

DOI:10.11918/j.issn.0367-6234.201807113

基于 Bayesian-LSSVM 和残差修正 的用户短期需水量预测

吴 珊,宋凌硕,侯本伟,寇晓霞

(北京工业大学 建筑工程学院,北京 100124)

摘 要: 为有效改善供水管网短期需水量预测模型在预测精度和稳定性方面存在的不足,提出在短期需水量预测模型基础上叠加残差预测模型的组合预测建模方法.首先采用贝叶斯最小二乘支持向量机法(Bayesian-LSSVM)建立管网用户需水量时间序列预测模型(BL模型),得到需水量预测初始值;对BL模型得到的需水量预测初始值的残差序列,构建基于贝叶斯最小二乘支持向量机法的混沌时间序列预测模型(RM模型),得到残差预测值;同时将RM模型得到的残差预测值实时补偿到BL模型的需水量预测初始值中,得到经过残差修正的需水量预测值.实例结果表明,RM模型可以准确捕获BL模型需水量预测初始值的残差变化趋势,对其残差序列进行准确预测;在短期需水量预测的精度和稳定性方面,由BL模型和RM模型叠加构成的组合预测模型(BL+RM模型)明显优于单一BL模型;BL+RM模型适用于平均需水量较小、水量波动性较大等不同特点用户的短期需水量预测,可有效满足实际工程的需要.

关键词: 短期需水量预测;残差修正;贝叶斯最小二乘支持向量机;混沌时间序列预测

中图分类号: TU991.33 **文献标志码:** A **文章编号:** 0367-6234(2019)08-0088-09

Short-term water demand forecast based on Bayesian least squares support vector machine and residual correction

WU Shan, SONG Lingshuo, HOU Benwei, KOU Xiaoxia

(College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: In order to effectively improve the short-term water demand forecasting model of water distribution networks in terms of prediction accuracy and stability, a novel combined prediction modeling method which can predict water demand and residuals simultaneously is proposed in this paper. First, the Bayesian least squares support vector machine method (Bayesian-LSSVM) was used to establish the time series prediction model of user's water demand (BL model) to obtain the initial water demand prediction values. Then, to predict the residual sequence of the initial water demand prediction values produced by the BL model, a chaotic time series prediction model (RM model) was constructed based on the Bayesian-LSSVM method. At the same time, the predicted residuals produced by the RM model were compensated to the water demand predictions by the BL model to correct the initial water demand prediction values. Results of the case study show that the RM model could accurately capture the change trend of the residual value of the BL model initial prediction values and the residual sequence of the BL model initial water demand prediction values. The combined forecast model consisting of the BL model and the RM model (BL+RM model) was superior to single BL model in the accuracy and stability of short-term water demand forecasting. BL+RM model was applicable for short-term water demand forecasting with different water demand characteristics such as small average water demand and large water volatility, and hence could effectively meet the needs of actual engineering.

Keywords: short-term water demand forecast; residual correction; Bayesian least squares support vector machine; chaotic time series prediction

城市供水系统水量预测分为中长期预测(1~10 a不等)和短期预测(1 h、1 d、1周等)两类.管网用户短期(15 min~1 h)需水量的准确预测是提高

供水管网水力模型精度的重要支撑.目前,短期需水量预测常用方法主要有时间序列三角函数法、灰色系统法、小波分析法、积分滑动平均法、Elman神经网络法、广义回归神经网络法、BP神经网络法、支持向量机预测法、最小二乘支持向量机法和贝叶斯最小二乘支持向量机等^[1-6].刘洪波等^[1]研究表明采用BP神经网络法的时需水量预测模型,与时间

收稿日期: 2018-07-12
基金项目: 国家水体污染控制与治理科技重大专项(2017ZX07108-002);国家自然科学基金(51508528)
作者简介: 吴 珊(1963—),女,博士,副教授
通信作者: 侯本伟,benweihou@bjut.edu.cn

序列三角函数分析法、灰色系统理论预测法、小波分析法的预测模型相比具有预测误差小和计算速度快的优点. 陈磊^[2]、王亮等^[3]基于历史需水量数据建立支持向量机(SVM)的时需水量预测模型, 预测结果的平均相对误差(E_{MAP})为 3.38%, 与基于 BP 神经网络的预测模型(E_{MAP} 为 10.25%)相比预测精度更高, 建模速度也更快. 陈磊等^[4]的研究还证明, 基于历史需水量数据以最小二乘支持向量机(LSSVM)法建立的时需水量模型, 其预测能力也明显高于基于 BP 神经网络建立的时需水量模型. 何必仕等^[5]基于温度、天气、风力、节假日和历史需水量数据, 比较了自回归积分滑动平均模型、Elman 神经网络、广义回归神经网络以及最小二乘支持向量机法(LSSVM)的时需水量预测模型, 结果显示 LSSVM 模型的需水量预测值 E_{MAP} 为 5.44%, 预测精度较高. 陈磊与张土乔^[6]提出利用贝叶斯置信框架推断最小二乘支持向量机的模型参数, 与 LSSVM 模型相比, 时需水量预测模型建模速度更快, 预测结果的 E_{MAP} 为 3.59%, 优于 LSSVM 模型的 4.83%.

模型预测值与实测值之间都会存在一定的误差(残差), 研究表明, 采用对需水量预测模型的残差建立残差预测模型, 同步修正需水量预测模型的预测值, 是一种提高模型预测精度的有效方法. 刘大同^[7]在进行时间序列预测时, 根据残差的混沌特性, 对模型的残差建立混沌时间预测模型, 对残差进行在线预测进而达到修正模型的预测值、提高模型预测精度的目的. 混沌时间序列预测在电力负荷预测方面得到广泛应用, 方仍存^[8]、雷绍兰^[9]、杨卓等^[10]根据电力负荷的混沌特性, 对电力负荷建立混沌时间序列预测模型, 均取得较好的预测效果.

除提高预测模型的精度之外, 如何提高预测模型对于不同特点用户的普遍适用性, 以及在不同需水工况下的稳定性(需水量变化时)也是需要关注的问题. 本文以贝叶斯最小二乘支持向量机法建立了需水量预测模型(BL 模型), 同时采用贝叶斯最小二乘支持向量机的方法对 BL 模型预测值与需水量观测值间的残差进行混沌时间序列预测(RM 模型), 以修正 BL 模型的预测结果. 经实例验证, 本文提出的组合预测(BL + RM)模型可显著提高短期(15 min)水量预测模型的精度和稳定性.

1 模型构建的基础理论

最小二乘支持向量机(LS-SVM)是标准支持向量机(SVM)的一种扩展^[11], LS-SVM 与标准 SVM 的区别主要在于采用不同的优化目标函数, 并且用等式约束代替不等式约束. LS-SVM 模型的正则化因

子和核宽度系数一般常采用交叉验证法、误差界方法和统计方法来确定, 容易产生过拟合现象, 而且耗时长、计算量大, 大大影响了建模的效率和所建模型的性能^[12]. 为此, Mackay^[13]提出一种贝叶斯证据框架, 已有不少学者将贝叶斯证据框架应用于 LS-SVM 模型参数的选择^[14-15].

1.1 最小二乘支持向量机原理

对于给定的输入数据集 $D = (\mathbf{x}_i, y_i)$, 其中 $\mathbf{x}_i \in \mathbf{R}^n$ 表示 n 维输入矢量, $\mathbf{x}_i$ 为第 i 个输入样本矢量, y_i 为第 i 个输出样本标量($i = 1, 2, \dots, l$), l 为样本输入个数, 设对样本集进行拟合的函数形式为

$$y(\mathbf{x}) = \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b. \quad (1)$$

式中: 非线性映射 $\boldsymbol{\varphi}$ 函数将输入数据映射到一个高维特征空间 $\mathbf{R}^n \rightarrow \mathbf{R}^{nh}$; b 为偏执量, $\boldsymbol{\omega} \in \mathbf{R}^{nh}$ 表示权值向量. 为求解 $y(\mathbf{x})$, 定义目标函数如下:

$$\min J(\boldsymbol{\omega}, \xi_i) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^l \xi_i^2, \quad (2)$$

$$\text{s. t } y_i = \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b + \xi_i.$$

式中: γ 为正规化参数, ξ_i 为松弛变量.

定义 Lagrange 函数:

$$L(\boldsymbol{\omega}, b, \xi_i, \alpha_i) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i [\boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b + \xi_i - y_i]. \quad (3)$$

式中: α_i 为拉格朗日乘子, 根据 KKT 条件:

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\omega}} = 0 &\rightarrow \boldsymbol{\omega} = \sum_{i=1}^l \alpha_i \boldsymbol{\varphi}(\mathbf{x}_i), \\ \frac{\partial L}{\partial b} = 0 &\rightarrow b = \sum_{i=1}^l \alpha_i = 0, \\ \frac{\partial L}{\partial \xi_i} = 0 &\rightarrow \alpha_i = \gamma \cdot \xi_i, \\ \frac{\partial L}{\partial \alpha_i} = 0 &\rightarrow \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b + \xi_i - y_i = 0. \end{aligned} \quad (4)$$

消去 $\boldsymbol{\omega}$ 和 ξ_i 得到优化问题的解析解

$$\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{M} + \gamma^{-1} \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}. \quad (5)$$

式中: $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_l]^T$; $\mathbf{1}$ 为元素取值全为 1 的 $1 \times l$ 维向量, $\mathbf{1} = [1, 1, \dots, 1]^T$; $\mathbf{I}$ 为 $l \times l$ 维对角线元素全为 1 的对角阵; $\mathbf{M}$ 为 $l \times l$ 维方阵, 其元素取值满足 $M_{ij} = \boldsymbol{\varphi}(\mathbf{x}_i)^T \cdot \boldsymbol{\varphi}(\mathbf{x}_j)$, 其中 $i, j = 1, 2, \dots, l$.

定义核函数:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}\|^2}{2\sigma^2}\right). \quad (6)$$

则采用径向核基函数(RBF)的 LS-SVM 表达式为

$$y(\mathbf{x}) = \sum_{i=1}^l \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b. \quad (7)$$

1.2 贝叶斯证据框架下的 LS-SVM 参数确定

证据框架将贝叶斯推断分为 3 个准则的推断,每个准则优化 LS-SVM 的不同参数,准则 1 利用贝叶斯规则对后验分布密度中的参数进行推断,并通过最大化后验概率来获得参数的估计值,可以推断出权值参数 ω 最优值;准则 2 通过最大化正规化参数 γ 的后验分布来求取正规化参数的最优值;准则 3 通过最大化后验概率来比较不同模型的优劣,从而选择最优核参数 $\sigma^{[16-17]}$.

准则 1 推断. 令 H 为模型空间, D 为样本训练集, 对于给定的超参数 $\lambda = 1/\gamma$, 由贝叶斯规则推断参数向量 ω 的后验概率为

$$P(\omega|D, \lambda, H) \propto P(D|\omega, \lambda, H)P(\omega|\lambda, H). \quad (8)$$

假设训练样本独立同分布, 且参数 ω 的先验分布 $P(\omega|\lambda, H)$ 服从正态分布, 可得 ω 的后验概率为

$$P(\omega|D, \lambda, H) \propto \exp\left(-\frac{1}{2} \|\omega\|^2 - \sum_{i=1}^l \xi_i^2\right). \quad (9)$$

式中: l 为训练样本的数目; ξ_i 为 LS-SVM 对每个训练样本的回归误差; 通过最大化式(9)得到 ω 的最优值 ω_{best} .

准则 2 推断. 通过最大化 λ 的后验概率 $P(\lambda|D, H)$, 即可求得 λ 的最优值 λ_{best} . 令 $E_\omega = \frac{1}{2} \|\omega\|^2$,

$E_D = \sum_{i=1}^l \xi_i^2$, 则有

$$P(\lambda|D, H) \propto \frac{(2\pi\lambda)^{\frac{k}{2}} \exp(-\lambda E_\omega^{\text{best}} - E_D^{\text{best}})}{(\det A)^{\frac{1}{2}}}. \quad (10)$$

式中: E_ω^{best} 、 E_D^{best} 分别为 E_ω 、 E_D 在 ω_{best} 处的值, $A = \partial^2(\lambda E_\omega + E_D)/\partial^2 \omega = \nabla^2(\lambda E_\omega + E_D)$, ∇ 为矢量微分算符, 通过最大化 $\ln P(\lambda|D, H)$ 可得到 λ 的最优值 λ_{best} .

准则 3 推断. 准则 3 通过检查模型的后验概率 $P(H|D) \propto P(D|H)P(H)$, 对不同模型的优劣进行比较, 从而选择最优核参数. 假定对于所有模型其先验概率 $P(H)$ 为平坦分布, 则 $P(D|H)$ 通过对参数 λ 的积分求得

$$P(H|D) \propto \int P(D|\lambda, H)P(\lambda|H)d\lambda \propto P(D|\lambda_{\text{best}}, H)/\sqrt{\delta}. \quad (11)$$

通过最大化 $\ln P(H|D)$ 可得 LS-SVM 的最优核参数 δ_{best} .

1.3 混沌时间序列预测

混沌时间序列预测是建立在 Takens 提出的嵌入定理和相空间重构理论基础上的, 其基本思想是

系统中的任一分量的演化是由与之相互作用着的其他分量决定的, 这些相关分量的信息就隐含在任一分量的发展过程中, 可以从仅仅与时间相关的混沌数据中提取和恢复出系统原来的规律, 实质上表达为高维空间下的一种轨迹^[7].

刘大同^[7]、高俊杰^[18]、朱志宇等^[19]采用 SVM 对混沌时间序列进行预测, 取得了较好的预测效果, 但 SVM 模型存在模型预测精度较低和计算速度慢的缺点^[6]. 本文研究采用贝叶斯优化最小二乘支持向量机法对混沌时间序列进行建模.

贝叶斯优化最小二乘支持向量机在混沌时间序列预测中的依据: 混沌序列 $\{e_1, e_2, \dots, e_k\}$, k 为序列总长度, 通过曹氏法(C-C 法)^[18]求得延迟时间 τ 和嵌入维数 m , 根据相空间重构原理, 进行相空间重构后共有 N 个相点 ($N = k - (m-1)\tau$), 依次为

$$\begin{aligned} e_1 &= \{e_1, e_{1+\tau}, e_{1+2\tau}, \dots, e_{1+(m-1)\tau}\}, \\ e_2 &= \{e_2, e_{2+\tau}, e_{2+2\tau}, \dots, e_{2+(m-1)\tau}\}, \\ &\vdots \\ e_N &= \{e_N, e_{N+\tau}, e_{N+2\tau}, \dots, e_{N+(m-1)\tau}\}. \end{aligned} \quad (12)$$

进一步演化后的相点 e_{N+1} 为

$$e_{N+1} = \{e_{N+1}, e_{N+1+\tau}, e_{N+1+2\tau}, \dots, e_{N+1+(m-1)\tau}\}. \quad (13)$$

上式的最后一维元素 $e_{N+1+(m-1)\tau}$ 正是需要预测的下一序列点 e_{k+1} .

贝叶斯最小二乘支持向量机在混沌预测中的应用具有两个性质: ①某个相点的演化行为与其邻近相点的演化行为类似; ②相点 e_i ($i = 1, 2, \dots, N$) 进一步演化成为 e_{i+1} , e_i 与 e_{i+1} 最后一维元素 $e_{i+1+(m-1)\tau}$ 之间有某种复杂的函数关系:

$$e_{i+1+(m-1)\tau} = f(e_i). \quad (14)$$

依据这两条性质, 找出 e_i 的 t 个邻近点组成的集合 e_{ij} ($j = 1, 2, \dots, t$), 再利用贝叶斯优化最小二乘支持向量机的逼近性能, 找出一个最优函数来拟合 $f(*)$, 即

$$e_{ij+1+(m-1)\tau} = f(e_{ij}). \quad (15)$$

若函数关系 $f(*)$ 确定, 则预测如下:

$$e_{k+1} = e_{N+1+(m-1)\tau} = f(e_N). \quad (16)$$

同理可进行 e_{k+2} 和 e_{k+3} 的预测值.

2 用户需水量预测步骤

2.1 模型输入参数选择

在短期需水量影响因素的研究方面, 信昆仑等^[20]进行日需水量预测时, 不考虑工作日和休息日的影响, 评估了温度、湿度、降水量与城市需水量的相关程度, 将温度、湿度作为需水量影响因子建立日需水量预测模型; Bakker^[21]、Al-Zahrani 等^[22]均未考

考虑工作日和休息日的差异,将温度、湿度和风速作为需水量影响因子建立日需水量预测模型。Brentan 等^[23]在进行实时(15 min)需水量预测时,考虑了工作日和休息日的影响,通过将温度、湿度、风速、降雨量等物理量与需水量做相关性分析,认为温度、湿度与需水量的相关性程度较高。

本文结合实际数据资料进行了综合分析,认为工作日与休息日在用水特征方面有差异,为提高模型预测的准确度,分别建立了工作日和休息日两种用水模式下的需水量预测模型(BL 模型);同时,选取了温度、湿度作为影响需水量变化的影响因子。

表 1 模型的输入变量

Tab. 1 Input variables of the model

模型	工作日 + 休息日模型	工作日模型	休息日模型
模型输入	$T, Q_{t-1}, Q_{t-96}, Q_{t-572}, T_t, W_t$	$T, Q_{t-1}, Q_{t-96}, T_t, Q_{t-480}, W_t$	$T, Q_{t-1}, Q_{t-96}, Q_{t-192}, T_t, W_t$

2.2 模型预测流程

基于贝叶斯最小二乘支持向量机和残差修正的用户短期需水量预测的流程如下:

- 1) 收集用户需水量历史数据及对应的温度、湿度历史记录,对历史数据中异常、缺失的数据进行处理;
- 2) 分析用户工作日与休息日需水量变化规律,建立初始化需水量预测模型;
- 3) 根据在线更新的输入数据 $\mathbf{x}_k$ 的观测值及需水量 y_k ,更新贝叶斯最小二乘支持向量机模型训练输入 $\mathbf{X} = \{\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}, \cdots, \mathbf{x}_{k-1}\}$ 为 $\mathbf{X} = \{\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \cdots, \mathbf{x}_k\}$ 进行下一步预测,预测值记为 $\bar{y}_{k+1}$;同时更新混沌时间序列残差预测模型(简称 RM 模型)输入 $\hat{\mathbf{e}} = [e_1, e_2, \cdots, e_{k-1}]$ 为 $\hat{\mathbf{e}} = [e_2, e_3, e_4, \cdots, e_k]$,对残差序列进行相空间重构,通过 RM 模型进行下一步残差序列预测,预测值记为 $\bar{e}_{k+1}$; (e_k 为残差,即 $e_k = y_k - \bar{y}_k$);
- 4) 通过 RM 模型和 BL 模型(简称 BL + RM 模型)计算第 $k + 1$ 步需水量预测值为 $y_{\text{predict}} = \bar{y}_{k+1} - \bar{e}_{k+1}$.

BL + RM 模型算法流程如图 1 所示,其中 RM 模型的算法结构示意如图 2 所示。

本文采用相对误差 (E_R)^[3,7,21]、平均相对百分误差 (E_{MAP})^[3,5,21-23]、均方误差 (E_{RMS})^[5,21,23]、皮尔逊相关系数 (R)^[5] 作为模型预测性能评价指标; E_R 、 E_{MAP} 和 E_{RMS} 的值越小模型预测精度越高; R 值衡量两个数据集的相关度, R 越接近于 1, 两个数据集合相关性越好,越接近在一条直线上。

为了预测用户在 t 时刻的需水量 Q_t ,模型建立过程中选取预测日前 n d 的历史数据训练模型,历史数据记录时间间隔为 15 min(采样频率为 4 次/h 即 0.25 h/次),选取 t 时刻(用 1, 2, 3, $\cdots$, 96 表示)、温度 T_t 、湿度 W_t ,前一个时刻的需水量 Q_{t-1} ,前 1 d 该时刻的需水量 Q_{t-96} ,前一周该时刻的需水量 Q_{t-572} (工作日和休息日考虑在一起的需水量预测模型)或 Q_{t-480} (工作日需水量预测模型)或 Q_{t-192} (休息日需水量预测模型)作为模型的输入, t 时刻的需水量 Q_t 作为模型的输出,具体见表 1。

$$E_R = \frac{|y_{o,i} - y_{m,i}|}{y_{o,i}} \times 100\%, \quad (17)$$

$$E_{\text{MAP}} = \frac{\frac{1}{n} \sum_{i=1}^n |y_{o,i} - y_{m,i}|}{y_{o,i}} \times 100\%, \quad (18)$$

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^n (y_{o,i} - y_{m,i})^2}{n}}. \quad (19)$$

式中: $y_{o,i}, y_{m,i}$ 分别为时间序列中第 i 时刻的需水量实测值和预测值, n 为时间序列的长度。

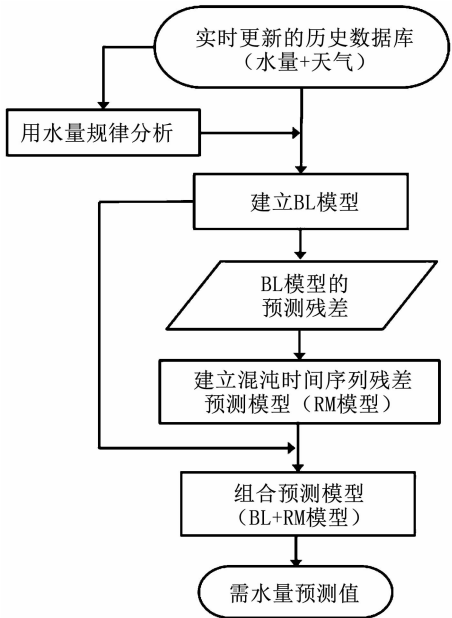


图 1 BL + RM 模型算法流程

Fig. 1 Flow chart of BL + RM model algorithm

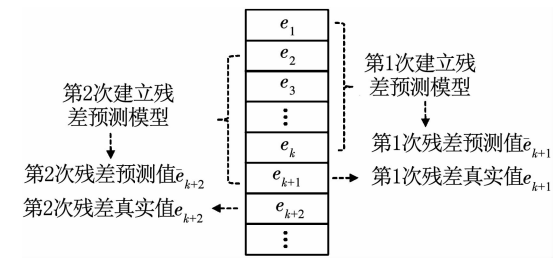


图 2 RM 模型算法流程

Fig. 2 Flow chart of RM model algorithm

3 实例分析

为了验证本文方法的有效性,以某大城市的 3 个小区需水量为例进行了水量预测应用验证. 首先收集并处理 2017 年该市 A、B、C 3 个小区的入口流量计所记录的每 15 min (采样频率为 4 次/h) 的需水量数据(即 0.25 h 记录一次,每天 24 h 共记录 96 个时刻的需水量数据). A、B、C 3 个小区有不同的

表 2 A、B、C 3 个小区 2017 年需水量分析

Tab. 2 Analysis of water demand of three communities (A, B, and C) in 2017

小区名称	最小值/(m ³ ·h ⁻¹)	最大值/(m ³ ·h ⁻¹)	均值/(m ³ ·h ⁻¹)	标准差/(m ³ ·h ⁻¹)	变异系数	平均周期
A	15.28	135.04	37.98	14.80	0.39	3 个小区需水量 平均周期相同
B	14.01	84.32	34.59	8.86	0.26	
C	15.68	169.92	90.20	32.80	0.36	

表 3 图 3-II 中各标志点坐标

Tab. 3 Coordinates of each marker point in Fig. 3-II

点号	a1	a1'	a1''	b1	b1'	c1	c1'	a2
坐标	(30,52.07)	(40,51.25)	(30,48.77)	(86,55.81)	(87,63.70)	(28,49.22)	(28,35.21)	(34,50.26)
点号	b2	b2'	c2	c2'	a3	b3	c3	c3'
坐标	(87,60.46)	(86,55.81)	(6,22.21)	(6,19.71)	(35,50.42)	(86,57.07)	(26,34.11)	(36,29.19)

表 4 图 3-II 用户用水规律分析

Tab. 4 Analysis of water use patterns in Fig. 3-II

时段	早高峰出现的 最大时间差/h	早高峰需水量的 最大误差/%	晚高峰出现的 最大时间差/h	晚高峰需水量的 最大误差/%	早、晚高峰时间段外相同 时刻需水量最大误差/%
周一—周日	(a1,a1')2.5	(a1,a1'')6.77	(b1,b1')0.25	(b1,b1')14.15	(c1,c1')39.81
周一—周五	(a1'',a2)1.25	(a1,a1'')6.77	(b1,b2)0.25	(b1,b2')12.71	(c2,c2')12.7
周六—周日	(a3,a1')1.25	(a3,a1')1.64	(b3,b1')0.25	(b3,b1')11.62	(c3,c3')16.87

图 3-I、图 3-II 分别为小区 7 月 3—9 日的 1 周需水量变化规律及 7 月 1—28 日连续 4 周对应各时刻需水量进行加权平均后的变化规律. 由表 4 可知,周一—周日、周一—周五(工作日)、周六—周日(休息日)早高峰出现的最大时间差分别为 2.5、1.25 和 1.25 h,早高峰需水量最大误差分别为 6.77%、6.77%、1.64%,分析可得休息日的早高峰与工作日早高峰相比明显不同,最多延迟 2.5 h 出现;晚高峰出现的最大时间差均为 0.25 h,晚高峰需水量最大

需水量特点,如表 2 所示,3 个小区的需水量平均周期相同;其中 A 和 B 小区平均需水量相近,但需水量随机波动性差别较大;A 和 C 小区的需水量随机波动性相近,但平均需水量有差异,其中 C 小区平均需水量较大;B 和 C 小区需水量的随机波动性及平均需水量均明显不同. 用预测日前 30 d 的数据来训练模型;在相同时段分别建立上述 3 个小区工作日和休息日的 BL + RM 模型预测未来 20 d(工作日)、未来 10 d(休息日)的需水量. 以下以 A 小区为例,进行具体应用分析说明.

3.1 用户用水模式分析

考虑到工作日和休息日对用水规律的影响,对 A 小区的用水规律进行分析如图 3 所示,其中横坐标为 1 d 24 h 内的 96 个时刻,纵坐标为各时刻的需水量. 表 3 为图 3 中各标志点的坐标,横坐标为时刻值,纵坐标为需水量. 用水规律分析结果如表 4 所示.

误差分别为 14.15%、12.71% 和 11.62%,可知休息日的晚高峰与工作日晚高峰在峰值出现时刻及峰值需水量均无明显差别;早高峰和晚高峰时间段外的需水量最大误差分别为 39.81%、12.7%、16.87%,因此,早高峰和晚高峰时间段外休息日需水量明显高于工作日需水量. 综上,研究中分别建立工作日和休息日需水量预测模型是正确的.

图 3-III 为法定节假日与其邻近前后休息日的需水量变化曲线,可以看出,节假日与休息日的用水

变化规律相似,早高峰及晚高峰的出现时刻及需水量也均无明显差别,因此,研究中将节假日划分到休息日预测模型中.

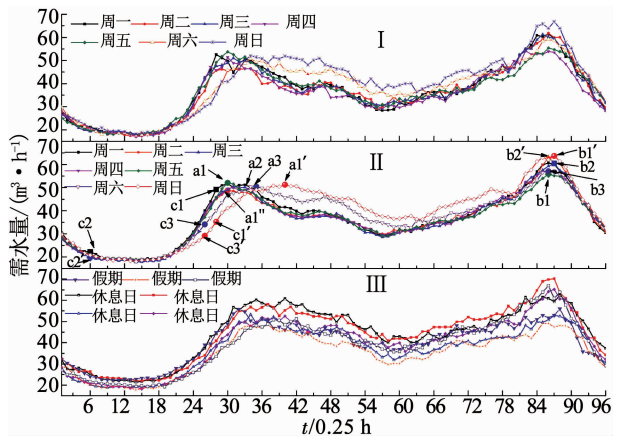


图 3 A 小区用户用水规律曲线

Fig. 3 Water use regularity curve of community A

3.2 需水量预测模型的建立

以 A 小区预测日前 30 d(仅工作日)的历史数据训练初始化的工作日 BL 模型,预测未来 20 d(6 月 21 日—7 月 20 日 1 个月左右)处于全年需水量高峰时的需水量;并同步对 BL 模型的残差建立 RM 模型,修正工作日 BL 模型预测值. 以 A 小区预测日前 30 d(仅休息日和法定节假日)的历史数据训练初始化的休息日需水量预测模型,预测未来 10 d(7 月 16 日—8 月 19 日 1 个月左右)处于全年需水量高峰时的需水量,并同步对 BL 模型的残差建立 RM 模型,修正休息日 BL 模型预测值.

3.3 预测结果分析

1)表 5,6 分别为将工作日与休息日考虑在一起、将工作日与休息日分开分别建立 BL 模型时,工作日、休息日的预测值与实际值相对误差占比. 由表 5,6 分析可知,将工作日和休息日分开时工作日和休息日 BL 模型的预测精度与将工作日和休息日考虑在一起时相比,BL 模型的预测精度得到显著提高,证实了依据工作日和休息日用水规律不同分别建立工作日和休息日水量预测模型的必要性.

2)工作日 6 月 21—23 日实际需水量、BL 模型、BL + RM 模型的预测结果如图 4,5,6 所示.

表 5 A 小区 BL 模型工作日预测误差占比			
Tab.5 Proportion of the forecast error of the BL model on working days			
时段	$E_R < 5\%$ 占比	$E_R < 10\%$ 占比	%
工作、休息日在一起	49.38	84.32	
工作、休息日分开	57.14	89.64	

表 6 A 小区 BL 模型休息日预测误差占比			
Tab.6 Proportion of the forecast error of the BL model on non-working days			
时段	$E_R < 5\%$ 占比	$E_R < 10\%$ 占比	%
工作日休息日在一起	48.44	76.77	
工作日休息日分开	55.00	86.25	

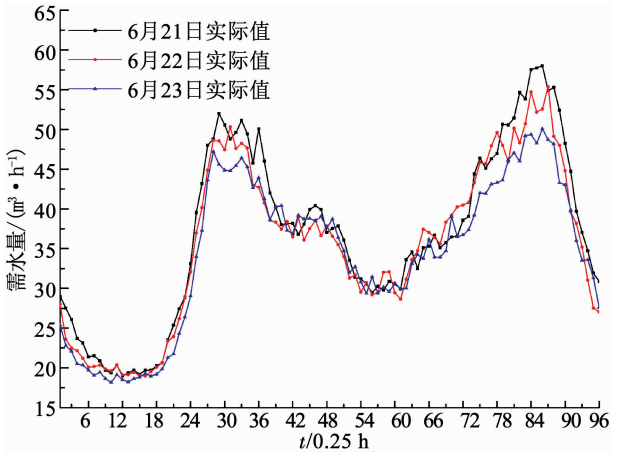


图 4 6 月 21—23 日实际需水量

Fig. 4 Actual water demand from June 21 to 23

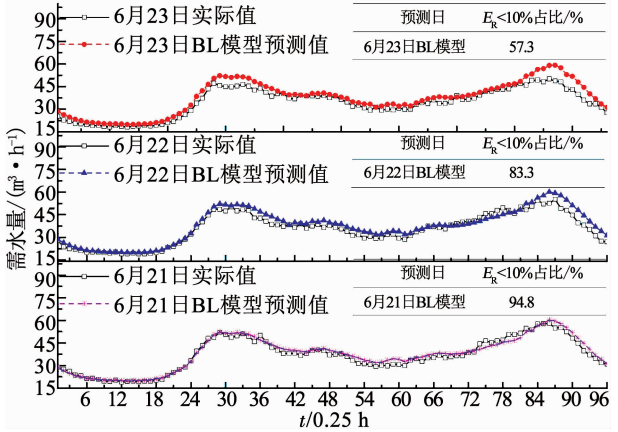


图 5 6 月 21—23 日实际需水量与 BL 模型预测值

Fig. 5 Actual water demand and BL model forecast on working days from June 21 to 23

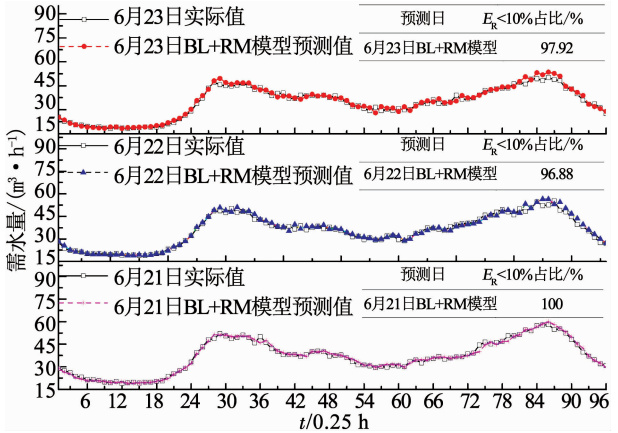


图 6 6 月 21—23 日实际需水量与 BL + RM 预测值

Fig. 6 Actual water demand and BL + RM model forecast on working days from June 21 to 23

由图 4 可知,6 月 23 日实际需水量较 6 月 21—22 日实际需水量有较大波动;分析图 5 可知,6 月 23 日 BL 模型的预测值与实际值之间存在较大的误差;由图 5,6 可知,6 月 23 日 BL + RM 模型的预测精度明显高于 BL 模型的预测精度. 表明 BL 模型不能快速适应需水量发生较大波动的情况,稳定性较差,而 RM 模型可以挖掘 BL 模型系统残差中蕴藏的规律,忽略系统残差背后众多影响因素和复杂的影响机理并预测系统的残差值,修正 BL 模型的预测值,BL + RM 组合模型的稳定性明显较高,对需水量波动较大的用户具有较强的适应能力.

3) 分别随机挑选工作日、休息日 BL 模型和 BL + RM 模型连续两天的预测结果,需水量预测值与实际观测值的对比分析如图 7,8 所示,其中图 7 为工作日 BL 模型和 BL + RM 模型连续 2 d 的预测结果,图 8 为休息日 BL 模型和 BL + RM 模型连续 2 d 的预测结果.

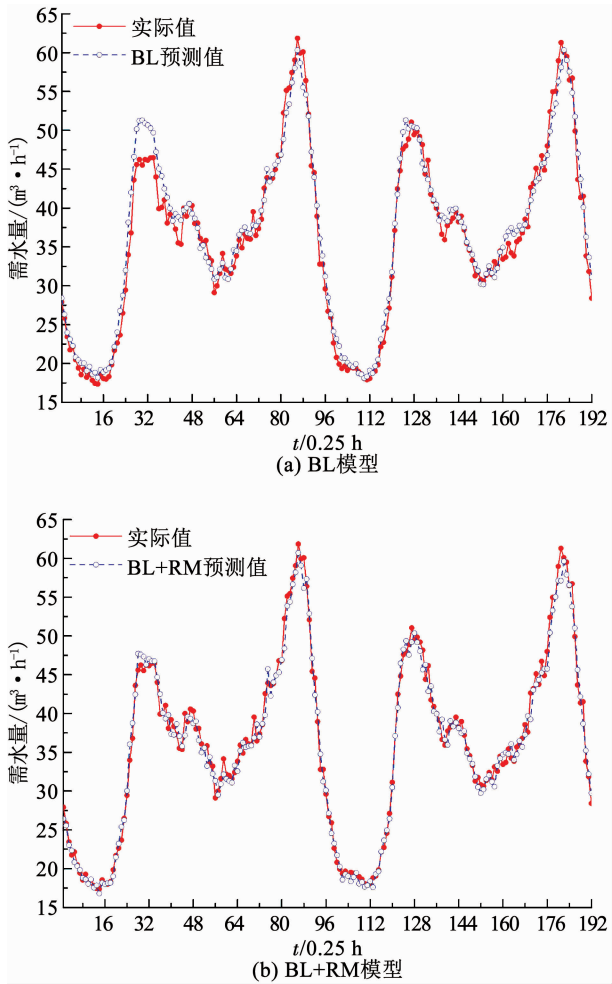


图 7 工作日模型预测结果

Fig. 7 Model prediction results on working days

由图 7(a),(b)及图 8(a),(b)分析可知,BL 模型可以准确预测需水量变化的周期性和趋势性,但不能对需水量变化的随机性做出准确的预测,仍存

较大的预测残差;通过对 BL 模型预测值与实际值间的残差建立 RM 模型,可以精确地捕获预测残差的变化趋势,预测未来的残差值,对 BL 模型的预测值合理地做出残差补偿,提高需水量预测精度.

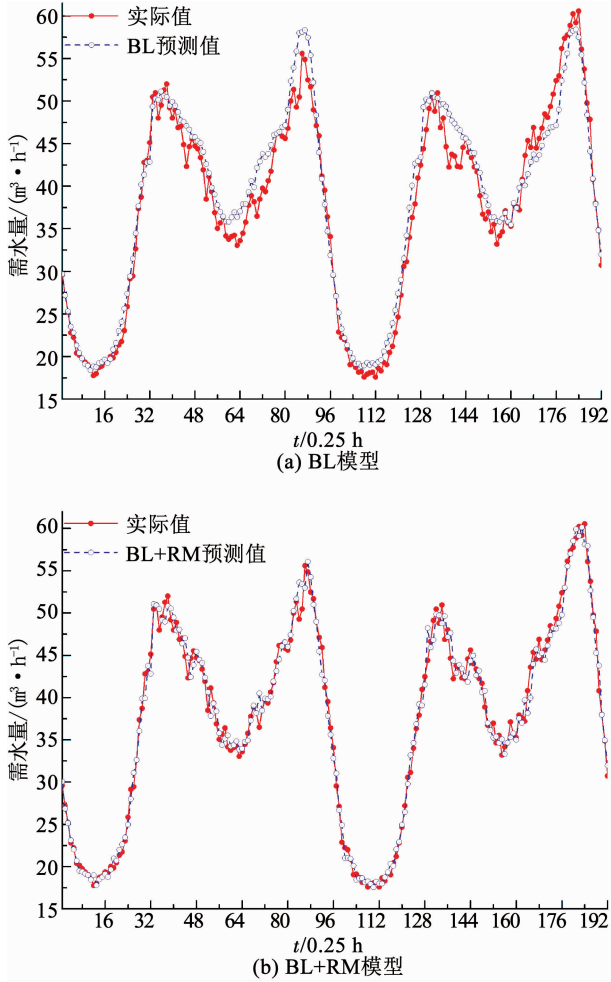


图 8 休息日模型预测结果

Fig. 8 Model forecast results on non-working days

表 7,8 分别为 A 小区工作日连续 20 d 和休息日连续 10 d 的预测结果,图 9 为工作日连续 20 d BL 模型和 BL + RM 模型预测结果与实际值的散点图,图 10 为休息日连续 10 d BL 模型和 BL + RM 模型预测结果与实际值的散点图.

由表 7,8 可知,与 BL 模型相比,BL + RM 模型可以显著提高需水量预测的精度;图 9,10 可以直观地看出,BL + RM 模型预测值与实际值之间的离群点数目更少,相关度更高.

表 7 工作日连续 20 d 预测结果

Tab. 7 Forecast results for 20 consecutive working days

模型	$E_R < 5\%$ 占比/%	$E_R < 10\%$ 占比/%	$E_{MAP}/\%$	E_{RMS}	相关系数 R
BL 模型	57.14	89.64	5.02	2.24	0.982
BL + RM 模型	82.97	98.59	2.94	1.45	0.991

表 8 休息日连续 10 d 预测结果

Tab. 8 Forecast results for 10 consecutive non-working days

模型	$E_R < 5\%$ 占比/%	$E_R < 10\%$ 占比/%	$E_{MAP}/\%$	E_{RMS}	相关系数 R
BL 模型	55.00	86.25	5.36	2.60	0.978
BL + RM 模型	82.92	99.06	2.91	1.48	0.992

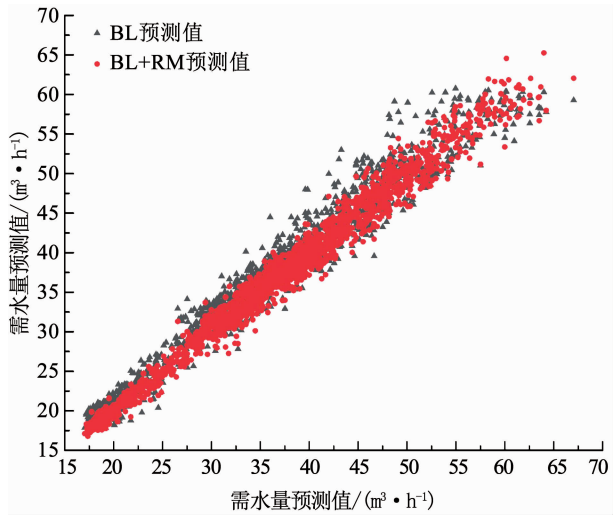


图 9 工作日连续 20 d 预测结果散点图

Fig. 9 Scatter plot for 20 consecutive working days prediction

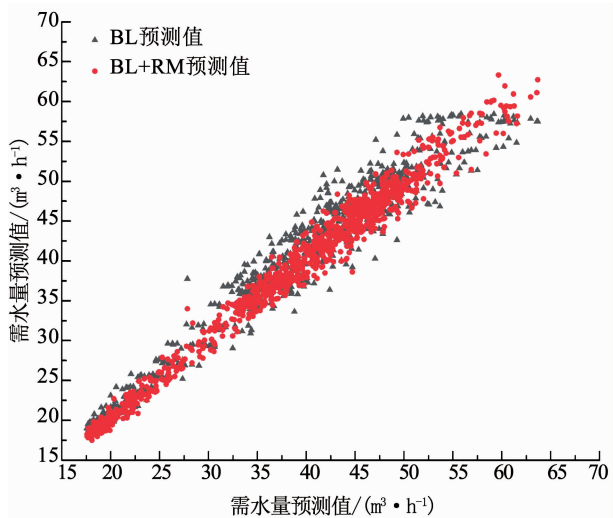


图 10 休息日连续 10 d 预测结果散点图

Fig. 10 Scatter plot for 10 consecutive non-working days prediction

4) A、B、C 3 个小区 BL + RM 模型和 BL 模型预测结果如表 9、10 所示。由表 9、10 可知,A、B、C 3 个小区工作日和休息日的 BL + RM 模型预测结果 E_R 在 10.0% 以下的比例均在 96.0% 以上, E_{MAP} 在 3.0% 左右,模型预测精度较高、精确有效,对多个用户适用性强。

表 9 BL + RM 模型和 BL 模型对 3 个小区工作日连续 20 d 的预测结果

Tab. 9 20-day forecast results of BL + RM model and BL model on working days of three communities

模型	$E_R < 5\%$ %	$E_R < 10\%$ %	$E_{MAP}/\%$	E_{RMS}	R
A 小区 BL + RM 模型	82.97	98.59	2.94	1.45	0.991
A 小区 BL 模型	57.14	89.64	5.02	2.24	0.982
B 小区 BL + RM 模型	80.47	96.51	3.20	1.48	0.985
B 小区 BL 模型	67.66	91.56	4.34	2.12	0.971
C 小区 BL + RM 模型	71.20	97.14	3.67	4.53	0.988
C 小区 BL 模型	55.83	86.67	5.32	7.14	0.973

表 10 BL + RM 模型和 BL 模型对 3 个小区休息日连续 10 d 的预测结果

Tab. 10 10-day forecast results of BL + RM model and BL model on non-working days of three communities

模型	$E_R < 5\%$ %	$E_R < 10\%$ %	$E_{MAP}/\%$	E_{RMS}	R
A 小区 BL + RM 模型	82.92	99.06	2.91	1.48	0.992
A 小区 BL 模型	55.00	86.25	5.36	2.60	0.978
B 小区 BL + RM 模型	76.46	97.71	3.37	1.52	0.984
B 小区 BL 模型	59.27	90.42	4.79	2.27	0.976
C 小区 BL + RM 模型	76.15	97.50	3.43	4.41	0.988
C 小区 BL 模型	58.23	89.79	4.92	6.45	0.977

4 结 论

- 1)RM 模型可以准确捕获 BL 模型预测值残差的变化趋势,修正 BL 模型的预测值,组合 BL 模型和 RM 模型的 BL + RM 模型具有较高的预测精度。
- 2)BL + RM 组合需水量预测模型具有较高的稳定性,适用于平均水量较小、水量波动较大等不同特点用户的需水量预测。
- 3)将工作日和休息日考虑在一起建模时,BL 模型的预测精度较低;根据用水规律的特点分别建立工作日、休息日需水量预测模型,可以较大地提高 BL 模型的预测精度。

参考文献

[1]刘洪波,张宏伟,田林. 人工神经网络法预测时用水量[J]. 中国给水排水,2002,18(12):39
LIU Hongbo, ZHANG Hongwei, TIAN Lin. Artificial neural network method for forecasting hourly water consumption[J]. China Water & Wastewater, 2002, 18(12):39. DOI: 10.3321/j.issn:1000-4602.2002.12.013

[2]陈磊,董志勇. 支持向量机在短期用水量预测中的应用[J]. 浙江工业大学学报,2007,35(4):448
CHEN Lei, DONG Zhiyong. Application of support vector machine to

- predict short-term water consumption [J]. Journal of Zhejiang University of Technology, 2007, 35(4): 448. DOI:10.3969/j.issn.1006-4303.2007.04.021
- [3] 王亮, 张宏伟, 牛志广. 支持向量机在城市用水量短期预测中的应用[J]. 天津大学学报(自然科学与工程技术版), 2005, 38(11): 1021
WANG Liang, ZHANG Hongwei, NIU Zhiguang. Application of support vector machines in short-term prediction of urban water consumption [J]. Journal of Tianjin University (Science and Technology), 2005, 38(11): 1021. DOI:10.3969/j.issn.0493-2137.2005.11.017
- [4] 陈磊, 张土乔. 基于最小二乘支持向量机的时用水量预测模型[J]. 哈尔滨工业大学学报, 2006, 38(9): 1528
CHEN Lei, ZHANG Tuqiao. Hourly water demand forecast model based on least squares support vector machine[J]. Journal of Harbin Institute of Technology, 2006, 38(9): 1528. DOI:10.3321/j.issn:0367-6234.2006.09.031
- [5] 何必仕, 熊晓锋, 蔡华强, 等. 城市供水 DMA 短期需水量预测比较研究[J]. 杭州电子科技大学学报, 2017, 37(2): 51
HE Bishi, XIONG Xiaofeng, CAI Huaqiang, et al. Comparative study on short-term water demand forecast of urban water supply DMA[J]. Journal of Hangzhou Dianzi University, 2017, 37(2): 51. DOI:10.13954/j.cnki.hdu.2017.02.011
- [6] 陈磊, 张土乔. 基于贝叶斯最小二乘支持向量机的时用水量预测模型[J]. 天津大学学报(自然科学与工程技术版), 2006, 39(9): 1037
CHEN Lei, ZHANG Tuqiao. Hourly water demand forecast model based on Bayesian least squares support vector machine[J]. Journal of Tianjin University (Science and Technology), 2006, 39(9): 1037. DOI:10.3969/j.issn.0493-2137.2006.09.005
- [7] 刘大同. 基于 Online SVR 的在线时间序列预测方法及其应用研究[D]. 哈尔滨: 哈尔滨工业大学, 2010
LIU Datong. Online SVR based online time series forecasting method and its application [D]. Harbin: Harbin Institute of Technology, 2010
- [8] 方仍存. 电力系统负荷区间预测[D]. 武汉: 华中科技大学, 2008
FANG Naicun. Power system load interval prediction[D]. Wuhan: Huazhong University of Science and Technology, 2008
- [9] 雷绍兰. 基于电力负荷时间序列混沌特性的短期负荷预测方法研究[D]. 重庆: 重庆大学, 2005
LEI Shaolan. Research on short-term load forecasting based on chaotic characteristics of power load time series [D]. Chongqing: Chongqing University, 2005
- [10] 杨卓. 基于负荷混沌特性和最小二乘支持向量机的短期负荷预测[D]. 西安: 西安理工大学, 2008
YANG Zhuo. Short-term load forecasting based on chaotic characteristics of load and least squares support vector machine [D]. Xi'an: Xi'an University of Technology, 2008
- [11] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer-Verlag Press, 1995
- [12] 郑小霞, 钱锋. 基于证据框架的最小二乘支持向量机在精对苯二甲酸生产中的应用[J]. 化工学报, 2006, 57(7): 1612
ZHENG Xiaoxia, QIAN Feng. Application of least squares support vector machine within evidence framework in PTA process [J]. Journal of Chemical Industry and Engineering (China), 2006, 57(7): 1612. DOI:10.3321/j.issn:0438-1157.2006.07.020
- [13] MACKAY D J C. Probable networks and plausible predictions: A review of practical Bayesian methods for supervised neural networks [J]. Network Computation in Neural Systems, 1995, 6(3): 469. DOI:10.1088/0954-898X_6_3_011
- [14] 阎威武, 常俊林, 邵惠鹤. 一种贝叶斯证据框架下支持向量机建模方法的研究[J]. 控制与决策, 2004, 19(5): 525, 533
YAN Weiwu, CHANG Junlin, SHAO Huihe. Modeling method based on support vector machines within the Bayesian evidence framework [J]. Control and Decision, 2004, 19(5): 525, 533. DOI:10.3321/j.issn:1001-0920.2004.05.010
- [15] 王振树, 李林川, 牛丽. 基于贝叶斯证据框架的支持向量机负荷建模[J]. 电工技术学报, 2009, 24(8): 127
WANG Zhenshu, LI Linchuan, NIU Li. Load modeling based on support vector machine based on Bayesian evidence framework[J]. Transactions of China Electrotechnical Society, 2009, 24(8): 127
- [16] 孙晓东, 陈龙, 杨泽斌, 等. 贝叶斯证据框架下 LS-SVM 的 BPMSM 磁链建模[J]. 浙江大学学报(工学版), 2012, (5): 873
SUN Xiaodong, CHEN Long, YANG Zebin, et al. Modeling of flux linkage for the BPMSM based on LS-SVM within the Bayesian evidence framework [J]. Journal of Zhejiang University (Engineering Science), 2012, (5): 873. DOI:10.3785/j.issn.1008-973X.2012.05.016
- [17] 张弦, 王宏力, 张金生, 等. 状态时间序列预测的贝叶斯最小二乘支持向量机方法[J]. 西安交通大学学报, 2010, 44(10): 42
ZHANG Xian, WANG Hongli, ZHANG Jinsheng, et al. A least squares support vector machine for condition time series prediction based on Bayesian evidence framework [J]. Journal of Xi'an Jiaotong University, 2010, 44(10): 42. DOI:10.7652/xjtub201010008
- [18] 高俊杰. 混沌时间序列预测研究及应用[D]. 上海: 上海交通大学, 2013
GAO Junjie. Research and application of chaotic time series prediction [D]. Shanghai: Shanghai Jiaotong University, 2013
- [19] 朱志宇, 姜长生, 张冰. 基于支持向量回归的混沌序列预测方法[J]. 电工技术学, 2005, 20(6): 57
ZHU Zhiyu, JIANG Changsheng, ZHANG Bing. Prediction of chaotic time series based on support vector regression [J]. Transactions of China Electrotechnical Society, 2005, 20(6): 57. DOI:10.3321/j.issn:1000-6753.2005.06.012
- [20] 信昆仑, 陶涛, 李树平, 等. 考虑气象因子的城市日用水量预测模型[J]. 武汉大学学报(工学版), 2009, 42(4): 461
XIN Kunlun, TAO Tao, LI Shuping, et al. Meteorological factors involved urban water demand forecast model [J]. Engineering Journal of Wuhan University, 2009, 42(4): 461
- [21] BAKKER M, DUIST H V, SCHAGEN K V, et al. Improving the performance of water demand forecasting models by using weather input[J]. Procedia Engineering, 2014, 70(70): 93. DOI:10.1016/j.proeng.2014.02.012
- [22] AL-ZAHRANI M A, ABO-MONASAR A. Urban residential water demand prediction based on artificial neural networks and time series models[J]. Water Resources Management, 2015, 29(10): 3651. DOI:10.1007/s11269-015-1021-z
- [23] BRENTAN B M, LUVIZOTTO JR E, HERRERA M, et al. Hybrid regression model for near real-time urban water demand forecasting [J]. Journal of Computational & Applied Mathematics, 2017, 309(C): 532. DOI:10.1016/j.cam.2016.02.009