第 5 次作业 – 上下文无关文法与下推自动机

2154312 郑博远

习题 5.1.5 设 $T = \{0, 1, (,), +, *, \phi, e\}$，可以把 T 看作字母表为 $\{0,1\}$ 的正则表达式所使用的符号的集合，惟一的不同是用 e 来表示符号 ε，目的是为了避免有可能出现的混淆。你的任务是以 T 为终结符号集合来设计一个 CFG，该 CFG 生成的语言恰好是字母表为 $\{0,1\}$ 的正则表达式。

解答：

设 CFG $G = (V, T, P, S)$ 生成的语言能够识别题目描述的正则表达式。其中，$V = \{S\}, T = \{0, 1, (,), +, *, \phi, e\}$，产生式 P 如下：

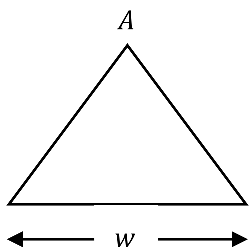
$$S \to S + S \mid SS \mid S^* \mid (S) \mid 0 \mid 1 \mid \phi \mid e$$

习题 5.2.2 假设 G 是一个 CFG，并且它的任何一个产生式的右边都不是 ε。如果 w 在 $L(G)$ 中，w 的长度是 n，w 有一个 m 步完成的推导，证明 w 有一个包含 $n+m$ 个节点的分析树。

证明：

对分析树进行归纳：

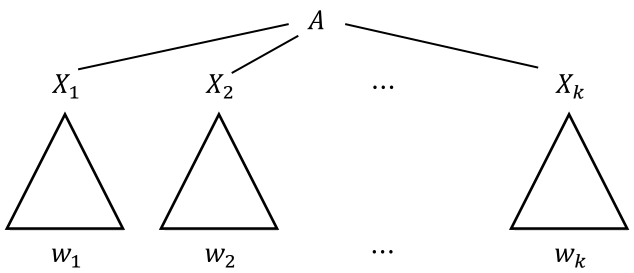
基础： 若 $m = 1$，即该推导只有一步。设 w 在变元 A 的语言中，一定存在产生式 $A \to w$。如下图所示，存在语法分析树满足文法 G 的条件。显然，它的根是 A，产物是 w，包含 $n + 1 = n + m$ 个节点（n 个叶子节点，1 个根节点）。



归纳： 假定在 $m+1$ 个推理步骤之后能够得出 w 在 A 的语言里这个事实，并且这个定理对于使得 B 的语言中的 x 成员用小于等于 m 步推理推得的所有串 x 和变元 B 成立。考虑得出 w 在 A 的语言里这个推理的最后一步，这一步使用了 A 的某个产生式，不妨设为$A \to X_1X_2 \cdots X_k$，其中 X_i 或者是一个终结符或者是一个变元。对应的将 w 划分为 $w_1w_2 \cdots w_k$。即每个 X_i 能通过 m_i 步（$m_i \le m$）推导出子字符串 w_i（串长为 n_i）。

1) 若 X_i 是终结符，则$X_i = w_i$。对应 $m_i = 0$。
2) 若 X_i 是变元，则该子树能通过 m_i 步（$m_i \le m$）推导出子字符串 w_i。由归纳法得到，其节点数为 $|w_i| + m_i = n_i + m_i$。

对于整棵语法分析树，其总推导步数 $m = 1 + \sum_{i=1}^{k} m_i$ 。其总节点数为各个子树的节点数之和加一（对应根节点），即 $\sum_{i=1}^{k}(n_i + m_i) + 1 = \sum_{i=1}^{k} n_i + \sum_{i=1}^{k} m_i + 1 = n + m$。



习题 5.2.3 假设在习题 5.2.2 中除了 G 中可能有右端为 ε 的产生式外其他所有的条件都满足，证明此时 w（ w 不是 ε ）的语法分析树有可能包含 $n+2m-1$ 个节点，但不可能更多。

证明：

对分析树进行归纳：

基础： 若 $m = 1$，即该推导只有一步。设 w 在变元 A 的语言中，一定存在产生式 $A \to w$。如下图所示，存在语法分析树满足文法 G 的条件。当 $w \ne \varepsilon$ ，语法分析树的根是 A，产物是 w，包含 $n+1 \le n+2m-1$ 个节点；若 $w = \varepsilon$，同样有节点个数$1 \le n+2m-1$。

归纳： 假定在 $m+1$ 个推理步骤之后能够得出 w 在 A 的语言里这个事实，并且这个定理对于使得 B 的语言中的 x 成员用小于等于 m 步推理推得的所有串 x 和变元 B 成立。考虑

得出 w 在 A 的语言里这个推理的最后一步，这一步使用了 A 的某个产生式，不妨设为$A \rightarrow X_1X_2 \cdots X_k$，其中 X_i 或者是一个终结符或者是一个变元。对应的将 w 划分为 $w_1w_2 \cdots w_k$。即每个 X_i 能通过 m_i 步（$m_i \leq m$）推导出子字符串 w_i（串长为 n_i）。

1) 若 X_i 是终结符，则 $X_i = w_i$。对应 $m_i = 0$，$c_i = n_i + m_i = n_i + 2m_i$。

2) 若 X_i 是变元，则该子树能通过 m_i 步（$m_i \leq m$）推导出子字符串 w_i。由归纳法得到，其节点个数（记为 c_i）满足 $c_i \leq\ n_i + 2m_i - 1$。

对于整棵语法分析树，其总推导步数 $m = 1 + \sum_{i=1}^{k} m_i$ 。假设 $A \rightarrow X_1X_2 \cdots X_k$ 中有 j 个 X_i （不妨设前 j 个）满足上述情况 2)，则语法分析树的总节点数 c 满足：

$$c = 1 + \sum_{i=1}^{k} c_i$$

$$\leq 1 + (n_1 + 2m_1 - 1) + \cdots + \left(n_j + 2m_j - 1\right) + \left(n_{j+1} + 2m_{j+1}\right) + \cdots + \ (n_k + 2m_k)$$

$$= 1 + \sum_{i=1}^{k} n_i + 2\sum_{i=1}^{k} m_i - j$$

$$= n + 2(m-1) + 1 - j$$

$$=\ n + 2m - 1 - j$$

$$\leq n + 2m - 1$$

综上所述，w 的语法分析树有可能包含 $n + 2m - 1$ 个节点（当前仅当 $j = 0$ 时取到此最大值），但不可能更多。

习题5.4.7 下面的文法生成的是具有 x 和 y 操作数、二元运算符 +、− 和 ∗ 的前缀表达式：

$$E\ \rightarrow\ +EE \mid *EE \mid -EE \mid x \mid y$$

a) 找到串 `+*-xyxy` 的最左推导、最右推导和一棵语法分析树。

b) 证明这个文法是无歧义的。

解答：

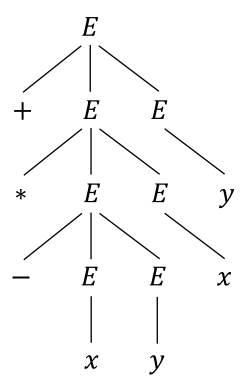
a) 最左推导如下：

$$
\begin{aligned}
E &\underset{lm}{\Rightarrow} +EE \\
&\underset{lm}{\Rightarrow} +*EEE \\
&\underset{lm}{\Rightarrow} +*-EEEE \\
&\underset{lm}{\Rightarrow} +*-xEEE \\
&\underset{lm}{\Rightarrow} +*-xyEE \\
&\underset{lm}{\Rightarrow} +*-xyxE \\
&\underset{lm}{\Rightarrow} +*-xyxy
\end{aligned}
$$

最右推导如下：

$$
\begin{aligned}
E &\underset{rm}{\Rightarrow} +EE \\
&\underset{rm}{\Rightarrow} +Ey \\
&\underset{rm}{\Rightarrow} +*EEy \\
&\underset{rm}{\Rightarrow} +*Exy \\
&\underset{rm}{\Rightarrow} +*-EExy \\
&\underset{rm}{\Rightarrow} +*-Eyxy \\
&\underset{rm}{\Rightarrow} +*-xyxy
\end{aligned}
$$

语法分析树如下图：



b) 证明如下：

首先，容易通过归纳法得到 $\forall w,\ E \overset{*}{\Rightarrow} w$ 都满足：w 中+、*、−的数量比 x、y 的数量

少一（记为命题 P_1），且其任何后缀中+、*、−的数量严格少于 x、y 的数量（记为命题 P_2）。下面进行归纳：

基础： $|w| = 1$时，有 $E \Rightarrow x$ 或 $E \Rightarrow y$。$w = x$ 或 $w = y$，符合上述命题。

归纳： $|w| > 1$时，递归推理最后一步一定使用了产生式 $E \rightarrow +EE \mid *EE \mid -EE$ 之一（不妨设使用 $E \rightarrow +EE$）。可将 w 表示为 $w = +w_1w_2$，满足 $E \overset{*}{\Rightarrow} w_1$ 和 $E \overset{*}{\Rightarrow} w_2$。由归纳法可得，$w_1$、$w_2$ 均满足上述命题。因而，w_1、w_2 中+、*、−的数量比 x、y 的数量少二；w 中+、*、−的数量比 x、y 的数量少一。同理易得 w 后缀中+、*、−数量严格少于 x、y 的数量。

接下来证明对于产生式 $E \rightarrow +EE \mid *EE \mid -EE$ 之一（以 $E \rightarrow +EE$为例），将 w 表示为 $w = +w_1w_2$的划分是唯一的。若 w 有两种不同的划分方法，假设方法 1 中$[1, m]$为 w_1，$[m+1, n]$为 w_2；方法 2 中$[1, k]$为 w_1，$[k+1, n]$为 w_2，且 $k > m$。为了满足 P_1，区间$[m+1, k]$中+、*、−的数量必须与 x、y 的数量相等，这与 P_2 矛盾。因此，将 w 表示为 $w = +w_1w_2$ 的划分是唯一的。

下面通过归纳法证明该文法的无歧义性：

基础： $|w| = 1$时，有 $E \Rightarrow x$ 或 $E \Rightarrow y$，显然有唯一的语法分析树。

归纳： $|w| > 1$时，则推导第一步一定使用了产生式 $E \rightarrow +EE \mid *EE \mid -EE$ 之一（不妨设使用 $E \rightarrow +EE$）。可将 w 可表示为 $w = +w_1w_2$，由于 w_1、w_2 的划分是唯一的。由于 $|w_1| < |w|, |w_2| < |w|$，由归纳法可得 w_1、w_2 对应的语法分析树是唯一的，因此 w 对应的语法分析树也是唯一的。

综上所述，该文法具有无歧义性。

补充习题 1 对于下列语言，分别构造接受它们的 PDA：

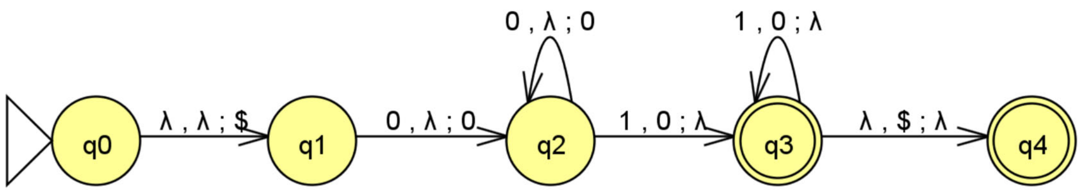
a) $\{0^n1^m \mid n \geq m \geq 1\}$

b) $\{1^n0^n1^m0^m \mid n, m \geq 1\}$

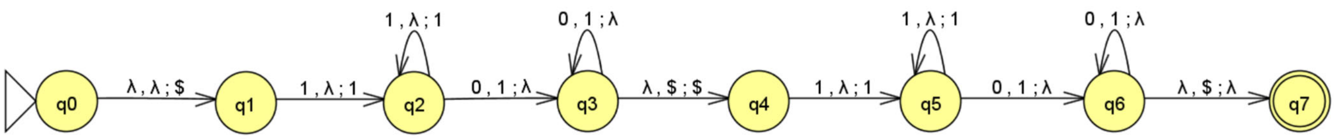
c) 含有 0 的个数和 1 的个数相同的所有 0, 1 串

解答：

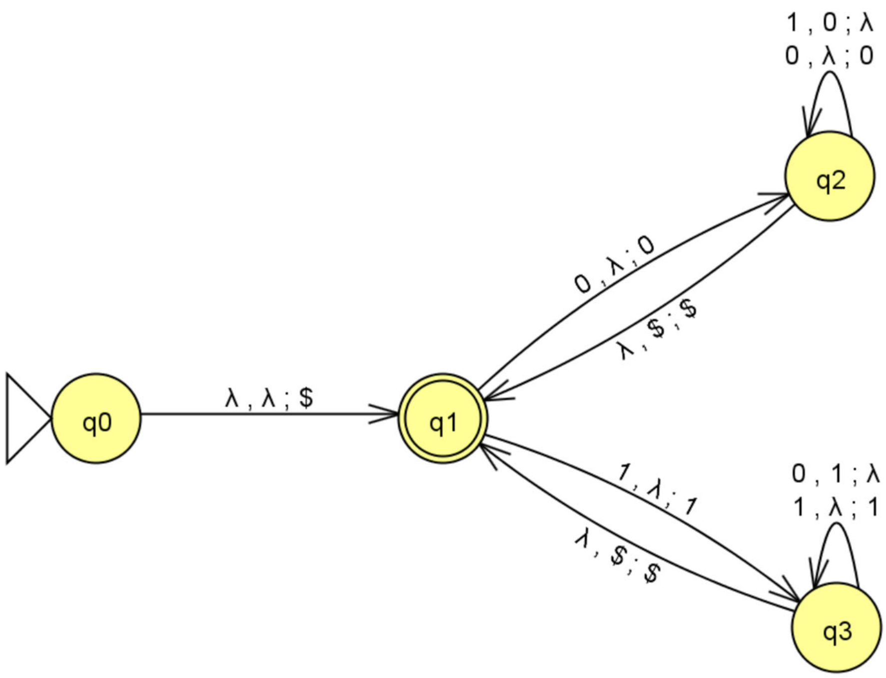
a) 构造 PDA 如下图所示：



b) 构造 PDA 如下图所示：



c) 构造 PDA 如下图所示：



补充习题 2 构造一个 PDA，使它等价于下列文法：

$$S \to aAA, \qquad A \to aS|bS|a$$

解答： 构造 PDA 如下图所示：