

DOI:10.11918/202406032

# GA-XGBoost 模型对路基压实质量的预测

赖建平<sup>1</sup>, 赵辉<sup>1</sup>, 王东升<sup>2</sup>, 冯怀平<sup>1,3</sup>

(1. 石家庄铁道大学 土木工程学院, 石家庄 050043; 2. 哈尔滨工业大学 交通科学与工程学院, 哈尔滨 150090;  
3. 省部共建交通工程结构力学行为与系统安全国家重点实验室(石家庄铁道大学), 石家庄 050043)

**摘要:** 为提升智能压实(intelligent compaction, IC)质量的实时检测与评价精度,提出一种基于 GA-XGBoost 模型的连续压实质量预测方法,以提高动态变形模量( $E_{vd}$ )的预测精度。模型以动态变形模量为目标,建立机器学习模型,主要采用决策树算法,构建 XGBoost 模型对压实质量进行预测分析。通过引入遗传算法(genetic algorithm, GA)对模型超参数寻优,以提高模型的预测精度和可靠性。首先,通过现场工程试验,测量压路机碾压时振动加速度,分析加速度信号,计算信号统计量并采用快速傅里叶变换(FFT)得出谐波频率,初步建立各项特征因子与  $E_{vd}$  之间的系统联系;其次,筛选各个时频域特征,进行相关性分析,选用相关性较高的特征来建立预测模型;最后,验证了 GA-XGBoost 预测模型可以较好的预测  $E_{vd}$ 。研究表明:遗传算法(GA)可以高效地确定 XGBoost 算法的超参数,且较单一的 XGBoost 模型表现出更优的收敛速度;通过优化特征因子,改变输入参数,提高了 GA-XGBoost 模型的预测精度,优化后均方误差为 3.9%,相关系数为 0.748;同时对比了传统 CMV 拟合  $E_{vd}$  的方法,该机器学习模型可以大幅度提高预测精度。

**关键词:** 智能压实;机器学习;XGBoost 算法;遗传算法;动态变形模量;时域特征

**中图分类号:** TU18 **文献标志码:** A **文章编号:** 0367-6234(2025)07-0033-09

## Prediction of dynamic deformation modulus of subgrade based on GA-XGBoost model

LAI Jianping<sup>1</sup>, ZHAO Hui<sup>1</sup>, WANG Dongsheng<sup>2</sup>, FENG Huaiping<sup>1,3</sup>

(1. School of Civil Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;  
2. School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China;  
3. State Key Laboratory of Mechanical Behavior and System Safety of Traffic Engineering Structures  
(Shijiazhuang Tiedao University), Shijiazhuang 050043, China)

**Abstract:** In order to improve the real-time detection and evaluation accuracy of intelligent compaction (IC) quality, a continuous compaction quality prediction method based on GA-XGBoost model was proposed to improve the prediction accuracy of dynamic deformation modulus ( $E_{vd}$ ). The model takes the dynamic deformation modulus as the goal, establishes a machine learning model, mainly uses the decision tree algorithm, and constructs the XGBoost model to predict and analyze the compaction quality. In order to improve the prediction accuracy and reliability of the model, genetic algorithm (GA) is introduced to optimize the hyperparameters of the model. Firstly, through the field engineering test, the vibration acceleration of the roller is measured, the acceleration signal is analyzed, the signal statistics are calculated and the harmonic frequency is obtained by fast Fourier transform (FFT), and the system connection between the characteristic factors and  $E_{vd}$  is preliminarily established. Secondly, the characteristics of each time-frequency domain are screened, the correlation analysis is carried out, and the characteristics with high correlation are selected to establish the prediction model. Finally, it is verified that the GA-XGBoost prediction model can better predict  $E_{vd}$ . The results show that the genetic algorithm (GA) can efficiently determine the hyperparameters of the XGBoost algorithm, and it shows better convergence speed than the single XGBoost model. By optimizing the feature factors and changing the input parameters, the prediction accuracy of the GA-XGBoost model is improved. The optimized mean square error is 3.9% and the correlation coefficient is 0.748. At the same time, compared with the traditional CMV fitting  $E_{vd}$  method, the machine learning model can greatly improve the prediction accuracy.

**Keywords:** intelligent compaction; machine learning; XGBoost algorithm; genetic algorithm; dynamic deformation modulus; time domain characteristics

收稿日期: 2024-06-14; 录用日期: 2024-08-29; 网络首发日期: 2025-07-08

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250708.1457.018>

基金项目: 国家自然科学基金联合基金项目(U22A20233)

作者简介: 赖建平(1999—),男,硕士研究生;冯怀平(1975—),男,教授,博士生导师

通信作者: 冯怀平, fenghuaiping@stdu.edu.cn

智能压实(IC)技术是人工智能技术应用于路基压实质量控制的成功典范,它提供了一种实时检测路基压实过程的方法。IC主要通过GPS定位技术,振动信息反馈以及信息化处理等手段对路基压实质量进行检测,已广泛应用于机场、铁路、公路、大坝等建设中。但现在智能压实测试精度问题依然是大家关注的热点,因此,探索提升压实质量预测精度方法对IC技术推广具有重要意义。

压实测试精度问题研究主要分为两个方面。一方面,在传统压实检测方法<sup>[1]</sup>的基础上提出新的检测思路。聂志红等<sup>[2]</sup>通过现场试验基于双曲面响应法,研究了压路机振动频率、振动幅度、行驶速度等因素对检测数据的影响显著性。张德等<sup>[3]</sup>通过工程现场试验分析不同压实情况及碾压参数对智能压实质量的影响。吴龙梁等<sup>[4]</sup>提出以压实合格率为指标建立多元线性回归模型和质量控制模型,采用简单克里金算法,分析了两种模型可以有效预测智能压实的压实情况。冯振刚等<sup>[5]</sup>基于多元回归模型可以有效预测沥青路面的压实情况。马源等<sup>[6]</sup>基于有限元仿真软件对影响CMV的一系列碾压参数进行了正交试验,对CMV变化规律定性分析,并从计算角度剖析其变化机理。蔡德钧等<sup>[7]</sup>基于现场试验从时域、频域、能量以及时频域的角度分析智能压实振动过程中谐波的变化情况。目前,大多数学者已经从多种渠道构建了路基压实度与智能压实测试值之间的关系,但是这些方法大都基于动力学模型基础上,由于现场作业时偶然性较大以及各种因素如填料含水率、压路机碾压参数等的共同影响,上述方法难以同时考虑各种因素的影响以及保证检测参数的稳定和量测结果的精度,具有一定的局限性。

另一方面,在人工智能技术的基础上探究新的检测方法。部分学者发现机器学习方法对于处理复杂多变的情况具有一定优势,因此,该方法在行业内迅速发展。其与工程实例已结合的越来越紧密。其中,李凌志等<sup>[8]</sup>采用自主采样的机器学习方法,引入K均值聚类算法对高速公路的裂缝进行预测。而在智能压实领域,通过建立机器学习模型,将碾压参数如振动速度、加速度、激振力、频率等引入其中,对现场作业时压实度进行预测。Wang等<sup>[9-12]</sup>基于人工神经网络模型,通过改变不同参数如土壤特性、激振力、振动频率等去预测压实度,并通过现场试验验证预测精度。Cao等<sup>[13]</sup>根据压路机的碾压参数基于人工神经网络模型对CMV进行预测,验证了神经网络模型预测CMV的可行性和潜力。安再展等<sup>[14]</sup>基于现场试验和建立径向基神经网络模型,预

测分析了碾压参数实时变化时的CMV,经过与试验结果对比,验证了该神经网络模型的精度。陈晓斌等<sup>[15]</sup>提出一种基于振动能量最小原则的优化方法改进神经网络模型并采用遗传算法寻出最佳参数,对优化高速铁路基碾压技术参数具有较好的改进。黄国卿等<sup>[16]</sup>基于BP神经网络模型,通过CMV、CCV、THD 3种指标作为输出参数预测土体压实度,相较于传统方法大幅度提升了连续压实均匀性表征能力和压实度的预测精度。宿辉等<sup>[17]</sup>基于GA-PSO混合优化BP神经网络模型预测堆石坝压实质量,经过对比分析该优化方式具有更好的性能,精度相对更高。但也可以看出尽管机器学习方法具有较好的效果,单一神经网络方法必然影响其普适性,且神经网络的隐藏层数量均由经验值取得,无法适应复杂的施工工况,需要耗费大量时间去提高精度。

近年来,部分研究者发现决策树算法对处理这类回归预测问题具有较好的能力,该算法具有可视化分析过程,相比于神经网络算法的“黑箱模型”具有更好的调控能力,另外其对于处理具有缺失值数据的样本也有更好的表现。根据文献<sup>[12]</sup>提出的XGBoost算法,主要将大量决策树组合,即使用一系列弱分类器形成强分类器,该算法保留上述决策树算法的优点,且弥补单一的决策树存在的不足。同时,为降低该算法过拟合的风险,决定采用遗传函数(GA),该方法是一种高效的全局搜索优化算法,可自适应搜索并寻找全局最优解,不会陷入局部最优,用该方法确定XGBoost算法的超参数并且对决策树进行有效剪枝以及加入正则化过程可以有效降低过拟合的风险,提高对 $E_{vd}$ 的预测精度。

本文采用大量智能压实试验数据,基于决策树算法建立的XGBoost预测模型,并引入GA确定XGBoost模型超参数。该模型将加速度信号特征与CMV及 $E_{vd}$ 统一结合,以 $E_{vd}$ 为目标,对智能压实质量进行预测分析,得到一种有效的压实质量预测模型,并传统压实质量分析方法对比,验证了该模型具有高效、准确的优点,具有一定的可行性以及实用性。

## 1 试验

### 1.1 工程背景与设备

现场试验在秦唐高速工程秦皇岛段开展,碾压设备采用徐工-XS263H型振动压路机。其中,工作质量为30 t;振动钢轮质量为13 t;振动轮宽度为2 170 mm;振动轮直径为1 600 mm;振动频率为27 Hz或32 Hz;激振力为209 kN或405 kN。路基填

料为含粗粒的粉质土,液限为 24.6%,塑限为 16.5%,最优含水率为 14.2%,最大干密度为 1.94 g/cm<sup>3</sup>,比重为 2.63,筑填厚度为 30 cm。 $E_{vd}$ 采用轻型落锤仪进行检测,并通过 RTK 测量仪记录其测点坐标。具体试验现场如图 1 所示。



(a) 压路机 (b) 压实过程



(c) RTK 标定坐标

图 1 试验现场

Fig. 1 Test field

本次试验共采用 4 个加速度传感器,采用 AB 胶黏结固定在振动轮轴心及车架上,双侧安装,安装示意图见图 2。传感器型号为美国 PCB 高灵敏度 365A16 triaxial ICP 三向加速度传感器,其参数为:灵敏度 25 mV/g;最大量程为 200 m/s<sup>2</sup>;设置加速度采样频率为 1 024 Hz。

本次试验所采用的常规检测指标  $E_{vd}$ ,是由德国进口 HMP 型动态变形模量测试仪测得,如图 3 所示,落锤式弯沉测试仪主要由落锤、承载板、测振传感器及沉陷测定仪组成,承载板直径为 300 mm,厚

度为 2 cm,有效测试深度为 40 ~ 50 cm,  $E_{vd}$  测试量程为 0 ~ 225 MPa,沉降测试量程为 0.1 ~ 2.0 mm,适用温度为 0 ~ 40 °C,该测试仪操作方法简单,检测时间较快,测试前进行 3 次预冲击,随后进行 3 次正式冲击测试,通过承载板内部压力和位移传感器的数值,读数仪自动显示  $E_{vd}$  数值。

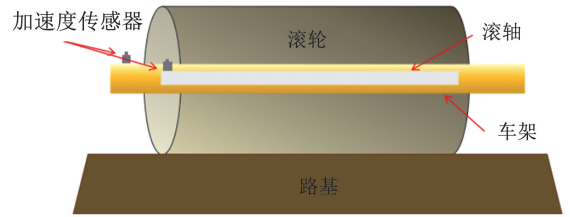


图 2 加速度传感器安装示意图(单侧)

Fig. 2 Acceleration sensor installation diagram (half)



图 3 动态变形模量测试仪

Fig. 3 Dynamic deformation modulus tester

试验测点布置见图 4。设置尺寸为 2 m × 20 m 的碾压带,间隔 2 m 布置一个  $E_{vd}$  测点,振动轮左右两侧布置,共 8 组;间隔 5 m 布置一个灌砂法测点,共 3 组。

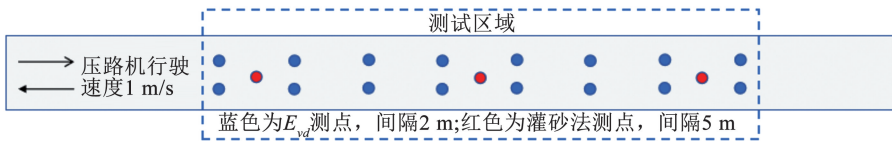


图 4 测点布置

Fig. 4 Measuring point arrangement

### 1.2 试验方案

试验场地为 20 m × 20 m 的矩形场地,压实条带为 2 m × 20 m,共设置 6 条碾压条带。试验振动压实 6 遍,分别是 3 遍弱振,3 遍强振。具体碾压方案如表 1 所示。每遍振动压实后,采用灌砂法对路基进行压实度检测;预先在碾压条带标定位置,采用轻型落锤仪进行检测,并记录加速度信号与压实度。

表 1 试验方案

Tab. 1 Test scheme

振压遍数	激振力/kN	频率/Hz	行车速度/(m·s <sup>-1</sup> )
1	209	27	1
2	209	27	1
3	209	27	1
4	405	31	1
5	405	31	1
6	405	31	1

### 1.3 压实度检测分析

根据室外智能压实试验方案,在每次碾压之后,根据条带预先布置的灌砂点位进行挖坑灌砂测试路基压实度,每条条带共 3 个测试点位,依次按照 1、2、3 顺序采样,将 3 次采样结果取平均值得到本次碾压后该条带的压实度值。压实度计算公式为

$$K = \frac{\rho_d}{\rho_c} \times 100\% \quad (1)$$

式中: $K$  为碾压材料的压实度; $\rho_c$  为碾压材料的最大干密度; $\rho_d$  为试坑材料的干密度, $\rho_d = \frac{m_d}{m_b} \times \rho_s$ ,其中, $m_b$  为填满试坑所需标准砂的质量, $m_d$  为试坑中取出材料的烘干质量, $\rho_s$  为标准砂的密度, $\rho_s = 1.529 \text{ g/cm}^3$ 。压实度与碾压遍数关系曲线如图 5 所示。

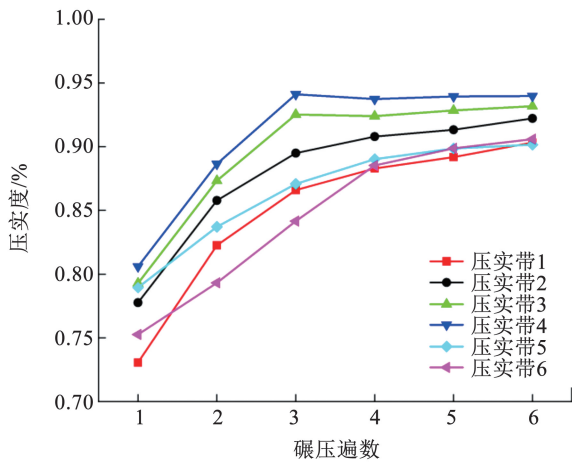


图 5 压实度与振压遍数关系

Fig. 5 The relationship diagram between compaction degree and vibration pressure times

由图 5 分析可得,经过 6 遍振动压实后,试验场地土体压实度逐渐增长且压实效果较好,压实度在 9% ~ 95% 之间,达到路基压实要求。

### 1.4 $E_{vd}$ 与压实度相关性分析

为了验证  $E_{vd}$  和压实度关系,本小节以压实度与  $E_{vd}$  的相关性研究出发,分析灌砂法所测得土体压实度和  $E_{vd}$  相关性,具体拟合曲线如图 6 所示。

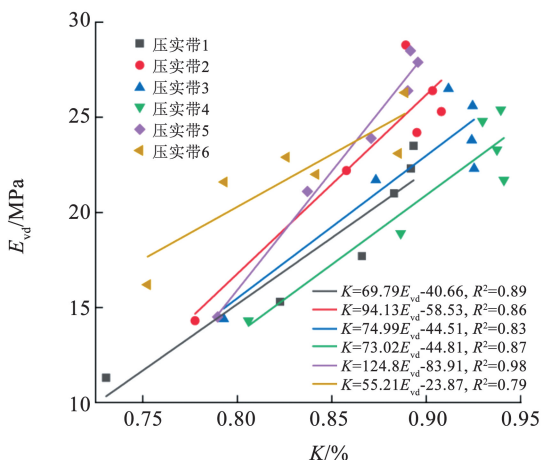


图 6 压实度- $E_{vd}$  相关性

Fig. 6 Correlation of between compaction and  $E_{vd}$

由图 6 可以发现, $E_{vd}$  和压实度存在较好的相关性,1 ~ 5 条带其  $R^2$  值都在 0.83 ~ 0.98 之间,表明常规压实指标  $E_{vd}$  可以替代压实度作为评判路基压实标准。第 6 条带压实度和  $E_{vd}$  相关性相较于其他条带,略微较弱,但  $R^2$  也接近 0.80,根据现行规范,该情况属于合理范畴。

## 2 GA-XGBoost 模型

### 2.1 XGBoost 算法原理介绍

XGBoost 算法是机器学习方法的一种,其本质为大量决策树组合而成,通过模型内嵌的函数计算方式,对数据进行特征提取,通过大量数据学习特征并建立损失函数,建立预测模型,直接进行输入、输出数据之间关联性的研究,对非线性的情况具有较高的预测精度。XGBoost 可以说是 GBDT (gradient boosting decision tree) 梯度提升树的一个改进版本。在文献[12]中,其算法原理如图 7 所示。

XGBoost 使用了一系列弱分类器(决策树)来构建一个强分类器。假设存在  $N$  个样本,每个样本包含  $M$  个特征,则 XGBoost 的目标函数可以表示为

$$F(\Theta) = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

式中: $l(y_i, \hat{y}_i)$  为损失函数,表示模型预测值  $\hat{y}_i$  和实测值  $y_i$  之间的差异; $\Omega(f_k)$  为正则化项,用于控制模型的复杂度。

模型的预测值计算公式为

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (3)$$

式中  $f_k(x_i)$  表示第  $k$  个决策树对于样本  $x_i$  的输出。

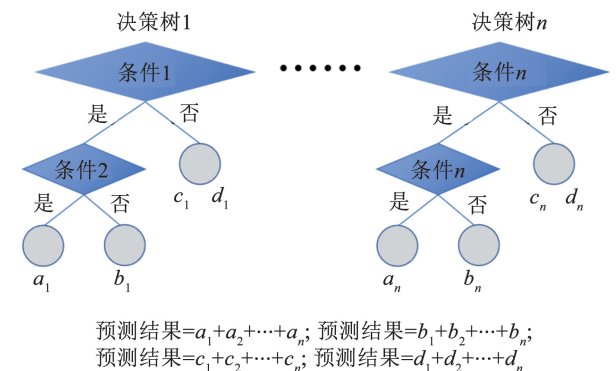


图 7 XGBoost 原理图

Fig. 7 XGBoost schematic diagram

### 2.2 遗传算法 (GA)

遗传算法是受自然进化理论启发的通过模仿自然选择和繁殖过程的搜索算法,主要分为选择、交叉、变异 3 个过程。在遗传算法中,每个个体都由代表基因集合的染色体构成,算法保留了大量的独立

个体,组成种群,即形成了针对给定问题的候选解集,通过算法迭代,利用适应度函数对个体进行评估,计算出适应度值,并以适应度高低形成对应概率,类似于轮盘赌,通过不同概率进行选择并将其遗传物质遗传给下一代。为了创建新的个体,将父本和母本部分基因交叉,传给子一代;在遗传过程可能产生基因突变,这就形成变异,其形式是通过随机改

变一个或多个基因型来实现。最后不断迭代,形成大量新个体,直到满足算法的终止条件。遗传算法(GA)具有智能式搜索、渐进式优化、全局最优解、通用性强、并行式运算等优点。基于此,选用遗传算法对 XGBoost 模型进行优化处理。基于遗传算法优化 XGBoost 模型主要流程如图 8 所示。

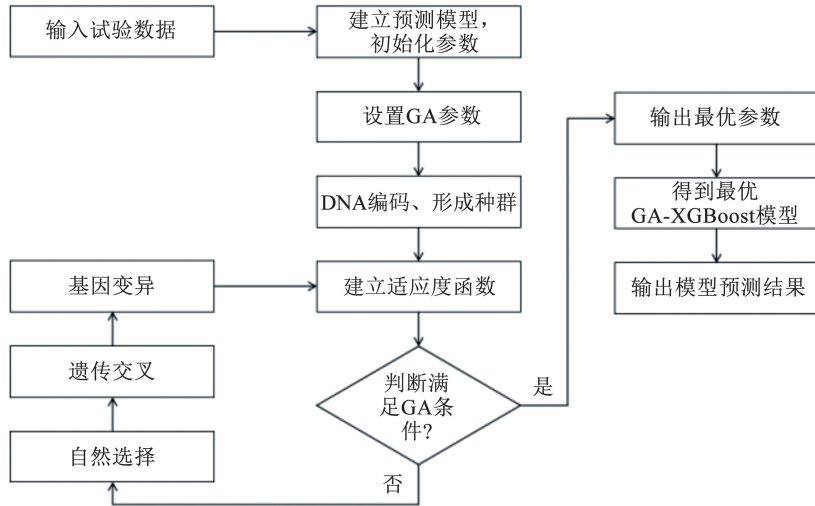


图 8 GA 优化 XGBoost 模型流程

Fig. 8 The flow of XGBoost model optimized by GA

### 2.3 数据预处理

#### 2.3.1 数据记录

在本研究中,基于高速公路压实工程的现场试验数据以及跟随压路机记录数据参数的统计,建立 GA-XGBoost 预测模型。影响路基压实的因素众多,压路机控制参数如碾压速度、碾压遍数、激振力等;土体的工程特性如颗粒级配、粒径、最大干密度、含水率等。以上因素对压实度的影响重要度不同,本文选用振动轮加速度信号预测压实度,加速度信号采集分为振动轮左右两侧和车架左右两侧,本次试验主要采用振动轮左右两侧 Y 轴方向上采集的加速度信号。研究人员记录每秒的加速度信号和该点对应的  $E_{vd}$ ,在大量加速度信号中,选取每秒 1 024 个加速度值以及对应的  $E_{vd}$ (图 9)。总共统计 318 组数据,且每组加速度信号都对应到该点位的  $E_{vd}$  值。

#### 2.3.2 信号处理

目前大多数研究人员对加速度信号特征分析多采用谐波法,因此,对于现场试验采集的加速度信号,利用 MATLAB 软件自编程序分析,编程主要代码功能为加速度信号部分统计量计算及快速傅里叶变换(FFT)下的时频域转换。其主要原理为

$$X(k) = \sum_{t=0}^{N-1} x(t) e^{-j\frac{2\pi kt}{N}}, 0 \leq k \leq N-1 \quad (4)$$

式中  $N$  表示时间序列。

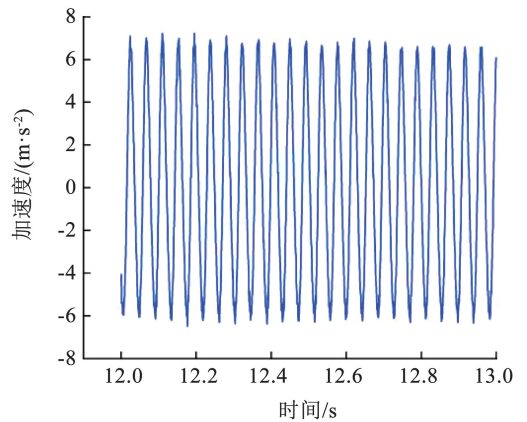


图 9 某一秒截取的加速度信号

Fig. 9 Acceleration signal intercepted in a second

统计量作为加速度信号的时域特征,各个计算方法见表 2,同时计算其谐波的总谐波失真(THD)、信噪比(SNR)、信号噪声和失真比(SINAD);FFT 其基本思想为将离散复杂的时域信号转化成频域信号,即加速度信号从随时间分布转化为随频率分布,纵轴对应为某一种频率下加速度幅值大小。压路机振动轮的加速度采集频率为 1 024 Hz。MATLAB 程序主要分为 4 个模块:加速度信号读取、统计量的计算、快速傅里叶变换与显示、导出数据。

表 2 各时域特征计算公式

Tab. 2 Calculation formula of each time domain feature

时域特征	计算方式
平均值	$\frac{1}{N} \sum_{i=1}^N s_i$
标准差	$\frac{1}{N} \sum_{i=1}^N \frac{(s_i - \bar{s})^2}{\rho_i^3}$
偏度	$\left( \frac{1}{N} \sum_{i=1}^N (s_i - \bar{s})^3 \right)^{\frac{1}{2}}$
峭度	$\frac{1}{N} \sum_{i=1}^N \frac{(s_i - \bar{s})^4}{\rho_i^4}$
峰峰值	$S_{\max} - S_{\min}$
均方根	$\left( \frac{1}{N} \sum_{i=1}^N s_i^2 \right)^{\frac{1}{2}}$
振幅因子	$S_{\max}/\text{RMS}$
波形因子	$\text{RMS} / \left( \frac{1}{N} \sum_{i=1}^N  s_i  \right)$
冲击因子	$S_{\max} / \left( \frac{1}{N} \sum_{i=1}^N  s_i  \right)$
裕度因子	$S_{\max} / \left( \frac{1}{N} \sum_{i=1}^N  s_i  \right)^2$

注:  $s_i$  为信号的采样数据点幅值,  $N$  为采样点样本个数,  $\bar{s}$  为数据点平均值,  $\rho_i$  为偏度,  $S_{\max}$  为最大值,  $S_{\min}$  为最小值。

## 2.4 GA-XGBoost 预测模型的建立

XGBoost 预测模型算法由 Scikit-learn 机器学习包实现, 训练集和测试集比例为 8:2, 并建立特征工程。建立搜索函数, 引入 GA, 构建适应度函数, 在经过选择、交叉、变异的不断循环后, 得到最优参数值, 让模型达到最优效果。其中, 模型参数对预测结果有较大的影响, 因此, 在 GA 优化后, 同时加入采用严谨的交叉验证, 并利用网格调参, 最优参数如表 3 所示。

表 3 XGBoost 模型最优参数

Tab. 3 Optimal parameters of XGBoost model

模型参数	最优值
Max depth	10
Learning rate	0.01
Random state	0
Min child weight	1
N estimators	502
Gamma	0
Subsample	0.8
Colsample bytree	0.75
Alpha	1

对预测模型的泛化能力进行评估, 不仅需要严谨可靠的试验方法, 还需要具有衡量泛化能力标准的评估指标。为量化评价模型预测结果的准确性, 采用相关系数 ( $R^2$ )、均方误差 (MSE)、平均绝对误差 (MAPE) 以及可解释方差 4 种指标共同描述 XGBoost 模型的预测结果。其评估指标数学上的计算方法如下:

$$R^2 = 1 - \frac{\sum (y - \hat{y}_i)^2}{\sum (y - \bar{y})^2} \quad (5)$$

$$\text{MSE} = \frac{1}{n} \times \sum (y_i - \hat{y}_i)^2 \quad (6)$$

$$\text{MAPE} = \frac{1}{n} \times \sum (|y_i - \hat{y}_i|) \quad (7)$$

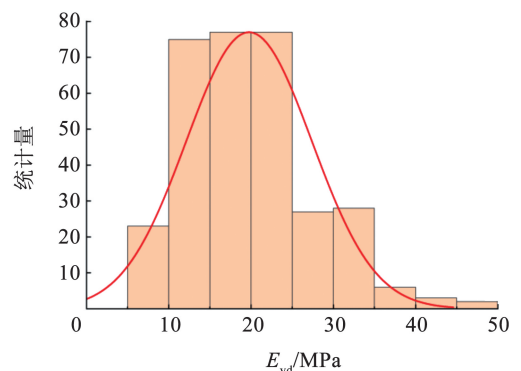
$$\text{可解释方差} = 1 - \frac{\text{Var}(y_i - \hat{y}_i)}{\text{Var}(y_i)} \quad (8)$$

式中: MSE 表示均方误差, MAPE 表示平均绝对误差,  $R^2$  表示相关系数,  $y_i$  为样本实测值,  $\hat{y}_i$  为样本预测值,  $\bar{y}_i$  为样本实测值均值,  $n$  表示样本的数量,  $\text{Var}(\cdot)$  表示方差计算。

## 3 数据分析与讨论

### 3.1 $E_{vd}$ 统计分析

GA-XGBoost 预测模型采用上述 19 个特征因子作为输入特征预测  $E_{vd}$  值。根据结果分析, 图 10 可以看出  $E_{vd}$  大致呈正态分布,  $E_{vd}$  在 10 ~ 25 MPa 之间的占比多, 在 30 ~ 45 MPa 之间的占比少; 同样, 图 11 箱型图全面综合地反映一组数据的集中趋势、离散程度、分布形状。箱内数据占总体数据 50%, 其中, 下限为 7.37 MPa, 上限为 38.4 MPa; 下四分位数  $Q_1$  为 13.77 MPa, 上四分位数  $Q_3$  为 23.65 MPa, 四分位数的差  $Q_3 - Q_1$  为 9.88 MPa, 此时箱子比较短, 可见数据比较集中; 中位数  $Q_2$  为 18.91 MPa, 平均水平比较低, 同时, 中位数在箱子的正中间, 数据呈正态分布; 整体均值高于中位数 19.75 MPa; 存在 5 个异常值。

图 10  $E_{vd}$  分布直方图Fig. 10  $E_{vd}$  value distribution histogram

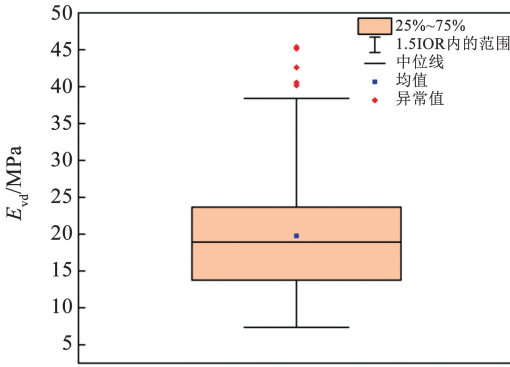


图 11  $E_{vd}$  数值分布情况箱形图

Fig. 11 Box plot of  $E_{vd}$  numerical distribution

### 3.2 相关性及重要性分析

由图 12 分析可得,在 19 个特征因子中,THD、

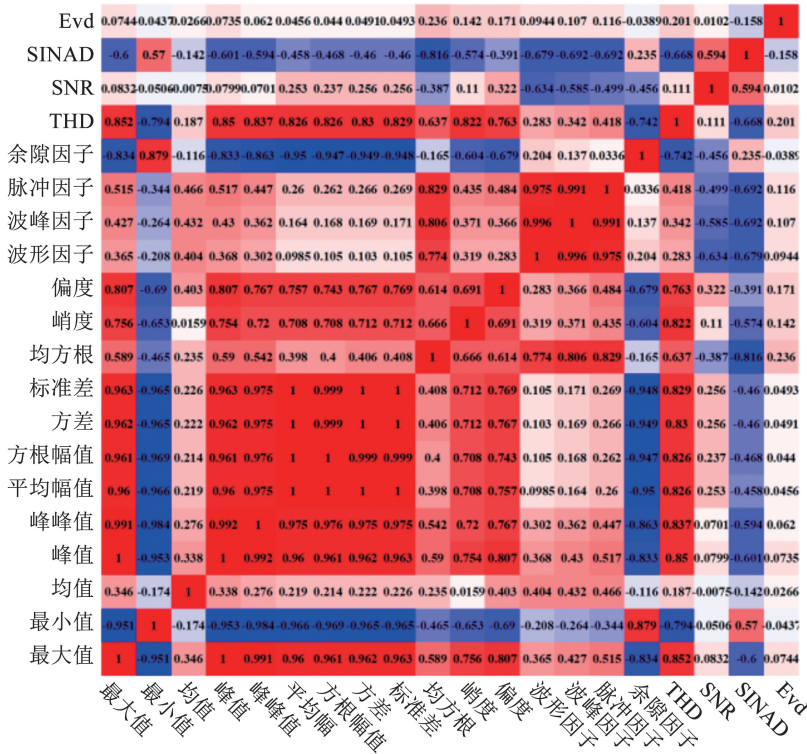


图 12 特征因子与  $E_{vd}$  相关性热力矩阵图

Fig. 12 Thermal matrix diagram of correlation between each factor and  $E_{vd}$

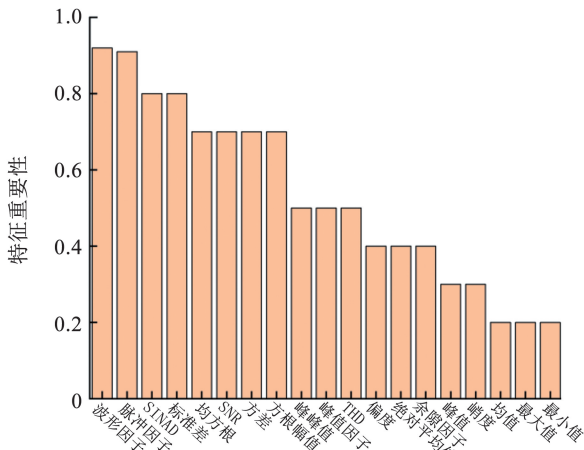


图 13 各特征因子重要性直方图

Fig. 13 Histogram of importance of each feature factor

均方根、峭度、偏度与  $E_{vd}$  的相关性最高,其他特征因子的相关性较差。图 13 是测试集中各个特征因子对  $E_{vd}$  的重要性程度分析,波形因子和脉冲因子这两种时域特征对  $E_{vd}$  的重要性最高,直接表明路基压实质量软硬情况可以直接反应到加速度的波形振幅。

### 3.3 初预测结果

图 14 是采用 XGBoost 预测模型对  $E_{vd}$  的预测结果,明显看出预测结果是非常接近真值的,两条曲线部分波动极为接近,但是也存在部分结果完全偏离真值,这种情况可能是因为此处某点的  $E_{vd}$  存在异常。其中,各评估指标可解释方差为 0.434, MAPE 为 0.303, MSE 为 0.116,  $R^2$  为 0.467。总体来说,预测值具有部分的可靠性。同时, XGBoost 预测模型学习特征需要进一步优化,使其达到更理想的效果。

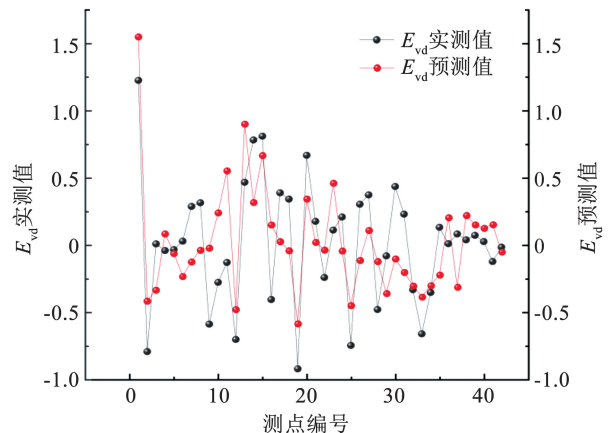


图 14  $E_{vd}$  实测值和预测值对比

Fig. 14 Comparison of measured and predicted values of  $E_{vd}$

### 3.4 GA-XGBoost 模型数据优化结果

在加速度信号中分析得出的 19 个特征因子对  $E_{vd}$  的预测有一定的可行性,但此时特征因子较多且存在多重共线性,对  $E_{vd}$  的预测效果不甚理想。因此,选择对上述模型数据进行筛选优化。

首先,从  $E_{vd}$  数值分布箱形图中发现,原始数据中存在数个异常值,出现偏离上极值线的情况,但由于异常值较少,并不影响数据的整体性,故选择删去;其次,从图 12 中发现,部分特征因子存在多重共线性,该情况会导致模型的泛化能力减弱。因此,需进一步对特征因子进行降维。根据 Pearson 相关系数理论,相关系数大于 0.8 为高度线性相关,本文选择对两两相关系数大于 0.8 的特征进行过滤,同时保留的特征因子需具有代表性且与  $E_{vd}$  的相关系数较高;结合图 13 重要性直方图,保留过滤后重要性程度不小于 0.4 的特征因子,综合考虑筛选结果是峭度、偏度、均方根、波形因子、脉冲因子、THD、峰峰值 7 种特征因子,该组合下,特征因子对  $E_{vd}$  的相关性和重要性皆较高且两两之间影响较小。

同时,考虑压路机的振动频率会对压实质量产生较大的影响,所以本次优化加入频率作为训练模型的特征因子,引入 0.5 倍频至 3 倍频共 6 种频率特征;最后,行业内各学者<sup>[13-18]</sup>已验证压实度测量值 CMV 也对  $E_{vd}$  有较强的相关性,因此,加入 CMV 作为训练预测模型的特征因子。优化结果见图 15。此时,4 种评估指标可解释方差为 0.732,MAPE 为 0.162,MSE 为 0.039, $R^2$  为 0.748。相较于之前,数据优化后有较明显的提高,相关性增强,说明这种特征因子组合下的  $E_{vd}$  数据预测具有可行性,但仍需要继续考虑是否存在其他特征情况产生更好的预测效果。图 16 和图 17 是分别以传统方法 CMV 与  $E_{vd}$  线性拟合曲线以及 GA-XGBoost 模型拟合出的曲线,由此可见,该预测模型相较于传统方法精确度具有一定的提高。

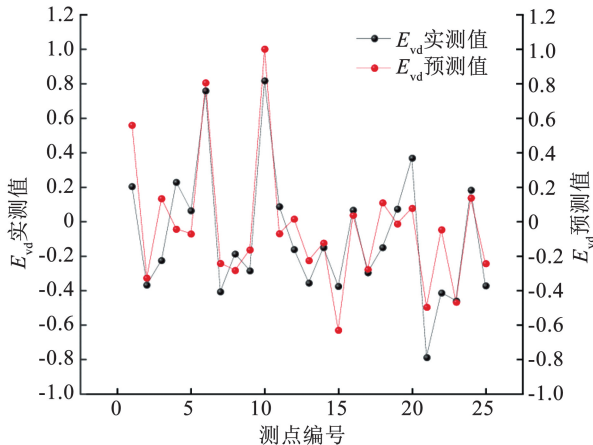


图 15 优化后实测值与预测值对比

Fig. 15 The comparison between the measured value and the predicted value after optimization

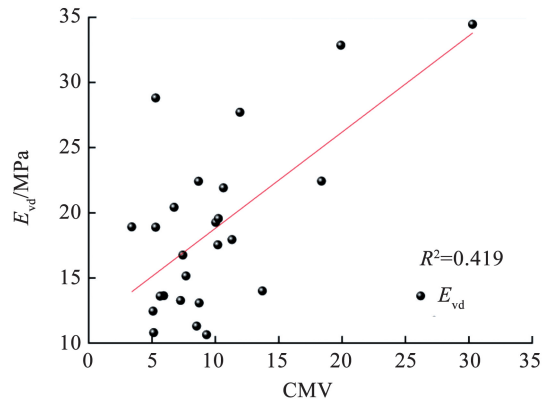


图 16 传统 CMV 拟合结果

Fig. 16 Traditional CMV fitting results

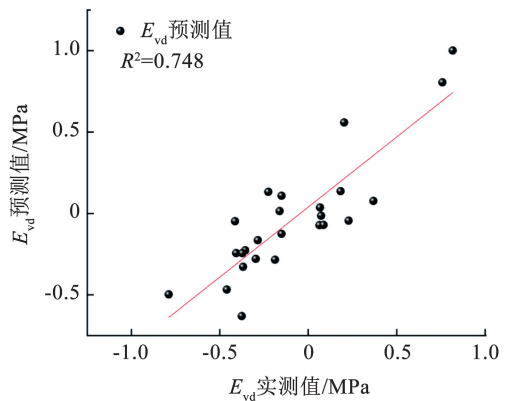


图 17 优化后 GA-XGBoost 模型拟合结果

Fig. 17 Fitting results of GA-XGBoost model after optimization

## 4 结论

本文通过现场压实试验研究,基于 GA-XGBoost 预测模型,分析了智能压实试验中不同的碾压参数对路基压实质量的影响,并优化了模型中各特征因子对  $E_{vd}$  的预测精度。得到的主要结论如下:

1) 由加速度信号分析得出的时域特征对智能压实指标  $E_{vd}$  的预测具有一定的可靠性,但需要进一步分析选取相关性较高的特征因子。

2) 采用遗传优化算法可以对 XGBoost 模型的超参数寻优,且寻优效果较好。GA-XGBoost 在测试集上的准确率较高,具有较优异的泛化能力。相较于传统方法基于倍频等单一指标,仅用 CMV 拟合  $E_{vd}$ ,客观性较差,影响了精度,本研究中基于 GA-XGBoost 预测模型采用多种特征因子分析压实度具有一定的可行性,可以提高模型的预测精度和可靠性。

3) 验证了 CMV 和频率的加入对  $E_{vd}$  的预测精度的影响,一定程度上提高了 GA-XGBoost 预测模型的精度。同时,模型结果表明, $E_{vd}$  数据值越集中,GA-XGBoost 预测模型的学习特征能力越明显,效果越好,预测的精度越高。

4) 实际工程试验验证了 GA-XGBoost 预测模型

具有一定的可行性,相较于传统方法对历史数据的学习和继承性较差,机器学习这方面有优势。在智能压实工程控制中,通过采集振动信号可以较快速反映路基压实质量,相较于传统压实后点式抽样检测路基压实质量,机器学习方法更加便捷、高效。

## 参考文献

- [1] 中国铁路总公司. 铁路路基填筑工程连续压实控制技术规程: Q/CR 9210—2015[S]. 北京: 中国铁道出版社, 2015  
China Railway Corporation. Technical specification for continuous compaction control of railway subgrade filling engineering: Q/CR 9210—2015[S]. Beijing: China Railway Publishing House, 2015
- [2] 聂志红, 谢扬, 焦倬. 基于双响应曲面法的压实质量连续检测参数稳健性设计[J]. 铁道学报, 2016, 38(6): 93  
NIE Zhihong, XIE Yang, JIAO Tan. Robust design of continuous compaction measured value based on dual response surface methodology[J]. Journal of the China Railway Society, 2016, 38(6): 93
- [3] 张德, 程志强, 谢胜加. 路基路面现场试验振动压实特性指标分析[J]. 上海公路, 2021(4): 1  
ZHANG De, CHENG Zhiqiang, XIE Shengjia. Analysis of vibration compaction characteristics index of subgrade and pavement field test[J]. Shanghai Highway, 2021(4): 1
- [4] 吴龙梁, 江辉煌, 唐建伟, 等. 基于多元回归分析的连续压实监控技术[J]. 岩土学报, 2020, 41(6): 2081  
WU Longliang, JIANG Huihuang, TANG Jianwei, et al. Continuous compaction monitoring technology based on multiple regression analysis[J]. Rock and Soil Mechanics, 2020, 41(6): 2081
- [5] 冯振刚, 舒金星, 曲建涛, 等. 沥青路面智能压实多元回归模型建立与验证[J]. 公路, 2023, 68(9): 66  
FENG Zhengang, SHU Jinxing, QU Jiantao, et al. Establishment and verification of intelligent compaction multiple regression model for asphalt pavement[J]. Highway, 2023, 68(9): 66
- [6] 马源, 方周, 韩涛, 等. 路基智能压实关键控制参数动态仿真及演变规律[J]. 中南大学学报(自然科学版), 2021, 52(7): 2246  
MA Yuan, FANG Zhou, HAN Tao, et al. Dynamic simulation and evolution of key control parameters for intelligent compaction of subgrade[J]. Journal of Central South University (Natural Science), 2021, 52(7): 2246
- [7] 蔡德钧, 叶阳升, 闫宏业, 等. 基于现场试验的高铁路基智能压实过程中振动波垂向传播机制[J]. 中国铁道科学, 2020, 41(3): 1  
CAI Degou, YE Yangsheng, YAN Hongye, et al. Vertical propagation mechanism of vibration wave during intelligent compaction of high-speed railway subgrade based on field test[J]. China Railway Science, 2020, 41(3): 1
- [8] 李凌志, 陈勇. 基于机器学习的高速公路水泥混凝土路面裂缝检测方法[J]. 建筑技术, 2023, 54(7): 808  
LI Lingzhi, CHEN Yong. Crack detection method of highway cement concrete pavement based on machine learning[J]. Building Technology, 2023, 54(7): 808
- [9] WANG Xuefei, CHENG Chi, LI Jiale, et al. Automated monitoring and evaluation of highway subgrade compaction quality using artificial neural networks[J]. Automation in Construction, 2023, 145: 0925
- [10] WANG Xuefei, DONG Xuping, ZHANG Zhishuai, et al. Compaction quality evaluation of subgrade based on soil characteristics assessment using machine learning[J]. Transportation Geotechnics, 2022, 32: 2214
- [11] WANG Xuefei, CHENG Chi, ZHANG Jianmin, et al. Real-time monitoring and quality assessment of subgrade compaction: key factors and ANN model[J]. Acta Geotechnica, 2022, 18(6): 3349
- [12] WANG Xuefei, DONG Xuping, LI Jiale, et al. Developing an advanced ANN-based approach to estimate compaction characteristics of highway subgrade[J]. Advanced Engineering Informatics, 2023, 56: 1474
- [13] CAO Liping, ZHOU Jie, LI Tao, et al. Influence of roller-related factors on compaction meter value and its prediction utilizing artificial neural network[J]. Construction and Building, 2021, 268: 121078
- [14] 安再展, 刘天云, 皇甫泽华, 等. 利用 CMV 评估堆石料压实质量的神经网络模型[J]. 水力发电学报, 2020, 39(4): 110  
AN Zaizhan, LIU Tianyun, HUANGFU Zehua, et al. Neural network model for evaluating compaction quality of rockfill materials by compaction meter value[J]. Journal of Hydroelectric Engineering, 2020, 39(4): 110
- [15] 陈晓斌, 谢康, 尧俊凯, 等. 基于能量最小原则的高铁填料压实过程振动参数优化[J]. 中南大学学报(自然科学版), 2023, 54(9): 3731  
CHEN Xiaobin, XIE Kang, YAO Junkai, et al. Vibration parameter optimization of high-speed railway filling compaction process based on the principle of minimum energy[J]. Journal of Central South University (Natural Science), 2023, 54(9): 3731
- [16] 黄国卿, 严筱, 杨永刚, 等. 智能压实间接指标试验分析与压实度预测[J]. 中外公路, 2020, 40(2): 24  
HUANG Guoqing, YAN Xiao, YANG Yonggang, et al. Intelligent compaction indirect index test analysis and compaction degree prediction[J]. Journal of China & Foreign Highway, 2020, 40(2): 24. DOI: 10.14048/j.issn.1671-2579.2020.02.006
- [17] 宿辉, 孙焯远, 赵宇飞, 等. 基于 GA-PSO 混合优化 BP 的面板堆石坝爆破料压实质量评价[J]. 人民黄河, 2023, 45(6): 137  
SU Hui, SUN Heyuan, ZHAO Yufei, et al. Compaction quality evaluation of blasting materials for concrete face rockfill dam based on GA-PSO hybrid optimization BP[J]. People's Yellow River, 2023, 45(6): 137
- [18] CHEN Tianqi, CARLO G. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Association for Computing Machinery, 2016