

气象统计预报模型的自适应方法

龙从彬

(四川省雅安地区气象台, 625000)

提 要

该文利用矩阵反演求逆定理, 导出预报模型建立时的统计参数估计递推公式, 从而解决了气象回归模型的自适应修正问题。使用该方法建立气象回归预报模型, 能够节省计算机的内存, 减少计算量, 并能很方便地对所建模型进行实时修正。

关键词: 预报模型 自适应方法

1 问题的提出

在基层气象台站, 建立气象回归预报模型以制作各类天气预报, 仍然是预报人员的常用技术。然而, 在实际业务工作中, 会出现这样的情况: 一个新建立的回归预报模型, 开始使用时效果较好, 随年代的推移, 模型的预报效果越来越差。此时, 我们自然会想到修改回归模型。按常规方法, 每当增加了一批新的数据, 要修改所建立的模型, 就需要把增加的数据和建模时的所有原始数据全部重新计算一遍。这需要占用计算机的大量内存, 花费许多机时, 而且许多计算是重复的。一个统计预报模型建立后, 总希望该模型能保持良好的使用效果, 这就提出了气象回归预报模型的自适应方法^[1]。本文利用矩阵反演求逆定理, 导出回归预报模型建立时的回归系数估计递推公式, 对 N 组数据建立的回归预报模型之统计参数, 在获得第 $N+1$ 组数据后, 只需将第 $N+1$ 组数据利用递推公式对预报模型的统计参数作适当的修正, 就能得到关于所有 $N+1$ 组数据的气象回归预报模型的统计参数。使用自适应方法来建立气象回归预报模型, 可以节省计算机内存, 减少计算量, 并使预报模型的统计参数随数据组序列的增加而及时修正。该方法计算简单, 即使在未使用计算机的台站, 也可以对所建模型很方便地用手工进行修正。

2 自适应方法递推公式的推导

多元回归模型的一般表达式为

$$Y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m + E \quad (1)$$

式中, E 是误差, m 是选取的预报因子个数, $x_1, x_2, \cdots, x_m$ 为预报因子, Y 是预报对象。

对一个数据序列

$$\{y(t), x_1(t), \cdots, x_m(t) \quad t=1, 2, \cdots, N\}$$

其中, N 是数据组数即资料年代数。我们引进如下向量和矩阵:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \quad \theta = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \quad (2)$$

$$\Phi = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ & & \cdots & & \\ 1 & x_{1N} & x_{2N} & \cdots & x_{mN} \end{bmatrix} \quad (3)$$

其中 $y_i; x_{1i}, \cdots, x_{mi}; e_i$ 分别是 $y(i); x_1(i), x_2(i), \cdots, x_m(i); e(i), i=1, 2, \cdots, N$ 。

利用上述向量和矩阵记号, 式(1)可以写成:

$$Y = \Phi\theta + E \quad (4)$$

θ 的最小二乘估计式为^[2]:

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y \quad (5)$$

式(5)是最小二乘估计的一般公式。其中 Φ 和 Y 都是由 $\{y(t), x_1(t), \cdots, x_m(t) \quad t=1, 2, \cdots, N\}$ 给出的, Φ^T 是 Φ 的转置矩阵。在实际预报业务工作中, 预报因子 $x_1, x_2, \cdots, x_m$ 的值与预报的对象 Y 的值是依序连

贯地得到的。如果我们取得了 N 组数据, 可以用式(5)估计出式(1)的回归系数 $a_0, a_1, \dots, a_m$ 。每当新增加一组数据, 按传统方法, 我们必须用式(5)将所有 $N+1$ 组数据全部重新计算一遍, 才能得到关于 $N+1$ 组数据的最小二乘估计之回归系数。本文提出气象回归预报模型的自适应方法, 对 N 组数据所得的最小二乘估计 $\hat{\theta}(N)$, 在获得了第 $N+1$ 组数据时作适当的修改, 就能得到关于所有 $N+1$ 组数据的最小二乘估计 $\hat{\theta}(N+1)$ 。

为了导出 θ 的递推公式, 我们把观测序列中的 N 作为变动参数并定义下列符号:

由 N 组数据求得的最小二乘估计记作: $\hat{\theta}(N)$; 由 $N+1$ 组数据求得的最小二乘估计记作: $\hat{\theta}(N+1)$; 相应地, 式(2)、式(3)中的 Y, Φ 在有 N 组数据时记为 $Y(N), \Phi(N)$ 。当增加了一组观测数据 $\{y(N+1), x_1(N+1), \dots, x_m(N+1)\}$ 时, $\Phi(N)$ 就增加一行, $Y(N)$ 则增加一个分量, 即

$$\Phi(N+1) = \begin{bmatrix} \Phi(N) \\ \varphi(N+1) \end{bmatrix} \quad (6)$$

$$Y(N+1) = \begin{bmatrix} Y(N) \\ y(N+1) \end{bmatrix}$$

其中, $\varphi(N+1) = (1, x_{1 \ N+1}, x_{2 \ N+1}, \dots, x_{m \ N+1})$ (7)

观测数据组数为 N 时, 由式(5)已知 θ 的最小二乘估计为:

$$\hat{\theta}(N) = [\Phi^T(N)\Phi(N)]^{-1}\Phi^T(N)Y(N) \quad (8)$$

相应地, 观测数据组数为 $N+1$ 时,

$$\begin{aligned} \hat{\theta}(N+1) &= [\Phi^T(N+1)\Phi(N+1)]^{-1} \cdot \\ &\quad \Phi^T(N+1)Y(N+1) \\ &= [\Phi^T(N)\Phi(N) + \varphi(N+1) \cdot \\ &\quad \varphi(N+1)]^{-1} \cdot [\Phi^T(N)Y(N) \\ &\quad + \varphi(N+1)y(N+1)] \quad (9) \end{aligned}$$

为了在推导中便于书写, 暂时省略 $Y(N), \Phi(N)$ 中的 N , 及 $\varphi(N+1)$ 中的 $N+1$, 这时, 式(9)变成:

$$\begin{aligned} \hat{\theta}(N+1) &= [\Phi^T\Phi + \varphi\varphi^T]^{-1} \cdot \\ &\quad [\Phi^TY + \varphi y(N+1)] \quad (10) \end{aligned}$$

在式(9)或式(10)的形式下, 我们清楚地看到, 当观测数据组由 N 个增加到 $N+1$ 个时, 重复计算的时间主要花在求逆矩阵 $[\Phi^T\Phi + \varphi\varphi^T]^{-1}$ 上面。为了简化计算, 我们利用矩阵反演求逆定理^[3], 可得到:

$$\begin{aligned} \hat{\theta}(N+1) &= \hat{\theta}(N) + K(N)[y(N+1) \\ &\quad - \varphi\hat{\theta}(N)] \quad (11) \end{aligned}$$

$$\text{其中, } K(N) = (\Phi^T\Phi)^{-1}\varphi[1 + \varphi(\Phi^T\Phi)^{-1}\varphi^T]^{-1} \quad (12)$$

在式(12)中, $(\Phi^T\Phi)^{-1}$ 是在求 $\hat{\theta}(N)$ 时已经计算了的, 而 $1 + \varphi(\Phi^T\Phi)^{-1}\varphi^T$ 是一个数量, 于是, $[1 + \varphi(\Phi^T\Phi)^{-1}\varphi^T]^{-1}$ 就变成了一个数的倒数运算。

式(11)具有很强的直观意义:

其一, 第 $N+1$ 次估计 $\hat{\theta}(N+1)$, 是第 N 次估计 $\hat{\theta}(N)$ 加上一个修正项 $K(N)[y(N+1) - \varphi\hat{\theta}(N)]$;

其二, 因为 $\varphi\hat{\theta}(N) = \hat{a}_0 + \hat{a}_1x_{1 \ N+1} + \dots + \hat{a}_2x_{2 \ N+1} + \dots + \hat{a}_mx_{m \ N+1}$, 所以, $\varphi\hat{\theta}(N)$ 表示估计得 $\hat{\theta}(N)$ 后, 用式(1)求得 y 的预报值, 而 $y(N+1)$ 则是新的实测值。

如果实测值和预报值相等, 即

$$y(N+1) - \varphi\hat{\theta}(N) = 0$$

那么, 就有 $\hat{\theta}(N+1) = \hat{\theta}(N)$, 表示不必修正 $\hat{\theta}(N)$ 。

反之, 若 $y(N+1) \neq \varphi\hat{\theta}(N)$, 此时, 要得出 $\hat{\theta}(N+1)$, 必须对 $\hat{\theta}(N)$ 加以修正。修正项与偏差项 $y(N+1) - \varphi\hat{\theta}(N)$ 成正比, $K(N)$ 就是增益因子。预报偏差越大, 则修正量越大, 增益因子越大, 修正量也越大。

为了进一步得出 $K(N)$ 的递推关系, 定义

$$P(N) = [\Phi^T(N)\Phi(N)]^{-1} \quad (13)$$

从而

$$K(N) = P(N)\varphi[1 + \varphi P(N)\varphi^T]^{-1} \quad (14)$$

$$P(N+1) = [I - K(N)\varphi]P(N) \quad (15)$$

式中, I 是单位矩阵(与 $P(N)$ 同阶)。

集中式(11)、(14)、(15), 重新添上推导过程中省略的 N 和 $N+1$, 得到如下一组递推公式:

$$\begin{cases} K(N) = P(N)\varphi(N+1)[1+\varphi(N+1) \cdot \\ \quad P(N)\varphi(N+1)]^{-1} \quad (16) \\ P(N+1) = [I - K(N) \cdot \\ \quad \varphi(N+1)]P(N) \quad (17) \\ \hat{\theta}(N+1) = \hat{\theta}(N) + K(N)[y(N+1) \cdot \\ \quad - \varphi(N+1)\hat{\theta}(N)] \quad (18) \end{cases}$$

利用(16)、(17)、(18)这组递推公式,可以递推计算 θ 的值。

用这组递推公式来估计未知参数,涉及到 $P(N)$ 、 $\hat{\theta}(N)$ 的初值选取问题。有两种方案可供采用:

方案1:先取一批观测值(相对 N 数量较小),用常规最小二乘法估计出 θ ,作为初值。此时,必须注意下面的情况,设

$$\Phi(N) = \begin{bmatrix} \varphi(1) \\ \varphi(2) \\ \vdots \\ \varphi(N) \end{bmatrix}$$

$$\text{那么 } \Phi^T(N)\Phi(N) = \sum_{t=1}^N \varphi(t)\varphi(t) \quad (19)$$

如果观测次数小于未知参数的个数,在模型(1)中,即 $N < m+1$,此时,由观测方程解出的 $a_0, a_1, \dots, a_m$ 不唯一, $\Phi^T(N)\Phi(N)$ 不是正定阵^[4]。为了找到正定的 P_0 ,我们必须取 $N_0 > m+1$ ($m+1$ 是未知参数的个数),这时就以

$$P_0 = P(N_0) = [\Phi^T(N_0)\Phi(N_0)]^{-1} \quad (20)$$

$$\hat{\theta}_0 = [\Phi^T(N_0)\Phi(N_0)]^{-1}\Phi^T(N_0)Y(N_0) \quad (21)$$

作为初值。

方案2:人为地给定初值

$$\begin{cases} P_0 = \frac{1}{\epsilon} I \end{cases} \quad (22)$$

$$\begin{cases} \hat{\theta}_0 = 0 \end{cases} \quad (23)$$

ϵ 取值足够小,一般取 $\epsilon = 10^{-4}$ 。可以证明,采用这组初值,经过一定次数的递推计算,就能得到满意的效果。

3 计算实例

盛夏低温灾害是四川省雅安地区农业生产的主要灾害性天气之一,低温往往和阴雨相伴出现,不仅使作物产生生理障碍,而且给

病虫害的滋生、繁殖创造了条件。两者综合作用对大春农业产量影响极大,根据资料统计,凡是雅安8月月平均气温与多年平均值比较偏低 1°C 以上时,该年盛夏均有1—2段低温阴雨天气,反之则无盛夏低温。因此,我们建立雅安8月月平均气温回归预报模型。

预报量 Y :雅安8月月平均气温。

预报因子: x_1 :前年12月500hPa西太平洋副热带高压强度指数; x_2 :上年1月500hPa北半球极涡中心强度指数; x_3 :上年5月500hPa北半球极涡中心强度指数; x_4 :上年2月500hPa亚洲地区纬向环流指数(查 I_2)

多元回归统计模型为:

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$$

矩阵表达式为:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad \theta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ 1 & x_{12} & x_{22} & x_{32} & x_{42} \\ & & \dots & & \\ 1 & x_{1N} & x_{2N} & x_{3N} & x_{4N} \end{bmatrix}$$

使用表1的资料,采用传统方法建立的多元回归方程为:

$$Y = 27.15 - 0.01x_1 - 0.03x_2 - 0.04x_3 - 1.60x_4$$

现在我们采用自适应方法建立回归预报模型,取第二方案的初值:

$$\hat{\theta}(N_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$P(N_0) = \frac{1}{\epsilon} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\epsilon = 0.0001$$

代入递推公式(16)、(17)、(18),用PASCAL语言编一程序,取N=22进行递推,得:

$$Y=27.15-0.01x_1-0.03x_2-0.04x_3-1.60x_4$$

表1 预报因子资料

年份	Y	x ₁	x ₂	x ₃	x ₄	年份	Y	x ₁	x ₂	x ₃	x ₄
1959	25.8	32	3	7	1.11	1970	25.6	0	3	5	1.00
1960	24.3	22	2	8	1.20	1971	24.7	21	3	3	0.96
1961	25.4	23	4	8	0.97	1972	26.4	7	-10	4	1.53
1962	24.5	32	0	3	1.22	1973	25.5	6	-10	5	1.43
1963	25.5	8	-8	6	1.35	1974	23.8	29	-5	0	1.74
1964	24.9	13	-3	4	0.96	1975	26.3	10	-5	6	0.80
1965	23.8	26	-11	-4	1.55	1976	24.2	14	-9	8	1.17
1966	24.0	0	3	2	1.28	1977	25.0	2	2	4	0.99
1967	26.2	16	-1	-1	1.45	1978	25.8	11	8	7	0.34
1968	23.3	9	-9	4	1.58	1979	24.8	14	-2	5	1.33
1969	25.4	7	4	3	0.61	1980	23.7	32	-1	5	1.64

注:x₁-x₄ 资料据中央气象台长期科提供。

表2 新得到的实测值

年份	Y	x ₁	x ₂	x ₃	x ₄	年份	Y	x ₁	x ₂	x ₃	x ₄
1981	25.1	12	5	10	0.85	1986	24.4	2	6	11	1.07
1982	24.8	13	-7	3	0.64	1987	24.8	20	-3	9	1.01
1983	24.3	25	0	6	1.43	1988	23.7	36	-12	1	1.55
1984	24.9	47	-8	5	1.17	1989	24.1	59	-4	8	1.13
1985	25.7	21	-2	1	0.56	1990	25.4	19	-10	1	1.57

从1981年开始,每得到新的实测值(见表2),我们即对模型进行递推修正(此时,只需计算新得到的一组实测值)。

求出逐年预报方程的系数(见表3):

递推预报方法与常规传统预报方法比较曲线见附图。

表3 1982-1990年逐年预报方程系数

年份	a ₀	a ₁	a ₂	a ₃	a ₄
1982	26.54	-0.02	0	-0.03	-1.07
1983	26.56	-0.02	0	-0.03	-1.07
1984	26.60	-0.01	-0.01	-0.03	-1.18
1985	26.61	-0.01	-0.01	-0.03	-1.19
1986	26.65	-0.01	-0.02	-0.04	-1.25
1987	26.65	-0.01	-0.02	-0.04	-1.25
1988	26.69	-0.01	-0.01	-0.04	-1.25
1989	26.70	-0.01	-0.01	-0.04	-1.24
1990	26.67	-0.02	-0.01	-0.04	-1.19

盛夏低温,而自适应递推方法建立的回归模型预报无盛夏低温,实况是1982年未出现盛夏低温。1988年,递推方法预报有盛夏低温,实况为1988年8月24-27日,雅安地区出现了一段低温阴雨天气,由于长期趋势预报准确、超前,全区大春播种期推迟,使水稻抽穗扬花关键期避开了盛夏低温时段,大大减轻了大春作物的损失。

本文在完成过程中承成都高原气象研究所徐裕华高级工程师指导,谨此致谢!

参考文献

1 宋健. 系统控制论. 系统工程理论与实践, 1989, (3)3.
2 黄嘉佑. 气象统计预报试用教材. 北京: 北京大学地球物理系. 1981, 42-44.
3 陈禹六. 大系统理论及其应用. 北京: 清华大学出版社, 1988, 14-15.
4 王萼芳. 高等代数. 上海: 上海科学技术出版社, 1982. 397.

附图 递推、常规两种方法比较

实线: 实测值; 点划线: 递推自适应预报值;
虚线: 常规法预报值

从上图可以看出, 使用自适应方法建立的递推预报模型, 其预报量较常规传统方法稳定, 预报精度有明显提高, 随年代的推移能保持良好的使用效果。经检验, 凡是预报无盛夏低温的年份, 实况均没有出现盛夏低温。1982年, 按传统方法建立的回归模型预报有

