

第四章 自顶向下语法分析方法



•4.1 确定的自顶向下分析思想

•4.2 LL(1)文法的判别

广东工业大学计算机学院

一、语法分析任务

- ✓ **词法分析阶段**，主要介绍了单词符号的结构、识别（用状态转换图），描述（通过正规式）以及有限自动机**DFA**和**NFA**。
- ✓ 在一个编译程序对某个源程序完成了词法工作以后,就进入了**语法分析阶段**。
- ✓ 由词法分析程序所产生的单词符号流，作为语法分析程序的输入串，按文法规则分析检查是否构成了合法的句子。

引言



一、语法分析任务

语法检查

- ✓ 根据语法规则对各种语法成分进行分析;
- ✓ 确定它们的语法关系以检查语法上的正确和错误;
- ✓ 指出错误的性质和出错位置。

如: **If B then S1 else S2** **正确**

若写成 **If B then else S2** **错误**

then后少一个S1

- 语法分析方法分类
 - ✓ 语法分析方法很多，但能够产生计算机程序并能得到广泛应用的主要有两大类，按照生成语法树的顺序：
 - 1. 自顶向下(自上而下)分析：也称面向目标的分析方法，就是从文法的开始符号出发，按最左推导方式向下推导，试图推导出要分析的输入串。
开始符号 \Rightarrow 输入符号串
 - 2. 自底向上(自下而上)分析：从输入符号串开始，按最左归约方式向上归约到文法的开始符号。
开始符号 $\xleftarrow{\text{归约}}$ 输入符号串

引言

Logo

- 两种方法反映了两种不同的语法树的构造过程：

- 自上而下的语法分析

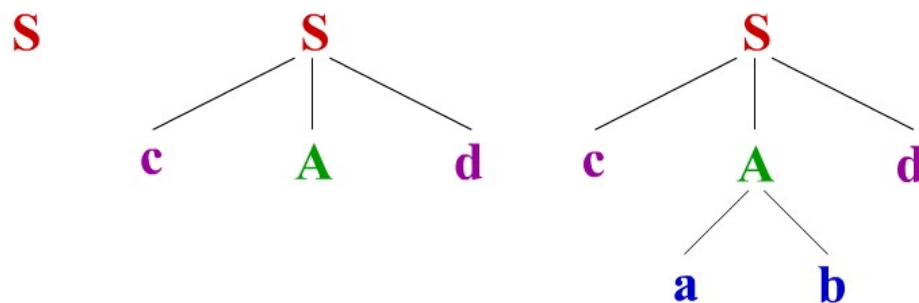
自下而上的语法分析

例：文法G： $S \rightarrow cAd$

$A \rightarrow ab$

$A \rightarrow a$

识别输入串 $w=cabd$ 是否该文法的句子



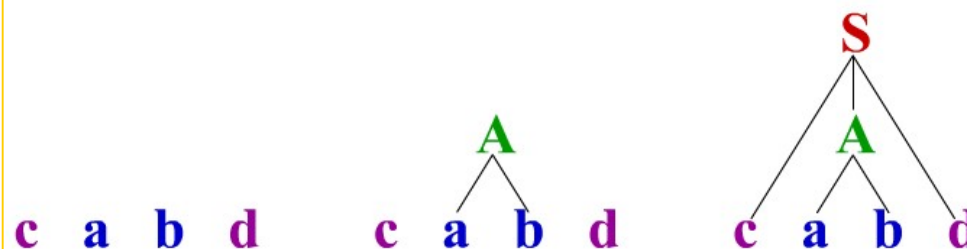
推导过程： $S \Rightarrow cAd \Rightarrow cabd$

例：文法G： $S \rightarrow cAd$

$A \rightarrow ab$

$A \rightarrow a$

识别输入串 $w=cabd$ 是否该文法的句子

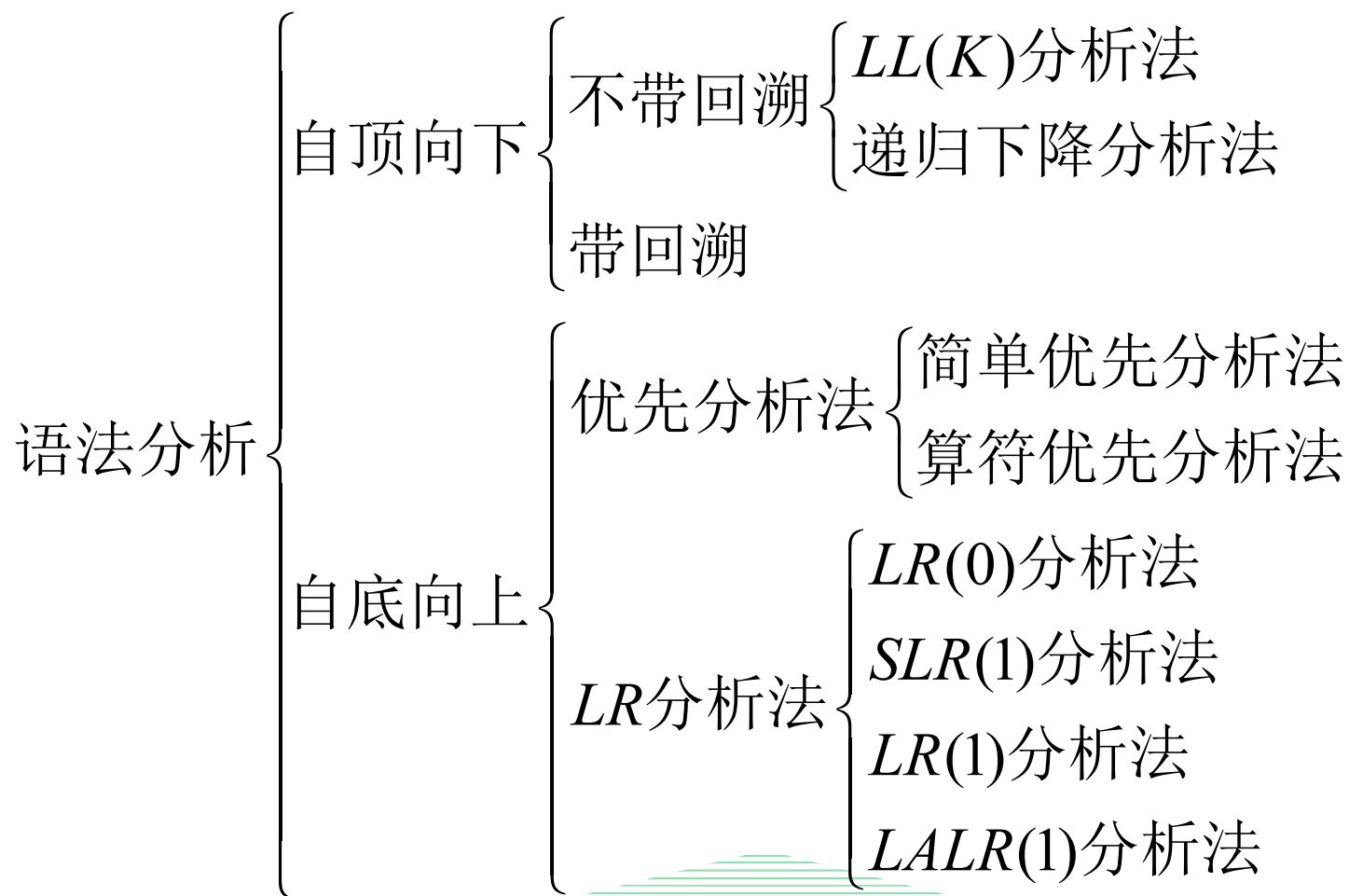


归约过程： $S \Leftarrow cAd \Leftarrow cabd$



引言

Logo





引言

Logo

- 本章主要学习内容

本章将介绍自顶向下语法分析思想，LL(1) 文法的判别，一些非LL(1) 文法到LL(1) 文法的等价转换，预测分析方法的工作机制。

本章重点

- 自顶向下分析思想
- 掌握**FIRST**集、**FOLLOW**集、**SELECT**集的计算，判别文法是否为**LL(1)** 文法
- 某些非**LL(1)** 文法向**LL(1)** 文法转换的技术*
- 掌握**LL(1)** 文法预测分析表的计算方法，并利用预测分析方法对符号串进行分析

预备知识：句子、句型、推导

Logo

- 在了解自顶向下分析方法之前，首先回顾在“文法和语言”一章的一些概念：
- **1. 句型**：文法 $G[S] = (V_N, V_T, P, S)$ ，若 $S \xRightarrow{*} x$ ，且 $x \in V^*$ 则称 x 是文法 $G[S]$ 的句型。
- **2. 句子**：文法 $G[S] = (V_N, V_T, P, S)$ ，若 $S \xRightarrow{*} x$ ，且 $x \in V_T^*$ ，则称 x 是文法 $G[S]$ 的句子。
- **例**： $G[S]: S \rightarrow 0S1, S \rightarrow 01$
可有推导 $S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 000S111 \Rightarrow 00001111$
- 说明 **00001111** 是 $G[S]$ 的句子。
- **3. 最左(最右)推导**：在推导的任何一步 $\alpha \Rightarrow \beta$ (其中 α 、 β 是句型)，都是对 α 中的最左(右)非终结符进行替换。
- 最右推导被称为**规范推导**。
- 由**规范推导**所得的句型称为**规范句型**。



要解决的基本问题

Logo

- 如何选择使用哪个产生式进行推导？
假定要被替换的最左非终结符号是**V**，且左部为**V**的规则有**n**条： $V \rightarrow A_1 | A_2 | \dots | A_n$ ，那么如何确定用哪个右部去替换**V**呢？



本课内容

Logo

- 5.1 确定的自顶向下分析思想
- 5.2 LL(1)文法的判别



确定的自顶向下分析方法

Logo

- 确定的自顶向下分析方法，首先从某文法的开始符号**S**出发，对给定的输入符号串**t**，如能根据当前的输入符号**a**(单词符号)唯一地确定选用某个产生式替换相应非终结符往下推导，或构造一棵相应的语法树，则分析是确定的；

自顶向下分析方法举例1

Logo

- 例4.1 若有文法 $G_1[S]$:

$S \rightarrow pA \mid qB \quad A \rightarrow cAd \mid a \quad B \rightarrow dB \mid c$

- 识别输入串 $w = pccadd$ 是否是 $G_1[S]$ 的句子。
- 试探推导过程: $S \Rightarrow pA \Rightarrow pcAd \Rightarrow pccAdd \Rightarrow pccadd$

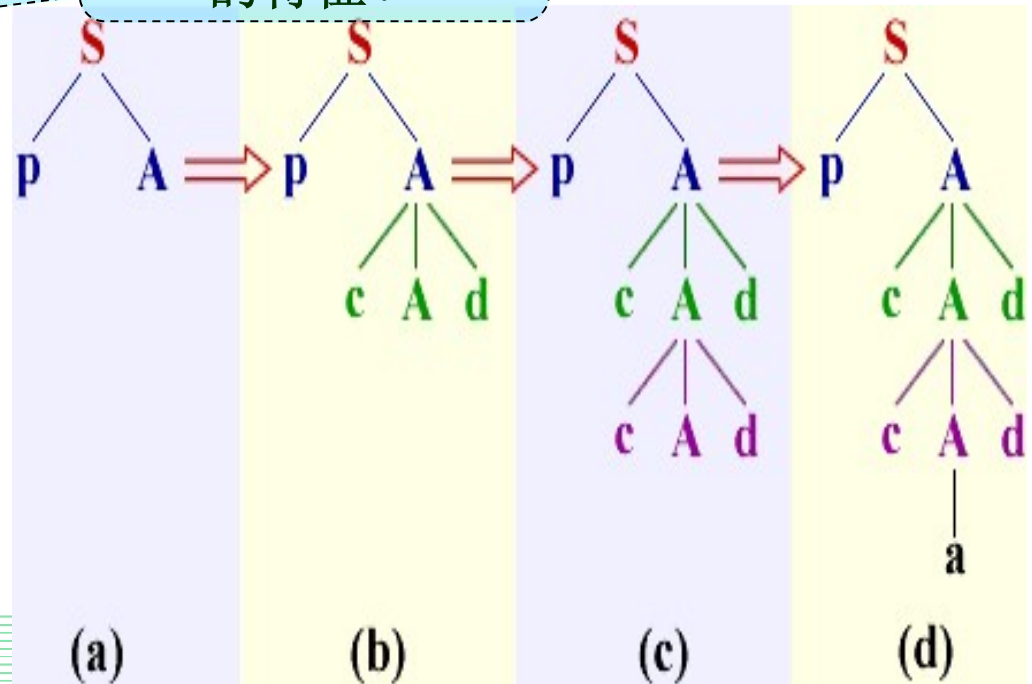
- 该文法有以下两个特点:

- ① 每个产生式的右部都由终结符号开始。

- ② 如果两个产生式有相同的左部, 那么它们的右部由不同的终结符号开始。

- 对于这样的文法, 在推导过程中完全可以**根据当前的输入符号决定选择哪个产生式往下推导**, 因此其分析过程是**唯一确定的**。

这是哪一种文法的特征?



自顶向下分析方法举例2

Logo

- 例 4.2 若有文法 $G_2[S]$:
 $S \rightarrow Ap|Bq$ $A \rightarrow a|cA$ $B \rightarrow b|dB$
这是哪一种文法?
- 识别输入串 $w = \text{ccap}$ 是否是 $G_2[S]$ 的句子。该文法的特点是:
 - ① 产生式的右部不全是由终结符开始。
 - ② 如果两个产生式有相同的左部，它们的右部是由不同的终结符或非终结符开始。
 - ③ 文法中无空产生式。
- 对于左部相同的产生式，其右部以非终结符开始时，在推导过程中应该如何选择?
- 这不像例 5.1 文法直观，对于 $w = \text{ccap}$ 为输入串时，其第一个符号是 c ，这时从 S 出发选择 $S \rightarrow Ap$ 还是选择 $S \rightarrow Bq$?
- 这时需要知道， Ap 或 Bq 它们的开始符号集合是什么，若 c 是包含在 Ap 的开始符号集合中，且不包含在 Bq 的开始符号集合中，则选择 $S \rightarrow Ap$ 往下进行推导。
 Ap 的开始符号集是: $\{a, c\}$
- 同样若 c 是包含在 Bq 的开始符号集合中，且不包含在 Ap 的开始符号集合中，则选择 $S \rightarrow Bq$ 往下推导。
 Bq 的开始符号集是: $\{b, d\}$
- 其它情况则为不确定推导或出错。

FIRST集的定义

Logo

- 设 $G[S] = (V_N, V_T, P, S)$ 是上下文无关文法，则
$$\text{FIRST}(\alpha) = \{a \mid \alpha \xRightarrow{*} a\beta, a \in V_T, \alpha, \beta \in V^*\}$$
- 称 $\text{FIRST}(\alpha)$ 为 α 的开始符号集或首符号集。
- 特别地，若 $\alpha \xRightarrow{*} \varepsilon$ ，则规定 $\varepsilon \in \text{FIRST}(\alpha)$

注意：仅要求 α 是含有非终结符的符号串

- 例如，在文法 $G2[S]$:

$S \rightarrow Ap|Bq \quad A \rightarrow a|cA \quad B \rightarrow b|dB$

- 中：
- $\text{FIRST}(Ap) = \{a, c\}$
- $\text{FIRST}(Bq) = \{b, d\}$

自顶向下分析方法举例2(续)

Logo

- 例 4.2 若有文法 $G_2[S]$:

$S \rightarrow Ap|Bq \quad A \rightarrow a|cA \quad B \rightarrow b|dB$

- 识别输入串 $w = \text{ccap}$ 是否是 $G_2[S]$ 的句子。

- 解: 试探推出输入串的推导过程为: Ap 的开始符号集是: $\{a, c\}$

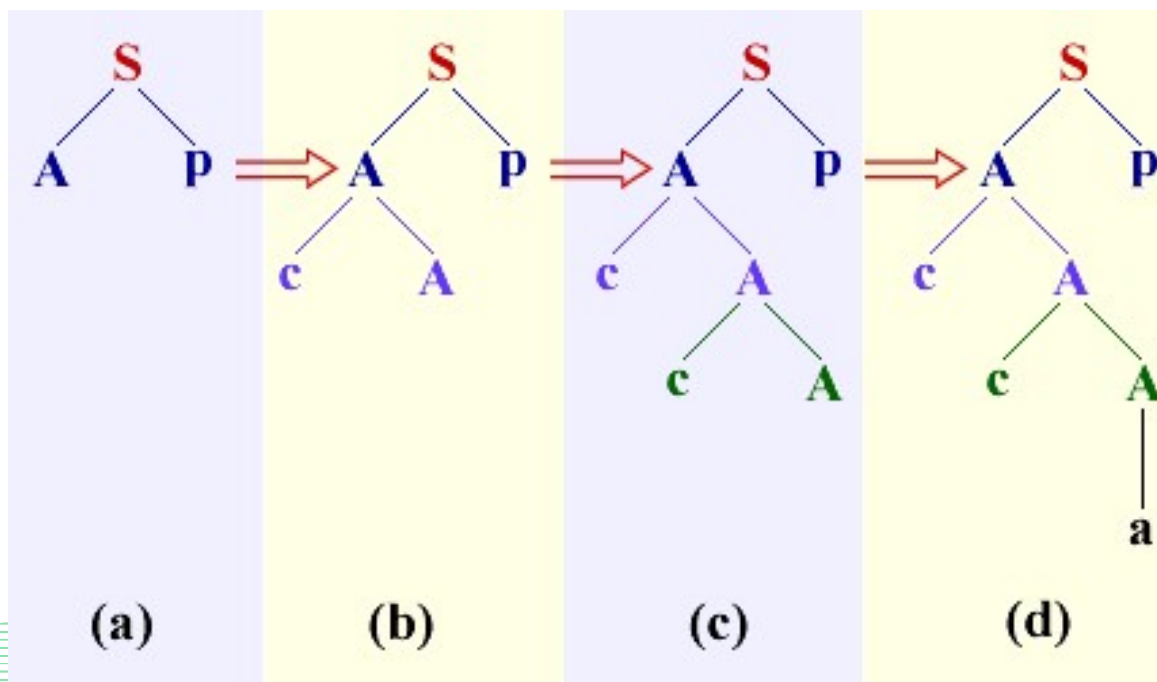
$S \Rightarrow Ap \Rightarrow cAp \Rightarrow ccAp \Rightarrow \text{ccap}$

在推导过程中读入 a 时,
选取 $S \rightarrow Ap$ 而不选择
 $S \rightarrow Bq$ 的原因是:

(1) $\text{FIRST}(Ap) = \{a, c\}$

(2) $\text{FIRST}(Bq) = \{b, d\}$

$\text{FIRST}(Bq)$ 中不含有 a



自顶向下分析方法举例3

Logo

- 若有文法 **G3[S]**:
 $S \rightarrow aA|d$ $A \rightarrow bAS|\epsilon$
- 识别输入串 **w = abd** 是否是 **G3[S]** 的句子。
- 此文法的特点是：文法中含有空产生式。
- 有如下结论：当某一非终结符的产生式中**含有空产生式**时，
- (1) 相同左部的**非空产生式右部的首符号集**两两不相交；
- (2) 并与在推导过程中“紧跟该**非终结符**后边可能出现的**终结符集**” (称作**FOLLOW集**) **也不相交**；
- 则仍可构造**确定的自顶向下分析**。
- 解**：试探推导出 **abd** 的推导过程为：
 $S \Rightarrow aA \Rightarrow \underline{abAS} \Rightarrow \underline{abS} \Rightarrow \underline{abd}$

由 $\text{FIRST}(bAS) \cap \text{FOLLOW}(A) = \emptyset$ 可确定选用 $A \rightarrow \epsilon$ 和 $S \rightarrow d$

(1) $\text{FIRST}(bAS) = \{b\}$
(2) $\text{FOLLOW}(A) = \text{FIRST}(S) = \{a, d\}$ /* 对于 $A \rightarrow bAS$ */
(#其实也是 $\text{Follow}(A)$ 的一部分)
 $\text{FIRST}(bAS) \cap \text{FOLLOW}(A) = \emptyset$

FOLLOW集的定义

*

Logo

- (1) 设 $G[S] = (V_N, V_T, P, S)$ 是上下文无关文法, $A \in V_N$, S 是开始符号, 则
$$FOLLOW(A) = \{a \mid S \xRightarrow{*} \mu A \beta \text{ 且 } a \in FIRST(\beta), \mu \in V_T^*, \beta \in V^+\}$$
- 另外, 若 $S \xRightarrow{*} \mu A \beta$, 且有 $\beta \xRightarrow{*} \varepsilon$, 则 $\# \in FOLLOW(A)$ 。
- (2) 也可定义为: 称 $FOLLOW(A) = \{a \mid S \xRightarrow{*} \dots A a \dots, a \in V_T\}$, 若有 $S \xRightarrow{*} \dots A$, 则规定 $\# \in FOLLOW(A)$ 。
- ‘ $\#$ ’ 是输入串的结束符, 或者称输入串括号。
- 因此当文法中含有形如 $A \rightarrow \alpha$, $A \rightarrow \beta$ 的产生式时 (其中 α 和 β 不能同时推导出空), 不妨假定 $\alpha \not\Rightarrow \varepsilon$, $\beta \xRightarrow{*} \varepsilon$, 则当 $FIRST(\alpha) \cap (FIRST(\beta) \cup FOLLOW(A)) = \emptyset$ 时, 对非终结符 A 的替换仍可确定采用哪一条产生式。
- 例如: 对应文法 $G_3[S]$: $S \rightarrow aA \mid d$ $A \rightarrow bAS \mid \varepsilon$
- abd 的推导过程为: $S \Rightarrow aA \Rightarrow \underline{abAS} \Rightarrow \underline{abS} \Rightarrow abd$
- 读入 d 时, 需要决定选取哪一条 A 为左部的产生式: $A \rightarrow bAS$ 或 $A \rightarrow \varepsilon$?
- 其中 $FIRST(bAS) = \{b\}$, $FIRST(\varepsilon) = \{\varepsilon\}$, $FOLLOW(A) = FIRST(S) \cup \{\#\} = \{a, d, \#\}$, 所以 $abAS \Rightarrow abS$ 中可以唯一确定产生式 $A \rightarrow \varepsilon$

FOLLOW集应用举例

Logo

- 当文法中含有形如 $A \rightarrow \alpha$, $A \rightarrow \beta$ 的产生式时(其中 α 和 β 不能同时推导出空), 不妨假定 $\alpha \not\Rightarrow \epsilon$, $\beta \xRightarrow{*} \epsilon$, 则当 $FIRST(\alpha) \cap (FIRST(\beta) \cup FOLLOW(A)) = \emptyset$ 时, 对非终结符 A 的替换仍可确定采用哪一条产生式。
- 例如: 对应文法 $G_3[S]'$: $S \rightarrow aA|b$ $A \rightarrow bAS|\epsilon$
- abb 的推导过程是?
- 可能性(1): $S \Rightarrow aA \Rightarrow \underline{abAS} \Rightarrow \underline{abS} \Rightarrow \underline{abb}$ 应用 $A \rightarrow \epsilon$ 和 $S \rightarrow b$
- 可能性(2): $S \Rightarrow aA \Rightarrow \underline{abAS} \Rightarrow \underline{ab}bASS \Rightarrow \dots$ 应用 $A \rightarrow bAS$
- (2) 不能推出 abb 。但在 $aA \Rightarrow abAS$ 之后, 能够确定可能性(1)吗?
- 读入最后一个 b 时, 如何决定选取 $A \rightarrow bAS$ 或 $A \rightarrow \epsilon$?
- 其中 $FIRST(bAS) = \{b\}$, $FIRST(\epsilon) = \{\epsilon\}$, $FOLLOW(A) = FIRST(S) \cup \{\#\} = \{a, b, \#\}$
- $FIRST(bAS) \cap (FIRST(\epsilon) \cup FOLLOW(A)) = \{b\}$,
- 所以对于 $abAS \Rightarrow abS$ 中 A 的替换, 不能确定使用产生式 $A \rightarrow \epsilon$

SELECT集的定义

Logo

- 给定上下文无关文法的产生式 $A \rightarrow \alpha$, $A \in V_n$, $\alpha \in V^*$, **SELECT**集的定义如下:
- 若 $\alpha \not\Rightarrow^* \varepsilon$, 则 $\text{SELECT}(A \rightarrow \alpha) = \text{FIRST}(\alpha)$ 。
- 若 $\alpha \Rightarrow^* \varepsilon$, 则 $\text{SELECT}(A \rightarrow \alpha) = (\text{FIRST}(\alpha) - \{\varepsilon\}) \cup \text{FOLLOW}(A)$ 。

• 例如: 对应文法 $G_3[S]$: $S \rightarrow aA|d$ $A \rightarrow bAS|\varepsilon$

• 可以求得各公式的**SELECT**集:

• $\text{SELECT}(S \rightarrow aA) = \text{FIRST}(aA) = \{a\}$

• $\text{SELECT}(S \rightarrow d) = \text{FIRST}(d) = \{d\}$

• $\text{SELECT}(A \rightarrow bAS) = \{b\}$

• $\text{SELECT}(A \rightarrow \varepsilon) = ? \quad ? \quad ?$
 $= (\text{FIRST}(\varepsilon) - \{\varepsilon\}) \cup \text{FOLLOW}(A)$
 $= \text{FOLLOW}(A) = \text{FIRST}(S) \cup \{\#\}$
 $= \{a, d, \#\}$

若 $S \Rightarrow^* \dots A$, 则 $\# \in \text{FOLLOW}(A)$

考察 $A \rightarrow bAS$

考察 $A \rightarrow \varepsilon$



LL(1)文法

Logo

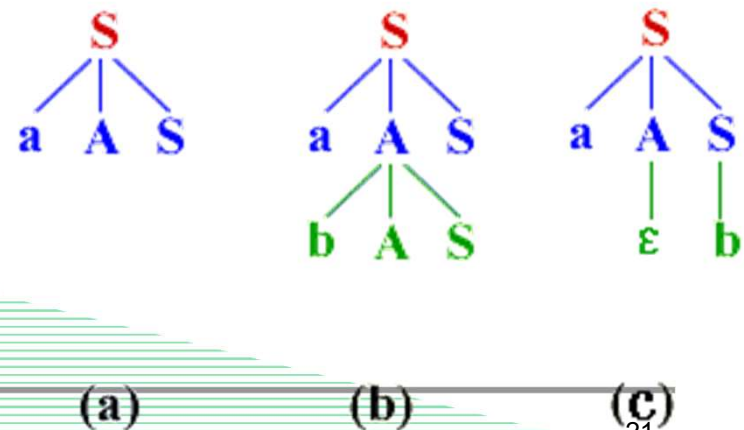
- 一个上下无关文法是LL(1)文法 \Leftrightarrow 对每个非终结符A的任意两个不同产生式 $A \rightarrow \alpha$ 和 $A \rightarrow \beta$, 满足:
$$\text{SELECT}(A \rightarrow \alpha) \cap \text{SELECT}(A \rightarrow \beta) = \emptyset$$
- 其中 α 、 β 不能同时 $\xRightarrow{*} \varepsilon$ 。
- 结论: 能够使用自顶向下分析技术的文法就是LL(1)文法。
- 例如: 对应文法G3[S]: $S \rightarrow aA|d$ $A \rightarrow bAS|\varepsilon$
- 可以求得各公式的SELECT集:
- $\text{SELECT}(S \rightarrow aA) = \{a\}$ $\text{SELECT}(S \rightarrow d) = \{d\}$
- $\therefore \text{SELECT}(S \rightarrow aA) \cap \text{SELECT}(S \rightarrow d) = \emptyset$
- $\text{SELECT}(A \rightarrow bAS) = \{b\}$ $\text{SELECT}(A \rightarrow \varepsilon) = \{a, d, \#\}$
- $\therefore \text{SELECT}(A \rightarrow bAS) \cap \text{SELECT}(A \rightarrow \varepsilon) = \emptyset$
- 所以G3[S]是LL(1)文法。

LL(1)文法举例

Logo

- 若有文法G4[S]:
(1) $S \rightarrow aAS$ (2) $S \rightarrow b$ (3) $A \rightarrow bAS$ (4) $A \rightarrow \epsilon$
- 则 $\text{SELECT}(S \rightarrow aAS) = \{a\}$ $\text{SELECT}(S \rightarrow b) = \{b\}$
- $\therefore \text{SELECT}(S \rightarrow aAS) \cap \text{SELECT}(S \rightarrow b) = \emptyset$ $= \text{FOLLOW}(A)$
 $= \text{FIRST}(S) =$
- $\text{SELECT}(A \rightarrow bAS) = \{b\}$ $\text{SELECT}(A \rightarrow \epsilon) = \{a, b\}$
- $\therefore \text{SELECT}(A \rightarrow bAS) \cap \text{SELECT}(A \rightarrow \epsilon) \neq \emptyset$
- \therefore G4[S]不是LL(1)文法, 不能使用自顶向下分析。
- 讨论: 对输入串ab#的试探推导, 会有什么问题?

- 推导可能性(一): $S \Rightarrow aAS$
- $\Rightarrow abASS \Rightarrow \dots$ 失败(选用(3))
- 推导可能性(二): $S \Rightarrow aAS$
- $\Rightarrow aS \Rightarrow ab$ 成功(选用(4))



与LL(1)文法有关的结论

Logo

- 结论：LL(**1**)文法是无二义的。
- LL(**1**)文法的含义是：
 - 第一个**L**表明自顶向下分析是从左向右扫描输入串；
 - 第二个**L**表明分析过程中将用最左推导。
- ‘**1**’表明只需向右看一个符号便可决定选择哪个产生式(规则)进行推导。
- 类似也可以有LL(**K**)文法，也就是需向前查看**K**个符号才可确定选用哪个产生式。通常采用**K = 1**，个别情况采用**K = 2**。



本课内容

Logo

- 4.1 确定的自顶向下分析思想
- 4.2 LL(1)文法的判别

判别LL(1)文法的基本思想

Logo

- 选用自顶向下分析技术时，首先必须判别所给文法是否是LL(1)文法。
 - 具体做法是：依次计算其FIRST、FOLLOW、SELECT集，进而判别是否是LL(1)文法。
- 要判别一个文法是否是LL(1)文法，此文法必须首先是压缩的，即不含有害规则和多余规则：
- 例如：文法G7[S]为：
$$\begin{array}{llllll} S \rightarrow AB & S \rightarrow bC & A \rightarrow \epsilon & A \rightarrow b & B \rightarrow \epsilon \\ B \rightarrow aD & C \rightarrow AD & C \rightarrow b & D \rightarrow aS & D \rightarrow c \end{array}$$
- 此文法就是一个经过压缩的文法。

判别LL(1)文法的步骤1

Logo

- 步骤1. 求出所有能推出 ϵ 的非终结符
- 首先建立一个以文法的非终结符个数为上界的一维数组 $X[]$ ，其数组元素用于记录各非终结符能否推出 ϵ 。
- 其值有三种情况：“未定”、“是”、“否”。
- 例如对于文法G7[S]为：

$S \rightarrow AB$ $S \rightarrow bC$ $A \rightarrow \epsilon$ $A \rightarrow b$ $B \rightarrow \epsilon$
 $B \rightarrow aD$ $C \rightarrow AD$ $C \rightarrow b$ $D \rightarrow aS$ $D \rightarrow c$

- 对应的 $X[]$ 为：

非终结符	S	A	B	C	D
初值	未定	未定	未定	未定	未定
第一次扫描		是	是		否
第二次扫描	是			否	

判别LL(1)文法的步骤2

Logo

- 步骤2. 对每一文法符号 $X \in V$, 计算 $FIRST(X)$ 。有两种方法(另外一种自己看书)：
- 方法一：根据定义计算(P80)
- (a) 若 $X \in V_T$, 则 $FIRST(X) = \{X\}$ 。
- (b) 若 $X \in V_N$, 且有产生式 $X \rightarrow a...$, $a \in V_T$, 则 $a \in FIRST(X)$ 。
- (c) 若 $X \in V_N$, $X \rightarrow \epsilon$, 则 $\epsilon \in FIRST(X)$ 。
- (d) 若 $X \in V_N$, 且有产生式 $X \rightarrow Y_1 Y_2 \dots Y_n$, $Y_1, Y_2, \dots, Y_i \in V_N$; 当 $Y_1 Y_2 \dots Y_{i-1} \Rightarrow \epsilon$ 时, (其中 $1 \leq i \leq n$), 则 $FIRST(Y_1), FIRST(Y_2), \dots, FIRST(Y_{i-1})$ 的所有非 ϵ 元素和 $FIRST(Y_i)$ 都包含在 $FIRST(X)$ 中。
- (e) 当(d)中所有 $Y_i \Rightarrow \epsilon$, ($i = 1, 2, \dots, n$), 则 $FIRST(X) = FIRST(Y_1) \cup FIRST(Y_2) \cup \dots \cup FIRST(Y_n) \cup \{\epsilon\}$
- 反复使用上述(b)~(e)步直到每个符号的 $FIRST$ 集合不再增大为止。

例如有: $X \rightarrow Y_1 Y_2 Y_3$
 $Y_1 \rightarrow a Y_1 \mid \epsilon$
 $Y_2 \rightarrow b Y_2 \mid \epsilon$ $Y_3 \rightarrow c$
则: $FIRST(X) = \{a, b, c\}$



计算FIRST集举例

Logo

- 例如对于文法G7[S]:

$S \rightarrow AB \quad S \rightarrow bC \quad A \rightarrow \epsilon \quad A \rightarrow b \quad B \rightarrow \epsilon$
 $B \rightarrow aD \quad C \rightarrow AD \quad C \rightarrow b \quad D \rightarrow aS \quad D \rightarrow c$

- 其各终结符的FIRST集为:

- $FIRST(S) = FIRST(A) \cup FIRST(B) \cup \{\epsilon\} \cup \{b\}$

- $FIRST(A) = \{b, \epsilon\} \quad FIRST(B) = \{a, \epsilon\}$

- $FIRST(C) = (FIRST(A) - \{\epsilon\}) \cup FIRST(D) \cup \{b\}$

- $FIRST(D) = \{a, c\}$

- 注意: 可在此基础上求一个字符串的集合。

- $FIRST(AB) = FIRST(A) \cup FIRST(B) \cup \{\epsilon\} = \{a, b, \epsilon\}$

- $FIRST(bC) = \{b\} \quad FIRST(\epsilon) = \{\epsilon\} \quad FIRST(b) = \{b\}$

- $FIRST(AD) = (FIRST(A) - \{\epsilon\}) \cup FIRST(D)$

- $= \{a, b, c\}$



计

- 有文法:
- $E \rightarrow TE'$
- $E' \rightarrow +TE'$
- $E' \rightarrow \epsilon$
- $T \rightarrow FT'$
- $T' \rightarrow *FT'$
- $T' \rightarrow \epsilon$
- $F \rightarrow (E) | i$
- 求文法中串的**FIRST**集

解: 该文法的非终结符号有E、E'、T、T'和F。

$$\begin{aligned} \text{FIRST}(E) &= \text{FIRST}(TE') \\ &= \text{FIRST}(FT'E') = \{ (, i \} \end{aligned}$$

$$\text{FIRST}(+TE') = \{ + \}$$

$$\text{FIRST}(\epsilon) = \{ \epsilon \}$$

$$\begin{aligned} \text{FIRST}(E') &= \text{FIRST}(+TE') \\ &\cup \text{FIRST}(\epsilon) = \{ +, \epsilon \} \end{aligned}$$

$$\text{FIRST}(T) = \text{FIRST}(FT') = \{ (, i \}$$

$$\text{FIRST}(*FT') = \{ * \}$$

$$\begin{aligned} \text{FIRST}(T') &= \text{FIRST}(*FT') \\ &\cup \text{FIRST}(\epsilon) = \{ *, \epsilon \} \end{aligned}$$

$$\text{FIRST}((E)) = \{ (\}$$

$$\text{FIRST}(i) = \{ i \}$$

$$\text{FIRST}(F) = \text{FIRST}((E))$$

$$\cup \text{FIRST}(i) = \{ (, i \}$$

计算First集练习2

解:

- $S ::= aABbcd \mid \varepsilon$ $\text{FIRST}(S) = \text{FIRST}(aABbcd) \cup \text{FIRST}(\varepsilon)$

- $A ::= ASd \mid \varepsilon$ $= \{a\} \cup \{\varepsilon\} = \{a, \varepsilon\}$

- $B ::= SAh \mid eC$ $\text{FIRST}(A) = \text{FIRST}(ASd) \cup \text{FIRST}(\varepsilon)$

- $C ::= Sf \mid Cg \mid \varepsilon$ $= \{a, d\} \cup \{\varepsilon\} = \{a, d, \varepsilon\}$

- 求此文法 $\text{FIRST}(B) = \text{FIRST}(SAh) \cup \text{FIRST}(eC) \cup \text{FIRST}(\varepsilon)$

$$= \{a, d, h\} \cup \{e\} \cup \{\varepsilon\} = \{a, d, h, e, \varepsilon\}$$

$$\text{FIRST}(C) = \text{FIRST}(Sf) \cup \text{FIRST}(Cg) \cup \text{FIRST}(\varepsilon)$$

$$= \{a, f\} \cup \{a, f, g\} \cup \{\varepsilon\} = \{a, f, g, \varepsilon\}$$

判别LL(1)文法的步骤3

Logo

- 步骤3. 计算FOLLOW集。也有两种方法(另外一种自己看书)：

例如有： $S \Rightarrow \alpha_1 A \beta_1 \Rightarrow \alpha_1 \alpha B \beta \beta_1$

- 方法一：根据定义计算(P82)。
- 对文法中每一 $A \in V_N$,
- 计算FOLLOW(A)

$FOLLOW(A) = FIRST(\beta_1)$

若 $\beta \xRightarrow{*} \epsilon$, 则

$FOLLOW(B) = FIRST(\beta_1)$

- (a) 设S为文法中开始符号，把{#}加入FOLLOW(S)中(这里“#”的意义是输入串的结束符，也称句子括号)。
- (b) 对每一个以A为左部的产生式 $A \rightarrow \alpha B \beta$, 把FIRST(β)的非空元素加入FOLLOW(B)中。
- 另外，如果 $\beta \Rightarrow \epsilon$ 则把FOLLOW(A)也加入FOLLOW(B)。
- (c) 反复使用(b)直到每个非终结符的FOLLOW集不再增大为止。

例1 设文法G[S]:

$S ::= SbA | aA$

$A ::= Bc$

$B ::= Sb$

求此文法的每一个非终结符号的**FOLLOW**集。

0

解: 1. 因为S为文法的开始符号, 所以 $\# \in \text{FOLLOW}(S)$;
由 $S ::= SbA$, 有 $\text{FIRST}(bA) = \{b\} \in \text{FOLLOW}(S)$;
由 $B ::= Sb$, 有 $\text{FIRST}(b) = \{b\} \in \text{FOLLOW}(S)$;
因此, $\text{FOLLOW}(S) = \{b, \#\}$ 。

2. 由 $S ::= SbA$ 或 $S ::= aA$, 有 $\text{FOLLOW}(S) \in \text{FOLLOW}(A)$ 。因此, $\text{FOLLOW}(A) = \{b, \#\}$ 。

3. 由 $A ::= Bc$, 有 $\text{FIRST}(c) = \{c\} \in \text{FOLLOW}(B)$ 。因此, $\text{FOLLOW}(B) = \{c\}$ 。

计算FOLLOW集举例2

Logo

- 例如对于文法G7[S]:

$S \rightarrow AB$ $S \rightarrow bC$ $A \rightarrow \epsilon$ $A \rightarrow b$ $B \rightarrow \epsilon$
 $B \rightarrow aD$ $C \rightarrow AD$ $C \rightarrow b$ $D \rightarrow aS$ $D \rightarrow c$

- 问题: 如何着手计算FOLLOW集? 要着重考察产生式右部
- $FOLLOW(S) = \{\#\} \cup FOLLOW(D) = \{\#\}$
- $FOLLOW(D) = FOLLOW(B) \cup FOLLOW(C)$
- $FOLLOW(B) = FOLLOW(S)$
- $FOLLOW(C) = FOLLOW(S)$
- $FOLLOW(A) = (FIRST(B) - \{\epsilon\}) \cup FOLLOW(S) \cup FIRST(D) = \{a, \#, c\}$

(a) S为开始符号, 则 $\{\#\} \in FOLLOW(S)$

- (b) 对每一个以A为左部的产生式 $A \rightarrow \alpha B \beta$, 把 $FIRST(\beta)$ 的非空元素加入 $FOLLOW(B)$ 中。
- 另外, 如果 $\beta \Rightarrow^* \epsilon$ 则把 $FOLLOW(A)$ 也加入 $FOLLOW(B)$ 中。



计算Follow集练习

Logo

- 有文法 $E \rightarrow TE'$, $E' \rightarrow +TE'$, $E' \rightarrow \varepsilon$, $T \rightarrow FT'$, $T' \rightarrow *FT'$, $T' \rightarrow \varepsilon$, $F \rightarrow (E)|i$, 求各非终结符号的FOLLOW集。

解:

$$\text{FOLLOW}(E) = \{\#\} \cup \text{FIRST}() = \{\}, \quad \{\#\}$$

$$\text{FOLLOW}(E') = \text{FOLLOW}(E) = \{\}, \quad \{\#\}$$

$$\begin{aligned} \text{FOLLOW}(T) &= (\text{FIRST}(E') - \{\varepsilon\}) \cup \text{FOLLOW}(E) \\ &= \{+, \quad), \quad \#\} \end{aligned}$$

$$\text{FOLLOW}(T') = \text{FOLLOW}(T) = \{+, \quad), \quad \#\}$$

$$\begin{aligned} \text{FOLLOW}(F) &= (\text{FIRST}(T') - \{\varepsilon\}) \cup \text{FOLLOW}(T) \\ &= \{*, +, \quad), \quad \#\} \end{aligned}$$

判别LL(1)文法的步骤4

Logo

- 步骤4. 计算相同左部产生式的SELECT集，如果这些SELECT集的交集为空，则是LL(1)文法。

- 例如对于文法G7[S]:

$S \rightarrow AB$ $S \rightarrow bC$ $A \rightarrow \epsilon$ $A \rightarrow b$ $B \rightarrow \epsilon$ $B \rightarrow aD$ $C \rightarrow AD$ $C \rightarrow b$ $D \rightarrow aS$
 $D \rightarrow c$

- 其FIRST集和FOLLOW集计算结果如表:

① 若 $\alpha \xrightarrow{*} \epsilon$, 则
 $SELECT(A \rightarrow \alpha) = FIRST(\alpha)$ 。
 ② 若 $\alpha \xrightarrow{*} \epsilon$, 则
 $SELECT(A \rightarrow \alpha) = (FIRST(\alpha) - \{\epsilon\}) \cup FOLLOW(A)$ 。

非终结符	是否 $\Rightarrow \epsilon$	FIRST集	FOLLOW集
S	是	{b, a, ϵ }	{#}
A	是	{b, ϵ }	{a, c, #}
B	是	{a, ϵ }	{#}
C	否	{b, a, c}	{#}
D	否	{a, c}	{#}

- $SELECT(S \rightarrow AB) = (FIRST(AB) - \{\epsilon\}) \cup FOLLOW(S)$
 $= (((FIRST(A) - \{\epsilon\}) \cup FIRST(B)) - \{\epsilon\}) \cup FOLLOW(S)$
 $= \{b, a\} \cup \{#\} = \{b, a, \#\}$

判别LL(1)文法的步骤4(续)

Logo

- 步骤4. 计算SELECT集。

此文法不是LL(1)

- 例如对于文法G7[S]:

$S \rightarrow AB$ $S \rightarrow bC$ $A \rightarrow \epsilon$ $A \rightarrow b$ $B \rightarrow \epsilon$ $B \rightarrow aD$ $C \rightarrow AD$ $C \rightarrow b$
 $D \rightarrow aS$ $D \rightarrow c$

- 其FIRST集和FOLLOW集计算结果如表:

① 若 $\alpha \not\Rightarrow^* \epsilon$, 则
 $SELECT(A \rightarrow \alpha) = FIRST(\alpha)$ 。
② 若 $\alpha \Rightarrow^* \epsilon$, 则
 $SELECT(A \rightarrow \alpha) = (FIRST(\alpha) - \{\epsilon\}) \cup FOLLOW(A)$ 。

非终结符	是否 $\Rightarrow \epsilon$	FIRST集	FOLLOW集
S	是	{b, a, ϵ }	{#}
A	是	{b, ϵ }	{a, c, #}
B	是	{a, ϵ }	{#}
C	否	{b, a, c}	{#}
D	否	{a, c}	{#}

- $SELECT(S \rightarrow AB) = \{b, a, \#\}$
- $SELECT(S \rightarrow bC) = FIRST(bC) = \{b\}$
- $\therefore SELECT(S \rightarrow AB) \cap SELECT(S \rightarrow bC) \neq \emptyset$

写出下列文法中各候选式的**FIRST**集和各非终结符的**FOLLOW**集

- $S \rightarrow aA|BA$
- $A \rightarrow cB|\varepsilon$
- $B \rightarrow bB|\varepsilon$



作业

Logo

- 章节介绍完后一起布置