文章编号：1674-2974（2025）04-0149-10 DOI：10.16339/j.cnki.hdxbzkb.2025274

融合 XGBoost 和 SVR 的滑坡位移预测

王惠琴[1†]，梁啸[1]，何永强[2]，李晓娟[2]，张建良[1]，郭瑞丽[1]，刘宾灿[3]
（1. 兰州理工大学 计算机与通信学院，甘肃 兰州 730050；
2. 西北民族大学 土木工程学院，甘肃 兰州 730030；
3. 陕西建工安装集团有限公司，陕西 西安 710068）

摘 要：利用极端梯度提升与支持向量回归，同时结合猎人猎物优化算法的优势，提出了一种融合极端梯度提升和支持向量回归的滑坡位移预测模型．首先采用极端梯度提升（extreme gradient boosting，XGBoost）进行滑坡位移初步预测，进一步利用猎人猎物优化算法（hunter-prey optimizer，HPO）优化支持向量回归（support vector regression，SVR）的超参数而构建了一种组合预测模型（HPO-SVR）以修正XGBoost的预测结果．两组滑坡位移实测数据表明：HPO算法通过不断更新猎人与猎物位置的动态寻优策略，获得了更加合理的SVR的超参数．相对于XGBoost、SVR，以及其与粒子群优化算法、遗传算法和HPO的组合预测模型而言，XGBoost-HPO-SVR组合模型在阳丌山滑坡和脱甲山滑坡位移预测中取得了良好的效果，其均方根误差和平均绝对误差分别为3.505和1.357，0.550和0.538.

关键词：极端梯度提升；支持向量回归；猎人猎物优化算法；滑坡位移预测
中图分类号：P694； **文献标志码**：A

Fusion of XGBoost and SVR for Landslide Displacement Prediction

WANG Huiqin[1†], LIANG Xiao[1], HE Yongqiang[2], LI Xiaojuan[2], ZHANG Jianliang[1], GUO Ruili[1], LIU Bincan[3]
（1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China;
2. School of Civil Engineering, Northwest Minzu University, Lanzhou 730030, China;
3. SCEGC Installation Group Company LTD, Xi'an 710068, China）

Abstract: In this paper, a landslide displacement prediction model integrating extreme gradient boosting and optimized support vector regression is proposed by using extreme gradient boosting and support vector regression, and combining the advantages of hunter-prey optimization algorithm. Firstly, extreme gradient boosting (XGBoost) is used for the preliminary prediction of landslide displacement, and then hunter-prey optimizer (HPO) is used to optimize support vector regression (SVR). A combined prediction model (HPO-SVR) is constructed by optimizing

* **收稿日期**：2024-03-04
基金项目：甘肃省交通厅科研项目（2022-14），Scientific Research Project of the Transportation Department of Gansu Province（2022-14）；甘肃省重点研发计划（21YF1GA381），The Key Research and Development Program of Gansu Province（21YF1GA381）；陕西省重点研发计划-工业领域（2024GXYBXM-42），Shaanxi Provincial Key R&D Program-Industrial Field（2024GX-YBXM-42）；甘肃省科技计划项目（25JRRA043），Gansu Provincial Science and Technology Program Project（25JRRA043）
作者简介：王惠琴（1971—），女，甘肃渭源人，兰州理工大学教授，博士
† 通信联系人，E-mail：Whq1222@lut.edu.cn

the hyperparameters of SVR using HPO to correct the prediction results of XGBoost. The validation of two sets of landslide displacement measured data shows that the HPO algorithm obtains a more reasonable hyperparameter of SVR through the dynamic optimization strategy of constantly updating the positions of the hunter and the prey. Relative to the combined prediction models of XGBoost, SVR, and its combination with particle swarm optimization algorithm, genetic algorithm, and HPO, the combined XGBoost-HPO-SVR model achieves good results in predicting the displacements of Yangwashan landslide and Tuojiashan landslide, with mean square errors of 3.505 and 0.550, and mean absolute errors of 1.357 and 0.538, respectively.

Key words: extreme gradient boosting; support vector regression; hunter prey optimization algorithm; landslide displacement prediction

滑坡是最频繁且分布范围广的地质灾害之一,会给人类的生命和财产带来巨大的损失.滑坡的发生是一个极其复杂的过程,不仅受地质等内部因素的影响,也受降雨等外部因素的影响.当滑坡发生时,其表面会产生形变,而形变最直观的表现就是位移.这是滑坡内部失稳变化最直接的表征,也为滑坡预警带来了重要的参考信息.实现滑坡位移高精度的预测是预防滑坡发生的重要手段,也是滑坡预警系统中重要的组成部分[1].

有关滑坡位移预测的研究起源比较早[2].早期的滑坡位移预测主要是基于统计学的方法,具体有灰色模型[3-4]和自回归移动平均法[5]等.这些方法较为准确地实现滑坡位移预测,揭示了滑坡的内在规律.但其仅适用于监测稳定、线性的滑坡位移数据.然而,滑坡的发生是一个非常复杂的过程,其位移数据往往呈现不稳定、非线性等特点.这就使得上述方法难以对滑坡数据进行有效处理.近年来,随着基于传感器、通信和计算机等技术的智能监测设备被广泛应用于各个方面,我国在许多易发生滑坡的地方都部署了智能监测设备,这为滑坡位移预测提供了更加精准的数据.同时,人工智能技术的出现,为不稳定、非线性的滑坡位移数据预测提供了更加有效的方法.为了对滑坡位移进行有效的预测,许多学者提出了不同的模型,主要有基于深度学习[6-8]和机器学习的模型[9-11]等.但是,这些预测模型在时序监测数据不稳定时存在模型参数难以确定、预测精度不高等问题,而且训练起来耗时、复杂.另外,基于深度学习的模型还需要大样本的数据集,而实际监测的滑坡数据通常具有非线性和样本少等特点.这就使得基于深度学习的位移模型难以广泛应用于实际滑坡位移预测.

随着智能优化算法与机器学习算法等进一步发展,基于二者的组合模型在滑坡位移预测中表现出了优异的性能,已成为当前滑坡位移预测中的研究热点之一.例如,优化支持向量回归算法的组合模型[12-14]和神经网络的组合模型[15-18]等.其中,文献[12]将20种智能优化算法用于优化SVR,以此来提升模型的性能.同时指出,最优超参数是调整机器学习方法性能的关键,而且没有一个群智能优化算法能够优化所有问题.因此,在地质灾害中要不断寻找和应用新模型来提升滑坡位移预测性能.文献[17]采用随机分形搜索和多元宇宙优化器优化了人工神经网络,进一步提升了滑坡位移预测的精度.由此可见,在这些组合模型中,均是通过引入群智能优化算法优化原有预测模型的超参数来提升预测模型的性能.但是,对于不稳定滑坡数据而言,其仍然表现出预测精度不高的问题.另外,群智能优化算法自身可能也会陷入局部最优值,从而影响预测精度.鉴于此,将新型群智能优化算法与滑坡位移预测模型结合,是寻找精准滑坡位移预测模型的有效手段之一.

为了实现滑坡位移的有效预测,本文利用极端梯度提升(extreme gradient boosting, XGBoost)模型的泛化能力、鲁棒性、灵活性等优势,先采用XGBoost对滑坡位移进行初步预测.其次,以支持向量回归(support vector regression, SVR)为基础,借助寻优能力强的猎人猎物优化算法(hunter-prey optimizer,

HPO)进一步优化支持向量回归的超参数,提出了一种融合XGBoost和优化SVR的组合预测模型.最后,通过滑坡北斗实测数据实验分析了本文所提模型(XGBoost-HPO-SVR)的预测效果,并与XGBoost、SVR、以及XGBoost和SVR与粒子群优化算法、遗传算法和HPO等组合预测模型的性能进行了对比.

1　相关算法介绍

1.1　极端梯度提升(XGBoost)

XGBoost算法是集成学习算法Boosting类的代表之一,具有较高的预测精度和效率,可用于解决回归、分类问题.其实质是通过多次迭代而获得训练模型.而且,每次迭代会训练出一个弱学习器.每个弱学习器是通过拟合上一轮弱学习器的残差而得到.之后,将所有弱学习器的预测值进行加权求和,从而得到最终的预测结果.该算法的详细原理可见文献[19].

对于滑坡位移训练样本数据集$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_k,y_k)\}$而言,XGBoost算法对第$k$个样本的预测过程可表示为

$$\hat{y}_k=\sum_{n=1}^{N}f_n(x_k) \tag{1}$$

式中:x_k为第k个样本;$\hat{y}_k$代表第k个样本的预测值;$f_n(x_k)$代表第n个弱评估器对第k个样本x_k的预测结果.

为了减少数据过拟合,提升模型的泛化能力,XGBoost算法在目标函数中引入了正则项$\sum_{n=1}^{N}\Omega(f_n(x_k))$,可由式(2)计算得到.即

$$\sum_{n=1}^{N}\Omega(f_n(x_k))=\alpha\eta+\sigma\sum_{j=1}^{\eta}w_j^2 \tag{2}$$

式中:α和σ为模型复杂度的相关变量;η表示叶节点个数;w_j表示叶节点j的权重.

引入后的目标函数为

$$O_{obj}=\sum_{k=1}^{T}l(y_k,\hat{y}_k)+\sum_{n=1}^{N}\Omega(f_n(x_k)) \tag{3}$$

式中:$\sum_{k=1}^{T}l(y_k,\hat{y}_k)$为$T$个样本总的损失函数值;$y_k$代表真实值;$T$为当引入第$n$个弱学习器时的总样本数;$N$为所有弱学习器的个数.

利用XGBoost模型预测滑坡位移时,先利用前k个滑坡数据完成XGBoost模型的训练.再将第$k+1$个滑坡数据输入训练好的XGBoost模型中进行预测.最后,利用真实值和预测值间的误差来评价模型.另外,弱学习器的个数是影响XGBoost模型预测结果的关键,XGBoost模型中弱学习器个数越多,模型的学习能力就越强,但越会带来计算资源的浪费.所以,可通过实验来获得弱学习器的最佳个数,旨在取得最好的预测效果.

1.2　支持向量回归算法(SVR)

SVR算法是一种基于支持向量机的回归类算法,在处理非线性回归问题上具有独特的优势[20].它利用核函数将低维空间上线性不可分数据转化为高维空间上的线性可分数据,很好地解决了数据之间的非线性问题,而且SVR在小样本数据集上的性能优于神经网络方法.因此,SVR被广泛应用在地质灾害、工业、农业等领域[20-22].其主要思想是,利用非线性映射将样本数据映射到高维特征空间进行训练计算,从而得到最优的决策函数,旨在取得最佳的拟合效果.其过程如下:

对于给定的滑坡位移训练样本数据集$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_k,y_k)\}$,SVR的估计函数为[23]

$$f(x)=\boldsymbol{W}^{\mathrm{T}}x+b \tag{4}$$

式中:$\boldsymbol{W}$和b是待确定的参数,$(\cdot)^{\mathrm{T}}$表示转置.

假设SVR通常容许$f(x)$和y之间最多存在一个偏差ε.即,只有当二者差别的绝对值大于ε时,才计算损失.因此,当训练样本落入中心为$f(x)$、宽度为2ε的间隔内时,则认为预测正确;否则,结果错误.所以,SVR算法估计函数的求解问题可转化为

$$\begin{cases}\min\limits_{W,b,\xi_k,\mathrm{s}_k}\dfrac{1}{2}\|\boldsymbol{W}\|^2+C\sum\limits_{k=1}^{m}(\xi_k+\mathrm{s}_k),\\ \mathrm{s.t.f}(x_k)-y_k\leqslant\varepsilon+\mathrm{s}_k,\\ y_k-f(x_k)\leqslant\varepsilon+\mathrm{s}_k,\\ \xi_k\geqslant 0,\mathrm{s}_k\geqslant 0,k=1,2,\cdots,m\end{cases} \tag{5}$$

式中:ξ_k和s_k是松弛变量;C是惩罚因子;y_k为对应样本x_k的真实值.利用SVR模型预测滑坡位移时,其过程和步骤与极端梯度提升(XGBoost)完全相同,只需要通过实验选取合适的核函数和超参数,即可提高SVR模型的预测效果.

1.3　猎人猎物优化算法(HPO)

HPO是一种群智能优化算法[24],其灵感来源于

狮子、狼等猎人以及鹿和羚羊等猎物.该算法具有超强的寻优能力,目前被广泛应用于电力、轴承故障识别、交通[25-27]等领域.其主要思想是:猎人捕食移动中的种群猎物时,通常会选择一个远离群体的猎物.当猎人发现猎物后,便开始追逐猎物.在追逐过程中,猎人朝着猎物的位置不断地调整自己的位置,而猎物也会向更安全的地方移动.通常认为,猎物最安全的地方则被认为是优化问题的最优解.该算法的主要技术步骤如下[24].

首先,随机初始化种群中每个成员的位置.根据此算法的规则与策略,在搜索空间中,通过控制和引导种群来不断更新成员的位置,并使用适应度函数的计算结果来动态评估新位置的优劣.该过程会使解决问题的方案在每次迭代中得到优化,从而获取最优位置.初始化种群的位置由式(6)随机生成得到.式(6)为

$$S_i = \mathrm{rand}(1, d) \cdot \left(b_{\max} - b_{\min}\right) + b_{\min} \tag{6}$$

式中:S_i代表猎人或猎物的位置;d代表问题变量的维数;$b_{\max}$和$b_{\min}$分别代表了变量的最大值和最小值;rand(·)代表随机函数.

初始化种群位置后,HPO算法的关键在于猎人和猎物的选择及其位置的不断更新,即为

$$S_i(t+1) = \begin{cases} S_i(t) + 0.5\left[\begin{matrix}\left(2QZP_{\mathrm{pos}} - S_i(t)\right) + \\ \left(2(1-Q)Z\mu - S_i(t)\right)\end{matrix}\right], \\ \quad R_5 < \beta; \end{cases} \tag{7a}$$

$$\begin{cases} H_{\mathrm{pos}} + QZ\cos(2\pi R_4) \times \left(H_{\mathrm{pos}} - S_i(t)\right), \\ \quad R_5 \geqslant \beta \end{cases} \tag{7b}$$

(7)

式中:$S_i(t)$和$S_i(t+1)$分别代表t和$t+1$时刻猎人和猎物的位置;P_{pos}是距离均值位置μ最远的猎物位置;H_{pos}是全局最优的位置;Z是自适应参数;Q是探索和开发之间的平衡参数;R_4和R_5分别是[-1, 1]和[0, 1]区间的随机数;β是调节参数,本文设置为0.1;μ是所有位置的平均值.当$R_5 < \beta$时,依据式(7a)更新猎人的位置,否则依据式(7b)更新猎物的位置.Q、Z和μ的计算公式分别为

$$Q = 1 - \lambda\left(\frac{0.98}{\lambda_{\max}}\right) \tag{8}$$

$$\begin{cases} \boldsymbol{V} = \boldsymbol{R}_1 < Q \\ \theta = (\boldsymbol{V} == \boldsymbol{0}) \\ Z = R_2 \otimes \boldsymbol{\theta} + \boldsymbol{R}_3 \otimes (\sim \theta) \end{cases} \tag{9}$$

$$\mu = \frac{1}{\omega}\sum_{i=1}^{\omega} S_i \tag{10}$$

式中:λ代表当前迭代次数;$\lambda_{\max}$代表最大迭代次数;$\boldsymbol{R}_1$和$\boldsymbol{R}_3$是[0, 1]之间的随机向量;$\boldsymbol{R}_2$是[0, 1]之间的随机数;$\boldsymbol{V}$是与变量数相关的随机向量;$\boldsymbol{\theta}$是满足条件($\boldsymbol{V} == \boldsymbol{0}$)的向量$\boldsymbol{R}_1$的索引值;⊗为克罗内克积;$\omega$为初始种群数.

2 XGBoost-HPO-SVR组合预测模型

对于复杂多样的滑坡灾害,当只采用单一滑坡位移预测模型时,预测结果常常不够理想.为此,本文利用鲁棒性强的XGBoost算法与高效率的SVR算法的组合方式来进行滑坡位移预测.其中,超参数的选取是提高SVR模型预测精度的关键.考虑到HPO算法的寻优能力强,需要设置的参数少,本文采用HPO来优化SVR的超参数,旨在提高SVR模型用于滑坡位移的预测精度.即,利用划分的训练集数据构建SVR模型,并将SVR预测模型的均方误差作为适应度函数输入HPO算法中,适应度函数的变量值就是需要的SVR的超参数C和g.随着迭代次数的完成,最小的适应度函数所对应的变量值就是SVR模型的最优超参数.

XGBoost-HPO-SVR组合模型主要由XGBoost滑坡位移初步预测模型和HPO-SVR残差预测模型两部分组成.其中,XGBoost滑坡位移初步预测模型进行滑坡位移初步预测时,首先将滑坡位移数据划分为训练集和测试集.利用训练集合理确定模型决策树的棵数,从而完成模型训练.同时,利用训练好的模型进行初步预测,并获得残差序列.在此基础之上,利用HPO-SVR残差预测模型进一步对XGBoost模型的残差序列进行修正,从而提高滑坡位移预测的精确性.具体步骤如下:

1)首先标准化实测的滑坡位移数据,然后将其划分为训练集和测试集.

2)利用训练集数据,在给定弱学习器的个数范围后,通过交叉验证的方式,先确定XGBoost模型中弱学习器的个数,并将其代入XGBoost模型中.在此

基础之上,利用训练集来完成XGBoost模型的训练.

3)基于训练好的XGBoost模型,进行滑坡位移的初步预测,并用实际位移值与预测值相减得到残差数据集.然后对其进行标准化处理,并按比例划分为残差训练集和残差测试集.

4)初始化HPO算法,设置HPO算法的种群数和迭代次数以及寻优范围.然后利用残差训练集通过HPO确定出SVR的最优超参数C和g.并将最优超参数值代入SVR模型中,再次利用残差训练集完成SVR模型的训练,并利用残差测试集进行滑坡位移残差的预测,从而得到残差的预测结果.

5)将XGBoost初步预测结果与HPO-SVR残差预测结果相加,得到最终的滑坡位移预测结果,并利用平均绝对误差(mean absolute error, MAE)、均方误差(mean square error, MSE)、均方根误差(root mean square error, RMSE)来评价模型的预测精度.

上述1)、2)和3)步为XGBoost-HPO-SVR组合模型的初步预测部分;4)和5)步为XGBoost-HPO-SVR组合模型的残差预测部分.XGBoost-HPO-SVR组合模型的流程如图1所示.

3　XGBoost-HPO-SVR模型实例测试

为了说明XGBoost-HPO-SVR组合模型的预测效果,本文选取了甘肃省渭源县不同的滑坡体类型,即土质推移式的阳山滑坡和牵引式黄土层的脱甲山的监测数据进行了实验.并与分别采用单一的XGBoost和SVR预测模型,以及XGBoost、SVR与遗传算法(genetic algorithm, GA)、粒子群优化算法(particle swarm optimization, PSO)和HPO构建的预测组合模型(即XGBoost-PSO-SVR、XGBoost-GA-SVR、XGBoost -SVR)进行了性能对比,其结果如图4~图7和表1~表2所示.

阳山滑坡受地震、风化、洪水冲蚀等累进性破坏的影响,坡体发育冲沟、森林植被破坏,总体坡向330°,是一处欠稳定的滑坡.脱甲山滑坡位于甘肃省定西市渭源县峡城乡脱甲山村,坡高70 m,滑坡长约220 m、宽100 m,平均坡度30°,总体坡向223°.本文利用阳山滑坡和脱甲山滑坡体的监测点累计位移

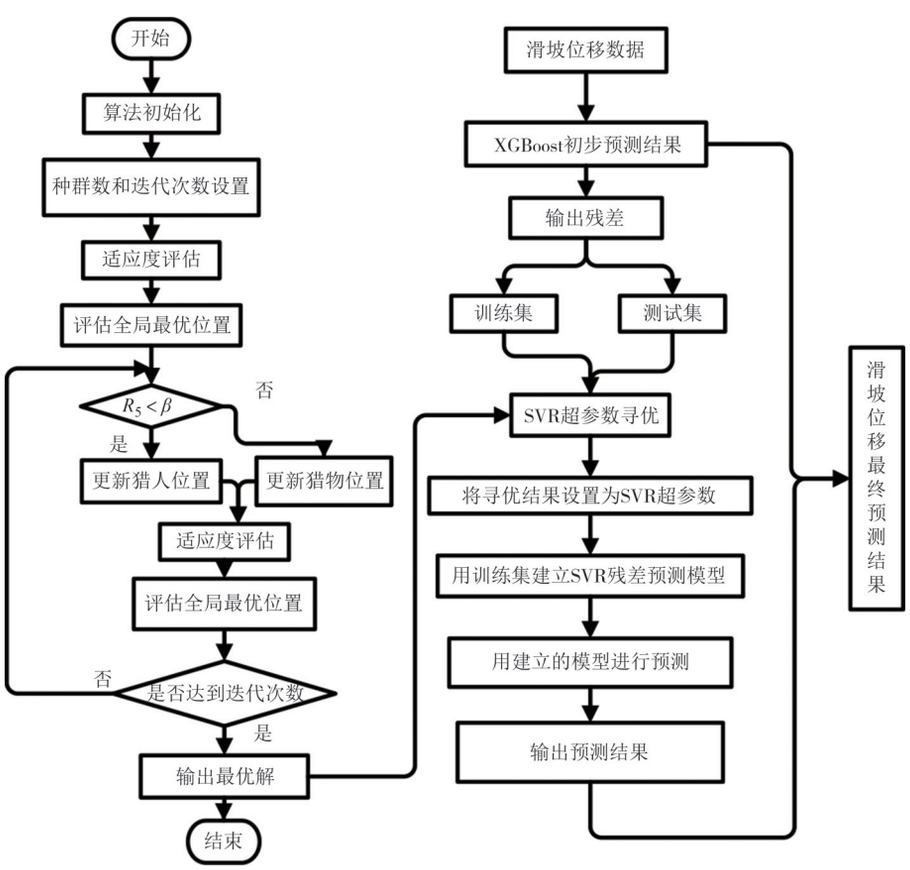
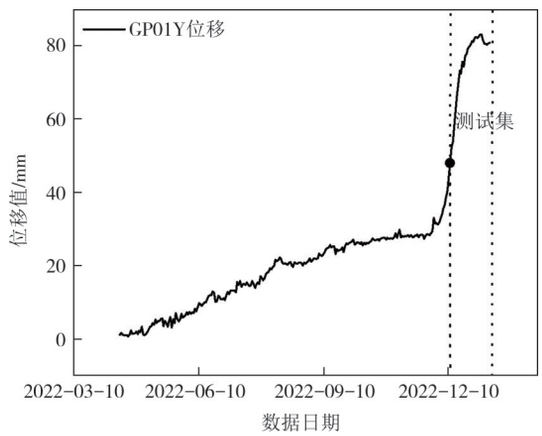


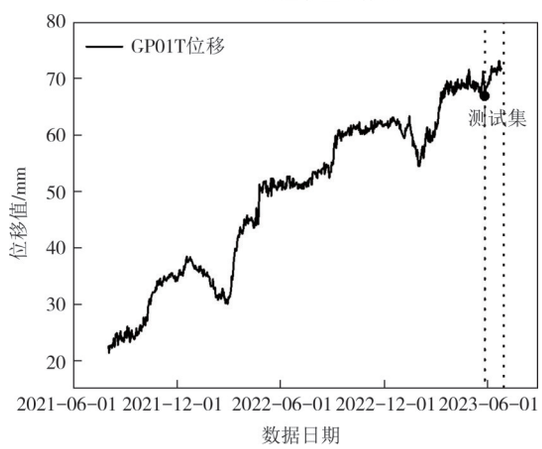
图1　融合极端梯度提升和优化支持向量回归的滑坡位移预测模型的流程

Fig.1　Landslide displacement prediction model incorporating extreme gradient boosting and optimal support vector regression

数据来测试XGBoost-HPO-SVR组合模型的预测性能.所用位移监测数据均由北斗/GNSS监测网络监测获得.其中,阳山山滑坡数据为2022年4月14日—2023年1月10日的累积位移值,脱甲山滑坡数据为2021年8月2日—2023年6月25日的累积位移值.阳山山滑坡和脱甲山滑坡每天的位移监测值如图2所示.



(a)阳山山滑坡位移监测数据



(b)脱甲山滑坡位移监测数据

图2 滑坡位移监测数据

Fig.2 Landslide displacement monitoring data

考虑到滑坡是受内外因素影响的复杂过程,其每天发生的位移,也会影响未来几天内滑坡的位移.故对位移值进行变换.也就是说,用第$t-1$,$t-2$和$t-3$天的位移预测第t天的位移.即在利用XGBoost-HPO-SVR进行初步预测和其余模型预测时,用前三天的位移去预测第四天的位移.为记录方便,后文使用GP01Y和GP01T分别表示阳山山滑坡和脱甲山滑坡的位移.

在利用GP01Y和GP01T滑坡监测数据验证各模型的预测效果时,均将数据分为训练集和测试集.利用XGBoost模型进行初步预测时,阳山山滑坡和脱甲山滑坡分别以2022年4月14日—2022年10月2日和2021年8月2日—2023年3月16的数据为训练集,以2022年10月3日—2023年1月10日和2023年3月17日—2023年6月25日的数据为测试集.在获得残差数据后,将残差数据按7∶3的比例划分为残差训练集和残差测试集,并利用HPO-SVR模型进行残差预测.即,在利用HPO-SVR进行残差预测时,阳山山滑坡分别以2022年10月3日—2022年12月11日和2022年12月12日—2022年1月10日的残差数据为训练集和测试集.脱甲山滑坡分别以2023年3月17日—2023年5月26日和2023年5月27日—2023年6月25日的残差数据为训练集和测试集.最后,综合初步预测和残差预测的结果,获得最终位移预测结果.对于其余模型预测,阳山山滑坡分别以2022年4月14日—2022年12月11日和2022年12月12日—2023年1月10日的数据为训练集和测试集.脱甲山滑坡分别以2021年8月2日—2023年5月26日和2023年5月27日—2023年6月25日的数据为训练集和测试集.其中,所划分的训练集数据来完成相应模型的参数选取和模型训练,测试集数据来完成模型的预测.

表1为XGBoost-HPO-SVR、XGBoost-PSO-SVR、XGBoost-GA-SVR等模型中SVR的最优超参数值.其中,SVR模型的核函数采用径向基核函数(rbf),其余参数采用Sklearn库中的默认值.HPO、PSO和GA算法的种群数为50,迭代次数为500次,寻优范围为0.01~200.

表1 不同模型优化的SVR超参数值

Tab. 1 Values of SVR hyperparameters optimized by different models

模型	数据集	超参数	
		C	g
XGBoost-HPO-SVR	GP01Y	1.015	0.01
	GP01T	0.173	6.900
XGBoost-PSO-SVR	GP01Y	10.152	0.010
	GP01T	0.010	11.070
XGBoost-GA-SVR	GP01Y	1.853	7.333
	GP01T	21.831	2.434

表1中,C为惩罚因子,代表了模型对误差的容忍程度.也就是说,C越大,说明模型在训练集中的表现越好(即其准确率越高),但其泛化能力越弱,越容易出现过拟合.反之,说明模型的泛化能力越强,但可能存在欠拟合现象,使得模型表现不佳.g代表模型核函数的系数,欠拟合时要增大该值,过拟合时要降低该值.SVR模型的性能受参数影响很大,因此

要获得性能好的模型需对模型的超参数进行合理的确定.本文采用 HPO 算法来确定 SVR 的超参数,即将 SVR 模型作为一个函数输入 HPO 算法中,超参数就是这个函数的变量值.通过设置 HPO 算法的种群数、迭代次数以及参数的寻优范围,让 SVR 模型在给定的超参数值范围内进行多次迭代来得到最优的超参数值.当 SVR 模型的均方误差最小时,模型的超参数值即为最优的超参数.其中,在利用 XGBoost 和 SVR 进行两步预测的过程中,阳山滑坡数据最优的 SVR 超参数 C 和 g 分别为 1.015、0.01,脱甲山滑坡数据最优的 SVR 超参数 C 和 g 分别为 0.173、6.900.

XGBoost 模型中树的数量决定了模型的学习能力.为使 XGBoost 模型有一个较高的学习能力,本文通过 5 折交叉验证的方式,利用 XGBoost 模型在训练集上的学习曲线来确定弱学习器的个数.其中,弱学习器的个数范围为 0~300,其结果如图 3 所示.由图 3 可知,弱学器个数在 55 左右,该模型的学习率较高.因此,所用弱学习器的个数选取为 55.其余参数均采用 XGBoost 库中的默认值.

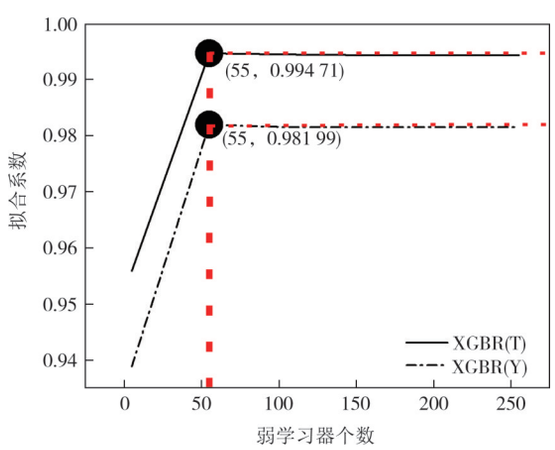


图 3　XGBoost 模型不同弱学习器个数的学习率

Fig. 3　Learning rate of XGBoost model with different number of weak learners

基于 GP01Y 和 GP01T 滑坡监测数据,采用不同模型的预测值和实际监测值的对比如图 4~图 7 所示.图 4 和图 5 分别为 XGBoost、SVR 与 XGBoost-HPO-SVR、XGBoost-PSO-SVR、XGBoost-GA-SVR、XGBoost-SVR 等模型分别在阳山滑坡和脱甲山滑坡预测结果对比图.从图 4 和图 5 可看出,XGBoost-HPO-SVR、XGBoost-PSO-SVR、XGBoost-GA-SVR、XGBoost-SVR 等组合模型的预测结果优于单一模型 XGBoost 和 SVR 的结果.特别是,XGBoost-HPO-SVR 的预测结果最好.

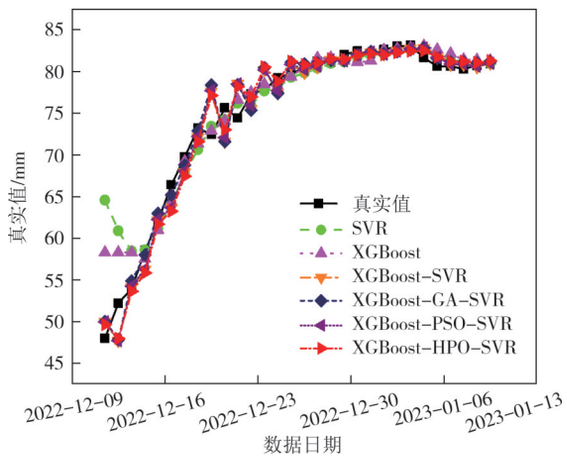


图 4　不同模型在 GP01Y 滑坡数据上预测结果对比图(测试集)

Fig.4　Comparison of prediction results of different models on GP01Y landslide data (test set)

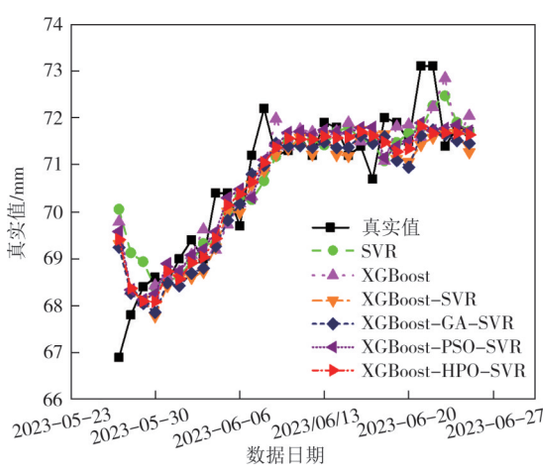


图 5　不同模型在 GP01T 滑坡数据上预测结果对比图(测试集)

Fig.5　Comparison of prediction results of different models on GP01T landslide data (test set)

表 2 给出了不同预测模型的 MAE、MSE 和 RMSE,以此来定量评估不同模型的滑坡位移预测精度.由表 2 可见:

1) 单一的预测 XGBoost 模型和 SVR 模型的 MSE 分别为 6.472 mm 和 13.480 mm. XGBoost-SVR、XGBoost-GA-SVR 和 XGBoost-PSO-SVR 等组合模型的 MSE 分别为 3.999 mm、3.891 mm、3.790 mm,而 XGBoost-HPO-SVR 组合模型的 MSE 仅为 3.505 mm.可以看出,组合模型的预测效果优于单一模型.尤其是 XGBoost-HPO-SVR 模型.XGBoost-HPO-SVR 相比于 XGBoost、SVR、XGBoost-SVR、XGBoost-GA-SVR 和 XGBoost-PSO-SVR 模型,MSE 分别下降了 2.967 mm、9.975 mm、0.494 mm、0.386 mm、0.285 mm,

其预测效果明显优于单一模型.对于模型的MAE来说,XGBoost-HPO-SVR的MAE仅为1.357 mm,而单一预测XGBoost模型和SVR模型的MAE分别为1.517 mm和1.810 mm.相比于单一预测XGBoost和SVR模型而言,XGBoost-HPO-SVR的MAE分别下降了10.55%和25.03%,其预测效果更优,更加接近真实值.综合而言,XGBoost-HPO-SVR模型具有最好的预测效果.

2)对于GP01T滑坡数据而言,单一的XGBoost模型和SVR模型的MSE分别为0.755 mm和0.784 mm,XGBoost-SVR、XGBoost-GA-SVR和XGBoost-PSO-SVR等组合模型的MSE分别为0.636 mm、0.574 mm和0.594 mm而XGBoost-HPO-SVR组合模型的MSE仅为0.550 mm.可以看出,组合模型的预测效果优于单一模型,尤其是XGBoost-HPO-SVR模型.XGBoost-HPO-SVR相比单一预测XGBoost和SVR模型,MSE分别下降了27.15%和29.85%.其预测效果明显优于单一预测模型.对于模型的MAE来说,XGBoost-HPO-SVR的MAE仅为0.538 mm,而单一预测XGBoost模型和SVR模型的MAE分别为0.644 mm和0.623 mm.相比于单一预测XGBoost和SVR模型而言,XGBoost-HPO-SVR的MAE分别下降了16.46%和13.64%,其预测效果比其他模型均有所提升.这同样表明,XGBoost-HPO-SVR的预测效果最好.

表2 不同模型预测结果对比(测试集)

Tab.2 Comparison of prediction results of different models (test set)

滑坡	预测模型	MAE/mm	MSE/mm	RMSE/mm
GP01Y	XGBoost	1.517	6.472	2.544
	SVR	1.810	13.480	3.671
	XGBoost-SVR	1.417	3.999	2.000
	XGBoost-GA-SVR	1.367	3.891	1.973
	XGBoost-PSO-SVR	1.364	3.790	1.947
	XGBoost-HPO-SVR	1.357	3.505	1.872
GP01T	XGBoost	0.644	0.755	0.869
	SVR	0.623	0.784	0.886
	XGBoost-SVR	0.571	0.636	0.798
	XGBoost-GA-SVR	0.577	0.574	0.758
	XGBoost-PSO-SVR	0.549	0.594	0.771
	XGBoost-HPO-SVR	0.538	0.550	0.742

为更直观地说明XGBoost-HPO-SVR组合模型与其他模型在不同滑坡数据集中的预测优势,本文进一步给出了不同模型预测结果的柱状图,如图6和图7所示.图6和图7为XGBoost-HPO-SVR与XGBoost-PSO-SVR、XGBoost-GA-SVR、XGBoost-SVR、XGBoost、SVR等模型分别在阳山滑坡和脱甲山滑坡预测精度对比图.其中RMSE和MAE参考左侧轴,MSE参考右侧轴.由图6和图7清晰可见,HPO算法明显提升了SVR模型的预测性能.采用XGBoost的组合模型的预测效果较好,而融合XGBoost、HPO和SVR算法两步预测的XGBoost-HPO-SVR模型明显优于只采用单一XGBoost、SVR模型.综合而言,本文所提XGBoost-HPO-SVR滑坡位移预测模型的性能更优.

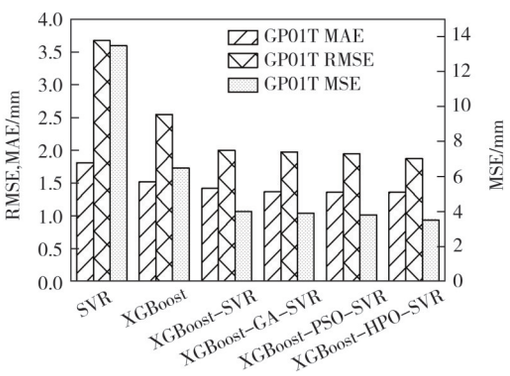


图6 GP01Y不同模型预测结果对比图

Fig. 6 Comparison of prediction results of different models of GP01Y

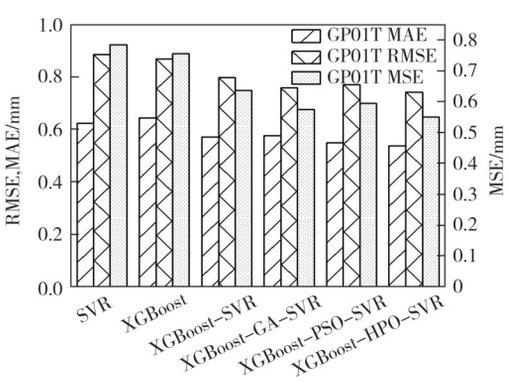


图7 GP01T不同模型预测精度对比图

Fig.7 Comparison of prediction accuracy of different models of GP01T

4 结论

本文将XGBoost、SVR和HPO相结合,提出了XGBoost-HPO-SVR滑坡位移预测模型.该模型利用XGBoost初步预测和SVR残差预测相结合的方式实现了滑坡位移的高精度预测.同时,采用猎人猎物优化算法(HPO)合理确定了SVR的超参数,进一步提

高了滑坡位移的预测精度.通过两起滑坡北斗/GNSS监测实例,对比验证了本文所提XGBoost-HPO-SVR组合模型的优势,结论如下:

1)利用HPO算法通过不断更新猎人与猎物位置的动态寻优策略,实现了SVR算法的精细化调参,并确定了超参数的取值.

2)XGBoost-HPO-SVR组合模型采取初步预测和残差预测相结合的方式,实现了滑坡位移预测精度的大幅提升.与XGBoost、SVR、HPO-SVR、PSO-SVR、GA-SVR以及XGBoost-PSO-SVR、XGBoost-GA-SVR、XGBoost -SVR等组合模型相比,其可获得精度更高的滑坡位移预测结果,在滑坡预警中具有较高的应用价值.

需注意的是,本文针对滑坡运动而产生的位移,构建了XGBoost-HPO-SVR模型.在应用本文所提出XGBoost-HPO-SVR模型进行滑坡位移预测时,需根据实际滑坡位移数据和实际需求确定残差预测的比例.此外,在XGBoost和SVR模型中还需通过实验确定一些更加合理的参数值,例如,弱学习器的个数以及核函数等.但正如引言所提到,滑坡失稳是受内外因素共同影响的复杂过程,顾及内外多因素的高精度滑坡位移预测模型将是下一步工作的重点

参考文献

[1] 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.

[2] SAITO M. Forecasting the time of occurrence of a slope failure [C]// Proceedings 6th International Conference Soil Mechanics and Foundation Engineering. 1965: 537-541.

[3] 谢锋,朱陆明,王立忠. 滑坡监控信息分析中的修正灰色系统预测模型及应用[J]. 岩石力学与工程学报,2005,24(22):4099-4105.
XIE F, ZHU L M, WANG L Z. Modified grey system forecasting model and its application for analyzing information of landslide monitory [J]. Chinese Journal of Rock Mechanics and Engineering, 2005, 24(22): 4099-4105.(in Chinese)

[4] 徐峰,汪洋, 杜娟, 等. 基于时间序列分析的滑坡位移预测模型研究[J]. 岩石力学与工程学报, 2011, 30(4):746-751.
XU F, WANG Y, DU J, et al. Study of displacement prediction model of landslide based on time series analysis [J]. Chinese Journal of Rock Mechanics and Engineering, 2011, 30(4): 746-751.(in Chinese)

[5] 孟蒙,陈智强, 黄达, 等. 基于H-P滤波法、ARIMA和VAR模型的库区滑坡位移综合预测[J]. 岩土力学,2016,37(增刊2):552-560.
MENG M, CHEN Z Q, HUANG D, et al. Displacement prediction of landslide in Three Gorges Reservoir area based on H-P filter, ARIMA and VAR models[J]. Rock and Soil Mechanics, 2016, 37(Sup.2): 552-560.(in Chinese)

[6] NIU X X, MA J W, WANG Y K, et al. A novel decomposition-ensemble learning model based on ensemble empirical mode decomposition and recurrent neural network for landslide displacement prediction[J]. Applied Sciences, 2021, 11(10): 4684.

[7] KUMAR P, SIHAG P, CHATURVEDI P, et al. BS-LSTM: an ensemble recurrent approach to forecasting soil movements in the real world[J]. Frontiers in Earth Science, 2021, 9: 716.

[8] 姜宇航,王伟,邹丽芳,等. 基于粒子群-变分模态分解、非线性自回归神经网络与门控循环单元的滑坡位移动态预测模型研究[J]. 岩土力学,2022,43(增刊1):601-612.
JIANG Y H, WANG W, ZOU L F, et al. Research on dynamic prediction model of landslide displacement based on particle swarm optimization-variational mode decomposition, nonlinear autoregressive neural network with exogenous inputs and gated recurrent unit[J]. Rock and Soil Mechanics, 2022, 43(Sup.1): 601-612.(in Chinese)

[9] HU X L, WU S S, ZHANG G C, et al. Landslide displacement prediction using kinematics-based random forests method: a case study in Jinping Reservoir Area, China[J]. Engineering Geology, 2021, 283: 105975.

[10] DU H, SONG D Q, CHEN Z, et al. Prediction model oriented for landslide displacement with step-like curve by applying ensemble empirical mode decomposition and the PSO-ELM method [J]. Journal of Cleaner Production, 2020, 270: 122248.

[11] 窦杰,向子林,许强, 等. 机器学习在滑坡智能防灾减灾中的应用与发展趋势[J]. 地球科学,2023,48(5):1657-1674.
DOU J, XIANG Z L, XU Q, et al. Application and development trend of machine learning in landslide intelligent disaster prevention and mitigation [J]. Earth Science, 2023, 48(5): 1657-1674.(in Chinese)

[12] MA J W, XIA D, WANG Y K, et al. A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: Insights from a case study of landslide displacement prediction[J]. Engineering Applications of Artificial Intelligence, 2022, 114: 105150.

[13] 张俊, 殷坤龙, 王佳佳, 等. 基于时间序列与PSO-SVR耦合模型的白水河滑坡位移预测研究[J]. 岩石力学与工程学报, 2015, 34(2): 382-391.
ZHANG J, YIN K L, WANG J J, et al. Displacement prediction of Baishuihe landslide based on time series and pso-svr model[J]. Chinese Journal of Rock Mechanics and Engineering, 2015, 34(2): 382-391.(in Chinese)

[14] 李麟玮,吴益平,苗发盛,等. 基于变分模态分解与GWO-MIC-SVR模型的滑坡位移预测研究[J]. 岩石力学与工程学报, 2018, 37(6): 1395-1406.
LI L W, WU Y P, MIAO F S, et al. Displacement prediction of landslides based on variational mode decomposition and GWO-

MIC-SVR model[J]. Chinese Journal of Rock Mechanics and Engineering, 2018, 37(6): 1395-1406.(in Chinese)

[15] MEHRABI M, MOAYEDI H. Landslide susceptibility mapping using artificial neural network tuned by metaheuristic algorithms[J]. Environmental Earth Sciences, 2021, 80(24): 804.

[16] HOSSEIN M, ATEFEH A D, HOSSEIN G M. A novel hybrid based on nature-inspired and Stochastic Fractal Search algorithms for optimizing of artificial neural network model in landslide susceptibility [J]. Engineering Applications of Artificial Intelligence, 2023, 117(PA): 105457-105457.

[17] GAO W, DAI S, CHEN X. Landslide prediction based on a combination intelligent method using the GM and ENN: two cases of landslides in the Three Gorges Reservoir, China [J]. Landslides, 2020, 17(1): 111-126.

[18] 王利, 许豪, 舒宝, 等. 利用互信息和IPSO-LSTM进行滑坡监测多源数据融合[J]. 武汉大学学报(信息科学版), 2021, 46(10): 1478-1488.
WANG L, XU H, SHU B, et al. A multi-source heterogeneous data fusion method for landslide monitoring with mutual information and IPSO-LSTM neural network [J]. Geomatics and Information Science of Wuhan University, 2021, 46(10): 1478-1488.(in Chinese)

[19] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016.

[20] 范勇, 胡名东, 杨广栋, 等. PCA-GWO-SVR机器学习用于边坡爆破振动速度峰值预测研究[J]. 振动工程学报, 2024, 37(8): 1431-1441.
FAN Y, HU M D, YANG G D, et al. PCA-GWO-SVR machine learning applied to prediction of peak vibration velocity of slope blasting [J]. Journal of Vibration Engineering, 2024, 37(8): 1431-1441.(in Chinese)

[21] 胡向东, 吕高飞, 白银. 基于优化支持向量回归的工业互联网安全态势预测方法[J]. 电子学报, 2023, 51(2): 446-454.
HU X D, LÜ G F, BAI Y. A method of security situation prediction for industrial Internet based on optimized support vector regression[J]. Acta Electronica Sinica, 2023, 51(2): 446-454.(in Chinese)

[22] 王晨辉, 赵贻玖, 郭伟, 等. 滑坡位移EEMD-SVR预测模型[J]. 测绘学报, 2022, 51(10): 2196-2204.
WANG C H, ZHAO Y J, GUO W, et al. Displacement prediction model of landslide based on ensemble empirical mode decomposition and support vector regression[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(10): 2196-2204.(in Chinese)

[23] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 133-137.
ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 133-137.(in Chinese)

[24] NARUEI I, KEYNIA F, SABBAGH MOLAHOSSEINI A. Hunter-prey optimization: algorithm and applications [J]. Soft Computing, 2022, 26(3): 1279-1314.

[25] ELSHAHED M, EL-RIFAIE A M, TOLBA M A, et al. An innovative hunter-prey-based optimization for electrically based single-, double-, and triple-diode models of solar photovoltaic systems[J]. Mathematics, 2022, 10(23): 4625.

[26] MA J, LIU F M. Bearing fault diagnosis with variable speed based on fractional hierarchical range entropy and hunter-prey optimization algorithm-optimized random forest [J]. Machines, 2022, 10(9): 763.

[27] 高雨虹, 曲昭伟, 宋现敏. 基于猎人猎物优化与双向长短时记忆组合模型的汽车出车率预测[J]. 交通运输系统工程与信息, 2023, 23(1): 198-206, 264.
GAO Y H, QU Z W, SONG X M. Car operation rate prediction based on combination model of hunter-prey optimizer algorithm and bi-directional long short-term memory neural network [J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(1): 198-206.(in Chinese)