2025年春季学期《编译工程》



语法分析 自顶向下-LL(1)文法

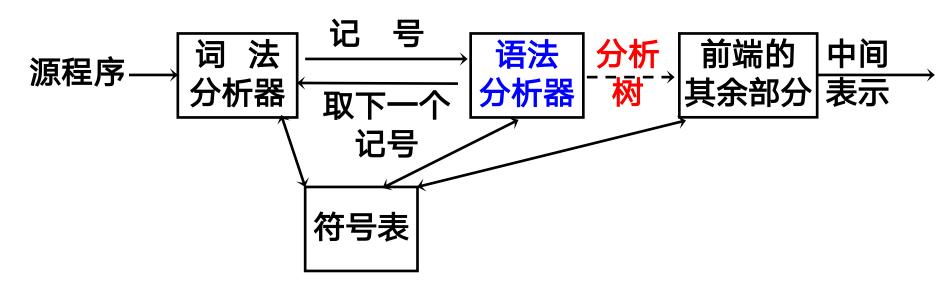
李 诚

国家高性能计算中心(合肥)、信息与计算机国家级实验教学示范中心 计算机科学与技术学院

2025年03月06日

❷ 本节提纲





• 自顶向下分析方法

- LL(1)文法
- 非递归预测分析方法



预测分析法(Predictive parsing)



• 与递归下降法相似,但

- 不会对若干产生式进行尝试
- 没有回溯
- 通过向前看一些记号来预测需要用到的产生式

·此方法接受LL(k)文法

- L-means "left-to-right" scan of input
- L-means "leftmost derivation"
- k-means "predict based on k tokens of lookahead"
- In practice, LL(1) is used



• 对文法加什么样的限制可以保证没有回溯?

- 先定义两个和文法有关的函数
 - FIRST(a) = { $a \mid a \Rightarrow * a..., a \in V_T$ } 意义: 可从a推导得到的串的首符号的集合
 - FOLLOW(A) = { $a \mid S \Rightarrow * ... Aa..., a \in V_T$ } 意义:可能在推导过程中紧跟在A右边的终结符号的集合



□ LL(1)文法: FIRST(X)



- 计算FIRST(X), $X \in V_T \cup V_N$
 - $X \in V_T$, FIRST(X) = {X}
 - $X \in V_N \perp \!\!\! \perp X \rightarrow \epsilon$ 则将 ε加入到FIRST(X)
 - $X \in V_N \perp X \rightarrow Y_1 Y_2 \dots Y_k$
 - 如果 $a \in FIRST(Y_i)$ 且 ϵ 在 $FIRST(Y_i)$, ..., $FIRST(Y_{i-1})$ 中,则将 a加入到 FIRST(X)
 - 如果 ε 在FIRST(Y₁), ..., FIRST(Y_k)中,则将ε 加入到FIRST(X)

FIRST集合只包括终结符和E



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow * FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

$$FIRST(F) = \{ (, id) \}$$

 $\Box X \in V_{\underline{T}}, FIRST(X) = \{X\}$ $\Box X \in V_{\underline{N}} \boxtimes X \to \varepsilon, \varepsilon \in FIRST(X)$ $\Box X \in V_{\underline{N}} \boxtimes X \to Y_{\underline{1}} Y_{\underline{2}} ... Y_{\underline{k}}$ $\Leftrightarrow \underline{w} \not = a \in FIRST(Y_{\underline{i}}) \underline{\mathbb{E}} \varepsilon FIRST(Y_{\underline{1}}), ..., \\
\underline{FIRST(Y_{\underline{i-1}})} \not = \underline{M} a \in FIRST(X)$ $\Leftrightarrow \underline{w} \not = \varepsilon FIRST(Y_{\underline{1}}), ..., FIRST(Y_{\underline{k}}) \not = \underline{N} \varepsilon \in FIRST(X)$





• 例
$$E \rightarrow TE'$$
 $E' \rightarrow + TE' \mid \varepsilon$
 $T \rightarrow FT'$
 $T' \rightarrow *FT' \mid \varepsilon$
 $F \rightarrow (E) \mid id$

$$FIRST(F) = \{ (, id) = FIRST(T) = FIRST(E) \}$$



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow * FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

FIRST(
$$F$$
) = { (, id } = FIRST(T) = FIRST(E)
FIRST(E ') = {+, ε }



• 例
$$E \rightarrow TE'$$
 $E' \rightarrow + TE' \mid \varepsilon$
 $T \rightarrow FT'$
 $T' \rightarrow *FT' \mid \varepsilon$
 $F \rightarrow (E) \mid id$

FIRST(
$$F$$
) = { (, id } = FIRST(T) = FIRST(E)
FIRST(E ') = {+, ε }
FRIST(T ') = {*, ε }



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow * FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id }
FIRST(E ') = {+, ε }
FRIST(T ') = {*, ε }



LL(1)文法: FOLLOW(A)



- 计算FOLLOW(A), $A \in V_N$
 - \$加入到FOLLOW(A), 当A是开始符号, \$是输入串的结束符号
 - 如果 $A \rightarrow a B\beta$, 则FIRST(β)-{ ϵ }加入到FOLLOW(B)
 - 如果 $A \rightarrow a \ B$ 或 $A \rightarrow a \ B\beta$ 且 $\epsilon \in FIRST(\beta)$,则FOLLOW(A)加入到FOLLOW(B)



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

- □ <u>当A是开始符号</u>, **§** ∈ FOLLOW(A)
- $\square A \rightarrow a B\beta$, FIRST(β)-{ ϵ } \subseteq FOLLOW(B)
- $\Box A \rightarrow a B 或 A \rightarrow a B \beta 且 \epsilon \in FIRST(\beta),$ $FOLLOW(A) \subseteq FOLLOW(B)$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id }
FIRST(E ') = {+, ϵ }
FRIST(T ') = {*, ϵ }
FOLLOW(E) = {), \$}



• 例
$$E \rightarrow TE'$$
 $E' \rightarrow + TE' \mid \varepsilon$
 $T \rightarrow FT'$
 $T' \rightarrow * FT' \mid \varepsilon$
 $F \rightarrow (E) \mid id$

```
□当A是开始符号,§ ∈ FOLLOW(A)
```

$$\square A \rightarrow a B\beta$$
, FIRST(β)-{ ϵ } \subseteq FOLLOW(B)

 $\Box \underline{A \rightarrow a \ B} \ \vec{\boxtimes} A \rightarrow \underline{a \ B\beta} \ \vec{\sqsubseteq} \ \in \ FIRST(\underline{\beta}),$ $FOLLOW(A) \subseteq \ FOLLOW(B)$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id }
FIRST(E ') = {+, ϵ }
FRIST(T ') = {*, ϵ }
FOLLOW(E) = {), \$} = FOLLOW(E ')



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

- □当A是开始符号,§ ∈ FOLLOW(A)
- $\square A \rightarrow a B\beta$, FIRST(β)-{ ϵ } \subseteq FOLLOW(B)
- $\Box \underline{A \rightarrow a \ B} \ \vec{\boxtimes} A \rightarrow a \ \underline{B\beta} \ \vec{\boxtimes} \ \in FIRST(\underline{\beta}),$ $FOLLOW(A) \subseteq FOLLOW(B)$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id }
FIRST(E ') = {+, ϵ }
FRIST(T ') = {*, ϵ }
FOLLOW(E) = {), \$} = FOLLOW(E ')
FOLLOW(T) = {+,), \$}



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow * FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

- □当A是开始符号,§ ∈ FOLLOW(A)
- $\square A \rightarrow a B\beta$, FIRST(β)-{ ϵ } \subseteq FOLLOW(B)
- $\Box \underline{A \rightarrow a \ B} \ \vec{\boxtimes} A \rightarrow a \ \underline{B\beta} \ \underline{\exists} \ \epsilon \in FIRST(\underline{\beta}),$ $FOLLOW(A) \subseteq FOLLOW(B)$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id }
FIRST(E ') = {+, ε }
FRIST(T ') = {*, ε }
FOLLOW(E) = {), \$} = FOLLOW(E ')
FOLLOW(T) = {+,), \$} = FOLLOW (T ')



• 例
$$E \rightarrow TE'$$

$$E' \rightarrow + TE' \mid \varepsilon$$

$$T \rightarrow FT'$$

$$T' \rightarrow *FT' \mid \varepsilon$$

$$F \rightarrow (E) \mid id$$

- □当A是开始符号,§ ∈ FOLLOW(A)
- $\square A \rightarrow a B\beta$, FIRST(β)-{ ϵ } \subseteq FOLLOW(B)
- $\Box \underline{A \rightarrow a \ B} \ \vec{\boxtimes} A \rightarrow \underline{a \ B\beta} \ \vec{\sqsubseteq} \ \in FIRST(\underline{\beta}),$ $FOLLOW(A) \subseteq FOLLOW(B)$

FIRST(
$$E$$
) = FIRST(T) = FIRST(F) = { (, id } FIRST(E ') = {+, ϵ }

FRIST(T ') = {*, ϵ }

FOLLOW(E) = {), \$} = FOLLOW(E ')

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

FOLLOW(T) = {*,+,), \$}



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset

② LL(1)文法



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset

• 该条件存在的必要性

- 容易理解
- · 每次通过输入词法单元记号和FIRST集合匹配产生式的时候,需要有唯一的选择



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset
- 若 $\beta \Rightarrow * \epsilon$, 那么 $FIRST(\alpha) \cap FOLLOW(A) = \emptyset$
- □假设FIRST(a) ∩ FOLLOW(A) = {a} $a \in \text{FIRST}(a)$: $A \Rightarrow *aa'$ $a \in \text{FOLLOW}(A)$: $B \Rightarrow *... A a...$

② LL(1)文法



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset
- 若 $\beta \Rightarrow * \epsilon$, 那么 $FIRST(\alpha) \cap FOLLOW(A) = \emptyset$
- □假设FIRST(a) \cap FOLLOW(A) = {a}

 $a \in FIRST(a)$: $A \Rightarrow *aa'$

 $a \in \text{FOLLOW}(A)$: $B \Rightarrow^* \dots A a \dots$

由于 $\beta \Rightarrow * \epsilon$, 所以遇到a时, 无法判断用哪一个产生式

- 可以用 $A \rightarrow a$ 来对A进行展开
- 亦可以用 $A \rightarrow \beta$ 和 $\beta \Rightarrow * ε$ 最后把A消掉

G LL(1)文法



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset

□例如,考虑下面文法

面临a...时, 第2步推导不知用哪个产生式

$$S \rightarrow A B$$

$$A \to a \ b \mid \varepsilon$$
 $a \in FIRST(ab) \cap FOLLOW(A)$

$$B \rightarrow a C$$

$$C \rightarrow \dots$$

② LL(1)文法



• LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha \mid \beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset

• LL(1)文法有一些明显的性质

- 没有公共左因子
- 不是二义的
- 不含左递归

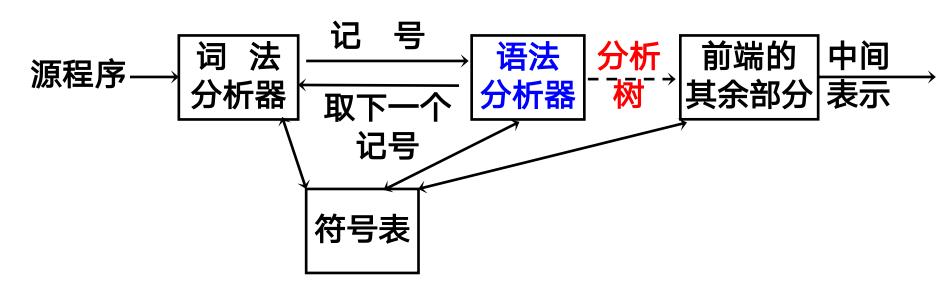


```
• 例 E \rightarrow TE'
E' \rightarrow + TE' \mid \varepsilon
T \rightarrow FT'
T' \rightarrow * FT' \mid \varepsilon
F \rightarrow (E) \mid id
```

```
FIRST(E) = FIRST(T) = FIRST(F) = { ( , id } FIRST(E') = {+, \epsilon}
FRIST(T') = {*, \epsilon}
FOLLOW(E) = FOLLOW(E') = { ), $}
FOLLOW(T) = FOLLOW (T') = {+, ), $}
FOLLOW(T) = {+, *, ), $}
```

❷ 本节提纲





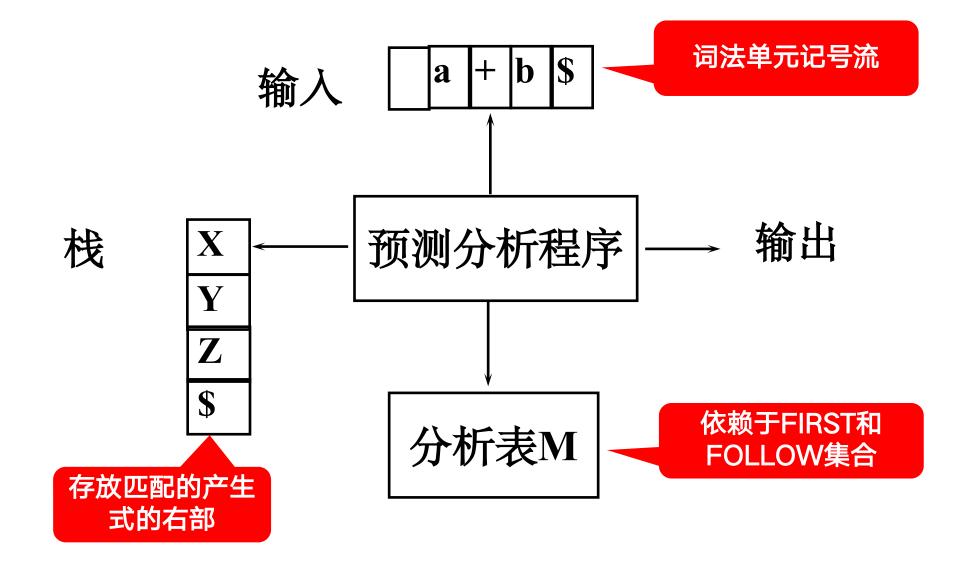
• 自顶向下分析方法

- LL(1)文法
- 非递归预测分析方法



非递归的预测分析







预测分析表M的构造



•行: 非终结符; 列: 终结符或\$; 单元: 产生式

非终	输入符号					
结符	id	+	*	()	\$
E	$E \rightarrow$			$E \rightarrow$		
	TE'			TE'		
<i>E</i> '		$E' \rightarrow$			$E' \rightarrow \varepsilon$	$E' \rightarrow \varepsilon$
		+ <i>TE</i> '				
T	$T \rightarrow$			$T \rightarrow$		
	FT'			FT'		
<i>T</i> '		$T' \rightarrow \varepsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \varepsilon$	$T' \rightarrow \varepsilon$
F	$F \rightarrow id$			$F \rightarrow (E)$		



预测分析表M的构造



- •对文法的每个产生式 $A \rightarrow a$,执行(1)和(2)
 - (1) $\forall FIRST(a)$ 的每个终结符a, $\forall A \rightarrow a m \land M[A, a]$
 - (2) 如果 ϵ 在FIRST(a)中,对FOLLOW(A)的每个终结符b (包括\$), 把 $A \rightarrow am \lambda M[A,b]$

M中其它没有定义的条目都是error





栈	输入	输出
\$ <i>E</i>	id * id + id\$	





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$
\$ <i>E</i> ' <i>T</i> ' id	id * id + id\$	$F \rightarrow id$





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$
\$ <i>E</i> ' <i>T</i> ' id	id * id + id\$	$F \rightarrow id$
\$E'T'	* id + id\$	匹配id





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$
\$E 'T' id	id * id + id\$	$F \rightarrow id$
\$E'T'	* id + id\$	
\$E'T'F*	* id + id\$	$T' \rightarrow *FT'$





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$
\$E 'T' id	id * id + id\$	$F \rightarrow id$
\$E 'T'	* id + id\$	
\$E 'T 'F *	* id + id\$	$T' \rightarrow *FT'$
\$E 'T 'F	id + id\$	





栈	输入	输出
\$ <i>E</i>	id * id + id\$	
\$E 'T	id * id + id\$	$E \rightarrow TE'$
\$E 'T 'F	id * id + id\$	$T \rightarrow FT'$
\$E 'T' id	id * id + id\$	$F \rightarrow id$
\$E 'T'	* id + id\$	
\$E 'T'F*	* id + id\$	$T' \rightarrow *FT'$
\$E 'T 'F	id + id\$	
\$ <i>E</i> ' <i>T</i> ' id	id + id\$	$F \rightarrow id$





栈	输入	输出
\$ <i>E</i> ' <i>T</i> ' id	id + id\$	$F \rightarrow id$
\$E 'T'	+ id\$	匹配id
\$E'	+ id\$	$T' \rightarrow \varepsilon$
\$E 'T+	+ id\$	$E' \rightarrow +TE'$
\$E 'T+	id\$	匹配+
\$E 'T 'F	id\$	$T \rightarrow FT'$
\$E 'T'id	id\$	$F \rightarrow id$
\$E 'T'	\$	匹配id





预测分析器接受输入id * id + id的所有动作

栈	输入	输出
\$E'T'	\$	$T' \rightarrow \varepsilon$
\$E'	\$	$E' \rightarrow \varepsilon$
\$	\$	Finished





 $e_part \rightarrow else \ stmt \mid \epsilon \quad expr \rightarrow b$

非终	输	入	符号	
结符	other	b	else	• • •
stmt	$stmt \rightarrow other$			
e_part			$\begin{array}{c} e_part \rightarrow \\ \text{else } stmt \\ e_part \rightarrow \varepsilon \end{array}$	
expr		$expr \rightarrow b$		

多重定义条目意味着文法左递归或者是二义的



多重定义的消除



例: 删去 $e_part \rightarrow \epsilon$, 这正好满足else和近的then配对

LL(1)文法: 预测分析表无多重定义的条目

非终	输	入	符号	
结符	other	b	else	• • •
stmt	$stmt \rightarrow other$			
e_part			$\begin{array}{c} e_part \rightarrow \\ else\ stmt \\ e_part \rightarrow \epsilon \end{array}$	
expr		$expr \rightarrow b$		

2025年春季学期《编译工程》



一起努力 打造国产基础软硬件体系!

李 诚

国家高性能计算中心(合肥)、信息与计算机国家级实验教学示范中心 计算机科学与技术学院

2025年03月06日