DOI:10.3880/j.issn.1006-7647.2023.04.014

基于 KPCA-ICFSFDP-MOGP 的拱坝多测点变形数据预处理及预测方法

葛盼猛[1],陈 波[1,2],刘庭赫[3],杨 帆[4]

(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.中水东北勘测设计研究有限责任公司,吉林 长春 130021;4.中国建筑金属结构协会,北京 100037)

摘要:为更高效地对高拱坝的多测点变形数据进行分析,引入了核主成分分析(KPCA)算法重构误差以识别多测点序列中的异常值;为解决用于测点聚类的密度峰值聚类(CFSFDP)算法高维表现较差的问题,利用 KPCA 算法对其进行降维操作,并提出了自动选取聚类中心与截断距离的改进 CFSFDP(ICFSFDP)算法;基于 KPCA-ICFSFDP 和多输出高斯过程(MOGP)算法,按分区建立了多测点变形 MOGP 回归模型。实例验证结果表明,对于簇内点数量较少的类,相较于不分区的 MOGP 模型,预测效果得到了一定的提升,同时在整体 MOGP 模型表现良好的测点,分区后仍然保持较高的预测精度,且与单输出高斯过程模型对比均有所提升。

关键词:高拱坝;多测点变形;KPCA;CFSFDP;MOGP

中图分类号:TV698.1 **文献标志码**:A **文章编号**:1006-7647(2023)04-0092-06

Preprocessing and prediction method for multi-point deformation data of arch dams based on KPCA-ICFSFDP-MOGP//GE Panmeng[1], CHEN Bo[1,2], LIU Tinghe[3], YANG Fan[4] (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. *China Water Northeastern Investigation, Design & Research Co., Ltd., Changchun* 130021, *China*; 4. *China Construction Metal Structure Association, Beijing* 100037, *China*)

Abstract: In order to analyze the multi-point data of high arch dams more efficiently, the kernel principal component analysis (KPCA) method for error reconstruction was introduced to identify the outliers in the multi-point sequence. To solve the problem of poor performance of the clustering by fast search and find of density peaks (CFSFDP) algorithm in high dimension, the KPCA method was used to reduce the dimension, and an improved CFSFDP (ICFSFDP) algorithm was proposed for automatic selection of cluster center and truncation distance. Based on KPCA-ICFSFDP method and the multiple output Gaussian process (MOGP) method, a multi-point deformation MOGP model was built according to the partition. The results show that for the small number of partitions within the cluster points, compared with the MOGP model without partition, the prediction effect is improved. Meanwhile, for the measurement points with good performance of the overall MOGP model, the prediction accuracy remains high after partitioning, which is improved compared with the single-output Gaussian process model.

Key words: high arch dam; multi-point deformation; KPCA; CFSFDP; MOGP

随着我国水电资源的不断开发,大批高拱坝陆续建成。对特高拱坝工程安全问题的监测与预防是一项长期且艰辛的工作[1]。拱坝作为整体结构,各测点变形之间存在着关联性,对拱坝的变形研究逐渐从单一测点研究转向了多测点研究[2-3]。而目前常用的面板模型仅能够挖掘环境量与响应量之间以及不同响应量之间的线性关系,缺乏对非线性关系的挖掘能力。随着监测系统的逐渐完善,积累了大量大坝变形监测序列数据,其中不乏由于设备和环境等因素导致的测量异常值。因此急需一种准确的

基金项目:国家自然科学基金面上项目(52079049);江苏省基础研究计划青年项目(BK20160872)

作者简介:葛盼猛(1997—),男,硕士研究生,主要从事水工结构安全监控研究。E-mail:gepanmeng@hhu.edu.cn

通信作者:陈波(1986—),男,教授,博士,主要从事水工结构安全监控研究。E-mail:chenbo@hhu.edu.cn

异常值识别方法对多测点变形序列进行分析。此外,特高拱坝不同区域变形性态的影响因素差异较大,但传统方法未考虑这些差异性的影响,难以刻画特高拱坝不同区域的实际变形规律,因此对于特高拱坝变形测点的分区研究逐渐得到重视[4]。

目前常用的聚类方法,如模糊灰色聚类[5]、K 均值聚类[6]等算法存在着聚类结果倾向于输出球型簇的问题,而密度峰值聚类(clustering by fast search and find of density peaks,CFSFDP)算法依靠对样本点密度和相对距离的定义解决了这个问题,在各种形状的数据集下都有良好的表现,且对于噪声点具有优越的识别能力,在工程数据挖掘领域得到了较为广泛的应用[6-8]。但 CFSFDP 算法自身存在诸如聚类中心确定依赖人工手动调整、超参数截断距离对聚类结果影响巨大等问题,因此很多学者对 CFSFDP 算法进行了研究与改进,如郑虹等[9]利用布谷鸟算法对 CFSFDP 算法的截断距离进行参数寻优。

本文采用聚类中心自动选取和基于轮廓系数确定截断距离的方法对 CFSFDP 算法进行了改进,提出了改进 CFSFDP(ICFSFDP)算法,并将其与核主成分分析(kernel principal component analysis,KPCA)算法结合用于大坝多测点数据的异常值识别,再按各测点序列聚类结果利用多输出高斯过程(multiple output Gaussian process,MOGP)对测点序列构建多测点变形 MOGP 回归模型,以期反映不同测点的变化趋势,提高其预测精度。

1 研究方法

1.1 基于 KPCA 算法的大坝多测点变形数据降维方法

CFSFDP 算法在高维数据中的表现常常较差,为了更好地挖掘环境量与变形之间的关系,故在对大坝多测点变形数据进行聚类之前,需对变形与环境量分别进行降维。依据大坝变形监测数据的环境量构建大坝变形模型的环境因子矩阵 $\boldsymbol{X}_t=(x_t)$,包含了水位(H_t^1、H_t^2、H_t^3)、温度(T_0、T_{1-7}、T_{8-15}、T_{16-30}、T_{31-60}、T_{61-120})和时效(θ、$\ln\theta$)3 个方面的影响因素,以及大坝多测点变形的面板数据 $\boldsymbol{Y}$。

KPCA 降维思想是将 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ 映射到高维特征空间之中。核矩阵 $\boldsymbol{K}$ 为

$$\boldsymbol{K}=(k_{ij})_{n\times n} \tag{1}$$

其中 $$k_{ij}=\exp\left(-\frac{\|y_i-y_j\|_2^2}{2\sigma^2}\right)$$

式中:σ 为用户定义参数;y_i 和 y_j 是多测点数据序列;n 为测点数据维度。

中心化核矩阵 $\tilde{\boldsymbol{K}}$ 为

$$\tilde{\boldsymbol{K}}=(\tilde{k}_{ij})_{n\times n} \tag{2}$$

其中 $$\tilde{k}_{ij}=k_{ij}-\frac{1}{n}\sum_{r=1}^{n}k_{ir}-\frac{1}{n}\sum_{r=1}^{n}k_{rj}+-\frac{1}{n^2}\sum_{r,s=1}^{n}k_{rs}$$

再求解中心化核矩阵的特征值并按顺序排列为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$,其对应的特征向量为 $\boldsymbol{V}=(V_1,V_2,\cdots,V_n)$,按设定的降维后维度 m,选取 m 个最大的特征值及对应的特征向量 $\boldsymbol{V}'_m=(V_1,V_2,\cdots,V_m)$,利用中心化核矩阵与 $\boldsymbol{V}'_m$ 相乘,即得到降维后的数据矩阵 $\boldsymbol{Y}_{\mathrm{new}}$:

$$\boldsymbol{Y}_{\mathrm{new}}=\tilde{\boldsymbol{K}}\boldsymbol{V}'_m=(\tilde{\phi}_1,\tilde{\phi}_2,\cdots,\tilde{\phi}_m)^{\mathrm{T}} \tag{3}$$

式中 $\tilde{\phi}$ 为非线性映射函数。

由于 KPCA 降维过程中对于噪声等数据问题十分敏感,而大坝自动化监测系统受到设备自身运行情况影响常常存在着一些不合理的离群值,这些值的存在将大大影响 KPCA 降维的精度,因此在 KPCA 降维的过程中,引入了重构误差的概念。令 f_{lz} 表示数据点 z 在第 l 个投影向量上的投影:

$$f_{lz}=\tilde{\phi}_z V_l \tag{4}$$

而重构误差为

$$E_{\mathrm{R}z}=\tilde{k}_{zz}-\sum_{l=1}^{q}f_{lz}^2 \tag{5}$$

当重构误差大于阈值时,数据点将会被认定为异常值进行排除,而阈值需要用户进行设置。

1.2 ICFSFDP 算法

1.2.1 传统 CFSFDP 算法

CFSFDP 算法因其能够输出非球型簇类的优点得到了广泛应用。其原理主要依据两个直观的假设:①聚类中心被一群密度较低的邻居点包围着;②聚类中心离比它密度更高的点的距离相对较大。因此该算法引入了局部密度 ρ_i 与样本间相对距离 d_{ij} 的概念[10]。

该算法所需用户设置的参数仅有截断距离 d_c,文献[10]中 d_c 选取为所有样本点的 d_i 由大到小排序的第 2%,但该值对该算法的聚类效果影响十分巨大,因此有众多学者对 d_c 的取值进行了讨论[11],本文则选择通过优化寻优算法的方式来确定 d_c 的取值。

1.2.2 CFSFDP 算法改进策略

a. 聚类中心质量排序方法。传统 CFSFDP 算法需要依靠人工对决策图进行判断进而决定聚类中心,无法做到自动选择聚类中心,且缺乏准确的定量方法。在决策图的基础上进行聚类中心选取时,存在主观因素,会对聚类结果造成重大影响,且决策图受到截断距离 d_c 的影响,因此提出了聚类中心的自动选择方法,即制定同时考虑样本间相对距离与局部密度的指标 γ,通过对 γ 进行由高至低的排序进而选取用户制定数量的聚类中心。为消除不同量纲

影响,对样本间相对距离和局部密度进行规范化处理,其数学原理如下:

$$\gamma_i = \delta_i' \rho_i' = \frac{(\delta_i - \delta_{\min})(\rho_i - \rho_{\min})}{(\delta_{\max} - \delta_i)(\rho_{\max} - \rho_i)} \tag{6}$$

式中:δ_i'为规范化后的样本间相对距离;ρ_i'为规范化后的局部密度;$\delta_{\max}$、$\delta_{\min}$分别为样本集中相对距离的最大值与最小值;$\rho_{\max}$、$\rho_{\min}$分别为样本集中局部密度的最大值与最小值。

b. 基于轮廓系数鲸鱼优化算法(whale optimization algorithm, WOA)的超参数寻优算法。CFSFDP 算法的 d_c 和聚类中心数需要用户自行设置,这些参数大大影响了其聚类效果,对于用户的先验知识要求较高。为降低算法的使用难度,引入 WOA 对其进行参数的自动寻优,寻优过程中以聚类结果的轮廓系数 S 为适应度函数值,其公式为

$$S = \frac{b - a}{\max(a, b)} \tag{7}$$

式中:a 为样本与同一类别中所有其他点之间的平均距离;b 为样本与下一个距离最近的簇中的所有其他点之间的平均距离。

1.3 多测点变形 MOGP 回归模型构建

作为整体结构,大坝各测点的变形在不同程度上相互影响,MOGP 模型针对多点建模,能够充分挖掘各测点间的相互关系。通过之前预处理与聚类分析,将拥有充分相关信息的测点归为一类,为 MOGP 模型建模做好数据准备。

MOGP 模型的基本原理可参考文献[12]。多测点变形 MOGP 回归模型算法流程如图 1 所示。

2 实例验证

某特高拱坝水库正常蓄水位 1 880 m,死水位 1 800 m,总库容 77.6 亿 m^3,最大坝高 305m,拱冠梁顶厚 16 m,拱冠梁底厚 63 m,最大中心角 93.12°,顶拱中心线弧长 552.23 m,厚高比 0.207,弧高比 1.811。

该水库现有正倒垂自动化测值数量总计 53 个测点,采用从 2015 年 8 月 20 日至 2021 年 3 月19 日起共 1 693 期测值,89 729 个数据。自动化测值常常会出现一些不合理的离群测值。

对于测值序列中存在的数值上相差较多的离群点,利用 KPCA 算法对其进行剔除。先选定 2015 年 8 月 20 日至 2017 年 8 月 20 日共 522 期数据作为 KPCA 异常点识别的训练集,利用人工方法对这 522 期数据的异常值进行识别,以供 WOA 算法在 KPCA 算法的超参数寻优过程中计算精准率,将认定的含有离群值的数据点标定为簇 1。

利用 WOA 算法对 KPCA 算法去除异常点所需的超参数进行寻优计算,WOA 种群数量设置为 50,最大迭代次数设置为 100,由于异常值识别属于二分类问题,因此采用 F_1 分数作为适应度函数值(式(8))。在二分类问题中的点可以分为 TP(真实值 1,预测值 1)、FP(真实值 0,预测值 1)、FN(真实值 1,预测值 0)、TN(真实值 0,预测值 0)4 种情况。参数寻优范围及结果如表 1 所示。

$$F_1 = 2\frac{pr}{p + r} \tag{8}$$

其中 $$p = \frac{n_{TP}}{n_{TP} + n_{FP}} \qquad r = \frac{n_{TP}}{n_{TP} + n_{FN}}$$

式中:p 为精准率;r 为召回率;n_{TP}、n_{FP}、n_{FN}分别为 TP、FP、FN 情况的点的数量。

表 1 WOA-KPCA 算法参数寻优范围及结果

参数	范围	结果
σ	[0.001,10]	9.516
q	[1,15]	5
重构误差阈值	[0.001,2]	0.628

在训练集中 WOA-KPCA 算法异常值识别结果

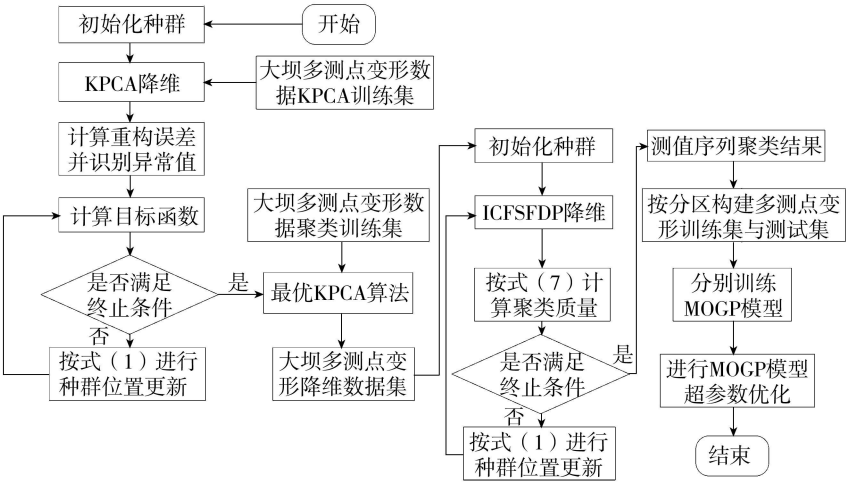


图 1 多测点变形 MOGP 回归模型算法流程

F_1 分数能够达到 0.936，能够准确地识别人为标记出的异常值，具体训练结果情况为 TP 类样本 204 个、TN 类 290 个、FP 类 26 个和 FN 类 2 个。

召回率达到了 99% 以上，精准率相对较低，为 88% 左右，可见 WOA-KPCA 算法能识别出绝大多数人工识别的离群点。由于人工与 WOA-KPCA 算法仅是依据变形序列本身对数值异常的点进行剔除，无法识别与挖掘出大坝运行过程中出现的状态异常，因此，将 2015 年 8 月 20 日至 2021 年 3 月 19 日共 1693 期数据输入训练好的 WOA-KPCA 算法初步剔除异常值，经由 KPCA 算法剔除异常值点后各测点序列的数值与平均周期变幅分布如表 2 所示。

表 2　KPCA 算法离群点剔除结果

测点	原始数据	KPCA 降噪后	平均周期变幅	测点	原始数据	KPCA 降噪后	平均周期变幅
PL1-1	0.30	0.23	0.91	PL13-3	6.22	2.68	8.43
PL1-2	7.73	0.58	0.64	PL13-4	6.39	3.58	11.23
IP1-1	13.65	0.58	1.23	PL13-5	4.78	3.13	8.12
PLR-1	3.24	2.89	0.51	IP13-2	6.03	6.04	3.31
IPR-3	10.31	0.34	0.31	IP13-1	1.32	1.14	2.64
IPR-2	5.66	1.36	0.24	PL11-1	1.81	1.71	5.71
PLR-2	0.22	0.23	0.55	PL11-2	9.66	2.83	8.61
PLR-3	9.44	0.22	0.27	PL11-3	3.03	3.03	10.15
IPR-1	0.40	0.07	0.21	PL11-4	8.50	3.75	12.13
PL23-1	9.19	1.83	1.01	PL11-5	7.62	2.49	7.48
PL23-2	0.89	0.74	0.51	IP11-1	2.03	1.32	2.32
PL23-3	2.16	0.52	1.39	PL9-1	2.68	2.35	6.85
IP23-1	0.21	0.15	0.29	PL9-2	4.15	3.43	6.48
PL19-1	13.78	0.82	2.38	PL9-3	6.32	3.33	10.18
PL19-2	1.36	1.33	3.93	PL9-4	3.39	3.44	10.64
PL19-3	1.23	1.18	3.89	PL9-5	9.27	1.10	2.88
PL19-4	1.40	1.42	4.31	IP9-1	5.76	1.15	0.61
PL19-5	9.21	0.69	2.12	PL5-1	2.04	1.98	6.01
IP19-1	6.97	0.74	0.53	PL5-2	4.24	2.95	8.32
PL16-1	1.18	1.05	1.83	PL5-3	2.56	2.59	7.62
PL16-2	5.88	3.45	11.38	PL5-4	1.43	1.35	3.96
PL16-3	1.77	1.75	5.53	IP5-1	5.63	0.85	0.58
PL16-4	8.80	2.62	7.63	PLL-1	7.52	0.70	0.64
PL16-5	2.09	2.07	6.51	PLL-2	0.65	0.34	0.68
IP16-1	2.18	0.86	2.31	PLL-3	0.77	0.28	0.52
PL13-1	1.52	1.48	4.19	IPL-1	5.09	0.48	0.22
PL13-2	5.44	5.13	14.70				

由表 2 可见，经由 WOA-KPCA 算法剔除离群点后，各测点数据序列的标准差均有所下降，说明序列中的异常值基本已被识别与剔除。

经由 KPCA 算法进行异常值识别后，在聚类过程中时间序列每一期数据都被视为样本的一个属性，如果直接进行聚类则属于高维问题，而维度过高会影响聚类精度，因此为了提升聚类精度，利用 KPCA 算法对变形序列沿时间维度进行降维操作，并对降维后数据利用 ICFSFDP 算法进行聚类分析，为满足后续高斯过程假设。将各序列数据进行标准化操作：

$$x^* = \frac{x - \mu}{\sigma} \tag{9}$$

式中：x^* 为标准化处理后的样本数据；x 为原序列样本数据；μ 为样本数据均值；σ 为样本的标准差。

本次寻优过程中，ICFSFDP 算法的超参数寻优范围设置见表 3，适应度函数收敛过程如图 2 所示。

表 3　ICFSFDP 算法超参数寻优范围及结果

超参数	范围	结果
σ	[0.001,15]	6.767
q	[5,30]	7
n	[2,8]	4
d_c	[0.1,2]	0.844

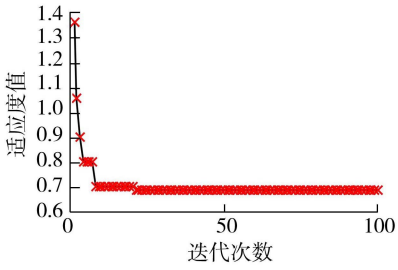


图 2　适应度函数收敛曲线

由图 2 可知，在经过大约 30 次迭代后，适应度值已经趋近收敛，寻优结束时 CFSFDP 算法的聚类决策图如图 3 所示。聚类的最终结果如表 4 所示。由表 4 可知，绝大多数测点被分为 4 个簇，且较为集中于簇 1 和簇 2，说明绝大多数测点变化规律高度相似，存在较强的关联性。

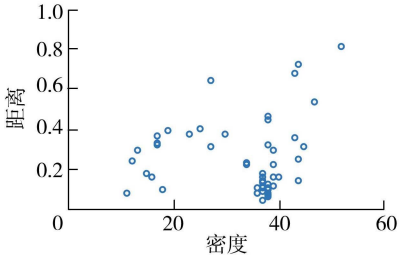


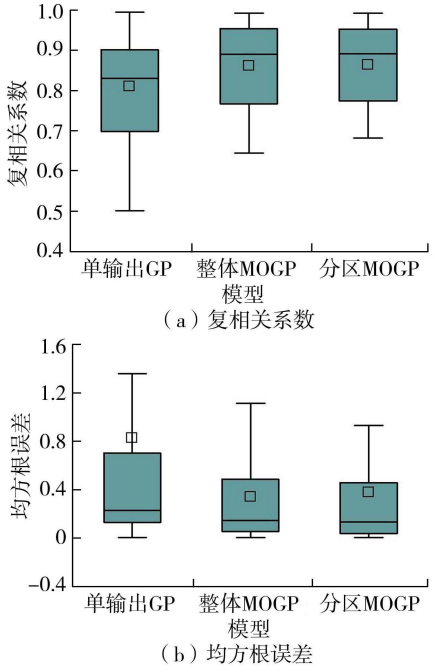
图 3　CFSFDP 算法聚类决策图

表 4　KPCA-ICFSFDP 算法聚类结果

簇类编号	聚类中心	簇内点数
1	PL1-1	30
2	PLR-3	12
3	IPL-1	5
4	IP11-1	6

利用 GP 算法为各测点构建单输出 GP 模型、利用 MOGP 算法为所有测点变形数据构建整体 MOGP 模型和按测点聚类结果分别构建分区 MOGP 模型。建模结果如图 4 所示。由图 4 可知，相较于单输出高斯过程（单输出 GP）模型，MOGP 模型的预测能力

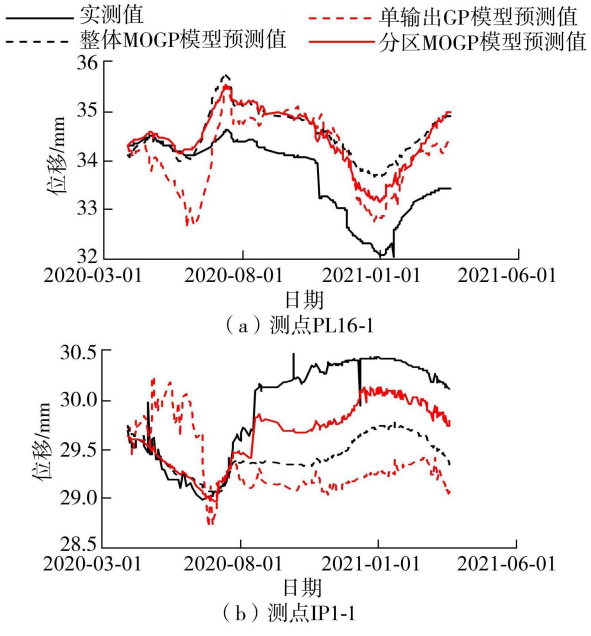
整体表现更好，利用 KPCA-ICFSFDP 算法对数据进行分区处理后，MOGP 模型的预测提升有限，但是对原本预测精度较差的模型有着相对较大的提升，如测点 PL16-1（簇 1）与 IP1-1（簇 2）等测点，其具体对比见图 5。



（a）复相关系数

（b）均方根误差

图 4　建模结果精度箱线图

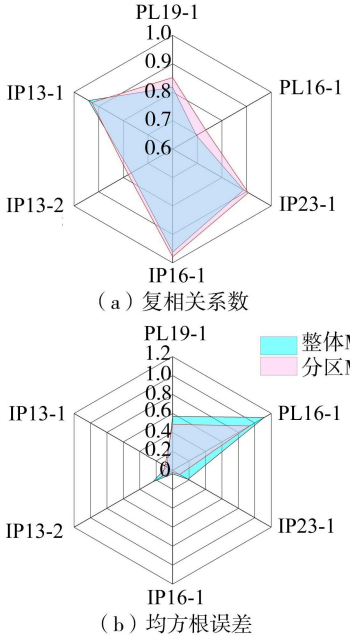


（a）测点PL16-1

（b）测点IP1-1

图 5　预测过程对比

由图 5 可知，当单输出 GP 模型与整体 MOGP 模型均表现得相对较差时，分区 MOGP 模型的表现有所提升，这种提升多出现在簇内样本较少的簇，如簇 1、2、3。可能的原因是，其本身数据序列规律较差，预测难度较高，且与整体主流规律有差异，因此通过聚类算法将与之相关性较高的序列提取出来，再进行 MOGP 建模，有助于提高其精度。例如，簇 3 的 6 个测点的建模结果比较如图 6 所示。由图 6 可知，分区 MOGP 模型各测点预测结果的复相关系数普遍高于整体 MOGP 模型，均方根误差则普遍较小，表明其精度更高。



（a）复相关系数

（b）均方根误差

图 6　簇 3 各序列建模精度

而在簇 4 中分区 MOGP 模型仍然表现优异，如典型测点 PL19-2，由图 7 可见，其预测效果非常好，预测结果的均值基本趋势与实测值保持一致，虽存在误差，但绝大多数能够落入 95% 的置信区间（$\pm 1.96\sigma$）内。

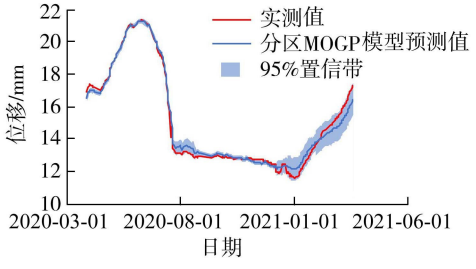


图 7　PL19-2 预测效果

本文提出的基于 KPCA-ICFSFDP-MOGP 构建的拱坝多测点变形 MOGP 回归模型能够有效识别出大坝变形监测数据中的异常值，提高数据质量，提升预处理效率；并能够通过挖掘拱坝各测点间变形之间的联系对各测点变形序列进行聚类分析。

3　结　语

针对大坝多测点变形自动化测值数据量庞大难以处理，以及分析预测中缺乏对测点序列之间相关性分析的问题，基于 KPCA-ICFSFDP-MOGP 算法，建

立了拱坝多测点变形 MOGP 回归模型。通过工程实例验证发现,利用 KPCA 算法能够有效识别大坝变形监测数据中的异常值,为接下来的数据挖掘工作打下良好基础;利用 KPCA-ICFSFDP 算法进行变形测点分区,能够更准确地提取测点序列间的内在关联,经由分区后构建的 MOGP 模型,在一些与主流测点相关性较差的测点的预测工作中有相较于整体 MOGP 建模与单输出 GP 建模更好的表现。与此同时,对于大量的规律相同的测点,分区后仍然能够保持相当高的预测精度。本文提出的拱坝多测点变形 MOGP 回归模型目前适用范围较窄,仅能够适用于整体性较高、测点间规律一致性较强的数据,且异常值识别时需要人为先验知识与介入。

参考文献:

[1] 顾冲时,苏怀智,王少伟. 高混凝土坝长期变形特性计算模型及监控方法研究进展[J]. 水力发电学报,2016,35(5):1-14. (GU Chongshi, SU Huaizhi, WANG Shaowei. Advances in calculation models and monitoring methods for long-term deformation behavior of concrete dams[J]. Journal of Hydroelectric Engineering,2016,35(5):1-14. (in Chinese))

[2] 周兰庭,柳志坤. 大坝变形实测数据的多重分形特征解析方法[J]. 水利水电科技进展,2021,41(6):18-24. (ZHOU Lanting,LIU Zhikun. Multifractal feature analysis method for measured data of dam deformation [J]. Advances in Science and Technology of Water Resources,2021,41(6):18-24. (in Chinese))

[3] 杨光,顾昊,刘尚蔚,等. 基于分数阶模式的高拱坝变形安全监控研究构想[J]. 水利水电科技进展,2022,42(5):85-93. (YANG Guang, GU Hao, LIU Shangwei, et al. Research proposal for deformation safety monitoring of high arch dams based on fractional-order modeling theory [J]. Advances in Science and Technology of Water Resources,2022,42(5):85-93. (in Chinese))

[4] 李子阳,李涵曼,李政勰,等. 基于多测点云相似的混凝土坝变形性态关联分析[J]. 水利水电科技进展,2021,41(6):13-17. (LI Ziyang,LI Hanman,LI Zhengxie,et al. Correlation analysis of deformation behavior of concrete dams based on similarity of multi-point cloud [J]. Advances in Science and Technology of Water Resources,2021,41(6):13-17. (in Chinese))

[5] 刘冲,沈振中,甘磊,等. 基于模糊灰色聚类-组合赋权的病险水库康复度综合评价方法[J]. 水利水电科技进展,2018,38(3):36-41. (LIU Chong,SHEN Zhenzhong, GAN Lei,et al. Comprehensive evaluation of rehabilitation degree of dangerous reservoirs based on fuzzy gray clustering and combination weighting method [J]. Advances in Science and Technology of Water Resources,2018,38(3):36-41. (in Chinese))

[6] 李俊,苏怀智,周仁练. 基于关联规则的土石坝渗流推理预测方法及应用[J]. 水利水电科技进展,2019,39(5):89-94. (LI Jun,SU Huaizhi,ZHOU Renlian. Seepage prediction method based on association rules and its application to earth-rock dams[J]. Advances in Science and Technology of Water Resources,2019,39(5):89-94. (in Chinese))

[7] 马福军,胡力勤. 密度峰值聚类算法在管廊大数据挖掘中应用[J]. 机电工程技术,2022,51(2):94-97. (MA Fujun,HU Liqin. Application of density peak clustering algorithm in big data mining of underground utility tunnel [J]. Mechanical & Electrical Engineering Technology,2022,51(2):94-97. (in Chinese))

[8] 王东风,李嘉宇,黄宇,等. 风电场风速预测的密度峰值聚类方法研究[J]. 太阳能学报,2021,42(12):110-118. (WANG Dongfeng, LI Jiayu, HUANG Yu, et al. Research on density peaks clustering method for wind speed prediction of wind farms[J]. Acta Energiae Solaris Sinica,2021,42(12):110-118. (in Chinese))

[9] 郑虹,周丽媛,韩旭明. 布谷鸟优化的密度峰值快速搜索聚类算法[J]. 长春工业大学学报,2018,39(3):253-260. (ZHENG Hong, ZHOU Liyuan, HAN Xuming. A cuckoo search optimization-based density peak clustering algorithm [J]. Journal of Changchun University of Technology,2018,39(3):253-260. (in Chinese))

[10] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science,2014,344(6191):1492-1496.

[11] WANG Shuliang, WANG Dakui, LI Caoyuan, et al. Clustering by fast search and find of density peaks with data field[J]. Chinese Journal of Electronics,2016,25(3):397-402.

[12] 邱小梦,欧阳亮酉,周世健. 多输出高斯过程在变形监测数据处理中的研究及应用[J]. 江西科学,2019,37(6):832-835. (QIU Xiaomeng, OUYANG Liangyou, ZHOU Shijian. Research and application of multi-output Gaussian process in deformation monitoring data processing. Jiangxi Science[J], 2019,37(6):832-835. (in Chinese))

(收稿日期:2022-07-22 编辑:俞云利)