

练习 4.1.1: 考虑上下文无关文法 $S \rightarrow SS+ \mid SS^* \mid a$ 以及串 $aa+a^*$

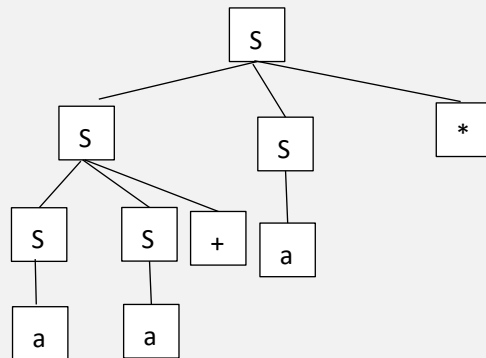
- 1) 给出这个串的一个最左推导
- 2) 给出这个串的一个最右推导
- 3) 给出这个串的一颗语法分析树
- 4) 这个文法是否是二义性的? 证明你的回答 (选作)
- 5) 这个文法生成的语言是什么?

答案

1) $S \Rightarrow SS^* \Rightarrow SS+S^* \Rightarrow aS+S^* \Rightarrow aa+S^* \Rightarrow aa+a^*$

2) $S \Rightarrow SS^* \Rightarrow Sa^* \Rightarrow SS+a^* \Rightarrow Sa+a^* \Rightarrow aa+a^*$

3)



4) 没有二义性。

附 4.1.1 (4) 的证明

由于上下文无关文法的二义性问题是不可判定的，所以这里只能具体问题具体分析。首先证明一个引理：题述上下文无关文法的任何一个句型都满足两个条件：

(1) 设变元 S 和终结符 a 的总数为 N ，运算符 $+$ 和 $*$ 的总数为 T ，则有 $N = T + 1$ ；

(2) 对句型的任一非空前缀，有 $N > T$ 。

证明：(1) 对于起始的 S ，有 $N = 1 = 0 + 1 = T + 1$ ；每一步推导都使得 N 和 T 都加一（采用前两个产生式）或都不变（采用最后一个产生式）；所以经过 $n \in \mathbb{N}^+$ 步推导后也满足 $N = T + 1$ 。(2) 用归纳法，起始的 S 显然满足 (2)；假设经过 $n \in \mathbb{N}^+$ 步推导后的句型满足 (2)，那么经过第 $n + 1$ 步推导后，如果采用最后一个产生式则不影响结论成立，如果采用前两个产生式，由于新产生的运算符在新产生的变元 S 之后，所以新句型仍然满足 (2)。

设 S 经过 $n \in \mathbb{N}^+$ 步推导后产生字符串 x ，其长度为 $l \in \mathbb{N}^+$ ，由于对于 $l = 1$ 的情况显然只有唯一的语法分析树，所以我们接下来只证明对于 $l \geq 2$ 的情况 x 也只有唯一的语法分析树。

既然 $l > 2$ ，所以显然 l 必然以 $+$ 或 $*$ 结尾。对 $i \in \{1, 2, \dots, n\}$ 定义两个函数 $n(i)$ 和 $t(i)$ ，分别表示 x 长为 i 的前缀中终结符 a 和运算符 $+, *$ 的个数。

既然 $l > 2$ ，所以由 S 派生出 x 的第一步推导必然是 $S \rightarrow S_1 S_2 *$ 或 $S \rightarrow S_1 S_2 +$ （下标用于在下文中指示前后两个变元）。那么一定存在 x 的一个长度小于等于 $l - 2$ 的非空前缀由 S_1 产生，而剩余部分除去末尾的运算符由 S_2 产生，我们设这个分界位置为 $d \leq n - 2$ （ x 的前 d 个字符由 S_1 产生，第 $d + 1$ 到 $l - 1$ 个字符由 S_2 产生）。

首先根据引理， d 一定是 $n(d) = t(d) + 1$ 的解，并且这个方程至少还有一个解是 $i = n$ 。干脆设它所有的解从小到大排列为 $\{i_1, \dots, i_m\}$ ， $m \geq 2$ ，则我们断言 $d = i_{m-1}$ 。若不然，假设 $d < i_{m-1}$ ，考虑 x 的第 $d + 1$ 到第 i_{m-1} 个字符组成的子串，它是 S_2 产生的字符串的一个非空前缀，其中 a 的个数应该是 $n(i_{m-1}) - n(d)$ ，而运算符的个数是 $t(i_{m-1}) - t(d)$ ，由于 d, i_{m-1} 都是上述方程的解，所以它们相等，这与引理矛盾。

这就表明，我们可以唯一确定这个分界点的位置。对 S_1, S_2 产生的两个子串递归地做上述分析直到某子串的长度为 1（也就是单个 a ），我们就可以唯一地构造出 x 的一棵语法分析树。

综上所述，题述上下文无关文法没有二义性。

□

5) 由字符 a 与运算符 $+, *$ 构成的后缀表达式。

练习 4.1.2: 为下面语言设计文法：所有由 0 和 1 组成的回文（palindrome）的集合，也就是从前面和从后面读结果都相同的串的集合。

$S \rightarrow 0S0 \mid 1S1 \mid 0 \mid 1 \mid \epsilon$

练习 4.1.3: 下面是一个只包含符号 a 和 b 的正则表达式的文法，其中用 $+$ 替代表示并运算的字符 $|$ ，以避免和文法中作为元符号使用竖线混淆

$\text{rexpr} \rightarrow \text{rexpr} + \text{rterm} \mid \text{rterm}$

$\text{rterm} \rightarrow \text{rterm} \text{rfactor} \mid \text{rfactor}$

$\text{rfactor} \rightarrow \text{rfactor} * \mid \text{rprimary}$

$\text{rprimary} \rightarrow a \mid b$

1) 对该文法提取左因子

2) 提取左因子的变换能使这个文法适用于自顶向下的语法分析技术吗？

- 3) 将提取了左公因子的文法继续消除左递归
4) 此时得到的文法适用于自顶向下的语法分析吗

1、没有左共因子

2、不能，因为文法还存在左递归

3、

```

$$\begin{aligned} rexpr &\rightarrow rterm \ rexpr' \\ rexpr' &\rightarrow +rterm \ rexpr' \mid \varepsilon \\ rterm &\rightarrow rfactor \ rterm' \\ rterm' &\rightarrow rfactor \ rterm' \mid \varepsilon \\ rfactor &\rightarrow rprimary \ rfactor' \\ rfactor' &\rightarrow *rfactor' \mid \varepsilon \\ rprimary &\rightarrow a \mid b \end{aligned}$$

```

4、合适