

基于主成分分析和随机森林回归模型的 工艺装备轴承剩余寿命预测

耿明 张海沧 康丽齐 黄林 张旭 高雅

(中车长春轨道客车股份有限公司工程技术中心, 130062, 长春//第一作者, 正高级工程师)

摘要 为解决传统随机森林回归模型对工艺装备轴承剩余寿命预测准确率偏低的问题,提出一种将 PCA(主成分分析)和随机森林回归模型相结合的工艺装备轴承剩余寿命预测方法。首先,应用时域分析法对特征集进行提取,并和样本对应的剩余寿命标签共同创建并形成训练集;然后,利用 PCA 算法对训练集中特征实施降维处理;最后,建立随机森林回归模型,输出工艺装备轴承剩余寿命。研究结果表明:基于 PCA 算法和随机森林回归模型的预测方法将预测准确度提高了约 10%,证实了该方法的有效性和准确性。

关键词 轨道交通车辆;工艺装备轴承;剩余寿命预测;主成分分析;随机森林回归模型

中图分类号 TH133.3

DOI:10.16037/j.1007-869x.2023.04.004

Prediction of Residual Life of Jig Rolling Bearing Based on PCA and Random Forest Model

GENG Ming, ZHANG Haicang, KANG Liqi, HUANG Lin, ZHANG Xu, GAO Ya

Abstract To solve the low prediction accuracy problem of the conventional random forest regression model for the residual life of jig rolling bearings, a method for predicting the residual life of jig rolling bearings combining PCA (principal component analysis) and random forest regression model is proposed. First, the feature set is drawn by time-domain analysis method, and jointly forms a training set with residual life label corresponding to samples; then, PCA algorithm is used to implement dimension reduction processing of the training set features; finally, a random forest regression model is established to predict the residual life of jig rolling bearings. Research results show that: the method based on the combination of PCA algorithm and random forest regression model improves the model prediction accuracy by about 10%, which confirms the effectiveness and accuracy of the method.

Key words rail transit vehicle; jig rolling bearing; prediction of residual life; PCA; random forest regression model

Author's address Engineering Technology Center, CRRC Changchun Railway Vehicles Co., Ltd., 130062, Changchun, China

工艺装备是制造轨道交通车辆车体的必备装备。轴承是工艺装备中承载焊接件翻转的最重要旋转支撑部件。当焊接件翻转时,一旦工艺装备轴承(以下简称“工装轴承”)发生故障,将导致工艺装备不工作,直接影响焊接工作的实施,特别是对于大型焊接件翻转工艺装备,还会存在较大的施工安全风险。因此,对工装轴承进行可靠性及剩余使用寿命预测具有现实意义。

轴承的剩余使用寿命,指的是轴承完全失效时对应时间点与最初检测时间点之间的差值。近年来,针对轴承剩余寿命预测,国内外学者开展了大量研究,提出了多维缩放、多尺度卷积神经网络等各种诊断方法^[1-6],从不同方面解决轴承寿命动态预测、故障诊断等问题。

在此基础上,本文提出基于 PCA(主成分分析)和随机森林回归模型的工装轴承剩余寿命预测方法,通过 PCA 算法对初始训练数据进行降维,生成 10 个新特征值,依次将新特征值与原时域特征值组建成新训练矩阵,输入到随机森林回归模型,进而较准确地预测工装轴承剩余寿命。

1 工装轴承寿命预测原理

1.1 PCA 原理

PCA 是一种无监督学习的算法,可用于数据降维,即将原数据集特征维度由 n 维降至 k 维^[7]。算法步骤如下:

1) 将原始特征矩阵进行以 0 为中心的去均值化操作,即数据标准化。设样本标准化后的特征矩阵为 \mathbf{X} ,其中数据集样本个数为 m ,工装轴承特征个

数为 n ,则有:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

2) 求 \mathbf{X} 的协方差矩阵 \mathbf{C}_{on} ,计算式为:

$$\mathbf{C}_{\text{on}} = \frac{1}{m} \mathbf{X}^T \mathbf{X} \quad (2)$$

3) 将协方差矩阵 \mathbf{C}_{on} 进行矩阵分解,求解出特征值 λ 及相应的特征向量,取数值较大的前 k 维特征对应的特征向量作为数据集的主成分。

1.2 随机森林回归原理

随机森林是一种以决策树作为基学习器的集成学习方法^[8]。随机森林回归模型预测过程:采集原始振动信号→生成时域特征信号→PCA 降维算法→形成组件训练集和测试集→训练随机森林回归器→输出轴承剩余寿命。

采用随机森林回归算法来预测工装轴承寿命,本质是通过工装轴承的振动信号去拟合其试验下的真实寿命。由此建立的基于随机森林回归工装轴承寿命预测数学模型为:

$$L_{\text{RF}} = \{N_{\text{tr}}, M_{\text{tr}}\} \quad (3)$$

$$M_{\text{tr}} = \{x_p, l_p\} \quad p = 1, 2, \cdots, m \quad (4)$$

式中:

L_{RF} ——应用随机森林回归算法预测的建模参数集;

N_{tr} ——模型基学习器个数;

M_{tr} ——建模特征数据集;

x_p ——第 p 组训练样本集;

l_p ——第 p 组样本集对应的工装轴承剩余寿命。

1.3 工装轴承寿命预测的思路及步骤

工装轴承的失效形式主要有振动导致的疲劳失效和高温烧伤点蚀失效。研究表明,工装轴承出现故障时,其振动信号会体现故障信息,在一定程度上反映工装轴承运行状态^[9]。由此,本文基于工装轴承振动数据,结合 PCA 和随机森林回归模型,实现对工装轴承剩余寿命的预测。主要步骤如下:

1) 根据轴承试验数据集,获取轴承原始的振动信号数据和剩余寿命数据。

2) 采用时域分析法从轴承振动信号中提取时域特征值,组成特征矩阵 \mathbf{X} ;将轴承剩余寿命作为样本标签 \mathbf{Y} ,将 \mathbf{X} 和 \mathbf{Y} 组合成训练集 \mathbf{T} 。

3) 利用 PCA 算法对时域特征降维,并与降维

前训练集 \mathbf{T} 进行融合,形成训练集 \mathbf{T}_1 ,按 30% 比例将 \mathbf{T}_1 划分为训练样本和测试样本。

4) 使用随机森林回归器对 \mathbf{T}_1 进行训练,得到 1 个可用来预测轴承剩余寿命的回归器。

5) 调整随机森林回归模型的可控参数,将调整后表现最佳的模型作为最终回归器,应用于轴承剩余寿命预测。

2 工装轴承寿命预测方法

2.1 工装轴承试验数据集及试验台装置

本文数据引自西安交通大学公开的 XJTU-SY 轴承试验数据集^[10],其中部分试验数据如表 1 所示。试验工装轴承型号为 LDK UER204,共 15 套该型轴承参与试验。该数据集有 3 种加载工况,不同工况的测试条件如表 2 所示。由于工装轴承在实际工况下只承受径向力,因此,试验设定工装轴承加载方向为径向液压加载,可调节测试条件为转速和径向力。采样频率为 25.6 kHz,记录时间间隔为 1 min,单次采样时长为 1.28 s。

表 1 XJTU-SY 轴承试验数据(部分)
Tab.1 XJTU-SY bearing test data (partial)

轴承编号	样本总数/个	实际寿命/min	失效位置
轴承 1	123	123	外圈
轴承 2	161	161	外圈
轴承 3	158	158	外圈
轴承 4	122	122	保持架
轴承 5	52	52	内外圈

表 2 不同工况的测试条件
Tab.2 Test condition under different working conditions

工况	转速/(r/min)	径向力/kN
工况 1	2 100	12
工况 2	2 250	11
工况 3	2 400	10

2.2 提取工装轴承的振动信号

工况 1 下试验工装轴承 1 原始振动信号如图 1 所示。由图 1 可见:当试验时间为 0 ~ 73 min 时,工装轴承振动信号波动幅度较小,说明工装轴承处于性能正常期;当试验时间为 73 ~ 119 min 时,工装轴承出现较大噪声,波动幅度较大,说明工装轴承出现异常,进入了性能退化期;当试验时间为 119 ~ 122 min 时,工装轴承振幅出现巨大波动,可确定工装轴承进入了失效期。

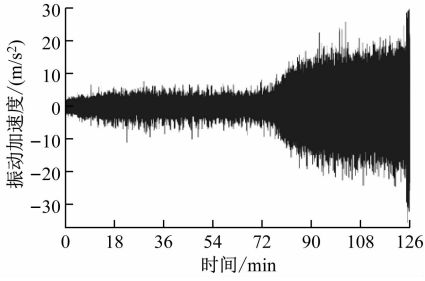


图1 工况1下试验工装轴承1的原始振动信号

Fig.1 Original vibration signal of test jig rolling bearing 1 under working condition 1

为了更准确地分析工装轴承的运行状态,需提取工装轴承的一些时域参数性能指标,其中,有量纲的时域特征指标包括加速度有效值 a_{rms} 、峭度值 a_k 、平均值 a_a 、标准差 a_{sd} 和极差 a_e ,量纲为一的指标包括峰度 a_f 、歪度 a_p 和脉冲值 a_i 。各指标为:

$$a_{\text{rms}} = \sqrt{\frac{1}{m} \sum_{p=1}^m x_p^2} \quad (5)$$

$$a_k = \frac{\beta}{a_{\text{rms}}^4} \quad (6)$$

$$a_a = \frac{1}{n} \sum_{q=0}^{n-1} Z_{pq} \quad (7)$$

$$a_{\text{sd}} = \sqrt{\frac{1}{n} \sum_{q=0}^{n-1} (Z_{pq} - \bar{Z}_p)^2} \quad (8)$$

$$a_e = \max x_p - \min x_p \quad (9)$$

$$a_f = E \left[\left(\frac{Z_{pq} - \mu}{\sigma} \right)^4 \right] \quad (10)$$

$$a_p = E \left[\left(\frac{Z_{pq} - \mu}{\sigma} \right)^3 \right] \quad (11)$$

$$a_i = \frac{a_{\text{rms}}}{|a_a|} \quad (12)$$

$$\beta = \frac{1}{m} \sum_{p=1}^m x_p^4 \quad (13)$$

式中:

p ——样本数, $p = 1, 2, \dots, m$;

\bar{Z}_p ——第 p 组样本集的均值;

Z_{pq} ——第 p 组样本集的第 q 条样本数据;

μ ——样本总体均值;

σ ——样本总体方差;

E ——期望值。

基于工况1下的轴承1数据集,对工装轴承进行全寿命振动试验,得到各性能参数指标的时域曲线如图2所示。

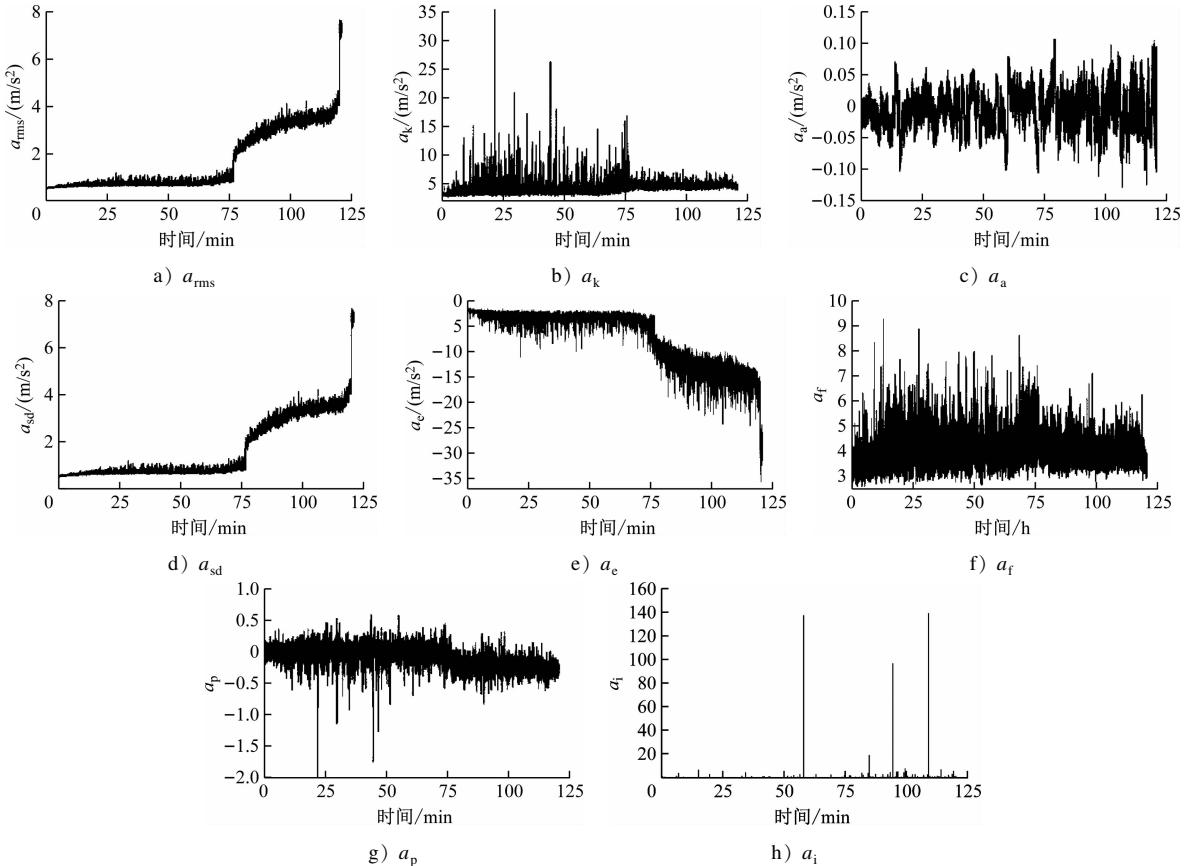


图2 各性能参数指标的时域曲线

Fig.2 Time-domain curve of each performance parameter index

3 试验验证

3.1 基于 PCA 算法的数据降维处理

利用 PCA 算法对 LDK UER204 型轴承的特征数据集进行降维,将时域特征原有的 11 个维度降至 10 个维度。经 PCA 算法降维后,特征参数没有实

际物理意义。将降维后的特征参数依次定义为 P_1, P_2, \cdots, P_{10} ,表 3 为 PCA 降维后的测试特征集。将表 3 所示特征集与降维前特征集进行合并,并与样本对应的剩余寿命组合成建模数据集。表 4 为建模所用数据集 T_1 ,其代表 PCA 降维后特征的示意性指标。

表 3 PCA 降维后的测试特征集
Tab.3 PCA test feature set after dimension reduction

样本序号	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
1	-336.32	-10.13	2.09	-0.59	0.31	0.07	0.01	-0.05	-0.02	-4.18
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
3 579	-422.23	62.83	16.36	5.08	0.39	2.07	0.08	0.13	-0.03	-1.46
3 580	-413.95	56.65	24.39	7.19	1.18	0.76	-0.04	-0.15	0.01	6.12

表 4 数据集 T_1
Tab.4 Dataset T_1

样本序号	$a_p/$ (m/s^2)	$a_r/$ (m/s^2)	$a_e/$ (m/s^2)	加速度方差/ (m/s^2) ²	加速度最小值/ (m/s^2)	加速度最大值/ (m/s^2)
1	0.02	3.29	-1.79	0.31	-1.88	-2.03
2	0.26	3.36	-1.78	0.30	-1.87	-1.02
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
3 579	-0.26	3.89	-31.32	53.53	-31.35	2.84
3 580	-0.17	2.87	-25.75	56.98	-25.74	3.82

3.2 工装轴承寿命的预测

选用工况 1 下的工装轴承试验数据作为建模数据集,基于 PCA 及随机森林回归模型对工装轴承寿命进行预测。工装轴承数据样本总数为 3 580 个,按 30% 比例将数据划分成 2 506 个训练样本和 1 074 个测试样本。其中,训练样本用于建立模型,测试样本用于验证模型预测的精确度。

图 3 为基于 PCA 和随机森林回归模型在测试集上得到的工装轴承寿命预测效果。由图 3 可以看出:基于 PCA 和随机森林回归模型的预测方法效果

较好,平均预测准确度达到了 99.3%。

4 结语

本文利用时域分析法提取失效工装轴承全寿命周期的时域振动信号,并将其作为建立机器学习模型的原始特征集,可以有效地建立工装轴承振动特征与剩余寿命的对应关系。

利用 PCA 算法对原特征集进行降维,并将降维后的数据与原特征集进行合并,得到新的建模数据集包含轴承信息更广,使得工装轴承剩余寿命预测的结果更准确。

本文所提基于 PCA 和随机森林回归模型的预测方法将工装轴承剩余寿命预测准确率提高到了 99.3%,验证了该方法的有效性和准确性。该方法可用于工装轴承剩余寿命预测的在线监测。

参考文献

[1] 王建国,李健,万旭东. 基于奇异值分解和局域均值分解的滚动轴承故障特征提取方法[J]. 机械工程学报, 2015, 51 (3): 104.
WANG Jianguo, LI Jian, WAN Xudong. Fault feature extraction

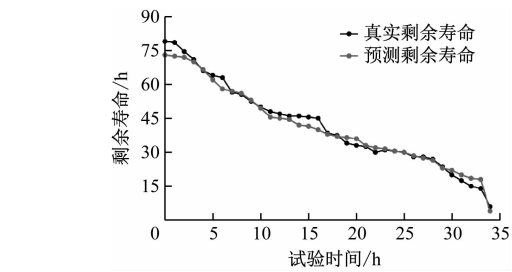


图 3 基于 PCA 和随机森林回归模型的工装轴承寿命预测
Fig.3 Prediction of jig rolling bearing lifespan based on PCA and random forest regression model

- method of rolling bearings based on singular value decomposition and local mean decomposition[J]. Journal of Mechanical Engineering, 2015, 51(3): 104.
- [2] 焦晓璇, 景博, 李娟, 等. 基于 ADEGWO-SVM 的机载燃油泵寿命预测研究[J]. 仪器仪表学报, 2018, 39(8): 43.
- JIAO Xiaoxuan, JING Bo, LI Juan, et al. Research on remaining useful life prediction of fuel pump based on adaptive differential evaluation grey wolf optimizer-support vector machine[J]. Chinese Journal of Scientific Instrument, 2018, 39(8): 43.
- [3] 张西宁, 张雯雯, 周融通, 等. 基于多维缩放和随机森林的轴承故障诊断方法[J]. 西安交通大学学报, 2019, 53(8): 1.
- ZHANG Xining, ZHANG Wenwen, ZHOU Rongtong, et al. Bearing fault diagnosis method based on multiple dimensional scaling and random forest[J]. Journal of Xi'an Jiaotong University, 2019, 53(8): 1.
- [4] 曲建岭, 余路, 袁涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. 仪器仪表学报, 2018, 39(7): 134.
- QU Jianling, YU Lu, YUAN Tao, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network[J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134.
- [5] 安宗文, 胡敏, 刘波. 随机载荷作用下风电齿轮箱轴承疲劳寿命预测方法[J]. 兰州理工大学学报, 2016, 42(1): 35.
- AN Zongwen, HU Min, LIU Bo. Fatigue life prediction method of bearing in wind turbine gearbox under random loads[J]. Journal of Lanzhou University of Technology, 2016, 42(1): 35.
- [6] SUN W Q, DENG A D, DENG M Q, et al. Multi-view feature fusion for rolling bearing fault diagnosis using random forest and autoencoder[J]. Journal of Southeast University (English Edition), 2019, 35(3): 302.
- [7] 许凡, 方彦军, 张荣. 基于 EEMD 模糊熵的 PCA-GG 滚动轴承聚类故障诊断[J]. 计算机集成制造系统, 2016, 22(11): 2631.
- XU Fan, FANG Yanjun, ZHANG Rong, et al. PCA-GG rolling bearing clustering fault diagnosis based on EEMD fuzzy entropy[J]. Computer Integrated Manufacturing Systems, 2016, 22(11): 2631.
- [8] ZHANG X T, TANG L W, WANG P, et al. Fault feature enhancement method based on multiscale orthogonal PCA-LPP manifold learning algorithm[J]. Journal of Vibration and Shock, 2015, 34(13): 66.
- [9] 李兆飞, 任小洪, 谭飞, 等. 滚动轴承振动信号的非线性分形特性研究[J]. 轴承, 2015(9): 53.
- LI Zhao-fei, REN Xiao-hong, TAN Fei, et al. Study on nonlinear fractal characteristics of vibration signals in rolling bearings[J]. Bearing, 2015(9): 53.
- [10] 雷亚国, 韩天宇, 王彪, 等. XJTU-SY 滚动轴承加速寿命试验数据集解读[J]. 机械工程学报, 2019, 55(16): 1.
- LEI Yaguo, HAN Tianyu, WANG Biao, et al. XJTU-SY rolling element bearing accelerated life test datasets: a tutorial[J]. Journal of Mechanical Engineering, 2019, 55(16): 1.

(收稿日期:2022-08-01)

(上接第 11 页)

- [2] 张喜燕, 赵永庆, 白晨光. 钛合金及应用[M]. 北京: 化学工业出版社, 2005: 287.
- ZHANG Xiyen, ZHAO Yongqing, BAI Chenguang. Titanium alloy and its application[M]. Beijing: Chemical Industry Press, 2005: 287.
- [3] 里维. 尤利格腐蚀手册[M]. 杨武, 译. 北京: 化学工业出版社, 2005: 380.
- REVIE R W. Uhlig's corrosion handbook[M]. YANG Wu, Trans. Beijing: Chemical Industry Press, 2005: 380.
- [4] 杨勇进, 张晓云, 刘明辉. TC4-DT 钛合金与异种材料接触腐蚀与防护研究[J]. 材料工程, 2012, 40(12): 55.
- YANG Yongjin, ZHANG Xiaoyun, LIU Minghui. Contact corrosion and protection between TC4-DT titanium alloy and dissimilar materials[J]. Journal of Materials Engineering, 2012, 40(12): 55.
- [5] 刘凤岭, 李金桂, 冯自修. 钛合金表面技术的进展[J]. 腐蚀与防护, 2001, 22(2): 54.
- LIU Fengling, LI Jingui, FENG Zixiu. Evolution of surface technologies for titanium alloys[J]. Corrosion & Protection, 2001, 22(2): 54.
- [6] 张晓云, 汤智慧, 孙志华, 等. 钛合金的电偶腐蚀与防护[J]. 材料工程, 2010, 38(11): 74.
- ZHANG Xiaoyun, TANG Zhihui, SUN Zhihua, et al. Galvanic corrosion and protection between titanium alloy and other materials[J]. Journal of Materials Engineering, 2010, 38(11): 74.
- [7] 曹文健, 汤智慧, 原玲, 等. 钛合金紧固件用铝涂层抗电偶腐蚀行为研究[J]. 装备环境工程, 2016, 13(1): 116.
- CAO Wenjian, TANG Zhihui, YUAN Ling, et al. Behaviour of galvanic corrosion resistance of aluminum coating on titanium alloy fastener[J]. Equipment Environmental Engineering, 2016, 13(1): 116.

(收稿日期:2022-07-31)