

基于预测平方和 PRESS 准则选取 预报因子的向前、向后算法及其应用*

俞善贤

沈锦花

(浙江省气象科学研究所)

(浙江省气象台)

本文提出了基于预测平方和 PRESS 准则选取预报因子的二个快速算法——向前、向后算法，这二个算法不仅使计算 PRESS 值的回归方程个数从原算法 $2^p - 1$ 个减少到至多为 $P(P+1)/2$ 个，而且使每个 PRESS 值的计算量减少很多。通过长期天气预报的实例计算，预测效果比逐步回归有明显提高。

1. 预测平方和 PRESS 准则

预测平方和 PRESS 准则是 D. M. Allen 于 1971 年提出的^[1]，基本思想如下：

在线性模型 $Y = X\beta + e$ 中

我们把第 i 个样本去掉，得到模型

$$Y^{(i)} = X^{(i)}\beta + e^{(i)}$$

其中 $Y^{(i)}$ 、 $X^{(i)}$ 和 $e^{(i)}$ 是分别从 Y 、 X 和 e 中删去第 i 行而得到的，对这个模型用最小二乘法得 β 的估计 $\hat{\beta}^{(i)}$ ，就可算出第 i 个试验点 x_i 处的预测值 $x_i'\hat{\beta}^{(i)}$ ，这时预测偏差为 $f_i = Y_i - x_i'\hat{\beta}^{(i)}$ ，对每个样本都这样做，其平方和

$$\text{PRESS} = \sum_{i=1}^n f_i^2$$

这个值的直观意义很明确，用它来衡量方程的预测能力是一个较好的数量指标。为了便于计算，已经证明了^[1]

$$\text{PRESS} = \hat{e}' D^{-1} \hat{e} \quad (1)$$

其中

$$\hat{e} = Y - X\beta = (I - X(X'X)^{-1}X')Y \quad (2)$$

$$D = \text{diag}(d_1, d_2, \dots, d_n) \quad (3)$$

d_i 为方阵 $I - X(X'X)^{-1}X'$ 的对角元 (i, i) 。

显然预测平方和 PRESS 之值与所选用的预报因子有关，这个准则就是在由预报因子 $X_1, X_2, \dots, X_p$ 可能产生的一切子集中 (共 $2^p - 1$ 个) 选取 PRESS 值达到最小的那个子集为最优回归子集。

原算法使用扫描运算。扫描运算如下：

设 A 为 n 阶方阵，以 a_{ii} 为枢轴的运算记为 $S_i A$ ，定义一个新方阵 $B = (b_{ij})_{n \times n}$ ，其中

$$\begin{cases} b_{jk} = a_{jk} - a_{ji}a_{ki}/a_{ii} & \text{当 } j \neq i, k \neq i \\ b_{ji} = -a_{ji}/a_{ii} & \text{当 } j \neq i \\ b_{ii} = a_{ii}/a_{ii} & \text{当 } j = i \\ b_{ii} = 1/a_{ii} & \end{cases} \quad (4)$$

自然假定 $a_{ii} \neq 0$ 。

为了计算各个预报因子子集的 PRESS 值，原算法计算每一个子集的 $(X'X)^{-1}$ 和 $\hat{\beta}$ 需作一次扫描运算，

然后用(1)式计算该子集的 PRESS 值, 整个算法共需 $2^P - 1$ 次扫描运算, 当 P 较大时, 计算量大得惊人。

2. 向前算法

在原算法中, 每引入或删除一个因子需作一次扫描运算, 这样做很不经济, 作者在文献[2]中证明的下列关系式, 可以减少很多计算量。

设预报因子为 P 个, $X = (X_1, X_2, \dots, X_P)$, 把 X 分成 $X_{P_1} = (X_{i_1}, X_{i_2}, \dots, X_{i_{P_1}})$, $X_{P_2} = (X_{i_{P_1+1}}, X_{i_{P_1+2}}, \dots, X_{i_P})$ 两部分, 对 $(X'X)$ 经过 P_1 次扫描运算, 记为 A

$$A = S_{i_1} S_{i_2} \cdots S_{i_{P_1}} (X'X) \quad i_j \in \{i_1, i_2, \dots, i_{P_1}\}$$

已求得残差向量

$$\hat{e}_{P_1} = Y - \hat{Y}_{P_1} = Y - X_{P_1} \hat{\beta}_{P_1}$$

和矩阵 $H_{P_1} = X_{P_1} (X'_{P_1} X_{P_1})^{-1} X'_{P_1}$ 的对角元。

若继续选入预报因子 X_{i_k} , $i_k \in \{i_{P_1+1}, i_{P_1+2}, \dots, i_P\}$

则

$$H_{P_1+1} = X_{P_1} (X'_{P_1} X_{P_1})^{-1} X'_{P_1} + \frac{1}{a_{i_k i_k}} (X_{P_1} A_{i_k} - X_{i_k}) (X_{P_1} A_{i_k} - X_{i_k})' \quad (5)$$

$$\hat{e}_{P_1+1} = \hat{e}_{P_1} - \frac{1}{a_{i_k i_k}} (X_{P_1} A_{i_k} - X_{i_k}) (X_{P_1} A_{i_k} - X_{i_k})' Y \quad (6)$$

其中

$$A_{i_k} = \begin{bmatrix} a_{i_1 i_k} \\ a_{i_2 i_k} \\ \vdots \\ a_{i_{P_1} i_k} \end{bmatrix}$$

$a_{i_k i_k}$ 为矩阵 A 的 (i_k, i_k) 元素。

由(1)、(5)、(6)式可知, 回归方程选进预报因子 X_{i_k} 后相应的 PRESS 值为

$$\text{PRESS} = \hat{e}_{P_1+1}' D^{-2} \hat{e}_{P_1+1} \quad (7)$$

其中 $\hat{e}_{P_1+1}$ 为(6)式, $D = \text{diag}(d_1, d_2, \dots, d_n)$, d_i 为矩阵 $I - H_{P_1+1}$ 的对角元 (i, i) 。

这时计算选入一个因子后的 PRESS 值, 只要增加计算 $\frac{1}{a_{i_k i_k}} (X_{P_1} A_{i_k} - X_{i_k}) (X_{P_1} A_{i_k} - X_{i_k})'$ 的对角元和 $\frac{1}{a_{i_k i_k}} (X_{P_1} A_{i_k} - X_{i_k}) (X_{P_1} A_{i_k} - X_{i_k})' Y$, 共需乘除运算约 $n(P_1 + 6)$ 次, 而原算法乘除运算约 $n(P_1^2 + 2P_1 + 2) + P(P + 2)$ 次。若 $n = 30$, $P = 10$, $P_1 = 5$, 乘除运算约减少 73.2%, 加减运算也类似。以上的关系式对 X_{P_k} 中的任何一个因子都适用, 总体上又减少很多计算量。向前算法每步选入一个因子, 使 PRESS 值极小, 直到 PRESS 值达最小为止。

3. 向后算法

设全模型的预报因子为 P 个, 对选定的其中 $P - 1$ 个因子, 定义与此相对应的矩阵 Q 使选模型的设计矩阵为 XQ , Q 为 $P \times (P - 1)$ 阶矩阵, 其秩为 $P - 1$, 对未选入的第 i 个因子, 定义矩阵 R , 使未选入因子为 XR , R 为 $P \times 1$ 阶矩阵, 第 i 行元素为 1, 其余为 0。

例如, 有预报因子 X_1, X_2, X_3, X_4 , 如果选取 X_1, X_2, X_4 , 这时相应的 Q 和 R 可选取为

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad R = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

有了给定的 Q , 这时选模型的回归系数 β 的最小二乘估计为(自然, 未选入因子的回归系数的估计值为 0)

$$\hat{\beta}_Q = (Q'X'XQ)^{-1}Q'X'Y$$

残差向量 $\hat{e}_Q$ 为

$$\hat{e}_Q = (I - XQ(Q'X'XQ)^{-1}Q'X')Y \quad (8)$$

由(1)、(2)、(3)式可知选模型的 PRESS(Q)值为

$$\text{PRESS}(Q) = \hat{e}_Q' D_Q^{-2} \hat{e}_Q \quad (9)$$

$$D_Q = \text{diag}(d_{Q_1}, \dots, d_{Q_k}, d_{Q_n}) \quad (10)$$

其中 d_{Q_k} 为方阵 $I - XQ(Q'X'XQ)^{-1}Q'X'$ 的对角元(k, k)。

下面对(8)、(10)式进行化简。因为矩阵 Q , R 满足条件 $R'Q = 0$, 则有下列矩阵关系式^[1]

$$Q(Q'X'XQ)^{-1}Q' = (X'X)^{-1} - (X'X)^{-1}R[R'(X'X)^{-1}R]^{-1}R'(X'X)^{-1} \quad (11)$$

把(11)式代入(8)式得

$$\hat{e}_Q = Y - X[(X'X)^{-1} - (X'X)^{-1}R[R'(X'X)^{-1}R]^{-1}R'(X'X)^{-1}]X'Y \quad (12)$$

考察式中 $[R'(X'X)^{-1}R]^{-1}$ 部分, 由 R 的定义可知 $R'(X'X)^{-1}R$ 是一个数, 它等于矩阵 $(X'X)^{-1}$ 的对角元(i, i), 记为 h_{ii} , 则

$$[R'(X'X)^{-1}R]^{-1} = \frac{1}{h_{ii}} \quad (13)$$

(12)式中的

$$R'(X'X)^{-1}X'Y = R'\hat{\beta} = \hat{\beta}_i \quad (14)$$

$\hat{\beta}$ 为全模型用最小二乘估计的回归系数, $\hat{\beta}_i$ 为 $\hat{\beta}$ 中第 i 个变量的回归系数。记

$$L = X(X'X)^{-1}R$$

事实上 L 为全模型的设计矩阵 X 乘 $(X'X)^{-1}$ 中的第 i 列。把(13)、(14)、(15)式代入(12)式, 得

$$\hat{e}_Q = Y - X\hat{\beta} + \frac{\hat{\beta}_i}{h_{ii}} L = \hat{e} + \frac{\hat{\beta}_i}{h_{ii}} L \quad (16)$$

同理(10)式中的 d_{Q_k} 为方阵 $I - X(X'X)^{-1}X' + \frac{1}{h_{ii}}LL'$ 的对角元。

若全模型已经计算了 PRESS 值, 这时从全模型中剔除某一个因子, 计算余下因子组成的选模型的 PRESS 值, 只要增加计算 $\frac{\hat{\beta}_i}{h_{ii}}L$ 和 $\frac{1}{h_{ii}}LL'$ 的对角元部分, 这样做需要乘除运算约 $n(P+5)$ 次, 若用原算法需要乘除运算约 $2P^2(n+1) - 3nP + 2(P+n)$ 次。若 $n=30$, $P=10$, 乘除运算下降 91.6%, 加减运算也类似。以上关系式对剔除 P 个因子中的任何一个都适用, 大大减少了计算量。

向后算法每步剔除一个因子, 使 PRESS 值极小, 直到 PRESS 值达最小为止。

4. 应用实例

在 IBM-PC/AT 微机上用编译 BASIC 语言实现了向前、向后和原算法的程序。

实例 1: 杭州地区 1—9 月份降水量预报

样本个数为 25, 候选因子为 12, 它们都是北半球 100 hPa 月平均网格点高度值, 选进方程的因子是:

X_1 : 上年 7 月, $30^\circ-40^\circ\text{N}$, $10^\circ-50^\circ\text{E}$ 区域;

X_3 : 上年8月, 60°N , $40^\circ\text{--}60^\circ\text{E}$ 区域;

X_7 : 上年9月, 80°N , $0^\circ\text{--}30^\circ\text{E}$ 区域;

X_{10} : 上年11月, $60^\circ\text{--}70^\circ\text{N}$, $100^\circ\text{--}130^\circ\text{E}$ 区域;

X_{11} : 上年11月, 30°N , $50^\circ\text{--}60^\circ\text{W}$ 区域。

用向前、向后算法和原算法计算时间分别 17 s, 23 s, 1 h 13 min 21 s, 结果相同, 向后算法比原算法减少 99.5%, 建立的方程为

$$\hat{y}_P = 2217.5 - 31.2482 X_1 + 12.7707 X_3 - 12.2234 X_7 + 7.4429 X_{10} + 19.6987 X_{11}$$

复相关系数 $R=0.7922$ 。

取选入和剔除变量的 F 检验信度 $\alpha=0.10$, $F_1=F_2=2.93$, 用逐步回归建立的方程为

$$\hat{y}_S = 2466.6 - 27.1269 X_1 + 6.9678 X_{10} + 19.0901 X_{11}$$

复相关系数 $R=0.7171$ 。

对 1983—1986 年进行试报, 结果见表 1。

实例 2: 绍兴地区 1—9 月份降水量预报

样本个数为 27, 候选因子为 6 个, 它们都是北半球 100 hPa 月平均网格点高度值, 选进方程的因子是:

X_1 : 上年8月, 40°N , 170°W 区域;

X_2 : 上年11月, $40^\circ\text{--}60^\circ\text{N}$, $170^\circ\text{E--}160^\circ\text{W}$ 区域;

X_4 : 上年11月, $20^\circ\text{--}30^\circ\text{N}$, $60^\circ\text{--}80^\circ\text{W}$ 区域。

用向前、向后算法都得出最优解, 方程为

$$\hat{y}_P = 2652 - 27.3038 X_1 - 18.7086 X_2$$

复相关系数 $R=0.701$ 。

取 $F_1=F_2=2.5$, 用逐步回归建立的方程为

$$\hat{y}_S = 2219 - 29.2176 X_1 - 16.7216 X_2 + 12.7365 X_4$$

复相关系数 $R=0.74$

对 1984—1986 年进行试报, 结果见表 2。

表 1 杭州地区 1—9 月份降水量
预报对照表

年份	降水量	向前、向后算法		逐步回归	
		$\hat{y}_P$	$ e_P $	$\hat{y}_S$	$ e_S $
1983	1 381	1 242	139	1 134	247
1984	1 432	1 096	336	1 137	295
1985	1 058	1 008	50	1 143	85
1986	1 159	1 146	13	1 091	68
平均	—	—	135	—	175

表 2 绍兴地区 1—9 月降水量
预报对照表

年份	降水量	向前、向后算法		逐步回归	
		$\hat{y}_P$	$ e_P $	$\hat{y}_S$	$ e_S $
1984	1 094	1 256	162	1 260	166
1985	957	1 200	243	1 232	275
1986	979	950	29	1 028	49
平均	—	—	145	—	163

5. 讨 论

向前、向后算法不保证得出的解是最优的, 但是这二个算法每选入或剔除一个因子使“局部”最优。在本文的二个实例和一些模拟例子中, 该算法得出的解与最优解有很好的近似, 大多数都是最优解, 当 P 大于 20 时, 原算法在一般微机上几乎不能实现, 此时向前、向后算法仍有效。

向前算法的特点是当矩阵 $X'X$ 奇异时, 仍可以使用, 但作者在模拟例子中发现, 对最优解的逼近

程度,向前算法比向后算法要差一些,有一个例子向前算法比最优解多选了一个因子,此时用向后算法就可以得出最优解。该例也启发我们去实现逐步算法,即向前算法选入一个因子后,就去考虑剔除多余因子,计算过程同逐步回归类似。用本文的一些算式是很容易实现的。

用PRESS准则来选取预报因子是非参数化的,不作任何假设检验,给实际使用带来方便。

用PRESS准则来选取因子同逐步回归有较大的不同。例如,实例1中当 F 检验的信度为0.1时,只选入3个因子,信度已很低,一般就不考虑继续选入因子,用PRESS准则选出了5个因子,试报四年的效果也有较明显提高。实例2中逐步回归选入3个因子,一般就不会考虑去剔除因子,而PRESS准则只选入2个因子,三年试报效果也说明2个因子的方程有好的预报能力。

通过上述讨论,可以认为向前、向后算法是有效的,具有实际使用价值。

参 考 文 献

- [1] 陈希孺、王松桂,近代实用回归分析,106—112,131—137,广西人民出版社,1984。
- [2] 俞善贤、沈锦花,基于预测平方和准则选取预报因子的一种快速算法——向前算法,数理统计与应用概率,3, 1, 37—42, 1988。

FORWARD AND BACKWARD ALGORITHMS FOR SELECTING PREDICTORS ON THE BASIS OF THE CRITERION FROM PREDICTION SUM OF SQUARES AND THEIR APPLICATION

Yu Shanxian

(Zhejiang Research Institute of Meteorological Science)

Shen Jinhua

(Meteorological Observatory of Zhejiang)

Abstract

In this paper, two fast algorithms for selecting predictors, the forward and backward algorithms, are proposed on the basis of the criterion from prediction sum of squares (PRESS). Not only would the numbers of the regression equation calculating PRESS value reduce from 2^p-1 to at most $p(p+1)/2$, but also the amount of calculation for every PRESS value is much decreased, when either algorithm is used. The application in the long-range weather forecast shows that these algorithms have higher prediction skill as compared with stepwise regression.