上下文无关语言的性质

- 上下文无关语言的泵引理
- 上下文无关语言的封闭性　　　　交、补不封闭
  - 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
  - 封闭性的应用
- 上下文无关语言的判定性质
- 乔姆斯基文法体系

代换

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 h 到语言, 对 $\forall L \subseteq \Sigma^*$,

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

定理 34

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

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. $V' = V \cup \left(\bigcup_{a \in T} V_a\right)$
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

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

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

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

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

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

因为 G' 中的每个 S_a 在 G 中是终结符 a，所以

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

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

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

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

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:

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

代换

封闭

CFL 对并/连接/闭包/同态封闭

定理 35

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

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

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

❸ 闭包和正比包运算封闭, 因为

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

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,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

❶ $L_1 \cup L_2$ 的文法为

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

❷ $L_1 L_2$ 的文法为

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

❸ 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 36

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

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

则 $L(G') = L^R$. 证明略.

CFL 对逆同态封闭

定理 37

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.

❶ $Q' \subset Q \times \Delta^*$: 状态 $[q, \overline{x}]$ 中的 $\overline{x}$ 为缓冲;

❷ 设 $q \in Q$, 那么 δ' 定义如下:

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

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

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

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

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

CFL 对交/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.

CFL 对补运算不封闭

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

$\cap$ $\bar{A}$

定理 38

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:

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

再往证 $\mathbf{L}(P') = L \cap R$, 略. $\square$

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:

❶ 因为 $\mathbf{a}^*\mathbf{b}^*\mathbf{c}^*$ 是正则语言,

❷ 而 $L \cap \mathbf{a}^*\mathbf{b}^*\mathbf{c}^* = \{a^n b^n c^n \mid n \geq 0\}$ 不是 CFL,

❸ 由 CFL 与正则语言的交还是 CFL, 所以 L 不可能是 CFL. □