第四章 上下文无关文法与下推自动机

- ✧推导树和文法的二义性
- ✧上下文无关文法的变换
  - Chomsky范式
  - Greibach范式
- ✧ 下推自动机
- ✧上下文无关语言的性质

本章要点

- 上下文无关文法（即2型文法）：

  产生式形如 $A \to \alpha$, $A \in N$, $\alpha \in (N \cup T)^*$ 所描述的语言称为上下文无关语言。

- 用途：

  可定义程序设计语言、进行语法分析、简化语言翻译…

- 2型文法对应的识别器——下推自动机

  PDA（Push Down Automata)由输入带、有限控制器和下推栈构成

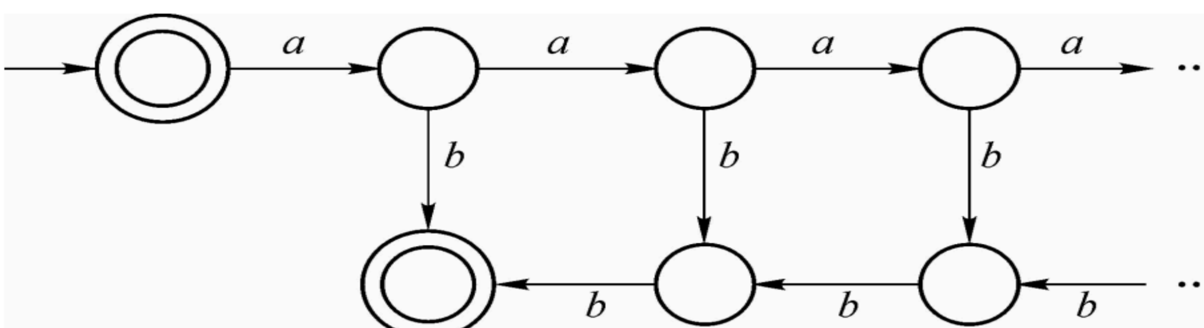
✧ 回顾：在第一讲中介绍过如下内容

设 T= { *0, 1* }， $L = \{ 0^n 1^n \mid n \geq 1 \}$,

如 *0011, 000111, 01* $\in$ *L*, 而 *10, 1001* , ε, *010* $\notin$ *L* .
如下是一个可接受该语言的上下文无关文法:

$S \rightarrow 01$
$S \rightarrow 0S1$

但没有任何有限自动机能够接受语言*L.*



4.1 推导树和二义性

归约与推导的概念:

✧ 推理字符串是否属于文法所定义的语言

一种是自下而上的方法，称为递归推理(***recursive inference***)，递归推理的过程习称为归约；

一种是自上而下的方法，称为推导（***derivation***）.

✧归约过程 将产生式的右部（***body***）替换为产生式的左部（ ***head*** ）.

✧ 推导过程 将产生式的左部（ ***head*** ）替换为产生式的右部（ ***body*** ）.

归约与推导

✧ 归约过程举例

对于CFG $G_{exp} = (\{E,O\}, \{ (,), +, *, v, d \}, P, E)$， P 为

（1） $E \to EOE$

（2） $E \to (E)$

（3） $E \to v$

（4） $E \to d$

（5） $O \to +$

（6） $O \to *$

递归推理出字符串 $v*(v+d)$ 的一个归约过程为

$$v*(v+d) \xrightarrow{(4)} v*(v+E) \xrightarrow{(6)} vO(v+E) \xrightarrow{(3)} vO(E+E)$$

$$\xrightarrow{(5)} vO(EOE) \xrightarrow{(1)} vO(E) \xrightarrow{(2)} vOE \xrightarrow{(3)} EOE \xrightarrow{(1)} E$$

归约与推导

✧ 推导过程举例

对于CFG G_{exp} = ({E,O}, { (,), + , *, v, d }, P, E)， P 为

(1) $E \rightarrow EOE$

(2) $E \rightarrow (E)$

(3) $E \rightarrow v$

(4) $E \rightarrow d$

(5) $O \rightarrow +$

(6) $O \rightarrow *$

从开始符号到字符串 $v*(v+d)$ 的一个推导过程为

$$E \xRightarrow{(1)} EOE \xRightarrow{(6)} E*E \xRightarrow{(2)} E*(E) \xRightarrow{(3)} v*(E)$$

$$\xRightarrow{(1)} v*(EOE) \xRightarrow{(5)} v*(E+E) \xRightarrow{(3)} v*(v+E) \xRightarrow{(4)} v*(v+d)$$

归约与推导

◇ **最左推导(leftmost derivations)**

若推导过程的每一步总是替换出现在最左边的非终结符，则这样的推导称为最左推导. 为方便，最左推导关系用$\underset{lm}{\Rightarrow}$表示，其传递闭包用$\overset{*}{\underset{lm}{\Rightarrow}}$表示.

如对于文法G_{exp}，下面是关于 $v*(v+d)$ 的一个最左推导：

$$E \underset{lm}{\Rightarrow} EOE \underset{lm}{\Rightarrow} vOE \underset{lm}{\Rightarrow} v*E$$

$$\underset{lm}{\Rightarrow} v*(E) \underset{lm}{\Rightarrow} v*(EOE) \underset{lm}{\Rightarrow} v*(vOE)$$

$$\underset{lm}{\Rightarrow} v*(v+E) \underset{lm}{\Rightarrow} v*(v+d)$$

$E \rightarrow EOE$
$E \rightarrow (E)$
$E \rightarrow v$
$E \rightarrow d$
$O \rightarrow +$
$O \rightarrow *$

归约与推导

✧ 最右推导(rightmost derivations)

若推导过程的每一步总是替换出现在最右边的非终结符，则这样的推导称为**最右推导**. 为方便，最右推导关系用 $\underset{rm}{\Rightarrow}$ 表示，其传递闭包用 $\underset{rm}{\overset{*}{\Rightarrow}}$ 表示.

如对于文法 G_{exp}，下面是关于 $v*(v+d)$ 的一个最右推导：

$$E \underset{rm}{\Rightarrow} EOE \underset{rm}{\Rightarrow} EO(E) \underset{rm}{\Rightarrow} EO(EOE)$$

$$\underset{rm}{\Rightarrow} EO(EOd) \underset{rm}{\Rightarrow} EO(E+d) \underset{rm}{\Rightarrow} EO(v+d)$$

$$\underset{rm}{\Rightarrow} E*(v+d) \underset{rm}{\Rightarrow} v*(v+d)$$

$E \to EOE$
$E \to (E)$
$E \to v$
$E \to d$
$O \to +$
$O \to *$

课堂练习

对于前缀表达式文法**G1:**

E::= – EE

E::= – E

E::= a | b | c

$E \rightarrow -EE \rightarrow --EE \rightarrow --aE \rightarrow --a-EE \rightarrow --a-bE \rightarrow --a-bc$

$E \rightarrow -EE \rightarrow -E-EE \rightarrow -E-Ec \rightarrow -E-bc \rightarrow --E-bc \rightarrow --a-bc$

$--a-bc \rightarrow --a-bE \rightarrow --a-EE \rightarrow --aE \rightarrow --EE \rightarrow -EE \rightarrow E$

推导树

用图的方法表示一个句型的推导，这种图称为推导树（也称语法树或语法分析树）。有助于理解语法结构的层次。

定义方法：

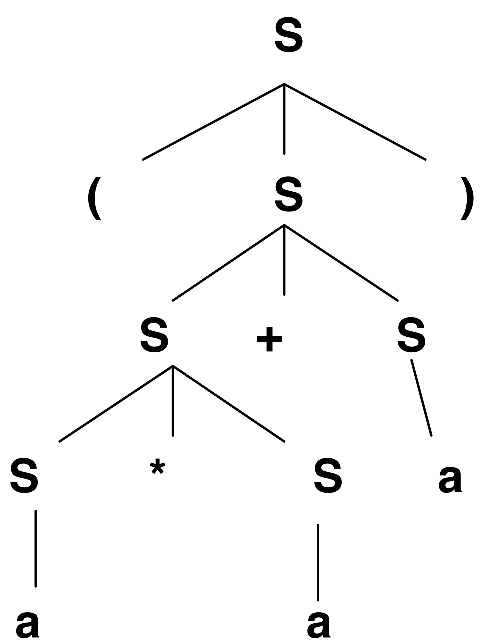
- 文法的起始符为根，树的枝结点标记是非终结符，叶结点标记为终结符或ε。
- 若枝结点有直接子孙$x_1, x_2, \ldots, x_k$，则文法中有生成式$A \rightarrow x_1 x_2 \ldots x_k$

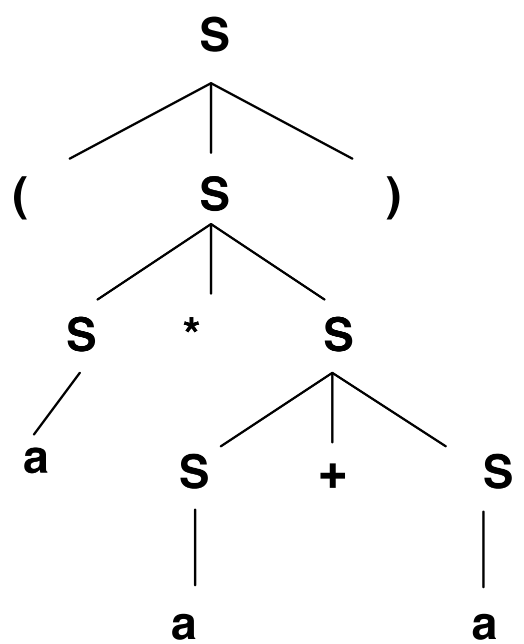
推导树举例

例：（书P94 例1）

文法S→S+S | S*S |(S)| a，

对句子 (a*a+a) 可有推导树



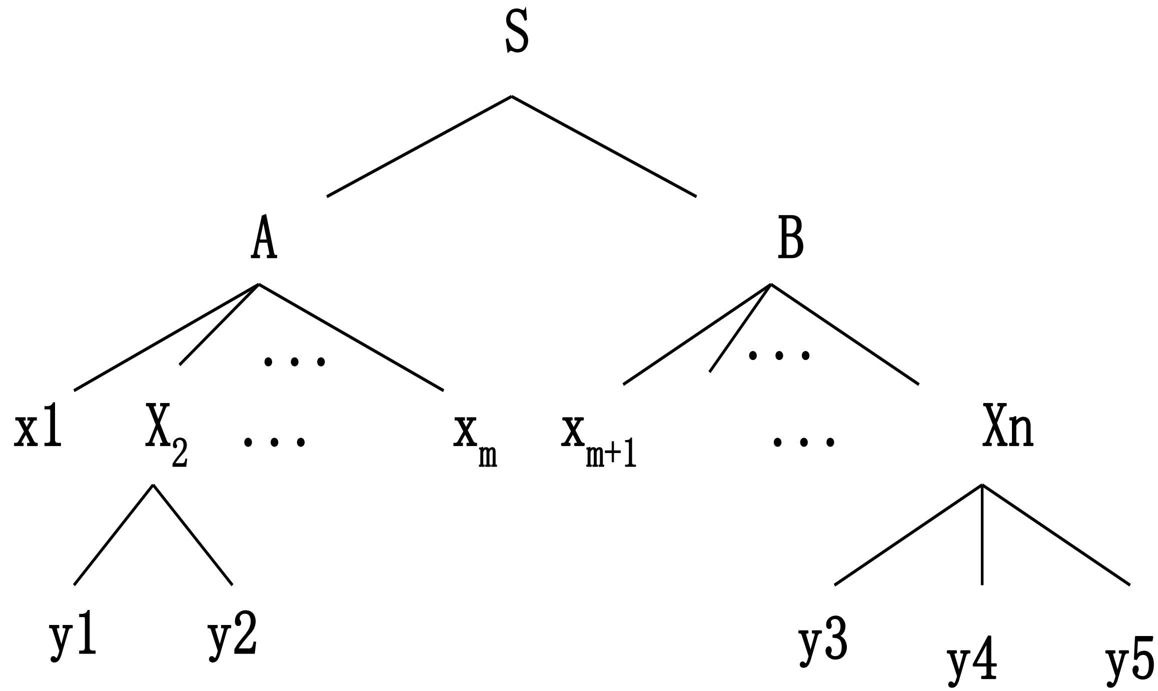


即：推导树是对文法G中一个<u>特定句子形式</u>的派生过程所做的一种自然描述。

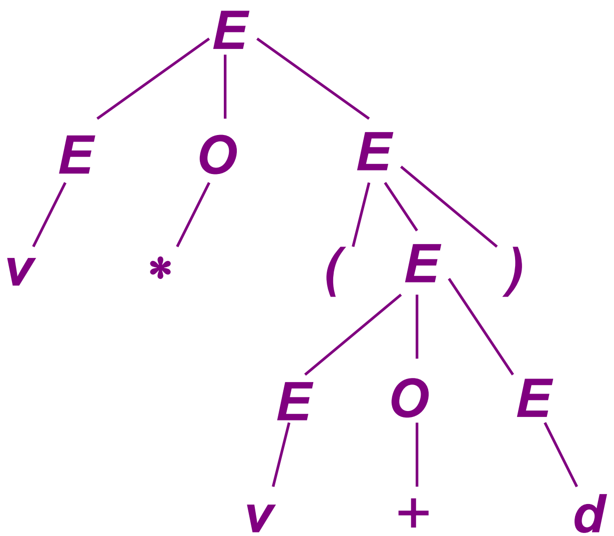
边缘

叶子从左向右组成的字符串称为推导树的边缘。如图

$x_1\ y_1\ y_2\ x_3 \cdots\ x_m\ x_{m+1} \cdots\ x_{n-1}\ y_3\ y_4\ y_5$是树的边缘



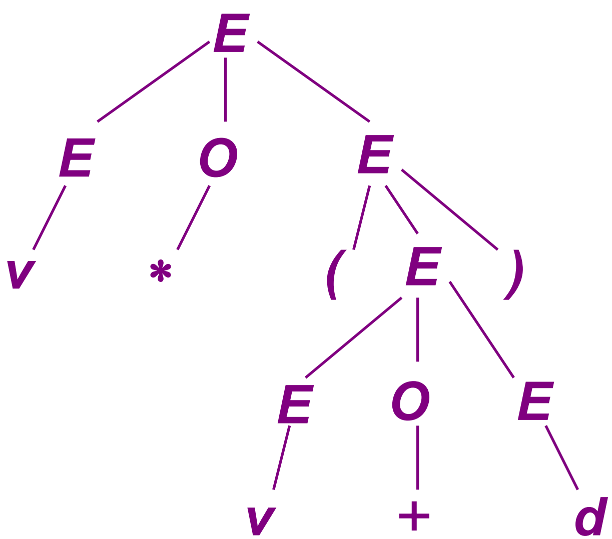
✧ 归约过程自下而上构造了一棵树 如对于文法 G_{exp}，关于 $v*(v+d)$ 的一个归约过程可以认为是构造了如下一棵树：



(1) $E \rightarrow EOE$
(2) $E \rightarrow (E)$
(3) $E \rightarrow v$
(4) $E \rightarrow d$
(5) $O \rightarrow +$
(6) $O \rightarrow *$

$$v*(v+d) \xrightarrow{(4)} v*(v+E) \xrightarrow{(6)} vO(v+E) \xrightarrow{(3)} vO(E+E)$$
$$\xrightarrow{(5)} vO(EOE) \xrightarrow{(1)} vO(E) \xrightarrow{(2)} vOE \xrightarrow{(3)} EOE \xrightarrow{(1)} E$$

✧ 推导过程自上而下构造了一棵树　如对于文法G_{exp}，关于 $v*(v+d)$ 的一个推导过程可以认为是构造了如下一棵树：



(1) $E \rightarrow EOE$
(2) $E \rightarrow (E)$
(3) $E \rightarrow v$
(4) $E \rightarrow d$
(5) $O \rightarrow +$
(6) $O \rightarrow *$

$$E \xrightarrow{(1)} EOE \xrightarrow{(6)} E*E \xrightarrow{(2)} E*(E) \xrightarrow{(3)} v*(E)$$

$$\xrightarrow{(1)} v*(EOE) \xrightarrow{(5)} v*(E+E) \xrightarrow{(3)} v*(v+E) \xrightarrow{(4)} v*(v+d)$$

归约、推导与分析树之间关系

✧ 三者之间的关系

设 CFG $G = (V, T, P, S)$. 以下命题是相互等价的：

(1) 字符串 $w \in T^*$ 可以归约（递归推理）到非终结符 A;

(2) $A \stackrel{*}{\Rightarrow} w$;

(3) $A \underset{lm}{\stackrel{*}{\Rightarrow}} w$;

(4) $A \underset{rm}{\stackrel{*}{\Rightarrow}} w$;

(5) 存在一棵根结点为 A 的分析树，其边缘为 w.

归约、推导与分析树之间关系

定理：

设2型文法G=（N，T，P，S），如果存在S=>+ ω，当且仅当文法G中有一棵边缘为ω的推导树。

证明：

需证明对任意枝结点B∈N，有B=>* ω当且仅当存在边缘为ω的B树（根为B的树）

子树概念：

一棵派生树的子树，是树中的某个顶点连同它的全部后裔，以及连接这些后裔的边。

证明步骤：

1. 证当ω是B树边缘时，有B =>* ω

   设B树边缘为ω，对树中枝结点数目m作归纳证明。

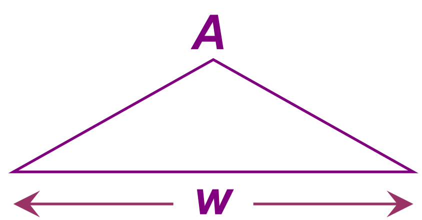
2. 设有B =>* ω，证明存在一棵边缘为ω的B树。

   对推导步数作归纳

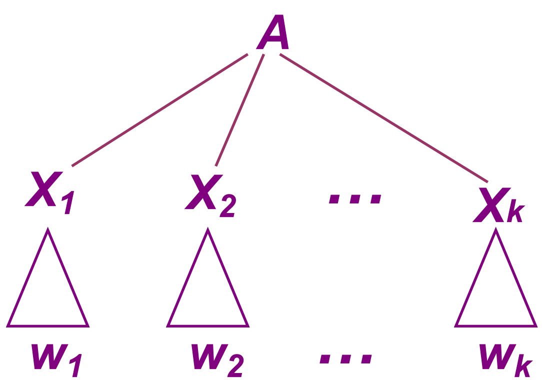
1. 证当ω是B树边缘时，有B =>* ω

设B树边缘为ω，对树中枝结点数目m作归纳证明。

基础 m为1. 分析树一定如右图所示，必定有产生式$A \to w$. 因此，$A \stackrel{*}{\Rightarrow} w$.



归纳 m大于1的分析树一定如右图所示，必定有产生式 $A \to X_1X_2\ldots X_k$. 存在$w_1,w_2,\ldots,w_k$，w_i是X_i子树的边缘（$1\le i \le k$），且$w = w_1w_2\ldots w_k$，由归纳假设，$X_i \stackrel{*}{\Rightarrow} w_i$（$1\le i \le k$）. 在此基础上易证得$A \stackrel{*}{\Rightarrow} w$.



2. 设有B =>* ω，证明存在一棵边缘为ω的B树。

对推导步数作归纳

基础 步数为 1. 一定有产生式 $A \rightarrow w$. w 可以归约到 A.

归纳 设步数大于 1，第一步使用了产生式$A \rightarrow X_1X_2...X_k$. 该推导如 $A \Rightarrow X_1X_2...X_k \overset{*}{\Rightarrow} w$. 可以将 w 分成 $w = w_1w_2...w_k$，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 $X_i \overset{*}{\Rightarrow} w_i$. 由归纳假设，$w_i$ 可以归约到 X_i .

这样，w_i 或者为X_i，或者可以归约到 X_i，使用产生式$A \rightarrow X_1X_2...X_k$，得出$w$ 可以归约到 A.

二义性

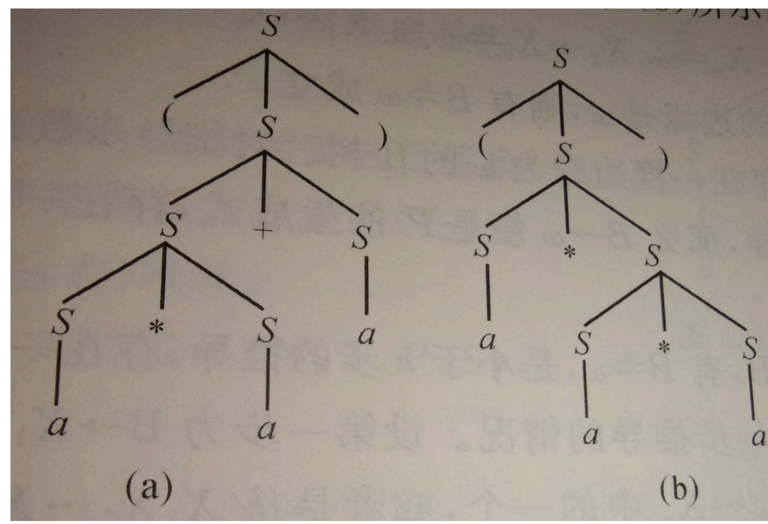
定义:

2型文法是二义的,当且仅当对于句子ω∈L(G),存在两棵不同的具有边缘为ω的推导树。

(即：如果文法是二义的, 那么它所产生的某个句子必然能从不同的最左(右)推导推出)。

例: (书P94 例1)

句子(a*a+a)有二棵不同的推导树. (相当于一个先算乘法,一个先算加法.)

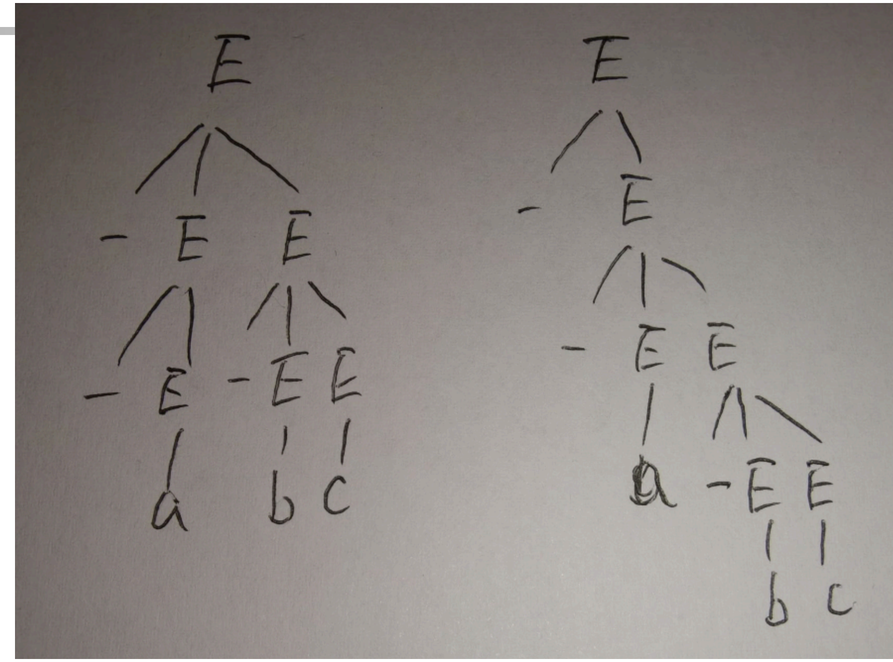


注意:

可有二个文法,一个有二义,一个无二义,但产生相同的语言.

可否通过变换消除二义性? —— 无一般的算法!

课堂练习



对于前缀表达式文法**G1:**

E::= – EE

E::= – E

E::= a | b | c

画出文法的句子 **– – a – bc** 的所有可能语法树，判断该文法是否具有二义性。

作业

Ch4 习题：1. 2. 3.