

上下文无关文法的应用与例题解析

4.8-4.9

第四章复习

- 推导树，二义性
- 上下文无关文法变换
- C范式/G范式
- 下推自动机
- 上下文无关文法与下推自动机相互变换

Chapter 4

推导树

二义性

上下文无关文法变换

消去无用符号 \rightarrow 可达 \downarrow

消去 ϵ

消去单产生式

消去左递归

C范式 / G范式

下推自动机

$M = (Q, T, \Gamma, \delta, q_0, z_0, F)$

$\delta: Q \times (T \cup \epsilon) \times \Gamma \rightarrow Q \times \Gamma^*$

上下文无关文法 \longleftrightarrow 下推自动机
[9.2.1]

Pumping 定理

编辑器中的形式语言

- 编译器中使用正则表达式、正则语言和DFA等工具实现了词法分析。
 - 词法分析就是采用正则语言等工具将输入的字符串分解成一个个的单词流，也就是诸如编程语言中的关键字、标识符这样有特定意义的单词。
- 编程语言必须定义出各种声明、语句和表达式的语法规则。
 - 回顾熟悉的编程语言，我们会发现其语法大都有某种递归的性质。
 - 正则表达式和正则语言无法描述这种递归结构。
 - 而是使用一种表现能力更强的语言——上下文无关语言。

上下文无关文法在语法分析中的应用 1

- 例1

- $G = (\{E\}, \{id, +, (,)\}, P, E)$
- //id是表示变量的标识符
- $E \rightarrow id$
 $E \rightarrow E + E$
 $E \rightarrow (E)$

- 表达式 $(a + b) + c$ 的推导过程 及推导树

上下文无关文法在语法分析中的应用2

- 例2

- $G = (\{E\}, \{id, +, *, (,)\}, P, E)$
- $//id$ 是表示变量的标识符
- $E \rightarrow E + E$
- $E \rightarrow E * E$
- $E \rightarrow (E)$
- $E \rightarrow id$

- 表达式 $a * b + c$ 的推导过程及推导树

- $G = (\{E, T, F\}, \{id, +, *, (,)\}, P, E)$
- $//id$ 是表示变量的标识符
- $E \rightarrow E + T$
- $E \rightarrow T$
- $T \rightarrow T * F$
- $T \rightarrow F$
- $F \rightarrow id$
- $F \rightarrow (E)$

上下文无关文法变换的应用-LL(k)文法

- 在编辑器的语法分析中常使用递归下降法。
 - 利用函数之间的递归调用模拟语法树自上而下的构造过程。
 - 支持递归下降的文法，必须能通过从左往右超前查看k个字符决定采用哪一个产生式，即LL(k)文法。
 - 左边文法无法猜测到底哪个E产生式，右边GFA可以消除这种情况。

```
G = ({E, F}, {id, *, /, (, )}, P, E)
E → F * F
E → F / F
F → id
F → ( E )
```

```
G = ({E, F}, {id, *, /, (, )}, P, E)
E → idG
E → (E)G
G → * F
G → / F
F → id
F → ( E )
```

上下文无关文法变换的应用-消除左递归

- 左递归，就是产生式产生的第一个符号有可能是该产生式本身的非终结符。
 - 如果在编写E的递归下降解析函数时，直接在函数的开头递归调用自己，输入字符串完全没有消耗，这种递归调用就会变成一种死循环。

$$G = (\{E, F\}, \{id, *, /, (,)\}, P, E)$$
$$E \rightarrow E + F$$
$$E \rightarrow F$$
$$F \rightarrow id$$
$$E \rightarrow FG$$
$$G \rightarrow + FG$$
$$F \rightarrow id$$
$$G \rightarrow \varepsilon$$

其他应用

- 计算机求解的很多问题的结构都可以用上下文无关语言表示，这些结构的实例生成都可以转化成文法的句子生成问题。
- 上下文无关文法广泛应用在在软件的测试、调试和实验等方面。
- 许多数据结构可以用文法表示
 - 如树和图。在信息处理和数据库中经常出现的层次数据，可以用文法表示，XML也是以文法为基础的。
- 除了在计算机科学的应用，文法还被应用在其他一些领域中。
 - 例如，文法被应用在基因序列的表示中。

典型例题1

例 1 考察下面的文法, 写出 $x+x/y \uparrow 2$ 的所有最左推导及对应的推导树, 并指出该文法是否具有二义性。

$$G: E \rightarrow E + E \mid E - E \mid E / E \mid E * E \mid E \uparrow E \mid (E) \mid \text{id}$$

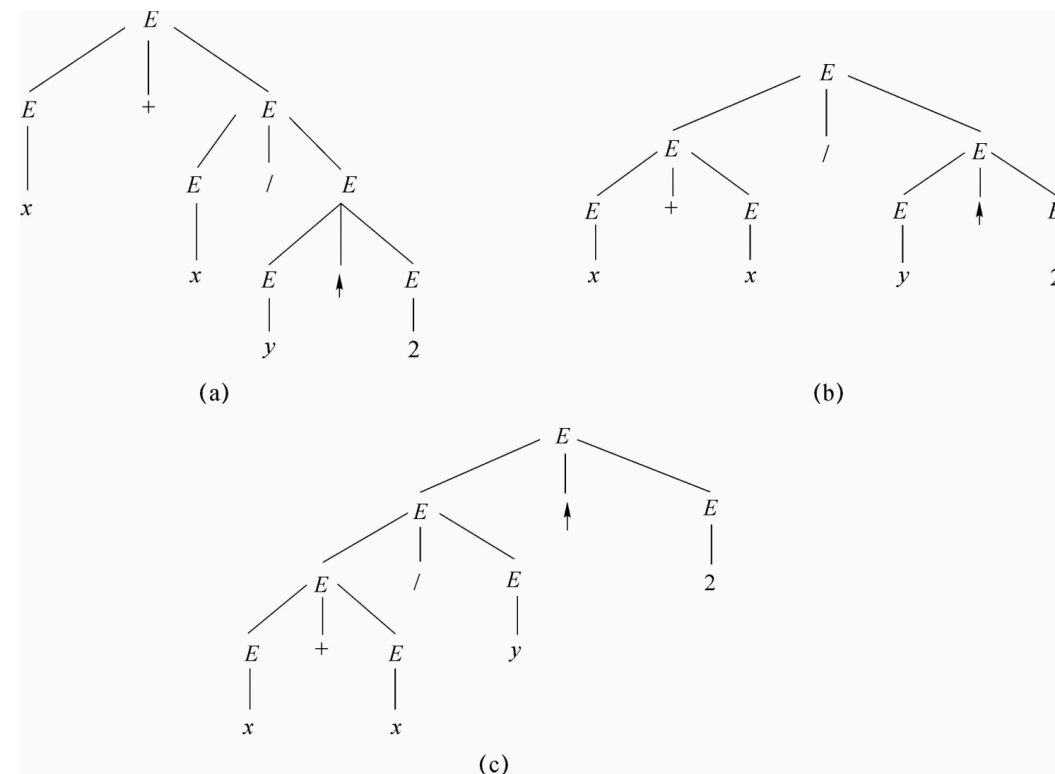
答: $x+x/y \uparrow 2$ 有如下 3 种最左推导:

(a) $E \Rightarrow E + E \Rightarrow x + E \Rightarrow x + E / E \Rightarrow x + x / E \Rightarrow x + x / E \uparrow E \Rightarrow x + x / y \uparrow E \Rightarrow x + x / y \uparrow 2$

(b) $E \Rightarrow E / E \Rightarrow E + E / E \Rightarrow x + E / E \Rightarrow x + x / E \Rightarrow x + x / E \uparrow E \Rightarrow x + x / y \uparrow E \Rightarrow x + x / y \uparrow 2$

(c) $E \Rightarrow E \uparrow E \Rightarrow E / E \uparrow E \Rightarrow E + E / E \uparrow E \Rightarrow x + E / E \uparrow E \Rightarrow x + x / E \uparrow E \Rightarrow x + x / y \uparrow E \Rightarrow x + x / y \uparrow 2$

对应下面 3 棵推导树, 如图 4.9.1 所示。

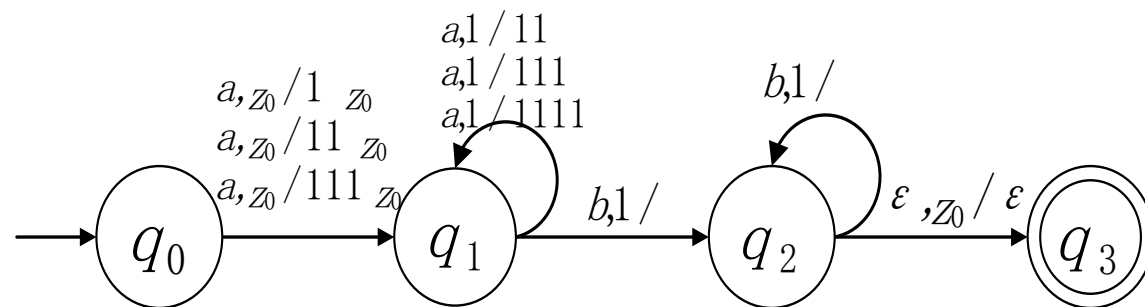


典型例题2

- 已知语言

$$L = \{a^n b^m \mid 0 \leq n \leq m \leq 3n\}$$

- 构造一个NPDA识别该语言。
- 该自动机接受aabbbbb的过程。



上述自动机接受 $aabbbbb$ 的迁移序列为

$$\begin{aligned}
 & (q_0, aabbbbb, z_0) \vdash (q_1, abbbb, 11z_0) \vdash (q_1, bbbb, 11111z_0) \vdash (q_2, bbbb, \\
 & 1111z_0) \vdash (q_2, bbb, 111z_0) \vdash (q_2, bb, 11z_0) \vdash (q_2, b, 1z_0) \vdash (q_2, \epsilon, z_0) \vdash (q_f, \epsilon, z_0)
 \end{aligned}$$

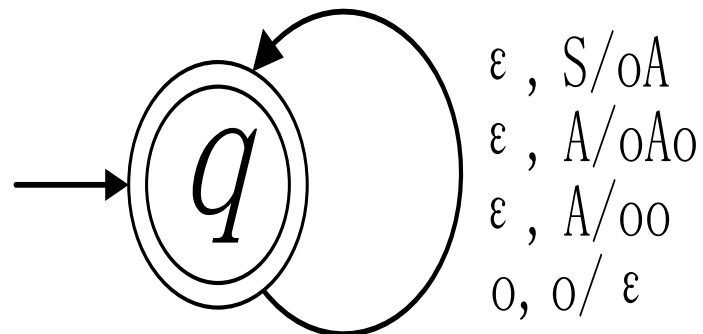
典型例题3

- 设文法**G**有如下的生成式

$$S \rightarrow oA$$

$$A \rightarrow oAo \mid oo$$

- 设计不确定的PDA **M**接受上面文法产生的语言，并采用格局方式写出句子ooo接受的过程。



$$(q, ooo, S) \vdash (q, ooo, oA) \vdash (q, oo, A) \vdash$$

$$(q, oo, oo) \vdash (q, o, o) \vdash (q, \epsilon, \epsilon)$$

典型例题4

- 证明下面的语言不是上下文无关语言

试证明语言 $L = \{a^n b^m a^n b^m \mid m, n \geq 0\}$ 不是上下文无关语言。

证明: 设 $L = \{a^n b^m a^n b^m \mid m, n \geq 0\}$ 是上下文无关语言, 令 $z = a^k b^k a^k b^k$, 其中 k 是泵浦引理

证明: 设 $L = \{a^n b^m a^n b^m \mid m, n \geq 0\}$ 是上下文无关语言, 令 $z = a^k b^k a^k b^k$, 其中 k 是泵浦引理要求的值。根据泵浦引理知, 拆分 $z = uvwxy$ 应满足泵浦引理的条件。

由于 $|vwx| \leq k$, 所以串 vwx 最多是 a 的串, 或 b 的串, 或它们的组合, 即

$$vwx \in a^* \text{ 或 } vwx \in b^*$$

$$vwx \in a^* b^* \text{ 或 } vwx \in b^* a^*$$

又因为 $|vx| \geq 1$, 则符号串 z 中对子串 v 和 x 的抽取仅在一个子串中增加了 a 或 b 的个数, 由 L 的定义知, $uv^2wx^2y \notin L$ 。与泵浦引理矛盾, 故语言 $L = \{a^n b^m a^n b^m \mid m, n \geq 0\}$ 不是上下文无关语言。