形式语言与自动机

李舟军　教授

北京航空航天大学计算机学院

lizj@buaa.edu.cn

第二章　文法

2 文法

2.1. 直观背景

自然语言包括词汇和文法两大部分，词汇是构成语言的素材，而文法是组合这些素材的一种可以接受的方法。例如，考虑下面的语句：

1 猫捉老鼠;

2 他给我一本书;

3 这个小孩跑得真快;

4 老鼠偷吃粮食;

都是有意义的正确语句.而下面的语句中：

5 粮食偷吃老鼠;

6 给一本书他我;

7 他送给我一本非常有趣的书;

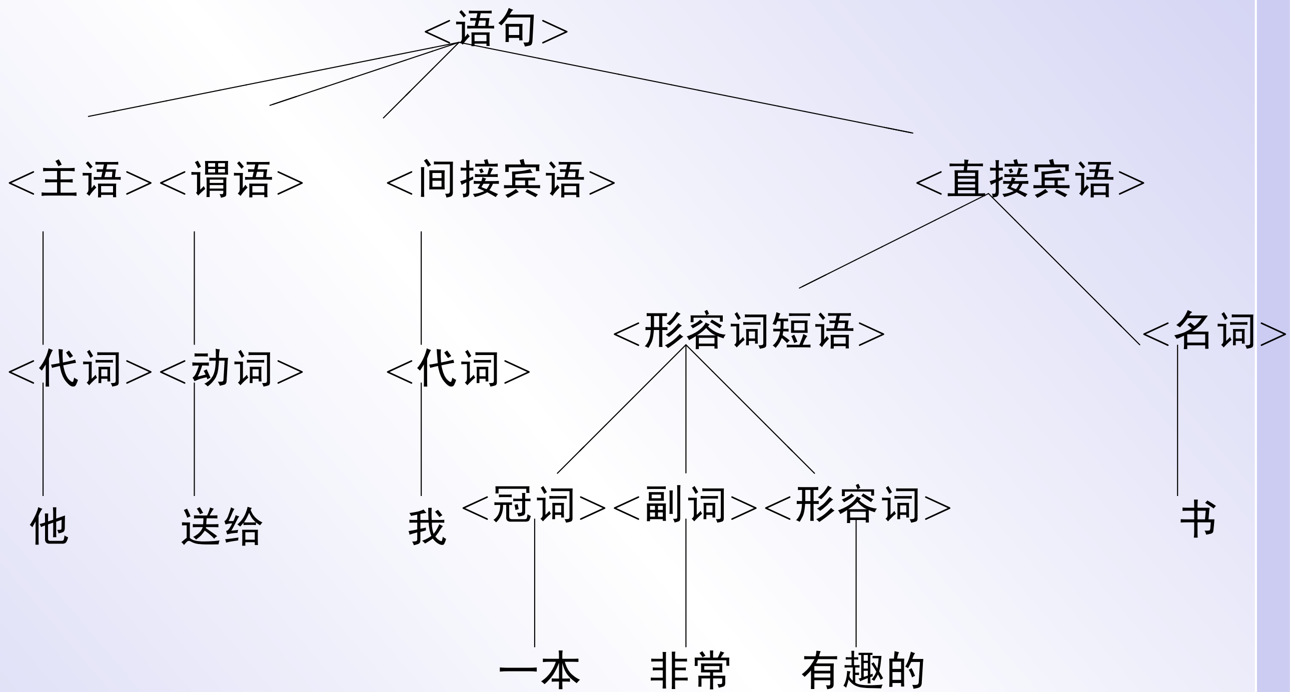
语句5和语句6都不是正确的语句.语句5的语义不正确;语句6的语序混乱;本书不讨论类似于语句5的这类错误,即不考虑语句的语义问题,只考虑如何避免出现语句6这类语法错误的问题.这就是本章引进的形式文法的任务.

当给定一个语句时，怎样判断它在语法上是正确的呢?通常的做法是对它进行**语法分析**，看其中的语法范畴是否都处于适当的位置。

如果所给的**语句不缺少必要的语法项目**，而且各**语法项目** 又都处于**适当的位置**，则认为该语句在语法上是正确的.

否则就认为是不正确的.

语法分析工作通常是借助于所给语句的**语法分析树**来进行的.



图：语句7的语法分析树

语法分析树把所给语句分解为它的各个组成部分，并以此来描述所给语句的语法结构.例如,在语句7的语法分析树中：

<语句>由<主语>,<谓语>,<间接宾语>,<直接宾语>依次组合而成.

<主语>由<代词>组成.

为了描述语句的语法结构，使用了称之为语法项目的新符号，如<主语>,<谓语>,<名词>等.为了把语法项目和语言中的词汇区分开来，用尖括号<和>把语法项目括起来.

如果用"$\Longrightarrow$"表示"由...组成的"，那么句7的语法分析树中所包含的语法规则为:

$<$ 语句 $>\Longrightarrow<$ 主语 $><$ 谓语 $><$ 间接宾语 $><$ 直接宾语 $>$;

$<$ 主语 $>\Longrightarrow<$ 代词 $>$;

$<$ 谓语 $>\Longrightarrow<$ 动词 $>$;

$<$ 动词 $>\Longrightarrow<$ 送给 $>$;

$<$ 间接宾语 $>\Longrightarrow<$ 代词 $>$;

$<$ 直接宾语 $>\Longrightarrow<$ 形容词短语 $><$ 名词 $>$;

$<$ 代词 $>\Longrightarrow$ 我;

$<$ 形容词短语 $>\Longrightarrow<$ 冠词 $><$ 副词 $><$ 形容词 $>$;

$<$ 冠词 $>\Longrightarrow$ 一本;

$<$ 副词 $>\Longrightarrow$ 很;

$<$ 形容词 $>\Longrightarrow$ 有趣的;

$<$ 名词 $>\Longrightarrow$ 书;

这类语法规则,称为"产生式"或"重写规则".

访问主页
标题页
第 7 页 共 47 页
返回
全屏显示
关闭
退出

有了一组重写规则之后,可以按照以下方式推导或产生语言的语句:

i) 从语法项目$<$ 语句 $>$开始;

ii) 对已得到的由语法项目和词汇所组成的符号串中的某个语法项目，用一个以此语法项目为左端的重写规则的右端来取代，就得到一个新的由语法项目和语言的词汇所组成的符号串;

iii) 重复ii)直到所得符号串中再无语法项目为止，最后所得到的结果就是语言的一个语句.

例如，语句”他送给我一本非常有趣的书”的全部推导过程为:

< 语句 > ⟹< 主语 >< 谓语 >< 间接宾语 >< 直接宾语 >
⟹< 主语 >< 谓语 >< 代词 >< 直接宾语 >
⟹< 主语 >< 谓语 >< 代词 >< 形容词短语 >< 名词 >
⟹< 主语 >< 谓语 >< 代词 >< 冠词 >< 副词 >< 形容词 >< 名词 >
⟹ 他 < 谓语 >< 代词 >< 冠词 >< 副词 >< 形容词 >< 名词 >
⟹ 他送给 < 代词 >< 冠词 >< 副词 >< 形容词 >< 名词 >
⟹ 他送给我 < 冠词 >< 副词 >< 形容词 >< 名词 >
⟹ 他送给我一本 < 副词 >< 形容词 >< 名词 >
⟹ 他送给我一本很 < 形容词 >< 名词 >
⟹ 他送给我一本很有趣的 < 名词 >
⟹ 他送给我一本很有趣的书

在推导产生语言的语句的步骤ii)中，由于对符号串中语法项目的选取及重写规则的选取都有一定的任意性，所以所给的语法规则可以描述许多不同的语句.例如，从下面七条重写规则:

$<$ 语句 $>\rightarrow<$ 主语 $><$ 谓语 $>$;

$<$ 主语 $>\rightarrow$ 我|你|他;

$<$ 谓语 $>\rightarrow$ 走了|来了|坐下;

可以推导出以下九个语句:

我走了	我来了	我坐下
你走了	你来了	你坐下
他走了	他来了	他坐下

访问主页
标题页
第 10 页 共 47 页
返回
全屏显示
关闭
退出

语法的目的之一就是用适当多条(一般是有限多条)把语言的全部语句都推导出来，这对有无穷多语句的语言来说，是有意义的.

在重写规则中出现了两类符号:

在最后形成的语句中不出现的语法项目,如$<$ 主语 $><$ 谓语 $><$ 名词 $><$ 动词 $>$等等, 我们称之为非终极符或变量.

所要描述的语言的词汇,即在最后形成的语句中出现的词语,如他,送给,我,有趣的,等等,我们称之为终极符.

此外,还有一个重写规则的集合,它一般是一个有限集.

最后,我们指出，语法上正确的语句并不一定是有意义的,例如下面的语句:

小米爱吃鸡.

访问主页
标题页
第 11 页 共 47 页
返回
全屏显示
关闭
退出

2.2. 文法的形式定义

本节给出文法和派生的定义，并给出文法所产生的字和语言

2.2. 文法的形式定义

本节给出文法和派生的定义，并给出文法所产生的字和语言

定义 2.2.1 文法： 设 V_N 和 V_T 都是非空有穷集，$S \in V_N$ 且 $V_N \cap V_T = \emptyset$, 记 $V = V_N \cup V_T$, 并称 $G =< V_N, V_T, P, S >$ 为文法，其中：

V_N– 称为 G 的变量集或非终极符字母表，并称 V_N 中的元素为 G 的变量或非终极符.

2.2. 文法的形式定义

本节给出文法和派生的定义，并给出文法所产生的字和语言

定义 2.2.1 文法：设 V_N 和 V_T 都是非空有穷集，$S \in V_N$ 且 $V_N \cap V_T = \emptyset$, 记 $V = V_N \cup V_T$, 并称 $G =< V_N, V_T, P, S >$ 为文法，其中：

V_N– 称为 G 的变量集或非终极符字母表，并称 V_N 中的元素为 G 的变量或非终极符.

V_T– 称为 G 的终极符字母表，并称 V_T 中的元素为 G 的终极符.

2.2. 文法的形式定义

本节给出文法和派生的定义，并给出文法所产生的字和语言

定义 2.2.1 文法： 设 V_N 和 V_T 都是非空有穷集，$S \in V_N$ 且 $V_N \cap V_T = \emptyset$, 记 $V = V_N \cup V_T$, 并称 $G =< V_N, V_T, P, S >$ 为文法，其中：

V_N– 称为 G 的变量集或非终极符字母表，并称 V_N 中的元素为 G 的变量或非终极符.

V_T– 称为 G 的终极符字母表，并称 V_T 中的元素为 G 的终极符.

P– 有穷集，称为 G 的产生式集，并称 G 中的元素为 G 的产生式.P 中的元素都具有如下的形式：

2.2. 文法的形式定义

本节给出文法和派生的定义，并给出文法所产生的字和语言

定义 2.2.1 文法：设 V_N 和 V_T 都是非空有穷集，$S \in V_N$ 且 $V_N \cap V_T = \emptyset$, 记 $V = V_N \cup V_T$, 并称 $G =< V_N, V_T, P, S >$ 为文法，其中：

V_N– 称为 G 的变量集或非终极符字母表，并称 V_N 中的元素为 G 的变量或非终极符.

V_T– 称为 G 的终极符字母表，并称 V_T 中的元素为 G 的终极符.

P– 有穷集，称为 G 的产生式集，并称 G 中的元素为 G 的产生式.P 中的元素都具有如下的形式：

$$\alpha \to \beta, \alpha \in V^*V_NV^* \text{且} \beta \in V^*$$

S– 称为 G 的句子符号。

通常用大写拉丁字母 $A, B, C, \ldots$ 等表示非终极符，用较前面的小写拉丁字母 $a, b, c, \ldots$ 等表示终极符，用较后面小写拉丁字母 t, u, v, w, x, y, z 等表示 V_T 上的字，即终极符号串，再用小写的希腊字母 $\alpha, \beta, \gamma, \ldots$ 等表示 V 上的字，即由终极符和非终极符组成的符号串.

通常用大写拉丁字母 $A, B, C, \ldots$ 等表示非终极符，用较前面的小写拉丁字母 $a, b, c, \ldots$ 等表示终极符，用较后面小写拉丁字母 t, u, v, w, x, y, z 等表示 V_T 上的字，即终极符号串，再用小写的希腊字母 $\alpha, \beta, \gamma, \ldots$ 等表示 V 上的字，即由终极符和非终极符组成的符号串.

例 1： 取 $G_1 =< \{S_0, S_1\}, \{0, 1\}, P_1, S_0 >$，其中 P_1 由以下的产生式组成：

$$\begin{array}{ll} S_0 \to 0; & S_1 \to 1; \\ S_0 \to 0S_0; & S_1 \to 0S_1; \\ S_0 \to 1S_1; & S_1 \to 1S_0; \end{array}$$

显然，G_1 就是一个文法.

为了给出文法$G =< V_N, V_T, P, S >$所产生的字和语言，先行引进V^*上的一个二元关系，即所谓的”派生”关系.

为了给出文法$G =< V_N, V_T, P, S >$所产生的字和语言，先行引进V^*上的一个二元关系，即所谓的"派生"关系.

定 义2.2.2设 文 法$G =< V_N, V_T, P, S >$且$\alpha, \beta \in V^*$. 如 果 存在$\alpha_1, \alpha_2, \alpha_3, \gamma \in V^*$使：

$$\alpha = \alpha_1\alpha_2\alpha_3, \beta = \alpha_1\gamma\alpha_3 \text{且} \alpha_2 \to \gamma \in P$$

则称β为α的直接派生或一步派生，记为：$\alpha \Longrightarrow_G \beta$.

即$\alpha \underset{G}{\Rightarrow} \beta$当且仅当$\exists_{v^*}\alpha_1\alpha_3\exists_p(\alpha_2 \to \gamma), \alpha = \alpha_1\alpha_2\alpha_3, \beta = \alpha_1\gamma\alpha_3$.

不强调文法G时，把$\Longrightarrow_G$简写为$\Longrightarrow$.

定理2.2.1若文法$G =< V_N, V_T, P, S >$，则$\underset{G}{\Rightarrow}$为$V^*$上的一个二元关系.

定义2.2.3设文法$G =< V_N, V_T, P, S >$,令：

i) $\stackrel{0}{\Longrightarrow}_G = I_v^*$, I_v^*表示v^*上的恒等关系;

ii) $\stackrel{n+1}{\Longrightarrow}_G = \stackrel{n}{\Longrightarrow}_G \circ \Longrightarrow_G$ ($\circ$表示关系的合成);

iii) $\stackrel{+}{\Longrightarrow}_G = \cup_{n=1}^{\infty} \stackrel{n}{\Longrightarrow}_G$;

iv) $\stackrel{*}{\Longrightarrow}_G = \cup_{n=0}^{\infty} \stackrel{n}{\Longrightarrow}_G$;

定理2.2.3设文法$G =< V_N, V_T, P, S >$，

i) 若$\alpha, \beta \in V^*$,则$\alpha \stackrel{0}{\Longrightarrow} \beta$当且仅当$\alpha = \beta$.

ii) $\stackrel{+}{\Longrightarrow}_G$为二元关系$\Longrightarrow$的传递闭包，因而为$V^*$上的一个传递二元关系.

iii) $\stackrel{*}{\Longrightarrow}_G$为二元关系$\Longrightarrow$的自反传递闭包，因而为$V^*$上的一个自反传递二元关系.

iv) 若$\alpha, \beta \in V^*$使$\alpha \stackrel{n+1}{\Longrightarrow} \beta$,则必有$\alpha_1, ..., \alpha_n \in V^*$使

v) $\alpha \Longrightarrow \alpha_1$;

$\alpha_i \Longrightarrow \alpha_{i+1}, 1 \leq i < n$;

$\alpha_n \Longrightarrow \beta$;

vi) 若$\alpha, \beta \in V^*$使$\alpha \stackrel{+}{\Longrightarrow} \beta$,则必有$n \in \mathbb{N}$使$\alpha \stackrel{n+1}{\Longrightarrow} \beta$.

若$\alpha, \beta \in V^*$使$\alpha \stackrel{*}{\Longrightarrow} \beta$,则必有$n \in \mathbb{N}$使$\alpha \stackrel{n}{\Longrightarrow} \beta$.

定义2.2.4设文法$G =< V_N, V_T, P, S >$，$w \in V_T^*$且$\alpha, \beta \in V^*$:

i) 若$\alpha \stackrel{*}{\Longrightarrow} \beta$,则称$\beta$为$\alpha$的一个派生，或由$\alpha$可派生出$\beta$.

ii) 若$S \stackrel{*}{\Longrightarrow} \alpha$,则称$\alpha$为$G$的一个句型.

iii) 若$S \stackrel{*}{\Longrightarrow} w$,则称$w$为$G$的一个字.

iv) 令$L(G) = \{w \in V_T^* | S \stackrel{*}{\Longrightarrow} w\}$,并称$L(G)$为$G$产生的语言.

访问主页
标题页
第 17 页 共 47 页
返回
全屏显示
关闭
退出

下面将给出一些文法的例子并求出它们各自的语言，但是在证明这些文法所产生的语言的过程中，往往要用到一种工具,即函数$\#(w,a)$,它表示字w中所含符号a的个数。

定理2.2.3设Σ为字母表且$a \in \Sigma$，则存在唯一的一个函数$f_a : \Sigma^* \to \mathbb{N}$满足:

i) $f_a(\epsilon) = 0$.

ii) 若$w \in \Sigma^*$且$b \in \Sigma$,则：

$$f_a(bw) = \begin{cases} f_a(w)+1 & ,\text{若} b = a \\ f_a(w) & ,\text{否则} \end{cases}$$

定 义2.2.5设Σ为字母表且$a \in \Sigma$，对每个$w \in \Sigma^*$,把$f_a(w)$记为$\#(w,a)$,并称$\#(w,a)$为字w中所含a的个数.

不难证明，$\#(w,a)$具有以下性质：

1) $\#(\epsilon, a) = 0$;

2) 若$w \in \Sigma^*$,则$\#(w,a) \geq 0$;

3) $\#(a,a) = 1$;

4) 若$b \in \Sigma, b \neq a$,则$\#(b,a) = 0$;

5) 若$w \in \Sigma^*, b \in \Sigma, b \neq a$,则$\#(bw,a) = \#(w,a)$;

6) 若$w \in \Sigma^*$,则$\#(aw,a) = \#(w,a) + 1$;

7) 若$w \in \Sigma^*$,则$\#(w,a) = \#(w^R,a)$;

8) 若$u, v \in \Sigma^*$,则$\#(uv,a) = \#(u,a) + \#(v,a)$;

9) 若$w \in \Sigma^*$,则$|w| = \Sigma_{a\in\Sigma}\#(w,a)$;

访问主页
标 题 页
第 19 页 共 47 页
返 回
全屏显示
关 闭
退 出

例2：对于例1中给出的下列文法G_1：

$$\begin{array}{ll} S_0 \longrightarrow 0; & S_1 \longrightarrow 1; \\ S_0 \longrightarrow 0S_0; & S_1 \longrightarrow 0S_1; \\ S_0 \longrightarrow 1S_1; & S_1 \longrightarrow 1S_0; \end{array}$$

我们有：

$$L(G_1) = \{w \in \{0,1\}^+ | \#(w,1)\text{为偶数}\}.$$

定义2.2.6设G_1和G_2为两个文法，若$L(G_1) = L(G_2)$,则称G_1和G_2等价，记为$G_1 \sim G_2$.

显然文法的等价具有以下性质：

设G_1, G_2, G_3为文法

1) 自反性：$G \sim G$;

2) 对称性：若$G_1 \sim G_2$，则$G_2 \sim G_1$;

3) 传递性：若$G_1 \sim G_2, G_2 \sim G_3$，则$G_1 \sim G_3$;

例3 取$G_2 =< \{S, A\}, \{0, 1, ..., 9\}, P_2, S >$，其中$P_2$由以下的产生式所组成：

$$S \to j; \qquad A \to j;$$

$$S \to iA; \qquad A \to jA;$$

$$i = 1, 2, ..., 9 \text{且} j = 0, 1, ..., 9.$$

显然G_2是一个文法，且不难验证:$L(G_2) = \mathbb{N}$.

例4 取$G_3 =< \{R_0, R_1, R_2, R_3\}, \{0, 1, ..., 9\}, P_3, R_0 >$，其中$P_3$由以下的一些产生式组成：

$$
\begin{array}{ll}
R_0 \to 0; & R_1 \to 0R_1; \\
R_0 \to i; & R_1 \to j; \\
R_0 \to iR_3; & R_1 \to jR_3; \\
R_0 \to kR_1; & R_1 \to iR_1; \\
R_0 \to jR_2; & R_1 \to kR_2; \\
R_2 \to 0R_2; & R_3 \to 0; \\
R_2 \to k; & R_3 \to 0R_3; \\
R_2 \to kR_3; & R_3 \to i; \\
R_2 \to iR_2; & R_3 \to iR_3; \\
R_2 \to jR_1; & R_3 \to kR_1; \\
 & R_3 \to jR_2;
\end{array}
$$

$i \in \{3, 6, 9\}, j \in \{1, 4, 7\}, k \in \{2, 5, 8\}.$

显然G_3是一个文法，且不难验证:$L(G_3) = \{3n|n \in \mathbb{N}\}$.

例5 取$G_4 =< \{C_0, C_1\}, V_T, P_4, C_0 >$，其中

$$V_T = \left\{ \begin{pmatrix} m \\ n \\ p \end{pmatrix} | m, n, p \in \{0, 1, 2, ...9\} \right\}$$

而P_4由以下的一些产生式组成：

若$m, n, p \in \{0, 1, ..., 9\}, m + n = p$,则有：

$$C_0 \to \begin{pmatrix} m \\ n \\ P \end{pmatrix} \qquad C_0 \to C_0 \begin{pmatrix} m \\ n \\ P \end{pmatrix}$$

若$m, n, p \in \{0, 1, ..., 9\}, m + n = 10 + p$,则有：

$$C_0 \to C_1 \begin{pmatrix} m \\ n \\ P \end{pmatrix}$$

续例5

若$m,n,p \in \{0,1,...,9\}, m+n+1=p$,则有：

$$C_1 \rightarrow C_0 \begin{pmatrix} m \\ n \\ P \end{pmatrix} \qquad C_1 \rightarrow \begin{pmatrix} m \\ n \\ P \end{pmatrix}$$

若$m,n,p \in \{0,1,...,9\}, m+n+1=10+p$,则有：

$$C_1 \rightarrow C_1 \begin{pmatrix} m \\ n \\ P \end{pmatrix}$$

显然G_3是一个文法，而且不难用归纳法证明:

续例5

i) $C_0 \overset{*}{\Longrightarrow} \begin{pmatrix} m_k \\ n_k \\ p_k \end{pmatrix} \dots \begin{pmatrix} m_0 \\ n_0 \\ p_0 \end{pmatrix}$ 当且仅当竖式加法 $\begin{array}{r} m_k \ \cdots \ m_0 \\ + \quad n_k \ \cdots \ n_0 \\ \hline p_k \ \cdots \ p_0 \end{array}$ 成立

ii) $C_1 \overset{*}{\Longrightarrow} \begin{pmatrix} m_k \\ n_k \\ p_k \end{pmatrix} \dots \begin{pmatrix} m_0 \\ n_0 \\ p_0 \end{pmatrix}$ 当且仅当竖式加法 $\begin{array}{r} m_k \ \cdots \ m_0 \\ n_k \ \cdots \ n_0 \\ + \qquad\qquad 1 \\ \hline p_k \ \cdots \ p_0 \end{array}$ 成立

因此，$L(G_4) = \left\{ \begin{pmatrix} m_k \\ n_k \\ p_k \end{pmatrix} \dots \begin{pmatrix} m_0 \\ n_0 \\ p_0 \end{pmatrix} \middle| \text{且竖式加法} \begin{array}{r} m_k \ \cdots \ m_0 \\ + \quad n_k \ \cdots \ n_0 \\ \hline p_k \ \cdots \ p_0 \end{array} \text{成立} \right\}$

例6 取$G_5 =< \{S\}, \{0, 1\}, P_5, S >$，其中$P_5$由以下的产生式所组成：

$$S \rightarrow 01; \qquad S \rightarrow 0S1;$$

显然G_5是一个文法，且不难验证:$L(G_5) = \{0^n1^n | n \geq 1\}$.

例7 取$G_6 =< \{S, B, C\}, \{a, b, c\}, P_6, S >$，其中$P_6$由以下的产生式所组成：

1) $S \to aSBC$;

2) $S \to aBC$;

3) $CB \to BC$;

4) $aB \to ab$;

5) $bB \to bb$;

6) $bC \to bc$;

7) $cC \to cc$;

显然G_5是一个文法，且不难验证:$L(G_6) = \{a^n b^n c^n | n \geq 1\}$.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

定义 2.3.1 ：设$G =< V_N, V_T, P, S >$为文法:

i) 若对P中生成式不加任何限制,则称G为0型文法,或短语结构文法(Phrase structure grammar),简记为psg.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

定义 2.3.1 ：设$G =< V_N, V_T, P, S >$为文法：

i) 若对P中生成式不加任何限制,则称G为0型文法,或短语结构文法(Phrase structure grammar),简记为psg.

ii) 若每个$\alpha \to \beta \in P$都满足$|\alpha| \le \beta$,则称G为1型文法, 或上下文有关文法(Context-sensitive grammar),简记为csg.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

定义 2.3.1 ： 设$G =< V_N, V_T, P, S >$为文法:

i) 若对P中生成式不加任何限制,则称G为0型文法,或短语结构文法(Phrase structure grammar),简记为psg.

ii) 若每个$\alpha \to \beta \in P$都满足$|\alpha| \le \beta$,则称G为1型文法, 或上下文有关文法(Context-sensitive grammar),简记为csg.

iii) 若每个$\alpha \to \beta \in P$都具有如下形式:

$$A \to \beta; \qquad \beta \in V^+, A \in V_N.$$

则称G为2型文法,或上下文无关文法(Context-sensitive grammar),简记为cfg.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

定义 2.3.1 ：设$G =< V_N, V_T, P, S >$为文法:

i) 若对P中生成式不加任何限制,则称G为0型文法,或短语结构文法(Phrase structure grammar),简记为psg.

ii) 若每个$\alpha \to \beta \in P$都满足$|\alpha| \le \beta$,则称G为1型文法, 或上下文有关文法(Context-sensitive grammar),简记为csg.

iii) 若每个$\alpha \to \beta \in P$都具有如下形式:

$$A \to \beta; \qquad \beta \in V^+, A \in V_N.$$

则称G为2型文法,或上下文无关文法(Context-sensitive grammar),简记为cfg.

iv) 若每个$\alpha \to \beta \in P$都具有如下形式:

$$A \to a; \qquad A \to aB; a \in V^T, A, B \in V_N.$$

则称G为3型文法,或正规文法(regular grammar),简记为rg.

2.3. 文法的分类

文法可用对其生成式施加的不同限制来进行分类.现行的分类是由Chomsky于1959年引进的，只是叙述上稍有变化.

定义 2.3.1 ： 设$G =< V_N, V_T, P, S >$为文法：

i) 若对P中生成式不加任何限制,则称G为0型文法,或短语结构文法(Phrase structure grammar),简记为psg.

ii) 若每个$\alpha \to \beta \in P$都满足$|\alpha| \le \beta$,则称G为1型文法, 或上下文有关文法(Context-sensitive grammar),简记为csg.

iii) 若每个$\alpha \to \beta \in P$都具有如下形式:

$$A \to \beta; \qquad \beta \in V^+, A \in V_N.$$

则称G为2型文法,或上下文无关文法(Context-sensitive grammar),简记为cfg.

iv) 若每个$\alpha \to \beta \in P$都具有如下形式:

$$A \to a; \qquad A \to aB; a \in V^T, A, B \in V_N.$$

则称G为3型文法,或正规文法(regular grammar),简记为rg.

由以上定义可知：3型文法必是2型文法,2型文法必是1型文法,1型文法必是0型文法.

例1 我们考虑§2.2的各例中所给出的文法：

1) G_1, G_2, G_3, G_4都是rg,因此也是cfg, csg, psg;

2) G_5为cfg,因此也是csg, psg;

3) G_5为csg,因此也是psg;

例2 $G_7 =< \{S, A, B\}, \{a, b\}, P, S >$,其中$P$由以下生成式构成：

$$S \rightarrow bA; \quad A \rightarrow a; \quad A \rightarrow aS; \quad A \rightarrow bAA;$$
$$S \rightarrow aB; \quad B \rightarrow b; \quad B \rightarrow bS; \quad B \rightarrow aBB;$$

显然G_7为cfg,并且$L(G_7) = \{w \in \{a, b\}^+ | w$中含有相同个数的$a$和$b\}$.

例3 $G_8 =< \{S\}, \{a, b, *\}, P, S >$,其中$P$由以下生成式构成：

$$S \to aSa|bSb|*;$$

显然G_8为cfg,且不难验证：$L(G_8) = \{w * w^R | w \in \{a, b\}^*\}$.

例4 $G_9 =< \{S\}, \{0, 1\}, P, S >$,其中$P$由以下生成式构成：

$$S \to 0S0; \quad S \to 1S1;$$

$$S \to 00; \quad S \to 11;$$

显然G_9为cfg,并且不难验证$L(G_9) = \{ww^R | w \in \{0, 1\}^+\}$.

例5 设Σ为任意一个字母表，令$L = \{ww | w \in \Sigma^+\}$,虽然$L$看起来与$\{ww^R | w \in \{0, 1\}^*\}$很象,实际上是有本质的差别，下面就来构造产生$L$的文法.

定义2.3.1 设Σ为字母表，如果$G=< V_N,\Sigma,P,S >$为$i(i=0,1,2,3)$型文法，则称$L(G)$为i型语言，或Σ上的i型语言.

对任意的字母表Σ，若用$\mathscr{L}_i(\Sigma)(i=0,1,2,3)$表示$\Sigma$上的所有$i$型语言的语言类,则有：

$$\mathscr{L}_3(\Sigma)\subset\mathscr{L}_2(\Sigma)\subset\mathscr{L}_1(\Sigma)\subset\mathscr{L}_0(\Sigma).$$

若用$\mathscr{L}_i$表示所有i型语言的语言类,则有：

$$\mathscr{L}_3\subset\mathscr{L}_2\subset\mathscr{L}_1\subset\mathscr{L}_0.$$

0型语言又称为递归可枚举集合，简记为$r.e$集合;

1型语言又称为上下文有关语言，简记为csl;

2型语言又称为上下文有关语言，简记为cfl;

3型语言又称为正规集或正规语言，简记为rl;

2.4. 空字

按定义，在csl, cfl和正规语言中都不含有ϵ.然而，ϵ为自由Monoid $<\Sigma^*, \cdot, \epsilon>$的幺元，具有特殊的重要性，必须允许它们可以含有$\epsilon$，否则将带来极大的不便。为此，对 § 2.3中引进的csg, cfg, rg作如下的修改：

允许P中含有如下的产生式：

$S \to \epsilon$,S为句子开始符号且不出现在任何的生成式之右端.

下面将讨论这一修改对csl, cfl和正规语言的影响.

引理2.4.1 若$G =< V_N, V_T, P, S >$为i型文法，则有同型文法$G_1 =< V_N', V_T, P', S' >$满足：

i) $L(G_1) = L(G)$;

ii) S'不出现在P'中任何生成式之右端.

定理2.4.1 若L为$i(i = 1, 2, 3)$型语言，则$L \cup \{\epsilon\}$和$L\ \{\epsilon\}$也是i型语言.

为了方便起见，今后常称满足定理2.4.1的$i(i = 1, 2, 3)$型文法为改进的i型文法.

访问主页
标题页
第 34 页 共 47 页
返回
全屏显示
关闭
退出

例2.4.1 §2.2中的例7中的文法G_6:

$G_6 =< \{S, B, C\}, \{a, b, c\}, P_6, S >$，其中$P_6$由以下的产生式所组成：

$S \to aSBC;\qquad S \to aBC;$

$CB \to BC;\qquad aB \to ab;$

$bB \to bb;\qquad bC \to bc;$

$cC \to cc;$

文法G_6的改进文法为$G_6' =< V_N', \{a, b, c\}, P', S' >$，其中$V_N' = \{S', S, B, C\}$,$P'$由以下的产生式所组成：

$S' \to aSBC;\qquad S' \to aBC;$

$S \to aSBC;\qquad S \to aBC;$

$CB \to BC;\qquad aB \to ab;$

$bB \to bb;\qquad bC \to bc;$

$cC \to cc;$

2.5. Kurode范式

定义 2.5.1 ： 设文法$G =< V_N, V_T, P, S >$，且$\alpha_0 \in V^*$.如果派生$\alpha_0 \stackrel{*}{\Longrightarrow} \alpha$具有以下形式：

$$\alpha_0 \Longrightarrow \alpha_1 \Longrightarrow \cdots \Longrightarrow \alpha_n \quad (\alpha_n = \alpha, n \in \mathbb{N})$$

而且对每个$i = 0, 1, ..., n-1$,皆有$\beta_i, \gamma_i, \theta_i, \phi_i \in V^*$，使：

$\alpha_i = \beta_i\theta_i\gamma_i$

$\alpha_{i+1} = \beta_i\phi_i\gamma_i$

$\theta_i = \phi_i \in P$

$|\beta_i| < |\beta_{i+1}\theta_{i+1}|$(或$|\gamma_i| < |\theta_{i+1}\gamma_{i+1}|$)

就称派生$\alpha_0 \stackrel{*}{\Longrightarrow} \alpha$为最左(最右)派生，记为$\alpha_0 \stackrel{*}{\Longrightarrow}_L \alpha$(或$\alpha_0 \stackrel{*}{\Longrightarrow}_R \alpha$).

定理 2.5.1 ： 设文法$G=<V_N,V_T,P,S>$，且$\alpha_0,a\in V^*$.若$n\in\mathbb{N}$,则$\alpha_0\overset{n}{\Longrightarrow}\alpha$当且仅当$\alpha_0\overset{n}{\Longrightarrow}_L\alpha$.

定理 2.5.2 ： 设文法$G=<V_N,V_T,P,S>$，且$\alpha_0,a\in V^*$,则$\alpha_0\overset{*}{\Longrightarrow}\alpha$当且仅当$\alpha_0\overset{*}{\Longrightarrow}_R\alpha$.

定义 2.5.2 ： 若文法$G=<V_N,V_T,P,S>$的P中产生式都具有如下形式：

i) $S\to\epsilon$;

ii) $A\to a$;

iii) $A\to B$;

iv) $A\to BC$;

v) $AB\to AC$;

vi) $AB\to CB$;

vii) $AB\to B$;

其中，$a\in V_T$且$A,B,C\in V_N$，而且，当$S\to\epsilon\in P$时，S不会出现在P中任何产生式的右端.

称满足上述条件的文法G为Kurode范式文法.

定理 2.5.3 ： 对每一个文法$G=<V_N,V_T,P,S>$，都有一个Kurode范式文法G'，使得$L(G)=L(G')$.

定理2.5.4 ：若$csgG =< V_N, V_T, P, S >$，则有文法$G' =< V_N', V_T, P', S' >$，使$L(G') = L(G)$,其中$P'$中产生式都具有如下形式：

i) $S \to \epsilon$;

ii) $A \to a$;

iii) $A \to B$;

iv) $A \to BC$;

v) $AB \to AC$;

vi) $AB \to CB$;

其中，$a \in V_T$且$A, B, C \in V_N$，当$S \to \epsilon \in P$时，S不会出现在P中任何产生式的右端.

定理2.5.5 ：若$cfgG =< V_N, V_T, P, S >$，则有文法$G' =< V_N', V_T, P', S' >$，使$L(G') = L(G)$,$P'$中产生式都具有如下形式：

i) $S \to \epsilon$;

ii) $A \to a$;

iii) $A \to B$;

iv) $A \to BC$;

其中，$a \in V_T$且$A, B, C \in V_N$，当$S \to \epsilon \in P$时，S不会出现在P中任何产生式的右端.

例2.5.1 取$G =< \{S, A, B, C\}, \{a, b\}, P, S >$,其中$P$由以下产生式组成:

$$S \to ASa; \quad AC \to aC; \quad S \to BSb; \quad Aa \to aA;$$
$$S \to C; \quad Ab \to bA; \quad C \to \epsilon; \quad BC \to bC;$$
$$Ba \to aB; \quad Bb \to bB;$$

不难证明$L(G) = \{ww|w \in \{a, b\}^*\}$.因为$S \Longrightarrow C \Longrightarrow \epsilon$,即$\epsilon \in L(G)$,所以可以取:

$V'_N = \{S', S, A, B, C, D_a, D_b, W_1, W_2\}$

$P' = \{S' \to \epsilon, S' \to S\} \cup \{D_a \to a, D_b \to b\} \cup \tilde{P}$.其中$\tilde{P}$由以下产生式组成:

$$S \to AW_1; \quad AC \to D_aC; \quad W_1 \to SD_a; \quad AD_a \to D_aA;$$
$$S \to BW_2; \quad AD_b \to D_bA; \quad W_2 \to SD_b; \quad BC \to D_bC;$$
$$S \to C; \quad BD_a \to D_aB; \quad BD_b \to D_bB;$$
$$CX \to X; \quad XC \to X, X \in \{S, A, B, C, D_a, D_b\}.$$

显然$G' =< V'_N, \{a, b\}, P', S' >$使$L(G') = L(G)$. 对$\tilde{P}$中形如$XY \to YX$的产生式可以用以下三个产生式代替:

$$XY \to X'Y; \quad X'Y \to X'X; \quad X'X \to YX.$$

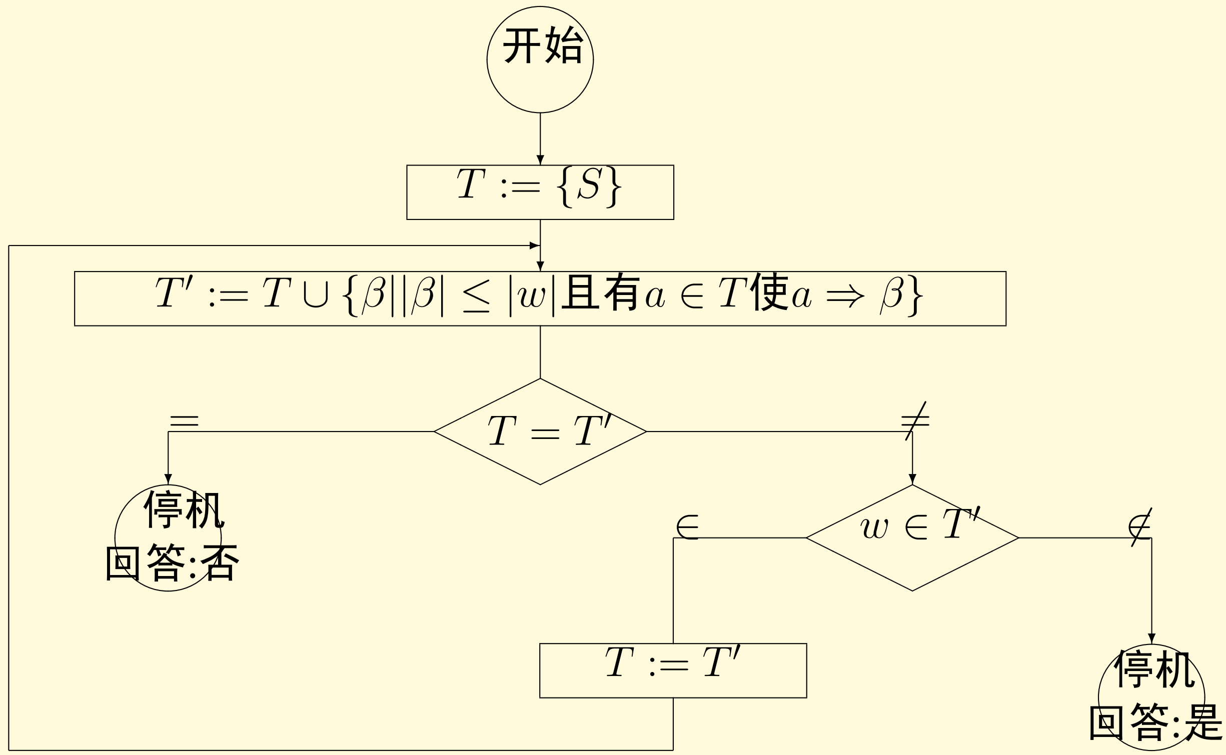
从而将文法P'改造为Kurode范式文法.

2.6. 上下文有关文法的递归性

本节主要讨论有关csg的字问题：对任意的csg $G =< V_N, V_T, P, S >$及$w \in V_T^*$，是否$w \in L(G)$.对一个判定问题$\boldsymbol{P}$,若有一个解决它的算法，就称$\boldsymbol{P}$为可解的，否则称$\boldsymbol{P}$为不可解的.

定理2.6.1 有关csg的字问题是可解的;

定理2.6.1 有关csg的字问题是可解的;



解csg字问题的算法

定理2.6.2 设$csg\ \ G=<V_N,V_T,P,S>$,$w\in L(G)$且$w\neq\epsilon$.当记$m=|w|,k=^{\#}V$,且n_w为w的最短派生的派生步骤数时,则有:

$$n_w\leq k+k^2+...+k^m$$

定理2.6.3 设$csg\ \ G=<V_N,V_T,P,S>$,$w\in L(G)$且$w\neq\epsilon$.当记$m=|w|,k=^{\#}V$,且n_w为w的最短派生的派生步骤数时,则有:

$$n_w\leq\frac{km(m+1)}{2}-\frac{(m-1)(m-2)}{2}$$

定理2.6.4 在定理2.6.3的条件下,$n_w\leq km^2$.

定理2.6.5 对$rg\ \ G$,有$n_w=m$.

2.7. 派生树

本节主要介绍一种描述cfg的派生过程的直观方法，就是所谓的派生树.

定义2.7.1 称序偶$D =< T, \varphi >$为$cfg\ G =< V_N, V_T, P, S >$的派生树.如果$T =< X, E >$为有序树,而且标记函数$\varphi : X \to V \cup \{\epsilon\}$还满足以下条件:

i) 若$v_0 \in X$为T的根,则$\varphi(v_0) = S$.

ii) 若$< v, u >\in E$,则$\varphi(v) \in V_N, \varphi(u) \in V \cup \{\epsilon\}$.

iii) 若v的所有直接子孙，从左至右依次为$v_1, ..., v_n$,则

$$\varphi(v) \to \varphi(v_1)...\varphi(v_n) \in P.$$

例2.7.1 取$G_{12} =< V_N, \{0,1\}, P, S >$,这里$V_N = \{S_1, S, A, B\}$,$P$由以下的产生式组成:

$$S \to \epsilon; \quad S \to S_1;$$

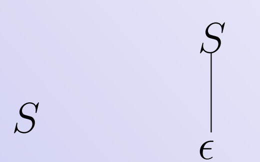
$$S_1 \to 0; \quad S_1 \to 0S_11;$$

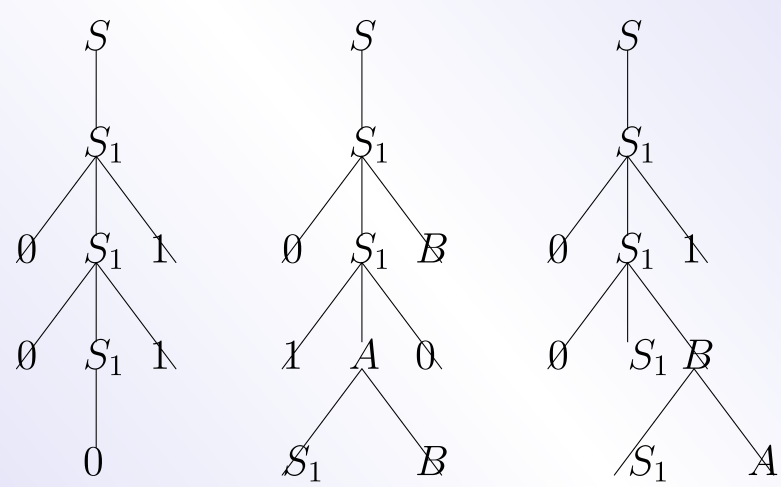
$$S_1 \to 0S_1B; \quad S_1 \to 1A0;$$

$$A \to 1B0; \quad A \to S_1B;$$

$$B \to 1BA; \quad B \to S_1A;$$

考察G_{12}的派生树.





(a) (b) (c) (d) (e)

访问主页
标题页
第45页共47页
返回
全屏显示
关闭
退出

定义2.7.2 设$D =< T, \varphi >$为$cfg\ \ G$的派生树.若T的所有叶子节点从左至右依次为$v_1, ..., v_n$.则称$a = \varphi(v_1)...\varphi(v_n) \in V^*$为$D$的结果.

引理2.7.1 设$G =< V_N, V_T, P, S >$为cfg,如果

$$X_1 \cdots X_n \overset{m}{\Longrightarrow} \alpha \qquad X_1, \cdots, X_n \in V_N \cup V_T$$

则有$m_1, \cdots, m_n \in \mathbb{N}$及$a_1, \cdots, a_n \in (V_N \cup V_T)^*$,使:

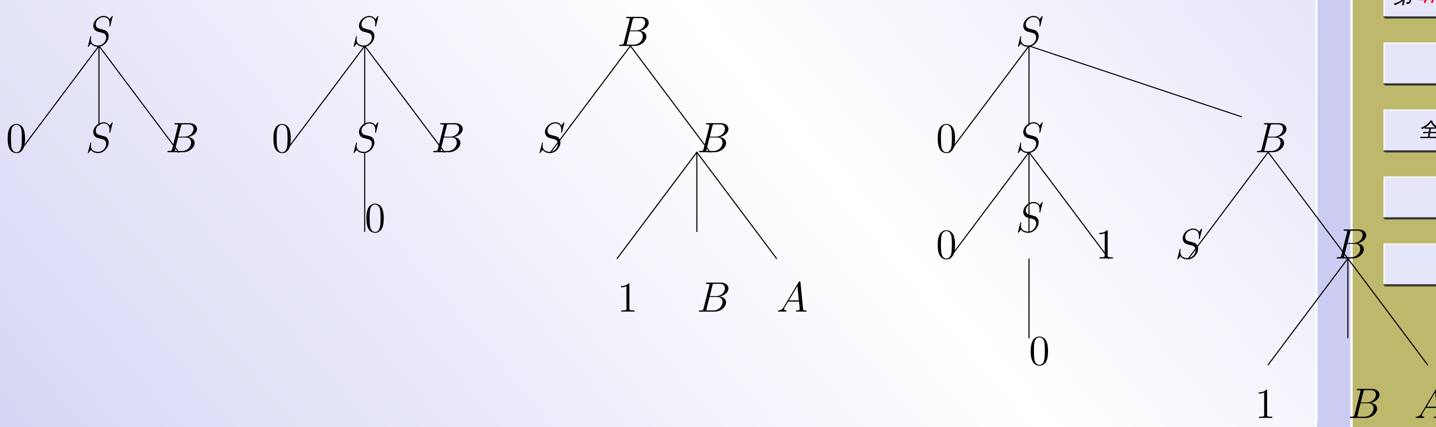
$$m = m_1 + \cdots + m_n \qquad \alpha = \alpha_1 \cdots \alpha_n$$

且$X_i \overset{m_i}{\Longrightarrow} \alpha_i \qquad 1 \leq i \leq n.$

定理2.7.1 设$G =< V_N, V_T, P, S >$为cfg,则$S \overset{*}{\Longrightarrow} \alpha$当且仅当有$G$的一棵以$\alpha$为结果的派生树.

引理2.7.2 设$G =< V_N, V_T, P, S >$为cfg且$A \in V_N$,则$A \stackrel{*}{\Longrightarrow}_G \alpha$当且仅当有$G_A$的一棵以$\alpha$为结果的派生树$D_{A,\alpha}$.

例2.7.2 对例1中的$cfg\ \ G_{12}$,考虑如下的派生$S \stackrel{*}{\Longrightarrow}_G 0001S1BA$.



访问主页
标题页
第 47 页 共 47 页
返回
全屏显示
关闭
退出