形式语言与自动机 作业 5

CFG & PDA

1951112　林日中

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. 记 CFG $G = (V,T,P,S) = (\{S\},T,P,S)$，其中，$V$ 是变元集合，T 是终结符号集合，S 是初始符号；P 是产生式

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

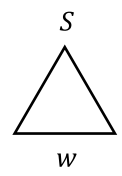
的集合. CFG G 生成的语言恰好是字母表为 $\{0,1\}$ 的正则表达式.

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. 归纳于 $S \overset{*}{\Rightarrow} w$，其中，$S$ 是 G 的开始符号.

基础：$m=1$. 此时一定有产生式 $S \rightarrow w$，因此存在一个包含 $n+1$ 个节点的分析树（$w \neq \varepsilon$），有 $n+m=n+1$，结果成立.



归纳：$m>1$. 设第一步使用了产生式 $S \rightarrow X_1X_2\cdots X_k$.

该推导如 $S \Rightarrow X_1X_2\cdots X_k \overset{*}{\Rightarrow} w$. 可以将 w 分成 $w = w_1w_2\cdots w_k$，其中

a)　若 X_i 为终结符，则 $w_i = X_i$.

b)　若 X_i 为非终结符，则 $X_i \overset{*}{\Rightarrow} w_i$，且其步数 $m_i < m$. 由归纳假设，存在根节点为

X_i 的子分析树，其节点数为 $|w_i| + m_i$.

对于上述情形 a)，没有进一步的关于 X_i 的推导，我们可以认为 $m_i = 0$. 那么，我们得到如下关系：

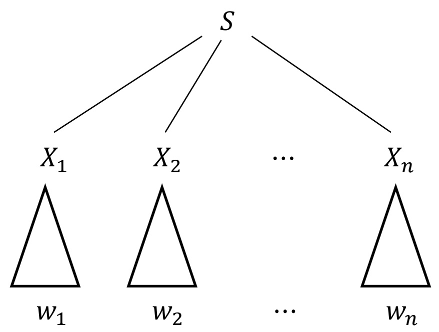
$$m = m_1 + m_2 + \cdots + m_k + 1$$

因此，存在一棵关于 w 的分析树，其节点数为

$$1 + (|w_1| + m_1) + (|w_2| + m_2) + \cdots + (|w_k| + m_k)$$

$$= (|w_1| + |w_2| + \cdots + |w_k|) + (m_1 + m_2 + \cdots + m_k + 1) = n + m$$

即 w 有一个包含 $n + m$ 个节点的分析树. □



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. 归纳于 $S \overset{*}{\Rightarrow} w$，其中，$S$ 是 G 的开始符号.

基础：$m = 1$. 此时一定有产生式 $S \to w$，因此存在一个包含 $n + 1$ 个节点的分析树（$w \neq \varepsilon$），有 $n + 2m - 1 = n + 1$，结果成立.

归纳：$m > 1$. 设第一步使用了产生式 $S \to X_1 X_2 \cdots X_k$.

该推导如 $S \Rightarrow X_1 X_2 \cdots X_k \overset{*}{\Rightarrow} w$. 可以将 w 分成 $w = w_1 w_2 \cdots w_k$，其中

a) 若 X_i 为终结符，则 $w_i = X_i$. 记终结符个数为 $j, 0 \leq j \leq k$，不妨设前 j 个 X_i 为终结符.

b) 若 X_i 为非终结符，则 $X_i \overset{*}{\Rightarrow} w_i$，且其步数 $m_i < m$. 由归纳假设，存在根节点为 X_i 的子分析树，其节点数为 $|w_i| + 2m_i - 1$.

对于上述情形 a)，没有进一步的关于 X_i 的推导，我们可以认为 $m_i = 0$. 那么，我们得到如下关系：

$$m = m_1 + m_2 + \cdots + m_k + 1$$

因此，存在一棵关于 w 的分析树，其节点数不超过

$$\begin{aligned}1 + (|w_1| + m_1) + (|w_2| + m_2) + \cdots + \left(|w_j| + m_j\right) + \left(|w_{j+1}| + 2m_{j+1} - 1\right) \\ + \left(|w_{j+2}| + 2m_{j+2} - 1\right) + \cdots + (|w_k| + 2m_k - 1) \\ = (|w_1| + |w_2| + \cdots + |w_k|) + (m_1 + m_2 + \cdots + m_k + 1) \\ + \left(m_{j+1} + m_{j+2} + \cdots + m_k\right) - j = n + 2m - 1 - 2j\end{aligned}$$

即 w 有一个最多包含 $n + 2m - 1 - 2j$ 个节点的分析树.

因为 $0 \leq j \leq k$，所以当 $j = 0$ 时，w 的分析树的节点数最大，为 $n + 2m - 1$.

因此，此时 w 的语法分析树有可能包含 $n + 2m - 1$ 个节点，但不能更多. □

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：

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

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.

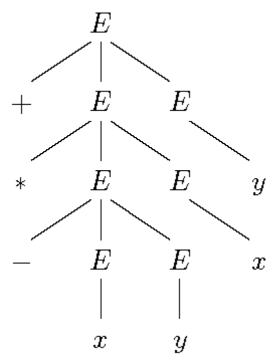
a) 串 `+*-xyxy` 的最左推导如下：

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

串 `+*-xyxy` 的最右推导如下：

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

串 `+*-xyxy` 的语法分析树如下图所示.



b)　先证明对任何终结符串 w，如下命题成立：

命题 P：$E\overset{*}{\Rightarrow}w \Leftrightarrow w$ 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1；并且 w 的任何真前缀中 $+$，$*$ 和 $-$ 的总数不少于 x 和 y 的总数.

（**充分性**）归纳于 $E\overset{*}{\Rightarrow}w$ 的步数 k.

基础：$k=1$. 即 $E\Rightarrow w$. 则必有 $w=x$ 或 $w=y$，充分性成立.

归纳：$k>1$. 则推导的第一步一定使用了三个产生式 $E\to +EE\mid *EE\mid -EE$ 之一，不妨设使用了产生式 $E\to +EE$. 此时，w 可表示为 $w=+w_1w_2$，并且有 $E\overset{*}{\Rightarrow}w_1$ 和 $E\overset{*}{\Rightarrow}w_2$. 而 $E\overset{*}{\Rightarrow}w_1$ 和 $E\overset{*}{\Rightarrow}w_2$ 的推导步数均小于 k，所以 w_1 和 w_2 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1，且其中的任何真前缀中 $+$，$*$ 和 $-$ 的总数不少于 x 和 y 的总数.

这样，我们可以推出：$w=+w_1w_2$ 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1，且其中的任何真前缀中 $+$，$*$ 和 $-$ 的总数不少于 x 和 y 的总数.

（**必要性**）归纳于 w 的长度 k.

基础：$k=1$. 因 w 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1，必有 $w=x$ 或 $w=$

y，所以 $E \overset{*}{\Rightarrow} w$ 成立.

归纳：$k > 1$. 因 w 的任何真前缀中 $+$，$*$ 和 $-$ 的总数不少于 x 和 y 的总数，所以 w 的第一个符号是 $+$，$*$ 或 $-$ 之一，不妨设为 $+$. 我们将 w 表示为 $w = +w_1w_2$，这里 $+w_1$ 是首次满足 w_1 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1 的 w 的真前缀（由于 w 中 x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1，这样的非空子串 w_1 和 w_2 总是可以找到的）. 我们能够推断：w_1 和 w_2 都满足：x 和 y 的总数比 $+$，$*$ 和 $-$ 的总数多 1，且其中的任何真前缀中 $+$，$*$ 和 $-$ 的总数不少于 x 和 y 的总数. 依归纳假设，我们得到 $E \overset{*}{\Rightarrow} w_1$ 和 $E \overset{*}{\Rightarrow} w_2$. 因此，$E \overset{*}{\Rightarrow} w$ 成立.

下面证明文法的无二义性，即证明对所有终结符串，其分析树或最左推导是唯一的. 我们归纳于终结符串的长度，以证明该文法所产生的任何终结符串的最左推导是唯一的.

设 w 表示该文法可推导出的任何字符串，现归纳于 w 的长度来证明其最左推导是唯一的.

基础：$|w| = 1$ 时，必有 $w = x$ 或 $w = y$；其最左推导是 $E \Rightarrow x$ 或 $E \Rightarrow y$，是唯一的.

归纳：设 $|w| < k, k > 1$ 时，w 有唯一的最左推导. 当 $|w| = k$ 时，产生 w 的第一步推导一定使用了三个产生式 $E \rightarrow +EE \mid *EE \mid -EE$ 之一（因为 $k > 1$）；不妨设 w 的第一个符号为 $+$，则第一步推导是唯一的，只能是 $E \Rightarrow +EE$；根据上下文无关文法的特性，存在 w_1、w_2，满足 $w = w_1w_2$（根据上述命题 P，这样的 w_1 和 w_2 的是唯一确定的），并且有 $E \overset{*}{\Rightarrow} w_1$ 和 $E \overset{*}{\Rightarrow} w_2$；因为 $|w_1| < k$ 以及 $|w_2| < k$，根据归纳假设，w_1 和 w_2 的最左推导是唯一的；综合唯一的第一步推导，就可得到一个 w 的最左推导，且是唯一的最左推导.

因此，该文法是无二义的，也即无歧义的. □

习题 1

对于下列语言，分别构造接受它们的 PDA：

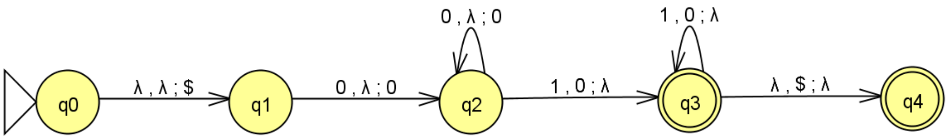
a) $\{0^n1^m \mid n \geq m \geq 1\}$

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

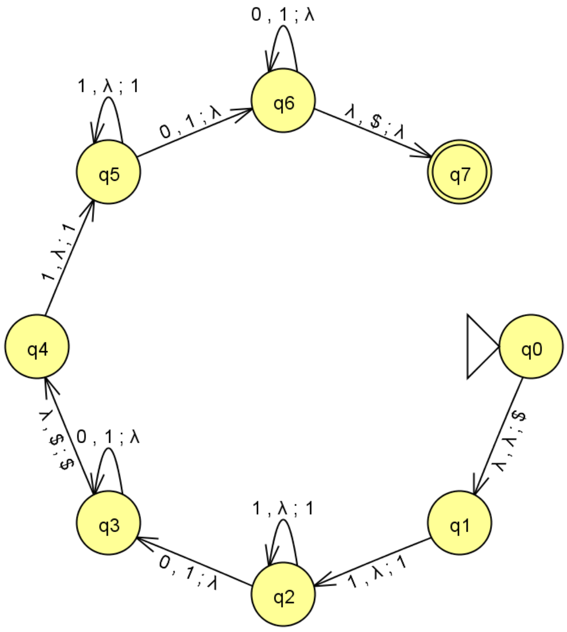
c) 含有 0 的个数和 1 的个数相同的所有 0, 1 串

解答.

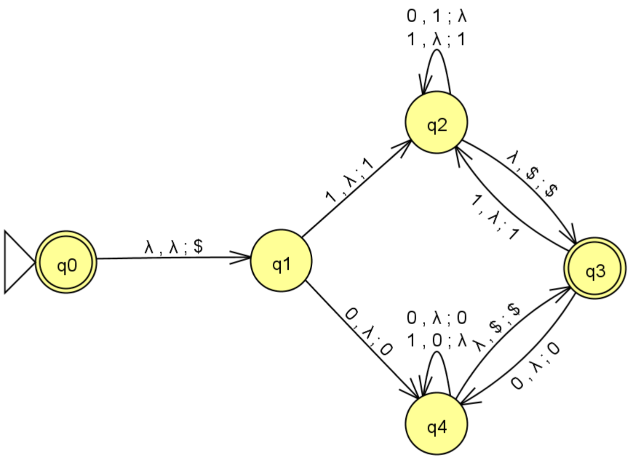
a) 接受语言 $\{0^n 1^m | n \geq m \geq 1\}$ 的 PDA 如下图所示.



b) 接受语言 $\{1^n 0^n 1^m 0^m | n, m \geq 1\}$ 的 PDA 如下图所示.



c) 接受语言 $\left\{\text{含有 0 的个数和 1 的个数相同的所有 0,1 串}\right\}$ 的 PDA 如下图所示.

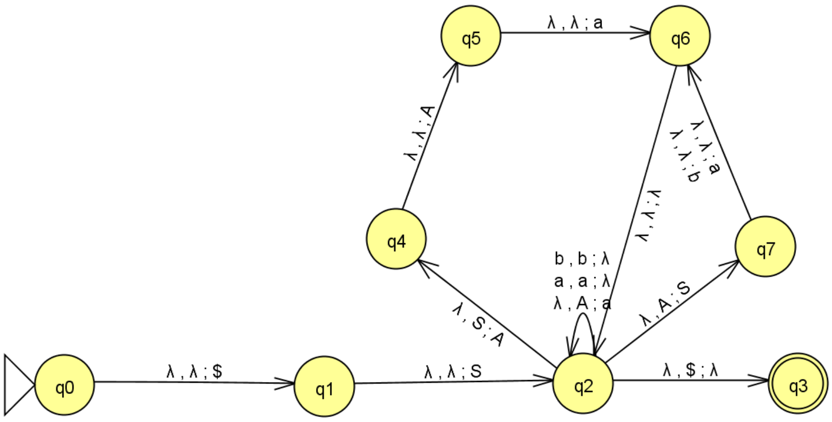


习题 2

构造一个 PDA，使它等价于下列文法：

$$S \to aAA \quad A \to aS \mid bS \mid a$$

解答. 等价于上述文法的 PDA 如下图所示.


q0
q1
q2
q3
q4
q5
q6
q7
λ , λ ; $
λ , λ ; S
λ , $; λ
λ , S ; A
λ , λ ; A
λ , λ ; a
λ , A ; S
λ , λ ; a
λ , λ ; b
λ , λ ; λ
b , b ; λ
a , a ; λ
λ , A ; a