Chapter 7

上下文无关语言的性质

7.1 上下文无关语言

直观来讲, 正则语言是从单独的字符串以并 *(union)*, 连接 *(concatenation)* 和重复 *(repetition)* 构建而来的; 而上下文无关语言则是以并, 连接和递归 *(recursion)* 构建而来的.

任何 Σ 上的所有语言

不妨设 $\Sigma = \{a\}$, 对任何 $0 \le x < 1$ 的实数 x, 定义语言

$$L_x = \left\{ a^n \,\middle|\, x \cdot 2^n \bmod 1 \ge \frac{1}{2} \right\},$$

即 $a^n \in L_x$ 当且仅当 x 二进制表示的第 $n+1$ 位为 1.

1. 如果 $x \ne y$, 则 x 和 y 一定有某些位不同, 所以 $L_x \ne L_y$;
2. 所以 Σ 上的所有语言, 至少与 0 和 1 之间的实数一样多;
3. 因此, Σ 上的所有语言是不可数的.

那么

“办法总比困难多!”——真的吗?

任何 Σ 上的所有 CFL

任何 CFG $G = (V, \Sigma, P, S)$ 可由符号集 $V \cup \Sigma \cup \{\varepsilon, \to, |, \diamondsuit\}$ 编码.

- 如文法 $S \to A \mid B$, $A \to aA \mid aC$, $B \to Bb \mid Cb$, $C \to \varepsilon \mid aCb$ 可编码为

$$S{\to}A|B\Diamond A{\to}aA|aC\Diamond B{\to}Bb|Cb\Diamond C{\to}\varepsilon|aCb;$$

- 用 0/1 编码这些符号

$$\begin{array}{lll} \varepsilon \mapsto 10 & a \mapsto 110 & S \mapsto 1110 \\ \to \mapsto 100 & b \mapsto 1100 & A \mapsto 11100 \\ | \mapsto 1000 & & B \mapsto 111000 \\ \Diamond \mapsto 10000 & & C \mapsto 1110000. \end{array}$$

- 文法编码再转换为 0/1 字符串

11101001110010001110001000011100100110111001000110
11100001000011100010011100011001000111000011001000
0111000010010100011011100001100;

- 当作二进制表示则为整数

$$248602534784558144413324333914267072 6924.$$

- 而 Σ 上两个文法如果不同, 这样编码会得到不同的整数;
- 因此 Σ 上所有 CFL 至多与正整数一样多, 是可数的.
- 因此, 并非所有的语言都是 CFL.

7.2 上下文无关语言的泵引理

7.2.1 上下文无关语言的泵引理

A→BC
A→a

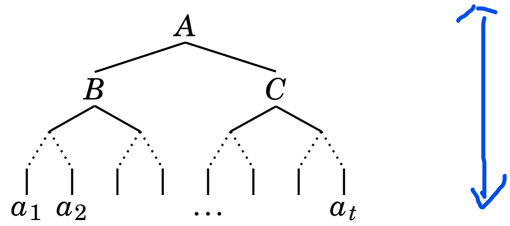
语法分析树的大小

定理 33. 对于乔姆斯基范式文法 $G = (V, T, P, S)$ 的语法树, 如果产物为终结符串 w, 且树的高度是 n, 那么 $|w| \le 2^{n-1}$.

证明: 对树的高度归纳.

归纳基础: 高度为 1 时, 只能是 $\begin{array}{c} A \\ | \\ a \end{array}$, 显然成立.

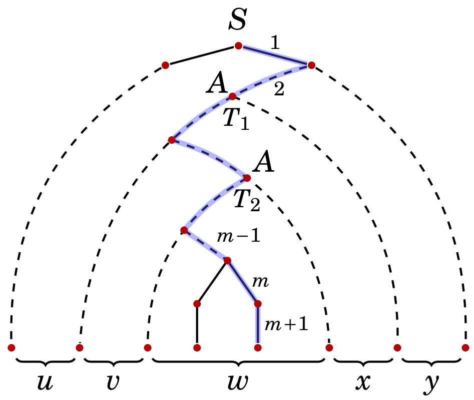
归纳递推: 高度为 n 时, 根节点一定是 $A \to BC$, 而 B 和 C 子树高度最多为 $n-1$, 由归纳假设, 产物最长都为 2^{n-2}. 因此整棵树产物最长 $2^{n-2} + 2^{n-2} = 2^{n-1}$. $\square$



上下文无关语言的泵引理

定理 34. *如果语言 L 是 CFL, 那么存在正整数 N, 它只依赖于 L, 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. $\forall i \geq 0$, $uv^iwx^iy \in L$.



证明:

1. 设 CNF 格式的 CFG G 接受语言 $L - \{\varepsilon\}$. 在 CNF 文法的派生树中, 如果从树根到叶子的最长路径长度为 k, 则该树产物的长度最多为 2^{k-1}. 因为该树内节点构成的是二叉树, 当最长路径为 k 时, 内节点二叉树部分最长路径长度为 $k-1$, 叶子最多为满二叉树时的 2^{k-1} 个. 如果设文法 G 中变元数 $|V| = m$, 则令 $N = 2^m$, 那么若有字符串 $z \in L(G)$, 且 $|z| \geq N$.
2. 则 z 的派生树内节点是二叉树, 树中最长路径长度至少也是 $m+1$, 该路径上节点至少有 $m+2$ 个, 除最后一个节点外, 其余标记都是变元.
3. 在该路径上接近树底部由下至上 $m+1$ 个内节点的标记中, 必有两个节点 T_2 和 T_1 标记了相同的变元 A. 不妨设这两棵 A 子树分别为 T_1 和 T_2, 且 T_1 比 T_2 更接近树根.

4. 那么, 若记 A 子树 T_2 的产物为 w, 且又因为它是 T_1 的子树, 那么 T_1 的产物可记为 vwx, 则有 $A \overset{*}{\Rightarrow} vAx$ 和 $A \overset{*}{\Rightarrow} w$.

5. 那么, 对 $\forall i \geq 0$, 有 $A \overset{*}{\Rightarrow} v^i w x^i$. 又因为 vwx 是 z 的一部分, 所以不妨设 $z = uvwxy$, 则 $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uv^i wx^i y$, 即 $\forall i \geq 0, uv^i wx^i y \in L$.

6. 又因为, 我们只考虑了接近树底部的 $m+1$ 个节点, 所以 T_1 子树中最长路径长度不超过 $m+1$, 那么 T_1 产物的长度不会超过 2^m, 所以 $|vwx| \leq 2^m = N$.

7. 而 T_1 子树派生 vwx 的第一个产生式必是 $A \to BC$, 那么 T_2 子树不是完全处于 B 中就是完全处于 C 中, 即 T_2 必定完全在 T_1 的左/右儿子中. 而不包括 T_2 的变元 B 或 C 都至少产生一个终结符, 所以 v 和 x 不可能同时为空, 即 $vx \neq \varepsilon$.

□

7.2.2 泵引理的应用

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

证明:

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 $vx \neq \varepsilon$.
4. 那么 vwx 只能包含一种或两种字符:
   - i 一种字符, 或为 0, 或为 1, 或为 2, 那么取 $i = 0$, 则 uv^0wx^0y 会只删除该字符的一部分, 所以 $uwy \notin L$;
   - ii 两种字符, 或为 0 和 1, 或为 1 和 2, 那么uv^0wx^0y 会只删除这两种字符的一部分, 所以也有 $uwy \notin L$;
5. 与泵引理 $uwy = uv^0wx^0y \in L$ 矛盾, 所以假设不成立.
6. L 不是上下文无关的. □

例 2. 证明 $L = \{ww \mid w \in \{0,1\}^*\}$ 不是上下文无关的.

(错误的) 证明: 假设 L 是 CFL. 取 $z = 0^N 1 0^N 1$, 那么 $z = uvwxy$ 为

$$z = \overbrace{\underbrace{00\cdots00}_{u}\ \underbrace{0}_{v}\ \underbrace{1}_{w}}^{0^N1}\overbrace{\underbrace{0}_{x}\ \underbrace{00\cdots01}_{y}}^{0^N1}$$

则对任意 $i \geq 0$, 有 $uv^i wx^i y \in L$, 满足泵引理. □

(正确的) 证明: 假设 L 是 CFL. 取 $z = 0^N 1^N 0^N 1^N$, 将 z 分为 $z = uvwxy$ 时

1. 若 vwx 在 z 中点的一侧, $uv^0 wx^0 y$ 显然不可能属于 L;
2. 若 vwx 包括 z 中点, 那么 $uv^0 wx^0 y$ 为 $0^N 1^i 0^j 1^N$, 也不可能属于 L.

所以假设不成立, L 不是 CFL. □

CFL 的泵引理同样只是必要条件

有些非 CFL, 泵引理对它们没有什么作用. 例如

$$L = \{ a^i b^j c^k d^l \mid i = 0 \text{ 或 } j = k = l \}$$

不是上下文无关的.

- 如果选 $z = b^j c^k d^l = uvwxy$, 则可以让 vwx 只含有 b, 那么对任何 m, 都有 $uv^m wx^m y \in L$;
- 如果选 $z = a^i b^j c^j d^j$, 则可以让 v 和 x 只含有 a, 那么对任何 m, 都有 $uv^m wx^m y \in L$.

所以无法使用泵引理证明 L 非 CFL.

Ogden 引理 (的较弱形式)

如果语言 L 是 CFL, 那么存在正整数 N, 它只依赖于 L, 对 $\forall z \in L$, 在 z 中至少 N 个任意位置作标记后, 就可以将 z 分为五部分 $z = uvwxy$ 满足:

1. v 和 x 一起至少含有一个标记位置;
2. vwx 中至多有 N 个标记位置;
3. $\forall i \geq 0$, $uv^i wx^i y \in L$.

7.3 上下文无关语言的封闭性

7.3.1 代换的封闭性

定义. 两个字母表 Σ 到 Γ 的函数 $s : \Sigma \to 2^{\Gamma^*}$ 称为代换 *(substitution)*. 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 $\forall L \subseteq \Sigma^*$,

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

定理 35. *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.*

构造方法

这门语言

设 CFL L 的文法 $G = (V, T^{\Sigma}, P, S)$, 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

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

$a \in T$, V_a

$a \in T$, T_a

1. $V' = V \cup \left(\bigcup_{a \in T} V_a\right)$
2. $T' = \bigcup_{a \in T} T_a$
3. P' 包括每个 P_a 和 P 中产生式, 但是要将 P 的产生式中每个终结符 a 均替换为文法 G_a 的开始符号 S_a.

证明: 对 $\forall w \in s(L)$, 那么一定存在某个 $x = a_1 a_2 \cdots a_n \in L$ 使

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

那么 w 可以分为 $w = w_1 w_2 \cdots w_n$ 且 $w_i \in s(a_i)$, 即

$$S_{a_i} \underset{G_{a_i}}{\overset{*}{\Rightarrow}} w_i.$$

由于 $S \underset{G}{\overset{*}{\Rightarrow}} x = a_1 a_2 \cdots a_n$, 所以

$$S \underset{G'}{\overset{*}{\Rightarrow}} S_{a_1} S_{a_2} \cdots S_{a_n} \underset{G'}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_n = w,$$

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

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

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

因为 G' 中的每个 S_a 在 G 中是终结符 a, 所以

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

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 $\mathbf{L}(G') \subseteq s(L)$. 因此 $\mathbf{L}(G') = s(L)$. □

例 3. 设 $L = \{\, w \in \{a,b\}^* \mid w \text{ 有相等个数的} a \text{和} b \,\}$, 代换

$$\begin{aligned} s(a) &= L_a = \{\, 0^n 1^n \mid n \ge 1 \,\} \\ s(b) &= L_b = \{\, ww^R \mid w \in (\mathbf{0}+\mathbf{1})^* \,\} \end{aligned}$$

求 $s(L)$ 的文法.

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

L_a 的文法为: $S_a \to 0S_a1 \mid 01$

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

那么 $s(L)$ 的文法为:

$$\begin{aligned} S &\to S_aSS_bS \mid S_bSS_aS \mid \varepsilon \\ S_a &\to 0S_a1 \mid 01 \\ S_b &\to 0S_b0 \mid 1S_b1 \mid \varepsilon \end{aligned}$$

7.3.2 并/连接/闭包/同态/逆同态/反转的封闭性

CFL 对并/连接/闭包/同态封闭

定理 36. 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

$$s(1) = L_1, \quad s(2) = L_2.$$

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

1. 由 $s\big(\{1, 2\}\big) = s(1) \cup s(2) = L_1 \cup L_2$, 所以并运算封闭;
2. 由 $s\big(\{12\}\big) = s(12) = s(\varepsilon)s(1)s(2) = L_1L_2$, 所以连接运算封闭;
3. 闭包和正比包运算封闭, 因为

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

$$= \{\varepsilon\} \cup s(1) \cup s(1)s(1) \cup \cdots$$
$$= L_1^*.$$

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

$$h(L) = \{\, h(w) \mid w \in L \,\} = \bigcup_{w \in L} \{h(w)\} = \bigcup_{w \in L} s'(w) = s'(L),$$

所以同态运算封闭. □

证明 2: 用文法证明并/连接/闭包的封闭性. 设 CFL L_1 和 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

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

2. $L_1 L_2$ 的文法为

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

3. L_1^* 的文法为

$$G_{\text{closure}} = (V_1 \cup \{S\},\ T_1,\ P_1 \cup \{S \to S_1 S \mid \varepsilon\},\ S).$$

再证明所构造文法的正确性, 略. □

CFL 对反转封闭

定理 37. *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

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

则 $L(G') = L^R$. 证明略. □

CFL 对逆同态封闭

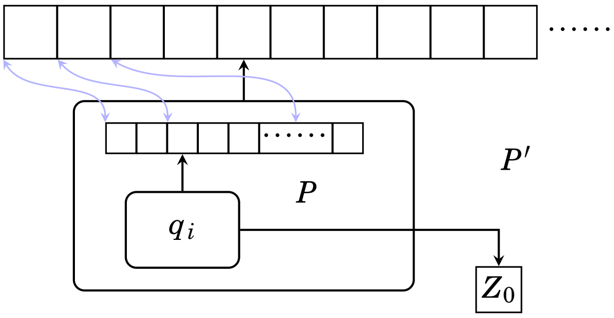
定理 38. 如果 L 是字母表 Δ 上的 *CFL*, h 是字母表 Σ 到 Δ^* 的同态, 那么 $h^{-1}(L)$ 也是 *CFL*.

证明:

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

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

在 P' 的状态中, 使用缓冲, 暂存字符 $a \in \Sigma$ 的同态串 $h(a)$ 的后缀.



1. $Q' \subset Q \times \Delta^*$: 状态 $[q,\bar{x}]$ 中的 $\bar{x}$ 为缓冲;
2. 设 $q \in Q$, 那么 δ' 定义如下:

   i $\forall [q,\bar{\varepsilon}] \in Q \times \{\bar{\varepsilon}\}$, $\forall a \in \Sigma$, $\forall X \in \Gamma$

   $$\delta'([q,\bar{\varepsilon}],a,X) = \{([q,h(a)],X)\}$$

   ii 若 $\delta(q,\bar{a},X) = \{(p_1,\beta_1),(p_2,\beta_2),\cdots,(p_k,\beta_k)\}$, 则

   $$\delta'([q,\overline{ax}],\varepsilon,X) = \{([p_1,\bar{x}],\beta_1),([p_2,\bar{x}],\beta_2),\cdots,([p_k,\bar{x}],\beta_k)\}$$

   这里 $\bar{a} \in \Delta \cup \{\varepsilon\}$, $\bar{x}$ 是某个 $h(a)$ 的后缀. □

7.3.3 交和补运算不封闭

CFL 对交运算不封闭

因为语言

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

都是 CFL, 而

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

不是 CFL.

思考

1. 显然存在 PDA P_1 和 P_2, 分别识别语言 L_1 和 L_2;
2. 是否可以同时运行 P_1 和 P_2, 来识别 $L_1 \cap L_2$?
3. 如果可以, 那为什么 $L_1 \cap L_2$ 仍然不是 CFL?

CFL 对补运算不封闭

因为

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

思考

$$L \text{ vs } \overline{L}.$$

直觉上一般认为, 一个问题和它的补, 似乎总是同一件事, 但其实不是.

- 如 $L = \{ww \mid w \in \Sigma^*\}$ 不是 CFL
- 而 $\overline{L} = \{x \mid x \text{ is not the form of } ww \text{ where } w, x \in \Sigma^*\}$ 是 CFL

定理 39. *若 L 是 CFL 且 R 是正则语言, 则 $L \cap R$ 是 CFL.*

证明: 设 DFA $D = (Q_1, \Sigma, \delta_1, q_1, F_1)$ 且 $\mathbf{L}(D) = R$, PDA $P = (Q_2, \Sigma, \Gamma, \delta_2, q_2, Z_0, F_2)$ 且 $\mathbf{L}(P) = L$, 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:

$$\delta([p,q], a, Z) = \begin{cases} \{([p,s], \beta) \mid (s, \beta) \in \delta_2(q, a, Z)\} & a = \varepsilon \\ \{([r,s], \beta) \mid r = \delta_1(p, a) \text{ 且 } (s, \beta) \in \delta_2(q, a, Z)\} & a \neq \varepsilon \end{cases}$$

再往证 $\mathbf{L}(P') = L \cap R$, 略. □

7.3.4 封闭性的应用

例 4. 请证明语言 L 不是 CFL

$$L = \{ w \in \{a, b, c\}^* \mid \mathrm{n}_a(w) = \mathrm{n}_b(w) = \mathrm{n}_c(w) \},$$

其中 $\mathrm{n}_a(w)$ 表示 w 中 a 的个数.

证明:

1. 因为 $\mathbf{a}^*\mathbf{b}^*\mathbf{c}^*$ 是正则语言,
2. 而 $L \cap \mathbf{a}^*\mathbf{b}^*\mathbf{c}^* = \{ a^n b^n c^n \mid n \geq 0 \}$ 不是 CFL,
3. 由 CFL 与正则语言的交还是 CFL, 所以 L 不可能是 CFL. □

?

例. 请证明语言 $L = \{ a^i b^j a^i b^j \mid i \geq 1, j \geq 1 \}$ 不是 CFL.

证明:

1. 因为 $\mathbf{a}^+\mathbf{b}^+\mathbf{a}^+\mathbf{b}^+$ 是正则语言,
2. 而 $L_{ww} = \{ ww \mid w \in \{a, b\}^* \}$ 不是 CFL,
3. 所以 $L_{ww} \cap \mathbf{a}^+\mathbf{b}^+\mathbf{a}^+\mathbf{b}^+ = L$ 也不可能是 CFL. □

7.4 上下文无关语言的判定性质

可判定的 *CFL* 问题

- 空性: 只需判断文法的开始符号 S 是否为非产生的.
- 有穷性和无穷性:
  1. 用不带无用符号的 CNF 的产生式画有向图;
  2. 变元为顶点, 若有 $A \to BC$, 则 A 到 B 和 C 各画一条有向边;
  3. 检查图中是否有循环.
- 成员性: 利用 CNF 范式, 有 CYK 算法检查串 w 是否属于 L.

CYK[1]算法

- CNF $G=(V,T,P,S)$, 以 $O(n^3)$ 时间检查 “$w=a_1a_2\cdots a_n\in\mathbf{L}(G)$?”
- 以动态规划方式, 在表中由下至上逐行计算 X_{ij}, 再检查 “$S\in X_{1n}$?”

$$X_{ij}=\left\{A\in V \,\middle|\, A\overset{*}{\Rightarrow}a_ia_{i+1}\cdots a_j,\ 1\le i\le j\le n\right\},$$

- 计算首行

$$X_{ii}=\left\{A \,\middle|\, A\to a_i\in P\right\}$$

共有 $O(n)$ 个 X_{ii} 需要计算.

- 计算其他

$$X_{ij}=\left\{A \,\middle|\, \begin{array}{l} i\le k<j, \\ BC\in X_{ik}X_{k+1,j}, \\ A\to BC\in P \end{array}\right\}$$

共计算 $O(n^2)$ 个 X_{ij}, 而每个要计算 $O(n)$ 组 $X_{ik}X_{k+1,j}$, 因此总时间复杂度为 $O(n^3)$.

$$\begin{array}{ccccccccc}
 & & & & X_{15} & & & & \\
 & & & X_{14} & & X_{25} & & & \\
 & & X_{13} & & X_{24} & & X_{35} & & \\
 & X_{12} & & X_{23} & & X_{34} & & X_{45} & \\
X_{11} & & X_{22} & & X_{33} & & X_{44} & & X_{55} \\
\hline
a_1 & & a_2 & & a_3 & & a_4 & & a_5
\end{array}$$

例 5. CNF G 如下, 用 CYK 算法判断 $bbabaa\in\mathbf{L}(G)$?

$$\begin{aligned}
S&\to AB\mid BC\\
A&\to BA\mid a\\
B&\to CC\mid b\\
C&\to AB\mid a
\end{aligned}$$

[1] J. Cocke, D. Younger, T. Kasami 分别独立发现了算法的基本思想

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

因为 $S \in X_{16} = \{S,A\}$, 所以 $bbabaa \in \mathbf{L}(G)$.

不可判定的 *CFL* 问题

1. 判断 CFG G 是否歧义的?
2. 判断 CFL 是否固有歧义的?
3. 两个 CFL 的交是否为空?
4. 两个 CFL 是否相同?
5. 判断 CFL 的补是否为空? 尽管有算法判断 CFL 是否为空
6. 判断 CFL 是否等于 Σ^*?

7.5 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:

1. 称 G 为 0 型文法或短语结构文法 (PSG, *Phrase Structure Grammar*), $L(G)$ 称为 0 型语言, 短语结构语言 (PSL, *Phrase Structure Language*), 或递归可枚举语言 (*Recursively Enumerable Language*);
2. 若 $|\beta| \ge |\alpha|$, 称 G 为 1 型文法或上下文有关文法 (CSL, *Context-Sensitive Language*), $L(G)$ 称为 1 型语言或上下文有关语言 (CSL, *Context-Sensitive Language*);
3. 若 $\alpha \in V$, 称 G 为 2 型文法或上下文无关文法, $L(G)$ 称为 2 型语言或上下文无关语言;

4. 若 $\alpha \to \beta$ 都形如 $A \to aB$ 或 $A \to a$, 其中 $A \in V, a \in T$, 称 G 为 3 型文法, 右线性文法或正则文法, $L(G)$ 称为 3 型语言或正则语言.

乔姆斯基文法体系的四种类型中, 0 型文法的能力等价于图灵机, 1 型文法的能力等价于线性界限自动机. 2 型文法能力等价于非确定的下推自动机. 3 型文法也称右线性文法, 能力等价于有穷自动机. 文法描述语言的能力, 0 型文法最强, 3 型文法最弱.

7.6 练习题

1. [Exercise 7.1.7] Suppose G is a CFG with p productions, and no production body longer than n. Show that if $A \overset{*}{\underset{G}{\Rightarrow}} \varepsilon$, then there is a derivation of ε from A of no more than $(n^p - 1)/(n - 1)$ steps. How close can you actually come to this bound?

chunyu@hit.edu.cn
http://iilab.net/chunyu