

一种民航发动机滚动轴承故障预警方法^{*}

刘 勇¹, 王 朝², 周 平³

(1. 中国民航大学 航空工程学院, 天津 300300;

2. 中国商飞北京民用飞机技术研究中心, 北京 102211;

3. 成都航利(集团)实业有限公司, 四川 成都 640041)

摘 要: 为了提前感知滚动轴承故障, 避免民用航空发动机非计划维护带来的损失, 提出了一种故障预警方法。对轴承振动信号进行特征工程, 提取其时域和频域特征, 引入梯度提升决策树(Gradient boosting decision tree, GBDT)算法, 量化了特征重要度; 在特征相关性分析的基础上, 利用核主成分分析(Kernel principal component analysis, KPCA)方法实现特征融合与主元提取, 再次结合GBDT构建了故障预警模型, 使用交叉验证法实现了模型泛化能力评估。结果表明: KPCA+GBDT模型的泛化性能显著, 模型的 F_1 分数高达0.991, 对应的受试者工作特性(Receiver operating characteristic, ROC)曲线下面积的值为0.998, 体现出该方法用于支撑航空发动机健康管理与维护决策工作的工程应用价值。

关键词: 航空发动机; 滚动轴承; 故障预警; 核主成分分析; 梯度提升决策树

中图分类号: V232 **文献标识码:** A **文章编号:** 1001-4055(2022)02-200284-10

DOI: 10.13675/j.cnki. tjjs. 200284

An Early Warning Method for Rolling Bearing Fault of Civil Aero-Engine

LIU Yong¹, WANG Chao², ZHOU Ping³

(1. College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China;

2. Beijing Aeronautical Science and Technology Research Institute of COMAC, Beijing 102211, China;

3. Chengdu Holy Industry & Commerce Corp. Ltd (Group), Chengdu 640041, China)

Abstract: In order to detect the rolling bearing fault in advance and avoid the loss caused by unscheduled maintenance of civil aero-engine, a fault early warning method is proposed in the present work. Feature engineering is applied to bearing vibration signals, and its time and frequency domain features are extracted. The gradient boosting decision tree (GBDT) algorithm is adopted to quantify the feature importance. On the basis of feature correlation analysis, the kernel principal component analysis (KPCA) method is used to achieve feature fusion and principal component extraction. Combining with GBDT algorithm again, the fault early warning model is established. The cross-validation method is used to evaluate the generalization ability of the proposed model. The results show that the generalization performance of the KPCA+GBDT model is significant. The F_1 score of the model is as high as 0.991, and the area under the receiver operating characteristic (ROC) curve, area under curve is 0.998, which means this model can be used to support the health management and maintenance decision-making work of civil aero-engine.

Key words: Aero-engine; Rolling bearing; Fault early warning; Kernel principal component analysis; Gradient boosting decision tree

^{*} 收稿日期: 2020-05-03; 修订日期: 2020-08-17。

基金项目: 中国民航大学科研启动基金项目(2020KYQD76)。

通讯作者: 刘 勇, 博士, 讲师, 研究领域为航空发动机结构强度及可靠性。

引用格式: 刘 勇, 王 朝, 周 平. 一种民航发动机滚动轴承故障预警方法[J]. 推进技术, 2022, 43(2):200284. (LIU Yong, WANG Chao, ZHOU Ping. An Early Warning Method for Rolling Bearing Fault of Civil Aero-Engine [J]. Journal of Propulsion Technology, 2022, 43(2):200284.)

1 引 言

航空发动机是一种复杂的旋转机械,现代民用航空发动机为满足高推重比的需求,往往采用高涵道比双转子结构布局,滚动轴承是其中的关键部件,在运行中起到负载支撑和动力传输的作用^[1],然而由于制造工艺、装配误差^[2]、结构耦合振动致使轴承产生磨损、脱落、裂纹等失效形式,进而出现振动异常乃至抱死现象,严重威胁到了航空发动机的可靠性和结构完整性。

对航空器设备进行维护,传统的方式有两种:一是等到有明显的故障特征出现后再维修,这种方式会导致非计划停机,经济损失较大,设备故障危险性增加;二是以固定的计划周期进行维护,但维修所需的人力成本和时间成本更高^[3]。早期轴承的弱故障成因复杂且难以检测,对轴承进行状态监测,实现提前故障预警具有重要意义。

基于轴承振动信号进行故障诊断工作,特征工程是最为关键的一步^[4-5],如何提取轴承从早期正常状态到严重故障期间的强噪声、非线性、非平稳信号特性,降低特征冗余,实现有效信息融合是一大难点^[6-9]。为充分挖掘轴承振动信号的非线性特征,可采用核主成分分析(Kernel principal component analysis, KPCA)算法^[10-11],它是主成分分析(Principal component analysis, PCA)算法的改进,通过引入非线性映射函数,将低维空间数据映射至高维,从而提高特征表达能力^[12-13]。在模型构建方面,近些年深度学习算法的强势崛起再一次推动了人工智能技术的浪潮,卷积神经网络(Convolutional neural network, CNN)、长-短期记忆(Long-short term memory, LSTM)神经网络等神经网络结构横空出世,体现出强大的非线性拟合能力^[14-15],深度神经网络虽然不需要高昂的特征工程成本,便能够自动提取特征^[16-18],但是,如果其网络结构设计、超参数调节不当,会出现优化结果震荡、过拟合、梯度消失、梯度爆炸等现象^[19],对工程技术人员的要求更高,同时,采用神经网络解决问题还需要海量数据和强大算力的支撑,相比之下,以梯度提升决策树(Gradient boosting decision tree, GBDT)为代表的集成学习算法,通过构建并结合多个基学习器来完成学习任务,常常能够获得比单一学习器显著优越的泛化性能^[20-21],相对于深度神经网络来讲,摆脱了对大量样本的需求和强大算力的依赖。

综上所述,航空发动机轴承寿命与具体工况密

切相关,可以用传感器将轴承振动信号采集,上传至数据平台,通过分析多工况历史数据,建立合适的机器学习模型,对轴承在未来一段工作时间内是否出现故障进行提前预警,并为维修人员提前进行维护提供参考,减少因航空发动机非计划停机而带来的损失,同时可避免因轴承故障而造成的灾难性后果。

2 方 法

2.1 轴承振动信号特征工程

算法具有一定的通用性,但无法在未经处理的数据集上发挥作用,能否从原始数据集中提取有效特征,决定了机器学习模型的上限,这些特征可以更好地向算法模型描述数据的潜在问题,以提高模型对未见数据的准确性。

(1) 时域特征提取

时域特征是衡量轴承振动信号的重要特征,分为有量纲时域特征和无量纲时域特征两种。有量纲时域特征是最为常用的,具有直观的统计学及物理含义。设轴承振动幅值时间序列为 $\{s_1, s_2, \dots, s_t\} (t = 1, 2, \dots, T)$,所要提取的有量纲时域特征包括:均值 x_μ ,方差 x_σ ,均方根 x_{rms} ,峭度 x_{kurt} ,偏度 x_{sk} ,最大值 x_{max} ,最小值 x_{min} ,极差 x_{ptp} ,绝对平均值 x_{abs} ,方根幅值 x_t 。各个有量纲时域特征的计算方法如表1所示。

Table 1 Dimensional time domain features	
Feature	Definition and representation
Mean	$x_\mu = \frac{1}{T} \sum_{t=1}^T s_t$
Variance	$x_\sigma = \frac{1}{T} \sum_{t=1}^T (s_t - x_\mu)^2$
Root mean square	$x_{rms} = \sqrt{\frac{1}{T} \sum_{t=1}^T s_t^2}$
Kurtosis	$x_{kurt} = [\frac{1}{T} \sum_{t=1}^T (s_t - x_\mu)^4] / x_\sigma^4$
Skewness	$x_{sk} = [\frac{1}{T} \sum_{t=1}^T (s_t - \bar{x} ^3)] / x_\sigma^3$
Maximum value	$x_{max} = \max(s_t)$
Minimum value	$x_{min} = \min(s_t)$
Range	$x_{ptp} = x_{max} - x_{min}$
Absolute mean	$x_{abs} = \frac{1}{T} \sum_{t=1}^T s_t $
Root mean square amplitude	$x_t = (\frac{1}{T} \sum_{t=1}^T \sqrt{ s_t })^2$

某些有量纲特征对故障早期的振动、冲击具有良好的敏感性,但随着故障的发生,敏感度会逐渐降低,特征稳定性较差;相反的,某些无量纲特征对轴承的转速、载荷等工况变化影响较小,特征稳定性较强,更适合实际的应用场景,无量纲特征是有量纲特征的衍生,因此常常将两类特征同时使用,需要提取的无量纲时域特征包括:峰值因子 x_{C-f} ,脉冲因子 x_{Ip-f} ,偏度因子 x_{P-f} ,峭度因子 x_{K-f} ,裕度因子 x_{Cl-f} ,波形因子 x_{S-f} 。各无量纲时域特征的计算方法如表2所示。

Table 2 Nondimensional time domain features

Feature	Definition and representation
Crest factor	$x_{C-f} = x_{\max}/x_{\text{rms}}$
Impulse factor	$x_{Ip-f} = x_{\max}/\bar{x}_{\text{abs}}$
Skewness factor	$x_{P-f} = x_{sk}/x_{\text{rms}}^3$
Kurtosis factor	$x_{K-f} = x_{\text{kurt}}/x_{\text{rms}}^4$
Clearance factor	$x_{Cl-f} = x_{\max}/x_t$
Waveform factor	$x_{S-f} = x_{\text{rms}}/\bar{x}_{\text{abs}}$

(2) 频域特征提取

除了对时域信号进行统计分析外,还可以采用快速傅里叶变换(Fast Fourier transform,FFT),将时域信号转换至频域,通过描述FFT谱中主频带位置变化、谱能量分布的分散程度来反映不同的故障模式。设经过快速傅里叶变换后,FFT谱在不同频率 $\{f_1, f_2, \dots, f_n\}$ ($n = 1, 2, \dots, N$)下,对应振动幅值为 $\{p_1, p_2, \dots, p_n\}$ ($n = 1, 2, \dots, N$),所要提取的频域特征包括:重心频率 f_{FC} ,均方频率 f_{MSF} ,均方根频率 f_{RMSF} ,频率方差 f_{VF} ,频率标准差 f_{RVF} 。各个频域特征的计算方法如表3所示。

Table 3 Frequency domain features

Feature	Definition and representation
Center of gravity frequency	$f_{FC} = (\sum_{n=1}^N f_n p_n) / \sum_{n=1}^N p_n$
Mean square frequency	$f_{MSF} = (\sum_{n=1}^N f_n^2 p_n) / \sum_{n=1}^N p_n$
Root mean square frequency	$f_{RMSF} = \sqrt{f_{MSF}}$
Variance of frequency	$f_{VF} = [\sum_{n=1}^N (f_n - f_{FC})^2 p_n] / \sum_{n=1}^N p_n$
Standard deviation of frequency	$f_{RVF} = \sqrt{f_{VF}}$

2.2 KPCA 特征融合方法

特征工程所得的高维特征容易出现特征之间的线性相关,为了降低特征信息冗余,提高特征区分度,

在构建模型前可尝试对数据进行降维。对非线性程度较高的特征进行主元提取,KPCA方法是一种更好的选择,它是对PCA方法的推广,引入核函数非线性映射,将原始特征映射至高维空间 F ,由低维线性不可分转换为高维线性可分,随后采用PCA方法进行线性降维。因此,KPCA方法能够有效保留原始数据特征,并能够抽取原始特征中蕴含的非线性关系,具体过程如下:

假设存在 M 个样本的数据集 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i\}$ ($i = 1, 2, \dots, M$), $\mathbf{x}_i \in R^N$ 且样本维度为 N ,对高维空间数据进行标准化,使其满足

$$\frac{1}{M} \sum_{i=1}^M \boldsymbol{\varphi}(\mathbf{x}_i) = 0 \quad (1)$$

$\boldsymbol{\varphi}$ 是将低维特征空间向量 $\mathbf{x}_i$ 映射至高维空间 F 的非线性映射函数,此时协方差矩阵 C 可表示为

$$C = \frac{1}{M} \sum_{i=1}^M \boldsymbol{\varphi}(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}_i)^T \quad (2)$$

协方差矩阵 C 的特征值 λ 和特征向量 $\boldsymbol{\nu}$ 满足

$$C\boldsymbol{\nu} = \lambda\boldsymbol{\nu} \quad (3)$$

寻找矩阵特征值 λ ($\lambda \geq 0$),以及对应的特征向量 $\boldsymbol{\nu}$ ($\boldsymbol{\nu} \in F$),所有特征向量的解 $\boldsymbol{\nu}$ 其实都在 $\boldsymbol{\varphi}(\mathbf{x}_1), \dots, \boldsymbol{\varphi}(\mathbf{x}_i)$ ($i = 1, 2, \dots, M$)的空间中,求解特征方程可得

$$\boldsymbol{\varphi}(\mathbf{x}_k) C \boldsymbol{\nu} = \lambda \boldsymbol{\varphi}(\mathbf{x}_k) \boldsymbol{\nu} \quad (4)$$

对特征向量 $\boldsymbol{\nu}$ 进行线性表示,即

$$\boldsymbol{\nu} = \sum_{i=1}^M \alpha_i \boldsymbol{\varphi}(\mathbf{x}_i) \quad (5)$$

则

$$\begin{aligned} & \frac{1}{M} \sum_{i=1}^M \alpha_i \sum_{j=1}^M [\boldsymbol{\varphi}(\mathbf{x}_k) \boldsymbol{\varphi}(\mathbf{x}_j)] [\boldsymbol{\varphi}(\mathbf{x}_j) \boldsymbol{\varphi}(\mathbf{x}_i)] = \\ & \lambda \sum_{i=1}^M \alpha_i [\boldsymbol{\varphi}(\mathbf{x}_k) \boldsymbol{\varphi}(\mathbf{x}_i)] \end{aligned} \quad (6)$$

从输入空间到高维特征空间的非线性映射通过核函数内积运算实现,若在低维空间存在函数满足 $k(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}_j)$,则称该函数为核函数,常见的有线性核函数、多项式核函数、高斯径向基核函数等,本文均选用高斯径向基核函数,其表达式为

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (7)$$

式中参数 γ 用于控制径向基函数的作用范围。

随后定义 $M \times M$ 维矩阵 K ,其中的元素均能使用核函数表示,即

$$K = \begin{bmatrix} \varphi(x_1)\varphi(x_1) & \cdots & \varphi(x_1)\varphi(x_m) \\ \vdots & \ddots & \vdots \\ \varphi(x_m)\varphi(x_1) & \cdots & \varphi(x_m)\varphi(x_m) \end{bmatrix} = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \cdots & k(x_m, x_m) \end{bmatrix} \quad (8)$$

采用核矩阵表示式(6),可简化为

$$K\alpha = M\lambda\alpha \quad (9)$$

根据该式可计算核矩阵的特征值与特征向量,进而求出协方差矩阵的归一化特征向量 $\nu^k(k=1, 2, \dots, M)$,则样本 x 的第 k 个线性主元为

$$h_k = \nu^k \varphi(x) = \sum_{i=1}^M \alpha_i^k K(x_i, x) \quad (10)$$

令 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 为核矩阵特征值,计算特征累积贡献率选取主元为

$$\sum_{k=1}^p \lambda_k / \sum_{i=1}^m \lambda_i \geq 0.9 \quad (11)$$

2.3 梯度提升决策树(GBDT)模型

梯度提升决策树(Gradient boosting decision tree, GBDT)是一种集成学习算法,它通过多轮迭代,每次迭代产生一个弱分类器,在上一轮分类器预测结果与实际值的残差基础上进行训练学习,通过多个弱分类器的线性组合,不断减少训练过程的残差,具体过程如下:

(1)首先初始化弱学习器,有

$$f_0(x) = \arg \min \sum_{i=1}^N L(y_i, c) \quad (12)$$

c 为经最小二乘法划分节点后的叶子节点取值。

(2)对每棵树 $m = 1, 2, \dots, M$

对每个样本 $i = 1, 2, \dots, N$,计算负梯度,即残差,有

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right] \quad (13)$$

式中 $f(x_i)$ 为弱学习器的预测值, y_i 为真实值。

将得到的残差值作为新样本真实值,数据 $(x_i, r_{im}), i = 1, 2, \dots, N$ 作为下棵树的训练数据,得到新的回归树 $f_m(x)$,其对应的叶子节点区域为 $R_{jm}, j = 1, 2, \dots, J, J$ 为回归树 t 的叶子节点个数,对叶子区域 $j = 1, 2, \dots, J$,计算最佳拟合值,有

$$\gamma_{jm} = \arg \min \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) \quad (14)$$

更新强学习器,则有

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}) \quad (15)$$

(3)得到最终学习器

$$f(x) = f_0(x) + \sum_{m=1}^M \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}) \quad (16)$$

2.4 模型性能度量

机器学习的目的是通过学习训练数据,更好地适用于“新样本”。在某些情况下由于模型的学习能力过强,以至于把训练样本的特有性质当成一般样本所具有的普遍性质进行学习,随之呈现出“过拟合”现象,与其对应的则是“欠拟合”,表明模型对样本的一般性质都未学习到位,有鉴于此,需要对模型泛化性能进行评估。

(1) 混淆矩阵

以轴承故障诊断任务为例,不光关心预测样本中有多少被诊断正确,即精准率 P (Precision),还需要关心真正的故障样本中是否存在漏检,即召回率 R (Recall),混淆矩阵能够实现此类需求下的模型性能评估。

以简单的二分类问题为例,可将样本真实类别与预测类别进行组合,划分为:真正例(True positive, TP)、假正例(False positive, FP)、假负例(False negative, FN)、真负例(True negative, TN),以上 4 种组合构成了整个测试数据集,混淆矩阵如表 4 所示。

Table 4 Confusion matrix

Truth	Predict	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

定义精准率 P 和召回率 R 为

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

在不同场景下,常采用 F_1 分数(F_1 score),综合考虑对精准率 P 与召回率 R 的重视程度,它是两者的调和平均值,记为

$$F_1 = \frac{2PR}{P + R} \quad (19)$$

(2) ROC 曲线及 AUC

很多分类器以概率为阈值界定最终的分类结果,大于阈值被认为是正例,小于阈值则是负例。除了采用前面提到的 F_1 分数作为性能度量外,还可针对任务需求采用不同阈值的截断点,受试者工作特性(Receiver operating characteristic, ROC)曲线便是从这个角度研究模型泛化性能的有力工具。其纵轴是真正率(True positive rate, TPR),横轴是假正率

(False positive rate, FPR), 分别定义为

$$TPR = \frac{TP}{TP + FN} \quad (20)$$

$$FPR = \frac{FP}{FP + TN} \quad (21)$$

绘制 ROC 曲线,需按照一定步长进行,对阈值区间 $[0,1]$ 进行等比例划分,计算每个阈值点下 TPR 与 FPR 值,若某分类器 ROC 曲线能将另一分类器 ROC 曲线包住,则可断定前者泛化性能更优,并用 ROC 曲线下面积(Area under curve, AUC)进行定量表示。

3 结果与讨论

基于 KPCA 与 GBDT 的轴承故障诊断流程主要包括以下步骤:轴承试验台振动信号采集、特征工程与特征融合、数据集组建与拆分、轴承故障预警模型建立、模型评估,如图 1 所示。研究所采用的数据集来源于某型航空发动机高压转子中介轴承的振动试验,其转子支撑方案如图 2 所示。

转子结构总共包含 5 个承力支点,其中高压转子采用 2 支点 3 轴承的 1-0-1 支撑方案,前支点为滚珠、滚棒双轴承并列结构,独特的结构设计使滚珠轴承、滚棒轴承分别承受轴向、径向载荷。因该型发动机高压压气机级数少,转速高,所以设计了单级涡轮结构,采用了中介轴承,将高压转子支撑于低压转子上。中介轴承的使用虽然可以缩短转子长度,节省一个承力框架,减轻发动机质量,从而提高推重比,但从动力学特征的角度分析,中介轴承会导致高低压转子间振动耦合,振动特性表现更为复杂,基于以上分析,本文选取 3-1 号滚珠轴承为对象进行研究。

3-1 号滚珠轴承如图 3 所示。

为验证所提出轴承故障预警方法的有效性,通过缩比验证模型模拟中介轴承工况环境,所搭建的实验台主要由滚动轴承、可拆卸轴承座、电动机、调速装置、数据采集传感器、数据存储计算机等设备组成。驱动电机转速为 1730r/min,使用联轴器与转子相连,传感探头采集轴承振动幅值并连接至数据采集设备,其采样频率为 6kHz,试验台结构如图 4 所示。

试验过程共生成样本 1320 组,其中包含外圈故障、内圈故障、滚珠故障 3 种故障模式,为了提升训练模型的鲁棒性,对样本类别进行扩充,每种故障模式设定 3 种不同尺寸的缺陷(编号为 No.1,2,3),共 9 类故障样本,外加正常运行的状态,共计 10 种工况。

图 5 选取了一些具有代表性的样本,展示了不同工况下,轴承的时域和频域信号。

观察轴承振动信号可知:不同工况下轴承的时域振动信号及频谱图具有显著差异,轴承正常工作时,其振动平稳且振动幅值维持在较低水平;故障模式下,振动信号可能伴有多次、短时、强烈的冲击成分,或呈现出振幅分布不均、频谱分散等特点。

采用 GBDT 算法,使用所提取的原始特征,建立初步故障预警模型,并计算不同特征重要度,其可视化结果如图 6(a)所示。图中横坐标为原始特征,纵坐标表示无量纲特征重要度,通过遍历 GBDT 模型的所有决策子树,计算其节点划分前后的信息增益熵,并做归一化处理,对不同特征的信息增益熵除以全部特征的信息增益熵总和,转换为百分比度量。图中结果表明,所提取的特征对模型建立均有不同程

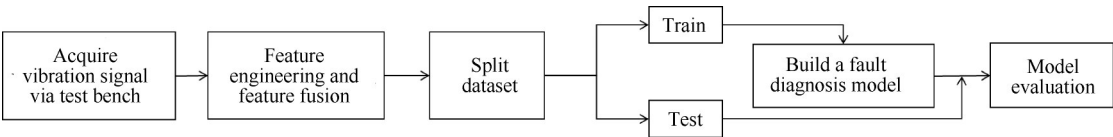


Fig. 1 Flow chart of bearing fault diagnosis

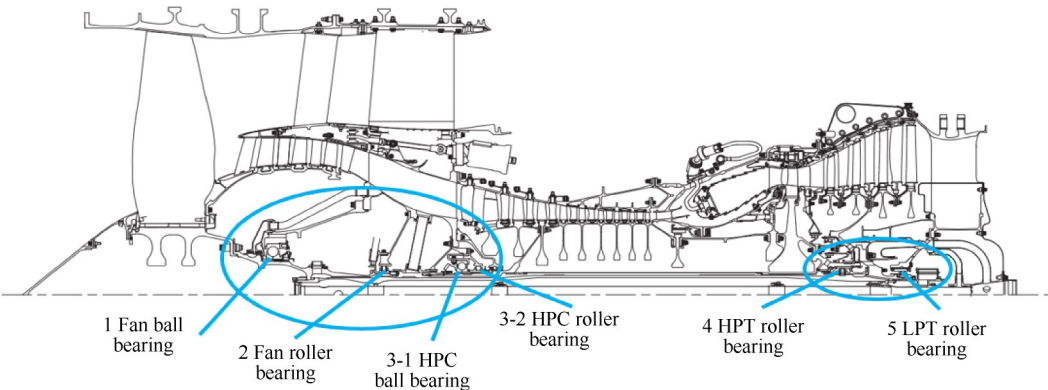


Fig. 2 Civil aviation engine rotor support scheme

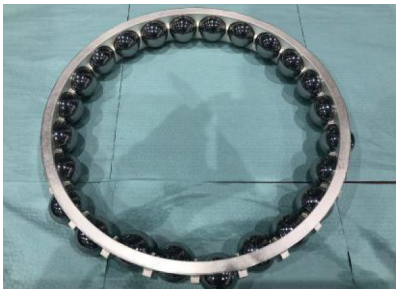


Fig. 3 Rolling parts of No.3-1 ball bearing

度的贡献,体现了特征工程结果的有效性。分析部分特征对模型影响的原因可能是:频率标准差 f_{RVF} ,频率方差 f_V ,方差 x_σ 均是振动能量离散程度的度量;均方根 x_{rms} ,绝对平均值 x_{abs} ,方根幅值 x_i 能够体现稳定的振动能量大小;重心频率 f_{FC} ,均方频率 f_{MSF} 代表了FFT谱中主频带的位置;峰度 x_{kurt} ,峰值因子 $x_{C,f}$,脉冲因子 $x_{Ip,f}$,极差 x_{pfp} 反映了轴承滚动体与故障点产生强烈撞击而带来的短时大幅度冲击振动;偏度 x_{sk} 能够衡量信号分布的不对称程度。

构造原始特征的协方差矩阵后,可求解不同特征间的相关系数矩阵,相关系数是不同特征之间线性相关程度的无量纲度量,值越接近1,说明正相关性越强;值越接近-1,说明负相关性越强,其可视化云图如图6(b)所示:部分特征之间存在一定的相关性,为了降低特征冗余,需采用KPCA算法实现特征融合。

经特征工程后生成的数据集包含样本1320条,样本特征21个,基于网格搜索算法,确定了径向基核函数中的超参数 $\gamma=0.054$,并生成核矩阵 K ,随后将核矩阵中心化,运用Jacobi迭代法计算核矩阵特征值及对应的特征向量,向量规范化后,将特征值降序排列,逐步计算特征值的累积贡献率,根据预先设定的90%累计贡献率阈值,提取了前5阶主元,用于GBDT模型构建,选取贡献率最高的前2阶主元Var_1和Var_2,可得其二维可视化结果如图7所示。

降维可视化结果表明:采用KPCA方法提取主元后,不同工况数据表现出良好的聚类特性,样本间区

分度较为清晰,但仍存在小部分重叠,分析其原因可能是可视化维度较低所致,特征被映射至高维时,聚类特性将会更加清晰。

在对特征使用KPCA算法进行特征融合与主元选择后,再次结合GBDT分类器构建轴承故障预警模型,采用5折交叉验证法实现模型评估,结果如图8和表5所示,图8中的数值表示样本个数。

KPCA+GBDT方法对于不同类型、不同尺寸缺陷的轴承故障均有良好的诊断效果,仅有小部分样本诊断存在问题,其精准率可达0.995,召回率可达0.987, F_1 分数可达0.991,一方面验证了前期特征工程的有效性,同时还体现了KPCA强大的非线性特征融合表达能力,两者为机器学习模型的构建创造了良好的前提条件。

Table 5 KPCA + GBDT generalization performance evaluation

Working condition	Precision P	Recall R	F_1 score
Normal operation	0.997	1.000	0.998
No.1 outer race fault	1.000	0.978	0.989
No.1 inner race fault	1.000	0.967	0.983
No.1 rolling parts fault	0.992	0.983	0.987
No.2 outer race fault	1.000	0.980	0.990
No.2 inner race fault	1.000	0.983	0.991
No.2 rolling parts fault	1.000	0.991	0.995
No.3 outer race fault	0.964	1.000	0.982
No.3 inner race fault	1.000	0.988	0.994
No.3 rolling parts fault	0.995	1.000	0.997
Avg	0.995	0.987	0.991

将本文所使用的梯度提升决策树(GBDT)算法与其他常用的分类器:逻辑回归(Logistic regression, LR),支持向量机(Support vector machines, SVM)进行了泛化性能对比,绘制了采用KPCA特征融合方法前后的ROC曲线,并计算对应的ROC曲线下面积,结果

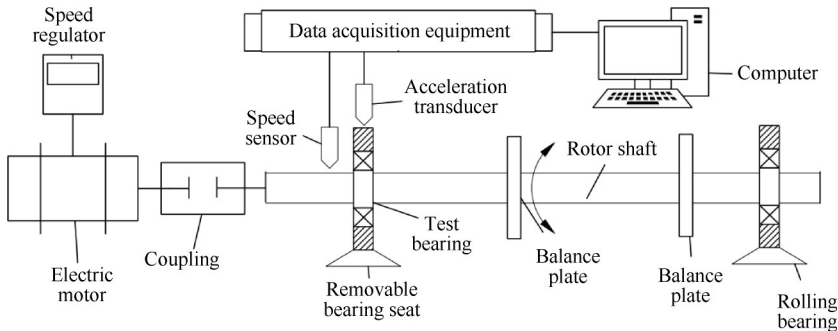


Fig. 4 Schematic diagram of vibration signal test platform

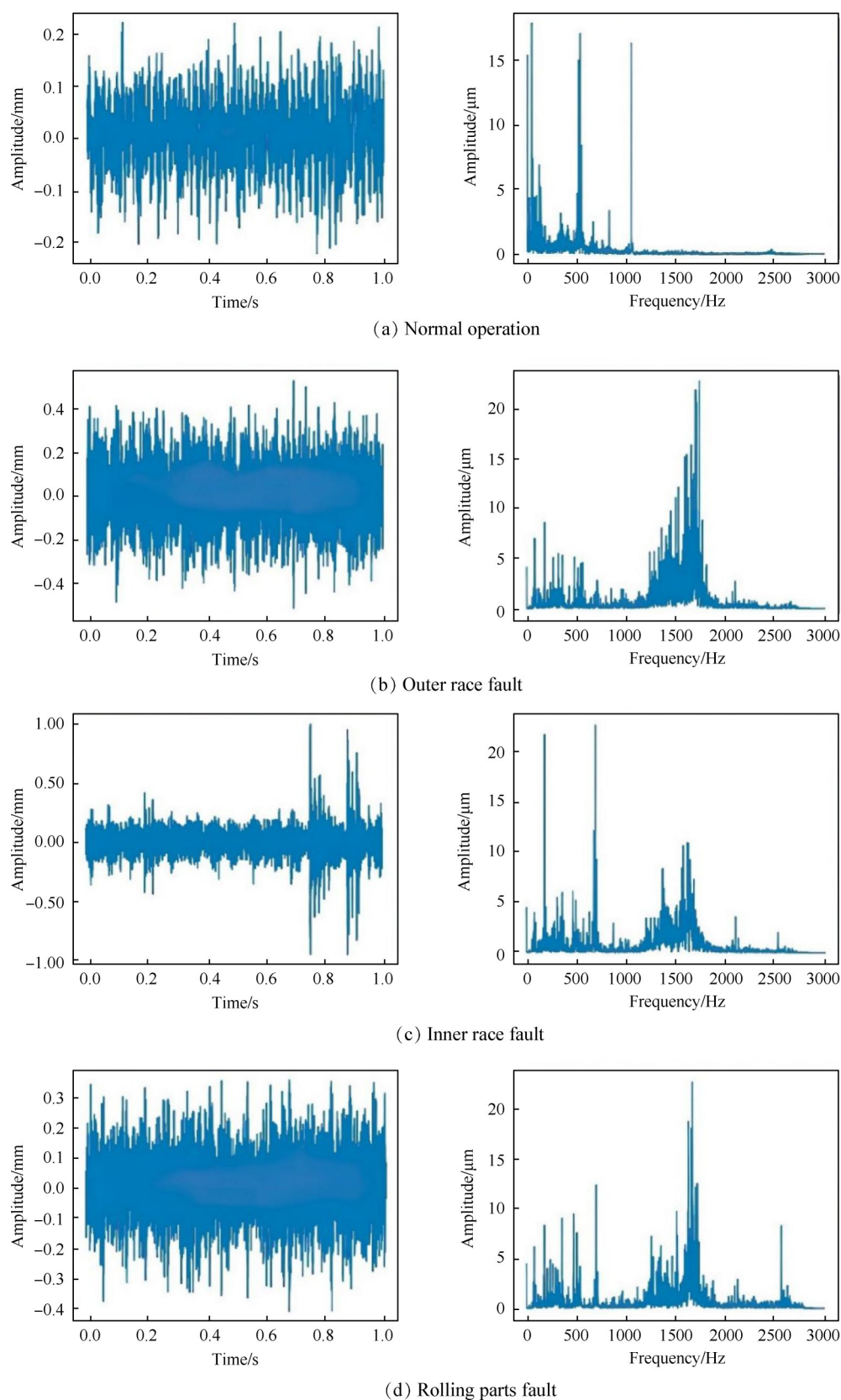


Fig. 5 Time and frequency domain signals of the tested bearing

如表6和图9所示。对比分析结果可知,未采用KPCA进行特征融合前,GBDT算法的泛化能力已十分显著,其 F_1 分数高达0.980,对应的AUC值为0.997,

相比于其他分类器具有明显的优势,这得益于集成学习方法强大的拟合能力;采用KPCA方法后,SVC,LR,GBDT分类器的 F_1 分数分别提升至0.979,0.822,

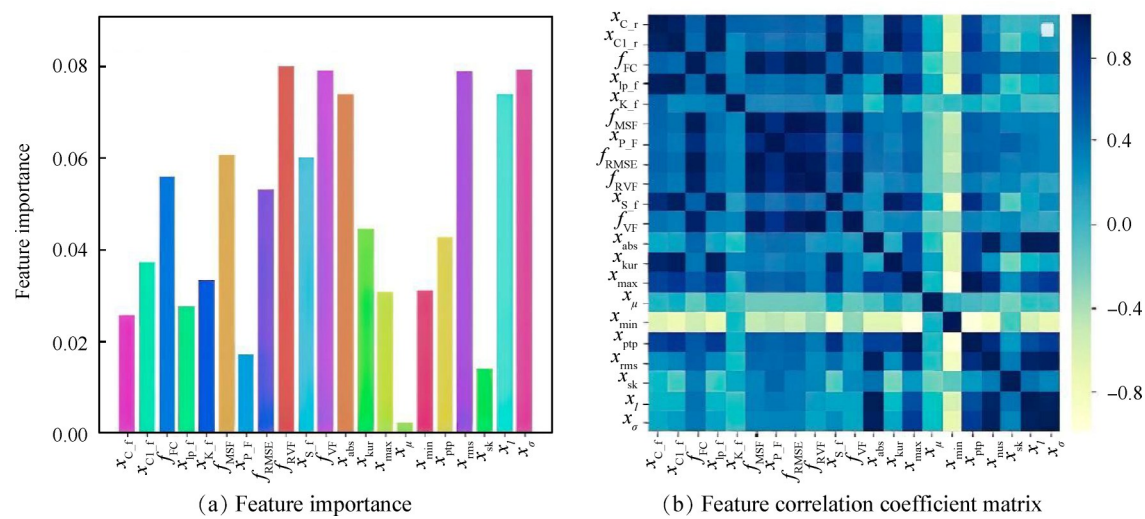


Fig. 6 Feature importance and feature correlation coefficient matrix

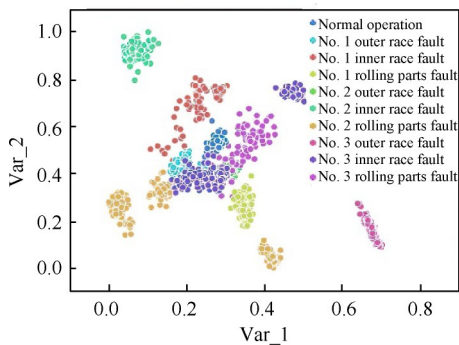
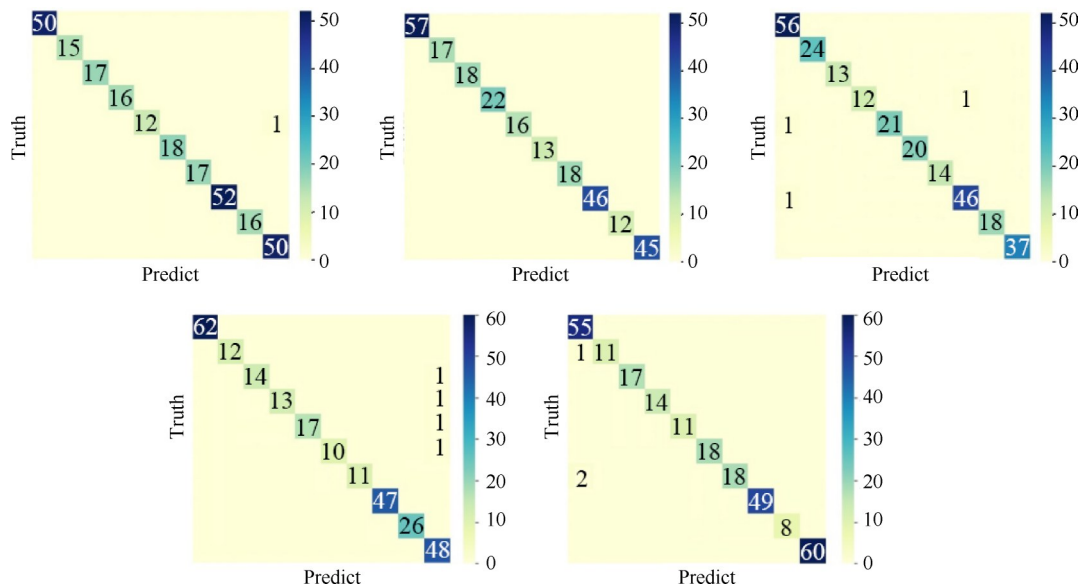


Table 6 Comparison of generalization performance of different classifiers

Classifier	Precision P	Recall R	F_1 score	Aera under ROC(AUC)
SVC	0.805	0.829	0.817	0.936
LR	0.724	0.773	0.720	0.864
GBDT	0.982	0.980	0.980	0.997
KPCA+SVC	0.980	0.980	0.979	0.995
KPCA+LR	0.827	0.818	0.822	0.884
KPCA+GBDT	0.995	0.987	0.991	0.998

0.991, 对应的 ROC 曲线下面积 AUC 值为 0.995, 0.884, 0.998, 各分类器的分类性能均得到了一定程度的改善, 分析其原因可能在于: 航空发动机滚动轴承是一种十分复杂非线性系统, KPCA 方法通过引入

非线性核函数, 能够有效进行主元提取和特征降维, 特征区分度提高, 模型泛化能力增强。



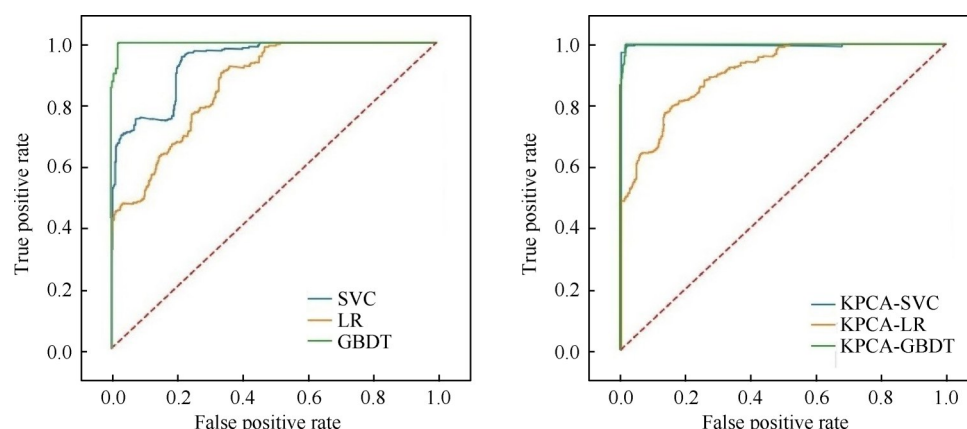


Fig. 9 ROC curve of fault diagnosis model

4 结 论

通过本文研究,得出以下结论:

(1)振动信号是轴承工作状态的反映,不同工况下轴承的振动信号、频谱图具有显著差异,良好的特征工程方法能为机器学习模型的构建创造良好的前提条件,同时,特征重要度分析结果也印证了所提取时域、频域特征的有效性。

(2)原始数据经特征工程后,特征之间往往具有线性相关性,采用KPCA算法进行特征融合与主元提取后,SVC,LR,GBDT分类器的 F_1 分数分别提升至0.979,0.822,0.991,对应的ROC曲线下面积AUC值为0.995,0.884,0.998,表明KPCA方法具有提高特征区分度的作用,并在模型泛化能力上得到体现。

(3)对于轴承的故障预警问题,KPCA+GBDT方法的 F_1 分数高达0.991,对应的AUC值为0.998,体现了该方法用于解决轴承故障预警问题所具备的工程应用价值,同时也可为其他系统的健康管理和维护决策工作提供参考。

致 谢:感谢中国民航大学科研启动基金项目的资助。

参考文献

- [1] 洪 杰,马艳红,张大义. 航空燃气轮机总体结构与动力学分析[M]. 北京:北京航空航天大学出版社,2014.
- [2] 陈志英,刘 勇,周 平,等. 基于改进Taguchi方法的航空发动机装配成功率计算方法[J]. 推进技术,2018,39(3): 653-659. (CHEN Zhi-ying, LIU Yong, ZHOU Ping, et al. Assembly Success Rate Calculation Method for Aero-Engine Based on Improved Taguchi
- Method[J]. *Journal of Propulsion Technology*, 2018, 39(3): 653-659.)
- [3] 陈志英,陈 光. 航空发动机维修性工程[M]. 北京:北京航空航天大学出版社,2013.
- [4] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.
- [5] 雷亚国,贾 峰,孔德同,等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报,2010,54(5): 99-104.
- [6] 程军圣,于德介,邓乾旺,等. 连续小波变换在滚动轴承故障诊断中的应用[J]. 中国机械工程,2003,14(23): 2037-2040.
- [7] 向 丹,岑 健. 基于EMD熵特征融合的滚动轴承故障诊断方法[J]. 航空动力学报,2015,30(5): 133-139.
- [8] Sinha J K, Elbhah K. A Future Possibility of Vibration Based Condition Monitoring of Rotating Machines [J]. *Mechanical Systems & Signal Processing*, 2013, 34(1-2): 231-240.
- [9] 鲁 峰,黄金泉,陈 煜,等. 基于SPSO-SVR的融合航空发动机传感器故障诊断[J]. 航空动力学报,2009,24(8): 1856-1865.
- [10] Navi M, Meskin N, Davoodi M. Sensor Fault Detection and Isolation of an Industrial Gas Turbine Using Partial Adaptive KPCA [J]. *Journal of Process Control*, 2018, 64: 37-48.
- [11] Deng F, Yang S, Liu Y, et al. Fault Diagnosis of Rolling Bearing Using the Hermitian Wavelet Analysis, KPCA and SVM[C]. *Shanghai: IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control*, 2017.
- [12] 么子云,朱丽娜,潘 彪,等. 一种基于小波包与KPCA的发动机多信号融合故障诊断方法[J]. 现代制造

- 工程, 2018, (6): 157-162.
- [13] Ben Ali J, Fnaiech N, Saidi L, et al. Application of Empirical Mode Decomposition and Artificial Neural Network for Automatic Bearing Fault Diagnosis Based on Vibration Signals [J]. *Applied Acoustics*, 2015, 89: 16-27.
- [14] Pan Honghu, He Xingxi, Tang Sai, et al. An Improved Bearing Fault Diagnosis Method Using One-Dimensional CNN and LSTM[J]. *Journal of Mechanical Engineering*, 2018, 64(7-8): 443-452.
- [15] 宋 亚, 夏唐斌, 郑 宇, 等. 基于 Autoencoder-BLSTM 的涡扇发动机剩余寿命预测[J]. 计算机集成制造系统, 2019, 25(7): 1611-1619.
- [16] 李 俊, 刘永葆, 余又红. 卷积神经网络和峭度在轴承故障诊断中的应用[J]. 航空动力学报, 2019, 34(11): 2423-2431.
- [17] 吴静然, 丁恩杰, 崔 冉, 等. 采用多尺度注意力机制的旋转机械故障诊断方法[J]. 西安交通大学学报, 2020, 54(2): 51-58.
- [18] Qian S, Yang X, Jie H, et al. Application of New Training Method Combined with Feedforward Artificial Neural Network for Rolling Bearing Fault Diagnosis [C]. *Nanjing: IEEE 2016 23rd International Conference on Mechatronics and Machine Vision in Practice*, 2016.
- [19] Lei Y, Jia F, Lin J, et al. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data[J]. *IEEE Transactions on Industrial Electronics*, 2016, 63(5): 3137-3147.
- [20] Zhang R, Li B, Jiao B. Application of XGboost Algorithm in Bearing Fault Diagnosis[J]. *IOP Conference Series: Materials Science and Engineering*, 2019, 490: 72062.
- [21] Kou L, Qin Y, Zhao X. An Integrated Model of kNN and GBDT for Fault Diagnosis of Wheel on Railway Vehicle [C]. *Chongqing: 2018 Prognostics and System Health Management Conference*, 2018.

(编辑:史亚红)