

基于变分模态分解和极端梯度提升的 公路边坡位移预测

龙志友^{1,2}, 万昭龙^{1,2}, 董 是^{1,2}, 杨 超³, 刘肖扬⁴

(1. 长安大学 运输工程学院, 西安 710064; 2. 道路基础设施数字化教育部工程研究中心, 西安 710064; 3. 陕西高速公路工程试验检测有限公司, 西安 710086; 4. 长安大学 公路学院, 西安 710064)

摘 要:针对公路边坡位移监测数据具有非线性、噪声多、不平稳的特点,导致边坡位移预测精度不足的缺点,提出了优化变分模态分解(VMD)和极端梯度提升(XGBoost)的回归预测算法,用于公路边坡位移数据的处理和预测。首先,利用粒子群算法寻优找到VMD的最优分解层数及惩罚因子,进而对边坡位移数据进行VMD,得到趋势位移、周期性位移和随机波动性位移特征;其次,加入其他监测数据作为回归预测特征变量,并利用SHAP对输入特征变量进行重要度解释,筛选出重要特征后输入XGBoost模型进行预测,同时利用网格搜索确定XGBoost模型最优参数;最后,基于实际案例分析,验证本文提出的边坡位移分解方法和回归预测模型的适用性和鲁棒性。结果表明:在本文的实际案例中,对于边坡位移数据分解,VMD较经验模态分解(EMD)和集合经验模态分解(EEMD)适用性更强;对于边坡位移预测,XGBoost预测精度较极限学习机(ELM)和支持向量机(SVM)分别提升4.37%和0.41%。本文回归模型具有较高的预测精度和较强的鲁棒性。同时,研究表明,根据VMD提取的边坡位移特征变量(周期性位移、随机波动性位移和趋势位移)对边坡位移预测的SHAP重要性程度较高。本文模型可为公路边坡位移预测及安全预警研究提供一定思路。

关键词:道路工程;公路边坡位移预测;变分模态分解;粒子群算法;极端梯度提升;鲁棒性检验

中图分类号:U41;X928.03 **文献标志码:**A **文章编号:**1671-5497(2025)07-2320-13

DOI:10.13229/j.cnki.jdxbgxb.20231197

Displacement prediction of highway slope based on variational mode decomposition and extreme gradient boosting

LONG Zhi-you^{1,2}, WAN Zhao-long^{1,2}, DONG Shi^{1,2}, YANG Chao³, LIU Xiao-yang⁴

(1. School of Transportation Engineering, Chang'an University, Xi'an 710064, China; 2. Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an 710064, China; 3. Shaanxi Expressway Engineering Testing Inspection & Testing Co., Ltd., Xi'an 710086, China; 4. School of Highway, Chang'an University, Xi'an 710064, China)

收稿日期:2023-11-04.

基金项目:国家自然科学基金项目(52478435);国家自然科学基金-联合基金重点项目(U2433210);陕西省青年科技新星项目(2025ZC-KJXX-46);中国博士后科学基金项目(2021M692427);陕西省交通运输厅科研项目(23-62X, 23-04K).

作者简介:龙志友(1998-),男,博士研究生.研究方向:道路基础设施数字化监测.E-mail: longzhiyou@chd.edu.cn

通信作者:董是(1989-),男,副教授,博士.研究方向:交通基础设施智能运维养护.E-mail: dongshi@chd.edu.cn

Abstract: Facing the nonlinear, high noisy, and unstable characteristics of highway slope displacement monitoring data lead to insufficient accuracy of slope displacement prediction, this paper proposes a regression prediction algorithm that optimizes the variational mode decomposition (VMD) and extreme gradient boosting (XGBoost) for the processing and prediction of highway slope displacement data. Firstly, the particle swarm optimization is used to find the optimal number of decomposition layers and the penalty factor of VMD, and then the slope displacement data are subjected to VMD to obtain the trend displacement, periodic displacement and random fluctuation displacement features. Secondly, other monitoring data are added as regression prediction feature variables, and the Shapley additive explanation (SHAP) is used to interpret the importance of the input feature variables, and then the important features are screened and inputted into XGBoost model for prediction, and the grid search is used to determine the optimal parameters of the XGBoost model. Finally, the applicability and robustness of the slope displacement decomposition method and regression prediction model proposed in this paper are verified based on actual case analysis. The results show that in the real cases of this paper, for slope displacement data decomposition, VMD has stronger applicability than empirical modal decomposition (EMD) and ensemble empirical modal decomposition (EEMD); for slope displacement prediction, the prediction accuracy of XGBoost is improved by 4.37% and 0.41% compared with the extreme learning machine (ELM) and support vector machine (SVM), respectively. The regression model proposed in this paper has high prediction accuracy and strong robustness. Meanwhile, it is shown that the slope displacement characteristic variables (periodic displacement, random fluctuation displacement and trend displacement) extracted by VMD have a greater degree of SHAP importance for slope displacement prediction. The method proposed in this paper can provide some ideas for highway slope displacement prediction and safety warning research.

Key words: road engineering; highway slope displacement prediction; variational mode decomposition; particle swarm optimization; extreme gradient boosting; robustness test

0 引言

随着我国高速公路建设的飞速发展,高速公路边坡的数量也日益增多。公路边坡服役的稳定性直接影响公路运营安全和管养成本,因此需要掌握公路边坡变形状况,进而开展路基边坡结构健康监测。张勤等^[1]指出,边坡位移的准确预测是边坡安全预警的重要前提。实现对边坡位移变化趋势的精准预测,可及时针对边坡问题采取应对措施,这对保障人民生命财产安全和高速公路安全运营具有重要意义。

近年来,众多研究围绕边坡位移预测模型展开,如极限学习机 (Extreme learning machine, ELM)^[2]、支持向量机 (Support vector machine, SVM)^[3]、轻量级梯度提升机 (Light gradient boosting machine, LightGBM)^[4]和极端梯度提升算法 (Extreme gradient boosting, XGBoost)^[5]等机器学习模型,在边坡位移预测领域得到广泛应

用。但是,这些算法在运算效率和预测精度上均存在一定的局限性。例如,ELM存在学习速率慢、学习率难以确定、易陷入局部最小值等缺点;SVM存在对参数设置、数据缩放、噪声数据敏感和计算复杂度高等缺点;LightGBM存在运算空间消耗大、运算时间长等缺点。尽管XGBoost也存在计算量大、预排序过程的空间复杂度高等缺点,但其预测精度高、灵活性强、防止过拟合和算法效率高等优点,使其在非线性、不平稳数据的预测上有着极强的优势,因而本文选用XGBoost对边坡位移进行预测。另外,大多数机器学习模型对内部机制及其输出结果的影响缺乏可解释性,因而本文引入沙普利加性解释 (Shapley additive explanation, SHAP)对XGBoost模型的输入特征变量进行解释。SHAP是由Lundberg等^[6]提出的一种基于博弈论的可解释机器学习的方法,具有较好的全局和局部解释能力。

由于边坡位移数据具有非线性、噪声多、不平

稳等特征,直接对其进行预测会降低模型的预测精度,因此需对边坡位移数据进行分解,将其转化为较为平稳的数据,以提高模型的预测精度。小波变换(Wavelet transform, WT)^[7]、经验模态分解(Empirical mode decomposition, EMD)^[8]、集合经验模态分解(Ensemble empirical mode decomposition, EEMD)^[9]和变分模态分解(Variational mode decomposition, VMD)^[2]在边坡位移数据分解中得到了广泛的应用。但是 WT 需要确定基函数,且自适应能力较差;EMD 虽然自适应能力强,但存在模态混叠和端点效应等问题;EEMD 虽然解决了模态混叠问题,但是其求解耗时且存在残留白噪声;VMD 不仅具有避免模态混叠、过包络、欠包络、边际效应等问题的优点,还具有复杂数据分解精度高和抗噪声数据能力强的优点,但是其分解层数 K 及惩罚因子 α 需要人为选取。为了解决这一问题,前人研究利用奇异值分解^[10]、蝙蝠算法^[11]和粒子群算法(Particle swarm optimization, PSO)^[12]等智能算法寻找最优的参数组合。相较于其他算法,粒子群算法具有收敛速度快、设置参数少等优点,能更快地找到最优参数组合,因而本文选用 PSO 来寻找 VMD 的最优参数。

综上所述,为解决公路边坡位移难预测的问题,本文提出了基于 PSO 寻优、VMD 分解、SHAP 解释、XGBoost 预测的公路边坡位移预测模型。该模型利用 PSO 寻找 VMD 的最优分解层数 K 和惩罚因子 α ,选取分解得到的各子序列变量作为特征输入至 XGBoost 模型进行回归预测,利用网格搜索找寻模型最优参数。同时,使用 SHAP 对输入特征进行重要度解释,选取重要度高的特征输入 XGBoost 模型以提高回归预测精度。为验证本文模型的适用性和鲁棒性,选取陕西省某高速公路 A 立交边坡和 B 立交边坡作为研究对象,以平均绝对百分比误差、均方根误差和拟合优度作为评价标准,选取 EMD、EEMD 和 VMD 作为边坡位移数据的分解方法,XGBoost、SVM 和 ELM 作为回归预测模型,将二者进行模型组合,对比其他文献提出的组合模型方法,以验证本文模型的鲁棒性和优越性。

1 研究方法

1.1 变分模态分解

VMD 是一种自适应、完全非递归的模态变

分和信号处理方法,由 Dragomiretskiy 等^[13]于 2014 年提出。通过实验证明,其在信号采样及噪声处理方面具有更高的鲁棒性,对于非线性、复杂度高、不平稳数据的处理更加有效。本文利用 VMD 算法将非线性、不平稳的高速公路边坡结构监测位移数据分解成较为平稳和具有实际物理意义的位移分量。VMD 算法可以分解成两部分,即变分问题的构造及变分问题的求解。VMD 的变分约束问题数学表达式为:

$$\begin{cases} \min_{\{u_k, \omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k u_k(t) = f(t) \end{cases} \quad (1)$$

式中: $U = \{u_1(t), \dots, u_k(t), \dots, u_K(t)\}$ 为各模态分量; $W = \{\omega_1, \dots, \omega_k, \dots, \omega_K\}$ 为各模态分量的中心频率; K 为分解得到的模态数; $f(t)$ 为原始信号。

应用二次惩罚项和拉格朗日乘子 λ 将上述问题转化为非约束问题进行求解,即:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

利用交替方向乘子法求解式(2)的最优解,不断更新 u_k 、 ω_k 和 λ ,将求解非约束问题转化为求解极小值问题,即:

$$\begin{aligned} u_k^{n+1} = & \arg \min \left\{ \alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \right. \\ & \left. \left\| f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\} \end{aligned} \quad (3)$$

模态 u_k 的最优解为:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(k) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (4)$$

式中: $\hat{u}_k^{n+1}(\omega)$ 、 $\hat{f}(\omega)$ 、 $\hat{\lambda}(\omega)$ 分别为 $u_k^{n+1}(t)$ 、 $f(t)$ 、 $\lambda(t)$ 的傅里叶变换; ω 为频率。

1.2 粒子群算法

在 VMD 过程中,分解层数 K 及惩罚因子 α 会对分解结果产生较大影响,由于这些参数是主观

设定的,因而不同的设置参数会导致数据欠分解或者过度分解现象的发生。为了得到最优的设置参数,本文利用PSO进行参数优化。

PSO优化原理如下:在 D 维空间中,随机生成一个粒子群,包含 N 个粒子,粒子 $i(i=1,2,\dots,N)$ 的位置可以表示为向量 $x_i=(x_{i1},x_{i2},\dots,x_{iD})$,利用 x_i 计算适应度,表示为 $f(x_i)$ 。粒子 i 的速度可以表示为向量 $v_i=(v_{i1},v_{i2},\dots,v_{iD})$,其个体局部极值 $pbest_i=(p_{i1},p_{i2},\dots,p_{iD})$,种群全局极值 $gbest=(g_1,g_2,\dots,g_D)$ 。粒子在移动的过程中,需要确定移动方向及长度,因此定义粒子 i 第 d 维速度及位置的更新公式分别为:

$$v_{id}^k = \omega v_{id}^{k-1} + c_1 r_1 (pbest_{id} - x_{id}^{k-1}) + c_2 r_2 (gbest_d - x_{id}^{k-1}) \quad (5)$$

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

式中: v_{id}^k 为第 k 迭代粒子 i 移动速度向量的第 d 维分量; x_{id}^k 为第 k 迭代粒子 i 位置向量的第 d 维分量; c_1 和 c_2 为学习因子,用于调节学习最大步长; r_1 和 r_2 为取值范围在0~1之间的随机数,用以增加搜索的随机性; ω 为惯性权重,用于调节空间的搜索范围。

在利用PSO确定VMD参数的过程中,需要确定粒子的适应度函数,为此唐贵基等^[12]提出了包络熵 E_p 的概念,长度为 N 的零均值信号 $x(j)$ 的包络熵 E_p 可表示为:

$$E_p = -\sum_{j=1}^N p_j \lg p_j \quad (7)$$

其中:

$$p_i = \frac{a(j)}{\sum_{j=1}^N a(j)} \quad (8)$$

式中: p_i 为 $a(j)$ 归一化后的结果; $a(j)$ 为 $x(j)$ 经Hilbert解调后得到的包络信号。当第 i 个粒子处于 x_i 位置时,计算经VMD得到的所有IMF分量的包络熵,所有分量包络熵的最小值称为局部极小熵值,记作 $\min_L E_p^{\text{IMF}}$ 。以局部极小熵值作为粒子群的适应度值,并以局部极小熵值最小化作为粒子群寻优的最终目标。

1.3 极端梯度提升算法

XGBoost是一种高效的梯度提升决策树(Gradient boosting decision tree,GBDT)算法,由原有的GBDT改进而来。该方法的核心是采用

集成思想,通过特定方法将多个弱学习器整合成一个强学习器,从而提升模型的精度。XGBoost具有精度高、灵活性强、防止过拟合的优点。XGBoost主要由损失函数和L2正则化项来构成目标函数^[14],即:

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (9)$$

式中: $\sum_{i=1}^n l(y_i, \hat{y}_i)$ 为损失函数,用以衡量模型的偏差; $\sum_{k=1}^K \Omega(f_k)$ 为L2正则化项,用以防止模型过拟合,其表达式如下所示:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (10)$$

式中: T 为叶子节点数量; ω 为叶子节点权重; γ 和 λ 为L2正则化参数。

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_i(x_i) \quad (11)$$

式中: $\hat{y}_i^{(t)}$ 为第 t 次迭代后样本 i 的预测值; $\hat{y}_i^{(t-1)}$ 为第 $t-1$ 次迭代后样本 i 的预测值; $f_i(x_i)$ 为第 t 棵树的模型预测结果。将式(11)代入式(9)中,并利用泰勒公式对目标函数进行转化,可得:

$$\text{Obj} \cong \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_k) + \text{constant} \quad (12)$$

式中: g_i 为损失函数的一阶梯度统计量; h_i 为损失函数的二阶梯度统计量。

1.4 模型解释

为了提升XGBoost模型的可解释性,本文引入SHAP对输入特征变量进行重要度解释,并筛选出重要特征变量输入模型以提高模型预测精度,计算公式如下^[15]:

$$\phi_h = \frac{|S|! (|M| - |S| - 1)!}{|M|!} (v(S \cup \{h\}) - v(S)) \quad (13)$$

式中: ϕ_h 为输入特征 h 的SHAP值; M 为需要解释的样本特征向量; S 为特征向量 M 中不包含特征 h 的特征组合向量; $v\{S\}$ 为 S 中所有特征向量共同作用所产生的贡献; $v(S \cup \{h\}) - v\{S\}$ 为特征 h 为该共同作用所带来的贡献。

1.5 评价标准

为了更好地评价本文模型的预测效果,采用平均绝对百分比误差(Mean absolute percentage error, MAPE)、均方根误差(Root mean square er-

ror, RMSE)、拟合优度 R^2 对模型预测结果进行评价,具体计算公式为:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

式中: y_i 为真实值; $\hat{y}_i$ 为模型预测值; n 为测试集样本个数; $\bar{y}$ 为真实值的平均值。

1.6 PSO-VMD-SHAP-XGBoost 算法

本文提出的 PSO-VMD-SHAP-XGBoost 模型具体流程图如图 1 所示。

(1) 将获取的边坡位移数据经数据预处理后,运用 VMD 算法进行位移数据分解。

(2) 利用 PSO 求解 VMD 的最优分解层数 K 和惩罚因子 α 。初始化粒子群,将需要求解的参数进行组合 $[\alpha, K]$ 作为粒子的位置。随机生成一些参数组合作为粒子的初始位置,并初始化粒子的移动速度。

(3) 计算每个粒子位置下 VMD 得到的所有 IMF 的包络熵,以其中最小的包络熵 $\min_L E_p^{\text{IMF}}$ 作为粒子的适应度,并更新粒子个体局部极值和种群全局极值,更新粒子的位置和速度,按照设定的迭代次数进行循环迭代。当达到最大的迭代次数时,输出最优的适应度及对应的参数组合。

(4) 以 PSO 输出的参数组合作为 VMD 的参数。将公路边坡位移数据分解得到的 IMF 划分为趋势项、周期项及随机波动项。

(5) 选取特征输入至 XGBoost 模型,对模型进行训练、调参和预测,并利用 SHAP 对输入特征进行重要度解释。选取重要度高的特征变量输入模

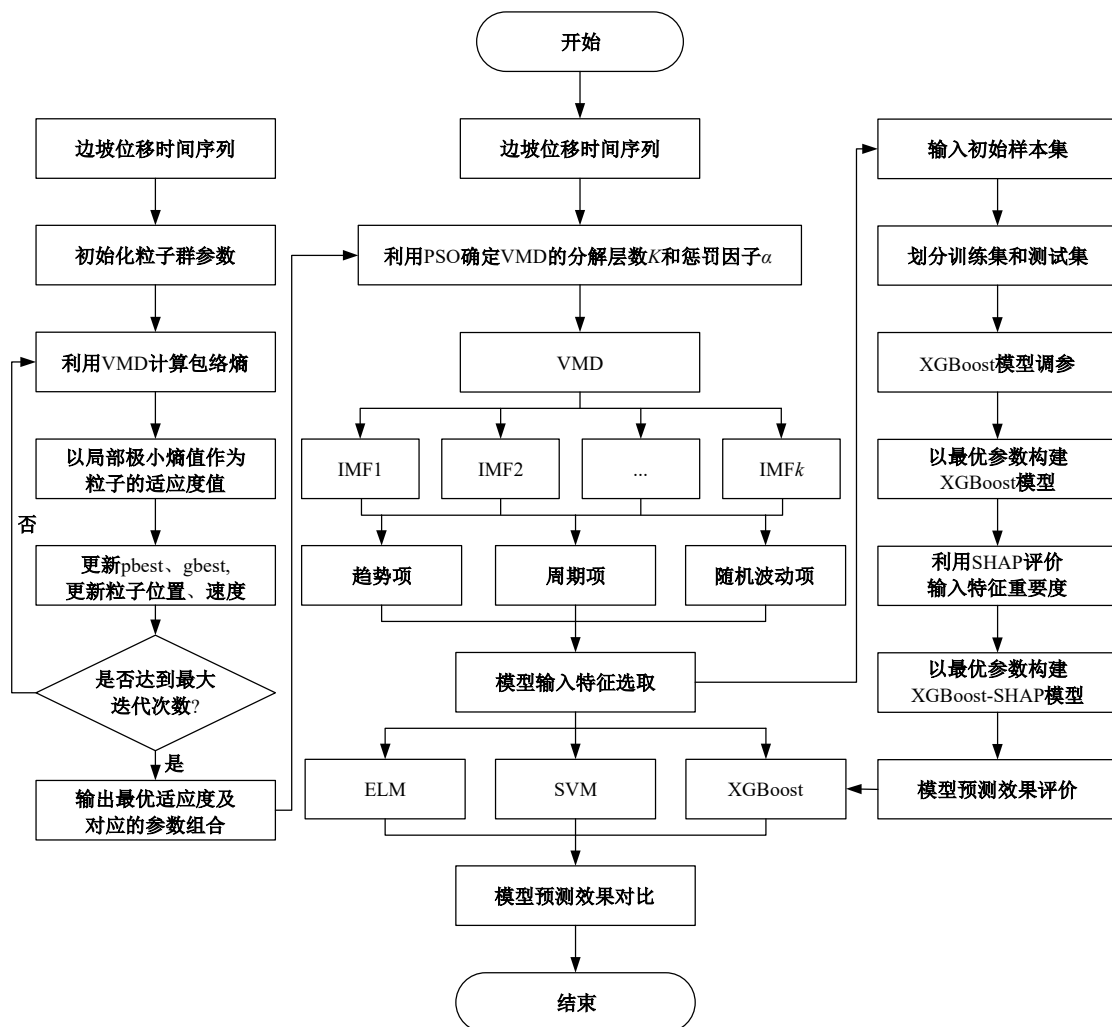


图1 PSO-VMD-SHAP-XGBoost 模型流程图

Fig. 1 Flow chart of PSO-VMD-SHAP-XGBoost model

型中,对模型进行训练、调参、预测和评价。

(6)选取参考文献[2,3,5]中的边坡位移预测模型对本文中的边坡位移数据进行预测,并与本文提出的边坡位移预测模型进行预测效果对比与评价。

2 工程实例

选取陕西省某高速公路 A 边坡作为本文模型算法研究的主要对象,选取 B 边坡作为本文模型算法鲁棒性检验的研究对象。其中,A 边坡位于 ST 高速某互通立交 E 匝道处,靠近某收费站,属于 6 级边坡;B 边坡位于 ST 高速某互通立交 A 匝道

处,属于 5 级边坡。两个边坡均布设了地表位移传感器(GPS-RTK)、深部位移传感器(固定式测斜仪)、渗压计及一体化雨量站,用于监测边坡地表位移、深部位移、地下水压力及雨量,具体监测内容如表 1 所示,传感器布设位置如图 2 所示。

表 1 边坡监测内容

Table 1 Slope monitoring content				
监测项目	仪器名称	仪器型号	记录频率	单位
地表位移	GPS-RTK	GSTP-RTK11-1	30 min/次	m
深部位移	固定式测斜仪	GSTP-ME610-2	30 min/次	°
地下水压力	渗压计	GSTP-ZX450	30 min/次	kPa
雨量	一体化雨量站	—	1 d/次	mm

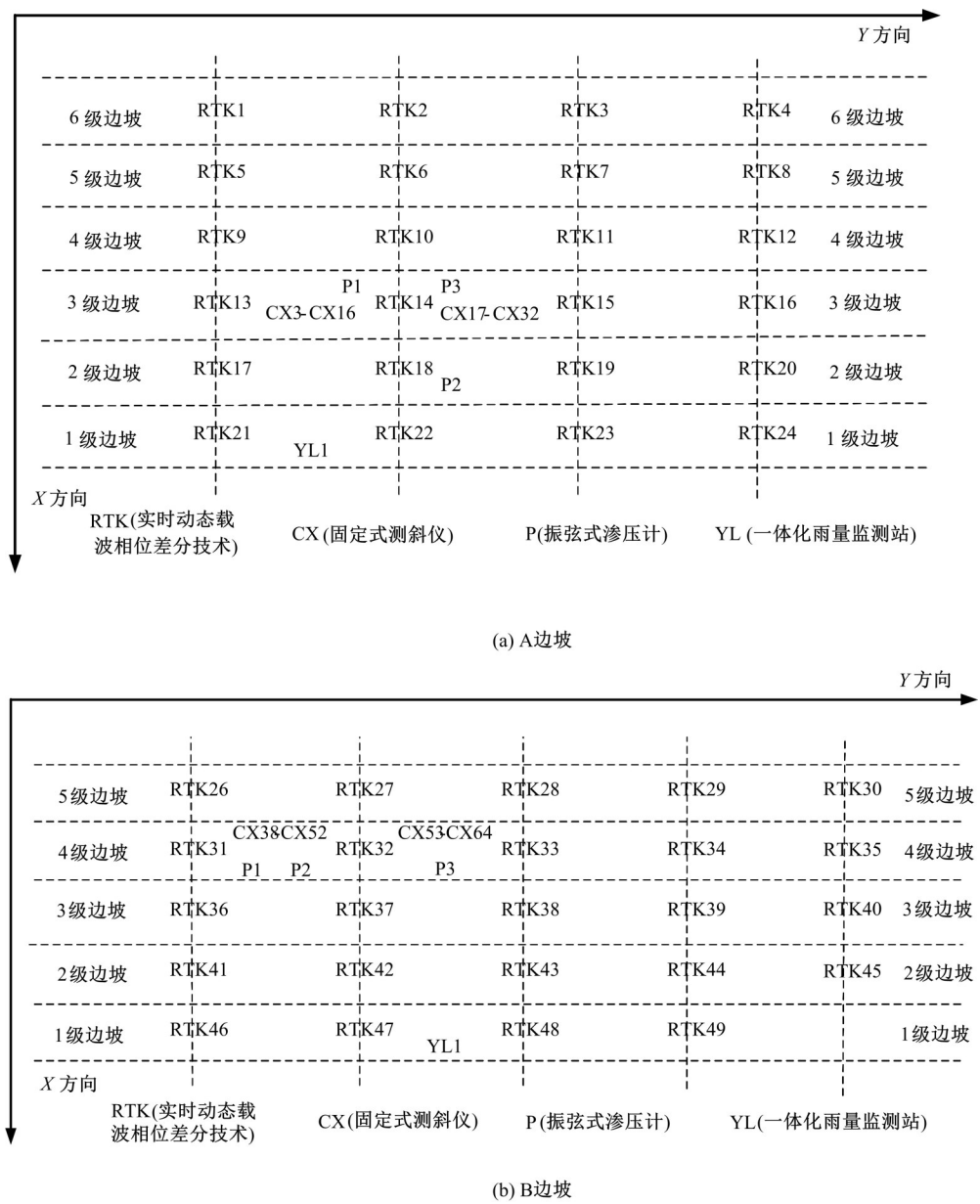


图 2 边坡位移监测传感器布设位置俯视图

Fig. 2 Overhead view of slope displacement monitoring sensors deployment location

2.1 数据处理

本文选取由GPS-RTK采集的A边坡2020年1月1日至2020年12月31日的边坡地表位移数据进行研究。GPS-RTK的数据采集频率为30 min/次,能够采集X、Y、Z方向的边坡地表位移数据。利用欧式范数 $\|x\|_2 = \sqrt{x^2 + y^2 + z^2}$ 将每个监测点 (x, y, z) 方向的数据进行融合^[16],得到边坡累积位移。

由于一年内GPS-RTK监测的数据量过大,难以在图中直观显示不同监测点间的关系,因此

选取其中1个月的数据进行细致研究。本文选取6月1日至6月30日每12 h重采样的数据进行分析,各传感器之间的数据分布如图3所示。由图3可知,RTK13监测的边坡地表位移数据与布设在X方向和Y方向的RTK监测的地表位移数据相似,变化趋势相同,监测误差相近,因而本文选取RTK13监测的数据对A边坡位移进行研究。由RTK13的年监测数据分布可知,该数据具有非线性、不平稳的特征。但该边坡总体变化范围较为稳定,累积位移最大值为5.11 mm,99%的累积位移在4.52 mm以下,80%以上的累积位移集中

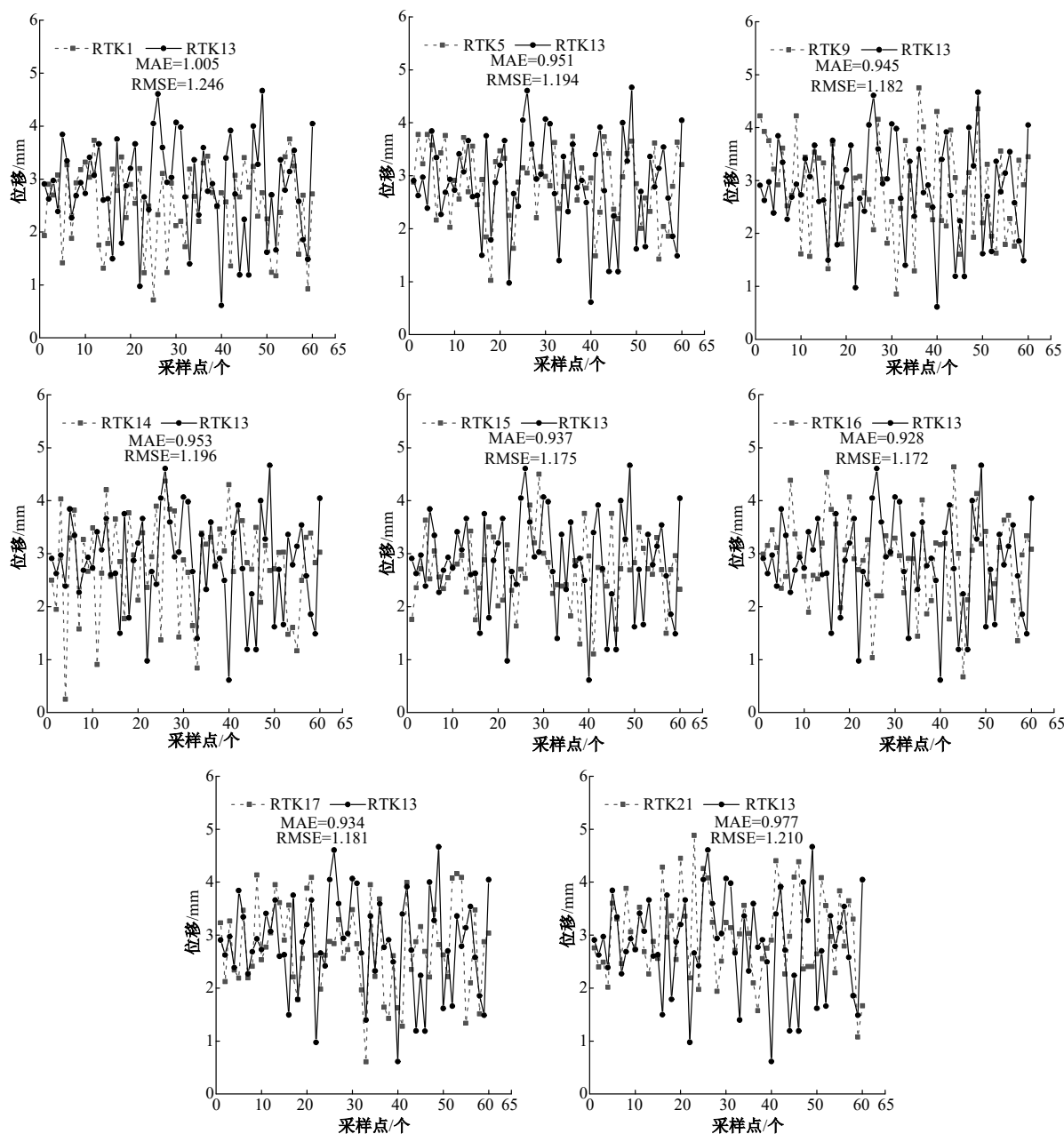


图3 RTK13与X、Y方向RTK监测数据的分布

Fig. 3 Distribution of RTK13 with RTK monitoring data in X and Y directions

在2~4 mm。

2.2 边坡位移分解

利用VMD算法对边坡位移数据进行分解,在此之前运用PSO求解最优分解层数 K 及惩罚因子 α 。结合参考文献[17]中的参数设置,并根据实际情况,本文粒子群参数设置如下:惯性权重 ω 为0.5,学习因子 $c_1=c_2=2$,种群规模 $O=30$,迭代次数 $H=20$ 。对于PSO寻优的VMD算法参数组合 $[\alpha, K]$,本文设定 α 的寻优范围为 $[100, 3\ 000]$, K 的寻优范围为 $[3, 12]$ 。由于PSO容易陷入局部最优的陷阱中,因此需利用PSO进行多次求解,进而找到全局最优解。本文选用CPU为AMD Ryzen 7 5800H with Radeon Graphic、GPU为NVIDIA GeForce RTX 3060、运行内存为32 G

的工作站进行了10次寻优求解。每一次寻优的分解层数 K 及惩罚因子 α 相同,参数组合 $[\alpha, K]=[1\ 505, 8]$ 。平均求解时长1 007 s, PSO未陷入局部最优,且运行时间较短,进一步说明了本文选取粒子群参数的合理性。

选取PSO寻优的参数组合对边坡位移数据进行VMD,分解结果如图4所示。由图4(a)可知,边坡位移数据的各IMF分量差异性较小,这是因为数据量过大,导致数据变化的细节在图中被掩盖。故进一步选取采样区间 $[6\ 000, 8\ 000]$ 的数据,并结合参考文献[18],对VMD得到的IMF分量进行研究。由图4(b)可以看出,IMF1的变化趋势与边坡总体位移数据的变化趋势相近,变化趋势较为缓和,反映了边坡位移的长期变化趋

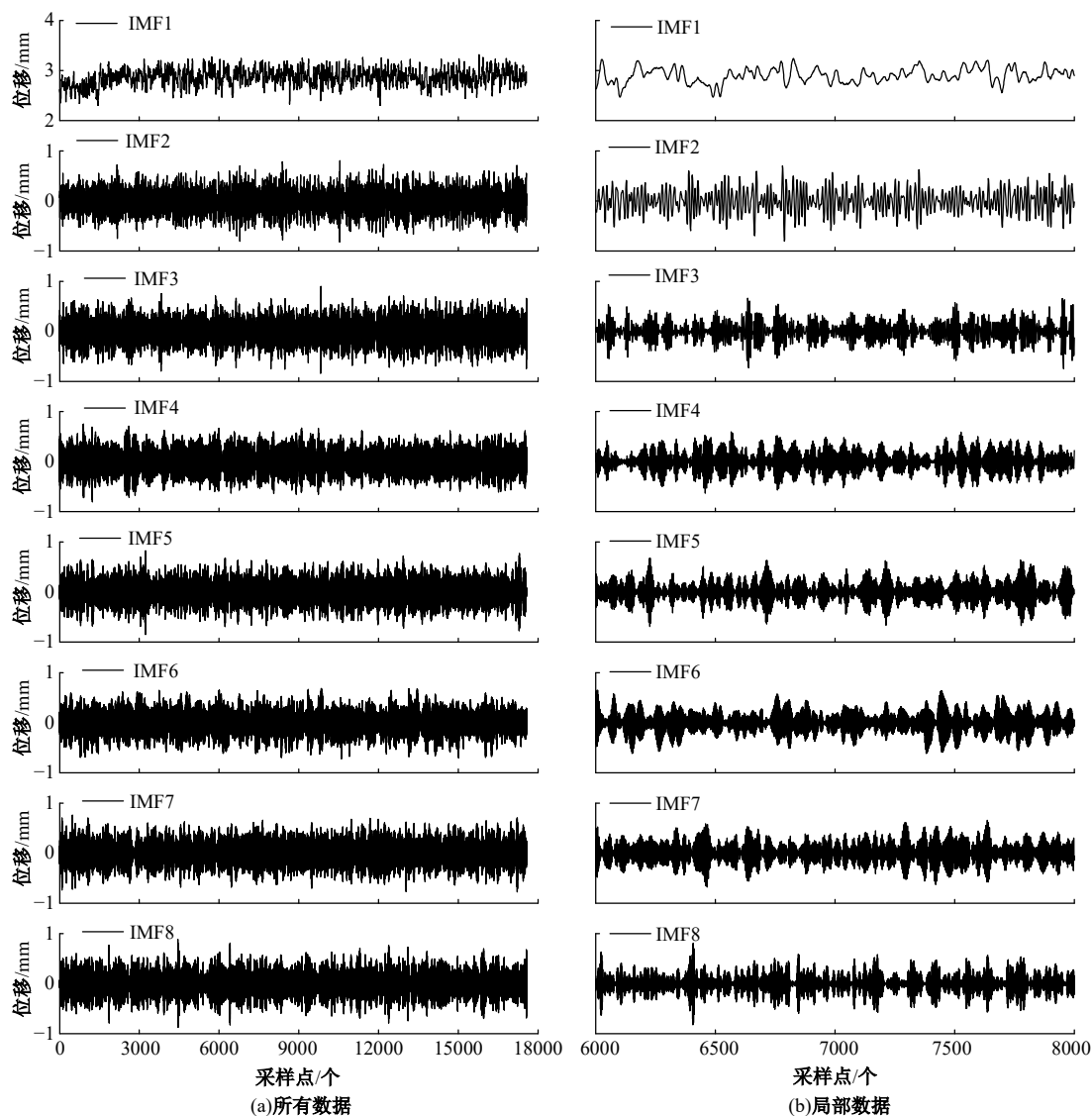


图4 RTK13边坡位移监测数据VMD

Fig. 4 VMD of RTK13 slope displacement monitoring data

势;IMF2反映了边坡受自身因素影响的周期性变化规律,IMF1和IMF2代表了边坡位移总体趋势,将其记为趋势位移;IMF3~IMF5具有相似的变化趋势,且数据波动较小,可以将其看作是受环境因素(如降雨、大风、温度等)影响的周期性位移;IMF6~IMF8的数据波动频率较大,规律性较弱,可将其看作是受仪器自身缺陷或者人为因素等弱相关因素影响的随机波动性位移。

2.3 边坡位移预测

选取VMD得到的IMF分量、趋势位移、周期性位移、随机波动性位移,RTK测得的X、Y、Z方向的边坡位移,固定式测斜仪及渗压计采集的数据作为特征输入预测模型中。为提升模型的收敛速度、稳定性及泛化能力,对上述输入特征进行归一化处理。为保证实验的可重复性,利用随机抽样方法并固定随机数种子为42,将数据按80%训练集和20%测试集的比例划分。

在模型训练过程中,利用网格搜索确定模型最优参数,得到XGBoost调参的最优结果如下:booster="gbtree", n_estimators=300, learning_rate=0.1, max_depth=5。XGBoost算法中,对于非线性、不平稳的边坡位移数据回归问题^[19],本文选用gbtree作为模型提升器^[20]。通过学习曲线和网格搜索调参,配合10折交叉验证,确定n_estimators为300, learning_rate为0.1, max_depth为5。这些参数的设置既防止了过拟合,又确保了模型的预测精度,满足数据的非线性特征和预测需求。XGBoost模型预测效果如图5所示。

如图5(a)所示,XGBoost模型得到的预测值与真实值构成的点与等值线偏离较小,散点较为均匀的分布在等值线两侧,且异常偏离点较少;如图5(b)所示,预测值与真实值的数值大小相近,变化趋势一致,预测误差较小,拟合程度较高;利用SHAP对输入特征变量进行重要度解释,如图5(c)所示趋势位移、周期性位移、随机波动性位移,以及RTK测得的X、Y、Z方向的边坡位移,对输出结果的影响重要度明显高于其他特征,其中周期性位移的影响重要度最高。为了提升模型的预测精度,保证模型的可解释性,本文选用经过SHAP值筛选后的特征作为输入变量。

为充分验证本文预测模型及特征选择方式在

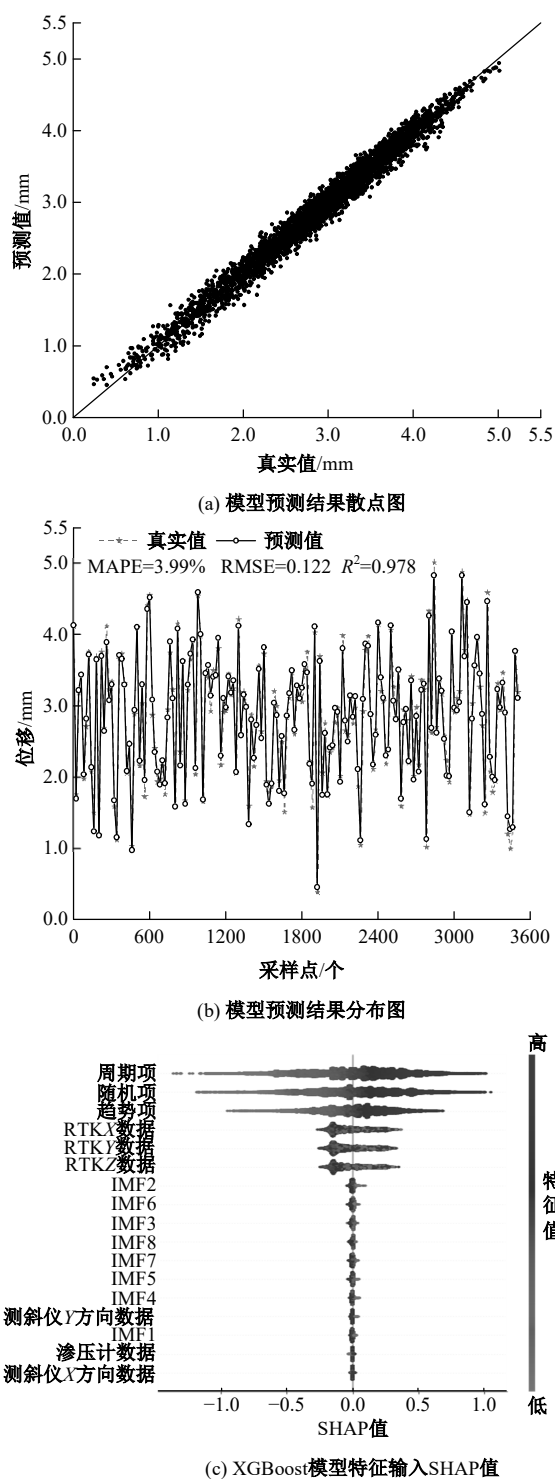


图5 XGBoost模型预测效果

Fig. 5 XGBoost model prediction effect

公路边坡位移预测中的优越性,选取参考文献[5]提出的XGBoost模型、参考文献[3]提出的SVM模型、参考文献[2]提出的ELM模型及其输入特征进行对比,结果如图6所示。从图6中可观察到,文献[5]提出的以渗压计数据和边坡位移 $t-1$ 、 $t-2$ 时刻的数据作为输入特征,用以预测 t 时刻

边坡位移的XGBoost模型,预测效果较差,大部分预测值在3 mm附近。其原因可能是本文中渗压计数据波动较小,大部分时间保持在同一数值,而边坡位移的变化波动较大,从而导致XGBoost模型预测效果降低,而经过本文方法选择特征后的XGBoost模型预测精度大幅度提升。文献[3]提出的将边坡位移数据转化为多维数据的SVM模型,在本文中的预测效果较好,总体的拟合效果较优。但是经过本文方法进行特征筛选后,其预测精度进一步提升。文献[2]提出的将边坡位移

数据经VMD后的各子序列作为输入特征的ELM模型,在本文中预测效果较差。根据本文利用SHAP对输入特征进行重要度解释(参考图5(c))可知,分解得到的IMF分量对模型预测的贡献度均较低,直接利用IMF分量进行预测会导致预测精度下降,而利用本文筛选后的特征进行预测,ELM模型的预测精度得到了大幅提升。这也进一步验证了特征选择的重要性,以及本文特征选取方式的合理性。

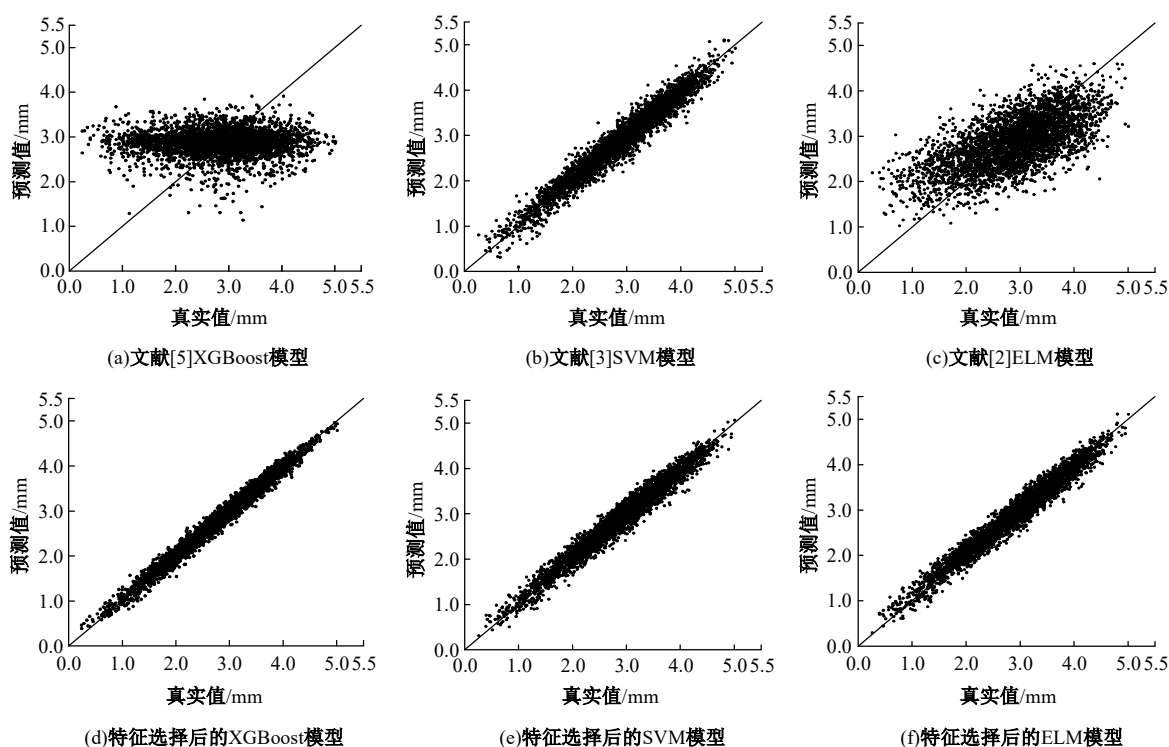


图6 不同输入特征模型预测结果散点图

Fig. 6 Scatterplot of prediction results of different input feature models

为提高真实值与预测值变化趋势的展示效果,对本文预测得到的3 514个采样点进行每20个点采样1次的重采样可视化,如图7所示。由图7可知,文献[5]XGBoost模型经过本文方法特征选择后,预测值与真实值的拟合程度更好,预测精度更高;文献[3]SVM模型经过本文方法特征选择后,预测精度更高,对某些极端点的预测效果更好;文献[2]ELM模型经过本文方法特征选择后,无论是变化趋势还是预测值大小均与真实值的情况基本一致。这些结果充分说明特征选择对于预测效果的重要性。

2.4 模型方法适用性及鲁棒性检验

为验证本文方法在高速公路边坡位移预测中

的适用性及鲁棒性,以陕西省境内的A边坡及B边坡为案例,选取常见的边坡位移数据分解算法如EMD^[8]、EEMD^[9]和VMD^[2],以及边坡位移预测算法如XGBoost^[5]、SVM^[3]和ELM^[2],进行回归预测模型组合分析。最终评价各组合预测模型的预测效果,并与其他文献提出的模型方法进行对比。

为了避免模型超参数寻优及计算环境等偶然因素对模型预测结果的影响,本文在相同运行环境下,对每种预测模型的10次运行结果取平均值,将其作为最终的预测值,相关评价结果如表2所示。

(1)对比分析边坡位移数据分解方式。在A

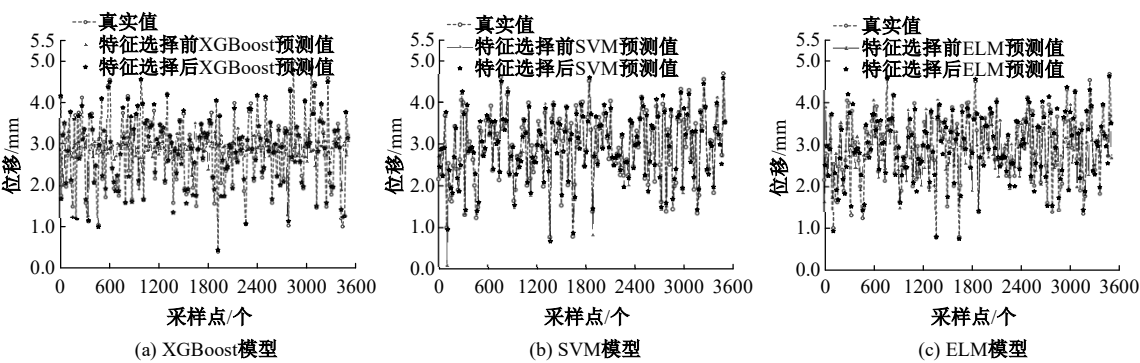


图7 不同输入特征模型预测结果点线图

Fig. 7 Dotted line plots of model prediction results for different input features

边坡和B边坡实际案例中,VMD较EMD和EEMD数据分解方式,在XGBoost、ELM和SVM回归预测模型中的多数情况下预测效果更优。这说明VMD作为公路边坡数据分解方式的优越性,同时也证明了本文选择VMD作为公路边坡位移数据分解方式的合理性及适用性。

(2)对比分析边坡位移预测模型。在两个边

坡的实际案例中,无论采用何种数据分解方式,XGBoost回归预测模型的预测效果均优于ELM和SVM回归预测模型。这说明XGBoost模型在公路边坡位移预测中具有更显著的优越性,同时也证明了本文选用XGBoost回归预测模型的合理性,以及该模型在公路边坡位移预测中的适用性。

表2 不同预测模型评价结果

Table 2 Different predictive modelling evaluation results

边坡类型	数据分解方式	预测模型	模型	R^2	RMSE/mm	MAPE/%
A 边坡	EMD	XGBoost	EMD-SHAP-XGBoost	0.969	0.148	4.93
		SVM	EMD-SHAP-SVM	0.968	0.149	12.03
		ELM	EMD-SHAP-ELM	0.935	0.212	6.57
	EEMD	XGBoost	EEMD-SHAP-XGBoost	0.979	0.123	4.00
		SVM	EEMD-SHAP-SVM	0.978	0.123	9.74
		ELM	EEMD-SHAP-ELM	0.967	0.152	4.77
	VMD	XGBoost	PSO-VMD-SHAP-XGBoost	0.980	0.119	3.74
		SVM	VMD-SHAP-SVM	0.958	0.166	5.17
		ELM	VMD-SHAP-ELM	0.959	0.170	5.46
	—	XGBoost	XGBoost ^[5]	0.010	0.843	32.84
	—	SVM	SVM ^[3]	0.944	0.197	6.27
	—	ELM	ELM ^[2]	0.417	0.636	22.32
B 边坡	EMD	XGBoost	EMD-SHAP-XGBoost	0.980	0.120	3.96
		SVM	EMD-SHAP-SVM	0.980	0.120	9.66
		ELM	EMD-SHAP-ELM	0.935	0.209	6.17
	EEMD	XGBoost	EEMD-SHAP-XGBoost	0.992	0.076	2.48
		SVM	EEMD-SHAP-SVM	0.992	0.077	6.12
		ELM	EEMD-SHAP-ELM	0.909	0.247	8.42
	VMD	XGBoost	PSO-VMD-SHAP-XGBoost	0.998	0.035	1.21
		SVM	VMD-SHAP-SVM	0.998	0.035	1.31
		ELM	VMD-SHAP-ELM	0.941	0.199	6.28
	—	XGBoost	XGBoost ^[5]	0.010	0.819	30.98
	—	SVM	SVM ^[3]	0.549	0.549	18.97
	—	ELM	ELM ^[2]	0.557	0.545	18.93

(3)对比分析公路边坡位移组合预测模型。在A边坡案例中,本文提出的PSO-VMD-SHAP-XGBoost模型的预测效果最好。在B边

坡案例中,PSO-VMD-SHAP-XGBoost和VMD-SHAP-SVM模型的预测效果优于其他模型,尽管二者的 R^2 和RMSE数值相同,但PSO-VMD-

SHAP-XGBoost模型的MAPE较VMD-SHAP-SVM模型更小。这说明 PSO-VMD-SHAP-XGBoost模型对于不同样本的相对误差较小,其预测效果更优。

从A、B两个边坡的预测结果可知,本文模型方法在高速公路边坡位移预测中具有良好的适用性和鲁棒性。在预测模型输入特征选择上,其他文献提出的特征选择方法在本文实际案例中大多表现较差,而经过本文特征选择后的预测模型的精度多数得到一定程度提升,说明了本文特征选择方式的优越性及合理性。

3 结 论

(1)对于非线性、不平稳的高速公路边坡位移数据,本文结合 PSO对其进行VMD,并将分解得到的各子序列划分为具有实际物理意义的趋势位移、周期性位移和随机波动性位移,这种有实际物理意义的位移数据对模型的预测起促进作用。

(2)在A边坡案例中,利用SHAP值对输入XGBoost模型的特征变量进行重要度解释,其中周期性位移、随机波动性位移和趋势位移的重要度位列前三;同时,根据SHAP值筛选输入特征变量后的XGBoost模型、SVM模型和ELM模型预测精度均得到一定程度的提升。这不仅说明了利用SHAP筛选输入特征的有效性,也说明了本文划分趋势位移、周期性位移和随机波动性位移的合理性。

(3)在本文公路边坡案例中,对于边坡位移数据的分解,VMD较EMD和EEMD具有更强的适用性;对于边坡位移数据的预测,XGBoost模型较ELM和SVM模型具有更高的预测精度;本文提出的 PSO-VMD-SHAP-XGBoost回归预测模型,较其他模型预测精度更高,且鲁棒性较强。

(4)其他文献提出的特征选取方法在本文边坡位移数据中的预测效果较差,而利用本文方法进行特征选择后的模型预测精度均有一定程度提升。这说明,在公路边坡位移预测过程中,需结合工程实际情况选择合适的输入特征和预测模型,才能使模型的预测效果达到最优。

参考文献:

- [1] 张勤,白正伟,黄观文,等. GNSS滑坡监测预警技术进展[J]. 测绘学报, 2022, 51(10): 1985-2000.
Zhang Qin, Bai Zheng-wei, Huang Guan-wen, et al. Review of GNSS landslide monitoring and early warning[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(10): 1985-2000.
- [2] 李博,李欣,芮红,等. 基于变分模态分解和灰狼优化极限学习机的隧道口边坡位移预测[J]. 吉林大学学报:工学版, 2023, 53(6): 1853-1860.
Li Bo, Li Xin, Rui Hong, et al. Displacement prediction of tunnel entrance slope based on variational modal decomposition and grey wolf optimized extreme learning machine[J]. Journal of Jilin University(Engineering and Technology Edition), 2023, 53(6): 1853-1860.
- [3] 李建新,刘小生,肖钢,等. 基于PSR-WSVM模型的边坡位移预测[J]. 大地测量与地球动力学, 2020, 40(6): 577-580.
Li Jian-xin, Liu Xiao-sheng, Xiao Gang, et al. Slope displacement prediction based on PSR-WSVM model[J]. Journal of Geodesy and Geodynamics, 2020, 40(6): 577-580.
- [4] 张凯,张科. 基于LightGBM算法的边坡稳定性预测研究[J]. 中国安全科学学报, 2022, 32(7): 113-120.
Zhang Kai, Zhang Ke. Prediction study on slope stability based on LightGBM algorithm[J]. China Safety Science Journal, 2022, 32(7): 113-120.
- [5] 黄智杰,沈佳,简文彬,等. 基于XGBoost模型的降雨诱发阶跃型滑坡位移预测[J]. 自然灾害学报, 2023, 32(2): 217-226.
Huang Zhi-jie, Shen Jia, Jian Wen-bin, et al. Displacement prediction of rainfall-induced step-like landslide based on XGBoost model[J]. Journal of Natural Disasters, 2023, 32(2): 217-226.
- [6] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]//Conference and Workshop on Neural Information Processing Systems. California, Long Beach, USA, 2017: 4765-4774.
- [7] 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.
Zhou Chao, Yin Kun-long, Huang Fa-ming. Application of the chaotic sequence WA-ELM coupling model in landslide displacement prediction[J]. Rock and Soil Mechanics, 2015, 36(9): 2674-2680.
- [8] 任超,梁月吉,庞光锋,等. 经验模态分解和遗传小波神经网络法用于边坡变形预测[J]. 测绘科学技术学报, 2014, 31(6): 551-555.
Ren Chao, Liang Yue-ji, Pang Guang-feng, et al. The empirical mode decomposition and genetic algo-

- rithm wavelet neural network for slope deformation prediction research[J]. Journal of Geomatics Science and Technology, 2014, 31(6): 551-555.
- [9] 谢博, 施富强, 廖学燕, 等. 边坡位移的 EEMD-PSO-ELM 模型预测方法[J]. 中国安全科学学报, 2020, 30(3): 157-162.
- Xie Bo, Shi Fu-qiang, Liao Xue-yan, et al. Slope displacement prediction method based on EEMD-PSO-ELM model[J]. China Safety Science Journal, 2020, 30(3): 157-162.
- [10] 周福成, 唐贵基, 何玉灵. 基于改进 VMD 的风电齿轮箱不平衡故障特征提取[J]. 振动与冲击, 2020, 39(5): 170-176.
- Zhou Fu-cheng, Tang Gui-ji, He Yu-ling. Unbalanced fault feature extraction for wind power gearbox based on improved VMD[J]. Journal of Vibration and Shock, 2020, 39(5): 170-176.
- [11] 蔡艳平, 范宇, 陈万, 等. 改进时频分析和特征融合在内燃机故障诊断中的应用[J]. 中国机械工程, 2020, 31(16): 1901-1911.
- Cai Yan-ping, Fan Yu, Chen Wan, et al. Applications of improved time-frequency analysis and feature fusion in fault diagnosis of IC engines[J]. China Mechanical Engineering, 2020, 31(16): 1901-1911.
- [12] 唐贵基, 王晓龙. 参数优化变分模态分解方法在滚动轴承早期故障诊断中的应用[J]. 西安交通大学学报, 2015, 49(5): 73-81.
- Tang Gui-ji, Wang Xiao-long. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing[J]. Journal of Xi'an Jiaotong University, 2015, 49(5): 73-81.
- [13] Dragomiretskiy K, Zosso D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [14] 黄钦, 谭翠, 杨波. 基于 XGBoost 算法的亚热带地区生态旅游适宜性评价方法研究[J]. 地球信息科学报, 2024, 26(2): 303-317.
- Huang Qin, Tan Cui, Yang Bo. Research on the evaluation method of ecotourism suitability in subtropical regions based on XGBoost algorithm[J]. Journal of Geo-information Science, 2024, 26(2): 303-317.
- [15] 闫孟婷, 毛玉鑫, 张天遥, 等. 基于 SHAP-XGBoost 混合模型的梯级水电站流量动态滞时研究[J]. 电力系统保护与控制, 2023, 51(11): 159-167.
- Yan Meng-ting, Mao Yu-xin, Zhang Tian-yao, et al. Dynamic hysteresis of cascade hydropower station discharge based on SHAP-XGBoost hybrid model[J]. Power System Protection and Control, 2023, 51(11): 159-167.
- [16] 杨川, 林日成, 季建勇, 等. 基于图深度学习与北斗监测的边坡位移预测研究[J]. 工程地质学报, 2024, 32(2): 612-622.
- Yang Chuan, Lin Ri-cheng, Ji Jian-yong, et al. Slope displacement prediction research based on the graph deep learning and beidou monitoring[J]. Journal of Engineering Geology, 2024, 32(2): 612-622.
- [17] 张俊, 张建群, 钟敏, 等. 基于 PSO-VMD-MCKD 方法的风机轴承微弱故障诊断[J]. 振动、测试与诊断, 2020, 40(2): 287-296, 418.
- Zhang Jun, Zhang Jian-qun, Zhong Min, et al. PSO-VMD-MCKD based fault diagnosis for incipient damage in wind turbine rolling bearing[J]. Journal of Vibration, Measurement & Diagnosis, 2020, 40(2): 287-296, 418.
- [18] 赵博超, 马嘉骏, 崔磊, 等. 基于改进 VMD-XGBoost-BILSTM 组合模型的光伏发电异常检测方法[J]. 计算机工程, 2024, 50(3): 306-316.
- Zhao Bo-chao, Ma Jia-jun, Cui Lei, et al. Photovoltaic anomaly detection based on improved VMD-XGBoost-BILSTM hybrid model[J]. Computer Engineering, 2024, 50(3): 306-316.
- [19] 徐开心, 戴宁, 汝欣, 等. 基于概率分布和 XGBoost 决策算法的织机异常数据处理方法[J]. 计算机集成制造系统, 2025, 31(6): 2021-2130.
- Xu Kai-xin, Dai Ning, Ru Xin, et al. Loom abnormal data processing method based on probability distribution and XGBoost decision algorithm[J]. Computer Integrated Manufacturing Systems, 2025, 31(6): 2021-2130.
- [20] 吴飞, 王鹏程, 杨康. 基于 PCA-SSA-XGBoost 的车辆驾驶性评估[J]. 吉林大学学报: 工学版, 2025, 55(1): 355-365.
- Wu Fei, Wang Peng-cheng, Yang Kang. Drivability evaluation model based on PCA-SSA-XGBoost[J]. Journal of Jilin University(Engineering and Technology Edition), 2025, 55(1): 355-365.