2 \dots a_m$

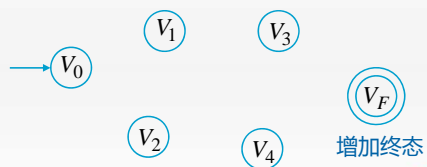
2022/3/22

School of Software

20

定理证明

构造NFA M ，每个变量 V_i 对应其中的状态：



2022/3/22

School of Software

定理证明

对产生式：

$$V_i \rightarrow a_1 a_2 \cdots a_m V_j$$

添加转移和中间状态：



2022/3/22

School of Software

定理证明

对产生式：

$$V_i \rightarrow a_1 a_2 \cdots a_m$$

添加转移和中间状态：

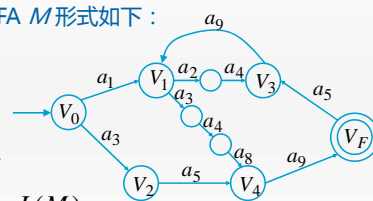


2022/3/22

School of Software

定理证明

因此NFA M 形式如下：



所以有
 $L(G) = L(M)$

2022/3/22

School of Software

定理证明

对左线性文法 G ：

构造一个右线性文法 G' ，使得

$$L(G) = L(G')^R$$

$$L(G') \xrightarrow{\text{正则语言}} L(G')^R \xrightarrow{\text{正则语言}} L(G)$$

2022/3/22

School of Software

定理证明

证明定理第二部分

$$\left\{ \begin{array}{l} \text{正则文法} \\ \text{产生的语言} \end{array} \right\} \supseteq \left\{ \text{正则语言} \right\}$$

即证正则语言 L 为一正则文法 G 产生的语言

2022/3/22

School of Software

定理证明

设 L 为正则语言，则存在NFA M ，使得：

$$L(M) = L$$

通过NFA M 构造正则文法 G ，使得：

$$L(G) = L(M) = L$$

2022/3/22

School of Software

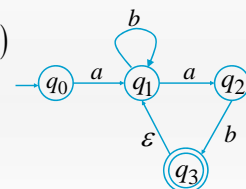
举例

设正则语言 L 为：

$$L = L(ab^*ab(b^*ab)^*)$$

对应的NFA M
如右图：

$$L = L(M)$$

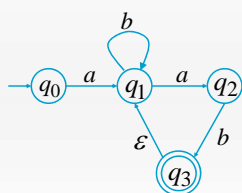


2022/3/22

School of Software

举例

将该NFA M
转为等价的右
线性文法：



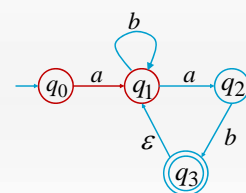
2022/3/22

School of Software

举例

将该NFA M
转为等价的右
线性文法：

$$q_0 \rightarrow aq_1$$



2022/3/22

School of Software

举例

G :

$$q_0 \rightarrow aq_1$$

$$q_1 \rightarrow bq_1$$

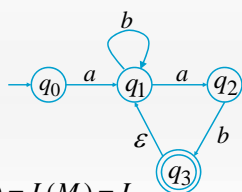
$$q_1 \rightarrow aq_2$$

$$q_2 \rightarrow bq_3$$

$$q_3 \rightarrow q_1$$

$$q_3 \rightarrow \varepsilon$$

$$L(G) = L(M) = L$$

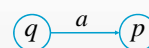


2022/3/22

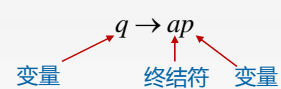
School of Software

一般情形

对状态转移：



添加产生式：



2022/3/22

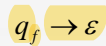
School of Software

一般情形

对状态转移：



添加产生式：

得到与 M 等价的右线性文法 G ，从而有：

$$L(G) = L(M) = L$$

2022/3/22

School of Software

15

正则文法和正则语言

- 文法
- 线性文法
- 正则文法和正则语言
- 自动机的积

2022/3/22

School of Software

16

积自动机

给定DFA

$$A = (Q_A, \Sigma, \delta_A, q_{0A}, F_A)$$

$$B = (Q_B, \Sigma, \delta_B, q_{0B}, F_B)$$

称自动机：

$$M = (Q_A \times Q_B, \Sigma, \delta, (q_{0A}, q_{0B}), F_A \times F_B)$$

为 A 和 B 的积，或称 M 为积自动机，记为 $M = A \times B$ 。其中状态转移函数 δ 为

$$\delta((q_A, q_B), a) = (\delta_A(q_A, a), \delta_B(q_B, a))$$

2022/3/22

School of Software

17

积自动机

例：图1中自动机

$$L_1 = \{\text{所有含0的串}\}$$

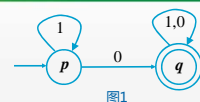


图1

图2中自动机

$$L_2 = \{\text{所有含1的串}\}$$

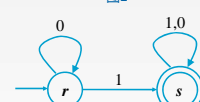


图2

2022/3/22

School of Software

18

积自动机

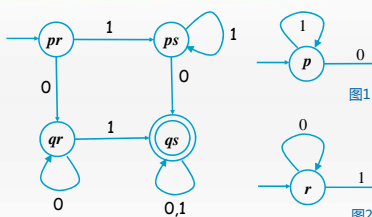


图1

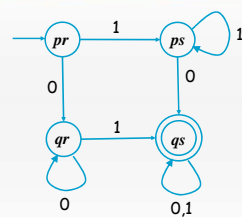
图2

2022/3/22

School of Software

19

积自动机



不难验证

该自动机接受的语言是前两个DFA的交：

{ 所有至少含一个0和一个1的字符串 }

2022/3/22

School of Software

20

练习

1. Construct a DFA that accepts the language generated by the grammar

$$S \rightarrow aaA, A \rightarrow baB, B \rightarrow aA|bb$$



For this language construct a left-linear grammar, too.

2. Find a regular grammar that generates the language
 $L(aa^*(ab+a)^*)$
3. Construct a right- and left-linear grammar for the language:
 $L = \{a^{2n}b^m \mid n \geq 0, m \geq 3\}$
4. L_1 and L_2 are regular languages. Using regular grammar to prove that:
 - (1) $L_1 \cup L_2$ is also a regular language.
 - (2) $L_1 L_2$ is also a regular language.

2022/3/22

School of Software



67

Thank you

2022/3/22

School of Software

68