

## 基于时空聚类挖掘的库岸边坡位移监测数据约简

陈波, 詹明强, 黄梓莘

### Reduction of multi-point displacement monitoring data of reservoir bank slope based on spatio-temporal clustering mining

CHEN Bo, ZHAN Mingqiang, HUANG Zishen

在线阅读 View online: <https://doi.org/10.12170/20210728002>

#### 您可能感兴趣的其他文章

#### Articles you may be interested in

##### 海上筒型基础的筒壁土压力计算

Earth pressure calculation for bucket wall of offshore bucket foundation

水利水电工程学报. 2018(3): 65 <https://doi.org/10.16198/j.cnki.1009-640X.2018.03.009>

##### 太湖流域湖西区暴雨时空分布特性分析

Analysis of spatio-temporal distribution of rainstorms in the western region of Taihu Lake Basin

水利水电工程学报. 2020(3): 43 <https://doi.org/10.12170/20190430002>

##### “烟花”台风影响下长江南通以下河段的增水分布特征

Spatio-temporal distribution characteristics of surge in the reach below Nantong of Yangtze Estuary under the influence of Typhoon In-Fa

水利水电工程学报. 2021(6): 25 <https://doi.org/10.12170/20210930003>

##### 水位骤降影响下岸坡稳定性简化分析

Simplified analysis method for the bank slope stability with the influence of water level plummet

水利水电工程学报. 2021(3): 16 <https://doi.org/10.12170/20200524001>

##### 基于位移的高桩码头地震易损性分析

Seismic vulnerability analysis of pile-supported wharves based on displacement

水利水电工程学报. 2019(3): 76 <https://doi.org/10.16198/j.cnki.1009-640X.2019.03.010>

##### 基于BP-ARIMA的混凝土坝多尺度变形组合预报模型

Multi-scale deformation combination forecast model for concrete dam based on BP-ARIMA

水利水电工程学报. 2018(2): 52 <https://doi.org/10.16198/j.cnki.1009-640X.2018.02.007>



扫码进入官网，阅读更多精彩文章



关注微信公众号，获得更多资讯信息

DOI:10.12170/20210728002

陈波, 詹明强, 黄梓莘. 基于时空聚类挖掘的库岸边坡位移监测数据约简 [J]. 水利水运工程学报, 2022(5): 94-101. (CHEN Bo, ZHAN Mingqiang, HUANG Zishen. Reduction of multi-point displacement monitoring data of reservoir bank slope based on spatio-temporal clustering mining[J]. Hydro-Science and Engineering, 2022(5): 94-101. (in Chinese))

# 基于时空聚类挖掘的库岸边坡位移监测数据约简

陈波<sup>1,2</sup>, 詹明强<sup>1,2</sup>, 黄梓莘<sup>3</sup>

(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. 中国电建集团中南勘测设计研究院有限公司, 湖南 长沙 410014)

**摘要:** 库岸边坡失稳会对工程自身效益和周边安全造成巨大损失, 而位移监测数据可以直接表征库岸边坡安全状况。传统变形位移分析一般仅考虑单个监测点, 不同监测点之间位移的相似性和关联性有待挖掘。基于时空数据挖掘领域的聚类方法, 综合考虑测点属性和空间特征, 采用 K-means 算法度量测点间的相似程度, 实现变形区域划分; 在变形区域划分基础上, 采用遗传算法优化的投影聚类算法, 将高维数据向低维空间映射, 通过提取测点数据特征, 筛选得到重点关注的测点和压缩数据量。经实例工程数据验证, 时空聚类挖掘方法便捷、有效, 逐步实现了边坡位移监测数据约简, 可用于类似库岸边坡的监测数据挖掘。

**关键词:** 库岸边坡; 位移; 时空数据挖掘; 划分聚类; 投影聚类

中图分类号: TV698.1

文献标志码: A

文章编号: 1009-640X(2022)05-0094-08

库岸边坡的运行稳定对水利工程的服役安全至关重要, 其失稳灾害会对工程自身效益和周边安全造成巨大损失。研究表明, 库岸边坡失稳破坏会经历渐变到突变的累进发展过程, 而边坡运行监测资料记录了失稳灾害孕育的全过程信息, 尤其是失稳破坏发生的前兆信息, 有必要结合监测资料开展库岸边坡的安全监控和灾变预警研究。

库岸边坡运行的影响因素众多且内在关系复杂, 考虑到传统边坡监测分析方法在应对时空数据结构时存在的局限性, 有学者引入数据挖掘领域的聚类分析算法, 针对边坡工程监测数据的类聚性提取特征和获得知识, 为边坡防护治理和科学决策提供依据。如王述虹等<sup>[1]</sup>将人工鱼群算法和 K-means 算法相结合, 提出一种用于岩体结构面产状分类的改进 AFSA-RSK 算法, 显著提升了运算速度和预测精度; 秦雨樵等<sup>[2]</sup>综合考虑边坡点位移及其对应的点安全系数, 提出一种基于 K-means 聚类算法的滑面搜索方法, 有效识别了边坡潜在的危险区域; 李佳伟等<sup>[3]</sup>在考虑边坡稳定性关键影响因素的基础上建立了投影寻踪聚类模型, 并进一步结合安全系数法综合评价边坡的稳定性; 徐哲等<sup>[4]</sup>融合 K-means 聚类及神经网络算法, 构建了边坡的稳定性评价模型, 并结合工程实例证明所建立模型的预测精确度; 王俊杰等<sup>[5]</sup>采用 K-means 算法对优势结构面的赤平投影交线进行划分聚类, 分类结果较为合理可靠; 王卓等<sup>[6]</sup>结合统计分类和 K-means 聚类方法对研究区裂缝段进行了危险性分级, 为区域的防灾减灾工作提供有力支持; Wang 等<sup>[7]</sup>结合 K-means 聚类算法、Alpha 形状、三次样条插值提出一种自动识别临界滑动面的方法, 根据极限状态下的测点位移准确识别边坡二维和三维临界滑动面; 金永强等<sup>[8]</sup>针对边坡运行监测数据的高维非线性特征, 采用投影寻踪及和声搜索相结合的算法实现了边坡稳定性的有效评价; Dyson 等<sup>[9]</sup>采用随机有限元法构建边坡稳定性分析模型, 针对随机场相似性采用层次聚类分析方法对边坡几何性状进行分类。

综上所述, 虽然聚类分析的引入推进了边坡监测数据挖掘的发展, 但目前针对动态多方位的实时监测

收稿日期: 2021-07-28

基金项目: 国家自然科学基金面上项目(52079049); 国家自然科学基金重点项目(51739003); 国家重点实验室基本科研业务费(522012272)

作者简介: 陈波(1986—), 男, 浙江绍兴人, 青年教授, 博士, 主要从事大坝安全监控研究。E-mail: chenbo@hhu.edu.cn

数据的挖掘工作开展较少,挖掘工作的开展深度及挖掘结果的应用频度尚处于浅尝辄止的阶段。基于此,针对边坡监测数据的多维时空特征引入适用于边坡海量监测信息的时空数据挖掘方法,采用 K-means 聚类算法划分测点区域和投影聚类算法提取数据特征,实现边坡监测数据约简的目的,将深层次的挖掘方法和有价值的挖掘信息应用于边坡安全监测。

## 1 基于时空聚类分析的库岸边坡监测数据挖掘

边坡监测项目繁多,测点监测信息丰富,采用单项目或单测点的信息评价边坡稳定具有片面性。在时空数据挖掘理论中,聚类分析是将数据集按照一定规则划分成不同类簇的方法,使划分结果具有“高内聚,低耦合”的显著特征,以达到类内数据相似度高、类间数据相似度低的目的。基于上述聚类思想,根据实际需求的不同又衍生了包括划分聚类、层次聚类、网格聚类、密度聚类和投影聚类等多种算法。

为最大程度挖掘边坡监测资料、反映数据时空特征,本文主要采用划分聚类和投影聚类算法开展时空数据挖掘,分别对边坡监测资料进行测点区域划分和特征提取,以达到边坡监测数据约简的目的。

### 1.1 基于划分聚类算法的边坡测点分区

K-means 算法是一种基于划分的无监督学习的经典聚类算法,通过穷举的方式寻找全局最优结果,并通过计算簇中对象的平均值实现划分目的<sup>[10]</sup>。算法步骤为:初始化类簇中心、初步划分数据集、重生成类簇中心、算法收敛判断。重复第 2~4 步不断更新类簇中心,直至类簇中心不再发生变化时,循环结束并输出聚类结果。

考虑到划分聚类的 K-means 算法简单易行、效果良好的特点,选取 K-means 算法进行边坡测点分区。利用 K-means 算法实现边坡测点分区目的,关键在于灵活构建测点的距离度量指标,综合考虑不同测点之间的相似程度。其中,空间距离指标可衡量不同测点之间空间位置的远近程度,属性距离指标可衡量不同测点之间的属性差异程度。同时,边坡的变形与其稳定性密切相关,变形作为边坡内部稳定状态动态演化的外部直接反映,可以捕捉到与边坡稳定密切相关的物理信息。因此,采用边坡位移测值计算属性距离,同时采用测点水平向和垂直向的空间坐标计算空间距离,不同的距离指标计算式如下:

$$d_1 = \sqrt{\sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^m |Z_{ij} - Z_{kj}|^2} \quad (1)$$

$$d_2 = \sqrt{\sum_{i=1}^n \sum_{k=1}^n (X_i - X_k)^2 + (Y_i - Y_k)^2 + (H_i - H_k)^2} \quad (2)$$

式中:  $d_1$  为测点的属性距离指标;  $Z_{ij}$  为第  $i$  测点第  $j$  时刻的位移测值;  $n$  为测点总数;  $m$  为最长时刻数;  $Z_{kj}$  为第  $k$  测点第  $j$  时刻的位移测值;  $d_2$  为测点的空间距离指标;  $X_i$ 、 $Y_i$ 、 $H_i$  表征  $i$  测点水平向和垂直向的空间坐标;  $X_k$ 、 $Y_k$ 、 $Z_k$  表征  $k$  测点水平向和垂直向的空间坐标。

综合考虑属性距离和空间距离,加权确定测点的综合距离指标:

$$d = w_1 d_1 + w_2 d_2 = w_1 \sqrt{\sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^m |Z_{ij} - Z_{kj}|^2} + w_2 \sqrt{\sum_{i=1}^n \sum_{k=1}^n (X_i - X_k)^2 + (Y_i - Y_k)^2 + (H_i - H_k)^2} \quad (3)$$

式中:  $d$  为测点的综合距离指标;  $w_1$  和  $w_2$  分别为属性距离和空间距离指标所占的权重,满足  $w_1 + w_2 = 1$ ,通常取  $w_1 = w_2 = 1/2$ 。

在构造测点综合距离指标的基础上,基于 K-means 算法原理实现边坡位移测点的区域划分。

### 1.2 基于投影聚类算法的边坡位移特征提取

**1.2.1 投影聚类算法原理** 投影聚类是处理高维、非线性及非正态数据的一类新型统计方法,通过寻找反映原始资料数据特征的投影,将高维样本数据映射到低维子空间中<sup>[3,11]</sup>。基于前面 K-means 聚类算法的测

点分区结果,采用投影聚类算法进一步提取位移数据特征并压缩数据量级。以下分步骤介绍投影聚类算法。

(1)数据无量纲处理。对测点数为 $n$ 、时序长度为 $m$ 的边坡位移监测数据集 $\{X_{ij}\}$ 进行无量纲处理,尽可能消除输入数据之间的量纲差异,并将处理后输入数据以 $x(i, j)$ 表示,其中, $i = 1, 2, \dots, n, j = 1, 2, \dots, m, x(i, j)$ 为第 $i$ 测点的第 $j$ 时刻值。

为充分提取边坡位移监测数据集的特征,通过采取年均值和年极值两个统计特征对数据集进行描述,以表征数据集的数值大小、极值分布和测值变化等情况。将边坡监测数据集 $\{X_{ij}\}$ 转化为均值集 $\{X\bar{r}_{ij}\}$ 和极值集 $\{Xx'_{ij}\}$ 后再进行无量纲处理,方便后续运算。

(2)构造投影函数。为实现线性空间高维数据的投影,需要构造投影函数,其中, $m$ 维数据 $\{x_{ij}|0 \leq i \leq n, 0 \leq j \leq m\}$ 的投影方向为 $a = \{a_1, a_2, \dots, a_j, \dots, a_n\}$ , $\{z_i|0 \leq i \leq n\}$ 则为投影方向上的投影聚类序列值,存在:

$$z(i) = \sum_{j=1}^m a(j)x(i, j) \quad (4)$$

(3)构建投影指标函数。投影指标函数有助于最优投影方向的选取,也是将高维数据向低维空间映射的关键。定义投影指标函数为:

$$Q(a) = S_z D_z \quad (5)$$

式中:投影点 $z_i$ 的标准差 $S_z = \sqrt{\sum_{i=1}^n (z(i) - E_z)^2 / (n-1)}$ , $E_z$ 为投影点 $z_i$ 的平均值;投影点 $z_i$ 的局部密度 $D_z = \sum_{i=1}^n \sum_{j=1}^m (R - r_{ij})u(R - r_{ij})$ , $r_{ij}$ 为样本间距, $R$ 为局部数据空间密度的窗口半径, $u(R - r_{ij})$ 为单位阶跃函数,当 $R \geq r_{ij}$ 时,其值为 0,反之其值为 1。

(4)优化投影指标函数。投影方向 $a$ 决定了投影指标函数值 $Q(a)$ 的大小,因此,最优投影方向的选取可以转化为投影指标函数极大值求解的问题,数学表达式为:

$$\begin{cases} \max Q(a) = S_z D_z \\ \text{s.t. } \sum_{j=1}^m a^2(j) = 1, \quad a(j) \in [0, 1] \end{cases} \quad (6)$$

(5)综合聚类分析。将步骤(4)求解得到的最优投影方向 $a^*$ 代入式(4),得到投影特征结果 $z^*(i)$ 。

**1.2.2 基于改进投影聚类算法的边坡特征提取** 投影聚类的关键在于最优投影方向的选取。然而,最优投影方向的求解本身是一个复杂的非线性优化问题,同时由于边坡监测数据量级较大,客观上限制了方法的实用性,有必要引入合适的优化算法帮助确定最优投影方向。因此,引入遗传算法<sup>[12]</sup>优化投影聚类的计算过程。遗传算法需要确定目标函数,算法目标函数见式(7),约束函数为 $\sum_{j=1}^m a^2(j) = 1$ 。

$$\max Q(a^*) = \sqrt{\sum_{i=1}^n (z(i) - E_z)^2 / (n-1)} \sum_{i=1}^n \sum_{j=1}^m (R - r_{ij})u(R - r_{ij}) \quad (7)$$

适应度函数是评价种群个体好坏的重要因素,要求算法结构简单且计算结果非负,以尽可能降低算法复杂度。根据目标函数设置适应度函数 $f = Q(a)$ ,由此计算出每个种群个体的适应度值。

## 2 工程实例

选取某拱坝库首左岸边坡为研究对象。此边坡属于该拱坝的近岸坝坡,距离大坝 600~1 300 m,顺河方

向长 700 m, 相对坡高 500~700 m。高程 1 400 m 以上平均坡度为 25°~45°, 高程 1 400 m 以下为 22°~25°, 并有多级缓坡地段。坡面走向约 S60°E, 岩层产状近 EW/S∠30°~35°, 边坡为二元结构的单斜顺向坡。由于边坡沿河各段的稳定程度不一, 因此在初步设计时, 根据边坡地质构造和失稳破坏模式的不同, 将 1 400 m 高程以下的边坡自上游向下游分为 I、II、III 区, 如图 1 所示。

2.1 边坡变形区域划分

基于 I 区、II 区和 III 区共 24 个地表位移测点的监测数据, 包括顺河向、顺坡向和垂直向 3 个方向, 采用 Matlab 自编 K-means 聚类算法进行测点区域划分。

在进行划分聚类之前, 首先计算测点空间距离和属性距离获得测点的综合距离指标。根据 X、Y 坐标和高程计算各个测点之间的空间距离, 同时, 根据顺河向、顺坡向和垂直向的三向位移测值计算各测点的属性距离。在对属性距离和空间距离进行标准化处理后, 加权计算获得测点的综合距离指标。

分区结果的正确率为聚类分区结果和勘测设计人员的参考分区结果相吻合的测点数与库岸边坡位移测点总数的比值。

设置初始聚类数目为 3, 在不指定聚类中心的情况下, 计算不同测点间的平方欧式距离以度量测点间的相似程度, 并采用循环迭代的方式实现测点区域划分的最优效果。如图 2 所示, 对比基于划分聚类算法的测点分区结果和基于初设资料的测点原始分区结果, 在 24 个位移测点中, 除测点交 5 和交 23 外, 其余 22 个测点的分区结果均与基于初设资料的测点分区结果吻合, 分区正确率达 91.7%, 这表明基于 K-means 算法的测点分区结果较为真实可信。

2.2 数据特征提取

在测点分区的基础上, 采用遗传算法改进的投影聚类算法, 对边坡监测信息做进一步的数据特征提取。首先针对 24 个地表位移测点 1997—2019 年的测值, 求取三向位移的年均值和年极值, 初步提取边坡监测数据集的数据特征, 形成 6 个  $n=23$ 、 $m=24$  的位移测值矩阵  $\{X\bar{r}_{ij}\}$ 、 $\{Xx'_{ij}\}$ ; 在对测值数据进行标准化的基础上, 设置遗传算法的运行参数, 包括: 迭代次数  $k=50$ 、窗口半径系数  $\alpha=0.1$ 、变量下界  $L_B=-1$ 、变量上界  $U_B=1$ 。根据投影聚类算法原理, 进一步设置相应目标函数、适应度函数及约束函数。

以顺坡区位移年均值为例, 经遗传算法的迭代计算, 求解获得最优投影方向  $B_a=(-0.188, 0.324, -0.258, 0.066, -0.206, -0.222, 0.140, 0.275, -0.105, -0.114, 0.165, 0.167, 0.400, 0.501, -0.025, 0.275, 0.229, 0.008, -0.039, -0.170, -0.129, 0.040, 0.185)$ ; 由此得到顺坡区位移年均值的投影聚类结果:  $Q(a)=(0.463, 0.468, 0.359, 0.237, 0, 0.479, 0.394, 0.352, 0.330, 0.332, 0.172, 0.274, 0.172, 0.052, 0.040, 0.893, 0.301, 0.322, 0.371, 0.201, 0.525, 0.037, 0.530, 1.000)$ ,  $Q(a)$  各列数值对应各个地表位移测点的投影聚类特征值。同理可得其余方向位移测点年均值、年极值的最优投影方向, 以及相应的测点投影聚类特征结果。

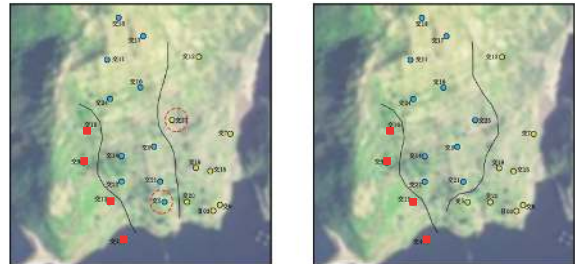
综上, 各区基于改进的投影聚类算法计算过程如图 3、4 所示, 投影聚类特征值汇总如表 1 所示。为方便直观对比分析, 将表 1 的投影聚类结果分区域、分方向整理成如图 5 所示的散点图。根据图 5 基于遗传算法的位移测点投影聚类结果, 以投影聚类特征值作为判断指标, 可以筛选出需要重点关注的测点。



图 1 边坡地理位置示意

Fig. 1 Schematic diagram of slope geographical position

■ I 区测点 ● II 区测点 ● III 区测点



(a) 初设资料测点分区结果 (b) 划分聚类测点分区结果

图 2 测点分区结果对比

Fig. 2 Comparison of zoning results of measuring points

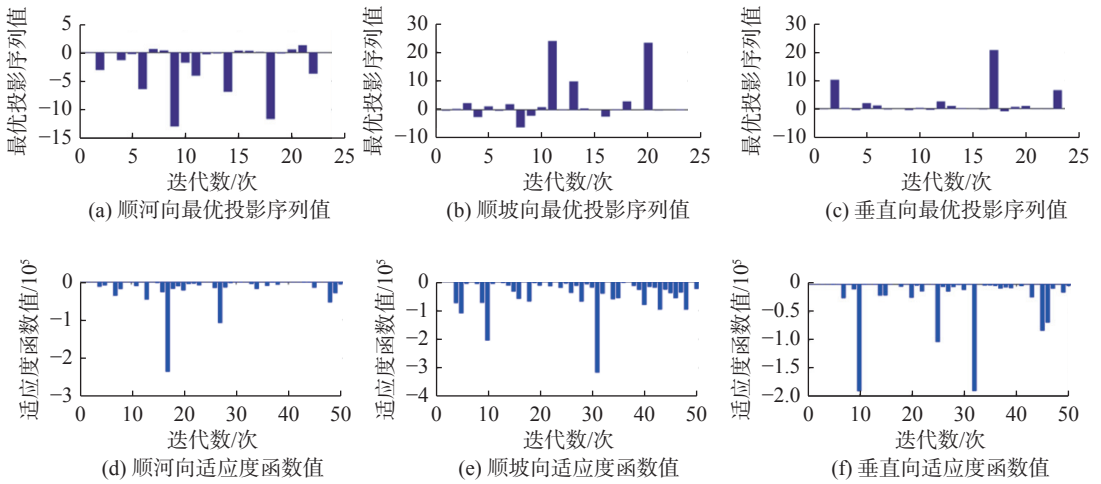


图 3 遗传算法改进的投影聚类计算过程(均值)

Fig. 3 Improved projection clustering calculation process diagram of genetic algorithm (mean value)

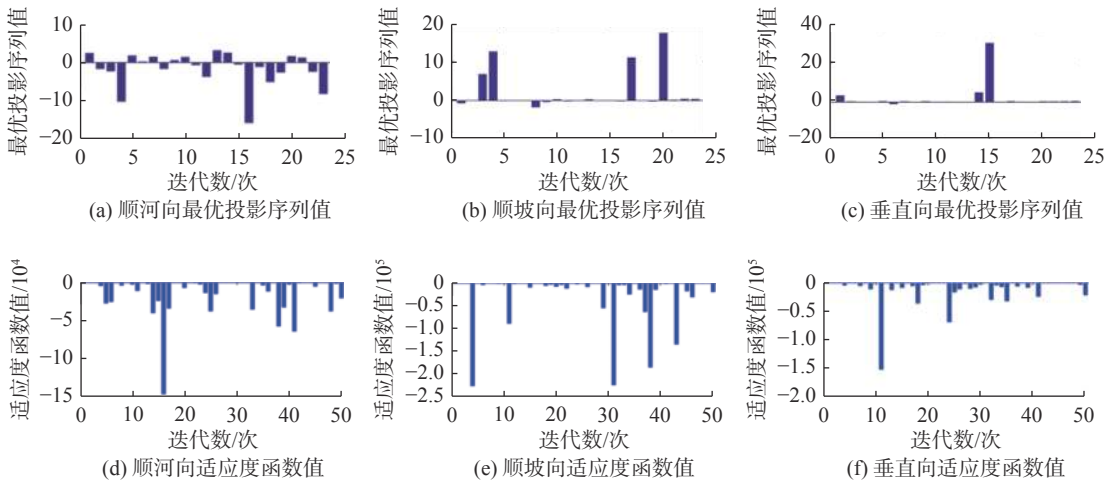


图 4 遗传算法改进的投影聚类计算过程(极值)

Fig. 4 Improved projection clustering calculation process diagram of genetic algorithm (extreme value)

表 1 基于遗传算法的分区位移测点投影聚类结果

Tab. 1 Projection clustering results of subarea displacement measurement points based on genetic algorithm

分区结果	交点	位移方向			分区结果	交点	位移方向		
		顺河	顺坡	垂直			顺河	顺坡	垂直
I 区	交4	0.331/0.348	0.463/0.119	0/0	II 区	交21	0.371/0.382	0.172/0.147	0.142/0.108
	交9	0.177/0.197	0.468/0.460	0.171/0.118		交22	0.398/0.435	0.052/0.066	0.102/0.062
	交13	0.437/0.485	0.359/0.334	0.178/0.132		交24	0.380/0.420	0.040/0.048	0.112/0.077
	交16	0.423/0.482	0.237/0.267	0.220/0.175		交6	0.266/0.325	0.893/0.891	0.526/0.476
	D03	0.368/0.417	0/0	0.090/0.054		交7	0/0.003	0.301/0.291	0.254/0.206
II 区	交5	0.596/0.652	0.479/0.468	0.223/0.183	交12	0.382/0.411	0.322/0.314	0.206/0.153	
	交8	0.288/0.318	0.394/0.378	0.213/0.159	交15	0.034/0	0.371/0.349	0.292/0.231	
	交10	0.460/0.494	0.352/0.342	0.260/0.213	III 区	交19	0.282/0.286	0.201/0.166	0.084/0.039
	交11	1.000/1.000	0.330/0.322	1.000/1.000		交20	0.438/0.498	0.525/0.507	0.285/0.286
	交14	0.388/0.439	0.332/0.314	0.154/0.096		交23	0.371/0.406	0.037/0.038	0.106/0.066
	交17	0.540/0.601	0.172/0.152	0.267/0.210	D01	0.341/0.360	0.530/0.520	0.223/0.239	
	交18	0.177/0.160	0.274/0.245	0.300/0.242	II 03	0.460/0.647	1.000/1.000	0.457/0.409	

注：“/”前为均值，“/”后为极值。

以顺河向位移测值的投影聚类计算结果为例,结合图 6(a)所示的典型测点的测值过程线,可以看出:年均值和年极值投影聚类特征值最大的都是测点交 11,最小的都是测点交 7,分别对应顺河向测值序列中的极大值和极小值,两个测点的位移时间曲线各自向正负两个方向延伸,而投影聚类特征值趋近于中值水平的测点,如测点交 10,测值波动幅度不大且不存在明显递增、递减趋势,测点运行状态比较安全。

根据上述规律筛选出其余两个方向需要重点关注的测点。在顺坡向测点中,大部分测点的投影聚类特征值均小于 0.6,测值变化比较平稳,而投影特征值趋近于 1 的Ⅲ区测点Ⅱ 03 和交 6 都呈现出明显递增趋势,20 年间的测值变幅达到 600 mm;在垂直向测点中,除Ⅱ区的测点交 11 外,其余测点的特征值均小于 0.6,测值变化规律相似且发展态势平稳。

由上述分析可见,基于遗传算法优化的投影聚类算法可以有效提高计算效率、压缩数据量级,同时根据计算结果能快速提取测点数据特征,直观反映出不同区域、不同方向的测值分布情况,筛选出其中需要重点关注的测点。

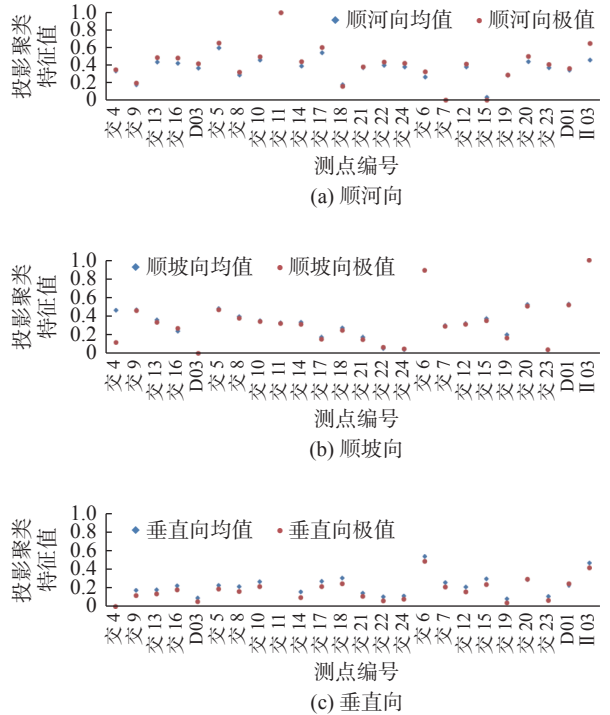


图 5 基于遗传算法的位移测点投影聚类结果  
Fig. 5 Projection clustering results of displacement measuring points based on genetic algorithm

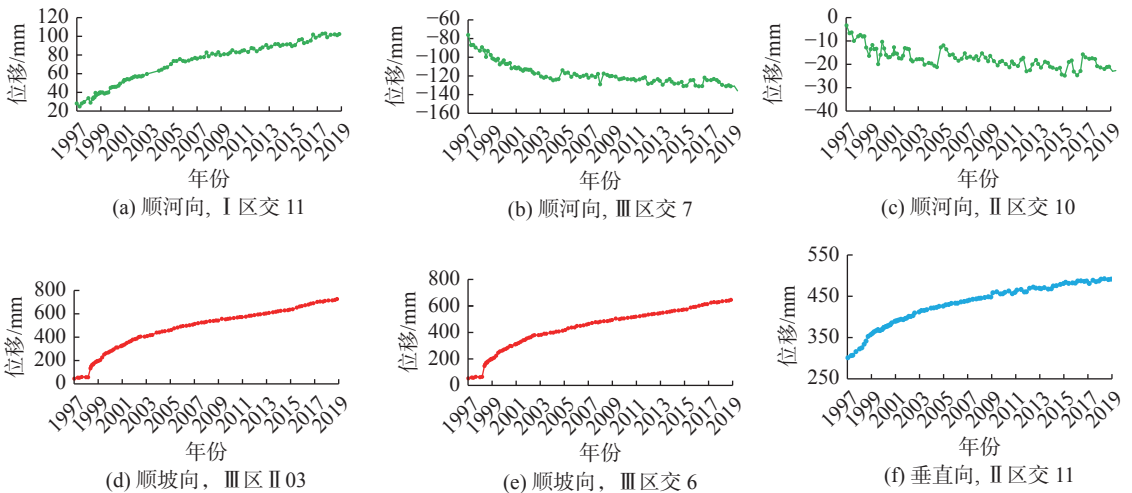


图 6 典型测点位移过程  
Fig. 6 Typical displacement process of measuring points

### 3 结 语

结合工程实际,针对边坡监测信息的多维特性和时空特征,引入时空数据挖掘领域的聚类方法,开展多测点多项目海量边坡监测信息的时空数据挖掘工作。结果表明,综合考虑库岸边坡测点属性特征和空间特征,采用 K-means 算法度量测点间的相似程度,可实现测点区域准确划分。在测点分区的基础上采用遗传

算法优化的投影聚类算法,将高维数据向低维空间进行映射,可以提取测点数据特征,从而压缩数据量级并筛选出需重点关注的测点,即基于聚类分析逐步实现了边坡位移监测数据的约简。

## 参 考 文 献:

- [1] 王述红,任艺鹏,陈俊智,等.一种改进鱼群聚类算法在结构面分组中的应用[J].*东北大学学报(自然科学版)*,2019,40(3):420-424.(WANG Shuhong,REN Yipeng,CHEN Junzhi,et al.An improved fish swarm clustering algorithm for structural grouping[J].*Journal of Northeastern University (Natural Science)*,2019,40(3):420-424.(in Chinese))
- [2] 秦雨樵,汤华,冯振洋,等.基于聚类分析的边坡稳定性研究[J].*岩土力学*,2018,39(8):2977-2983,2990.(QIN Yuqiao,TANG Hua,FENG Zhenyang,et al.Slope stability evaluation by clustering analysis[J].*Rock and Soil Mechanics*,2018,39(8):2977-2983,2990.(in Chinese))
- [3] 李佳伟,李婷,顾中明,等.基于投影寻踪聚类方法的边坡稳定性分析[J].*水电与新能源*,2019,33(6):27-30.(LI Jiawei,LI Ting,GU Zhongming,et al.Slope stability analysis based on projection pursuit clustering method[J].*Hydropower and New Energy*,2019,33(6):27-30.(in Chinese))
- [4] 徐哲,胡焕校,邓超.K-means聚类神经网络在边坡稳定性评价中的应用探究[J].*水资源与水工程学报*,2017,28(3):198-204.(XU Zhe,HU Huanxiao,DENG Chao.Study on the application of K-means clustering algorithm and neural network in slope stability evaluation[J].*Journal of Water Resources and Water Engineering*,2017,28(3):198-204.(in Chinese))
- [5] 王俊杰,冯登,柴贺军,等.基于赤平极射投影和K-均值聚类算法的优势结构面分析[J].*岩土工程学报*,2018,40(1):74-81.(WANG Junjie,FENG Deng,CHAI Hejun,et al.Dominant discontinuities based on stereographic projection and K-means clustering algorithm[J].*Chinese Journal of Geotechnical Engineering*,2018,40(1):74-81.(in Chinese))
- [6] 王卓,许强,魏勇,等.基于统计和聚类分析的黄土台塬边缘裂缝危险性分级[J].*安全与环境工程*,2020,27(1):62-68,74.(WANG Zhuo,XU Qiang,WEI Yong,et al.Risk classification of cracks in the loess platform edge based on statistics and cluster analysis[J].*Safety and Environmental Engineering*,2020,27(1):62-68,74.(in Chinese))
- [7] WANG Y K,HUANG J S,TANG H M. Automatic identification of the critical slip surface of slopes[J].*Engineering Geology*,2020,273:105672.
- [8] 金永强,苏怀智,李子阳.基于和声搜索的边坡稳定性投影寻踪聚类分析[J].*水利学报*,2007,38(增刊1):682-686.(JIN Yongqiang,SU Huaizhi,LI Ziyang.Slope stability projection pursuit clustering analysis based on harmony search[J].*Journal of Hydraulic Engineering*,2007,38(Suppl 1):682-686.(in Chinese))
- [9] DYSON A P,TOLOOIYAN A. Prediction and classification for finite element slope stability analysis by random field comparison[J].*Computers and Geotechnics*,2019,109:117-129.
- [10] 王振武.数据挖掘算法原理与实现[M].2版.北京:清华大学出版社,2017.(WANG Zhenwu.Principle and implementation of data mining algorithm[M].2nd ed.Beijing:Tsinghua University Press,2017.(in Chinese))
- [11] 司俊燕,郭海庆,徐飞,等.PSO-PP模型在岩石高边坡稳定性评价中的应用[J].*水电能源科学*,2011,29(1):79-82.(SI Junyan,GUO Haiqing,XU Fei,et al.PSO-PP model for high rock slope stability evaluation and its application[J].*Water Resources and Power*,2011,29(1):79-82.(in Chinese))
- [12] GOLDBERG D E. Genetic algorithms in search, optimization and machine learning[M]. Boston: Addison-Wesley Publisher, 1989.

## Reduction of multi-point displacement monitoring data of reservoir bank slope based on spatio-temporal clustering mining

CHEN Bo<sup>1,2</sup>, ZHAN Mingqiang<sup>1,2</sup>, HUANG Zishen<sup>3</sup>

(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. *PowerChina Zhongnan Engineering Corporation Limited, Changsha 410014, China*)

**Abstract:** The instability disaster of reservoir bank slope will cause huge losses to the benefit of the project and the safety of surrounding life and property, and the displacement monitoring data can directly characterize the safety status of reservoir bank slope. In view of the traditional deformation and displacement analysis, only a single monitoring point is considered, and the similarity and relevance of displacement between different monitoring points still need to be excavated. Based on the clustering method in the field of spatio-temporal data mining, considering the attribute characteristics and spatial characteristics of measuring points, K-means algorithm is used to measure the similarity between measuring points and realize the division of measuring points; based on the division of measuring points, the projection clustering algorithm optimized by genetic algorithm is used to map the high-dimensional data to the low-dimensional space. By extracting the characteristics of measuring point data, the purpose of screening the measuring points needing attention and compressing the data order is achieved. Based on practical engineering data, it is shown that the spatio-temporal clustering mining method is convenient and effective, and gradually reduces the monitoring data of slope displacement. The method can be used for monitoring data mining of similar reservoirs.

**Key words:** reservoir bank slope; displacement; spatio-temporal data mining; partition clustering; projection clustering