

编译原理

第四章 语法分析——自上而下分析

丁志军

dingzj@tongji.edu.cn

内 容 线索

1. 语法分析器的功能

2. 自上而下分析方法概述

3. LL (1) 分析方法

4. 递归下降分析程序

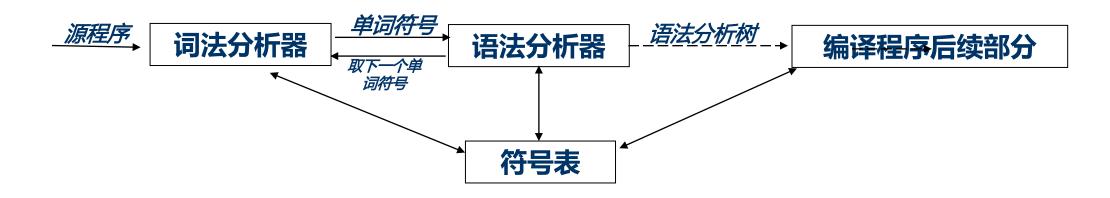
5. 预测分析程序

语法分析器

■ 语法分析的任务:

对任一给定 $w \in V_T^*$,判断 $w \in L(G)$?

■ 语法分析器:按照产生式规则,做识别w的工作



语法分析器在编译程序中的地位

语法分析方法

- 自上而下分析
- > 构建语法树
- > 下推自动机
- > LL (K) 分析法

从文法的开始符号出发,反复使用各种产生式,寻找与输入符号匹配的最左推导。

语法分析方法

- 自下而上分析
- > 移进-规约方法
- > CYK算法
- > early算法
- > 算符优先分析法
- > LR分析法

从输入符号串开始,逐步进行归约(最右推导的逆过程),直至归约到文法的开始符号。

内 容 线索

√. 语法分析器的功能

2. 自上而下分析方法概述

3. LL (1) 分析方法

4. 递归下降分析程序

5. 预测分析程序

自上而下分析

- 从文法的开始符号出发,向下推导,推出句子
- 对任何的输入串(单词符号),试图用一切可能的办法,从文法的开始符号出发,自上而下地为输入串建立一棵语法树,即为输入串寻找一个最左推导。

语法分析示例

G[S]:
$$S \rightarrow id := E = E \rightarrow E + E = E \rightarrow E * E$$

$$E \rightarrow -E = E \rightarrow (E) = E \rightarrow id$$

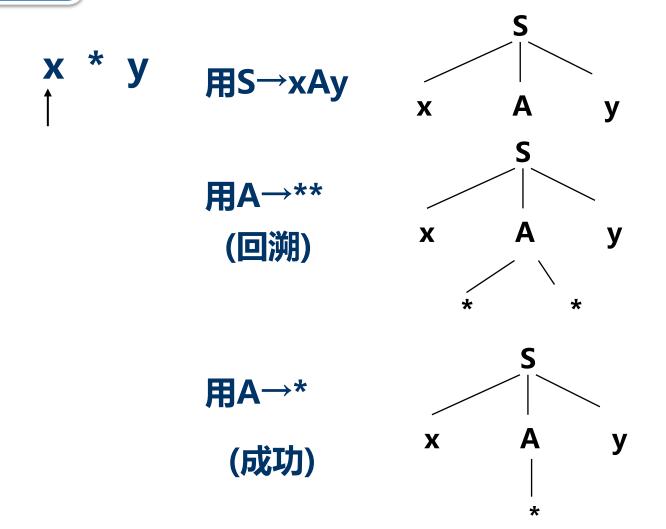
id:=-id 是否为G[S]的句子?

 $S \Rightarrow id := E \Rightarrow id := -id$

基本思想

最左推导过程中,为当前字符选择一个产生式,根据产生式右部首 字符进行匹配

设文法G[S]: S→xAy, A→** | * , 判定输入串 x * y是否为它的句子?



推导过程:

$$S \Rightarrow xAy$$

语法分析示例

$$G[S]: S \rightarrow id := E \to E + E = E \rightarrow E + E$$
 $E \rightarrow -E = E \rightarrow (E) = E \rightarrow id$
 $id := -id + id + id$ 是否为 $G[S]$ 的句子?
 $S \Rightarrow id := E \Rightarrow id := -E \Rightarrow id := -id$ (错误)
 $S \Rightarrow id := E \Rightarrow id := -E \Rightarrow id := -E + E$
 $\Rightarrow id := -E + E + E \Rightarrow id := -E + E + E \Rightarrow ...$

带回溯自上而下分析面临的问题

- 虚假匹配问题
- **■** 回溯
 - > 回溯会引起时间和空间的大量消耗
- 文法的左递归问题
 - → 一个文法是含有左递归的,如果存在非终结符P
 P → Pα
 - > 含有左递归的文法将使自上而下的分析过程陷入无限循环
- 报告分析不成功时,难于知道输入串中出错的确切位置。
- 实际上采用了一种穷尽一切可能的试探法,因此效率很低,代价很高

内 容 线索

√. 语法分析器的功能

√. 自上而下分析方法概述

3. LL (1) 分析方法

4. 递归下降分析程序

5. 预测分析程序

LL(1)分析法

- 从左(Left)到右扫描输入串;构造最左(Leftmost)推导;分析时每步向前看一个(1)字符。
- 目的: 构造不带回溯的自上而下分析算法
 - > 左递归的消除
 - > 消除回溯,提左因子
 - > FIRST集合, FOLLOW集合
 - > LL(1)分析条件
 - > LL(1)分析方法

左递归文法

- 一个文法含有下列形式的产生式时,
 - a)直接递归

$$A \rightarrow A\beta$$
 $A \in V_N$, $\beta \in V^*$

b)间接递归

$$A \rightarrow B\beta$$

$$B \rightarrow A\alpha$$
 $A, B \in V_N, \alpha, \beta \in V^*$

称为左递归文法。

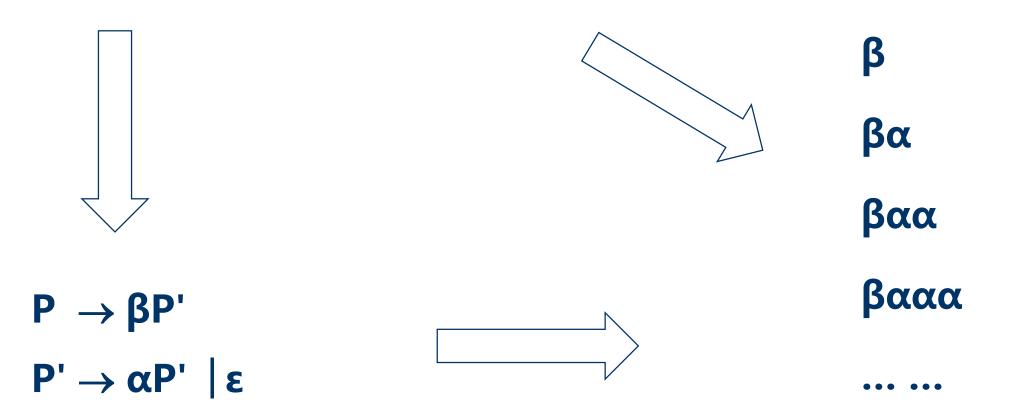
■ 如果一个文法是左递归时,则不能采用自顶向下分析法。

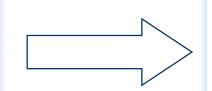
示例2

文法
$$P_1 \rightarrow aP_2$$
 $P_1 \rightarrow P_2b$ $P_2 \rightarrow P_1c$ $P_2 \rightarrow d$ 是间接左递归

消除直接左递归

$P \rightarrow P\alpha \mid \beta$ (α≠ε, β不以P开头)





$$E \rightarrow TE'$$
 $E' \rightarrow + TE' \mid \epsilon$
 $T \rightarrow FT'$
 $T' \rightarrow *FT' \mid \epsilon$
 $F \rightarrow (E) \mid i$

■ 一般地,假定关于P的产生式是

$$P \rightarrow P\alpha_1 \mid P\alpha_2 \mid ... \mid P\alpha_m \mid \beta_1 \mid \beta_2 \mid ... \mid \beta_n$$

其中:
$$α_i ≠ ε$$
, $β_i$ 不以P开头,

则改写为:
$$P \rightarrow \beta_1 P' \mid \beta_2 P' \mid \dots \mid \beta_n P'$$

$$P' \rightarrow \alpha_1 P' \mid \alpha_2 P' \mid ... \mid \alpha_m P' \mid \epsilon$$

消除左递归的算法思想

文法
$$P_2 \rightarrow P_1 b$$

$$P_3 \rightarrow P_2 c$$

$$P_1 \rightarrow P_3 d$$

$$P_1 \leftarrow P_2 \leftarrow P_3$$

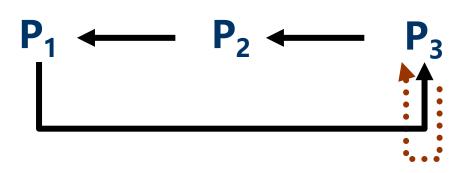
文法
$$P_1 \rightarrow P_3 d$$

$$P_2 \rightarrow P_1 b$$

$$P_3 \rightarrow P_2 c$$

$$\Rightarrow P_1 b c$$

$$\Rightarrow P_3 d b c$$



消除左递归算法

- (1) 排序: P₁、P₂、...、P_n
- (2) FOR i := 1 TO n DO BEGIN

FOR j:= 1 TO i - 1 DO

把形如 P_i \rightarrow P_i γ 的规则改写为:

 $P_i \rightarrow \delta_1 \gamma | \delta_2 \gamma | \dots | \delta_k \gamma$

其中: $P_j \rightarrow \delta_1 | \delta_2 | ... | \delta_k$ 是关于 P_j

的所有规则;

消除关于Pi规则的直接左递归。

END

(3) 化简: 删除永不使用的产生式

算法示意

文法
$$P_2 \rightarrow P_1 b$$

$$P_3 \rightarrow P_2 c$$

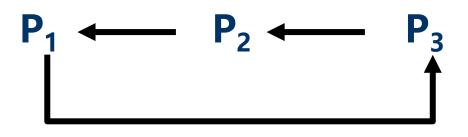
$$P_1 \rightarrow P_3 d$$
文法 $P_1 \rightarrow P_3 d$

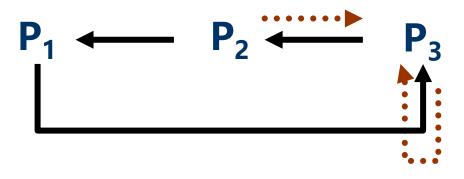
$$P_2 \rightarrow P_1 b$$

$$\Rightarrow P_3 db$$

$$P_3 \rightarrow P_2 c$$

$$\Rightarrow P_3 dbc$$





```
文法G[P<sub>3</sub>]: P_3 \rightarrow P_2 c | c \quad P_2 \rightarrow P_1 b | b \quad P_1 \rightarrow P_3 a | a
有推导: P_3 \Rightarrow P_2 c \Rightarrow P_1 bc \Rightarrow P_3 abc, 存在左递归。
按P<sub>1</sub>(1)、P<sub>2</sub>(2)、P<sub>3</sub>(3)排序,执行算法得:
i=1, j从1至0, P₁的产生式P₁→P₃a|a无直接左递归, 无需消除直接左递归。
i=2, j从1至1, P₂的候选式含P₁, P₁的产生式代入P₂的产生式得: P₂→P₃ab|ab|b,
   无直接左递归。
i=3, j从1至2:
   j=1, P₃的候选式不含P₁, 所以无需替换;
   j=2, P₃的候选式含P₂, 将P₂→P₃ab|ab|b代入S的候选式得:
                              P<sub>3</sub>→P<sub>3</sub>abc|abc|bc|c
   再消除直接左递归得:
                             P_3 \rightarrow abcP_3'|bcP_3'|cP_3'
                              P<sub>3</sub>'→abcP<sub>3</sub>'|ε
消除无用产生式: P<sub>2</sub>→P<sub>3</sub>ab|ab|b, P<sub>1</sub>→P<sub>3</sub>a|a,
                   P_3 \rightarrow abcP_3'|bcP_3'|cP_3'
得文法:
                          P<sub>3</sub>'→abcP<sub>3</sub>'|ε
    文法对应的正规式: V1=(abc|bc|c)(abc)*。
```

文法G[S]: S →Qc|c Q→Rb|b R→Sa|a

解: 1) 排序: S(1)、Q(2)、R(3)

2) 代入得: S →Q c c

 $Q \rightarrow R b \mid b$

R →Rbca | bca | ca | a

消除直接左递归:

$$S \rightarrow Q c | c$$
 $Q \rightarrow R b | b$
 $R \rightarrow bcaR' | caR' | aR'$
 $R' \rightarrow bcaR' | \epsilon$

消除隐含的左递归算法与非终极符排序方法无关

回溯问题

■ 例如, 有产生式:

语句→ if 条件 then 语句 else 语句

| while 条件 do 语句

| begin 语句表 end

若要寻找一个语句,那么关键字 if, while, begin就提示某个替换式是唯一的替换式。

无回溯!

回溯问题

■ 例如,有产生式:

语句→ if 条件 then 语句 else 语句

| if 条件 then 语句

| while 条件 do 语句

| begin 语句表 end

产生回溯!

回溯原因

若当前符号 = a, 下一步要展开A,

而 A $\rightarrow \alpha_1 |\alpha_2| ... |\alpha_n$,

怎样选择 α_i ?

- (1) 以a为头的α_i如果只有一个,则替换唯一;
- (2) 若以a为头有多个 α_i 的,则替换不唯一,可能需要回溯,这是文法的问题,应该变换文法。

文法的要求

- (1) 不含左递归
- (2) 对每个非终结符的候选式,其任何推导的头符号(终结符) 集合两两不相交。
 - 符号串α的终结首符集FIRST (α) 定义为:

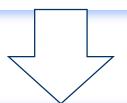
FIRST(α) = {
$$a \mid \alpha^* \Rightarrow a..., a \in V_T$$
 }
特别地,若α $\stackrel{*}{\Rightarrow} \epsilon$,则规定ε∈FIRST (α)。

以上条件(2)可表示为: 对文法的任一非终结符号A, 若 A→α₁ | α₂ | ... | α_n
 则应有 FIRST(α_i)∩FIRST(α_i) = Φ, i ≠ j

回溯解决方法

■ 提取公共左因子,将文法改造成任何非终结符的所有候选首符集 两两不相交。

$$A \rightarrow \delta \beta_1 | \delta \beta_2 | \dots | \delta \beta_n | \gamma_1 | \gamma_2 | \dots | \gamma_m$$
(其中 γ_1 、 γ_2 、 ...、 γ_m 不以δ开头)



$$A \rightarrow \delta A' \mid \gamma_1 \mid \gamma_2 \mid ... \mid \gamma_m$$

$$A' \rightarrow \beta_1 \mid \beta_2 \mid ... \mid \beta_n$$

文法G: S→aSb|aS|ε

解: 提取: S →aS(b|ε)

S→_E

引入新符: S →aSA ε

 $A \rightarrow b|\epsilon$

一般地,经过反复提取左因子可把每一个非终结符的所有候选首符集变成两两不相交。

LL (1) 分析条件

- 若文法已经消除了左递归,且对每个非终结符满足 FIRST(α_i)∩FIRST(α_i) = Φ
 - > 否就解决了虚假匹配和回溯的问题?
- 对某个输入符号a,及待匹配的非终结符
 - A $(A\rightarrow \alpha_1 | \alpha_2 | ... | \alpha_n)$
 - > a属于某个候选式的FIRST集合
 - a不属于任何候选式的FIRST集合,即对任意α_i, a ∉ FIRST (α_i)

 $G(S): S \rightarrow aA|d$

 $A \rightarrow bAS|d|\epsilon$

FIRST(S)= $\{a, d\}$ FIRST(A) = $\{b, d, \epsilon\}$

输入符号串abd是否为句子?

 $S \Rightarrow aA$

 \Rightarrow abAS

 \Rightarrow abdS

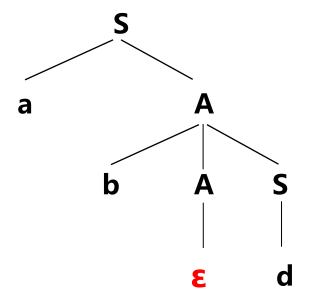
0 0 0

 $S \Rightarrow aA$

 \Rightarrow abAS

 \Rightarrow abS

 \Rightarrow abd



这是因为A有产生式A→ε,而从开始符号S可以得出 $S \stackrel{*}{\Rightarrow} ...$ Ad...

FOLLOW

■ 设S是文法G的开始符号,对G的任何非终结符A,定义A的后继终 结符号集为:

FOLLOW (A) = { a
$$\mid S \Rightarrow ... Aa..., a \in V_T$$
 }

■ 特别地,若S ⇒ ...A,则规定

 $\# \in FOLLOW(A)_{\circ}$

FOLLOW(A)是所有句型中出现在紧接A之后的终结符或"#"。

结论

- 当非终结符A面临输入符号a,且a \notin FIRST(α_i) (对任意i)时,如果A的某个候选首符集包含 ϵ (即 ϵ CFIRST (A)),那么,当a ϵ FOLLOW (A) 时,就允许A自动匹配(即选用A ϵ 工作),否则,认为a的出现是一种语法错误。
- 要正确进行不带回溯的语法分析,文法应满足的第三个条件可表示为: 若A的 候选首符集中包含ε,则

FIRST (A) \cap FOLLOW (A) = Φ

LL (1) 文法

- 如果文法G满足以下条件:
 - (1) 文法消除了左递归;
 - (2) 文法中每个非终结符A的各个产生式的候选首符集两两 不相交,即

(3) 对文法中的每个非终结符A, 若它存在某个候选首符集中包含ε, 则FIRST(A)∩ FOLLOW(A) = Φ,

则称该文法G为LL(1)文法。

LL (1) 分析方法

- 对一个LL (1) 文法,可以对某个输入串进行有效的无回溯的自上而下分析。
- 设面临的输入符号为a,要用非终结符A进行匹配,且 $A \to \alpha_1 \mid \alpha_2 \mid \mid \alpha_n$,则可如下分析:
 - (1) 若a \in FIRST (α_i),则指派 α_i 执行匹配任务;
 - (2) 否则
 - ① 若ε∈FIRST(A),且a∈FOLLOW (A),则让A 与ε自动匹配;
 - ②否则,a的出现是一种语法错误。

G(S):S → aA|d, A → bAS|d|ε是否为LL(1)文法?

解: (1) 不含左递归;

(2) 对于S, First(aA)={a} First(d)={d} 对于A, First(bAS)={b} First(d)={d}

First(ε)={ε}
 任一非终结符的候选首符集两两不相交 满足条件 (2)

(3) FIRST(S)={a, d} FIRST(A) ={b, d, ε}
Follow(S)={#, a, d} FIlow(A)={a, d, #}
所以 FIRST(A) ∩FIlow(A)={d} ≠Φ
不满足条件 (3)

所以,该文法不是LL(1)文法

如果文法为G(S):S \rightarrow aA|d, A \rightarrow bAS|ε是否为LL(1)文法?

解: (1) 不含左递归;

- (2) 对于S, First(aA)={a} First(d)={d}
 对于A, First(bAS)={b} First(ε)={ε}
 任一非终结符的候选首符集两两不相交 满足条件(2)
- (3) FIRST(S)={a, d} FIRST(A) ={b, ε}
 Follow(S))={#, a, d} Follow(A)={#, a, d}
 满足条件 (3)

所以,该文法是LL(1)文法

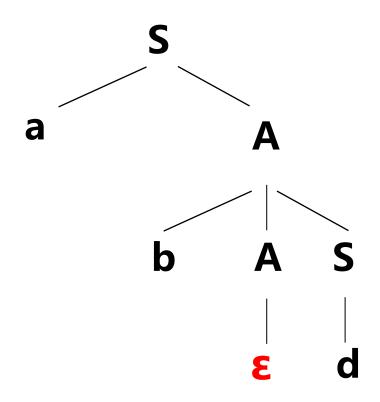
修改文法为G(S):S → aA|d, A → bAS|ε 输入符号串abd是否为句子?

 $S \Rightarrow aA$

 \Rightarrow abAS

 \Rightarrow abS

 \Rightarrow abd



FIRST (α) 构造

对于符号串 $\alpha = X_1X_2...X_n$,构造 FIRST (α)

- (1) 置 FIRST(α) = FIRST (X_1) {ε};
- (2) 若对 1<=j<= i -1, ε∈FIRST (X_j), 则把 FIRST(X_i) -{ε}加到FIRST(α)中;
- (3) 若对1<= j <=n, ε∈FIRST (X_j), 则把 ε加到FIRST(α)中。

FIRST (X) 构造

对于文法G的每个文法符号 $X \subset V_T \cup V_N$,构造FIRST (X) 的方法是:

- (1) 若X∈V_T, 则FIRST (X) ={X};
- (2) 若X \in V_N,且有产生式X \rightarrow a...,a \in V_T,则把a加入到FIRST(X)中,若有 X \rightarrow ε,则把ε加入FIRST(X);
 - (3) 若 $X \in V_N$, 且 $X \to Y \dots$, $Y \in V_N$, 则把

FIRST (Y) - {ε}加到FIRST (X)中,

若X →Y₁Y₂ ... Y_k , Y₁, Y₂, ... ,Y_{i-1} \in V_N, ε \in FIRST (Y_j),则把 (1<= j <= i -1)

FIRST (Y_i) - {ε}加到FIRST (X)中。

特别地,若ε∈FIRST (Y_i) (1<=j<= k),则 ε∈FIRST (X)

解

FIRST (E) = FIRST (T)
$$= FIRST (F)$$

$$= \{ (, i \}$$

$$FIRST (E') = \{ +, \epsilon \}$$

$$FIRST (T') = \{ *, \epsilon \}$$

解

$E \rightarrow TE'$; $E' \rightarrow +TE' \mid \varepsilon$; $T \rightarrow FT'$; $T' \rightarrow *FT' \mid \varepsilon$; $F \rightarrow (E) \mid i$

```
FIRST(E)
             FIRST(E')
            =\{+,\epsilon\}
FIRST(T)
             FIRST(T') = \{ *, \epsilon \}
FIRST(F)
            = { (, i }
FIRST(TE') = FIRST(T)-\{\epsilon\} \biguplus FIRST(E')-\{\epsilon\}
FIRST(+TE') = { + }
FIRST(FT') = FIRST(F)-\{\epsilon\} \sqcup F[RST(FT')-\{\epsilon\}
FIRST(*FT') = { * }
FIRST((E)) = \{ ( \}
```

FOLLOW(A)构造

对于文法G的每个非终结符,构造FOLLOW(A)的方法是:

- (1) 若A为文法开始符号,置#于FOLLOW(A) 中;
- (2) 若有产生式B→αAβ, 则将FIRST(β) - {ε}加到FOLLOW (A)中;
- (3) 若有B→ α A或B → α A β , 且 $\beta \stackrel{*}{\Rightarrow} \epsilon$ 则将FOLLOW(B)加到FOLLOW(A)中;
- (4) 反复使用以上规则, 直至 FOLLOW(A)不再增大为止。

解

= {*, +, #,) }

```
E \rightarrow TE'; E' \rightarrow +TE' \mid \varepsilon; T \rightarrow FT'; T' \rightarrow *FT' \mid \varepsilon; F \rightarrow (E) \mid i
                   FIRST (E) = FIRST (T) = FIRST (F) = \{(, i)\}
                   FIRST (E') = \{+, \epsilon\}
                   FIRST (T') = \{*, \epsilon\}
FOLLOW(E)
                       = \{ \# \} \cup \{ \} \} = \{ \# , \} \}
FOLLOW(E')
                        = FOLLOW(E)-\{\varepsilon\} \cup FOLLOW(E')-\{\varepsilon\} = \{\#, \}
FOLLOW(T)
                        = FIRST(E')-\{\epsilon\} \cup FOLLOW(E)-\{\epsilon\} \cup FOLLOW(E')-\{\epsilon\} = \{+, \#, \}
FOLLOW(T')
                        = FOLLOW(T)-\{\varepsilon\} \cup FOLLOW(T')-\{\varepsilon\} = \{+, \#, \}
FOLLOW(F)
                        = FIRST(T')-\{\epsilon\} \cup FOLLOW(T)-\{\epsilon\}\cup FOLLOW(T')-\{\epsilon\}
```