

基于互补集合经验模态分解和支持向量回归机的城市轨道交通线路轨距劣化预测

贾清天 林海剑 金 忠

(上海地铁维护保障有限公司, 200235, 上海)

摘 要 [目的] 为了加强城市轨道交通区间线路质量的状态管理, 需要对轨距在空间上的整体劣化趋势进行预测。

[方法] 引入 CEEMD(互补集合经验模态)理论, 提取轨道区间几何形位的 IMF(本征模态函数), 利用 PSO(改进粒子群)算法优化 SVR(支持向量回归机), 对提取数据进行训练, 标定预测模型最优参数后进行测试集验证, 构建 CEEMD-PSO-SVR 预测模型。通过上海轨道交通 16 号线上行轨道区间 K12+134—K15+743 内的 1 128 组轨检样本数据对预测模型进行了试验。[结果及结论] CEEMD-PSO-SVR 预测模型同 PSO-SVR 模型、ARIMA(自回归移动平均模型)相比, 在均方根误差、平均绝对误差、平均相对误差绝对值等 3 项性能评价指标上具有优势。

关键词 城市轨道交通线路; 轨距劣化; 互补集合经验模态分解; 支持向量回归机

中图分类号 U216.3

DOI:10.16037/j.1007-869x.2025.01.009

Gauge Deterioration Prediction of Urban Rail Transit Lines Based on CEEMD and SVR

JIA Qingtian, LIN Haijian, JIN Zhong

(Shanghai Metro Maintenance Support Co., Ltd., 200235, Shanghai, China)

Abstract [Objective] In order to strengthen the status management of urban rail transit line sections, it is necessary to predict the overall deterioration trend of the gauge in space.

[Method] CEEMD (complementary ensemble empirical mode decomposition) theory is introduced to extract the IMF (intrinsic mode function) of the geometric alignment of the track section. The PSO (particle swarm optimization) algorithm is utilized to optimize the SVR (support vector regression machine) to train and test the extracted data after calibrating the optimal parameters of the prediction model. Thus, the CEEMD-PSO-SVR prediction model is constructed. The prediction model is tested with 1,128 sets of track inspection sample data within the upward track section from K12+134 to K15+743 on Shanghai Metro Line 16. [Result & Conclusion] Compared with the PSO-SVR model and the ARIMA (autoregressive in-

tegrated moving average) model, the CEEMD-PSO-SVR prediction model has advantages in three performance evaluation indicators, namely root mean square error, mean absolute error, and absolute value of mean relative error.

Key words urban rail transit line; gauge deterioration; CEEMD; SVR

轨道几何形位是反映轨道质量好坏的重要指标。在轨道交通车辆长时间反复作用下, 轨道会出现一定的纵向或横向变形, 当轨道几何形位达到一定阈值后会影响到列车的通过平稳性和乘车舒适性, 同时加速钢轨的磨耗, 导致轨距劣化加剧, 进一步降低了轨道使用寿命^[1], 为城市轨道交通行车安全带来风险隐患。

TQI(轨道质量指数)常用来表征轨道几何形位的劣化程度, 一些学者在 TQI 预测方面开展研究并取得了一些成果。文献[2]认为 TQI 属于灰色系统, 提出了 TQI 非等时距预测模型。文献[3]认为轨道几何形位变化具有非线性, 呈现浴盆曲线特点, 为此建立了多阶段线性预测模型。近年来随着人工智能的兴起, 一些学者将自学习模型引入到轨道几何形位领域, 并得到了验证。文献[4]通过 PCA(主成分分析)筛选了影响轨道几何形位变化的重要指标, 结合 SVM(支持向量机)实现对轨道几何形位的预测。

在轨道几何形位预测研究之外, 一些学者在轨道几何形位的指标处理方面也做了研究。文献[5]针对采集几何形位数据中存在的异常值, 利用 EMD(经验模态分解)法结合灰色关联分析对其进行了处理。文献[6]通过 WVD(维格纳-威尔分布)时频能量密度联合 EMD 法成功提取了轨道状态表征信息。通过大量的文献梳理发现, 针对轨道几何形位的预测多集中在综合性指标 TQI 上, 且基于历史性 TQI 多期数据开展预测, 但该方法未能给出某一要

素的劣化趋势。目前,针对轨道几何形位中的单项要素,以及轨道区间内轨距劣化趋势预测的研究较少。

本文以轨道几何形位的轨距要素作为对象,研究轨道区间内轨距状态整体劣化趋势。通过对检测数据中存在的异常值进行处理,利用 CEEMD(互补集合经验模态分解)提取 IMF(本征模态函数),并与 PSO-SVR(改进粒子群-支持向量回归机)模型组合,建立了具有能够实现轨道区间内的轨距劣化预测 CEEMD-PSO-SVR 模型,并利用采集的上海某轨道交通线路上行区间内轨检数据进行了验证,结果表明本模型有效且具有较高精度。

1 轨距劣化分解及预测模型构建

1.1 轨距劣化分解

1.1.1 EMD 算法

EMD 算法是由黄锬等科学家提出,最早是为了帮助解决自适应时频信号的处理问题,核心思想是用任意不同尺度的原始信号分解若干个独立的 IMF 和 Res(残余分量)^[7]。

采集的轨距信息原始序列通过 EMD 算法分解为:

$$x(t) = \sum_{i=1}^m F_{IM,i}(t) + R_{es}(t) \quad (1)$$

式中:

t ——时间;

$x(t)$ ——采集的轨距序列样本;

$F_{IM,i}(t)$ ——第 i 个 IMF 分量;

$R_{es}(t)$ ——残余分量,为各 IMF 残余之和。

1.1.2 CEEMD 算法

在对轨距的 EMD 实际测试过程中,发现该算法极易产生模态分解不彻底的结果,具体表现在分解后的模态分量之间存在大量混叠现象,此结果增加了轨距序列的混沌性,降低了预测准确性。本文在 EMD 算法的基础上利用改进的算法(CEEMD 算法)进行处理。其过程为:

在 $x(t)$ 中加入 N 组辅助白噪声, N 组白噪声以互为相反数的形式进行添加,得到 $2N$ 个信号集,对重构后所有的轨距序列利用 EMD 算法进行运算,分解的结果使每个信号点都能得到 1 组 $F_{IM,i}$,最后对多组 $F_{IM,i}$ 均值化。

CEEMD 算法的好处在于模型迭代 1 次可以获得 1 组 F_{IM} ,提取出有用的轨距状态信息,减少了高

频模态随机噪声的干扰,完成轨距数据频域分解,有利于实现对轨距劣化趋势研究。

1.2 轨距劣化预测模型构建

1.2.1 SVR 预测

SVR 是 SVM 的分支形式之一,用来解决回归预测问题^[8]。假设 x_i 为 v 维输入样本向量, y_i 为对应真实值,在实数空间 \mathbf{R} 范围内,轨距分解后的样本集合为: $\{(x_i, y_i) | x_i \in \mathbf{R}^v, y_i \in \mathbf{R}\}$ 。

本文采用线性回归函数,引入不敏感损失函数后,可得到:

$$f(x) = \omega^T x + b \quad (2)$$

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{其他} \end{cases} \quad (3)$$

式中:

x ——输入样本向量;

ω, b ——参数;

ε ——不敏感损失系数;

y ——真实值;

$f(x)$ ——回归函数。

为求得式(2)的结果,需要寻求 ω 的最小值。为降低求解难度,采用最小化欧几里德空间范数,对问题转化为求最优目标函数。引入拉格朗日函数后,对各参数求偏导处理,并将偏导后的结果带入拉格朗日函数 L 中,经对偶形式转换求拉格朗日乘子 α^*, α 的最小值,优化 SVR 模型的新形式。求解上述轨距劣化趋势预测问题时,需设定非线性的映射值 $\phi(x)$,保证低维的轨距序列在高维空间上的映射关系,引入核函数 $K(x_i, x_j)$ 取代拉格朗日函数 L 对偶形式中内积 $\phi(x_i) \cdot \phi(x_j)$,在本次轨距劣化预测模型构建过程中通过多次试验确定多项式为本次映射函数。

结合约束条件,可求得:

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

式中:

l ——训练集中采集的轨距样本总数。

针对式(4)回归函数参数 b 的求解,引入 KKT (Karush-Kuhn-Tucker) 优化条件,可简化为:

$$b = y_i \pm \varepsilon - \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) \quad (5)$$

对于式(4)和式(5)中 SVR 核函数中最佳参数的取值求解,目前没有统一的方法,考虑到 PSO 算法具有分散式搜索、连续搜索等优点,可以很好地

为 SVR 模型计算最优参数服务。本文利用 PSO 算法对轨距劣化趋势预测 SVR 模型参数进行优化,选取均方误差 E_{MS} 作为 PSO 算法的适应度函数,其形式为:

$$E_{MS} = \frac{1}{n} \sum_{w=1}^n (y_w - \hat{y}_w)^2 \quad (6)$$

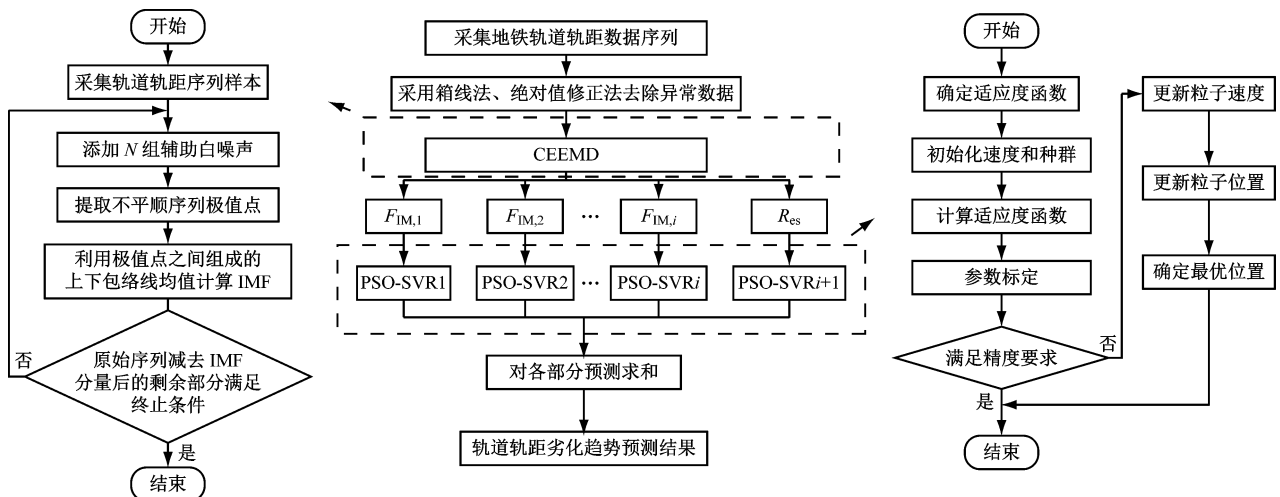
式中:

y_w ——训练集中采集的第 w 个轨距真实值;

\hat{y}_w ——训练集中采集的第 w 个轨距预测值。

1.2.2 预测模型计算流程

轨距变化具有动态性,表现在轨道区间劣化程度上。针对轨距序列的混沌特点,提出了基于 CEEMD 和 SVR 的轨距劣化趋势预测模型。计算流程如图 1 所示。



注: PSO-SVR i 代表 $F_{IM,i}$ 构成的预测模型。

图 1 基于 CEEMD 和 SVR 的轨距劣化趋势预测模型计算流程

Fig. 1 Calculation process of the track gauge deterioration prediction model based on CEEMD and SVR

2 实例验证与分析

2.1 数据来源

本文研究数据来源于上海轨道交通 16 号线的轨检仪检测结果,上线检测时间 2022 年 1 月 23 日,检测轨道区间为上行线 K12 + 134—K15 + 743,涉及长度为 3 609 m,经整理获得检测连续数据 1 128 组。为便于分析,在考虑曲线加宽等因素的前提下,将上述数据处理后换算成与标准轨距的偏差值。

2.2 数据处理

2.2.1 异常值初步判断

在轨距指标测量过程中,受个人经验、设备灵敏度等因素影响,采集到的数据存在一定的异常值。为准确判断轨距劣化趋势,须对异常值进行判别、剔除。常见异常值判断有箱线图、 3δ (δ 为标准偏差) 准则及绝对值修正法等。

本文采用箱线图结合绝对值修正法对原始检测数据进行异常值判断。

2.2.1.1 绝对值修正法

根据统计学原理,具有动态性的数据波动范围

往往在一定的阈值内。设阈值为 Z , 轨距序列样本为 c , 则序列中的异常值分布范围为 $|c_j| \geq Z(|c_j|$ 为第 j 个样本绝对值)。采用绝对值修正法求得 $\bar{c} \neq 0$ (\bar{c} 为样本均值) 的序列样本绝对值的均值 $|\bar{c}|$ 。轨距序列信息异常值初步判断阈值 Z 为:

$$Z = k \left(\frac{1}{n} \sum_{j=1}^n |c_j| \right) = k |\bar{c}| \quad (7)$$

式中:

k ——放大系数,为经验值,一般取 4.0 ~ 5.0, 本文取平均值 4.5。

2.2.1.2 箱线图

箱线图是一种常见的异常值判断方法,无须预设数据分布属性,直观反映给定序列的分布情况。通常箱线图包含 5 个统计量信息,从上到下依次是上限值、上四分位数 Q_1 、中位数、下四分位数 Q_3 、下限值。初步将处于统计量信息上下限之外的离群值判断为异常值,再结合轨距的轨道区间复核确认。

将轨道区间内采集的轨距数据序列按照线型划分为圆曲线(6 段)、直线(5 段)及缓和曲线(10 段)。对 21 段的连续轨距序列可能存在的异常值

通过箱线图判断。轨距箱线图如图 2 所示。

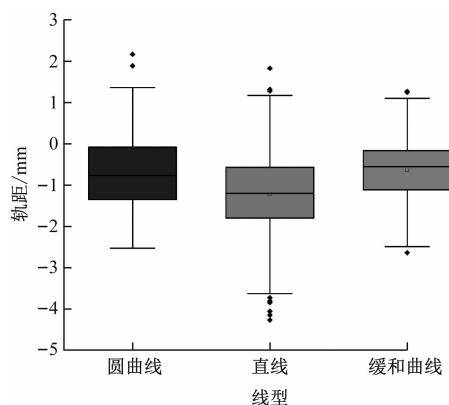


图 2 轨距箱线图

Fig. 2 Box plot of track gauge

从图 2 可以看出:缓和曲线段检测结果最佳,离群值较少,样本检测中误差概率最低;直线段检测状态最差且离群值较多,样本检测中误差概率最高,是否将离群值确定为异常值需结合绝对值修正法阈值 Z 及现场数据来综合考虑。

2.2.2 异常值处理步骤

经上述过程分析初步确认后纳入异常值管理,是否最终确认为异常值,需根据综合分析的结果来确定。轨距序列检测数据中对异常数据的判断结合箱线法、绝对值修正法的综合分析过程如下:

步骤 1 收集轨检仪检测的原始数据。

步骤 2 上海申通地铁集团有限公司编制的 Q/SD-WBZ-FB-SS-GWG22—2021《轨道设备维护规程》中关于轨检仪检测数据的管理规定的具体内容,见表 1。结合表 1 超限等级标准,对异常值进行初步判断。

表 1 轨检仪检测数据管理规定

Tab. 1 Regulations on the management of track inspection

instrument detection data				单位:mm	
超限等级	偏差值				
	轨距	水平	高低	轨向	三角坑
I 级	$\geq +6, \leq -3$	6	6	6	6
II 级	$\geq +8, \leq -5$	8	8	8	8
III 级	$\geq +10, \leq -7$	10	10	10	10

步骤 3 将箱线图法确定的离群值范围与绝对值修正法确定的异常值取平均值,作为最终的异常值阈值标准。

步骤 4 对超出评价异常值阈值标准且达到超限值范围内的数据进行现场复核,采用实测结果进

行替换。

步骤 5 对超出评价异常值阈值标准但未达到超限值范围内的数据,均视为异常值处理。

通过上述分析,锁定异常值 c_j ,并采用其里程前后点的平均值将其替换为 c'_j 。

2.3 分解结果分析

对研究范围内的 1 128 组轨距序列数据进行 CEEMD 分解处理,结果见图 3。从轨道区间内轨距的原始序列看,轨距分布是一种复杂多耦合条件共同作用的结果,无规律可循,非线性较强。经添加辅助白噪声,利用 CEEMD 算法对经检测并处理异常数据后的轨道几何轨距指标进行分解,各频域分量逐渐平缓,原始具有较强混沌特性的轨距序列分解成了波动较小的 9 个 F_{IM} 及 R_{es} 。 F_{IM} 分量一定程度上反映了在某一作用条件下轨距的变化规律; R_{es} 属于最低频分量,揭示了轨距在轨道区间内的总体变化趋势。通过对轨道几何轨距指标进行 CEEMD 分解,降低了非线性程度,提升了轨距劣化趋势发展的可预测性。

2.4 预测结果

选用自学习能力较强的 SVR 模型进行预测,将前 95% 样本数据(1 075 组)划分为输入训练集,后 5% 样本数据划分为测试集(53 组),用于验证预测模型的准确性。在各分解分量的训练过程中利用 PSO 算法对 SVR 的惩罚因子 a 、核参数 g 最佳组合方式进行优化。

优化后的最佳参数,用于测试集的模型标定,以提升预测模型的拟合精度。经 CEEMD 法分解后 SVR 模型预测结果见表 2。

由表 2 可以看出: $F_{IM,1}$ 预测效果较差,相关系数 ξ^2 仅为 0.35,主要原因在于 $F_{IM,1}$ 属于高频模态,序列非线性程度较高;其余分解分量预测的 ξ^2 均超过 0.95, R_{es} 预测结果的 ξ^2 甚至达到了 1,一定程度上反映了 CEEMD 分解相对比较彻底,很好地处理了模态混叠现象。对各分解分量预测结果叠加得到最终预测结果,如图 4 所示。

2.5 性能评价

为了验证本文构建的 CEEMD-PSO-SVR 预测模型的有效性,利用 PSO-SVR、ARIMA(自回归移动平均模型)与同样轨道区间内的轨距指标数据中后 5% 的样本进行预测对比。各模型对轨距劣化预测结果对比见图 5。

通过各模型的预测结果来看,本文构建的预测

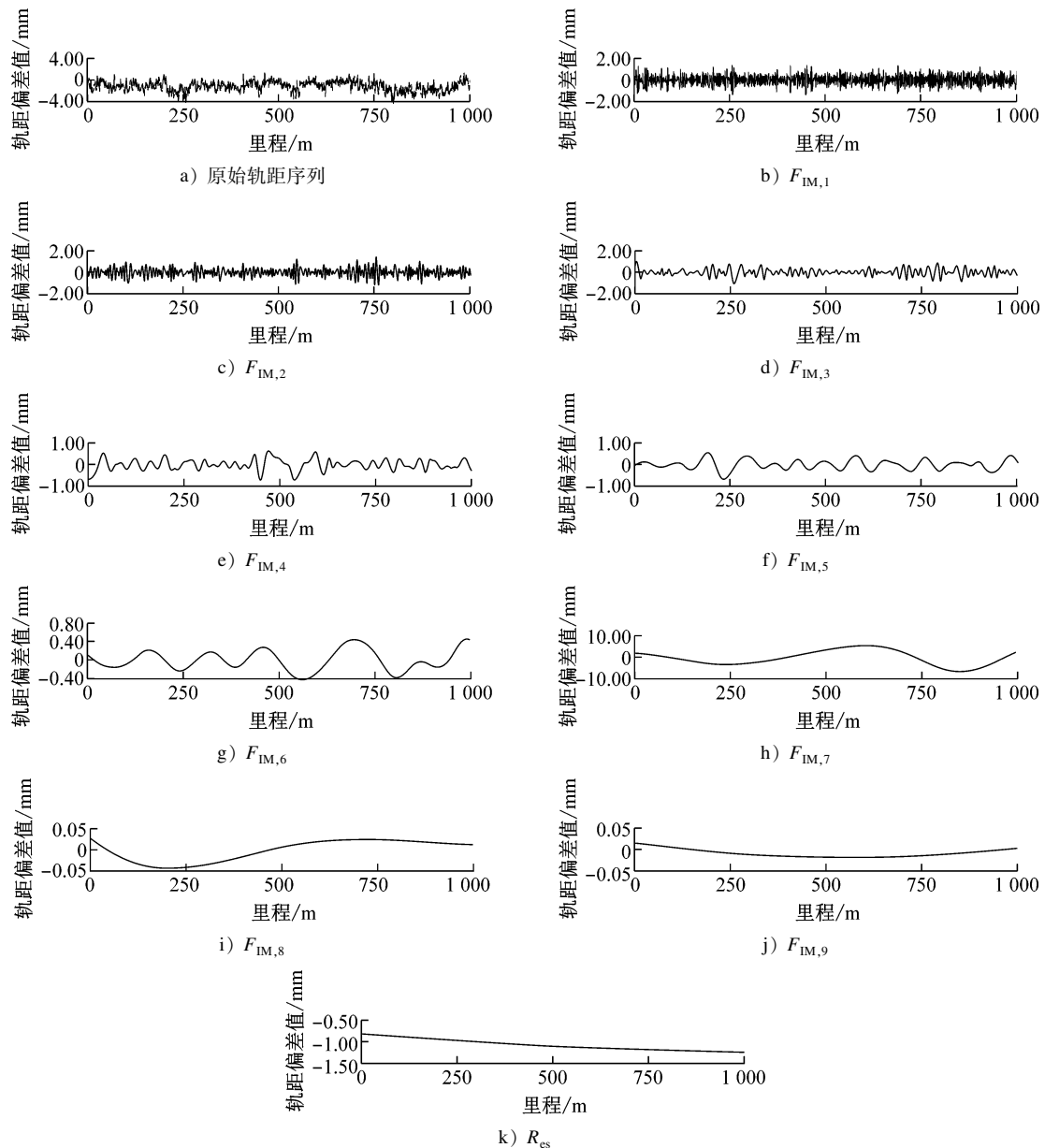


图3 轨距序列数据 CEEMD 分解结果

Fig. 3 CEEMD decomposition result of the track gauge series data

模型与轨道区间内轨距分布真实值相比,吻合度较好,预测偏差量在三者中波动小,CEEMD-PSO-SVR 预测结果与对比模型相比,表明本文构建的组合模型能够准确对轨道区间内轨距劣化发展规律进行预测。

对各模型预测结果进行量化,选用均方根误差 E_{RMS} 、平均绝对误差 E_{MA} 、平均相对误差绝对值 E_{MAP} 等 3 项指标进行评价。各预测模型性能评价结果见表 3。 E_{RMS} 反映了样本之间的离散程度,3 种预测模型中 CEEMD-PSO-SVR 的离散程度最小,仅为 0.432 4; E_{MAP} 衡量的是误差总体水平,可以看出

CEEMD-PSO-SVR 总体误差率最小,PSO-SVR 高于 ARIMA 模型; E_{MA} 反映预测误差的实际情况,可以看出 CEEMD-PSO-SVR 的预测误差最小,ARIMA 预测误差最大。从上述 3 个性能评价指标结果看,本文构建的模型均表现出最优性能,且评价结果较为稳定。

3 结语

1) 本文通过 CEEMD 算法对轨道区间内的轨距信息进行分解,产生了具有表征不同尺度内的轨距序列数据,降低了轨距变化的非线性,可为轨道

表 2 经 CEEMD 法分解后 SVR 模型预测结果

Tab.2 Prediction results of SVR model after CEEMD decomposition

分量	E_{RMS}	最佳参数组合	ξ^2
$F_{IM,1}$	0.421 2	$a = 2.026\ 4, g = 0.01$	0.35
$F_{IM,2}$	0.095 6	$a = 100.000\ 0, g = 0.01$	0.89
$F_{IM,3}$	0.031 8	$a = 100.000\ 0, g = 0.01$	0.99
$F_{IM,4}$	0.009 6	$a = 73.537\ 1, g = 0.01$	0.99
$F_{IM,5}$	0.002 9	$a = 6.261\ 3, g = 0.01$	0.99
$F_{IM,6}$	0.004 7	$a = 9.747\ 5, g = 0.01$	0.99
$F_{IM,7}$	0.010 4	$a = 84.223\ 9, g = 0.01$	0.95
$F_{IM,8}$	0.000 3	$a = 100.000\ 0, g = 0.01$	0.99
$F_{IM,9}$	0.000 4	$a = 100.000\ 0, g = 0.01$	0.99
R_{es}	0.004 5	$a = 95.178\ 3, g = 0.01$	1.00

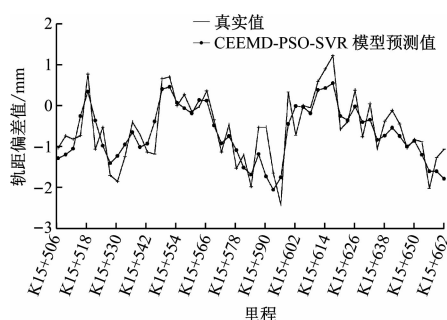


图 4 CEEMD-PSO-SVR 模型预测结果

Fig.4 Prediction results of CEEMD-PSO-SVR model

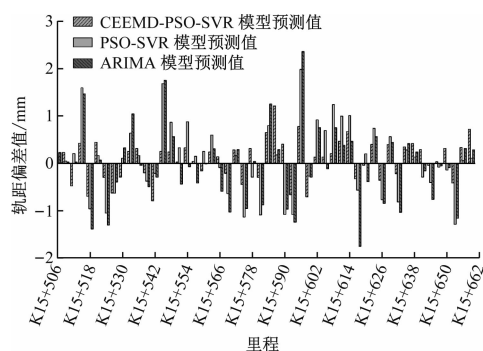


图 5 各模型对轨距劣化预测结果对比

Fig.5 Prediction result comparison of the gauge deterioration by different models

表 3 各预测模型性能评价结果

Tab.3 Performance evaluation results by different prediction models

预测模型	E_{MA}	E_{RMS}	E_{MAP}
PSO-SVR	0.590 2	0.754 4	1.750 8
ARIMA	0.606 7	0.800 9	1.336 2
CEEMD-PSO-SVR	0.366 8	0.432 4	0.931 2

注: 自回归移动平均模型 ARIMA 中自回归阶数为 2, 差分次数为 2, 平均移动阶数为 2。

区间内轨距劣化机理研究提供参考。

2) 针对具有不同特征尺度的轨距序列数据, 构

建立了相应的预测模型, 并采用 PSO 算法来优化 SVR 核参数, 从而实现预测目标, 降低了具有高度非线性数据预测的难度。

3) CEEMD-PSO-SVR 预测模型的结果在各性能评价指标上均优于对比模型, 并具有较强的精度。本模型可以对轨道区间内轨距劣化趋势进行预测应用, 对后续动态优化轨道养护维修计划具有指导意义。

参考文献

- [1] APEZETXEA I S, PEREZ X, ALONSO A. Experimental validation of a fast wheel wear prediction model[J]. Wear, 2021, 486: 204090.
- [2] 曲建军, 高亮, 田新宇, 等. 基于灰色理论的轨道几何状态中长期时变参数预测模型的研究[J]. 铁道学报, 2010, 32(2): 55.
QU Jianjun, GAO Liang, TIAN Xinyu, et al. Study on the mid & long term prediction model of track geometry state based on the grey time-varying parameters theory[J]. Journal of the China Railway Society, 2010, 32(2): 55.
- [3] 常欢, 刘仍奎, 方圆. 改进的轨道质量指数线性预测模型研究[J]. 铁路计算机应用, 2010, 19(6): 1.
CHANG Huan, LIU Rengkui, FANG Yuan. Research of improved linear prediction model of Track Quality Index[J]. Railway Computer Application, 2010, 19(6): 1.
- [4] LASISI A, ATTOH-OKINE N. Principal components analysis and track quality index: a machine learning approach[J]. Transportation Research Part C: Emerging Technologies, 2018, 91: 230.
- [5] 李再韩, 雷晓燕, 高亮. 轨道不平顺检测数据的预处理方法分析[J]. 铁道科学与工程学报, 2014, 11(3): 43.
LI Zaiwei, LEI Xiaoyan, GAO Liang. Analysis of preprocessing methods of track irregularity inspection data[J]. Journal of Railway Science and Engineering, 2014, 11(3): 43.
- [6] 徐磊, 翟婉明. 铁路轨道不平顺的时-频能量联合分析方法[J]. 铁道学报, 2017, 39(4): 9.
XU Lei, ZHAI Wanming. Time-frequency energy joint analysis method for railway track irregularity[J]. Journal of the China Railway Society, 2017, 39(4): 9.
- [7] WANG T, ZHANG M, YU Q, et al. Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal[J]. Journal of Applied Geophysics, 2012, 83: 29.
- [8] DA SILVA SANTOS C E, SAMPAIO R C, DOS SANTOS COELHO L, et al. Multi-objective adaptive differential evolution for SVM/SVR hyperparameters selection[J]. Pattern Recognition, 2021, 110: 107649.

· 收稿日期:2022-09-08 修回日期:2023-01-14 出版日期:2025-01-10

Received:2022-09-08 Revised:2023-01-14 Published:2025-01-10

· 通信作者: 贾清天, 工程师, 13029818701@163.com

· ©《城市轨道交通研究》杂志社, 开放获取 CC BY-NC-ND 协议

© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license