

基于 Kmeans - EMD 与 IWOA - Elman 的碾压速度异常值检测与修正

乔天诚, 佟大威, 王佳俊, 关涛, 吴斌平

(天津大学 水利工程仿真与安全国家重点实验室, 天津 300072)

摘要: 碾压速度是评价压实质量的重要指标,但在监控过程中,碾压速度易受施工环境、定位漂移等干扰而出现异常检测值,影响压实质量的评价精度,但目前还缺乏对碾压速度异常值检测与修正的相关方法研究。为保障碾压速度的数据质量,结合碾压速度的时序变化特征,利用 Kmeans 算法初步定性检测异常值,弱化异常值对经验模态分解 (EMD) 结果的影响,并基于 EMD 实现对异常值的精细定量检测,提高异常值检测的精度;进而利用经混沌种群初始化、非线性收敛因子、自适应惯性权重与鲛鱼效应 - 黄金正弦改进的鲸鱼优化算法 (IWOA) 优化 Elman 神经网络,并构建碾压速度异常值修正模型,实现对碾压速度异常值的修正。将本文方法应用于西南某大型水电工程,结果表明:Kmeans 算法与 EMD 的联合作用相比箱线图法可更高精度地检测碾压速度中的异常值;IWOA - Elman 神经网络预测值与真实值的相关系数达到 0.907 75,相比常规模型不仅可以更好地确保数据的完整性与可靠性,还可以为压实质量的高精度评价奠定良好的数据基础。

关键词: 碾压速度; 异常值检测; Kmeans 算法; 经验模态分解; 异常值修正; IWOA - Elman 神经网络

中图分类号: TV641

文献标识码: A

文章编号: 1672-643X(2022)03-0124-08

Outlier detection and correction for rolling speed based on Kmeans - EMD and IWOA - Elman

QIAO Tiancheng, TONG Dawei, WANG Jiajun, GUAN Tao, WU Binping

(State Key Laboratory of Civil Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China)

Abstract: Rolling speed is an important indicator for evaluating compaction quality, but in the monitoring process, the rolling speed is easily disturbed by construction environment, positioning drift and other disturbances, resulting in outliers which will affect the accuracy of compaction quality evaluation. However, methods for detecting and correcting outliers of rolling speed are not reported in current research. To ensure the data quality of the rolling speed, based on the characteristics of the time series of rolling speed, we adopted Kmeans algorithm to detect outliers preliminarily and weaken the influence of outliers on the results of empirical mode decomposition (EMD). According to the results of EMD, the fine quantitative detection of outliers was achieved, which in turn improved the accuracy of outlier detection. Furthermore, whale optimization algorithm (WOA) improved by chaos population initialization, nonlinear convergence factors, adaptive inertial weights and catfish effect - golden sine was used to optimize the Elman neural network and then a correction model was established for the correction of outliers of rolling speed. According to the application of the proposed method to a large hydropower project in Southwest China, the combined effect of Kmeans algorithm and EMD can detect outliers of rolling speed with higher accuracy than the box plot method; the correlation coefficient between the predicted value of the IWOA - Elman neural network and the actual value can reach up to 0.907 75, compared to conventional models, the

收稿日期: 2021-09-04; 修回日期: 2021-12-13

基金项目: 国家自然科学基金雅砻江联合基金项目 (U1965207)

作者简介: 乔天诚 (1997 -), 男, 黑龙江齐齐哈尔人, 硕士研究生, 主要从事大坝智能施工研究。

通讯作者: 佟大威 (1982 -), 男, 山西太原人, 博士, 副教授, 硕士生导师, 主要从事水利水电工程施工进度仿真与实时控制分析技术研究。

IWOA - Elman neural network can not only ensure better data integrity and reliability, but also lay a good data foundation for the high-precision evaluation of compaction quality.

Key words: rolling speed; outlier detection; Kmeans algorithm; empirical mode decomposition (EMD); outlier correction; improved whale optimization algorithm (IWOA) - Elman neural network

1 研究背景

目前,水利工程建设已逐步趋于智能化。由于在大坝建设过程中的压实质量实时评价可有效弥补事后评价的低时效性,极大提升工程质量的控制水平与效率,因而其评价方法得到了广泛研究^[1-2]。筑坝碾压监控系统可实现对施工过程的有效控制,为压实质量评价奠定数据基础,因此该系统已普遍应用于大坝施工过程^[3-5]。但碾压监控系统仅侧重对碾压速度的整体把控,即动态采集与调整,却未实现对碾压速度异常值的检测与修正,因而无法保障所采集的碾压速度数据的质量。

大坝填筑碾压速度是评价其压实质量的重要指标^[6-7],在监控过程中,碾压速度的本质是依托 GPS (global positioning system) 定位设备采集的连续时间序列,但 GPS 定位设备存在定位漂移、数传延迟等问题,且人为因素与仓面平整度也会导致数据异常。异常值不仅会影响碾压速度的数据质量,还易降低压实质量的评价精度,但当前尚缺乏相关分析方法,无法有效实现碾压速度的异常值检测与修正,故开展基于 Kmeans - EMD 与 IWOA - Elman 的碾压速度异常值检测与修正研究具有重要意义。

异常值是导致数据质量降低的重要因素,在各行业内均得到极大重视,其相关研究也取得了丰硕成果,如孤立森林算法^[8]、移动小波树^[9]、LOF 指数法^[10]、箱线图法^[11]等均被成功应用于异常值检测。但由于碾压速度具有较强的时序变化特征,平稳性较差,其加速与减速阶段的交互影响不利于基于统计学方法的异常值检测,故本文通过 Kmeans 算法^[12]与 EMD(empirical mode decomposition)^[13]从定性与定量两个角度检测碾压速度中的异常值,以保障异常值检测的精度。

尽管异常值检测是确保数据质量的基础,但异常值检测却会破坏原有的数据序列结构,因此异常值修正同样是确保数据质量的重要部分。异常值修正可确定异常位置下的合理取值,保障数据的完整性,提高数据可靠性^[14]。碾压速度作为一维时间序列,其异常值修正通常需依托同序列的其他数据并结合预测模型实现,如支持向量机^[15]、径向基神经

网络^[16]、极限学习机^[17-18]、长短期记忆神经网络^[19]、Elman 神经网络^[20]等均可用于异常值修正。而其中,Elman 神经网络承接层的设定使其具有更高的敏感性和更强的动态信息记忆与处理能力^[21],故本文将基于 Elman 神经网络实现碾压速度的异常值修正,并利用鲸鱼优化算法(whale optimization algorithm, WOA)优化 Elman 神经网络的初始权值与阈值,同时针对 WOA 存在的不足进行改进,形成 IWOA - Elman 神经网络。

综上所述,本文将基于 Kmeans 算法、EMD、IWOA - Elman 神经网络开展有关碾压速度异常值检测与修正的研究:首先,结合碾压速度的时序变化特征,利用 Kmeans 算法与 EMD 从定性与定量两个角度检测碾压速度中的异常值;进而,为保障碾压速度数据的完整性与可靠性,确保良好的数据序列结构,利用 IWOA - Elman 神经网络构建碾压速度异常值修正模型,从而提高碾压速度的数据质量,为压实质量高精度评价奠定良好的数据基础。

2 研究方法

本文结合 Kmeans 算法与 EMD 对碾压速度进行异常值检测,并利用 IWOA - Elman 神经网络构建碾压速度异常值修正模型,以提高碾压速度的数据质量,研究框架如图 1 所示,具体研究方法如下:

(1) 本文研究基于西南地区某心墙堆石坝工程开展,考虑到心墙堆石坝心墙区压实质量对大坝整体建设质量有重要影响,故选择心墙区凸块碾压速度监控数据为研究对象;

(2) 本文通过以下 4 方面内容实现对碾压速度异常值的检测与修正:①结合碾压速度的时序变化特征,先后利用 Kmeans 算法与 EMD 对碾压速度中的异常值做初步定性与精细定量检测,保障异常值检测的精度;②利用混沌种群初始化、非线性收敛因子、自适应惯性权重与鲶鱼效应 - 黄金正弦改进 WOA,弥补 WOA 在解决复杂优化问题时收敛速度慢、收敛精度低的不足;③利用 IWOA(improved whale optimization algorithm)优化 Elman 神经网络的初始阈值与权值,提高 Elman 神经网络的泛化能力与预测精度,强化网络稳定性,避免输出结果陷入局部最优;④基于 IWOA - Elman 神经网络构建碾压速

度异常值修正模型,保障数据的完整性与可靠性;

(3)将本文提出的碾压速度异常值检测与修正方法应用于选定的心墙区凸块碾压机速度监控数据中,验证本文方法的有效性,并通过与箱线图法及Elman神经网络、BP神经网络的对比,验证本文异常值检测与修正方法的优越性。

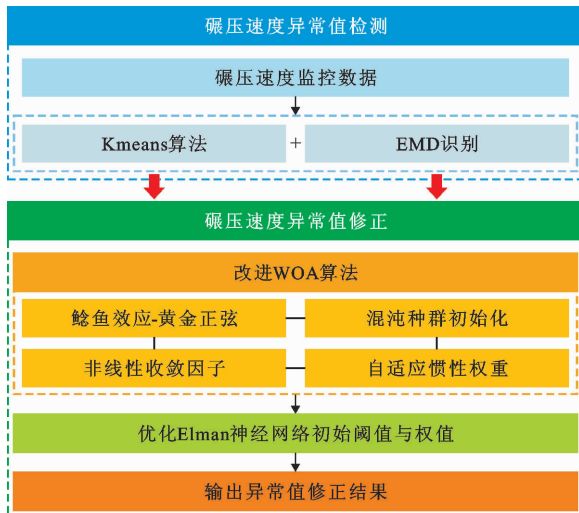


图1 筑坝碾压速度异常值检测与修正的研究框架

3 碾压速度异常值检测与修正方法

3.1 基于 Kmeans - EMD 的碾压速度异常值检测

3.1.1 基于 Kmeans 算法的异常值初步定性检测

Kmeans 算法又称 K 均值算法,是一类典型的基于距离的聚类算法,因结构清晰、收敛速度快而被广泛应用于聚类及异常值检测问题中^[22-23]。

Kmeans 算法的基本思想是在碾压速度数据集 $D = \{X_1, X_2, \dots, X_m\}$ 中选取 K 个样本作为簇心,即质心 $Z_j (j = 1, 2, \dots, K)$,进而计算各样本 $X_i (i = 1, 2, \dots, m)$ 与 K 个质心间的欧式距离 d_{ij} ,并将各样本划分至所距最近质心的所在簇,得到 K 个簇 $C_j (j = 1, 2, \dots, K)$,然后逐步计算新簇质心,并重复上述步骤,直到聚类准则函数收敛,聚类准则函数如式(1)所示。

$$E = \sum_{j=1}^K \sum_{X_i \in C_j} \|X_i - Z_j\|_2^2 \quad (1)$$

E 值表征了聚类后所得各簇的最小化平方误差, E 值越小则簇中碾压速度数据围绕质心越紧密,簇内相似度越高,聚类效果越好,因此各簇内碾压速度数据均应表现出聚集性,但异常值会因不满足碾压速度的基本变化特征而偏离质心位置,表现出离群性,故可结合这一特性实现异常值的初步定性检

测,从而弱化异常值对 EMD 分解结果的影响。

3.1.2 基于 EMD 的异常值精细定量检测 EMD 是一种针对非线性、非平稳时间序列的自适应分解方法,在使用时无需设定任何基函数,仅依托序列自身的时间尺度即可将数据分解为若干不同频率的子序列,即本征模态函数(intrinsic mode function, IMF)和代表趋势项序列的残差余项(residual, R)^[24]。

EMD 在应用时需确定碾压速度序列 $v(t)$ 的均值 $v_m(t)$ 及 $v(t)$ 与 $v_m(t)$ 间的差值 $h_1(t)$,若 $h_1(t)$ 满足:①极值点的数目与过零点的数目相等或至多相差 1;②上、下包络线均值为 0。则令 $c_1(t) = h_1(t)$,并以 $c_1(t)$ 为 IMF 第一分量,若不满足,则视 $h_1(t)$ 为待分解序列 $v(t)$,重复上述过程直至满足要求。而后,从 $v(t)$ 中分离出 $c_1(t)$,得到余项序列 $r_1(t) = v(t) - c_1(t)$,再令 $r_1(t)$ 为待分解序列,并依次重复上述步骤,得到 $c_2(t), c_3(t), \dots, c_n(t)$,若残差余项 $r_n(t)$ 为常数或单调函数,则分解结束,得到 n 个 IMF 分量和残差余项 $r_n(t)$ 。

IMF 分量较原碾压速度序列有更简单的波动规律,其中,高频分量反映了碾压速度序列中的高频与噪声信息,残差余项 R 与低频分量则分别粗略和细致地描述了碾压速度序列的变化趋势。

为实现异常值的精细定量检测,本文首先利用经 Kmeans 算法剔除异常值后的碾压速度序列 $v(t)$ 组成序列 $v'(t)$;其次,利用 EMD 对 $v'(t)$ 进行分解,得到 $v'(t)$ 的各 IMF 分量及余项 R ;再次,由后至前选择 m 个分量进行叠加,构成序列 $m(t)$,以弱化噪声影响,保留变化趋势;最后,分析 $v'(t)$ 与 $m(t)$ 间的偏离程度 $d(t)$, $d(t)$ 的计算如式(2)所示,并定义偏离程度阈值 d_0 ,当 $d(t) > d_0$ 时认为该时刻对应的值为异常值。

$$d(t) = \frac{|m(t) - v'(t)|}{v'(t)} \quad (2)$$

3.2 基于 IWOA - Elman 神经网络的碾压速度异常值修正模型

3.2.1 IWOA 简介 IWOA 是 WOA 的改进,WOA 是受座头鲸狩猎行为启发提出的一种智能算法,具体可参考 Mirjalili 等^[25]的描述。WOA 通过模拟座头鲸狩猎时的包围式捕猎、气泡网式捕猎和随机式捕猎行为搜索最优解,在寻优过程中所需参数少、全局收敛性强,相比粒子群算法、遗传算法等优化算法,不仅易于实现,亦能满足精度要求,适用于本文对 Elman 神经网络初始阈值和权值的优化^[25-26]。

但本文对 Elman 神经网络初始阈值和权值的优

化属多参数复杂优化问题,而 WOA 在解决此类问题时,仍存在收敛速度慢、收敛精度低、后期开发能力较弱的缺陷,且算法本身的趋同性易导致搜索陷入局部最优。

针对上述问题:

(1) 本文利用混沌机制初始化 WOA 种群^[27],通过随机遍历搜索空间的方式强化种群个体的多样性,提高初始种群质量,强化算法的搜索效率与寻优能力,以避免因随机初始化种群造成的可行解搜索范围变大、搜索速度降低、寻优精度下降等问题。该方法在实现过程中首先随机生成实数向量 X_i ,而后根据公式(3)生成其余个体初始位置。

$$X_{i+1} = \text{mod}(X_i + 1/5 - (4\pi)^{-1} \cdot \sin(2\pi \cdot X_i), 1) \quad (3)$$

式中: X_i 为第 i 个体的位置; mod 为求余函数。

(2) 本文引入非线性收敛因子 a ^[28],以解决收敛因子若通过线性方式由 2 递减至 0 则难以适应实际优化搜索过程的问题,进而有效平衡全局与局部搜索,非线性收敛因子如公式(4)所示。

$$a = 2 \times \left(1 - \left(\frac{e^{t/t_{\max}} - 1}{e - 1}\right)^\mu\right) \quad (4)$$

式中: a 为收敛因子; t 为当前迭代次数; t_{\max} 为最大迭代次数; μ 为衰减系数, $\mu > 0$, 本文取 0.6。

(3) 本文提出一种自适应惯性权重策略,使种群可以根据搜索情况改变收敛速度,以快速收敛至全局最优解。该策略可在搜索前期获得较大步长,提高搜索效率,而在后期获得较小步长,提高寻优精度,自适应惯性权重如公式(5)所示。

$$w = (w_{\max} (1 + \cos \frac{t \cdot \pi}{t_{\max}}) + w_{\min} (1 - \cos \frac{t \cdot \pi}{t_{\max}})) / 2 \quad (5)$$

式中: w 为惯性权重; t 为当前迭代次数; t_{\max} 为最大迭代次数; w_{\min} 为惯性权重的最小值, 本文取 0.4; w_{\max} 为惯性权重的最大值, 本文取 0.9。

(4) 本文结合鲶鱼效应策略^[29]对 WOA 进行改进,若最优解在规定迭代次数内没有进化,则更新种群内适应度较差的 90% 个体的位置,以恢复种群活力,提高算法跳出局部最优的能力,避免算法早熟。其中,位置更新过程引入黄金正弦算法实现^[30],如公式(6)~(9)所示。

$$X(t+1) = X(t) \cdot |\sin(r_1)| - r_2 \cdot \sin(r_1) \cdot D \quad (6)$$

$$D = |x_1 \cdot P_t - x_2 \cdot X(t)| \quad (7)$$

$$x_1 = -\pi + (1 - V) \cdot 2\pi \quad (8)$$

$$x_2 = -\pi + V \cdot 2\pi \quad (9)$$

式中: $X(t+1)$ 和 $X(t)$ 分别为鲸鱼个体第 $t+1$ 次和第 t 次迭代的位置; P_t 为鲸鱼个体在 t 次迭代中的最优位置; x_1 和 x_2 为黄金分割系数; $V = (\sqrt{5} - 1)/2$; r_1 和 r_2 分别为 $[0, 2\pi]$ 间和 $[0, \pi]$ 间的随机数。

3.2.2 IWOA - Elman 神经网络模型 本文均以均方根误差(root mean square error, RMSE)作为适应度函数评价 IWOA 的寻优情况,并利用寻优结果优化 Elman 神经网络的初始阈值与权值^[31-32],以提高 Elman 神经网络的泛化能力,弥补传统 Elman 神经网络通过随机方式生成初始阈值与权值易导致模型训练过程收敛稳定性下降的不足,避免输出结果陷入局部最优,提高神经网络的预测精度, IWOA - Elman 神经网络模型的构建步骤如图 2 所示,具体解释如下:

(1) 初始化 WOA 参数并根据 Elman 神经网络拓扑结构确定鲸鱼群搜索维度, 本文鲸鱼群搜索维度为 Elman 神经网络阈值与权值的数量和:

$$K = E \cdot F + F \cdot H + F + H \quad (10)$$

式中: K 为鲸鱼群搜索维度; E 为输入层神经元个数; F 为隐含层神经元个数; H 为输出层神经元个数。

(2) 利用混沌机制初始化鲸鱼个体位置;

(3) 计算各鲸鱼个体初始位置适应度,并记录当前适应度最优个体的所在位置 X' 及其适应度 Y' ;

(4) 引入非线性收敛因子及自适应惯性权重,更新鲸鱼个体的 a, w 等参数;

(5) 依据 WOA 包围式捕猎、气泡网式捕猎和随机式捕猎行为的执行条件,确定当前行为模式,得到鲸鱼个体更新后的位置;

(6) 经位置更新后,再次确定各鲸鱼个体的适应度,并记录具有最优适应度个体的位置 X 及其适应度 Y ,若 Y 优于已有的 Y' 则更新 Y' 与 X' ,反之则进入步骤(7);

(7) 判断 Y' 在规定迭代次数下是否进化,若进化则进入步骤(8),反之则引入鲶鱼效应 - 黄金正弦对适应度较差的 90% 个体的位置进行更新,再进入步骤(8);

(8) 则重复执行步骤(4)~步骤(7),直到满足迭代终止条件时,认为算法已生成最优解,此时最优解保存于 X' ;

(9) 利用最优解 X' 优化 Elman 神经网络的初始阈值与权值,实现 IWOA - Elman 神经网络模型的构建。

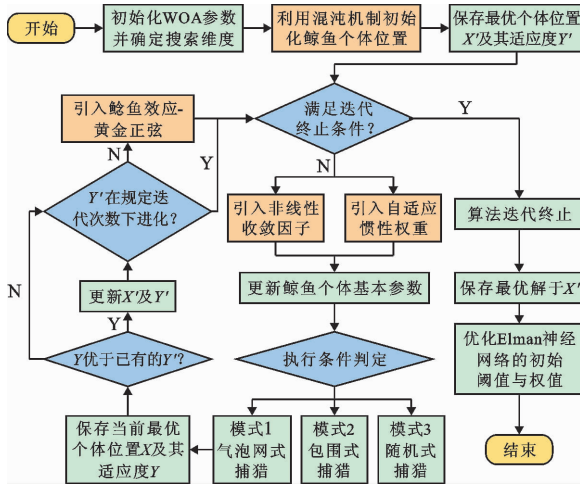


图2 IWOA-Elman神经网络模型的构建步骤

4 实例工程应用

以西南地区某心墙堆石坝心墙区凸块碾压机速度监控数据为研究对象,共采集到1423个连续的碾压速度序列点,其碾压速度数据值=实际碾压速度(km/h)×100,在此基础上,利用本文所提方法开展碾压速度的异常值检测与修正。

4.1 异常值检测结果

对碾压速度而言,其异常值通常表现为异常大值或异常小值,而异常小值多由变换条带的正常工作流程导致,且异常大值对应的超速情况会使单位时间内填筑料的碾压功输入不足,相比变换条带的减速阶段更易导致填筑料压实不充分,从而影响压实质量评价结果^[33],故本文的异常值检测主要面向碾压速度中的异常大值,并基于Kmeans算法与EMD实现。

4.1.1 基于Kmeans算法的异常值检测结果 本文通过对碾压机工作状态(速度稳定、加速与减速)的区分与试验分析,判断当Kmeans算法簇个数为3时即可解释不同的数据情况,亦可确保聚类结果稳定,Kmeans算法异常值检测结果如图3所示。

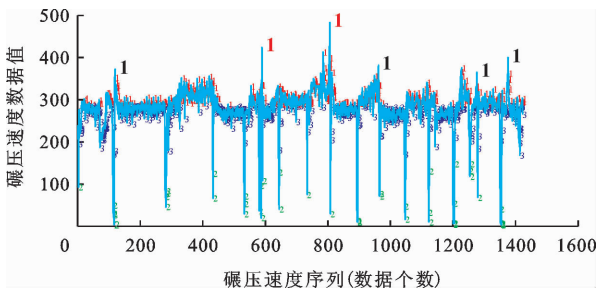


图3 基于Kmeans算法的碾压速度异常值检测结果

由图3可知,碾压速度中的异常大值均体现在1号簇,且图中放大的红色标号位置,即序列589和

806位置为明显异常大值,可将其定义为异常值;而放大的黑色标号位置则为模糊异常大值,无法直接通过Kmeans算法判断,需经EMD的二次分析。

4.1.2 基于EMD的异常值检测结果 由于Kmeans算法仅能定性检测异常值,以弱化明显异常大值对EMD分解结果的影响,却无法实现对模糊异常大值的判断,故本文在将Kmeans算法检测出的异常值剔除后,利用EMD对组成序列 $v'(t)$ 进行分解,共获得10个分量,包括9个IMF分量与余项 R ,如图4所示,进而利用选定的IMF分量与余项 R 构建新序列,并结合公式(2)进一步判断异常值位置,实现异常值的精细定量检测。

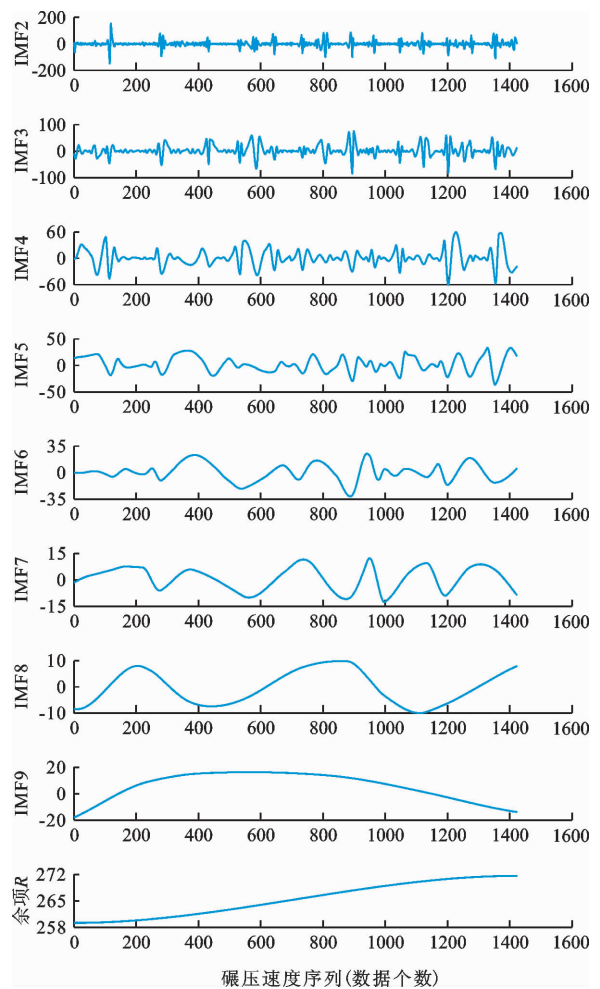


图4 碾压速度序列的EMD分解结果

根据前文3.1.2节中的描述与试验分析,本文选定 m 值为5,即利用后4个IMF分量与余项 R 构建序列 $m(t)$,并确定偏离程度阈值 d_0 为0.35。EMD异常值检测结果如图5所示,图5中圆圈对应的序列位置为异常值所在位置,黑色实线为利用IMF分量与余项 R 构建的序列 $m(t)$ 。

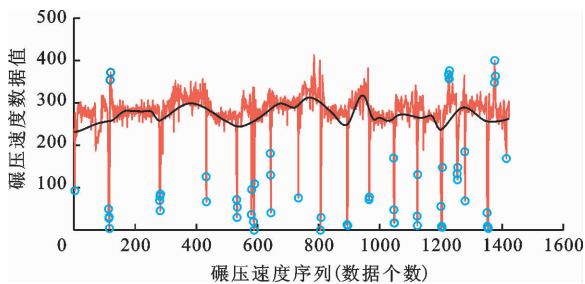


图 5 基于 EMD 的碾压速度异常值检测结果

如图 5 所示,EMD 异常值检测通过去除高频与噪声分量实现对碾压速度的滤波处理,有效避免了对异常大值的误判。EMD 异常值检测结果表明碾压速度序列中 118、120、1 225、1 226、1 227、1 228、1 375、1 376 和 1 378 位置均为异常值,应予以剔除。

4.1.3 Kmeans - EMD 对异常值检测的优势 本文以箱线图异常值检测方法^[34]作为对比,论证 Kmeans 算法与 EMD 联合运用于异常值检测的优势。箱线图异常值检测结果如图 6 所示。本文所用碾压速度数据对应的箱线图上限与下限分别为 351.5 和 219.5,即图 6 中两虚线之外的序列位置均为箱线图所检测的异常值。由图 6 可知,尽管箱线图可基于统计学原理实现异常值检测,却没有考虑碾压速度的时序变化特征,易将连续超速值均检测为异常值,故本文异常值检测方法具有更高的检测精度。

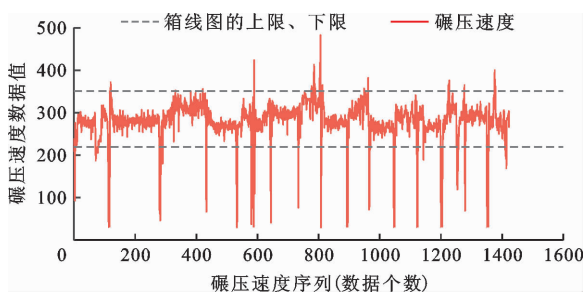


图 6 基于箱线图的碾压速度异常值检测结果

4.2 IWOA - Elman 神经网络异常值修正结果

为确定异常值位置下碾压速度的合理取值,以保障数据的完整性,强化数据的可靠性,本文构建了 IWOA - Elman 神经网络碾压速度异常值修正模型。设定鲸鱼种群数为 10,最大迭代次数为 500,规定迭代次数为 20,Elman 神经网络隐含层神经元为 5,训练函数为 traingdx,碾压速度数据总数为 1 423。为实现对碾压速度序列的预测,并考虑到碾压速度序列预测组构建时不应过多影响原序列的完整性^[35],且前 10 个数据中不含异常值,本文以 10 个碾压速度序列为一组,将各组的 10 个碾压速度序列作为输

入数据,预测下一组中第 1 个碾压速度数据值。经上述处理,本文共得到 1 413 组数据,规定前 1 300 组为训练集,后 113 组为测试集,输出序列中不包括原碾压速度序列中的前 10 个数据值。IWOA - Elman 神经网络测试集输出结果如图 7 所示。

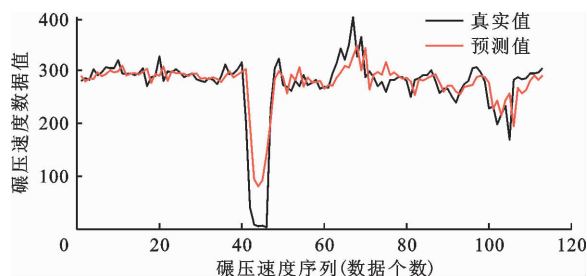


图 7 IWOA - Elman 神经网络碾压速度测试集输出结果

由图 7 可知,IWOA - Elman 神经网络可较好地拟合真实值情况,尽管变更条带的减速过程在数值表现上的时间相关性较弱,在一定程度上影响了 IWOA - Elman 神经网络的拟合效果,但相关性分析表明测试集相关系数仍达到 0.907 75,且在与 Elman 神经网络及 BP 神经网络的对比中体现出优越性,相关系数对比结果如表 1 所示,即 IWOA - Elman 神经网络在碾压速度序列的拟合上具有较高的鲁棒性与拟合精度,其修正结果亦可有效体现碾压速度的时间相关性。

表 1 不同神经网络形式的测试集相关系数对比

网络形式	测试集相关系数
IWOA - Elman	0.90775
Elman	0.85332
BP	0.81047

利用 IWOA - Elman 神经网络碾压速度异常值修正模型获取碾压速度序列中 118、120、589、806、1 225、1 226、1 227、1 228、1 375、1 376 和 1 378 位置的预测值并对原碾压速度序列进行替换,替换后各位置下碾压速度数据值分别为 310.2、279.6、339.9、361.2、322.3、304.3、310.8、328.7、312.5、323.7 及 303.0。

为评价替换后的数据质量,对替换后的碾压速度再次进行 Kmeans 聚类,结果如图 8 所示。由图 8 可知,经异常值检测与修正后,碾压速度数据质量得到明显提升,不存在异常大值,可为压实质量高精度评价提供良好的数据基础。

4.3 讨论

本文研究在综合考虑碾压速度时序变化特征的

基础上,实现了基于 Kmeans - EMD 的碾压速度异常值检测和基于 IWOA - Elman 神经网络的异常值修正,相较于传统方法,本文方法具有更高的异常值检测与修正精度,不仅可确保更好的数据完整性与可靠性,还可为压实质量高精度评价提供良好的数据基础,能有效弥补当前有关碾压速度数据异常值检测与修正研究的空白,提高了碾压速度的数据质量。但本文异常值检测与修正方法不适用于碾压速度缺失较多的状况,且由于碾压施工过程工艺情况复杂,其超速与减速阶段间可能存在复杂的相互影响,而本文仅通过单位时间内填筑料碾压功的输入情况判断异常值检测主要面向的数据类型,后续研究还需深入探讨碾压速度与压实质量间的相关关系,以准确界定碾压速度异常值检测的数据范围。

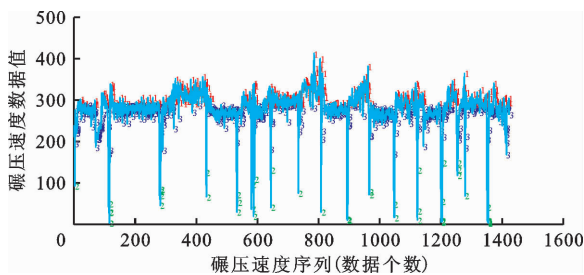


图8 以碾压速度修正值替换后 Kmeans 异常值检测结果

5 结论

目前在大坝坝体填筑施工中,尚无碾压速度异常值的检测与修正方法,因而难以确保碾压速度的数据质量,还易导致压实质量评价精度的降低。针对这一现状,本文开展了基于 Kmeans - EMD 与 IWOA - Elman 的碾压速度异常值检测与修正研究,得到的主要结论如下:

(1) 提出基于 Kmeans 算法与 EMD 的异常值检测方法:首先利用 Kmeans 算法对碾压速度进行聚类,通过将明显异常大值定义为异常值,实现对异常值的初步定性检测;再利用 EMD 对处理后的碾压速度序列进行分解,并通过选用的 IMF 分量与余项 R 构建新序列,实现基于偏离程度阈值 d_0 的异常值精细定量检测。实例工程应用结果表明, Kmeans 算法和 EMD 的联合运用相比箱线图法可更高精度地检测碾压速度中的异常值,有助于提高数据质量。

(2) 构建 IWOA - Elman 神经网络碾压速度异常值修正模型:首先,利用混沌种群初始化、非线性收敛因子、自适应惯性权重与鲑鱼效应 - 黄金正弦改进 WOA,弥补 WOA 在解决复杂优化问题时收敛

速度慢、收敛精度低的不足;其次,利用 IWOA 优化 Elman 神经网络的初始阈值与权值,提高 Elman 神经网络的泛化能力,强化网络稳定性,避免输出结果陷入局部最优;最后,基于 IWOA - Elman 神经网络实现碾压速度的异常值修正,确保数据的完整性和可靠性。实例工程应用结果表明, IWOA - Elman 神经网络的预测结果与原碾压速度数据间拟合情况良好,相关系数达到 0.907 75,相比常规模型不仅可实现更有效的异常值修正,保障数据序列结构,还可为压实质量高精度评价奠定良好的数据基础。

参考文献:

- [1] 王佳俊,钟登华,关涛,等. 基于 KM 和 AC - BFA 模糊逻辑的土石坝压实质量实时评价[J]. 水力发电学报, 2019, 38(3): 165 - 178.
- [2] 崔博,张慧高,王佳俊,等. 基于施工全过程参数的堆石坝掺砾土心墙压实质量实时评价[J]. 水利水电技术(中英文), 2021, 52(2): 109 - 118.
- [3] HUANG Shengxiang, ZHANG Wen, WU Gen, et al. Research on real-time supervisory system for compaction quality in face rockfill dam engineering[J]. Journal of Sensors, 2018(5): 6487405. <https://doi.org/10.1155/2018/6487405>.
- [4] 倪仕文,彭卫平,王蝉,等. 土石坝填筑碾压施工质量实时监控研究与应用[J]. 水力发电, 2019, 45(7): 80 - 84.
- [5] 马洪琪,钟登华,张宗亮,等. 重大水利水电工程施工实时控制关键技术及其工程应用[J]. 中国工程科学, 2011, 13(12): 20 - 27 + 2.
- [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 - 2257.
- [7] 吴龙梁,江辉煌,彭新平,等. 土石坝压实质量连续检测指标影响因素分析[J]. 水资源与水工程学报, 2021, 32(3): 203 - 209.
- [8] 张海龙,范振东,陈敏. 孤立森林算法在大坝监测数据异常识别中的应用[J]. 人民黄河, 2020, 42(8): 154 - 157 + 168.
- [9] 夏景,梁薇,吴珠瑛. 基于移动小波树的电力监控异常数据自动识别算法研究[J]. 电子设计工程, 2020, 28(18): 148 - 152.
- [10] 郭战坤,金永威,梁小珍,等. 基于异常值检测的港口集装箱吞吐量预测模型[J]. 数学的实践与认识, 2019, 49(17): 26 - 34.
- [11] 周建民,王发令,张龙,等. 基于 RBF 神经网络与模糊评价的滚动轴承退化状态定量评估[J]. 机械设计与研究, 2019, 35(6): 116 - 122 + 127.
- [12] 钱宇骋,甄超,季坤,等. 变压器在线监测数据异常值检测与清洗[J]. 哈尔滨理工大学学报, 2020, 25(5): 15 - 22.

- [13] 方海泉,薛惠锋,蒋云钟,等. 基于 EEMD 的水资源监测数据异常值检测与校正[J]. 农业机械学报,2017,48(9):257-263.
- [14] HARADA R, SHIGETA Y. Selection rules for outliers in outlier flooding method regulate its conformational sampling efficiency[J]. Journal of Chemical Information and Modeling, 2019, 59(9): 3919-3926.
- [15] 郭霖,白丹,王新端,等. 基于支持向量机的滴灌灌水器流量预测模型建立与验证[J]. 农业工程学报,2018,34(2):74-82.
- [16] BELDERRAR A, HAZZAB A. Hierarchical genetic algorithm and fuzzy radial basis function networks for factors influencing hospital length of stay outliers[J]. Healthcare informatics research, 2017, 23(3): 226-232.
- [17] 李军,李大超. 基于优化核极限学习机的风电功率时间序列预测[J]. 物理学报,2016,65(13):39-48.
- [18] GUO Wei, XU Tao, TANG Keming. M-estimator-based online sequential extreme learning machine for predicting chaotic time series with outliers[J]. Neural Computing and Applications, 2017, 28(12):4093-4110.
- [19] YANG Haimin, PAN Zhisong, TAO Qing. Robust and adaptive online time series prediction with long short-term memory[J]. Computational Intelligence and Neuroscience, 2017, 2017: 9478952.
- [20] 李广儒,张新,朱庆辉. 基于 Adaboost 的改进 Elman 神经网络港口吞吐量预测方法[J]. 重庆交通大学学报(自然科学版),2021,40(5):1-5.
- [21] DOU Chunxia, QI Hang, LUO Wei, et al. Elman neural network based short-term photovoltaic power forecasting using association rules and kernel principal component analysis[J]. Journal of Renewable and Sustainable Energy, 2018, 10(4):043501.
- [22] 王丝丝,张敬磊,王建兴,等. 基于改进群智能 K-means 优化算法的侵犯性驾驶行为异常点检测[J]. 科学技术与工程,2018,18(34):208-214.
- [23] 李昌华,卜亮亮,刘欣. 基于聚类和神经网络对建筑节能气候数据清洗的算法[J]. 计算机应用,2018,38(S1):83-86+111.
- [24] HUANG N E, SHEN Zheng, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings Mathematical Physical & Engineering Sciences, 1998, 454(1971): 903-995.
- [25] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95(5): 51-67.
- [26] 龙文,蔡绍洪,焦建军,等. 求解大规模优化问题的改进鲸鱼优化算法[J]. 系统工程理论与实践,2017,37(11):2983-2994.
- [27] 史文涛,周林荣. 基于改进鲸鱼算法的黔北地区农产品配送问题研究[J]. 计算机应用与软件,2020,37(10):91-98+124.
- [28] 孙鹤旭,张维,雷兆明,等. 基于改进鲸鱼优化算法的风电消纳调度研究[J]. 高技术通讯,2020,30(3):298-304.
- [29] 丁相毅,张博. 基于改进鲑鱼效应粒子群算法的电站优化调度[J]. 人民黄河,2018,40(4):128-131.
- [30] TANYILDIZI E, DEMIR G. Golden sine algorithm: A novel math-inspired algorithm[J]. Advances in Electrical and Computer Engineering, 2017, 17(2):71-78.
- [31] 贾依达尔·热孜别克,袁逸萍,孙文磊,等. 根据 PCA-GA-Elman 的风机齿轮箱轴承温度预测[J]. 机械科学与技术,2020,39(11):1671-1675.
- [32] 刘子英,钱超,朱琛磊. 基于 IPSO-Elman 的锂电池剩余寿命预测[J]. 现代电子技术,2020,43(12):100-105.
- [33] 许成帮. 压路机碾压速度对路基压实效果研究[J]. 公路交通科技(应用技术版),2019,15(2):28-29.
- [34] 寇元宝,汪崇建,负瑞光. 采煤机运行数据预处理算法研究[J]. 煤矿机械,2020,41(10):41-43.
- [35] STOJANOVIC M, BOZIC M, STANKOVIC M, et al. A methodology for training set instance selection using mutual information in time series prediction[J]. Neurocomputing, 2014, 141: 236-245.