

基于机器学习的高速列车转向架振动信号监测

蔡俊鹏^{1,2}, 吴炳福^{1,2}, 陈德旺^{1,2}

(1. 福州大学 数学与计算机科学学院, 福建 福州 350116;

2. 福州大学 轨道交通研究院, 福建 福州 350116)

摘要: 为了提高列车运行时的稳定性和安全性, 对列车运行时安全加速进行实时监控与预测。根据动车实际运营保存的大量转向架振动信号数据, 不同于传统的基于动力学的研究, 提出了基于机器学习的高速列车转向架振动信号的研究。首先通过小波变换去噪对原始数据进行数字滤波处理, 强化所需部分的数据, 增加了转向架振动信号数据的精确度。其次, 通过实验对几种常用的机器学习算法模型进行参数拟合, 经分析和对比实验结果表明, 随机森林算法在转向架枕梁振动信号监测中性能表现最好且预测精度最优, 均方根误差达到最低 0.069, 其稳定性达到最高。最后总结得出, 在不同里程、不同速度下, 转向架振动信号的有效值和最大值的变化在转向架服役性能、安全运行方面具有可推广性, 能够说明基于随机森林算法的转向架振动信号的监测可以有效地监控和预测高速列车运行速度的安全阈范围并优化列车转向架轮对镟修周期。

关键词: 机器学习; 小波变换去噪; 随机森林算法; 列车转向架

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2019)08-0130-06

doi: 10.3969/j.issn.1673-629X.2019.08.025

High-speed Train Bogie Vibration Signal Monitoring Based on Machine Learning

CAI Jun-peng^{1,2}, WU Bing-fu^{1,2}, CHEN De-wang^{1,2}

(1. School of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China;

2. Institute of Rail Transport, Fuzhou University, Fuzhou 350116, China)

Abstract: In order to improve the stability and safety of high-speed train vehicles, real-time monitoring and prediction of safe acceleration during train operation are carried out. In this paper, according to the data of a large number of bogie vibration signals stored and maintained in the actual operation of the motor vehicle, different from the traditional research based on dynamics, a vibration signal based on machine learning for high-speed train bogies is proposed. First of all, the raw data is filtered digitally by wavelet transform denoising to strengthen the required part of the data and increase the precision of the bogie vibration signal data. Secondly, parameter fitting of several commonly used machine learning algorithm models through experiments, the results of analysis and comparison show that the random forest algorithm has the best performance and the best prediction accuracy in the monitoring of bogie bolster vibration signals. The root-mean-square error reaches the minimum 0.069, and its stability is the highest. Finally, it is concluded that the changes in the effective value and the maximum value of the bogie vibration signal at different mileage and at different speeds can be generalized in terms of the service performance and safe operation of the bogie, and can explain the vibration signal of the bogie based on the random forest algorithm. The monitoring can effectively monitor and predict the safety threshold range of the high-speed train operating speed and optimize the train bogie repair cycle.

Key words: machine learning; wavelet transform denoising; random forest algorithm; train bogie

收稿日期: 2018-08-15

修回日期: 2018-12-19

网络出版时间: 2019-03-21

基金项目: 国家自然科学基金(71671044); 空间数据挖掘与信息共享教育部重点实验室项目(2018LSDMIS09); 工信部倍增计划资助项目(00101522)

作者简介: 蔡俊鹏(1994-), 男, 硕士, 从事数据挖掘、数据分析、机器学习、智能交通系统信息与控制技术的研究; 陈德旺, 博士, 教授, 博导, 从事交通信息工程及控制、机器学习、计算智能、最优化及其应用和智能交通系统的研究。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190321.0942.068.html>

0 引言

中国高速铁路具有高速度、高密度和大载客量等特点,这些指标都体现了中国高速铁路的综合水平已经达到了世界前列。中国高速铁路运营最高速度已经达到了350公里每小时,保持着高速铁路历史最高运营速度的记录。在列车运行中,列车转向架是高速列车的重要部件,在长期的运营中保存了大量的转向架振动信号数据,其振动信号很大程度上决定车辆运行的稳定性和安全性。经实验表明,对转向架振动信号提取的特征进行监测与分析能够准确识别出列车故障的状态,识别率高达92%^[1]。随着高速列车运营里程的增长,转向架的服役性能与安全状态始终处于时变状态,这对转向架的安全与维护带来极大的挑战。因此准确地分析转向架的振动数据信号,能有效指导高速列车的安全提速,优化转向架的维修周期,有利于缓解中国高速铁路运营和交通运输的压力,提高中国高速铁路的运营竞争力,是一个重大而有意义的课题^[2-3],对高速列车运行的经济性和安全性都具有重要意义。

在转向架部件研究方面,国内外科研工作人员在监控技术的基础上,使用增强实验仿真的方式,以不同的方法对转向架的故障检测诊断问题进行了深入研究,并取得了许多成果。例如,金炜东等^[4]对转向架故障数据进行分类训练,训练模型的准确率达到90%以上。秦娜等^[5-6]采用集合经验模态分解EEMD对列车振动信号进行分析,进一步提高了支持向量机的转向架故障特征的识别率。侯亚东等^[7-8]采用小波变换处理转向架齿轮故障监测信号。Hinton等^[9-10]基于人脑学习思想提出深度学习的机器学习方法,Dahl等^[11]验证了基于深度学习训练转向架故障功率谱数据的可行性。庞荣等^[12]利用深度学习的方法对转向架的故障特征提取,对其中关键部件的识别正确率有了很大的提升。总的来说,目前国内外对转向架的研究方案都有其理论和实际意义,但是也存在一定的缺陷。基于故障诊断的方法,基于研究现代信号处理方法(EEMD、Hilbert变换等)和数据挖掘分类算法(支持向量机、K最近邻、隐马尔可夫模型等)^[13-15]的研究方法,能有效对转向架故障进行诊断,提高维修效率。但是这些方法都采用增强实验的方法来研究,存在实验成本高、难验证的问题。而且增强实验与实际相比存在性能偏离,实际中的转向架结构参数是缓慢变化的。因此,文中以动车长期运行过程当中保存下来的转向架振动信号为基础,对其分析研究,该方法不存在发生性能偏离和数据量不足等缺点,实验结果更加契合实际运行情况,且实验成本低,能真实反映转向架的服役性能变化规律,具有较好的指导意义。

文中分析对比了国内外关于转向架的研究现状,以实际动车运营数据作为主要研究对象,对复杂的转向架振动信号数据进行小波变换去噪处理。在此基础上,研究转向架部件的振动速度、振动加速度和历史运营里程与振动值的服役性能规律,通过对多元线性回归、反向传播神经网络、径向基函数神经网络、回归决策树和随机森林算法实验进行对比分析,最终得出随机森林算法在动车转向架振动数据的信号监测中精度最优。验证了基于随机森林算法的高速列车转向架振动信号的研究在转向架服役性能和安全运行方面起到有效实时监测的效果,具有一定的研究价值。

1 转向架振动信号数据

文中的转向架数据集是通过专业的轨道交通车辆安全监测系统对运营动车的长期跟踪,实验采集得到的。数据集包括5个转向架关键部件:枕梁、轴箱、构架、弹簧筒和冷却风机。这些数据集包含了部件的横向振动有效值、横向振动最大值、垂向振动有效值、垂向横向振动最大值和高速列车运营速度、运营加速度、历史运营里程等信息。经过实验得知,转向架其余部件的实验结果对比分析过程与枕梁相类似,并且实验结果所采用的模型与枕梁是一致的,为了避免篇幅冗余,文中以枕梁数据为例进行实验分析。

动车转向架枕梁垂向振动中的训练集共有18370个训练样本,囊括了6种不同的历史运营里程的样本,其中11.7万公里的样本3433个,15万公里的样本2889个,20万公里的样本2947个,25万公里的样本2908个,30万公里的样本3092个,35万公里的样本3101个,每种运营里程下的数据记录着转向架枕梁运营过程中的振动有效值和振动最大值随运营速度、运营加速度的变化关系。

图1是11.7万公里运营里程下动车转向架枕梁垂向振动有效值和振动最大值的数据。

图2是在不同历史运营里程下列车从启动到加速到最大速度的转向架枕梁垂向振动最大值和振动有效值的平均值数据。

无论从单个历史运营里程还是从多个历史运营里程分析,该动车转向架振动信号随着运营里程和时间的增长,其振动值上升均具有一定的规律。由于转向架的振动的信号十分复杂,其信号频谱宽大,存在近似于白噪声的干扰噪声频谱几乎覆盖了整个振动信号阈值,很难确定信号的滤波频率,且转向架振动信号中夹杂着大量的随机噪声,无法严格按周期对数据进行采样。转向架振动激励源多,转向架的某个部件的振动信号数据往往会与大量的随机噪声数据相互影响,导致各种振动数据的信噪比特别低^[16]。为了提升原始

信号信噪比及样本数据曲线光滑度,需要对数据进行去噪处理。

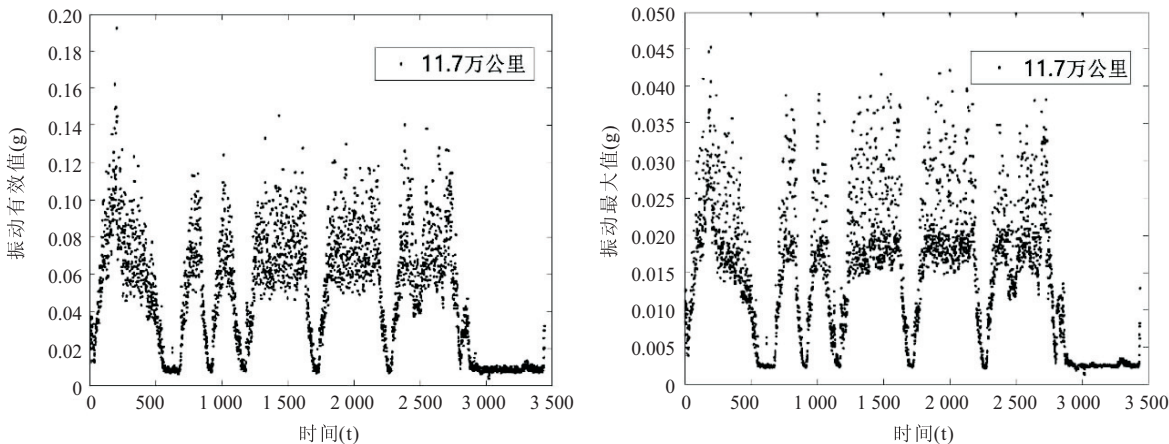


图 1 11.7 万公里运营里程振动散点图

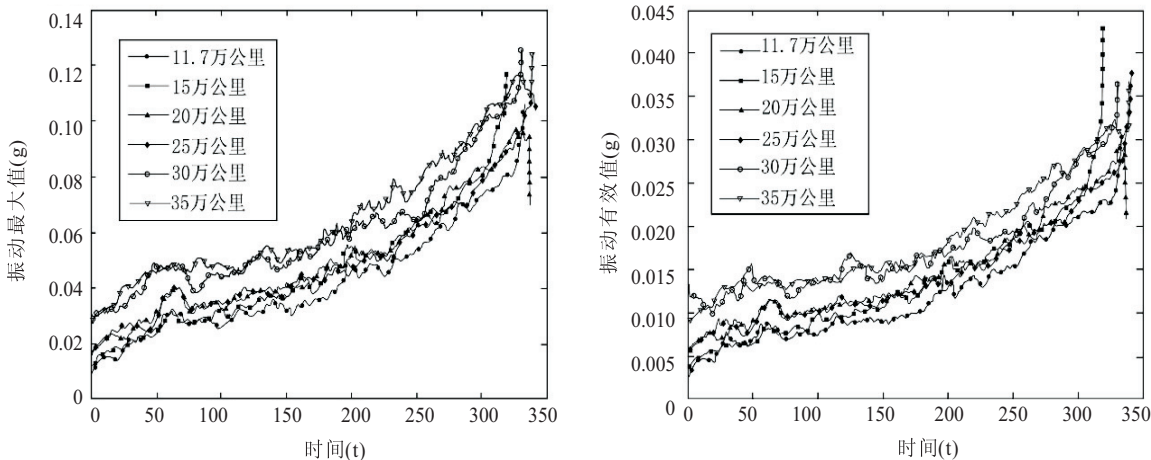


图 2 不同运营里程的振动平均值

2 基于小波变换的转向架数据去噪

数据的去噪滤波是信号数据处理领域中一个重要的组成部分^[17]。数据去噪的方法有很多种,但是有些去噪方法会过度的平滑,造成谱线中的峰值失真,从而丢失信号的重要信息成分。为了得到更高的信噪比,通过仿真实验,对几种常用的滤波去噪方法(包括巴特沃斯低通滤波器去噪、FIR 低通滤波器去噪、移动平均滤波去噪、小波去噪等)进行滤波前后的波形图片分析对比。从小波去噪的结果来看,由于小波去噪能自动适应时频信号分析的要求,得到波形不是很平滑,但是去噪效果明显,且上升沿和下降沿保持的非常高,对于转向架数据信号的演变规律的信息得到了较好的保存,相比之下,它很好地解决了巴特沃斯低通滤波器和 FIR 低通滤波器无法刻画信号的非平稳特性,信噪比低的缺点。因此,最终选定小波变换对数据信号进行去噪。

小波变换(wavelet transform, WT)是近年来十分热门的去噪分析方法,它是短时傅里叶变换局部化思

想的继承和发展,对含有等效噪声或人为噪声有很好的分割效果^[18],并且克服了窗口大小不随频率变化和缺乏离散正交性等缺点,是进行信号时频分析和数据处理的理想方法^[19]。它的主要优点是通过对时间频率的局部化分析,伸缩平移运算对信号逐步进行多尺度细化,最终达到高频处时间细分,低频处频率细分,从而充分突出问题某些方面的特征,自动适应时频信号分析的要求。

文中对 11.7 万公里运营里程列车从启动到加速到最大速度的转向架枕梁垂向振动有效值和振动最大值的数据进行小波去噪处理,所选数据的小波去噪滤波器分解层数为两层,选择阈值方法为 rigorsure 方式,选择软阈值去噪,噪声结构为 unscaled white noise,可以给出小波去噪滤波器的运行结果,如图 3 所示。

结果表明,去噪效果明显,能自动适应时频信号分析的要求,且上升沿和下降沿保持的非常高,没有过度平滑而造成振动信号线谱峰值失真,对于转向架数据信号振动规律的信息得到了较好的保存,可以得到较高的振动信号的信噪比。

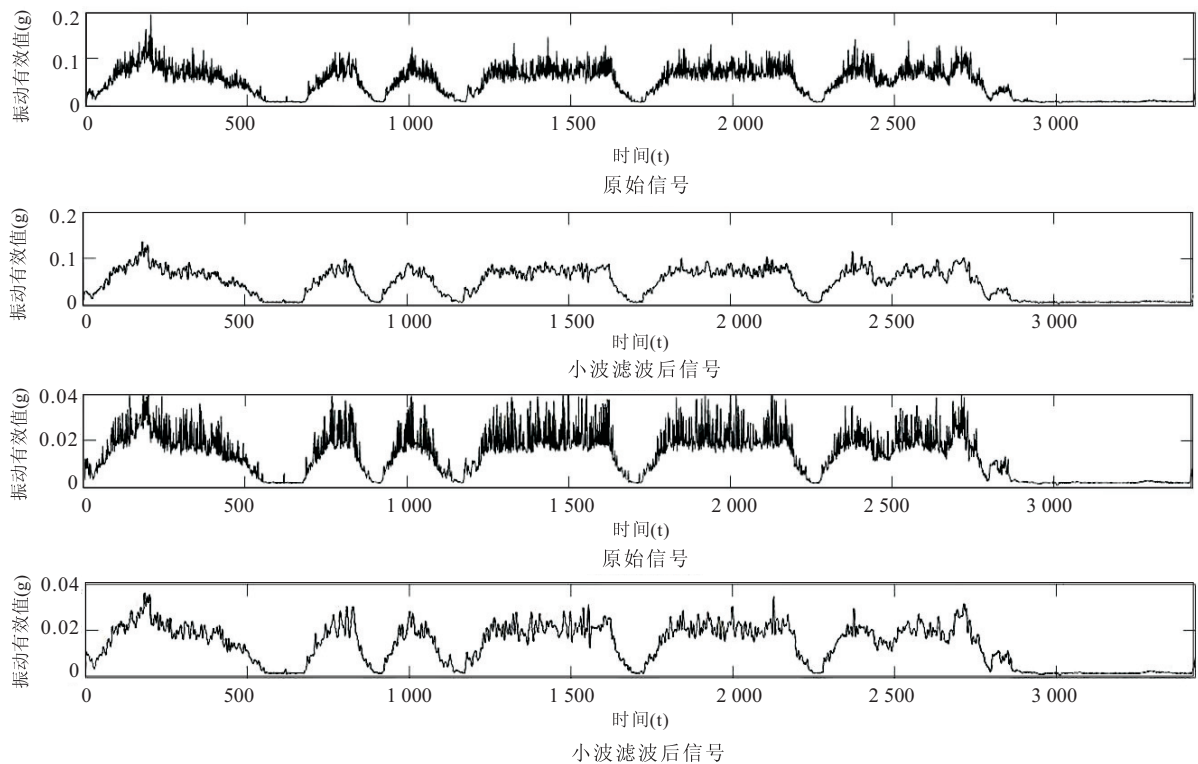


图 3 小波变换去噪结果

3 基于机器学习的转向架振动信号的比较分析

在转向架系统中,其参数包括速度、加速度和历史运营里程,其中速度的取值范围是[0,350],加速度的取值范围是[10,10],而历史运营里程的取值范围为[117 000,350 000],为避免因其输入参数数量纲差异过大,导致每个输入参数对输出影响的权重不同,在数据去噪的基础上,还需要对数据进行归一化处理,将输入参数的值域归一化为[0,1],使各个参数具有同等的权重。

在评价指标上,使用均方差、标准差和算法稳定性来评价转向架部件的服役性能模型。稳定性则是将同一个机器学习算法模型训练 M 次,得到 M 个模型,则假设每一次得到的测试集均方差为 R_1, R_2, \dots, R_M ,那么稳定性 S 为:

$$S = \frac{1}{N} \sum_{i=1}^M (R_i - \bar{R})^2$$

其中, \bar{R} 为多次均方差的平均数:

$$\bar{R} = \frac{1}{M} \sum_{i=1}^M R_i$$

对所有训练样本进行提取特征、数据去噪、数据归一化和数据划分后,针对动车转向架枕梁垂向振动数据集,文中使用多元线性回归、反向传播神经网络、径向基函数神经网络、回归决策树和随机森林算法建立

转向架振动信号监测模型。一共选用了 12 859 组枕梁振动数据作为训练集,另一组作为验证集,共 5 511 组数据,训练集和测试集的比例根据经验取为 7 : 3。分别比较各个模型的训练集、测试集的拟合误差来验证模型的正确性和稳定性,避免求得的模型存在过拟合的问题。

另外通过比较各个模型测试集误差模型的稳定性,验证模型是否存在不稳定的问题。

利用 5 511 组枕梁垂向振动数据来验证五个模型的有效性。验证集在五个模型上的拟合均方差、标准差和稳定性指标对比如表 1 所示。

表 1 测试集验证结果

模型	均方根 误差	均方差	稳定性	拟合效 果排名
多元线性回归模型	0.085	0.007 02	0	5
BP 神经网络模型	0.070	0.004 88	1.54e-08	3
RBF 神经网络模型	0.070	0.004 84	7.54e-09	2
梯度提升回归树模型	0.083	0.006 92	2.12e-09	4
随机森林回归模型	0.069	0.004 73	2.00e-09	1

从表 1 中的测试集误差可以直观看出,相对于其他所建立模型的精度,随机森林回归模型比两个神经网络提高了 1.4%,比梯度提升树模型提高了 16.9%,比多元线性回归模型 18.8%。而在模型的稳定性上,随机森林也是优于除多元线性回归之外的其他三个模型,多元线性回归由于每次训练的结果都是一样的,所

以其稳定性是最好的,但是其精度确是最差的。

通过对训练集和测试集的实验对比,选取了随机森林模型来预测动车转向架枕梁的垂向振动数据。在

11.7 万公里的固定里程下,列车的运行速度和运行加速度变化以及所得模型的振动最大值的结果如图 4 所示。

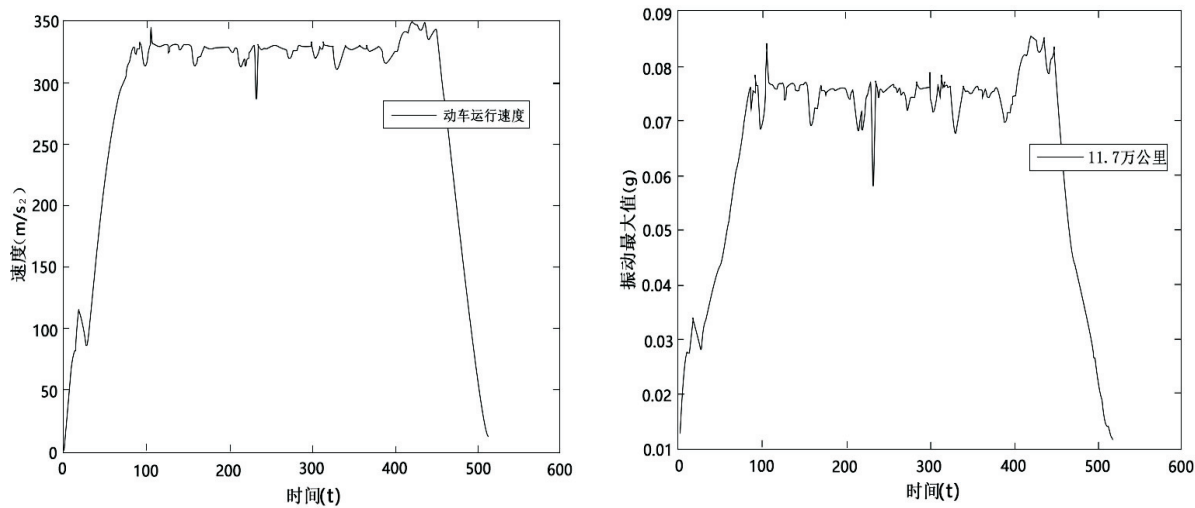


图 4 11.7 万公里下的运行速度及枕梁振动最大值预测结果

从列车枕梁的振动数据监测结果看出;当列车处于牵引阶段时,列车的振动最大值是随着运营速度的增加而增加的。当列车处于惰行阶段时,列车会有减速和加速状态,因此振动值是处于震荡变化的;当列车处于减速阶段,随着运营速度的降低,振动值也随之减小。总的来说,模型的预测监测结果与高速列车实际

运行情况是十分吻合的。
图 5 是根据所得模型,预测高速列车在 11.7 万公里、15 万公里、20 万公里、25 万公里、30 万公里、35 万公里、40 万公里的运营里程下,列车枕梁的振动最大值随列车运行状态的变换关系。

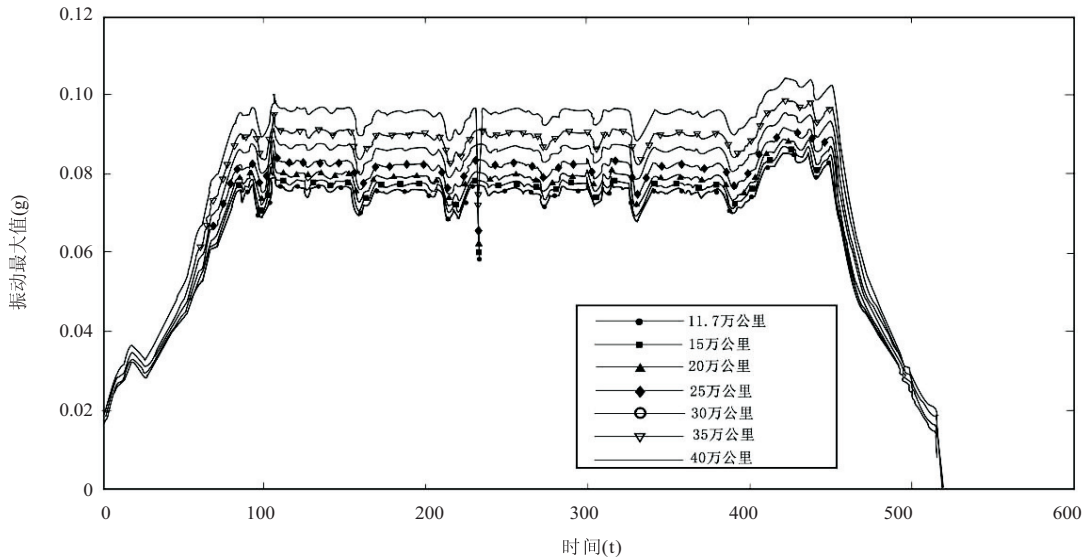


图 5 各里程枕梁振动最大值预测结果

从最终选取的模型来看,随机森林模型具备一定的鲁棒性能力,对数据的监测符合其实际的服役性能,能实时预测出在特定里程下,动车转向架枕梁垂向振动值随运行速度和加速度的变化规律。在特定里程下,当列车速度或者加速度超出限速范围,应用该模型可以监测到振动值超出安全阈值。其次,该模型还能监测转向架的振动趋势是否异常。当列车转向架服役

里程达到使用年限时,列车转向架的振动值也会超出安全阈值,该模型能制定更加合理的镟修周期,减少列车运营成本。基于机器学习高速列车转向架振动信号的研究对转向架的服役性能分析和安全运行具有重要的指导意义。

实验结果表明,针对动车转向架枕梁垂向振动数据集的监测模型,随机森林回归模型在回归性能上是

最佳的。随机森林算法相比于BP神经网络算法、RBF神经网络算法、多元线性回归算法,能更好地保证在转向架部件上复杂的振动信号的数据表现出良好的线性拟合性、数据集适应性以及抗噪性^[20-21]。在实现方面,随机森林比其他算法具有更良好的实现性。在处理复杂且多维的实际问题中,随机森林采用特征空间划分学习方式,适应高维数据的处理^[22-23],面对振动信号与动车运营速度、加速度的分析方面具有更好的优化能力。随机森林中各个决策树相互独立,进行学习和预测的效率更高。经过实验验证,在小波变换数据去噪的基础上,基于随机森林算法的实验模型能有效地反映高铁转向架部件服役性能演变规律和安全运行,具有较高的工程应用价值。

4 结束语

列车转向架振动信号是一个振动源复杂,噪音强,数量庞大的数据集,仅靠去噪处理,虽然噪声减少,数据显示效果较好,但是依然具有误差。因此,文中应用具有抗噪性的随机森林算法对数据进行拟合。经过实验表明,基于随机森林的高速列车转向架振动信号监测模型在高速列车安全提速和有效指导合理的维修保养应用中,具有一定的工程价值。为提高监测模型的精准性和有效性,在高速列车转向架振动的安全阈值设置方面需要做进一步的研究。

参考文献:

[1] 李 潭,谭献海. 高速列车转向架振动信号多重分形特征提取与故障诊断[J]. 工业控制计算机,2015,28(7):68-69.

[2] 张红军,黄成荣,陈国胜,等. 高速列车转向架技术[J]. 机车电传动,2004(3):1-4.

[3] 赵明花,梁树林,宋春元. 高速列车转向架技术研究[J]. 中国工程科学,2015,17(4):53-62.

[4] 段旺旺,金炜东. 基于AR功率谱的高速列车转向架故障信号分析[J]. 噪声与振动控制,2015,35(2):51-56.

[5] 秦 娜,金炜东,黄 进,等. 基于EEMD样本熵的高速列车转向架故障特征提取[J]. 西南交通大学学报,2014,49(1):27-32.

[6] 李海涛,王成国,许跃生,等. 基于EEMD的轨道一车辆系统垂向动力学的时频分析[J]. 中国铁道科学,2007,28(5):24-30.

[7] 侯亚东. 基于小波系数的高速列车监测数据特征分析[D]. 成都:西南交通大学,2014.

[8] 郑晓龙,杨贵营,黄 晗. 基于小波包特征熵的高速列车监测数据的特征分析[J]. 青岛科技大学学报:自然科学版,2015,36(2):213-217.

[9] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2015, 313(5786):504-507.

[10] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7):1527-1554.

[11] DAHL G E, YU Dong, DENG Li, et al. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech, & Language Processing, 2012, 20(1):30-42.

[12] 庞 荣,余志斌,熊维毅,等. 基于深度学习的高速列车转向架故障识别[J]. 铁道科学与工程学报,2015,12(6):1283-1288.

[13] LEI Yaguo, HE Zhengjia, ZI Yanyang. Application of a novel hybrid intelligent method to compound fault diagnosis of locomotive roller bearings[J]. Journal of Vibration & Acoustics, 2008, 130(3):569-583.

[14] 冯雪冰. 基于优化SVM的列车转向架故障挖掘模型仿真[J]. 计算机仿真,2015,32(10):182-185.

[15] 尹怀仙,王 凯,张铁柱,等. 基于PSO-BP神经网络的城轨列车转向架轮对轴箱故障预测[J]. 复杂系统与复杂性科学,2015,12(4):97-103.

[16] 章 欣. 基于振动信号的高速铁路伤损探测方法[D]. 哈尔滨:哈尔滨工业大学,2011.

[17] 徐科军. 信号分析与处理[M]. 北京:清华大学出版社,2006:206-265.

[18] 刘洲峰,徐庆伟,李春雷. 基于小波变换的图像分割研究[J]. 计算机应用与软件,2009,26(4):62-64.

[19] 魏 松,李 琦,赵仁才. 基于短时傅立叶变换语言信号分析算法[J]. 电子测量技术,2006,29(1):16-17.

[20] 吴俊杰,赵 鹏. 非线性噪声数据集上基于随机森林的空缺值填补算法[J]. 计算机应用与软件,2013,30(7):51-53.

[21] DAUBECHIES I. The wavelet transform, time-frequency localization and signal analysis[J]. Journal of Renewable & Sustainable Energy, 2015, 36(5):961-1005.

[22] 吕 兵,王华珍. 基于随机森林的高维数据可视化[J]. 计算机应用,2014,34(6):1613-1617.

[23] 刘 凯,郑山红,蒋 权,等. 基于随机森林的自适应特征选择算法[J]. 计算机技术与发展,2018,28(9):101-104, 111