

编译原理和技术 作业 4

习题 3.4. 文法

$$R \rightarrow R \text{'|'} R \mid RR \mid R* \mid (R) \mid a \mid b \quad (4.1)$$

产生字母表 $\{a, b\}$ 上所有不含 ε 的正规式. 注意, 第一条竖线加了引号, 它是正规式的或运算符, 而不是文法产生式右部各选择之间的分割符, 另外 $*$ 在这是一个普通的终结符. 该文法是二义的.

(b) 为该文法写一个等价的非二义文法. 它给予算符 $*$, 连接和 $|$ 的优先级和结合性同 2.2 节中定义的一致.

解: (b) 如下

$$R \rightarrow R \text{'|'} S \mid S \quad (4.2)$$

$$S \rightarrow ST \mid T \quad (4.3)$$

$$T \rightarrow T* \mid (R) \mid a \mid b \quad (4.4)$$

习题 3.10. 构造下面文法的 LL(1) 分析表.

$$\begin{aligned} D &\rightarrow TL \\ T &\rightarrow \text{int} \mid \text{real} \\ L &\rightarrow \text{id} R \\ R &\rightarrow , \text{id} R \mid \varepsilon \end{aligned} \quad (10.1)$$

解: 各个非终结符的开始符号集和后继符号集为

$$\begin{aligned} FIRST(D) &= FIRST(T) = \{\text{int}, \text{real}\} \\ FIRST(L) &= \{\text{id}\} \\ FIRST(R) &= \{, , \varepsilon\} \\ FOLLOW(D) &= FOLLOW(L) = FOLLOW(R) = \{\$ \} \\ FOLLOW(T) &= FIRST(L) = \{\text{id}\} \end{aligned} \quad (10.2)$$

构造出 LL(1) 分析表如下

非终结符	输入符号				
	int	real	id	,	\$
D	$D \rightarrow TL$	$D \rightarrow TL$			
T	$T \rightarrow \text{int}$	$T \rightarrow \text{real}$			
L			$L \rightarrow \text{id}$		
R				$R \rightarrow , \text{id} R$	$R \rightarrow \varepsilon$

习题 3.12. 下面的文法是否为 LL(1) 文法, 说明理由.

$$\begin{aligned}
 S &\rightarrow AB \mid PQx \\
 A &\rightarrow xy \\
 B &\rightarrow bc \\
 P &\rightarrow dP \mid \varepsilon \\
 Q &\rightarrow aQ \mid \varepsilon
 \end{aligned} \tag{12.1}$$

解: 该文法不是 LL(1) 文法. 理由如下:

因为 $x \in FIRST(A)$, 所以 $x \in FIRST(AB)$. 而对于 PQx , 因为 $\varepsilon \in FIRST(P)$, $\varepsilon \in FIRST(Q)$, 所以 $x \in FIRST(PQ)$, 因此有 $x \in FIRST(AB) \cap FIRST(PQx) \implies FIRST(AB) \cap FIRST(PQx) \neq \emptyset$, 故该文法不是 LL(1) 的.

习题 3.19. 考虑下面的文法:

$$\begin{aligned}
 E &\rightarrow E + T \mid T \\
 T &\rightarrow TF \mid F \\
 F &\rightarrow F* \mid a \mid b
 \end{aligned} \tag{19.1}$$

(b) 构造 LALR 分析表.

解: 拓广文法如下

$$\begin{aligned}
 E' &\rightarrow E \\
 E &\rightarrow E + T \mid T \\
 T &\rightarrow TF \mid F \\
 F &\rightarrow F* \mid a \mid b
 \end{aligned} \tag{19.2}$$

首先构造 LR(1) 项目集规范族如下, 其中项目集标号前的箭头代表了它们在 DFA 之中的转换关系.

$$\begin{aligned}
 I_0 : E' &\rightarrow \cdot E, \$ & I_0 \xrightarrow{T} I_2 : E &\rightarrow T \cdot, +/\$ \\
 E &\rightarrow \cdot E + T, +/\$ & T &\rightarrow T \cdot F, +/a/b/\$ \\
 E &\rightarrow \cdot T, +/\$ & F &\rightarrow \cdot F*, +/* /a/b/\$ \\
 T &\rightarrow \cdot TF, +/a/b/\$ & F &\rightarrow \cdot a, +/* /a/b/\$ \\
 T &\rightarrow \cdot F, +/a/b/\$ & F &\rightarrow \cdot b, +/* /a/b/\$ \\
 F &\rightarrow \cdot F*, +/* /a/b/\$ & I_0, I_6 \xrightarrow{F} I_3 : T &\rightarrow F \cdot, +/a/b/\$ \\
 F &\rightarrow \cdot a, +/* /a/b/\$ & F &\rightarrow F \cdot *, +/* /a/b/\$ \\
 F &\rightarrow \cdot b, +/* /a/b/\$ & I_0, I_2, I_6, I_9 \xrightarrow{a} I_4 : F &\rightarrow a \cdot, +/* /a/b/\$ \\
 I_0 \xrightarrow{E} I_1 : E' &\rightarrow E \cdot, \$ & I_0, I_2, I_6, I_9 \xrightarrow{b} I_5 : F &\rightarrow b \cdot, +/* /a/b/\$ \\
 E &\rightarrow E \cdot + T, +/\$
 \end{aligned} \tag{19.3}$$

$$\begin{aligned}
I_1 \xrightarrow{+} I_6 : E \rightarrow E + \cdot T, +/\$ & \quad I_3, I_7 \xrightarrow{*} I_8 : F \rightarrow F * \cdot, +/\$ /a/b/\$ \\
T \rightarrow \cdot TF, +/a/b/\$ & \quad I_6 \xrightarrow{T} I_9 : E \rightarrow E + T \cdot, +/\$ /a/b/\$ \\
T \rightarrow \cdot F, +/a/b/\$ & \quad T \rightarrow T \cdot F, +/a/b/\$ \\
F \rightarrow \cdot F*, +/\$ /a/b/\$ & \quad F \rightarrow \cdot F*, +/\$ /a/b/\$ \\
F \rightarrow \cdot a, +/\$ /a/b/\$ & \quad F \rightarrow \cdot b, +/\$ /a/b/\$ \\
F \rightarrow \cdot a, +/\$ /a/b/\$ & \quad F \rightarrow \cdot b, +/\$ /a/b/\$ \\
I_2, I_9 \xrightarrow{F} I_7 : T \rightarrow TF \cdot, +/a/b/\$ & \\
F \rightarrow F \cdot *, +/\$ /a/b/\$ &
\end{aligned} \tag{19.4}$$

可以看出 LR(1) 项目规范族中没有同心的项目集, 因此可直接构造 LALR 分析表. 先对产生式编号如下

$$\begin{aligned}
(1) E \rightarrow E + T & \quad (5) F \rightarrow F* \\
(2) E \rightarrow T & \quad (6) F \rightarrow a \\
(3) T \rightarrow TF & \quad (7) F \rightarrow b \\
(4) T \rightarrow F &
\end{aligned} \tag{19.5}$$

构造出 LALR 分析表如下

表 19.1: 由文法 19.1 构造的 LALR 分析表.

状态	动作					转移		
	<i>a</i>	<i>b</i>	+	*	\$	<i>E</i>	<i>T</i>	<i>F</i>
0	<i>s4</i>	<i>s5</i>				1	2	3
1			<i>s6</i>		<i>acc</i>			
2	<i>s4</i>	<i>s5</i>	<i>r2</i>		<i>r2</i>			7
3	<i>r4</i>	<i>r4</i>	<i>r4</i>	<i>s8</i>	<i>r4</i>			
4	<i>r6</i>	<i>r6</i>	<i>r6</i>	<i>r6</i>	<i>r6</i>			
5	<i>r7</i>	<i>r7</i>	<i>r7</i>	<i>r7</i>	<i>r7</i>			
6	<i>s4</i>	<i>s5</i>					9	3
7	<i>r3</i>	<i>r3</i>	<i>r3</i>	<i>s8</i>	<i>r3</i>			
8	<i>r5</i>	<i>r5</i>	<i>r5</i>	<i>r5</i>	<i>r5</i>			
9	<i>s4</i>	<i>s5</i>	<i>r1</i>		<i>r1</i>			7

习题 3.27. 文法 G 的产生式如下

$$\begin{aligned}
S &\rightarrow I \mid R \\
I &\rightarrow d \mid Id \\
R &\rightarrow WpF \\
W &\rightarrow Wd \mid \varepsilon \\
F &\rightarrow Fd \mid d
\end{aligned} \tag{27.1}$$

(b) 该文法是 LR(1) 文法吗? 为什么?

解: 拓广文法为

$$\begin{aligned}
 S' &\rightarrow S \\
 S &\rightarrow I \mid R \\
 I &\rightarrow d \mid Id \\
 R &\rightarrow WpF \\
 W &\rightarrow Wd \mid \varepsilon \\
 F &\rightarrow Fd \mid d
 \end{aligned} \tag{27.2}$$

可以构造出 I_0 如下

$$\begin{aligned}
 S' &\rightarrow \cdot S, \$ \\
 S &\rightarrow \cdot I, \$ \\
 S &\rightarrow \cdot R, \$ \\
 I &\rightarrow \cdot d, d/\$ \\
 I &\rightarrow \cdot Id, d/\$ \\
 R &\rightarrow \cdot WpF, \$ \\
 W &\rightarrow \cdot WpF, p/d \\
 W &\rightarrow \cdot, p/d
 \end{aligned} \tag{27.3}$$

只关注这两个项目: $[I \rightarrow \cdot d, d]$ 和 $[W \rightarrow \cdot, d]$. 在分析器还未读入符号, 面临 d 的情况下, 前者表明分析器可以移进 d ; 后者表明分析器可以按 $W \rightarrow \varepsilon$ 进行规约. 因此出现了移进-规约冲突, 所以文法不是 LR(1) 的.