

基于多尺度相关向量机的城市日用水量预测

白云^{1a}, 谢晶晶^{1b}, 王晓雪², 李川³

(1. 安徽科技学院 a. 建筑学院; b. 资源与环境学院, 安徽 233100; 2. 重庆市南岸区环境监测站, 重庆 400060;
3. 重庆工商大学 装备系统服役健康保障国际联合研究中心, 重庆 400067)

摘要: 为提高日用水量预测精度, 提出一种基于多尺度相关向量机的预测模型。通过静态小波分解将用水量非平稳时间序列分解为不同尺度的平稳时间序列, 然后在分解后的各子序列分别建立相关向量机回归模型进行预测, 最后通过小波逆变换将各子序列预测结果整合得出原始用水量时间序列的预测值。在实例分析中分别利用多尺度关联向量机模型和单尺度相关向量机预测模型对实际用水量进行预测分析。结果表明, 前者具有更高的预测精度, 可应用于城市日用水量的预测。

关键词: 多尺度; 相关向量机; 日用水量预测; 小波逆变换; 重庆市

中图分类号: TU991.31

文献标识码: A

文章编号: 1672-643X(2016)03-0039-04

Forecast of urban daily water consumption based on multi-scale relevance vector machine

BAI Yun^{1a}, XIE Jingjing^{1b}, WANG Xiaoxue², LI Chuan³

(1. a. College of Architecture; b. College of Resource and Environment, Anhui Science and Technology University, Anhui 233100, China; 2. Nanan District Environmental Monitoring Station of Chongqing, Chongqing 400060, China; 3. Research Center of System Health Maintenance, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: In order to improve the forecast precision of daily water consumption, the paper proposed a forecast model based on multi-scale relevance vector machine (MSRVM). The non-stationary time series of daily water consumption are decomposed into different scales stationary time series by stationary wavelet transform. Then it established the regression model of relevance vector machine in each scale to predict respectively, and finally employed the forecast results of the RVM outputs at all the scales to reconstruct the forecast value of the original daily water consumption time series through the inverse wavelet transform. Application examples show that compared with mono-scale RVM model, the MSRVM model has the better precision and can be applied in the forecast of daily water consumption.

Key words: multi-scale; relevance vector machine; daily water consumption forecast; inverse wavelet transform; Chongqing City

1 研究背景

重庆市为全国400个缺水城市之一,属于中度缺水地区,其中12个区县属于重度缺水地区。自直辖以来,随着经济高速发展,城市化进程加快和人民生活水平提高,城市用水量大大增加,现有水资源和供水设施不能满足城市的用水量需求,供需矛盾日

益突出,所以,城市供水系统的优化调度尤为重要。而有效、准确的城市用水量预测是城市给水系统工程模拟和调度决策的基础和前提。

传统用水量预测模型多为线性模型系统,算法简单、速度较快,但无法解决非线性的实际用水量时间序列问题。因此,混沌理论^[1]、灰色理论^[2]、人工神经网络^[3]、支持向量机^[4]等非线性系统黑箱建模

收稿日期:2015-09-28; 修回日期:2015-12-03

基金项目:安徽高校自然科学研究重点项目(KJ2016A168);安徽科技学院校级重点学科建设(AKZDXK2015B01);教育部留学回国人员科研启动基金(教外司留[2013]693号);重庆市研究生教育教学改革研究项目(yjg143015)

作者简介:白云(1985-),男,山西阳泉人,博士,研究方向:水资源预测与管理。

通讯作者:谢晶晶(1985-),女,安徽安庆人,硕士,工程师,研究方向:资源管理及污染防治研究。

方法引入用水量预测领域。为提高预测精度,也提出了多种组合模型^[5-7]。但以上模型立足于全局预测,对局部时间序列规律的变化未做考虑,模型结构未最大化的表达时间序列内在规律,从而导致预测结果不太理想。为了克服以上缺点,笔者从局部时间序列入手,将用水量时间序列分解为多尺度的子序列,然后在各个子序列中建模预测,最后将各子序列预测值整合为最终用水量预测值。

2 多尺度相关向量机预测模型

多尺度相关向量机预测模型(MSRVM)基本原理是通过小波分解将用水量时间序列分解为多个尺度的子序列,然后在各个子序列中应用相关向量机(RVM)回归建模进行预测,最后通过小波逆变换整合各子序列,得出最终预测值。

2.1 多尺度分解

静态小波变换(SWT)与其他基于 Mallat 的小波分解方法不同,是通过滤波器的抽样代替对时间序列的抽样,得到与原始序列长度一样的细节系数,能够实现平移不变性的小波分解方法。因此,静态小波变化是一种冗余的、非降维的、平稳的小波变换,有利于时间序列的分解和重构。所以,在建模过程中,利用静态小波分解对用水量时间序列 $X = (x_1, x_2, \dots, x_N)$ 进行尺度分解。分解尺度为 n 时, X 被分解为 n 层细节系数和 $n+1$ 层近似系数,为方便描述,每个分解尺度下 X 的分解系数为 $X_j = (x_{j1}, x_{j2}, \dots, x_{jN}), j \in [1, n+1]$ 。

2.2 相关向量机回归算法

相关向量机^[8]与支持向量机一样,是基于核函数映射将低维空间非线性问题转化为高维空间的线性问题。相关向量机的训练是在贝叶斯框架下进行的,在先验参数的结构下基于主动相关决策理论来移除不相关的点,从而获得稀疏化的模型。由于在样本数据的迭代学习过程中,大部分参数的后验分布趋于零,而非零参数所对应的学习样本,与决策域的样本并不相关,只代表数据中的原型样本,因此称这些样本为相关向量,体现了数据中最核心的特征。同支持向量机相比,相关向量机最大的优点就是极大地减少了核函数的计算量,并且也克服了所选核函数必须满足 Mercer 条件的缺点。

通过多尺度分解后,在各子序列中建立相关向量机回归模型,如下:

$$\hat{y}_j = f_j(X_j; \omega_j) + \varepsilon_j \quad (1)$$

式中: \hat{y}_j 为 j 尺度预测值向量, $\hat{y}_j = (\hat{y}_{j1}, \hat{y}_{j2}, \dots, \hat{y}_{jN})$;

ω_j 为 j 尺度相关向量机回归模型权值向量, $\omega_j = (\omega_{j1}, \omega_{j2}, \dots, \omega_{jM})$; ε_j 为 j 尺度均值为 0, 方差为 σ^2 的噪声集, $\varepsilon_j = (\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{jM})$; f_j 为 j 尺度所建立的相关向量机回归模型函数,可由下式表示:

$$f_j(X_j; \omega_j) = \sum_{i=1}^M \omega_{ji} \Phi(X_{ji}) + \omega_0 \quad (2)$$

式中: Φ 为非线性基函数。

根据先验概率分布和似然分布,利用贝叶斯式计算权值的后验概率分布,得:

$$p(\omega_j | \hat{y}_j, \alpha, \sigma^2) = (2\pi)^{-(N+1)/2} | \sum_j |^{-1/2} \cdot \exp\left\{-\frac{1}{2}(\omega_j - \mu_j)^T \sum_j^{-1} (\omega_j - \mu_j)\right\} \quad (3)$$

$$\sum_j = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \quad (4)$$

$$\mu_j = \sigma^{-2} \sum \Phi^T \hat{y}_j \quad (5)$$

式中: $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$, α 和 σ^2 均当成超参数来处理,通过对权值积分,可以得出超参数的边缘似然分布:

$$p(\omega_j | \hat{y}_j, \alpha, \sigma^2) = (2\pi)^{-N/2} | \sigma^2 I + \Phi A^{-1} \Phi^T |^{-1/2} \cdot \exp\left\{-\frac{1}{2} \hat{y}_j^T (\sigma^2 I + \Phi A^{-1} \Phi^T)^{-1} \hat{y}_j\right\} \quad (6)$$

通过最大化上式求超参数的可能值,由于不能得出超参数的解析形式,故采用迭代法进行求解,当超参数的估计过程收敛时,则根据权值的后验概率分布和最优超参数 α_{MP} 和 σ_{MP}^2 ,对新的数据 X_j^* 进行回归分析,计算新的 \hat{y}_j^* 预测分布

$$P(\hat{y}_j^* | \hat{y}_j, \alpha_{MP}, \sigma_{MP}^2) = N(\hat{y}_j^*, \sigma_*^2) \quad (7)$$

$$\hat{y}_j^* = \mu^T \Phi(X_j^*) \quad (8)$$

2.3 多尺度整合

各尺度系数经过 RVM 建模预测后,得出 $\hat{y}_j = (\hat{y}_{j1}, \hat{y}_{j2}, \dots, \hat{y}_{jN})$,利用小波逆变换将各尺度预测值进行整合,如下:

$$\hat{y} = \text{ISWT}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_j) \quad (9)$$

2.4 MSRVM 模型建模步骤

多尺度相关向量机预测模型考虑了时间序列尺度的影响,将复杂的单一时间序列分解为简单的多时间尺度的序列,使用用水量时间序列变化规律简单化。同时,每一个子序列的参数可单独训练,解决了单一时间建模方法运算复杂度高的问题。多尺度关联向量机日用水量预测模型结构如图 1,建模具体步骤如下:

第一步,确定时间序列 X 的输入-输出结构。本文采用混沌相空间重构技术来确定,则第 i 天的模型输入向量是 $x_{(i-\tau)}, x_{(i-2\tau)}, \dots, x_{(i-m\tau)}$,期望响应为

$x_{(i)}$ (m 代表嵌入维数和 τ 代表延迟时间)。

第二步,确定最佳分解尺度 n 。本文利用母波函数和尺度函数中心频率^[9] 来确定,则分解尺度 n 的计算公式为:

$$n = \log_2\left(\frac{N}{m} \cdot f_s f^* f_a^*\right) - 1 \quad (10)$$

式中: f_s 为采样频率,取1; f^* 为母波函数中心频率; f_a^* 为尺度函数中心频率。

第三步,运行 SWT 程序,此程序在 Matlab2011 中执行。

第四步,对 j 个因子序列分别建立 RVM 预测模型。则第 i 天建模的输入 - 输出结构为 $[X_j = (x_{j(i-\tau)}, x_{j(i-2\tau)}, \dots, x_{j(i-m\tau)})]$;期望输出 $x_{j(i)}$]。

第五步,第 j 个尺度训练好的 RVM 模型用来预测第 j 个 \hat{y}_j 。

第六步,运行 SWT 逆运算 (ISWT),将 j 个因子预测值整合为最终用水量 \hat{y} 。

第七步,输出预测结果。

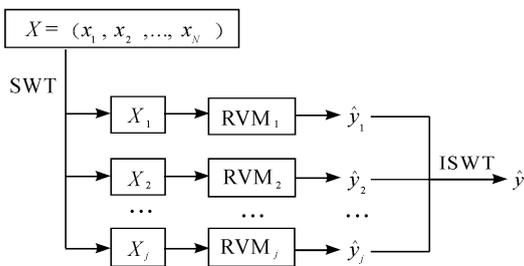


图1 多尺度关联向量机日用水量预测模型结构

3 实例分析

本文以重庆市永川区某自来水公司 2012 年 1 - 6 月的用水量进行预测,统计数据共 180 个,如图 2 所示。选取前 150 d 数据建模,后 30 d 数据作模型有效性检验。

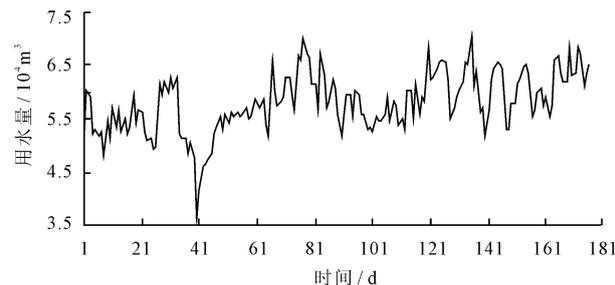


图2 日用水量时间序列

3.1 模型参数确定

RVM 模型训练输入和输出向量结构中有两个

参数,延迟时间 τ 和嵌入维数 m 。本文分别采用自相关函数法和饱和关联维数法来确定^[10],根据计算得出 $\tau = 1, m = 7$, 则 RVM 回归模型输入量为预测日前 7 d 用水量数据,输出量为预测当日用水量,即 7 个输入,1 个输出(输入 $x_{(i-1)}, x_{(i-2)}, \dots, x_{(i-7)}$;期望输出 $x_{(i)}$)。RVM 回归模型中非线性基函数 Φ 选用高斯核函数,其核宽度通过混沌粒子群优化算法迭代寻优。

最佳分解尺度 n 根据公式 (10) 来确定。笔者选用 db4 小波作为 SWT 分解母波,在软件 Matlab 中计算母波函数的中心频率为 0.714,利用功率谱计算尺度函数中心频率为 0.4,则训练集长度 $N = 150$ 的用水量时间序列最佳分解尺度为 $n = 2$ 。

3.2 日用水量预测结果

为了进行比较,分别用基于多尺度相关向量机的预测方法和单尺度的相关向量机预测方法对 2012 年 6 月的日用水量进行预测。图 3 为两种预测方法结果与实际用水量对比图。值得注意的是,图中结果代表的是利用前 150 d 用水量数据建模后,一步预测结果。

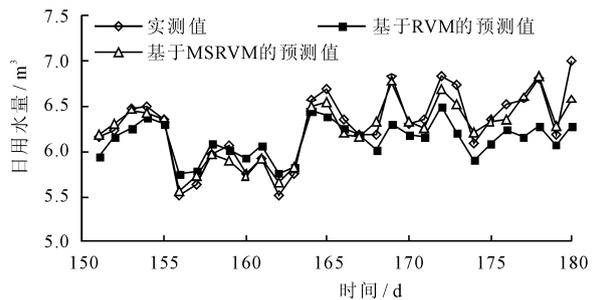


图3 预测结果与实际用水量的比较

3.3 评价检验

为了评价检验该预测结果的准确性,从模型精度评价检验和个体预测结果的检验两方面来验证。

(1)模型精度评价检验。基于多尺度相关向量机模型的精度评价通常用后验差方法进行检验。其基本公式如下:

经过预测,已得到日用水量预测序列,和用水量实际记录数据序列得到残差序列

$$e = x - y \quad (11)$$

用水量记录数据序列的均值、协方差为 S_1 ,残差序列的协方差为 S_2 ,则

$$S_1 = \sqrt{\frac{\sum_{i=1}^N [x(i) - \bar{x}]^2}{n - 1}} \quad (12)$$

$$S_2 = \sqrt{\frac{\sum_{i=1}^N [e(i) - \bar{e}]^2}{n-1}} \quad (13)$$

然后,利用式(12)和(13),根据

$$C = S_2/S_1 \quad (14)$$

$$p = p\{|e(i) - \bar{e}| < 0.6745S_1\} \quad (15)$$

计算后验差比值 C 和小误差概率 p 共同验证模型精度,其标准为模型精度级别 = $\max\{C \text{ 所处级别}, p \text{ 所处级别}\}$ 。

模型的精度通常分为四级,分级标准及相应的 C 和 p 值见表 1^[11]。

表 1 模型精度等级

模型精度等级	p	C
1级(好)	≥ 0.95	≤ 0.35
2级(合格)	$0.80 \leq p < 0.95$	$0.35 < C \leq 0.50$
3级(勉强)	$0.70 \leq p < 0.80$	$0.50 < C \leq 0.65$
4级(不合格)	< 0.70	> 0.65

用多尺度建模(MSRVM)和单尺度建模(RVM)进行预测,经后验差比值 C 和小误差概率 p 对模型验证,相应评价结果见表 2。

表 2 模型精度结果

预测模型	p	C
RVM	0.87	0.43
MSRVM	0.97	0.31

从表 2 中的数据 and 表 1 中精度等级划分可以看出 MSRVM 的模型精度等级为好,RVM 的模型精度等级为合格。

(2) 预测结果评价检验。根据整体评价预测方法,采用平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和标准均方根误差(NRMSE)作为预测效果评价指标,评价结果见表 3。

表 3 预测评价结果

预测模型	MAE/m ³	MAPE/%	NRMSE/m ³
RVM	2225.19	0.0346	0.0443
MSRVM	898.46	0.0141	0.0196

从图 3 可以看出,整体上 RVM 的预测精度要低于 MSRVM 的预测精度,同时结合表 3 可以得出,MSRVM 模型的 3 种评价指标均优于 RVM 模型,主要原因是由于多尺度相关向量机预测模型较好地分

解了复杂多变的用水量时间序列,提取出明显有效的细节因子和趋势因子,整合各子模型的样本数据特性信息,故对用水量预测效果好。

4 结 论

笔者对城市的日用水量进行了探讨,提出了采用多尺度相关向量机模型预测城市日用水量,并与单尺度相关向量机模型进行比较,结果显示前者具有更高的精度,是预测城市日用水量的较理想方法。同时为了检验该预测结果的准确性,从模型精度评价检验和个体预测结果的检验两方面来验证分析,证明多尺度相关向量机模型是一种行之有效的模型。

参考文献:

- [1] 张琴,汪雄海,朱庆建. 基于联合时序的混沌时用水量短期预测调度[J]. 排灌机械工程学报,2011,29(4): 352-358,346.
- [2] 傅金祥,潘海宾,马兴冠. 灰色残差模型与递补模型对比及在城市需水量预测中的应用[J]. 给水排水,2008,34(6):111-113.
- [3] Firat M, Turan M E, Yurdusev M A. Comparative analysis of neural network techniques for predicting water consumption time series [J]. Journal of Hydrology, 2010, 384(1-2): 46-51.
- [4] 王圃,白云,李川,等. 基于变结构支持向量回归的城市日用水量预测[J]. 应用基础与工程科学学报,2015,23(5): 895-901.
- [5] 王圃,王以知,张晋,等. 基于用水量驱动因子的水量预测模型[J]. 排灌机械工程学报,2014,32(12): 1051-1056.
- [6] 冯天梅,张鑫. 基于修正组合模型的包头市用水量预测分析[J]. 西北农林科技大学学报(自然科学版),2014,42(3): 226-234.
- [7] 章恒全,何薇. 基于主成分回归与灰色神经网络模型的水资源承载力需水量预测[J]. 水资源与水工程学报,2014,25(1): 103-108.
- [8] Tipping M E. The relevance vector machine[J]. Advances in Neural Information Processing System, 2000, 12(3): 652-658.
- [9] 樊计昌,刘明军,海燕,等. 计算尺度函数和小波函数中心频率的 GUI 及其应用[J]. 科技导报,2007,25(24): 36-39.
- [10] 吕金虎,陆君安,陈士华. 混沌时间序列分析及其应用[M]. 武汉:武汉大学出版社,2005.
- [11] 尹学康. 城市需水量预测[M]. 北京:中国建筑工业出版社,2006.