

引用格式 唐菲菲,胡佳赢,马英,等.基于递归特征消除选择的粒子群优化-支持向量机回归模型的滑坡位移预测方法[J].工业建筑,2024,54(11):50-60. TANG F F, HU J Y, MA Y, et al. A Landslide Displacement Prediction Method of Particle Swarm Optimization Combined with Support Vector Machine Regression Based on Recursive Feature Elimination Selection [J]. Industrial Construction, 2024, 54: 50-60 (in Chinese). DOI: 10. 3724/j. gyjzG23071806

基于递归特征消除选择的粒子群优化-支持向量机回归模型的滑坡位移预测方法*

唐菲菲¹ 胡佳赢¹ 马英¹ 周泽林² 王俊³ 郝亚飞⁴

(1. 重庆交通大学智慧城市学院, 重庆 402260; 2. 中国十九冶集团有限公司, 成都 610031;

3. 中信科智联科技有限公司, 重庆 400030; 4. 中国葛洲坝集团易普力股份有限公司, 重庆 401121)

摘要 在季节性降水及水库水位变化的影响下,部分山体会产生滑坡,滑坡位移的累积位移-时间曲线表现为明显的“阶跃型”动态变形特征。针对阶跃型滑坡位移,提出基于递归特征消除(RFE)算法的粒子群优化算法(PSO)-支持向量机回归(SVR)阶跃型滑坡位移预测模型,并以新铺滑坡为例展开研究。探究了滑坡位移数据的异常值剔除及缺失值填充方法,采用基于中位数法与集合经验模态分解的方法进行异常值剔除,采用基于统计学变量的方法进行缺失值填充;然后采用指数平滑法将阶跃型滑坡的累积位移拆分为趋势项和周期项。其中对趋势项位移采用傅里叶曲线进行拟合预测;对周期项位移通过基于SVR的RFE筛选出与周期项位移相关性高的影响因子,建立周期项位移预测模型,采用PSO对预测模型参数进行优化;最后,叠加周期项与趋势项位移预测结果,得到滑坡累积位移预测值,所提模型拟合优度为0.999,均方根误差为9.974 mm,平均绝对误差为7.037 mm。与网络搜索交叉验证算法-优化支持向量机模型(GSCV、SVR模型)、遗传算法优化-支持向量机模型(GA-SVR模型)对比,该模型对于突变位移的预测能力较强,适用于阶跃型滑坡中位移加速变化时期的风险预警。

关键词 滑坡;位移预测;递归特征消除;支持向量机;粒子群优化算法

A Landslide Displacement Prediction Method of Particle Swarm Optimization Combined with Support Vector Machine Regression Based on Recursive Feature Elimination Selection

TANG Feifei¹ HU Jiaying¹ MA Ying¹ ZHOU Zhelin² WANG Jun³ HAO Yafei⁴

(1. College of Smart City, Chongqing Jiaotong University, Chongqing 402260, China; 2. China 19th Metallurgical Group Corporation Limited, Chengdu 610031, China; 3. CICT Connected and Intelligent Technologies Co., Chongqing 400030, China; 4. Explosives Corporation Limited, CGGC, Chongqing 401121, China)

Abstract: Affected by the seasonal precipitation and changes in water levels of reservoirs, parts of mountains will generate landslides in which the cumulative displacement-time curve of some landslides exhibits obvious “step type” dynamic deformation characteristics. In the light of that displacement, a landslide displacement prediction model was proposed, which was combined the particle swarm optimization (PSO) algorithm with the support vector regression (SVR) based on the recursive feature elimination (RFE) algorithm. and the research was performed combined with the Xinpu landslide. Firstly, the methods of outlier removal and missing value filling for landslide displacement data were explored. Outlier rejection

* 重庆市技术创新与应用发展专项重点项目(CSTB2022TIAD-KPX0098);重庆市研究生联合培养基地资助项目(JDLHPYJD2021002);重庆交通大学研究生科研创新项目(2023S0131)。

第一作者:唐菲菲,博士,教授,主要从事三维激光扫描技术、无人机倾斜摄影技术、实景三维地理信息技术在灾害监测预警、智能交通、国土自然资源规划等方面研究,fftang80@126.com。

通信作者:胡佳赢,硕士研究生,主要从事地质灾害监测预警的研究,tsyww0001@163.com。

收稿日期:2023-07-18

tion was adopted a median-based method with ensemble empirical mode decomposition, and missing value filling was adopted a statistical variable-based method. Then, the “step type” cumulative displacement of landslides was split into the trend and period terms by the exponential smoothing method. The displacement of the trend term was fitted by Fourier curve. For the displacement of the periodic term, the influence factors with high correlation with the displacement of the periodic term were selected based on the SVR-RFE method, and the prediction model for the displacement of the periodic term was established. Subsequently, the prediction model parameters were optimized by the PSO algorithm. Finally, the predicted displacement of the period term was superimposed on the predicted displacement of the trend term to obtain the final predicted cumulative displacement of landslides, in which the goodness of fit was 0.999, the root mean square error was 9.974 mm, and the average absolute error was 7.037 mm. Comparing the proposed model with the GSCV-SVR model and the GA-SVR model, the proposed model was of a strong predictive capability for sudden displacement and was suitable for risk warning during periods of accelerated displacement changes of “step type” displacement of landslides.

Keywords: landslide; displacement prediction; PFE; SVR; PSO

滑坡灾害会导致资源破坏和经济损失。深入研究滑坡的成因及变形规律,在滑坡发生之前制定有效的预警预报措施,对于滑坡灾害的防治及风险分析和决策有着重要的指导意义。针对滑坡位移的预测,国内外学者开展了大量的研究,从最初的经验性预测,到数理统计预测,再到非线性预测及现在的综合预测,滑坡位移预测理论得到不断发展和完善。^[1]

目前,滑坡位移预测模型是将神经网络、人工智能等现代科学引入的非线性智能预测模型。Kalan-tar 等分别采用支持向量机模型、人工神经网络模型以及逻辑回归模型对不同滑坡数据集进行预测,并对比分析了模型预测精度。^[2] Shihabudheen 等在经验模态分解的基础上,建立了极限学习自适应神经模糊推理系统,对阶跃式滑坡位移进行预测。^[3] 彭令等采用核主成分分析和支持向量机模型对库岸边坡位移进行预测,并利用粒子群算法进行模型参数的优化。^[4] 周超等将极限学习机与小波分解算法相结合,建立了小波分析-极限学习机(WA-ELM)模型^[5],预测结果表明该模型预测效果较优。黄健等综合灰色预测 GM(1,1)模型、改进的自适应遗传算法(IAGA)与小波神经网络(WNN)建立了 GM-IAGA-WNN 滑坡预测耦合模型^[6],并以垮梁子滑坡为例展开位移预测研究,模型精确性较高。Zhang 等提出了时间序列-灰色模型 TGM(1,n,1,p,q)灰色模型^[7],通过在 TGM 灰色模型中加入非线性时间校正因子和遗传算法来优化背景值的权重系数,并以马家沟库岸滑坡为检验实例。

受季节性降雨和水库水位变化等因素的影响,滑坡累积位移-时间曲线表现为明显的“阶跃式”变

化特征^[8]。对于阶跃型滑坡,国内外学者将滑坡的累积位移拆分为趋势项和周期项,并分别对两项位移构建预测模型。王朝阳等通过移动平均法将累积位移进行拆分,通过多项式拟合及 CNN-LSTM 模型^[9]对趋势项及周期项位移进行预测,并以八字门滑坡为例进行验证,预测效果较好。Li 等应用小波分解法将三峡水库新滩滑坡累积位移分解为低频和高频分量,并用 Volterra 滤波模型构建预测模型^[10],结果表明经过小波分解后得到的预测结果更好,说明了解析预测方法的优势。陈曦等以白水河滑坡为例,采用经验模态分解对滑坡累积位移进行分解,并对利用改进门限自回归模型对分量单独预测^[11],最终预测效果良好。Guo 等通过变分模态分解方法(VMD)将阶跃式位移分为趋势项位移、周期项位移及随机位移,并提出一种灰狼优化器算法与反向传播神经网络算法相结合的预测模型^[12]对试验区滑坡进行了预测。Deng 等以三峡库区白水河滑坡为例,通过集合经验模态分解(EEMD)方法将滑坡位移分解为趋势项和周期项,并使用最小二乘拟合对趋势项进行预测,以基于粒子群优化的支持向量回归模型对周期项位移进行预测^[13],通过对三峡库区其他五个滑坡进行试验,验证了该方法的有效性。

然而,目前确定滑坡影响因子的方法比较单一,通常是将与预测变量相关度大的影响因子提取出来作为特征子集,由于没有与后续预测模型相结合,缺乏对影响因子与预测值之间的真正关联度的分析,导致这类方法筛选得到的影响因子并不一定是预测模型的最优自变量,从而影响模型的预测精度。

基于上述问题,提出基于递归特征消除(RFE)选择的粒子群优化算法(PSO)-支持向量机回归(SVR)的阶跃型滑坡位移预测模型。在该模型中,首先通过指数平滑法将滑坡累计位移拆分为趋势项和周期项;同时,对单调性强的趋势项位移采用傅里叶曲线进行拟合并预测;对于周期项位移,则首先采用基于SVR的RFE算法(SVR-RFE),并同步进行特征筛选和模型预测,再通过预测模型的目标函数计算特征的重要性,筛选出与周期项位移相关性高的影响因子;然后通过PSO算法对预测模型进行参数优化,完成周期项位移的预测;最后将周期项预测结果与趋势项预测结果叠加,实现阶跃型位移滑坡的累积位移预测。

1 滑坡位移预测模型构建

1.1 基于指数平滑法的滑坡位移项分解

滑坡位移的产生由滑坡体内在因素和外部环境因素共同作用,文中研究的滑坡位移诱发因素主要有两方面:一是滑坡体所受的重力、岩土性质等引起的滑坡自然位移趋势变化;二是降雨引起的周期性位移变化。

滑坡累计位移序列为具有一定增长趋势的非稳定时间序列,根据滑坡历史监测数据,可将滑坡变形运动过程视为位移随时间变化的时间序列模型。基于时间序列加法模型的相关研究成果^[14-15],可将滑坡累计位移分解为趋势性位移、周期性位移及随机性位移。由于目前的监测手段不能够对随机项位移进行很好的监测,且随机项对累计位移的预测影响较小,所以,暂不考虑该项。因此,可将滑坡累积位移分解为趋势项和周期项进行研究,其表达式见式(1):

$$y_t = p_t + q_t \quad (1)$$

式中: y_t 为 t 时刻的滑坡位移实际值; p_t 为滑坡位移趋势项; q_t 为滑坡位移周期项。

大多数研究通常使用移动平均法^[16-17]对趋势性位移进行提取,但该方法对历史数据全部加以同等权重利用,无法对未来的波动进行很好的估计。指数平滑法^[18]是在移动平均法基础上发展起来的一种时间序列分析预测法,它通过计算指数平滑值,给予时间序列数据逐渐减弱的影响程度,即随着数据的远离,赋予逐渐收敛为零的权数,对现象的未来进行预测,其算式为:

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1} \quad (0 < \alpha < 1) \quad (2)$$

式中: S_t 为 t 时刻的平滑值; S_{t-1} 为 $t-1$ 时刻的平滑

值; α 为平滑系数,其决定了平滑水平以及对预测值与实际值之间差异的响应速度,取值范围为 $[0, 1]$ 。

1.2 趋势项位移预测

滑坡趋势项位移随时间表现为单调递增曲线,可以采取傅里叶曲线拟合对趋势项位移进行预测。傅里叶曲线拟合^[19]主要是采用一对正交函数对离散数据进行统计分析,并通过傅里叶级数逼近实现曲线拟合,其保真度高,拟合度好,表达式如式(3):

$$y(t) = a_0 + \sum_{i=1}^n a_i \cos \omega_i t + \sum_{i=1}^n b_i \sin \omega_i t \quad (3)$$

式中: a_0 、 ω 为常系数; a_i 与 b_i 为各级系数; n 为拟合阶数。

1.3 基于RFE的PSO-SVR周期项位移预测模型

采用基于RFE算法的SVR模型来对降雨特征因子进行筛选,并用PSO对模型参数进行优化,从而精准预测滑坡周期项位移。步骤为:1)对降雨因子进行分析变换,构建初始影响因子集合;2)基于预测模型本身,通过以SVR为基学习器的RFE算法对初始影响因子进行重要性排序,得到最优影响因子集合;3)利用PSO算法对模型参数进行优化,构建周期项位移预测模型。

1.3.1 SVR-RFE算法

支持向量机回归(SVR)^[20]是美国学者Vapnik提出的一种用于非线性问题的回归预测方法。其主要思想是:将所有的样本点整理到一个超平面上,以样本点离超平面的距离最小为目标,寻找最优的回归超平面。SVR采用一个误差控制函数 ε (不敏感损失函数),当模型输出值 $f(x)$ 与真实值 y 之间的偏差大于 ε ,才会计算模型损失。在不敏感损失函数 ε 的控制下,通过引入惩罚参数 C 和松弛因子 ξ_i ($\xi_i > 0$)来求解超平面。该模型以统计学习理论为基础,结构简单,分类准确率高,泛化能力强。SVR预测模型算式为:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (4)$$

式中: $K(x, x_i)$ 为核函数; b 为偏置量。

递归特征消除(RFE)^[21]是典型的wrapper包裹式特征选择方法。其原理是反复构建机器学习模型,按一定的特征评价准则对特征进行排序,每次迭代删除一个重要性最低的特征,越早被删除的特征被认为重要性最小,排序靠后,反复进行此过程,直至最后一个特征,由此实现特征因子的排序,得到最优特征子集。其主要用于降低特征维度,消除特征之间

的冗余。由于 RFE 算法与 SVR 本身结合得较好,且这样构造出的特征子集与支持向量机密切相关,可以带来较高的学习能力,因此,将两者结合,构建基于支持向量回归的递归特征消除算法(SVR-RFE)对影响因子进行排序以得到最优影响因子组合。

以线性 SVM 为例,它是以特征空间上每个特征对目标函数的影响力大小作为排序的系数,若超平面表示为 $y=wx+b$,则 SVM-RFE 特征筛选模型就是按照超平面上每个特征因子的权重 w_i 大小进行排序,特征因子的权值越大则它对目标函数的影响越大,所具有的判别信息就越多,对超平面的贡献越大。以此对所有的特征因子进行排序,算法每迭代一次就会删除 $\|w_i\|$ 最小的那个特征,然后重新迭代,直至所有特征都排序完成。算法流程如下:

输入:训练样本集 $X_0 = [x_1, x_2, \dots, x_i]^T, y = [y_1, y_2, \dots, y_i]^T$ 。

初始化:初始化一个空的特征排序表 $r = [r_i]$, $r_i \in \Phi$,用于存储最终的特征重要性排序。 $s = [1, 2, \dots, d]$ 为初始化当前特征序列索引, d 是特征的总数, s 包含了所有特征的索引。通过循环迭代,直到 s 为空列表,即所有特征都被评估并排序。

1) 获取当前训练样本 $X = X_0(\text{rand}(i):, s(-i))$, $\text{rand}(i)$ 指随机行。

2) 训练 SVR 回归器: $\text{SVRtrain}(X, y)$, $X = X_0(\text{rand}(i):, s(-i))$ 。

3) 计算排序系数: $\text{Rank}(i)$ 。

4) 找出最小排序系数的特征 $f = \text{argmin}(\text{Rank}(i))$ 。

5) 更新特征排序表 $r = [s(f), r]$ 。

6) 删除排序系数最小的特征 $s = s(1:f-1, f+1: \text{length}(s))$ 。

输出:特征排序表。

1.3.2 PSO-SVR 算法

采用径向基函数作为 SVR 的核函数,惩罚参数 C 及核函数参数 γ 是决定预测模型好坏的关键。粒子群优化算法(PSO)^[22]是一种群体智能优化算法,通过迭代寻找最优解,具有强大全局搜索能力。因此可采用粒子群优化算法寻找 SVR 的最佳参数,构建 PSO-SVR 模型。

使用 PSO 算法进行 SVR 参数调优的步骤如下:

1) 初始化解空间。确定惩罚参数 C 及核函数参数 γ 选取范围,种群规模 a ,粒子的初始速度 v_i ,初始位置 x_i ,学习因子 c_1, c_2 ,权重因子 ω 以及迭代

次数 M 。

2) 定义目标函数。在 SVR 预测中,通常使用均方误差或均方根误差作为目标函数。

3) 初始化粒子群的位置和速度。每个粒子代表一组参数,粒子的位置表示参数的值,速度表示参数的变化方向和速率。初始化时随机生成一组粒子,并将其位置和速度随机分配在参数范围内。

4) 计算每个粒子的适应度。将每个粒子的位置作为 SVR 模型的参数,运行模型并计算相应的目标函数值。

5) 更新粒子的位置和速度。根据每个粒子的适应度和全局最优粒子的位置,更新每个粒子的速度和位置。并与历史值进行比较确定个体极值 p_{best} ,将当前的个体极值 p_{best} 与全局极值 g_{best} ,进行比较,更新 p_{best} 与 g_{best} 值;式(5)、(6)分别为速度更新算式、位置更新算式。

$$v_{id}^k = w_i v_{id}^{k-1} + c_1 r_1 (p_{\text{best}_{id}} - x_{id}^{k-1}) + c_2 r_2 \cdot (g_{\text{best}_{id}} - x_{id}^{k-1}) \quad (5)$$

式中: $v_{id}^{k-1}, x_{id}^{k-1}$ 分别为第 i 个粒子在第 d 维空间中第 k 次迭代的速度和位置; r_1, r_2 分别为 $[0, 1]$ 上的随机数。

$$x_{id}^k = x_{id}^{k-1} + v_{id}^k \quad (6)$$

6) 重复步骤 4) 和 5),直到更新后的全局最优位置满足设定的最小误差或者达到最大迭代次数,终止计算。

2 滑坡位移预测

2.1 研究区概况

新铺滑坡位于重庆市奉节县长江南岸斜坡地带,下游距三峡大坝 182 km,整体斜长约为 2 km,相对高差为 634.4 m,斜坡坡度大约在 $15 \sim 20^\circ \text{C}$ 。整体地形是以构造侵蚀剥蚀作用为主的低中山河谷地形。由于滑坡的多期活动,整个斜坡坡面形成多级台地地貌,按当地自然名称从上至下依次为(175 m 水位以上):大坪台地、上二台台地、下二台台地。滑坡位置及滑坡全貌如图 1 所示。研究区多年平均降雨量为 1 145 mm,主要集中在 7—9 三个月,约占全年降雨量的 42%。地下水主要以松散岩类孔隙水和基岩风化裂隙水为主,水文地质特征明显。

2.2 监测方案及数据分析

根据新铺滑坡的实际情况,整体根据三纵五横剖面布置监测设施,采用 GNSS 及雨量监测设备对新铺滑坡进行监测,在滑坡的三条横剖面共布设

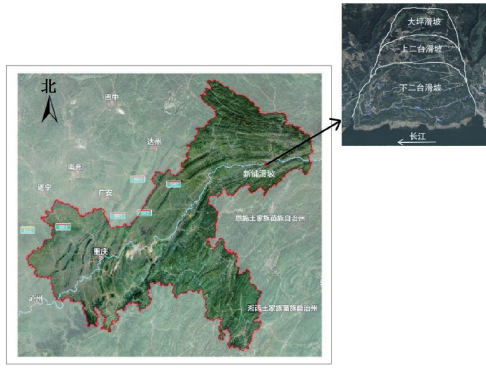


图1 新铺滑坡位置及滑坡结构示意图

Fig.1 A schematic diagram of the location of Xinpu landslide and landslide structure

9个GNSS自动监测点:GNSS01~GNSS09,用以监测地表位移;在滑坡中部易于保护的地方安装一套雨量自动监测设备,全天候监测滑坡区域的降雨量情况。监测设备布置如图2所示。

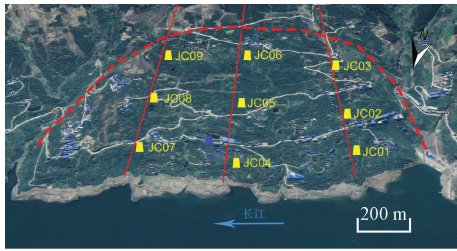


图2 新铺滑坡监测点布控

Fig.2 Arrangements of observation points for Xinpu landslide

经监测的2014年7月—2018年7月该滑坡区域的点位日位移-降雨量关系如图3所示(由于GNSS09点位未出现明显变形,图中没有显示)。

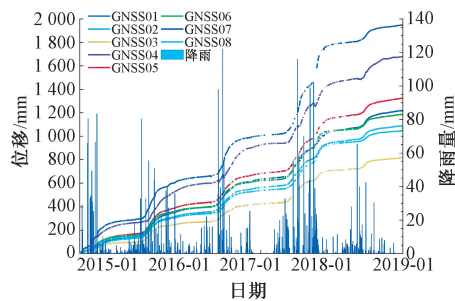


图3 各监测点日位移与降雨量关系

Fig.3 Relations between daily displacement and rainfall at each observation point

各点位累积位移均呈现比较明显的“阶跃型”增长,其中GNSS07监测点变形较大,且该点位对降雨的反应更为敏感,因此,以点GNSS07为研究对象,选取2014年7月3日—2018年12月20日每天为单位的Xinpu滑坡累积位移时间序列及降雨量数据进行研究。滑坡位移预测流程如图4所示。

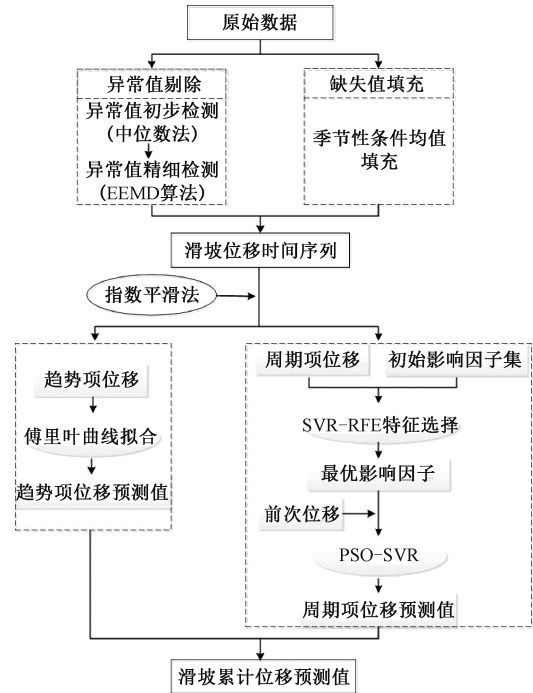


图4 滑坡位移预测流程

Fig.4 The flow chart for displacement prediction of landslide

2.3 数据预处理

2.3.1 基于中位数与集合经验模态分解的组合异常值剔除

由于传感器本身误差、人为错误等导致数据集中存在异常值。异常值的剔除主要是针对滑坡位移时间序列数据中所存在的噪声、离群点进行检测。常见的异常值检测算法主要有基于统计、聚类、距离以及密度等异常值检测算法。这些方法都是从数据全集考虑,删除超出判断标准的离群点。对于阶跃型滑坡,一般很难发现位移变形较小时期的局部异常值。因此,将统计分析方法与集合经验模态分解(EEMD)结合,提出了基于中位数与EEMD组合的方法,从整体和局部上检测滑坡数据的异常值。步骤如下:

1) 基于中位数法的异常值初步检测。由于中位数是通过排序得到的,不受数据中极值的影响,从而在一定程度上可以作为分布数列的代表,因此选取中位数作为阶跃型滑坡数据离群异常值的筛选标准。记日位移时间序列为 $\{a_i\}$,序列 a_i 的中位数为 $\text{Med}\{a_i\}$,序列 a_i 中每个数与中位数 $\text{Med}\{a_i\}$ 的比值记为 ε_i 。

$$\varepsilon_i = \frac{a_i}{\text{Med}\{a_i\}} \quad (7)$$

若 $\varepsilon_i > 1$, ε_i 越大,则该日位移数据为过分大异常值的可能性越大;若 $\varepsilon_i < 1$, ε_i 越小,则该日位移数据

为过分小异常值的可能性越大。结果显示:大部分 ε_i 处于 0.01~20, 因此将阈值设置为中位数的 0.01~20 倍, 删除超出阈值的数据, 构成新序列 $\{b_i\}$ 。

2) 基于 EEMD 算法的异常值精细检测。首先,

对得到的新序列 $\{b_i\}$ 进行分解, 取参数添加噪声序列数目 $N=2\ 000$, 最终标准偏差为 0.2, 从高频到低频依次得到 8 个本征模函数和 1 个余量(图 5), 并以波动频率从高到低依次排列。

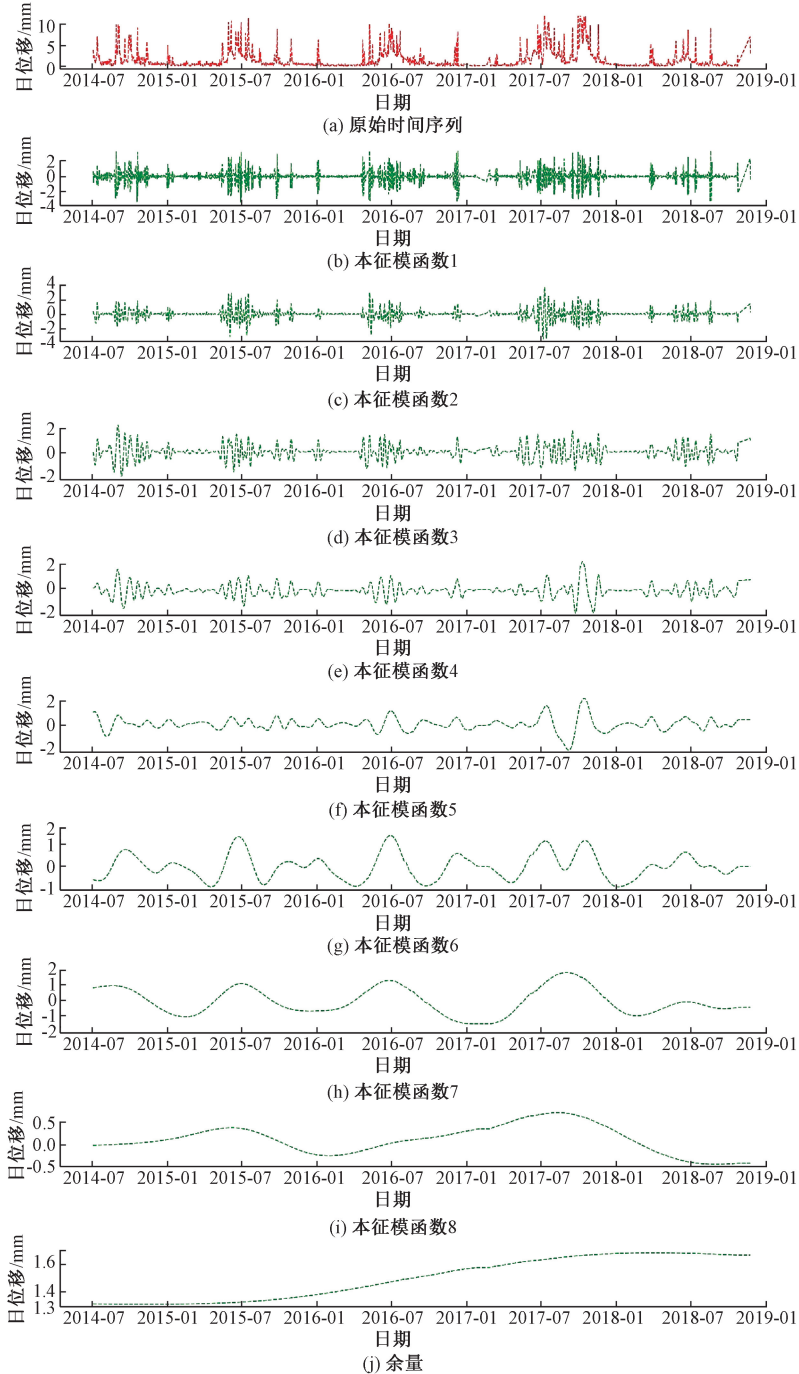


图 5 原始时间序列的 EEMD 分解

Fig. 5 EEMD decomposition of original time series

其次, 采用相关系数法进行不同频率分量层级的划分, 对初筛后滑坡时间序列进行重构, 分离出低频分量, 并计算重构序列, $\{c_i\}$ 与初筛后序列 $\{b_i\}$ 间的偏差系数 δ_i 。

$$\delta_i = \frac{|\{c_i\} - \{b_i\}|}{\{c_i\}} \quad (8)$$

δ_i 越大, 该点的数据为异常值的可能性越大。由于实验中绝大部分的偏差比率小于 5, 于是设置

偏差阈值为 5,即偏差比率大于 5 的数据为异常值,将其剔除。异常点偏差比率示意图如图 6 所示。图中通过精细检测共删除了 6 个异常值。这 6 个异常点均出现于非雨季,该阶段滑坡位移变化较为平稳,仅通过中位数法无法从位移曲线中发现隐藏的异常点,通过 EEMD 算法可以对缓慢变形期的异常点进行精细化检测。

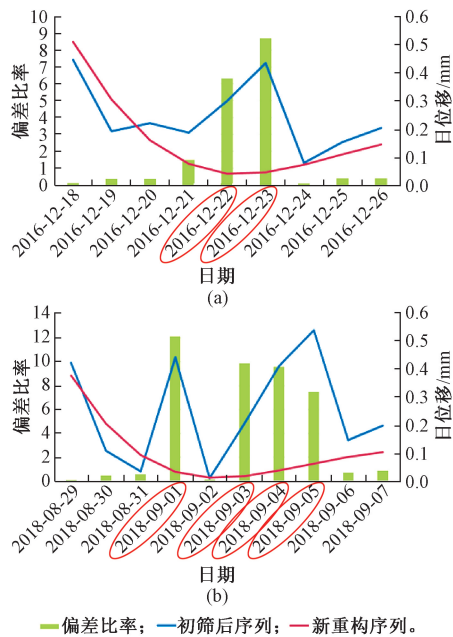


图 6 异常点偏差比率示意

Fig. 6 A schematic diagram for deviation ratios of outliers

2.3.2 基于季节性条件均值填充方法的缺失值填充

在对滑坡进行监测的过程中,难免遇到各类不确定的情况,如仪器故障、天气恶劣等,导致滑坡监测数据不完整,使得数据集时序不连续,不利于后续滑坡位移模型的建立,因此,在滑坡位移模型建立前须对缺失值进行填充。

通过分析可知:各年度的位移变化呈现一定的规律:受降雨影响的七八月间,位移数据变化较大,而未受外部因素影响的其他月份位移数据变化较为平稳。根据这一特性,将季节性因素作为条件均值填充的条件,以缺失值均值填充作为前一天的位移作为参考值,并以一年为周期,计算与前 n 个周期当天位移值的均值作为预测值进行填充。2017 年 1—9 月填充效果如图 7 所示。

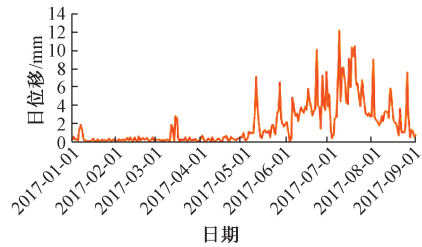


图 7 季节性均值填充效果可视化

Fig. 7 Visualization of filling effect for seasonal means

2.4 趋势项位移预测

考虑内外因素对滑坡的不同影响,运用时间序列加法模型将滑坡位移分解为趋势性位移和周期性位移。一般来说,如果时间序列波动比较平稳, α 的取值应在 0.5 以下,使模型能尽可能多地包含较长时间序列信息。鉴于监测到的数据量较大,经过试算,当 $\alpha = 0.04$ 时,趋势项位移既具有平滑性又具有分段性,因此确定该值为平滑系数进行滑坡累计位移分解。

分解出趋势项和周期项位移后,采用傅里叶曲线拟合对趋势项位移进行预测,设定拟合阶数 $n=3$,各拟合参数及精度指标计算结果如表 1 所示。经计算,趋势项位移的拟合优度 R^2 为 0.999,均方根误差为 0.535 mm,平均绝对误差为 0.492 mm,拟合效果良好。

表 1 趋势项拟合参数

Table 1 Fitting parameters of the trend term

a_1	a_2	a_3	b_1	b_2	b_3	a_0	w
-55.637 81	37.012 33	1.639 635	100.352 4	17.363 32	-9.604 428	2 277.128	0.017 21

2.5 周期项位移影响因子提取和位移预测

2.5.1 初始影响因子集构建

基于降雨对滑坡的作用规律,建立降雨影响因子集合。将日平均降雨、月平均降雨、前 5、10、15、20、30、60 d 的平均降雨,以及前 10、15、20、30、45、60 d 的累积降雨 14 个因子作为初始影响因子。其中,对前期累积降雨量的计算通常是根据 Crozier 提

出的计算式^[23]:

$$P_a = \sum_{i=1}^n K_i R_i \quad (9)$$

式中: P_a 为滑坡发生前 n 天的累积降雨量; R_i 为第 i 天降雨量; K_i 为第 i 天降雨衰减系数,可根据相关地区的具体条件以及所取得前期累积降雨时段采用,经多次测算,取 $K_i = 0.84$ 。

2.5.2 影响因子筛选

由于初始影响因子集中的影响因子之间存在线性相关性,故对初始影响因子进行筛选,以减少数据冗余。因此将采用 3 种方法对影响因子进行筛选,并将获得的 3 组影响因子集输入到相同参数设置的支持向量回归模型中进行训练,通过最终模型预测准确度来定量评价影响因子筛选方法的优劣。

1) Pearson 相关系数^[24]。Pearson 相关系数可以用来度量变量 X 与 Y 之间的相关程度,算式如式(10)所示:

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}} \quad (10)$$

Pearson 相关系数的取值范围为 $-1 \sim 1$, 越靠近 1 或 -1 , 两个因子的正相关性或负相关性越高。选取与周期项位移的相关系数 $|R| \geq 0.8$ 的高相关度影响因子: 20、30、60 d 的平均降雨。

2) 灰色关联度分析。将初始影响因子序列作为比较序列, 周期项位移序列作为参考序列。首先对其进行均值化和归一化处理; 然后计算影响因子序列 $X_i(t)$ 与周期项位移序列 $Y_0(t)$ 之间的灰色关联系数, 算式如式(11)所示:

$$\zeta_{0,i}(t) = \frac{\min_i \min_t |Y_0(t) - X_i(t)| + \rho \max_i \max_t |Y_0(t) - X_i(t)|}{|Y_0(t) - X_i(t)| + \rho \max_i \max_t |Y_0(t) - X_i(t)|} \quad t = 1, 2, \dots, N \quad (11)$$

式中: ρ 为决定灰色关联度分辨率的分辨系数, 取值范围为 $0 \sim 1$, 通常取 0.5。

最后, 计算灰色关联系数的均值。

$$\gamma_{0,i}(t) = \frac{\sum_{t=1}^N \zeta_{0,i}(t)}{N} \quad (12)$$

一般来说, 取其关联度或关联度排序的平均值作为阈值, 即以关联度 0.9 或关联度排序数 8 为标准, 根据计算结果, 选取 15 d 的平均降雨、20 d 的累积及平均降雨、30 d 的累积及平均降雨、45 d 的累积降雨和 60 d 的累积及平均降雨作为预测模型的输入变量。

3) SVR-RFE 算法。由于使用非线性核与线性核基本上结果相同, 为降低算法复杂度及运行时间, 将核函数为线性函数的支持向量回归模型作为基学习器。特征优先级排序结果为 10、1、6、7、2、4、1、5、1、8、1、3、1、9, 最优特征所在位置为第 2、7、9、11、13 位, 分别对应为月平均降雨、15、20、30、60 d 的平均降雨。对模型进行交叉验证, 交叉验证得分达到了 0.94 以上。

2.5.3 周期项滑坡位移预测

对以上三种方法筛选所得到的影响因子输入支持向量回归模型, 对 GNSS07 号点的周期项位移进行预测, 将数据集按照 9:1 的比例拆分, 以 2014 年 7 月—2018 年 7 月的数据作为训练集, 剩下的作为测试集。在对影响因子进行标准化处理的前提下, 对支持向量回归模型的参数设置均采用默认值, 设定核函数为 RBF 径向基核函数, 维度设置为 3, 惩罚参数 C 为 1, 不限制最大迭代次数。在相同参数设

置下, 三种方法预测效果可视化如图 8 所示。方法 1 为通过 Pearson 相关系数筛选的影响因子输入的 SVR 模型, 方法 2 为通过灰色关联度分析筛选的影响因子输入的 SVR 模型, 方法 3 为 SVR-RFE 模型。三种筛选方法的精度对比见表 2。

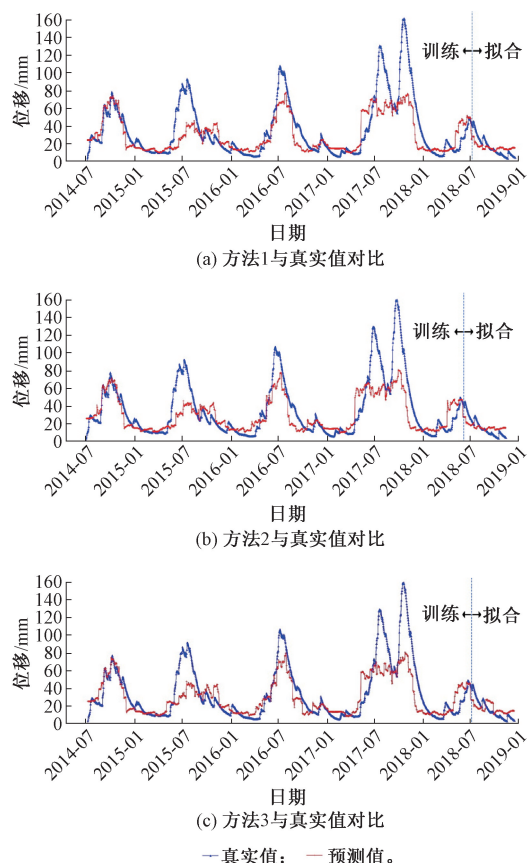


图 8 不同影响因子筛选方法的位移预测结果对比

Fig. 8 Comparisons between displacement prediction results by screening methods with different influence factors

表 2 三种筛选方法的精度对比

Table 2 Precision comparisons of three screened methods

筛选方法	均方根误差/mm	平均绝对误差/mm	拟合优度
1	20.519	12.980	0.577
2	20.557	13.168	0.575
3	19.612	12.701	0.613

从表 2 可以看出:方法 3 的三项评价指标均优于另外两种方法,即支持向量回归模型在经过方法 3 对影响因子进行筛选后,较之传统单一筛选方法 1 和 2 而言,模型精度更高。因此,可认为基于预测模型算法本身而建立起的递归特征消除算法对影响因子的筛选效果较好。

2.5.4 模型参数优化

采用粒子群优化算法对 SVR 预测模型中的惩罚参数 C 以及径向基核函数 γ 进行寻优,并与常用的网格搜索交叉验证算法(GSCV)、遗传算法(GA)进行对比分析,各优化方法的参数设置如表 3 所示,将周期项预测值与真实值进行对比,如图 9、10 所示。

表 3 三种优化方法的参数设置

Table 3 Parameter setting of three optimization methods

优化方法	参数设置
GSCV-SVR	交叉验证折数:10,迭代次数:100
GA-SVR	种群大小:100,交叉概率:0.6,变异概率:0.01,迭代次数:100
PSO-SVR	个体学习因子 C_1 :2,群体学习因子 C_2 :2,惯性权重因子:0.4,初始种群规模:100,迭代次数:100

从图 9、图 10 可以看出:使用文中方法筛选出的影响因子所构建的 SVR 预测模型,经三种超参数优化方法后,都得到了较好的预测效果,其中 PSO-SVR 模型预测曲线更加贴近实际曲线,相较于其他两种模型具有更优秀的预测性能。尤其在波动幅度最大的 2017 年 7 月—2018 年 1 月,通过粒子群优化算法优化出的预测模型表现更加出色,其预测精度最高。

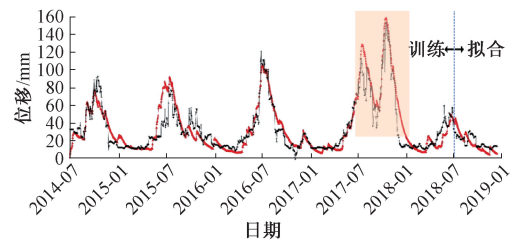
2.6 模型评价

为定量比较三种优化方法的效果,选择指标均方根误差、相关系数及平均绝对误差,来计算模型预测精度,结果如表 4 所示。

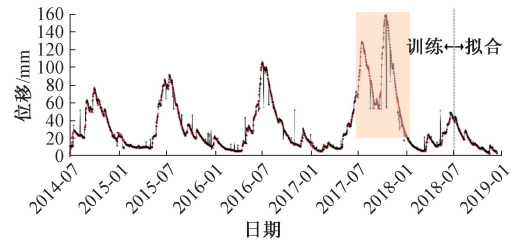
表 4 三种优化方法预测精度评价

Table 4 Evaluation on prediction precision of three optimization methods

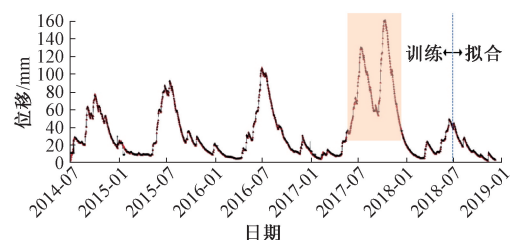
预测模型	均方根误差/mm	平均绝对误差/mm	相关系数
GSCV-SVR	12.338	8.354	0.847
GA-SVR	5.202	1.250	0.973
PSO-SVR	1.172	0.271	0.999



(a) GS-SVR 预测值与真实值对比



(b) GA-SVR 预测值与真实值对比



(c) PSD-SVR 预测值与真实值对比

— 真实值; — 预测值。

图 9 三种优化方法预测结果

Fig. 9 Prediction results by three optimization methods

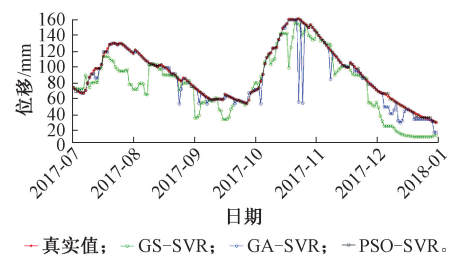


图 10 三种优化方法预测结果(细节)

Fig. 10 Prediction results by three optimization methods (in details)

由表 4 可见:基于 PSO-SVR 预测模型的均方根误差、平均绝对误差均为最小,相关系数达到了 0.995,即预测值与真实值相关性极强,表明 PSO-SVR 模型在经过 SVR-RFE 特征选择后具有很好的预测能力,对于阶跃型滑坡位移预测具有较高可靠性。

2.7 总位移预测

将趋势项位移预测值与经基于 RFE 的 PSO-SVR 预测的周期项位移预测值进行叠加,得到 GNSS07 号点位的总预测位移,如图 11、12 所示。

计算训练集及测试集的预测误差,得到新铺滑

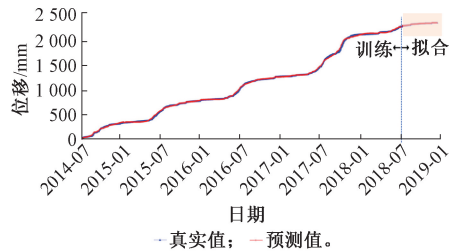


图 11 GNSS07 号点累积位移预测结果

Fig. 11 Prediction results of cumulative displacement for point GNSS07

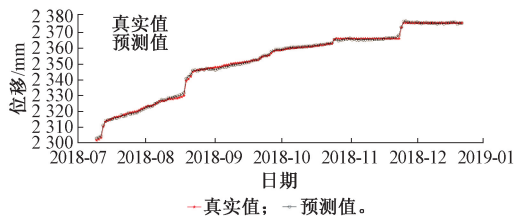


图 12 GNSS07 号点累积位移预测结果(测试集)

Fig. 12 Prediction results of cumulative displacement for point GNSS07 (test sets)

坡总位移的预测误差指标:相关系数为 0.999,均方根误差为 9.974 mm,平均绝对误差为 7.037 mm。实验结果表明:提出的基于 RFE 特征选择的 PSO-SVR 模型具有较好的位移预测和曲线拟合能力,在滑坡变形较大的情况下,仍可较为精确地对滑坡位移进行预测,适用于阶跃型滑坡位移剧烈变化时期的风险预警。

3 结束语

针对阶跃型滑坡位移难以精准预测的问题,将累积位移拆分成趋势项位移和周期项位移分别进行预测。主要成果如下:

1)提出了基于中位数法与 EEMD 结合的异常值剔除方法,该方法通过偏差可筛选出隐藏在局部的异常值,对于阶跃型滑坡的异常值检测有很好的适用性。

2)提出了基于 RFE 特征选择的 PSO-SVR 滑坡位移预测模型,对周期项位移进行预测。首先,通过 SVR-RFE 算法筛选影响因子,建立周期项位移预测模型。与采用 Pearson 相关分析以及灰色关联度分析的影响因子筛选方法相比较,结果表明:在相同参数设置下,使用 SVR-RFE 方法筛选出的影响因子所建立的模型精度更高;之后,采用 PSO 算法对预测模型参数进行优化。与基于 GSCV 算法及 GA 算法优化的支持向量回归模型相比较,PSO-SVR 模型

精度更高,预测值与真实值之间相关性更强,在阶跃型滑坡的加速变形期,该模型能够提供更为精准的位移预测值。最后,叠加由傅里叶曲线拟合预测得到的趋势项位移,得到最终的阶跃型滑坡累积位移。

模型的相关系数为 0.999,均方根误差为 9.974 mm,平均绝对误差为 7.037 mm,整体预测效果较好,对于阶跃型滑坡位移预测具有较高可靠性。

参 考 文 献

- [1] 许强,黄润秋,李秀珍. 滑坡时间预测预报研究进展[J]. 地球科学进展,2004(3):478-483.
- [2] KALANTAR B, UEDA N, AL-NAJJAR H A H, et al. A comparison between three conditioning factors dataset for landslide prediction in the Sajadrood catchment of Iran[C]//Proceedings of XXIV ISPRS Congress: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2020:625-635.
- [3] SHIHABUDHEEN K V, PILLAI G N, PEETHAMBARAN B. Prediction of landslide displacement with controlling factors using extreme learning adaptive neuro-fuzzy inference system (ELAN-FIS)[J]. Applied Soft Computing, 2017, 61: 892-904.
- [4] 彭令,牛瑞卿,赵艳南,等. 基于核主成分分析和粒子群优化支持向量机的滑坡位移预测[J]. 武汉大学学报(信息科学版), 2013,38(2):148-152,161.
- [5] 周超,殷坤龙,黄发明. 混沌序列 WA-ELM 耦合模型在滑坡位移预测中的应用[J]. 岩土力学, 2015, 36(9): 2674-2680.
- [6] 黄健,李桥,巨能攀,等. 基于主控因子分析与 GM-IAGA-WNN 联合模型的平推式滑坡位移预测研究:以垮梁子滑坡为例[J]. 工程地质学报, 2019, 27(4):862-872.
- [7] ZHANG W, XIAO R, SHI B, et al. Forecasting slope deformation field using correlated grey model updated with time correction factor and background value optimization[J/OL]. Engineering Geology, 2019, 260[2023-07-18]. <https://doi.org/10.1016/j.enggeo.2019.105215>.
- [8] 宋丽伟. 基于经验模态分解和 LSTM 模型的滑坡位移预测[J]. 人民长江, 2020, 51(5): 144-148.
- [9] 王朝阳,李丽敏,温宗周,等. 基于时间序列和 CNN-LSTM 的滑坡位移动态预测[J]. 国外电子测量技术, 2022, 41(3): 1-8.
- [10] LI Y Y, SUN R L, YIN K L, et al. Forecasting of landslide displacements using a chaos theory based wavelet analysis-Volterra filter model[J/OL]. 2019[2023-07-18]. <chrome-extension://bnjoienjhclcabnbkhhfndecoipmcdg/background/jgpdf/layout/index.html?file=https://www.nature.com/articles/s41598-019-56405-y.pdf>.
- [11] 陈曦,高雅萍,涂锐. 基于 EMD-TAR 组合模型的滑坡位移预测研究[J]. 人民珠江, 2022, 43(3): 96-101,108.
- [12] GUO Z Z, CHEN L X, GUI L, et al. Landslide displacement prediction based on variational mode decomposition and WA-GWO-BP model[J]. Landslides, 2020, 17(3): 567-583.
- [13] DENG D M, LIANG Y, WANG L Q, et al. Displacement predic-

- tion method based on ensemble empirical mode decomposition and support vector machine regression: a case of landslides in Three Gorges Reservoir area[J]. *Yantu Lixue/Rock and Soil Mechanics*, 2017, 38(12):3660-3669.
- [14] 李秀珍,许强. 滑坡预报模型和预报判据[J]. *灾害学*, 2003, 18(4):71-78.
- [15] 李强,李端有. 滑坡位移监测动态预报时间序列分析技术研究[J]. *长江科学院院报*, 2005, 22(6):22-25.
- [16] XING Y, YUE J P, CHENG C. Interval estimation of landslide displacement prediction based on time series decomposition and long short-term memory network[J]. *IEEE Access*, 2020, 8:3187-3196.
- [17] SHARIFI S, HENDRY M, MACCIOTTA R, et al. Evaluation of filtering methods for use on high-frequency measurements of landslide displacements[J]. *Natural Hazards and Earth System Sciences*, 2022, 22(2):411-430.
- [18] 罗伟沅,蒋亚楠,许强,等. GPS 滑坡位移监测时序分析与组合建模预测[J]. *防灾科技学院学报*, 2020, 22(4):20-28.
- [19] LIN H, LU Y, DING R, et al. A multi-feature fusion apple classification method based on image processing and improved SVM[J]. *Agricultural Biotechnology*, 2022, 11(5):84-91.
- [20] VAPNIK V N. The nature of statistical learning theory[M]. Berlin: Springer, 1995.
- [21] GUYON I, WESTON J, BARNHILL S, et al. Gene Selection for cancer classification using support vector machines[J]. *Machine Learning*, 2002, 46(1/2/3):389-422.
- [22] YIN Y H, JANG-JACCARD J, XU W, et al. IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset[J/OL]. *Journal of Big Data*, 2023, 10 [2023-07-18]. <https://doi.org/10.1186/s40537-023-00694-8>.
- [23] CROZIER M J, EYLES R J. Assessing the probability of rapid mass movement[C]//*Proceedings of the 3rd Australia-New Zealand Conference on Geomechanics*. Wellington: New Zealand Institution of Engineers, 1980:47-51.
- [24] QI L P. A course recommender system of MOOC based on collaborative filtering algorithm with improved pearson correlation coefficient[D]. 武汉:华中师范大学, 2019.