

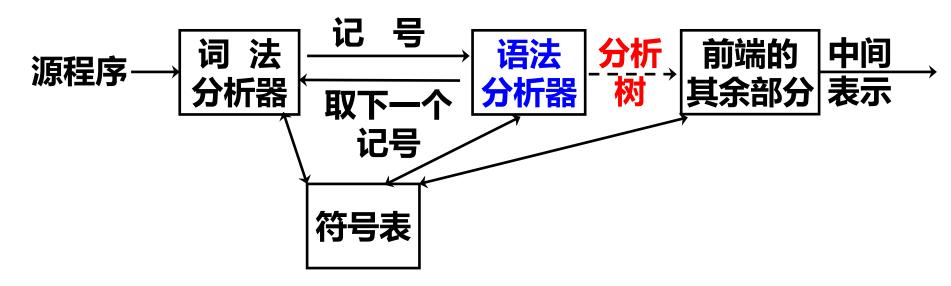
第4讲 语法分析-自顶向下-LL(1)文法

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 2023年09月18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(α) = { $a \mid \alpha \Rightarrow^* a..., a \in V_T$ } 意义:可从 α 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 & 加入到FIRST(X)
 - $X \in V_N \perp \!\!\! \perp X \rightarrow Y_1 Y_2 \dots Y_k$
 - •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
$$E \rightarrow TE'$$
 $E' \rightarrow + TE' \mid \varepsilon$
 $T \rightarrow FT'$
 $T' \rightarrow *FT' \mid \varepsilon$
 $F \rightarrow (E) \mid id$

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

- $\square X \in V_T$, FIRST(X) = {X}
- $\square X \in V_N \sqsubseteq X \to \varepsilon, \varepsilon \in FIRST(X)$
- $\Box X \in V_N \blacksquare X \to Y_1 Y_2 \dots Y_k$
 - **◇** 如果 $a \in FIRST(Y_i)$ 且 £ 在 $FIRST(Y_i)$, ..., $FIRST(Y_{i-1})$ 中,则 $a \in FIRST(X)$
 - **◇**如果 ε 在FIRST(Y_I), ..., FIRST(Y_k)中,则 ε ∈ 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
$$E o TE'$$
 $E' o + TE' \mid \varepsilon$
 $T o FT'$
 $T' o *FT' \mid \varepsilon$
 $F o (E) \mid \mathrm{id}$

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 $A \rightarrow \alpha B\beta$, 则FIRST(β)-{ ϵ }加入到FOLLOW(B)
 - 如果 $A \rightarrow \alpha B$ 或 $A \rightarrow \alpha B\beta$ 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)
- $\square A \rightarrow \alpha B\beta$, FIRST(β)-{ ε } \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)
```

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

 $\Box \underline{A} \rightarrow \alpha \underline{B} \ \overline{\mathbf{g}} A \rightarrow \alpha \underline{B} \beta \underline{\mathbf{g}} \mathbf{\epsilon} \in \mathrm{FIRST}(\underline{\beta}),$ $\overline{\mathrm{FOLLOW}(A)} \subseteq \overline{\mathrm{FOLLOW}(B)}$

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



• 例
$$E o TE'$$
 $E' o + TE' \mid \varepsilon$
 $T o FT'$
 $T' o *FT' \mid \varepsilon$
 $F o (E) \mid \mathrm{id}$

- □当A是开始符号, \$ ∈ FOLLOW(A)
- $\square A \rightarrow \alpha B\beta$, FIRST(β)-{ ϵ } \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)
- $\square A \rightarrow \alpha B\beta$, FIRST(β)-{ ϵ } \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
$$E o TE'$$
 $E' o + TE' \mid \varepsilon$
 $T o FT'$
 $T' o *FT' \mid \varepsilon$
 $F o (E) \mid \mathrm{id}$

- □当A是开始符号, \$ ∈ FOLLOW(A)
- $\square A \rightarrow \alpha B\beta$, FIRST(β)-{ ϵ } \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 $A \rightarrow \alpha/\beta$ 都满足下列条件:

- FIRST(α) \cap FIRST(β) = \emptyset



·LL(1)文法的定义

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

- FIRST(α) \cap FIRST(β) = \emptyset
- $□ 假设FIRST(\alpha) \cap FOLLOW(A) = \{a\}$

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

 $a \in FOLLOW(A)$: $B \Rightarrow^* ... A a...$

② LL(1)文法



·LL(1)文法的定义

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

- FIRST(α) \cap FIRST(β) = \emptyset
- 口假设 $FIRST(\alpha) \cap FOLLOW(A) = \{a\}$

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

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

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 $A \rightarrow \alpha$ 来对A进行展开
- 亦可以用 $A \rightarrow \beta \pi \beta \Rightarrow * \epsilon$ 最后把A消掉



·LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha/\beta$ 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

$$S \rightarrow A B$$

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

$$B \rightarrow a C$$

$$C \rightarrow \dots$$



·LL(1)文法的定义

任何两个产生式 $A \rightarrow \alpha/\beta$ 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

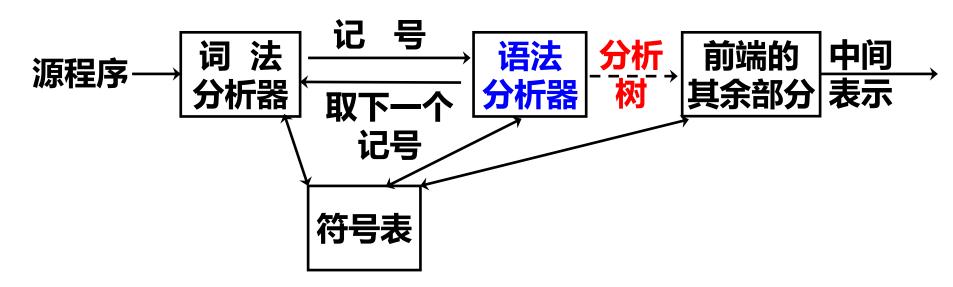


```
• 例 E \rightarrow TE'
E' \rightarrow + TE' \mid \varepsilon
T \rightarrow FT'
T' \rightarrow *FT' \mid \varepsilon
F \rightarrow (E) \mid id
```

```
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





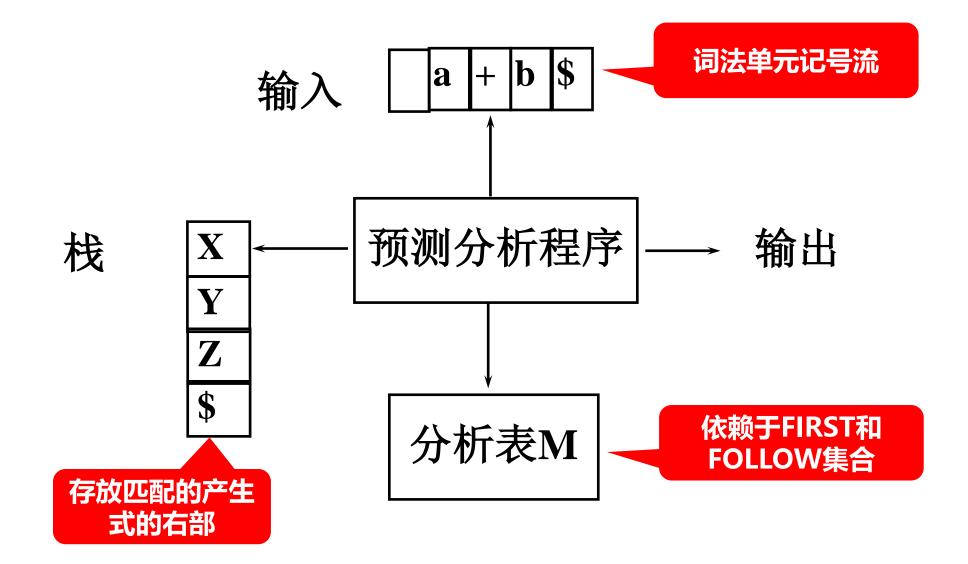
・自顶向下分析方法

- 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



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

M中其它没有定义的条目都是error



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>		$E' \rightarrow$			$E' \to \varepsilon$	$E' o \epsilon$
		+ TE '				
T	$T \rightarrow$			$T \rightarrow$		
	FT'			FT'		
T'		$T' \rightarrow \varepsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \varepsilon$	$T' \rightarrow \varepsilon$
F	$F \rightarrow id$			$F \rightarrow (E)$		





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\$	$T \rightarrow FT'$
\$E 'T' id	id * id + id\$	$F \rightarrow id$





栈	输入	输出
\$E	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\$	匹配id





栈	输入	输出
\$E	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\$	$F \rightarrow \mathrm{id}$
\$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\$	
\$E 'T' id	id + id\$	$F \rightarrow \mathrm{id}$





栈	输入	输出
\$E 'T' 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\$	$F \rightarrow \mathrm{id}$
\$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





例: $stmt \rightarrow if expr then stmt e_part$ other

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

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



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 2023年09月18日