

网络出版时间:2016-11-24 13:52 DOI:10.13207/j.cnki.jnwafu.2017.01.029
网络出版地址:http://www.cnki.net/kcms/detail/61.1390.S.20161124.1352.058.html

基于 KPCA-RVM 的土石坝沉降预测模型研究

马春辉^{1,2}, 杨 杰^{1,2}, 程 琳^{1,2}, 李 婷^{1,2}, 陈 容³

(1 西安理工大学 水利水电学院, 陕西 西安 710048; 2 西北旱区生态水利工程国家重点实验室培育基地, 陕西 西安 710048;
3 甘肃大唐碧口水力发电厂, 甘肃 文县 746412)

[摘 要] 【目的】针对土石坝坝体沉降存在多变量、强耦合、强干扰的复杂问题, 建立基于 KPCA-RVM 的土石坝沉降预测模型。【方法】利用核主元分析(KPCA)对输入向量进行降维处理, 以减少因子个数, 随后利用相关向量机(RVM)模型对土石坝沉降进行预测, 并以平均相对误差为指标对预测精度进行评价。【结果】实例应用表明, KPCA-RVM 模型将输入向量由 14 个降低到 7 个, 预测结果的平均相对误差仅为 0.9%, 预测效果得到明显提升。【结论】利用 KPCA-RVM 模型对土石坝进行沉降预测, 不仅可以减少输入向量个数, 而且可以提高预测精度, 可在实际工程中推广应用。

[关键词] 土石坝; KPCA-RVM 模型; 沉降预测; 核主元分析; 相关向量机

[中图分类号] TV641; TV698.1⁺1 **[文献标志码]** A **[文章编号]** 1671-9387(2017)01-0211-07

KPCA-RVM based prediction model for settlement of earth-rockfill dam

MA Chunhui^{1,2}, YANG Jie^{1,2}, CHENG Lin^{1,2}, LI Ting^{1,2}, CHEN Rong³

(1 Institute of Water Resources and Hydro-electric Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China;
2 State Key Laboratory Base of Eco-hydraulic Engineering in Arid Area, Xi'an, Shaanxi 710048, China;
3 Datang Bikou Hydropower Plant, Wen county, Gansu 746412, China)

Abstract: 【Objective】A KPCA-RVM based prediction model for settlement of earth-rockfill dam was established aiming at the complex characteristics of multi variables, strong coupling and strong interference in settlement of earth-rockfill dams. 【Method】The kernel principal component analysis (KPCA) was used to reduce the number of the input vectors. Then, the settlement of earth-rockfill dam was predicted using the relevant vector machine (RVM) model, and the prediction accuracy was evaluated using average relative error. 【Result】The number of input vectors was reduced from 14 to 7 by KPCA-RVM model. The average relative error of prediction results was only 0.9%, indicating the prediction was significantly improved. 【Conclusion】Using KPCA-RVM model to predict settlement of earth dam not only reduced the number of input vectors, but also improved the prediction accuracy. The KPCA-RVM model has great application in practical projects.

Key words: earth-rockfill dam; KPCA-RVM model; settlement prediction; kernel principal component analysis; relevance vector machine

〔收稿日期〕 2015-09-09

〔基金项目〕 国家自然科学基金项目(51409205); 陕西省重点科技创新团队项目(2013KCT-15); 博士后自然科学基金项目(2015M572656XB); 水文水资源与水利工程科学国家重点实验室开放研究基金项目(2014491011); 西安理工大学水利水电学院青年科技创新团队项目(2016ZZKT-14)

〔作者简介〕 马春辉(1993—), 男, 山西晋中人, 在读硕士, 主要从事水工结构数值仿真及安全监控研究。
E-mail: shanximachunhui@foxmail.com

〔通信作者〕 杨 杰(1971—), 男, 四川大邑人, 教授, 博士, 主要从事水工结构、水库大坝安全及除险加固理论研究。
E-mail: lgdajs@163.com

由于土石坝是由多种材料组成的散粒体,在荷载作用下,其竖直位移要较混凝土坝大得多,变位的大小和时间效应对坝的安全以及裂缝的出现等都是重要影响因素^[1-2],因此通过已有的监测数据,建立可靠的大坝沉降预测模型,对大坝的安全监控与预警具有十分重要的意义。目前,常用的大坝变形预测模型主要有统计模型、灰色预测模型、偏最小二乘回归、支持向量机模型(SVM)和相关向量机模型(RVM)等。其中统计模型简单直观,应用广泛,但精度和泛化能力较差^[3];灰色预测模型若所选变量不当,矩阵及其逆矩阵会出现病态,致使参数失真^[4-5];偏最小二乘回归存在因子多重相关性干扰和模型拟合效果不佳的问题^[6];SVM虽有较好的数据回归与分类能力,但存在参数影响大、对缺失数据敏感和核函数需满足 Merce 条件等问题^[7-8]。RVM 是在贝叶斯理论框架之下的概率模型,其适用范围和泛化能力与 SVM 相当,且具有模型结构稀疏性、计算复杂度相对较低、可提供方差和所需参数少等优势^[9]。当输入变量之间存在着非线性特征及噪声时,预测模型的精度、泛化能力和计算时间会受到较大影响^[10]。因此,采用核主元分析法(KPCA)可从原始数据集中提取非线性特征量,缩短计算时间。相较于主元分析(PCA),KPCA 模型能够有效处理变量之间的非线性强耦合问题,起到数据校正、降噪功能^[11]。

为此,本研究建立 KPCA-RVM 模型,利用 KPCA 的非线性特征提取能力,将原始数据通过核主元分析进行降维、重组,然后利用 RVM 良好的函数逼近能力进行土石坝沉降预测,以期为大坝沉降的准确预测提供新的思路。

1 KPCA-RVM 模型的原理

1.1 核主元分析(KPCA)的原理

首先给定训练样本 $h = \{h_1, h_2, \dots, h_l\}$, 其中第 k 个样本 $h_k \in \mathbf{R}$, $k=1, 2, 3, \dots, l$, l 为样本总数, h_k 可由多个自变量组成。定义非线性映射 $\varphi(h_k)$, $\mathbf{R} \rightarrow F$, 通过函数 $\varphi(h_k)$ 可将样本从原始输入空间映射到高维特征空间 F 中^[10]。假设映射后其均值为 0, 即满足 $\frac{1}{l} \sum_{k=1}^l \varphi(h_k) = 0$, 则 F 空间中的样本协方差矩阵 $\mathbf{C}$ 为:

$$\mathbf{C} = \frac{1}{l} \sum_{k=1}^l \varphi(h_k) \varphi(h_k)^T. \quad (1)$$

其对应的特征方程为:

$$\lambda \mathbf{V} = \mathbf{C} \mathbf{V}. \quad (2)$$

式中: λ 为特征值组成的向量, $\mathbf{V}$ 为特征向量。

所有特征值非 0 的特征向量 $\mathbf{V}$, 均可由映射到特征空间的样本矢量线性表示。因此, 存在系数 ∂_k , 使下式成立:

$$\mathbf{V} = \sum_{k=1}^l \partial_k \varphi(h_k). \quad (3)$$

将式(1)、(3)代入式(2)中, 在等式两边同乘内积 $\varphi(h_j)$ ($j=1, 2, \dots, l$, j 与 k 的取值可相等, 也可不相等), 同时定义 $l \times l$ 的核矩阵 $\mathbf{K}$, 则转换为求矩阵 $\mathbf{K}$ 的特征值和特征向量:

$$l \lambda \partial = \mathbf{K} \partial. \quad (4)$$

式中: $\partial = [\partial_1, \partial_2, \dots, \partial_l]$, 核矩阵 $\mathbf{K}$ 的各元素 K_{kj} 为:

$$K_{kj} = K(h_k, h_j) = [\varphi(h_k) \cdot \varphi(h_j)]. \quad (5)$$

式中: $K(h_k, h_j)$ 为核函数, 常用的核函数形式有多项式、径向基和 Sigmoid 核函数。

某一特征值占特征值总和的比率, 反映其相应主元对整体方差的贡献率。一般选取前几个较大特征值对应的特征向量构成特征子空间。假设提取 m 个非零特征值的特征向量 $\beta^1, \beta^2, \dots, \beta^m$, 其中 $\beta^i = (\beta^i_1, \beta^i_2, \dots, \beta^i_l)$, 其中 $i=1, 2, \dots, m$ 。根据下式对 F 空间的相应主元 v_i 进行归一化处理, 使其作为 F 空间的基矢量^[12]。

$$\lambda_i (\beta^i \cdot \beta^i) = 1. \quad (6)$$

式中: λ_i 为相应主元 v_i 对应的特征值。

则对任一样本数据 h_k , 均可映射至 F 空间中, 表示为 $(r_1, r_2, \dots, r_m)$, 其投影分量 r_i ($i=1, 2, 3, \dots, m$) 为:

$$r_i = v_i \varphi(h_k) = \sum_{j=1}^m \beta_j^i K(h_k, h_j). \quad (7)$$

在上述推导过程中, 均以假设变量的均值为 0 作为前提, 实际情况中样本数据并不满足这个条件。因此, 将核矩阵 $\mathbf{K}$ 替换为 $\tilde{\mathbf{K}}$, 其形式如下:

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{L}_{l \times l} \mathbf{K} - \mathbf{K} \mathbf{L}_{l \times l} + \mathbf{L}_{l \times l} \mathbf{K} \mathbf{L}_{l \times l}. \quad (8)$$

式中: $\mathbf{L}_{l \times l}$ 为 $l \times l$ 单位矩阵, 其系数为 $1/l$ 。

1.2 相关向量机模型(RVM)的原理

RVM 是基于贝叶斯框架的概率学习模型, 其基本原理为: 首先定义受超参数控制的 Gaussian 先验概率, 其在贝叶斯框架下进行机器学习, 并采用主动相关判定理论(ARD)移除不相关点, 从而得到稀疏化模型^[13]。RVM 不仅可以定量预测, 而且能够提供方差。

给定 N 组训练样本 $\{z_n, c_n\}_{n=1}^N$, 其中 z_n 为输入参数值, 可由多个自变量组成, c_n 为真实目标值。设 $\{z_e, c_e\}$ 为训练样本中的某一组样本, c_n, c_e 中的 e 与 n 的取值可相等, 也可不相等。

$$c_e = q(z_e; w) + \epsilon_e = \sum_{n=1}^N \omega_n K(z_e, z_n) + \omega_0 + \epsilon_e. \quad (9)$$

式中: q 为模型拟合值; w 为参数向量, $w = (\omega_0, \omega_1, \dots, \omega_N)^T$, 其中 ω_0 为基础量参数, $\omega_n (n=1, 2, \dots, N)$ 为参数; ϵ_e 为误差, 服从均值为 0、方差为 σ^2 的高斯分布; $K(z_e, z_n)$ 为核函数。设 c_n 为独立分布, 则似然函数 p 可表示为:

$$p(c|w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{\|c - \Phi w\|^2}{2\sigma^2}\right). \quad (10)$$

式中: c 为目标向量, $c = (c_1, c_2, \dots, c_N)^T$; Φ 为基函数, $\Phi = [\varphi(z_1), \varphi(z_2), \dots, \varphi(z_N)]$, 其中 $\varphi(z_n) = [1, K(z_n, z_1), K(z_n, z_2), \dots, K(z_n, z_N)]^T$ 。

若利用最大化似然函数估计参数向量 w , 可导致因参数过多而使模型过学习。因此, 假设参数 ω_n 服从均值为 0、方差为 α_n^{-1} 的高斯分布 G , 则有:

$$p(w|\alpha) = \prod_{n=0}^N G(\omega_n | 0, \alpha_n^{-1}). \quad (11)$$

式中: α 为决定参数向量 w 先验分布的超参数, 只与其对应参数向量 w 相关。

假定 α, σ^2 均服从 Gamma 先验概率, 根据定义的似然分布和先验分布, 可得参数向量 w 的后验分布为:

$$p(w|c, \alpha, \sigma^2) = \frac{p(c|w, \sigma^2) p(w|\alpha)}{p(c|\alpha, \sigma^2)} = (2\pi)^{-M/2} |\Sigma|^{1/2} \exp\left(-\frac{1}{2}(w - \mu)^T \Sigma^{-1} (w - \mu)\right). \quad (12)$$

式中: Σ 为后验协方差, $\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}$, 其中 $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$; $\mu = \sigma^{-2} \Sigma \Phi^T t$, 为后验均值; M 为相关向量的个数。

对训练样本似然函数中的参数向量 w 进行积分, 可得到由超参数 α, σ^2 控制的边缘分布为:

$$p(c|\alpha, \sigma^2) = (2\pi)^{-N/2} |\Omega| \exp\left(\frac{-c^T \Omega^{-1} c}{2}\right). \quad (13)$$

式中: $\Omega = \sigma^2 E + \Phi A^{-1} \Phi^T$, 其中 E 为单位向量。

采用迭代的方法对上式求偏导, 以得到边缘分布最大化的 α 和 σ^2 。在计算过程中, 多出现某组样本超参数 $\alpha_n \rightarrow \infty$ 的情况。若某组样本 μ_n 趋于 0, 则将其对应的基向量删除, 令对应的 $\omega_n = 0$, 从而得到稀疏化的 RVM。考虑到 RVM 需要计算协方差矩阵 Σ , 计算量比较大, 运行时间长, 本研究采用改进后的快速序列稀疏贝叶斯学习算法^[14]。

在获得超参数 α 、噪声方差 σ^2 后, 由于权值最优估计的不确定性, 给定任意一个输入样本 z^* , 则可描述预测不确定度的均值 q^* 和方差 σ^{2*} ^[13], 计算公式如下:

$$q^* = \mu^T \varphi(z^*), \quad (14)$$

$$\sigma^{2*} = \sigma_{MP}^2 + \varphi(z^*)^T \Sigma \varphi(z^*). \quad (15)$$

式中: σ_{MP} 为超参数的最优值。

2 基于 KPCA-RVM 的建模

建立基于 KPCA 和 RVM 的土石坝沉降预测模型: 先采用 KPCA 对原始数据进行预处理, 提取贡献率较高的主元, 同时达到降低数据维数的效果; 再利用 RVM 对重新映射后的数据进行训练, 建立预测模型; 通过调整相关参数, 选取拟合效果最佳的参数组合作为模型的最终参数; 根据 RVM 提供的方差, 以概率分布的形式描述沉降预报的不确定度, 确定预报区间。图 1 为基于 KPCA-RVM 的土石坝沉降预测模型的工作流程图。

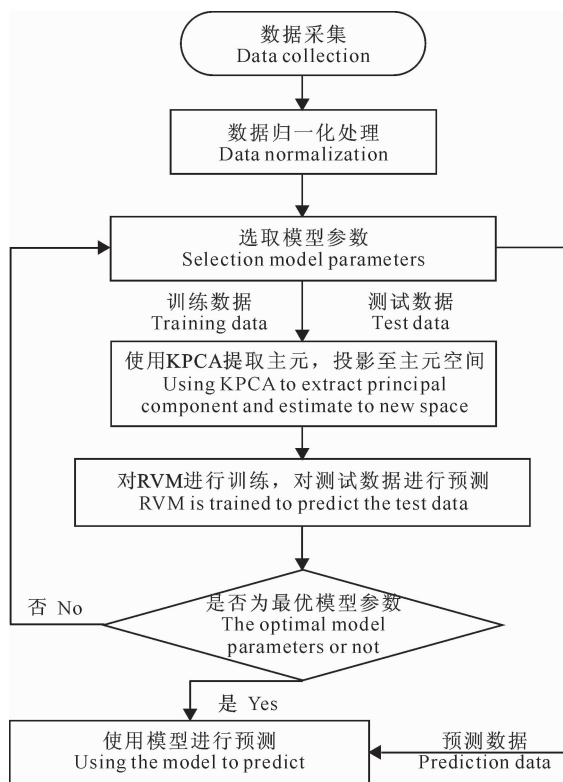


图 1 基于 KPCA-RVM 的土石坝沉降预测模型的工作流程

Fig. 1 Flow chart of the KPCA-RVM based prediction model for settlement of earth-rockfill dam

从图 1 可知, KPCA-RVM 模型的详细建模步骤为:

(1) 设 S 组训练数据集为 $\{X_n, Y_n\}_{n=1}^S$, 其中每组输入数据均由 a 个因子组成, 即 $X_n = \{x_{n1}, x_{n2}, \dots, x_{na}\}$, 输出数据为 Y_n 。采用线性函数法或标准正态法均可对样本输入数据 $x_{nz} (z=1, 2, 3, \dots, a)$ 进行归一化处理, 本研究采用线性函数法进行归一化处理, 得到 x'_{nz} :

$$x'_{nz} = 0.1 + 0.8(x_{nz} - x_{\min}) / (x_{\max} - x_{\min}). \quad (16)$$

式中: $x_{\max}$ 为 $\max_{1 \leq z \leq a}(x_{nz})$, $x_{\min}$ 为 $\min_{1 \leq z \leq a}(x_{nz})$ 。

(2) 选用径向基函数作为 KPCA 的核函数, 表示为:

$$K(X_b, X_u) = \exp\left[-\frac{\|X_b - X_u\|^2}{g}\right] \quad (17)$$

式中: X_b, X_u 均为任意一组输入数据, b, u 的取值可相等, 也可不相等, $b, u = 1, 2, 3, \dots, S$; g 为 KPCA 核函数的参数。

为计算训练数据的核函数矩阵, 首先利用 KPCA 对其进行非线性主元分析, 将主元按贡献率由大到小排序, 本研究选取累积贡献率达到 70% 左右的前 n 个主元, 其余可认为是噪声成分。

(3) RVM 核函数同样选用径向基函数, 以 g' 为 RVM 核函数的参数。将降维后的训练数据作为 RVM 的输入样本, 对 RVM 模型进行训练, 随后采用训练完成的模型对测试数据进行预测。

(4) 模型的主要参数包括: KPCA 的核参数 g 和 RVM 的核参数 g' 。通过调整参数, 循环代入步骤(2)与(3)中进行计算。最终选取测试数据拟合效果最佳的参数组合作为模型的最终参数^[15]。

(5) 使用最终建立的 KPCA-RVM 模型对测试数据进行预测。

针对测试数据的处理需注意以下问题:

(1) 对测试数据进行 KPCA 计算时, 无需重新进行核主元分析。使用测试数据和训练数据共同建立核矩阵, 再与训练数据的转化矩阵相乘, 即可得到测试数据在训练数据特征空间中的投影。

(2) 通过对比核主元分析计算的结果可知, 核主元分析时, 若数据范围包含训练数据, 则相较于不包含训练数据的计算结果, 模型精度有大幅度的提高。这是因为在进行数据归一化处理时, 若数据范围包

含训练数据, 相较于仅采用测试数据, 其归一化结果更加合理。

(3) 在采用已训练完毕的 RVM 进行预测时, 数据范围仅为测试数据, 无需包含训练数据。

3 基于 KPCA-RVM 土石坝沉降模型的应用

某水电枢纽工程位于我国甘肃省境内, 属大 II 型工程, 水库校核洪水位 708.80 m, 正常高水位 704.00 m, 水库总库容 $5.21 \times 10^8 \text{ m}^3$, 为季调节水库。坝体为碾压式壤土心墙土石坝, 最大坝高 101.8 m, 坝顶长度 297.36 m。测点 D13-3 位于坝左 0+304.00 m, 坝下 0+041.50 m 处。选取 D13-3 测点的竖向位移人工监测数据为研究样本, 共有监测数据 205 组, 以前 190 组作为训练数据, 后 15 组为测试数据。

首先建立输入向量。大坝的沉降影响因素通常为温度、水压和时效 3 个因子, 每个因子由若干向量组成, 本模型共有 14 个输入向量^[16], 分别为: 温度因子 $\sin \frac{2\pi t}{365} - \sin \frac{2\pi t_0}{365}$ 、 $\cos \frac{2\pi t}{365} - \cos \frac{2\pi t_0}{365}$ 、 $\sin \frac{4\pi t}{365} - \sin \frac{4\pi t_0}{365}$ 和 $\cos \frac{4\pi t}{365} - \cos \frac{4\pi t_0}{365}$, 水压因子 $H - H_0$ 、 $(H - H_0)^2$ 、 $(H - H_0)^3$ 、 $(H - H_0)^4$ 、 $\bar{H}_{60-31}$ 、 $\bar{H}_{30-11}$ 、 $\bar{H}_{10-4}$ 和 $\bar{H}_{3-1}$, 时效因子 $\theta - \theta_0$ 和 $\ln \theta - \ln \theta_0$ 。其中: t_0 为训练样本的第 1 个监测日至始测日(1977-01-02)累计时间, d ; t 为监测日至始测日累计时间, d ; H, H_0 分别为监测日和始测日所对应上游水头; $\bar{H}_{60-31}$ 、 $\bar{H}_{30-11}$ 、 $\bar{H}_{10-4}$ 、 $\bar{H}_{3-1}$ 分别为对应监测日前 60~31 d、30~11 d、10~4 d 和 3~1 d 上游水位的平均值; θ 和 θ_0 为分别用 t 和 t_0 除以 100。水压作为大坝沉降的重要影响因素, 实测样本数据的上游水位变化如图 2 所示。

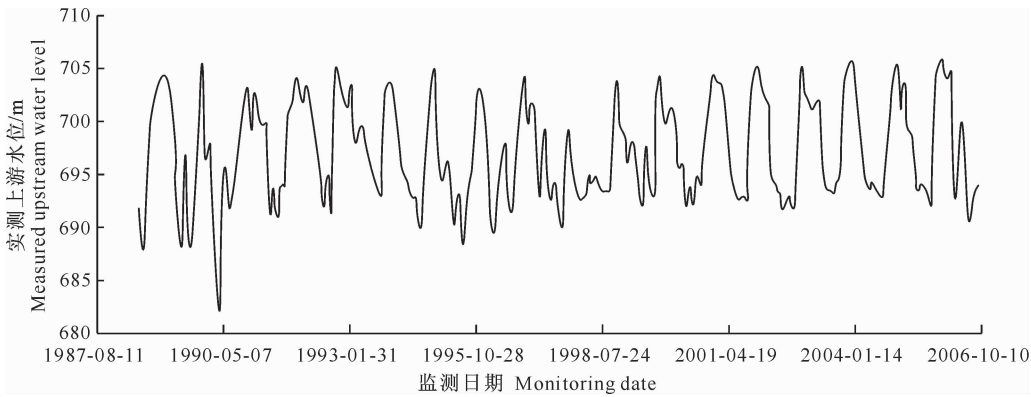


图 2 实测样本数据上游水位的变化

Fig. 2 Measured change of upstream water level

在 KPCA 计算核矩阵过程中, 将输入矩阵(190×14)变为核矩阵(190×190), 随后对核矩阵进行主元

提取。因此,KPCA 的主元提取数可大于 14 个,最高达到 190 个,主元提取个数应通过分析研究确定。当核参数固定为 $g=25.6, g'=5.76$ 时,KPCA 提取主元数目与模型检验结果的关系如图 3 所示。

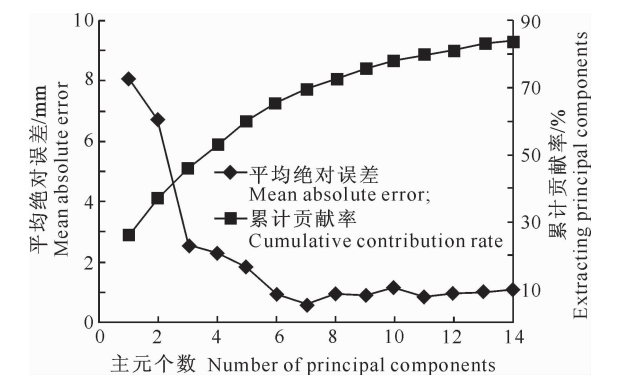


图 3 KPCA 提取的主元个数与模型平均绝对误差及累计贡献率的关系

Fig. 3 Relationship between the number of principal components extracted by KPCA and the mean absolute error and cumulative contribution rate

由图 3 可知,若仅改变提取的主元数目,随着 KPCA 提取主元个数的增加,KPCA-RVM 模型的平均绝对误差(MAE)先呈下降趋势,当主元个数为 7 时达到最小;当主元个数增加到 8 时,MAE 又有小幅增大;其后 MAE 略有波动但逐渐趋于平稳。由此表明,合理采用 KPCA 提取主元,可达到消除输入数据中的噪声、降低数据维数、提高模型预测精度的目的,可用数据维数较低的主元完整替代原始数据。计算表明,被提取的主元个数并不是越多越好,多余的主元反而会夹带更多的噪声,进而影响模型的预测精度^[17]。本研究的主元数取为 7 个,对应累计贡献率为 69.3%,相较于原始数据的 14 个输入向量,数据维数降低效果十分明显。

图 4 为传统模型(HST)和 KPCA-RVM 模型训练数据拟合值与大坝沉降实测值的对比。由图 4 可知:相较于传统的 HST 模型,KPCA-RVM 模型的拟合效果提升明显,为大坝沉降的准确预测奠定了基础。

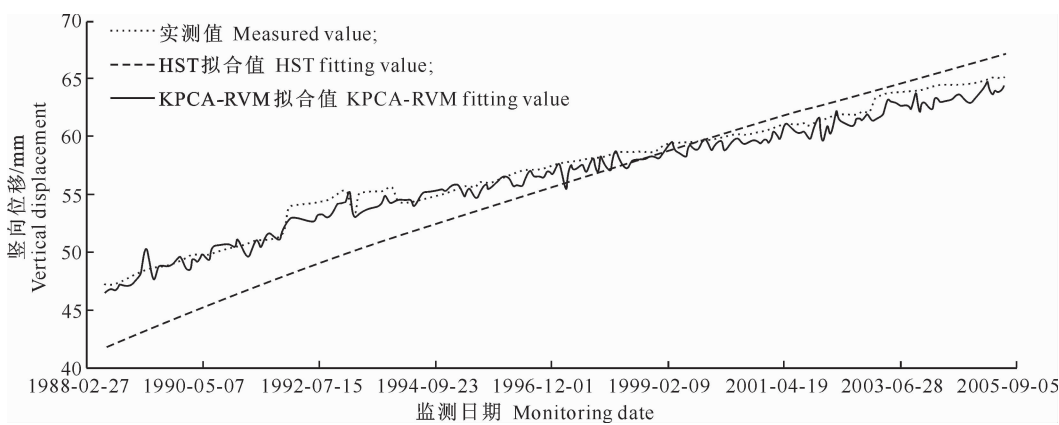


图 4 基于 HST 和 KPCA-RVM 模型大坝沉降拟合值与实测值的对比

Fig. 4 Comparison of predicted dam settlement values by HST and KPCA-RVM models and the measurements

为证明 KPCA-RVM 模型的学习和泛化能力,本研究同时建立了 HST、SVM、RVM、PCA-SVM、PCA-RVM、KPCA-SVM 等模型。采用相同的训练数据对各模型进行训练,并针对同一测试数据进行

预测,随后使用最大相对误差、平均相对误差和平均绝对误差对预测效果进行评价,其预测精度的对比如表 1 所示。

表 1 基于训练数据的不同预测模型参数及误差的比较

模型 Model	KPCA 核参数 KPCA parameter	SVM 或 RVM 参数 SVM or RVM parameter		最大相对误差/% Maximum relative error	平均相对误差/% Average relative error	平均绝对误差/mm Mean absolute error
		C	g'			
HST	—	—	—	4.11	3.55	2.325 6
SVM	—	600	41.0	2.87	2.60	1.702 5
RVM	—	—	11.0	3.92	2.34	1.535 7
PCA-SVM	—	100	2.0	2.84	2.56	1.678 9
PCA-RVM	—	—	5.4	3.30	1.93	1.264 1
KPCA-SVM	21.0	60	8.0	2.98	1.45	0.953 0
KPCA-RVM	25.6	—	5.76	2.44	0.90	0.591 5

由表 1 可知,传统的 HST 模型有较大的误差;而 SVM、RVM 模型的预测效果明显好于 HST 模型,预测误差大幅度降低;使用 PCA 模型对输入数据进行主元提取后,再使用 SVM、RVM 模型进行预测,精度得到了小幅度的提高;使用 KPCA 模型对输入数据进行非线性主元提取后,KPCA-SVM 和 KPCA-RVM 模型的预测精度均有较大幅度提高,主要原因是 KPCA 对原始数据的非线性提取效果十分显著。综上所述,基于 KPCA-RVM 的土石

坝沉降预测模型有较好的预测精度,且在计算过程中 RVM 模型的计算速度明显快于 SVM。

图 5 为 HST、RVM、PCA-RVM、KPCA-SVM、KPCA-RVM 模型测试数据的预测值与大坝沉降实际观测值的对比。由图 5 可知:KPCA-RVM 模型的预测精度较传统 HST 模型有十分明显的提高,与 RVM、PCA-RVM、KPCA-SVM 相比,预测精度和预测稳定性也有明显提高,但也有部分预测值偏离明显,尚有待于继续改进。

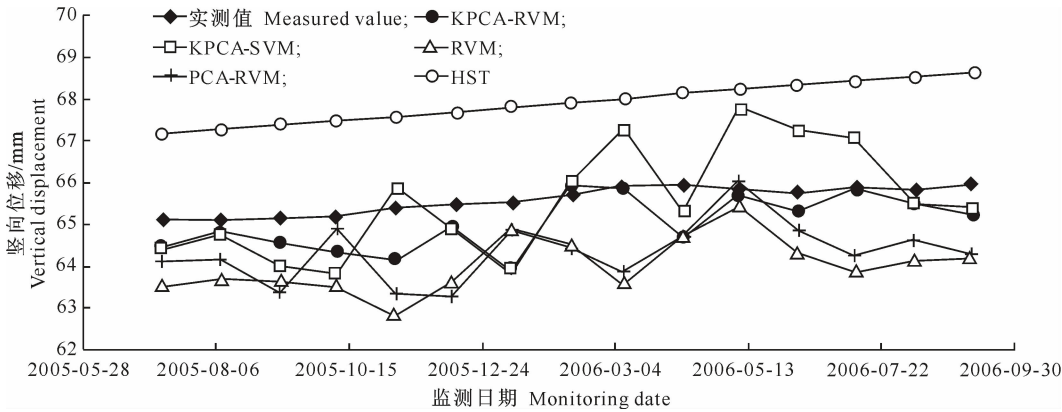


图 5 基于测试数据的 KPCA-RVM 等模型预测值与实测值的比较

Fig. 5 Comparison of the prediction values by KPCA-RVM and other models and the measured values based on test data

将预报值的均值和方差作为预报的后验概率分布函数,对预报值的不确定性进行分析,以进一步考虑沉降预报的不确定性。发生概率为 99.73% 的区

间预报结果及实测值如图 6 所示。由图 6 可知:沉降预报区间可将实测沉降值完全包含,因此区间预报结果可靠。

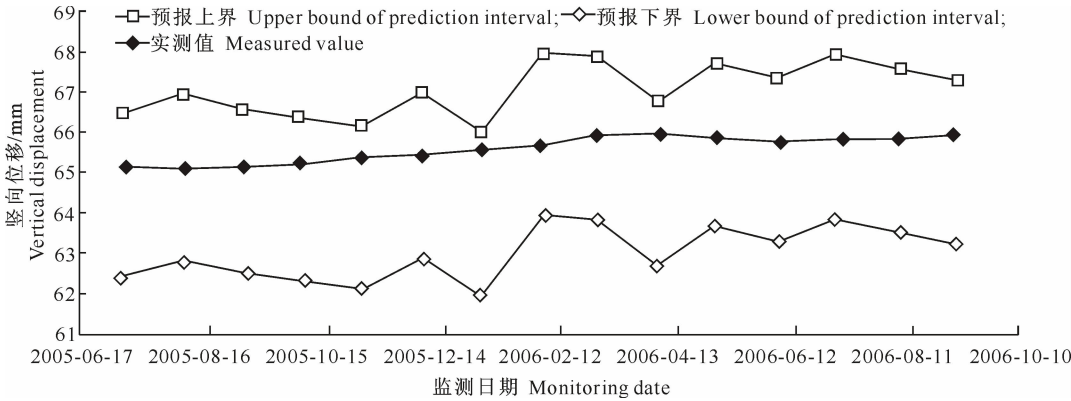


图 6 KPCA-RVM 模型预测区间与实测值的对比

Fig. 6 Comparison of prediction intervals by KPCA-RVM model and the measured values

4 结 论

本研究建立 KPCA-RVM 模型对土石坝沉降进行预测。首先,采用 KPCA 对原始数据的主元进行非线性提取,显著降低了预测模型的输入维数,并消除噪声,提高了后续数据的处理速度;随后,采用基

于贝叶斯框架的 RVM 学习模型进行预测;所建立的 KPCA-RVM 模型具有预测精度高、运行速度快等优势。由于 RVM 模型不仅可以定量预测,而且能够计算预测值的方差,由此建立预测区间,以概率分布的形式描述沉降预报的不确定度,可为实际水利工程的大坝监测提供更多可利用信息。因此,

KPCA-RVM 模型在实际工程中具有较大的实用性。但需要说明的是,由于 KPCA 与 RVM 均采用了核函数,而目前对核函数及核参数的选择并无统一标准,关于如何更合理地选择核函数及核参数,尚有待进一步深入研究。

参考文献

- [1] 吴中如. 水工建筑物安全监控理论及其应用 [M]. 北京:高等教育出版社,2003.
Wu Z R. Safety monitoring theory & its application of hydraulic structures [M]. Beijing: Higher Education Press, 2003.
- [2] 顾冲时, 吴中如. 大坝与坝基安全监控理论和方法及其应用 [M]. 南京: 河海大学出版社, 2006.
Gu C S, Wu Z R. Safety monitoring of dams and dam foundations: theories & methods and their application [M]. Nanjing: Hohai University Press, 2006.
- [3] 王彭煦, 宋文晶. 水布垭面板坝实测沉降分析与土石坝沉降统计预报模型 [J]. 水力发电学报, 2009, 28(4): 81-85.
Wang P X, Song W J. Analysis on the measured settlement and prediction modeling of statistic settlement of Shuibuya CFRD [J]. Journal of Hydroelectric Engineering, 2009, 28(4): 81-85.
- [4] 刘国华, 何勇兵, 汪玉树. 土石坝沉降预测中的多变量灰色预测模型 [J]. 水利学报, 2003(12): 84-88.
Liu G H, He Y B, Wang S Y. Application of grey multi-variable forecasting model for the settlement of earth dam [J]. Journal of Hydraulic Engineering, 2003(12): 84-88.
- [5] 何自立, 杨建国, 靳国云, 等. 基于新维无偏灰色马尔可夫模型的大坝沉降预测研究 [J]. 西北农林科技大学学报(自然科学版), 2013, 41(5): 213-218.
He Z L, Yang J G, Jin G Y, et al. Dam settlement prediction based on metabolism unbiased Grey-Markov model [J]. Journal of Northwest A&F University(Nat Sci Ed), 2013, 41(5): 213-218.
- [6] 杨杰, 杨丽, 李建伟, 等. 基于改进遗传算法-偏最小二乘回归的大坝变形监测模型 [J]. 西北农林科技大学学报(自然科学版), 2010, 38(2): 206-210.
Yang J, Yang L, Li J W, et al. Monitoring model for dam deformation based on partial least-squared regression and improved genetic algorithm [J]. Journal of Northwest A&F University(Nat Sci Ed), 2010, 38(2): 206-210.
- [7] 韩立伟, 李宗坤, 王复明, 等. 基于反演的 LS-SVM 及其在土石坝沉降中的应用 [J]. 水电能源科学, 2008, 26(1): 104-106.
Han L W, Li Z K, Wang F M, et al. Application of LS-SVM based on back-analysis in rockfill dam settlement analysis [J]. Water Resources and Power, 2008, 26(1): 104-106.
- [8] 李波, 顾冲时, 李智录, 等. 基于偏最小二乘回归和最小二乘支持向量机的大坝渗流监控模型 [J]. 水利学报, 2008, 39(12): 1390-1394.
Li B, Gu C S, Li Z L, et al. Monitoring model for dam seepage based on partial least-squares regression and partial least square support vector machine [J]. Journal of Hydraulic Engineering, 2008, 39(12): 1390-1394.
- [9] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 2001, 1: 211-244.
- [10] 徐晔, 杜文莉, 钱锋. 基于核主元分析和最小二乘支持向量机的软测量建模 [J]. 系统仿真学报, 2007, 19(17): 3873-3875.
Yan Y, Du W L, Qian F. Soft sensor modeling based on KPCA and least square SVM [J]. Journal of System Simulation, 2007, 19(17): 3873-3875.
- [11] Hofmann T, Schölkopf B, Smola A J. Kernel methods in machine learning [J]. Annals of Statistics, 2008, 36: 1171-1220.
- [12] 高海华, 杨辉华, 王行愚. 基于 PCA 和 KPCA 特征抽取的 SVM 网络入侵检测方法 [J]. 华东理工大学学报(自然科学版), 2006, 32(3): 321-326.
Gao H H, Yang H H, Wang X Y. PCA /KPCA feature extraction approach to SVM for anomaly detection [J]. Journal of East China University of Science and Technology(Nat Sci Ed), 2006, 32(3): 321-326.
- [13] 颜学峰, 陈佳, 胡春平, 等. KPCA-RVM 组合建模方法及其在软测量中的应用 [J]. 石油化工高等学校学报, 2009, 22(1): 82-85.
Yan X F, Chen J, Hu C P, et al. KPCA-RVM modeling method and its application for soft sensor [J]. Journal of Petrochemical Universities, 2009, 22(1): 82-85.
- [14] Tipping M E, Faul A C. Fast marginal likelihood maximisation for sparse bayesian models [C]//Avenuse J J T. Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. Florida: Key West, 2003: 1-13.
- [15] 李雪红, 徐洪钟, 顾冲时, 等. 主元神经网络模型在大坝观测资料分析中的应用 [J]. 大坝观测与土工测试, 2001, 25(5): 14-16.
Li X H, Xu H Z, Gu C S, et al. Application of principal component fuzzy neural network model to analysis of observation data of dams [J]. Dam Observation and Geotechnical Tests, 2001, 25(5): 14-16.
- [16] 杜传阳, 郑东健, 陈敏, 等. 基于 APPSO-RVM 与 APPSO-SVM 的大坝安全预警模型的应用比较研究 [J]. 水电能源科学, 2009, 22(1): 45-49.
Du C Y, Zheng D J, Chen M, et al. Application and comparison of SVM and RVM algorithm in dam safety modeling [J]. Water Resources and Power, 2009, 22(1): 45-49.
- [17] 邵年华, 沈冰, 黄领梅, 等. KPCA_SVM 水文时间序列预测模型的建立与应用 [J]. 西北农林科技大学学报(自然科学版), 2009, 37(9): 213-219.
Shao N H, Shen B, Huang L M, et al. Establishment and application of hydrological time series forecasting model based on KPCA_SVM [J]. Journal of Northwest A&F University (Nat Sci Ed), 2009, 37(9): 213-219.