

文章编号: 1672-3961(2011)04-0119-06

基于小波神经网络的水文时间序列预测

朱跃龙, 李士进, 范青松, 万定生
(淮海大学计算机与信息学院, 江苏 南京 210098)

摘要: 复杂时间序列预测是时间序列分析的主要研究内容之一, 已成为一个具有重要理论和实际应用价值的热点研究领域。基于小波和神经网络组合模型, 提出一种多因子小波预测模型以提高水文时间序列的预测精度。并根据不同小波函数对水文时间序列数据的适应性, 提出了一种基于加权相关系数的小波函数选择准则。以国家重要水文站淮河王家坝站汛期的日流量时间序列预测为例, 对各种常用小波函数进行了实验。结果发现选择得到的 Haar 小波和 B3 Spline 小波函数预测精度较高, 从而验证了小波函数选取准则的有效性; 通过和传统单序列小波神经网络模型比较, 发现提出的多因子小波神经网络模型的预测合格率在不同预见期均提高了 10% 以上, 并且对洪水高流量方向预测合格率提高了 15%。

关键词: 时间序列预测; 小波神经网络; 小波选择

中图分类号: TN 911 **文献标志码:** A

Wavelet neural network model based complex hydrological time series prediction

ZHU Yue-long, LI Shi-jin, FAN Qing-song, WAN Ding-sheng
(School of Computer & Information Engineering, Hohai University, Nanjing 210098, China)

Abstract: Time series prediction is one of the main research topics in time series analysis, which is of great importance both in theoretical and application aspects. To improve the performance of the wavelet neural network model on a complex time series, a novel multi-factor prediction model was proposed. According to the adaptability of different wavelet functions to hydrological time series, a new criterion for the selection of different wavelet functions was also put forward, which was based on weighted correlation coefficients. Finally, the newly proposed method was tested on predicting the daily flow of Wangjiaba Station, which is a very important observation site on Huaihe River. It was found that the chosen Haar wavelet and B3 Spline wavelet can produce higher prediction accuracy, which validates the effectiveness of the selecting principle of wavelet function. Compared with the traditional wavelet neural network for single time series, at least 10% improvement was observed for different predicting periods, and 15% improvement in forecasting the high flow direction during the disastrous flood period. All the experimental results showed that the proposed multi-factor prediction model is effective for complex hydrological time series prediction.

Key words: Time series prediction; wavelet neural network; wavelet selection

0 引言

时间序列数据挖掘研究主要包括: 预测、分类、相似性搜索及序列模式挖掘^[1], 而预测是科学决

策、规划的重要前提。文献[2-8]基于小波分解对不同领域的时间序列预测进行研究, 但只是利用小波分解对单一时间序列进行分解。仅仅依靠预测时间序列自身提供的信息, 对提高预测精度具有一定局限性。通常在物理意义上, 一个系统内的多时间序

列间往往具有相关性。相关的时间序列不仅在原始数据上具有相关性,在不同尺度下也具有相关性。在水文领域,径流实际上跟系统内的其他水文要素具有相关关系,如上下游关系站或相关站的水文要素(包括流量、水位、气温和降水等)都会对径流产生较大影响。

近年来,小波和神经网络结合使用的方法得到广泛关注,利用此组合模型对径流进行预测成为研究热点^[3]。径流序列自身的组成比较复杂,在一个径流中含有多种频率成份,所以有必要对其进行分频率研究,而小波变换则提供了一种便利的时频分析技术。不同领域的时间序列数据具有不同特性,仅利用同一种小波函数进行分解是不合适的。文献[9]从序列重构角度,应用蒙特卡罗方法,通过模拟数据以及黄河利津站和浙江白溪流域的实测水文序列,探讨小波分析中影响小波函数选择的因素,进而建立选择小波函数的依据。研究结果表明序列自身的变化特性是小波函数选择的重要影响因素。文献[10]通过训练多尺度预测模型的实验方法,对比了几种小波函数对多尺度框架下径流预测的影响。对于不同数据,特别是大量数据情况下,使用这种类似特征选择中的包裹器方法确定最优的小波函数比较耗时。在文献[5]基础上,提出了一种多因子小波神经网络预测模型,该模型以多时间序列信息作为输入,不仅包括预测目标时间序列的当前小波系数,还包括与此时间序列相关的其它时间序列的当前小波系数。多因子模型继承了小波网络模型的优点,利用了当前可获得的信息并提取出当前时刻多道时间序列时变特征。针对多因子小波预测模型,提出了一种基于加权相关系数的小波函数自动选择标准,实验结果表明该方法是有效的。

1 模型与方法

1.1 时间序列小波分解

小波分析是 Fourier 分析发展史上一个里程碑式的进展,具有时、频同时局部化的优点,因此被誉为数学“显微镜”。利用离散小波分解可以对时间序列进行多尺度分解,从中提取不同频率区间的成分序列,实现对时间序列的分频研究。

小波分解的快速算法有 Mallat 塔式算法^[11]和 A Trou 算法^[12]。相对于 Mallat 算法二抽一规则, A Trou 算法是无抽取的算法,具有平移不变性^[13],通过提取每个时间点上不同频率带的小波系数信息,可提供最大化的信息量;当数据序列经常更新

时,已经提取的小波系数不需要再重新计算。这种方法更加有利于时间序列的预测,因此本文采用 A Trou 算法进行冗余小波变换。

另外,需要注意的是,由于分解后的系数用于预测时间序列的未来值,因此变换时为了计算 t 时刻的系数,不能用到 t 时刻之后的观测值 $\{X(i), i > t\}$ 。

1.2 小波与神经网络耦合模型

借助于小波的多分辨分解功能,可以从原始序列得到分辨率“从粗到细”的多尺度特征。通过对这些小波特征的学习可以对时间序列潜在的多因素变化过程有更加显性的描述,网络也更加容易“抓住”输入和输出数据之间的内在的联系和规律。

在给出本文提出的多因子小波网络模型之前,首先介绍文献[5]中提出的小波网络模型。该小波网络模型,是在三层神经网络的基础上,以 t 时刻原始信号分解的小波系数 $\{c_1^t, c_2^t, \dots, c_k^t\}$ 作为输入,以 $t+T$ 时刻的原始信号 $f(t+T)$ 为输出而构建的模型(T 是预见期)。小波变换将复杂的时间序列分解成若干不同频率的细节信号和背景信号,细节信号是高频部分,背景信号是低频部分,它们是原始时间序列在不同频率区间上的表现,相对原始序列的比重不同,因而对原始时间序列的预测起到的作用不同。小波网络模型结合了小波分析的在信号处理上的时间域—频率域的多分辨性,和神经网络的强大的非线性逼近特性,因此兼有二者的优点。以上述方式建立的模型称为小波网络模型(wavelet network model WNN)。

如果仅利用单时间序列分解出来的小波系数,对未来进行预测,模型的预测能力有限。从原始时间序列分解出来的小波系数,已经包含对时间序列前期变化中长时间记忆和短时间记忆,加入更多的前期小波系数往往不能对预测目标起到更好的辅助作用。而且前期小波系数的加入,使得模式输入个数增加,相应训练模式集的复杂度也增加,对神经网络结构的确定和训练时间也会有更高要求。对于更加复杂时间序列,仅仅根据本时间序列的前期小波系数,不能对时间序列未来的演变进行更好地刻画。所以自然考虑到将与本时间序列相关的其它时间序列信息也包括到对本时间序列的预测模型的输入中,让神经网络学习多个时间序列的小波系数的时变特征和本时间序列预测目标之间的映射关系,期望可以改善模型的预测能力。文献[14]考虑了河道流量的一种非线性预测方法,将单变量的非线性时间序列预测扩展为多变量的形式,包括与此流量

相关的其它时间序列的信息。阐述了多变量方法的概念基础, 并应用于河道流量预测。结果显示其具有良好的预测性能, 且此模型由于采用不同信息源作为输入而具有伸缩性, 这种技术在水文领域更加具有现实意义。

将小波网络模型应用到径流预测, 不仅能充分利用本站和与本站有关 (一般是上下游关系或者平行关系) 的其它站的流量和水位信息, 而且能进一步利用小波变换多尺度变换得到的特征向量, 可以提高预测模型的精度。本文将此多因子模型称为河道小波网络模型 (river wavelet network model, RWNN)。下面给出利用各相关序列当前小波系数进行预测的模型结构, 假设要预测的站点为 A 与 A 相关的站点为 B 至 I , 则神经网络预测模型结构见图 1。

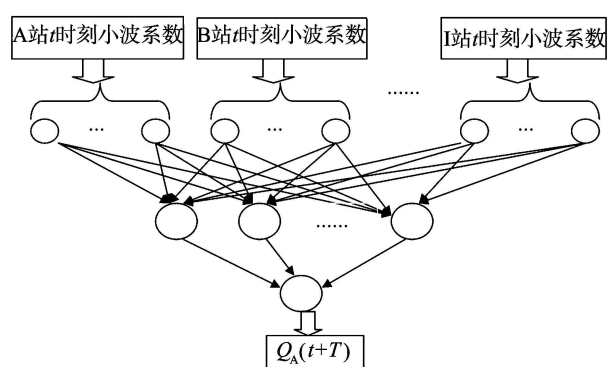


图 1 河道小波网络预测模型

Fig 1 Forecasting model of river wavelet network

2 小波函数的选择

目前, 小波神经网络模型, 或者小波与其他预测工具的浅耦合模型在许多领域中得到关注^[2-8, 10], 但这些研究几乎都没有对小波函数进行选择 (文献 [10] 除外)。已报道研究^[2-8, 10]所采用的小波函数和特性如表 1 所示。表 1 中所列小波均用于离散小波变换, 都是紧支撑的。使用非紧支撑小波, 必须进行截断, 因而会产生截断误差, 而紧支撑小波不存在这样的误差。短的紧支撑小波可以节省计算时间, 实时性好。紧支撑小波具有整数长度的低通滤波系数。

表 1 常用小波性质表

Table 1 Properties of several commonly used wavelets

小波	正交性	滤波系数长度	支撑宽度	对称性
Haar	是	2	1	否
Db2	是	4	3	否
Db3	是	6	5	否
biort_4	否	9	8	是
B3_spline	否	5	4	是

在利用小波分解数据进行预测时对小波性质具有如下要求: (1) 小波的低通滤波系数长度不能太长。由于 d 在特征向量中唯一表示短时间变化信息, 所以小波函数的低通滤波系数的长度不能太大, 一般小于 10 就足够, 更长时间的变化信息可通过高水平的系数表达。(2) 分解后的小波系数应当不失真, 不同水平的序列会比较平滑, 适于建模。据此应选用对称小波或接近对称的小波。在信号分析中, 对称小波或接近对称的小波滤波系数往往具有线性相位或至少具有广义线性相位, 能够避免信号在小波分解和重构时的失真。(3) 时间序列在不同的尺度上往往存在一些明显的或潜在的跃变现象, 因此所选小波应能够提取时间序列中的跃变特征。由于 Haar 小波和 Daubechies 小波适合于提取时间序列跃变特征, 以往的研究中也考虑了这两种非对称正交小波。除了表 1 中所列的小波函数, 本文还考虑了近似对称的 coif 小波和双正交系的 biort_3 和 biort_2 以及反双正交系的 rbior_4 小波。

小波分解作为一种有效的问题分解和特征提取工具, 因此基于小波分解的预测模型在多个领域的研究中被广泛应用。但在具体领域中, 存在不同性质的时间序列数据, 用不同的小波函数对这些数据进行分解, 所产生的成分序列对建立的预测模型性能会有不同程度的影响。小波滤波器由于自身的滤波性质和其具有不同支撑长度的特点, 在对具体的时间序列分解时, 应该选择一个合适的小波, 使分解后的小波系数序列能够更好的提高预测模型的性能。本文针对所研究的河道小波网络模型, 提出了一种基于加权相关系数的小波选择标准。

本文研究的小波网络模型, 使用 t 时刻的不同时间序列的小波系数作为模型的输入, 输出为预测时间序列的 $t+T$ 原始数据。对于多因子预测模型, 设 t 时刻的模型输入包括预测时间序列的小波系数 $d^1(t), \dots, d^1(t), d^2(t), \dots, d^2(t), d^3(t), \dots, d^3(t), d^4(t), \dots, d^4(t), d^5(t), \dots, d^5(t), d^6(t), \dots, d^6(t), d^7(t), \dots, d^7(t), d^8(t), \dots, d^8(t), d^9(t), \dots, d^9(t), d^{10}(t), \dots, d^{10}(t)$, 输出为原始时间序列的未来值 $x_s(t+T)$ 。预测目标 $x_s(t+T)$ 也含有各个频率区间上的对应成分, 即不同水平上的小波系数 $d^1(t+T), \dots, d^1(t+T), d^2(t+T), \dots, d^2(t+T), d^3(t+T), \dots, d^3(t+T), d^4(t+T), \dots, d^4(t+T), d^5(t+T), \dots, d^5(t+T), d^6(t+T), \dots, d^6(t+T), d^7(t+T), \dots, d^7(t+T), d^8(t+T), \dots, d^8(t+T), d^9(t+T), \dots, d^9(t+T), d^{10}(t+T), \dots, d^{10}(t+T)$ 。

根据时间序列分析中关于平稳时间序列预测理论^[15], 设有预测目标时间序列 $\{X(t), t=1, \dots, n\}$ 的 T 时间间隔的自相关系数 ρ_T^{xx} , 又存在与 $\{X(t), t=1, \dots, n\}$ 相关的时间序列 $\{Y(t), t=1, \dots, n\}$, 且两者之间的 T 时间间隔的相关系数 ρ_T^{xy} , 则 ρ_T^{xx} 和 ρ_T^{xy} 越大, 预测目标时间序列 $\{X(t), t=1, \dots, n\}$ 的预测方差就越小。

基于这个思想提出多因子小波预测模型小波选

择的评判方法,即对不同水平的小波分解序列、统计预测目标时间序列与自身不同时间偏移的输入小波序列自相关系数以及预测目标时间序列与多因子输入时间序列的小波系数序列相关系数,最后采用加权的方法综合各个水平上的相关系数。不同的小波滤波器,产生的小波系数序列频率区间也是不同的,所以必须采用同一小波对序列进行分解。例设对输入为 $d^a_1(t), \dots, d^a_3(t)$ 和 $d^b_1(t), \dots, d^b_3(t)$, $c^a_3(t)$ 输出为 $x_k(t+1)$ 的预测模型,计算方法如表 2 本文利用最后一行的加权向量来比较不同小波函数的结果。

表 2 二因子小波网络输入的小波函数选择指标计算方法

Table 2 Calculation method of wavelet function selecting criteria with two factors

输入	d_1 系数	d_2 系数	d_3 系数	c_3 系数
序列 b	$\rho^{ab}_{d_1 1}$	$\rho^{ab}_{d_2 1}$	$\rho^{ab}_{d_3 1}$	$\rho^{ab}_{c_3 1}$
序列 a	$\rho^{aa}_{d_1 1}$	$\rho^{aa}_{d_2 1}$	$\rho^{aa}_{d_3 1}$	$\rho^{aa}_{c_3 1}$
加权系数	$\rho^{ab}_{d_1 1} + \beta \rho^{aa}_{d_1 1}$	$\rho^{ab}_{d_2 1} + \beta \rho^{aa}_{d_2 1}$	$\rho^{ab}_{d_3 1} + \beta \rho^{aa}_{d_3 1}$	$\rho^{ab}_{c_3 1} + \beta \rho^{aa}_{c_3 1}$

下面给出表 2 中相关系数计算方法。假设时间序列为 $\{X(1), X(2), \dots, X(n)\}$, 另一相关时间序列为 $\{Y(1), Y(2), \dots, Y(n)\}$, n 为序列长度。则两者相关系数理论定义如公式 (1), 具体计算方法如公式 (2)。

$$\rho_T = \rho_{t+t-T} = \frac{\text{Cov}(X(t), Y(t+T))}{\sqrt{\text{Var}(X(t)) \text{Var}(Y(t+T))}} \quad (1)$$

$$\rho_T = \frac{\frac{1}{N-T} \sum_{t=1}^{N-T} (X_t - \bar{X})(Y_{t+T} - \bar{Y})}{\sqrt{\frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2 \cdot \frac{1}{N} \sum_{t=1}^N (Y_t - \bar{Y})^2}} = \frac{\frac{N}{N-T} \sum_{t=1}^{N-T} (X_t - \bar{X})(Y_{t+T} - \bar{Y})}{\sqrt{\sum_{t=1}^N (X_t - \bar{X})^2 \sum_{t=1}^N (Y_t - \bar{Y})^2}} \quad (2)$$

3 实验结果与分析

3.1 小波函数选择实验结果

小波函数的选择结果和具体研究对象有关, 本文的研究对象是国家重要水文观测站王家坝站的汛期日流量过程。对比不同年份王家坝站的汛期日流量过程, 发现该站 2003 年日流量序列较为典型, 选择 2003 年汛期日流量序列作为小波函数适应性选择的数据资料, 实验资料还包括 3 个主要上游站(班台、息县和潢川)的 2003 年同期汛期日流量数据。本实验选用如下小波对上述 4 个站点的 2003 年日流量序列数据进行 3 水平小波分解, 参与比较

的小波与其相应的低通滤波器如表 3 所示。

表 3 小波函数及其低通滤波器

Table 3 Wavelet function and its low-pass filter

小波名	低通滤波器系数					
B3 spline	0.062 5	0.250 0	0.375 0	0.250 0	0.062 5	
Haar	0.500 0	0.500 0				
db2	-0.091 5	0.158 5	0.591 5	0.341 5		
db3	0.024 9	-0.060 4	-0.095 5	0.325 2	0.570 6	0.235 2
coif	-0.011 1	-0.051 4	0.272 1	0.602 9	0.602 9	-0.051 4
bio1.3	-0.062 5	0.062 5	0.500 0	0.500 0	0.062 5	-0.062 5
bio2.2	-0.125 0	0.250 0	0.750 0	0.250 0	-0.125 0	
bio4.4	0.026 7	-0.016 9	-0.078 2	0.266 9	0.602 9	0.266 9
	-0.078 2	-0.016 9	0.026 7			
rbio4.4	-0.045 6	-0.028 8	0.295 6	0.557 5	0.295 6	-0.028 8

因为王家坝站点的日流量序列跟自身的成分相关性较大, 所以令其权重系数 $\alpha_{wb} = 0.6$ 而班台和潢川站均在小支流上, 息县站在大支流上, 所以令 $\alpha_{xb} = 0.2$, $\alpha_{bt} = 0.1$, $\alpha_{hc} = 0.1$ 即权重分配为 $\{\alpha_b, \alpha_{xb}, \alpha_{bt}, \alpha_{hc}, \alpha_{wb}\} = \{0.1, 0.2, 0.1, 0.6\}$ 。令 $T=1$ 时, 各个不同的小波函数对时间序列分解结果的相关性统计结果见表 4 (由于篇幅所限, 省略了加权之前的具体数据, 文中只给出加权后的相关系数)。

表 4 各小波分解后加权相关系数

Table 4 Weighted correlation coefficient of wavelet decomposition

小波函数	d_1	d_2	d_3	c_3
B3 spline	0.493 3	0.768 1	0.853 8	0.962 7
Haar	0.381 3	0.707 5	0.829 0	0.957 2
db2	0.228 6	0.629 9	0.800 9	0.949 2
db3	0.165 2	0.621 7	0.803 0	0.942 7
coif	0.183 5	0.617 2	0.798 2	0.947 0
bio1.3	0.287 5	0.684 3	0.821 0	0.951 8
bio2.2	0.326 8	0.499 7	0.762 7	0.924 5
bio4.4	0.166 9	0.561 1	0.787 3	0.941 9
rbio4.4	0.173 0	0.647 2	0.808 0	0.944 0

从表 4 可知, B3 spline 小波分解后序列与目标序列各对应子序列相关系数最大, Haar 小波比 B3 spline 的结果略差。发现, db2 比 Haar 小波的统计结果要差, 这说明同一家族中支撑宽度大的小波分解出来的小波系数, 并不会使预测更加准确。如果选择一个不适合的小波对王家坝等 4 站的日流量数据进行分解, 将导致很差的统计结果, 例如 bio2.2 小波, 在各个水平上的统计结果都较差, 说明其对此 3 站日流量数据进行分解未能有效提取出对预测有利的小波系数特征。

3.2 预测结果及分析

以王家坝站日流量作为预测对象, 选取与之主

要相关的 3 个观测站: 班台、息县、潢川。此 3 站分别为上游三大支流上的站点, 数据时间跨度为 2000~2007 年, 仅包括每年汛期的日流量数据, 具体为 2000~2005 年数据区间均为 6 月 1 日至 10 月 1 日, 2006 年数据区间 5 月 1 日至 10 月 1 日, 2007 年数据区间 5 月 15 日至 10 月 1 日。本文将 2000~2005 年的数据作为训练集, 2006 年和 2007 年的数据作为测试集。

(1) 河道小波网络模型的建立

根据 3.1 节小波函数选择结果可以看出, Haar 小波和 B3 spline 小波分解的数据对王家坝站的河道小波网络模型的适应度在相关系数的意义上是较好的, 所以本文采用这两种小波对王家坝、班台、息县和潢川 4 站的日流量序列的分别做 3 水平分解。然后以此 4 站的当日小波系数作为神经网络的输入, 以王家坝站的预测目标 (1 d 2 d 3 d 预测期的日流量) 作为输出, 对两种小波分解的预测结果进行比较。

在进行神经网络模型训练之前, 首先对输入做 Mid-Range 标准化处理, 对输出流量做自然对数变换后也将其规范化到 $[-1, 1]$, 然后利用 Matlab 的神经网络工具箱来实现神经网络模型的训练。在标准 BP 算法的基础上, 使用 LM 数值优化方法来加快神经网络训练速度, 并使用贝叶斯正则化方法保证训练网络的泛化能力。采用试错法确定最优的神经网络隐层节点数目。对于应用 Haar 小波的神经网络模型, 1 d 2 d 和 3 d 预测期最优隐层节点分别为 4、5 和 5 而对于应用 B3 spline 小波的神经网络模型, 1 d 2 d 和 3 d 预测期最优隐层节点分别为 5、4 和 6。然后对不同的初始权值, 采用多次训练的方法选取最优模型。最后用训练得到的网络模型对测试集进行模拟, 产生预测输出。网络的原始输出需要进行反 Mid-Range 标准化, 得到数据 Y 。由于原始流量作为输出做了自然对数变换, 所以需要对 Y 进行反自然对数变换, 最终得到原始数据的预测结果。

(2) 结果评价指标

对于时间序列预测, 常用的评价标准有平均绝对误差 (E_A) 和平均相对误差 (E_R)。其计算公式如公式 (3) 和 (4):

$$E_A = \frac{1}{T} \sum_{i=1}^T |P_i - a_i| \tag{3}$$

$$E_R = \frac{1}{T} \sum_{i=1}^T \frac{|P_i - a_i|}{a_i} \tag{4}$$

其中 P_i 、 a_i 分别是 T 个观测中第 i 个观测的预测值

和实际值。

另外, 按照《水文情报预报规范》合格率定义, 计算相对误差在 20% 之内的合格预测数目, 得到合格率, 记为 E_{20} 。又由于是洪水预报, 我们还考虑了预测方向正确率以及发生洪水时的高流量方向正确率。方向正确率, 又叫方向变差对称值 (V_{DS}), 如式 (5), 其中 P 为预测期。为了比较高流量方向判断的准确度, 对流量大于 500 的预测统计其方向正确率, 记为 V_{DSH} 。

$$V_{DS} = \frac{1}{T - P} \sum_{t=P}^T \Phi((a_t - a_{t-P})(P_t - P_{t-P}))$$
$$\Phi(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{5}$$

(3) Haar 小波和 B3 spline 小波分解建模结果比较

表 5 给出了基于 Haar 小波分解和 B3 spline 小波分解的小波网络模型在测试集上的预测结果。

表 5 Haar 和 B3 spline WNN 在测试集上不同预测期结果的比较

Table 5 Comparison of Haar and B3 spline wavelet over different forecast periods on the test set

评价指标	1 d 预测期		2 d 预测期		3 d 预测期	
	Haar	B3 spline	Haar	B3 spline	Haar	B3 spline
E_A	40.4622	52.3562	91.6689	108.8331	149.8930	194.4273
E_R	0.0910	0.0978	0.2086	0.1895	0.3219	0.3464
E_{20}	0.8885	0.9015	0.6051	0.6298	0.4051	0.3846
V_{DS}	0.7256	0.7414	0.7418	0.7308	0.6813	0.6809
V_{DSH}	0.8256	0.8605	0.6977	0.6279	0.6163	0.6279

从表 5 可知, 在 1 d 预测期上, B3 spline 的 E_{20} 指标要好于 Haar 的指标; 在 E_A 和 E_R 指标上基于 B3 spline 小波分解的小波网络模型要比基于 Haar 小波分解的模型稍差; B3 spline 的 V_{DS} 和 V_{DSH} 更好些; 在 2 d 预测期上, B3 spline 的 E_R 和合格率指标均要好于 Haar 的指标, 但是在合格率仅约为 60% 的情况下, 从 V_{DS} 和 V_{DSH} 指标上看, Haar 小波的预测结果稍好于 B3 spline 小波的预测结果; 在 3 d 预测期上, 在合格率很低的情况下, V_{DSH} 显示 B3 spline 小波预测结果稍优于 Haar 小波的预测结果。

总体上 B3 spline 小波的预测结果要好于 Haar 小波, 这也验证了上一节提出的小波函数选择方法的正确性。

(4) 与单序列的小波网络模型结果比较

表 6 显示了测试集上 RWNN 和 WNN 在不同评价指标上的统计结果。

从表 6 可知, RWNN 模型在不同预测期上的所有指标均优于 WNN 模型, 这充分体现了扩展模型预测性能的提高。按照《水文情报预报规范》评价

标准,对于 1 d 预测期,河道小波网络模型满足甲级预报的标准,小波网络模型仅满足乙级预报的标准;对于 2 d 的预测期,RWNN 满足丙级预报标准,可用于参考性预报,而 WNN 精度低于丙级;对于 3 d 预测期情况,两个模型精度均低于丙级,但是从 V_D 和 V_{DSH} 可以看出,河道小波网络模型在方向预测性能上表现更好,特别是在高流量的方向预报正确率(V_{DSH})上,RWNN 高出 WNN 15 个百分点,适合于做 3 d 预测期的参考性估报。同时,在 V_{DSH} 指标上,1 d 预测期的 RWNN 比 WNN 高出 12 个百分点,达到 86%,能更加可靠地回答在洪峰变化过程中所关注的“次日洪水流量是升高还是降低”的问题。

表 6 测试集上 WNN 和 RWNN 在不同预测期上的结果比较
Table 6 Comparison of WNN and RWNN on the test set with different forecast periods

评价 指标	1 d 预见期		2 d 预见期		3 d 预见期	
	WNN	RWNN	WNN	RWNN	WNN	RWNN
E_A	56.9227	52.3562	126.3932	108.8331	215.650	194.4273
E_R	0.1164	0.0978	0.2537	0.1895	0.4088	0.3464
E_D	0.8147	0.9015	0.4754	0.6298	0.2558	0.3846
V_{DS}	0.6702	0.7414	0.6466	0.7308	0.6109	0.6809
V_{DSH}	0.7442	0.8605	0.5930	0.6279	0.4767	0.6279

4 结束语

本文在小波预测模型的基础上,建立了适应水文领域多时间序列数据的多因子小波网络预测模型。多因子小波网络预测模型在原有数值计算模型基础上,适当考虑了河道物理影响因素,计算快捷,具有良好的可扩展性和应用价值。以国家重点水文观测站王家坝站的日流量序列作为研究对象,验证了小波函数选取准则的有效性;并利用王家坝及上游的班台、息县和潢川三站日流量数据,建立河道小波网络模型,与传统单序列小波神经网络模型比较实验证明河道小波网络模型预测性能明显优于小波网络模型,提高了 1 d 2 d 预报期的预报精度等级;特别在高流量数据变化方向预测显示出河道小波网络模型对 1 d 2 d 和 3 d 预测期洪水流量的变化预测具有指导性作用,从而为正确进行防洪决策提供有效信息。

参考文献:

[1] ODED Maimon, LOR Rokach. Data mining and knowledge discovery handbook[M]. New York: Springer, 2005.
[2] CHEN Yuehui, YANG Bo, DONG Jiwen. Time series

prediction using a local linear wavelet neural network[J]. Neurocomputing, 2006, 69(6): 449-465.
[3] DASH P K, NAYAK M, SENAPATIM R, et al. Mining for similarities in time series data using wavelet based feature vectors and neural networks[J]. Engineering Applications of Artificial Intelligence, 2007, 20(2): 185-201.
[4] GAN Xiaobing, LIU Ying, and AUSTIN FR. A Prediction method for time series based on wavelet neural network[J]. //Proceedings of CIS2005. New York: Springer, 2005, 902-908.
[5] WANG Wensheng, XIONG Huakang, DING Jing. Study on wavelet network model and its application to the prediction of daily discharge[J]. Advances In Water Science, 2004, 15(3): 382-386.
[6] BENAOUA D, MURTAGH F, STARCK J L, et al. Wavelet based nonlinear multiscale decomposition model for electricity load forecasting[J]. Neurocomputing, 2006, 70(1-3): 139-154.
[7] CHEN Di, FENG Hahang, LIN Qingjia, et al. Multi-scale Internet traffic prediction using wavelet neural network combined model[J]. //Proceedings of First International Conference on Communications and Networking in China, Piscataway: IEEE, 2006, 1-5.
[8] LIA Guo, QIN Zheng. Moving windows quadratic autoregressive model for predicting nonlinear time series[J]. Chinese Journal of Computers, 2004, 27(07): 1004-1008.
[9] SANG Yanfang, WANG Dong. Wavelets selection method in hydrologic series wavelet analysis[J]. Journal of Hydraulic Engineering, 2008, 39(3): 295-300.
[10] LIU Suyi, QUAN Xianzhang, ZHANG Yongchuan. Influence on result of runoff analysis using different wavelet function[J]. Hydroelectric Energy, 2003, 21(1): 29-31.
[11] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.
[12] SHENSAM J. The discrete wavelet transform: wedding the A-Trous and Mallat algorithms[J]. IEEE Transactions on Signal Processing, 1992, 40(10): 2464-2482.
[13] SARISARRAF H, BRZAKOVIC D. Shift invariant discrete wavelet transform[J]. IEEE Transactions on Signal Processing, 1997, 45(10): 2621-2630.
[14] PORPORATO A, RDOLFI L. Multivariate nonlinear prediction of river flows[J]. Journal of Hydrology, 2001, 248(1-4): 109-122.
[15] BROCKWELL P, DAVIS A. Time series: theory and method[M]. New York: Springer, 2002.

(编辑:陈燕)