



编译原理

第五章 语法分析——自下而上分析

第五章 语法分析——自下而上分析

- 自下而上分析的基本问题
- 算符优先分析算法
- LR 分析法

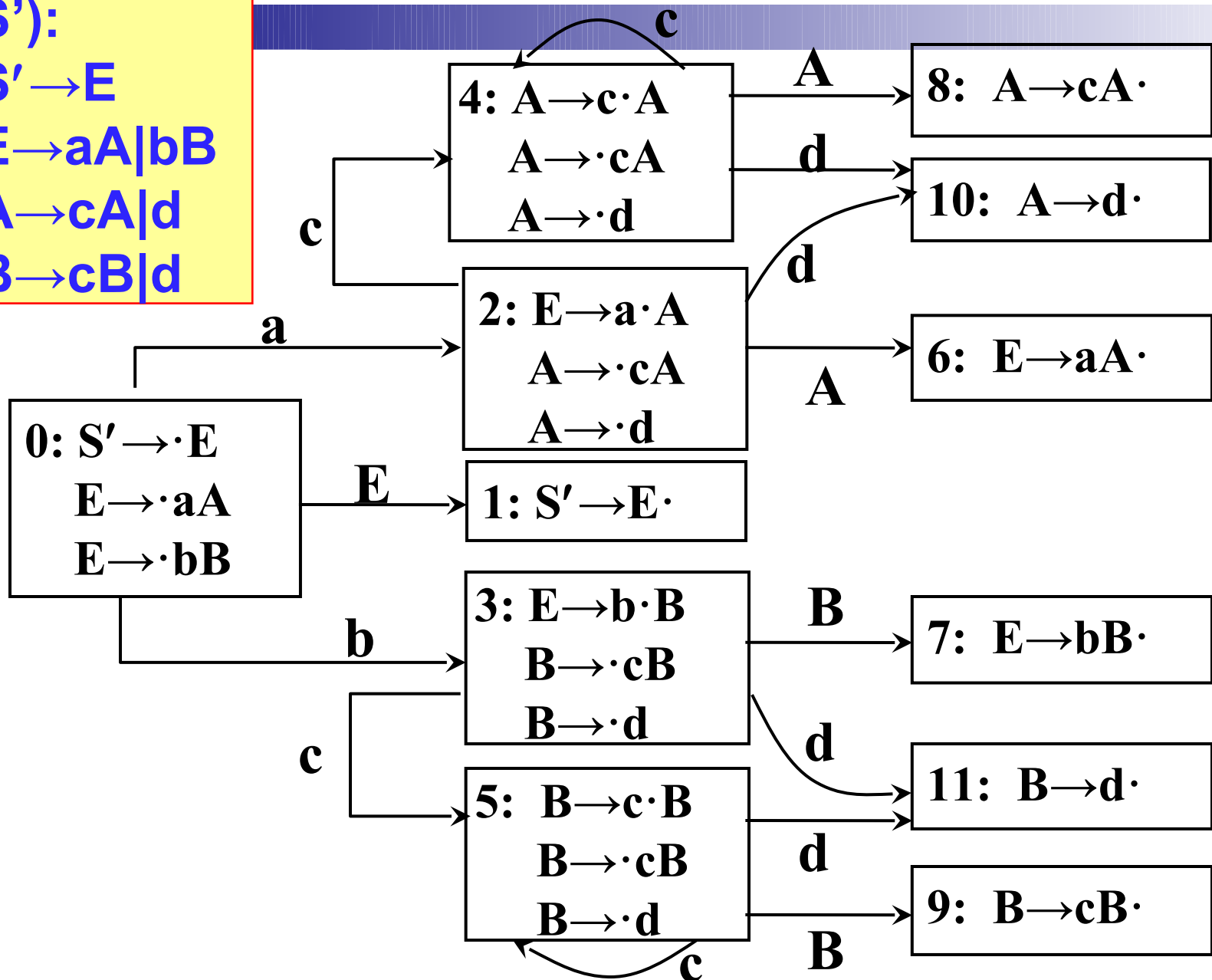
第五章 语法分析——自下而上分析

- 自下而上分析的基本问题
- 算符优先分析算法
- LR 分析法
 - LR(0) 项目集规范族和 LR(0) 分析表的构造
 - SLR 分析表的构造

LR(0) 分析表的构造

- 假若一个文法 G 的拓广文法 G' 的活前缀识别自动机中的每个状态（项目集）不存在下述情况：
 - 1) 既含移进项目又含归约项目，
 - 2) 含有多个归约项目则称 G 是一个 **LR(0) 文法**。

$G'(S')$:
 $S' \rightarrow E$
 $E \rightarrow aA | bB$
 $A \rightarrow cA | d$
 $B \rightarrow cB | d$



构造 LR(0) 分析表的算法

- 令每个项目集 I_k 的下标 k 作为分析器的状态，包含项目 $S' \rightarrow \cdot S$ 的集合 I_k 的下标 k 为分析器的初态。
- 在任何时候，分析栈中的活前缀 $X_1 X_2 \cdots X_m$ 的有效项目也正是从识别活前缀的 DFA 的初态出发，读出 $X_1 X_2 \cdots X_m$ 后到达的那个项目集（状态）
- 也正是栈顶状态 S_m 所代表的那个集合

LR(0) 分析表的 ACTION 和 GOTO 子表构造

1. 若项目 $A \rightarrow \alpha \cdot a\beta$ 属于 I_k 且 $GO(I_k, a) = I_j$, a 为终结符, 则置 $ACTION[k, a]$ 为 “sj”。
2. 若项目 $A \rightarrow \alpha \cdot$ 属于 I_k , 那么, 对任何终结符 a (或结束符 #), 置 $ACTION[k, a]$ 为 “rj” (假定产生式 $A \rightarrow \alpha$ 是文法 G' 的第 j 个产生式)。
3. 若项目 $S' \rightarrow S \cdot$ 属于 I_k , 则置 $ACTION[k, \#]$ 为 “acc”。
4. 若 $GO(I_k, A) = I_j$, A 为非终结符, 则置 $GOTO[k, A] = j$ 。
5. 分析表中凡不能用规则 1 至 4 填入信息的空白格均置上 “报错标志”。

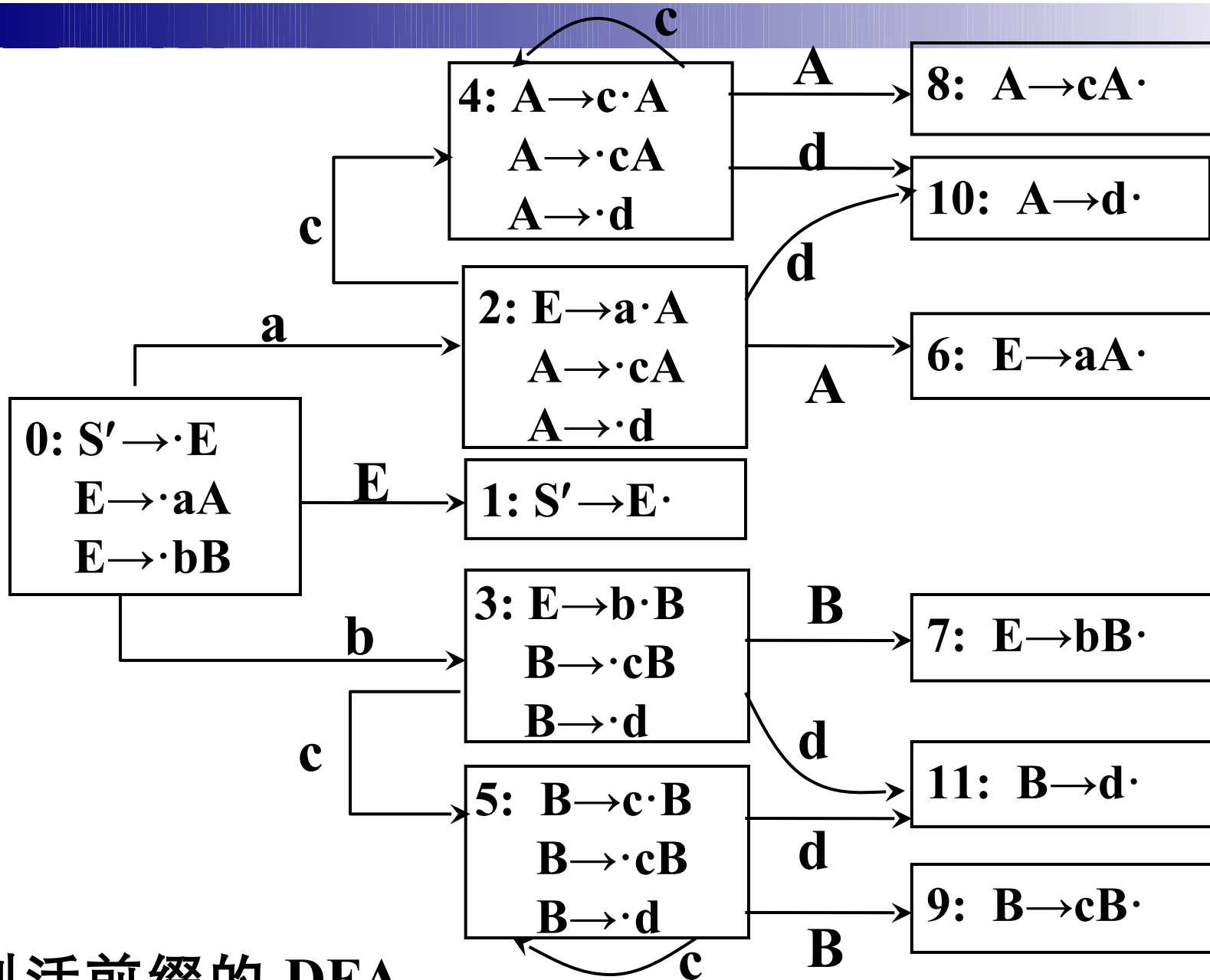
■ 文法 $G(S')$

$S' \rightarrow E$

$E \rightarrow aA | bB$

$A \rightarrow cA | d$

$B \rightarrow cB | d$

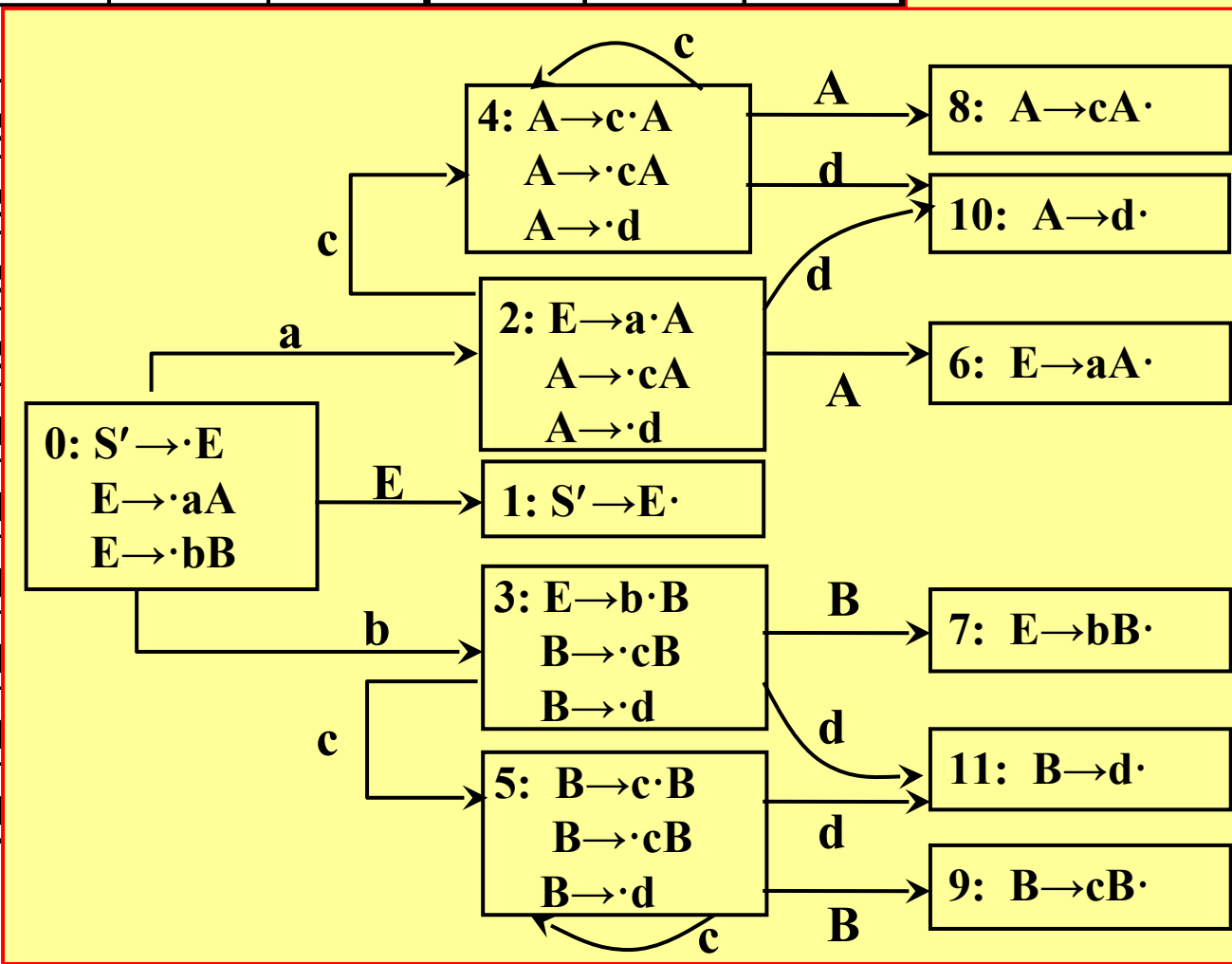


识别活前缀的 DFA

■ LR(0) 分析表为

状态	ACTION					GOTO		
	a	b	c	d	#	E	A	B
0	s2	s3				1		
1								
2								
3								
4								
5								
6	r1	r1						
7	r2	r2						
8	r3	r3						
9	r5	r5						
10	r4	r4						
11	r6	r6						

0: $S' \rightarrow E$
 1: $E \rightarrow aA$
 2: $E \rightarrow bB$
 3: $A \rightarrow cA$
 4: $A \rightarrow d$
 5: $B \rightarrow cB$
 6: $B \rightarrow d$



■ 例：按上表对 bcd 进行分析

步骤	状态	符号	输入串
1	0	#	bcd#
2	03	#b	cd#
3	035	#bc	d#

0: $S' \rightarrow E$
 1: $E \rightarrow aA$
 2: $E \rightarrow bB$
 3: $A \rightarrow cA$
 4: $A \rightarrow d$
 5: $B \rightarrow cB$
 6: $B \rightarrow d$

状态	ACTION					GOTO		
	a	b	c	d	#	E	A	B
0	s2	s3				1		
1					acc			
2			s4	s10			6	
3			s5	s11				7
4			s4	s10			8	
5			s5	s11				9
6	r1	r1	r1	r1	r1			
7	r2	r2	r2	r2	r2			
8	r3	r3	r3	r3	r3			

■ 例：按上表对 bccd 进行分析

步骤	状态	符号	输入串
3	035	#bc	d#
4	035 <u>11</u>	#bcd	#
5	0359	#bcB	#

0: $S' \rightarrow E$
 1: $E \rightarrow aA$
 2: $E \rightarrow bB$
 3: $A \rightarrow cA$
 4: $A \rightarrow d$
 5: $B \rightarrow cB$
 6: $B \rightarrow d$

	ACTION					GOTO		
状态	a	b	c	d	#	E	A	B
0	s2	s3				1		
1					acc			
2			s4	s10			6	
3			s5	s11				7
4			s4	s10			8	
5			s5	s11				9
6	r1	r1	r1	r1	r1			
7	r2	r2	r2	r2	r2			
8	r3	r3	r3	r3	r3			
9	r5	r5	r5	r5	r5			
10	r4	r4	r4	r4	r4			
11	r6	r6	r6	r6	r6			

■ 例：按上表对 bccd 进行分析

<u>步骤</u>	<u>状态</u>	<u>符号</u>	<u>输入串</u>
5	0359	#bcB	#
6	037	#bB	#
7	01	#E	#
8	接受		

状态	ACTION					GOTO		
	a	b	c	d	#	E	A	B
0	s2	s3				1		
1					acc			
2			s4	s10			6	
3			s5	s11				7
4			s4	s10			8	
5			s5	s11				9
6	r1	r1	r1	r1	r1			
7	r2	r2	r2	r2	r2			
8	r3	r3	r3	r3	r3			
9	r5	r5	r5	r5	r5			
10	r4	r4	r4	r4	r4			

0: $S' \rightarrow E$
 1: $E \rightarrow aA$
 2: $E \rightarrow bB$
 3: $A \rightarrow cA$
 4: $A \rightarrow d$
 5: $B \rightarrow cB$
 6: $B \rightarrow d$

第五章 语法分析——自下而上分析

- 自下而上分析的基本问题
- 算符优先分析算法
- LR 分析法
 - LR(0) 项目集族和 LR(0) 分析表的构造
 - SLR 分析表的构造

5.3.3 SLR 分析表的构造

- LR(0) 文法太简单，没有实用价值

- 例 5.11 考察下面的拓广文法：

(0) $S' \rightarrow E$

(1) $E \rightarrow E + T$

(2) $E \rightarrow T$

(3) $T \rightarrow T * F$

(4) $T \rightarrow F$

(5) $F \rightarrow (E)$

(6) $F \rightarrow i$

■ 这个文法的 LR(0) 项目集规范族为：

I_0 : $S' \rightarrow \cdot E$
 $E \rightarrow \cdot E + T$
 $E \rightarrow \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_1 : $S' \rightarrow E \cdot$
 $E \rightarrow E \cdot + T$

I_2 : $E \rightarrow T \cdot$
 $T \rightarrow T \cdot * F$

I_3 : $T \rightarrow F \cdot$

I_4 : $F \rightarrow (\cdot E)$
 $E \rightarrow \cdot E + T$
 $E \rightarrow \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_5 : $F \rightarrow i \cdot$

I_6 : $E \rightarrow E + \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

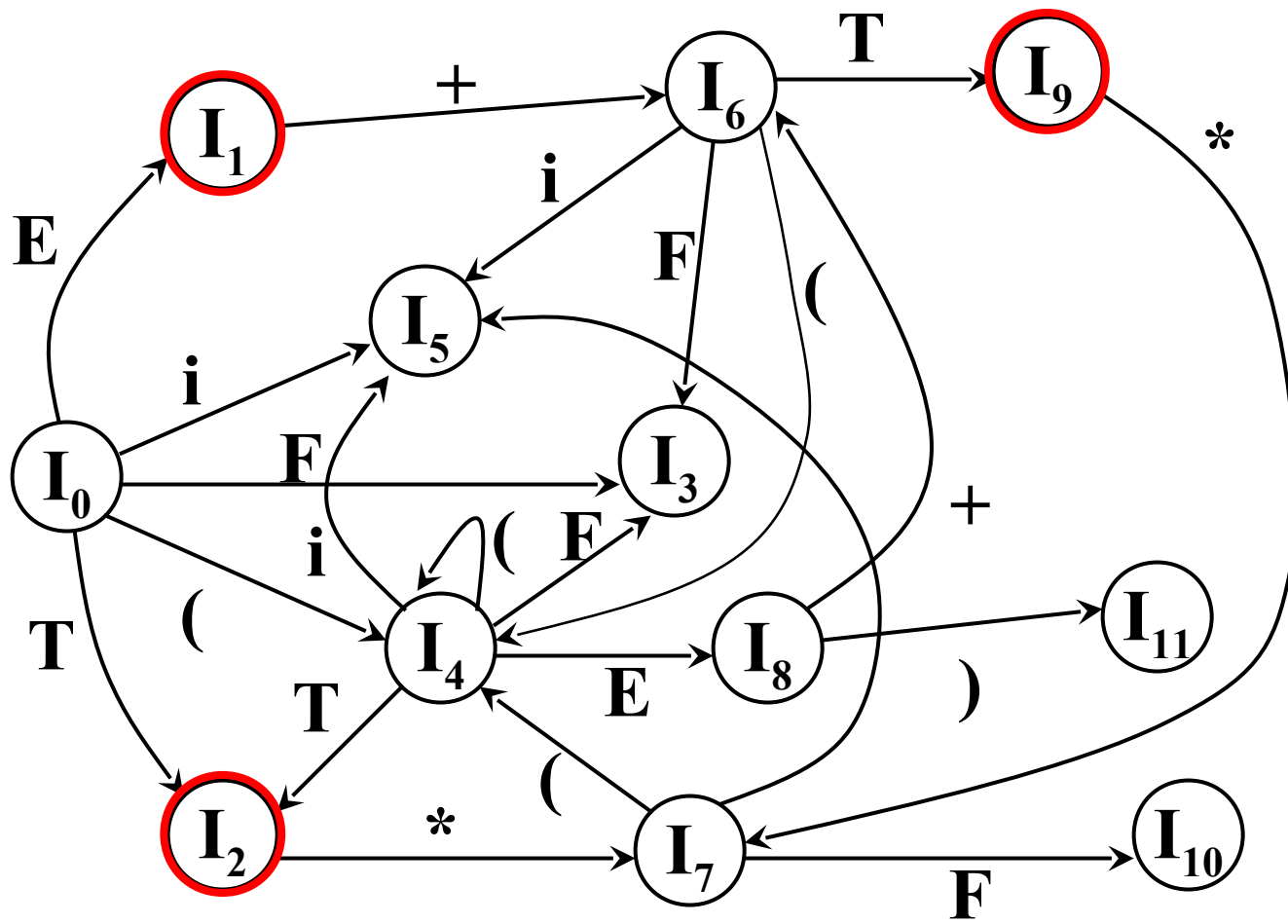
I_7 : $T \rightarrow T * \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_8 : $F \rightarrow (E \cdot)$
 $E \rightarrow E \cdot + T$

I_9 : $E \rightarrow E + T \cdot$
 $T \rightarrow T \cdot * F$

I_{10} : $T \rightarrow T * F \cdot$

I_{11} : $F \rightarrow (E) \cdot$



■ 这个文法的 LR(0) 项目集规范族为：

$I_0: S' \rightarrow \cdot E$

$E \rightarrow \cdot E + T$

$E \rightarrow \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_1: S' \rightarrow E \cdot$

$E \rightarrow E \cdot + T$

$I_2: E \rightarrow T \cdot$

$T \rightarrow T \cdot * F$

$I_3: T \rightarrow F \cdot$

$I_4: F \rightarrow (\cdot E)$

$E \rightarrow \cdot E + T$

$E \rightarrow \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_5: F \rightarrow i \cdot$

$I_6: E \rightarrow E + \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_7: T \rightarrow T * \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_8: F \rightarrow (E \cdot)$

$E \rightarrow E \cdot + T$

$I_9: E \rightarrow E + T \cdot$

$T \rightarrow T \cdot * F$

$I_{10}: T \rightarrow T * F \cdot$

$I_{11}: F \rightarrow (E) \cdot$

- I_1 、 I_2 和 I_9 都含有 “移进 - 归约” 冲突

$I_1: S' \rightarrow E \cdot$

$E \rightarrow E \cdot + T$

$I_2: E \rightarrow T \cdot$

$T \rightarrow T \cdot * F$

$I_9: E \rightarrow E + T \cdot$

$T \rightarrow T \cdot * F$

讨论：有没有办法消除冲突？

$$FOLLOW(A) = \{a \mid S \xRightarrow{*} \dots Aa \dots, a \in V_T\}$$

$$FOLLOW(S') = \{\#\}$$

$$FOLLOW(E) = \{\#,), +\}$$

5.3.3 SLR 分析表的构造

- LR(0) 文法太简单，没有实用价值
- 假定一个 LR(0) 规范族中含有如下的一个项目集 (状态) $I = \{X \rightarrow \alpha \cdot b\beta, A \rightarrow \alpha \cdot, B \rightarrow \alpha \cdot\}$ 。 $\text{FOLLOW}(A)$ 和 $\text{FOLLOW}(B)$ 的交集为 \emptyset ，且不包含 b ，那么，当状态 I 面临任何输入符号 a 时，可以：
 1. 若 $a=b$ ，则移进；
 2. 若 $a \in \text{FOLLOW}(A)$ ，用产生式 $A \rightarrow \alpha$ 进行归约；
 3. 若 $a \in \text{FOLLOW}(B)$ ，用产生式 $B \rightarrow \alpha$ 进行归约；
 4. 此外，报错。

- 假定 LR(0) 规范族的一个项目集
 $I = \{A_1 \rightarrow \alpha \cdot a_1 \beta_1, A_2 \rightarrow \alpha \cdot a_2 \beta_2, \dots, A_m \rightarrow \alpha \cdot a_m \beta_m, B_1 \rightarrow \alpha \cdot, B_2 \rightarrow \alpha \cdot, \dots, B_n \rightarrow \alpha \cdot\}$ 如果集合 $\{a_1, \dots, a_m\}$, $\text{FOLLOW}(B_1)$, \dots , $\text{FOLLOW}(B_n)$ 两两不相交 (包括不得有两个 FOLLOW 集合有 #), 则:
 1. 若 a 是某个 a_i , $i=1,2,\dots,m$, 则移进;
 2. 若 $a \in \text{FOLLOW}(B_i)$, $i=1,2,\dots,n$, 则用产生式 $B_i \rightarrow \alpha$ 进行归约;
 3. 此外, 报错。

构造 SLR(1) 分析表方法

- 首先把 G 拓广为 G' ，对 G' 构造 LR(0) 项目集规范族 C 和活前缀识别自动机的状态转换函数 GO
- 然后使用 C 和 GO ，按下面的算法构造 SLR 分析表
 - 令每个项目集 I_k 的下标 k 作为分析器的状态，包含项目 $S' \rightarrow \cdot S$ 的集合 I_k 的下标 k 为分析器的初态。

SLR(1) 分析表的 ACTION 和 GOTO 子表构造

1. 若项目 $A \rightarrow \alpha \cdot a\beta$ 属于 I_k 且 $GO(I_k, a) = I_j$, a 为终结符, 则置 $ACTION[k, a]$ 为 “sj” ;
2. 若项目 $A \rightarrow \alpha \cdot$ 属于 I_k , 那么, 对任何终结符 a , $a \in FOLLOW(A)$, 置 $ACTION[k, a]$ 为 “rj” ; 其中, 假定 $A \rightarrow \alpha$ 为文法 G' 的第 j 个产生式;
3. 若项目 $S' \rightarrow S \cdot$ 属于 I_k , 则置 $ACTION[k, \#]$ 为 “acc” ;
4. 若 $GO(I_k, A) = I_j$, A 为非终结符, 则置 $GOTO[k, A] = j$;
5. 分析表中凡不能用规则 1 至 4 填入信息的空白格均

LR(0) 分析表的 ACTION 和 GOTO 子表构造

1. 若项目 $A \rightarrow \alpha \cdot a\beta$ 属于 I_k 且 $GO(I_k, a) = I_j$, a 为终结符, 则置 $ACTION[k, a]$ 为 “sj”。
2. 若项目 $A \rightarrow \alpha \cdot$ 属于 I_k , 那么, 对任何终结符 a (或结束符 #), 置 $ACTION[k, a]$ 为 “rj” (假定产生式 $A \rightarrow \alpha$ 是文法 G' 的第 j 个产生式)。
3. 若项目 $S' \rightarrow S \cdot$ 属于 I_k , 则置 $ACTION[k, \#]$ 为 “acc”。
4. 若 $GO(I_k, A) = I_j$, A 为非终结符, 则置 $GOTO[k, A] = j$ 。
5. 分析表中凡不能用规则 1 至 4 填入信息的空白格均置上 “报错标志”。

SLR(1) 分析表的 ACTION 和 GOTO 子表构造

1. 若项目 $A \rightarrow \alpha \cdot a\beta$ 属于 I_k 且 $GO(I_k, a) = I_j$, a 为终结符, 则置 $ACTION[k, a]$ 为 “sj” ;
2. 若项目 $A \rightarrow \alpha \cdot$ 属于 I_k , 那么, 对任何终结符 a , $a \in FOLLOW(A)$, 置 $ACTION[k, a]$ 为 “rj” ; 其中, 假定 $A \rightarrow \alpha$ 为文法 G' 的第 j 个产生式;
3. 若项目 $S' \rightarrow S \cdot$ 属于 I_k , 则置 $ACTION[k, \#]$ 为 “acc” ;
4. 若 $GO(I_k, A) = I_j$, A 为非终结符, 则置 $GOTO[k, A] = j$;
5. 分析表中凡不能用规则 1 至 4 填入信息的空白格均

SLR(1) 文法

- 按上述方法构造出的 ACTION 与 GOTO 表如果不含多重入口，则称该文法为 **SLR(1) 文法**
- 使用 SLR 表的分析器叫做一个 **SLR 分析器**
- 每个 SLR(1) 文法都是无二义的。但也存在许多无二义文法不是 SLR(1) 的
- $LR(0) \subset SLR(1) \subset$ 无二义文法

■ 例 5.11 考察下面的拓广文法：

$$(0) S' \rightarrow E$$

$$(1) E \rightarrow E + T$$

$$(2) E \rightarrow T$$

$$(3) T \rightarrow T * F$$

$$(4) T \rightarrow F$$

$$(5) F \rightarrow (E)$$

$$(6) F \rightarrow i$$

■ 这个文法的 LR(0) 项目集规范族为：

I_0 : $S' \rightarrow \cdot E$
 $E \rightarrow \cdot E + T$
 $E \rightarrow \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_1 : $S' \rightarrow E \cdot$
 $E \rightarrow E \cdot + T$

I_2 : $E \rightarrow T \cdot$
 $T \rightarrow T \cdot * F$

I_3 : $T \rightarrow F \cdot$

I_4 : $F \rightarrow (\cdot E)$
 $E \rightarrow \cdot E + T$
 $E \rightarrow \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_5 : $F \rightarrow i \cdot$

I_6 : $E \rightarrow E + \cdot T$
 $T \rightarrow \cdot T * F$
 $T \rightarrow \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

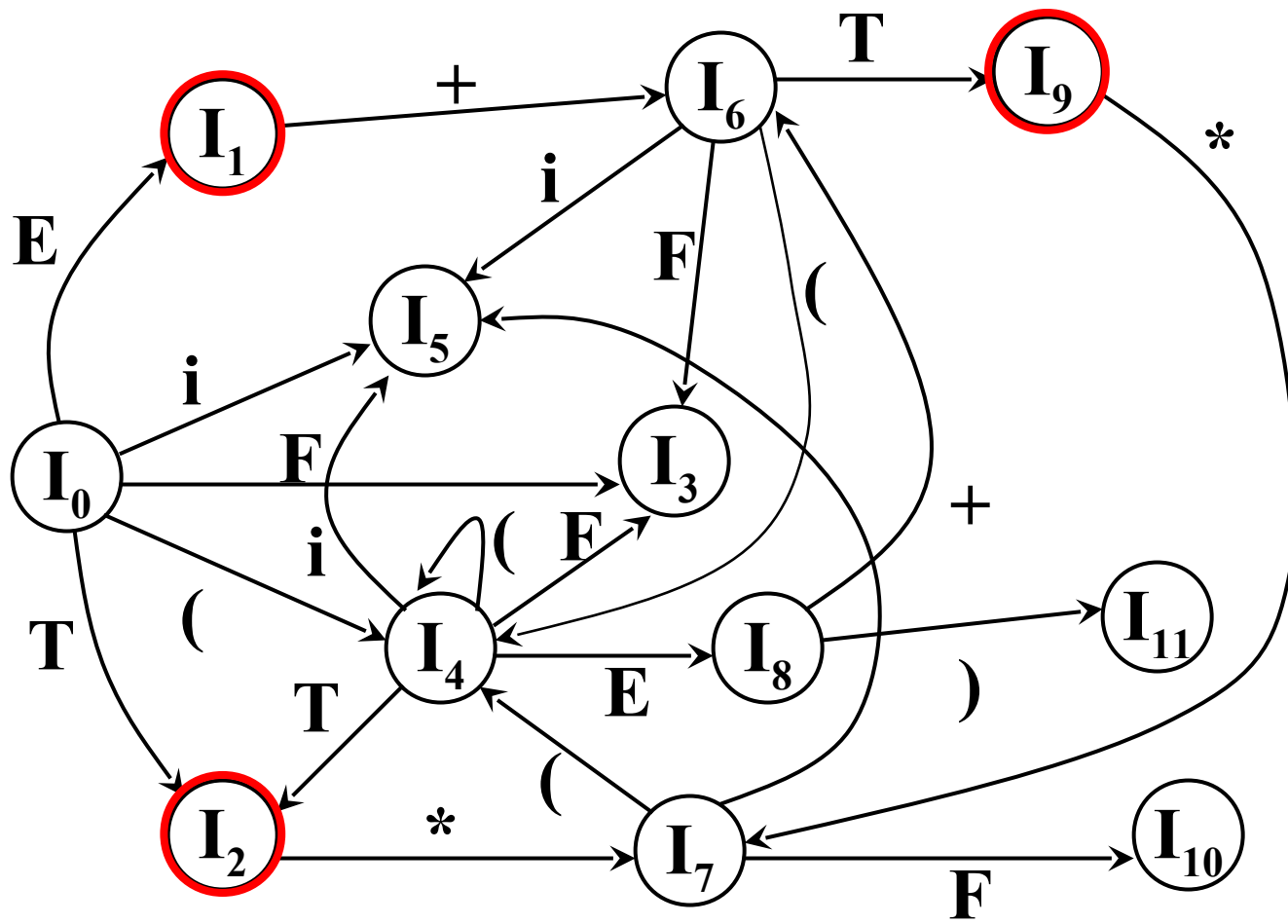
I_7 : $T \rightarrow T * \cdot F$
 $F \rightarrow \cdot (E)$
 $F \rightarrow \cdot i$

I_8 : $F \rightarrow (E \cdot)$
 $E \rightarrow E \cdot + T$

I_9 : $E \rightarrow E + T \cdot$
 $T \rightarrow T \cdot * F$

I_{10} : $T \rightarrow T * F \cdot$

I_{11} : $F \rightarrow (E) \cdot$



■ 这个文法的 LR(0) 项目集规范族为：

$I_0: S' \rightarrow \cdot E$

$E \rightarrow \cdot E + T$

$E \rightarrow \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_1: S' \rightarrow E \cdot$

$E \rightarrow E \cdot + T$

$I_2: E \rightarrow T \cdot$

$T \rightarrow T \cdot * F$

$I_3: T \rightarrow F \cdot$

$I_4: F \rightarrow (\cdot E)$

$E \rightarrow \cdot E + T$

$E \rightarrow \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_5: F \rightarrow i \cdot$

$I_6: E \rightarrow E + \cdot T$

$T \rightarrow \cdot T * F$

$T \rightarrow \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_7: T \rightarrow T * \cdot F$

$F \rightarrow \cdot (E)$

$F \rightarrow \cdot i$

$I_8: F \rightarrow (E \cdot)$

$E \rightarrow E \cdot + T$

$I_9: E \rightarrow E + T \cdot$

$T \rightarrow T \cdot * F$

$I_{10}: T \rightarrow T * F \cdot$

$I_{11}: F \rightarrow (E) \cdot$

- I_1 、 I_2 和 I_9 都含有 “移进 - 归约” 冲突

I_1 : $\overset{\circ}{S'} \rightarrow E \cdot$
 $E \rightarrow E \cdot + T$

I_2 : $E \rightarrow T \cdot$
 $T \rightarrow T \cdot * F$

I_9 : $E \rightarrow E + T \cdot$
 $T \rightarrow T \cdot * F$

采取 SLR (1) 冲突消解

FOLLOW(S') = {#}

$I_1: S' \rightarrow E \cdot$

$E \rightarrow E \cdot + T$

其分析表如下：

状态	ACTION						GOTO		
	i	+	*	()	#	E	T	F
0	s5			s4			1	2	3
1		s6				acc			
2		r2	s7		r2	r2			
3		r4	r4		r4	r4			
4	s5			s4			8	2	3
5		r6	r6		r6	r6			
6	s5			s4				9	3
7	s5			s4					10
8		s6			s11				
9		r1	s7		r1	r1			
10		r3	r3		r3	r3			
11		r5	r5		r5	r5			

**FOLLOW(E) = {#,),
+}**

**$I_2: E \rightarrow T \cdot$
 $T \rightarrow T \cdot * F$**

**$I_9: E \rightarrow E + T \cdot$
 $T \rightarrow T \cdot * F$**

状态	ACTION						GOTO		
	i	+	*	()	#	E	T	F
0	s5			s4			1	2	3
1		s6				acc			
2		r2	s7		r2	r2			
3		r4	r4		r4	r4			
4	s5			s4			8	2	3
5		r6	r6		r6	r6			
6	s5			s4				9	3
7	s5			s4					10
8		s6			s11				
9		r1	s7		r1	r1			
10		r3	r3		r3	r3			
11		r5	r5		r5	r5			

小结

- 根据识别活前缀的 DFA、LR(0) 项目集规范族
 - 构造 LR(0) 分析表
 - 构造 SLR(1) 分析表

作业

- P134—5 (1) , (2) , (3), (4)