

第三章：语法分析

LR方法
LR(0) 自动机

1.1 LR方法基本思想

符号栈 $\# \alpha$ 有终极符也有非终极符

输入流 $\beta \#$ 终极符

- ◆ 假如要分析的串没有语法错误，则 $\alpha\beta$ 一定是文法的一个句型
- ◆ LR方法的主要思想是，从输入流依此把符号移入符号栈，直至栈顶出现一个句柄；之后对句柄进行归约，直至栈顶不出现句柄；重复上述过程，直至最终归约为一个开始符，且输入流为空。

1.2 LR类分析方法相关定义

- ◆ 规范句型：用最右推导导出的句型（也称右句型）。
- ◆ 规范前缀：若存在规范句型 $\alpha\beta$ ，且 β 是终极符串或空串，则称 α 为规范前缀。
- ◆ 规范活前缀：若某一规范句型的规范前缀 α 不含句柄或含一个句柄并且具有形式 $\alpha=\alpha'\pi$ （ π 是句柄），则称规范前缀 α 为规范活前缀（简称活前缀）。
- ◆ 归约规范活前缀：若 α 是含句柄的规范前缀，且句柄在 α 的最右端，即有 $\alpha=\alpha'\pi$ ，且 π 是句柄，则称活前缀 α 为归约规范活前缀（简称可归前缀）。

1.3 LR类分析方法的关键问题

- ◆ 如何判断分析栈内容是否为可归前缀
- ◆ 能唯一的确定可归前缀中的句柄

1.4 可归前缀的判断

派生定理：

- ◆ 开始符产生式的右部是可归前缀。
- ◆ 如果 $\alpha_1 A \alpha_2$ 是可归前缀，且 $A \rightarrow \beta$ 是产生式，则 $\alpha_1 \beta$ 也是可归前缀。

设有文法 $G[S]$ ：

$S \rightarrow aAc$ [1]

$A \rightarrow dbb$ [2]

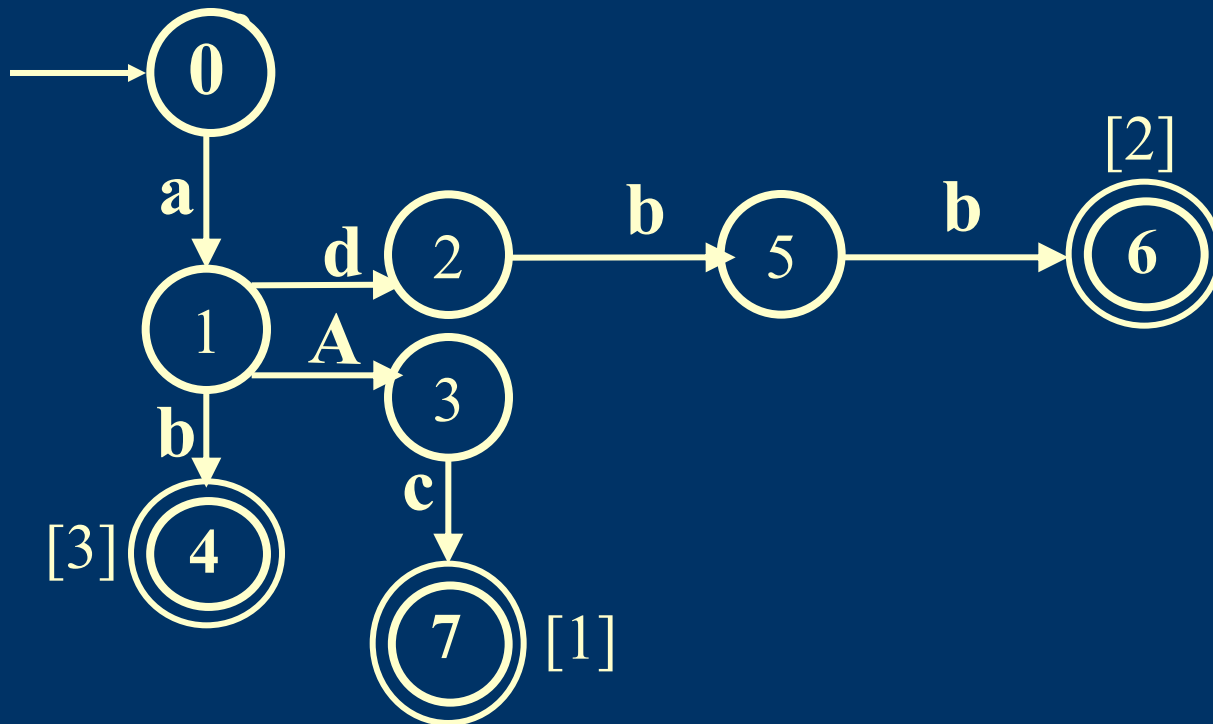
$A \rightarrow b$ [3]

可归前缀为： aAc , $adbb$, ab

1.5 识别归约活前缀的自动机

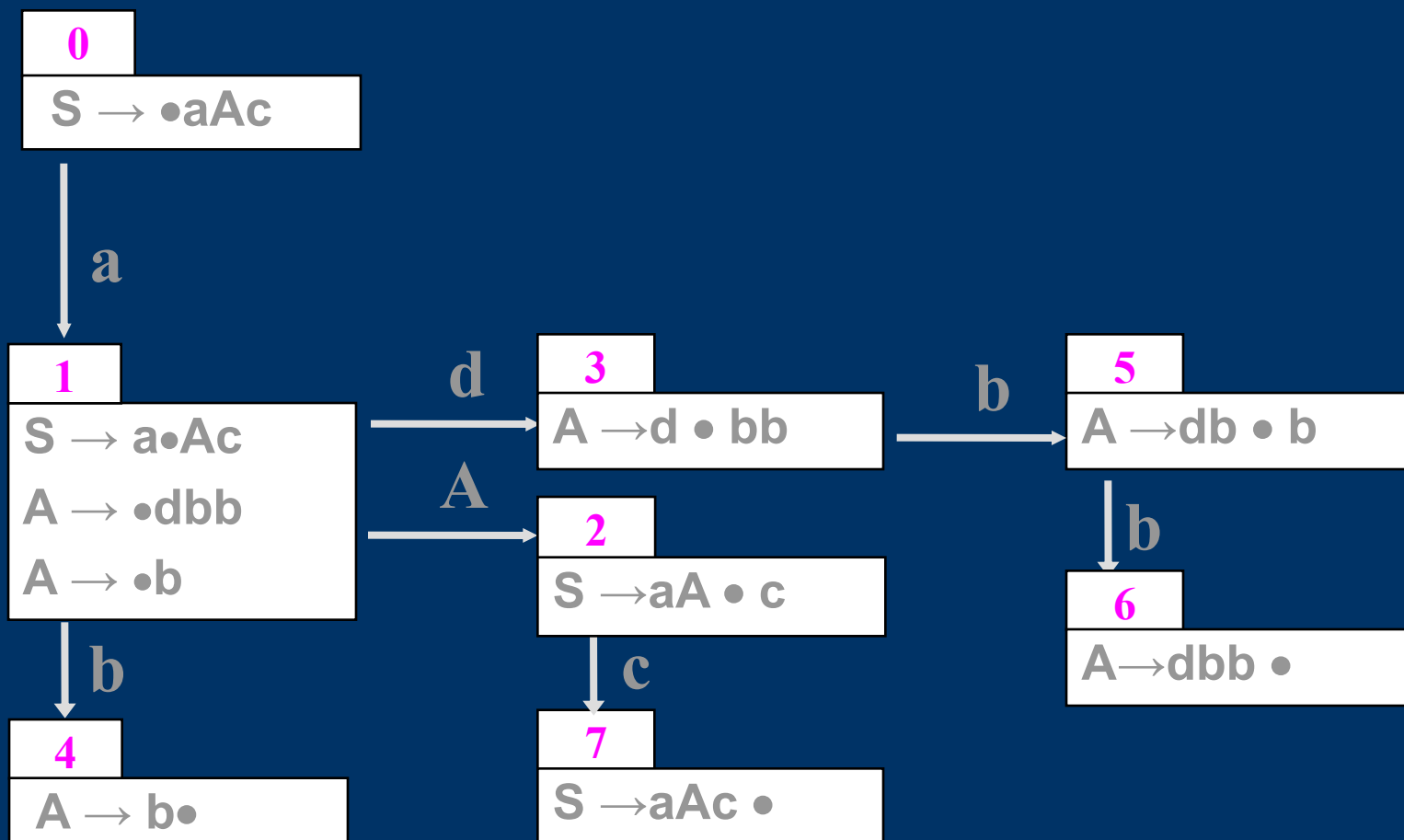
- ◆ 识别可归前缀的自动机表示为：

aAc, adbb, ab



加强版识别可归前缀的自动机表示为：

aAc, adbb, ab



2. LR(0)分析法基本概念

LR(0) 项目：

- ◆ LR(0) 项目是带一个圆点'•'的产生式。每个项目都标志分析时的某一状态。

项目的含义：分析过程中某一个步骤的状态。

例：

$S \rightarrow aAc$ 对应四个LR(0)项目：

- | | | |
|----|-----------------------------|----------------|
| 1. | $S \rightarrow \bullet aAc$ | 符号栈中不包含句柄的任何部分 |
| 2. | $S \rightarrow a\bullet Ac$ | 符号栈中包含句柄中的a |
| 3. | $S \rightarrow aA\bullet c$ | 符号栈中包含句柄中的aA |
| 4. | $S \rightarrow aAc\bullet$ | 符号栈中包含完整句柄aAc |

2. LR(0)分析法基本概念

- ◆ 项目集的投影：假设 IS 是 LR(0) 项目集，则称下面 $IS_{(X)}$ 为 IS 关于 X 的投影集：

$$IS_{(X)} = \{A \rightarrow \alpha X \bullet \beta \mid A \rightarrow \alpha \bullet X \beta \in IS, X \in (V_T \cup V_N)\}$$

项目集的投影 IS 关于 X 的投影集的含义：项目集中每一个项目所描述的状态，处理完一个符号后所对应的状态。

2. LR(0)分析法基本概念

例: $G[S]$:

$$S \rightarrow aAc[1]$$
$$A \rightarrow dbb[2]$$
$$A \rightarrow b[3]$$

令项目集 $IS = \{S \rightarrow a \bullet Ac[1], A \rightarrow \bullet dbb[2], A \rightarrow \bullet b[3]\}$

则: $IS_{(A)} = \{S \rightarrow aA \bullet c[1]\}$

$$IS_{(S)} = \{ \}$$
$$IS_{(d)} = \{A \rightarrow d \bullet bb[3]\}$$
$$IS_{(b)} = \{A \rightarrow b \bullet [3]\}$$
$$IS_{(a)} = \{ \}$$
$$IS_{(c)} = \{ \}$$

2. LR(0)分析法基本概念

项目集的闭包：假设IS是LR(0)项目集，则称下面
CLOSURE (IS) 为IS的闭包集：

CLOSURE (IS)

$$= IS \cup \{A \rightarrow \bullet \pi \mid Y \rightarrow \beta \bullet A \eta \in \text{CLOSURE}(IS), \\ A \rightarrow \pi \text{ 是产生式} \}$$

闭包的含义：满足某个条件的所有东西。项目集的闭包则表示分析到了项目集中每一个项目所描述的状态，还需要做的所有归约所用的产生式。

2. LR(0)分析法基本概念

例: $G[S]$:

$$S \rightarrow aAc[1]$$
$$A \rightarrow dbb[2]$$
$$A \rightarrow b[3]$$

令项目集 $IS = \{S \rightarrow a \bullet Ac[1]\}$

则: $CLOSURE(\{S \rightarrow a \bullet Ac[1]\})$

$$= \{S \rightarrow a \bullet Ac[1], A \rightarrow \bullet dbb[2], A \rightarrow \bullet b[3]\}$$

2. LR(0)分析法基本概念

- ◆ 项目集的转换函数（GO函数）：若IS是一个LR(0)项目集， $X \in (V_T \cup V_N)$ ，函数 $GO(IS, X)$ 定义为：

$$GO(IS, X) = \text{CLOSURE}(IS_{(X)}),$$

其中 $IS_{(X)}$ 为LR(0)项目集IS关于X的投影。

项目集的转换函数的含义：项目集中每一个项目所描述的状态，当处理完一个符号后还需做的所有归约。

2. LR(0)分析法基本概念

例: $G[S]$:

$$S \rightarrow aAc[1]$$
$$A \rightarrow dbb[2]$$
$$A \rightarrow b[3]$$

令项目集 $IS = \{S \rightarrow \bullet aAc[1]\}$

则: $GO(IS, a) = CLOSURE(IS_{(a)})$

$$= CLOSURE(\{S \rightarrow a\bullet Ac[1]\})$$
$$= \{S \rightarrow a\bullet Ac[1], A \rightarrow \bullet dbb[2], A \rightarrow \bullet b[3]\}$$

3. 构造LR(0)活前缀状态机LRSM₀

- ✧ 为了使“成功”状态易于识别，通常LR文法要求文法的开始符唯一且不出现于产生式的右部，因此要增加一个新的产生式 $Z \rightarrow S$ （称拓广产生式），其中S是原文法的开始符，而Z则是新符号。
- ✧ LRSM₀可以识别所有的可归活前缀
- ✧ LRSM₀中的每个状态将对应一个饱和项目集：
 - (1) 其中一部分是由先驱状态分出来（称为基本项目）；
 - (2) 一部分则是由基本项目扩展出来的（称为扩展项目或派生项目）。派生部分项目的特点是其中的“•”出现在产生式右部的最左侧。

3. 构造LR(0)活前缀状态机LRSM₀

- ◆ Step1. 构造初始状态IS₀: IS₀=CLOSURE({Z→•S}), 并给IS₀标上NO。
- ◆ Step2. 从已构造的LRSM₀部分图选择被标为NO的任一状态IS, 删除NO, 对每个符号 $X \in V_T \cup V_N$, 做下面动作:
 - 1) 令IS_j = CLOSURE(IS_(X)) 。
 - 2) 若IS_j非空:
 - ①如果在LRSM₀部分图中已有IS_j项目集,
则在IS和IS_j之间画有向X边: IS \xrightarrow{X} IS_j 。
 - ②如果在LRSM₀部分图中没有IS_j项目集,
则将IS_j作为LRSM的一个新的状态节点, 并给IS_j标上NO,
同在IS和IS_j之间画有向X边:
IS \xrightarrow{X} IS_j 。
- ◆ 重复Step2 , 直至没有被标记为NO的状态结点为止。

3. 构造LR(0)活前缀状态机LRSM₀

- ☆ 形如 $A \rightarrow \pi \cdot [P]$ 的项目称为归约型项目
- ☆ 形如 $A \rightarrow \alpha \cdot \beta [P]$ 的项目称为移入型项目
- ☆ 移入—归约冲突
- ☆ 归约—归约冲突
- ☆ 终止状态
- ☆ 非终止状态

例子

例1：计算下列文法的LR(0)状态机。

文法G[S]:

$S \rightarrow aAbB$

$A \rightarrow c \mid Ac$

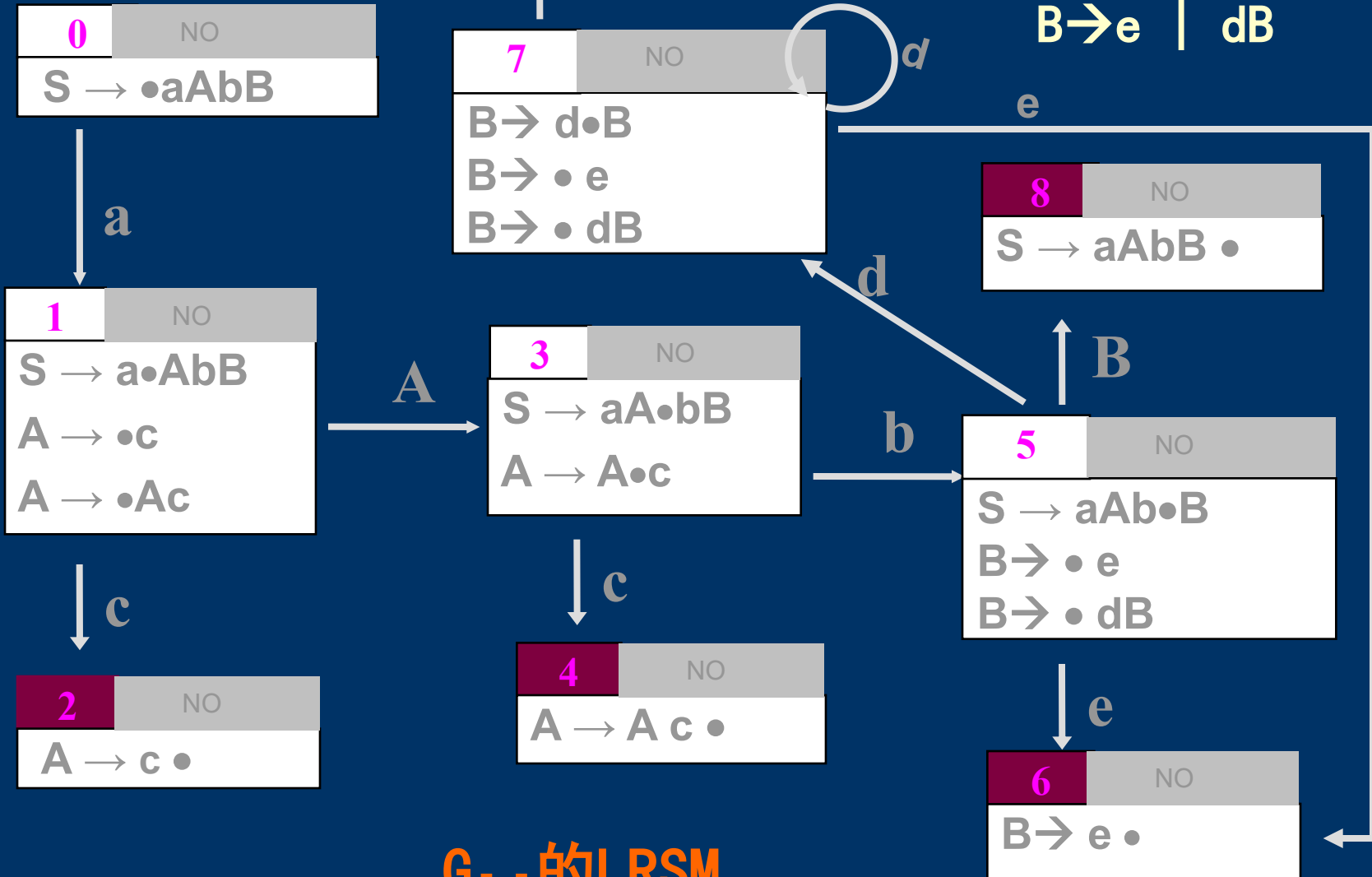
$B \rightarrow e \mid dB$

文法G[S] :

S → aAbB

A → c | Ac

B→e | dB



$G_{[s]}$ 的 LRSM_0

类型题

计算下列文法的LR(0)状态机。

文法G(S) :

$$S \rightarrow E \$$$
$$E \rightarrow E + T$$
$$E \rightarrow T$$
$$T \rightarrow id | (E)$$

