DOI：10.13890/j.issn.1000-128X.2023.06.009　　文章编号：1000-128X(2023)06-0071-08

基于GA-岭回归分析的机车车轮踏面磨耗量预测算法研究

方　鑫[1]，刘　通[2]，程亚萍[2]，孙宇铎[2]，王菲儿[2]
（1.中国铁路上海局集团有限公司 上海机辆段，上海　200070；
2.中国铁道科学研究院集团有限公司 金属及化学研究所，北京　100081）

摘　要：车轮踏面磨耗量是评价机车运行安全的重要参数，但多数车轮运维现场的条件尚不能对踏面磨耗量及时、准确地进行监测。为了解决上述问题，文章提出了基于“GA-岭回归”分析的机车车轮踏面磨耗量预测算法（简称“GA-岭回归”预测算法）。“GA-岭回归”预测算法分为数据前处理和数据预测分析2部分。对于数据前处理，首先依据不同的测量方式对采集到的踏面磨耗量数据进行分类，基于车轮的实际运维情况，分析不同类型数据的特点；随后以镟修周期作为数据划分标准，对分类后的数据进行切片处理；最后采用相关标准和主成分分析法对相应的动态测量数据进行清洗、降噪。对于数据预测分析，首先划分数据集，进行数据整合，创建训练集数据的时间滑动窗口；随后采用岭回归算法对训练集数据进行回归分析训练，并结合遗传算法和验证集数据进行模型参数的调优，以提高预测准确性。基于测试集数据，分别采用传统预测算法、岭回归线性预测算法和“GA-岭回归”预测算法，对3类预测算法的预测效果进行对比分析；随后采用相同的测试方法，扩大车轮样本的数量，继续进行预测效果的对比分析。测试结果表明，采用“GA-岭回归”预测算法的预测误差和误差标准差均相对较低。经分析可知，“GA-岭回归”预测算法具有较高的预测精度，同时可保证更好的预测稳定性。

关键词：机车车轮；车轮踏面磨耗量预测；主成分分析；岭回归；遗传算法
中图分类号：U260.331^{+}.1　　**文献标识码：**A

Research on the prediction algorithm of tread wear for locomotive wheels based on GA-ridge regression analysis

FANG Xin[1], LIU Tong[2], CHENG Yaping[2], SUN Yuduo[2], WANG Feier[2]

(1. Shanghai Depot, China Railway Shanghai Group Co., Ltd., Shanghai 200070, China;
2. Metals and Chemistry Research Institute, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

Abstract: Wheel tread wear is an important parameter to evaluate the operational safety of locomotives, yet timely and accurate monitoring is often lacking at wheel operation and maintenance sites. To this end, this paper proposed a prediction algorithm of tread wear for locomotive wheels based on GA-ridge regression analysis (hereinafter referred to as the "GA-ridge regression" prediction algorithm). This algorithm consisted of two steps: data pre-processing and data-based prediction analysis. In the first step, collected tread wear data was classified according to different measurement methods, and characteristics of different data types were analyzed considering the actual operation and maintenance of wheels. The classified data was then sliced using the profiling cycle as the data partitioning criterion, followed by cleaning and noise reduction of the corresponding dynamic measurement data by relevant criteria and principal

基金项目：中国国家铁路集团有限公司科技研究开发计划项目（P2023J001）；中国铁道科学研究院集团有限公司科研项目（2022YJ245）
通信作者：刘　通（1985—），男，博士，高级工程师，主要从事机车车辆故障预测及健康管理（PHM）方面的研究；E-mail: liutongtky@163.com
引用格式：方鑫, 刘通, 程亚萍, 等. 基于GA-岭回归分析的机车车轮踏面磨耗量预测算法研究[J]. 机车电传动, 2023(6): 71-78.
Citation: FANG Xin, LIU Tong, CHENG Yaping, et al. Research on the prediction algorithm of tread wear for locomotive wheels based on GA-ridge regression analysis[J]. Electric drive for locomotives,2023(6): 71-78.

component analysis. In the second step, data was integrated into datasets, and a time-sliding window was created for the training set data. The ridge regression algorithm was used to train the training set data for regression analysis, and the model parameters were tuned using a combination of the genetic algorithm and the validation set data to improve the prediction accuracy. The test set data was used for prediction by the traditional prediction algorithm, ridge regression linear prediction algorithm, and GA-ridge regression prediction algorithm respectively to compare and analyze their prediction effects. Additionally, a comparative analysis was conducted using the same evaluation method and sample wheels in an expanded size to further assess the prediction effects. The results indicate relatively lower prediction errors and standard deviations of errors when using the GA-ridge regression prediction algorithm. This research concludes that the GA-ridge regression prediction algorithm provides higher prediction accuracy and better prediction stability.

Keywords: locomotive wheels; wheel tread wear prediction; principal component analysis; ridge regression; genetic algorithm

0 引言

机车车轮承载了机车的绝大部分载荷，并通过轮轨黏着为机车提供制动力和牵引力[1]。踏面磨耗量作为车轮服役的重要安全评价指标，其变化不仅改变了车轮踏面的型面特征，也对机车整体系统的响应产生影响[2]。在现场运维过程中，工程师常根据踏面磨耗量预估车轮轮径，作为判断轮径差的重要依据。因此，踏面磨耗量直接影响机车的行车安全，工程师需及时、准确地掌握踏面磨耗量的变化情况，这对测量数据采集的及时性和准确性提出了较高的要求。

针对上述情况，机务段需组织大量人员测量车轮的踏面磨耗量，由镟修车床或人工测量的数据虽然可保证一定的准确性，但由于段内机车的配属量通常较大，若要保证较高的测量频率，需要耗费大量的人力、物力。目前，多数机务段已安装“铁路机车车辆出入库检测系统（TDDS）”，可对车轮的踏面磨耗量等数据进行回段监测，获取其“动态测量数据”。该类型数据虽然在一定程度上缩短了测量周期，但其准确性和稳定性均不及镟修车床或人工测量的数据，孤立的动态测量数据无法准确反映相应时刻的踏面磨耗量。若能在一定时间内对车轮踏面磨耗量进行较为准确的预测，便可有效避免因踏面磨耗量、轮径差超限而造成的行车安全问题。

目前，很多学者已对车轮踏面磨耗量的预测进行了深入研究。文献[3]采用UM和MATLAB软件，建立了“车辆–轨道”的动力学耦合模型和车轮踏面磨耗预测模型，通过数值仿真的方式预测车轮踏面磨耗的变化规律。文献[4]采用ABAQUS建立三维动态有限元模型，对轮轨滚动接触过程中的车轮踏面磨耗量进行数值仿真分析。以上方法需要完善的机车参数、线路条件等数据，准确预测则要建立复杂的仿真模型，较大的计算量也导致了较长的计算时间，现场适用性普遍较差。调研发现，机车运维单位多采用传统预测方法，通过计算上一镟修周期的踏面磨耗率，以及其在预测时刻与当前时刻间的差值，推测未来的踏面磨耗量，但机车运用情况通常较为复杂，该方法很难保证预测的准确性。

很多领域已开始采用数据回归分析的方式，通过探索自变量与因变量间的关系，在一定时间内完成因变量的预测分析[5-7]。经调研可知，在同一镟修周期内，车轮踏面磨耗量与运用时间存在一定的关系，故可采用回归分析模型结合遗传算法（Genetic Algorithm，GA）[8]，预测一定时间段内的车轮踏面磨耗量。

本文依据不同的测量方式对机车车轮踏面磨耗量数据进行分类，基于现场运维情况，分析不同类型数据的特点，随后以镟修周期作为基础计算区间对分类后的数据进行切片处理；依据动态测量数据的特点，首先采用相关标准和主成分分析法（Principal Component Analysis，PCA）对其进行清洗、降噪，接着划分数据集，进行数据整合，并建立训练集的时间滑动窗口；随后采用岭回归算法预测镟修周期内的车轮踏面磨耗量，并结合遗传算法进行模型的参数调优，提高预测准确性；最后对比传统预测算法、岭回归线性预测算法和“GA-岭回归”预测算法的预测误差，评估“GA-岭回归”预测算法的准确性。

1 数据前处理

1.1 数据分类与时间差值计算

根据《铁路技术管理规程（普速铁路部分）》（TG/01A—2017）（简称“技术规程”），以数据的测量方式为依据，将采集的机车车轮踏面磨耗量数据分为4类，分别为镟前测量数据、镟后测量数据、日常测量数据、动态测量数据。提取某一车轮的踏面磨耗量数据，对其分类，并进行数据切片，如图1所示。由于镟前、镟后测量数据和日常测量数据均由人工进行测量和采集，故可统称为人工测量数据。

镟后测量数据和镟前测量数据是在机车车轮镟修时获得，测量周期相对较长，由于影响机车车轮镟修的因素较多，实际镟修周期并不稳定。根据车轮的现场运维和管理情况，将本次镟后测量数据作为本镟修周期内的首个测量数据，由于镟修通常已消除车轮踏

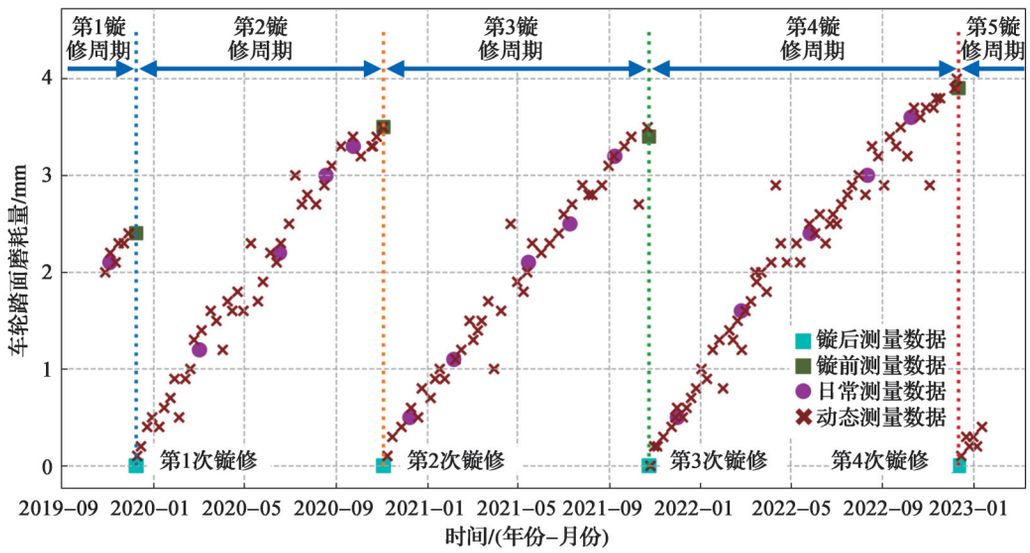

图1　车轮踏面磨耗量测量数据

Fig. 1　Wheel tread wear measurement data

面磨耗，因此该值默认为0 mm；将下一次的镟前测量数据作为本镟修周期内的末尾数据，该值理论上为本镟修周期内的最大值。

如图1所示，该车轮数据共存在4组镟修测量数据，可将其分为5个镟修周期（包含3个数据类型完整的镟修周期）。由于镟修通常已消除车轮失圆、擦伤等影响踏面磨耗量测量准确性的因素，因此镟后测量数据的准确性普遍高于镟前测量数据。

日常测量数据在机车部分整备和小辅修时获取，测量周期相较于镟前、镟后测量数据较短；若机务段已安装可测量车轮踏面磨耗量的镟修车床，则日常测量数据的准确性不及镟前测量数据；若机务段未安装该镟修车床，则日常测量数据与镟前测量数据均采用人工手持第四种检查器测得[9]，因此两者在准确性上并无差异。

动态测量数据由机务段内安装的TDDS测量所得，机车每次回段整备通过该设备时，均会采集相应的动态测量数据。如图1所示，动态测量数据的采集周期远低于日常测量数据，可完整地记录车轮踏面磨耗量的变化情况。但动态测量数据容易受到设备性能、环境和被测物体等客观因素的影响，因此准确性相对较低，数据值波动较大。通过数据前处理，可有效提高动态测量数据的质量，有助于提高车轮踏面磨耗量的预测准确性。

如图1所示，首先以车轮的镟前/镟后测量数据所对应的测量时间作为车轮各镟修周期的划分依据；随后提取车轮的日常测量数据、动态测量数据及其测量时间，以测量时间为基准进行排序，将上述2类数据切片后划分至相应的镟修周期中。经统计可知，该车轮共存在镟前、镟后测量数据各4条，日常测量数据15条，动态测量数据151条，可划分5个镟修周期。由于第4镟修周期数据种类齐全、分布较均匀，故选取该镟修周期进行预测算法的训练和测试。将时间数据转化为该镟修周期的时间差值，以“天”作为单位，镟后测量数据所对应的时间差值默认为0，如图2所示。

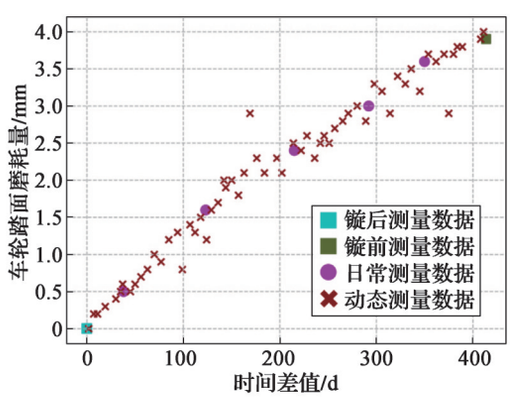

图2　第4镟修周期车轮踏面磨耗量测量数据

Fig. 2　Wheel tread wear measurement data for the 4th profiling cycle

1.2　动态测量数据清洗

1.2.1　基于相关标准的数据清洗

根据“技术规程”的要求，机车车轮踏面磨耗量的限值不可大于7 mm。依据限值要求，在数据清洗算法中加入相应的判断逻辑，当动态测量数据已显示踏面磨耗量大于7 mm时，按以下逻辑进行处理：

①若后续的人工测量数据并未超限，则认为动态测量数据存在问题，应清洗掉；

②若后续的人工测量数据同样超限，则认为动态测量数据正常，予以保留。

1.2.2　基于PCA的数据清洗

PCA是一种较为常用的降维方法[10]，可将高维的数据通过线性变换投影至低维空间[11]。在维度变换的过程中，对相应的特征值进行排序，数据变量的方差越大，则表明包含的原始信息越多[12]。

首先，将人工测量数据和其对应的测量时间差值合并为二维数组$\boldsymbol{X}$；随后，计算$\boldsymbol{X}$的协方差矩阵$\boldsymbol{C}$，由于$\boldsymbol{C}$为实对称矩阵，故一定存在2个单位正交的特征向量e_1和e_2，及其对应的特征值λ_1和λ_2，则特征值λ_1、λ_2即为数据变量的方差。其中，二维数组$\boldsymbol{X}$、协方差矩阵$\boldsymbol{C}$、协方差矩阵$\boldsymbol{C}$的特征值λ_1和λ_2，及其对应的特征向量e_1、e_2关系为

$$\boldsymbol{X}=\begin{bmatrix} z_1 & z_2 & \cdots & z_m \\ \Delta t_1 & \Delta t_2 & \cdots & \Delta t_m \end{bmatrix} \tag{1}$$

$$\boldsymbol{C}=\frac{1}{m}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}=\begin{bmatrix} \dfrac{1}{m}\sum_{i=1}^{m}z_i^2 & \dfrac{1}{m}\sum_{i=1}^{m}\left(z_i\Delta t_i\right) \\ \dfrac{1}{m}\sum_{i=1}^{m}\left(z_i\Delta t_i\right) & \dfrac{1}{m}\sum_{i=1}^{m}\Delta t_i^2 \end{bmatrix}= \begin{bmatrix} \operatorname{cov}(z,z) & \operatorname{cov}(z,\Delta t) \\ \operatorname{cov}(z,\Delta t) & \operatorname{cov}(\Delta t,\Delta t) \end{bmatrix} \tag{2}$$

$$\boldsymbol{E}=\begin{bmatrix} e_1 & e_2 \end{bmatrix} \tag{3}$$

$$\boldsymbol{E}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{E}=\Lambda=\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \tag{4}$$

式中：z为人工测量数据，mm；Δt为测量时间差值，d；m为人工测量数据的数量。

由于2组特征向量e_1、e_2需满足线性无关的要求，故特征向量e_1、e_2相互垂直，于是建立以e_1、e_2为基准的坐标系。此时将动态测量数据和其测量时间差值合并为二维数组$\boldsymbol{Y}$，通过式(5)将$\boldsymbol{Y}$中的数据点映射至e_2坐标轴上，采用PCA法完成动态测量数据的降维，如图3所示。

$$\boldsymbol{Y}'=e_2\boldsymbol{Y} \tag{5}$$

式中：$\boldsymbol{Y}'$为动态测量数据在e_2坐标轴的映射集。

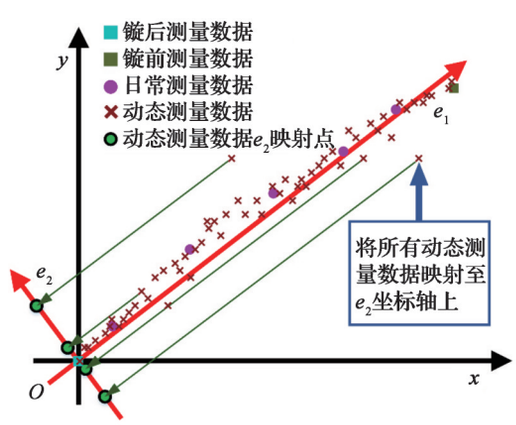

图3 动态测量数据映射至e_2坐标轴

Fig. 3 Mapping of data from dynamic measurement to e_2 axis

对映射集$\boldsymbol{Y}'$中的数据进行K-S检验，检验P值为0.33，若取显著性水平为0.05，认为数据符合正态分布规律[13-14]。计算映射集$\boldsymbol{Y}'$的平均值$\mu=0.12$，标准差$\sigma=0.27$，计算置信度为95%的动态测量数据的置信区间为$[-0.41,0.66]$[15]，清洗掉位于置信区间外的离群数据，如图4所示。

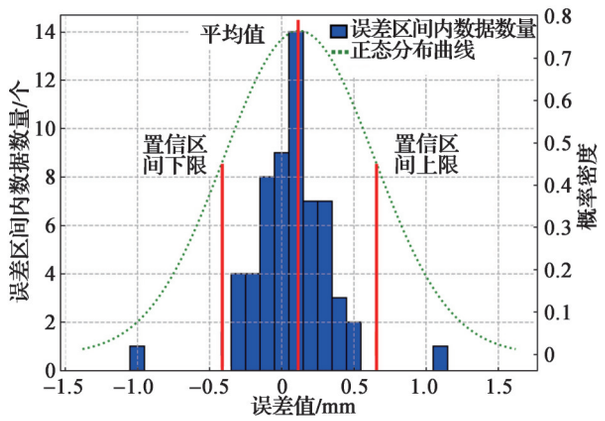

图4 映射集数据误差值分布情况

Fig. 4 Distribution of error values of mapping set data

依据“技术规程”，该车轮第4镟修周期的动态测量数据不存在符合清洗条件的数据；经PCA和相应的离群点分析，共清洗2条置信区间外的离群动态测量数据，如图5所示。

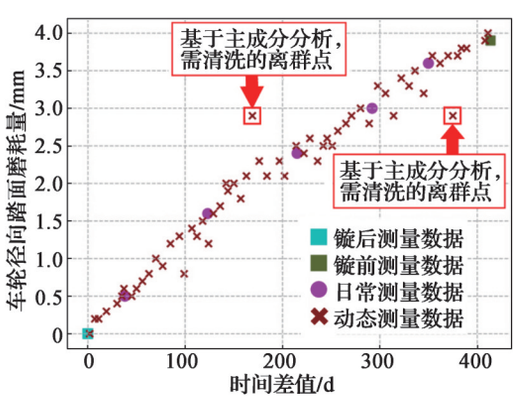

图5 清洗后的数据分布情况

Fig. 5 Distribution of cleaned data

2 数据预测分析

2.1 数据集划分

为提高预测算法的准确性和可靠性，并实现对预测结果的量化评估，需要对踏面磨耗量数据进行数据集划分。由于存在4类踏面磨耗量数据，并且准确性和分布情况等存在较大差异，若直接进行数据整合，随机划分数据集，预测算法易出现较大误差，或陷入局部最优的困境。因此，在保证评估客观的前提下，依据不同数据类型的特点，进行具有针对性的数据集划分。

①测试集：用于评估不同预测算法的准确性和可靠性，需要数据具有较高的准确性。镟后测量数据虽然准确性较高，但数值通常为0 mm，测试效果不佳，故不取用；镟前测量数据位于镟修区间的末尾，可较好地反映踏面磨耗情况，准确性较高，因此可以取用。为丰富测试结论，并保证测试集的分布均匀，取用日常测量数据的中值（以测量时间差值为序）作为测试

集数据。

②验证集：作为预测算法参数调整的重要依据，需结合优化算法，以寻求最优的数据回归模型。验证集首先需保证数据准确性和充足数据量，除此之外，数据分布应较为均匀，同时取值具有一定的随机性。因此，以测量时间差值为序，固定取日常测量数据的最后1个数据作为验证集数据，并将日常测量数据均分为2部分，每部分中随机取1个数据作为验证集数据（不可与最后1个数据点重复）；同样以测量时间差值为序，将动态测量数据均分为4部分，每部分中随机取4个数据作为验证集数据。

③训练集：用于数据回归模型的训练，需要保证一定的数据量。因此，将所有剩余的日常测量数据、动态测量数据和镟后测量数据划分至训练集。

完成数据集划分后，对各数据集的数据进行整合（不再按照测量方式进行数据分类），如图6所示。

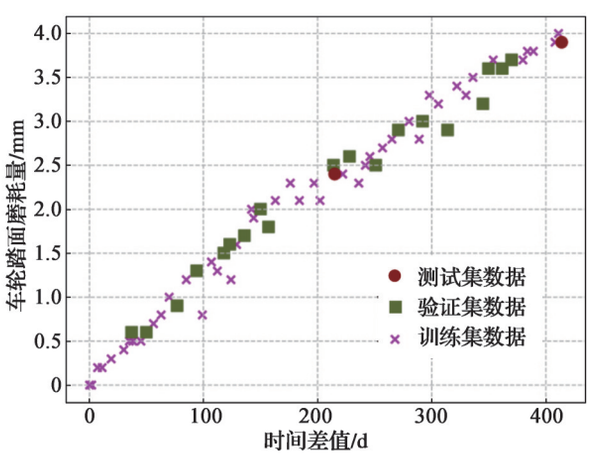

图6　数据集划分情况

Fig. 6　Partion into datasets

2.2　建立训练集的滑动窗口

基于训练集$\boldsymbol{X}^{(0)}=\left[x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)}\right]$创建滑动窗口数据集$\boldsymbol{X}^{(1)}$，具体过程为

$$\left.\begin{aligned}x_1^{(1)}&=\frac{2x_1^{(0)}+x_2^{(0)}}{3}\\x_n^{(1)}&=\frac{x_{n-1}^{(0)}+2x_n^{(0)}}{3}\\x_i^{(1)}&=\frac{x_{i-1}^{(0)}+x_i^{(0)}+x_{i+1}^{(0)}}{3},\quad i=2,3,\cdots,n-1\end{aligned}\right\}\tag{6}$$

如图7所示，设置训练集的滑动窗口长度为3，滑动步长为1。创建滑动窗口可有效降低动态测量数据的波动，保证后续回归分析的准确性[16]。

2.3　GA-岭回归模型构建

岭回归是一种经过改良的最小二乘法，通过在代价函数中添加正则化项，而放弃无偏性。虽然损失了一定量的原始信息特征，但较易获得更优的拟合效果，

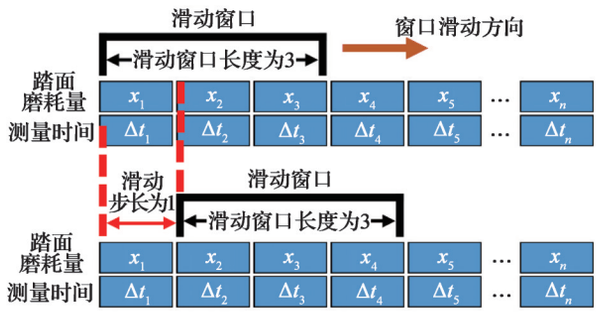

图7　滑动窗口数据集创建

Fig. 7　Dataset creation of sliding window

岭回归算法比较适用于变量间存在一定共线性的问题[17]。

在车轮的现场运维过程中，工程师需要快速、准确地对一定时间内的车轮踏面磨耗量进行预测。岭回归算法不仅能够快速寻找到机车车轮踏面磨耗量与测量时间差值间的关系，同时可有效防止过拟合问题[18]，因此该方法较为适用于车轮运维现场的踏面磨耗量预测。

由于车轮的运用时间差值与踏面磨耗量呈正相关，而且踏面磨耗量的镟后测量数据为镟修周期内的首个数据，通常为0 mm，故建立截距值为0的幂函数模型，如式(7)。

$$z_{\mathrm{pn}}=a\Delta t^{n}\tag{7}$$

式中：z_{pn}为“GA-岭回归”分析预测算法的车轮踏面磨耗量预测值，mm；a, n均为“GA-岭回归”分析预测算法的相关系数。

遗传算法的思想取自于生物进化学中优胜劣汰的规律，检索方法的迭代过程是存活和测试[19]，利用遗传算法可较为准确地寻找到目标函数的最优解[20]。

基于验证集数据，通过式(8)可获取预测算法计算的踏面磨耗量与实际验证集数据间的残差平方和（RSS），评估每次岭回归分析的拟合效果；结合遗传算法，可在一定区间范围内，寻找到拟合效果最优的指定幂值n及其对应的相关参数a，计算流程见图8。为兼顾寻优的准确性和效率，设置DNA长度为30（控制指定幂值n的优化精度），迭代次数为2 000次；依据车轮踏面磨耗量的变化规律，将目标函数最优解的搜索区间设定为[0.5,1.5]。

$$\left[v^2\right]=\sum_{i=1}^{n_{\mathrm{v}}}\left(z_{\mathrm{v}}-\tilde{y}_i\right)^2\tag{8}$$

式中：z_{v}为车轮踏面磨耗量验证集数据，mm；$\tilde{y}_i$为基于验证集数据的测量时间，由算法预测的车轮踏面磨耗量，mm；n_{v}为验证集数据数量。

完成计算后可知，该车轮踏面磨耗数据的最优预测模型为$z_{\mathrm{pn}}=0.037\Delta t^{0.775}$，即$a$=0.037, n=0.775，如图9所示。“GA-岭回归”预测算法分析后获得的回归曲线能够较好地反映车轮踏面磨耗量的变化趋势。

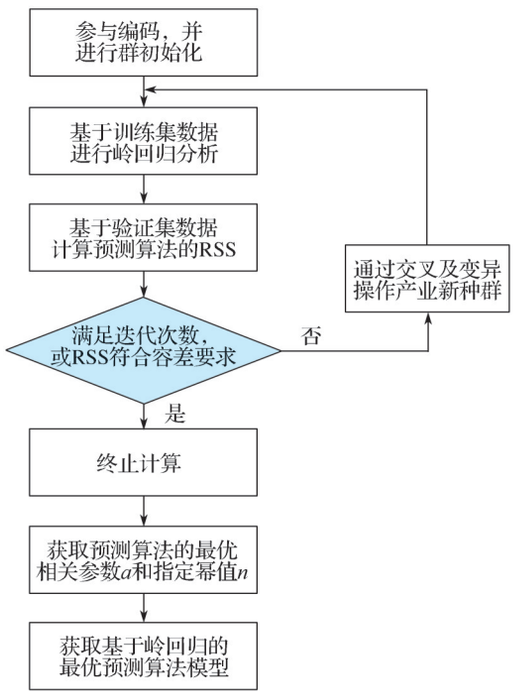

图8 “GA-岭回归”预测算法流程图

Fig. 8 Flow chart of "GA-ridge regression" prediction algorithm

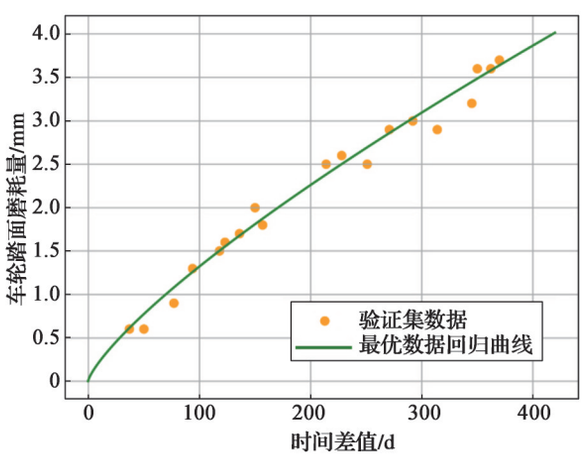

图9 最优预测模型

Fig. 9 Optimal prediction model

3 实例分析

3.1 传统预测算法

传统预测算法多用于车轮运维现场，可快速预测一定时间内的踏面磨耗量。

首先，通过2次人工测量的车轮踏面磨耗量和其对应的测量时间（通常为当前和历史的最近一次），计算该时间段内的车轮踏面磨耗率，计算公式为

$$\delta_t = \frac{\Delta z}{\Delta t} \tag{9}$$

式中：δ_t为车轮踏面磨耗率，mm/d；Δz为车轮踏面磨耗量差值，mm。

传统方法默认在相应的时间区间内，车轮踏面磨耗量与时间呈线性关系，于是根据过去最近时间范围内的踏面磨耗速率，预测下一时刻的踏面磨耗量，计算公式为

$$z_{pt} = z_{now} + \delta_t \Delta t' \tag{10}$$

式中：z_{pt}为传统预测算法的车轮踏面磨耗量预测值，mm；z_{now}为车轮踏面磨耗量当前值，mm；$\Delta t'$为预测值所在时间与当前时间的差值，d。

3.2 岭回归线性预测算法

为了评估遗传算法在“GA-岭回归”预测算法中的作用，采用岭回归线性预测算法对踏面磨耗量进行预测。首先，建立截距值为0的幂函数模型，如式(11)所示；然后，利用岭回归算法对训练集数据进行线性回归分析，获得最优系数a'，即可对未来时刻的车轮踏面磨耗量进行预测。

$$z_{po} = a' \Delta t \tag{11}$$

式中：z_{po}为岭回归线性预测算法的车轮踏面磨耗量预测值，mm；a'为回归分析预测方法的相关系数。

3.3 预测算法的评估和分析

基于测试集数据，对传统预测算法、岭回归线性预测算法和“GA-岭回归”预测算法的准确性进行量化对比评估，分别采用3种预测算法对某车轮镟修周期的中位和末位的踏面磨耗量进行预测，如表1所示。采用传统预测算法的误差值分别为0.40 mm, 0.36 mm，采用岭回归线性预测算法的误差分别为0.17 mm, 0.40 mm，而采用“GA-岭回归”预测算法的误差分别降至0.01 mm, 0.07 mm。由此可知，“GA-岭回归”预测算法的预测误差明显减小。

表1 3种预测方法的误差值对比

Table 1 Comparison of error values among three prediction methods mm

算法类型	踏面磨耗量	误差值	算法类型	踏面磨耗量	误差值
测试集实测值	2.40		测试集实测值	3.90	
传统预测算法	2.80	0.40	传统预测算法	4.26	0.36
岭回归线性预测算法	2.23	0.17	岭回归线性预测算法	4.30	0.40
“GA-岭回归”预测算法	2.39	0.01	“GA-岭回归”预测算法	3.97	0.07

随后扩大验证样本的数量，分别采用传统预测算法、岭回归线性预测算法和“GA-岭回归”预测算法，对5个车轮的10个镟修区间中的踏面磨耗量进行预测。预测误差如图10所示，横坐标标签首位为车轮编号，次位为镟修周期编号，末位为测试集数据编号，$\bar{E}_{pt}$, $\bar{E}_{po}$, $\bar{E}_{pn}$分别表示传统预测算法、岭回归线性预测算法

和“GA-岭回归”预测算法的预测误差均值。由图 10 可知，传统预测算法的误差均值和标准差分别为 0.30 mm, 0.15 mm，岭回归线性预测算法的误差均值和标准差分别为 0.24 mm, 0.14 mm，而“GA-岭回归”预测算法预测误差均值和标准差均为 0.09 mm。由此可知，采用 GA-岭回归预测算法的预测误差较低，预测误差浮动较小。

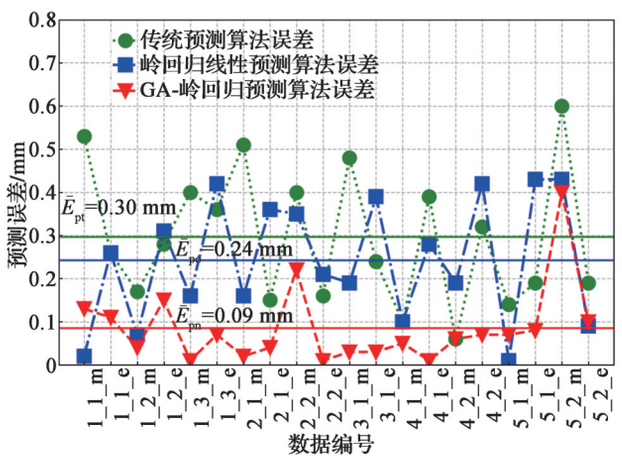

图 10　3 种预测方法的误差值对比

Fig. 10　Comparison of error values among three prediction methods

传统预测方法存在一定的局限性，一方面车轮踏面磨耗量的变化规律受材质、线路和天气等多种客观因素影响，与运用时间并非为简单的线性关系，另一方面人工测量数据的采集周期较长，难以全面地反映车轮踏面磨耗量的增长规律。因此，传统预测算法的误差相对较高。由于以上问题，将已完成前处理的动态测量数据引入踏面磨耗量的预测分析中，若仅采用岭回归算法对训练集数据进行回归分析，以预测踏面磨耗量，预测准确性的提升较为有限。因此，引入遗传算法，结合验证集数据，对岭回归算法的参数进行调优。采用“GA-岭回归”预测算法可快速、准确地找到踏面磨耗量与测量时间差值的最优关系，进行一定时间内的踏面磨耗量预测，克服传统预测算法和岭回归线性预测算法的局限性，有效减小了预测误差，提高预测的准确性和稳定性。

4　结论

为解决机车运维现场无法及时、准确地监测车轮踏面磨耗量的问题，本文提出了“GA-岭回归”预测算法，算法主要分为 2 部分：数据前处理和数据预测分析。“GA-岭回归”预测算法的训练数据在人工测量数据的基础上，引入动态测量数据，并对其进行前处理；随后采用岭回归分析算法预测镟修周期内的车轮踏面磨耗量，并结合遗传算法进行模型的参数调优，获得的回归曲线能够较好地反映车轮踏面磨耗量的变化趋势。

基于测试集数据，分别采用传统预测算法、岭回归线性预测算法和“GA-岭回归”预测算法，对预测效果进行对比分析。传统预测算法的误差分别为 0.40 mm, 0.36 mm，岭回归线性预测算法的误差分别为 0.17 mm, 0.40 mm，而“GA-岭回归”预测算法的误差分别降至 0.01 mm, 0.07 mm，误差减小明显。随后将样本的数量扩大至 10 个，采用相同测试方法，传统预测算法的误差均值和标准差分别为 0.30 mm, 0.15 mm，岭回归线性预测算法的误差均值和标准差分别为 0.24 mm, 0.14 mm，而“GA-岭回归”预测算法的预测误差和标准差均为 0.09 mm。综上所述，相较于传统预测算法和岭回归线性预测算法，“GA-岭回归”预测算法具有较高的预测精度和稳定性。经调研可知，“GA-岭回归”预测算法的预测精度可满足运维单位对机车车轮踏面磨耗量监测和轮径差判断的需求。

参考文献：

[1] 杨艳丽, 孙传喜, 张军, 等. 我国机车车辆车轮磨耗型踏面外形及设计方法[J]. 内燃机车, 2011(1): 6-9.
YANG Yanli, SUN Chuanxi, ZHANG Jun, et al. Review of profile and design of all kinds of domestic worn profile treads for locomotive and rolling stock[J]. Diesel locomotives, 2011(1): 6-9.

[2] 韩鹏, 张卫华. 高速列车轮对磨耗统计规律及预测模型[J]. 机械工程学报, 2016, 52(2): 144-149.
HAN Peng, ZHANG Weihua. Prediction model and verification of wheel wear in high-speed trains[J]. Journal of mechanical engineering, 2016, 52(2): 144-149.

[3] 王红兵, 李艺, 李国芳, 等. 高速车辆车轮踏面磨耗预测及系统参数影响分析[J]. 机械科学与技术, 2022, 41(5): 786-794.
WANG Hongbing, LI Yi, LI Guofang, et al. Prediction of wheel tread wear of high speed vehicles and analysis of influence of system parameters[J]. Mechanical science and technology for aerospace engineering, 2022, 41(5): 786-794.

[4] 常崇义, 王成国, 金鹰. 基于三维动态有限元模型的轮轨磨耗数值分析[J]. 中国铁道科学, 2008, 29(4): 89-95.
CHANG Chongyi, WANG Chengguo, JIN Ying. Numerical analysis of wheel/rail wear based on 3D dynamic finite element model[J]. China railway science, 2008, 29(4): 89-95.

[5] 钟绍林, 秦梦瑶, 廖时元, 等. 基于数据拟合的铁路车站旅客最高聚集人数计算研究[J]. 铁路计算机应用, 2022, 31(8): 1-6.
ZHONG Shaolin, QIN Mengyao, LIAO Shiyuan, et al. Calculation of maximum number of gathering passengers in railway station based on data fitting[J]. Railway computer application, 2022, 31(8): 1-6.

[6] 张贺, 范梦萱. 基于 Lasso 回归模型的青岛海洋经济和海洋产业分析[J]. 海洋开发与管理, 2022, 39(8): 22-28.

ZHANG He, FAN Mengxuan. The marine economy and marine industry in Qingdao based on Lasso regression model [J]. Ocean development and management, 2022, 39(8): 22-28.

[7] 李宏伟, 赵杨坤. 350 km/h复兴号动车组制动闸片在服役过程中的磨损行为及其使用寿命预测[J]. 铁道机车车辆, 2021, 41(2): 22-27.
LI Hongwei, ZHAO Yangkun. Wear performance and life prediction of brake pad in service process of 350 km/h high-speed train[J]. Railway locomotive & car, 2021, 41(2): 22-27.

[8] 杨晓炳, 闫泽鹏, 尹升华, 等. 基于GA-SVM的钢渣基胶凝材料开发及料浆配比优化[J]. 工程科学学报, 2022, 44(11): 1897-1908.
YANG Xiaobing, YAN Zepeng, YIN Shenghua, et al. Development of steel-slag-based cementitious material and optimization of slurry ratio based on genetic algorithm and support vector machine (GA-SVM)[J]. Chinese journal of engineering, 2022, 44(11): 1897-1908.

[9] 程中国. 地铁车辆轮对外形尺寸在线检测系统[J]. 城市轨道交通研究, 2021, 24(9): 228-231.
CHENG Zhongguo. Online detection system of metro vehicle wheelset profile size[J]. Urban mass transit, 2021, 24(9): 228-231.

[10] 颜湘武, 李君岩. 基于主成分分析法的直驱式风电场分群方法[J]. 电力系统保护与控制, 2020, 48(5): 127-133.
YAN Xiangwu, LI Junyan. Grouping method of direct drive wind farm based on principal component analysis[J]. Power system protection and control, 2020, 48(5): 127-133.

[11] 张晓英, 吴宏强, 舒子江, 等. 基于PCA和改进ABC-K聚类算法的双馈机组风电场等值建模[J]. 电气工程学报, 2023, 18(1): 201-210.
ZHANG Xiaoying, WU Hongqiang, SHU Zijiang, et al. Equivalent modeling of DFIG wind farm based on PCA and improved ABC-K clustering algorithm[J]. Journal of electrical engineering, 2023, 18(1): 201-210.

[12] VASWANI N, NARAYANAMURTHY P. Static and dynamic robust PCA and matrix completion: a review[J]. Proceedings of the IEEE, 2018, 106(8): 1359-1379.

[13] 李雪. 一种基于正态分布密度函数的模糊查询方法[J]. 计算机技术与发展, 2018, 28(9): 1-6.
LI Xue. A fuzzy matching method based on normal distribution density function[J]. Computer technology and development, 2018, 28(9): 1-6.

[14] 王巍, 陈崇双, 周宏宇, 等. 基于KS检验的铁路车站日常货运量预测[J]. 综合运输, 2018, 40(12): 93-96.
WANG Wei, CHEN Chongshuang, ZHOU Hongyu, et al. Daily freight volume prediction of railway station based on KS test[J]. China transportation review, 2018, 40(12): 93-96.

[15] 何小年, 段凤华. Python在统计学课程实训中的应用探索[J]. 软件, 2021, 42(9): 32-34.
HE Xiaonian, DUAN Fenghua. Exploration on the application of Python in the practical training of statistics course [J]. Computer engineering & software, 2021, 42(9): 32-34.

[16] 孟建军, 潘彦龙, 陈晓强, 等. 基于灰色BP神经网络的高速列车轴箱轴承温度预测方法[J]. 轴承, 2022(4): 77-82.
MENG Jianjun, PAN Yanlong, CHEN Xiaoqiang, et al. Prediction method for temperature of axle box bearings for high speed train based on grey BP neural network[J]. Bearing, 2022(4): 77-82.

[17] 朱旺, 毛宝俊, 谢强. 1 100 kV特高压变压器套管震后力学性能快速评估方法[J]. 高电压技术, 2022, 48(12): 4904-4914.
ZHU Wang, MAO Baojun, XIE Qiang. Fast evaluation method on post-earthquake mechanical performance of 1 100 kV UHV transformer bushing[J]. High voltage engineering, 2022, 48(12): 4904-4914.

[18] 刘恩江, 宋云胜, 梁吉业. 基于数据划分的核岭回归加速算法[J]. 中国科学技术大学学报, 2018, 48(4): 284-289.
LIU Enjiang, SONG Yunsheng, LIANG Jiye. An accelerator for kernel ridge regression algorithms based on data partition[J]. Journal of university of science and technology of China, 2018, 48(4): 284-289.

[19] 马砺, 张鹏宇, 郭睿智, 等. 巷道火灾密闭过程烟气温度预测的GA-SVM模型[J]. 中国矿业大学学报, 2021, 50(4): 641-648.
MA Li, ZHANG Pengyu, GUO Ruizhi, et al. GA-SVM model for prediction flue gas temperature of roadway fire under sealing process[J]. Journal of China university of mining & technology, 2021, 50(4): 641-648.

[20] 薛赛, 贺青川, 潘骏, 等. 基于GA-SVM的永磁同步电机动态偏心故障诊断[J]. 组合机床与自动化加工技术, 2022(9): 99-103.
XUE Sai, HE Qingchuan, PAN Jun, et al. Research on dynamic eccentricity fault diagnosis method of permanent magnet synchronous motor based on GA-SVM[J]. Modular machine tool & automatic manufacturing technique, 2022(9): 99-103.

收稿日期：2023-07-21
修回日期：2023-09-06