

DOI:10.19344/j. cnki. issn1671-5276.2024.03.031

高斯过程回归在轴承健康状态预测中的应用

马浩^{1,2}, 石永进^{1,2}, 李伟^{1,2}

(1. 西安中车永电捷力风能有限公司, 陕西 西安 710086; 2. 轨道交通牵引电机山西省重点实验室, 山西 永济 044502)

摘要: 轴承作为列车牵引电机的重要零部件, 准确评估其后续健康状态对列车的安全运行至关重要。基于轴承台架试验及实车数据, 将主成分分析 (PCA) 和高斯过程回归 (GPR) 应用于轴承的健康状态评估及预测中。通过对试验数据进行特征提取, 获得能够表征轴承衰退规律的特征数据, 应用 PCA 将提取的数据进行降维并建立轴承状态评估数据, 应用 GPR 对评估数据进行学习和预测。通过预测值与真实值的对比验证, GPR 可以实现轴承的健康状态预测, 并在低采样率下保持了较高的准确率。

关键词: 轴承状态预测; 高斯过程回归; 主成分分析; 牵引电机

中图分类号: TH133.33; U279.3 **文献标志码:** A **文章编号:** 1671-5276(2024)03-0146-05

Application of Gaussian Process Regression in Prediction of Bearing Health

MA Hao^{1,2}, SHI Yongjin^{1,2}, LI Wei^{1,2}

(1. CRRC Xi'an Yongdianjeli Wind Energy Co., Ltd., Xi'an 710086, China;

2. Shanxi Key Laboratory of Traction Motor for Rail Transit, Yongji 044502, China)

Abstract: Bearing is an important part of train traction motor and accurate evaluation of its subsequent health status is very important for train safe operation. Based on the bearing bench test and real vehicle data, principal component analysis (PCA) and Gaussian process regression (GPR) are applied to assess and predict bearing health status. Features are extracted from the experimental data, obtaining the characteristic data that can characterize the law of bearing degradation. PCA is used to reduce the dimension of extracted data and establish the bearing state assessment data, and GPR is applied to learn and predict the assessment data. Through the comparison of the predicted value and the true value, GPR can predict the health of bearing and maintain a high accuracy rate at a low sampling rate.

Keywords: bearing condition prediction; Gaussian process regression; principal component analysis; traction motor

0 引言

随着大数据与机器学习算法的发展, 越来越多的业主将产品的全寿命状态监测纳入车辆的维保中^[1], 同时对基于数据驱动的产品状态评估与寿命预测提出了更高的需求。而轴承作为牵引电机的核心部件, 其当前健康状态及未来状态预测对列车安全运行具有重要意义。目前在轨道交通领域应用较多的故障诊断系统来自于唐智公司基于共振解调技术开发的自诊断系统^[2]。该系统具备轴承状态判断、故障部位报警的功能, 但缺少业主更加关心的轴承状态预测功能, 无法为业主延长修程修制提供参考。本研究基于科技部国家重点研发计划课题“高性能电机绝缘轴承运维监测技术”项目需求, 对目前牵引电机最常用的圆柱滚子轴承和深沟球轴承开展技术研究, 通过 PCA 方

法对轴承的特征指标进行融合并建立健康度指标; 根据健康度指标评估轴承状态并通过高斯过程回归 (GPR) 算法完成轴承的健康状态预测; 使用台架试验数据及实车数据对该方法的准确性进行验证。

1 轴承健康度指标建立

当轴承出现故障时, 传感器采集到的振动信号将出现一定的变化, 该变化将同时反映到相关的特征指标中。通过提取振动信号的特征指标, 根据轴承衰退规律选择合适的特征指标, 使用主成分分析 (PCA) 完成指标融合并根据轴承全寿命周期数据, 建立合理的轴承健康指标。

1.1 振动信号特征指标提取

振动信号特征指标包含了时域指标、频域指标及无量纲指标。时域指标中, 有效值、方差的稳

基金项目: 国家重点研发计划项目 (2020YFB2007905)

第一作者简介: 马浩 (1989—), 男, 山西永济人, 工程师, 硕士, 研究方向为轴承应用技术研究, mahaochina@126.com。

定性较好,对冲击不敏感;峰值、峰峰值、峭度及偏斜度对冲击的敏感性较好。频域指标中,谱均值和谱有效值用来衡量幅值的平均水平,稳定性好;谱中心用来衡量频谱中心的迁移情况;谱方差用来衡量幅值的波动性,部分特征指标如表1所示。

表1 特征指标

特征指标	表达式	说明
平均值	$T_0 = \frac{1}{N} \sum_{i=1}^N u(i)$	$u(i)$ 表示时域信号样本点; N 表示样本点总数
峭度	$T_1 = \frac{1}{N} \sum_{i=1}^N (u(i) - T_0)^4$	
方差	$T_2 = \frac{1}{N-1} \sum_{i=1}^N (u(i) - T_0)^2$	
谱均值	$T_3 = \frac{2}{N} \sum_{i=1}^{N/2} s(j)$	$s(j)$ 表示经过傅里叶变换后的频域序列
谱有效值	$T_4 = \sqrt{\frac{2}{N} \sum_{i=1}^{N/2} s(j)^2}$	

1.2 基于PCA的特征指标融合

通过对振动信号的特征提取,完成从低信息量到高信息量的转变,但各个特征指标之间仍然存在冗余,过高的冗余度会降低模型精度,因此还需使用PCA对提取的特征信息做进一步融合。

PCA本质是一种基于统计特征的多维正交线性变换,其基本流程如下^[3]。

设轴承振动信号的特征向量矩阵为 $\mathbf{Z}$,其列向量 $\mathbf{Z}_k = (z_{1k}, z_{2k}, \dots, z_{nk})^T$ 为 n 维特征向量, $\mathbf{Z}_k$ 的协方差矩阵为

$$\mathbf{R}_z = \sum_{k=1}^M (\mathbf{Z}_k - \bar{\mathbf{z}})(\mathbf{Z}_k - \bar{\mathbf{z}})^T \quad (1)$$

式中: M 为训练样本的数量; $\bar{\mathbf{z}}$ 为各样本向量的均值。

求解 $\mathbf{R}_z$ 的全部特征值 $\lambda_i (i=1, 2, \dots, m)$ 和特征向量 $\mathbf{v}_i$ 。将特征值 λ_i 按照从大到小的顺序排列: $\lambda_1 > \lambda_2 > \dots > \lambda_m$,则对应的特征向量为 $\mathbf{v}_i (i=1, 2, \dots, m)$ 。样本向量 $\mathbf{z}_i$ 投影到特征向量 $\mathbf{v}_i$ 得到该方向上对应的主成分为

$$w_{ij} = \mathbf{v}_i^T (\mathbf{z}_j - \bar{\mathbf{z}}), \quad i=1, 2, \dots, m \quad (2)$$

特征向量所对应的特征值越大,它在重构时的贡献也越大,前 p 个主成分的累计方差贡献率为

$$h(p) = \sum_{i=1}^p \lambda_i / \sum_{k=1}^m \lambda_k \quad (3)$$

1.3 轴承健康度指标建立

轴承的健康度指标是一种自比较的评估方

法,该方法剔除了因装配、外在环境对状态评估的影响。以融合后指标的最大值与最小值作为参照,计算其相对距离并归一化至 $[0, 1]$ 范围内,计算公式如下:

$$C_v = 1 - d_2 / d_1 \quad (4)$$

式中: C_v 为轴承健康度指标; d_1 为融合指标最大值与最小值之间的欧氏距离; d_2 为各时刻下融合指标与最小值之间的欧氏距离,均为无量纲指标。

2 轴承健康状态预测

根据建立的轴承健康度指标,使用高斯回归算法对轴承的健康度进行预测以评价未来一段时间内的健康度能否满足运行要求。

高斯过程回归是指对输入和输出变量均为连续变量的情况进行预测^[4],给定训练样本集 $D = \{(x_i, y_i) | i=1, 2, \dots, n\}$,其中,样本集中的 x_i 为输入量且有 $x_i \in R^d$, y_i 为输出量且有 $y_i \in R$ 。随机高斯过程产生训练样本集,在输入 x_i 时对应随机变量 $f(x_i)$,考虑含噪声 ε_i 产生的 y_i 即为

$$y_i = f(x_i + \varepsilon_i) \quad (5)$$

这里假设均值函数为 0 , ε_i 为独立同分布,则有 $\varepsilon_i \sim N(0, \sigma_n^2)$,在训练集 y_i 和测试点 x_* 的 y_* 的后验分布是:

$$P(y_* | X, y_i, x_*) \sim N(u_*, \Sigma_*) \quad (6)$$

$$u_* = \mathbf{k}(x_*, X) [\mathbf{k}(X, X) + \sigma_n^2 \mathbf{I}_n]^{-1} y_i \quad (7)$$

$$\Sigma_* = \mathbf{k}(x_*, x_*) - \mathbf{k}(x_*, X) \times [\mathbf{k}(X, X) + \sigma_n^2 \mathbf{I}_n]^{-1} \mathbf{k}(X, x_*) \quad (8)$$

式中: u_* 为均值; Σ_* 为方差; $\mathbf{I}_n$ 为 n 维单位矩阵; $\mathbf{k}(X, X)$ 为对称的协方差矩阵。

采用平方指数协方差函数作为高斯过程回归的核函数:

$$k(x_p, x_q) = \sigma_f^2 \exp \left[-\frac{1}{2} (x_p - x_q) \times M^{-1} (x_p - x_q) \right] \quad (9)$$

$\theta = \{M, \sigma_n^2, \sigma_f^2\}$ 为超参数,采用最大似然估计函数对其进行优化^[5],公式如下:

$$L = \log P(\mathbf{y} | \mathbf{x}, \theta) = \frac{1}{2} \log [\det(\mathbf{k} + \sigma_n^2 \mathbf{I}_n)] -$$

$$\frac{1}{2} \mathbf{y}^T [\mathbf{k} + \sigma_n^2 \mathbf{I}_n]^{-1} \mathbf{y} - \frac{N}{2} \log 2\pi \quad (10)$$

式中 $\mathbf{y}$ 为输出量 y_i 的集合,即有 $\mathbf{y} = y_1, y_2, \dots, y_i$ 。

针对超参数 θ 对似然函数 L 求导,最终计算得到最优超参数 $M, \sigma_n^2, \sigma_f^2$,代入式(7)、式(8)中得到预测的均值和方差,即为GPR模型的预测值及置信度。方差越小,则预测结果越好。

最后,采用方均根误差(root mean square

error, RMSE) 来对算法的有效性进行评估^[6], 公式如下:

$$R_{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta_i^2} \quad (11)$$

式中: n 为预测样本数; Δ_i 为各预测值与真实值之间的误差。

3 台架试验验证

3.1 数据来源及说明

本文采用在杭州某实验中心进行的轴承衰退试验数据进行验证, 轴承型号分别为 NU214 圆柱滚子轴承及 6311 深沟球轴承。试验转速 2 500 r/min, 轴承承受径向载荷 28 kN。试验数采设备通过振动传感器实时采集被试轴承的振动数据。采样频率 25 600 Hz, 采样间隔 5 min, 单次采样时长 4 s。

3.2 轴承健康指标建立

根据前文的方法对轴承的试验数据进行特征指标的提取, 部分结果如图 1—图 3 所示。

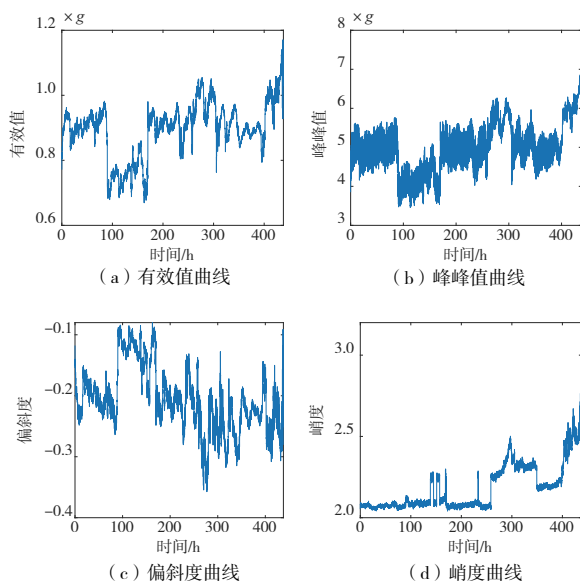


图 1 轴承的时域指标

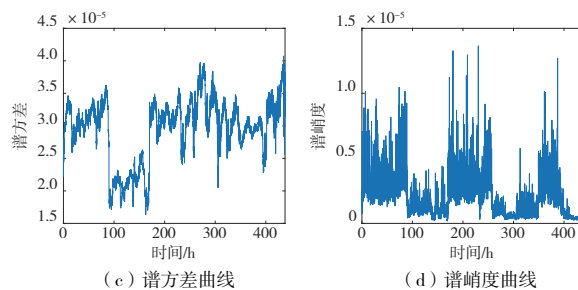
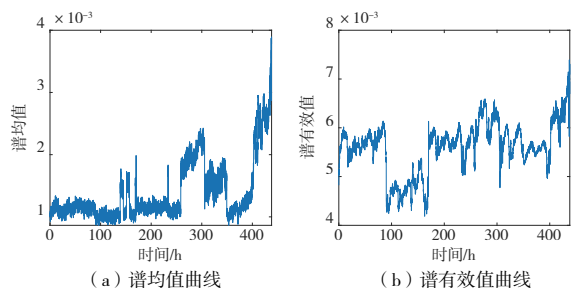


图 2 轴承的频域指标

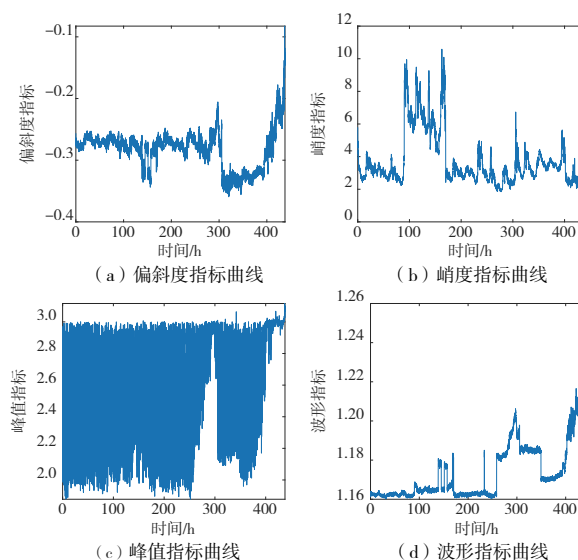


图 3 轴承的无量纲指标

从图 1—图 3 中可知, 峰峰值及峰值指标的信息嘈杂, 且未见明显的衰退变化; 偏斜度、谱峭度、峭度指标未见明显衰退变化。将上述指标剔除后进行 PCA 特征融合, 选择贡献率大于 95% 的数据建立轴承健康度指标, 如图 4 所示。

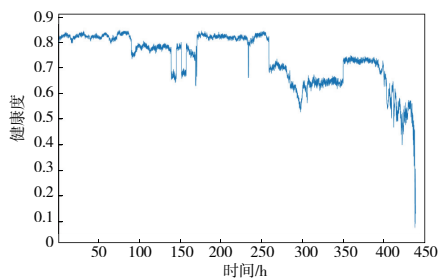


图 4 轴承的健康度指标

从图 4 可以看出, 通过轴承的健康度指标可以更加直观地了解轴承状态的衰退变化情况, 在 260 h 前后时间段内有一个健康度指标降低的情况, 因此可将 260 h 前的阶段定义为轴承正常运行阶段, 将 260 h~400 h 之间的阶段定义为轴承的劣

化阶段。在劣化阶段,轴承的健康度指标有一个先下降,后上升的过程。这是因为当轴承出现初始故障时,连续的运转会导致轴承出现一定的剥落或裂纹,此时轴承的健康度随之下降;随着轴承的运行,损伤扩展到更广的区域后,健康度会有一定程度的上升,这种现象被称为“愈合”现象^[7]。随着轴承运行,轴承的健康度会在较短的时间内快速衰退至 0.2,达到完全失效状态。

3.3 轴承健康状态预测

根据获得的轴承健康度衰退变化规律,选择平方指数协方差函数构建 GPR 模型,对轴承未来一段时间的健康度进行预测。以第 41 个点作为健康度预测的起点,保持建模数据间隔为 5 min,预测未来 30 个点的健康度,预测结果如图 5 所示。

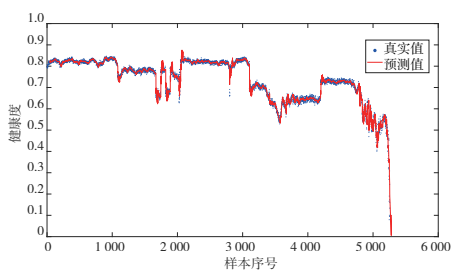


图 5 轴承的健康度预测

利用 RMSE 值对预测准确性进行计算,其中,全局预测的 RMSE 值为 0.02,停机前最后 30 个点的 RMSE 值为 0.32。原因为停机前轴承已进入失效状态,短时间内采样点波动较大,轴承的健康度指标在小范围内出现一定的波动,造成预测值与真实值出现了一定的差异。为验证模型在低采样率下的准确性,对原始数据进行降采样。将原先 5 min 的采样间隔提升为 50 min,轴承的健康度预测结果如图 6 所示。

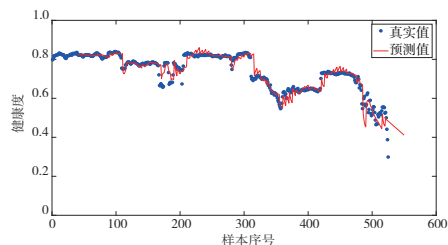


图 6 降采样后健康度预测

对降采样后的预测结果求 RMSE 值,其全局预测 RMSE 值为 0.03,停机前最后 30 个点的 RMSE 值为 0.10。通过对比降采样前后的 RMSE 值,停机前的预测准确性较降采样之前有一定的

提升,原因为降采样在一定程度上剔除了轴承寿命最终阶段较多的干扰数据,使预测结果与真实结果更为接近。

4 实车验证

为进一步验证该方法的准确性,使用某地铁车辆车载 PHM 诊断单元的数据进行评估及预测。PHM 诊断单元的采集机制为定转速触发采集,振动传感器安装于牵引电机的非传动端,采样频率 25 600 Hz,采样时长 2 s,采样触发转速为 $(2\,500 \pm 5\%)$ r/min。

选取其中一台牵引电机数据进行分析,部分特征指标如图 7—图 10 所示。

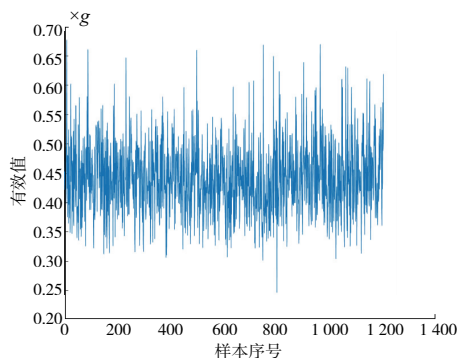


图 7 2 车 1 轴电机振动有效值曲线

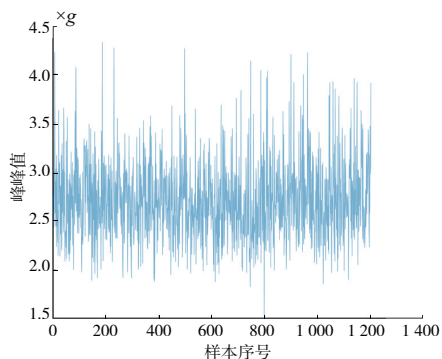


图 8 2 车 1 轴电机峰峰值曲线

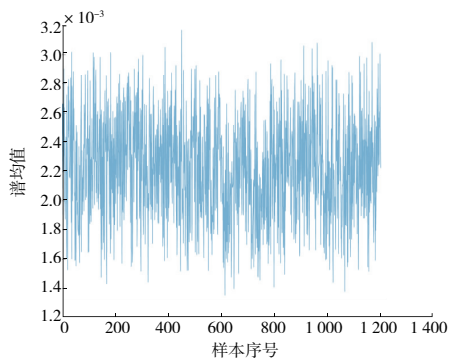


图 9 2 车 1 轴电机谱均值曲线

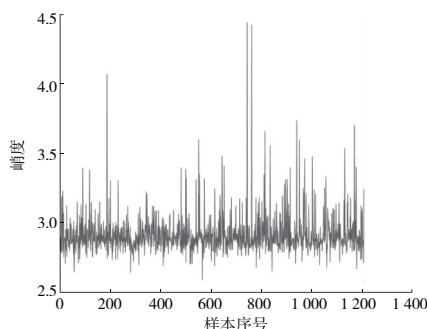


图 10 2 车 1 轴电机峭度曲线

对特征指标进行 PCA 融合并建立的健康度曲线如图 11 所示。

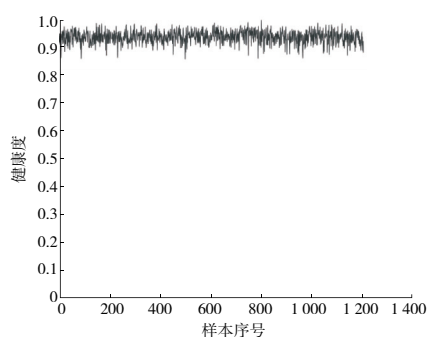


图 11 2 车 1 轴电机健康度曲线

从图 11 可以看出,虽然该电机轴承的振动有效值、峭度存在一定的变化,但其总体健康度维持在 0.9 以上。由于该车为全新运营车辆,运行里程较短,其健康度计算结果与电机轴承的实际状态吻合。

最后使用 GPR 模型对现有数据进行学习,并预测未来一段时间内的轴承健康度变化情况,同时与真实数据进行对比,其预测健康度变化如图 12 所示。

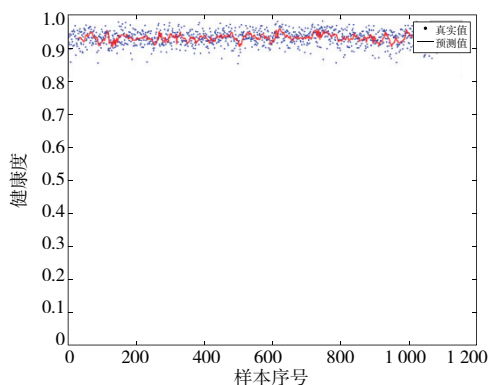


图 12 2 车 1 轴电机健康度预测曲线

根据 RMSE 计算结果,2 车 1 轴电机的全局 RMSE 值及最后 30 个预测点的 RMSE 值均为 0.07,这表明实车验证的准确性已接近台架试验验证的准确性,具备进一步推广的基础条件。

5 结语

针对现阶段只能对轴承是否存在故障进行判断、无法预测轴承后续运行状态的问题,通过使用 PCA 技术对轴承振动的时频域指标进行融合并建立一个可以表征轴承衰退性能的健康度指标,进而利用 GPR 技术完成轴承的健康状态预测研究,再根据 RMSE 值完成预测准确性评估。试验结果表明,使用健康度评估方法能够较为准确地评估轴承状态,通过 GPR 技术预测台架试验轴承健康度的全局 RMSE 值为 0.02,预测实车牵引电机轴承健康度的全局 RMSE 值为 0.07,验证了所提方法的可行性和准确性,为后续进一步开展基于轴承健康度的状态评估及寿命预测奠定了理论基础。

参考文献:

- [1] 曹勇,张玉文,龚艳. 基于大数据和云计算的车辆智能运维模式[J]. 城市轨道交通研究,2020,23(4): 69-73.
- [2] 许艳华,黄贵发,唐德尧. 城轨交通车辆走行部故障自动诊断技术应用[J]. 都市快轨交通,2012,25(5): 131-135.
- [3] 刘永斌,何清波,孔凡让,等. 基于 PCA 和 SVM 的内燃机故障诊断[J]. 振动、测试与诊断,2012,32(2): 250-255,342.
- [4] 王巍. 基于稀疏高斯过程回归的锂电池剩余寿命预测[D]. 北京:北京交通大学,2018.
- [5] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning[M]. Cambridge, Mass.: MIT Press,2006.
- [6] 魏孟,李嘉波,叶敏,等. 基于高斯混合回归的锂离子电池 SOC 估计[J]. 储能科学与技术,2020,9(3): 958-963.
- [7] WILLIAMS T, RIBADENEIRA X, BILLINGTON S, et al. Rolling element bearing diagnostics in Run-to-failure lifetime testing[J]. Mechanical Systems and Signal Processing,2001,15(5):979-993.

收稿日期:2022-10-13