

引用格式:欧白羽,杨勐,韩旭. 基于互信息的主成分分析结合支持向量回归的滚动轴承剩余寿命预测研究[J]. 北京化工大学学报(自然科学版), 2021, 48(6): 108-117.

OU BaiYu, YANG Meng, HAN Xu. Remaining useful life prediction of rolling bearings based on mutual information based principal component analysis combined with support vector regression[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2021, 48(6): 108-117.

# 基于互信息的主成分分析结合支持向量回归的滚动轴承剩余寿命预测研究

欧白羽<sup>1</sup> 杨 勐<sup>2\*</sup> 韩 旭<sup>3</sup>

(1. 陆军航空兵学院 陆军航空兵研究所, 北京 101121; 2. 陆军航空兵学院 航空机械工程系, 北京 101123; 3. 93146 部队, 北京 100000)

**摘 要:** 滚动轴承作为旋转机械设备中的关键部件,影响着设备的可靠性运行。为了智能开展设备维护工作,提高设备的运转效率,提出一种基于互信息(mutual information, MI)的主成分分析(principal component analysis, PCA)(MI-PCA)结合支持向量回归(support vector regression, SVR)的滚动轴承剩余寿命预测方法。首先利用小波包降噪算法剔除原始振动信号中的异常数据点和噪声,并基于降噪数据提取其时域、频域和时频域特征;然后结合特征与剩余寿命的互信息值进行特征筛选,再通过 PCA 降维算法获得可表征轴承退化状态的敏感特征,用于 SVR 的输入;最后构建并训练 SVR 剩余寿命预测模型,并将其应用于滚动轴承全寿命试验数据。试验结果表明与基于 MI 和基于 PCA 的 SVR 回归预测模型(MI-SVR 模型、PCA-SVR 模型)相比,基于 MI-PCA 的 SVR 模型具有更高的预测精度(预测精度可达 97%),能够实现滚动轴承剩余寿命的精准预测,为开展及时有效的设备维护工作提供了决策依据。

**关键词:** 滚动轴承; 剩余寿命预测; 支持向量回归; 主成分分析; 互信息

**中图分类号:** TP391 **DOI:** 10.13543/j.bhxbzr.2021.06.014

## 引 言

随着现代工业的快速发展,旋转机械设备越来越大型化、复杂化和智能化<sup>[1]</sup>,如何保证旋转设备的高可靠性安全运行,提高其工作效率,已成为当前亟需解决的关键问题之一。滚动轴承作为旋转机械设备的重要部件,在复杂的工作条件下易发生疲劳损伤<sup>[2-3]</sup>,进而影响设备的运行效率与安全。因此,对滚动轴承进行有效、可靠、准确的剩余使用寿命预测具有重要的实际意义。基于预测的剩余寿命开展及时的维护工作,可提高设备的整体运转效率,避免了在传统的维护工作中存在的设备得不到及时维护、设备发生故障的风险增加以及健康状态下维护

次数过多而造成的资源浪费等问题<sup>[4]</sup>。

鉴于轴承的振动信号中包含大量的运行状态信息<sup>[5]</sup>,当前多以设备的振动监测数据为基础进行剩余寿命的预测研究<sup>[6]</sup>。通常,采集的振动信号中包含数据异常点和噪声,而这些噪声会给预测结果带来误差,因此需要对原始振动信号进行去噪处理。Beale 等<sup>[7]</sup>提出了自适应小波包阈值降噪算法用于消除风力涡轮机叶片的振动噪声。许志华等<sup>[8]</sup>研究了基于变分模态分解的风机齿轮箱振动信号降噪方法。王大兴<sup>[9]</sup>针对轴承振动信号提出了经验模态分解与小波包多阈值相结合的降噪方法,有效剔除了信号中的异常数据点。

在实现振动信号的降噪处理后,就需要提取表征轴承退化状态的特征,目前主要使用信号处理的方式提取常见的时域特征、频域特征和时频域特征等<sup>[10]</sup>。在提取的众多特征中,存在着部分冗余特征和对轴承退化状态不敏感的特征,因此需要对其进行约简,以获得有效的敏感特征作为预测模型的输

收稿日期: 2021-04-29

第一作者: 男,1978 年生,高级工程师

\* 通信联系人

E-mail: yangmengniu@163.com

入,从而提高模型的预测精度。常用的特征约简方法包括相关系数法、互信息法 (mutual information, MI)、流形学习<sup>[11]</sup> 和主成分分析 (principal component analysis, PCA)<sup>[12]</sup> 等。然而,相关系数法和互信息法只是对提取的多维特征进行筛选<sup>[13]</sup>,并没有考虑特征之间的冗余性,若特征之间存在信息重叠,则容易产生预测模型的“过拟合”现象,进而影响模型的预测精度<sup>[14]</sup>,而 PCA 可以消除特征之间的冗余信息<sup>[12]</sup>。近年来,智能算法的持续改进与发展为剩余寿命的预测研究提供了新思路,如双向长短时记忆神经网络<sup>[15]</sup>、改进隐马尔可夫模型<sup>[16]</sup>、基于 Bayesian optimization and hyperband (BOHB) 的深度置信网络<sup>[17]</sup> 以及深度卷积神经网络<sup>[18]</sup> 等都取得了较好的预测效果。但这些算法的结构较为复杂且超参数众多,而支持向量回归 (support vector regression, SVR) 算法的泛化能力和非线性映射能力强且结构简单<sup>[19]</sup>,能有效解决小样本问题<sup>[20]</sup>。

基于以上分析,本文提出了一种基于 MI-PCA 结合 SVR 的滚动轴承剩余寿命预测方法。通过小波包降噪算法实现原始振动信号的去噪处理,并提取时

域、频域和时频域特征;然后计算特征与剩余寿命之间的互信息值进行特征筛选,进而结合 PCA 算法获取敏感特征用于 SVR 的输入;最后建立 SVR 预测模型,实现轴承剩余寿命的精准预测。滚动轴承的全寿命试验数据验证表明,与基于 MI 和基于 PCA 的 SVR 模型相比,所提模型具有更高的预测精度,有利于开展及时的设备维护工作,提高设备的运转效率。

## 1 剩余寿命预测建模

### 1.1 建模流程

为了实现滚动轴承剩余使用寿命的精准预测,本文以轴承的振动信号数据为基础,提出了基于 MI-PCA 结合 SVR 的剩余寿命预测建模方法。流程包括 3 个部分:(1) 利用小波包降噪算法实现原始振动信号的预处理,并提取表征轴承退化状态的时域、频域和时频域特征;(2) 基于特征与剩余寿命的互信息值进行特征筛选,再通过 PCA 算法获取敏感特征用于 SVR 模型的输入;(3) 建立并训练 SVR 回归预测模型,实现轴承的剩余寿命预测。具体的建模流程如图 1 所示。

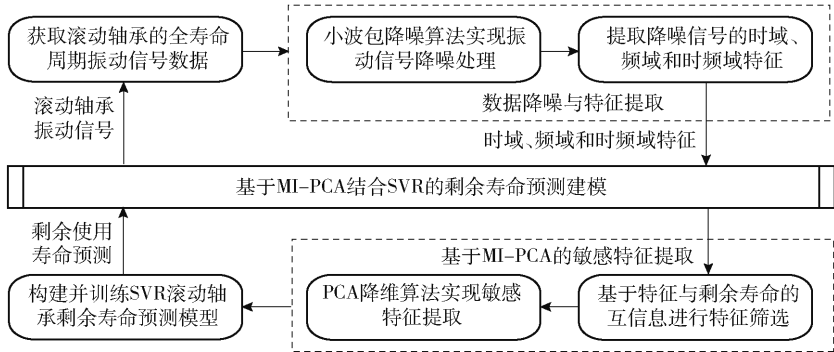


图 1 剩余寿命预测建模流程

Fig. 1 The process of remaining useful life prediction modeling

### 1.2 数据降噪与特征提取

由于旋转机械设备结构的复杂性以及工作条件的多样性,导致采集的滚动轴承振动信号中包含大量的噪声<sup>[21]</sup>,染噪后的信号数据会影响轴承剩余寿命预测的精度。因此,本文首先采用小波包降噪算法对原始振动信号进行去噪处理。小波包降噪的关键在于信号的分解与重构,相比于传统的小波变换,小波包分析不仅对每层中的低频部分进行分解,而且对高频部分也进行分解<sup>[22]</sup>,能够有效提高信号高频段的分辨率,从而使得信号的表达更为精确。为了进一步挖掘振动信号中的隐含状态信息,使用信号处理的方法从不同角度提取表征轴承退化状态的

特征。一般可以从时域、频域和时频域中提取相关特征,以充分利用振动信号数据实现剩余寿命的精准预测。

在提取频域特征时,首先利用快速傅里叶变换将去噪后的时域振动信号  $x(t)$  ( $t = 1, 2, \dots, n$ ) 转换为频域振动信号  $y(k)$  ( $k = 1, 2, \dots, m$ ),然后提取相关的频域特征。

在提取时频域特征时,首先通过经验模态分解将时域振动信号从高频至低频分解为若干个本征模态函数 (intrinsic mode function, IMF),即 IMF 分量  $c(t)$  ( $t = 1, 2, \dots, n$ )。为了保证分量信号的有效性,选取前 3 阶 IMF 分量的有效值作为时频域特征。

本文提取的时域、频域和时频域特征如表 1 所示。

表 1 提取的时域、频域和时频域特征  
Table 1 The time-domain, frequency-domain and time-frequency-domain features

| 特征  | 计算方式   | 特征   | 计算方式 |
|-----|--|--|------|
| 时域  | 均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$                      | 峭度因子 $K_r = \frac{\frac{1}{n} \sum_{i=1}^n ( x_i  - \bar{X})^4}{X_{rms}^4}$                |      |
|     | 均方根 $X_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$            | 峰值因子 $C = \frac{\max(X)}{X_{rms}}$   |      |
|     | 方差 $X_s = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$            | 脉冲因子 $I = \frac{X_{max}}{X}$   |      |
|     | 标准差 $X_{ss} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ | 波形因子 $S = \frac{X_{rms}}{X}$   |      |
| 频域  | 整流平均 $\bar{X}_r = \frac{1}{n} \sum_{i=1}^n  x_i $                | 偏度因子 $C_L = \frac{\max(X)}{\left  \sqrt{\frac{1}{n} \sum_{i=1}^n \sqrt{ x_i }} \right ^2}$ |      |
|     | 峰峰值 $X_{pp} = \max(x_i) - \min(x_i)$                             | 边缘因子 $M = \frac{\max(X)}{\left( \frac{1}{n} \sum_{i=1}^n  x_i  \right)^2}$                 |      |
|     | 平均频率 $f_1 = \frac{\sum_{k=1}^m y(k)}{m}$                         | 频率均方根 $f_2 = \sqrt{\frac{1}{m} \sum_{k=1}^m y(k)^2}$                                       |      |
|     | 频率峰值 $f_3 = \max(y(k))$  |  |      |
| 时频域 | IMF1 $I_1 = \sqrt{\frac{1}{n} \sum_{i=1}^n c_i^2}$               | IMF2 $I_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n c_i^2}$   |      |
|     | IMF3 $I_3 = \sqrt{\frac{1}{n} \sum_{i=1}^n c_i^2}$               |  |      |

1.3 敏感特征提取

1.3.1 互信息计算

互信息是两个随机变量之间相互依赖的度量,用来衡量两个变量的相互关联程度,表示两个变量之间含有相同信息的部分<sup>[23]</sup>。假设两个随机变量 **X** 和 **Y**,其各自的边缘概率密度函数和联合概率密度函数分别为  $p(x)$ 、 $p(y)$  和  $p(x,y)$ ,则它们的互信息值  $I$  如式(1)所示。

$$I(\mathbf{X};\mathbf{Y}) = \sum_{x_i \in \mathbf{X}, y_j \in \mathbf{Y}} p(x_i, y_j) \lg \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (1)$$

若 **X** 和 **Y** 的互信息值为 0,表示两个变量为无关变量或相互独立变量,即二者没有相互包含的信息;反之,互信息值越大,表示两个变量之间相互关

联的程度越高,包含的相同信息量也就越多。本文通过计算特征与剩余寿命之间的互信息值进行特征筛选,保留与剩余寿命关联度较高的特征。

1.3.2 主成分分析

主成分分析法是一种常用的数据降维方法,其基本原理是将具有一定相关性的原始变量经线性变化转换为一组互不相关的变量,且第一变量集中了原始变量所包含的主要信息<sup>[24]</sup>。基于 PCA 降维的转换公式如式(2)所示。

$$\mathbf{Z}_{m \times 1} = \mathbf{L}_{m \times s} \mathbf{X}_{s \times 1} = \begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1s}x_s \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2s}x_s \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{ms}x_s \end{cases} \quad (2)$$

式中,**X** 和 **Z** 分别为原始变量和转换后的变量,**L** 为荷载矩阵,即转换矩阵, $s$  为原始变量维数, $m$  为转换后的变量维数,且  $m \leq s$ 。本文首先基于互信息法筛选主要特征,然后利用 PCA 减少主要特征之间的冗余和重叠信息,从提取的多维特征中获取关键敏感特征,作为预测模型的输入。

1.4 支持向量回归预测模型

支持向量机(support vector machine, SVM)是一种用于解决小样本、非线性及高维模式识别下统计估计和预测问题的机器学习方法<sup>[25]</sup>。支持向量回归是 SVM 函数在回归拟合领域的应用。SVR 的基本思想是在高维特征空间上寻找一个最优超平面,能够使训练集样本距离该超平面最近,并将该超平面作为回归模型对未知数据进行预测。假设训练样本为  $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \cdots, (\mathbf{x}_n, \mathbf{y}_n)$ ,  $\mathbf{x}_i \in R^n$  为模型输入,  $\mathbf{y}_i \in R$  为模型期望输出,  $n$  为训练样本个数,则回归函数可表示为

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + \mathbf{b} \quad (3)$$

式中,**b** 为偏置向量; $\alpha_i$  和  $\alpha_i^*$  为拉格朗日算子; $K(\mathbf{x}_i, \mathbf{x})$  为核函数。核函数可以将低维空间的线性不可分问题转化为高维特征空间的线性可分问题。考虑到滚动轴承的退化趋势呈非线性,本文采用参数较少且非线性映射能力较强的高斯核函数解决 SVR 的回归预测问题。高斯核函数的表达式为

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2} \right) \quad (4)$$

式中, $\sigma$  为高斯核函数的指数,影响着 SVR 模型的预测精度。



1.5 模型评估指标

本文采用平均相对泛化误差  $E_{\text{ARG}}$  (average relative generalization error, ARGE)、均方根误差  $E_{\text{RMS}}$  (root mean square error, RMSE) 和相对误差  $E_{\text{R}}$  (relative error, RE) 来评估模型的回归预测能力, 计算公式分别如下。

$$E_{\text{ARG}} = \frac{1}{n} \sum_{k=1}^n \frac{|P_{\text{out},k} - E_{\text{out},k}|}{E_{\text{out},k}} \tag{5}$$

$$E_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{k=1}^n (P_{\text{out},k} - E_{\text{out},k})^2} \tag{6}$$

$$E_{\text{R},k} = \frac{|P_{\text{out},k} - E_{\text{out},k}|}{E_{\text{out},k}} \times 100\% \tag{7}$$

式中,  $P_{\text{out},k}$  为模型的预测输出,  $E_{\text{out},k}$  为模型的期望输出,  $n$  为模型输入样本个数,  $k$  表示第  $k$  个样本。

2 试验验证

2.1 试验数据

以 IEEE PHM12 轴承加速性能退化试验的全寿命数据<sup>[26]</sup>为例验证所提方法的有效性。试验数据来自如图 2 所示的 PRONOSTIA 试验台。通过加速度传感器采集轴承在水平和垂直两个方向上的振动信号, 每间隔 10 s 采集一次数据, 设置采样频率为 25.6 kHz, 每次采样 0.1 s, 因此每次可采集 2 560 个振动加速度。该数据集共包含 3 种工况下 17 个滚动轴承的全寿命周期试验数据, 工况划分如表 2 所示。由于轴承水平方向振动信号包含更多的状态信息<sup>[26]</sup>, 因此本文以工况 1 下轴承 1 水平方向的振动信号为研究对象, 构建基于 MI-PCA 结合 SVR 的滚动轴承剩余寿命预测模型。

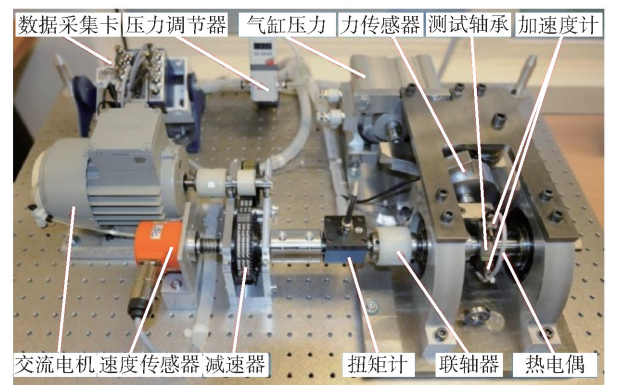


图 2 PRONOSTIA 试验台

Fig. 2 The PRONOSTIA test rig

2.2 数据降噪与敏感特征提取

首先, 采用小波包降噪算法处理轴承 1 的原始振动信号, 其原始振动信号及降噪后的振动信号如

图 3 所示。从图中可以看出降噪后的数据更加平滑, 剔除了异常数据点, 并保留了原始振动信号中的有效信息。

表 2 工况划分设置

| Table 2 The working condition division settings |                           |       |
|---|---------------------------|-------|
| 工况  | 转速/(r·min <sup>-1</sup> ) | 载荷/N  |
| 1   | 1 800                     | 4 000 |
| 2   | 1 650                     | 4 200 |
| 3   | 1 500                     | 5 000 |

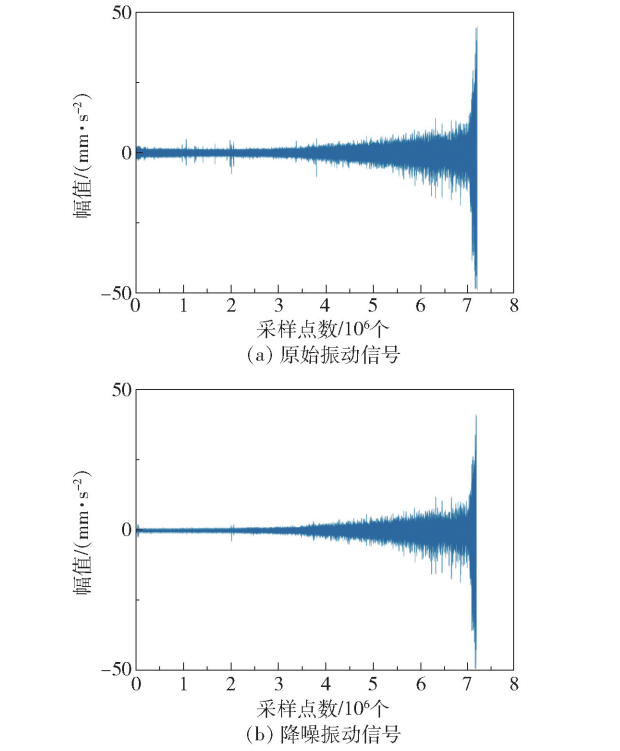


图 3 原始信号和小波包降噪信号的对比

Fig. 3 Comparison of raw signals and wavelet packet denoising signals

为了更好地突出显示小波包降噪的效果, 选取某一次采集的振动信号进行分析, 同时将时域振动信号转换为频域信号, 以便更直观地分析降噪效果。图 4 为小波包降噪前后的频域信号对比, 可以看出小波包降噪方法可以在一定程度上滤除噪声, 而且能够保留原始信号的有效信息, 即降噪后的频域信号含有峰值, 可表征轴承的磨损状况<sup>[27]</sup>, 从而可提取出对轴承退化敏感的特征。

图 5 为提取的 18 个时域、频域和时频域特征。从图 5 可以看出, 部分特征的退化趋势不明显, 无法表征轴承的退化状态, 因此首先采用互信息法筛选对轴承退化状态较为敏感的特征。表 3 为提取的特

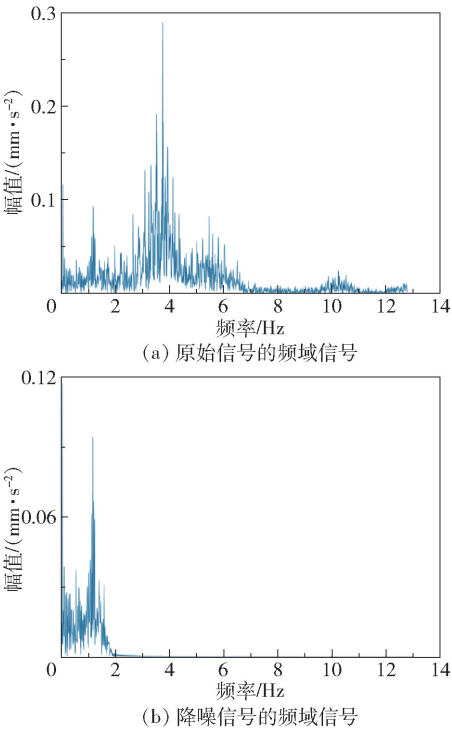


图4 原始信号和小波包降噪信号的频域信号对比

Fig.4 Frequency-domain signals comparison of raw signals and wavelet packet denoising signals

表3 特征与真实剩余寿命的互信息值

| Table 3 Mutual information values of features and true remaining useful life |      |          |      |          |      |
|--|------|----------|------|----------|------|
| 特征   | 互信息  | 特征       | 互信息  | 特征       | 互信息  |
| 均值   | 6.54 | 标准差      | 7.38 | 峰峰值      | 7.05 |
| 均方根  | 7.37 | 整流平均     | 7.51 | 方差       | 5.72 |
| 偏度因子   | 7.51 | 峭度因子     | 7.28 | 峰值因子     | 7.83 |
| 波形因子   | 7.56 | 脉冲因子     | 7.71 | 边缘因子     | 7.08 |
| 频率峰值   | 7.44 | 平均频率     | 7.00 | 频率均方根    | 7.37 |
| IMF1 有效值   | 7.21 | IMF2 有效值 | 7.20 | IMF3 有效值 | 7.33 |

征与真实剩余寿命的互信息计算结果,将互信息的平均值设置为阈值,保留互信息值大于阈值的特征。本文案例计算的阈值为 7.28,剔除 7 个对轴承退化状态不敏感的特征,然后基于保留的 11 个特征进行 PCA 降维,选取累积贡献率大于 90% 的第一主元变量作为敏感特征<sup>[28]</sup>,结果如图 6 所示。从图 6 可以看出,敏感特征可以很好地表征轴承的退化状态,即正常状态、轻微退化、严重退化和失效状态。

2.3 预测结果分析

将提取的敏感特征作为 SVR 模型的输入,将轴承的真实剩余寿命作为 SVR 模型的期望输出,对模

型进行训练和测试。选取 4 种退化状态中的各 50 个样本进行测试,为了验证基于 MI-PCA 的 SVR 模型预测结果的准确性,分别建立基于 MI 和基于 PCA 的 SVR 回归预测模型对相同的样本进行测试。

表 4 为 3 种预测模型的评估指标结果。从表 4 可以看出,3 种模型对前 3 种退化状态剩余寿命的预测误差较低,对失效状态的预测误差较大,然而 MI-PCA-SVR 模型的预测误差要远小于另外两种模型。总体来看,MI-SVR 模型、PCA-SVR 模型以及 MI-PCA-SVR 模型的平均预测准确率分别为 90%、93% 和 97%,MI-PCA-SVR 模型的预测准确率最高,预测误差最低,说明采用 MI-PCA 方法对输入样本进行降维处理可提高模型的预测精度。

表4 3种模型评估结果

Table 4 Evaluation results for three models

| 退化<br>状态 | MI-SVR  |          | PCA-SVR |          | MI-PCA-SVR |          |
|----------|---------|----------|---------|----------|------------|----------|
|          | ARGE    | RMSE     | ARGE    | RMSE     | ARGE       | RMSE     |
| 正常状态     | 0.025 5 | 63.855 6 | 0.005 7 | 12.882 0 | 0.002 4    | 5.726 3  |
| 轻微退化     | 0.028 0 | 34.323 5 | 0.052 7 | 55.936 4 | 0.029 7    | 31.671 9 |
| 严重退化     | 0.048 6 | 39.087 4 | 0.029 0 | 22.069 3 | 0.007 6    | 6.083 0  |
| 失效状态     | 0.276 9 | 15.414 5 | 0.177 5 | 10.876 8 | 0.072 3    | 5.877 6  |

图 7 为 3 种模型的相对误差曲线。从图 7 可以看出,MI-PCA-SVR 模型在严重退化和失效状态下的预测误差均小于 MI-SVR 模型和 PCA-SVR 模型。此外,MI-SVR 模型的相对误差曲线波动较大,其相对误差最高可达 50%,而 MI-PCA-SVR 模型的相对误差曲线较平稳,总体误差均在 13% 以内。结合表 4 的 RMSE 计算结果和图 7 的误差曲线,说明 MI-PCA-SVR 模型的稳定性要优于另外两种模型。基于以上分析,采用基于 MI-PCA 的 SVR 模型能够实现滚动轴承剩余寿命的精准预测。

2.4 IMS 轴承数据案例验证

为进一步验证所提方法的有效性,采用美国辛辛那提大学智能维护系统中心(Intelligent Maintenance System, IMS)的滚动轴承全寿命周期试验数据进行案例分析与验证。试验数据来自图 8 所示的试验台,通过加速度传感器采集轴承的振动信号,每间隔 10 min 采集一次数据,设置采样频率为 20 kHz,采样点数为 204 800。试验结束后共取得 3 个数据集。本文以数据集 2 中轴承 1 运行的全寿命周期加速疲劳振动数据作为研究对象,构建基于 MI-PCA 结合 SVR 的滚动轴承剩余寿命预测模型。

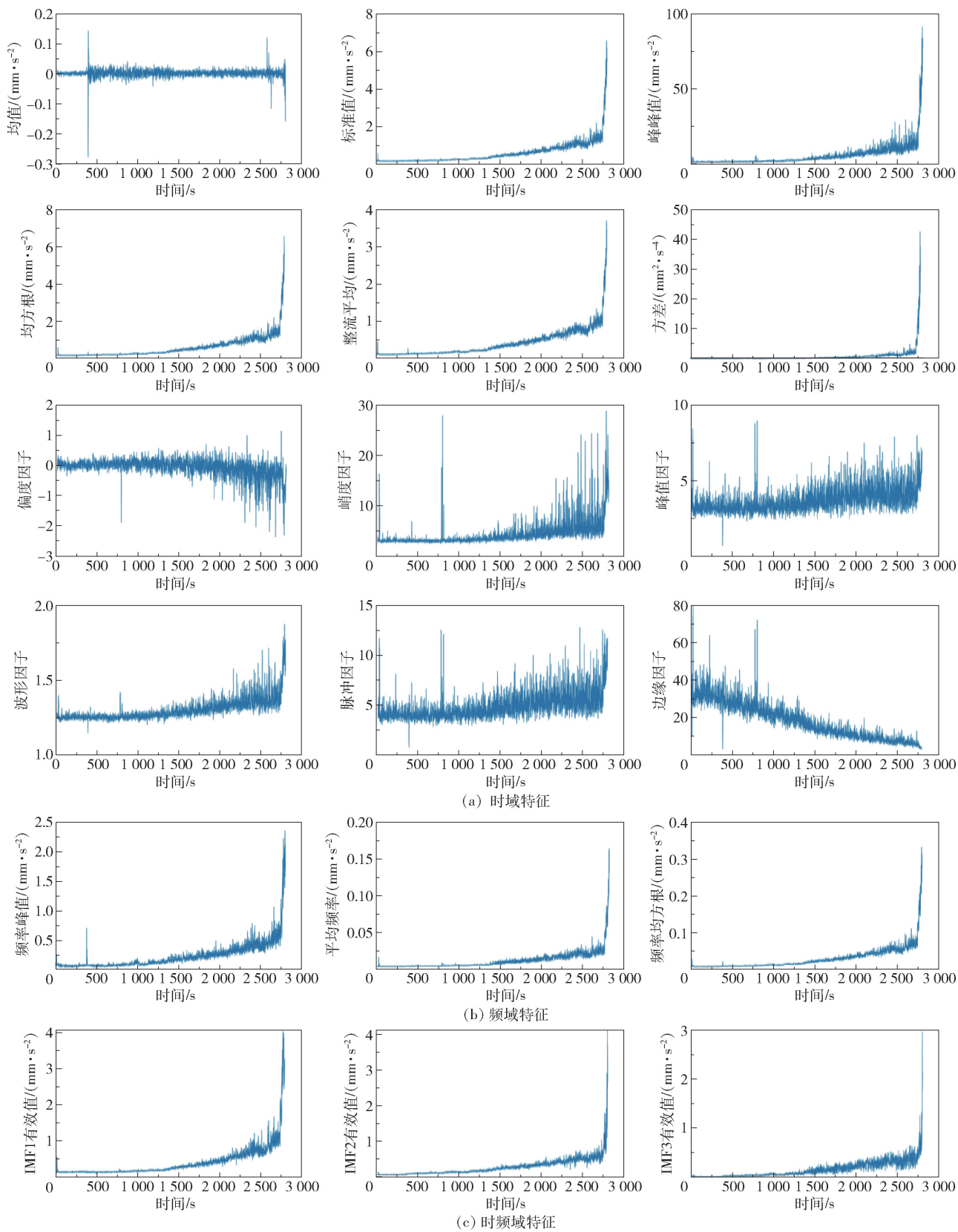


图 5 降噪信号的时域、频域和时频域特征

Fig. 5 Time-domain, frequency-domain and time-frequency-domain features of denoising signals

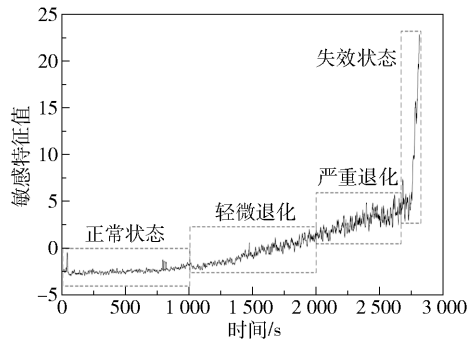


图 6 敏感特征  
Fig. 6 The sensitive feature

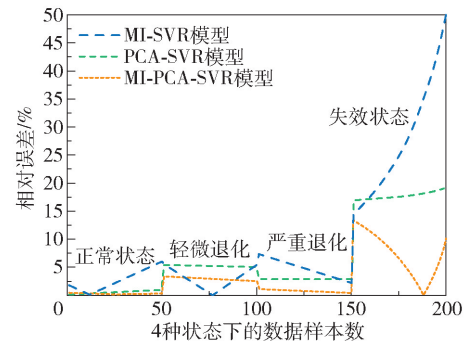


图 7 3 种模型的相对误差曲线  
Fig. 7 Relative error curves for the three models

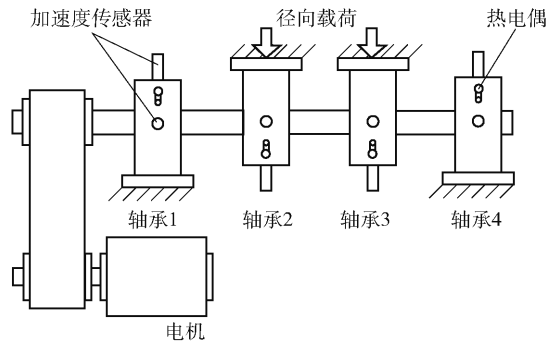


图 8 IMS 试验台  
Fig. 8 The IMS test rig

为体现本文所提方法的全局优势,分别使用 MI-SVR 模型、PCA-SVR 模型和 MI-PCA-SVR 模型预测轴承 1 的全过程剩余使用寿命。3 种方法的预测结果如图 9 所示。从图 9 可以看出,MI-SVR 模型和 PCA-SVR 模型在轴承的正常运行及初始退化阶段预测值和实际值之间的误差较大,随着时间的推移,在轴承的退化后期,误差逐渐缩小,而 MI-PCA-SVR 模型的整体预测误差均较小,预测值在实际值上下波动。此外,为了衡量模型预测结果的稳定性,计算得到 MI-SVR 模型、PCA-SVR 模型和 MI-PCA-SVR 模型的 RMSE 值分别为 14.922 7、20.065 2

和 2.609 0,表明本文所提方法的剩余寿命预测结果更加稳定。通过分析 3 种模型的预测结果曲线以及 RMSE 评估指标值可以得出,本文提出的 MI-PCA-SVR 模型的预测精度更高,而且预测结果的稳定性最优。

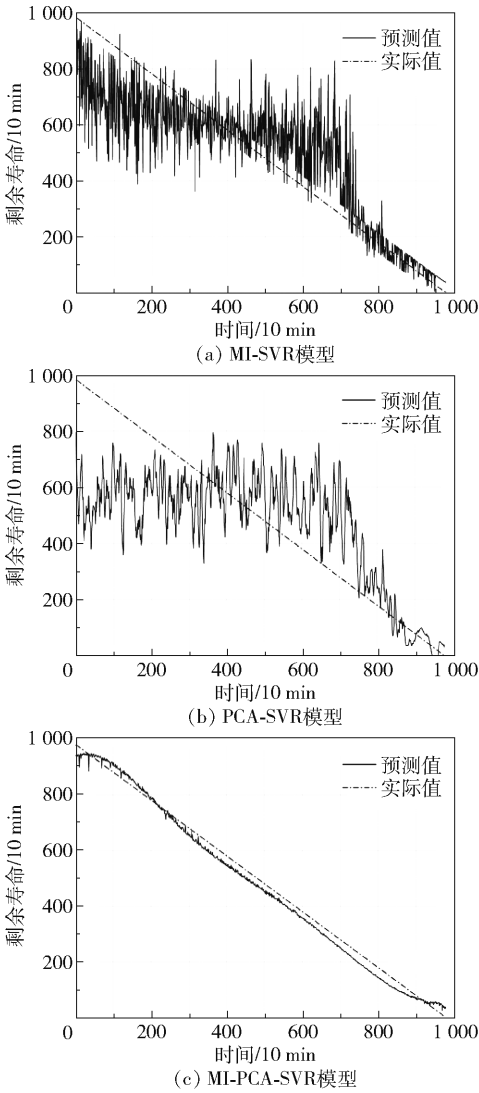


图 9 3 种模型的剩余寿命预测结果  
Fig. 9 Remaining useful life prediction results for the three models

### 3 结论

为了提高旋转机械设备的运转效率,实现设备的智能维护,本文提出了一种基于 MI-PCA 结合 SVR 模型的滚动轴承剩余寿命预测方法。首先利用小波包降噪算法消除原始振动信号中的噪声点对预测精度的影响,并基于降噪数据提取时域、频域和时频域特征。然后计算提取的特征与轴承真实剩余



寿命的互信息值进行特征筛选,进而通过 PCA 降维算法剔除筛选特征之间的冗余信息,获取表征轴承退化状态的敏感特征。最后以敏感特征作为模型的输入数据实现 SVR 模型的训练和测试。试验结果表明,基于 MI-PCA 的敏感特征提取方法可以充分挖掘振动信号中的状态信息,有效区分轴承退化过程中的不同状态;且基于 MI-PCA 构建的 SVR 模型可实现正常状态、轻微退化、严重退化和失效状态这 4 种退化状态下剩余寿命的精准预测,为开展及时有效的设备维护工作提供了有力依据。

符 号 说 明

- $x(t)$ —时域振动信号,  $\text{mm/s}^2$
- $y(t)$ —频域振动信号,  $\text{mm/s}^2$
- $c(t)$ —IMF 分量,  $\text{mm/s}^2$
- $\bar{X}$ —均值,  $\text{mm/s}^2$
- $X_{ss}$ —标准差,  $\text{mm/s}^2$
- $X_{pp}$ —峰峰值,  $\text{mm/s}^2$
- $X_{rms}$ —均方根,  $\text{mm/s}^2$
- $\bar{X}_r$ —整流平均,  $\text{mm/s}^2$
- $X_s$ —方差,  $\text{mm}^2/\text{s}^4$
- $C_L$ —偏度因子
- $K_r$ —峭度因子
- $C$ —峰值因子
- $S$ —波形因子
- $I$ —脉冲因子
- $M$ —边缘因子
- $f_1$ —平均频率,  $\text{mm/s}^2$
- $f_2$ —频率均方根,  $\text{mm/s}^2$
- $f_3$ —频率峰值,  $\text{mm/s}^2$
- $I_1$ —IMF1 有效值,  $\text{mm/s}^2$
- $I_2$ —IMF2 有效值,  $\text{mm/s}^2$
- $I_3$ —IMF3 有效值,  $\text{mm/s}^2$

参考文献：

[1] SHAO K X, FU W L, TAN J W, et al. Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing [J]. Measurement, 2021, 173: 108580.

[2] HU Z X, WANG Y, GE M F, et al. Data-driven fault diagnosis method based on compressed sensing and improved multi-scale network[J]. IEEE Transactions on Industrial Electronics, 2020, 67(4): 3216–3225.

[3] 王琇峰, 文俊. 基于噪声信号和改进 VMD 的滚动轴

承故障诊断[J]. 噪声与振动控制, 2021, 41(2): 118–124.

WANG X F, WEN J. Fault diagnosis of rolling bearings based on noise signal and improved VMD[J]. Noise and Vibration Control, 2021, 41(2): 118–124. (in Chinese)

[4] 朱丽娟. 机械设备运行可靠性评估的发展与思考[J]. 山东工业技术, 2014(23): 45.

ZHU L J. The development and thinking of mechanical equipment operation reliability evaluation[J]. Shandong Industrial Technology, 2014(23): 45. (in Chinese)

[5] LI Z H, TAO Y, ABU-SIADA A, et al. A new vibration testing platform for electronic current transformers [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68: 704–712.

[6] 王银刚. 基于振动数据分析的轴承典型故障预测研究[D]. 天津:中国民航大学, 2020.

WANG Y G. Research on typical failure prediction of bearing based on vibration data analysis [D]. Tianjin: Civil Aviation University of China, 2020. (in Chinese)

[7] BEALE C, NIEZRECKI C, INALPOLAT M. An adaptive wavelet packet denoising algorithm for enhanced active acoustic damage detection from wind turbine blades [J]. Mechanical Systems and Signal Processing, 2020, 142: 106754.

[8] 许志华, 潘庭龙. 基于变分模态分解的风机齿轮箱振动信号降噪方法研究[J]. 机电工程, 2021, 38(1): 129–132.

XU Z H, PAN T L. Noise reduction method of fan gearbox vibration signal based on variational mode decomposition [J]. Journal of Mechanical & Electrical Engineering, 2021, 38(1): 129–132. (in Chinese)

[9] 王大兴. 滚动轴承振动信号降噪及故障识别研究[D]. 兰州:兰州理工大学, 2020.

WANG D X. Research on de-noising and fault identification of rolling bearing vibration signal [D]. Lanzhou: Lanzhou University of Technology, 2020. (in Chinese)

[10] WANG F T, LIU X F, DENG G, et al. Remaining life prediction method for rolling bearing based on the long short-term memory network [J]. Neural Processing Letters, 2019, 50(2): 2437–2454.

[11] SHI M K, ZHAO R Z, WU Y C, et al. Fault diagnosis of rotor based on local-global balanced orthogonal discriminant projection[J]. Measurement, 2021, 168: 108320.

[12] 戴邵武, 陈强强, 丁宇. 基于时域特征的滚动轴承寿命预测[J]. 计算机测量与控制, 2019, 27(10): 60–63.



- DAI S W, CHEN Q Q, DING Y. Prediction for rolling bearing remaining life based on time domain features[J]. Computer Measurement & Control, 2019, 27(10): 60 – 63. (in Chinese)
- [13] 孟建军, 胡文涛. 基于互信息和 SVR 的滚动轴承剩余寿命预测[J]. 机械设计与研究, 2020, 36(6): 92 – 95.
- MENG J J, HU W T. Prediction of the remaining useful life of rolling bearings based on mutual information and SVR [J]. Machine Design and Research, 2020, 36(6): 92 – 95. (in Chinese)
- [14] 丁公博, 农振学, 王超, 等. 基于 MI-PCA 与 BP 神经网络的石羊河流域中长期径流预报[J]. 中国农村水利水电, 2019(10): 66 – 69.
- DING G B, NONG Z X, WANG C, et al. Long-term runoff forecasting model based on MI-PCA and BP neural network in Shiyang River Basin[J]. China Rural Water and Hydropower, 2019(10): 66 – 69. (in Chinese)
- [15] ELSHEIKH A, YACOUT S, OUALI M S. Bidirectional handshaking LSTM for remaining useful life prediction [J]. Neurocomputing, 2019, 323: 148 – 156.
- [16] CHEN Z, LI Y P, XIA T B, et al. Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy[J]. Reliability Engineering and System Safety, 2019, 184: 123 – 136.
- [17] HARIS M, HASAN M N, QIN S Y. Early and robust remaining useful life prediction of supercapacitors using BOHB optimized deep belief network[J]. Applied Energy, 2021, 286: 116541.
- [18] 马忠, 郭建胜, 顾涛勇, 等. 基于改进卷积神经网络的航空发动机剩余寿命预测[J]. 空军工程大学学报(自然科学版), 2020, 21(6): 19 – 25.
- MA Z, GUO J S, GU T Y, et al. A remaining useful life prediction for aero-engine based on improved convolution neural networks [J]. Journal of Air Force Engineering University (Natural Science Edition), 2020, 21(6): 19 – 25. (in Chinese)
- [19] KAYA D. Optimization of SVM parameters with hybrid CS-PSO algorithms for Parkinson's disease in LabVIEW environment [J]. Parkinson's Disease, 2019, 2019: 2513053.
- [20] SHI Y, GONG F R, WANG M Y, et al. A deep feature mining method of electronic nose sensor data for identifying beer olfactory information[J]. Journal of Food Engineering, 2019, 263: 437 – 445.
- [21] 杨旭, 邱明, 陈立海, 等. 基于 PSO-RWE 的自适应小波阈值函数滚动轴承振动信号去噪方法[J]. 航空动力学报, 2020, 35(11): 2339 – 2347.
- YANG X, QIU M, CHEN L H, et al. Adaptive wavelet threshold function based on PSO-RWE for vibration signal denoising of rolling bearing[J]. Journal of Aerospace Power, 2020, 35(11): 2339 – 2347. (in Chinese)
- [22] 陈长征, 魏巍. 基于改进 LMD 与小波包降噪对故障弱信号的提取[J]. 机械设计与制造, 2020(1): 165 – 168, 172.
- CHEN C Z, WEI W. Extraction of weak fault signal based on improved LMD and wavelet packet de-noising [J]. Machinery Design & Manufacture, 2020(1): 165 – 168, 172. (in Chinese)
- [23] KRASKOV A, STÖGBAUER H, GRASSBERGER P. Estimating mutual information [J]. Physical Review E, 2004, 69: 066138.
- [24] 余忠永, 黄俊, 许二敏, 等. 基于 PCA 降维的多特征行人再识别[J]. 信息通信, 2019(4): 13 – 16.
- YU Z Y, HUANG J, XU E M, et al. Multi-feature pedestrian re-identification based on PCA dimension reduction [J]. Information & Communications, 2019(4): 13 – 16. (in Chinese)
- [25] 陈芳, 程献宝, 黄安民, 等. 基于人工蜂群算法优化 SVM 的 NIR 杉木弹性模量预测[J]. 林业科学, 2021, 57(1): 161 – 168.
- CHEN F, CHENG X B, HUANG A M, et al. Elastic modulus prediction of *Cunninghamia lanceolata* based on artificial bee colony algorithm SVM and NIR[J]. Scientia Silvae Sinicae, 2021, 57(1): 161 – 168. (in Chinese)
- [26] SOUALHI A, MEDJAHHER K, ZEROUNI N. Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression[J]. IEEE Transactions on Instrumentation and Measurement, 2015, 64(1): 52 – 62.
- [27] 纪俊卿, 张亚靓, 孟祥川, 等. 自适应小波阈值滚动轴承故障振动信号降噪方法[J]. 哈尔滨理工大学学报, 2021, 26(2): 124 – 130.
- Ji J Q, ZHANG Y L, MENG X C, et al. Adaptive wavelet threshold rolling bearing fault vibration signal denoising method[J]. Journal of Harbin University of Science and Technology, 2021, 26(2): 124 – 130. (in Chinese)
- [28] 王佳瑜. 数据驱动的缓变微小故障早期诊断及剩余寿命预测[D]. 开封: 河南大学, 2017.
- WANG J Y. Data-driven early diagnosis and residual useful life prediction of slowly varying small fault [D]. Kaifeng: Henan University, 2017. (in Chinese)

# Remaining useful life prediction of rolling bearings based on mutual information based principal component analysis combined with support vector regression

OU BaiYu<sup>1</sup> YANG Meng<sup>2\*</sup> HAN Xu<sup>3</sup>

(1. Army Aviation Research Institute of Army Aviation Institute, Beijing 101121; 2. Department of Aviation Machinery Engineering of Army Aviation Institute, Beijing 101123; 3. 93146 Troops, Beijing 100000, China)

**Abstract:** As a key component in rotating machinery, rolling bearings affect its reliable operation. In order to carry out equipment maintenance intelligently and improve the operation efficiency of the equipment, a remaining useful life prediction method for rolling bearings using mutual information (MI) based principal component analysis (PCA) (MI-PCA) combined with support vector regression (SVR) is proposed. First, the wavelet packet denoising algorithm is used to eliminate the abnormal data points and noise in the original vibration signal, and the time-domain, frequency-domain and time-frequency-domain features are then extracted based on the denoising data. The feature selection is subsequently conducted according to the mutual information between the features and the remaining useful life, and the sensitive features representing the bearing degradation state are obtained using the PCA algorithm, which is used as the input of the SVR. Finally, the remaining useful life prediction model of the whole life test data of rolling bearings based on the SVR is established and trained. Compared with the SVR regression prediction models based on MI and PCA (MI-SVR model, PCA-SVR model), the test results show that the SVR model based on MI-PCA has higher accuracy in predicting the remaining useful life of rolling bearings. The prediction accuracy can reach 97%, which provides a basis for carrying out timely and effective equipment maintenance work.

**Key words:** rolling bearings; remaining useful life prediction; support vector regression; principal component analysis; mutual information

(责任编辑:吴万玲)