

## 基于 KPCA-TFPSO-BL 的泥石流预测研究<sup>\*</sup>

徐根祺<sup>1</sup> 曹宁<sup>2</sup> 李璐<sup>3</sup> 谢国坤<sup>1</sup> 党文博<sup>4</sup>

(1. 西安交通工程学院机械与电气工程学院 西安 710300;

2. 西安交通工程学院土木工程学院 西安 710300;

3. 铜川职业技术学院机电工程学院 铜川 727000;

4. 陕西建工第二建设集团有限公司市政路桥工程公司 宝鸡 721000)

**摘要:**针对当前研究中泥石流诱发因子敏感度各异导致的预测准确度不高、数据集样本有限造成的模型训练和预测效果不佳、非线性过程严重引起的参数难以确定等问题,利用改进的核主成分分析算法(kernel principal component analysis,KPCA)筛选出相关性一般的因子,结合宽度学习(broad learning,BL)建立泥石流概率预测模型,再通过引入正弦因子的粒子群算法(TFPSO)对模型进行优化,最终建立基于 KPCA-TFPSO-BL 的泥石流预测模型。通过实验对比了经典 BL 模型、KPCA-PSO-BL 模型以及 KPCA-TFPSO-BL 模型的性能,结果表明,KPCA-TFPSO-BL 的均方根误差为 4.92,平均绝对误差为 4.60,训练时间为 7.22 s,该模型在预测误差和训练时间方面综合表现最佳。本研究为泥石流预测领域提供了一种新的思路和借鉴。

**关键词:**泥石流;预测模型;宽度学习

**中图分类号:** P642;TN919.5 **文献标识码:** A **国家标准学科分类代码:** 520.20

## Study on debris flow prediction based on KPCA-TFPSO-BL

Xu Genqi<sup>1</sup> Cao Ning<sup>2</sup> Li Lu<sup>3</sup> Xie Guokun<sup>1</sup> Dang Wenbo<sup>4</sup>

(1. School of Mechanical and Electrical Engineering, Xi'an Traffic Engineering Institute, Xi'an 710300, China;

2. Civil Engineering College, Xi'an Traffic Engineering Institute, Xi'an 710300, China;

3. School of Mechanical and Electrical Engineering, Tongchuan Vocational and Technical College,

Tongchuan 727000, China; 4. Municipal Road and Bridge Engineering Company, Shaanxi Construction

Second Construction Group Co., Ltd., Baoji 721000, China)

**Abstract:** In response to the problems of low prediction accuracy caused by the varying sensitivities of debris flow triggering factors in current research, poor model training and prediction performance due to limited dataset samples, and difficulty in determining parameters caused by severe nonlinear processes, an improved kernel principal component analysis (KPCA) algorithm was used to screen out factors with general correlation, combined with broad learning (BL) to establish a debris flow probability prediction model. Then, a particle swarm optimization (PSO) based on sine factors was introduced to optimize the model, and finally, a debris flow prediction model based on KPCA-TFPSO-BL was established. The performance of the classic BL model, KPCA-PSO-BL model, and KPCA-TFPSO-BL model was compared through experiments. The results showed that the root mean square error of KPCA-TFPSO-BL was 4.92, the average absolute error was 4.60, and the training time was 7.22 seconds. This model showed the best comprehensive performance in terms of prediction error and training time. This study provides a new approach and reference for the field of debris flow prediction.

**Keywords:** debris flow; predictive model; broad learning

收稿日期:2024-08-29

<sup>\*</sup> 基金项目:陕西省自然科学基金基础研究计划(2023-JC-YB-464)、西安交通工程学院中青年基金项目(2023KY-02)资助

## 0 引言

我国幅员辽阔,宽广的陆地面积和复杂的地理环境为各种自然灾害的孕育提供了可能,我国自然灾害发生频率和类型常年居于世界前列<sup>[1]</sup>。不规则的地形条件、沟壑纵横的高原山区和破碎的岩体结构,加之人类工程活动的肆虐,造成了植被破坏严重、地质灾害频发的现状<sup>[2-3]</sup>。泥石流作为发生较频繁、破坏性较强的一种地质灾害,长期受到国家和政府有关部门的高度关注,国务院专门成立了防灾减灾部门对灾害进行应对<sup>[4]</sup>。

泥石流的突发性极强,造成的损失极大,给人民生命财产安全带来了很大的影响<sup>[5]</sup>。为了减少乃至消除泥石流造成的伤亡,对泥石流发生情况及时作出准确预测是防灾减灾的有效方式。因此,如何对泥石流的发生做出精准预测一直以来都是该领域的研究热点问题。泥石流的诱发因素众多,致灾机理复杂,针对泥石流形成机理的研究大都基于特定的边界条件,很难建立普适性模型。因此,一部分研究者将目光转向人工智能技术,试图利用人工智能算法预测泥石流发生的可能<sup>[6-7]</sup>。

李璐等<sup>[8]</sup>借助松鼠觅食算法(sparrow search algorithm, SSA)对最小二乘支持向量机进行改进,将优化后的算法用于泥石流预测并获得了较高的准确性,该模型在处理复杂的非线性问题方面存在收敛速度慢的缺点。翟淑花等<sup>[9]</sup>提出了一种基于遗传规划的泥石流多因子融合预测方法,克服了仅以雨量作为指标的传统模型的弊端,但该方法中泥石流暴发影响因子的选取过于主观。李丽敏等<sup>[10]</sup>利用布谷鸟优化轻量梯度提升机对泥石流作出预测并取得了较好的效果,但该模型的局部搜索能力较差,可能存在收敛速度慢且收敛精度不足的风险。曾鼎等<sup>[11]</sup>通过一种具有噪声的基于密度的聚类算法(density-based spatial clustering of applications with noise, DBSCAN)对传统极限学习机进行优化并对泥石流的发生进行预测,预测结果具有较高的准确性,但研究未考虑到 DBSCAN 聚类算法调参过程比较复杂,不同参数组合对预测效果会产生较大影响。张建伟等<sup>[12]</sup>提出了一种基于贝叶斯网络的泥石流预测方法,该方法预测准确率较高,但过于依赖降雨量而忽视了影响泥石流的其他重要因素。邓玖林<sup>[13]</sup>基于因子分析法建立了泥石流预测模型,该模型消除了各影响因子间多重共线性的问题,但预测准确率还有待进一步提高。周伟等<sup>[14]</sup>基于 Fisher 判别法建立了泥石流预测模型并提高了泥石流的预报精度,但该方法局限性较大,推广具有较大难度。

为了解决研究中存在的问题,本文利用改进的核主成分分析算法(kernel principal component analysis, KPCA)对泥石流诱发因子的贡献率进行排序,删除贡献率较低的因子,仅保留影响程度较高的因子,提高模型预测准确性;鉴于一般的传统模型参数寻优困难,借助参数较少的宽度学习(broad learning, BL)建立泥石流概率预测模型;为了

进一步提升模型的参数寻优能力,再通过改进粒子群算法(particle swarm optimization, PSO)对模型参数进行优化,扩大经典 PSO 中粒子的搜寻范围,增加避免陷入局部最优的概率,更进一步改善模型预测准确性、减小预测误差并缩短模型的训练时间。

## 1 相关理论

### 1.1 改进主成分分析

主成分分析(principal component analysis, PCA)是数据降维中常用的方法之一,借助经典主成分分析方法将非线性问题线性化的思想,提出了 KPCA 方法用于特征抽取。传统 KPCA 在通过核函数处理初始数据相关性的过程中利用了各样本间相互独立这一条件<sup>[15]</sup>。在工程中具体应用该方式时,各个特征量所赋予的权值并不相同。因此,在综合了数据合理性以及特征抽取和数据降维可靠性的条件下,对传统 KPCA 进行了改进。为了优化经典 PCA 原始数据的特征权值,在协方差矩阵中引入样本权值。

用于训练的样本集记为  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ ,  $x_i \in \mathbf{R}^p$ ,  $i = 1, 2, \dots, n$ , 其中,  $\mathbf{R}^p$  为输入信息。将输入信息的非线性映射表示为  $\varphi: \mathbf{X} \rightarrow \mathbf{F}$ , 其中,  $\mathbf{F}$  是特征空间。

在协方差矩阵中引入权值,如式(1)所示。

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n h_i \cdot \varphi(x_i) \varphi^T(x_i) \quad (1)$$

式中:  $\mathbf{C}$  是协方差矩阵;  $n$  为训练样本数;  $h_i$  为样本权值。

$$\omega_i = \sqrt{h_i}, \text{ 满足 } \sum_{i=1}^n h_i \varphi(x_i) = 0.$$

将引入权值后的协方差矩阵进行特征分解,结果如下:

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

式中:  $\lambda$  表示特征值;  $\mathbf{v}$  是特征向量。

$$\mathbf{v}_r = \sum_{i=1}^n C_r^i \omega_i \varphi(x_i) \quad r = 1, 2, \dots, n \quad (3)$$

将式(2)改写成:

$$\lambda (\omega_i \varphi(x_i), \mathbf{v}_r) = (\omega_i \varphi(x_i), \mathbf{C} \mathbf{v}_r) \quad (4)$$

将  $\mathbf{v}_r$  和  $\mathbf{C}$  的表达式代入式(4),同时令核矩阵  $\mathbf{K} = (\mathbf{K}_{ij})_{n \times n} = (\omega_i \varphi(x_i), \omega_j \varphi(x_j))$ , 可得:

$$\mathbf{K} \boldsymbol{\alpha}_i = n \lambda \boldsymbol{\alpha}_i \quad i = 1, 2, \dots, n \quad (5)$$

式中:  $\boldsymbol{\alpha}_i$  和  $n \lambda$  分别表示核矩阵的特征向量和特征值。

取前  $m$  个特征值对应的归一化特征向量  $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m$ , 将  $\varphi(x_j)$  投影至  $\mathbf{v}_r$  上,可得:

$$\mathbf{g}_r(x_j) = \varphi(x_j) \cdot \mathbf{v}_r = \sum_{i=1}^m \boldsymbol{\alpha}_i^i (\omega_i \varphi(x_j), \omega_i \varphi(x_j)) \quad (6)$$

式中:  $\mathbf{g}_r(x_j)$  表示第  $r$  个非线性主元。

将投影后的  $\mathbf{g}(x_j)$  作为样本特征值,以核函数替代点积运算,可得:

$$\mathbf{g}_r(x_j) = \varphi(x_j) \cdot \mathbf{v}_r = \sum_{i=1}^m \boldsymbol{\alpha}_i^i (\omega_i x_j, \omega_i x_j) \quad (7)$$

个体贡献率如下:

$$ICR_i = \frac{\lambda_r^i}{\sum_{r=1}^m \lambda_r^i} \quad i = 1, 2, \dots, n \quad (8)$$

累计贡献率如下:

$$CCR = \frac{\sum_{r=1}^b \lambda_r}{\sum_{r=1}^m \lambda_r} \quad (9)$$

式中:  $b$  是主成分数;  $m$  是初始因子数。

### 1.2 宽度学习

将输入矩阵映射成特征节点矩阵,经过增强变换生成增强节点矩阵,再以特征映射节点和增强节点共同作为隐层输入,同时借助伪逆寻求隐层至输出的连接关系。整个过程中,仅对隐层至输出的连接进行更新。若特征映射节点数和增强节点数无法达到要求的精度,BL可通过增量学习的方式重新对数据进行训练。宽度学习初始结构如图1所示。

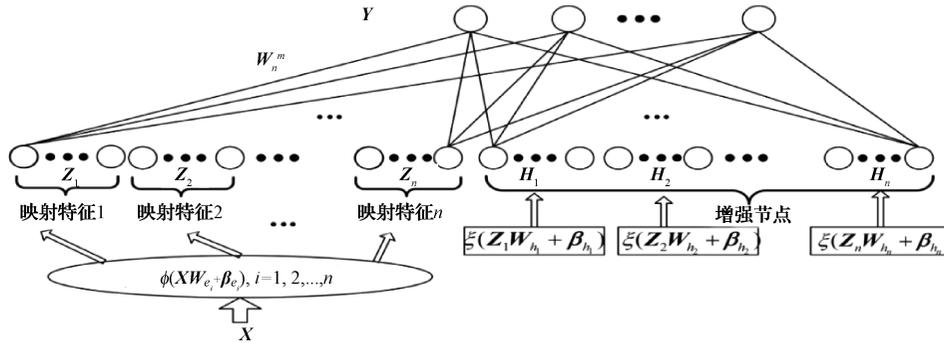


图1 宽度学习初始结构

Fig. 1 Initial structure of broad learning

令第  $i$  个特征映射  $Z_i = \Phi_i(XW_{e_i} + \beta_{e_i})$ , 第  $j$  个增强节点  $H_j = \xi_j(Z_jW_{h_j} + \beta_{h_j})$ , 定义  $Z^i = [Z_1, Z_2, \dots, Z_i]$ ,  $H^j = [H_1, H_2, \dots, H_j]$ 。对于  $n$  个特征映射,每个映射节点生成  $k$  个增强节点,如下:

$$Z_i = \Phi(XW_{e_i} + \beta_{e_i}) \quad i = 1, 2, \dots, n \quad (10)$$

宽度学习可表示为:

$$Y = [Z_1, Z_2, \dots, Z_n | \xi(Z^1W_{h_1} + \beta_{h_1}), \dots, \xi(Z^mW_{h_m} + \beta_{h_m})]W^m = [Z^n | H^m]W^m \quad (11)$$

式中:  $W^m = [Z^n | H^m]^+ Y$ ,  $[Z^n | H^m]^+$  可由伪逆岭回归算法求得,伪逆岭回归算法如下:

$$A^+ = \lim_{\lambda \rightarrow 0} (\lambda I + AA^T)^{-1} A^T Y \quad (12)$$

若在初始结构的基础上插入  $p$  个增强节点,令  $A^m =$

$[Z^n | H^m]$ , 定义  $A^{m+1} = [A^m | H^{m+1}]$ 。则  $A^{m+1}$  的伪逆如下:

$$(A^{m+1})^+ = \begin{bmatrix} (A^m)^+ - DB^T \\ B^T \end{bmatrix} \quad (13)$$

其中,  $W^{m+1}$  为:

$$W^{m+1} = \begin{bmatrix} W^m - DB^T Y \\ B^T Y \end{bmatrix} \quad (14)$$

类似上述过程,可对新的特征映射节点和新的输入进行更新。更新后结构如图2所示。宽度学习算法流程如图3所示。

### 1.3 改进粒子群优化算法

对于维度为  $D$  的空间,以该空间中的粒子当做所求

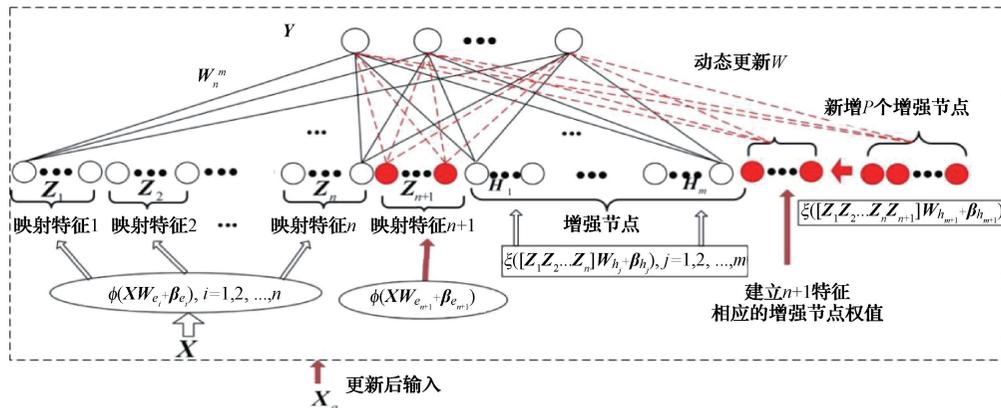


图2 宽度学习更新结构

Fig. 2 Update structure of broad learning

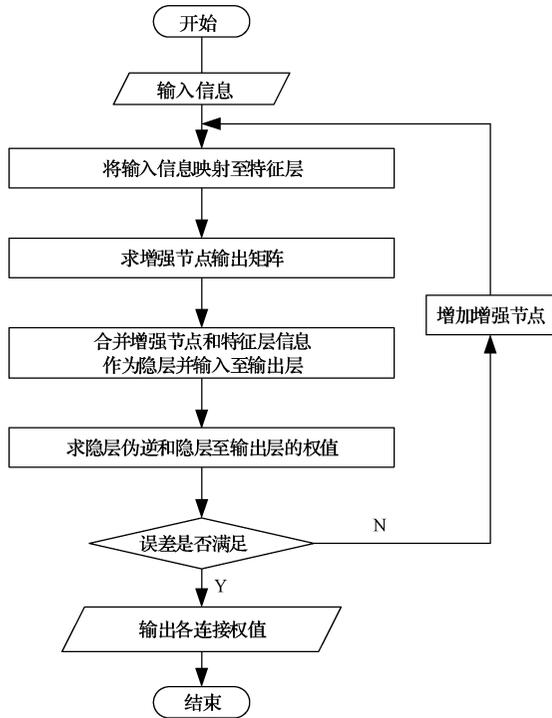


图3 宽度学习流程

Fig. 3 Flowchart of broad learning

问题的解,所有粒子都有相应的适应度和速度,以寻求问题最优解为目标,空间中的每个粒子会根据自身与最优粒子间的位置关系适时改变运动速度和当前位置<sup>[16]</sup>。

由  $n$  个粒子构成的种群记为  $X = (X_1, X_2, \dots, X_n)$ , 其中,  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$ , 运动速度记做  $V_i = (v_{i1}, v_{i2}, \dots, v_{id})^T$ , 则该粒子的位置和运动速度迭代如下:

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (15)$$

$$v_{id}(t+1) = v_{id}(t) + c_1 \times r_1() \times [p_{id}(t) - x_{id}(t)] + c_2 \times r_2() \times [p_{gd}(t) - x_{id}(t)] \quad (16)$$

式中:  $x_{id}(t+1)$  表示下一时刻粒子  $i$  第  $d$  维的速度;  $x_{id}(t)$  表示当前时刻粒子  $i$  第  $d$  维的位置;  $c_1, c_2$  为加速度参数。

PSO 算法存在收敛速度慢且准确率低的问题,为了平衡不局部和整体的关系,在式(16)中引入惯性权重  $\omega$ , 可得:

$$v_{id}(t+1) = \omega \times v_{id}(t) + c_1 \times r_1() \times [p_{id}(t) - x_{id}(t)] + c_2 \times r_2() \times [p_{gd}(t) - x_{id}(t)] \quad (17)$$

粒子的位置迭代依然按照式(15)进行。

引入惯性权重  $\omega$  是要限制粒子的搜索空间,在一定程度上弱化了最大运动速度对算法性能的影响,粒子的搜索方式由  $\omega$  决定。为了做到全局搜索和局部搜索之间的平衡,必须选择合适的  $\omega$  值。

Eberhart 等<sup>[17]</sup>采用了线性递减惯性权重,但该方法

中权重的选取过于依赖经验,可能导致算法难以收敛;董丽丽等<sup>[18]</sup>提出了步长较小的线性递减权重,但该方法收敛较慢;万忠海等<sup>[19]</sup>选择微分递减权重,该方法也可能出现不收敛的情况;马艇等<sup>[20]</sup>提出非线性微分更新权重,该方法虽然能够收敛,但计算量较大。因此,权值如下:

$$\omega = \frac{(\omega_{\max} - \omega_{\min}) \times (s - i)}{s} + \omega_{\min} \quad (18)$$

粒子的速度迭代公式如下:

$$v_{id}(t+1) = \left[ \frac{(\omega_{\max} - \omega_{\min}) \times (s - i)}{s} + \omega_{\min} \right] \times v_{id}(t) + c_1 \times r_1() \times [p_{id}(t) - x_{id}(t)] + c_2 \times r_2() \times [p_{gd}(t) - x_{id}(t)] \quad (19)$$

以上优化后的算法为标准 PSO 算法。

此外,为了避免寻参过程陷入局部最优,对上述标准 PSO 算法再次改进,利用正弦函数因子  $\sinh$  使粒子位置更新过程中获得振动而扩大搜索范围,从而增大寻优概率。粒子的当前位置更新公式如下:

$$x_{id}(t+1) = x_{id}(t) \times (1 - \sinh) + v_{id}(t+1) \times \sinh \quad (20)$$

改进后算法为正弦因子的粒子群算法(TFPSO),其流程如图4所示。

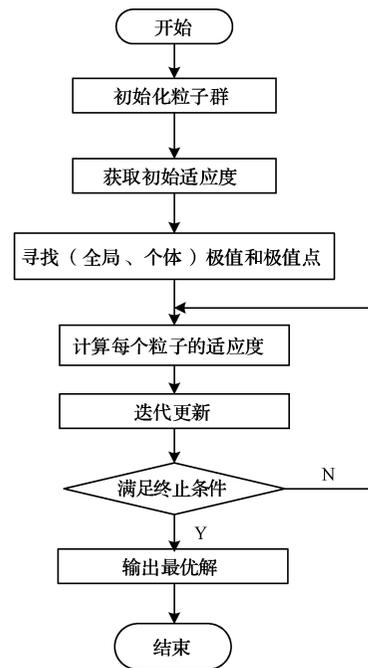


图4 TFPSO 流程

Fig. 4 Flowchart of TFPSO

#### 1.4 基于粒子群优化的宽度学习

为了提高标准粒子群优化算法在泥石流灾害预测中的准确度和可靠性,本文使用 TFPSO 对 BL 的参数进行优化,以确保 BL 的预测效率和精度并改善其性能表现。KPCA-TFPSO-BL 流程如图5所示。

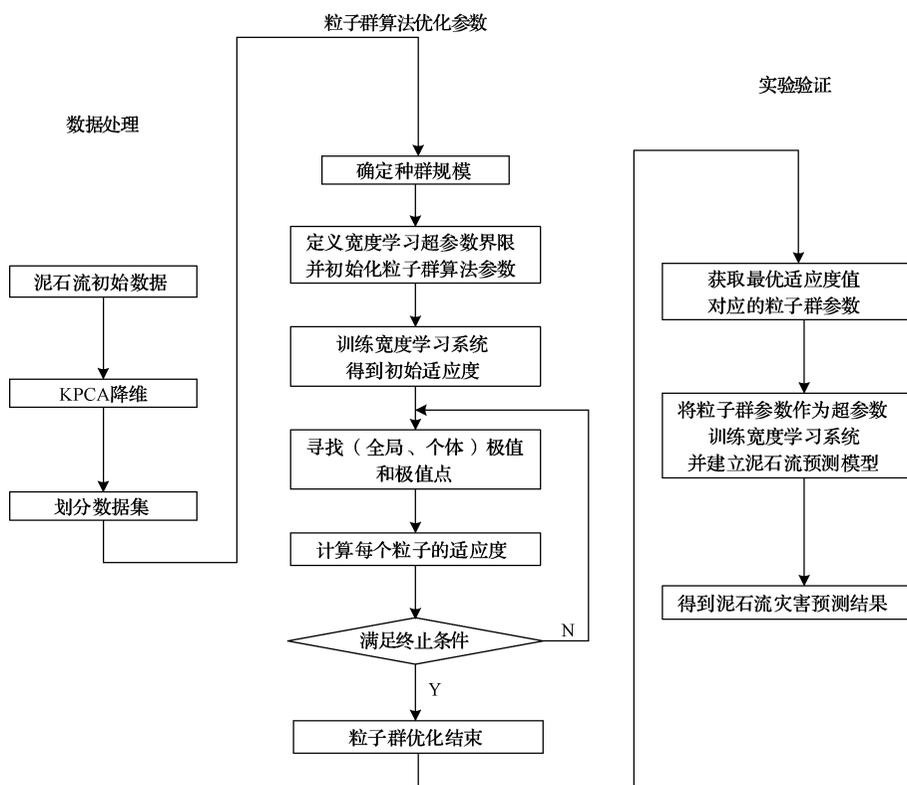


图5 KPCA-TFPSO-BL 流程  
Fig. 5 Flowchart of KPCA-TFPSO-BL

## 2 基于 KPCA-TFPSO-BL 的泥石流灾害预测模型

基于 KPCA-TFPSO-BL 泥石流灾害预测模型的具体实现过程如下。

- 1) 对获取到的初始数据进行处理, 然后使用 KPCA 核主成分分析筛选出满足需求的致灾因子。
- 2) 划分数据集, 确定合理的训练集、测试集和验证集比例。
- 3) 初始化寻优参数及 BL 参数。
- 4) 根据步骤 1) 中提取的影响因子建立泥石流预测模型, 同时利用训练数据集中的数据训练最佳适应度函数。
- 5) 通过测试集数据对不同预测模型的预测效果进行分析, 评估出最优模型并得出其预测结果。
- 6) 通过验证集数据对步骤 5) 中评估出的最优模型预测效果进行验证。

## 3 实验分析

### 3.1 研究区概况及数据来源

陕西山阳中村镇地处秦岭山麓, 山地错落沟壑纵横, 周围多为中低山地, 土石量高达近  $2 \times 10^6 \text{ m}^3$ , 山地占总面积 4/5 以上。因所处山区高差较大, 加之该地区河流隶属汉江流域, 支流众多, 水源充足。尤至夏秋时节, 降雨量

丰富, 年均雨量可达 700 mm 左右, 如遇长时连续降雨, 极易造成土质疏松而引发泥石流等地质灾害<sup>[21]</sup>。

依据《泥石流灾害防治工程勘察规范》<sup>[22]</sup>, 结合山阳县当地实地勘探和泥石流监测数据, 对泥石流成灾机理及孕育演化特征进行了研究。该区域以泥石流形成三要素中的水源和物源条件可将其划归为崩塌型泥石流, 崩塌、滑坡等灾害主要为泥石流的形成提供固体物来源。同时考虑到该区域的地质勘察数据和监测点的传感器采集数据, 选取出 16 个初始影响因子, 各因子参数如表 1 所示。

由于泥石流的暴发受前期雨量、24 h 最大雨量、1 h 最大雨量和 10 min 最大降雨等多因素影响, 因此综合考量这些因素, 最终选取暴雨强度  $R$  作为模型参数之一,  $R$  的计算公式如下:

$$R = K \left( \frac{H_{24}}{H_{24(D)}} + \frac{H_1}{H_{1(FD)}} + \frac{H_{1/6}}{H_{1/6(D)}} \right) \quad (21)$$

式中:  $K$  为前期雨量修正系数,  $K = 1$  表示前期雨量为 0,  $K > 1$  表示有前期雨量;  $H_{24}$  为 24 h 最大降雨量, 其临界值取为  $H_{24(D)} = 30 \text{ mm}$ ;  $H_1$  为 1 h 最大降雨量, 其临界值取为  $H_{1(FD)} = 15 \text{ mm}$ ;  $H_{1/6}$  为 10 min 最大雨量, 其临界值取为  $H_{1/6(D)} = 6 \text{ mm}$ 。

监测数据易受环境变化的影响, 如出现丢失、逸出或维度不一等影响, 这些异常数据都可能对模型预测结果造成较大影响, 因此将这些数据输入模型前有必要对原始数

表1 影响因子及表征意义

Table 1 The impact factors and significance

| 序号 | 影响因子    | 表征意义                |
|----|---------|---------------------|
| 1  | 暴雨强度    | 降雨集中程度,灾害形成的水源条件    |
| 2  | 孔隙水压力   | 土壤地下水的压力            |
| 3  | 土壤含水率   | 土壤所含水分的量            |
| 4  | 相对高差    | 两点间海拔高度差,提供灾害形成的势能  |
| 5  | 岩性因素    | 反映岩石特征,影响流域产沙量      |
| 6  | 植被覆盖率   | 影响流域产沙产水量           |
| 7  | 流域面积    | 反映地表汇流的因素之一         |
| 8  | 沟床比降    | 河床落差与长度之比,提供灾害形成的势能 |
| 9  | 沿沟松散物储量 | 灾害形成的物源条件           |
| 10 | 流域发育程度  | 反映流域地貌的演化阶段         |
| 11 | 流域完整系数  | 反映流域地表径流的汇流条件       |
| 12 | 冲淤变幅    | 泥沙影响河床的幅度           |
| 13 | 补给段长度比  | 反映流体沿程泥沙补给情况        |
| 14 | 区域构造影响  | 差异性抬升形成区内特殊地质环境     |
| 15 | 沟槽横断面   | 反映流体通畅情况            |
| 16 | 山坡坡度    | 流体势能转变为动能的条件        |

据进行处理。此处参考文献[23]的方法依次对以上提及的异常数据做出相应的缺失值处理、离群值处理和归一化处理。

### 3.2 影响因子筛选

鉴于样本集中各数据间的相关性会对模型预测结果的准确性造成较大干扰,利用前文所述的 KPCA 方法对 16 个初始影响因子进行筛选。筛选出的主要影响因子贡献率和累计贡献率如图 6 所示。

由图 6 可知,采用 KPCA 分析出的前 6 个主成分的累积贡献率分别为 38.19%、56.83%、68.45%、79.25%、88.20%和 95.52%。此外,第 6 个主成分的累积贡献率已高于 95%,达到 95.52%。所以,当应用 KPCA 方法进行主成分筛选时,前 6 个分量可以全面反映原始指标体系,并将初始的 16 维数据缩减为 6 维数据。

### 3.3 性能分析

对于本文提出的改进粒子群优化算法中粒子当前位置更新公式中的正弦函数  $\sinh$ ,自变量定义域越宽,曲线

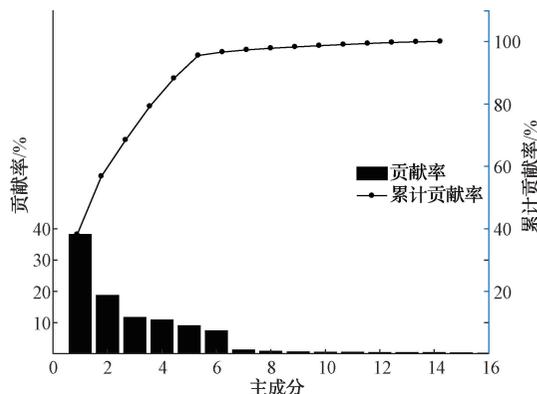


图6 KPCA分析结果

Fig. 6 The analysis result of KPCA

的极值数就越多,表明函数值的波动越强,则粒子更可能达到最优解,换句话说,抑制局部最优的可能性也就越高。正弦函数在 100 次迭代后可得到 100 个对应的函数值,迭代数与其函数值的关系如图 7 所示。为了进一步研究当粒子位置发生改变时函数的振动特性,分别以 Sphere 函数在 PSO 和 TFPSO 中粒子位置变化为对象进行实验分析,结果如图 8 所示。

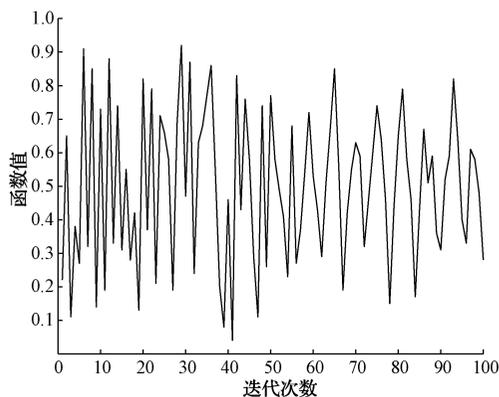


图7 正弦函数振动

Fig. 7 Vibration of the sine function

迭代过程中自变量的取值为随机产生,图 8 表明,在生成随机自变量的过程中,函数值会发生较大幅度的振动。由图 8 可以看出,引入  $\sinh$  因子后的改进 PSO 算法粒子位置范围大幅拓宽,因此该实验可以验证,引入  $\sinh$  确实能够使粒子在一定范围内产生较大程度的振动,有利于拓展解空间。

### 3.4 仿真实验

以陕西省山阳县内地质灾害多发区域为研究对象,将当地地形地貌、地理环境因素以及传感器采集信息用于研究中提出的模型,以经 KPCA 筛选出的累计贡献率前 6 的因子作为输入,泥石流发生概率为输出进行预测。从所有数据中抽取 2 005 个作为数据样本,再选取其中 1 800 个用于模型训练,200 个数据用于模型测试,剩余 5 个用

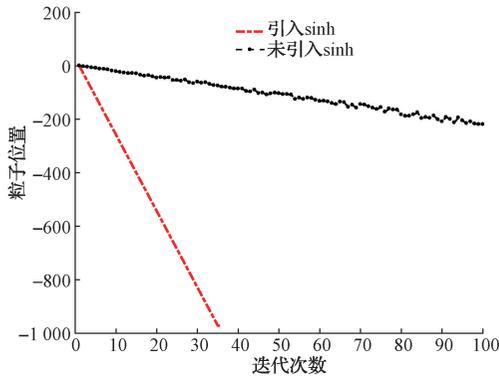


图8 粒子位置变化对比

Fig. 8 The comparison of particle positions changes

于模型验证。部分传感器数据如表2所示。

表2 传感器信息

Table 2 The data of sensors

| 样本编号  | 土壤含水率/% | 孔隙水压力/kPa | 降雨量/mm |
|-------|---------|-----------|--------|
| 1     | 5.3     | 2.1       | 10.1   |
| 2     | 5.8     | 2.2       | 13.5   |
| 3     | 5.2     | 2.5       | 20.4   |
| 4     | 6.6     | 3.2       | 21.1   |
| ⋮     | ⋮       | ⋮         | ⋮      |
| 2 005 | 41.7    | 66.5      | 89.6   |

采用传统模型进行泥石流预测时,考虑到输入信息之间具有较大相关性且维数过大,故采用KPCA配合BL模型并采用TFPSO算法进行优化,以期得到较准确的泥石流发生概率。TFPSO算法参数如表3所示。

表3 参数设置

Table 3 Parameter settings table

| 参数 | $\omega$ | $c_1$ | $c_2$ | max_iter | $N$ | $D$ |
|----|----------|-------|-------|----------|-----|-----|
| 设置 | 0.7      | 1.4   | 1.4   | 100      | 15  | 2   |

为了分析模型的预测效果,需要与其他模型进行性能比较。分别选取回归预测中常用的平均绝对误差(MAE)、均方根误差(RMSE)两个指标作为衡量模型预测性能的标准,各指标分别如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y(i) - \bar{y}(i)| \quad (22)$$

$$RMSE = \left( \frac{1}{n} \sum_{i=1}^n (y(i) - \bar{y}(i))^2 \right)^{\frac{1}{2}} \quad (23)$$

式中:  $y(i)$  表示实际值;  $\bar{y}(i)$  代表预测值;  $n$  为样本数。

将选取的1800组数据输入搭建的KPCA-TFPSO-BL模型进行训练,采用测试样本预测泥石流灾害发生情

况,为了验证KPCA降维的必要性,将经KPCA处理后的数据和未经KPCA处理的数据依次输入模型,根据预测情况进行比较,结果如图9所示。

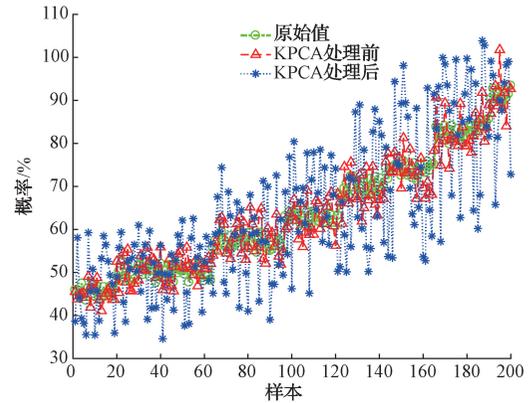


图9 降维效果对比

Fig. 9 Comparison of dimensionality reduction effects

从图9可以明显看出,KPCA处理后的预测值比KPCA处理前整体上更接近实际泥石流概率值。从KPCA降维前后的数据均方根误差来看,KPCA处理后的预测值均方根误差为3.52,而KPCA处理前的预测值均方根误差为11.09,表明经KPCA处理后的模型预测准确性更高,效果更好,采用KPCA对输入数据进行降维是很有必要的。

接下来本文提及的所有模型的输入均为经KPCA处理后的数据,为了语言描述简洁,均省略了KPCA的前缀,如KPCA-TFPSO-BL直接表示为TFPSO-BL。

利用训练集分别训练TFPSO-BP、TFPSO-BL和TFPSO-SVM3种模型,比较3种模型的训练误差RMSE,结果如图10所示,不同训练样本数对应的测试样本与灾害概率关系如图11所示。

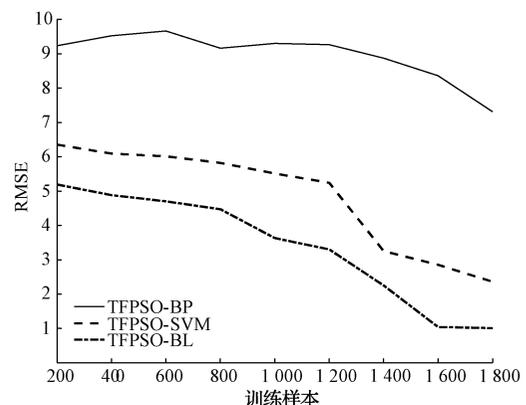


图10 3种模型RMSE

Fig. 10 RMSE for three models

图10中,TFPSO-BP、TFPSO-SVM和TFPSO-BL分别表示用TFPSO算法优化BP网络、支持向量机和宽

度学习 3 种模型的训练误差曲线,可以看出,当训练样本较少时,3 个模型的 RMSE 均较大,随着样本数量的扩充 RMSE 均呈逐渐减小的趋势。当训练样本增加至 1 800 个时,TFPSO-BL 的误差已经开始趋于稳定且数值较小,而 TFPSO-BP 的 RMSE 值较大,TFPSO-SVM 的 RMSE 值介于二者之间且未趋于稳定。说明 TFPSO-BL 模型在应对小样本方面优势比 SVM 更加明显,对于样本数据获取困难的泥石流灾害预测更为适合。

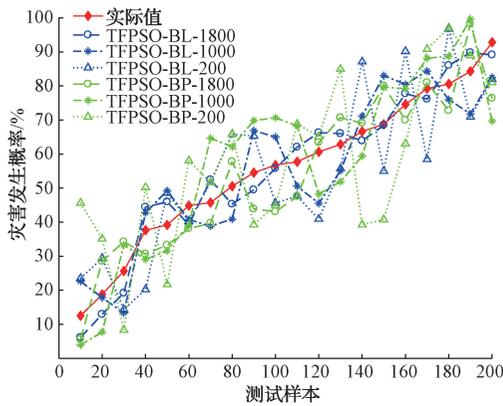


图 11 两种模型预测对比

Fig. 11 The comparison of forecasts of two models

图 11 中,TFPSO-BL-1800、TFPSO-BL-1000、TFPSO-BL-200 分别表示训练样本数量为 1 800 个、1 000 个、200 个时 TFPSO-BL 模型的预测结果,TFPSO-BP-1800、PSO-BP-1000、PSO-BP-200 分别表示训练样本数量为 1 800 个、1 000 个、200 个时 PSO-BP 模型的预测结果。从整体上看,训练集的规模越大,TFPSO-BL 的预测值越接近于实际值,而 TFPSO-BP 则表现的不如 TFPSO-BL 明显,这也是受样本数量限制所致,再次说明 TFPSO-BL 适合应对小样本数据集。同时,TFPSO-BL-1800 的预测结果比较接近实际值,也说明将 TFPSO-BL 模型用于泥石流预测是可行的。

为进一步评估该模型在泥石流概率预测中的表现,采用相同样本依次训练 BL 模型、PSO-BL 模型和 GA-BL 模型,以 MAE、RMSE 和训练时间为指标与 TFPSO-BL 模型的预测结果进行分析对比,结果如图 12 和表 4 所示。

图 12 结果表明,整体上 TFPSO-BL 的预测结果与实际值较为贴合,PSO-BL 的预测效果稍弱,GA-BL 的预测效果较弱,BL 的预测值偏离实际值最远。

从表 4 可以看出,以 RMSE 为指标进行分析,GA-BL 和 BL 的表现较为接近,二者均没有表现出优势,TFPSO-BL 的 RMSE 最小,PSO-BL 次之且和 TFPSO-BL 相差不大;以 MAE 为指标进行分析,GA-BL 和 BL 两种模型的表现和 RMSE 相似,PSO-BL 相比 TFPSO-BL 略有优势但优势非常微弱,可以忽略不计;从训练时间来看,TFPSO-BL 仅是 PSO-BL 的 14.68%,优势非常显著,综合以

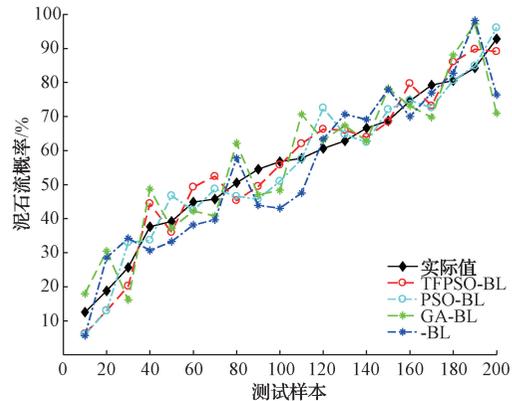


图 12 模型预测情况对比

Fig. 12 The comparison of forecast results of models

表 4 4 种模型评价指标

Table 4 The evaluation indexes of four models

| 模型       | MAE  | RMSE | 时间/s  |
|----------|------|------|-------|
| GA-BL    | 8.07 | 8.65 | 77.33 |
| PSO-BL   | 4.57 | 5.40 | 49.19 |
| TFPSO-BL | 4.60 | 3.92 | 7.22  |
| BL       | 7.73 | 9.41 | 87.26 |

上分析不难看出 TFPSO-BL 的性能最优。

通过验证样本集中 5 处泥石流沟数据对模型性能进行验证,再次对研究中提出的模型进行 10 次预测可行性实验,将其与 PSO-BL 模型的平均均方根误差和平均绝对误差进行比较,结果分别如表 5 和 6 所示。

表 5 平均均方根误差

Table 5 The mean root mean square error

| 序号 | 测试点 | PSO-BL | TFPSO-BL |
|----|-----|--------|----------|
| 1  | 杜家沟 | 3.24   | 1.34     |
| 2  | 椿树沟 | 4.03   | 2.11     |
| 3  | 杨峪沟 | 3.66   | 1.15     |
| 4  | 碾家沟 | 4.25   | 2.07     |
| 5  | 桥耳沟 | 3.88   | 0.44     |

表 6 平均绝对误差

Table 6 The mean absolute error

| 序号 | 测试点 | PSO-BL | TFPSO-BL |
|----|-----|--------|----------|
| 1  | 杜家沟 | 4.41   | 2.65     |
| 2  | 椿树沟 | 4.59   | 3.02     |
| 3  | 杨峪沟 | 3.02   | 1.98     |
| 4  | 碾家沟 | 3.28   | 2.26     |
| 5  | 桥耳沟 | 2.70   | 1.17     |

表 5 数据表明,TFPSO-BL 在 5 个测试点处的平均 RMSE 和平均 MAE 均更小,说明预测结果更接近实际情况,再次验证了研究所提模型在泥石流预测中的可行性和有效性。

#### 4 结论

本文利用 KPCA 算法结合 TFPSO 算法和宽度学习建立了基于 KPCA-TFPSO-BL 的泥石流概率预测模型,通过该模型对泥石流灾害发生情况进行预测和分析。分析了导致灾害发生的各影响因素与灾害发生可能性之间的关系,借助陕西省山阳县泥石流多发区的历史数据进行了仿真实验,同时将研究提出的模型与经典粒子群优化宽度学习的预测模型、基于遗传算法优化的宽度学习模型以及未经优化的宽度学习模型的预测效果分别进行了比较。

1) 泥石流形成过程受到水源条件、地理环境、人类活动等多种因素共同作用的影响,预测模型的建立属于严重非线性问题。本文利用 KPCA 对泥石流诱发因子的贡献率进行了排序,删除贡献率较低的因子,仅保留影响程度较高的因子;再利用参数较少的宽度学习框架搭建泥石流概率预测模型,然后通过 TFPSO 算法对模型参数进行了优化,提出了基于 KPCA-TFPSO-BL 的泥石流预测模型。

2) 通过对比 4 种模型在样本集下的误差和训练时间等指标,表明了研究中提出的模型相比其他 3 种模型优势主要表现在,首先一方面减少了模型输入矩阵的冗余度,另一方面可以有效应对各影响因子的相关性和敏感性问题,比较客观地反映出诱发因子与灾害发生概率之间的关系,最大可能减弱了研究者的主观性对模型预测效果的干扰;其次克服了传统模型参数寻优困难的缺点,能够有效缩短模型的训练时间;最后通过扩大经典 PSO 中粒子的搜寻范围,增加了其避免陷入局部最优的概率,进一步提升了模型的参数寻优能力,同时也更进一步改善了模型的预测准确性、减小了预测误差。

3) 研究所用的数据仅来自于某一时间段的单一研究区域内,因此数据具有明显的时间局限性和空间局限性,要使模型具有更广的普适性和更强的泛化性,还需通过更多不同时间段和不同研究区的数据进一步对模型进行验证,并根据结果进行优化。

4) 研究引入了 sinh 因子,通过提升粒子的振动幅度来改善搜寻范围,但迭代过程中函数自变量通过随机过程自动生成,如若初始赋值不理想可能影响算法的收敛速度,如何赋予自变量合适的初始以应对可能出现的收敛过慢问题,有待后续进一步研究。

#### 参考文献

- [1] 张晓飞, 吕中虎, 杨秀元, 等. 弱反射光栅滑坡监测系统的研究与应用[J]. 电子测量技术, 2022, 45(6): 119-123.  
ZHANG X F, LYU ZH H, YANG X Y, et al. Research and application of landslide monitoring system based on weak-reflection fiber grating sensing network [J]. Foreign Electronic Measurement Technology, 2022, 45(6): 119-123.
- [2] 王超, 申祎, 仇星, 等. 基于优化视觉词典的震后高分遥感影像震害建筑物检测[J]. 电子测量与仪器学报, 2020, 34(10): 40-50.  
WANG CH, SHEN Y, QIU X, et al. Damaged building detection based on optimized visual dictionary from post-earthquake high-resolution remote sensing images[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(10): 40-50.
- [3] 林君. 透视地球, 发展地球探测仪器, 探测地球深部奥秘[J]. 仪器仪表学报, 2016, 37(12): 2647.  
LIN J. Perspective on the Earth, development of Earth exploration instruments, and exploration of the mysteries of the deep Earth[J]. Chinese Journal of Scientific Instrument, 2016, 37(12): 2647.
- [4] 苏轶君. 基于高分遥感影像的泥石流潜在形成区识别方法探讨[J]. 中国地质灾害与防治学报, 2020, 31(5): 22-31.  
SU Y J. Probing into the techniques recognizing potential debris flow formation regions [J]. The Chinese Journal of Geological Hazard and Control, 2020, 31(5): 22-31.
- [5] ZHAO H J, WEI A H, MA F S, et al. Comparison of debris flow susceptibility assessment methods: Support vector machine, particle swarm optimization, and feature selection techniques [J]. Journal of Mountain Science, 2024, 21(2): 397-412.
- [6] DAUD H, TANOLI I J, ASIF M S, et al. Modelling of debris-flow susceptibility and propagation: A case study from Northwest Himalaya [J]. Journal of Mountain Science, 2024, 21(1): 200-217.
- [7] LI Y, XU L, SHANG Y, et al. Debris flow susceptibility evaluation in meizoseismal region: A case study in Jiuzhaigou, China[J]. Journal of Earth Science, 2024, 35(1): 263-279.
- [8] 李璐, 张永强, 李丽敏, 等. 基于松鼠觅食算法优化 LSSVM 的泥石流预测[J]. 计算机测量与控制, 2023, 31(8): 238-244.  
LI L, ZHANG Y Q, LI L M, et al. Prediction of debris flow based on squirrel foraging algorithm optimized LSSVM [J]. Computer Measurement & Control, 2023, 31(8): 238-244.
- [9] 翟淑花, 冒建, 南赞, 等. 基于遗传规划的泥石流多因子融合预测方法[J]. 中国地质灾害与防治学报, 2020, 31(6): 111-115.  
ZHAI SH H, MAO J, NAN Y, et al. Multi-factors fusion method of debris flow prediction based on genetic programming [J]. The Chinese Journal of Geological Hazard and Control, 2020, 31(6): 111-115.
- [10] 李丽敏, 张俊, 温宗周, 等. 基于布谷鸟优化轻量梯

- 度提升机的泥石流预测[J]. 科学技术与工程, 2021, 21(30): 13177-13184.
- LI L M, ZHANG J, WEN Z ZH, et al. Debris flow prediction model based on cuckoo search-light gradient boosting machine [J]. Science Technology and Engineering, 2021, 21(30): 13177-13184.
- [11] 曾鼎, 曾勇. 基于改进极限学习机的泥石流发生预测[J]. 计算机与现代化, 2020(9): 95-99.
- ZENG D, ZENG Y. Prediction of debris flow based on improved extreme learning machine [J]. Computer and Modernization, 2020(9): 95-99.
- [12] 张建伟, 雷霆, 陈二阳, 等. 降雨量贝叶斯分类器的泥石流预测研究[J]. 成都大学学报(自然科学版), 2015, 34(1): 94-96.
- ZHANG J W, LEI L, CHEN ER Y, et al. Debris flow prediction based on rainfall Bayes classifier [J]. Journal of Chengdu University (Natural Science Edition), 2015, 34(1): 94-96.
- [13] 邓玖林. 基于因子分析法的泥石流预测模型[J]. 甘肃水利水电技术, 2018, 54(7): 26-30, 42.
- DENG J L. Debris flow prediction model based on factor analysis method [J]. Gansu Water Resources and Hydropower Technology, 2018, 54(7): 26-30, 42.
- [14] 周伟, 邓玖林. 基于 Fisher 判别法的台风暴雨泥石流预测模型[J]. 水科学进展, 2019, 30(3): 392-400.
- ZHOU W, DENG J L. Predictive model of typhoon-induced debris flow based on Fisher discriminant analysis [J]. Advances in Water Science, 2019, 30(3): 392-400.
- [15] 黄志礼, 蔺红. 基于 LSTM 与 KPCA 的退役动力电池分选方法[J]. 国外电子测量技术, 2023, 42(12): 152-159.
- HUANG ZH L, LIN H. Sorting method for retired power batteries based on LSTM and KPCA [J]. Foreign Electronic Measurement Technology, 2023, 42(12): 152-159.
- [16] 张宗堂, 肖天祥, 高文华, 等. 交通荷载下煤矸石路基填料累积变形 PSO-BP 神经网络预测模型[J]. 水利水电科技进展, 2024, 44(2): 87-91.
- ZHANG Z T, XIAO T X, GAO W H, et al. PSO-BP neural network prediction model for cumulative deformation of coal gangue subgrade filler under traffic loading [J]. Advances in Science and Technology of Water Resources, 2024, 44(2): 87-91.
- [17] EBERHART R C, SHI Y. Particle swarm optimization: Developments, applications and resources [C]. Proceedings of Congress on Evolutionary Computation 2021. IEEE, 2021: 81286.
- [18] 董丽丽, 于苗, 徐淑琴. 线性递减权重粒子群优化投影寻踪模型及应用[J]. 排灌机械工程学报, 2015, 33(12): 1044-1048.
- DONG L L, YU M, XU SH Q. Application of particle swarm with linearly decreasing weight optimized projection pursuit model [J]. Journal of Drainage and Irrigation Machinery Engineering, 2015, 33(12): 1044-1048.
- [19] 万忠海, 叶生进, 郑姣. 线性微分递减自适应粒子群算法在水电站优化调度中的应用[J]. 水力发电, 2017, 43(9): 85-88.
- WAN ZH H, YE SH J, ZHENG J. Application of particle swarm optimization with linear differential decline adaptive in operation optimization of hydropower station [J]. Water Power, 2017, 43(9): 85-88.
- [20] 马艇, 刘春生, 叶青, 等. SA-PSO 算法在飞行器逃逸中的应用[J]. 飞行力学, 2014, 32(6): 570-573.
- MA T, LIU CH SH, YE Q, et al. Application of SA-PSO algorithm in the pursuit-evasion problem of vehicles [J]. Flight Dynamics, 2014, 32(6): 570-573.
- [21] HUANG Y, ZHANG B. Challenges and perspectives in design-ning engineering structures against debris-flow disaster [J]. European Journal of Environmental and Civil Engineering, 2022, 26(10): 4476-4497.
- [22] 国土资源部国际合作与科技司. 泥石流灾害防治工程勘查规范: DZ/T 0220—2006 [S]. 北京: 中国标准出版社, 2006.
- Department of International Cooperation and Science and Technology, Ministry of Land and Resources. Specification of geological investigation for debris flow stabilization: DZ/T 0220—2006 [S]. Beijing: Standards Press of China, 2006.
- [23] 徐根祺, 曹宁, 李璐, 等. 基于改进粒子群优化支持向量机的泥石流灾害预测模型[J]. 国外电子测量技术, 2022, 41(9): 73-81.
- XU G Q, CAO N, LI L, et al. Debris flow disaster prediction model based on improved particle swarm optimization support vector machine [J]. Foreign Electronic Measurement Technology, 2022, 41(9): 73-81.

## 作者简介

徐根祺, 副教授, 主要研究方向为机器学习、群智能算法。

E-mail: xugenqi@xjy.edu.cn