



中国科学技术大学
University of Science and Technology of China



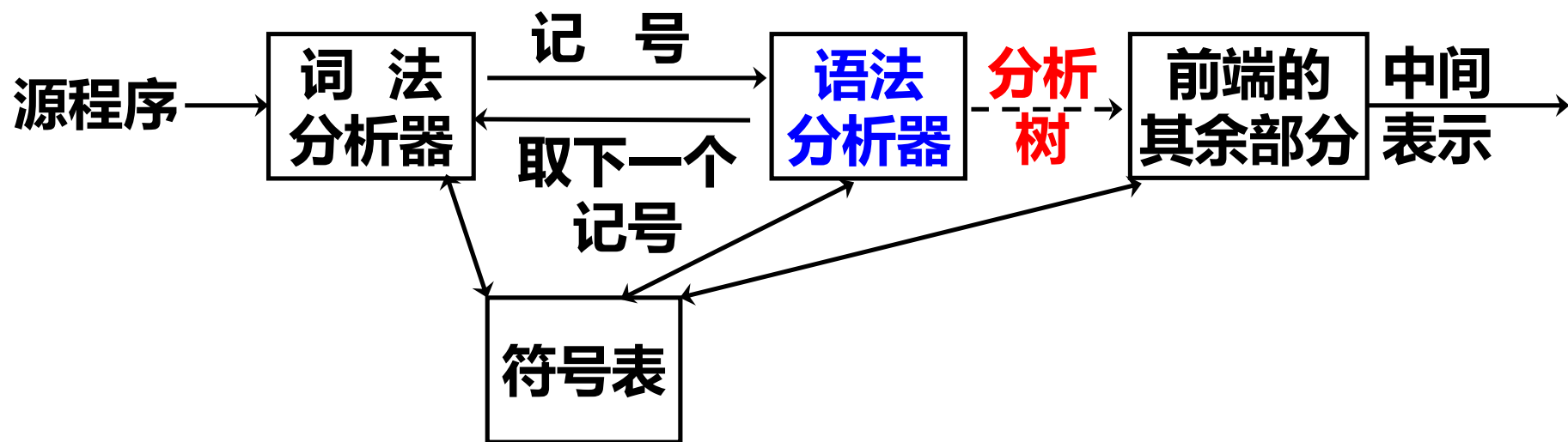
《编译原理与技术》

语法分析 V

计算机科学与技术学院

李 诚

30/09/2019



□LR(k)分析技术

- ❖ 规范的LR方法(LR(1))
- ❖ 向前看的LR方法(LALR)
- ❖ 分析器的生成器(Bison) – 课后Tutorial

□语法分析技术总结



$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S$
 $S \rightarrow V = E$
 $S \rightarrow \cdot E$
 $V \rightarrow \cdot * E$
 $V \rightarrow \cdot \text{id}$
 $E \rightarrow \cdot V$

V

$I_2:$
 $S \rightarrow V \cdot = E$
 $E \rightarrow V \cdot$

$=$

$I_6:$
 $S \rightarrow V = \cdot E$
 $E \rightarrow \cdot V$
 $V \rightarrow \cdot * E$
 $V \rightarrow \cdot \text{id}$



$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S$
 $S \rightarrow V = E$
 $S \rightarrow \cdot E$
 $V \rightarrow \cdot * E$
 $V \rightarrow \cdot \text{id}$
 $E \rightarrow \cdot V$

V

$I_2:$
 $S \rightarrow V \cdot = E$
 $E \rightarrow V \cdot$

$=$

$I_6:$
 $S \rightarrow V = \cdot E$
 $E \rightarrow \cdot V$
 $V \rightarrow \cdot * E$
 $V \rightarrow \cdot \text{id}$

项目 $S \rightarrow V \cdot = E$ 使得
 $\text{action}[2, =] = \text{s6}$

项目 $E \rightarrow V$ 使得
 $\text{action}[2, =] = \text{r5}$
因为 $\text{Follow}(E) = \{=, \$\}$

产生移进-归约冲突，
但该文法不是二义的。



□目标：在识别活前缀DFA的状态中，增加信息，排除一些不正确的归约操作



□目标：在识别活前缀DFA的状态中，增加信息，排除一些不正确的归约操作

□方法：添加了前向搜索符

❖一个项目 $A \rightarrow \alpha \beta$ ，如果最终用这个产生式进行归约之后，期望看见的符号是 a ，则这个加点项的前向搜索符是 a 。

❖上述项目可以写成： $A \rightarrow \alpha \beta, a$



□目标：在识别活前缀DFA的状态中，增加信息，排除一些不正确的归约操作

□方法：添加了前向搜索符

❖一个项目 $A \rightarrow \alpha \beta$ ，如果最终用这个产生式进行归约之后，期望看见的符号是 a ，则这个加点项的前向搜索符是 a 。

❖上述项目可以写成： $A \rightarrow \alpha \beta, a$

□与SLR(1)分析的区别

❖项目集的定义发生了改变

❖closure(I) 和GOTO函数需要修改



□LR(1)项目:

$$[A \rightarrow \alpha \cdot \beta, a]$$

- ❖ 当项目由两个分量组成，第一分量为SLR中的项，第二分量为搜索符（向前看符号）
- ❖ LR(1)中的1代表了搜索符 a 的长度



□LR(1)项目:

$$[A \rightarrow \alpha \cdot \beta, a]$$

- ❖ 当项目由两个分量组成，第一分量为SLR中的项，第二分量为搜索符（向前看符号）
- ❖ LR(1)中的1代表了搜索符 a 的长度

□使用注意事项:

- ❖ 当 β 不为空时， a 不起作用
- ❖ 当 β 为空时，如果下一个输入符号是 a ，将按照 $A \rightarrow \alpha$ 进行归约



□LR(1)项目:

$$[A \rightarrow \alpha \cdot \beta, a]$$

- ❖当项目由两个分量组成，第一分量为SLR中的项，第二分量为搜索符（向前看符号）
- ❖LR(1)中的1代表了搜索符 a 的长度

□LR(1)项目 $[A \rightarrow \alpha \cdot \beta, a]$ 对活前缀 γ 有效:

- ❖如果存在着推导 $S \Rightarrow_{rm}^* \delta A w \Rightarrow_{rm} \delta \alpha \beta w$ ，其中：
 - $\gamma = \delta \alpha$;
 - a 是 w 的第一个符号，或者 w 是 ϵ 且 a 是 $\$$



□例 $S \rightarrow BB$

$B \rightarrow bB \mid a$

从最右推导 $S \Rightarrow_{rm}^* bbBba \Rightarrow_{rm} bbbBba$ 看出：

$[B \rightarrow b \cdot B, b]$ 对活前缀 $\gamma = bbb$ 是有效的



□例 $S \rightarrow BB$

$B \rightarrow bB \mid a$

从最右推导 $S \Rightarrow_{rm}^* bbBba \Rightarrow_{rm} bb**b**B**b**a$ 看出：

令 $A = B$, $\alpha = b$, $\beta = B$, $\delta = bb$, $\gamma = \delta\alpha = bbb$, $w = ba$

$[B \rightarrow b \cdot B, b]$ 对活前缀 $\gamma = bbb$ 是有效的



□构造LR(1)项目集规范族

❖也就是构造识别活前缀的DFA

□构造规范的LR分析表

❖状态之间的转换关系



□基础运算1：计算闭包CLOSURE(I)

- ❖ I中的任何项目都属于CLOSURE(I)
- ❖ 若有项目 $[A \rightarrow \alpha B \beta, a]$ 在CLOSURE(I)中，而 $B \rightarrow \gamma$ 是文法中的产生式， b 是FIRST(βa)中的元素，则 $[B \rightarrow \gamma, b]$ 也属于CLOSURE(I)

保证在用 $B \rightarrow \gamma$ 进行归约后，

- 出现的输入字符 b 是句柄 $\alpha B \beta$ 中 B 的后继符号
- 或者是 $\alpha B \beta$ 归约为 A 后可能出现的终结符。



□基础运算2：通过GOTO(I,X)算CLOSURE(J)

- ❖ 将J置为空集
- ❖ 若有项目 $[A \rightarrow \alpha X \beta, a]$ 在I中，那么将项目 $[A \rightarrow \alpha X \beta, a]$ 放入J中
- ❖ 计算并返回CLOSURE(J)

注意：GOTO(I,X)中的X可以是终结符或非终结符



□具体算法

❖ 初始项目集 I_0 :

$I_0 = \text{CLOSURE}([S' \rightarrow S, \$])$ 将\$作为向前的搜索符

❖ 设C为最终返回的项目集族, 初始为 $C = \{I_0\}$

❖ 重复以下步骤

➤ 对C中的任意项目集I, 重复

- 对每一个文法符号X(终结符或非终结符)

- 如果 $\text{GOTO}(I, X) \neq \emptyset$ 且 $\text{GOTO}(I, X) \notin C$, 那么将 $\text{GOTO}(I, X)$ 放入C

- 注: 上述 $\text{GOTO}(I, X)$ 是上一页ppt中计算闭包的GOTO

➤ 当C中项目集不再增加为止



构造LR(1)项目集族： 举例



中国科学技术大学
University of Science and Technology of China

$S' \rightarrow S, \$$ I_0

步骤一：从初始项开始



构造LR(1)项目集族：举例



中国科学技术大学
University of Science and Technology of China

$S' \rightarrow S, \$$ I_0
 $S \rightarrow BB$

步骤二：计算非核心项目的第一个分量



构造LR(1)项目集族：举例



中国科学技术大学
University of Science and Technology of China

$S' \rightarrow S, \$$ I_0
 $S \rightarrow BB, \$$

步骤三：通过FIRST($\epsilon \$$) 计算非核心项目的第二个分量



$S' \rightarrow S, \$$ $S \rightarrow BB, \$$ $B \rightarrow bB$ $B \rightarrow a$	I_0
---------------------------------------------------------------------------------------------	-------

步骤二：计算非核心项目的
第一个分量



构造LR(1)项目集族：举例



中国科学技术大学
University of Science and Technology of China

$S' \rightarrow S, \$$ I_0

$S \rightarrow BB, \$$

$B \rightarrow bB, b/a$

$B \rightarrow a, b/a$

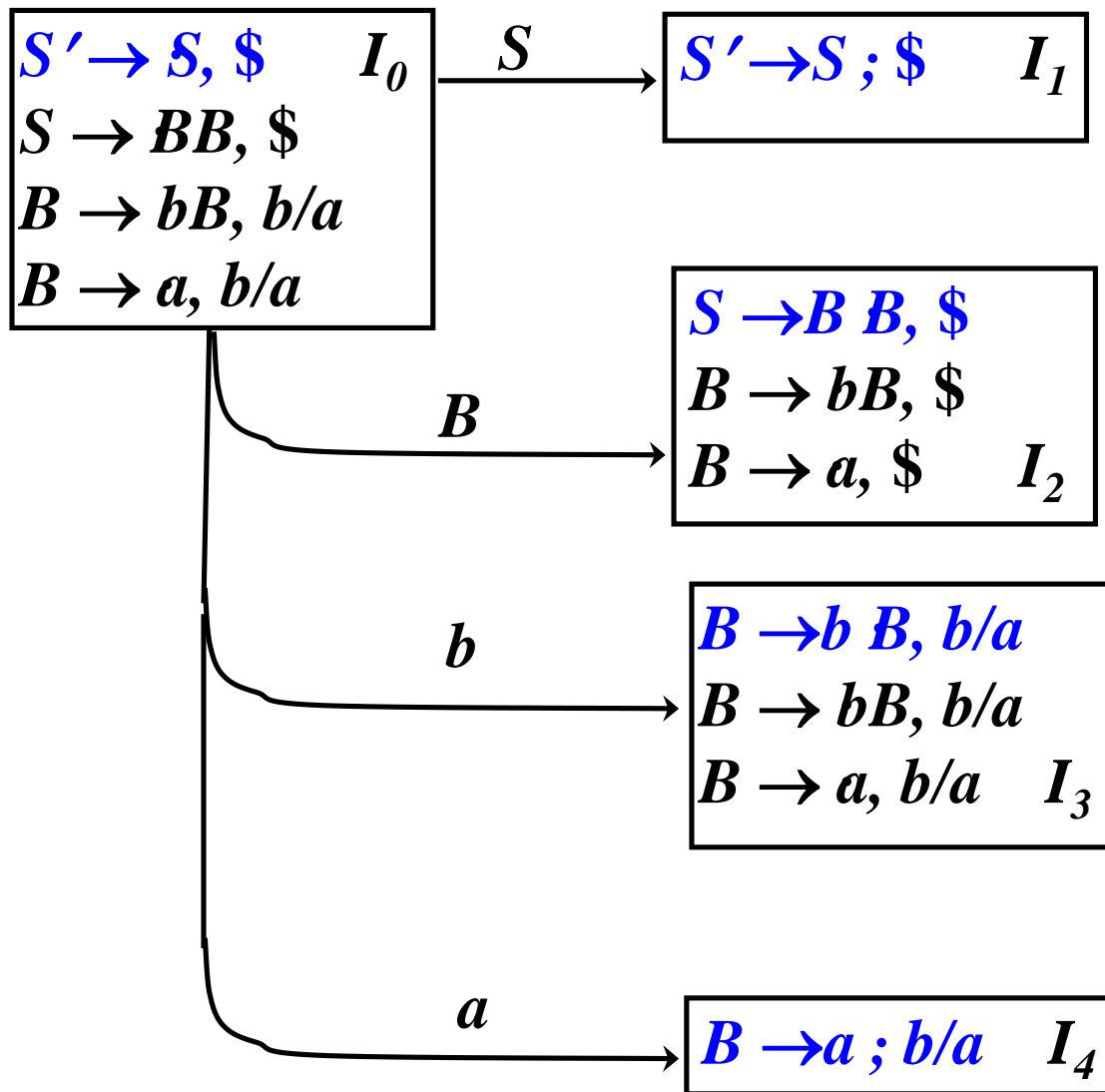
步骤三：通过FIRST(B\$) 计算非核心项目的第二个分量



构造LR(1)项目集族：举例



中国科学技术大学
University of Science and Technology of China

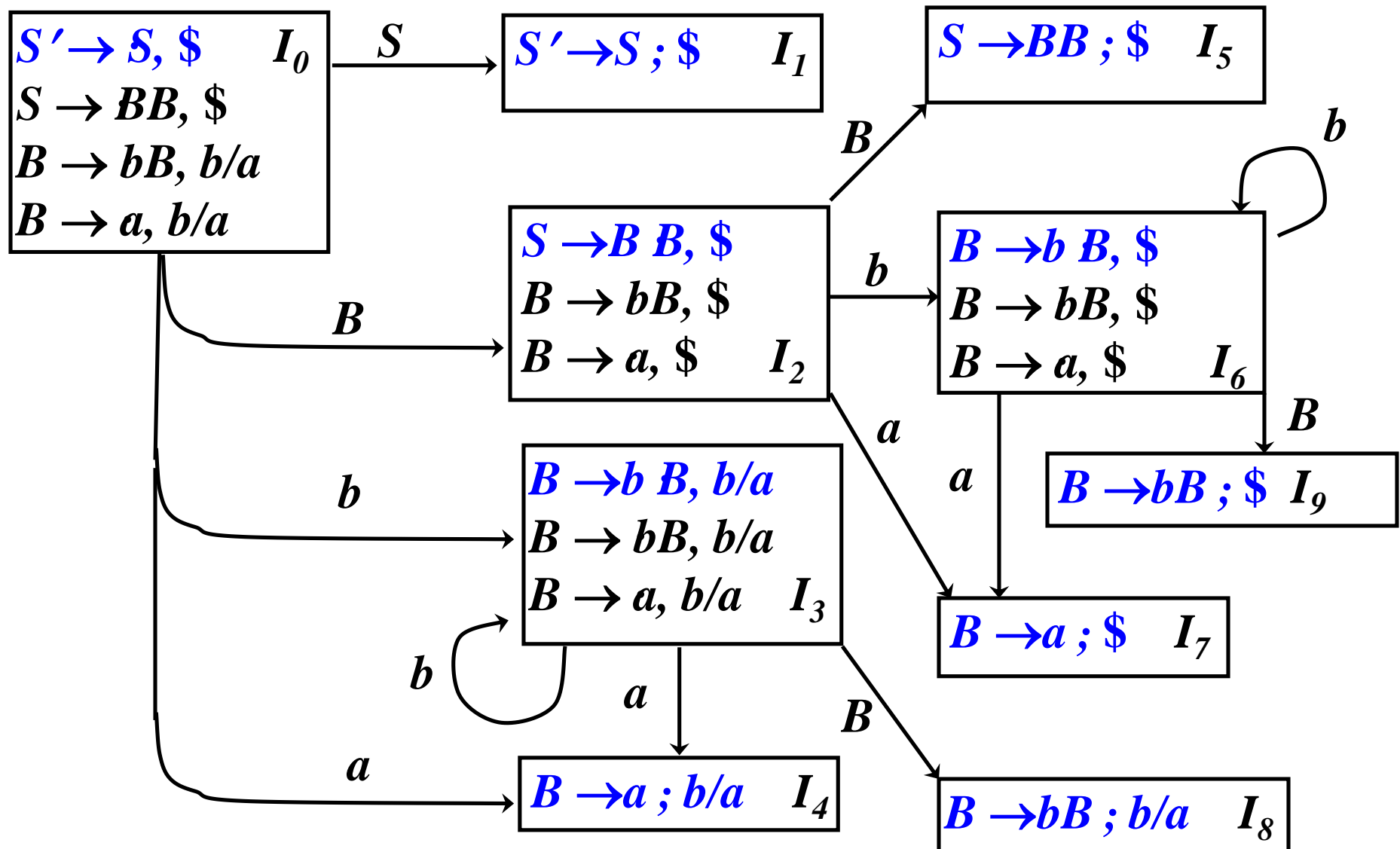




构造LR(1)项目集族：举例



中国科学技术大学
University of Science and Technology of China





□构造识别拓广文法 G' 活前缀的DFA

❖基于LR(1)项目族来构造

□状态 i 的 $action$ 函数如下确定:

❖如果 $[A \rightarrow \alpha \ a\beta, b]$ 在 I_i 中, 且 $goto(I_i, a) = I_j$, 那么置 $action[i, a]$ 为 sj (此时, 不看 b)

❖如果 $[A \rightarrow \alpha \cdot, a]$ 在 I_i 中, 且 $A \neq S'$, 那么置 $action[i, a]$ 为 rj

❖如果 $[S' \rightarrow S ; \$]$ 在 I_i 中, 那么置 $action[i, \$] = acc$

如果上述构造出现了冲突, 那么文法就不是LR(1)的



□构造识别拓广文法 G' 活前缀的DFA

□状态 i 的 $action$ 函数如下确定：

❖参见上页ppt

□状态 i 的 $goto$ 函数如下确定：

❖如果 $goto(I_i, A) = I_j$, 那么 $goto[i, A] = j$



□构造识别拓广文法 G' 活前缀的DFA

□状态 i 的 $action$ 函数如下确定:

❖参见上页ppt

□状态 i 的 $goto$ 函数如下确定:

❖如果 $goto(I_i, A) = I_j$, 那么 $goto[i, A] = j$

□分析器的初始状态是包含 $[S' \rightarrow \cdot S, \$]$ 的项目集对应的状态

用上面规则未能定义的所有条目都置为**error**

$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S$
 $S \rightarrow V = E$
 $S \rightarrow \cdot E$
 $V \rightarrow * E$
 $V \rightarrow \cdot \text{id}$
 $E \rightarrow \cdot V$

V

$I_2:$
 $S \rightarrow V \cdot = E$
 $E \rightarrow V \cdot$

$=$

$I_6:$
 $S \rightarrow V = \cdot E$
 $E \rightarrow \cdot V$
 $V \rightarrow * E$
 $V \rightarrow \cdot \text{id}$

项目 $S \rightarrow V \cdot = E$ 使得
 $\text{action}[2, =] = \text{s6}$

项目 $E \rightarrow V$ 使得
 $\text{action}[2, =] = \text{r5}$
 因为 $\text{Follow}(E) = \{=, \$\}$

产生移进-归约冲突，
但该文法不是二义的。



非SLR(1)但是LR(1)文法



中国科学技术大学
University of Science and Technology of China

$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S, \$$
 $S \rightarrow V = E, \$$
 $S \rightarrow \cdot E, \$$
 $V \rightarrow \cdot * E, =$
 $V \rightarrow \cdot \text{id}, =$
 $E \rightarrow \cdot V, \$$
 $V \rightarrow \cdot * E, \$$
 $V \rightarrow \cdot \text{id}, \$$

$V \longrightarrow$

$I_2:$
 $S \rightarrow V \cdot = E, \$$
 $E \rightarrow V \cdot, \$$

无移进归约冲突

计算闭包:

定义里: $[A \rightarrow \alpha B \beta, a]$

这里: $[S \rightarrow E, \$]$

$\text{FIRST}(\beta a)$



$\text{FIRST}(\epsilon \$) = \{\$ \}$

注: 十分感谢王灿、范睿两位同学帮助我纠正对于该问题的讲解错误。



非SLR(1)但是LR(1)文法



中国科学技术大学
University of Science and Technology of China

$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S, \$$
 $S \rightarrow V = E, \$$
 $S \rightarrow \cdot E, \$$
 $V \rightarrow * E, =/\$$
 $V \rightarrow \cdot \text{id}, =/\$$
 $E \rightarrow \cdot V, \$$

$V \longrightarrow$

$I_2:$
 $S \rightarrow V \cdot = E, \$$
 $E \rightarrow V \cdot, \$$

可通过合并搜索符简化

计算闭包:

定义里: $[A \rightarrow \alpha B \beta, a]$

这里: $[S \rightarrow E, \$]$

$\text{FIRST}(\beta a)$



$\text{FIRST}(\epsilon \$) = \{ \$ \}$



非SLR(1)但是LR(1)文法



中国科学技术大学
University of Science and Technology of China

$S \rightarrow V = E$
 $S \rightarrow E$
 $V \rightarrow * E$
 $V \rightarrow \text{id}$
 $E \rightarrow V$

$I_0:$
 $S' \rightarrow \cdot S, \$$
 $S \rightarrow V = E, \$$
 $S \rightarrow \cdot E, \$$
 $V \rightarrow * E, =/\$$
 $V \rightarrow \cdot \text{id}, =/\$$
 $E \rightarrow \cdot V, \$$

$V \longrightarrow$

$I_2:$
 $S \rightarrow V \cdot = E, \$$
 $E \rightarrow V ; \$$

可通过合并搜索符简化

计算闭包:

定义里: $[A \rightarrow \alpha B \beta, a]$

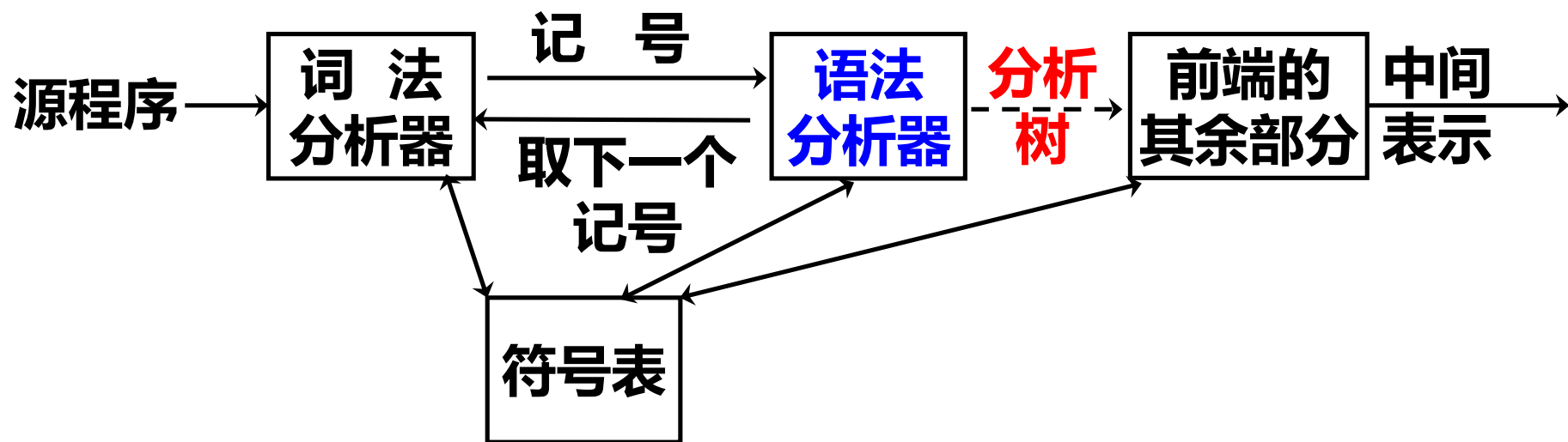
这里: $[S \rightarrow E, \$]$

$\text{FIRST}(\beta a)$



$\text{FIRST}(\epsilon \$) = \{ \$ \}$

每一个SLR(1)文法都是LR(1)的



□LR(k)分析技术

- ❖ 规范的LR方法(LR(1))
- ❖ 向前看的LR方法(LALR)
- ❖ 分析器的生成器(Bison) – 课后Tutorial

□语法分析技术总结

□研究LALR的原因

规范LR分析表的状态数偏多

□LALR特点

- ❖ LALR和SLR的分析表有同样多的状态，比规范LR分析表要小得多
- ❖ LALR的能力介于SLR和规范LR之间
- ❖ LALR的能力在很多情况下已经够用

□LALR分析表构造方法

- ❖ 通过合并规范LR(1)项目集来得到



□ 合并识别 LR(1)文法的活前缀的DFA中的相同
核心项目集(**同心项目集**)

□ 同心的LR(1)项目集

❖ **核心：项目集中第一分量的集合**

❖ 略去搜索符后它们是相同的集合

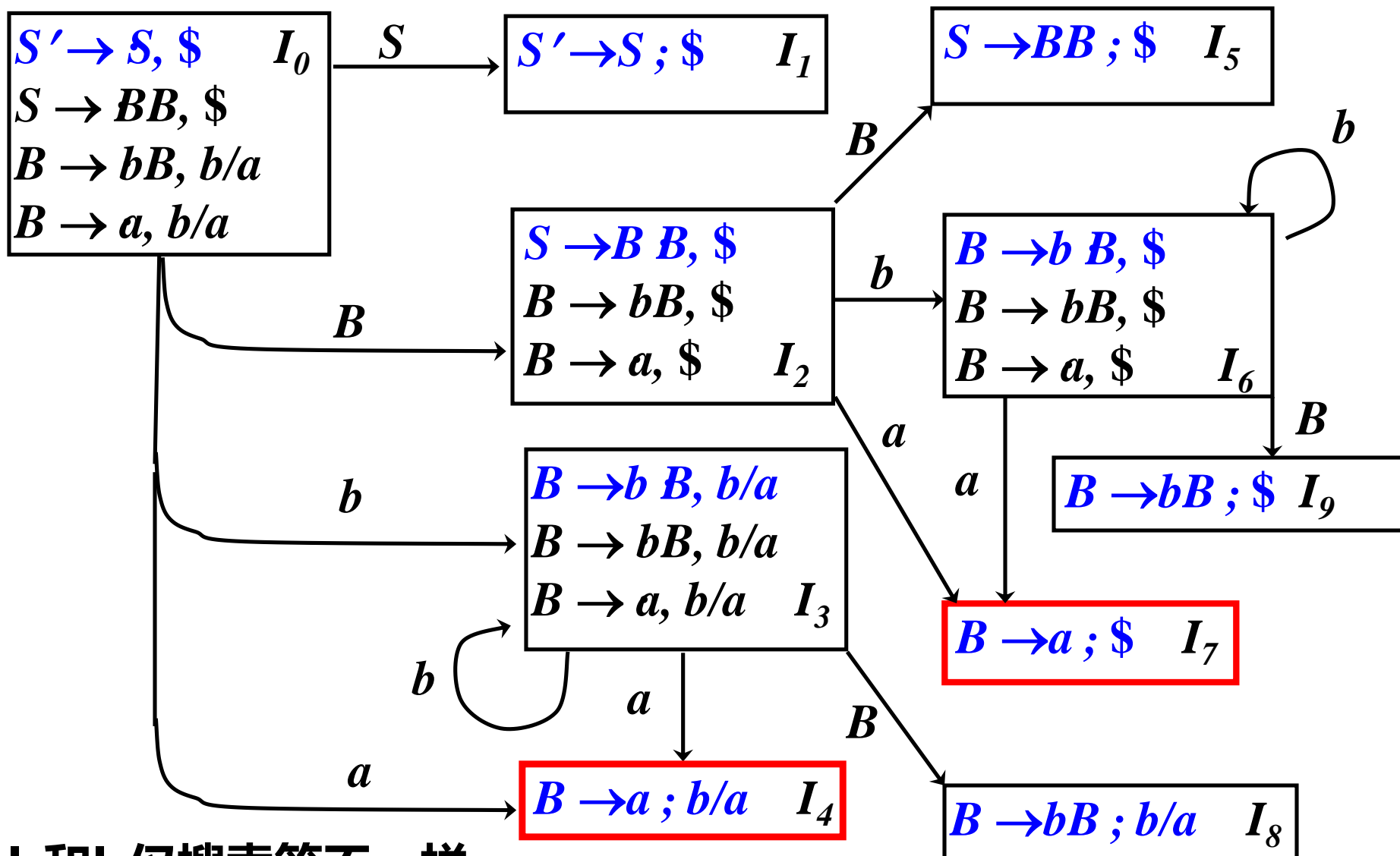
❖ 例：[**B** → **bB**, **\$**] 与 [**B** → **bB**, **b/a**]

第一分量

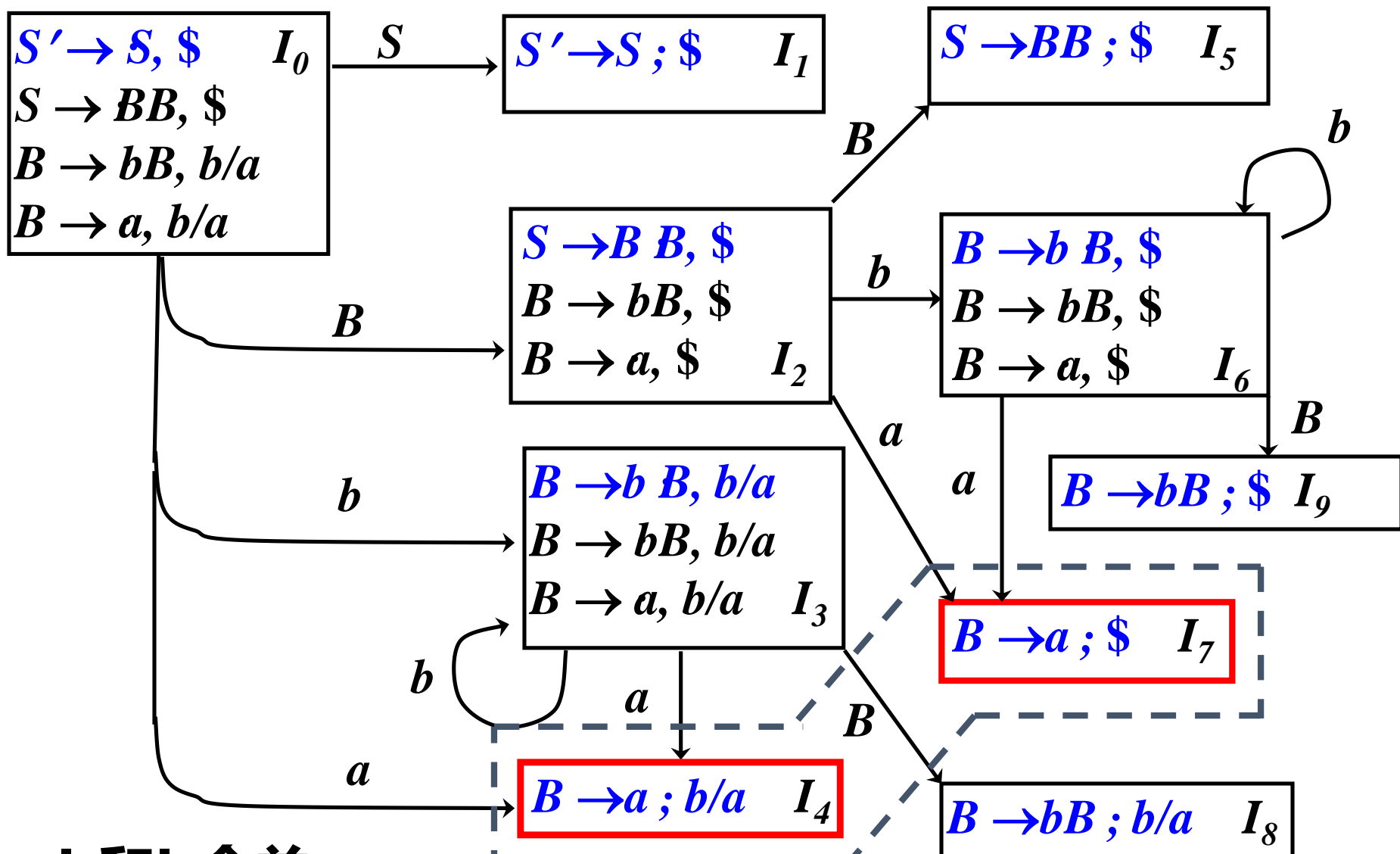
第二分量

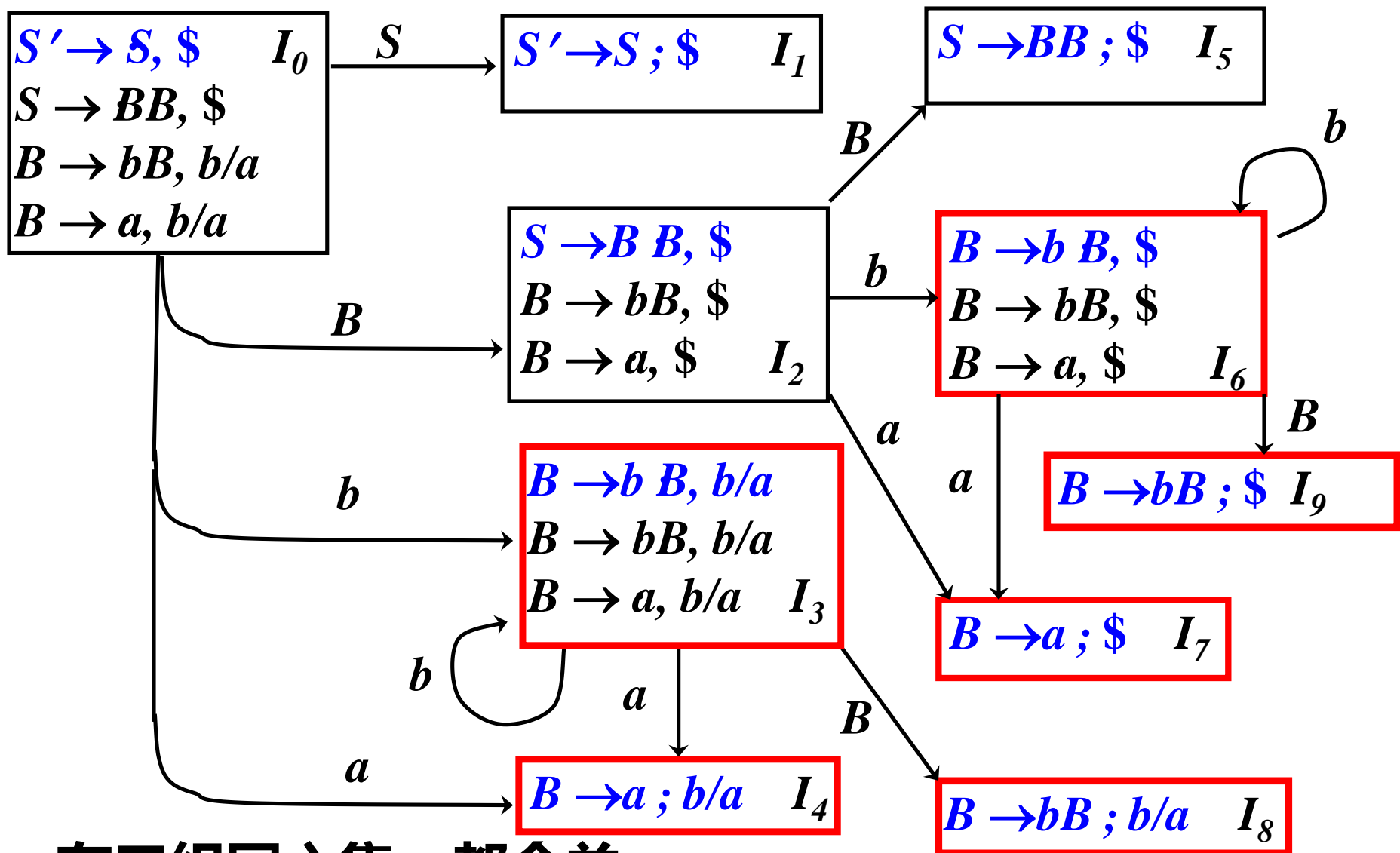


识别活前缀的DFA

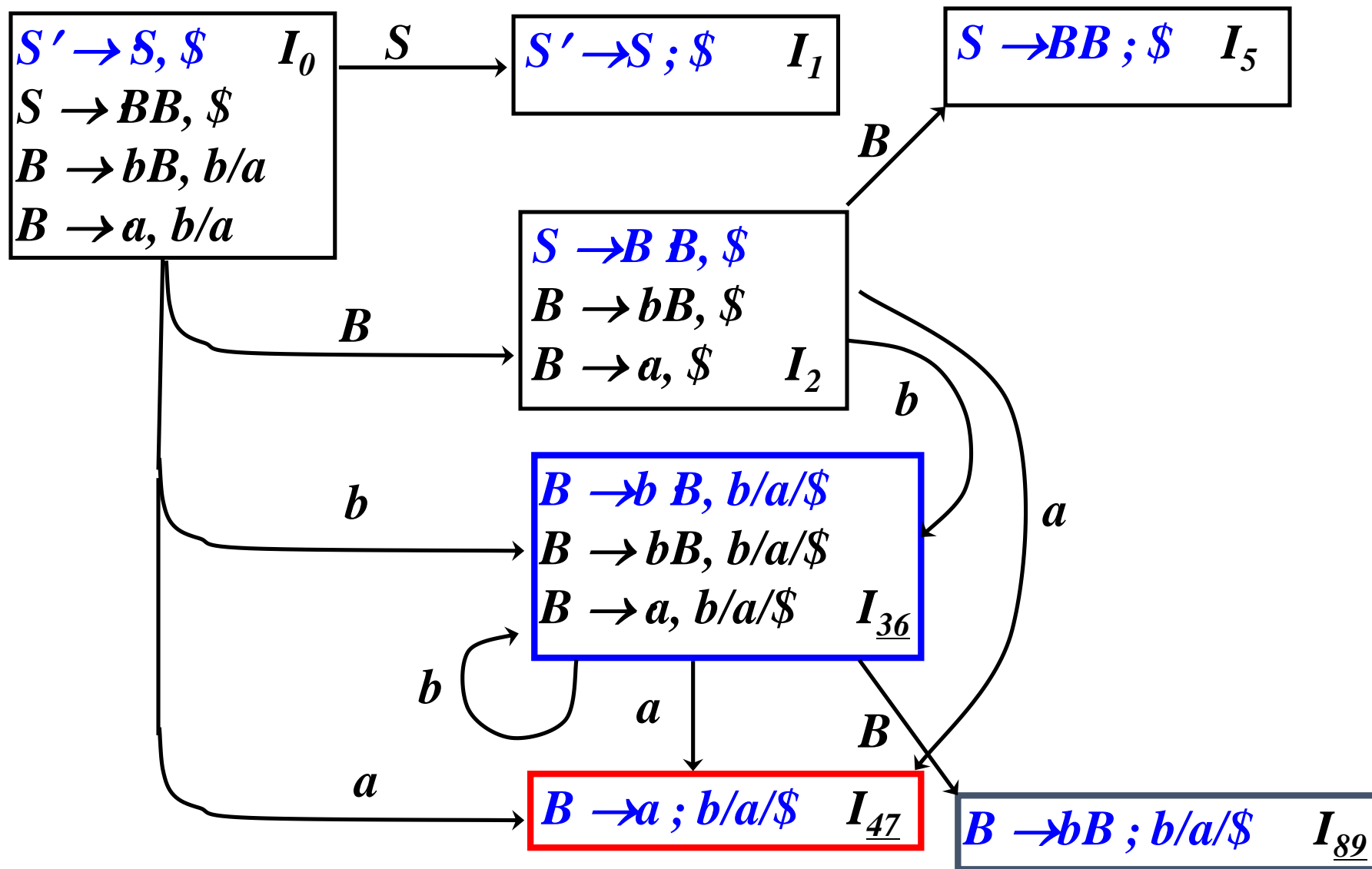


I_4 和 I_7 仅搜索符不一样



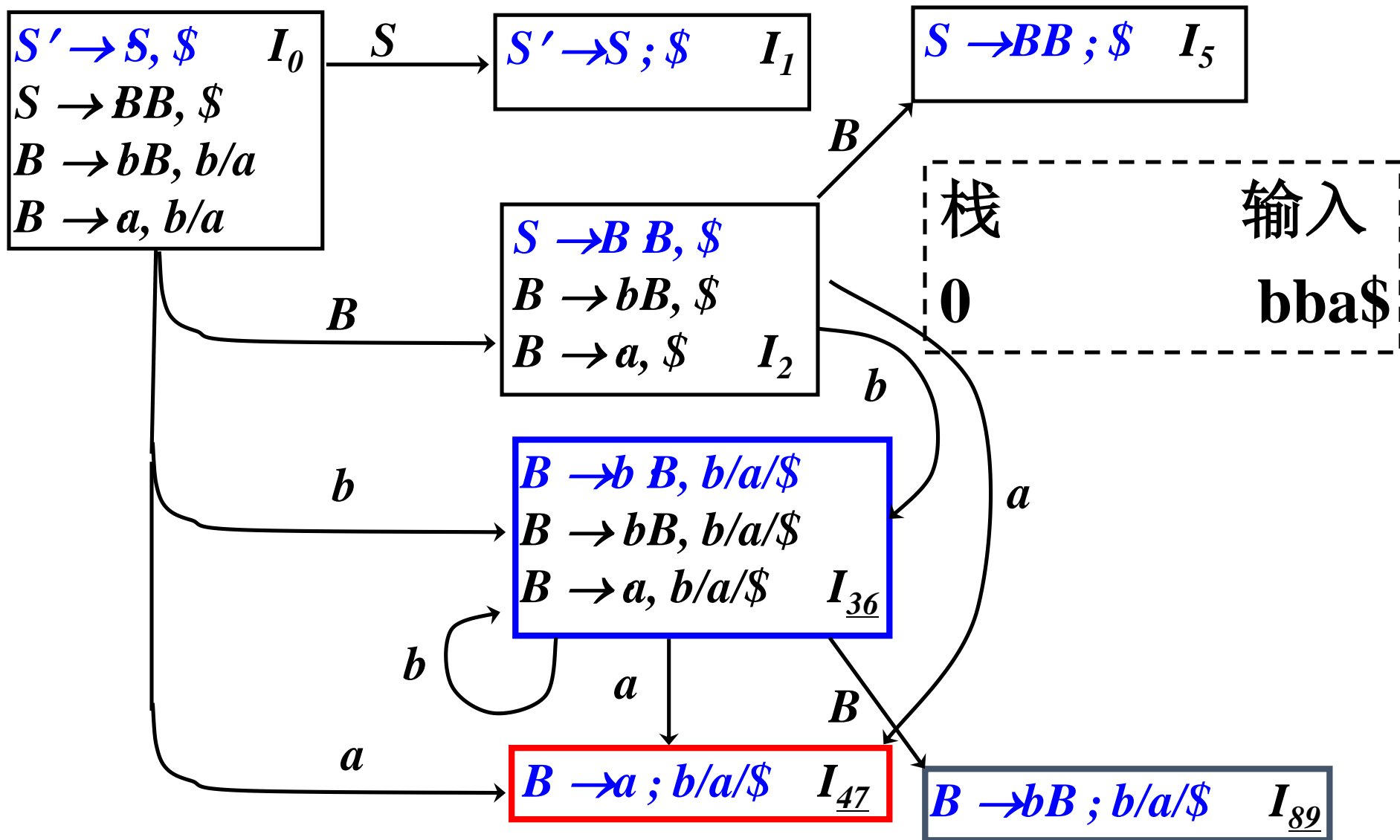


有三组同心集，都合并



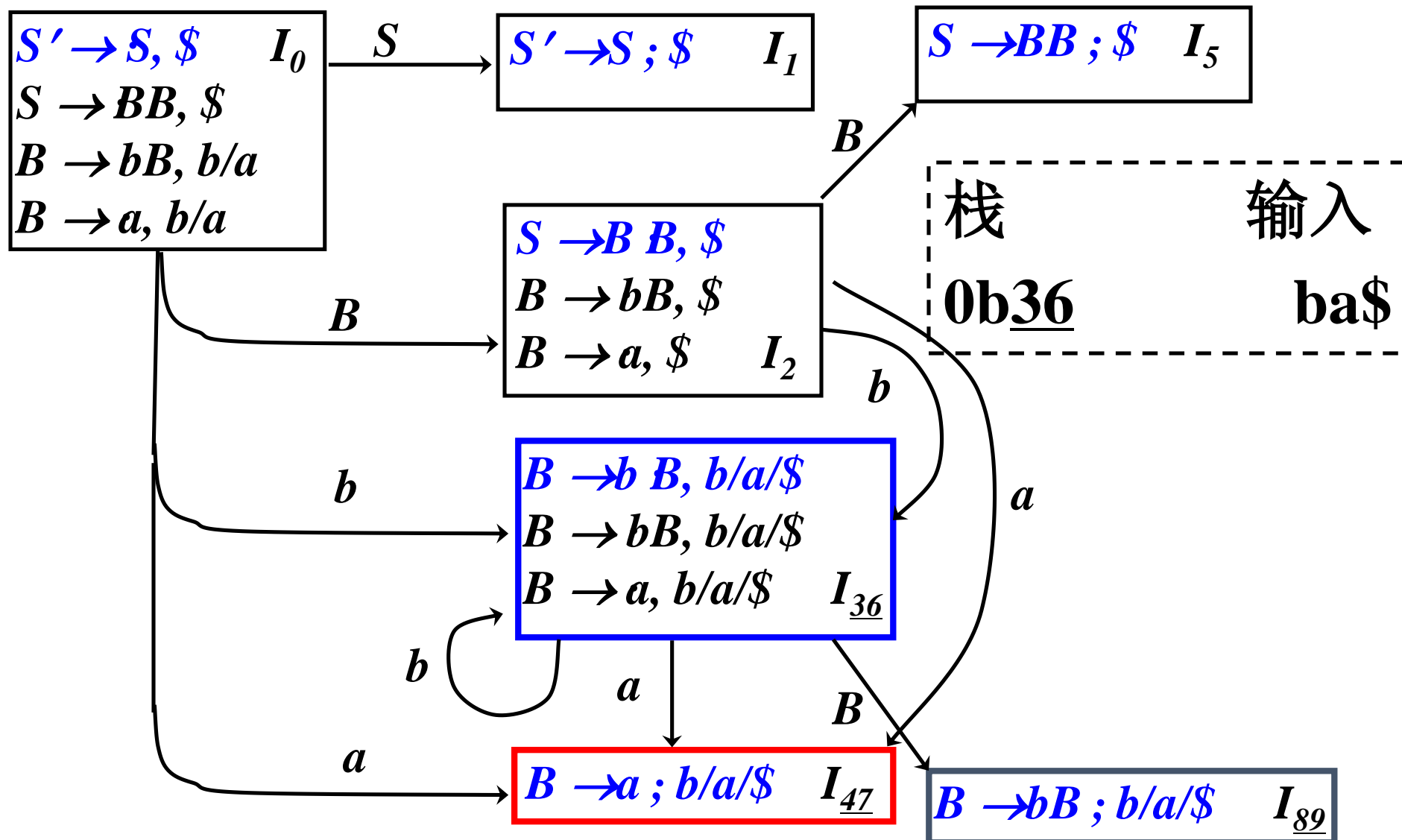


LALR分析：举例



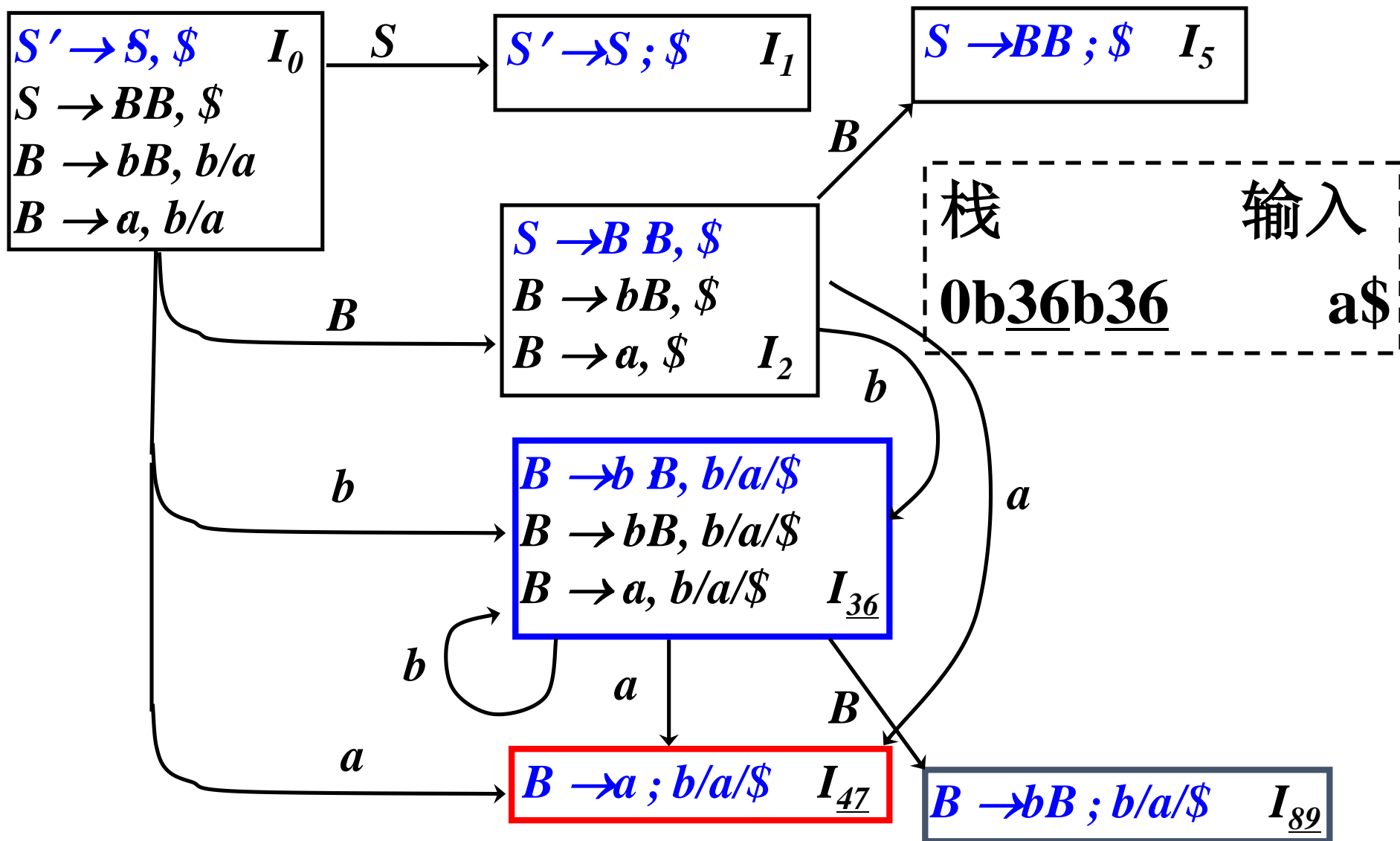


LALR分析：举例



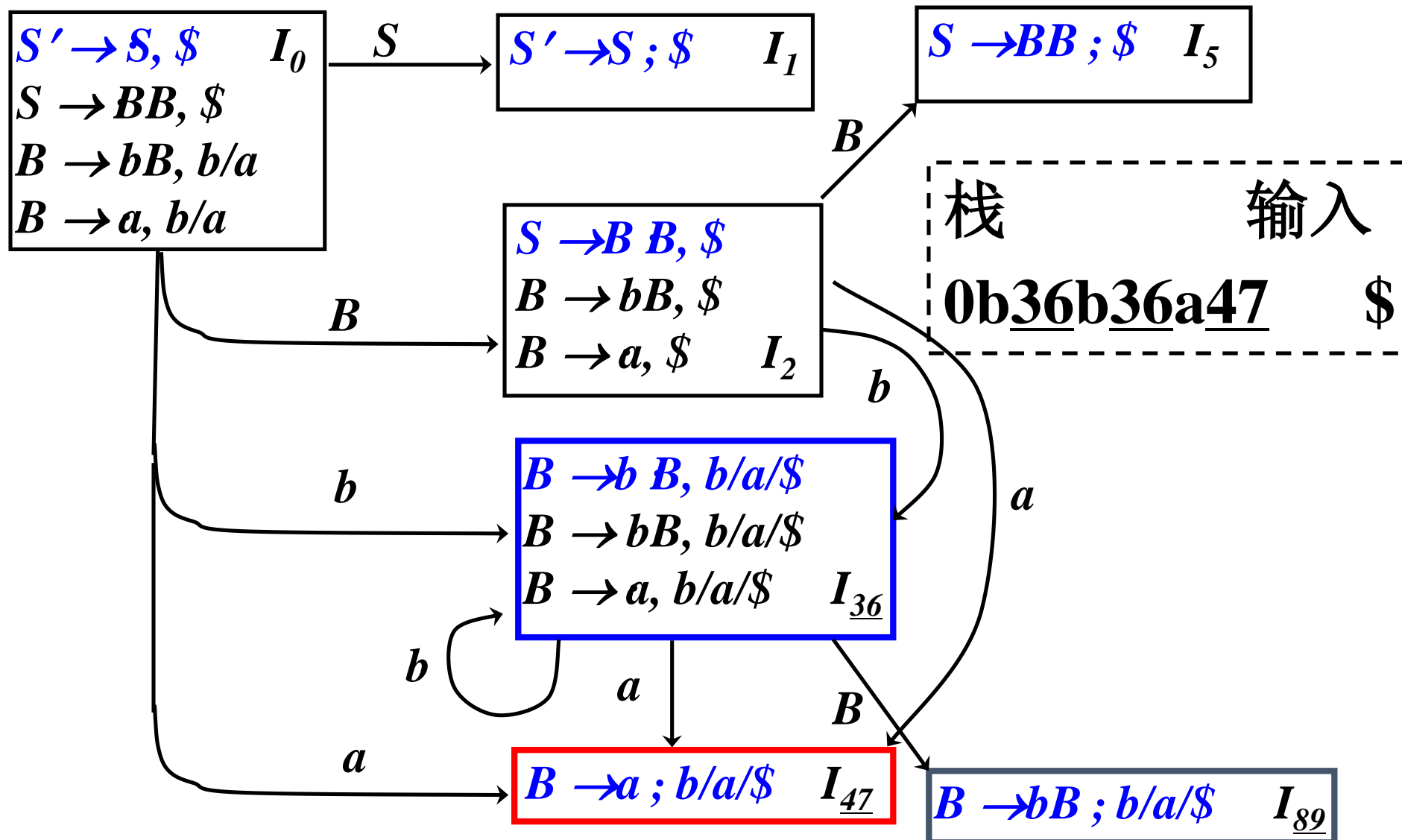


LALR分析：举例



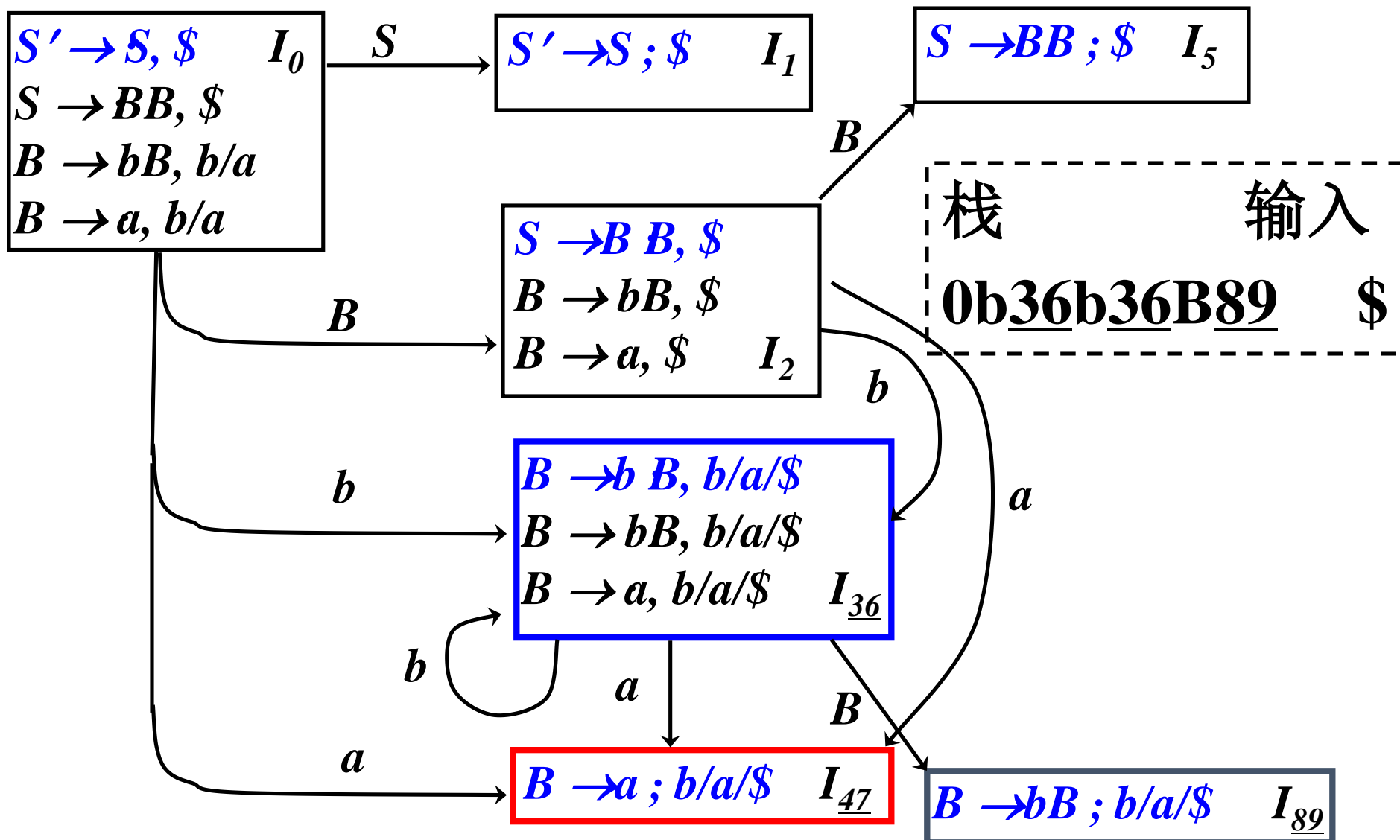


LALR分析：举例



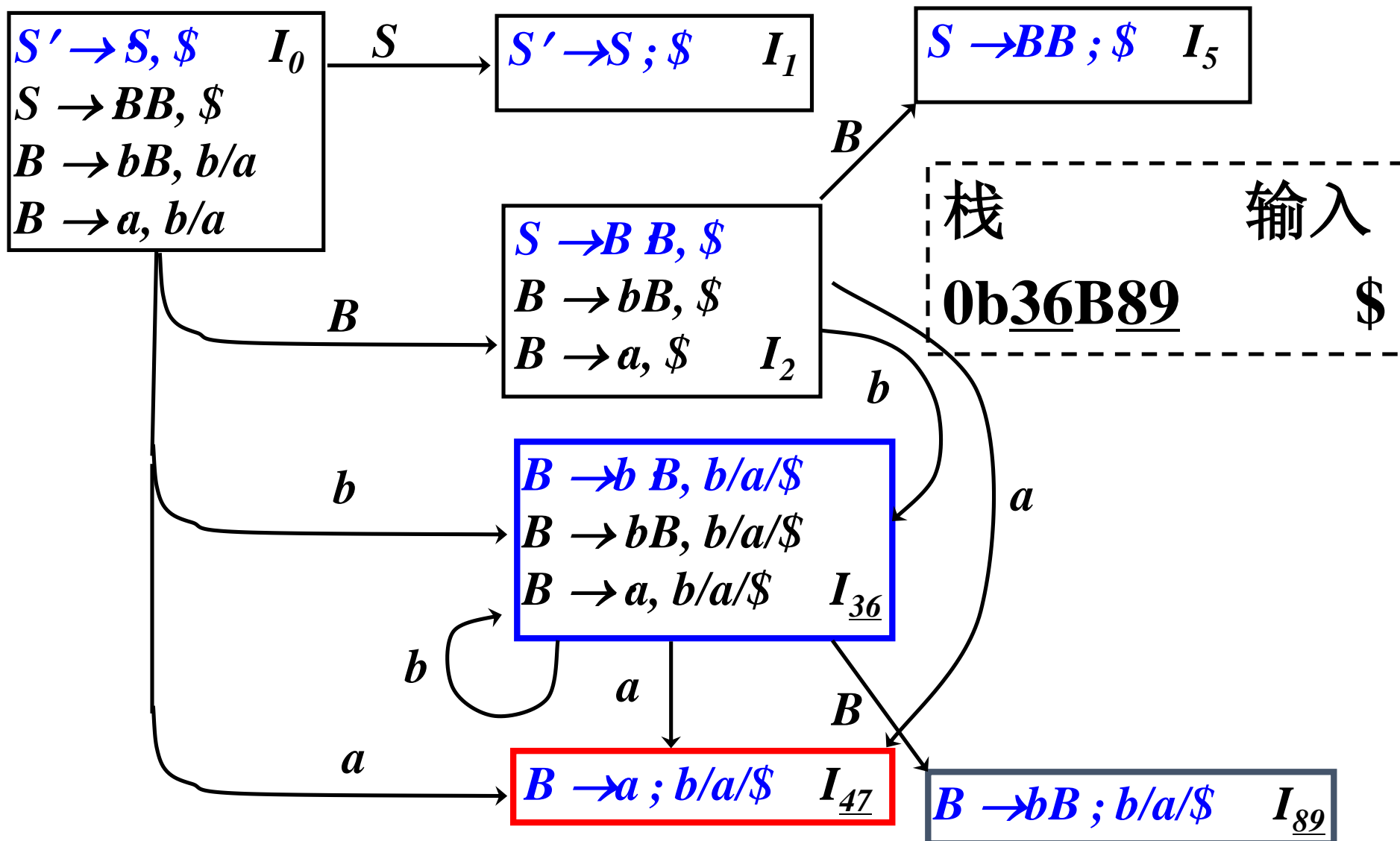


LALR分析：举例



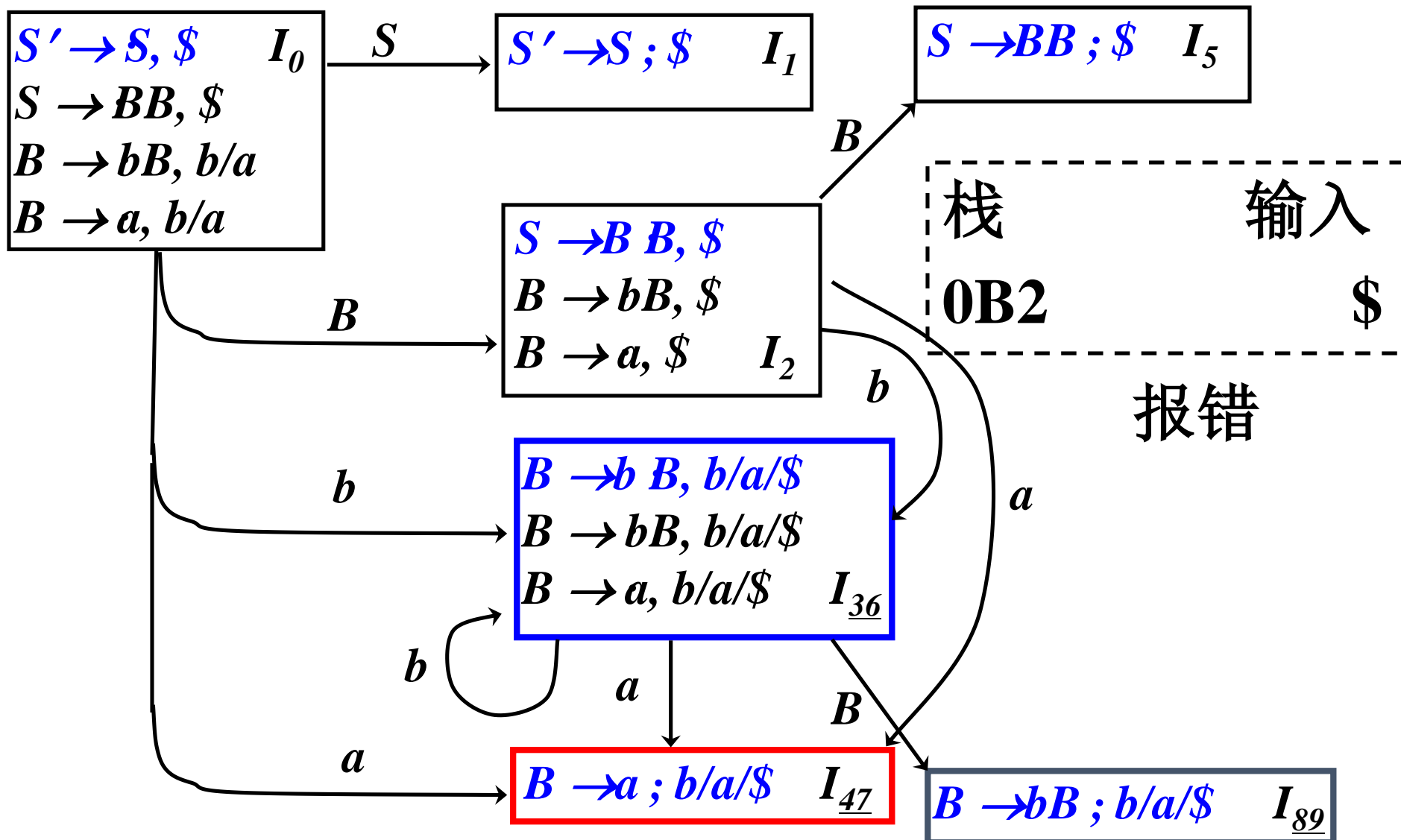


LALR分析：举例





LALR分析：举例





2、构造LALR(1)分析表

- ❖ 构造LR(1)项目集规范族 $C = \{I_0, I_1, \dots, I_n\}$
- ❖ 构造LALR(1)项目集规范族 $C' = \{J_0, J_1, \dots, J_k\}$,
其中任意项目集 $J_i = I_n \cup I_m \cup \dots \cup I_t$
 - $I_n, I_m, \dots, I_t \in C$ 且具有共同的核心
- ❖ 按构造规范LR(1)分析表的方式构造分析表

如没有语法分析动作冲突，那么给定文法就是
LALR(1)文法

□合并同心项目集可能会引起冲突

❖同心集的合并不会引起新的移进-归约冲突

项目集1

$[A \rightarrow \alpha ; a]$

$[B \rightarrow \beta a \gamma, c]$

...

项目集2

$[B \rightarrow \beta a \gamma, b]$

$[A \rightarrow \alpha ; d]$

...

如果有移进归约冲突，则合并前就有冲突



□合并同心项目集可能会引起冲突

- ❖同心集的合并不会引起新的移进-归约冲突
- ❖同心集的**合并有可能产生**新的归约-归约冲突

$S' \rightarrow S$

$S \rightarrow aAd / bBd /$
 aBe / bAe

$A \rightarrow c$

$B \rightarrow c$

对 ac 有效的项目集

$A \rightarrow c ; d$
 $B \rightarrow c ; e$

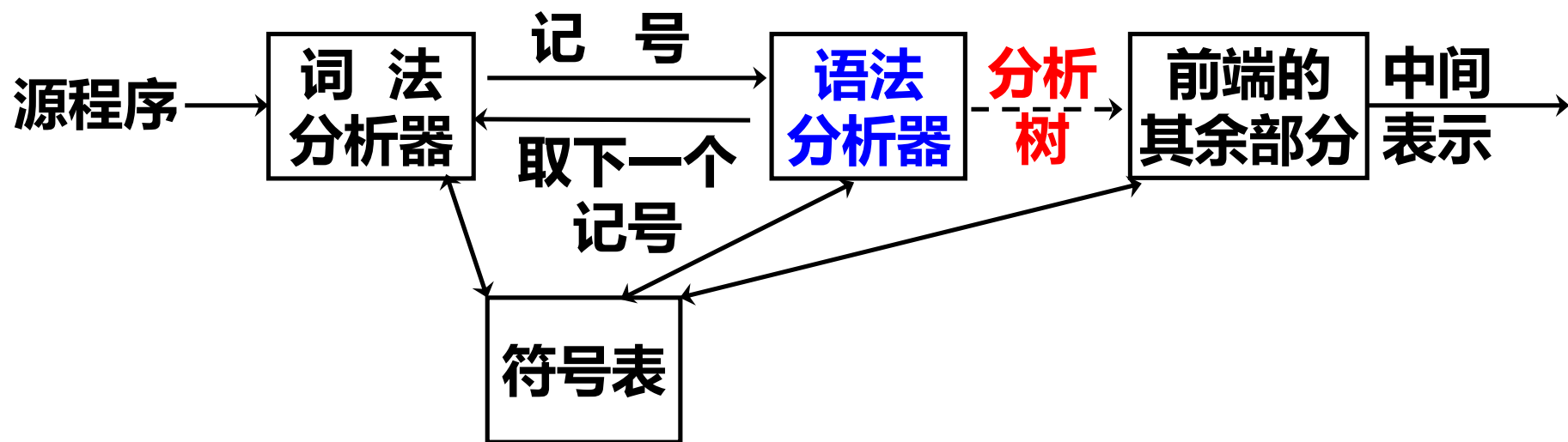
对 bc 有效的项目集

$A \rightarrow c ; e$
 $B \rightarrow c ; d$

合并同心集后

$A \rightarrow c ; d/e$
 $B \rightarrow c ; d/e$

该文法是LR(1)的
但不是LALR(1)的



□LR(k)分析技术

- ❖ 规范的LR方法(LR(1))
- ❖ 向前看的LR方法(LALR)
- ❖ 分析器的生成器(Bison) – 课后Tutorial

□语法分析技术总结



LR分析法总结



中国科学技术大学
University of Science and Technology of China

		SLR	LALR	LR(1)
初始状态		$[S' \rightarrow S]$	$[S' \rightarrow S, \$]$	$[S' \rightarrow S, \$]$
项目集		LR(0) CLOSURE(I)	合并LR(1)项目集 族的同心项目集	LR(1), CLOSURE(I) 搜索符考虑 FISRT (βa)
动作	移进	$[A \rightarrow \alpha a \beta] \in I_i$ $GOTO(I_i, a) = I_j$ ACTION $[i, a] = sj$	与LR(1)一致	$[A \rightarrow \alpha a \beta, b] \in I_i$ $GOTO(I_i, a) = I_j$ ACTION $[i, a] = sj$
	归约	$[A \rightarrow \alpha] \in I_i, A \neq S'$ $a \in FOLLOW(A)$ ACTION $[i, a] = rj$	与LR(1)一致	$[A \rightarrow \alpha; a] \in I_i$ $A \neq S'$ ACTION $[i, a] = rj$
	接受	$[S' \rightarrow S \cdot] \in I_i$ ACTION $[i, \$] = acc$	与LR(1)一致	$[S' \rightarrow S; \$] \in I_i$ ACTION $[i, \$] = acc$
	出错	空白条目	与LR(1)一致	空白条目
GOTO		$GOTO(I_i, A) = I_j$ GOTO $[i, A] = j$	与LR(1)一致	$GOTO(I_i, A) = I_j$ GOTO $[i, A] = j$
状态量		少(几百)	与SLR一样	多(几千)



上下文无关文法

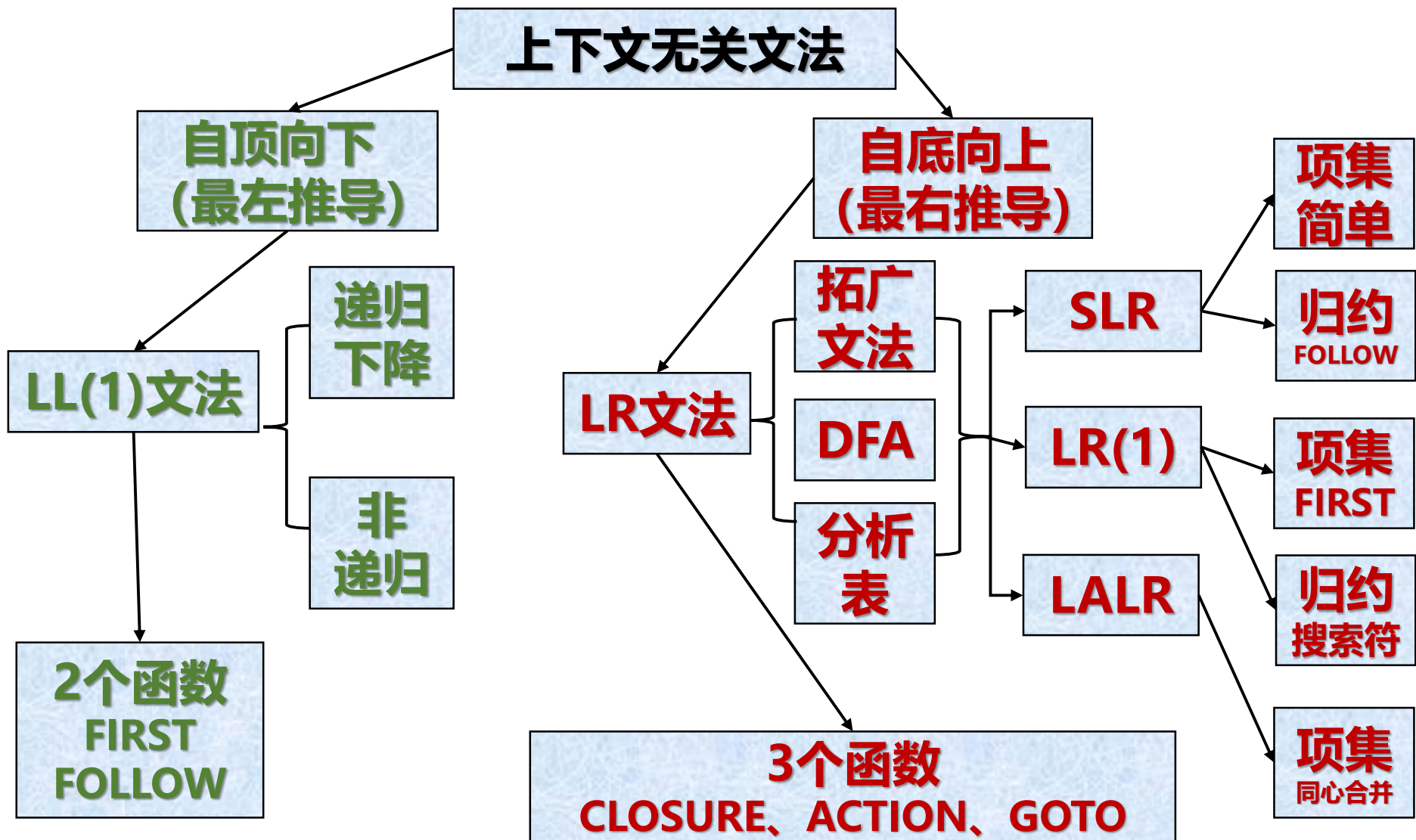
自顶向下
(最左推导)

LL(1)文法

递归
下降

非
递归

2个函数
FIRST
FOLLOW





LR和LL分析方法的比较



中国科学技术大学
University of Science and Technology of China

	LR(1)方 法	LL(1)方 法
建立分析树	自底而上	自顶而下
归约or推导	规范归约	最左推导
决定使用产生式的时机	看见产生式整个右部推出的串后(句柄)	看见产生式推出的第一个终结符后
对文法的限制	无	无左递归、无公共左因子
分析表	状态 \times 文法符号, 大	非终结符 \times 终结符, 小
分析栈	状态栈, 信息更多	文法符号栈
确定句柄	根据栈顶状态和下一个符号便可以确定句柄和归约所用产生式	无句柄概念
语法错误	决不会将出错点后的符号移入分析栈	和LR一样, 决不会读过出错点而不报错



□**颁奖词**: For significant contributions in the design and theory of compilers, the architecture of large systems and the development of **reduced instruction set computers (RISC)**; for discovering and systematizing many fundamental transformations now used in optimizing compilers including reduction of operator strength, **elimination of common subexpressions**, **register allocation**, **constant propagation**, and **dead code elimination**.

❖ https://amturing.acm.org/award_winners/cocke_2083115.cfm

□**演讲**:

❖ THE SEARCH FOR PERFORMANCE IN SCIENTIFIC PROCESSORS

❖ https://dl.acm.org/ft_gateway.cfm?id=1283945&type=pdf



杜克大学学士、博士
供职于IBM T.J.
Watson Research
Center



中国科学技术大学
University of Science and Technology of China



《编译原理与技术》

语法分析 V

谋定而后动，知止而有得！

—— 孙子兵法 计篇