形式语言与自动机

Formal Languages and Automata Theory

上下文无关语言的性质

计算机科学与技术学院
哈尔滨工业大学（深圳）

郑宜峰

主要内容

- 上下文无关语言的泵引理及应用
- 上下文无关语言的封闭性
- 上下文无关语言的判定性质
- 乔姆斯基文法体系

主要内容

- 上下文无关语言的泵引理及应用
- 上下文无关语言的封闭性
- 上下文无关语言的判定性质
- 乔姆斯基文法体系

- **泵引理(pumping lemma)：如果语言L是正则的，那么存在正整数 N，对$w \in L$, 只要$|w| \geq N$，就可以将w分为三部分$w = xyz$满足:**
  - $y \neq \varepsilon$（或$|y| > 0$）；
  - $|xy| \leq N$;
  - $\forall k \geq 0, xy^k z \in L$;
    - 中间的串不论循环几次，所得的串仍属于该正则语言。

正则语言的泵引理

- **泵引理(pumping lemma)：如果语言L是正则的，那么存在正整数 N，对$w \in L$, 只要$|w| \geq N$，就可以将w分为三部分$w = xyz$满足：**
  - $y \neq \varepsilon$ （或$|y| > 0$）；
  - $|xy| \leq N$;
  - $\forall k \geq 0, xy^k z \in L$;
    - 中间的串不论循环几次，所得的串仍属于该正则语言。

泵引理可以用来确定特定语言不是正则语言，泵引理成立是正则语言判定的必要条件，但不是充分条件。

上下文无关语言的泵引理

- **定理: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:**

（1） $vx \neq \varepsilon$（或$|vx| > 0$）;

（2） $|vwx| \leq N$;

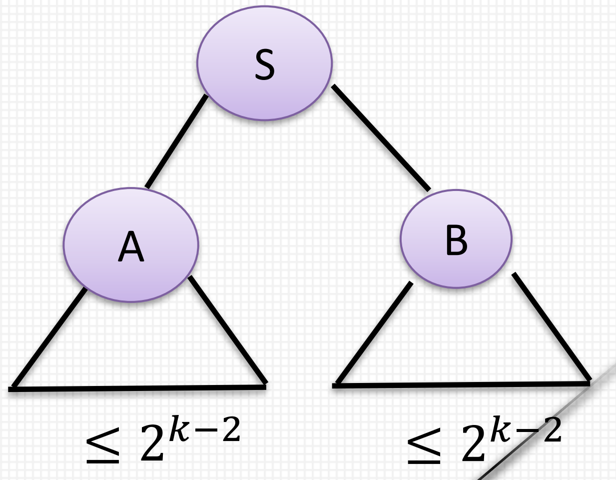
（3） $\forall i \geq 0, uv^iwx^iy \in L$。

- 设CFG $G=(V,T,P,S)$为接受$L-\{\varepsilon\}$的CNF（乔姆斯基范式）文法。
- 乔姆斯基范式文法（回顾）：每个不带ε的CFL都可以由这样的CFG G定义，G中每个产生式的形式都为

$$A \to BC \text{ 或 } A \to a$$

这里的A, B和C是变元，a是终结符。

- **首先，在文法G的派生树中，若最长路径为k，则产物w（派生出的句子）的长度最多为2^{k-1}，即$\leq 2^{k-1}$。**
- **归纳法证明：**
  - 归纳基础：当$k=1$，一定是S直接连接一个终结符，产物长度为1，所以为$\mathbf{2^{(1-1)}}$
  - 假设路径长度为$k-1$时成立，即$|w| \leq 2^{(k-1-1)} = 2^{(k-2)}$
  - 则路径长度为k时，第一步派生一定是调用了$S \rightarrow AB$，然后由A和B的子树去派生。A, B子树的可能最长路径为$k-1$，故每个子树的产物长度最多为2^{k-2}，合一起为2^{k-1}

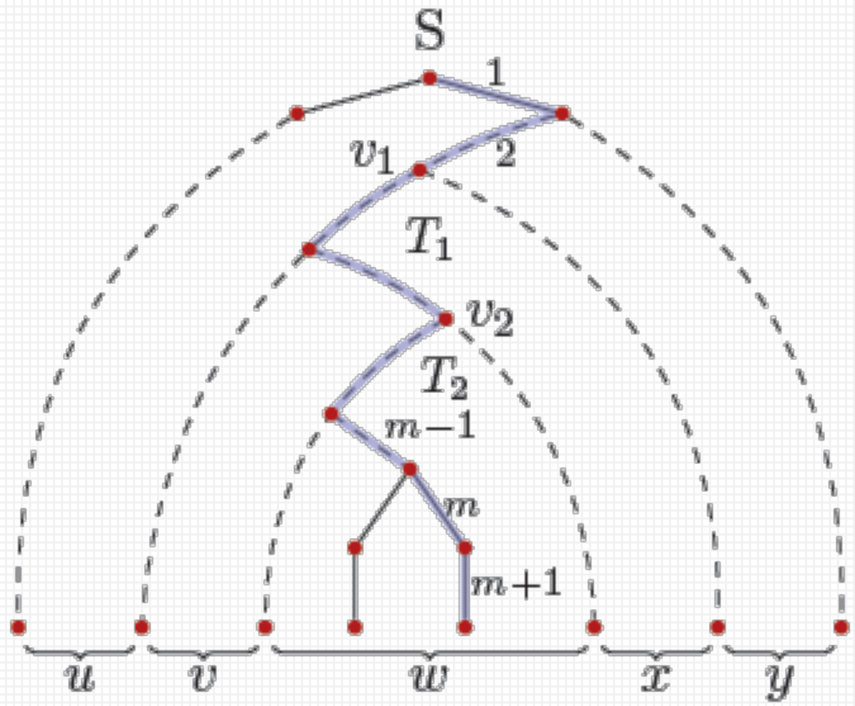


（归纳法已证）在文法G的派生树中，若最长路径为k，则产物w（派生出的句子）的长度最多为2^{k-1}，即$\leq 2^{k-1}$。

- 设G中变元数$|V| = m$，$N = 2^m$，那么若有$z \in \mathbf{L}(G)$，$|z| \geq N = 2^m$，则z的派生树中最长路径长度至少也是$m+1$
- 因此，这条路径上节点至少 $m+2$ 个。除了最后一个节点外，其余均为变元，即至少$m+1$个变元
- 该路径由下至上$m + 1$个内节点中，至少有两个节点标记了相同的变元。

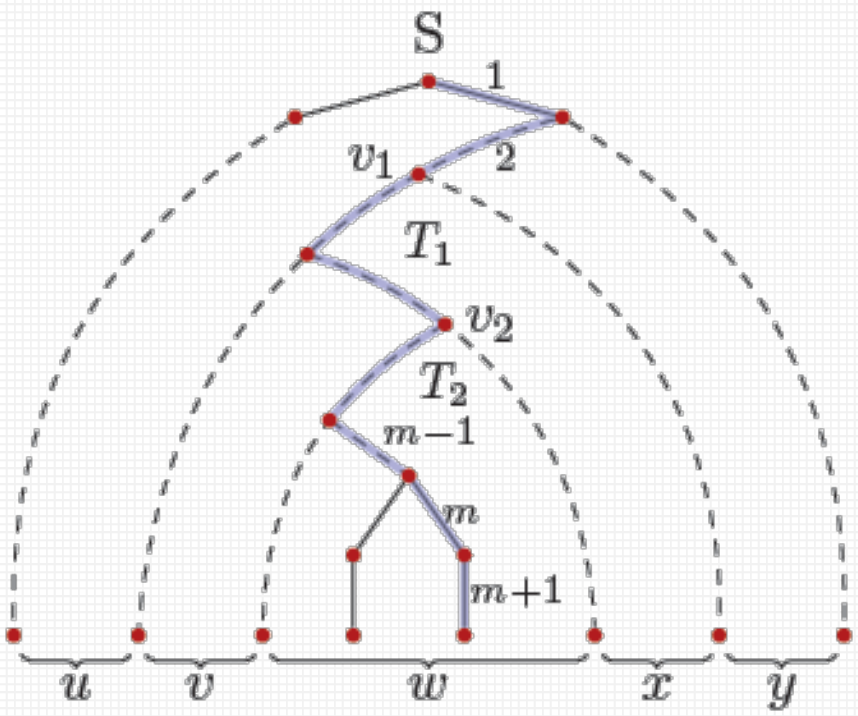
泵引理证明

- 如果这两个节点分别是v_1和v_2，标记均为A，设v_1比v_2更接近树根。
- 设以v_1为根的子树为T_1。
- 设以v_2为根的子树为T_2。



由前，派生树中最长路径长度至少是$m+1$

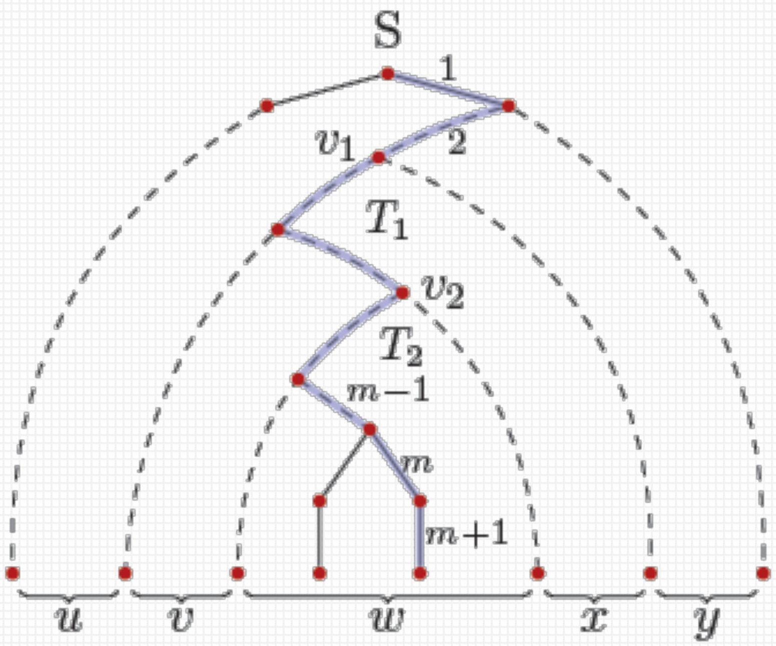
- **设T_1的产物为z_1，T_2的产物为w。由于T_2是T_1的子树，则T_1的产物z_1必包含w，不妨记为$z_1 = vwx$，则有$A \overset{*}{\Rightarrow} vAx$和$A \overset{*}{\Rightarrow} w$。**
- **那么对任意$i \geq 0$，$A \overset{*}{\Rightarrow} v^i w x^i$。**
  - 当$i = 0$时，即$A \overset{*}{\Rightarrow} w$
  - 当$i > 0$时，即通过迭代调用$A \overset{*}{\Rightarrow} vAx$，产生$A \overset{*}{\Rightarrow} v^i A x^i$, 最后调用$A \overset{*}{\Rightarrow} w$，即得到$A \overset{*}{\Rightarrow} v^i w x^i$



（注：前面假设这两个节点v_1和v_2的标记均为A）

泵引理证明

- 对任意$i \geq 0$，$A \overset{*}{\Rightarrow} v^i w x^i$。
- 设整棵树的产物为$z = uvwxy$(即T_1 产物vwx的前面部分为u，后面部分为y)，则
  $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uv^i w x^i y$。（条件3）
- T_1路径长度不超过$m+1$，其产物vwx长度不超过2^m（即N），即$|vwx| \leq N$。（条件2）
- T_1 派生出vwx的第一个产生式必须是$A \to BC$。T_2必定处于B的树或C的树中，并且B和C都至少产生一个终结符，所以v和x不可能同时为空，即$vx \neq \varepsilon$。（条件1）



定理：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：
（1） $vx \neq \varepsilon$（或$|vx| > 0$）；
（2） $|vwx| \leq N$；
（3） $\forall i \geq 0, uv^i w x^i y \in L$。

- **例1：证明** $L = \{0^n 1^n 2^n \mid n \ge 1\}$ **不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z \in L\ (|z| \ge N)$满足泵引理；

对 $\forall z \in L$，只要 $|z| \ge N$，就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \ne \varepsilon$;
(2) $|vwx| \le N$;
(3) $\forall i \ge 0$, $uv^iwx^iy \in L$

- **例1：证明** $L = \{0^n1^n2^n | n \geq 1\}$ **不是上下文无关语言。**

(1) 假设L**是CFL，那么存在整数**N**，对**$\forall z \in L\ (|z| \geq N)$ **满足泵引理；**
(2) 从L**中取**$z = 0^N1^N2^N$**，则显然**$z \in L$**且**$|z| = 3N \geq N$**；**

对 $\forall z \in L$, 只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例1：证明** $L = \{0^n 1^n 2^n \mid n \geq 1\}$ **不是上下文无关语言。**

(1) 假设L**是CFL，那么存在整数**N**，对**$\forall z \in L\ (|z| \geq N)$ **满足泵引理；**

(2) 从L**中取**$z = 0^N 1^N 2^N$**，则显然**$z \in L$**且**$|z| = 3N \geq N$**；**

(3) 由泵引理，z**可被分为**$z = uvwxy$**，且有** $|vwx| \leq N$ **和**$vx \neq \varepsilon$；

对 $\forall z \in L$，只要$|z| \geq N$，就可以将z分为五部分$z = uvwxy$满足：

(1) $vx \neq \varepsilon$;

(2) $|vwx| \leq N$;

(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例1：证明** $L = \{0^n1^n2^n | n \geq 1\}$ **不是上下文无关语言。**

(1) 假设L**是CFL，那么存在整数**N**，对**$\forall z \in L\ (|z| \geq N)$ **满足泵引理；**

(2) 从L**中取**$z = 0^N1^N2^N$**，则显然**$z \in L$**且**$|z| = 3N \geq N$**；**

(3) 由泵引理，z**可被分为**$z = uvwxy$**，且有** $|vwx| \leq N$ **和**$vx \neq \varepsilon$**；**

(4) 那么vx**可能：**

i. 只包含0或1或2，那么$uwy \notin L$**；（**uwy**即删掉部分0，或部分1，或部分2，那么都不会在语言中）**

ii. 只包含0和1，或只包含1和2，那么也有$uwy \notin L$**；(**uwy**即删掉部分0和部分1，或部分1和部分2，那么都不会在语言中）**

对 $\forall z \in L$，只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例1：证明 $L = \{0^n1^n2^n | n \geq 1\}$ 不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z \in L\ (|z| \geq N)$ 满足泵引理；
(2) 从L中取$z = 0^N1^N2^N$，则显然$z \in L$且$|z| = 3N \geq N$；
(3) 由泵引理，z可被分为$z = uvwxy$，且有 $|vwx| \leq N$ 和$vx \neq \varepsilon$；
(4) 那么vx可能：
i. 只包含0或1或2，那么$uwy \notin L$；（uwy即删掉部分0，或部分1，或部分2，那么都不会在语言中）
ii. 只包含0和1，或只包含1和2，那么也有$uwy \notin L$；（uwy即删掉部分0和部分1，或部分1和部分2，那么都不会在语言中）
(5) 与泵引理$uwy = uv^0wx^0y \in L$矛盾，假设不成立；
(6) L不是上下文无关的。

对 $\forall z \in L$，只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例2：证明** $L = \{ww | w \in \{0,1\}^*\}$ **不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z \in L$ ($|z| \geq N$) 满足泵引理；

(2) 取$z = 0^N 1^N 0^N 1^N$，则显然$z \in L$且$|z| = 4N \geq N$；

对 $\forall z \in L$, 只要 $|z| \geq N$， 就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^i wx^i y \in L$

- **例2:证明** $L=\{ww|w\in\{0,1\}^*\}$ **不是上下文无关语言。**

(1) 假设L**是CFL，那么存在整数**N**，对**$\forall z\in L\ (|z|\geq N)$ **满足泵引理;**

(2) 取$z=0^N1^N0^N1^N$**，则显然**$z\in L$**且**$|z|=4N\geq N$**;**

(3) 由泵引理, z**可分为**$z=uvwxy$**，且有** $|vwx|\leq N$ **和** $vx\neq\varepsilon$

对 $\forall z\in L$, 只要 $|z|\geq N$, 就可以将z分为五部分 $z=uvwxy$满足:
(1) $vx\neq\varepsilon$;
(2) $|vwx|\leq N$;
(3) $\forall i\geq 0$, $uv^iwx^iy\in L$

- **例2:证明** $L=\{ww|w\in\{0,1\}^*\}$ **不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z\in L$ ($|z|\geq N$) 满足泵引理;

(2) 取$z=0^N1^N0^N1^N$，则显然$z\in L$且$|z|=4N\geq N$;

(3) 由泵引理, z可分为$z=uvwxy$，且有 $|vwx|\leq N$ 和 $vx\neq\varepsilon$

(4) 那么vx可能：

i. 只包含0, 或1；ii. 包含部分0和部分1；

iii. 包含部分1和部分0；

对该三种情况，uv^0wx^0y显然都不属于L；

所以假设不成立，L不是CFL。

对 $\forall z\in L$, 只要 $|z|\geq N$，就可以将z分为五部分 $z=uvwxy$满足:

(1) $vx\neq\varepsilon$;

(2) $|vwx|\leq N$;

(3) $\forall i\geq 0$, $uv^iwx^iy\in L$

- **例3:证明 $L = \{a^n|n为素数\}$ 不是上下文无关语言。**

(1)假设L是CFL，那么存在整数N，对$\forall z \in L\ (|z| \geq N)$ 满足泵引理；(2)取$z = a^p\ (p$为不小于N的素数)则显然$z \in L$且 $|z| = p \geq N$;

(3)由泵引理, z可分为$z = uvwxy$，且有 $|vwx| \leq N$ 和$vx \neq \varepsilon$。

对 $\forall z \in L$, 只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足:
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例3：证明** $L = \{a^n | n\text{为素数}\}$ **不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z \in L\ (|z| \geq N)$ 满足泵引理；(2) 取$z = a^p$ (p为不小于N的素数) 则显然$z \in L$且$|z| = p \geq N$;

(3) 由泵引理, z可分为$z = uvwxy$，且有 $|vwx| \leq N$ 和$vx \neq \varepsilon$。

(4) 考虑$|uv^0wx^0y| = |uwy| = m$，即当循环次数$i = 0$时，除了v和x之外的串的长度为$m\,(0 \leq m \leq p - 1)$。

则对于任意循环次数$i > 0$, 有$|uv^iwx^iy| = m + i(p - m)$

对 $\forall z \in L$, 只要$|z| \geq N$，就可以将z分为五部分$z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例3：证明** $L = \{a^n | n\text{为素数}\}$ **不是上下文无关语言。**

(1) 假设L**是CFL，那么存在整数**N**，对**$\forall z \in L$ $(|z| \geq N)$ **满足泵引理；(2) 取**$z = a^p$ **(**p**为不小于**N**的素数) 则显然**$z \in L$**且** $|z| = p \geq N$；

(3) 由泵引理, z**可分为**$z = uvwxy$**，且有** $|vwx| \leq N$ **和**$vx \neq \varepsilon$**。**

(4) 考虑$|uv^0wx^0y| = |uwy| = m$**，即当循环次数**$i = 0$**时，除了**v和x**之外的串的长度为**m $(0 \leq m \leq p-1)$。

则对于任意循环次数$i > 0$, 有$|uv^iwx^iy| = m + i(p-m)$

取$i = p+1$**，则泵出的串长度为**$m + (p+1)(p-m)$.

注意到$m + (p+1)(p-m) = m + p^2 - pm + p - m$

$$= p^2 - pm + p$$

$$= p(p - m + 1)$$

对 $\forall z \in L$, 只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足：
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

- **例3:证明 $L = \{a^n | n\text{为素数}\}$ 不是上下文无关语言。**

(1) 假设L是CFL，那么存在整数N，对$\forall z \in L$ ($|z| \geq N$) 满足泵引理；(2) 取$z = a^p$ (p为不小于N的素数) 则显然$z \in L$且 $|z| = p \geq N$;

(3) 由泵引理, z可分为$z = uvwxy$，且有 $|vwx| \leq N$ 和$vx \neq \varepsilon$。

(4) 考虑$|uv^0wx^0y| = |uwy| = m$，即当循环次数$i = 0$时，除了v和x之外的串的长度为m ($0 \leq m \leq p-1$)。

则对于任意循环次数$i > 0$, 有$|uv^iwx^iy| = m + i(p-m)$

取$i = p+1$，则泵出的串长度为$m + (p+1)(p-m)$.

注意到$m + (p+1)(p-m) = m + p^2 - pm + p - m$

$$= p^2 - pm + p$$

$$= p(p - m + 1)$$

显然该长度不是素数，那么$uv^{p+1}wx^{p+1}y$也就不属于L。

所以假设不成立， L不是CFL。

对 $\forall z \in L$, 只要 $|z| \geq N$，就可以将z分为五部分 $z = uvwxy$满足:
(1) $vx \neq \varepsilon$;
(2) $|vwx| \leq N$;
(3) $\forall i \geq 0$, $uv^iwx^iy \in L$

主要内容

- 上下文无关语言的泵引理
- **上下文无关语言的封闭性**
  - 代换的封闭性
  - 并/连接/闭包/同态/逆同态/反转的封闭性
  - 交和补运算不封闭
- 上下文无关语言的判定性质
- 乔姆斯基文法体系

- **定义：两个字母表Σ到Γ的函数$s:\Sigma \to 2^{\Gamma^*}$称为代换。 Σ中的一个字符a在s的作用下为Γ上的一个语言L_a，即**

$$s(a) = L_a$$

- **扩展s的定义到字符串，**

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

- **再扩展s到语言，对$\forall L \subset \Sigma^*$，**

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

- **例:对字母表**$\Sigma = \{0,1\}$**，有代换**$s(0) = \{a^n b^n | n \geq 1\}, s(1) = \{aa, bb\}$**，则**

(1)对于符号串$w = 01, s(w)$为两个**语言**$s(0)$**和**$s(1)$**的连接，**$s(w)$**含有所有形式为**$a^n b^n aa$**和**$a^n b^{n+2}$**的串。**

(2) 设$L = \mathbf{L}(\mathbf{0}^*)$**，则**$s(L) = (s(0))^*$**，其包含所有形式为**
$a^{n_1} b^{n_1} a^{n_2} b^{n_2} \ldots a^{n_k} b^{n_k}$, **这里**$k \geq 0$，$n_1$，$n_2$, ⋯，$n_k$为任意的非负整数序列。比如像$\varepsilon$，$aabbaaabbb, abaabbabab$这些符号串。

- **定理:如果有Σ上的CFL L和代换s，且每个 $a \in \Sigma$的$s(a)$都是CFL，那么$s(L)$也是CFL。**
- **证明：构造方法:**

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

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

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

(2) $T' = \bigcup_{a \in T} T_a$

P'包括(i)每个P_a中的产生式；(ii) P的产生式，但是要替换产生式中的每个终结符a为文法G_a的开始符号S_a。

- **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：** $S_b \to 0S_b0|1S_b1|\varepsilon$

那么，$s(L)$**的文法为：**

$$S \to S_aSS_bS|S_bSS_aS|\varepsilon$$
$$S_a \to 0S_a1|01$$
$$S_b \to 0S_b0|1S_b1|\varepsilon$$

P'包括(i)每个P_a中的产生式；(ii) P的产生式，但是要替换产生式中的**每个终结符**$\boldsymbol{a}$**为文法**$\boldsymbol{G_a}$**的开始符号**$\boldsymbol{S_a}$。

- **证明（充分性$s(L) \subseteq \mathbf{L}(G')$）：**

对$\forall w \in s(L)$，**那么一定存在某个**$x = a_1a_2 \cdots a_n \in L$ **使**

$$w \in s(x) = s(a_1)s(a_2) \cdots s(a_n)。$$

那么w **可以分为**$w = w_1w_2 \cdots w_n$，**且**$w_i \in s(a_i)$，**则有**

$$S_{a_i} \overset{*}{\underset{G_{a_i}}{\Longrightarrow}} w_i$$

P'包括P的产生式，但是要替换产生式中的每个终结符a为文法G_a的开始符号S_a。

由于$x \in L$，**有** $S \overset{*}{\underset{G}{\Rightarrow}} x = a_1a_2 \cdots a_n$，**所以**

$$S \overset{*}{\underset{G'}{\Rightarrow}} S_{a_1}S_{a_2} \cdots S_{a_n}$$

因为$S_{a_i} \overset{*}{\underset{G_{a_i}}{\Longrightarrow}} w_i$，$S \overset{*}{\underset{G'}{\Rightarrow}} S_{a_1}S_{a_2} \cdots S_{a_n} \overset{*}{\underset{G'}{\Rightarrow}} w_1w_2 \cdots w_n$

所以$w \in \mathbf{L}(G')$，**即**$s(L) \subseteq \mathbf{L}(G')$。

P'包括每个P_a中的产生式

$T' = \bigcup_{a \in T} T_a$

- 证明（必要性$\mathbf{L}(G') \subseteq s(L)$）:
- 因为G'的终结符仅能由每个S_a派生，因此对$\forall w \in \mathbf{L}(G')$有

$$S \overset{*}{\underset{G'}{\Rightarrow}} \alpha = S_{a_1} S_{a_2} \cdots S_{a_n} \overset{*}{\underset{G'}{\Rightarrow}} w$$

因为G'中的每个S_a在G中是终结符a, 所以有

$$S \overset{*}{\underset{G}{\Rightarrow}} a_1 a_2 \cdots a_n = x \in L$$

又因为$\alpha = S_{a_1} \cdots S_{a_n} \overset{*}{\underset{G'}{\Rightarrow}} w = w_1 \cdots w_n$，所以 $S_{a_i} \overset{*}{\underset{G'}{\Rightarrow}} w_i$，即 $w_i \in s(a_i)$, 那么$w = w_1 w_2 \cdots w_n \in s(a_1) s(a_2) \cdots s(a_n) = s(a_1 a_2 \cdots a_n) = s(x) \subseteq s(L)$

所以$w \in s(L)$, 即$\mathbf{L}(G') \subseteq s(L)$，又由前$s(L) \subseteq \mathbf{L}(G')$，因此$\mathbf{L}(G') = s(L)$。

P'包括(i)每个P_a中的产生式；(ii) P的产生式，但是要替换产生式中的**每个终结符a为文法G_a的开始符号S_a**。

- **定理:上下文无关语言在并，连接，闭包，正闭包，同态下封闭。**
- **证明1:设**$\Sigma = \{1,2\}$, L_1, L_2**是任意CFL。定义代换**

$$s(1) = L_1, s(2) = L_2$$

语言$\{1,2\}, \{12\}, \{1\}^*$**和**$\{1\}^+$**显然都是CFL，那么**

（1）由$s(\{1,2\}) = s(1) \cup s(2) = L_1 \cup L_2$**，所以并运算封闭；**

（2）由$s(\{12\}) = s(12) = s(\varepsilon)s(1)s(2) = L_1L_2$**，所以连接运算封闭；**

- (3) 闭包运算封闭，因为

$$
\begin{aligned}
s(\{1\}^*) &= s(\{\varepsilon, 1, 11, \cdots\}) \\
&= s(\varepsilon) \cup s(1) \cup s(11) \cup \cdots \\
&= \{\varepsilon\} \cup s(1) \cup s(1)s(1) \cup \cdots \\
&= \{\varepsilon\} \cup L_1 \cup L_1 L_1 \cup \cdots \\
&= {L_1}^*
\end{aligned}
$$

- (4) 正闭包运算封闭的证明类似。

CFL对并/连接/闭包/同态封闭

- (5) 若h是Σ上的同态， L是Σ上的CFL， 对$\forall a \in \Sigma$ 令代换$s(a) = \{h(a)\}$, 则

$$h(L) = \{h(w) | w \in L\} = \bigcup_{w \in L} \{h(w)\} = \bigcup_{w \in L} s(w) = s(L)$$

所以同态运算封闭。

定义:若 Σ 和Γ是两个字母表, 同态定义为函数 $h: \Sigma \rightarrow \Gamma^*$ （即将Σ上的一个符号映射为Γ上的符号串）

- (5) 若h是Σ上的同态，L是Σ上的CFL，对$\forall a \in \Sigma$令代换$s(a) = \{h(a)\}$, 则

$$h(L) = \{h(w) | w \in L\} = \bigcup_{w \in L} \{h(w)\} = \bigcup_{w \in L} s(w) = s(L)$$

所以同态运算封闭。

对于某个 $w \in L$，令$w = w_1 w_2 \cdots w_n (w_i \in \Sigma)$，则有$s(w) = s(w_1 w_2 \cdots w_n)$。
由代换性质，有:

$$\begin{aligned} s(w) &= s(w_1)s(w_2)\cdots s(w_n) \\ &= \{h(w_1)\}\{h(w_2)\}\cdots\{h(w_n)\} \\ &= \{h(w_1)h(w_2)\cdots h(w_n)\} \\ &= \{h(w)\} \end{aligned}$$

定义:若 Σ 和Γ是两个字母表, 同态定义为函数 $h: \Sigma \rightarrow \Gamma^*$ （即将Σ上的一个符号映射为Γ上的符号串）

- 证明2(用文法证明):并/连接/闭包的封闭性。设CFL L_1, L_2的文法分别为

$$G_1 = (V_1, T_1, P_1, S_1), G_2 = (V_2, T_2, P_2, S_2)$$

那么，分别构造

（1） $L_1 \cup L_2$的文法为

$$G_{union} = (V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, P_1 \cup P_2 \cup \{\boldsymbol{S \to S_1 | S_2}\}, S)$$

（2） $L_1 L_2$的文法为

$$G_{concat} = (V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, P_1 \cup P_2 \cup \{\boldsymbol{S \to S_1 S_2}\}, S)$$

（3） ${L_1}^*$的文法为

$$G_{closure} = (V_1 \cup \{S\}, T_1, P_1 \cup \{\boldsymbol{S \to S_1 S | \varepsilon}\}, S)$$

CFL对反转封闭

- **定理：如果L是CFL，那么L^R也是CFL。**
- **证明：设L的文法$G = (V,T,P,S)$，构造文法**

$$G' = (V,T,\{A \to \alpha^R | A \to \alpha \in P\},S)$$

则$L(G') = L^R$**。**

例：对于一个有产生式$\boldsymbol{S \to 0S1}$|10的文法G，反转语言的文法的产生式为$S \to 1S0|10$

- **定理:如果L是CFL，那么L^R也是CFL。**
- **证明:设L的文法$G = (V, T, P, S)$, 构造文法**

$$G' = (V, T, \{A \to \alpha^R \mid A \to \alpha \in P\}, S)$$

则$L(G') = L^R$。

用归纳法证明由此形式构造的文法G'，其对应的语言是L^R。具体需要证明的是对于任意的变元$A \in V$, $u \in (V \cup \Sigma)^*$，如果$A \overset{*}{\Rightarrow}_G u$需要$n$步派生的话，那么当且仅当$A \overset{*}{\Rightarrow}_{G'} u^R$需要$n$步派生。

- **定理:如果L是CFL，那么L^R也是CFL。**
- **证明:设L的文法$G = (V, T, P, S)$, 构造文法**

$$G' = (V, T, \{A \to \alpha^R | A \to \alpha \in P\}, S)$$

则$L(G') = L^R$。

归纳基础:$n = 0$时，若$A \overset{0}{\Rightarrow}_G u$，则有$u = u^R = A$，因此$A \overset{0}{\Rightarrow}_{G'} u^R$。反之亦成立;

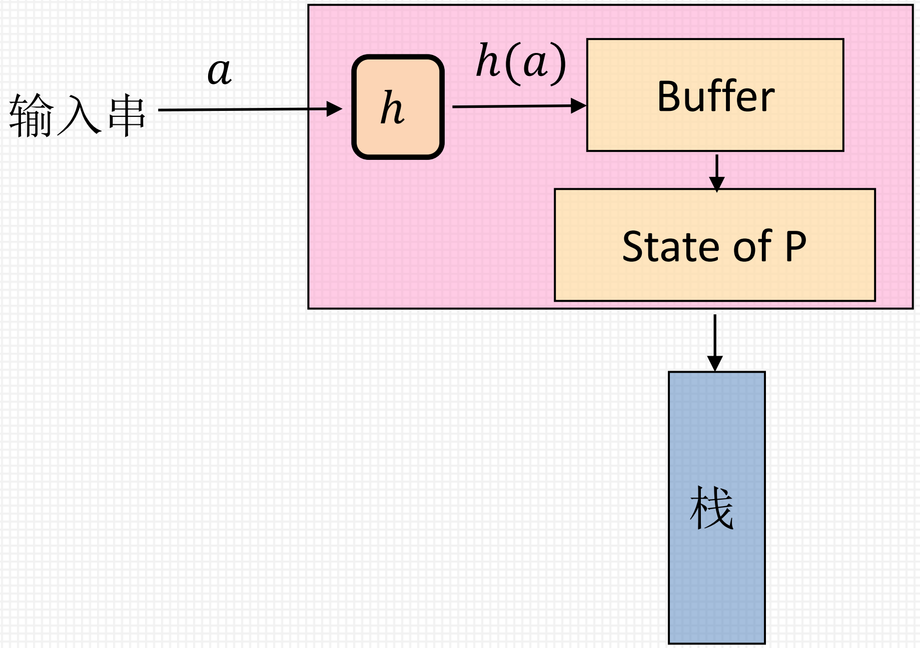
归纳假设：对于任意的$n > 0, A \in V, u \in (V \cup \Sigma)^*$，$A \overset{n-1}{\Longrightarrow}_G u$当且仅当$A \overset{n-1}{\Longrightarrow}_{G'} u^R$。

派生步数为n步时，即若$A \overset{n}{\Rightarrow}_G u$，则有变元$C \in V$和$x, y, z \in (V \cup \Sigma)^*$, 使得$u = xyz$, $A \overset{n-1}{\Longrightarrow}_G xCz, C \Rightarrow_G y$。由归纳假设，有$A \overset{n-1}{\Longrightarrow}_{G'} z^R C x^R$，由$G'$的构造，有$C \Rightarrow_{G'} y^R$,因此推出$A \overset{n}{\Rightarrow}_{G'} z^R y^R x^R = (xyz)^R = u^R$。相似的，由$A \overset{n}{\Rightarrow}_{G'} u^R$能推出$A \overset{n}{\Rightarrow}_{G'} u$。

因此，$A \overset{*}{\Rightarrow}_G u$需要$n$步派生的话，那么当且仅当$A \overset{*}{\Rightarrow}_{G'} u^R$需要$n$步派生。

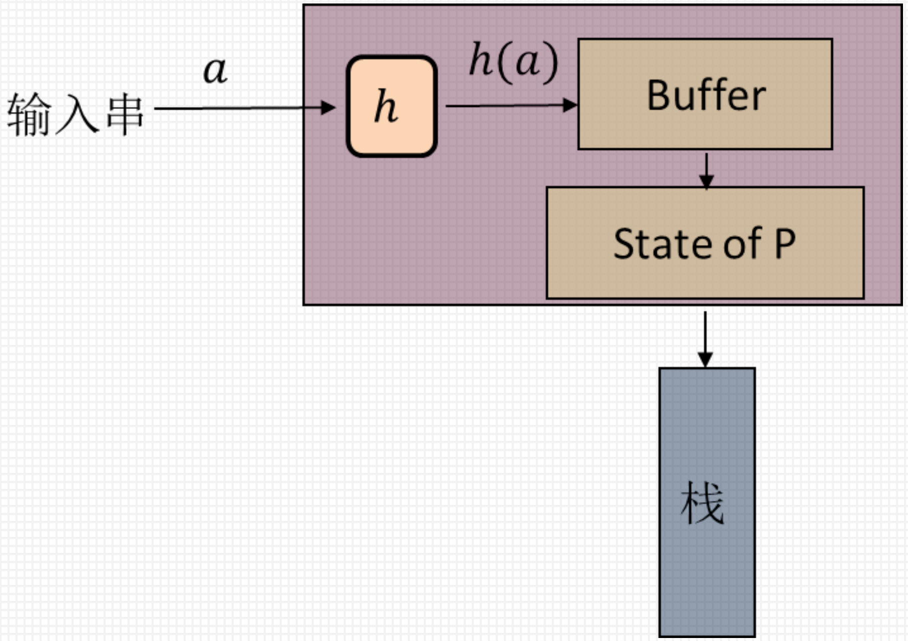
CFL对逆同态封闭

- **定理：如果L是字母表Δ上的CFL，h是字母表Σ 到 Δ^*的同态，那么$h^{-1}(L)$也是CFL。**
- **证明：基本思路：设PDA $P = (Q, \Delta, \Gamma, \delta, q_0, Z_0, F)$，$\mathbf{L}(P) = L$。构造$\mathbf{L}(P') = h^{-1}(L)$的PDA。在$P'$的状态中，使用缓冲(Buffer)，暂存字符$a \in \Sigma$的同态串$h(a)$。**



- P'的输入带上要消耗一个符号a时，先生成符号串$h(a)$，放入buffer中（buffer相当于P的输入带），然后调用P的状态转移函数去消耗$h(a)$。
- 当消耗完P'的输入带上的符号串w的时候，P是消耗$h(w)$。如果$h(w)$能被P接受，则说明w能被P'接受，即$h^{-1}(w)$ 能被P'接受。

- **证明：设PDA** $P=(Q,\Delta,\Gamma,\delta,q_0,Z_0,F)$，$\mathbf{L}(P)=L$。**构造**$\mathbf{L}(P')=h^{-1}(L)$**的PDA** P'：$P'=(Q',\Sigma,\Gamma,\delta',[q_0,\varepsilon],Z_0,F\times\{\varepsilon\})$



(1) Q'中的状态为$[q\in Q,x]$。x为Σ中某个终结符a的$h(a)$的后缀，用于表示P'缓冲中还未被消耗的串。初始时，buffer上没有字符，是ε。因此初始状态是$[q_0,\varepsilon]$

(2) δ'的定义如下：

即对P'输入带上的任意字符a, 把$h(a)$装载到buffer中，该buffer即P输入带上待消耗的符号串。

- $\delta'([q,\varepsilon],a,X)=\{([q,h(a)],X)\}$, 对$\forall q\in Q,\forall a\in\Sigma,\forall X\in\Gamma$。

即P'模拟P处理存在缓冲区中的符号串，在这个过程中，P'的输入带上不消耗新的符号。

- 若$\delta(q,b,X)$包含(p,γ)，这里$b\in\Delta$或者$b=\varepsilon$，则$\delta'([q,bx],\varepsilon,X)$包含$([p,x],\gamma)$

- **证明：设PDA** $P = (Q, \Delta, \Gamma, \delta, q_0, Z_0, F)$，$\mathbf{L}(P) = L$。

构造$\mathbf{L}(P') = h^{-1}(L)$**的PDA** P':

$$P' = (Q', \Sigma, \Gamma, \delta', [q_0, \varepsilon], Z_0, F \times \{\varepsilon\})$$

(3)终态定义为$[q \in F, \varepsilon]$**, 即**q**为**P**的终态，缓冲中已无待消耗的符号，即为空串**ε

P'**和**P**之间的关系可以这样表示：**

$(q_0, h(w), Z_0) \vdash_P^* (p, \varepsilon, \gamma)$**当且仅当**$([q_0, \varepsilon], w, Z_0) \vdash_{P'}^* ([p, \varepsilon], \varepsilon, \gamma)$

CFL 对交/补运算不封闭

- **CFL在交运算下不封闭**
- **例如，语言**

$$L_1 = \{0^n 1^n 2^i \mid n \geq 1, i \geq 1\}$$
$$L_2 = \{0^i 1^n 2^n \mid n \geq 1, i \geq 1\}$$

都是CFL, 而

$$L_1 \cap L_2 = \{0^n 1^n 2^n \mid n \geq 1\}$$

不是CFL。

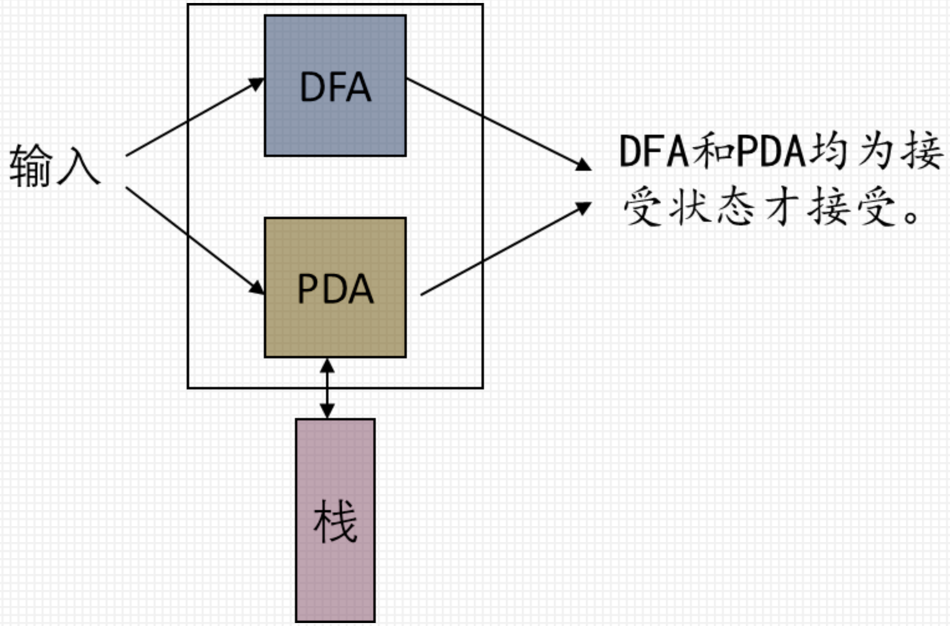
- **CFL在补运算下不封闭**

因为$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$

如果补运算封闭，则由该关系式得出交运算封闭（因为CFL在并运算下也封闭），但是交运算是不封闭的。因此补运算补封闭。

- **定理:若L是CFL且R是正则语言，则$L \cap R$是CFL。**

- 设DFA $D = (Q_1, \Sigma, \delta_1, q_1, F_1)$且$\mathbf{L}(D) = R$
- 设PDA $P = (Q_2, \Sigma, \Gamma, \delta_2, q_2, Z_0, F_2)$且$\mathbf{L}(P) = L$
- $L \cap R$表示同时被D和P接受的语言，即从D的开始状态经过字符串w到终止状态F_1，且从P的开始状态经过字符串w到终止状态F_2。



- 设DFA $D=(Q_1,\Sigma,\delta_1,q_1,F_1)$且$\mathbf{L}(D)=R$
- 设PDA $P=(Q_2,\Sigma,\Gamma,\delta_2,q_2,Z_0,F_2)$且$\mathbf{L}(P)=L$
- $L\cap R$的PDA P'可以构造如下：

构造PDA $P'=(Q_1\times Q_2,\Sigma,\Gamma,\delta,[q_1,q_2],Z_0,F_1\times F_2)$

其中δ为:

$\delta([p,q],a,Z)=\{([p,s],\beta)|(s,\beta)\in\delta_2(q,a,Z)\}$，若$a=\varepsilon$。

$\delta([p,q],a,Z)=\{([r,s],\beta)|r=\delta_1(p,a),(s,\beta)\in\delta_2(q,a,Z)\}$，若$a\neq\varepsilon$。

- **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$L \cap a^* b^* c^* = \{a^n b^n c^n | n \geq 0\}$不是CFL。

(3) 由CFL与正则语言的交还是CFL,所以L不可能是CFL。

主要内容

- 上下文无关语言的泵引理
- 上下文无关语言的封闭性
- **上下文无关语言的判定性质**
- 乔姆斯基文法体系

- **1. 空性：只需判断文法的开始符号S是否为非产生的。**
- **回顾非产生符号定义：**
- **对于CFG $G = (V, T, P, S)$, 符号$X \in (V \cup T)$，如果$\alpha X \beta \overset{*}{\Rightarrow} w$ ($w \in T^*$)，称X是产生的(generating)。**
- **计算“产生的”符号集**
  - 每个T中的符号都是产生的
  - $A \to \alpha \in P$ 且α中符号都是产生的, 则A是产生的

- **2. 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, 若有 A → BC, 则 A 到 B 和 C 各画一条有向边;
  - 检查图中是否有循环。
  - 当且仅当图中没有循环时，L是有穷的；否则，L是无穷的。

- 3. 成员性：利用CNF范式， 有CYK算法检查串w是否属于L。
- CYK算法的基本思想由J. Cocke， D. Younger， T. Kasami三个人独立发现

- **对于CNF** $G=(V,T,P,S)$, **CYK算法可以检查**$w=a_1a_2\dots a_n$**是否属于**$\mathbf{L}(G)$?
- **令**$X_{ij}=\{A\in V|A\overset{*}{\Rightarrow}a_i\dots a_j\}$, **即**$X_{ij}$**为能派生出**$a_i\dots a_j$**的变元的集合。计算出每个变元**$X_{ij}$。
- **判断**S**是否在**X_{1n}**中。如果在的话，则说明由**S**可以派生出**$w=a_1a_2\dots a_n$，**即**w**属于**$\mathbf{L}(G)$。

- **构造三角形状的表格：每一行元素是变元X_{ij}，对应于某种长度的子串。**
  - 最下面的一行对应于长度为1的子串
  - 由下往上的第二行对应于长度为2的子串
  - …
  - 最上面的一行对应于长度为n的子串（即w本身）

- **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。

- **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：$X_{ij} = \{A|i \leq k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$
    - 比如，对于X_{12}， 1$\leq k < 2$,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$\leq k < 3$，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

- 例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？

$$S \to AB|BC$$
$$A \to BA|a$$
$$B \to CC|b$$
$$C \to AB|a$$

- 例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他：$X_{ij} = \{A|i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$$S \to AB|BC$$
$$A \to BA|a$$
$$B \to CC|b$$
$$C \to AB|a$$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$$X_{1,1} = \{B\}, X_{2,2} = \{A, C\}, X_{3,3} = \{A, C\},$$
$$X_{4,4} = \{B\}, X_{5,5} = \{B\}$$

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A | A \to a_i \in P\}$
其他：$X_{ij} = \{A | i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$S \to AB | BC$
$A \to BA | a$
$B \to CC | b$
$C \to AB | a$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$\{B\}$	$\{A, C\}$	$\{A, C\}$	$\{B\}$	$\{A, C\}$
b	a	a	b	a

$$X_{1,1} = \{B\}, X_{2,2} = \{A, C\}, X_{3,3} = \{A, C\},$$
$$X_{4,4} = \{B\}, X_{5,5} = \{B\}$$

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他：$X_{ij} = \{A|i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$S \to AB|BC$
$A \to BA|a$
$B \to CC|b$
$C \to AB|a$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$\{B\}$	$\{A,C\}$	$\{A,C\}$	$\{B\}$	$\{A,C\}$
b	a	a	b	a

X_{12}，需要考虑$X_{11}X_{2,2} = \{\mathrm{B}\}\{\mathrm{A,C}\} = \{\mathrm{BA,BC}\}$

寻找能产生BA或者BC的产生式，得到

$X_{12} = \{S, A\}$

CYK算法

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \rightarrow a_i \in P\}$
其他：$X_{ij} = \{A|i \leq k < j, BC \in X_{ik}X_{k+1,j}, A \rightarrow BC\}$

$S \rightarrow AB|BC$
$A \rightarrow BA|a$
$B \rightarrow CC|b$
$C \rightarrow AB|a$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$\{S, A\}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$\{B\}$	$\{A, C\}$	$\{A, C\}$	$\{B\}$	$\{A, C\}$
b	a	a	b	a

X_{12}，需要考虑$X_{11}X_{2,2} = \{\mathrm{B}\}\{\mathrm{A}, \mathrm{C}\} = \{\mathrm{BA}, \mathrm{BC}\}$
寻找能产生BA或者BC的产生式，得到$X_{12} = \{S, A\}$

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他：$X_{ij} = \{A|i \leq k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$$S \to AB|BC$$
$$A \to BA|a$$
$$B \to CC|b$$
$$C \to AB|a$$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$\{S,A\}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$\{B\}$	$\{A,C\}$	$\{A,C\}$	$\{B\}$	$\{A,C\}$
b	a	a	b	a

X_{23}，需要考虑$X_{22}X_{33} = \{\text{A, C}\}\{\text{A, C}\} = \{\text{AA, AC, CA, CC}\}$
寻找能产生集合中变元串的产生式，得到
$X_{23} = \{B\}$

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他： $X_{ij} = \{A|i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$S \to AB|BC$
$A \to BA|a$
$B \to CC|b$
$C \to AB|a$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$\{S,A\}$	$\{B\}$	$X_{3,4}$	$X_{4,5}$	
$\{B\}$	$\{A,C\}$	$\{A,C\}$	$\{B\}$	$\{A,C\}$
b	a	a	b	a

X_{23}， 需要考虑$X_{22}X_{33} = \{A, C\}\{A, C\} = \{AA, AC, CA, CC\}$
寻找能产生集合中变元串的产生式，得到$X_{23} = \{B\}$

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他：$X_{ij} = \{A|i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$S \to AB|BC$
$A \to BA|a$
$B \to CC|b$
$C \to AB|a$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$\{S,A,C\}$				
$\emptyset$	$\{S,A,C\}$			
$\emptyset$	$\{B\}$	$\{B\}$		
$\{S,A\}$	$\{B\}$	$\{S,C\}$	$\{S,A\}$	
$\{B\}$	$\{A,C\}$	$\{A,C\}$	$\{B\}$	$\{A,C\}$
b	a	a	b	a

以此类推，可计算出每个$X_{i,j}$。在计算每个$X_{i,j}$时，如果找不到相应的变元，则为$\emptyset$。

CYK算法

- **例：对于如下的CNF文法，判断$w = baaba$是否在该文法表示的语言中？**

基础: $X_{ii} = \{A|A \to a_i \in P\}$
其他：$X_{ij} = \{A|i \le k < j, BC \in X_{ik}X_{k+1,j}, A \to BC\}$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
b	a	a	b	a

$\{S,A,C\}$				
$\emptyset$	$\{S,A,C\}$			
$\emptyset$	$\{B\}$	$\{B\}$		
$\{S,A\}$	$\{B\}$	$\{S,C\}$	$\{S,A\}$	
$\{B\}$	$\{A,C\}$	$\{A,C\}$	$\{B\}$	$\{A,C\}$
b	a	a	b	a

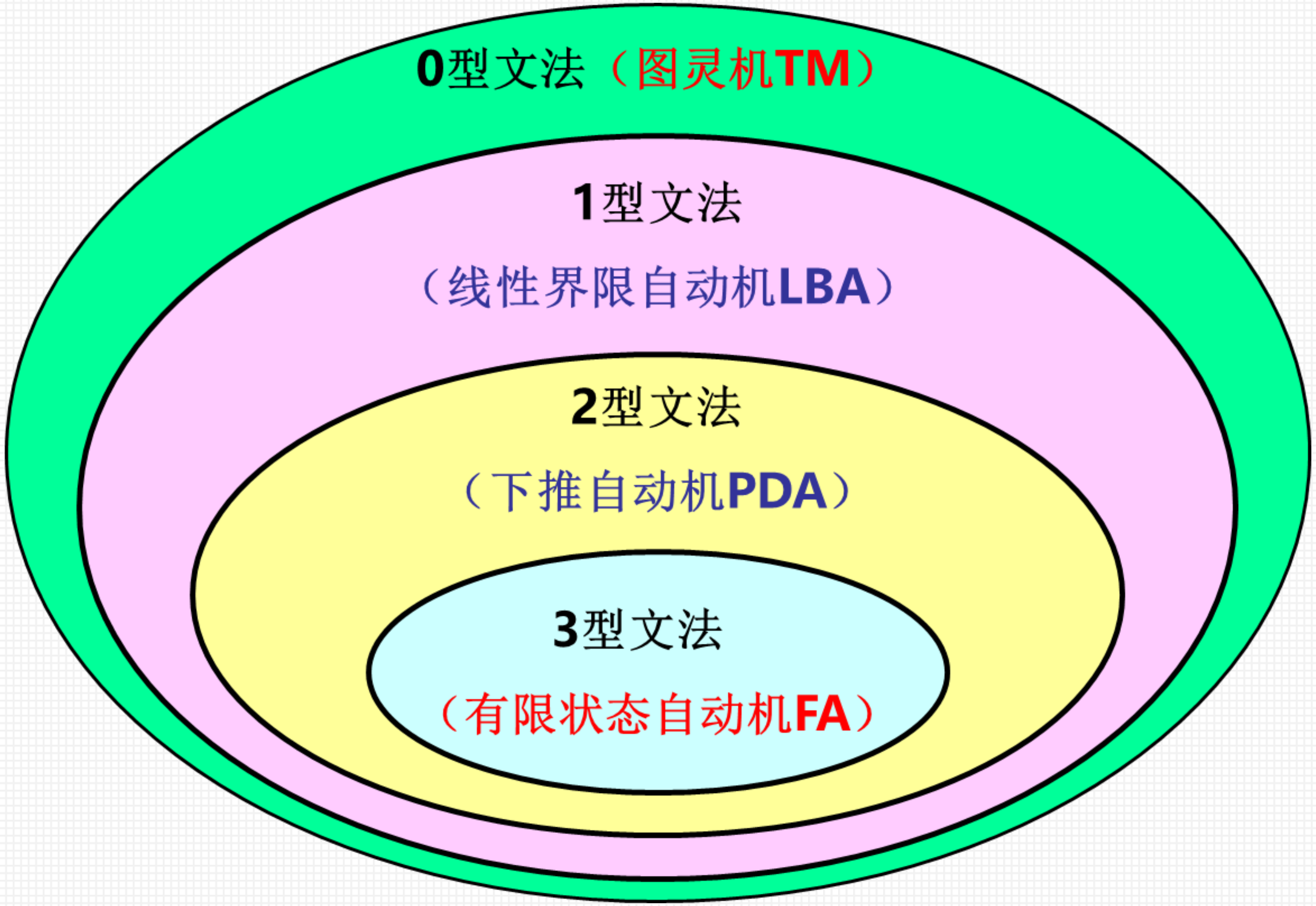
$S \to AB|BC$
$A \to BA|a$
$B \to CC|b$
$C \to AB|a$

由表格中结果可见，$X_{1,5} = \{S,A,C\}$,因此$S \in X_{1,5}$，可以派生出$w = baaba$。

不可判定的CFL问题

- 判断 CFG G 是否歧义的?
- 判断 CFL 是否固有歧义的?
- 两个 CFL 的交是否为空?
- 两个 CFL 是否相同?
- 判断 CFL 的补是否为空? （尽管有算法判断 CFL 是否为空）
- 判断 CFL 是否等于Σ^*?

乔姆斯基文法体系



- **如果文法** $G = (V, T, P, S)$, **符号串** $\alpha \in (V \cup T)^* V (V \cup T)^*, \beta \in (V \cup T)^*$, **产生式都形如**

$$\alpha \to \beta,$$

即每个产生式的左部α**中至少要有一个变元，那么：**

3型文法：

若$\alpha \to \beta$ **都形如**$A \to aB$ **或**$A \to a$**或**$A \to \varepsilon$ $(A, B \in V, \ a \in T)$, **称**G**为3型文法或正则文法。**

例：

S→ ε, S→ aT, S → bT

T → a, T → b, T → aS, T → bS

所表示的语言包含由任意个a,b组成，且长度为偶数的符号串。

- **如果文法** $G=(V,T,P,S)$, **符号串** $\alpha\in(V\cup T)^*V(V\cup T)^*, \beta\in(V\cup T)^*$，**产生式都形如**

$$\alpha\rightarrow\beta,$$

即每个产生式的左部α**中至少要有一个变元，那么:**

2型文法:

若$\alpha\in V$, **称**G**为2型文法或上下文无关文法。**

- **如果文法** $G = (V,T,P,S)$, **符号串** $\alpha \in (V \cup T)^* V (V \cup T)^*, \beta \in (V \cup T)^*$， **产生式都形如**

$$\alpha \rightarrow \beta,$$

即每个产生式的左部α**中至少要有一个变元，那么：**

1型文法：

若$|\alpha| \leq |\beta|$, 称G为1型文法或上下文有关文法。

例：

AB → CDB

AB → CdEB

ABcd → abCDBcd

B → b

- **如果文法** $G = (V, T, P, S)$, **符号串** $\alpha \in (V \cup T)^* V (V \cup T)^*$, $\beta \in (V \cup T)^*$，**产生式都形如**

$$\alpha \to \beta,$$

即每个产生式的左部α**中至少要有一个变元，那么：**

0型文法：

没有任何限制的文法(unrestricted grammar)。

小结

- **CFL的泵引理及其应用：证明不是CFL**
- **CFL的封闭性**
  - 封闭运算：并、乘、闭包、代换、同态映射、逆同态映射
  - 不封闭运算：交、补
- **CFL的判定性质**
  - 可判定性：空性，无穷性，成员性
  - 不可判定性
- **乔姆斯基文法体系**