作业 L2.14 - L3.2：CFL 的性质与 CFL 泵引理

2025 年 12 月 16 日

题目 1

证明下述语言是 CFL：

$$L_1 = \{a^n b^n \mid n \geq 0 \text{ 且} n \text{ 不是} 5 \text{ 的倍数}\}.$$

Solution.

直接构造 CFG 生成该语言。注意 n 不是 5 的倍数当且仅当

$$n = 5k + r, \qquad k \geq 0,\ r \in \{1, 2, 3, 4\}.$$

构造文法：

$$\boxed{\begin{aligned} S &\to S_1 \mid S_2 \mid S_3 \mid S_4, \\ S_1 &\to aUb, \\ S_2 &\to aaUbb, \\ S_3 &\to aaaUbbb, \\ S_4 &\to aaaaUbbbb, \\ U &\to aaaaaUbbbbb \mid \epsilon. \end{aligned}}$$

U 生成 $a^{5k}b^{5k}$；S_r 在两侧各补 r 个符号，生成 $a^{5k+r}b^{5k+r}$（$r = 1, 2, 3, 4$），正好对应 n 不是 5 的倍数的所有情形。因此 L_1 为 CFL。

题目 2

证明下述语言是 CFL：

$$L_2 = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w),\ w \text{ 不含子串} \mathsf{aab}\}.$$

Solution.

设

$$L_{bal} = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}, \qquad R = \{w \in \{a, b\}^* \mid w \text{ 不含 } \mathsf{aab}\}.$$

L_{bal} 是 CFL：可由经典文法 $S \to aSbS \mid bSaS \mid \epsilon$ 生成所有 a, b 数目相等的串。R 是正则语言（“不含固定子串”可由 DFA 识别）。于是

$$L_2 = L_{bal} \cap R$$

为 CFL（CFL 与正则语言交封闭）。

题目 3

证明存在算法，可判定一个 CFL 是否包含长度小于 n 的串。

Solution.

给定 CFL 的 CFG G 与整数 n。构造正则语言

$$R_{<n} = \{w \mid |w| < n\}.$$

$R_{<n}$ 可由一个 n 个状态的 DFA 识别（逐字符计数到 n 即转入拒绝陷阱）。

由于 CFL 与正则语言交封闭，$L(G) \cap R_{<n}$ 仍是 CFL。并且 CFL 的空性问题可判定：对 CFG 可通过“可生成符号/可达符号”分析判断是否能生成某个终结串。

因此算法为：

$$\text{判断} L(G) \cap R_{<n} \stackrel{?}{=} \emptyset.$$

若非空，则存在长度 $< n$ 的串；否则不存在。

题目 4

L_1 为 CFL，L_2 为正则语言，证明存在算法可判定 L_1 和 L_2 是否包含相同元素（即 $L_1 \cap L_2 \neq \emptyset$）。

Solution.

题意为判定 L_1 与 L_2 是否**有公共元素**（即 $L_1 \cap L_2 \neq \emptyset$）。

由于 L_1 是 CFL，L_2 是正则语言，且 CFL 与正则语言的交仍是 CFL，所以 $L_1 \cap L_2$ 是 CFL。而 CFL 的空性问题可判定（例如对 CFG 进行 generating/reachable 分析，或对 PDA 判定是否存在可接受路径）。

因此算法为：构造识别 $L_1 \cap L_2$ 的 CFG/PDA，并判定其语言是否为空。非空则两者包含相同元素，空则不包含。

CFL 非封闭性与 CFL 泵引理

题目 5

证明下述语言不是 CFL：

$$K_1 = \{a^n b^j \mid n = j^2\}.$$

Solution.

用 CFL 泵引理（Bar-Hillel 引理）。假设 K_1 是 CFL，设泵长度为 p。取

$$w = a^{p^2} b^p \in K_1.$$

将 $w = uvwxy$，满足 $|vwx| \le p$、$|vx| \ge 1$ 且对所有 $i \ge 0$ 有 $uv^iwx^iy \in K_1$。由于 $|vwx| \le p$，vwx 至多覆盖 b 段的一部分或跨越 a/b 分界，但无法同时覆盖到“足够多的 a 与 b”以维持二次关系。

分情况：

- 若 vx 全在 b^p 中，则泵 $i = 0$ 会减少 b 的个数但不改变 a 的个数，得到 $a^{p^2} b^{p-t}$。此时应满足 $p^2 = (p-t)^2$，只能推出 $t = 0$，与 $|vx| \ge 1$ 矛盾。
- 若 vx 全在 a^{p^2} 中，则泵 $i = 2$ 会增加 a 的个数但不改变 b 的个数，得到 $a^{p^2+t} b^p$。此时应满足 $p^2 + t = p^2$（因为 b 仍是 p 个，平方仍为 p^2），矛盾。
- 若 vx 跨越分界，则泵会同时改变 a 与 b 的个数，但 $|vx|$ 与 $|vwx|$ 都至多为 p，因此 b 的变化量至多为 p，使得 b 从 p 变为 $p + \Delta$（$|\Delta| \le p$），而平方变化为

  $$(p+\Delta)^2 - p^2 = 2p\Delta + \Delta^2,$$

  其绝对值至少为 $2p - 1$（取 $\Delta = \pm 1$ 时），远大于 a 段可被改动的幅度（至多 p），因此不可能仍保持 $n = j^2$。

三种情况均导致矛盾，故 K_1 不是 CFL。

题目 6

证明下述语言不是 CFL：

$$K_2 = \{a^n \mid n \text{ 为质数}\}.$$

Solution.

用 CFL 泵引理，设泵长度为 p，取一个质数 $q > p$，令 $w = a^q \in K_2$。任取分解 $w = uvwxy$ 满足泵条件。由于字母表单一，v 与 x 都是若干个 a，记 $|vx| = t \ge 1$。泵 $i = q + 1$，得到串长度

$$|uv^{q+1}wx^{q+1}y| = q + q \cdot t = q(1+t),$$

这是合数（$q \ge 2$ 且 $1 + t \ge 2$），因此不在 K_2，与泵引理矛盾。故 K_2 不是 CFL。

题目 7

证明下述语言不是 CFL:

$$K_3 = \{ww^R w \mid w \in \{a,b\}^*\}.$$

Solution.

设反证 K_3 是 CFL，泵长度为 p。取

$$w_0 = a^p b^p, \qquad w = w_0 w_0^R w_0 = a^p b^p b^p a^p a^p b^p \in K_3.$$

考虑任意分解 $w = uvwxy$ 且 $|vwx| \le p$。由于 $|vwx|$ 的长度限制，它只能落在 w 的某个局部片段中，无法同时跨越左右两侧的两处“镜像对应区域”。泵 $i=0$ 或 $i=2$ 将在某一局部改变字符个数/结构，破坏“前后两段相同且中间为反转”的整体结构，从而使得泵后串不再能写成 tt^Rt 的形式。因此与 CFL 泵引理矛盾，K_3 不是 CFL。

题目 8

用泵引理证明下述语言不是 CFL:

$$\{0^n1^n0^n1^n \mid n \ge 0\}, \qquad \{0^n\#0^{2n}\#0^{3n} \mid n \ge 0\}, \qquad \{w\#t \mid w,t \in \{a,b\}^*, \text{且}w\text{ 是}t\text{ 的子串}\}.$$

Solution.

下面给出每个语言的经典泵引理证明思路（均取 CFL 泵长度 p，构造一个“重复块足够长”的串，使 $|vwx| \le p$ 的局部泵动无法保持全局约束）。

(1) $L = \{0^n1^n0^n1^n\}$ 非 CFL

取 $w = 0^p1^p0^p1^p$。任意 $|vwx| \le p$ 的片段只能落在至多两个相邻块中（无法同时影响四段计数）。泵后将改变某一段或两段的长度，使得四段长度不再全相等，从而不在 L。

(2) $L = \{0^n\#0^{2n}\#0^{3n}\}$ 非 CFL

取 $w = 0^p\#0^{2p}\#0^{3p}$。$|vwx| \le p$ 只能改变三个 0-块中的一个或跨越一个分隔符，但泵后无法同时维持三个块的比例 $1:2:3$，矛盾。

(3) $L = \{w\#t \mid w$ 是t 的子串$\}$ 非 CFL

取 $w = a^pb^p\#a^pb^p$（左侧 w 与右侧 t 相同，显然满足“子串”）。泵 $i=0$ 会在左侧（或右侧）的某个局部删除一段字符，导致左侧 w 变成 $a^{p-r}b^p$ 或 a^pb^{p-r}，而右侧 t 仍为 a^pb^p（或结构被局部扰动），从而 w 不可能再是 t 的子串（因为 t 只有一个从 a 到 b 的分界位置）。因此与 CFL 泵引理矛盾。