

翟宇梅, 赵瑞星, 高建春, 等. 2014. 遗忘因子自适应最小二乘算法及其在气温预报中的应用. 气象, 40(7): 881-885.

遗忘因子自适应最小二乘算法及其 在气温预报中的应用^{* 1}

翟宇梅¹ 赵瑞星² 高建春¹ 王力维¹ 韩海东¹

1 北京应用气象研究所, 北京 100029

2 中国人民解放军 61741 部队, 北京 100094

提 要: 海量数据的利用是建立自适应预报模型的基础, 但随着数据的不断增加, 新引入数据的作用会逐渐降低, 有可能导致预报模型失效。为克服因数据量增加引起的所谓“数据饱和”现象对天气预报效果的影响, 本文给出了考虑遗忘因子的线性自适应最小二乘建模算法的原理和方法, 并利用该算法进行了最高气温和最低气温预报试验。结果表明, 考虑遗忘因子的线性自适应建模算法优于传统的线性自适应建模算法, 加入遗忘因子可以避免产生“数据饱和”现象, 适当地选择遗忘因子有助于提高模型的预报准确率。

关键词: 遗忘因子, 自适应建模, 最小二乘法, 气温预报

中图分类号: P456

文献标志码: A

doi: 10. 7519/j. issn. 1000-0526. 2014. 07. 012

Forgetting-Factor Adaptive Least Square Algorithm and Its Application in Temperature Forecasting

ZHAI Yumei¹ ZHAO Ruixing² GAO Jianchun¹ WANG Liwei¹ HAN Haidong¹

1 Beijing Institute of Applied Meteorology, Beijing 100029

2 61741 Troops of People's Liberation Army, Beijing 100094

Abstract: A mass of data is the foundation of adaptive forecasting model. However, the role of new incoming data will be gradually reduced and the performance of the model will become poor with the data increasing. In order to overcome the influence of “data saturation” on the weather forecast, the method of adaptive linear least square modeling algorithm considering forgetting factors is developed and applied in max-min temperature forecast. The results show that this adaptive linear least square modeling algorithm considering forgetting factors is superior to the traditional adaptive linear modeling algorithm, it can reduce the effect of “data saturation” by using the forgetting factor, and it is possible to improve the model's forecast accuracy by choosing the appropriate forgetting factors.

Key words: forgetting factor, adaptive modeling, least square algorithm, temperature forecast

引 言

天气系统是时变系统, 了解掌握天气系统演变规律需要建立合理的预报模型, 并对模型参数进行

准确估计。由于天气系统的复杂性, 是否能够建立起与天气系统时变特点最优匹配的预报模型成为提高天气预报准确率的关键, 这就涉及到对预报模型结构的辨识和模型参数的最优估计问题。随着数值天气预报技术的迅速发展以及数值天气预报模式的

* 省部级 2009 年重点科研项目“概率天气预报业务系统”资助

2013 年 7 月 18 日收稿; 2013 年 9 月 16 日收修定稿

第一作者: 翟宇梅, 主要从事天气与气候分析预测技术研究. Email: zhaim1019@sina.com

不断更新,建立在大量数据基础上的自适应建模技术在数值天气预报产品解释方面得到成功应用,明显提高了天气预报准确率(龙从彬,1992;谢炯光等,2003)。常用的自适应建模技术主要有 Kalman 滤波(陆如华等,1994;李日华等,2003;白永清等,2013)、多层递阶(孔玉寿,1988;吴钟浚等,1989)、人工神经网络(林健玲等,2006;孙军波等,2010;钱莉等,2010)等。通过不断引入新数据修改原有模型系数,即通过跟踪时变参数动态地把握天气系统变化特征,使模型适应新的数据特性,是这类建模技术的共同特点。然而,研究发现,随着数据的不断增加,引入的新数据对模型系数的修正能力会逐渐降低,即新信息被旧信息逐渐淹没,这种现象通常称之为“数据饱和”(方崇智等,1988)。在方崇智给出的仿真例子中,当数据总量超过 200 时就会出现“数据饱和”现象,此时参数的估计值趋于不变,系统参数辨识失效。

克服“数据饱和”现象的影响,通常采用降低老数据的作用来实现。常用方法有两种,其一为限定记忆法,其二为遗忘因子法。关于限定记忆法在气象领域的应用已有讨论(赵瑞星等,1999),而遗忘因子法虽然在水位、洪峰流量等水文参数的预报中得到成功应用(袁晶等,2004;2006),但在气象要素预报方面的应用还不多见。

最小二乘法因其算法简单、理论成熟、通用性强、精度较高而成为系统参数辨识的最基本最常用方法,在天气预报领域也被广泛应用。因此,本文主要讨论考虑遗忘因子的线性自适应最小二乘建模算法,并利用该方法进行了最高和最低气温预报试验,比较了遗忘因子不同取值对预报结果的影响。实例计算表明,与传统的建模方法相比,该方法对预报模型参数的估计精度较高,因而为提高天气预报准确率提供了一条新思路。

1 遗忘因子线性自适应最小二乘算法

自适应建模技术的核心是不断引入新的数据修改原有预报模型系数,以使其适应新的数据集合,此过程是通过一个被称作系数估计误差阵的方阵 $\mathbf{P}(k)$ (第 k 步的系数估计误差阵)来实现的,通过不断修正该方阵的值来实现修正预报模型的系数。但是系数估计误差阵 $\mathbf{P}(k)$ 是一个随 k 变化的单调减函数,这一点可从 $\mathbf{P}(k)$ 的表达式看出:

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \mathbf{\Gamma}(k)[\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}(k)^T + \mathbf{I}]\mathbf{\Gamma}(k)^T \quad (1)$$

式中, $\mathbf{X}(k)$ 为因子向量, $\mathbf{\Gamma}(k)$ 为通常所称的卡尔曼增益矩阵, $\mathbf{P}(k)$ 定义为:

$$\mathbf{P}(k) = \left[\sum_{i=1}^k \mathbf{X}(i)\mathbf{X}(i)^T \right]^{-1}$$

按照 $\mathbf{P}(k)$ 的定义,不难看出 $\mathbf{P}(k)$ 是一个正定方阵,因此,式(1)中的 $\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}(k)^T$ 和 $[\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}(k)^T + \mathbf{I}]$ 也是正定方阵。当卡尔曼增益 $\mathbf{\Gamma}(k)$ 不为零矩阵时 $\mathbf{\Gamma}(k)[\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}(k)^T + \mathbf{I}]\mathbf{\Gamma}(k)^T$ 也是正定方阵。因此由式(1)可以得到 $\mathbf{P}(k)$ 是递减的正定方阵,即当 $k \rightarrow \infty$ 时, $\mathbf{P}(k)$ 的元素逐渐趋于零,亦即 $\mathbf{P}(k)$ 将逐渐变成奇异阵。但因 $\mathbf{P}(k)$ 的秩不会随 k 变化,它始终是满秩的,这就意味着,只能是 $\mathbf{P}(k)$ 逐渐变为一个满秩的零矩阵。由 $\mathbf{\Gamma}(k)$ 和 $\mathbf{P}(k)$ 的关系知,当 $\mathbf{P}(k)$ 变为零矩阵时,有 $\mathbf{\Gamma}(k)$ 也变为零矩阵。由于模型系数的修正与 $\mathbf{\Gamma}(k)$ 有关,当 $\mathbf{\Gamma}(k)$ 变为零矩阵时,模型系数修正等于零,这时便会产生“数据饱和”现象。所谓“数据饱和”现象是指随着引入的数据越来越多,新数据的作用会越来越小,其所提供的信息被淹没在老数据的海洋之中。如果自适应算法对新老数据给予相同的信度,那么随着从新数据中获得的信息量相对下降,对算法的修正能力将会慢慢失去。由于常用的自适应算法一般对新老数据都给予相同的信度,因此,在引入数据较多时,往往就会出现这种“数据饱和”现象。此时模型系数失去了修正能力,即使预报误差较大,也无法更新模型系数进而改进预报效果。特别对于天气变化这种时变过程,这种现象将导致模型参数估计不能及时跟踪系统时变特点,更重要的是,当系数估计误差阵 $\mathbf{P}(k)$ 逐渐变为零矩阵时,由于计算机的舍入误差,可能会使 $\mathbf{P}(k)$ 过早地失去正定性和对称性,造成参数估计远离真值。

考虑遗忘因子是一种克服“数据饱和”现象的方法。考虑遗忘因子的线性自适应建模算法,是在自适应建模算法的基础上通过赋予老数据遗忘因子,降低老数据所提供的信息量,相对增加新数据作用的一种能够克服“数据饱和”现象的线性动态建模方法。

通常情况下,最小二乘准则意义下的动态模型为:

$$\mathbf{U}(k) = \mathbf{V}(k)\mathbf{B}(k) + \mathbf{e}(k) \quad (2)$$

式中, $\mathbf{V}(k)$ 是因子数为 m 的前 k 个因子向量 $\mathbf{X}(i)$,

$i=1,2,\dots,k$ 组成的矩阵, $\mathbf{B}(k)$ 是前 k 个模型系数组成的矩阵, $\mathbf{U}(k)$ 是前 k 个预报量 $\mathbf{Z}(i), i=1,2,\dots,k$ 组成的矩阵或向量, $\mathbf{e}(k)$ 是前 k 个随机扰动组成的矩阵或向量。在因子矩阵和预报量矩阵中, 加上衰减因子 $\beta(0 < \beta \leq 1)$ 后因子矩阵和预报量矩阵记作:

$$\mathbf{V}^*(k) = [\beta^{k-1}\mathbf{X}(1), \beta^{k-2}\mathbf{X}(2), \dots, \beta\mathbf{X}(k-1), \mathbf{X}(k)]^T \quad (3)$$

$$\mathbf{U}^*(k) = [\beta^{k-1}\mathbf{Z}(1), \beta^{k-2}\mathbf{Z}(2), \dots, \beta\mathbf{Z}(k-1), \mathbf{Z}(k)]^T \quad (4)$$

通常定义:

$$\mathbf{P}(k) = [\mathbf{V}^*(k)^T \mathbf{V}^*(k)]^{-1} \quad (5)$$

称之为逆相关阵。如果因子数为 m , 那么它是一个 $m \times m$ 的方阵。由最小二乘法很容易得到 $\mathbf{B}(k)$ 的估计:

$$\hat{\mathbf{B}}(k) = \mathbf{P}(k) \mathbf{V}^*(k)^T \mathbf{U}^*(k) \quad (6)$$

由式(3)和(5)可以得到:

$$\mathbf{P}^{-1}(k) = \sum_{i=1}^k \mu^{k-i} \mathbf{X}(i)^T \mathbf{X}(i) \quad (7)$$

式中, $\mu = \beta^2, 0 < \mu \leq 1, \mu$ 称为遗忘因子, 其大小可表征“遗忘速度”。当遗忘因子 $\mu = 1$ 时, 考虑遗忘因子线性自适应建模算法就是普通的线性自适应建模算法。显然由式(7)可以得到:

$$\mathbf{P}^{-1}(k) = \mu \mathbf{P}^{-1}(k-1) + \mathbf{X}(k)^T \mathbf{X}(k) \quad (8)$$

由式(4)、(6)和(8), 有

$$\begin{aligned} \hat{\mathbf{B}}(k) &= \mathbf{P}(k) \mathbf{V}^*(k)^T \mathbf{U}^*(k) \\ &= \hat{\mathbf{B}}(k-1) + \mathbf{P}(k) \mathbf{X}(k)^T [\mathbf{Z}(k) - \mathbf{X}(k) \hat{\mathbf{B}}(k-1)] \end{aligned} \quad (9)$$

通常令

$$\mathbf{\Gamma}(k) = \mathbf{P}(k) \mathbf{X}(k)^T \quad (10)$$

因此, 式(9)可以写成:

$$\hat{\mathbf{B}}(k) = \hat{\mathbf{B}}(k-1) + \mathbf{\Gamma}(k) [\mathbf{Z}(k) - \mathbf{X}(k) \hat{\mathbf{B}}(k-1)] \quad (11)$$

另外, 我们可以利用矩阵反演公式将式(5)改写成:

$$\begin{aligned} \mathbf{P}(k) &= [\mu \mathbf{P}^{-1}(k-1) + \mathbf{X}(k)^T \mathbf{X}(k)]^{-1} \\ &= \frac{1}{\mu} \left[\mathbf{I} - \frac{\mathbf{P}(k-1) \mathbf{X}(k)^T \mathbf{X}(k)}{\mathbf{X}(k) \mathbf{P}(k-1) \mathbf{X}(k)^T + \mu \mathbf{I}} \right] \mathbf{P}(k-1) \end{aligned} \quad (12)$$

由式(10)和(12)可得到:

$$\begin{aligned} \mathbf{\Gamma}(k) &= \mathbf{P}(k-1) \mathbf{X}(k)^T [\mathbf{X}(k) \mathbf{P}(k-1) \times \\ &\quad \mathbf{X}(k)^T + \mu \mathbf{I}]^{-1} \end{aligned} \quad (13)$$

这里式(11)~(13)即为遗忘因子线性自适应最小二乘建模算法公式, 其中, $\mathbf{P}(k)$ 为系数估计误差

方阵, 是一对称矩阵, $\mathbf{\Gamma}(k)$ 为卡尔曼增益矩阵。

一般情况下, 模型系数估计初值阵 $\hat{\mathbf{B}}(0)$ 可由有限样本的回归方法或其他方法确定, 系数估计误差方阵的初值 $\mathbf{P}(0)$ 假定是一对角阵或零矩阵。当期望有较快的收敛速度时, $\mathbf{P}(0)$ 为相对较大的对角阵。在实际计算时遗忘因子 μ 必须选择接近于 1 的正数, 通常不小 0.95。如果系统是线性系统, 且时变特征明显, μ 还应选择大些的值, 一般应选 $0.98 \leq \mu < 1$ (方崇智等, 1988)。当 $\mu = 1$ 时, 上述建模公式就是一般的自适应最小二乘建模算法公式。

2 预报试验

最高、最低气温预报是天气预报的难点, 为了研究遗忘因子线性自适应最小二乘算法在天气预报中的应用效果, 利用式(11)~(13)进行了预报试验。预报因子为欧洲数值预报中心发布的山西岢岚气象观测站周围最近四个格点的 850 hPa 72 h 气温预报值, 预报量为岢岚站未来第三天的最高气温和最低气温。试验所用资料为岢岚气象站的实测数据, 其中最高气温为 1996 年 6—10 月, 最低气温为 1996 年 9 月至 1997 年 1 月。为方便计算, 两组数据去掉缺测后统一取样本数为 150 个。

具体方案如下:

(1) 初值计算: 用前 20 个数据作为初值估计数据, 计算自适应最小二乘模型的系数估计初值 $\hat{\mathbf{B}}(0)$ 和系数估计误差方阵的初值 $\mathbf{P}(0)$;

(2) 模型系数估计: 用前 80 个数据作为训练数据, 对模型系数进行自适应学习训练;

(3) 遗忘因子确定: 用第 81~130 个数据作为预报模型试验数据, 进行预报模型训练和检测, 择优选择预报模型外参数(遗忘因子), 并用此模型预报最后 20 个数据;

(4) 独立样本检验: 用最后 20 个数据作为独立样本检验数据, 对预报结果进行评价。

图 1 为考虑遗忘因子的最高气温预报结果与实际结果的对比, 图 2 为最高气温预报误差。图 3 为考虑遗忘因子的最低气温预报结果与实际结果的对比, 图 4 为最低气温预报误差。

计算了 μ 取值 0.91~1.0 的预报情况。为了综合评价预报效果, 分析不同遗忘因子预报效果差异, 除计算了预报平均绝对误差外, 还用以下两种预报评分标准对预报结果进行了评价。评价结果见表 1

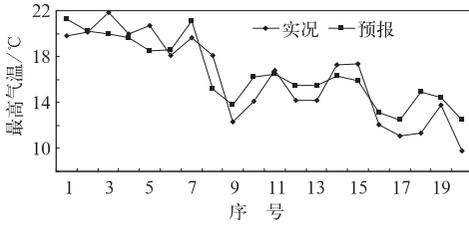
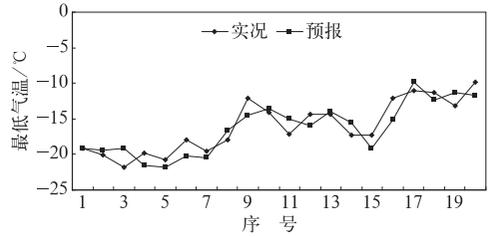
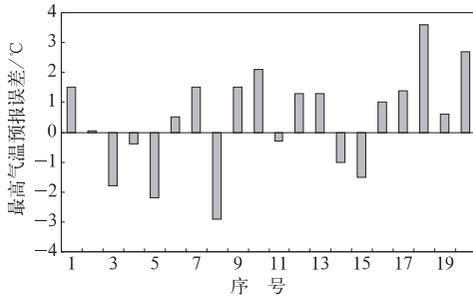
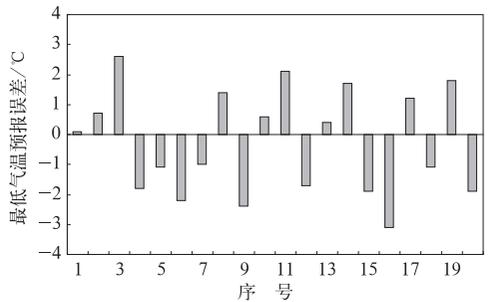
图1 最高气温预报与实况($\mu=0.989$)Fig. 1 Maximum temperature forecasting and observation ($\mu=0.989$)图3 最低气温预报与实况($\mu=0.989$)Fig. 3 Minimum temperature forecasting and observation ($\mu=0.989$)图2 最高气温预报误差($\mu=0.989$)Fig. 2 Maximum temperature forecasting error ($\mu=0.989$)图4 最低气温预报误差($\mu=0.989$)Fig. 4 Minimum temperature forecasting error ($\mu=0.989$)

表1 不同遗忘因子取值时预报结果评价比较

Table 1 Evaluation comparison of forecast results on different forgetting factor values

遗忘因子(μ)	最高气温			最低气温		
	评分标准 1/%	评分标准 2/%	平均绝对误差/°C	评分标准 1/%	评分标准 2/%	平均绝对误差/°C
1.0	60.0	70.0	1.380	55.0	70.0	1.983
0.999	60.0	70.0	1.371	55.0	70.0	1.916
0.998	60.0	70.0	1.362	56.0	70.0	1.837
0.997	60.0	70.0	1.353	56.0	70.0	1.704
0.996	61.5	75.0	1.344	56.0	70.0	1.687
0.995	61.5	75.0	1.335	56.0	70.0	1.642
0.994	61.5	75.0	1.326	57.5	75.0	1.619
0.993	63.0	75.0	1.317	57.5	75.0	1.573
0.992	63.0	75.0	1.308	57.5	75.0	1.570
0.991	63.0	75.0	1.299	59.5	75.0	1.569
0.990	63.0	75.0	1.291	59.5	75.0	1.567
0.989	65.0	75.0	1.282	61.0	75.0	1.540
0.988	63.0	75.0	1.274	59.5	75.0	1.528

(限于篇幅,只给出了 μ 取值0.988~1.0的情况)。

评分标准 1:采用业务预报评定规范中对气温的评分规定,即预报值在实际值的 $\pm 1^\circ\text{C}$ 范围内得1分,在实际值的 $\pm 2^\circ\text{C}$ 范围内得0.6分,在实际值的 $\pm 3^\circ\text{C}$ 范围内得0.3分,其他不得分。

评分标准 2:预报值在实际值的 $\pm 2.0^\circ\text{C}$ 范围内(含 $\pm 2.0^\circ\text{C}$)为正确,得1分,否则不正确,得0分。

由表1可见,取不同的遗忘因子,预报评价结果

不同。但不论遗忘因子取值如何,考虑遗忘因子后($\mu \neq 1$)的线性自适应最小二乘建模算法的预报准确率总是不低于未考虑遗忘因子($\mu = 1$)的预报准确率。特别是当 $\mu = 0.989$ 时,考虑遗忘因子后的预报准确率明显优于不考虑遗忘因子的情况。

上述最高气温和最低气温两个预报实例均表明,遗忘因子的选取是影响参数识别结果的一个重要因素,不同遗忘因子的预报结果不同(本文给出的

两个预报试验中 μ 的最优取值为0.989),但考虑遗忘因子总是有助于提高模型的预报能力。

3 结 论

对于目前天气预报中常用的自适应建模算法,当引入数据较多时,可能会产生“数据饱和”现象,此时模型系数失去修正能力,即使预报误差较大,也会因无法更新模型系数而自适应新的数据特性,进而无法改善预报效果。“数据饱和”和计算机舍入误差的共同作用,还很有可能会使系数估计误差阵失去正定性和对称性,造成预报模型参数估计远离真值。为克服这种现象的影响,可采用降低老数据作用的方法来建模。

考虑遗忘因子的自适应最小二乘算法是降低老数据作用、克服“数据饱和”现象影响的方法之一。该方法能够科学利用系统当前及历史数据所含的信息,通过赋予新、旧数据不同的信度,避免新信息被旧信息淹没,使预测模型更有效地自适应于天气系统的时变特征。本文给出了考虑遗忘因子的线性自适应最小二乘建模算法,并用两个实例进行进一步验证,考虑遗忘因子的线性自适应建模算法总是优于传统的线性自适应建模算法($\mu=1$),加入遗忘因子不仅可以避免“数据饱和”现象的产生,适当地选择遗忘因子有助于提高预报准确率。

本文的研究是初步的。鉴于天气系统变化的复杂性和多样性,可考虑研究能够根据系统时变特征自动调整遗忘因子的实时辨识算法,如对于慢变化的稳定性天气过程和快变化的转折性天气过程可分别选用不同的遗忘因子,既保证有较快的动态响应

速度,又保持有较高的参数辨识精度,通过得到匹配天气过程演变的预报模型参数估计,使模型自适应天气系统动态特性的能力更强,从而达到减小预报误差的目的。

参考文献

- 白永清,林春泽,陈正洪,等. 2013. 基于 LAPS 分析的 WRF 模式逐时气温精细化预报释用. 气象,39(4):460-465.
- 方崇智,萧德云. 1988. 过程辨识. 北京:清华大学出版社,159-165.
- 孔玉寿. 1988. 多层递阶方法在短期天气预报中的应用试验. 气象,14(1):17-22.
- 李日华,李刚. 2003. Klamann 滤波在天气预报中的应用. 数学的实践与认识,33(10):88-92.
- 林健玲,金龙,彭海燕. 2006. 区域降水数值预报产品人工神经网络释用预报研究. 气象科技,34(1):12-17.
- 龙从彬. 1992. 气象统计预报模型的自适应方法. 气象,18(12):29-32.
- 陆如华,何于班. 1994. 卡尔曼滤波方法在天气预报中的应用. 气象,20(9):41-43,40.
- 钱莉,兰晓波,杨永龙. 2010. 最优子集神经网络在武威气温客观预报中的应用. 气象,36(5):102-107.
- 孙军波,钱燕珍,陈佩燕,等. 2010. 登陆台风站点大风预报的人工神经网络方法. 气象,36(9):81-86.
- 吴钟浚,姚嘉玲. 1989. 多层递阶方法在中期天气预报中的应用. 热带气象学报,5(4):378-382.
- 谢炯光,曾琮,纪忠萍. 2003. 中国近 30 年来气象统计预报进展. 气象科技,31(2):67-79,83.
- 袁晶,张小峰. 2004. 基于遗忘因子的 BP 神经网络水文实时预报方法. 水科学进展,15(6):787-792.
- 袁晶,张小峰. 2006. 基于遗忘因子和误差修正的水文实时预报方法研究. 中国农村水利水电,(9):32-35.
- 赵瑞星,翟宇梅,潘剑翔,等. 1999. 概率天气预报技术与应用. 北京:解放军出版社,40-48.