

# 基于耦合主成分法的小波网络模型城市需水预测

段 凯,梅亚东,周研来,颜世杰

(武汉大学 水资源与水电工程国家重点实验室,湖北 武汉 430072)

**摘要:**为了尽可能准确预测城市未来的需水量,以上海市为例,对年用水量的 9 个相关因子进行主成分分析,得到两个综合因子。将两综合因子与用水量的历史数据一起作为输入项,建立一种小波网络模型,以 1980~2005 年的数据为训练样本,采用引入了附加动量项和自适应学习率的 BP 算法进行模型率定,并以 2006~2008 年的数据对模型进行了检验。结果表明:所建模型结构简洁,收敛速度与预测精度均较为理想,在城市需水预测中有着广阔的应用前景。而如何选择最佳隐层数和隐层节点数以及获得更快的收敛速度仍将是今后研究的重点问题

**关键词:**小波网络;主成分分析;改进 BP 算法;需水预测;上海

**中图分类号:**TV211.1 **文献标志码:**A

城市需水量受到城市规模、社会经济发展水平和政府政策等诸多相关因素的影响,是一个典型的多层次复杂非线性系统,合理的需水量预测对于城市产业形态布局,供水系统优化等决策有着重要的意义。传统的需水预测方法主要有时间序列分析法和因果分析法两类。时间序列分析法主要依靠对数据结构和演变特征的分析来挖掘其内部蕴涵的信息,并以此为依据来对未来的情况作出预测,如博克思-詹金斯法、GM(1,1)模型等<sup>[1-2]</sup>。此类方法结构较简单,但仅仅依靠历史数据的发展趋势进行类推,缺乏物理成因上的依据,精度难有保证。而因果分析法主要是建立在供需水量序列与其影响因素之间的物理成因进行分析的基础之上,此类方法需要大量的统计数据,模型结构亦较复杂,如多元回归法、投影寻踪回归、系统动力学等<sup>[3]</sup>。

小波网络模型最早是由法国科学研究机构 IRISA 的 Q. H. Zhang 和 B. Albert 等人提出的<sup>[4]</sup>,近十几年来得到了广泛的应用<sup>[5-7]</sup>。其基本思想是通过以小波元代替神经元来实现小波变换与神经网络的融合,相较于传统的 BP 神经网络具有更强的自学习与逼近能力,而且由于目标函数关于权值是严格凸的,所以其全

局极小解是唯一的。事实上 V. Kreinovich 等于 1994 年已经证明了小波神经网络是逼近单变量函数的最佳逼近器<sup>[8]</sup>。为结合时间序列分析与因果分析两类方法的优点,本文尝试将优选出的影响因子经过主成分分析后,与用水量历史序列一起作为小波网络的输入项,通过改进的 BP 算法来确定小波网络的参数值,从而建立起一种耦合了主成分分析的小波网络模型,解决需水预测中的非线性与模糊性问题。

## 1 年用水量影响因子的主成分分析

本文以上海市为例,参阅历年上海市统计年鉴,收集了上海市 1980~2008 年的年用水量及相关数据。以年用水总量为输出项,结合实际情况遴选出当年的户籍人口、城市化水平(以城市人口占总人口的百分比估算)、人均 GDP、农业总产值、工业增加值、建筑业产值、第三产业产值、人均可支配收入、全社会固定资产投资总额等 9 个因素作为相关因子,它们与用水总量均有较强的相关性,相关系数分别为 0.91,0.93,0.92,0.99,0.89,0.95,0.88,0.93 和 0.94。但这些拟定的因子之间存在着不同程度的多重相关性,一方面会造成运算不稳定、矩阵病态等问题,对分析效果产生

收稿日期:2010-07-01

基金项目:国家自然科学基金重点项目资助(50839005)

作者简介:段 凯,男,博士研究生,主要从事水文学及水资源研究。E-mail:duankaiwu@yahoo.com.cn

一定的负面影响,另一方面,对于这类多维的复杂非线性问题,网络结构的选择也会出现一些困难:网络规模若太小则可能无法收敛,若过大,则在训练中效率不高,而且还会由于过度拟合造成网络性能脆弱,容错性下降。为了避免多重相关性的影响并简化小波网络的拓扑结构,提高模型的效率与稳定性,本文先采用主成分分析法进行信息的提取。

主成分分析法是一种将多维指标转化为少数几个不相关的综合指标的多元统计分析方法,这些综合指标是原指标的线性组合,且互相之间线性无关,即通过降维,以期用最少的变量反映尽可能多的信息。主成分分析的基本步骤可归纳如下。

(1) 归一化处理。为消除量纲的影响,并使分析后所得数据控制在 0~1 之间,以能够满足后文隐层函数的范围要求 (sigmoid 函数的值域为 (0,1)), 对原始数据进行如下归一化处理:

$$x_{it}^* = 0.9(x_{it} - \min_{t=1}^T x_{it}) / (\max_{t=1}^T x_{it} - \min_{t=1}^T x_{it}) + 0.05 \quad (i = 1, 2, \dots, p; t = 1, 2, \dots, T) \quad (1)$$

式中,  $x_{it}$  为第  $i$  个指标的第  $t$  个原始数据。

(2) 求协方差矩阵:

$$Z = \begin{bmatrix} \text{cov}(1,1) & \text{cov}(1,2) & \dots & \text{cov}(1,j) \\ \text{cov}(2,1) & \text{cov}(2,2) & \dots & \text{cov}(2,j) \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}(i,1) & \text{cov}(i,2) & \dots & \text{cov}(i,j) \end{bmatrix} \quad (2)$$

其中,  $\text{cov}(i,j)$  为第  $i$  个指标与第  $j$  个指标的协方差,可按下式计算:

$$\text{cov}(i,j) = \frac{1}{T-1} \sum_{t=1}^T (x_{it}^* - \bar{x}_i^*)(x_{jt}^* - \bar{x}_j^*) \quad (3)$$

式中,  $\bar{x}_i^*$  与  $\bar{x}_j^*$  分别为归一化处理后第  $i,j$  个指标的均值。

(3) 特征分解:

$$Z = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (4)$$

式中,  $\mathbf{\Lambda}$  为  $\mathbf{Z}$  的特征值组成的对角阵;  $\mathbf{U}$  为  $\mathbf{Z}$  的特征向量按列组成的正交阵,它构成了新的向量空间,作为新变量(主成分)的坐标轴,又称为载荷轴。

(4) 根据累积贡献率来确定主成分的个数。前  $m$  个主成分的累积贡献率可按下式计算:

$$\eta_m = (\lambda_1 + \lambda_2 + \dots + \lambda_m) / (\lambda_1 + \lambda_2 + \dots + \lambda_p) \quad (5)$$

式中,  $\lambda_m$  为  $\mathbf{\Lambda}$  中特征值由大到小排列后的第  $m$  个值。

本文中代入相关因子  $x_{it}$  ( $i = 1, 2, \dots, 9; t = 1, 2, \dots, 29$ ) 计算,当提取两个主成分时,累积贡献率已达到 98.76%,可以代表原因子的主要信息,即取  $m = 2$ 。

(5) 提取相应的投影方向,得到  $m$  个新的综合因

子:

$$\mathbf{F}^{T \times m} = \mathbf{x}_{T \times p}^* \mathbf{U}_{p \times m} \quad (6)$$

本文中相应的投影方向即对应的特征向量为 (0.750 8, 0.106 7, -0.201 9, 0.399 1, -0.291 4, -0.071 1, -0.324 6, -0.125 4, -0.117 4) 和 (0.274 3, 0.336 1, 0.342 5, 0.3466, 0.3357, 0.3472, 0.3237, 0.3395, 0.3479), 代入式(6)即可提取出两个新的综合因子。另外,为了将用水量数据的演变特征信息引入模型,本文将历史序列,即预测值前两年和前一年的用水量数据作为模型的第 3、4 个输入项。

## 2 小波网络模型的构造

### 2.1 小波网络原理

传统的 BP 网络模型把一组样本的输入输出问题转化为一个非线性映射问题,从理论上来说,其拟合误差可以达到任意小的程度,但事实上因为其隐层基函数一般为 Sigmoid 函数、双极性 Sigmoid 函数这类互不正交的函数,权重的学习容易出现峡谷型误差曲面,不可避免地存在着收敛速度慢和易陷入局部最优的问题。小波网络是小波变换与神经网络的一种紧致型结合,它用小波函数  $\Psi$  来代替传统神经网络中的隐层节点激励函数,相应的输入层到隐层的权值及隐层阈值则由小波函数的尺度参数  $a$  与平移参数  $b$  所代替,隐节点的输出即为不同尺度  $a$  及不同平移量  $b$  下的一族小波函数值,因而充分继承了二者的优点,能够自适应地调整小波基的形状大小,使得模型具有更强的局部逼近能力。另外,小波基函数是正交或者近似正交的,权重之间相关冗余度很小,故而有着更快的收敛速度,而且小波神经网络的误差函数关于权值是严格凸的,不存在局部极小点<sup>[9]</sup>。小波网络的输出层通常使用线性函数,它将隐层节点的输出进行线性叠加以得到模型的输出。3 层小波网络可表示为:

$$f(x) = \sum_{k=1}^H \omega_k \Psi\left(\frac{x_k - b_k}{a_k}\right) \quad (7)$$

$$x_k = \sum_{j=1}^J \lambda_j x_{kj} \quad (8)$$

式中,  $J$  为输入节点数,  $\lambda_j$  为第  $j$  个输入节点的权重值,  $H$  为隐层节点数,  $\omega_k$  表示第  $k$  个隐层节点的权重值,  $f(x)$  为模型输出。本文的小波函数选择常用的 Morlet 小波<sup>[10]</sup>, 该小波为有限支撑、余弦调制的高斯波,具有良好的时频域分辨性能,其实部为:

$$\Psi(x) = \cos(1.75x) \exp(-x^2/2) \quad (9)$$

### 2.2 模型参数的学习

常见的小波网络模型的参数学习算法有共轭梯度

下降法、BP 算法等。标准 BP 算法存在着收敛速度慢、易陷入局部极小值及出现长时间平台、易振荡等问题。针对 BP 算法的这些缺点,人们提出了各种改进算法,如附加动量法、自适应学习率、牛顿法及弹性 BP 法等<sup>[11]</sup>。小波网络能够有效避免局部极小的问题,且附加动量项 BP 算法能够有效平滑网络的训练路径,提高收敛速度,自适应学习率 BP 算法则能够根据网络误差曲面的情况自适应调整网络的学习步长,从而解决了步长选择上的困难。所以,为了提高算法的效率,减轻训练过程中的振荡,本文将附加动量项与自适应学习率引入到 BP 算法中来实现参数的优选。

定义误差函数为:

$$E = \frac{1}{2} \sum_{p=1}^T [Y_p - f(x_p)]^2 \quad (10)$$

式中,  $T$  为样本数,  $Y_p$  与  $f(x_p)$  分别为样本  $p$  的实际值与模型输出值。

各参数的梯度计算可参考文献[12],在引入附加动量项与自适应学习率后,网络各参数可按下式进行调整:

$$\omega(t+1) = \omega(t) - \eta(1-\alpha) \frac{\partial E}{\partial \omega(t)} + \alpha[\omega(t) - \omega(t-1)] \quad (11)$$

$$\lambda(t+1) = \lambda(t) - \eta(1-\alpha) \frac{\partial E}{\partial \lambda(t)} + \alpha[\lambda(t) - \lambda(t-1)] \quad (12)$$

$$\alpha(t+1) = \alpha(t) - \eta(1-\alpha) \frac{\partial E}{\partial \alpha(t)} + \alpha[\alpha(t) - \alpha(t-1)] \quad (13)$$

$$b(t+1) = b(t) - \eta(1-\alpha) \frac{\partial E}{\partial b(t)} + \alpha[b(t) - b(t-1)] \quad (14)$$

式中,  $t$  为迭代步数;  $\alpha$  为动量因子(本文取为 0.95);  $\eta$  为学习率,其初始值取为 0.006。在模型训练过程中按下式进行调整:

$$\eta(t+1) = \begin{cases} 1.015\eta(t), & E(t) < E(t-1) \\ 0.85\eta(t), & E(t) > 1.02E(t-1) \\ \eta(t), & \text{其他} \end{cases} \quad (15)$$

### 2.3 模型结构的优化

为简化模型的结构,本文采用单隐层网络结构。对于 3 层的神经网络, G. Cybenko 已经证明只要隐节点足够多<sup>[13]</sup>,即可以任意精度逼近一个非线性函数。但一般来说,若隐层神经元过少,网络无法很好地学习,训练速度慢,甚至无法收敛,而如果隐层神经元过多,会造成自由参数过多,网络结构冗余,性能下降。目前尚无较明确的理论来指导隐层节点数的选取,实

际应用中多采用试算法来确定。本文将拟合容许误差设为 0.01,最大迭代次数设为 30 000 次,以 1980 ~ 2005 年的数据作为训练样本,以 2006 ~ 2008 年 3 a 的数据对模型进行检验。通过反复试算发现,对于耦合了主成分分析的小波网络,在隐层节点数取为 8 时,模型已达到容许误差,并有较高且稳定的收敛速度,当取更多的节点时模型的效率开始下降,故模型结构可定为 4-8-1,同理可将 BP 网络(PCA-BPNN)和 9 个相关因子直接作为输入项的小波网络(WN)的模型结构分别定为 4-8-1 和 9-14-1。

### 3 结果比较与分析

本文通过相对误差(relative error, RE)及平均相对误差(average relative error, ARE)来对模型预测精度进行评价,计算公式为:

$$RE_i = |Y_i - f(x_i)| / Y_i \quad (16)$$

$$ARE = \sum_{i=1}^3 RE_i / 3 \quad (17)$$

另外,通过最佳隐层节点数、达到容许误差的迭代次数及训练时间来对模型的整体性能进行评价,对 3 种模型比较结果如表 1 所示。

表 1 模型拟合与预测结果比较

方法	$RE_i / \%$			$ARE / \%$	迭代/次	训练时间/s
	2006 年	2007 年	2008 年			
PCA-WN	1.4033	1.7632	1.6355	1.6007	305	3.70
PCA-BPNN	2.6684	0.4641	5.9550	3.0292	29076	503.39
WN	4.4795	7.6098	10.6095	7.5663	268	5.41

从表 1 可以看到,PCA-WN 模型的平均相对误差为最小,3 a 的相对误差均控制在 2% 以内,且结构精简,所需训练时间为最短。PCA-BPNN 模型也可达到较为理想的精度,但从其训练情况来看,该模型的局部逼近能力有限,引入附加动量项与自适应学习率算法后虽在开始阶段能够迅速收敛,但在达到一定误差范围后便很难进一步逼近,所以该模型效率较低。此外,通过对初始  $E$  值的比较可以看到,小波网络的初始寻优能力要远强于 BP 网络。而 WN 模型收敛速度也较理想,但因输入项较多导致了其结构的臃肿,预测精度也难有保障,存在着较大偏差。所以综合预测精度、模型结构和拟合效率各项评价指标来看,PCA-WN 模型的整体性能为最佳。

#### 参考文献:

- [1] 卫小英,霍丽骊. 博克思-詹金斯预测方法简介[J]. 预测,1984,(增):35-39.
- [2] 李振全,徐建新,邹向涛,等. 灰色系统理论在农业需水量预测中的应用[J]. 中国农村水利水电,2005,(11):24-26.