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

第四章　上下文无关文法

4 上下文无关文法

4.1. 上下文无关文法的简化

在cfg的形式定义中,允许它含有某些不经济的成分，例如，它可能含有冗余的变量，终极符或产生式。怎样鉴别一个文法中存在的这种冗余成分，并把它们删除呢?这就是本节所要解决的问题.

为方便起见，先引进一些概念和记号.设cfg G $=<V_N, V_T, P, S>$:

1) 若对于$x \in V$,有$\alpha_1, \alpha_2 \in V^*$使$\alpha = \alpha_1 x \alpha_2$,就称$x$在$\alpha$中出现;

2) 如果$S \overset{*}{\Rightarrow} \alpha_1 x \alpha_2 \overset{*}{\Rightarrow} \beta$, 则称$x$在句型$\beta$的一个派生中出现;当$\beta \in V_T^*$时,还称$x$在字$\beta$的一个派生中出现;

3) 对树T,用$D(T)$表示T的高,$|T|$表示T的结点数.

定理4.1.1 以下问题是可解的：对任意的$cfg\ G$,是否$L(G)=\emptyset$?

证明：设$cfg\ G=<V_N,V_T,P,S>$,令

$$V_0=V_T;$$
$$V_{i+1}=V_i\cup\{A\in V_N|有\alpha\in V_i^*使A\to\alpha\in P\},$$
$$(i=0,1,2,\dots)$$

则有$V_i\subseteq V_{i+1}\subseteq V_N\cup V_T,(i=0,1,2,\dots)$

因此$\cup_{i=0}^{\infty}V_i\subseteq V_N\cup V_T$.但$V_N\cup V_T$为有穷集,所以必有$m\in N$,使$V_{m+1}=V_m$.否则$\cup_{i=0}^{\infty}V_i$必为无穷集.这与$V_N\cup V_T$为有穷集矛盾,从而得到$V_m\subseteq\cup_{i=1}^{\infty}V_i\subseteq V_N\cup V_T$.

为了完成本定理的证明,还须证明以下的命题：

若$A\in V_N$,则$A\in V_m$当且仅当有$w\in V_T^*$使$A\overset{*}{\Rightarrow}w$.

1) 必要性:

用关于$n_A = min\{i \in \mathbb{N} | A \in V_i\}$的第二归纳法来证.

当$n_A = 1$时,则$A \in V_1$,因此有$\alpha \in V_T^*$,使$A \to \alpha \in P$, 这时取$w = \alpha$,即知此时命题为真.

假定对任意的正整数$k > 1$,当$n_A < k$时命题为真.当$n_A = k$时,则有$\alpha \in V_{k-1}^*$使$A \to \alpha \in P$.由于$k - 1 \geq 1$,所以有$\alpha \notin V_T^*$.因此可设

$$\alpha = X_1 \dots X_l \quad X_1, \dots, X_l \in V_{k-1}$$

对每个$i(1 \leq i \leq l)$,若$X_i \in V_T$,则取$w_i = X_i$;若$X_i \notin V_T$,则由$n_{X_i} \leq k - 1 < k$及归纳假设知道,必有$w_i \in V_T^*$使$X_i \overset{*}{\Rightarrow} w_i$.从而得到

$$A \Rightarrow X_1 \dots X_l \overset{*}{\Rightarrow} w_1 X_2 \dots X_l \overset{*}{\Rightarrow} \cdots \overset{*}{\Rightarrow} w_1 \dots w_l \in V_T^*$$

即当$n_A = k$时命题为真.

访问主页
标题页
第 5 页 共 108 页
返回
全屏显示
关闭
退出

2) 充分性:

用关于派生$A \overset{*}{\Rightarrow} w$的派生步数$n$之第二归纳法来证.

当$n = 1$时,则$A \rightarrow w \in P$,且$w \in V_T^*$.因此$A \in V_1 \subseteq V_m$.此时命题为真.

假定对任意的正整数$k > 1$,当$n < k$时命题皆真.当$n = k > 1$时,$A \overset{k}{\Rightarrow} w$必有如下形式

$$A \Rightarrow X_1 \ldots X_l \overset{k-1}{\Rightarrow} w$$

根据引理2.2,必有$w_1, \ldots, w_l \in V_T^*$,使

$$w = w_1 \ldots w_l \qquad k_1 + k_2 + \cdots + k_l = k - 1$$

且$X_i \overset{k_i}{\Rightarrow} w_i, \qquad 1 \leq i \leq l$

从而由归纳假设知,对每个$i(1 \leq i \leq l)$,必有$X_i \in V_T$或$X_i \in V_m$.因此$A \in V_{m+1}$即$A \in V_m$.

这表明,当$n = k$时,命题为真.

根据以上证明的命题,有

$$L(G) = \emptyset\text{当且仅当}S \notin V_m$$

从而可以得到给定判定问题的判定算法.

对于$cfg\ G$的$L(G)$是否为空,还可以从派生树的角度来考虑.

定理4.1.2 若$G=<V_N, V_T, P, S>$为cfg,则$L(G)\neq\emptyset$当且仅当有$w\in L(G)$及T_w使$D(T_w)\leq \#V_N$.

证明:

定理4.1.3 设$G = < V_N, V_T, P, S >$为cfg且$x \in V$.有$w \in L(G)$使x在w的一个派生中出现,当且仅当有$u \in L(G)$及T_u,使$D(T_u) \leq 3(^\# V_N + 1)$且$x$在$T_u$对应的$u$的派生中出现.

证明: 充分性是显然的,下面只证必要性.

假定x在w的某个派生中出现且这棵派生树为T_w.

1) 若$D(T_w) \leq 3(^\# V_N + 1)$,则取$u = w$即可.

2) 若$D(T_w) > 3(^\# V_N + 1)$,首先由x在T_w对应的派生中出现知道,在T_w中必有一条如下形式的通路:

$$S \to A_1 \to \cdots \to A_m \to a, a \in V_T, \text{且} A_1, \ldots, A_m \in V_N$$

当记$A_0 = S$及$A_{m+1} = a$时,还有$A_i = x(0 \leq i \leq m+1)$.这时,$m$只能有以下两种情况之一出现:

情况一:$m \geq 2(^\# V_N + 1)$

这时必有$i \geq ^\# V_N + 1$或$m - i \geq ^\# V_N + 1$.因此,在$A_0, \ldots, A_{i-1}$或$A_{i+1}, \ldots, A_m$中必有相同者.不妨设为$A_s = A_t(0 \leq s < t \leq i-1$或$i+1 \leq s < t \leq m)$

情况二:$m < 2(^{\#}V_N + 1)$

因为$D(T_w) > 3(^{\#}V_N + 1)$,所以T_w中还必有一条如下通路:

$A_j \to B_1 \to \cdots \to B_n \to b \quad b \in V_T$且$B_1, \ldots, B_n \in V_N, n \geq^{\#} V_N + 1$

其中$j \leq m$.从而知道在$B_1, \ldots, B_n$中必有相同者,不妨设为$B_{s'} = B_{t'}(1 \leq s' < t' \leq n)$.

若用以T_w的上述通路上的以$A_t(B_{t'})$为根的子树代替其中以$A_s(B_{s'})$为根的子树时,就可以得到一棵以$w' \in L(G)$为结果的新的G的派生树$T_{w'}$,显然$T_{w'}$比T_w至少少一个非叶结点,且x仍为$T_{w'}$的一个结点,即x仍在$T_{w'}$对应的w'的派生中出现.

3) 对w'再重复上述的对w的论证1)及2),再每次选择时,都尽量多地消去多余结点,则至多经过有限次(实际上,不会超过$|T_w| - 3(^{\#}V_N + 1)$次),就会得到一个字$u \in L(G)$及T_u,使$D(T_u) \leq 3(^{\#}V_N + 1)$且$x$在$T_u$对应的$u$的派生中出现.

定理4.1.4 以下的问题都是可解的：

i) 对任意的cfg $G = < V_N, V_T, P, S >$及$x \in V$，是否有$w \in L(G)$使x在w的某个派生中出现?

ii) 对任意的cfg $G = < V_N, V_T, P, S >$及$A \in V_N$，是否有$w \in L(G)$使A在w的某个派生中出现?

iii) 对任意的cfg $G = < V_N, V_T, P, S >$及$a \in V_T$，是否有$w \in L(G)$使a在w的某个派生中出现?

证明： 因为问题ii)和iii)都是问题i)的子问题,所以若问题i)可解,则问题ii)和iii)自然都有解.根据定理4.1.3,x在某个字$w \in L(G)$的一个派生中出现,当且仅当有$u \in L(G)$及T_u存在,使得$D(T_u) \le 3(^{\#}V_N + 1)$,且$x$在$T_u$对应的$u$的派生中出现,即$x$为$T_u$的一个结点. 因此,只须检查所有高度不超过$3(^{\#}V_N + 1)$且结果为$L(G)$中字的派生树,看$x$是否为某棵这样的派生树的一个结点即可.由于$G$的高度不超过$3(^{\#}V_N + 1)$的派生树只有有限棵,所以这可用一个统一的算法来实现.这表明问题i)是可解的.

定 理4.1.5 若$G =< V_N, V_T, P, S >$为cfg且$L(G) \neq \emptyset$,则有cfg $G' =< V'_N, V'_T, P', S >$使:

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

ii) $V'_N \subseteq V_N, V'_T \subseteq V_T$且$P' \subseteq P$;

iii) 若$x \in V'$,则有$w_1, w_2, w_3 \in V'^*_T$使$S \overset{*}{\Rightarrow}_{G'} w_1 x w_3 \overset{*}{\Rightarrow}_{G'} w_1 w_2 w_3$.

证明:令

$$V_N' = \{A \in V_N | 有 w_1, w_2, w_3 \in V_T^* 使 S \stackrel{*}{\Rightarrow} w_1 A w_3 \stackrel{*}{\Rightarrow} w_1 w_2 w_3\}$$

$$V_T' = \{a \in V_T | 有 w_1, w_3 \in V_T^* 使 S \stackrel{*}{\Rightarrow} w_1 a w_3\}$$

$$P' = \{A \to \alpha \in P | A \in V_N' 且 \alpha \in (V_N' \cup V_T')^*\}$$

则$G' =< V_N', V_T', P', S >$为$cfg$,且还满足定理的条件ii)和iii).下面证明$G'$还满足定理的条件i),即$L(G') = L(G)$.

首先由$P' \subseteq P$得:$L(G') \subseteq L(G)$.

其次,任取$w \in L(G)$,则$S \stackrel{*}{\Rightarrow} w$.对这个派生过程中出现的每个$x \in V$,必有$\alpha_1, \alpha_2 \in V^*$,使$S \stackrel{*}{\Rightarrow} \alpha_1 x \alpha_2 \stackrel{*}{\Rightarrow} w$. 这时,根据2.7节中引理2.3知道,必有$w_1, w_2, w_3 \in V_T^*$,使

$$w = w_1 w_2 w_3 \quad \alpha_1 \stackrel{*}{\Rightarrow} w_1 \quad x \stackrel{*}{\Rightarrow} w_2 且 \alpha_3 \stackrel{*}{\Rightarrow} w_3$$

所以$S \stackrel{*}{\Rightarrow} \alpha_1 x \alpha_3 \stackrel{*}{\Rightarrow} w_1 x \alpha_3 \stackrel{*}{\Rightarrow} w_1 w_2 \alpha_3 \stackrel{*}{\Rightarrow} w_1 w_2 w_3$

因此$x \in V_N' \cup V_T'$.从而由P'定义可知$S \stackrel{*}{\Rightarrow}_{G'} w$,即$w \in L(G')$.

推论4.1.6 若L为cfl,则有cfg $G =< V_N, V_T, P, S >$使:

i) $L(G) = L$;

ii) 若$x \in V_N \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$使:

$$S \overset{*}{\Rightarrow} w_1 x w_3 \overset{*}{\Rightarrow} w_1 w_2 w_3.$$

证明: 因为L为cfl,所以有cfg G'使$L(G') = L$.根据定理4.1.5,这时有cfg $G =< V_N, V_T, P, S >$满足:

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

ii 若$x \in V_N \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$,使

$$S \overset{*}{\Rightarrow} w_1 x w_3 \overset{*}{\Rightarrow} w_1 w_2 w_3.$$

定义4.1.1 设$G =< V_N, V_T, P, S >$为上下文无关文法.

1) 若派生$\alpha \Rightarrow \beta$具有如下形式:

$$xA\tilde{\alpha} \Rightarrow xy\tilde{\beta}\tilde{\alpha}(\text{或})\tilde{\alpha}Ax \Rightarrow \tilde{\alpha}\tilde{\beta}yx$$

其中$x \in V_T^*, A \in V_N$且$\tilde{\alpha}, \tilde{\beta} \in (V_N \cup V_T)^*$,就称$\alpha \Rightarrow \beta$为$G$的一步最左(或最右)派生,记为$\alpha \Rightarrow_L \beta$(或$\alpha \Rightarrow_R \beta$).

2) 若派生$\alpha \stackrel{*}{\Rightarrow} \beta$中的每一步派生都是$G$的一步最左(或最右)派生,则称$\alpha \stackrel{*}{\Rightarrow} \beta$为$G$的一个最左(或最右)派生,记为$\alpha \stackrel{*}{\Rightarrow}_L \beta$或$\alpha \stackrel{*}{\Rightarrow}_R \beta$.

定理4.1.7 设$G=<V_N,V_T,P,S>$为cfg,$A\in V_N$且$w\in V_T^*$,则以下各条件等价:

i) $A\overset{*}{\Rightarrow}w$;

ii) $A\overset{*}{\Rightarrow}_L w$;

iii) $A\overset{*}{\Rightarrow}_R w$;

证明: 显然只须证明$i)\Rightarrow ii)$和$i)\Rightarrow iii)$即可.由于这两者的证明是类似的,所以只证前者.

用关于派生$A\overset{*}{\Rightarrow}w$的派生步数$n$的第二归纳法来证明.

1) 当$n=0$时,命题显然为真;

2) 假定对任意的自然数$k>0$,当$n<k$时令题皆为真,则当$n=k$时,由$k>0$知,$A\overset{*}{\Rightarrow}w$必有如下形式:

$$A\Rightarrow u_1\dots u_m\overset{k-1}{\Rightarrow}w\qquad u_1\dots u_m\in V_N\cup V_T$$

根据2.7节中引理2.3知, 必有$w_1, \ldots, w_m \in V_T^*$,使

$w = w_1 \ldots w_m$且$u_i \stackrel{*}{\Rightarrow} w_i \quad 1 \leq i \leq m$

因为每个$u_i \stackrel{*}{\Rightarrow} w_i \quad 1 \leq i \leq m$的派生步数都不超过$k-1$,所以,由归纳假设有$u_i \stackrel{*}{\Rightarrow}_L w_i \quad 1 \leq i \leq m$,因此

$A \Rightarrow_L u_1 \ldots u_m \stackrel{*}{\Rightarrow}_L w_1 u_2 \ldots u_m \stackrel{*}{\Rightarrow}_L \cdots \stackrel{*}{\Rightarrow}_L w_1 \ldots w_{m-1} u_m \stackrel{*}{\Rightarrow}_L w_1 w_2 \ldots w_m$

亦即$A \stackrel{*}{\Rightarrow}_L w$.

下面讨论如何在cfg中消去形如$A \rightarrow B$的单位产生式.

定理4.1.8 对每个cfg G $=<$ V_N, V_T, P, S $>$,皆有cfg G' $=<$ V_N, V_T, P', S $>$满足:

1) $L(G) = L(G')$;

2) 若$A, B \in V_N$,则$A \rightarrow B \notin P'$.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 18 页 共 108 页
返回
全屏显示
关闭
退出

证明：令

$P' = (P \setminus \{A \to \alpha \in P | \alpha \in V_N\}) \cup \{A \to \beta | \beta \notin V_N$, 且有$B \in V_N$及自然数$n <^{\#} V_N$使$A \overset{n}{\Rightarrow}_G B \Rightarrow_G \beta\}$

则$G' =< V_N, V_T, P', S >$显然满足条件ii).由P'的定义可知,如果$\alpha \Rightarrow_{G'} \beta$,则必有$\alpha \overset{*}{\Rightarrow}_G \beta$,所以$L(G') \subseteq L(G)$.

反之,若$w \in L(G)$,则由定理4.1.7知,必有$S \overset{*}{\Rightarrow}_{G\quad L} w$.取一个最短的最左派生,记为

$S \Rightarrow_{G\quad L} x_1A_1\alpha_1 \Rightarrow_{G\quad L} x_2A_2\alpha_2 \Rightarrow_{G\quad L} \cdots \Rightarrow_{G\quad L} x_mA_m\alpha_m \Rightarrow_{G\quad L} w$

在这个派生中,连续使用形如$A \to B \in P(A, B \in V_N)$的生成式的次数必定会小于$^{\#}V_N$,否则上述派生就不是最短的. 从而由$P'$的定义而得$S \overset{*}{\Rightarrow}_{G\quad L} w$

推论4.1.9 对每个$cfl\ \ L$,都存在一个$cfg\ \ G=<V_N,V_T,P,S>$满足:

1) $L(G)=L$;

2) 若$A\rightarrow\alpha\in P$,则$\alpha\notin V_N$;

3) 若$x\in V$,则有$w_1,w_2,w_3\in V_T^*$使:

$$S\overset{*}{\Rightarrow}w_1xw_3\overset{*}{\Rightarrow}w_1w_2w_3$$

4.2. $Chomsky$范式文法

定义4.2.1 如果$cfg\ G=<V_N,V_T,P,S>$的P中之产生式都具有如下形式:

$$A\rightarrow BC\text{或}A\rightarrow a, a\in V_T\text{且}A,B,C\in V_N$$

则称G为$Chomsky$范式文法.

例1 取$G_1 =< \{S_1, S_2, S_3, A, B\}, \{a, b, \circledast\}, P, S_1 >$,其中$P$由以下产生式组成:

$$S_1 \to \circledast \quad S_2 \to S_1A \quad A \to a$$

$$S_1 \to AS_2 \quad S_3 \to S_1B \quad B \to b$$

$$S_1 \to BS_3$$

则G_1是$Chomsky$范式文法,而且不难证明:$L(G_1) = \{w \circledast w^R | w \in \{a, b\}^*\}$.

例2 取$G_2 =< \{S\}, \{a, b, \circledast\}, \{S \to \circledast, S \to aSa, S \to bSb\}, S >$, 则$G_2$不是$Chomsky$范式文法,而且不难证明:$L(G_2) = L(G_1)$.

定理4.2.1 若G为cfg且$\epsilon \notin L(G)$,则有$Chomsky$范式文法$\tilde{G} =< V_N, V_T, P, S >$满足:

1) $L(\tilde{G}) = L(G)$;

2) 若$x \in V_N \cup V_T$,则必有$w_1, w_2, w_3 \in V_T^*$使:

$$S \overset{*}{\Rightarrow}_{\tilde{G}} w_1 x w_3 \overset{*}{\Rightarrow}_{\tilde{G}} w_1 w_2 w_3.$$

证明：分以下四步来完成.

1) 因为G为cfg,所以由定理4.1.9知道,必有 $cfg\ G_1 =< V_N^{(1)}, V_T, P^{(1)}, S >$ 满足:

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

   ii) 若$A \to \alpha \in P^{(1)}$,则$\alpha \notin V_N^{(1)}$.

2) 取$V_N^{(2)} = V_N^{(1)} \cup \{X_a | a \in V_T\}$,并定义同构$h : V_N^{(1)} \cup V_T \to V_N^{(2)}$如下:

$$h(u) = \begin{cases} u & u \in V_N^{(1)} \\ X_u & u \in V_T \end{cases}$$

并令$P^{(2)} = \{A \to a \in P^{(1)} | a \in V_T\} \cup \{X_a \to a | a \in V_T\} \cup \{A \to h(\alpha) | \alpha \notin V_T$且$A \to \alpha \in P^{(1)}\}$

则$G_2 =< V_N^{(2)}, V_T, P^{(2)}, S >$为$cfg$,而且有$L(G_1) \subseteq L(G_2)$.

为了证明$L(G_2) \subseteq L(G_1)$,只须证明以下的命题:

若$A \in V_N^{(1)}$且$A \overset{*}{\Rightarrow}_{G_2} w \in V_T^*$,则$A \overset{*}{\Rightarrow}_{G_1} w$.

实际上,这可用关于派生$A \overset{*}{\Rightarrow}_{G_2} w$的派生步数$n$之第二归纳法来证明.

当$n = 1$时,命题显然为真;

假定对任意正整数$k > 1$,当$n < k$时命题皆真,当$n = k > 1$时,
$A \overset{k}{\Rightarrow}_{G_2} w$必有如下形式:

$A \Rightarrow_{G_2} C_1 \dots C_m \overset{k-1}{\Rightarrow}_{G_2} w$,即,$A \rightarrow C_1 \dots C_m \in P^{(2)}$

且每个C_j都能推出终结符串,即存在$u_1, \dots, u_m \in V_T^*$,使

$$w = u_1 \dots u_m \quad C_j \overset{k_j}{\Rightarrow}_{G_2} u_j \ \ 1 \le j \le m$$

且$k - 1 = k_1 + k_2 + \cdots + k_m$

另外,由$P^{(2)}$的定义知$A \rightarrow h^{-1}(C_1 \ldots C_m) \in P^{(1)}$,再由$h$的定义还知道有

$$h^{-1}(C_1 \ldots C_m) = h^{-1}(C_1) \ldots h^{-1}(C_m)$$

且$h^{-1}(C_j) = \begin{cases} a & C_j = X_a \text{且} a \in V_T \\ A & C_j = A \in V_N^{(1)} \end{cases}$

所以有$h^{-1}(C_j) \overset{l_j}{\Rightarrow}_{G_2} u_j$且$l_j \le k_j, 1 \le j \le m$

因 此,$A \Rightarrow_{G_1} h^{-1}(C_1) \ldots h^{-1}(C_m) \overset{*}{\Rightarrow}_{G_1} u_1 h^{-1}(C_2) \ldots h^{-1}(C_m) \overset{*}{\Rightarrow}_{G_1} \cdots \overset{*}{\Rightarrow}_{G_1} u_1 \ldots u_m$

3) 最后,取$G_3 =< V_N^{(3)}, V_T, P^{(3)}, S >$,其中

$l = max\{|\alpha||A \to \alpha \in P^{(2)}\}$

$V_N^{(3)} = V_N^{(2)} \cup \{D_\beta|\beta \in (V_N^{(2)})^*$且$2 \le |\beta| < l\}$

$P^{(3)} = \{A \to \alpha \in P^{(2)}||\alpha| \le 2\} \cup \{A \to BD_\alpha|A \to B\alpha \in P^{(2)}$且$|\alpha| \ge 2\} \cup \{D_{B_1 \ldots B_t} \to B_1 D_{B_2 \ldots B_t}|B_1, \ldots, B_t \in V_N^{(2)} 2 < t < l\} \cup \{D_{BC} \to BC|B, C \in V_N^{(2)}\}$

显然G_3为$Chomsky$范式文法,且有$L(G_2) \subseteq L(G_3)$.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 28 页 共 108 页
返回
全屏显示
关闭
退出

为了证明$L(G_3) \subseteq L(G_2)$,只须证明以下的命题:

若$A \in V_N^{(2)}$且$A \stackrel{*}{\Rightarrow}_{G_3} \;_L\; w \in V_T^*$,则$A \stackrel{*}{\longrightarrow}_{G_2} w$

对此,仍用关于派生$A \stackrel{*}{\Rightarrow}_{G_3} \;_L\; w$的派生步数$n$之第二归纳法来证明.

当$n = 1$时,命题显然为真;

假定对任意的正整数$k > 1$,当$n < k$时命题皆真,当$n = k > 1$时,按$P^{(3)}$的定义, $A \stackrel{*}{\Rightarrow}_{G_3} \;_L\; w$只有以下两种可能:

1. $A \Rightarrow_{G_2} BC \stackrel{k-1}{\Rightarrow}_{G_3} \;_L\; w$

   这时必有$w_1, w_2 \in V_T^*$,使

   $w = w_1 w_2 \quad B \stackrel{k_1}{\Rightarrow}_{G_3} \;_L\; w_1 \quad C \stackrel{k_2}{\Rightarrow}_{G_3} \;_L\; w_2$且$k - 1 = k_1 + k_2$

   从而由归纳假设得$B \stackrel{*}{\Rightarrow}_{G_2} w_1$且$C \stackrel{*}{\Rightarrow}_{G_2} w_2$

   所以$A \Rightarrow_{G_2} BC \stackrel{*}{\Rightarrow}_{G_2} w_1 C \stackrel{*}{\Rightarrow}_{G_2} w_1 w_2$

2. $A \Rightarrow_{G_3} B_1 D_{B_2 \ldots B_t} \overset{k-1}{\Rightarrow}_{G_3} \quad _L\, w$

根据2.6节中引理2.3,这时必有$w_1, w_2 \in V_T^*$,使

$$w = w_1 \tilde{w}_2 \quad B_1 \overset{k_1}{\Rightarrow}_{G_3} \quad _L\, w_1 \quad D_{B_2 \ldots B_t} \overset{\tilde{k_2}}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_2 \quad k-1 = k_1 + \tilde{k_2}$$

对任意的正整数i,当$2 \leq i < t-1$时, $D_{B_i \ldots B_t} \overset{\tilde{k_i}}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_i$必有如下形式:

$$D_{B_i \ldots B_t} \Rightarrow_{G_3} B_i D_{B_{i+1} \ldots B_t} \overset{\tilde{k_i}-1}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_i$$

从 而 由2.6节 中 的 引 理2.3知 道,必 有$w_i, \tilde{w}_{i+1} \in V_T^*$,使$\tilde{w}_i = w_i \tilde{w}_{i+1} \quad B_i \overset{k_i}{\Rightarrow}_{G_3} \quad _L\, w_i \quad D_{B_{i+1} \ldots B_t} \overset{\tilde{k}_{i+1}}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_{i+1}$且$\tilde{k_i} - 1 = k_i + \tilde{k}_{i+1}$

若$i = t-1$,则$D_{B_{t-1} B_t} \overset{\tilde{k}_{t-1}}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_{t-1}$必有如下形式:

$$D_{B_{t-1} B_t} \Rightarrow_{G_3} B_{t-1} B_t \overset{\tilde{k}_{t-1}-1}{\Rightarrow}_{G_3} \quad _L\, \tilde{w}_{t-1}$$

访问主页
标题页
第 30 页 共 108 页
返回
全屏显示
关闭
退出

从而由2.6节中的引理2.3知道,必有$w_{t-1}, w_t \in V_T^*$,使

$\tilde{w}_{t-1} = w_{t-1}\tilde{w}_t \quad B_{t-1} \overset{k_{t-1}}{\Rightarrow}_{G_3}\ _L\, w_{t-1} \quad B_t \overset{\tilde{k}_t}{\Rightarrow}_{G_3}\ _L\, \tilde{w}_t$且$\tilde{k}_{t-1} - 1 = k_{t-1} + k_t$

总结以上即得到

$w = w_1 \ldots w_t \quad B_j \overset{k_j}{\Rightarrow}_{G_3}\ _L\, w_j \quad 1 \leq j \leq t$且$k - (t-1) = k_1 + \cdots + k_t$

从而由归纳假设可得$B_j \overset{*}{\Rightarrow}_{G_2} w_j \quad 1 \leq j \leq t$

因此, $A \Rightarrow_{G_2} B_1 \ldots B_t \overset{*}{\Rightarrow}_{G_2} w_1 B_2 \ldots B_t \overset{*}{\Rightarrow}_{G_2} \cdots \overset{*}{\Rightarrow}_{G_2} w_1 \ldots w_t$

4) 根据定理4.1.5,存在$cfg\ \tilde{G} =< V_N, V_T, P, S >$满足:

i) $L(\tilde{G}) = L(G_3)$,即$L(\tilde{G}) = L(G)$;

ii) $V_N \subseteq V_N^{(3)}$且$P \subseteq P^{(2)}$;

iii) 若$x \in V_n \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$,使

$S \overset{*}{\Rightarrow} w_1 x w_3 \overset{*}{\Rightarrow} w_1 w_2 w_3$

由于G_3为$Chomsky$范式文法,根据ii),$\tilde{G}$也是$Chomsky$范式文法,所以$\tilde{G}$满足定理的全部要求.

定理4.2.2 若L为cfl且$\epsilon \notin L$,则有$Chomsky$范式文法$G = < V_N, V_T, P, S >$满足:

1) $L(G) = L$;

2) 若$x \in V_N \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$使:

$$S \stackrel{*}{\Rightarrow} w_1 x w_3 \stackrel{*}{\Rightarrow} w_1 w_2 w_3.$$

对给定的$cfl\ \ G$,求出一个$Chomsky$范式文法$\tilde{G}$使得$L(G) = L(\tilde{G})$,称为把G化为$Chomsky$范式.

例3 取$G_3 =< \{S, T, L\}, \{a, b, +, -, *, /, (,)\}, P_3, S >$,其中$P_3$由以下产生式组成:

$$\begin{array}{lll} S \to T + S & T \to L * T & L \to (S) \\ S \to T - S & T \to L/T & L \to a \\ S \to T & T \to L & L \to b \end{array}$$

则G_3不 是$Chomsky$范 式 文 法,若 分 别 用a和b代 表 数 和 标 识符,则$L(G)$就是所有算术表达式的集合.

下面把G_3化为$Chomsky$范式,分为以下几个步骤完成:

1) 删去表达式$S \to T$和$T \to L$.

2) 删去终极符.

3) 化为$Chomsky$范式.

访问主页
标题页
第 33 页 共 108 页
返回
全屏显示
关闭
退出

4.3. $Greibach$范式文法

定义4.3.1 如果$cfg\ G=<V_N,V_T,P,S>$的产生式都具有如下形式:

$$A\rightarrow a\alpha, a\in V_T, A\in V_N\text{且}\alpha\in V_N^*$$

则称G为$Greibach$范式文法.

例1 取$G_1 =< \{S, A, B\}, \{a, b, \circledast\}, P, S >$,其中$P$由以下的产生式组成:

$S \longrightarrow \circledast \qquad A \longrightarrow a \quad B \longrightarrow b$

$S \longrightarrow aSA$

$S \longrightarrow bSB$

则G_1为$Greibach$范式文法,显然$L(G_1) = \{w \circledast w^R | w \in \{a, b\}^*\}$.

例2 取$G_2 =< \{S, A, B\}, \{a, b, c\}, P, S >$,其中$P$由以下的产生式组成:

$$
\begin{array}{lll}
S \rightarrow aSb & A \rightarrow BSc & B \rightarrow AS \\
S \rightarrow AB & A \rightarrow b & B \rightarrow S \\
S \rightarrow C & A \rightarrow B &
\end{array}
$$

则G_2不是$Greibach$范式文法.

定理4.3.1 设$G =< V_N, V_T, P, S >$为cfg且$A \to \alpha B\beta \in P$,若令$G_1 =< V_N, V_T, P_1, S >$,其中:

$$P_1 = \{P \setminus \{A \to \alpha B\beta\}\} \cup \{A \to \alpha\gamma\beta | B \to \gamma \in P\}$$

则$L(G_1) = L(G)$.

证明：显然有$L(G_1) \subseteq L(G)$.为了证明$L(G) \subseteq L(G_1)$,只须证明

若$D \overset{*}{\Rightarrow}_G w \in V_T^*$,则$D \overset{*}{\Rightarrow}_{G_1} w$

对此,用关于派生$D \overset{*}{\Rightarrow}_G w$的派生步数$n$的第二归纳法进行证明：

当$n = 1$时,命题显然为真;

假定对于任意的正整数$k > 1$,当$n < k$时命题皆真,当$n = k$时,$D \overset{*}{\Rightarrow}_G w$必有如下的形式：

$D \Rightarrow_G D_1 \ldots D_t \overset{k-1}{\Rightarrow}_G w$

这时必有$u_1, \ldots, u_t \in V_T^*$,使

$$w = u_1 \ldots u_t \quad D_j \overset{k_j}{\Rightarrow}_G u_j \quad 1 \leq j \leq t$$

且$k - 1 = k_1 + \cdots + k_t$

派生$D \Rightarrow_G D_1 \ldots D_l \overset{k-1}{\Rightarrow}_G w$的第一步所用之产生式$D \rightarrow D_1 \ldots D_t$只能有以下两种可能：

1) $D \rightarrow D_1 \ldots D_t$不是$A \rightarrow \alpha B \beta$,因此, $D \rightarrow D_1 \ldots D_t \in P_1$.但由归纳假设知$D_j \overset{*}{\Rightarrow}_{G_1} u_j (1 \leq j \leq t)$,所以

$D \Rightarrow_{G_1} D_1 \ldots D_t \overset{*}{\Rightarrow}_{G_1} u_1 D_2 \ldots D_t \overset{*}{\Rightarrow}_{G_1} \cdots \overset{*}{\Rightarrow}_{G_1} u_1 \ldots u_t$

即$D \overset{*}{\Rightarrow}_{G_1} w$

2) $D \to D_1 \ldots D_t$就是$A \to \alpha B \beta$,因此,$D = A$,且有$i(1 \le i \le t)$使$D_i = B$.所以$D_i \Rightarrow_G \gamma \overset{*}{\Rightarrow}_G u_i$.如果记$\gamma = C_1 \ldots C_l$时,则必有$v_1, \ldots, v_l \in V_T^*$,使

$$u_i = v_1 \ldots v_l \quad C_p \overset{s_p}{\Rightarrow}_G v_p \quad 1 \le p \le l$$

且$k_i - 1 = s_1 + \cdots + s_l$

但由归纳法知道$D_j \overset{*}{\Rightarrow}_{G_1} u_j \quad 1 \le j \le t$

且$C_P \overset{*}{\Rightarrow}_{G_1} v_p \quad 1 \le p \le l$

所以

$$D_1 \ldots D_{i-1} \overset{*}{\Rightarrow}_{G_1} u_1 D_2 \ldots D_{i-1} \overset{*}{\Rightarrow}_{G_1} \ldots D_{i-1} \overset{*}{\Rightarrow}_{G_1} \cdots \overset{*}{\Rightarrow}_{G_1}$$

$$C_1 \ldots C_l \overset{*}{\Rightarrow}_{G_1} v_1 C_2 \ldots C_l \overset{*}{\Rightarrow}_{G_1} \cdots \overset{*}{\Rightarrow}_{G_1} v_1 \ldots v_l$$

$$D_{i+1} \ldots D_t \overset{*}{\Rightarrow}_{G_1} u_{i+1} D_{i+2} \ldots D_t \overset{*}{\Rightarrow}_{G_1} \cdots \overset{*}{\Rightarrow}_{G_1} u_{i+1} \ldots u_t$$

从而得到

$$D \Rightarrow_{G_1} D_1 \ldots D_{i-1} C_1 \ldots C_l D_{i+1} \ldots D_t \overset{*}{\Rightarrow}_{G_1} u_1 \ldots u_{i-1} v_1 \ldots v_l u_{i+1} \ldots u_t$$

亦即$D \overset{*}{\Rightarrow}_{G_1} w$

定理4.3.2 设$G=<V_N,V_T,P,S>$为cfg且$A\in V_N$,如果:$\{A\rightarrow\alpha|A\rightarrow\alpha\in P\}=\{A\rightarrow A\alpha_i|1\leq i\leq t\}\cup\{A\rightarrow\beta_j|1\leq j\leq s\}$,其中每个$\beta_j\notin A(V_N\cup V_T)^*$. 则当令:

$$G_1=<V_N\cup Z,V_T,P_1,S>$$

$$P_1=(P\setminus\{A\rightarrow A\alpha_i|1\leq i\leq t\})\cup\{A\rightarrow\beta_jZ|1\leq j\leq s\}\cup\{Z\rightarrow a_i,Z\rightarrow a_iZ|1\leq i\leq t\}$$

时,必有$L(G_1)=L(G)$.

证明：1)任取$w \in L(G)$,则有最左派生如下：

$S \Rightarrow_{G\ L} x_1A_1y_1 \Rightarrow_{G\ L} \cdots \Rightarrow_{G\ L} x_mA_my_m \Rightarrow_{G\ L} w$

依次来检查每个$A_i(1 \le i \le m)$.

i) 若$A_i \ne A$,则按P_1的定义有$x_iA_iy_i \Rightarrow_{G\ L} x_{i+1}A_{i+1}y_{i+1}$.

ii 若$A_i = A$,因为是最左派生,故有：

$x_iAy_i \Rightarrow_{G\ L} x_iA\alpha_{j_1}y_i \Rightarrow_{G\ L} \cdots \Rightarrow_{G\ L} x_iA\alpha_{j_p}\ldots\alpha_{j_1}y_i \Rightarrow_{G\ L} x_i\beta_k\alpha_{j_p}\ldots\alpha_{j_1}y_i \quad 1 \le k \le s$

从而得$x_{i+p+1}A_{i+p+1}y_{i+p+1} = x_i\beta_k\alpha_{j_p}\ldots\alpha_{j_1}y_i$

而且还有$x_iAy_i \Rightarrow_{G_1} x_i\beta_kZy_i \Rightarrow_{G_1} x_i\beta_k\alpha_{j_p}Zy_i \Rightarrow_{G_1} \cdots \Rightarrow_{G_1} x_i\beta_k\alpha_{j_p}\ldots\alpha_{j_1}y_i$

所以$x_iA_iy_i \stackrel{*}{\Rightarrow}_{G_1} x_{i+p+1}A_{i+p+1}y_{i+p+1}$

这表明,必有$S \stackrel{*}{\Rightarrow}_{G_1} w$.

2) 任取$w \in L(G_1)$,则有最右派生如下:

$$S \Rightarrow_{G_1\ R} y_1A_1x_1 \Rightarrow_{G_1\ R} \cdots \Rightarrow_{G_1\ R} y_mA_mx_m \Rightarrow_{G_1\ R} w$$

也依次来检查每个$A_j(1 \le j \le n)$.

若$A_j \neq A$,按P_1定义有$y_jA_jx_j \Rightarrow_G y_{j+1}A_{j+1}x_{j+1}$.

若$A_j = A$,因为是最右派生,故有

$$y_jAx_j \Rightarrow_{G_1} y_j\beta_kZx_j \Rightarrow_{G_1} y_j\beta_k\alpha_{j_1}Zx_j \Rightarrow_{G_1} \cdots \Rightarrow_{G_1} y_j\beta_k\alpha_{j_l}\ldots\alpha_{j_1}x_j \quad 1 \le k \le s$$

从而得$y_{j+l+1}A_{j+l+1}x_{j+l+1} = y_j\beta_k\alpha_{j_l}\ldots\alpha_{j_1}x_j$

所以$y_jA_jx_j \stackrel{*}{\Rightarrow}_G y_{j+l+1}A_{j+l+1}x_{j+l+1}$

这表明必有$S \stackrel{*}{\Rightarrow}_G w$.

定理4.3.3 对每个cfg G且$\epsilon \notin L(G)$,皆有$Greibach$范式文法$\tilde{G} =< V_N, V_T, P, S >$满足:

1) $L(\tilde{G}) = L(G)$;

2) 若$x \in V_N \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$使:

$$S \stackrel{*}{\Rightarrow}_G w_1 x w_3 \stackrel{*}{\Rightarrow}_G w_1 w_2 w_3.$$

证 明： 因为G为cfg,所以由定理4.2.1知道,存在$Chomsky$范式$G_1 =< \{A_1, ..., A_m\}, V_T, P_1, A_1 >$,使$L(G_1) = L(G)$.以下的证明分两步来完成.

1) 通过引入新变量$Z_1, ..., Z_m$,把P_1中产生式改造为如下形式:

$A_{i'} \to a\alpha \quad a \in V_T, \alpha \in \{A_1, ..., A_m, Z_1, ..., Z_m\}^*$

或$A'_i \to A'_j\gamma, j > i, \gamma \in \{A_1, ..., A_m, Z_1, ..., Z_m\}^+$

其中$A'_1 = Z_m, ..., A'_m = Z_1, A'_{m+1} = A_1, ..., A'_{2m} = A_m$

对此,可以用关于i的归纳法,对A_i逐个进行改造.

当$i = 1$时,因无变量A_0,故可认为对A_1已改造成功.

对任意的正整数$k > 0$,假定$i < k$时,对每个A_i皆已改造成功,当$i = k$时, $A_k \to \alpha$只有以下三种类型:

$A_k \to a \quad a \in V_T$

$A_k \to A_iA_l, i < k, 1 \leq l \leq m$

$A_k \to A_jA_t, j \geq k, 1 \leq t \leq m$

其中第一种类型(即$A_k \to a$型)产生式已合要求,所以只需改造后两种类型的产生式.

$A_k \to a\alpha \quad a \in V_T$

$A_k \to A_{i'}\beta \ \ i < i' < k$

$A_k \to A_j\beta \quad j \geq k$

这里$\alpha \in (V_N \cup \{Z_1, \ldots, Z_{k_1}\})^*, \beta \in V_N(V_N \cup \{Z_1, \ldots, Z_{k_1}\})^*$

再重复上述过程,至多经过$k-1$次,就会使所有以A_k为左端的产生式变为

$A_k \to a\alpha \quad a \in V_T$

$A_k \to A_k\beta$

$A_k \to A_j\beta \quad j > k$

这里$\alpha \in (V_N \cup \{Z_1, \ldots, Z_{k_1}\})^*, \beta \in V_N(V_N \cup \{Z_1, \ldots, Z_{k_1}\})^*$

应用定理4.3.2,并引进新变量Z_k,就使所有以A_k为左端的产生式变为

$A_k \to a\alpha \quad a \in V_T$

$A_k \to A_j\beta \ \ j > k$

$A_k \to a\alpha Z_k \quad j > k$

$A_k \to A_j\beta Z_k \quad j > k$

以及$Z_k \to \beta$

$Z_k \to \beta Z_k$

2) 标准化

根据上步的结果,对每个$i(1 \le i \le 2m)$,所以以A'_i为左端的产生式只有以下两种形式

$A'_i \rightarrow a\alpha \quad a \in V_T$

$A'_i \rightarrow A'_j\beta \quad j > i$

因此,以A_{2m}为左端的产生式就只能有以下一种形式

$A'_{2m} \rightarrow a\alpha \quad a \in V_T$

下面按i的归纳法,对A'_{2m-i}逐个进行标准化

当$i = 0$时,A'_{2m}的标准化已经完成.

假定对任意的自然数$k > 0$,当$i < k$时, A'_{2m-1}的标准化都已完成,那么只要对每个$A'_{2m-i} \rightarrow A'_j\beta(j > 2m-k)$应用一次定理4.3.1(取$B = A'_j$)后,则所有以$A'_{2m-k}$为左端的产生式也都变成为如下形式

$A'_{2m-k} \rightarrow a\beta' \quad a \in V_T$

这就完成了对A'_{2m-i}的标准化.

例3 取$G=<\{A_1,A_2,A_3\},\{a,b,c\},P,A_1>$,其中$P$由以下产生式组成:

$$
\begin{aligned}
&A_1 \rightarrow aA_1b \quad A_1 \rightarrow A_2A_3 \quad A_1 \rightarrow c \\
&A_2 \rightarrow A_3A_1c \quad A_2 \rightarrow b \\
&A_3 \rightarrow A_2A_1
\end{aligned}
$$

求G的$Greibach$范式.

下面介绍$Greibach$范式文法的另一种求法:
设$G =< V_N, V_T, P, S >$为cfg且$\epsilon \notin L(G)$,P中产生式可分为以下两类:

1) $A \rightarrow a\alpha \;\; a \in V_T$且$\alpha \in (V_N \cup V_T)^*$
2) $A \rightarrow B\beta \;\; B \in V_N$且$a \in (V_N \cup V_T)^*$

以任意次序给V_N中元素进行编号,不妨设为$V_N = \{S_1, ..., S_n\}$,则可构造一个正规表达式方程组如下:

$$(S_1, \cdots, S_n) = (S_1, \cdots, S_n)\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} + (a_1, \cdots, a_n) \qquad (*)$$

使:

1) 对每个字$x \in a_i$,则$S_i \rightarrow x$为一个1)型产生式.
2) 对每个字$y \in \overline{S_1 A_{1i} + \cdots + S_n A_{ni}}$,则$S_i \rightarrow y$为一个2)型产生式.

正规表达式方程组(*)的解为:

$$(S_1,\cdots,S_n)=(a_1,\cdots,a_n)\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}^*$$

$$=(a_1,\cdots,a_n)+(a_1,\cdots,a_n)\begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \cdots & \cdots & \cdots \\ B_{n1} & \cdots & B_{nn} \end{pmatrix}$$

其中 $\begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \cdots & \cdots & \cdots \\ B_{n1} & \cdots & B_{nn} \end{pmatrix}=\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}^*=$

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}+\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}\begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \cdots & \cdots & \cdots \\ B_{n1} & \cdots & B_{nn} \end{pmatrix}$$

由此可见,对每个B_{ij}产生式,其右端第一个符号不是终极符, 则必有某个$S_k(1\le k\le n)$,所以只要对它应用一次定理4.3.1(取$B=S_k$),即可化为1)型产生式.

对在产生式右端非第一符号处出现的每个终极符a，都用一个相应的新变量X_a代替，并增加一个新产生式$X_a\to a$.

这样就得到了一个$Greibach$范式文法G'使$L(G')=L(G)$.对G'再应用一次定理4.1.5,即可获得所要之$Greibach$范式文法.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 49 页 共 108 页
返回
全屏显示
关闭
退出

例4 对例3中所给之$cfg\ G$,其正规表达式方程组为

$$(A_1, A_2, A_3) = (A_1, A_2, A_3)\begin{pmatrix} \emptyset & \emptyset & \emptyset \\ A_3 & \emptyset & A_1 \\ \emptyset & A_1 c & \emptyset \end{pmatrix} + (aA_1b + c, b, \emptyset)$$

该方程组的解为

$$(A_1, A_2, A_3) = (aA_1b + c, b, \emptyset) + (aA_1b + c, b, \emptyset)\begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

$$\begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} = \begin{pmatrix} \emptyset & \emptyset & \emptyset \\ A_3 & \emptyset & A_1 \\ \emptyset & A_1 c & \emptyset \end{pmatrix} + \begin{pmatrix} \emptyset & \emptyset & \emptyset \\ A_3 & \emptyset & A_1 \\ \emptyset & A_1 c & \emptyset \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

访 问 主 页
标 题 页
第 51 页 共 108 页
返 回
全 屏 显 示
关 闭
退 出

因此,$B_{11}=B_{12}=B_{13}=\emptyset$

从而得到

$$(A_1,A_2,A_3,A_4,A_5)=(aA_1A_4+bB_{21}+c,b+bB_{22},bB_{23},b,c)$$

$$\begin{pmatrix} B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}=\begin{pmatrix} bB_{23} & \emptyset & aA_1A_4+bB_{21}+c \\ \emptyset & aA_1A_4A_5+bB_{21}A_5+cA_5 & \emptyset \end{pmatrix}$$
$$\begin{pmatrix} \emptyset & aA_1A_4+bB_{21}+c \\ aA_1A_4A_5+bB_{21}A_5+cA_5 & \emptyset \end{pmatrix}\begin{pmatrix} B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

这就是所要的$Greibach$范式文法.

定理4.3.4 若L为cfl且$\epsilon \notin L$,则有$Greibach$范式文法$\tilde{G} =< V_N, V_T, P, S >$满足:

1) $L(G) = L$;

2) 若$x \in V_N \cup V_T$,则有$w_1, w_2, w_3 \in V_T^*$使:

$$S \stackrel{*}{\Rightarrow} w_1 x w_3 \stackrel{*}{\Rightarrow} w_1 w_2 w_3.$$

4.4. 迭代定理

给定一个语言L,如果L是cfl则可以构造一个$cfg\ \ G$使$L(G)=L$,从而证明L确实是一个cfl.如果L不是cfl,到目前尚无方法加以证明.第三章曾用$Pumping$引理有效地解决了许多非rl的证明问题.能否把第三章的$Pumping$引理加以推广,用以证明给定语言不是cfl的问题呢?

本节首先证明迭代定理，然后给出它的两个重要特例,即著名的$Ogden$引理和$uvwxy$字定理,它们可以看作是$Pumping$引理在cfl情况下的推广，再接着描述用上述定理证明一个给定的语言不是cfl的具体做法.此外，还要以$uvwxy$字定理为工具，讨论cfg产生的语言的有穷性的可解性问题，以及先天歧义文法的存在问题。

定理4.4.3 (uvwxy字定理)对每个$cfl\ \ L$,都有正整数p和q,使得当$z \in L$且$|z| > p$时,皆有:$z = uvwxy$,满足:

i) $vx \neq \epsilon$;

ii) $|vwx| \leq q$;

iii) 若$k \in \mathbb{N}$,则$uv^k wx^k y \in L$.

证明： 设L为一个cfl,并且不妨假设$\epsilon \notin L$,根据定理4.2.2,从而存在$Chomsky$范式文法$G =< V_N, V_T, P, S >$使得$L = L(G)$.对于任意得$z \in L$,当k是z的派生树的高度时(即,$k = D(T_z)$),则有：

$$|z| \leq 2^{k-1}$$

实际上,仅当z的派生树是满二叉树时,等号才成立,$|z| = 2^{k-1}$,其它时候均有$|z| < 2^{k-1}$.当z的派生树$D(T_z)$是满二叉树时,此时每一条通路的长度均为k,并且每一条通路上恰有k个用非终结符号标记的非叶子结点,1个用终结符号标记的叶子结点.

现在取$p = 2^{|V_N|} = 2^{|V_N|+1-1}$, $z \in L$, $|z| \geq N$,此时z的语法树中至少有一条长度大于等于$|V_N| + 1$的通路,该通路上的非叶子结点的个数大于等于$|V_N| + 1$.取该树中最长的一条通路p,p中的非叶子结点的数目大于等于$|V_N| + 1$,它们的标记都是非终结符号.由于$|V_N| + 1 \geq |V_N|$,这些非叶子结点中必定有不同的结点标有相同的语法变量. 现在取通路p中最接近叶子的两个结点v_1, v_2,它们都标记有相同的非终结符号A. 为了明确起见,不妨设v_1是v_2的祖先结点.显然,v_1到叶结点的通路长度小于等于$|V_N| + 1$.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 55 页 共 108 页
返回
全屏显示
关闭
退出

设结点v_1左边的所有叶子结点的标记从左到右构成的字符串为u.

以结点v_1为根的子树中,结点v_2左边的所有叶子结点的标记从左到右构成的字符串为x.

结点v_1右边的所有叶子结点的标记从左到右构成的字符串为y.

此时有:$z = uvwxy$

注意到以v_1为根的A的子树的最大路长小于等于$|V_N| + 1$,所以v_1的结果vwx满足

$$|vwx| \leq 2^{|V_N|+1-1} = 2^{|V_N|} = N.$$

再注意到G是cfg,而v_1的后代v_2标记为变量A,所以$|vx| \geq 1$.

此时有

$S \overset{*}{\Rightarrow} uAy \overset{+}{\Rightarrow} uvAxy \overset{+}{\Rightarrow} uvwxy$

显然,对于$i = 0, 1, 2, 3, \ldots, A \overset{*}{\Rightarrow} v^iAx^i \overset{*}{\Rightarrow} v^iwx^i$

所以$S \overset{*}{\Rightarrow} uAy \overset{+}{\Rightarrow} uv^iAx^iy \overset{+}{\Rightarrow} uv^iwx^iy$

这就是说,对于$i = 0, 1, 2, 3, \ldots, uv^iwx^iy \in L$,定理得证.

定理4.4.2 (Ogden引理)对每个$cfl\ \ L$,都有一个正整数n_L使得当$z \in L$且$|z| > n_L$时，对任意正整数$P(n_L \leq P \leq |z|)$及z的任意P个指定位置$i_1, i_2, \cdots, i_p(1 \leq i_1 < i_2 < \cdots < i_p \leq |z|)$,皆有:$z = uvwxy$,满足:

i) v和x至少有一个指定的位置.

ii) vwx至多含有n_L个指定的位置.

iii) 若$k \in \mathbb{N}$,均有$uv^kwx^ky \in L$.

证明： 设L为一个cfl.不妨假设$\epsilon \notin L$,从而存在$Chomsky$文法$G=<V_N,V_T,P,S>$使得$L=L(G)$,取$n_L=2^{|V_N|}=2^{|V_N|+1-1}$.

设$z\in L$,并且z中指定位置的数目不少于n_L.定义z的语法树中满足下列条件的非叶子结点为分支点：该结点有两个子结点,并且它的这两个子结点具有指定位置的后代.

按照如下方式构造通路p:先将树根放入p.设r是刚刚被放入通路p中的结点,如果r是叶子,则完成p的构造;如果r只有一个子结点具有指定位置的后代,则将这个子结点放入p;如果r 的两个子结点都有具有指定位置的后代,则将含有较多指定位置的后代的子结点加入p.

重复上述过程,直到结束.

显然,p中至少有n_L+1个分支点,在这些分支结点中,至少有两个不同的结点用相同的非终结符号标记.设结点v_1,v_2是p中最接近叶子的两个都标有相同变量A的分支点.为了明确起见,不妨设v_1是v_2的祖先结点.

与$uvwxy$字定理的证明类似,设:

设结点v_1左边的所有叶子结点的标记从左到右构成的字符串为u.

以结点v_1为根的子树中,结点v_2左边的所有叶子结点的标记从左到右构成的字符串为x.

结点v_1右边的所有叶子结点的标记从左到右构成的字符串为y.

此时有:$z = uvwxy$

注意到通路p在v_1子树部分所含的分支点的个数小于等于$|V_N| + 1$,所以v_1的结果vwx所含的指定位置最多为n_L个.再注意到v_1是分支点,并且v_2是它的后代之一,所以,vx中至少有一个指定位置.

此时有:

$S \stackrel{*}{\Rightarrow} uAy \stackrel{+}{\Rightarrow} uvAxy \stackrel{+}{\Rightarrow} uvwxy$

显然,对于$i = 0, 1, 2, 3, \ldots, S \stackrel{*}{\Rightarrow} uAy \stackrel{+}{\Rightarrow} uv^iAx^iy \stackrel{+}{\Rightarrow} uv^iwx^iy$

这就是说,对于$i = 0, 1, 2, 3, \ldots, uv^iwx^iy \in L$,定理得证.

定理4.4.1 (迭代定理)对每个$cfl\ \ L$,都有一个正整数n_L使得当$z \in L$且$|z| > n_L$时，对任意正整数$P(n_L \leq P \leq |z|)$及z的任意P个指定位置$i_1, i_2, \cdots, i_p (1 \leq i_1 < i_2 < \cdots < i_p \leq |z|)$,皆有:$z = uvwxy$,满足:

i) 或者u, v和w都各至少含有一个指定的位置，或者w, x和y都各至少有一个指定的位置.

ii) vwx至多仅含n_L个指定的位置.

iii) 对任意$k \in \mathbb{N}$,均有$uv^k wx^k y \in L$.

下面通过一些例子来说明用迭代定理,Ogden引理和$uvwxy$字定理证明一个给定的语言不是cfl的具体作法.

例1 试证$L_1 = \{a^n b^n c^n | n \geq 1\}$不是$cfl$.

证明： 对任意的正整数p和q,显然$a^{p+q+1}b^{p+q+1}c^{p+q+1} \in L_1$,所以当

$a^{p+q+1}b^{p+q+1}c^{p+q+1} = uvwxy, vx \neq \epsilon$,且$|vwx| \leq q$

时,必有$vx \in \{a,b\}^+ \cup b, c^+$.因此$uwy \notin L_1$.从而由$uvwxy$字定理知道,$L_1$不是$cfl$.

访问主页
标题页
第 61 页 共 108 页
返回
全屏显示
关闭
退出

例2 试证$L_2 = \{a^i b^j c^i d^j | i, j \geq 1\}$不是$cfl$.

证明：对任意的正整数p和q,令$n = 1 + p + q$.显然$a^n b^n c^n d^n \in L_2$.所以当

$a^n b^n c^n d^n = uvwxy, vx \neq \epsilon$,且$|vwx| \leq q$

时,必有$vx \in \{a, b\}^+ \cup b, c^+ \cup c, d^+$.因此$uwy \notin L_2$.从而由$uvwxy$字定理知道,$L_1$不是$cfl$.

例3 试证$L_3=\{a^ib^jc^k|i\neq j\neq k\}$不是$cfl$.

证明： 对任意的正整数d令$n=1+d, m=n+n!$,则显然有$a^nb^mc^{m+m!}\in L_3$,取$a^nb^mc^{m+m!}$的前n个位置即n个a所处的位置为指定位置.这时,如果$a^nb^mc^{m+m!}=uvwxy$满足以下条件:

i) v或x至少含有一个指定的位置;

ii) vwx至多含有d个指定位置;

则必有$k_0\in\mathbb{N}$使$uv^{k_0}wx_{k_0}y\notin L_3$.下面来证明这点:

首先,若v或x含有两种以上的符号,则$uv^2wx^2y\notin L_3$.因此$v,x\in a^*\cup b^*\cup c^*$.再根据条件i),这时$v$和$x$ 只能有以下五种可能:

1) $v=\epsilon$

根据条件i),这时必有正整数$p<n$,使得$x=a^p$.因为$p|n!$,所以有正整数q, 使得$n!=pq$.但这时uwy中恰含$n-p$个a,故有$uv^{q+1}wx^{q+1}y=a^{n-p}a^{p(q+1)}b^mc^{m+m!}=a^mb^mc^{m+m!}\notin L_3$

2) $x=\epsilon$

根据条件i),这时必有正整数$p<n$,使$v=a^p$.因为$p|n!$,所以有正整数q使$n!=pq$.但这时uwy恰含$n-p$个a,故有

$$uv^{q+1}wx^{q+1}y=a^{n-p}a^{p(q+1)}b^mc^{m+m!}=a^mb^mc^{m+m!}\notin L_3$$

3) $v, x \in \{a\}^+$

根据条件i),这时必有正整数p_1和p_2,使得$v = a^{p_1}, x = a^{p_2}$且$1 \le p_1 + p_2 < n$.因为$(p_1 + p_2)|n!$,所以有正整数q使$n! = (p_1 + p_2)q$.但这时uwy中有$n - p_1 - p_2$个a,故有

$$uv^{q+1}wx^{q+1}y = a^{n-p_1-p_2}a^{p_1(q+1)}b^m c^{m+m!} = a^m b^m c^{m+m!} \notin L_3$$

4) $x \in \{b\}^+$

根据条件i),这时$v \in \{a\}^+$,因此有正整数$p_1 < n$使$v = a^{p_1}$.但$x \in \{b\}^+$,所以有正整数$p_2 \le m$使$x = b^{p_2}$.由$p_2|m!$,故有正整数q使$m! = p_2 q$.但这时uwy中恰含$n - p_1$个a和$m + m! - p_2$个c,所以有

$$uv^{q+1}wx^{q+1}y = a^{n-P_1}a^{p_1(q+1)}b^{m-p_2}b^{p_2(q+1)}c^{m+m!} = a^{n+p_1q}b^{m+m!}c^{m+m!}$$

5) $x \in \{c\}^+$

根据条件i),这时$v \in \{a\}^+$,因此有正整数$p_1 < n$使$v = a^{p_1}$.但$x \in \{c\}^+$,所以有正整数$p_2 \le m + m!$使$x = c^{p_2}$.因为$p_1|n!$,所以有正整数q使$n! = p_1 q$.但这时uwy中恰含有$n - P_1$个a和$m + m! - P_2$个c,所以有

$$uv^{q+1}wx^{q+1}y = a^{n-p_1}a^{p_1(q+1)}b^m c^{m+m!-p_2 q}c^{p_2(q+1)} = a^m b^m c^{m+m!+p_2 q} \notin L_3$$

总之,皆有自然数$k_0 = q + 1$,使$uv^{k_0}wx^{k_0}y \notin L_3$.从而再根据$Ogden$引理知道,$L_3$不是上下文无关语言.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 66 页 共 108 页
返回
全屏显示
关闭
退出

例4 试证$L_4 = \{a^p | p$为素数$\}$不是cfl.

证明： 因为素数的个数是无穷的,所以对任意的正整数n,都有素数$p > 3n$,对字$a^p \in L_4$,若$a^p = uvwxy$,使$vx \neq \epsilon$,则当记$s = |vx|$及$t = |uwy|$时,必有

$p = s + t$且$uv^k wx^k y = a^{t+ks}$

但因$vx \neq \epsilon$,所以$s > 0$.因此$p(s+1)$不是素数,从而得:

$uv^{p+1}wx^{p+1}y = a^{t+(p+1)s} = a^{p(s+1)} \notin L_4$

根据$uvwxy$字定理知道,L_4不是cfl.

定理4.4.4 设L为cfl,则有常数p和q存在,使

$${}^{\#}L=\infty\text{当且仅当有}z\in L\text{使}p<|z|\leq p+q.$$

证明:因为L为cfl,故有$Chomsky$范式文法
$G=<\{A_1,\ldots,A_n\},V_T,P,S>$使$L(G)=L$.假定$p$和$q$为由$uvwxy$字定理所决定的正整数(实际上,可取$p=2^{\#V_N}$和$q=2^{\#V_N+1}$),那么,当$z\in L$且$p<|z|\leq p+q$时,就有

$z=uvwxy\ \ vx\neq\epsilon\ \ |vwx|\leq q$

使$uv^iwx^iy\in L\ \ i\in L$,所以L必为无穷集.

若L为无穷集,则必有$z\in L$,使$|z|>p$:

1) 若$|z|\leq p+q$,则无须再证.

2) 若$|z|>p+q$,取$z'=uwy\in L$,由$1\leq|vx|\leq|vwx|\leq q$知道,$p<|z'|<|z|$.对z'再重复上述论述,至多经过$(|z|-(p+q))$次,即可得到一个$\bar{z}\in L$使$p<|\bar{z}|\leq p+q$.

定理4.4.5 下面关于cfg的判定问题是可解的：

对任意的cfg $G =< V_N, V_T, P, S >$,是否$\#L(G) < \infty$?

证明：对cfl $L(G)$,根据定理4.4.4,有正整数p和q存在,使:

$^{\#}L(G) = \infty$当且仅当有$z \in L(G)$,使$p < |z| \leq p + q$.

当令

$T_0 = \{S\}$

$T_{m+1} = \{\beta | 有\alpha \in T_m 使\alpha \overset{i}{\Rightarrow} \beta, 1 \leq i \leq 2^{p+q}, \beta \in V_T^* 或|\alpha| < |\beta|\}, m = 0, 1, 2, \ldots.$

定理4.4.6 对每个cfg $G =< V_N, V_T, P, S >$,皆有cfg $\overline{G} =< \overline{V_N}, V_T, \overline{P}, S >$满足:

i) $L(\overline{G}) = L(G)$;

ii) 若$A \in V_N$且$A \neq S$,则$L(\overline{G}_A)$为无穷集;

证明: 令

$V_N' = \{A \in V_N | A \neq S$且$^{\#}(L(G_A)) < \infty\}$

$\overline{V}_N = V_N \setminus V_N'$

$\overline{P} = \{A \rightarrow h(B_1) \ldots h(B_t) | A \in \overline{V}_N$且$A \rightarrow B_1 \ldots B_t \in P\}$

其中关系(或多值函数)h定义如下:

$h = (\cup_{A \in V_N'} \{A\} \times L(G_A)) \cup \{< A, A > | A \in \overline{V}_N\} \cup \{< a, a > | a \in V_T\}$

则$\overline{G} =< \overline{V}_N, V_T, \overline{P}, S >$就是合乎定理要求之$cfg$.

实际上,由$\overline{P}$的定义,显然有$L(\overline{G}) \subseteq L(G)$.为了证明$L(G) \subseteq L(\overline{G})$,只须证明以下的命题即可:

若$A \in \overline{V}_N$且$A \overset{*}{\Rightarrow}_G w \in V_T^*$,则$A \overset{*}{\Rightarrow}_{\overline{G}} w$

对此用关于派生$A \overset{*}{\Rightarrow}_G w$的派生步数$n$之第二归纳法来证.

当$n = 0$时,命题显然为真;

假定对任意的正整数$k \geq 1$,当$n < k$时命题皆真.

当$n = k$时,$A \overset{*}{\Rightarrow}_{\overline{G}} w$必有如下形式:

$A \Rightarrow_G B_1 \ldots B_t \overset{*}{\Rightarrow}_G w, B_1, \ldots B_t \in V_N \cup V_T$

因此,必有$u_1, \ldots, u_t \in V_T^*$,使

$w = u_1 \ldots u_t \quad B_i \overset{*}{\Rightarrow}_G u_i \quad 1 \leq i \leq t$

从而由h的定义及归纳假设可得

$h(B_i) = u_i$或$h(B_i) \overset{*}{\Rightarrow}_G u_i \quad 1 \leq i \leq t$

因此$A \Rightarrow_{\overline{G}} h(B_1) \ldots h(B_t) \overset{*}{\Rightarrow}_{\overline{G}} u_1 \ldots u_t$

即$A \overset{*}{\Rightarrow}_{\overline{G}} w$

访问主页
标题页
第 70 页 共 108 页
返回
全屏显示
关闭
退出

例5 取$G =< \{S, A, B\}, \{a, b, c, d\}, P, S >$,其中$P$由以下产生式组成:

$$\begin{array}{ll} S \to aBSA & S \to bAaaB \\ A \to ab & A \to aab \\ B \to c & B \to dc \end{array}$$

因为$L(G_A) = \{ab, aab\}, L(G_B) = \{c, dc\}$,所以$V'_N = \{A, B\}, V_N = \{S\}$.
这时二元关系h为:

$$h = \{< S, S >, < A, ab >, < A, aab >, < B, c >, < B, dc >, < a, a >, < b, b >, < c, c >, < d, d >\}$$

此时$\overline{P}$由以下的产生式组成:

1) 由$S \to h(a)h(B)h(S)h(A)$得:

$$\begin{array}{ll} S \to acSab & S \to acSaab \\ S \to adcSab & S \to adcSaab \end{array}$$

2) 由$S \to h(b)h(A)h(a)h(a)h(B)$得:

$$\begin{array}{ll} S \to babaac & S \to babaadc \\ S \to baabaac & S \to baabaadc \end{array}$$

4.5. 自嵌套特性

定义4.5.1 设$G=< V_N, V_T, P, S >$为cfg,若有$A \in V_N$及$a_1, a_2 \in (V_N \cup V_T)^*$使得$A \overset{*}{\Rightarrow} a_1Aa_2$,就称$G$为自嵌套的,否则,称$G$为非自嵌套的.

正是cfg的自嵌套特性,才导致了cfl的$uvwxy$字定理,但这并不是说,由自嵌套的cfg G产生的语言$L(G)$,一定不是rl.

例1 取$G=<\{S\},\{a,b\},P,S>$,其中P由以下产生式组成:

$$
\begin{array}{ll}
S \rightarrow aSa & S \rightarrow aS \\
S \rightarrow bS & S \rightarrow a \\
S \rightarrow b &
\end{array}
$$

则G为自嵌套的cfg,但$L(G)=\{a,b\}^{+}$却是正规语言.

证明： $cfg\ \ G$是非自嵌套的,存在非自嵌套的$Greibach$范式文法$G' =< V'_N, V_T, P', S >$,使$L(G') = L(G)$.

令$m =^{\#} V'_N$

及$l = max\{|\alpha||A \rightarrow \alpha \in P'\}$

若$S \overset{*}{\Rightarrow}_{G'\ \ L} x\alpha, x \in V_T^*, \alpha \in (V'_N)^*$

则$|\alpha| \leq ml$.

否则G'必为自嵌套的,再令:

$\overline{G} =< \overline{V}_N, V_T, \overline{P}, [S] >$

其中

$\overline{V}_N = \{[\alpha]|\alpha \in (V'_N)^*$且$1 \leq |\alpha| \leq ml\}$

$\overline{P} = \{[A\beta] \rightarrow a[\alpha\beta]|A \rightarrow a\alpha \in P', \beta \in (V'_N)^*$且$|\alpha\beta| \leq ml\}$

则$\overline{G}$为rg.

如果$x \in V_T^*, \alpha, \beta \in (V'_N)^*$,且$|\alpha\beta| \leq ml$,则显然有:

$xA\beta \Rightarrow_{G'} xa\alpha\beta$当且仅当$x[A\beta] \Rightarrow_{\overline{G}} xa[\alpha\beta]$

所以

$S \overset{*}{\Rightarrow}_{G'\ \ L} w \in V_T^*$当且仅当$S \overset{*}{\Rightarrow}_{\overline{G}} w$

访问主页
标题页
第 74 页 共 108 页
返回
全屏显示
关闭
退出

4.6. 广义cfg

对$cfg\ G =< V_N, V_T, P, S >$,$P$中不能含有形如$A \to \epsilon$的产生式,即使对修改后的$cfg\ G$,当$A \neq S$时,$P$中仍不能含有形如$A \to \epsilon$的产生式,本节将推广$cfg$的概念,解除上述限制.

定义4.6.1 若文法$G =< V_N, V_T, P, S >$的P中之产生式都具有如下形式

$A \to \alpha \quad A \in V_N$ 且$\alpha \in (V_N \cup V_T)^*$,就称$G$为广义$cfg$.

引理4.6.1 若$G=<V_N,V_T,P,S>$为广义cfg,则有一个算法,它能判断对任意的$A\in V_N$,是否有$A\overset{*}{\Rightarrow}\epsilon$.

证明: 令

令

$T_0=\{B|B\to\epsilon\in P\}$

$T_{m+1}=\{C|\text{有}B_1,\ldots,B_t\in T_m\text{使}C\to B_1\ldots B_t\in P\}\cup T_m(m=0,1,2,\ldots)$

时,显然有

$T_m\subseteq T_{m+1}\subseteq V_N(m=0,1,2,\ldots)$

由于V_N为有限集,所以必有n为自然数,使$T_n=T_{n+1}$.从而知道:

1) 若$A\in T_n$,则$A\overset{*}{\Rightarrow}\epsilon$;

2) 若$A\notin T_n$,则$A\overset{*}{\Rightarrow}\epsilon$不成立;

定理4.6.1 L为cfl,当且仅当有广义cfg G,使$L(G)=L$.

证明: 因为每个cfg都是广义cfg,所以必要性是显然的.下面来证充分性.设$G=<V_N,V_T,P,S>$为广义cfg,且$L(G)=L$.任取$S_1\notin V_N\cup V_T$,令

$V'_N=\{A\in V_N|A\overset{*}{\Rightarrow}\epsilon\}$

$\overline{P}=\{A\rightarrow h(B_1)\ldots h(B_t)|A\rightarrow B_1\ldots B_t\in P$且$h(B_1)\ldots h(B_t)\neq\epsilon\}\cup\{S_1\rightarrow\epsilon|S\in V'_N\}\cup\{S_1\rightarrow\alpha|S_1\rightarrow\alpha\in P$且$\alpha\neq\epsilon\}$

其中二元关系h定义如下:

$h=\{<a,a>|a\in V_T\}\cup\{<A,A>|A\in V_N\}\cup\{<A,\epsilon>|A\in V'_N\}$

则$\overline{G}=<V_N\cup\{S_1\},V_T,\overline{P},S_1>$为$cfg$,而且还有$L(\overline{G})\subseteq L(G)$.

为了证明$L(G) \subseteq L(\overline{G})$,只须证明

若$A \in V_N$且$A \overset{*}{\Rightarrow}_G w \in V_T^+$,则$A \overset{*}{\Rightarrow}_{\overline{G}} w$

用关于派生$A \overset{*}{\Rightarrow}_G w$的派生步数$n$之第二归纳法证明如下:

当$n = 1$时,命题显然为真;

假定对任意正整数$k > 1$,当$n < k$时命题皆真;

当$n = k$时,$A \overset{k}{\Rightarrow}_G w$必有如下的形式:

$A \Rightarrow_G B_1 \ldots B_t \overset{k-1}{\Rightarrow}_G w$

这时,有$u_1, \ldots, u_t \in V_T^*$使

$w = u_1 \ldots u_t \;\; B_j \overset{k_j}{\Rightarrow}_G u_j \;\; 1 \le j \le t$

且$k - 1 = k_1 + \cdots + k_t$

若令

$$h(B_j) = \begin{cases} B_j & \text{当} u_j \neq \epsilon \\ \epsilon & \text{否则} \end{cases} \; j = 1, 2, \ldots, t$$

则有$h(B_j) = u_j$或$h(B_j) \overset{k_j}{\Rightarrow}_G u_j \quad 1 \leq j \leq t$

从而由归纳假设及$k_j \leq k-1 (j = 1, \ldots, t)$得到

$h(B_j) \overset{*}{\Rightarrow}_{\overline{G}} u_j \quad 1 \leq j \leq t$

所以$A \Rightarrow_{\overline{G}} h(B_1)\ldots h(B_t) \overset{*}{\Rightarrow}_{\overline{G}} u_1 h(B_2)\ldots h(B_t) \overset{*}{\Rightarrow}_{\overline{G}} \cdots \overset{*}{\Rightarrow}_{\overline{G}} u_1 \ldots u_{t-1} h(B_t) \overset{*}{\Rightarrow}_{\overline{G}} u_1 \ldots u_t$

亦即$A \overset{*}{\Rightarrow}_{\overline{G}} w$.

例1 取$G=<\{S,A,B\},\{0,1\},P,S>$,其中$P$由以下产生式组成:

$$\begin{array}{lll} S\rightarrow AB & A\rightarrow 1 & B\rightarrow \epsilon \\ & A\rightarrow BB & B\rightarrow 0 \\ & A\rightarrow BSB & B\rightarrow 0A1 \end{array}$$

试求一个cfg $\overline{G}$,使$L(\overline{G})=L(G)$.

4.7. 特殊类型的cfg和cfl

本节将介绍几种常见的特殊类型的cfg和cfl,为方便起见,总假定$G =< V_N, V_T, P, S >$为广义cfg.

定义4.7.1 若P中产生式都具有以下形式:

$$A \rightarrow u\text{或}A \rightarrow uBv \quad B \in B_N \quad uv \neq \epsilon\text{且}u, v \in V_T^*$$

就称G为线性文法,记为$lg \ \ G$.若有$lg \ \ G$,使$L = L(G)$,则称L为线性语言,记为$ll \ \ L$.

显然rg都是lg,所以rl也都是ll.但是ll并不都是rl.

例1 取$G_1 =< \{S\}, \{0,1\}, P_1, S >$,其中$P_1$由以下产生式组成:

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

这时有$L(G_1) = \{ww^R | w \in \{0,1\}^*\}$.所以$L_1 = L(G_1)$为$ll$,但$L_1$不是$rl$.

例2 取$G_2 =< \{S\}, \{0, 1\}, P_2, S >$,其中$P_2$由以下产生式组成:

$S \to 01 \quad S \to 0S1$

这时有$L(G_2) = \{0^n1^n | n \geq 1\}$.所以$L_2 = L(G_2)$为$ll$,但$L_2$不是$rl$.

定义4.7.2 若存在V_N中元素的一种编号方法,亦即有$V_N = \{A_1, \cdots, A_n\}$,使得$P$中每个产生式都满足以下条件:

$$若A_i \to \alpha \in P,则\alpha \in (V_T \cup \{A_i, A_{i+1}, \cdots, A_n\})^* \quad 1 \le i \le n$$

就称G为顺序文法,记为$sg\ \ G$.若有$sg\ \ G$使$L = L(G)$,就称L为顺序语言,记为$sl\ \ L$.

前面例子给出的文法G_1和G_2都是sg,所以,L_1和L_2也都是sl.

例3 取$G_3 =< \{S, A\}, \{a, b\}, P_3, S >$,其中$P_3$由以下产生式组成:

$S \rightarrow AA \quad A \rightarrow aAb \quad A \rightarrow \epsilon$

当令$A_1 = S, A_2 = A$时,可知G_3确是sg.因此$L(G_3) = \{a^n b^n a^m b^m | n, m \geq 0\}$为$sl$,可以证明$L_3 = L(G_3)$不是$ll$.

证明：假定L_3是ll,则有$lgG' =< V'_N, \{a,b\}, P', S'\} >$,使$L_3 = L(G')$.记

$l = max\{|\beta||A \to \beta \in P'\}, k =^{\#} V'_N$

则当$n > kl$时,在派生$S' \overset{*}{\Rightarrow}_{G'} a^n b^n a^n b^n$中, 必有一个非终结符变量出现至少两次以上.设$A$为第一个出现两次以上的变量. 那么派生$S' \overset{*}{\Rightarrow}_{G'} a^n b^n a^n b^n$必有以下形式:

$S' \overset{*}{\Rightarrow} a^{i_1} b^{i_2} A a^{j_2} b^{j_1} \overset{*}{\Rightarrow} a^{i_1} b^{i_2} a^{i_3} b^{i_4} A a^{j_4} b^{j_3} a^{j_2} b^{j_1} \overset{*}{\Rightarrow} a^{i_1} b^{i_2} a^{i_3} b^{i_4} w a^{j_4} b^{j_3} a^{j_2} b^{j_1}$

这里$a^{i_1} b^{i_2} a^{i_3} b^{i_4} w a^{j_4} b^{j_3} a^{j_2} b^{j_1} = a^n b^n c^n d^n$且$i_3 + i_4 + j_3 + j_4 > 0$

因此,$a^{i_1} b^{i_2} (a^{i_3} b^{i_4})^l w (a^{j_4} b^{j_3})^l a^{j_2} b^{j_1} \in L, l = 0, 1, 2, \ldots$

这与$L_3 = \{a^n b^n a^m b^m | n, m \geq 0\}$矛盾,所以$L_3$不是$ll$.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 86 页 共 108 页
返回
全屏显示
关闭
退出

例4 取$G_4 =< \{S_1, S_2, S_3\}, \{a, b, c, d, e\}, P_4, S_1 >$,其中$P_4$由以下产生式组成:

$$
\begin{array}{lll}
S_1 \to cS_2d & S_1 \to cS_3d & \\
S_2 \to cS_2ae & S_2 \to cS_3d & \\
S_3 \to cbe & S_3 \to aS_3be & S_3 \to cS_2d
\end{array}
$$

显然G_4为lg,且$L(G_4) = \{c^{m_1+\cdots+m_p+p}u_1^{m_1}d \cdots u_p^{m_p}d | p, m_1 \geq 1, m_2, m_3, \cdots, m_p \geq 0\}$且$u_{2i} = ae, u_{2i-1} = be, 1 \leq i \leq \frac{p+1}{2}$.因此$L_4 = L(G_4)$为$ll$, 但是可以证明$L_4$不是$sl$.

例5 取$G_5 =< \{S, A\}, \{a, b\}, P_5, S >$,其中$P_5$由以下产生式组成:

$$S_1 \to aA \quad A \to bS$$
$$S \to ab \quad A \to b$$

显然G_5为rg,但G_5不是sg.因为$L_5 = \{(ab)^n | n \geq 1\}$可以由$sl$ $G_5' =< \{S\}, \{a, b\}, \{S \to ab, S \to abS\}, S >$产生,所以$L_5$仍是$sl$.

访问主页
标题页
第 88 页 共 108 页
返回
全屏显示
关闭
退出

定义4.7.3 若L为cfl,且有有限个字$w_1, \cdots, w_k$使$L \subseteq w_1^* w_2^* \cdots w_k^*$,就称$L$为界限语言,记为$bl\ \ L$.

显然,L_2, L_3和L_5都是bl,而L_1和L_4都不是bl.

例6 取$G_6 =< \{S, A\}, \{a, b\}, P_6, S >$,其中$P_6$由以下产生式组成:

$$S \to AS \quad A \to aAb$$
$$S \to \epsilon \quad\quad A \to \epsilon$$

显然G_6为sg,且$L(G_6) = \{(a^{i_1}b^{i_1})\cdots(a^{i_n}b^{i_n})|i_1,\cdots,i_n, n \geq 0\}$,因此$L_6 = L(G_6)$为$sl$.但是可以证明$L_6$不是$bl$.

例7 显然rl $L_7 = \{a, b\}^*$不是bl.

定义4.7.4 如果存在$w \in L(G)$,使w至少有两个不同的最左派生,则称$cfg\ \ G$为歧义文法,否则称为非歧义文法.设L为cfl,若有$cfg\ \ G$使$L(G)=L$,则G为歧义文法,就称L为先天歧义语言.

例8 取$G_7 =< \{S, A, B\}, \{a, b\}, P_7, S >$,其中$P_7$由以下产生式组成:

$$\begin{array}{lll} S \rightarrow aB & S \rightarrow bA & \\ A \rightarrow bAA & A \rightarrow aS & A \rightarrow a \\ B \rightarrow aBB & B \rightarrow bS & B \rightarrow b \end{array}$$

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

因为$aabbab \in L(G_7)$有以下两个不同的最左派生:

$$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabB \Rightarrow aabbS \Rightarrow aabbaB \Rightarrow aabbab$$

$$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabSB \Rightarrow aabbAB \Rightarrow aabbaB \Rightarrow aabbab$$

所以G_7为歧义文法.

但是$L_8 = L(G_7)$却不是先天歧义语言，因为当取$G_8 =< \{S, A, B\}, \{a, b\}, P_8, S >$,其中$P_8$由以下产生式组成:

$$
\begin{array}{llll}
S \longrightarrow aB & S \longrightarrow bA & S \longrightarrow aBS & S \longrightarrow bAS \\
A \longrightarrow a & A \longrightarrow bAA & & \\
B \longrightarrow b & B \longrightarrow aBB & &
\end{array}
$$

时,则$L(G_8) = L_8$,且G_8是非歧义的,所以L_8不是先天歧义语言.

例9 取$G_9 =< \{S, A, B, C, D\}, \{a, b, c\}, P_9, S >$,其中$P_9$由以下产生式组成:

$$
\begin{array}{ll}
S \longrightarrow AD & S \longrightarrow BC \\
A \longrightarrow \epsilon & A \longrightarrow aA \\
B \longrightarrow \epsilon & B \longrightarrow aBb \\
C \longrightarrow \epsilon & C \longrightarrow cC \\
D \longrightarrow \epsilon & D \longrightarrow bDc
\end{array}
$$

显然G_9为sg且$L(G_9) = \{a^n B^n c^m, a^n b^m c^m | m, n \geq 0\}$. 因此, $L_9 = L(G_9)$为sl.由$L_9 \subseteq a^* b^* c^*$可知,$L_9$是$bl$,但$L_9$并不是$ll$.

4.8. 先天歧义语言

作为迭代定理的应用,本节将证明先天歧义语言的存在性.

定义4.8.1 设Σ为字母表,且$w, u, v \in \Sigma^*$.

1) 若有$x, y \in \Sigma^*$,使$w = xuy$,则称u为w的一个子字.

2) 若有$x, y, z \in \Sigma^*$,使$w = xuyvz$,则称w的两个子字u和v是不相交的.

3) 若u和v都是w的子字,且有$x, y, z \in \Sigma^+$, 使$u = xy$且$v = yz$,则称u和v部分重叠.

若u和v都是w的子字,则u和v部分重叠,当且仅当

1) u和v相交;

2) u不是v的子字,v不是u的子字;

定义4.8.2 设$cfg\ G=<V_N,V_T,P,S>$,且$\alpha,\beta,\gamma\in(V_N\cup V_T)^*$.

1) 若$S\overset{*}{\Rightarrow}\alpha A\gamma$且$A\Rightarrow\beta$,则称$\beta$为句型$\eta=\alpha\beta\gamma$的一个简单短语.

2) 若$S\overset{*}{\Rightarrow}\alpha A\gamma$且$A\overset{+}{\Rightarrow}\beta$,则称$\beta$为句型$\eta=\alpha\beta\gamma$的一个短语.

若β为句型$\eta=\alpha\beta\gamma$的一个(简单)短语,则当$\eta\in V_T^*$时,必有$\beta\in V_T^*$.这时称β为字$\eta=\alpha\beta\gamma$的一个(简单)短语.

定理4.8.1 设cfg $G =< V_N, V_T, P, S >$为非歧义的,$w \in L(G)$且u, v都是w的短语,则以下三个条件必有一个成立.

1) u和v不相交;

2) u为v的子字;

3) v为u的子字;

证明: 因为$w \in L(G)$,所以有最左派生如下:

$S \Rightarrow_L \beta_0 \Rightarrow_L \beta_1 \Rightarrow_L \cdots \Rightarrow_L \beta_n$且$\beta_n = w$

由于u为w的一个短语, 故有$i(0 \leq i < n)$,使$\beta_i = x_i A \alpha_i$且$A \overset{+}{\Rightarrow} u$;又因$v$也是$w$的短语,故又有$j(0 \leq j < n)$,使$\beta_j = x_j B \alpha_j$ 且$B \overset{+}{\Rightarrow} v$.这里,$x_i, x_j \in V_T^*$,且$\alpha_i, \alpha_j \in (V_N \cup V_T)^*$.

显然,这时i和j仅有以下三种可能:

1) $i = j$

因为G为非歧义,所以必有$x_i = x_j$, $A = B$且$\alpha_i = \alpha_j$, 因此$u = v$.

2) $j < i$

因为$\beta_j \overset{+}{\Rightarrow}_L \beta_i$,所以由引理2.2知道有

$B \overset{+}{\Rightarrow}_L xA\gamma \;\; x \in V_T^*$且$\gamma \in (V_N \cup V_T)^*$

或$\alpha_j \overset{+}{\Rightarrow}_L x'A\gamma' \;\; x' \in V_T^*$且$\gamma' \in (V_N \cup V_T)^*$

在前一种情况下,有

$B \overset{+}{\Rightarrow}_L xA\gamma \overset{*}{\Rightarrow}_L xu\gamma \overset{*}{\Rightarrow} xuy$且$v = xuy$

即u为v的一个子字.

在后一种情况下,有

$S \overset{+}{\Rightarrow}_L x_jB\alpha_j \overset{+}{\Rightarrow}_L x_jv\alpha_j \overset{+}{\Rightarrow}_L x_jv\alpha'A\gamma' \overset{+}{\Rightarrow}_L x_jvx'u\gamma' \overset{+}{\Rightarrow}_L x_jvx'uy'$

且$w = x_jvx'uy'$

即u和v不相交;

3) $j > i$

与2)类似,这时有v为u的一个子字或u与v不相交.

定理4.8.2 设$cfg\ \ G =< V_N, V_T, P, S >$.若有$w \in L(G)$及$w$的两个短语$u$和$v$部分重叠,则$G$为歧义的.

定理4.8.3 设$cfg\ \ G$非歧义,则有$Chomsky$范式文法G'满足:

1) $L(G') = L(G)$;

2) G'非歧义;

定理4.8.4 $L=\{a^ib^jc^k|i=j或j=k\}$为先天歧义语言.

证 明 : 因为L为cfl,所以有$Chomsky$范式文法$G=<V_N,V_T,P,S>$使$L(G)=L$.令

$L_0=\{a^ib^ic^j|i,j\geq 0\}$

$L_1=\{a^ib^jc^j|i,j\geq 0\}$

则$L=L_0\cup L_1$.

设p_0为由迭代定理所确定的正整数,再取$p=p_0!$.当取$z_0=a^{2p}b^{2p}c^{3p}$时, 则$z_0\in L_0$Ł$_1$,则$|z_0|=7p>p_0$.指定z_0中最左边的p个b所占位置为指定位置,则必有$z_0=u_0v_0w_0x_0y_0$满足:

i) u_0,v_0和w_0都各至少占有一个指定位置,或w_0,x_0和y_0都各至少占有一个指定位置;

ii) $v_0w_0x_0$至多占有p_0个指定位置;

iii) 若$k\in\mathbb{N}$,则$u_0v_0^kw_0x_0^ky_0\in L$;

iv 有$A_0\in V_N$使得在派生$S\overset{*}{\Rightarrow}z_0$中有

$S\overset{*}{\Rightarrow}u_0A_0y_0\ \ A_0\overset{*}{\Rightarrow}v_0A_0x_0$且$A_0\overset{*}{\Rightarrow}w_0$

首先,v_0和x_0都仅能含一种符号,否则有$u_0v_0^2w_0x_0^2y_0\notin L$.

访问主页
标题页
◀◀ ▶▶
◀ ▶
第 102 页 共 108 页
返回
全屏显示
关闭
退出

其次,根据i),只能有以下两种可能:

1) u_0, v_0和w_0各都至少占有一个指定位置,这时有

$u_0 = a^{2p}b^{p+l}$

$v_0 = b^k(j > 0, k > 0$且$j + k < p)$

显然,x_0这时只能有以下两种可能:

i) $x_0 \in b^*$

这时必有

$w_0 = b^l(l > 0)$

$x_0 = b^m(m \geq 0)$

$y_0 = b^{p-(j+k+l+m)}c^{3p}(j + k + l + m \leq p)$

从而由$k > 0$及$m \geq 0$知道,$u_0w_0y_0 = a^{2p}b^{2p-(k+m)}c^{3p} \notin L$

ii) $x_0 \in c^*$

这时

$$w_0 = b^{p-(j+k)}c^l(l \geq 0)$$

$$x_0 = c^m(m \geq 0)$$

$$y_0 = c^{3p-(m+l)}(m + l \leq 3p)$$

从而由$0 < k < p$得

$$u_0v_0^2w_0x_0^2y_0 = a^{2p}b^{2p+k}c^{3p+m} \notin L.$$

2) 假定w_0, x_0和y_0各都至少占有一个指定的位置,这时有

$$y_0 = b^i c^{3p} (i > 0)$$

$$x_0 = b^j (j > 0, i + j < p)$$

显然,这时v_0只能有以下两种可能:

i) $v_0 \in b^*$

这时必有

$$w_0 = b^k (k > 0)$$

$$u_0 = b^l (l \geq 0)$$

$$u_0 = a^{2p} b^{2p-(i+j+k+l)} (j + k + l \leq p_0 < p)$$

从而由$j > 0$知道$u_0 w_0 y_0 = a^{2p} b^{2p-(j+l)} c^{3p} \notin L$.

ii) $v_0 \in a^*$

这时$w_0 = a^l b^{2p-(i+j)} (l \geq 0)$

$v_0 = a^k$

$u_0 = a^{2p-(l+k)} (0 \leq k \leq 2p-l)$

因此$u_0 v_0^2 w_0 x_0^2 y_0 = a^{2p+k} b^{2p+j} c^{3p}$

若$k \neq j$,则由$2p + j < 3p$得：$u_0 v_0^2 w_0 x_0^2 y_0 \notin L$.因此,必有$k = j$.亦即

$u_0 = a^{2p-(j+l)} (l \geq 0)$

$v_0 = a^j (j > 0)$

$w_0 = a^l b^{2p-(i+j)} (i > 0)$

$x_0 = b^j (i + j < p)$

$y_0 = b^i c^{3p}$

由于$j = |v_0| < |v_0 w_0 x_0| \leq p_0$且$p = p_0!$,所以$j|p$.因此有正整数$q$使$p = jq$.从而得$S \overset{*}{\Rightarrow} u_0 A_0 y_0 \overset{*}{\Rightarrow} u_0 v_0^{q+1} A_0 x_0^{q+1} y_0 \overset{*}{\Rightarrow} u_0 v_0^{q+1} w_0 x_0^{q+1} y_0$ 且$u_0 v_0^{q+1} w_0 x_0^{q+1} y_0 = a^{3p} b^{3p} c^{3p}$

对字$z_1 = a^{3p}b^{3p}c^{3p}$,当指定z_1中最左边p个b所占的位置时,通过类似的论证可得

$u_1 = a^{3p'}b^{i'}(i' > 0)$

$v_1 = b^{j'}(j' > 0)$

$w_1 = b^{2p-(i'+j')}c^{k'}(k' \geq 0)$

$x_1 = c^{j'}(i' + j' < p_0)$

$y_1 = c^{2p-(i'+j')}$

且有$A_1 \in V_N$,使得在派生$S \stackrel{*}{\Rightarrow} z_1$中有

$S \stackrel{*}{\Rightarrow} u_1 A - 1 y_1 \;\; A_1 \stackrel{*}{\Rightarrow} v_1 A_1 x_1$且$A_1 \stackrel{*}{\Rightarrow} w_1$

由于$j' = |v_1| < |v_1 w_1 x_1| \leq p_0$且$p = p_0!$,所以$j'|p$,因此有正整数$q'$使$p = j'q'$.从而得

$S \stackrel{*}{\Rightarrow} u_1 A_1 y_1 \stackrel{*}{\Rightarrow} \stackrel{*}{u}_1 \; v_1^{q'+1} A_1 x_1^{q'+1} y_1 \Rightarrow u_1 v_1^{q'+1} w_1 x_1^{q'+1} y_1$

且$u_1 v_1^{q'+1} w_1 x_1^{q'+1} y_1 = a^{3p}b^{3p}c^{3p}$

访 问 主 页
标 题 页
◀◀ ▶▶
◀ ▶
第 107 页 共 108 页
返 回
全 屏 显 示
关 闭
退 出

这样一来,就得到了字$a^{3p}b^{3p}c^{3p}$的两个短语

$v_0^{q+l}w_0x_0^{q+l} = a^{p+j+l}b^{3p-i}$

及$v_1^{q'+1}w_1x_1^{q'+1} = b^{3p-i'}c^{p+j'+k'}$

因为$3p-i > 3p-p = 2p, 3p-i' > 3p-p = 2p$

而且短语$a^{p+j+l}b^{3p-i}$中的b为$a^{3p}b^{3p}c^{3p}$中的b^{3p}的右边之b.所以这两个短语既不是不相交,又不是互为子字,从而由定理4.8.2知道G为歧义文法.

由于对G的选取是随意的.因此,再由定理4.8.3知道,任何产生语言L的cfg必为歧义的cfg. 故L为先天歧义的.