形式语言与自动机 作业 6

上下文无关文法的性质

1951112　林日中

习题 7.2.1

用 CFL 泵引理来证明下面的语言都不是上下文无关的：

a)　$\{a^i b^j c^k | i < j < k\}$.

b)　$\{a^n b^n c^i | i \le n\}$.

c)　$\left\{0^p \middle| p \text{ 是素数}\right\}$. 提示：使用和例 4.3 中证明不是正则语言时采用相同的思想.

d)　$\{0^i 1^j | j = i^2\}$.

e)　$\{a^n b^n c^i | n \le i \le 2n\}$.

f)　$\left\{ww^R w \middle| w \text{ 是 0 和 1 的串}\right\}$. 也就是说，由某个串 w 和它的反向串再和它本身连接起来的串（比如 001100001）构成的集合.

证明. 对于每个问题，假设语言 L 是上下文无关的. 那么由上下文无关语言的泵引理，存在一常数 $n > 0$，对于语言 L 中每个满足 $|z| \ge n$ 的字符串 z，存在一组 u, v, w, x, y 使得 $z = uvwxy$ 且有 $|vwx| \le n, vx \ne \varepsilon$ 且 $\forall k \ge 0, uv^k wx^k y \in L$.

a)　考察串 $z = a^n b^{n+1} c^{n+2}$. 对于串 vwx，可能有 $vwx \in a^+, a^+b^+, b^+, b^+c^+$ or c^+.

当 $c \notin vwx$ 时（即前 3 种可能性），对于串 uv^3wx^3y，它有不少于 $n+2$ 个 a 或 b，而只有 $n+2$ 个 c. 此时，显然有 $uv^3wx^3y \notin L$.

当 $c \in vwx$ 时（即后 2 种可能性），因为 $|vwx| \le n$，故有 $a \notin vwx$. 对于串 $uv^0wx^0y = uwy$，它有不多于 $2n+1$ 个 b 和 c 以及 n 个 a，故它不可能同时有多于 n 个 b 和多于 n 个 c. 那么也就有 $uv^0wx^0y \notin L$.

因此，该语言不是上下文无关的. □

b)　考察串 $z = a^n b^n c^n$.

当 $c \notin v, x$ 时，有 $vwx \in a^+, b^+$ or a^+b^+. 对于串 uv^0wx^0y，它一定有少于 n 个 a 和/或少于 n 个 b，因而有 $uv^0wx^0y \notin L$.

当 $a, b \notin vwx$ 时，有 $vwx \in c^+$. 对于串 uv^3wx^3y，它一定有多于 n 个 c，因而有 $uv^3wx^3y \notin L$. 当 $a, c \notin vwx$ 或 $b, c \notin vwx$ 时，分别会有相同的矛盾出现.

因此，该语言不是上下文无关的. $\square$

c) 考察串 $z=0^p, p\ge n$，其中 p 是素数. 令 $|vwx|\le n, |vx|=m\ge 1$，此时有 $uv^kwx^ky=0^{p+m(k-1)}, k\ge 0$. 当 $k=p+1$ 时，$uv^kwx^ky=0^{p+pm}=0^{p(1+m)}$. 显然，$p(1+m)$ 不是素数，故有 $uv^kwx^ky\notin L$.

因此，该语言不是上下文无关的. $\square$

d) 考察串 $z=0^n1^{n^2}$.

显然，当 v 或 x 横跨 0^n 和 1^{n^2} 的边界时，由泵引理产生的串 $uv^kwx^ky\notin L$.

当 $vwx\in 0^+$ 时，对于串 $uv^2wx^2y=0^{n+|vx|}1^{n^2}, 0<|vx|\le p$，显然有 $(n+|vx|)^2\ne n^2$，故有 $uv^2wx^2y\notin L$.

当 $vwx\in 1^+$ 时，对于串 $uv^2wx^2y=0^n1^{n^2+|vx|}, 0<|vx|\le p$，显然有 $n^2\ne n^2+|vx|$，故有 $uv^2wx^2y\notin L$.

当 $v\in 0^+, x\in 1^+$ 时，对于串 $uv^2wx^2y=0^{n+|v|}1^{n^2+|x|}, |vx|>0$，因为 $(p+1)^2=p^2+2p+1, |x|<n$，所以有 $n^2+|x|<(n+1)^2$，进而一定有 $(n+|v|)^2\ne n^2+|x|$. 故有 $uv^2wx^2y\notin L$.

因此，该语言不是上下文无关的. $\square$

e) 考察串 $a^{\hat{n}}b^{\hat{n}}c^{2\hat{n}}$，其中 $\hat{n}$ 是泵长度.

当 $c\notin v,x$ 时，有 $vwx\in a^+, b^+$ or a^+b^+. 对于串 $uv^mwx^my, m=\hat{n}$，它仍属于该语言. 但当 $m\ge\hat{n}+1$ 时，有 $n\ge 2\hat{n}+1, i\le n-1$，故产生的串不再属于该语言.

当 $a\notin v,x$ 时，有 $vwx\in b^+, c^+$ or b^+c^+. $vwx\in b^+$ or b^+c^+ 时，对于串 uv^0wx^0y，它一定含有少于 $\hat{n}$ 个 b，故产生的串不再属于该语言；$vwx\in c^+$ 时，对于串 uv^2wx^2y，它一定含有多于 $2\hat{n}$ 个 c，故产生的串不再属于该语言.

因此，该语言不是上下文无关的. $\square$

f) 考察串 $0^n110^n0^n1$. 与 b) 中讨论的情况类似，3 段 0^n 不可能同时存在于 vwx 中，将他们分别记为 $0^n_{(1)}, 0^n_{(2)}, 0^n_{(3)}$. 对于串 vwx，可能有 $vwx\in 0^+_{(1)}, 0^+_{(2)}, 0^+_{(3)}, 0^+_{(1)}0^+_{(2)}$ or $0^+_{(2)}0^+_{(3)}$. 在任何一种可能性中，对于串 $uv^iwx^iy, \forall i\in\mathbb{N}, i\ne 1$，都会出现 $0_{(1)}$、$0_{(2)}$ 和 $0_{(3)}$ 的数量不相等的情况，即产生的串不再属于该语言.

因此，该语言不是上下文无关的. $\square$

习题 7.2.5

使用奥格登引理（习题 7.2.3）来证明下列语言不是 CFL：

a) $\{0^i 1^j 0^k | j = \max(i,k)\}$.

b) $\{a^n b^n c^i | \ i \neq n\}$. 提示：如果 n 是奥格登引理的常数，考虑串 $z = a^n b^n c^{n+n!}$.

证明.

a) 考察串 $z = 0^{2n} 1^{2n} 0^n$，并标记所有最后的 0 为 $\hat{0}$（即 $z = 0^{2n} 1^{2n} \hat{0}^n$）.

显然，当 v 或 x 横跨 0^{2n}、1^{2n} 和 $\hat{0}^n$ 的边界时，由奥格登引理产生的串已不再属于语言 L.

当 $v \in 0^+$ or 1^+ 时，对于由奥格登引理产生的串，题设条件 $j = \max(i,k)$ 将不再成立，因为其中一个数会增长，同时另一个会保持不变.

当 v 和 x 完全包含 $\hat{0}$ 时，在对 z 应用奥格登引理时，必然在某时有 $k > i, j$，题设条件 $j = \max(i,k)$ 将不再成立.

因此，该语言不是上下文无关的. □

b) 考察串 $z = a^n b^n c^{n+n!}$，其中 n 是奥格登引理中的常数.

我们标记所有的 a 和 b. 注意到，当 v 或 x 包含 a 和 b 的混合时，在 $i = 2$ 的情况下，由奥格登引理生成的串 $uv^i wx^i y$ 不再属于该语言.

当 $v = a^\alpha$ 和 $x = b^\beta$ 时，假设 $\alpha \neq \beta$，那么 a 和 b 的数量将不相等，因此，一定有 $\alpha = \beta$. 令 $\gamma = \alpha = \beta$，那么最终的字符串将是这样的形式

$$a^{n+\gamma(i-1)} b^{n+\gamma(i-1)} c^{n!+n}$$

因此，我们假设 a 或 b 的指数等于 c 的指数，有

$$n + \gamma(i-1) = n! + n$$

$$\gamma(i-1) = n!$$

$$i - 1 = \frac{n!}{\gamma}$$

由于 $\gamma \leq n$，上述等式右侧能够整除. 故我们能够选择一个满足这个约束的 i，从而得出由奥格登引理生成的串 $uv^i wx^i y$ 与我们先前的约束 $\{a^n b^n c^i | i \neq n\}$ 矛盾.

因此，该语言不是上下文无关的. □

习题补充 1

构造与下列文法等价的 CNF.

$$
\begin{aligned}
&S \to ABB \mid bAA \\
&B \to aBa \mid aa \mid \varepsilon \\
&A \to bbA \mid \varepsilon
\end{aligned}
$$

解答. 消除 ε-产生式，有

$$
\begin{aligned}
&S \to \varepsilon \mid A \mid B \mid AB \mid BB \mid ABB \mid b \mid bA \mid bAA \\
&B \to aBa \mid aa \\
&A \to bba \mid bb
\end{aligned}
$$

消除单一产生式，有

$$
\begin{aligned}
&S \to \varepsilon \mid bbA \mid bb \mid aBa \mid aa \mid AB \mid BB \mid ABB \mid b \mid bA \mid bAA \\
&B \to aBa \mid aa \\
&A \to bba \mid bb
\end{aligned}
$$

引入新变量 C_1, C_2, C_3, C_4, C_5，有

$$
\begin{aligned}
&S \to \varepsilon \mid C_3A \mid C_5C_5 \mid C_4BC_4 \mid C_4C_4 \mid AB \mid BB \mid AC_1 \mid b \mid C_5A \mid C_5C_2 \\
&B \to C_4BC_4 \mid C_4C_4 \\
&A \to C_3A \mid C_5C_5 \\
&C_1 \to BB \\
&C_2 \to AA \\
&C_3 \to C_5C_5 \\
&C_4 \to a \\
&C_5 \to b
\end{aligned}
$$

引入新变量 C_6（$C_6 \to BC_4$），有

$$
\begin{aligned}
&S \to \varepsilon \mid b \mid C_3A \mid C_5C_5 \mid C_4C_6 \mid C_4C_4 \mid AB \mid BB \mid AC_1 \mid C_5A \mid C_5C_2 \\
&B \to C_4C_6 \mid C_4C_4 \\
&A \to C_3A \mid C_5C_5 \\
&C_1 \to BB \\
&C_2 \to AA \\
&C_3 \to C_5C_5 \\
&C_4 \to a
\end{aligned}
$$

$C_5 \to b$

$C_6 \to BC_4$