



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
 - 计算 $FIRSTVT$ 和 $LASTVT(P)$ 集合
 - 构造算符优先关系表
- LR 分析法

算符优先分析算法

- 可归约串，句型，短语，直接短语，句柄，规范归约
- 一个文法 G 的句型的**素短语**是指这样一个短语，它至少含有一个终结符，并且，除它自身之外不再含任何更小的素短语
- **最左素短语**是指处于句型最左边的那个素短语

■ 考虑下面的文法 $G(E)$:

- (1) $E \rightarrow E + T \mid T$
- (2) $T \rightarrow T * F \mid F$
- (3) $F \rightarrow P \uparrow F \mid P$
- (4) $P \rightarrow (E) \mid i$

对于句型: $T + F * P + i$

短语: $T, F, P, i, F * P,$

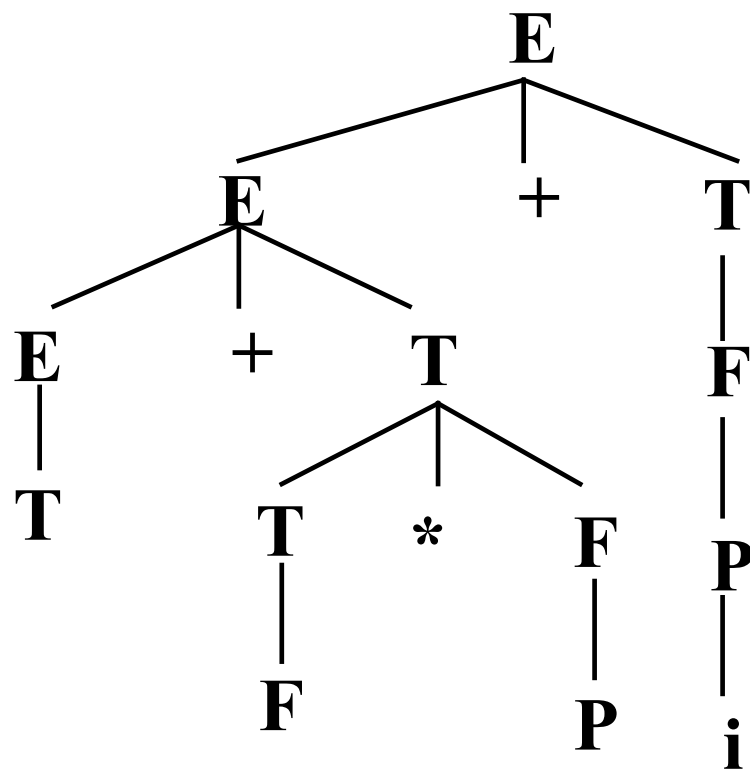
$T + F * P, T + F * P + i$

直接短语: T, F, P, i

句柄: T

素短语: $F * P, i$

最左素短语: $F * P$



- 算符优先文法句型（括在两个 # 之间）的一般形式写成：

$$\#N_1a_1N_2a_2\cdots N_na_nN_{n+1}\#$$

其中，每个 a_i 都是终结符， N_i 是可有可无的非终结符。

- 定理：一个算符优先文法 G 的任何句型的最左素短语是满足如下条件的最左子串 $N_ja_j\cdots N_ia_iN_{i+1}$ ，

$$\begin{array}{c} a_{j-1} \blacklozenge a_j \\ a_j \blacklozenge a_{j+1}, \quad \cdots, \quad a_{i-1} \blacklozenge a_i \\ a_i \square a_{i+1} \end{array}$$

算符优先分析算法

- 使用一个符号栈 S ，用它寄存终结符和非终结符， k 代表符号栈 S 的使用深度

```

k:=1;
S[k]:=' #' ;
REPEAT

```

把下一个输入符号读进 a 中

```

  IF S[k] ∈ VT THEN
  WHILE S[j] ≠ a DO
  BEGIN

```

```

    REPEAT

```

```

      Q:=S[j];

```

```

      IF S[j-1] ∈ VT THEN j:=j-1 ELSE j:=j-2

```

```

    UNTIL S[j] ≠ Q;

```

把 S[j+1]...S[k] 归约为某个 N;

```

    k:=j+1;

```

```

    S[k]:=N

```

```

  END OF WHILE;

```

```

  IF S[j] ≠ a OR S[j] ≠ a THEN

```

```

    BEGIN k:=k+1; S[k]:=a END

```

```

  ELSE ERROR /* 调用出错诊察程序 */

```

```

UNTIL a=' #'

```

自左至右，终结符对终结符，非终结符对非终结符，而且对应的终结符相同。

$$N \rightarrow X_1 X_2 \dots X_{k-j}$$


S[j+1] S[j+2] ... S[k]

j:=j-1 ELSE j:=j-2

算符优先分析算法

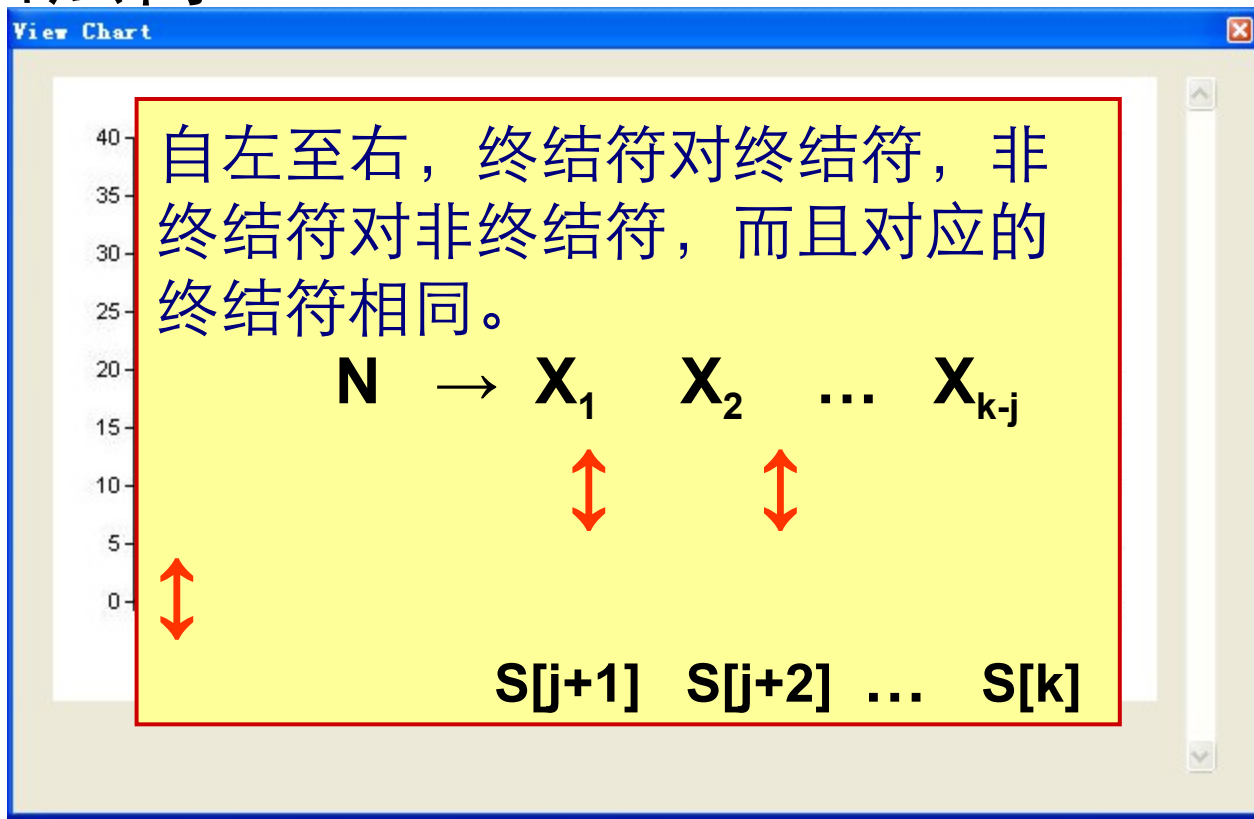
- 在算法的工作过程中，若出现 j 减 1 后的值小于等于 0 时，则意味着输入串有错。在正确的情况下，算法工作完毕时，符号栈 S 应呈现： $\# N \#$ 。
- 由于非终结符对归约没有影响，因此，非终结符根本可以不进符号栈 S 。

分析树 vs. 语法树

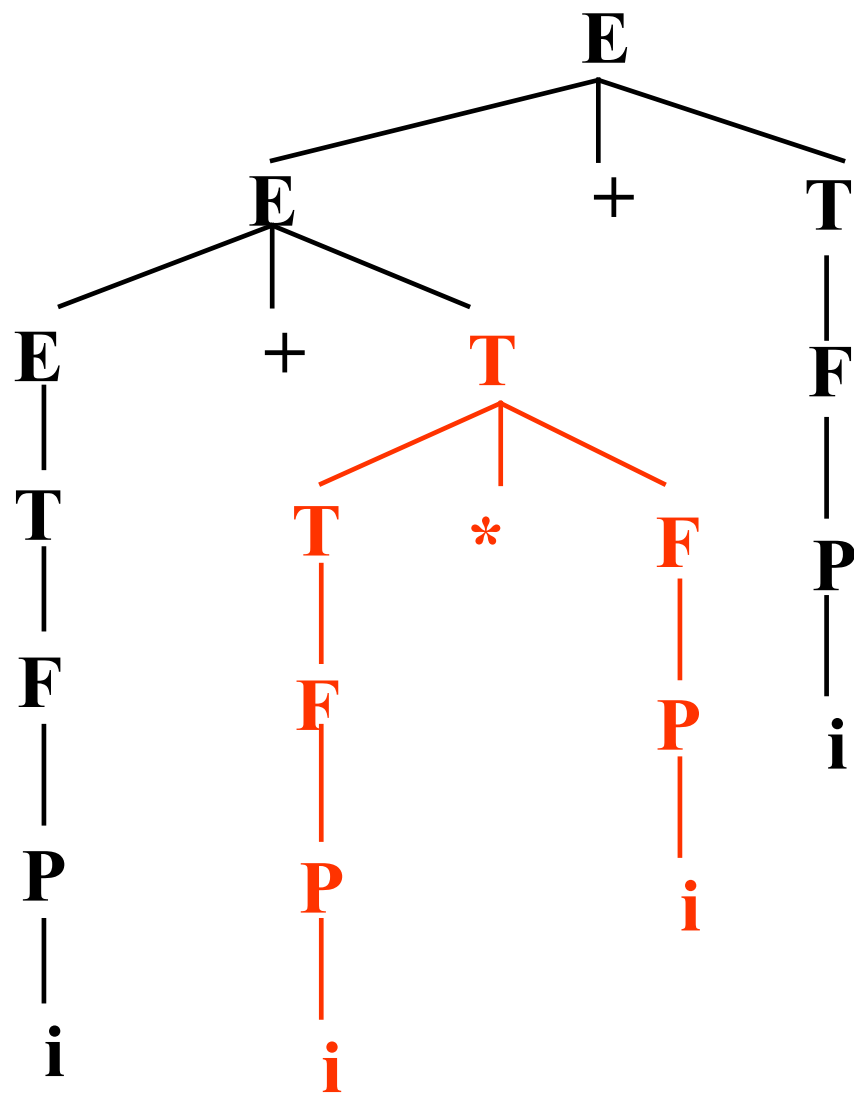
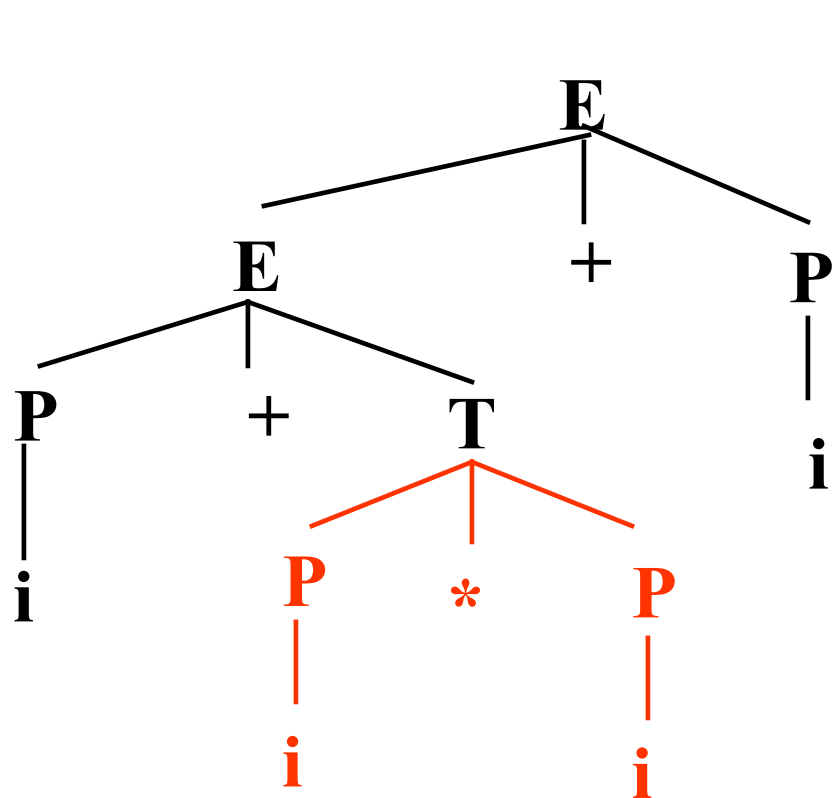
- 对于文法的句子来说，它的算符优先分析的结果就是语法树

A. 正确

B. 错误



■ 算符优先分析一般并不等价于规范归约



■ 考虑下面的文法 $G(E)$:

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

(2) $T \rightarrow T * F \mid F$

(3) $F \rightarrow P \uparrow F \mid P$

(4) $P \rightarrow (E) \mid i$

的句子 $i+i*i+i$

算符优先分析算法

- 算符优先分析法特点

- 优点：简单，快速

- 缺点：可能错误接受非法句子，能力有限

- 算符优先分析法是一种广为应用、行之有效的方

- 用于分析各类表达式

- ALGOL 60

优先函数

- 每个终结符 θ 与两个自然数 $f(\theta)$ 与 $g(\theta)$ 相对应，使得

若 $\theta_1 < \theta_2$ ，则 $f(\theta_1) < g(\theta_2)$

若 $\theta_1 = \theta_2$ ，则 $f(\theta_1) = g(\theta_2)$

若 $\theta_1 > \theta_2$ ，则 $f(\theta_1) > g(\theta_2)$

f 称为入栈优先函数， g 称为比较优先函数。

- 优点

- 便于比较，节省空间

- 缺点

- 原来不存在优先关系的两个终结符，由于自然数相对应，变成可以比较的
- 要进行一些特殊的判断

■ 文法 $G(E)$

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

$$(2) T \rightarrow T * F \mid F$$

$$(3) F \rightarrow P \uparrow F \mid P$$

$$(4) P \rightarrow (E) \mid i$$

的优先函数如下表

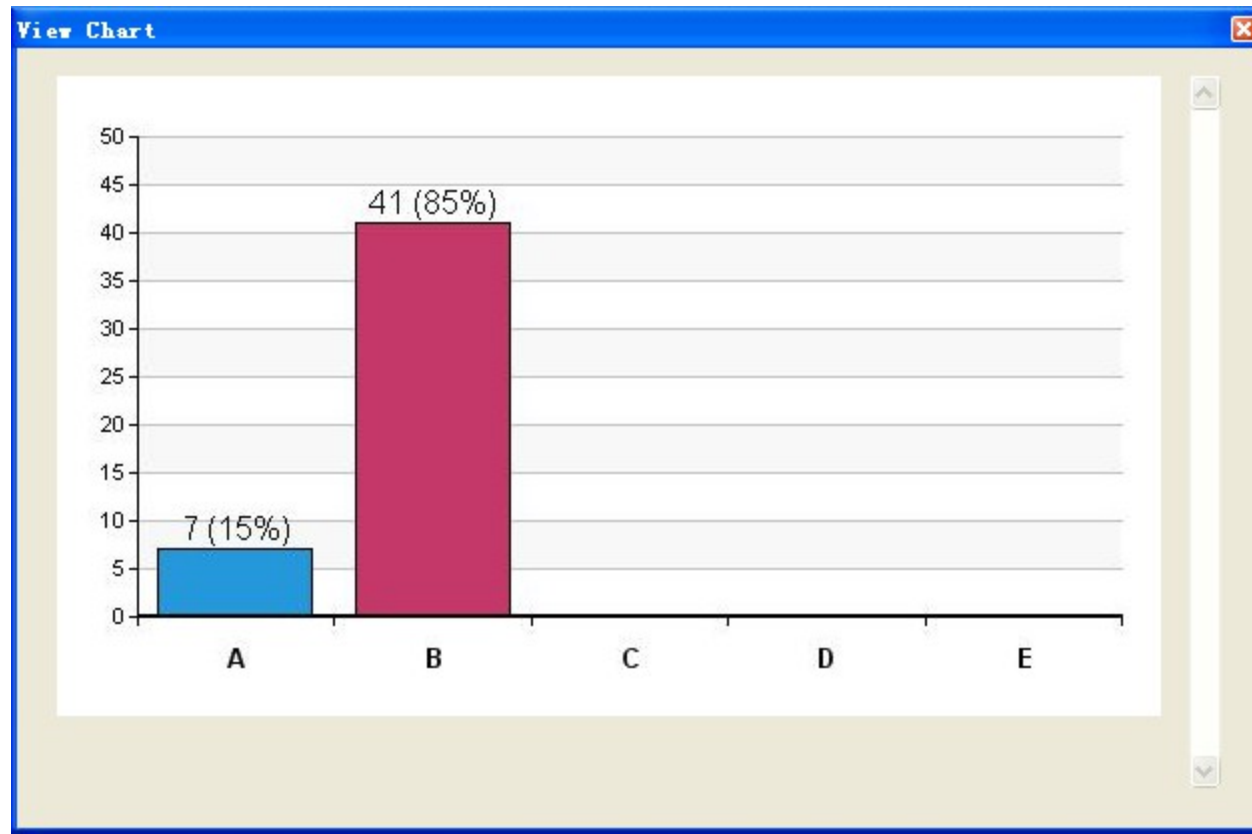
	+	*	\uparrow	()	i	#
F	2	4	4	0	6	6	0
G	1	3	5	5	0	5	0

讨论：优先函数

- 对于任何无冲突优先关系表都存在优先函数？

A. 是

B. 不是



- 有许多优先关系表不存在优先函数，如：

	a	b
a		
b		

不存在对应的优先函数 f 和 g

假定存在 f 和 g ，则有

$$f(a)=g(a), \quad f(a)>g(b),$$

$$f(b)=g(a), \quad f(b)=g(b)$$

导致如下矛盾：

$$f(a) > g(b) = f(b) = g(a) = f(a)$$

讨论：优先函数

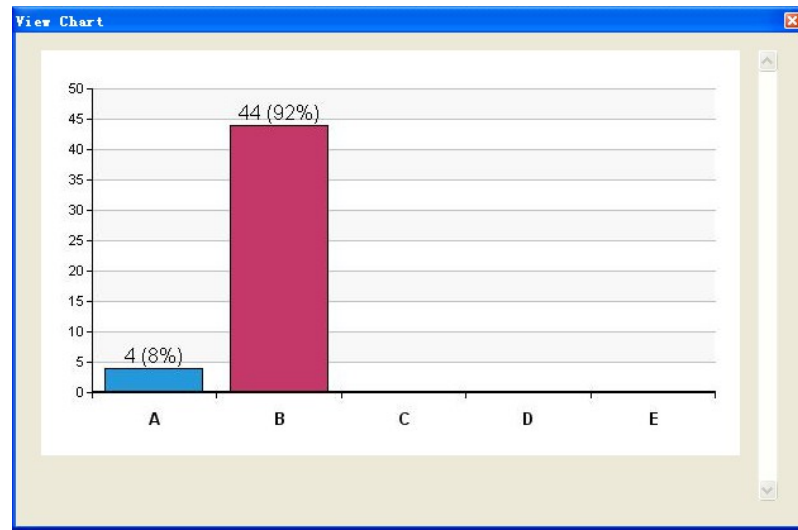
■ 优先函数如果存在，唯一吗？

A. 是

B. 不是

如果优先函数存在，则不唯一（无穷多）

！



■ 如果优先函数存在，则可以通过以下三个步骤从优先表构造优先函数

1. **画图**：对于每个终结符 a ，令其对应两个符号 f_a 和 g_a ，画一以所有符号和为结点的方向图。如果 $a \prec b$ ，则从 f_a 画一条弧至 g_b ，如果 $a \succ b$ ，则画一条弧从 g_b 至 f_a 。
2. **数数**：对每个结点都赋予一个数，此数等于从该结点出发所能到达的结点（包括出发点自身）。赋给 f_a 的数作为 $f(a)$ ，赋给 g_a 的数作为 $g(a)$ 。
3. **验证**：检查所构造出来的函数 f 和 g 是否与原来的关系矛盾。若没有矛盾，则 f 和 g 就是要求的优先函数，若有矛盾，则不存在优先函数。

1 对于每个终结符 a ，令其对应两个符号 f_a 和 g_a ，画一以所有符号和为结点的方向图。如果 $a \sqsubseteq b$ ，则从 f_a 画一条弧至 g_b ，如果 $a \blacklozenge b$ ，则画一条弧从 g_b 至 f_a 。

■ 现在必须证明

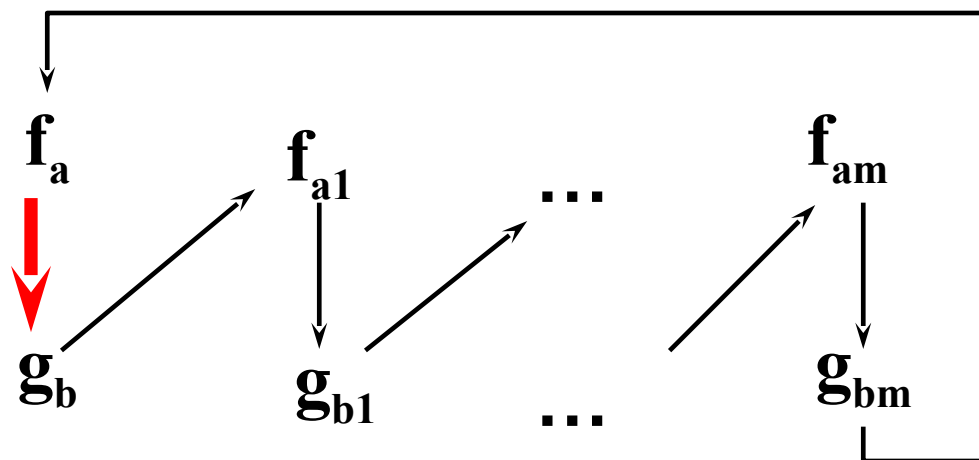
- 若 $a \blacklozenge b$ ，则 $f(a) = g(b)$
- 若 $a \blacklozenge b$ ，则 $f(a) < g(b)$
- 若 $a \sqsubseteq b$ ，则 $f(a) > g(b)$

■ 第一个关系可从函数的构造直接获得

- 若 $a \blacklozenge b$ ，则既有从 f_a 到 g_b 的弧，又有从 g_b 到 f_a 的弧。所以， f_a 和 g_b 所能到达的结是相同的。

■ 至于 $a \sqsubseteq b$ 和 $a \blacklozenge b$ 的情形，只须证明其一。

- 如果 $a \sqsubseteq b$ ，则有从 f_a 到 g_b 的弧。也就是， g_b 能到达的任何结点 f_a 也能到达。因此， $f(a) \geq g(b)$
- 我们所需证明的是，在这种情况下， $f(a) = g(b)$ 不应成立
- 我们将指出，如果 $f(a) = g(b)$ ，则根本不存在优先函数
- 假若 $f(a) = g(b)$ ，那么必有如下的回路：



因此有

$$a \square b, a_1 \blacklozenge b, a_1 \square \blacklozenge b_1, \dots, a_m \square \blacklozenge b_m, a \blacklozenge b_m$$

对任何优先函数 f' 和 g' 来说，必定有

$$f'(a) > g'(b) \geq f'(a_1) \geq g'(b_1) \geq \dots \geq f'(a_m) \geq g'(b_m) \geq f'(a)$$

从而导致 $f'(a) > f'(a)$ ，产生矛盾。因此，不存在优先函数 f 和 g 。

■ 例：取前面文法 $G(E)$

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

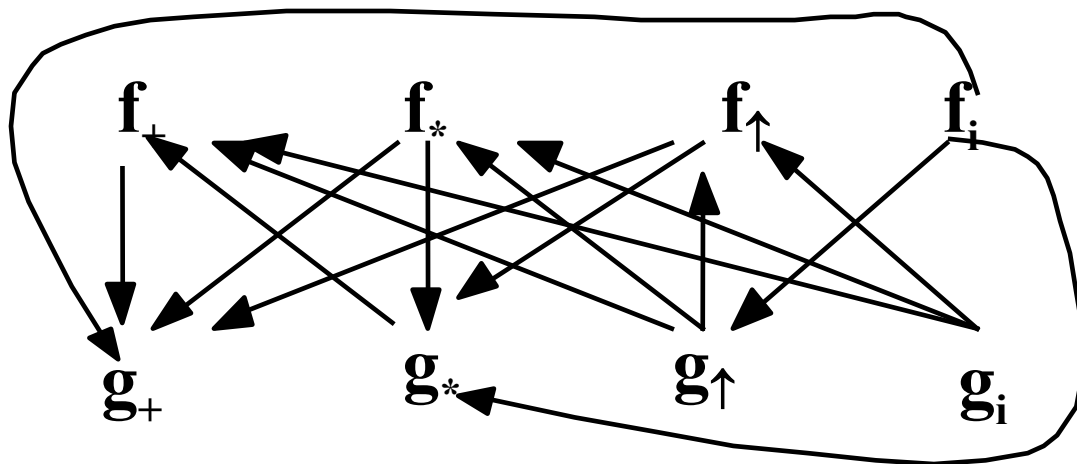
$$(2) T \rightarrow T * F \mid F$$

$$(3) F \rightarrow P \uparrow F \mid P$$

$$(4) P \rightarrow (E) \mid i$$

的终结符 $+$, $*$, \uparrow , i

	+	*	↑	i
+				
*				
↑				
i				



	+	*	↑	i
f	2	4	4	7
g	1	3	6	6

小结

- 算符优先分析算法
 - 最左素短语
- 算符优先关函数及其构造方法

作业

■ P133—3