上下文无关语言的性质

泵引理及应用

- 如果语言L是CFL，那么存在正整数N，对$\forall z \in L$只要$|z| \geq N$，就可以将z分为五部分$z = uvwxy$，满足：
  - $vx \neq \varepsilon$（或$|vx| \geq 0$）
  - $|vwx| \leq N$
  - $\forall i \geq 0, uv^i wx^i y \in L$
- **使用步骤**

1. 假设L是CFL，那么存在正整数N，对$\forall z \in L(|z| \geq N)$满足泵引理
2. 取满足条件的z
3. 由泵引理，将z分为$z = uvwxy$，且有$|vwx| \leq N$且$vx \neq \varepsilon$
4. 对i取值，使得$uv^i wx^i y \notin L$接受（i通常可取0）
5. 假设不成立，故L不是CFL

同样的，泵引理也只是CFL的**必要条件**而不是**充分条件**

上下文无关语言的封闭性

基础：代换

- 定义：两个字母表Σ到Γ的函数$s: \Sigma \to 2^{\Gamma^*}$，即$\Sigma$中的一个字符$a$在$s$的作用下为$\Gamma$上的一个语言$L_a$，即$s(a) = L_a$
- 扩展s的定义到字符串：

$$s(\varepsilon) = \varepsilon$$
$$s(xa) = s(x)s(a)$$

- 扩展s的定义到语言，对$\forall L \subset \Sigma^*$

$$s(L) = \bigcup_{x \in L} s(x)$$

可以看出，对字符串的扩展定义无非就是把**两种语言进行拼接**，对语言的扩展定义无非是把语言里的**所有字符串经**s映射后取并集**

- 构造方法

定理：Σ上的$CFL \ \ L$和代换s，且每个$a \in \Sigma$的$s(a)$都是CFL，则$s(L)$也是CFL

设$CFL \ \ L$的文法$G = (V, T, P, S)$，每个$s(a)$的文法$G_a =$（V_a, T_a, P_a, S_a)，则$s(L)$的文法可以构造为

$$G' = (V', T', P', S)$$

其中：

1. $V' = V \cup (\bigcup_{a \in T} V_a)$
2. $T' = \bigcup_{a \in T} T_a$

实际上，这两个条件在写题的时候不需要过多关注，主要关注的是产生式的集合P'，其包括

- 每个P_a中的产生式
- P的产生式，但是将其中的每个终结符a替换成文法G_a的开始符号S_a

附例：

- **例：设**$L = \{w \in \{a,b\}^* | w$ **有相等个数的** a **和** $b\}$，**代换**

$$s(a) = L_a = \{0^n 1^n | n \geq 1\}$$
$$s(b) = L_b = \{ww^R | w \in (\mathbf{0}+\mathbf{1})^*\}$$

求$s(L)$**的文法。**

解：设计L**的文法为：** $S \to aSbS|bSaS|\varepsilon$

L_a**的文法为：** $S_a \to 0S_a1|01$

L_b**的文法为：** $S_b \to 0S_b0|1S_b1|\varepsilon$

封闭性运算

上下文无关语言对**「并/连接/闭包/同态/逆同态/反转」**是封闭的

这部分的证明比较繁琐，但大多都是基于「代换」进行证明的，然后要注意的就是这些运算的定义，可以参考「正则语言的封闭性」一节，这里注重讨论「交补」运算的封闭性

- CFL在交运算下不封闭，典型的例子是

$$\{0^n1^n2^i\} \cap \{0^i1^n2^n\} = \{0^n1^n2^n\} \neq CFL \quad (n \geq 1, i \geq 1)$$

- CFL在补运算下不封闭，因为在$L1 \cap L2 = \overline{\overline{L1} \cup \overline{L2}}$中，如果补运算封闭，那么交运算也封闭，而交运算是不封闭的
- 若L是CFL而R是正则语言，则$L \cap R$是CFL
  - 这个定理通常用来证明某个语言不是CFL
  - 使用方法是，令该L与构造的RL相交，使得结果明显不是一个CFL
  - 说明，因为$CFL \cap RL = CFL$，根据定理，原L不是CFL

- **例：证明语言**L**不是CFL。**

$$L = \{w \in \{a,b,c\}^* | n_a(w) = n_b(w) = n_c(w)\}$$

其中$n_a(w)$**表示**w**中**a**的个数。**

- **证明：**

(1) 因为$\boldsymbol{a^*b^*c^*}$**是正则语言，**

(2) 而$L \cap a^*b^*c^* = \{a^nb^nc^n | n \geq 0\}$**不是CFL。**

(3) 由CFL与正则语言的交还是CFL，所以L**不可能是CFL。**

上面的例子使用的是典型的非CFL的$RL : a^*b^*c^*$

上下文无关语言的判定问题

- 空性：只需要判断文法的开始符号S是否是**非产生**的
- 有穷性和无穷性：用不带无用符号和ε-产生式的CNF范式文法的产生式画有向图，变元为顶点，如果有$A \to BC$，则从A到B和C各画一条有向边，如果图中没有循环，则L有穷；如果有循环，则L无穷
- 成员性：CYK算法

CYK算法

对于$CNF\ \ G$，CYK算法可以检查字符串$w=a_1a_2...a_n$是否属于语言$L(G)$，步骤如下（没错我偷懒了）：

1. 画三角形表格

   三角形表格示例（假设w的长度为5）：

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
a_1	a_2	a_3	a_4	a_5

   $X_{ij}=\{A\in V|A\overset{*}{\Rightarrow}a_i \dots a_j\}$，即$X_{ij}$为能派生出$a_i \dots a_j$的变元的集合。

2. - **计算表格中的每一行的变元：**
     - 基础: $X_{ii}=\{A|A\to a_i\in P\}$
     - 其他：$X_{ij}=\{A|i\le k<j, BC\in X_{ik}X_{k+1,j}, A\to BC\}$
       - 比如，对于X_{12}，$1\le k<2$,因此k只能取1，则考虑$X_{11}X_{2,2}$，找出能产生$X_{11}X_{2,2}$中变元串的变元
       - 对于X_{13}，$1\le k<3$，因此k可以取1，2，则需要考虑$X_{11}X_{2,3}$和$X_{12}X_{3,3}$，取$X_{11}X_{2,3}$和$X_{12}X_{3,3}$的并集$(X_{11}X_{2,3})\cup(X_{12}X_{3,3})$。找出能产生该集合中变元串的变元

3. 最后，检查是否满足$S\in X_{1,5}$

乔姆斯基文法体系

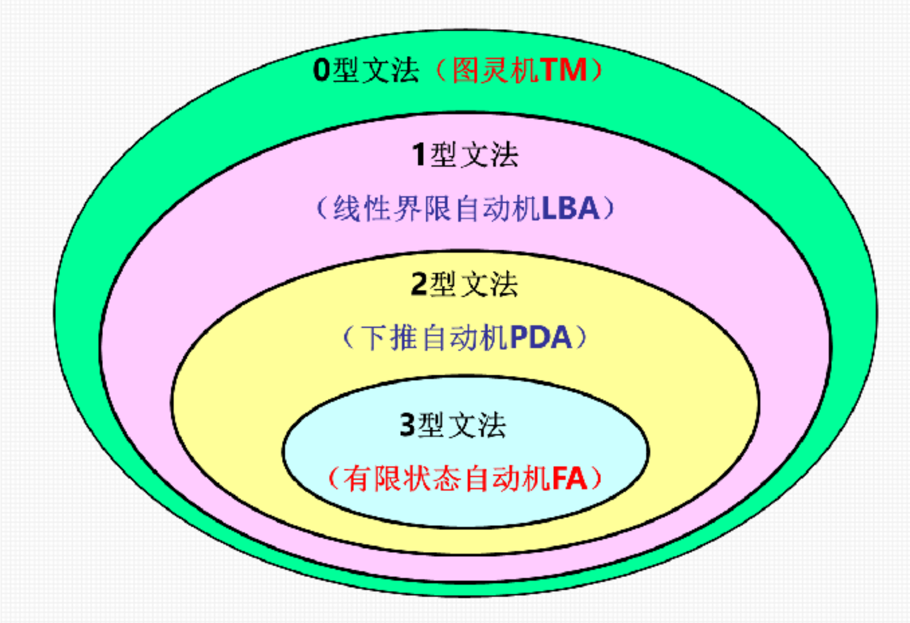
对于文法$G=(V,T,P,S)$，符号串$\alpha\in(V\cup T)^*V(V\cup T)^*,\beta\in(V\cup T)^*$，产生式形如

$$\alpha\to\beta$$

即每个产生式左部α中至少有一个变元，定义：

- 3型文法：$\alpha\to\beta$都形如$A\to aB$或$A\to a$或$A\to\varepsilon$（$A,B\in V;a\in T$）
- 2型文法：$\alpha\in V$
- 1型文法：$|\alpha|\le|\beta|$
- 0型文法：没有限制

其中3型文法是正则文法，2型文法是上下文无关文法，1型文法是上下文有关文法


0型文法（图灵机TM）
1型文法
（线性界限自动机LBA）
2型文法
（下推自动机PDA）
3型文法
（有限状态自动机FA）