

基于聚类分区和 MO-LSSVR 的高拱坝变形预测模型

刘伟琪^{1,2}, 陈波^{1,2}, 葛盼猛^{1,2}, 张晓玲³

(1. 河海大学水利水电学院, 江苏 南京 210098;
2. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098;
3. 山东省华诚工程咨询监理有限公司, 山东 高密 261500)

摘要:为解决传统单测点监控模型未考虑多测点间的内在关联,难以反映高拱坝变形区域分布特征的问题,提出了基于聚类分区和多输出最小二乘支持向量回归机(MO-LSSVR)的高拱坝变形预测模型。模型基于测点之间的复合相似性指标,借助层次凝聚聚类(HAC)算法实现空间测点的聚类分区,再利用融合测点关联特性的 MO-LSSVR 对分区内多测点进行建模。工程实例验证表明,模型聚类分区结果与坝体变形空间分布特征较吻合,具有较高的准确性和稳健性,为从多测点关联维度预测坝体变形和监控大坝整体安全性态提供了一种新方法。

关键词:变形预测;高拱坝;聚类分区;MO-LSSVR

中图分类号:TV698.1 **文献标志码:**A **文章编号:**1006-7647(2023)02-0102-07

Deformation prediction model of a high arch dam based on clustering and MO-LSSVR//LIU Weiqi^{1,2}, CHEN Bo^{1,2}, GE Panmeng^{1,2}, ZHANG Xiaoling³(1. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China; 2. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China; 3. Shandong Huacheng Engineering Consulting and Supervision Co., Ltd., Gaomi 261500, China)

Abstract: To solve the problem that the internal correlation between multiple measuring points cannot be considered by traditional single measuring point monitoring models, which is difficult to reflect the regional characteristics of high arch dam deformation in space. A high arch dam deformation prediction model based on clustering partition and a multi-output least square support vector regression machine (MO-LSSVR) algorithm is proposed. Based on the composite similarity index between the measuring points, the clustering partition of spatial correlation measuring points is realized by hierarchical agglomerative clustering (HAC) algorithm. The MO-LSSVR algorithm integrating the correlation characteristics of measuring points is then used to model the points in the partition. The engineering example results show that the clustering partition results are consistent with the spatial distribution characteristics of dam deformation. The MO-LSSVR model based on the reasonable partition results has high accuracy and robustness, which provides a new method to accurately predict the dam deformation and monitoring the overall safety state of the dam from the multi-measuring points correlation dimension.

Key words: deformation prediction; high arch dam; spatial clustering; MO-LSSVR

目前我国已建有锦屏、二滩和小湾等超高拱坝,其变形在空间上具有明显且复杂的区域性特征,亟须研究适用于高拱坝空间变形特性的监测模型。然而,常见的监控模型通常仅针对单个测点进行建模,忽略了隐含于各测点监测数据中的整体信息,虽可对大坝全部测点进行单测点建模,但由于缺少对各个测点之间关联性的考虑,无法从空间多测点的机理性关联层面准确反映大坝整体安全性态。传统单测点监控模型有统计模型、确定性模型和混合模型。已有学者基于传统模型建立了空间多测点位移场的时空分布模型,如何金平等^[1]推求出多测点水压分量和温度分量的预置因子集,并以此为基础构建了多测点的统计模型;顾冲时等^[2]在传统方法中融入测点空间坐标,建立了大坝变形的三维多测点正、反分析模型。然而,传统多测点模型本身的局限性制约了其推广和应用,如统计模型存在影响因子间的

多重共线性问题,一定程度会影响预测效果,且当监测序列过短时,可能会出现过拟合问题^[3];确定性模型与混合模型对结构计算的精度依赖程度较高。而随着人工智能技术的发展,以人工神经网络为代表的智能算法已被应用到水工安全监控领域并取得了较好的预测效果,如反向传播网络(back propagation, BP)^[4]、长短期记忆网络(long short-term memory, LSTM)^[5-6]、极限学习机(extreme learning machine, ELM)^[7]等。此外,极端梯度提升(extreme gradient boosting, XGBoost)^[8]、随机森林(random forest, RF)^[9]等算法也得到了广泛应用。然而,神经网络等算法普遍存在高计算复杂度、模型参数烦琐等缺陷,影响模型的实际应用和泛化。融合先进智能算法建立更有效的多测点变形预测模型,仍值得进一步研究。支持向量回归机(support vector regression, SVR)克服了神经网络的缺点,具有理论基础坚实、泛化性能优越等优点,已成为最实用和成熟的机器学习方法之一。最小二乘支持向量回归机(least square support vector regression, LSSVR)采用等式约束取代了经典 SVR 中的不等式约束,解决了求解耗时的凸二次规划问题,在大坝变形监控领域亦得以广泛应用^[10-12]。然而,传统的 SVR 仅限于一维输出,对所有测点依次 SVR 单输出建模仍无法解决多输出之间的潜在关联性,难以准确评估大坝整体性状。多输出最小二乘支持向量回归机(multioutput-least square support vector regression, MO-LSSVR)是一种较为新颖的 SVR 改进算法,能够考虑多个输出之间的潜在联系并进行多测点建模,将该算法应用于构建变形监测模型,能有效克服上述现有监控模型的局限性。

高拱坝作为一种大型水工混凝土结构,由于其坝高高超、拱圈跨度大,且根据坝体材料参数分为不同分区,因此不同区域的变形有明显差异,而同一区域多测点的变形具有较强关联性。K-均值、层次凝聚聚类、密度峰值聚类等聚类方法已在水工安全监控领域得到成熟应用,能够从众多测点的海量监测数据中深度挖掘有效信息,对高拱坝的变形区域进行划分,使得同一分区内的测点具有高度关联和相似性,而不同分区间的测点具有明显相异性。在此基础上构建不同分区的多测点模型,能够充分发挥 MO-LSSVR 考虑多测点间潜在关联的优越性,为识别高拱坝变形的区域性特征提供有效帮助。

因此,本文提出一种基于空间聚类分区和 MO-LSSVR 的高拱坝变形预测模型,该模型基于复合相似性指标,利用层次凝聚聚类(hierarchical agglomerative clustering, HAC)算法对高拱坝测点进

行空间维度划分;基于聚类结果,对于各分区内的多个测点,利用能够反映各测点关联性的 MO-LSSVR 进行多输出建模,同时与单测点 LSSVR 和统计模型进行比较,以验证本文方法的准确性与有效性;最后加入干扰测点探究不同建模环境下 MO-LSSVR 的稳健性。

1 基于复合相似性指标与 HAC 算法的空间聚类分区

高拱坝坝体变形在空间上具有明显区域性特征,为定量分析不同测点的相似程度,需构建测点间的复合相似性指标;之后借助聚类算法实现对大坝变形区域的有效划分,进而对大坝每个分区进行多测点建模。

1.1 复合相似性指标及标准化

为充分考虑测值序列的变化趋势,挖掘不同测点之间的相似(异)性特征,从而达到准确、合理的聚类效果,参考已有研究成果^[13],定义 3 种变形测值指标如表 1 所示。

表 1 3 种变形测值指标

指标	表达式	表征意义
变量 x_{it}	$x_{it} = \delta_{it}$	变形的数值大小
变形增幅 y_{it}	$y_{it} = x_{it} - x_{i,t-1} = \delta_{it} - \delta_{i,t-1}$	变形的数值变化幅度
变形增速 z_{it}	$z_{it} = \frac{y_{it}}{x_{i,t-1}} = \frac{\delta_{it} - \delta_{i,t-1}}{\delta_{i,t-1}}$	变形的数值变化速度

注: δ_{it} 和 $\delta_{i,t-1}$ 分别为第 i 个测点在 t 时刻和 $t-1$ 时刻的变形值。

表 1 中 3 种指标性质不同,其量纲和数量级也存有差异,不适合直接用原始值进行计算分析。因此采用 Z-score 标准化方法对原始指标值进行标准化处理。以 x_{it} 为例,若有 n 个测点及 T 个时间点,先计算出 x_{it} 的均值 μ_x 和标准差 σ_x ,再利用下式进行标准化操作:

$$Z(x_{it}) = \frac{x_{it} - \mu_x}{\sigma_x} \tag{1}$$

同理可对其他两种指标进行标准化处理,标准化后的指标均值为 0,标准差为 1。基于上述 3 种标准化后的指标,定义测点 i 和测点 j 之间的 3 种距离如表 2 所示。

表 2 表征相似性的 3 种距离

名称和符号	表达式	表征意义
绝对距离 $d_{ij,AD}$	$d_{ij,AD} = \sum_{t=1}^T [Z(x_{it}) - Z(x_{jt})]^2$	测点 i 和测点 j 测值数值大小的差异
增量距离 $d_{ij,ID}$	$d_{ij,ID} = \sum_{t=1}^T [Z(y_{it}) - Z(y_{jt})]^2$	测点 i 和测点 j 测值变化幅度的差异
相对增速距离 $d_{ij,RGD}$	$d_{ij,RGD} = \sum_{t=1}^T [Z(z_{it}) - Z(z_{jt})]^2$	测点 i 和测点 j 测值变化速度的差异

将表 2 中 3 种距离指标结合可得到描述空间测点 i 和测点 j 之间的复合相似性指标 $d_{ij,CD}$,其定义为

$$d_{ij,CD} = \alpha_1 d_{ij,AD} + \alpha_2 d_{ij,ID} + \alpha_3 d_{ij,RGD} \quad (2)$$

式中 α_1 、 α_2 、 α_3 分别为 3 种距离指标的权重,且 $\alpha_1 + \alpha_2 + \alpha_3 = 1$,本文拟定 $\alpha_1 = 0.4$, $\alpha_2 = \alpha_3 = 0.3$ 。

1.2 HAC 算法

K-均值、最大期望算法等平面聚类方法输出的是平面无结构聚类集合,其缺陷在于计算前需首先定义初始点及聚类数目,且聚类效果对于初始参数的选择较为敏感。考虑到大坝实测数据在时序长度及测点数量上数据量庞大,不能对聚类数量等初始参数进行精准划分,故本文采用 HAC 算法进行空间测点的聚类分区。该算法的聚类策略表现为:先将各个对象分别作为一个典型簇,通过不断重复地合并最近的两个典型样本点/簇进行聚类,最终生成聚类层次树结构,并根据迭代终止条件(如阈值切线)划分层次树,输出聚类结果。

2 基于 MO-LSSVR 的高拱坝变形预测模型构建

2.1 经典单输出 LSSVR

单输出 LSSVR 由 Suykens 等^[14-15]提出,用等式约束代替经典支持向量回归中的不等式约束,并进一步简化约束条件,将原始二次规划问题转换为线性方程组求解问题。假设样本数量为 N ,对于给定训练集 $\{(\mathbf{x}_i, y_i) \mid i=1, 2, \dots, N\}$,构造下列最小化目标函数:

$$\min J(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \boldsymbol{\xi}^T \boldsymbol{\xi} \quad (3)$$

$$\text{s. t.} \quad y_i = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b + \xi_i \quad (4)$$

式中: $\boldsymbol{\varphi}(\mathbf{x}_i)$ 为低维输入空间到高维特征空间的映射函数; $\mathbf{w}$ 、 b 分别为权重向量和偏置; γ 为正则化参数; $\boldsymbol{\xi}$ 为由松弛变量构成的向量。

引入拉格朗日函数,上述问题求解可转化成对以下线性方程组的求解:

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

其中 $\mathbf{e} = (1, 1, \dots, 1)$

式中: $\boldsymbol{\alpha}$ 为拉格朗日乘子向量; $\mathbf{I}$ 为单位矩阵; $\mathbf{K}$ 为核函数,其作用体现在训练数据映射到高维度特征空间后能够代替回归运算,从而将非线性问题转换为线性问题。径向基核函数是最常用且效果较好的核函数之一,其表达式为

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

式中: $\mathbf{x}$ 、 $\mathbf{x}_i$ 为输入值; σ 为核函数作用范围; ρ 为径向基核函数唯一的超参数, $\rho = 1/2\sigma^2$ 。

结合 KKT 约束条件,求得式(5)的解 $\boldsymbol{\alpha}^*$ 和 b^* ,故最小二乘支持向量回归函数 $f(\mathbf{x})$ 表示为

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^* \quad (7)$$

2.2 关联多输出 MO-LSSVR

MO-LSSVR 将输入测点视作整体并考虑多测点之间潜在的关联性,将单输出场景推广到多输出场景。假设共有 M 个输出, N_m 为第 m 个输出的样本数量,给定训练集 $\{(\mathbf{x}_{m,i}, y_{m,i}) \mid i=1, 2, \dots, N_m\}$,其函数可表示为 $f_m(\mathbf{x}) = \mathbf{w}_m^T \boldsymbol{\varphi}(\mathbf{x}) + b_m$,其中 $\mathbf{w}_m$ 和 b_m 分别为模型的第 m 个输出的权重向量和偏置。

为了将测点间的潜在关联性考虑在内,基于正则化多目标学习^[16]的思想,假设权重向量由两部分组成, $\mathbf{w}_m = \mathbf{w}_0 + \mathbf{v}_m$,其中 $\mathbf{w}_0$ 为公共权重向量,而 $\mathbf{v}_m$ 为个体权重向量,即认为各测点权重向量 $\mathbf{w}_m$ 以 $\mathbf{w}_0$ 为基准,以 $\mathbf{v}_m$ 为幅度而发生波动。两测点之间关联性越强,则 $\mathbf{w}_0$ 所占权重越大;关联性越弱,则 $\mathbf{v}_m$ 所占权重越大。构造下列最小化目标函数:

$$\min J(\mathbf{w}_0, \mathbf{v}_m, \boldsymbol{\xi}_m) = \frac{1}{2} \mathbf{w}_0^T \mathbf{w}_0 + \frac{1}{2} \frac{\lambda}{M} \sum_{m=1}^M \mathbf{v}_m^T \mathbf{v}_m + \frac{1}{2} \gamma \sum_{m=1}^M \boldsymbol{\xi}_m^T \boldsymbol{\xi}_m \quad (8)$$

$$\text{s. t.} \quad y_{m,i} = (\mathbf{w}_0 + \mathbf{v}_m)^T \boldsymbol{\varphi}(\mathbf{x}_{m,i}) + b_m + \xi_{m,i} \quad (9)$$

式中: $\boldsymbol{\xi}_m$ 为由松弛变量构成的向量; λ 为超参数。

类似于 LSSVR,通过引入式(8)和式(9)所对应的拉格朗日函数,可转化为对其对应线性方程组的求解。记方程组的解为 $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \dots, \boldsymbol{\alpha}_M^*)$, $\boldsymbol{\alpha}_m = (\alpha_{m,1}^*, \alpha_{m,2}^*, \dots, \alpha_{m,N_m}^*)$ 和 $\mathbf{B}^* = (b_1^*, b_2^*, \dots, b_M^*)$,则对于单个输出 m ,其最小二乘支持向量回归函数为

$$f_m(\mathbf{x}) = \sum_{m=1}^M \sum_{i=1}^{N_m} \alpha_{m,i}^* K(\mathbf{x}_{m,i}, \mathbf{x}) + \frac{M}{\lambda} \sum_{i=1}^{N_m} \alpha_{m,i}^* K(\mathbf{x}_{m,i}, \mathbf{x}) + b_m^* \quad (10)$$

将上述两种 SVR 方法进行对比,结果如表 3 所示。相比于经典 LSSVR,MO-LSSVR 适用于多输出场景,且同时考虑了不同输出间的相关性,只比 LSSVR 多了 1 个超参数 λ 。假设以 n 个测点的变形为输出目标,MO-LSSVR 仅需建模 1 次,需求解的参数共 3 个;而 LSSVR 需要对 n 个测点分别建模,共需求解参数 $2n$ 个,故在模型训练的精简度上 MO-LSSVR 占

表 3 两种 SVR 方法的对比

方法	类型	相关性考量	参数
LSSVR	多输入+单输出	否	γ, ρ
MO-LSSVR	多输入+多输出	是	λ, γ, ρ

优。由此可见,MO-LSSVR 具有 LSSVR 所不具备的优势。

2.3 基于 MO-LSSVR 的高拱坝变形预测模型构建步骤

结合上文基于 HAC 算法的测点空间聚类分区和关联多输出的 MO-LSSVR,可以实现高拱坝变形预测模型的构建,主要步骤如下:

步骤 1 变形实测数据预处理。主要完成各测点原始测值序列中的粗差剔除和缺失值填补工作。

步骤 2 复合相似性指标计算。计算坝体各变形测点之间的复合相似性指标,得到复合距离矩阵。

步骤 3 坝体测点聚类分区。基于各测点的复合距离矩阵,对所有测点进行层次凝聚聚类,得到空间分区结果。

步骤 4 定义输入与输出。模型的输出变量为同一分区 n 个测点的变形测值 $(\delta_1, \delta_2, \cdots, \delta_n)$,输入变量为基于高拱坝统计模型中的各分量因子。

根据传统监控理论,大坝变形 δ 主要由水压分量 δ_H 、温度分量 δ_T 和时效分量 δ_θ 构成^[17],即

$$\delta = \delta_H + \delta_T + \delta_\theta \tag{11}$$

对于本文所研究的高拱坝,其水压分量通常与上游水位的 1、2、3、4 次方,即 H 、 H^2 、 H^3 和 H^4 有关,可表示为

$$\delta_H = \sum_{i=1}^4 [a_{1i}(H^i - H_0^i)] \tag{12}$$

式中: H 、 H_0 分别为 t 时刻的上游水位和坝基高程; a_{1i} 为拟合系数。

温度分量通常采用组合多种谐波的方法进行模拟,可表示为

$$\delta_T = \sum_{p=1}^2 \left[b_{1p} \left(\sin \frac{2\pi pt}{365} - \sin \frac{2\pi pt_0}{365} \right) + b_{2p} \left(\cos \frac{2\pi pt}{365} - \cos \frac{2\pi pt_0}{365} \right) \right] \tag{13}$$

式中: t 、 t_0 分别为监测序列某时刻和基准时刻; b_{1p} 、 b_{2p} 为拟合系数; p 为时间参数, $p=1$ 表示年周期, $p=2$ 表示半年周期。

时效分量一般用多项式和对数函数来描述,可表示为

$$\delta_\theta = c_1(\theta - \theta_0) + c_2(\ln\theta - \ln\theta_0) \tag{14}$$

式中: θ 、 θ_0 分别为 t 时刻和基准时刻的时效因子; c_1 、 c_2 为拟合系数。

综上所述,混凝土高拱坝统计模型表达式可描述为^[18]

$$\delta = a_0 + \sum_{i=1}^4 [a_{1i}(H^i - H_0^i)] + \sum_{p=1}^2 \left[b_{1p} \left(\sin \frac{2\pi pt}{365} - \sin \frac{2\pi pt_0}{365} \right) + \right.$$

$$\left. b_{2p} \left(\cos \frac{2\pi pt}{365} - \cos \frac{2\pi pt_0}{365} \right) \right] + c_1(\theta - \theta_0) + c_2(\ln\theta - \ln\theta_0) \tag{15}$$

式中 a_0 为常系数。

由此确定 MO-LSSVR 的输入变量 x 共有 10 项,可表示为

$$x = (H, H^2, H^3, H^4, T_{\sin 1}, T_{\cos 1}, T_{\sin 2}, T_{\cos 2}, \Theta_1, \Theta_2) \tag{16}$$

式中: $T_{\sin 1}$ 、 $T_{\sin 2}$ 分别为年周期和半年周期正弦谐波因子; $T_{\cos 1}$ 、 $T_{\cos 2}$ 分别为年周期和半年周期余弦谐波因子; Θ_1 、 Θ_2 分别为多项式时效因子和对数时效因子。

基于式(15)可确定各变量的计算公式如下:

$$\begin{cases} T_{\sin 1} = \sin \frac{2\pi t}{365} - \sin \frac{2\pi t_0}{365} \\ T_{\sin 2} = \sin \frac{4\pi t}{365} - \sin \frac{4\pi t_0}{365} \\ T_{\cos 1} = \cos \frac{2\pi t}{365} - \cos \frac{2\pi t_0}{365} \\ T_{\cos 2} = \cos \frac{4\pi t}{365} - \cos \frac{4\pi t_0}{365} \\ \Theta_1 = \theta - \theta_0 \\ \Theta_2 = \ln\theta - \ln\theta_0 \end{cases} \tag{17}$$

步骤 5 参数寻优与模型训练。将各测点测值序列前 80% 作为训练集进行训练。MO-LSSVR 的性能较很大程度上取决于模型参数的取值,本文采用支持向量机类算法最为常用的十倍交叉验证结合网格搜索算法^[19],按照最小均方误差原则确定模型最优参数 λ^* 、 γ^* 和 ρ^* 。由于网格搜索本质上为穷举算法,故能保证参数寻优的精度要求,避免计算结果出现随机性。可采用相同方法确定 LSSVR 的最优参数 γ^* 和 ρ^* 。

步骤 6 模型预测与效果评估。以各测点测值序列后 20% 作为验证集进行预测,并将预测结果与各测点实际变形进行比较。为进一步量化两种模型的建模效果,考虑相关系数、均方根误差 RMSE 及平均绝对误差 MAE 这 3 项指标以评估模型精度。

3 模型验证

3.1 工程概况

锦屏一级水电站位于四川省雅砻江干流,工程属大(1)型一等工程,大坝为混凝土双曲拱坝。坝顶高程 1 885.00 m,建基面高程 1 580.00 m,最大坝高 305.00 m,为当今世界最高坝。坝体共布置 29 个正垂线测点进行坝体变形监控,测点分布较为均匀,可较为全面地反映大坝整体变形。

选取大坝自 2016 年 7 月 23 日至 2021 年 3 月 19

日的正垂线监测数据和环境量数据进行分析,并完成原始数据的预处理。由于各变形测点的原始时间序列都较为完整、平滑,不存在测值残缺、错误异常等数据污染问题,因此只对极个别粗差进行人为剔除,采用相邻测值对缺失值进行线性差值填补。将2016年7月23日至2020年3月18日的数据作为训练集,2020年3月19日至2021年3月19日的数据为测试集,其比例约为4:1。

3.2 基于 HAC 算法的聚类分区

计算坝体全部测点之间表征相似性的3种距离,并基于式(2)求得复合相似性指标,得到复合距离矩阵。根据复合距离矩阵,以平均链距离为聚类划分的判定标准,对坝体测点进行 HAC 聚类并将其划分为3个区域,如图1所示。其中,Ⅰ号分区测点在大坝靠近两岸处呈对称分布,Ⅱ号分区测点集中在拱坝的中部中、低高程部位,Ⅲ号分区测点分布在高高程部位。综合来看,聚类分区结果与锦屏大坝依据坝体材料形成的分区具有一定关联,且与大坝径向位移分布特征较为契合,证明分区结果较为合理。聚类之后,同一分区的测点变形变化规律具有较高相似性,能够综合反映坝体变形在空间维度上的区域分布特征。

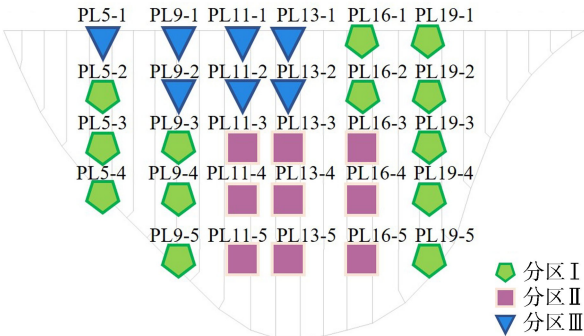


图1 坝体测点 HAC 聚类分区结果

3.3 模型的预测结果评估

以坝体中部Ⅱ号分区为例,给出 MO-LSSVR 模型的构建过程及预测结果。首先利用十倍交叉验证和网格搜索算法选参,参数 λ 、 γ 和 p 分别从 $\{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}$ 、 $\{2^{-10}, 2^{-8}, \dots, 2^8, 2^{10}\}$ 和 $\{2^{-5}, 2^{-3},$

$\dots, 2^8, 2^{10}\}$ 中进行取值,按最小均方差原则确定模型3个最优超参数分别为 $\lambda^* = 2^9$ 、 $\gamma^* = 2^0$ 、 $\rho^* = 2^{-1}$ 。基于最优超参数构建预测模型,对测试数据集进行预测并验证建模效果。为突出 MO-LSSVR 模型的优越性,对分区内各测点分别构建 LSSVR 模型和统计模型两种单输出模型作为对比。

针对Ⅱ号分区的9个测点,3种模型的预测效果及精度评估如图2(为节约篇幅,仅列出代表测点预测结果)和表4所示。可以看出,3种模型中,传统模型的预测值与实际变形的偏差最大;相比之下,两种基于 SVR 的模型能更为准确地反映大坝变形及其影响因素之间的非线性关系。传统单输出 LSSVR 模型的精度略高于统计模型,而本文提出的

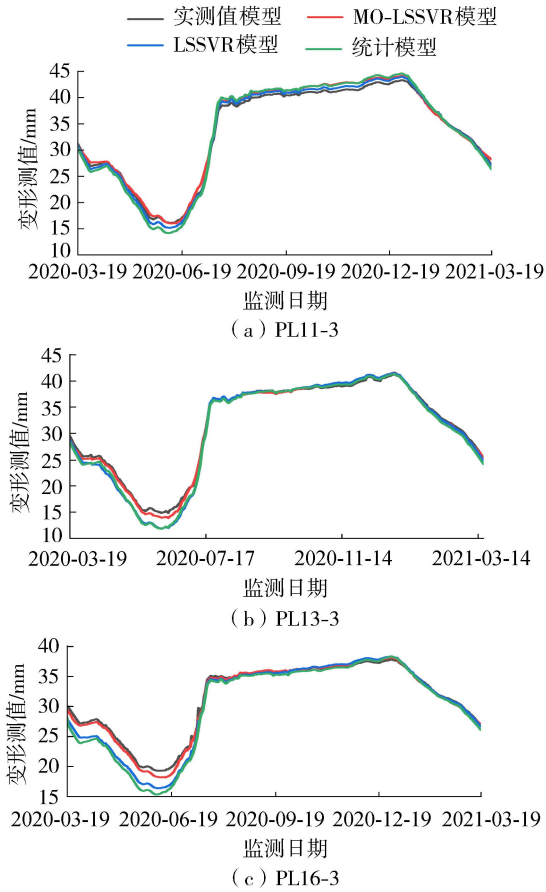


图2 3种模型预测过程线对比

表4 3种模型预测精度评估

测点	RMSE			MAE			相关系数		
	统计模型	LSSVR 模型	MO-LSSVR 模型	统计模型	LSSVR 模型	MO-LSSVR 模型	统计模型	LSSVR 模型	MO-LSSVR 模型
PL16-3	2.196	1.975	1.303	1.872	1.599	1.051	0.9249	0.9327	0.9973
PL16-4	2.189	1.744	0.990	1.944	1.395	0.758	0.9203	0.9395	0.9989
PL16-5	2.631	2.257	1.592	2.433	1.842	1.335	0.8997	0.9238	0.9924
PL13-3	1.947	1.448	1.146	1.762	1.287	1.034	0.9219	0.9374	0.9956
PL13-4	2.028	1.692	1.047	1.769	1.358	0.837	0.9322	0.9396	0.9946
PL13-5	2.469	2.042	1.120	2.321	1.607	0.934	0.9017	0.9367	0.9919
PL11-3	2.211	2.176	1.399	1.993	1.924	1.152	0.9089	0.9134	0.9952
PL11-4	1.821	1.791	1.039	1.328	1.414	0.797	0.9249	0.9521	0.9982
PL11-5	1.599	1.329	0.428	0.903	0.875	0.347	0.9259	0.9388	0.9976

MO-LSSVR 模型得到的变形预测值最接近实测值,各测点相关系数均超过 0.99,且 MAE、RMSE 指标最小,预测性能明显优于其他两种单输出模型,由此验证了本文方法的有效性。

3.4 干扰测点对 MO-LSSVR 模型精度的影响

为探究引入干扰项对 MO-LSSVR 模型精度的影响,仍选择Ⅱ号分区的 9 个测点作为建模对象,考虑加入与该分区相似性最差的测点进行建模并评估预测结果。经过比对,选取与Ⅱ号分区各测点复合距离最远的 PL9-1 作为干扰测点。从空间分布看,Ⅱ号分区位于大坝中部的中、低高程部位,而 PL9-1 测点位于大坝靠近左岸坝肩的高高程部位,两者所在的坝体部位属不同材料分区,在荷载下的变形响应特征具有明显差异。观察上述 10 个测点测值变化曲线,同样可以发现 PL9-1 与其他测点的测值变化趋势有明显区别。

因此,建模测点集合选取具有强空间关联的Ⅱ号分区 9 个测点及与其复合相似性最差的 PL9-1 测点,以此构建 MO-LSSVR 模型,模型超参数寻优结果为 $\lambda^* = 2^{-3}$ 、 $\gamma^* = 2^{-10}$ 、 $\rho^* = 2^{-1}$ 。同时为形成对比,对 PL9-1 测点所属Ⅲ号分区进行 MO-LSSVR 建模,模型预测结果如图 3 和表 5 所示。由表 5 可知,加入 PL9-1 干扰测点,对原分区内 9 个测点的建模精度基本无影响,但干扰测点自身的预测效果相对较

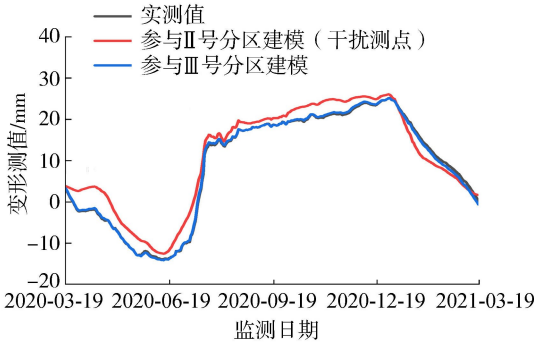


图 3 PL9-1 测点在两种建模环境下的模型预测结果

表 5 PL9-1 与Ⅱ号分区测点 MO-LSSVR 模型预测精度评估

测点	建模情况	RMSE	MAE	相关系数
PL16-3	参与Ⅱ号分区建模	1.4723	1.2250	0.9944
PL16-4	参与Ⅱ号分区建模	1.7759	1.5112	0.9919
PL16-5	参与Ⅱ号分区建模	1.3177	1.2246	0.9905
PL13-3	参与Ⅱ号分区建模	1.0876	0.8365	0.9956
PL13-4	参与Ⅱ号分区建模	1.0784	0.8139	0.9946
PL13-5	参与Ⅱ号分区建模	1.1894	0.9789	0.9866
PL11-3	参与Ⅱ号分区建模	1.6686	1.4812	0.9944
PL11-4	参与Ⅱ号分区建模	1.2868	1.1133	0.9952
PL11-5	参与Ⅱ号分区建模	0.3980	0.3137	0.9973
PL9-1	参与Ⅱ号分区建模	3.1377	2.7472	0.9149
PL9-1	参与Ⅲ号分区建模	1.5980	1.2311	0.9892

差,其原因可能是Ⅱ号分区内的测点具有更强关联性,而 PL9-1 与其他测点的复合距离最远,它们之间的关联性也稍弱,一定程度上影响了本身的建模精度。然而,当 PL9-1 参与所在Ⅲ号分区的多测点建模时,由图 3 可以看出其预测曲线比作为Ⅱ区干扰测点时的预测曲线更逼近实测值,能够达到 MO-LSSVR 应有的预测精度。因此验证 MO-LSSVR 能够充分挖掘不同测点之间的关联特性,并具有较强的稳健性和抗干扰能力;同时进一步说明了基于复合距离指标进行聚类分区的合理性,通过聚类分区将相似性最大的测点归为一类,保证了分区内各测点在空间变形特征上具有更高相似性和关联性,之后再对其进行 MO-LSSVR 多输出建模,从而有助于保证预测精度。

4 结 论

- a. 基于复合相似性指标和 HAC 算法得到的聚类分区结果符合坝体一般变形规律,能够反映坝体变形的空间分布特征。
- b. 基于聚类分区结果构建的 MO-LSSVR 模型能够挖掘分区多测点的内在关联性,比传统统计模型、LSSVR 模型具有更高精度和合理性,能够准确反映大坝整体安全性能。
- c. 通过加入干扰测点,验证了 MO-LSSVR 模型具有较强稳健性和抗干扰能力的结论;同时说明建模环境对预测效果有一定影响,聚类分区有助于提升多测点建模精度。
- d. 在各算法中,MO-LSSVR 的预测精度和泛化性能在一定程度上受超参数取值的影响,本文选择了最常用的交叉验证和网格搜索算法确定其最优参数。然而,采用 MO-LSSVR 对多测点进行建模需要大量的训练样本,一定程度上增加了参数寻优的计算量。因此,兼具效率和准确性的智能算法应在下一步的工作中进行深入研究。

参考文献:

[1] 何金平,李珍照. 大坝结构性态多测点数学模型研究 [J]. 武汉水利电力大学学报,1994,27(2):134-136. (HE Jinping, LI Zhenzhao. Research on the mathematical model of multiple survey points for dam structure behaviour [J]. Engineering Journal of Wuhan University, 1994,27(2):134-136. (in Chinese))

[2] 顾冲时,吴中如,蔡新. 探讨混凝土坝空间位移场的正反分析模型 [J]. 工程力学,1997,14(1):138-144. (GU Chongshi, WU Zhongru, CAI Xin. Analysis and back analysis model of 3d displacement field of concrete dams [J]. Engineering Mechanics, 1997,14(1):138-144. (in

- Chinese))
- [3] SU Huaizhi, HU Jiang, WU Zhongru. A study of safety evaluation and early-warning method for dam global behavior[J]. Structural Health Monitoring, 2012, 11(3): 269-279.
 - [4] 闫滨,高真伟,强丽峰. 基于 L-M 算法的 BP 神经网络在大坝安全监控预报中的应用[J]. 沈阳农业大学学报, 2009, 40(4): 506-509. (YAN Bin, GAO Zhenwei, QIANG Lifeng. Application of BP neural network based on levenberg-marquardt algorithm in prediction of dam safety monitoring[J]. Journal of Shenyang Agricultural University, 2009, 40(4): 506-509. (in Chinese))
 - [5] 欧斌,吴邦彬,袁杰,等. 基于 LSTM 的混凝土坝变形预测模型[J]. 水利水电科技进展, 2022, 42(1): 21-26. (OU Bin, WU Bangbin, YUAN Jie, et al. LSTM-based deformation prediction model of concrete dams [J]. Advances in Science and Technology of Water Resources, 2022, 42(1): 21-26. (in Chinese))
 - [6] 侯回位,郑东健,刘永涛,等. 基于 EEMD-SE-LSTM 的混凝土坝变形监测模型[J]. 水利水电科技进展, 2022, 42(1): 61-66. (HOU Huiwei, ZHENG Dongjian, LIU Yongtao, et al. Deformation monitoring model of concrete dams based on EEMD-SE-LSTM[J]. Advances in Science and Technology of Water Resources, 2022, 42(1): 61-66. (in Chinese))
 - [7] 胡德秀,屈旭东,杨杰,等. 基于 M-ELM 的大坝变形安全监控模型[J]. 水利水电科技进展, 2019, 39(3): 75-80. (HU Dexiu, QU Xudong, YANG Jie, et al. A safety monitoring model of dam deformation based on M-ELM [J]. Advances in Science and Technology of Water Resources, 2019, 39(3): 75-80. (in Chinese))
 - [8] 徐钊,苏怀智,杨立夫. 基于 GP-XGBoost 的大坝变形预测模型[J]. 水利水电科技进展, 2021, 41(5): 41-46. (XU Ren, SU Huaizhi, YANG Lifu. Dam deformation prediction model based on GP-XGBoost[J]. Advances in Science and Technology of Water Resources, 2021, 41(5): 41-46. (in Chinese))
 - [9] 田菊飞,苏怀智. 基于随机森林算法的大坝应力预测模型的构建及其应用[J]. 水电能源科学, 2018, 36(5): 54-56. (TIAN Jufei, SU Huaizhi. Development and application of dam stress prediction model based on random forest algorithm[J]. Water Resources and Power, 2018, 36(5): 54-56. (in Chinese))
 - [10] 林楠,李伟东,张文春,等. 最小二乘支持向量机在深基坑变形预测中的应用[J]. 辽宁工程技术大学学报(自然科学版), 2014, 33(11): 1471-1474. (LIN Nan, LI Weidong, ZHANG Wenchun, et al. Deformation prediction of deep foundation pit with least square support vector machine [J]. Journal of Liaoning Technical University (Natural Science), 2014, 33(11): 1471-1474. (in Chinese))
 - [11] 宋志宇,李俊杰. 最小二乘支持向量机在大坝变形预测中的应用[J]. 水电能源科学, 2006, 24(6): 49-52. (SONG Zhiyu, LI Junjie. Research on dam displacement forecasting model based on least squares support vector machine[J]. Water Resources and Power, 2006, 24(6): 49-52. (in Chinese))
 - [12] 冯腾飞,钟钰,刘小生,等. 基于自适应人工蜂群算法优化的最小二乘支持向量机在变形预测中的应用[J]. 江西理工大学学报, 2018, 39(3): 35-39. (FENG Tengfei, ZHONG Yu, LIU Xiaosheng, et al. Application in deformation prediction based on LSSVR model optimized by AABC[J]. Journal of Jiangxi University of Science and Technology, 2018, 39(3): 35-39. (in Chinese))
 - [13] CHEN Bo, HU Tianyi, HUANG Zishen, et al. A spatio-temporal clustering and diagnosis method for concrete arch dams using deformation monitoring data [J]. Structural Health Monitoring, 2019, 18(5/6): 1355-1371.
 - [14] SUYKENS J. Least squares support vector machines[J]. International Journal of Circuit Theory & Applications, 2002, 27(6): 605-615.
 - [15] EVGENIOU T, PONTIL M. Regularized multi-task learning [C]//Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, [s. n.], 2004.
 - [16] 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.
 - [17] 王继敏,顾冲时,张晨,等. 基于面板时空模型的锦屏一级大坝变形性态分析[J]. 水力发电学报, 2020, 39(11): 21-30. (WANG Jimin, GU Chongshi, ZHANG Chen, LIU Jian. Deformation behavior analysis of Jinping arch dam based on spatiotemporal model of variable intercept panel data [J]. Journal of Hydroelectric Engineering, 2020, 39(11): 21-30. (in Chinese))
 - [18] 袁玉萍,胡亮,周志坚. 基于遗传算法对支持向量机模型中参数优化[J]. 计算机工程与设计, 2008, 29(19): 5016-5018. (YUAN Yuping, HU Liang, ZHOU Zhijian. Optimizing parameters of support vector machine's model based on genetic algorithm[J]. Computer Engineering and Design, 2008, 29(19): 5016-5018. (in Chinese))
 - [19] 徐硕,乔晓东,朱礼军,等. 多任务最小二乘支持向量回归机及其在近红外光谱分析技术中的应用研究[J]. 光谱学与光谱分析, 2011, 31(5): 1208-1211. (XU Shuo, QIAO Xiaodong, ZHU Lijun, et al. Multi-task least-squares support vector regression machines and their applications in nir spectral analysis [J]. Spectroscopy and Spectral Analysis, 2011, 31(5): 1208-1211. (in Chinese))

(收稿日期:2022-05-10 编辑:骆超)