

DOI: 10.7652/xjtuxb201706001

主分量分析和隐马尔科夫模型结合的轴承监测诊断方法

张西宁, 雷威, 李兵

(西安交通大学机械制造系统工程国家重点实验室, 710049, 西安)

摘要: 为了快速识别轴承的故障模式以及性能退化状态,提出了一种基于主分量分析和隐马尔科夫模型的轴承监测诊断方法。该方法首先提取了轴承振动信号的混合域故障特征集,使用主分量分析对混合域故障特征集降维,然后使用降维后的特征训练隐马尔科夫模型,最后用降维后的测试样本测试模型的性能,根据隐马尔科夫模型输出的对数似然概率,确定轴承故障模式以及轴承的性能退化状态。开展了不同状态滚动轴承振动测试实验,数据分析结果表明,提出的方法诊断准确率均能达到100%,相比基于补偿距离选择特征降维及隐马尔科夫模型诊断方法,最高将分类离散度提高123.74%,并且在轴承的性能退化实验中,提出的方法能在故障早期给出故障预警,证明了该方法的有效性和准确性。

关键词: 混合域故障特征集;主分量分析;隐马尔科夫模型;轴承监测诊断

中图分类号: TH17 **文献标志码:** A **文章编号:** 0253-987X(2017)06-0001-07

Bearing Fault Detection and Diagnosis Method Based on Principal Component Analysis and Hidden Markov Model

ZHANG Xining, LEI Wei, LI Bing

(State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Aiming at accurately and rapidly recognizing bearing fault pattern and performance degradation, a bearing fault detection and diagnosis method based on principal component analysis and hidden Markov model is proposed. The mixed domain fault feature set of bearing vibration signal, which corresponds to different bearing conditions, is extracted with principal component analysis to reduce the dimension of the feature set, then hidden Markov models are trained with part of the reduced feature set. The performances of trained models are verified with the remaining parts in this feature set. The bearing fault patterns are recognized and bearing performance degradation is assessed by comparing the logarithmic likelihood probability value of the hidden Markov models. Experiments under different bearing conditions are carried out, vibration signals are collected, and the correct classification rate of the proposed method reaches 100%. Compared with the compensation distance evaluation based feature dimension reducing technique and hidden Markov model, the classification dispersion of the proposed method is increased by 123.74%. In the bearing performance degradation monitoring, the proposed method exhibits better effectiveness and accuracy for early bearing degradation warning.

Keywords: mixed domain fault feature set; principal component analysis; hidden Markov model; bearing fault detection and diagnosis

收稿日期: 2016-12-23。 作者简介: 张西宁(1965—),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(51275379);国家自然科学基金创新研究群体资助项目(51421004)。

网络出版时间: 2017-03-15

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20170315.1646.004.html>

滚动轴承是旋转机械中应用最为广泛的基础零部件,也是最易损坏的元件之一^[1]。据不完全统计,旋转机械的故障约有30%是由滚动轴承的失效引起。因此,轴承一直都是设备状态监测与故障诊断领域的热门研究对象^[2]。

由于轴承故障机理的复杂性,单一特征或单域特征难以全面刻画轴承运行状态,构建混合域故障特征集成为近年来轴承故障特征提取的发展趋势之一^[3]。然而,并不是特征维数越高故障诊断精度越高,由于提取的特征中有些是非敏感特征甚至是干扰特征,将严重影响诊断效果。因此,有必要对高维故障特征集进行降维。文献[4]使用补偿距离评估技术对故障特征集降维,建立隐马尔科夫模型,对轴承信号进行故障诊断。文献[5]提取时域、频域以及时频域的29维故障特征集,并利用近邻成分分析对故障特征进行降维,结合隐马尔科夫模型实现轴承的状态监测。文献[6]提取时域参数以及幅值谱熵作为特征集,并使用线性判别分析进行降维,结合隐马尔科夫模型进行轴承性能退化预测。

采用不同的特征提取方法提取出来的特征子集可能存在较大的差别,进而影响故障诊断结果。目前,对高维特征集进行降维的方法大致可以分为基于特征融合和基于特征选择两大类^[3]。基于特征选择的降维方法只保留少数特征,并没有消除特征间的相关关系。主分量分析作为一种基于特征融合的降维方法,不仅能够提取数据的有效信息,而且特征之间不存在冗余,在特征压缩与数据降维方面应用广泛。

本文提出基于主分量分析和隐马尔科夫模型的轴承监测诊断方法,提取了滚动轴承振动信号的时域、频域以及时频域的故障特征集,并使用主分量分析对故障特征集进行降维,结合隐马尔科夫模型,实现了对滚动轴承的故障诊断。对轴承的寿命数据建模,在轴承故障早期给出预警,验证了方法的有效性。

1 理论基础

1.1 特征提取

轴承的振动信号含有丰富的状态信息,当轴承性能退化时,其时域信号的幅值以及概率分布会随之变化,频域中不同频率成分的能量以及主能量谱峰位置同时发生变化,其时频结构中能量分布也会发生变化。本文构造了基于时域、频域以及时频域的故障特征集,时域指标包括均方值、方差、标准差、方根幅值、平均幅值、均方幅值、峰值、峰峰值、偏

斜度指标、峭度指标、波形指标、峰值指标、脉冲指标和裕度指标;频域定义了13个频谱的特征参数^[7]以及7个功率谱密度函数的特征参数^[8]。在时频域,本文提取了信号的连续小波尺度能量谱为

$$E(a) = \int_{-\infty}^{\infty} |W_f(a, b)|^2 db \quad (1)$$

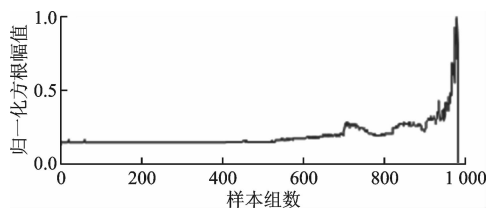
式中: $E(a)$ 为信号在尺度为 a 时的能量值,也称为小波能谱; $W_f(a, b)$ 为小波变换的幅值。

设 $E = \{E_1, E_2, \dots, E_n\}$ 为信号在 n 个尺度上的小波能谱, $\tilde{E} = \{\tilde{E}_1, \tilde{E}_2, \dots, \tilde{E}_n\}$ 为归一化的能量谱,其中

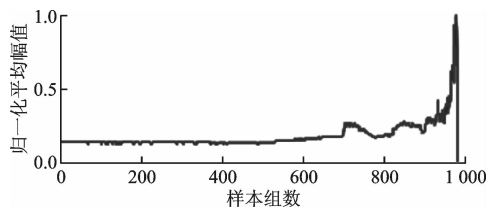
$$\tilde{E}_l = E_l / \sum_{i=1}^n E_i, \quad l \in [1, n] \quad (2)$$

连续小波能量谱在尺度空间上对信号能量进行划分,能够同时反映信号在时域与频域上的能量分布特征。

不同的指标能够表征的信息不一致,单一特征或单域特征难以全面刻画轴承运行状态,但并不是特征越多越好。在提取的特征指标中有部分非敏感特征甚至是干扰特征量,导致特征集包含大量的冗余信息。图1为从轴承性能退化数据中提取出的部分特征归一化的结果,图1a、图1b为振动信号的时域参数方根幅值和平均幅值归一化后的结果。由图可以看出,图1a与图1b随状态变化的趋势基本相同,两时域参数的线性相关系数为0.9988。这些冗余信息不仅对识别无用,而且会增加计算量以及模型的复杂度,某些干扰信息甚至会造成误判,因此很有必要对特征进行降维,提取出有效的特征子集。



(a) 归一化的方根幅值



(b) 归一化的平均幅值

图1 部分特征归一化后的结果

1.2 主分量分析和隐马尔科夫模型

目前,常用的降维方法有距离测度法、线性判别

分析、独立成分分析、主分量分析等。主分量分析(PCA)是一种基于特征融合的信息压缩和数据降维方法^[9]。

隐马尔科夫模型(HMM)是在 Markov 链的基础上发展起来的。由于实际问题比 Markov 链模型所描述的更为复杂,观察到的事件并不是与状态一一对应的,而是通过一组概率分布相联系,这样的模型称为 HMM。HMM 是一个双重随机过程:马尔科夫链和一般随机过程,其中马尔科夫链用于描述状态之间的转移,一般随机过程描述状态和观测变量之间的统计关系^[10]。

将 HMM 应用于工程实践中需要解决 3 个问题:评估问题、解码问题和学习问题。伴随着 3 个问题的解决逐渐形成了 3 个基本算法:前-后向算法、Viterbi 算法和 Baum-Welch 算法。

2 基于 PCA 和 HMM 的轴承监测诊断方法

根据上述分析,本文提出了基于 PCA 和 HMM 的滚动轴承监测诊断方法。该方法流程如图 2 所示,步骤如下:

- (1)首先提取滚动轴承训练信号故障特征,构造故障特征集;
- (2)使用 PCA 对提取出的特征集进行维数约简;
- (3)使用 Lloyds 算法^[11]对约简后的特征集进行标量化;
- (4)使用 Baum-Welch 算法训练各种故障模式下的 HMM,建立模型库;
- (5)将待诊断信号同样经过特征提取、PCA 降维以及标量量化的步骤处理之后,输入到训练好的

模型中,使用前-后向算法计算模型输出的对数似然概率,似然概率最大的模型对应待诊断轴承的故障模式。

如果只根据正常状态下的数据训练模型,将轴承的全寿命数据作为测试数据输入到模型中,输出的对数似然概率越大,说明数据是轴承在正常状态下产生的概率越大。当轴承的性能退化,即脱离正常状态时,该模型下输出的似然概率也会随之减小,从而反映滚动轴承所处的状态,实现轴承的状态监测,及时发现轴承的早期故障。

3 轴承的故障诊断实验

本次实验使用的实验台如图 3 所示。该轴承实验台包括直流电机、滚动轴承安装架、加载装置和滚动轴承等部分。实验轴承选用了 4 种状态的 6308 深沟球轴承,分别为正常、内圈剥落、外圈剥落和滚动体剥落。实验时,轴承内圈转速为 1 200 r/min,传感器为 IMI 的 601A11 加速度传感器,数采卡为 UA300 系列数据采集卡。采样频率为 10 kHz,采样时间持续为 10 s。

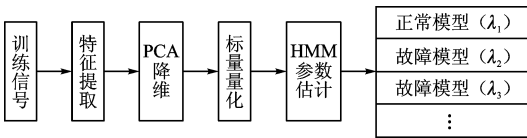


图 3 滚动轴承实验台

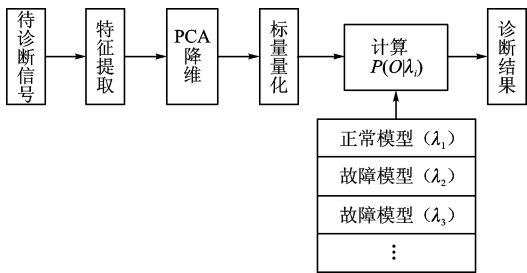
3.1 轴承故障特征提取以及降维

每种状态采集 10 组数据,共 40 组。在进行数据处理时,首先将每组数据分成 10 段,每段取 10 000 点,分别对每段数据提取上述混合域故障特征,获得混合域故障特征集。

使用 PCA 对混合域特征集进行降维处理,图 4 为主分量保存的信息量,其 3 维保存的信息量就达到 99.16%,因此取降维后的特征维数为 3。图 5 所示为经过 PCA 降维后的 3 维特征,其中任意两维之



(a) 训练信号处理流程



(b) 待诊断信号处理流程

图 2 基于 PCA 和 HMM 的监测诊断方法流程图

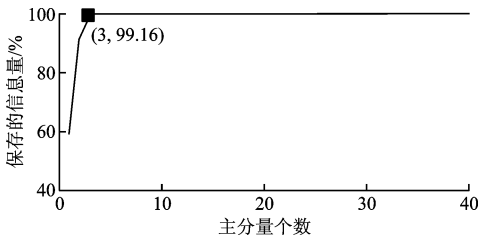


图 4 主分量保存的信息量

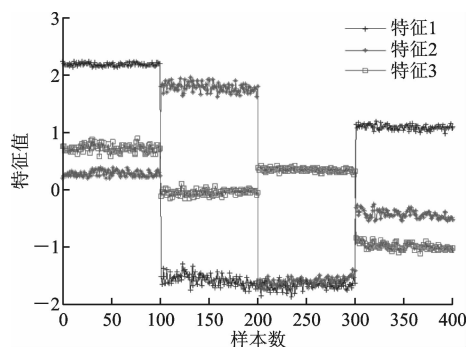


图5 PCA降维后的3维特征

间的相关系数均为 0,说明使用 PCA 降维方法所提取的特征不包含冗余信息。

按照补偿距离评估技术(CDET)^[4]计算各个归一化特征的评估指标值,同样从中选取指标值最大的 3 个归一化特征,见图 6。这 3 个敏感特征分别为归一化的峰值、归一化的平均频率、归一化的均方幅值,图 7 所示为 CDET 处理后的 3 维特征。其中,处理后的平均频率与处理后的峰值、处理后的均方幅值的相关系数分别为 $-0.874\ 3$ 、 $-0.803\ 0$,而处理后的峰值与处理后的均方幅值的相关系数高达 $0.973\ 7$ 。由此可见,使用 CDET 所提取的 3 个特征中包含有较多的冗余信息。

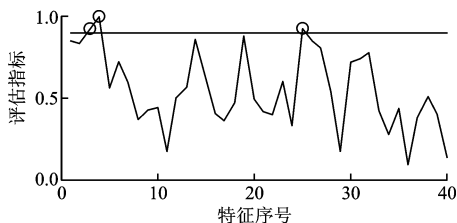


图6 各种归一化特征的距离评估指标

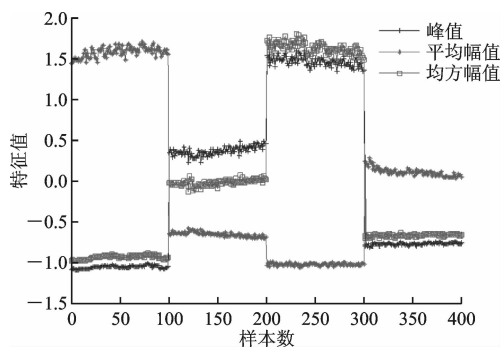


图7 CDET选择的3维特征

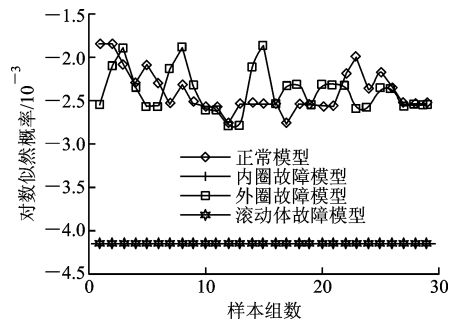
3.2 模型训练和概率输出结果

对降维后的特征进行标量量化,并选取每种状态的第 1 组作为训练数据,HMM 的状态数设为 4,量化级数设为 30,并将最大迭代步数设为 20。使用 4 种状态的训练数据,分别训练正常模型、内圈故障

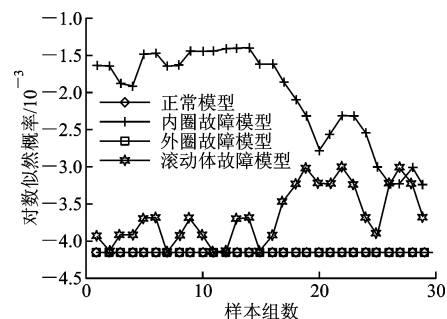
模型、外圈故障模型和滚动体故障模型。

为了验证 PCA-HMM 的诊断结果,本文分别使用 CDET 和 PCA 提取原始故障特征集的 3 维故障特征,结合 HMM 对轴承的测试数据进行诊断。

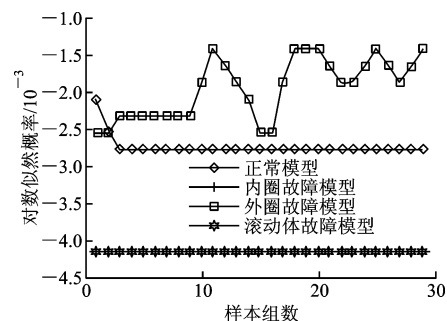
CDET-HMM 的诊断结果如图 8 所示,其中图 8a 为正常轴承数据的诊断结果,识别的准确率为



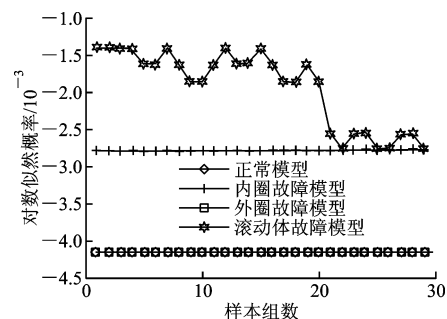
(a) 正常轴承



(b) 内圈故障轴承



(c) 外圈故障轴承



(d) 滚动体故障轴承

图8 CDET-HMM 的诊断结果

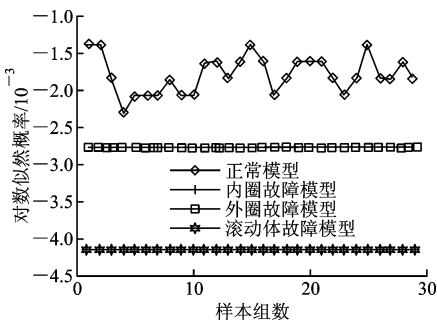
62.07%;图 8b 为内圈故障轴承数据的诊断结果,识别的准确率为 93.1%;图 8c 为外圈故障轴承数据的诊断结果,识别的准确率为 93.1%;图 8d 为滚动体故障轴承数据的诊断结果,识别的准确率为 100%。从图 8 中可以看到,滚动体故障轴承能被准确识别,正常、内圈、外圈故障的识别存在误判。

PCA-HMM 的诊断结果如图 9 所示,其中图 9a 为正常轴承数据的诊断结果;图 9b 为内圈故障轴承数据的诊断结果;图 9c 为外圈故障轴承数据的诊断结果;图 9d 为滚动体故障轴承数据的诊断结果。从图中可以看到诊断结果良好,4 种故障类型的轴承振动信号均能被有效地区分开,说明 PCA-HMM 模型在轴承故障诊断应用中具有良好的效果,并且 PCA-HMM 的区分度明显高于使用 CDET-HMM 的区分度。

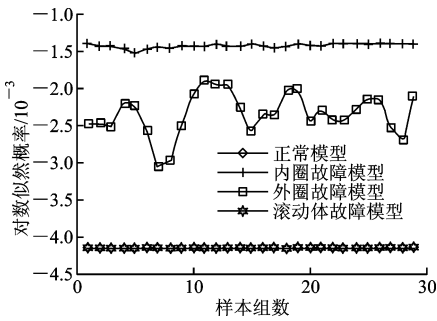
为充分对比实验结果,多次从距离评估指标前 10 的指标中随机选择 3 个特征,按照上述步骤训练、测试模型,此处不再赘述。实验结果显示,PCA-HMM 对滚动轴承故障识别效果最佳。

为了充分对比 CDET-HMM 与 PCA-HMM 的分类效果和可靠性,定义了描述分类可靠性的量——分类离散度^[12],即诊断结果中最大对数似然概率与次大对数似然概率之差除以最大对数似然概率与最小对数似然概率之差

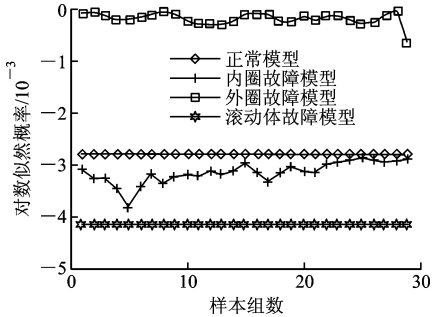
$$C_i = \frac{\lg P_{\max}(O|\lambda_i) - \lg P'_{\max}(O|\lambda_i)}{\lg P_{\max}(O|\lambda_i) - \lg P_{\min}(O|\lambda_i)} \times 100 \quad (3)$$



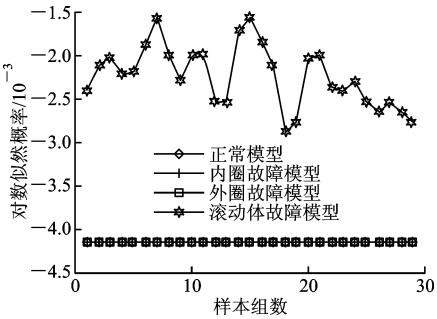
(a) 正常轴承



(b) 内圈故障轴承



(c) 外圈故障轴承



(d) 滚动体故障轴承

图 9 PCA-HMM 的诊断结果

式中: λ_i 为第 i 种模式下的 HMM 模型参数; $\lg P_{\max}(O|\lambda_i)$ 、 $\lg P'_{\max}(O|\lambda_i)$ 、 $\lg P_{\min}(O|\lambda_i)$ 分别为在该模型下的最大、次大、最小对数似然概率。

当分类离散度越接近 100,说明分类效果最好;分类离散度越接近 0,说明分类效果越差。现仅计算分类正确样本的分类离散度,表 1 为 4 种模式下 CDET-HMM 与 PCA-HMM 分类离散度平均值的对比,从表中可以看出 PCA-HMM 分类效果更好。

表 1 CDET-HMM 与 PCA-HMM 分类离散度平均值

状态	分类离散度		提高比例/%
	CDET-HMM	PCA-HMM	
正常	18.41	41.19	123.74
内圈故障	36.53	65.87	80.32
外圈故障	24.48	48.78	99.26
滚动体故障	73.38	100	36.28

4 轴承的状态监测实验

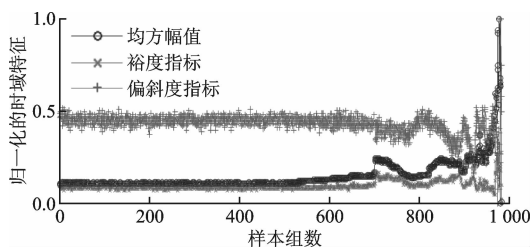
在工程实际中,采集各种状态下的轴承全寿命数据往往是难以实现的。如果仅使用正常状态下的轴承数据训练 HMM,并将轴承的全寿命数据作为测试数据,则可以根据输出的对数似然概率反映轴承偏离正常状态的程度。

采用美国辛辛那提大学智能维护系统中心(IMS

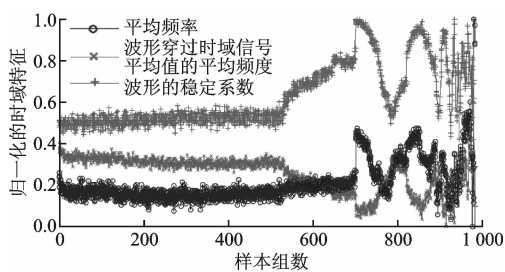
center)的轴承疲劳寿命实验数据^[13],实验轴承为莱克斯诺的 ZA-2115,传感器为 PCB 353B33 型加速度传感器,数采卡型号为 NI 的 6062E。实验中轴承转速为 2 000 r/min,径向载荷为 26.6 kN,采样频率为 20 480 Hz。实验过程中每隔 10 min 记录 1 次,每次采样持续时间为 1 s,最终轴承 1 出现外圈故障,共记录 984 组数据。

4.1 特征提取

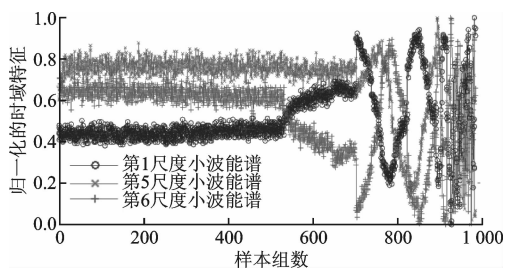
将上述的轴承全寿命数据按照前文所述的混合域故障特征集的构造方法,分别对每组数据提取时域、频域以及时频域故障特征,得到轴承故障特征集。由于轴承故障特征集的维数较高,分别用归一化后的时域特征、频域特征以及时频域特征 3 张图表示,如图 10a、图 10b、图 10c 所示,从图 10 可以观察到,随着样本组数增加,各条特征曲线均有一定的变化,但是变化趋势不相同,说明每个特征都包含一定的轴承状态变化信息,但所包含的信息各不相同。因此,需要对混合域故障特征集进行 PCA 降维提取有用信息,然后使用 HMM 进行轴承的状态监测。



(a) 时域特征



(b) 频域特征



(c) 时频域特征

图 10 轴承全寿命数据归一化特征

4.2 维数约简

使用 PCA 对混合域故障特征集降维,提取有用特征,剔除特征集的冗余信息。前 3 组主成分保存的信息量就达到 90.2%,前 6 组主成分保存的信息量超过 95%。因此,选取前 3 组作为主成分,对特征向量进行降维,得到降维后的 3 维特征向量,如图 11 所示。

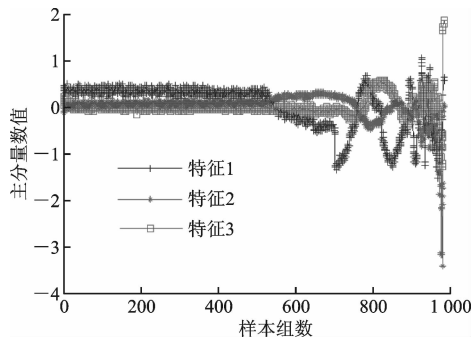


图 11 降维后的 3 维特征

4.3 模型训练和概率输出结果

对 PCA 降维后得到的特征向量进行标量量化。从 984 组降维后的数据中选取前 50 组作为正常状态的训练数据,训练 HMM。根据上述分析将状态数设为 4,量化级数为 50,并将最大迭代步数设为 20。得到 HMM 的参数后,使用前-后向算法计算观测序列的对数似然概率。当轴承处于同一退化阶段时,其观测序列的对数似然概率服从相同的分布,以某一值为中心上下波动。当轴承处于不同的退化阶段时,其分布不同,因此可以通过对数似然概率的变化体现状态的变化。

将数据输入到训练的正常模型中,得到图 12 所示的对数似然概率曲线。从图 12 中可以看到,在 1~533 组观测序列的对数似然概率在 -0.3 左右波动,说明在此阶段轴承处于同一状态,也就是正常状态。在 533~630 组时,观测序列的对数似然概率脱离正常值,并持续减小,说明轴承开始出现故障且开

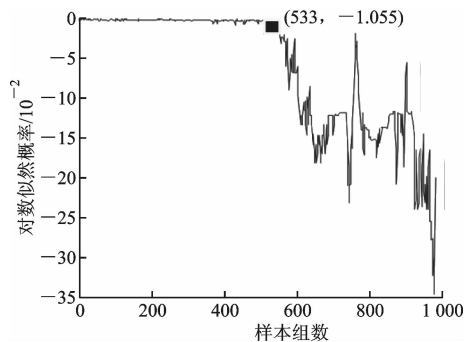


图 12 模型输出的对数似然概率

始加深。在 630~790 组时,观测序列的对数似然概率逐渐升高,轴承故障部位的尖锐部分在运行过程中逐渐被磨平。在 790 组之后的,观测序列的对数似然概率再次减小,并出现剧烈波动,轴承故障反复加深磨平直至失效。

通过上述分析可知,随着轴承性能的退化,观测值序列在正常 HMM 模型中输出的对数似然概率呈现减小的趋势。因此,可以设置某一值作为对数似然概率的阈值,当对数似然概率超过阈值时给出故障警报。本文使用拉依达法则确定阈值,检测异常点。当数据点与其集合的数学期望之差的绝对值大于 3 倍标准差时,即 $|x_i - \bar{x}| > 3\sigma$ 时,将该数据点判定为异常数据。为避免偶然误差,定义对数似然概率连续 5 次出现异常时,表示设备出现异常状态,给出故障警报。如图 13 所示,本文提出的方法在第 529 组数据时给出了数据异常的警报,说明对数似然概率已经连续 5 次超出阈值范围。本文方法对早期故障更加敏感,说明本文方法能在故障早期给出准确的故障发生时刻,从而给出故障预警。

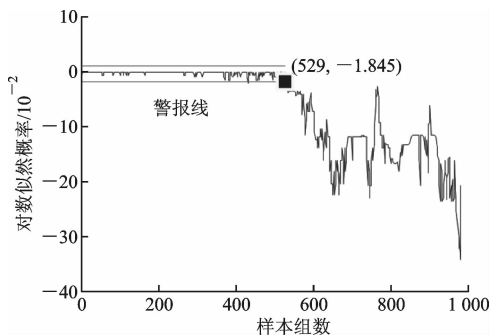


图 13 轴承性能退化报警线

5 结 论

本文提出基于 PCA 和 HMM 的滚动轴承监测诊断方法。首先提取滚动轴承的振动信号混合域故障特征集,并使用 PCA 对特征集进行维数约简,减小 HMM 模型训练时间和运算量。为验证所提方法的有效性和可靠性,在轴承故障诊断实验中与 CDET-HMM 作对比,并且对轴承的正常状态下的振动数据建模,对轴承的性能状态进行监测,得到结论如下。

(1)使用 PCA-HMM 对 4 种故障状态的滚动轴承进行故障诊断,其诊断准确率均能达到 100%,说明了本文方法的有效性。

(2)在与 CDET-HMM 对比实验中,PCA-HMM

分类效果优于 CDET-HMM,最高将分类离散度提高 123.74%。

(3)对轴承的正常数据建模,模型能够准确地给出早期故障预警,并且较好地刻画了轴承性能退化过程。实验结果验证了本文所提方法在轴承状态监测中的可行性和有效性。

参考文献:

- [1] 屈梁生,何正嘉. 机械故障诊断学 [M]. 上海: 上海科学技术出版社, 1986: 81.
- [2] PATIL M S, MATHEW J, RAJENDRAKUMAR P K. Bearing signature analysis as a medium for fault detection: a review [J]. Journal of Tribology, 2008, 130(1): 014001.
- [3] 苏祖强. 基于泛化流形学习的风电机组传动系统早期故障诊断方法研究 [D]. 重庆: 重庆大学, 2015: 17-18.
- [4] 蒋会明, 陈进, 董广明, 等. 基于集成隐马尔可夫模型的轴承故障诊断 [J]. 振动与冲击, 2014, 33(10): 92-96.
- JIANG Huiming, CHEN Jin, DONG Guangming, et al. Integrated HMM-based bearing fault diagnosis [J]. Journal of Vibration and Shock, 2014, 33(10): 92-96.
- [5] ZHOU H, CHEN J, DONG G, et al. Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model [J]. Mechanical Systems and Signal Processing, 2016, 66: 568-581.
- [6] LIU T, CHEN J, ZHOU X N, et al. Bearing performance degradation assessment using linear discriminant analysis and coupled HMM [J]. Journal of Physics: Conference Series, 2012, 364(1): 12028-12039.
- [7] 雷亚国, 何正嘉, 訾艳阳. 基于混合智能新模型的故障诊断 [J]. 机械工程学报, 2008, 44(7): 112-117.
- LEI Yaguo, HE Zhengjia, ZI Yanyang. Fault diagnosis based on novel hybrid intelligent model [J]. Chinese Journal of Mechanical Engineering, 2008, 44(7): 112-117.
- [8] CHEN P, TANIGUCHI M, TOYOTA T, et al. Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming [J]. Mechanical Systems & Signal Processing, 2005, 19(1): 175-194.
- [9] 屈梁生, 张西宁, 沈玉娣. 机械故障诊断理论与方法 [M]. 西安: 西安交通大学出版社, 2009: 170-172.

(下转第 109 页)