Ch5 语法分析

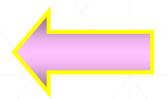
- 5.1 "移近—归约"分析法
- 5.3 LR分析鸟瞰
- 5.4 LR (0)分析
- 5.5 SLR (1)分析
- 5.6 LR (1)分析
- 5.7 LALR (1)分析
- 5.8 LR分析对二义文法的应用
- 5.9 LR分析的错误处理与恢复
- 5.10 语法分析器的自动生成与YACC(自学)





5.7 LALR (1)分析

5.7.1 LALR(1)分析实现思想



5.7.2 LALR(1)分析表的构造

分析功能:

 $LR(1) \supset SLR(1) \supset LR(0)$

存储开销:

 $LR(1) \uparrow SLR(1) LR(0) \downarrow$



存储开销 ⇒ LR(0)、SLR(1)

分析功能 ⇒ LR(1)





LALR(1) Look -Ahead LR(1)

LALR(1)分析表状态数 = LR(0)的状态数

LALR(1)分析功能 ⊆ LR(1)



例: 设有文法G(S)为

$$S \rightarrow L = R \mid R$$

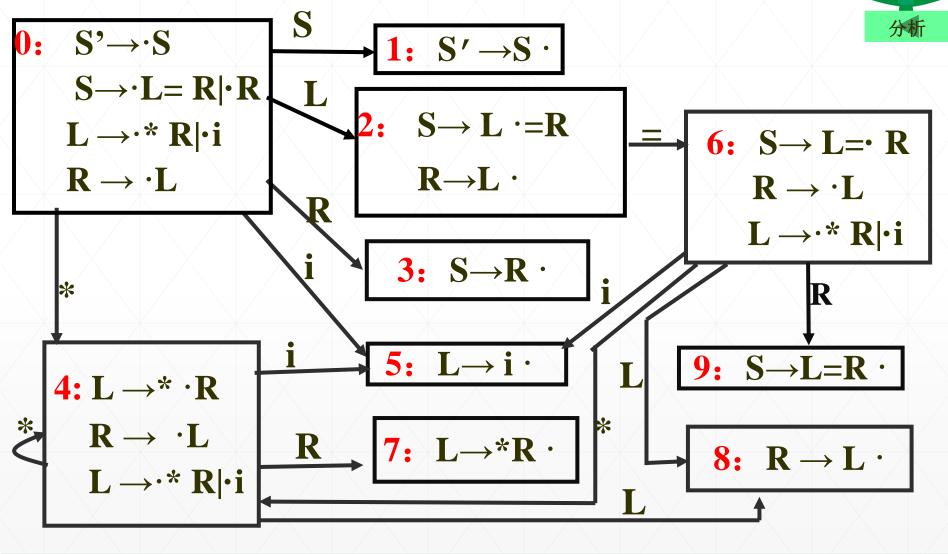
$$L \rightarrow *R \mid i$$

$$R \to L$$

 $S' \rightarrow S$

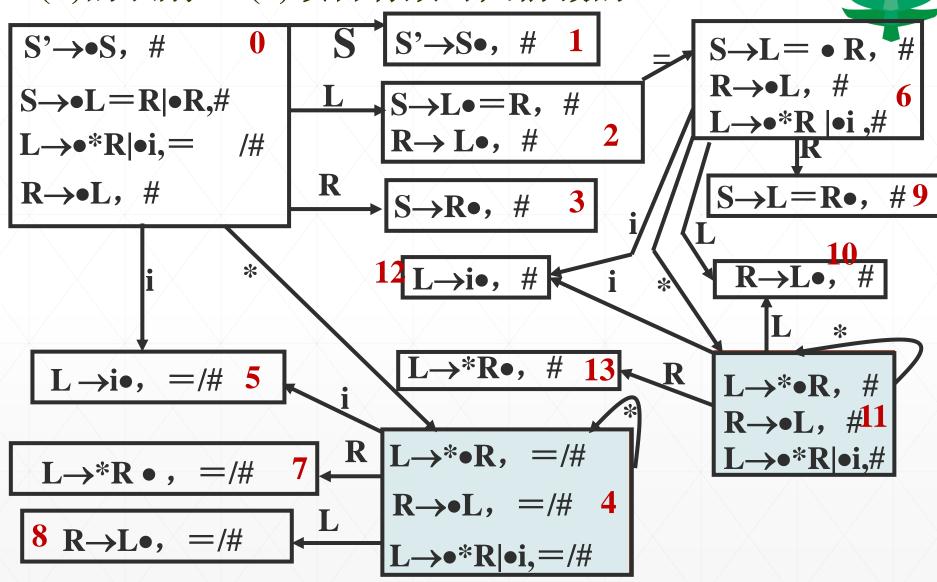
G(S)的识别LR(0)项目有效可归前缀的DFA:

 $S \rightarrow L = R \ \ \bigcirc \mid R \ \bigcirc \mid$

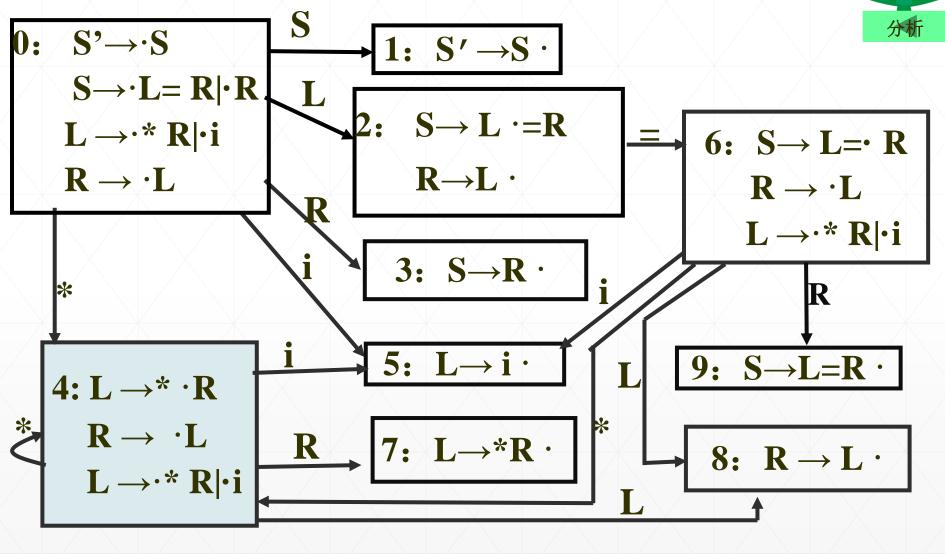


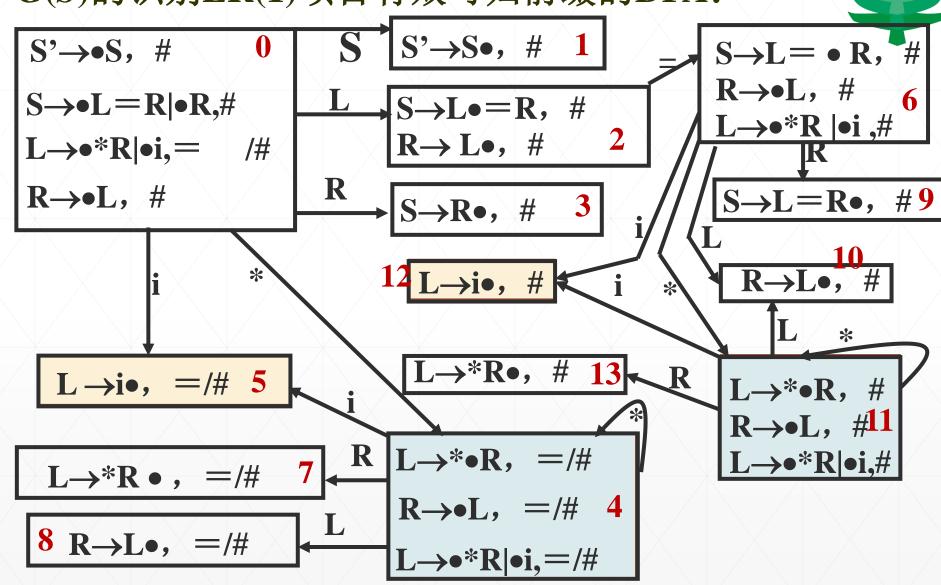
 $L \rightarrow *R \otimes |i \otimes 4|$





G(S)的识别LR(0)项目有效可归前缀的DFA:



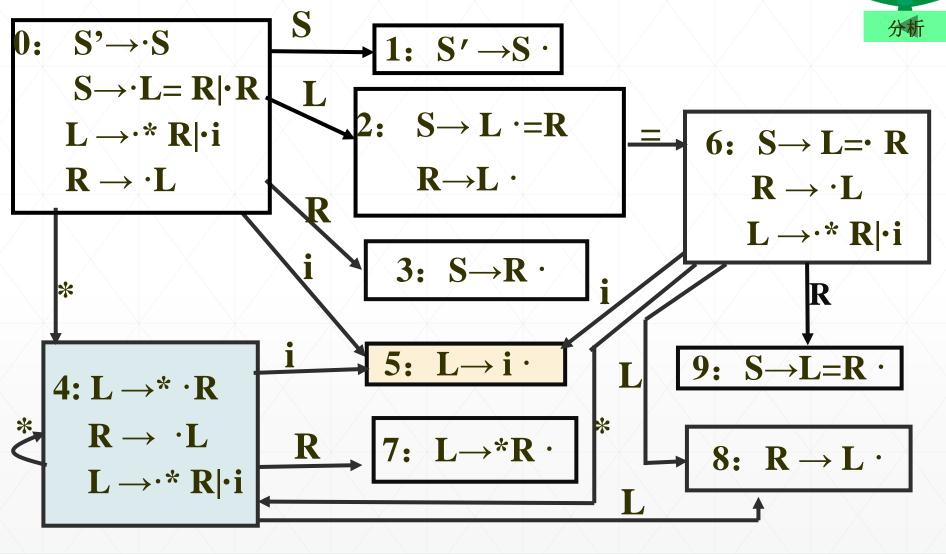


 $R \rightarrow L(5)$

第 10 页

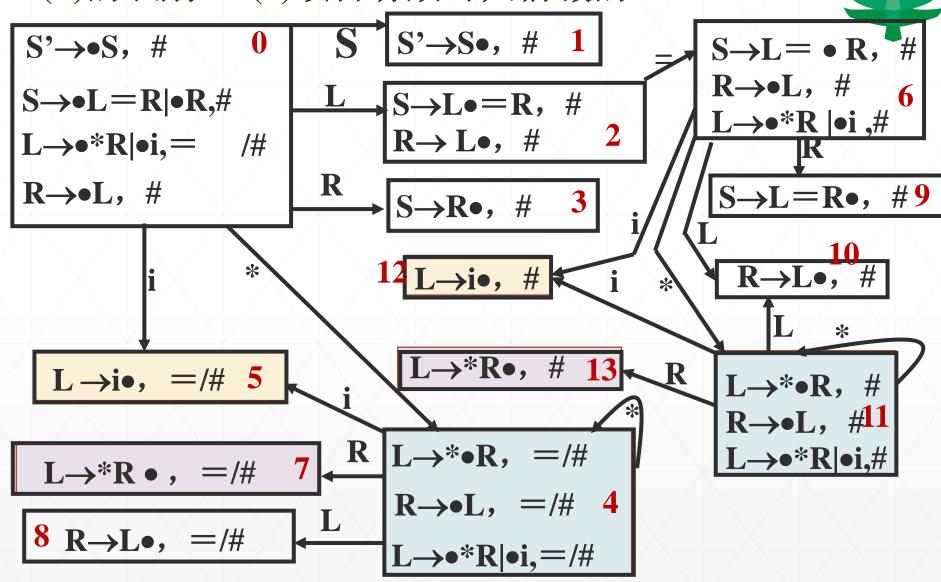
G(S)的识别LR(0)项目有效可归前缀的DFA:

 $S \rightarrow L = R \ \ \bigcirc \mid R \ \bigcirc \mid$

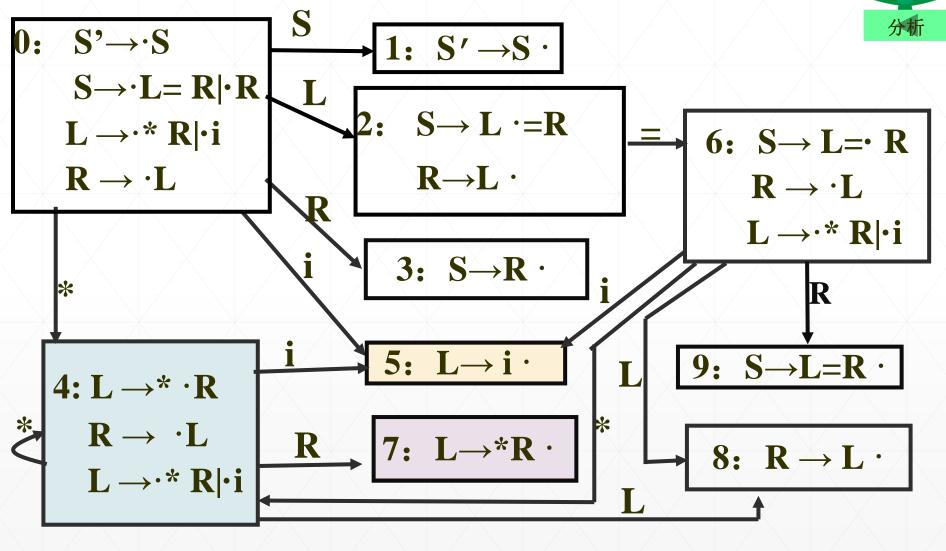


 $L \rightarrow *R \otimes |i \otimes 4|$



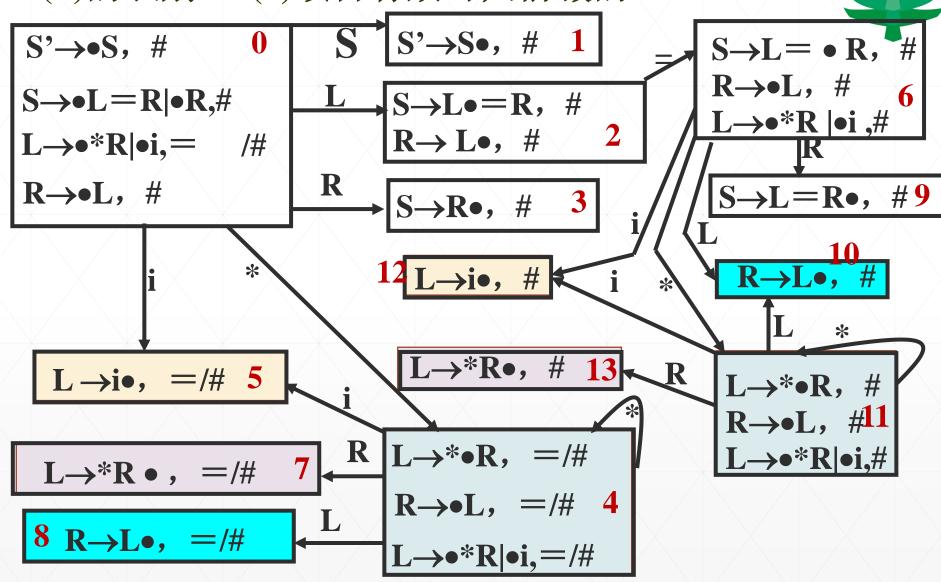




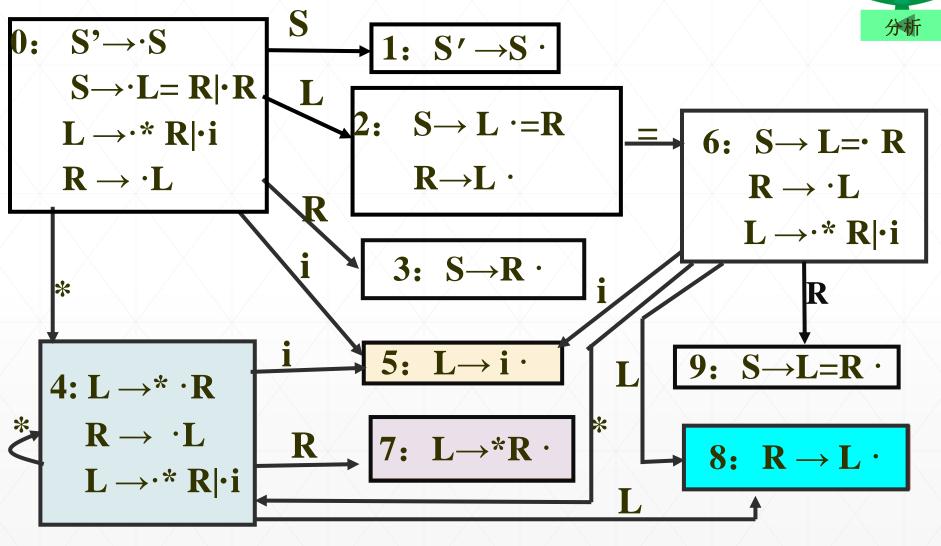


$$S$$
' $\rightarrow S$









$$S$$
' $\rightarrow S$

恩想。一缕

对比文法G(S)的识别LR(1)项目有效可归前缀 DFA和识别LR(0)项目有效可归前缀的DFA

LR(0)的DFA状态

LR(1)的DFA状态

4 分裂为

4, 11

5 分裂为

5, 12

分裂为

7、13

8

分裂为

8, 10

每个状态中 项目仅搜索符 不同, LR(0) 项目相同

■ 定义(同心状态、同心项目集)

对文法G的LR(1)项目集规范族(识别可归前缀的DFA), 若存在两个(或两个以上)项目集(状态)I0、I1, 其中的 LR(0)项目相同,仅搜索符不同,则称 I_0 、 I_1 为G的LR(1)的同心项目集(同心状态).或称Io、I1具有相同的心.

合并LR(1)的同心项目集(状态)=> LALR (1)的C或DFA



DFA的状态转移或GO函数做相应的修改。



1.LR(1)的C(DFA)无冲突 LALR(1)的C(DFA)

LALR(1)的C(DFA) { 无冲突:构造LALR(1)分析表 有冲突:构造LR(1)的分析表

2. LALR(1)的C无冲突,分析能力同LR(1)。 确定串中错误前有时会比LR(1)多产生若 干步归约。



5.7 LALR (1)分析

- 5.7.1 LALR(1)分析实现思想
- 5.7.2 LALR(1)分析表的构造



■ 算法5.11 (LALR(1)分析表构造)



输出: 文法G'的LALR(1)分析表

方法: 设 $C = \{I_0, I_1, ..., I_n\}$,分析表的状态= $\{0, 1, ..., n\}$ 。

①若GO $(I_k, a) = I_j$, $a \in V_T$, 则置action $(K, a) = S_j$; 若GO $(I_k, A) = I_j$, $A \in V_N$, 则置GOTO[K, A] = j。

②若 $[A \rightarrow \alpha \cdot ,a] \in I_k$,则置 $action(K,a) = r_i$,其中 $A \rightarrow \alpha$ 为文法G的第i个产生式并且A不是文法的开始符号;

③ 若接受项目 $[S' \rightarrow S', #] \in I_k$, S'是文法的开始符号,

则置action(K, #) = acc;

④分析表中不能用①至③规则填入信息的元素,则置"出错标志"

■ 算法(LALR(1)分析表构造)

输入: 拓广的文法G'及文法G'的识别LALR(1)项目有效 可归前缀的DFA

输出: 文法G'的LALR(1)分析表

方法: 设状态集 $Q = \{0, 1, ..., n\} =$ 分析表的状态

- ②若 $[A \rightarrow \alpha \cdot ,a] \in k$,则置 $action(K,a) = r_i$,其中 $A \rightarrow \alpha$ 为文法G的第j个产生式并且A不是文法的开始符号;
- ③ 若接受项目 $[S' \rightarrow S', \#] \in k$, S'是文法的开始符号, 则置action(K, #) = acc;
- ④分析表中不能用①至③规则填入信息的元素,则置"出错标志"

Ch5 语法分析 5.7 LALR(1)分析 5.7.1 LALR(1)分析实现思想 法G(S')的识别LALR(1)项目有效可归前 缀的DFA $S \rightarrow L = R \bullet , \#$ $S \rightarrow S^{\bullet}$, # 1 $S' \rightarrow \bullet S, \#$ $S \rightarrow L = \bullet R, \#$ $L S \rightarrow L \bullet = R, \#$ $S \rightarrow \bullet L = R/\bullet R, \#$ $R \rightarrow \bullet L$, # $R \rightarrow L \bullet$, # $L\rightarrow \bullet *R/\bullet i,\# 6$ $L \rightarrow \bullet *R/\bullet i, =/#$ $R \rightarrow \bullet L$, # $L \rightarrow i \bullet$, =/# 8+10 $R \rightarrow L \bullet$, =/# $S \rightarrow R \bullet$, # |3 $L \rightarrow * \bullet R, = /\#$ 4+11 $R \rightarrow \bullet L$, =/# $L \rightarrow \bullet *R/\bullet i, =/#$ $L\rightarrow *R \bullet , =/# 7+13R$ 第 22 页



文法G(S)LALR(1)分析表

state	Action				Goto		
		*	i	#	L	R	S
0		S ₄₊₁₁	S ₅₊₁₂		2	3	1
1				acc			
2	S ₆			r ₅			
3				$\mathbf{r_2}$			
4+11		S ₄₊₁₁	S_{5+12}		8+10	7+13	
5+12	r_4			r_4			
6		S ₄₊₁₁	S ₅₊₁₂		8+10	9	
7+13	\mathbf{r}_3			r ₃			
8+10	r ₅			r ₅			
9				\mathbf{r}_1			

S	$S \rightarrow S$
S	$S \rightarrow L = R \ (1) R (2)$
I	$L \rightarrow *R \otimes i \otimes 4$

 $R \rightarrow L$ (5)





定义

按照LALR(1)的项目集规范族构造的文法G的LALR(1)分析表,如果每个入口不含多重定义,则称文法G为LALR(1)文法。使用 LALR(1)分析表的语法分析器称作 LALR(1)分析器。



构造LALR(1)的分析表综述:

1. 构造文法G的LR(1)的项目集规范族C(DFA)

$$C = \{I_0, I_1, I_2, \dots, I_n\}$$

2. 合并C中的所有同心的LR(1) 项目集,用 M表示合并后的新项目集

 $M = \{M_0, M_1, M_2, ..., M_m\}$ 其中: m < n M 称为LALR(1)的项目集规范族。

3. M若没有冲突,据M构造文法G的LALR(1)的分析表。

LALR(1)的项目集规范族若存在冲突,只

能是归约— 归约冲突。 (尽管原来的LR(1)的C不冲突)

证明:

设原LR(1)的项目集规范族C中有如下项目集

$$I_{k}: [A \rightarrow \alpha \cdot , W_{1}] \quad I_{j}: [A \rightarrow \alpha \cdot , W_{2}]$$

$$[B \rightarrow \beta \cdot a \gamma, b] \quad [B \rightarrow \beta \cdot a \gamma, c]$$

由于 I_k 与 I_i 均无冲突,故有

$$W_1 \cap \{ a \} = \emptyset \qquad W_2 \cap \{ a \} = \emptyset$$

从而
$$(W_1 \cup W_2) \cap \{a\} = \emptyset$$



现将 I_k 与 I_i 合并,有

 $I_{k/i}$:

$$[A \rightarrow \alpha \cdot , W_1 \cup W_2]$$
$$[B \rightarrow \beta \cdot a \gamma, \{b\} \cup \{c\}]$$

若此时 I_{kli} 有"移进—归约"冲突,则必有

$$(W_1 \cup W_2) \cap \{a\} \neq \emptyset$$

与 I_k 和 I_i 无冲突的假设矛盾。

:. 合并同心集后不会引入新的"移进—归约" 冲突。 证毕。



例:设有下列文法

$$S \rightarrow Aa \mid bAc \mid Bc \mid bBa$$

$$A \rightarrow d$$

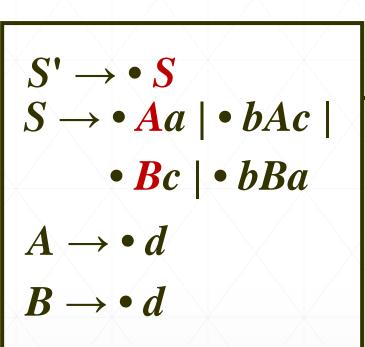
$$B \rightarrow d$$

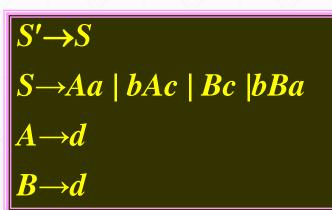
判断该文法是哪类LR文法?

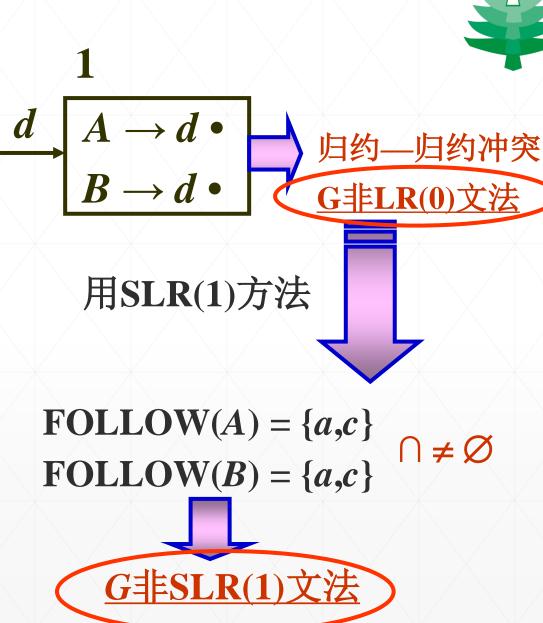
没有发现有二义性,所以从LR(0)开始判断。 拓广文法:

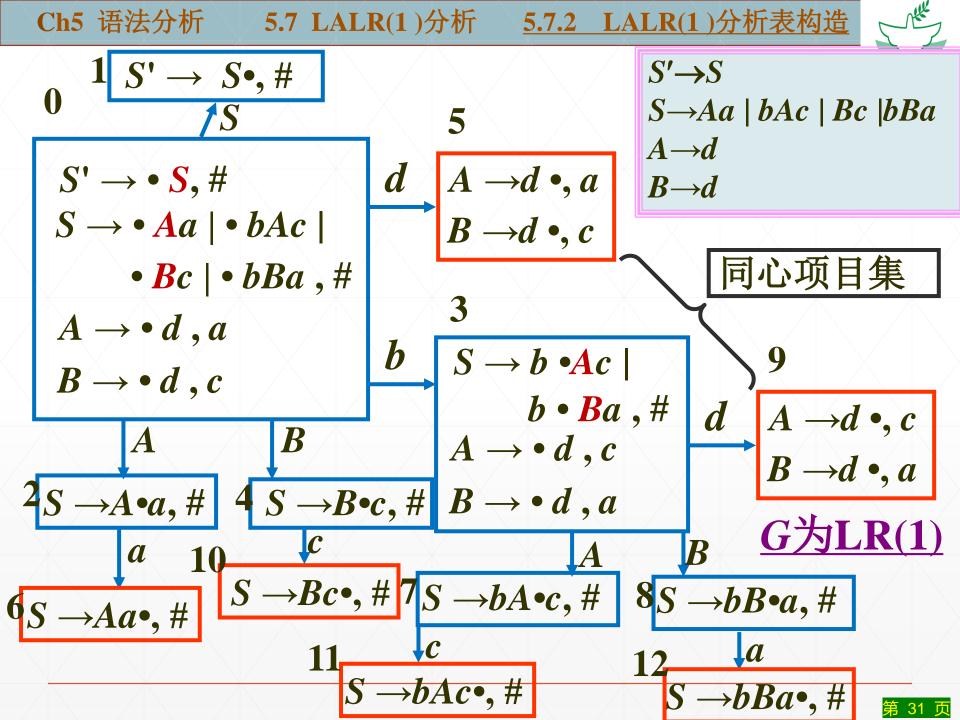
增加一产生式: $S' \rightarrow S$













状态5和状态9合并

$$5+9$$

$$A \rightarrow d \cdot, a/c$$

$$B \rightarrow d \cdot, c/a$$

归约—归约冲突。

所以G不是LALR(1)



Ch5 语法分析

- 5.1 "移近—归约"分析法
- 5.3 LR分析鸟瞰
 - 5.4 LR (0)分析
 - 5.5 SLR (1)分析
 - 5.6 LR (1)分析
 - 5.7 LALR (1)分析
 - 5.8 LR分析对二义文法的应用



- 5.9 LR分析的错误处理与恢复
- 5.10 语法分析器的自动生成与YACC(自学)



定理

任何一个二义文法都不是一个LR文法。



二义文法会导致语法分析的二义性。



二义文法的有用之处在于可以缩小文法的规模。



例如,
$$G: E \rightarrow E + E \mid E^*E \mid (E) \mid i$$

G':
$$E \rightarrow T \mid E+T$$

$$T \rightarrow F \mid T*F$$

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

例如, $G: S \rightarrow iSeS \mid iS \mid a$

 $G': S \rightarrow \{iSeS\} \mid \{iS\} \mid a$



例:设有文法G(E):

$$E \rightarrow E + E$$

$$E \rightarrow E^*E$$

$$E \rightarrow (E)$$

$$E \rightarrow i$$

拓广文法:

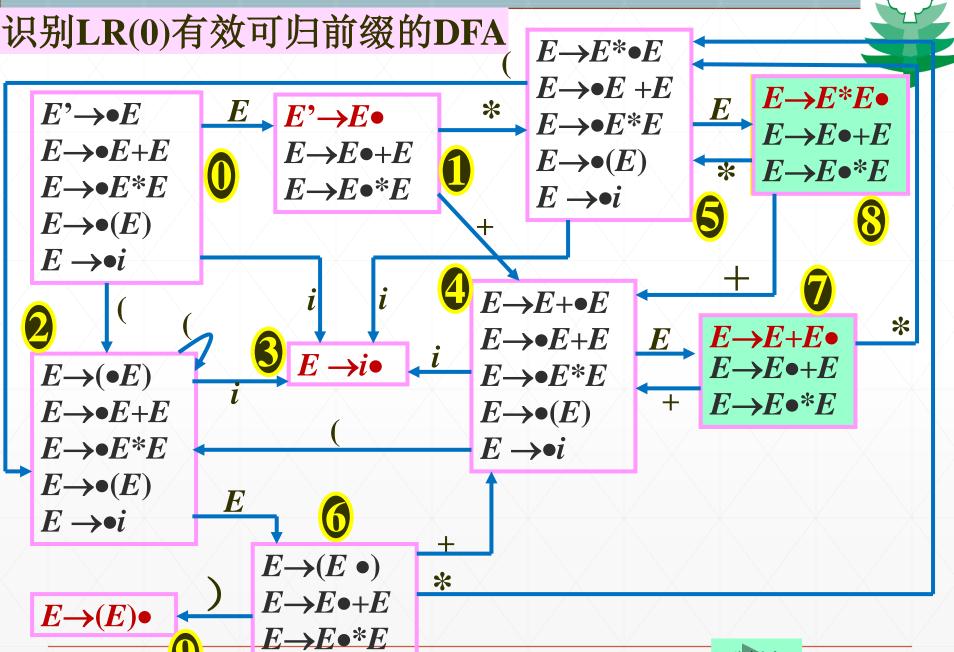
$$E' \rightarrow E$$

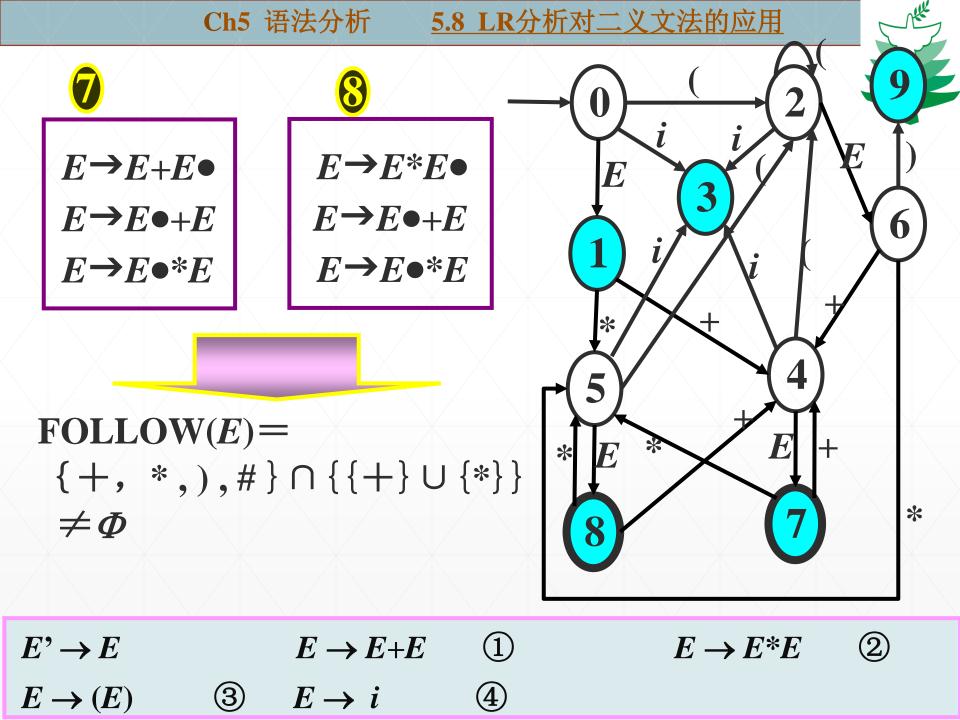
$$E \rightarrow E + E$$

$$E \rightarrow E^*E$$

$$E \rightarrow (E)$$

$$E \rightarrow i$$



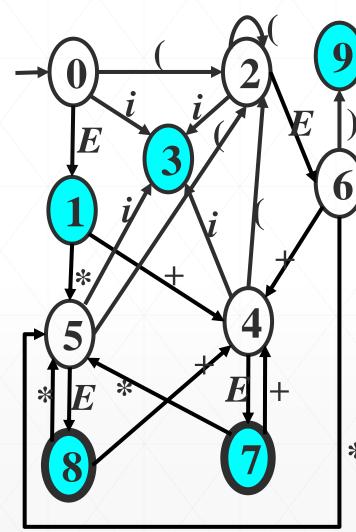


Ch5 语法分析 5.8 LR分析对二义文法的应用





 $E' \rightarrow E$ 1 $E \rightarrow E + E$ 2 $E \rightarrow E * E$ 3 $E \rightarrow (E)$ 4 $E \rightarrow i$



文法G(E)的LR分析表

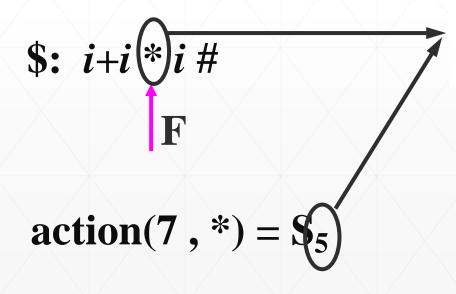
状			STAT	ION			GOT
状态	+	*	(()	i	#	E
0			S ₂		S ₃		1
1	S ₄	S ₅				acc	
2			S ₂		S_3		6
3	r ₄	r ₄	r ₄	r_4	r ₄	r ₄	
4			$\mathbf{S_2}$		S ₃		7
5			S ₂		S_3		8
6	S ₄	S ₅		S ₉			
7	\mathbf{r}_{1}	S ₅	\mathbf{r}_{1}	\mathbf{r}_1	\mathbf{r}_1	\mathbf{r}_1	
8	\mathbf{r}_{2}	r ₂	\mathbf{r}_{2}	r ₂	\mathbf{r}_{2}	r ₂	
9	r ₃	\mathbf{r}_3	r_3	\mathbf{r}_3	r_3	r_3	

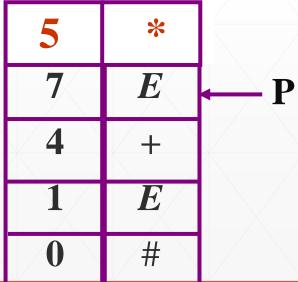
 $1:E' \rightarrow E \bullet \quad 3:E \rightarrow i \bullet \quad 7:E \rightarrow E + E \bullet$ $\delta: E \rightarrow E^*E^{\bullet}$ $9:E \rightarrow (E)^{\bullet}$



$$action(7, +) = r_2$$

7	E	P
4	+	E
1	E	
0	#	







例: 设有文法G:

$$S \rightarrow iSeS \mid iS \mid a$$
 ① | ② | ③

其中:

i: if e_r then

e: else

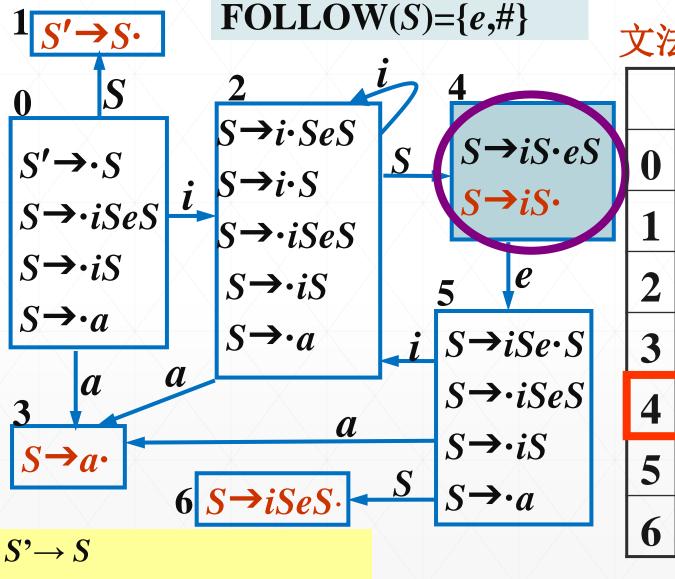
a, S: 语句

拓广文法G:

$$S' \rightarrow S$$

$$S \rightarrow iSeS \mid iS \mid a$$
 ① | ② | ③

文法G的识别LR(0)项目有效可归前缀的DFA



文法G的LR分析表

	i	e	a	#	S
0	S ₂		S_3		1
1				acc	
2	S_2		S_3		4
3	r ₃	r ₃	$\mathbf{r_3}$	r ₃ /	
4	\mathbf{r}_2	S ₅	\mathbf{r}_{2}	$\mathbf{r_2}$	
5	S ₂		S_3		6
6	\mathbf{r}_1	\mathbf{r}_1	\mathbf{r}_1	$\mathbf{r_1}$	

 $S' \rightarrow S$



$$S \rightarrow iSeS \mid iS \mid a$$
 ① | ② |③

if
$$e_1$$
 then if e_2 then S_1 else S_2

i a e a

Ch5 语法分析



- 5.1 "移近—归约"分析法
- 5.3 LR分析鸟瞰
- 5.4 LR (0)分析
- 5.5 SLR (1)分析
- 5.6 LR (1)分析
- 5.7 LALR (1)分析
- 5.8 LR分析对二义文法的应用
- 5.9 LR分析的错误处理与恢复
- 5.10 语法分析器的自动生成与YACC(自学)





编译中的错误种类

- 1. 词法错误;
- 2. 语法错误;
- 3. 语义错误(静态、动态);
- 4. 违反环境限制的错误;



- 1. 清晰准确地报告错误(错误的定性和定位);
- 2. 迅速从每个错误中恢复过来,以便诊断后面的错误;



错误处理与恢复策略

- 1. 局部化恢复
- 出错产生式
 扩充语言的文法,增加产生错误结构的产生式。分析中可以直接识别处理错误。
 - 3. 全局纠正 获取全局最小代价纠正。





1. 紧急恢复方式

编译器每次发现错误时,抛弃一个输入符号,直到输入符号属于某个指定的同步符号集合为止。

例如,界限符:

; \ }\ end ...



思想: 在分析到某一含有错误的短语时,

分析程序认为含有错误的短语是由某一非终 结符A所推导出的,

短语的一部分已处理,放在栈顶部,

剩下未处理的在输入串,

分析程序跳过这些剩余符号,

直至找到A的跟随字符为止,

同时把栈顶内容逐个移去,

直至找到某个状态q,GOTO(q,A)对应一新状态,将GOTO(q,A),A压入栈。

第 49 页



2. 短语级恢复

发现错误时,对剩余符号串作局部校正。 使用可以使编译器继续工作的输入串代替剩余输入的前缀。

在输入串的出错点采用插入、删除或修改的方法。

*关键:选择合适的替换串。

$$(i+i; \longrightarrow (i+i);$$

$$i i^*i;$$
 \longrightarrow $i(+)i^*i;$



- 1. 清晰准确地报告错误(错误的定性和定位);
- 2. 迅速从每个错误中恢复过来,以便诊断后面的错误;
- 错误恢复的困难
 - 1. 源程序书写的灵活性;
 - 2. 错误发生的随机性;
 - 3. 不使正确程序的处理效率降低。
 - 4. 可能校正途径的多样性。



例:设有文法G(E):

$$E' \rightarrow E$$

$$E \rightarrow E + E$$

$$E \rightarrow E * E$$

2

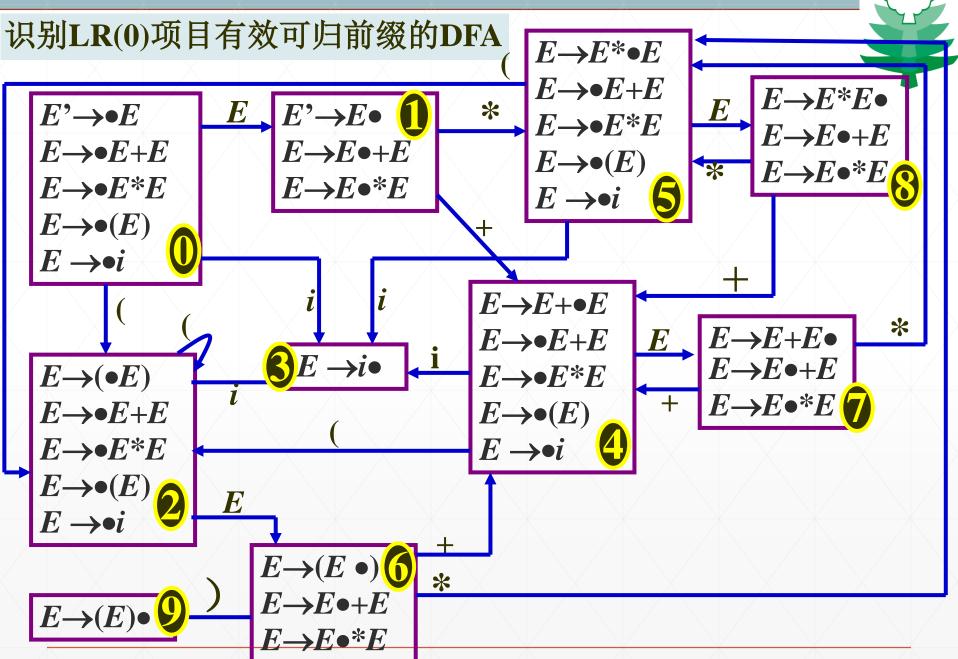
$$E \rightarrow (E)$$

3

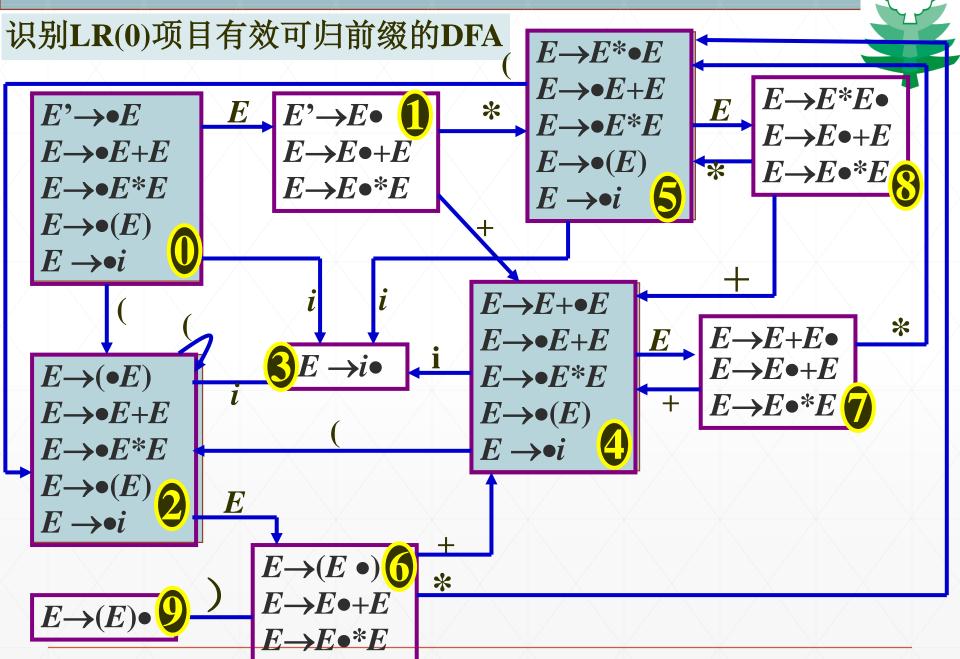
$$E \rightarrow i$$

4

构造文法G(E)的带出错处理的LR分析表



AD X			ACT	ION			GOTO
状态	i	+	*	(#	E
0	S3			S2			1
1		S4	S5			acc	
2	S3			S2			6
3	r4	r4	r4	r4	r4	r4	
4	S3			S2			7
5	S3			S2			8
6		S4	S5		S9		
7	r1	r1	S5	r1	r1	r1	
8	r2	r2	r2	r2	r2	r2	
9	r3	r3	r3	r3	r3	r3	





e1:

处于状态0、2、4、5时,要求输入符号为运算对象i或(E),此时若遇到+、*、#,则调用e1。

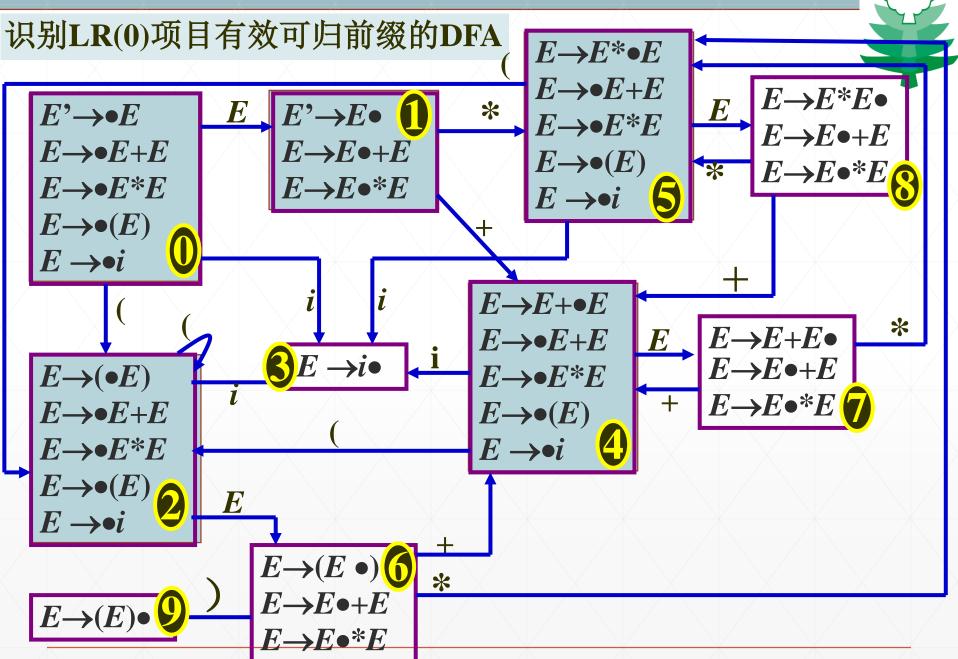
e1功能:

将假设的 i 和状态3入栈; (shift)

e1出错信息:

缺少运算对象。

小下子			ACT	ION		GOTO	
状态	i	+	*			#	E
0	S3	e1	e1	S2		e1	1
1		S4	S5			acc	
2	S3	e1	e1	S2		e1	6
3	r4	r4	r4	r4	r4	r4	
4	S3	e1	e1	S2		e1	7
5	S3	e1	e1	S2		e1	8
6		S4	S5		S9		
7	r1	r1	S5	r1	r1	r1	
8	r2	r2	r2	r2	r2	r2	
9	r3	r3	r3	r3	r3	r3	





e2:

处于状态0、1、2、4、5时, 若遇到")",则调用e2。

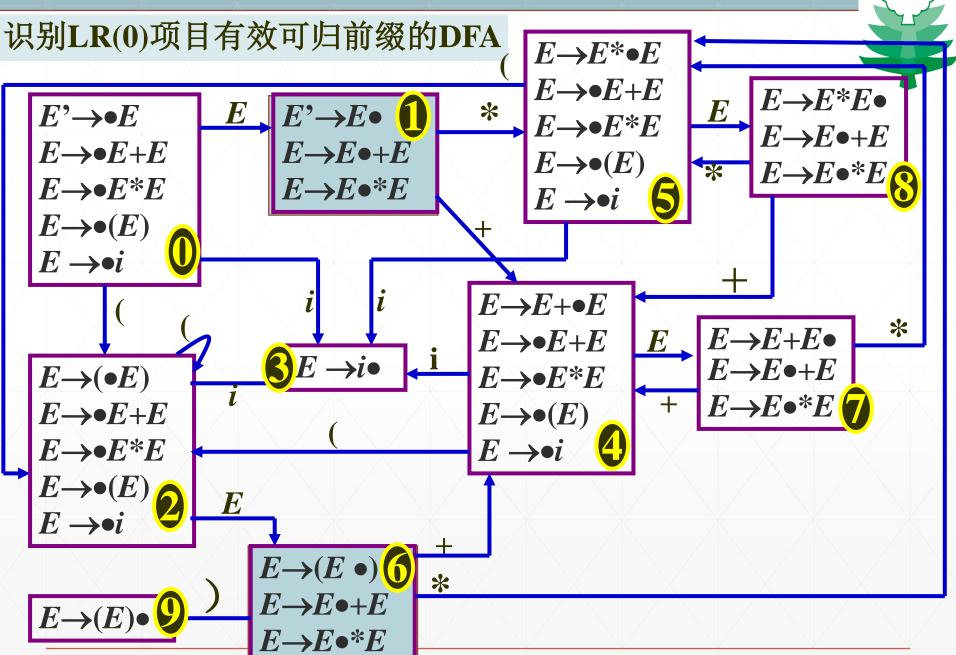
e2功能:

删除输入的")"; (F++)

e2出错信息:

")"不配对。

小中子			ACT	ION			GOTO
状态	i	+	*			#	E
0	S3	e1	e1	S2	e2	e1	1
1		S4	S5		e2	acc	
2	S3	e1	e1	S2	e2	e1	6
3	r4	r4	r4	r4	r4	r4	
4	S3	e1	e1	S2	e2	e1	7
5	S3	e1	e1	S2	e2	e1	8
6		S4	S5		S9		
7	r1	r1	S5	r1	r1	r1	
8	r2	r2	r2	r2	r2	r2	
9	r3	r3	r3	r3	r3	r3	





e3:

处于状态1或6时,期望下面输入符为运算符或")",但遇到"i"或"("时,则调用e3。

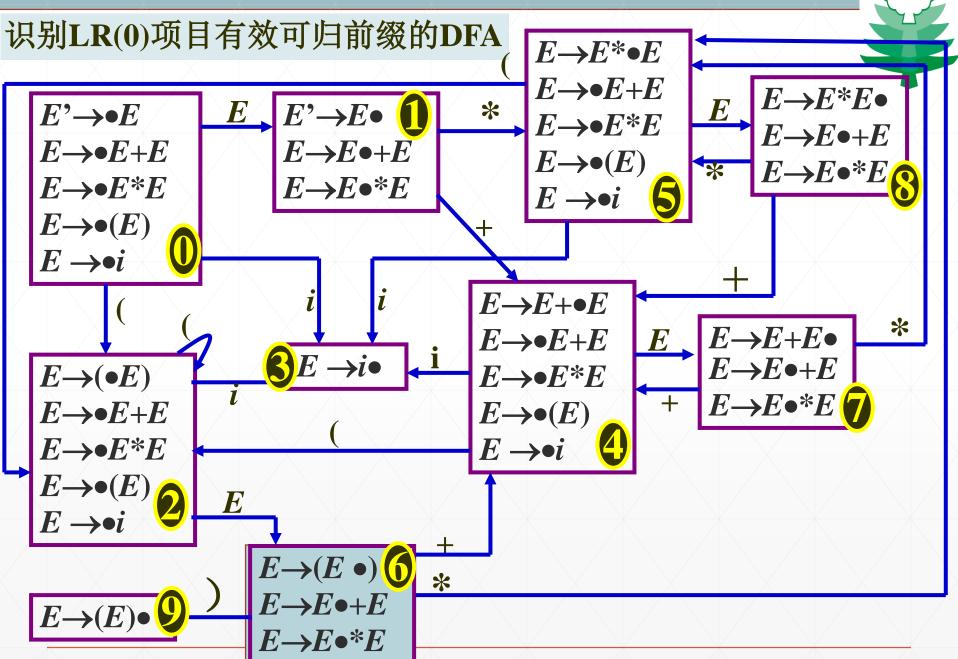
e3功能:

将假设的"+"和状态 4入栈, (P++);

e3出错信息:

缺少运算符。

状态			ACT	ACTION			GOTO
11/25	i	+	*			#	E
0	S3	e1	e1	S2	e2	e1	1
1	e3	S4	S5	e3	e2	acc	
2	S3	e1	e1	S2	e2	e1	6
3	r4	r4	r4	r4	r4	r4	
4	S3	e1	e1	S2	e2	e1	7
5	S3	e1	e1	S2	e2	e1	8
6	e3	S4	S5	e3	S9		
7	r1	r1	S5	r1	r1	r1	
8	r2	r2	r2	r2	r2	r2	
9	r3	r3	r3	r3	r3	r3	





e4:

处于状态6时,期望下面输入符为运算符或")",但遇到"#"时,则调用e4。

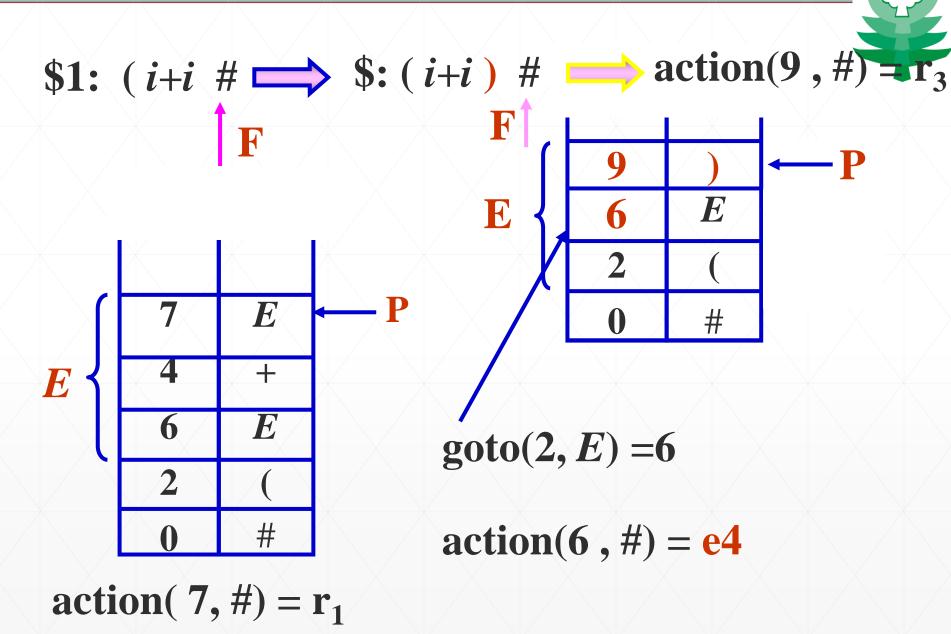
e4功能:

将假设的")"和状态9入栈,(P++);

e4出错信息:

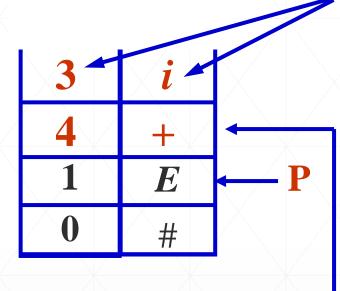
缺少")"。

状态			ACT	ION			GOTO
1/125	i	+	*	(#	E
0	S3	e1	e1	S2	e2	e1	1
1	e3	S4	S5	e3	e2	acc	
2	S3	e1	e1	S2	e2	e1	6
3	r4	r4	r4	r4	r4	r4	
4	S3	e1	e1	S2	e2	e1	7
5	S3	e1	e1	S2	e2	e1	8
6	e3	S4	S5	e3	S9	e4	
7	r1	r1	S5	r1	r1	r1	
8	r2	r2	r2	r2	r2	r2	
9	r3	r3	r3	r3	r3	r3	









$$action(1, i) = e3$$

Ch5 语法分析 (自下而上)



end

