第 28卷第 1期　　南 京 气 象 学 院 学 报　　Vol 28 No 1
2005年 2月　　Journal of Nanjing Institute of Meteorology　　Feb 2005

文章编号: 1000-2022(2005) 01-0058-06

基于主分量的神经网络水位预报模型应用研究

黄海洪,　孙崇智,　金　龙
(广西自治区气象台,广西 南宁　530022)

摘　要: 根据气象和水文资料,以上游面雨量、水位值为预报因子,以西江流域的梧州水位为预报量,发现预报因子与预报量有很好的相关性。采用人工神经网络与主分量分析相结合的方法,建立了梧州水位的预报模型。结果表明,该预报模型对历史样本拟合精度高,试报效果及预报稳定性明显好于传统的神经网络预报模型,可在预报业务中应用。

关键词: 水位预报; 面雨量; 神经网络; 主分量
中图分类号: P338　　文献标识码: A

江河水位的变化对人类生产活动具有重要影响。由于影响水位变化的机制和因素十分复杂,目前江河水位预报工作主要还是依靠统计预报方法。应用较多的有回归、相似等预报方法[1-2]。这些方法的本质大多属于线性的顺序处理技术,由于线性的顺序处理技术对复杂变化过程的拟合有较大的局限性,所以这些预报方法的准确率难以令人满意。人工神经网络预报模型结构灵活并具有模拟非线性关系的能力,因而在气象、水文等领域有较广泛的应用[3-6]。但该预报方法在预报建模中易出现“过拟合”现象,这与神经网络学习矩阵的构造密切相关。本文尝试采用主分量分析方法构造神经网络的学习矩阵,建立基于主分量的神经网络水位预报模型,探索提高江河水位预报准确率的途径,以期为水位预报研究工作开拓新的思路和方法。

1　面雨量的定义和计算

面雨量是指某一时段内一定面积上的平均雨量,一般某一特定区域的面雨量可表示为

$$P = \left(\int_A p\,\mathrm{d}A\right) / A。\tag{1}$$

其中, P 为面雨量, A 为特定区域的面积, p 为有限元 $\mathrm{d}A$ 上的雨量。

流域面雨量计算采用泰森多边形法,先以泰森多边形法求得各雨量站的面积权重系数,然后用各站点雨量与该站所占面积权重相乘后累加得到区域面雨量。计算区域的面雨量为

$$P = f_1 p_1 + f_2 p_2 + \cdots + f_n p_n。\tag{2}$$

收稿日期: 2003-12-09; 改回日期: 2004-03-01
基金项目: 国家自然科学基金资助项目(40075021)
作者简介: 黄海洪(1963-),男,广西南宁人,高工,硕士,主要从事环境气象学研究, E-mail: huanghh63@163.com.

式中, $f_1, f_2, \cdots, f_n$分别为各雨量站用多边形面积计算的权重系数; $p_1, p_2, \cdots, p_n$ 分别为各测站同时期降雨量, P 为流域平均雨量。

本文所计算的面雨量插值到各格点上, 网格距为 15 km × 15 km, 计算区域为 104~ 113 °E, 21~ 27°N。

2　预报建模的基本原理和方法

2 1　主分量分析方法 (PCA, Principal Component Analysis)

主分量分析方法是利用降维的思想, 把原来多指标转化为少数几个相互独立的综合指标的一种多元统计分析方法, 该方法被广泛应用于大气科学的预报和分析研究中。其基本思想是: 设某一预报对象有 n 个预报因子, $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$, 可以通过主分量分析方法构造 n 个新的综合因子变量, $\boldsymbol{Z} = (z_1, z_2, \cdots, z_n)^{\mathrm{T}}$。每一个新的综合因子变量是原因子变量的线性组合, 新的因子变量之间是相互正交的, 即各个因子之间的相关系数为零, 各个新的综合因子变量的特征值明确表示了它对原因子组的贡献大小。

2 2　神经网络方法

人工神经网络是一个可自动提取一组预报变量和另一组自变量之间非线性关系的数据处理系统, 该网络的建立过程称为人工神经网络的训练过程, 是用自适应算法递归迭代求解因变量与自变量之间的非线性关系。训练后的神经网络模型可用来估算或预报预测变量。到目前为止, 已经出现许多神经网络模型及相应的学习方法, 其中反向传播神经网络模型 (简称 BP 模型)是应用较广泛的一种, 这种模型在输入层和输出层之间插入若干个隐含层, 相邻层次之间的神经网络元之间用连接权系数作相互连接, 而各层内的神经元之间没有连接, 本文使用的神经网络模型只含 1个隐含层, $f(x)$为网络激活函数, 采用 S型函数

$$f(x) = 1/(1 + \mathrm{e}^{-x})。\tag{3}$$

具体的算法可参阅文献 [7]。

3　预报模型的建立

3 1　预报因子的初步选取

通常流域的水位与上游面雨量、水位的关系可用一个离散的函数关系式描述

$$H(t) = F[A(t), A(t - \Delta t), \cdots, A(t - n\Delta t), H_1(t), H_1(t - \Delta t), \cdots, H_1(t - n\Delta t)]。\tag{4}$$

式中, H 表示水位, A 表示面雨量, H_1 表示上游水位, F 表示某种函数关系, Δt是数据时间间隔, 单位为 d, n为整数。本文讨论的预报问题是以水位为预报量, 并以其上游面雨量及水位为预报因子。本文从 1999—2002年 75个超过警戒水位 15 m的历史样本中选取 60个样本用于预报建模, 15个样本用于方程检验。通过相关场的普查, 共获得 22个预报因子, 其中 10个因子为梧州以西珠江流域水文站点的水位, 另外 12个因子是梧州以西珠江流域的格点面雨量。根据上述预报因子与梧州距离的远近, 分别取梧州水位出现前的 3 d、2 d或 1 d的量值。预报因子与预报量的相关系数在 0. 30~ 0. 75之间, 达到 0. 05的显著性水平。这说明所选的预报因子与预报量有较好的相关。

3 2　主分量学习矩阵模型的建立

根据 2 1节的主分量分析方法, 对上述各因子作主分量分析。表 1给出了由初选的全部 22个预报因子作主分量计算得出的各主分量特征值及各主分量与预报量之间的相关系数。以主分量特征值相对较大, 并考虑与预报量相关系数高作为取舍标准, 确定神经网络学习矩阵

的输入。从表 1可以看出, 相对来说, 前 4个主分量与预报量的相关系数较高, 并且相对应的特征值也较大, 因此用这 4个主分量来构造神经网络的学习矩阵。本文采用 3层前馈网络计算方法, 隐节点数为 3, 学习因子和动力因子分别取 0. 9和 0. 7。利用 2. 2节给出的公式 (3), 对标准化的学习矩阵作反复的学习训练, 当学习矩阵训练达 10 000次时, 误差函数趋于稳定。由网络模型所确定的连接权系数及阀值可以得到相应的预报模型, 并以此确定的参数和网络结构作为未来独立样本和实际预报的模型。

表 1 各主分量特征值及主分量与预报量的相关系数

Table 1 The characteristic value of the principal components and the correlation coefficients between principal components and predictands

序号	特征值	相关系数	序号	特征值	相关系数	序号	特征值	相关系数	序号	特征值	相关系数
1	7. 33	0. 62	7	0. 70	0. 14	13	0. 18	- 0. 08	19	0. 06	0. 03
2	3. 43	0. 58	8	0. 61	0. 08	14	0. 16	- 0. 07	20	0. 04	- 0. 03
3	2. 71	0. 24	9	0. 38	0. 11	15	0. 12	0. 11	21	0. 03	- 0. 10
4	2. 29	0. 20	10	0. 33	0. 02	16	0. 08	0. 03	22	0. 00	0. 02
5	2. 01	0. 05	11	0. 29	0. 21	17	0. 07	- 0. 06			
6	0. 89	- 0. 05	12	0. 23	0. 06	18	0. 07	0. 01			

4 预报结果分析

4. 1 主分量神经网络预报模型的预报检验分析

由图 1可以看到, 主分量神经网络预报模型利用前 4个主分量对历史样本有很好的拟合效果, 尤其是对历史样本的水位极大、极小值均有较好的拟合效果。为了进行较客观定量的分析, 采用如下 3种统计评价指标进行分析。

(1) 平均绝对百分比误差 (相对误差)

$$E_{MAP} = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - y_t'|}{|y_t|}。 \quad (5)$$

(2) 均方误差

$$E_{MS} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - y_t')^2}。 \quad (6)$$

(3) 平均绝对误差

$$E_{MA} = \frac{1}{n}\sum_{t=1}^{n}|y_t - y_t'|。 \quad (7)$$

计算结果表明, 该预报模型对 60个历史样本的拟合平均绝对误差 (E_{MA}) 为 0. 616 m, E_{MAP} 为 3. 4%, E_{MS}为 0. 717 m。

进一步利用该预报模型, 对 15个独立样本进行了逐次的预报计算 (表 2), 发现该预报模型对独立样

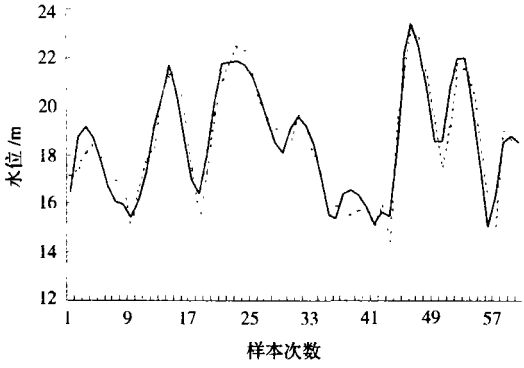

图 1 梧州水位变化曲线

(实线为实测值, 虚线为主分量神经网络预报模型拟合值)

Fig 1 Water level variation of Wuzhou

(solid line: observations; dashed line: the fitting values of the neural network forecast model based on PCA)

本具有较好的预报能力, 预报的最大相对误差为 8 20%, 平均相对误差为 3 05%, 平均绝对误差为 0. 57 m。

表 2　主分量神经网络预报模型对独立样本的预报结果

Table 2　The forecast results of the neural network forecast model based on PCA

样本	观测值 /m	预报值 /m	相对误差 /%	预报误差 /m	样本	观测值 /m	预报值 /m	相对误差 /%	预报误差 /m
1	16. 82	17. 33	2 84	0 48	9	22. 53	23 41	3 89	0 88
2	15. 10	15. 61	3 35	0 51	10	22. 73	23 29	2 47	0 56
3	15. 34	15. 31	0 19	- 0 03	11	22. 55	21 76	3 48	- 0. 79
4	15. 90	15. 84	0 35	- 0 06	12	21. 44	21 40	0 18	- 0. 04
5	16. 50	17. 01	3 10	0 51	13	19. 55	20 01	2 35	0 46
6	17. 60	18. 43	4 69	0 83	14	17. 43	17 95	2 93	0 52
7	20. 32	19. 06	6 18	- 1 26	15	15. 05	16 29	8 20	1 24
8	21. 73	22. 05	1 49	0 32					

4 2　不同预报方法的对比分析

为了进一步对比分析, 根据这 22个预报因子 (表 1), 采用常规的神经网络预报建模方法进行计算分析。本文尝试采用常规的逐步回归筛选因子的方法, 对全部 22个预报因子进行筛选, 初选出 10个主要预报因子, 其中面雨量因子为 7个, 水位因子为 3个, 再对这 10个预报因子用神经网络方法进行预报建模。为了进行较客观的对比分析, 回归因子神经网络预报模型采用与主分量神经网络预报模型相同的 3层前馈网络, 并且学习因子和动量因子同样取 0 9 和 0. 7。对回归因子神经网络预报模型的学习矩阵也作 10 000次训练, 当误差函数趋于稳定时, 训练结束。图 2是由最终的网络权系数和阀值计算得出的预报模型对历史样本的拟合计算结果。

由图 2可以看到, 回归因子神经网络预报模型对历史样本也有很好的拟合效果, 而式 (5)、(6)、(7) 3种统计评价指标的计算结果分别为: E_{MA} = 0 579 m, E_{MAP} = 3 2%, E_{MS} = 0 743 m。由上述计算结果不难看出, 回归因子神经网络预报模型对历史样本的拟合精度与主分量神经网络预报模型的相应结果较相近, 前者的拟合效果略好于后者。

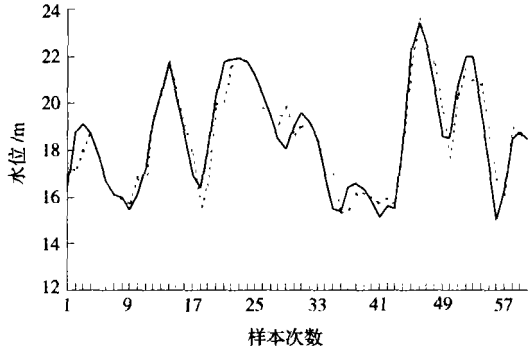

图 2　梧州水位变化曲线
(实线为实测值, 虚线为回归因子
神经网络预报模型拟合值)
Fig 2　Water level variations of Wuzhou
(solid line: observations; dashed line: the fitting values of
the neural network forecast model based on PCA)

在一般情况下, 预报模型对历史样本的拟合精度很高并不一定说明其预报效果也一定很好。神经网络预报方法在实际预报建模的应用中, 最大的障碍就是存在明显的“过拟合”现象, 因此神经网络预报

方法预报模型是否存在“过拟合”现象,是检验其应用价值的重要标准。

表 3为回归因子神经网络预报模型对 15个独立样本进行逐次预报的计算结果,分析表明,该预报模型对独立样本具有一定的预报能力,预报的最大相对误差为 14.32%,15次预报的平均绝对误差为 0.67 m,平均相对误差为 3.73%,这几项指标均大于主分量神经网络预报模型的相应结果。同样利用(5)、(6)、(7)式 3种统计评价指标对两种预报模型进行计算分析(表 4),可以发现回归因子神经网络预报模型对 15个独立样本的 3种统计评价指标的预报误差,均大于该预报模型对历史样本的拟合误差,即回归因子神经网络预报模型比较明显地存在拟合精度高于独立样本预报精度的情况。

表 3 回归因子神经网络预报模型对独立样本的预报结果

Table 3 The forecast results of the neural network forecast model based on regression

样本	观测值/m	预报值/m	相对误差/%	预报误差/m	样本	观测值/m	预报值/m	相对误差/%	预报误差/m
1	16.85	17.60	4.47	0.75	9	22.53	23.52	4.40	0.99
2	15.10	16.40	8.62	1.30	10	22.73	23.01	1.23	0.28
3	15.34	15.57	1.51	0.23	11	22.55	21.77	3.47	-0.78
4	15.90	15.89	0.07	-0.01	12	21.44	21.22	1.03	-0.22
5	16.50	16.50	0.04	0.00	13	19.55	20.04	2.53	0.49
6	17.60	17.86	1.47	0.26	14	17.43	18.37	5.38	0.98
7	20.32	20.94	3.04	0.62	15	15.05	17.21	14.32	2.16
8	21.73	22.69	4.41	0.96					

然而,与此不同的是,主分量神经网络预报模型对历史样本的拟合精度与对独立样本的预报精度差异不大。更为重要的是,主分量神经网络预报模型对独立样本的预报误差明显小于回归因子神经网络预报模型(表 4),这对实际预报是十分重要的。

表 4 两种预报方法的 3种评价指标的比较

Table 4 Comparision of the two prediction models with three evaluation methods

		E_{MAP}/%	E_{MS}/m	E_{MA}/m
主分量神经网络预报模型	模拟	3.4	0.717	0.616
	预报	3.1	0.675	0.565
回归因子神经网络预报模型	模拟	3.2	0.743	0.579
	预报	3.7	0.863	0.667

上述分析表明,回归因子神经网络预报模型存在较明显的“过拟合”现象,主分量神经网络预报模型不但没有明显的“过拟合”现象,而且预报精度也明显好于前者。其主要原因是,主分量神经网络预报模型所采用的预报因子是原因子群作 EOF展开后的 4个主分量,它们不仅浓缩包含了全部预报因子的主要预报信息,同时这 4个主分量是正交的,没有复共线性关系。根据 4个主分量构造的神经网络预报模型,其网络结构规模较小,应用研究成果表明,在能够包含预报量与预报因子间主要相关关系的条件下,网络规模越小预报模型的预报性能越好。相反,回归因子神经网络预报模型不仅网络结构较大,并且预报因子之间存在复共线性关系,所以预报模型出现了历史样本拟合效果好,独立样本预报精度低的问题。

5 小　结

(1)根据统计计算结合实践经验, 挑选出的上游面雨量及水位预报因子与预报量——西江梧州水位有较好的相关性。

(2)主分量分析方法构造神经网络学习矩阵, 可以浓缩众多预报因子的有用信息, 起到显著的降维作用, 使得网络结构规模变小; 同时又由于各主分量的正交性可以使学习矩阵的输入节点之间减少重复信息和噪声。

(3)实例计算结果表明, 主分量神经网络预报模型与回归因子神经网络预报模型相比较, 对历史样本的拟合前者略差于后者, 对独立样本的预报效果前者明显优于后者, 这说明主分量神经网络预报模型使预报泛化能力有显著的提高。

参考文献:

[1] 张素亭, 李致家. 试析卡尔曼滤波与水力学方法结合在洪水实时预报中的应用[J]. 海河水利, 2002, 20(1): 50-52.

[2] 朱定真, 陈玉石. 汛期长江中下游降水量与下游水位增幅的预测[J]. 气象科学, 2002, 22(3): 351-355.

[3] Jin Long, Lou Ying, Lin Zhenshan. Study on mixed model of neural network for farmland flood/drought prediction[J]. Acta Meteor Sinica, 1997, 11(3): 364-373.

[4] 金　龙, 陈　宁, 林振山. 基于人工神经网络的集成预报方法研究和比较[J]. 气象学报, 1999, 57(2): 198-207.

[5] 袁美英, 潘华盛. 利用人工神经网络方法预报松花江汛期水位[J]. 气象科技, 2002, 30(1): 43-46.

[6] 熊立华, 郭生练. 神经网络方法在洪水实时预报中的应用研究[J]. 水电能源科学, 2002, 20(3): 28-31.

[7] 李学桥, 马　莉. 神经网络—工程应用[M]. 重庆: 重庆大学出版社. 1996: 37-44.

Neural Network Prediction Model of Xijiang River Water Level Based on Principal Component Analysis

HUANG Hai-hong　　SUN Chong-zhi,　　JIN Long

(Guangxi Meteorological Observatory, Nanning　530022, China)

Abstract The analysis of the meteorological and hydrological data shows that there is close correlation between the water level of the Xijiang river and the upper reach water level and areal mean rainfall. The new neural network prediction model of water level is established based on the principal component analysis (PCA). The comparison between the new model based on PCA and the traditional neural network model indicates that the new model is significantly superior to the traditional model in prediction accuracy and prediction stability, thus having a good prospect in operational application.

Key words: water level prediction; areal mean rainfall; neural network; principal component analysis