

DOI: 10.13382/j.jemi.B2002967

位置优化 Fisher 测度在轴承故障特征选择中的应用*

刘浩炜 刘 韬 涂文涛 陈 庆

(昆明理工大学 机电工程学院 昆明 650504)

摘要:为了提高滚动轴承故障诊断率,充分利用时域、频域及时频域特征对轴承运行状态的识别能力的差异性,并考虑到特征之间易出现不相关、冗余干扰等问题以及实际工程对简单、快速、有效的特征评估方法的需求,在构建轴承混合域特征集的基础上,提出了一种位置优化 Fisher 测度(POFDM)方法并将其应用于轴承故障特征选择。该方法基于 Fisher 准则,引入中值法通过多类样本的位置关系修正特征对状态分离聚合敏感程度的评估系数,从而筛选出能抑制状态间重合度的特征。此外,针对智能诊断模型确定最优特征集效率低的问题,提出了多维空间测度-Fisher 的特征集评估方法,通过计算不同维数候选特征集在多维空间中的距离测度指标,基于极大值原则筛选出最优特征集。最后,通过轴承故障实验对所提算法进行验证,实验结果表明,提出方法得到的最优低维特征集可以有效诊断轴承故障,在特征组合数为3时支持向量机分类器诊断正确率达到了99.17%。

关键词:混合域;位置关系修正;位置优化 Fisher 测度;多维空间距离测度-Fisher

中图分类号: TH165.3; TN911

文献标识码: A

国家标准学科分类代码: 510.40

Application of position optimized Fisher measure in bearing fault feature selection

Liu Haowei Liu Tao Tu Wentao Chen Qing

(Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650504, China)

Abstract: In order to improve the fault diagnosis rate of rolling bearings, make full use of the difference in the recognition ability of the bearing operating state by the time domain, frequency domain and frequency domain features, and take into account that the features are prone to irrelevance, redundant interference and other issues, as well as the simple, fast and effective feature evaluation of the actual project method needs. Position optimized Fisher distance measure (POFDM) method is proposed and applied to bearing fault characteristic select. The method is based on Fisher's criterion, and the positional relationship between multi-class samples is used to correct the evaluation coefficient by the median method, which could reflect sensitivity of the state separation and aggregation. Thus, the features that can suppress the degree of state coincidence are selected. In addition, aiming at the problem that the intelligent diagnosis model is inefficient in seeking optimal feature set, feature set evaluation method based on multi-dimensional spatial measure-Fisher is proposed. The optimal feature set is selected based on the maximum value principle by calculating the distance measure index of different dimension candidate feature sets in multidimensional space. Finally, the proposed algorithm is verified by the bearing fault experiment. The experimental results show that the optimal low-dimensional feature set obtained by the proposed method achieves 99.17% diagnostic accuracy of the SVM classifier when the number of feature combinations is 3, which can effectively diagnose bearing faults.

Keywords: mixed domain; position relationship correction; position optimized Fisher measure (POFDM); multi-dimensional spatial measure-Fisher

0 引言

机械设备的状态监测与故障诊断一直是研究热点,而轴承作为旋转机械设备中最重要的一类部件,一直是该领域的热门研究对象^[1]。在轴承的状态监测和故障诊断技术中,基于振动的监测是较为成熟和可靠的一种方法^[2],其中,如何基于采集到的振动信号提取有效的敏感特征则是故障诊断的关键。为了更有效地识别和诊断故障,包括有效值、峭度等时域统计特征,频率谱熵、幅值谱熵等频域特征,小波包能量等时频域特征都被广泛讨论并用于轴承的故障特征提取。这些特征从不同的角度反映信号对故障的识别能力,但是,维数过高的特征集中可能会有冗余特征与不相关特征,增加分类器的学习与计算复杂度,甚至降低分类准确率^[3]。因此,约减和筛选高维特征集,获取低维敏感特征集对于轴承故障状态的识别具有重大意义。

诸如主成分分析(PCA)^[4]、线性判别分析(LDA)^[5]等特征约减方法已被引入轴承故障诊断,但这些约减算法难以保持原始数据非线性流形的特点,而轴承的振动信号具有非线性^[6]。局部保留投影(LPP)虽然可以保持原始数据的非线性流形^[7],但在面对数据集的非线性及各类特征不明显等问题时,不能有利辨识^[3]。徐卫晓等^[8]基于 KPCA-BP 神经网络对轴承振动信号进行诊断,具有较好的诊断效果。虽然 LE(Laplacian Eigenmaps)、LPP 等流形学习方法以及基于核方法的 KPCA、K LDA 等这些特征约减方法得到的主要主成分能保留原始特征集的大部分信息,但考虑到实际工程的需要,对于高维特征集来说,特征选择方法更具实际意义。特征选择方法一般使用评价准则来增强特征与类的相关性,削减特征之间的相关性,一般将评价函数分成 4 类^[9-10],距离度量、依赖性度量、信息度量、以及一致性度量。距离度量因其具有简单、直观、高效的优势^[11]非常适合用于工程实际。Relief 系列算法属于典型的距离度量特征选择算法,近几年被广泛应用于机械故障诊断^[12-14],但 Relief 算法易忽略小类样本,不能排除冗余特征^[15-16]。

FDA(Fisher discriminant analysis)一般被用来线性分类,通过寻找最佳投影适量使得映射后同类样本的样本方差最小,不同类样本的样本均值之差最大^[17],基于这样的思想,基于 Fisher 准则的特征评估方法已经应用在语音识别^[18-19]等领域中。本文基于 Fisher 准则评估特征对于轴承状态类型的敏感程度,考虑到真实数据可能会有野点存在的情况,引入中值法^[20],通过多类样本位置关系修正评估系数。针对通过模型确定最优特征集维数时复杂且效率低的问题,提出了多维空间测度-Fisher 的特征集评估方法,并基于极大值筛选出最佳特征组合,旨

在为最优特征集的确定提供一种简单有效的解决思路。

1 混合域特征集的构造

不同特征对轴承故障状态识别的敏感程度不同,时域特征是一种直观的信号特征,可通过时域的统计指标来辨识轴承故障,本文主要引进常见的 14 个时域特征量构成 14 维的时域特征集。

时域指标提供的信息量有限,且对轴承故障部位、故障类型及损伤程度无法做出准确判断^[21],频域参数包含了大量有利于故障诊断的信息^[22],本文引进了 15 个频域特征量构成 15 维的频域特征集。时域、频域特征量如表 1 所示。

表 1 时域、频域特征集

Table 1 Feature set of time domain and frequency domain

时域特征	频域特征
均值、标准差	谱总值
有效值、峰峰值	频率方差
零峰值、绝对峰值	均值频率
平均幅值、方根幅值	幅值谱熵
歪度指标、峭度指标	频谱分散度(7个)
峰值指标、裕度指标	主频位置表征(2个)
脉冲指标、波形指标	重心频率、均方根频率

单一的时域、频域特征只能从整体上描述信号,而联合时频特征可以兼顾时域和频域上的信息,对信号进行局域性分析。小波包分析可以对信号的低频段和高频段部分进行多层次辨析,从不同频段对信号的局部特征进行描述,能够准确刻画各个状态信号的时频特性。轴承出现故障时,振动信号的某一频带的能量可能会增加,而另一频带的能量可能会减小,采用小波包能量^[23]的概念对不同频带的时间序列提取能量占比作为特征量,振动信号在各频带投影序列的能量占比定义为:

$$p_{Nj} = E_{Nj} / \sum_{j=1}^{2^N} E_{Nj}$$

$$E_{Nj} = \int |S_{Nj}(t)|^2 dt = \sum_{l=1}^M |d_l^{i,N}|^2 \quad j = 1, 2, \dots, 2^N \quad (1)$$

式中: E_{Nj} 为小波包分解第 N 层从低频到高频第 j 个频带成分重构后的时间序列信号的能量; $d_l^{i,N}$ 为第 N 层第 j 个时间序列信号 $S_{Nj}(t)$ 第 l 个离散点幅值; M 为时间序列信号 $S_{Nj}(t)$ 的离散点个数。

经验模态分解(EMD)基于信号的局部特征时间尺度,将信号分解为若干个基本模态函数(IMF),每个 IMF 分量包含了不同时间尺度的信息^[24],而且随信号本身的变化而变化,具有自适应性。在分解得到 IMF 分量的基础上,可以提取各个 IMF 分量的能量作为特征。信号分

解如下:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (2)$$

式中:IMF 分量 $c_i(t)$ 表示信号高频段到低频段的各个成分; $r_n(t)$ 表示残余分量。

2 基于 Fisher 准则的特征选择

2.1 位置优化 Fisher 测度方法 (position optimized Fisher measure, POFM)

FDA 算法在进行分类时,首先要计算类间散度矩阵 S_B 和类内散度矩阵 S_W ,如式(3),通过式(4)寻找最佳映射向量 w ,最终将原始特征样本映射到 w 而实现分类。

$$S_W = \sum_{c=1}^C \sum_{i=1}^{S_c} (x_i - \mu_c)(x_i - \mu_c)^T$$

$$S_B = \sum_{c=1}^C S_c(\mu_c - \mu)(\mu_c - \mu)^T \quad (3)$$

式中: C 为类别数; S_c 为第 c 类样本的样本数; x_i 为样本; μ_c 为第 c 类样本的样本均值; μ 为全类样本的样本均值。得到的 S_B 和 S_W 为一个矩阵。

$$\max f(w) = \frac{w^T S_B w}{w^T S_W w} \quad (4)$$

在特征敏感性评估过程中,没有必要去求映射向量 w ,它基于 FDA 方法的思想,通过类间散度矩阵 S_B 和类内散度矩阵 S_W 引入一个类可分度指标 r 来判断不同特征对不同类别的区分能力,类间散度越大,类内散度越小, r 越大,表明该特征对类分离度贡献越大,据此判断该特征对分类的影响程度,如式(5)。

$$r = \frac{S_B + S_W}{S_W} \quad (5)$$

设含有 C 个类别的特征集为:

$$\{m_{c,s,f}, c = 1, 2, \dots, C; s = 1, 2, \dots, S_c; f = 1, 2, \dots, J\} \quad (6)$$

式中: $m_{c,s,f}$ 为第 c 类第 f 维特征的第 s 个样本值; S_c 为第 c 类总的样本数; J 为特征数; C 为总类别数。位置优化 Fisher 距离测度对特征敏感性评估的实现过程如下,流程如图 1 所示:

1) 规整特征集,抑制不同特征量纲的影响:

$$m_{c,s,f} = \frac{m_{c,s,f} - \min(m_f)}{\max(m_f) - \min(m_f)} \quad (7)$$

式中: $\min(m_f)$ 和 $\max(m_f)$ 为第 f 维特征样本数据的最小值和最大值。

2) 计算第 f 维特征下第 c 类样本的类中心 μ_{cf} :

$$\mu_{cf} = \frac{1}{S_c} \sum_{s=1}^{S_c} m_{c,s,f} \quad (8)$$

FDA 在求类内散度时采用均值作为其类中心,表征

某一类样本的平均水平。

3) 根据 μ_{cf} 计算第 f 维特征对应的类内散度系数 μ_{fw}

$$\mu_{fw} = \sum_{c=1}^C \sum_{s=1}^{S_c} (m_{c,s,f} - \mu_{cf})^2 \quad (9)$$

由 μ_{fw} 可充分获取该特征下同类样本至类中心的紧密度信息。文献[19]采用类内样本至类中心距离的中值去衡量整个类别的类内距,不可避免的会丢失样本距离信息,相比之下,该方法更能全面描述类内紧密度水平。

4) 计算第 f 维特征全类的样本中心 μ_f :

$$\mu_f = \frac{1}{C} \frac{1}{S_c} \sum_{c=1}^C \sum_{s=1}^{S_c} m_{c,s,f} \quad (10)$$

同上,这里依然采用全类样本的均值表征该特征全类样本的平均水平。

5) 根据 μ_f 计算第 f 维特征对应的类间散度系数 μ_{fb} :

$$\mu_{fb} = \sum_{c=1}^C S_c(\mu_{cf} - \mu_f)^2 \quad (11)$$

μ_{fb} 衡量的是该特征下各类类中心至全类中心的距离水平。计算过程中乘以各类特征数目可以消除不同类样本数目不同的影响。因此,该特征评估算法适用于不平衡样本。

6) 计算第 f 维特征第 c 类样本的等效半尺度 Rad_{cf} :

$$Rad_{cf} = med(\{|m_{c,s,f}^n - med(m_{c,s,f})|\}) \quad (12)$$

式中: $n=1, 2, \dots, N$ 。 med 表示取中值,中值法可以有效避免野点干扰,增强算法抗干扰能力^[19]。该方法中,需要计算每一个样本至其样本中值的距离,选取前 N 个最远点,并取前 N 个最远距离的中值作为第 f 特征下第 c 类样本的等效半尺度 Rad_{cf} 。

N 的选取由野点数目 y 来决定:

$$N = 2 \times y + 1 \quad (13)$$

因为一般情况下真实数据的野点数目未知,若选取的最远点数目等于野点数目,则会导致等效半尺度过长,故使用式(13),并取前 N 个最远距离的中值作为等效半尺度(式(12)),可更好的规避野点误选带来的风险,也更贴近该特征样本下的有效半尺度。

7) 计算第 f 维特征的位置修正系数 R_f

定义 $R_{ce,f}$ 为第 f 维特征中第 c 类和第 e 类的位置关系因子,描述如下:

$$R_{ce,f} = |med(m_{cf}) - med(m_{ef})| - (Rad_{cf} + Rad_{ef}) \quad (14)$$

式中: $c, e=1, 2, \dots, C, c \neq e$ 。此位置关系因子描述的是两类样本的分离、相交关系,若 $R_{ce,f} > 0$,则说明两类样本完全可分,且该值越大,分离程度越高。若 $R_{ce,f} \leq 0$,则说明两类样本有重合,且该值越小,分离程度越低。计算第 f 维特征下任意两类的位置关系因子,并求其均值,作为第 f 维特征的位置修正系数,其数学描述如下:

$$R_f = \frac{\sum_{c,e=1,c \neq e}^{C \times (C-1)} R_{ce,f}}{C \times (C-1)} \quad (15)$$

将获取的位置修正系数进行规整,方便后续计算,如式(16)。

$$R_f^* = \frac{R_f}{\max(R_f)} \quad (16)$$

8) 计算第 f 维特征的特征敏感性评估系数 P_f :

$$P_f = R_f^* \times \frac{\mu_{fb}}{\mu_{fi}} \quad (17)$$

其中,类间散度和类内散度指标包含了全部的样本距离信息,为了降低样本数据野点的干扰,本文考量了每一维特征下不同类别的位置关系,并用其修正特征评估系数。

9) 归一化处理:

$$P_f^* = \frac{P_f}{\max(P_f)} \quad (18)$$

将特征评估系数进行归一化处理, P_f^* 越大,该特征对状态分类越敏感。

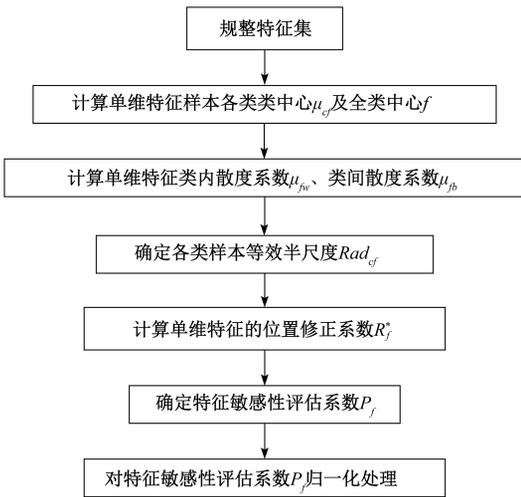


图 1 位置优化 Fisher 距离测度特征敏感性评估算法流程

Fig. 1 Flowchart of position optimized Fisher distance measure feature sensitivity evaluation algorithm

2.2 多维空间测度-Fisher 的特征集评估方法 (multidimensional space measure-Fisher,MSMF)

MSMF 的特征集评估方法实现过程如下。

1) 选择评估系数最大值对应的特征作为初始特征集 S_1 。

2) 以特征数逐步增 1 的方式将其余特征按其评估系数大小(递减排序)逐一加入前一个特征集构造 $S_i (i=2, \dots, J)$ 等 $J-1$ 个候选特征集,即保证候选特征集维度大于 1。记候选特征集 S_i 的特征样本为:

$$\{m_{c,s,j}^i, c=1,2,\dots,C; s=1,2,\dots,S_c; j=1,2,\dots,J\} \quad (19)$$

式中: $m_{c,s,j}^i$ 为候选特征集 S_i 第 c 类第 j 维的第 s 个样本值。

3) 计算候选特征集 S_i 的类内散度系数 μ_{wi} :

$$\mu_{cj}^i = \frac{1}{S_c} \times \sum_{s=1}^{S_c} m_{c,s,j}^i \quad (20)$$

$$\mu_{wi} = \sum_{c=1}^C \sum_{s=1}^{S_c} \sqrt{\sum_{j=1}^i (m_{c,s,j}^i - \mu_{cj}^i)^2} \quad (21)$$

式(20)依然采用均值 $\mu_{cj}^i (j=1, \dots, i)$ 作为候选特征集 S_i 第 c 类样本第 j 维的类中心。候选特征集 S_i 中每一个样本的特征维数为 i ,样本类中心也是在 i 维空间进行度量,式(21)通过计算样本至其类中心在 i 维空间中的欧氏距离 μ_{wi} 表征 i 维空间中特征样本的紧密程度, μ_{wi} 越小,表明 i 个特征的综合聚类效果越好。

4) 计算候选特征集 S_i 的类间散度系数 μ_{bi} :

$$\mu_j^i = \frac{1}{C} \frac{1}{S_c} \sum_{c=1}^C \sum_{s=1}^{S_c} m_{c,s,j}^i \quad (22)$$

$$\mu_{bi} = \sum_{c=1}^C S_c \sqrt{\sum_{j=1}^i (\mu_{cj}^i - \mu_j^i)^2} \quad (23)$$

式中: μ_j^i 表示候选特征集 S_i 第 j 维特征的全类中心,同样采用各类中心至全类中心的欧氏距离 μ_{bi} 度量 i 维空间中不同类特征样本的分离程度(式(23)), μ_{bi} 越大,表明 i 个特征对不同状态的可分性越好。

5) 计算候选特征集 S_i 的特征集评估系数 R_i

$$R_i = \frac{\mu_{bi}}{\mu_{wi}} \quad i=2, \dots, J \quad (24)$$

式中: R_i 的值越大,表明候选特征集 S_i 对状态的分离聚合越敏感。

6) 确定最优特征集

$$\text{extre} \{R_i\} \quad i=2, \dots, J \quad (25)$$

式中: extre 表示取特征集评估系数 R_i 极大值。同样地,为了降低野点数据存在带来的特征集评估不准确的风险,取 R_i 极大值对应的候选特征集作为最优特征集 S^* 的参考。

3 实验与分析

3.1 轴承故障实验

为了验证算法的有效性,设计了如下实验:实验装置如图 2 所示,内圈安装在转轴上,被转子驱动设备驱动,转速为 720 r/min,外圈被夹装装置固定,加速度传感器安装在轴承 x, y 方向,获取的振动信号由 NI9234 采集卡采集,采样率为 65 536 Hz。试验轴承采用 GB6203 滚动轴承,具体参数如表 2 所示。本文对试验轴承的外圈、内

圈、滚动体使用电火花加工故障,故障尺寸为 $0.5\text{ mm} \times 4\text{ mm} \times 0.5\text{ mm}$,得到正常轴承、外圈故障轴承、内圈故障轴承、滚动体故障轴承每种轴承状态的数据各 3 组,各个状态的轴承时域波形如图 3 所示。



图 2 轴承故障实验设备安装图

Fig. 2 Installation drawing of bearing fault test equipment

表 2 试验轴承参数

Table 2 Parameters test bearings

型号	内径/ mm	节径/ mm	滚动体 数目	滚动体直 径/mm	接触 角/(°)
GB6203	17	28.5	8	6.747	0

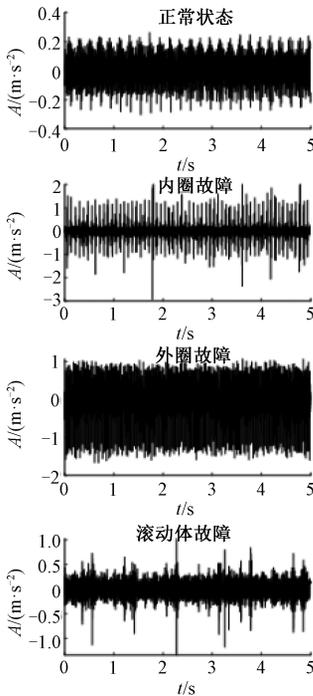


图 3 轴承各状态时域波形

Fig. 3 The time-domain waveform of bearing different states

3.2 特征敏感性评估

首先根据第 1 节从时域、频域、时频域进行混合域的

特征集构造。从 4 类原始数据中分别抽取 100 个样本,共得到 400 个样本。基于 db20 小波对每个样本进行小波包分解,分解层数为 3,得到 8 段分信号的能量占比,分别为小波包能量谱指标 1~8;将每个样本进行 EMD 分解,考虑到轴承故障信息主要包含在高频带,取信号前 7 个 IMF 分量,取其能量占比作为特征,分别为 EMD 分量能量谱指标 1~7;从时域、频域中分别提取表 1 所提特征,共 29 个。因此,3 个批次试验构造的特征样本集维度为 44×400 。

随后,按 1.2 节提出的 POFM 对特征敏感性进行评估。经试验观察,野点数目大致在 10 个以内,为得到有效的等效半尺度,根据式(13),将最远点数目设置为 21,各特征评估系数结果如图 4 所示。

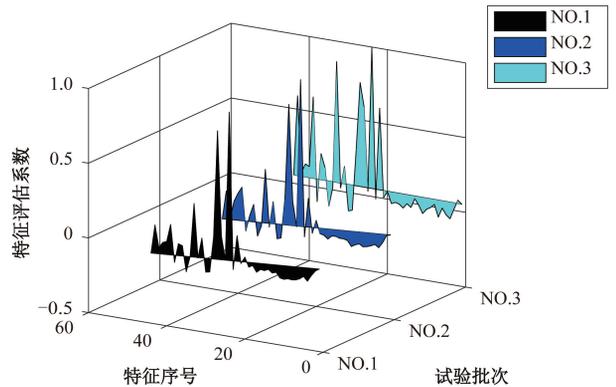


图 4 试验特征评估结果

Fig. 4 Feature evaluation result graph

由图 4 可见,特征评估系数在这三次试验中大致相近,其中特征评估系数越高,表明该特征越有利于轴承状态的分类,反之表明该特征对分类越不敏感。随后,将所有批次试验得到的特征评估系数取平均,作为 3.3 节构造候选特征集的参考标准。按评估系数从大到小顺序对原始特征集中的特征进行排序,1~8 分别为频谱分散度指标 1、谱总值、主频位置指标 1、频谱分散度指标 2、均值频率、有效值、标准差、频谱分散度指标 7,9~16 为频谱分散度指标 3,绝对峰值、平均幅值、零峰值、峰峰值、裕度指标、脉冲指标、峰值指标,17~24 为方根幅值、波形指标、重心频率、均方根频率、主频位置指标 2、小波包能量谱指标 7、峭度,小波包能量谱指标 3,25~32 为 EMD 分量能量谱指标 1、小波包能量谱指标 8、小波包能量谱指标 1、小波包能量谱指标 5、均值、小波包能量谱指标 4、歪度指标、幅值谱熵,33~40 为小波包能量谱指标 2、EMD 分量能量谱指标 3、小波包能量谱指标 6、EMD 分量能量谱指标 2、频谱分散度指标 4、EMD 分量能量谱指标 5、EMD 分量能量谱指标 4、EMD 分量能量谱指标 7、41~44 为 EMD 分量能量谱指标 6、频谱分散度指标 6、频谱分散度指标 5、频率方差。

为了使评估效果更直观,本文分别应用了 PCA、KPCA、ReliefF 等方法对原始特征样本集进行降维处理,ReliefF 和 POFM 方法选取特征评估系数最大的两个特征,PCA 和 KPCA 方法选取第 1 主成分和第 2 主成分,测试样本降维效果如图 5(a)~(d)所示,其中○表示内圈故障,◇表示滚动体故障,▽表示外圈故障,☆表示正常状态。从图 5 可以看出,PCA 方法对轴承 4 种状态的评估中滚动体故障和内圈故障状态有交叉部分,ReliefF 方法不能有效区分冗余特征,故两个特征表现出近似的直线关系,且轴承 3 种故障状态皆存在交叉部分,KPCA 方法与 POFM 方法皆可以对轴承的 4 种状态做出有效区分,其中两种方法在轴承正常状态、滚动体故障、外圈故障下的类内散度表现相差不多,但对于内圈故障的类内散度表现来说,POFM 方法要优于 KPCA。

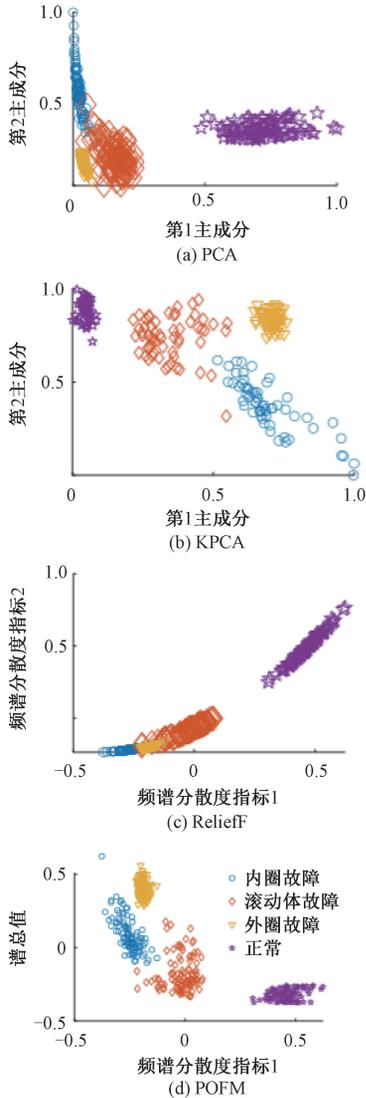


图 5 基于不同方法的降维效果

Fig. 5 Dimensionality reduction effect diagram based on different methods

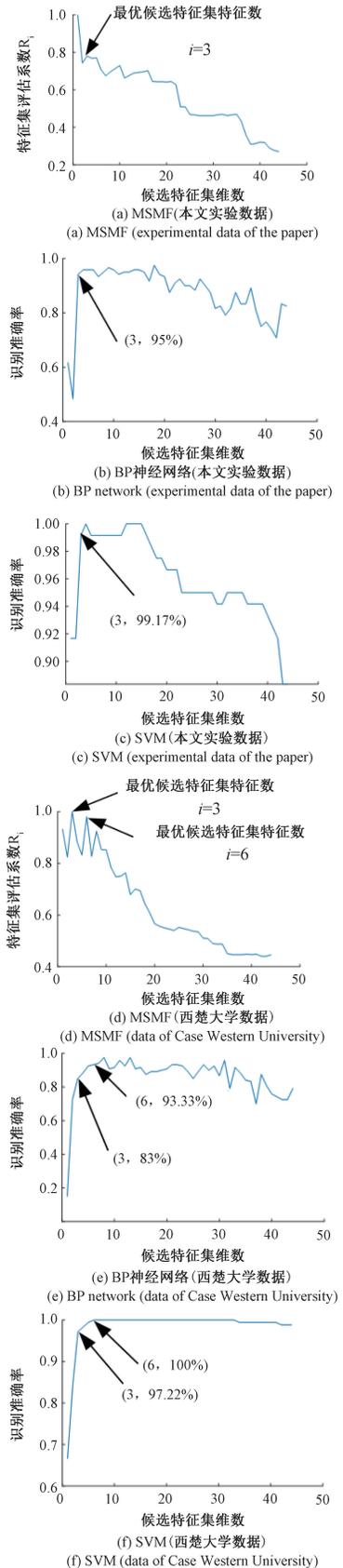


图 6 候选特征集评估结果及模型验证

Fig. 6 Feature set evaluation result and model verification graph

3.3 特征集评估及模型对比验证

特征敏感性的评估可以比较单一特征的判别能力,但真正的关键是判断特征组合对轴承状态的联合识别能力进而找出最优特征集,该特征集应满足保证较高识别准确率的同时包含特征数目最少。

按 2.2 节所述办法进行候选特征集的构造,并采用多维空间距离测度-Fisher 的特征集评估方法对候选特征集评估。为了证明本文所提方法的有效性,采用常见的模型确定最优集的方案对比验证,该方案步骤如下:

1) 构造候选特征集,同上。

2) 数据划分,为了避免人为设定训练样本和测试样本盲目乐观的情况发生,对原始特征样本集进行三七折划分,即每次试验时采取随机抽样的方式任意抽取 70% 的样本作为训练集,其余样本作为测试集。

3) 按照 MSMF 方法对不同维度的特征组合进行综合评价,评估结果如图 6(a) 所示。

4) 模型验证,采用简单实用的 BP 神经网络对按照 2.2 节构造出的多个候选特征集进行分类识别精度验证。模型包含输入层,隐含层,输出层 3 层。输入层节点数为候选特征集包含的特征个数 n ,依据经验,隐含层节点数为 $2n+1$ 。隐含层采用 sigmoid 传递函数,输出层采用 purelin 传递函数,动量因子设为 0.9,学习率为 0.01。为了增加模型识别精度的可靠性,取 10 次试验识别精度的平均值作为各个候选特征集最终的识别精度值,模型识别准确率随特征数目的变化情况如图 6(b) 所示。为了使 MSMF 方法更具说服力,避免单一模型带来的偶然性问题,本文使用支持向量机(SVM)模型对构造出的候选特征集进行进一步的分类识别验证。本文在 python 中使用 sklearn 库,SVM 模型采用一对一策略,核函数为高斯核,依照经验参数 c 和 g 分别设为 2 和 10,模型识别准确率随候选特征集维数的变化情况如图 6(c) 所示。

为了证明所提方法的适用性,采用凯斯西楚大学轴承数据中心提供的轴承实验数据进行验证,对故障尺寸为 0.177 8 mm 的内圈故障、0.177 8 mm 的外圈故障、0.177 8 mm 的滚动体故障及故障尺寸为 0.533 4 mm 的内圈故障、0.533 4 mm 的外圈故障及正常轴承的振动信号提取特征,同样采用 BP 神经网络和 SVM 模型进行验证,模型设置同上。采用多维空间测度-Fisher 的特征集评估方法对候选特征集评估结果如图 6(d) 所示,BP 模型和 SVM 模型验证结果如图 6(e) 和(f) 所示。

轴承状态的识别准确率与输入特征、选取的模型及模型参数都有很大关系。从图 6(b)、(c)、(e)、(f) 看出,对于本文实验数据(4 种轴承状态)和西楚大学数据(6 种轴承状态),BP 神经网络和 SVM 模型都表现出识别准确率不随特征数目的增加而增加,甚至因为特征数目的增加而下降。因此,寻找可以保证较高识别准确率

且特征数目较少的最优特征集具有实际意义。针对本文设计实验 4 种状态的数据(内圈、外圈、滚动体故障尺寸相同),使用 MSMF 方法得到多维特征集 S_i 的特征集评估系数 R_i ,基于 R_i 极大值选出最优特征集 S_3 ,其特征维度为 3(图 6(a)),包含由 3.2 节 POFM 方法得到的评估系数排名前三的特征,分别为频谱分散度指标 1、谱总值、主频位置指标 1,将这 3 个特征输入模型,BP 神经网络模型平均识别准确率达到 95%(图 6(b)),SVM 模型平均识别准确率达到 99.17%(图 6(c)),和最高识别准确率相差不大,说明 MSMF 方法在寻找最优特征组合具有有效性。针对西楚大学轴承数据(包含内圈轻微、严重故障,外圈轻微、严重故障等 6 种状态),MSMF 方法基于极大值筛选出最优特征集的参考特征集 S_3 和 S_6 (图 6(d)),分别包含由 POFM 方法得到的评估系数排名前三和前六的特征。图 6(e)中,将候选特征集 S_3 、 S_6 输入 BP 神经网络模型时,其平均识别准确率达到 83%、93.33%。图 6(f)中,将候选特征集 S_3 、 S_6 输入 SVM 模型时,平均识别准确率达到 97.22%、100%。可以看出,使用 BP 神经网络和 SVM 对类别数更多的轴承数据进行识别诊断,由 MSMF 方法确定的最优候选特征集可以满足最优特征组合在保证较高识别率的同时包含特征数目较少的要求,这进一步说明了在类别数目较多、样本数据复杂情况下 MSMF 方法寻找最优特征组合的有效性和鲁棒性,在工程实际中可以直接选择 MSMF 方法确定的最优候选特征集对故障轴承进行诊断。

4 结 论

本文提出了位置优化 Fisher 测度特征敏感性评估算法,解决了由于真实数据存在野点导致评估不准确进而影响轴承故障识别精度的问题,为工程实际提供了一种简单有效的特征评估方法。

考虑到确定最优特征组合的实际需要以及模型确定最优特征组合效率较低的问题,本文提出了多维空间测度-Fisher 的特征集评估算法,基于特征集评估系数极大值为最优特征集的确定提供参考,研究表明,该方法确定的最优特征集在保证较高轴承故障识别精度的同时,特征集维数也较低。同时针对更加复杂轴承故障数据时,该方法也表现出了很强的通用性和有效性,为最优特征集的确定提供了一个有效的解决思路。

参考文献

- [1] 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 H M, CHEN J, DONG G M, et al. Bearing fault diagnosis based on Integrated hidden markov model[J].

- Journal of Vibration and Shock, 2014, 33(10): 92-96.
- [2] 李兵, 张培林, 任国全, 等. 基于互信息的滚动轴承故障特征选择方法[J]. 测试技术学报, 2009, 23(2): 183-188.
- LI B, ZHANG P L, REN G Q, et al. Fault feature selection method for rolling bearings based on mutual Information [J]. Journal of Test and Measurement Technology, 2009, 23(2): 183-188.
- [3] 王雪冬, 赵荣珍, 邓林峰. 基于 KSLPP 与 RWKNN 的旋转机械故障诊断[J]. 振动与冲击, 2016, 35(8): 219-223.
- WANG X D, ZHAO R ZH, DENG L F. Fault diagnosis of rotating machinery based on KSLPP and RWKNN[J]. Journal of Vibration and Shock, 2016, 35(8): 219-223.
- [4] SUN W, CHEN J, LI J. Decision tree and PCA-based fault diagnosis of rotating machinery [J]. Noise & Vibration Worldwide, 2007, 21(3): 1300-1317.
- [5] 黄大荣, 陈长沙, 孙国玺, 等. 复杂装备轴承多重故障的线性判别分析与反向传播神经网络协作诊断方法[J]. 兵工学报, 2017, 38(8): 1649-1657.
- HUANG D R, CHEN CH SH, SUN G X, et al. Linear discriminant analysis of multiple faults in complex equipment bearings and collaborative diagnosis method of back propagation neural networks [J]. Acta Armamentarii, 2009, 38(8): 1649-1657.
- [6] JANJARASJITT S, OCAK H, LOPARO K A. Bearing condition diagnosis and prognosis using applied nonlinear dynamical analysis of machine vibration signal [J]. Journal of Sound & Vibration, 2008, 317(1): 112-126.
- [7] YU J B. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models [J]. Expert Systems with Applications, 2011, 25(7): 2573-2588.
- [8] 徐卫晓, 宋平, 谭继文. 基于 KPCA-BP 网络模型的滚动轴承故障诊断方法研究[J]. 煤矿机械, 2014, 35(8): 265-267.
- XU W X, SONG P, TAN J W. Research on fault diagnosis method of rolling bearing based on KPCA-BP network model [J]. Mining Machinery, 2014, 35(8): 265-7.
- [9] DASH M, LIU H. Feature Selection for Classification [M]. Amsterdam: IOS Press, 1997.
- [10] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述[J]. 控制与决策, 2012, 27(2): 161-166.
- YAO X, WANG X D, ZHANG Y X, et al. Summary of feature selection methods [J]. Control and Decision, 2012, 27(2): 161-166.
- [11] 蔡哲元, 余建国, 李先鹏, 等. 基于核空间距离测度的特征选择 [J]. 模式识别与人工智能, 2010, 23(2): 235-240.
- CAI ZH Y, YU J G, LI X P, et al. Feature selection based on kernel space distance measurement [J]. Pattern Recognition & Artificial Intelligence, 2010, 23(2): 235-240.
- [12] 朱丽娜, 王学杰, 张进杰. 基于 ReliefF-PCA 和 SVM 的发动机故障诊断方法研究[J]. 北京化工大学学报(自然科学版), 2018, 45(1): 55-59.
- ZHU L N, WANG X J, ZHANG J J. Research on engine fault diagnosis method based on ReliefF-PCA and SVM [J]. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2018, 45(1): 55-59.
- [13] 赵荣珍, 李坤杰, 邓林峰. 基于 ReliefF 和遗传算法的故障特征筛选方法 [C]. 全国设备监测诊断与维护学术会议、全国设备故障诊断学术会议, 2014.
- ZHAO R ZH, LI K J, DENG L F. Fault feature screening method based on ReliefF and genetic algorithm [C]. National Conference on Equipment Monitoring Diagnosis and Maintenance, National Conference on Equipment Fault Diagnosis, 2014.
- [14] 姜万录, 王友荣, 王振威, 等. 基于 ReliefF 算法和相关度计算结合的故障特征降维方法及其应用 [J]. 液压与气动, 2015(12): 18-24.
- JIANG W L, WANG Y R, WANG ZH W, et al. Fault feature reduction method based on ReliefF algorithm and correlation calculation and its application [J]. Hydraulics and Pneumatics, 2015(12): 18-24.
- [15] GUAN X, CHEN G. Sharing pattern feature selection using multiple improved genetic algorithms and its application in bearing fault diagnosis [J]. Journal of Mechanical Science and Technology, 2019, 33(1): 129-138.
- [16] 屈尔庆, 刘坤, 陈海永, 等. 基于 P-ReliefF 特征选择方法的带钢表面缺陷识别 [J]. 电子测量与仪器学报, 2017, 31(7): 1053-1060.
- QU ER Q, LIU K, CHEN H Y, et al. Surface defect identification of strip steel based on P-ReliefF feature selection method [J]. Journal of Electronic Measurements and Instrumentation, 2017, 31(7): 1053-1060.
- [17] DÍAZ-VICO D, DORRONSORO J R. Deep least squares fisher discriminant analysis [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, doi: 10.1109/TNNLS.2019.2906302.
- [18] 谢小娟, 曾以成, 熊冰峰, 等. 说话人识别中基于 Fisher 比的特征组合方法 [J]. 计算机应用, 2016, 36(5): 1421-1425.

XIE X J, ZENG Y CH, XIONG B F, et al. Feature combination method based on Fisher ratio in speaker recognition [J]. Journal of Computer Applications, 2016, 36(5): 1421-1425.

- [19] 李晋徽, 杨俊安, 项要杰. 基于高斯滤波器及费舍尔准则的特征提取方法[J]. 电路与系统学报, 2013, 18(2): 400-404.
- LI J H, YANG J AN, XIANG Y J. Feature extraction method based on gaussian filter and Fisher criterion[J]. Journal of Circuits and Systems, 2013, 18(2): 400-404.
- [20] 王之海, 伍星, 柳小勤. 基于位置补偿系数距离估计的滚动轴承特征损伤敏感性评估算法研究[J]. 振动与冲击, 2019, 38(1): 65-72.
- WANG Z H, WU Xing, LIU X Q. Research on damage sensitivity evaluation algorithm of rolling bearing based on distance compensation coefficient distance estimation[J]. Journal of Vibration and Shock, 2019, 38(1): 65-72.
- [21] 蒋超, 张应红, 徐晋勇, 等. 基于振动的滚动轴承故障诊断技术研究[J]. 煤矿机械, 2012, 33(2): 235-237.
- JIANG CH, ZHANG Y H, XU J Y, et al. Research on fault diagnosis technology of rolling bearing based on vibration [J]. Mining Machinery, 2012, 33(2): 235-237.
- [22] LONG B. Diagnostics of filtered analog circuits with tolerance based on LS-SVM using frequency features[J]. Journal of Electronic Testing, 2012, 28(3): 291-300.
- [23] RAMIREZ-VILLEGAS J F, RAMIREZ-MORENO D F. Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions [J]. Neurocomputing, 2012, 77(1): 82-100.
- [24] SINGH S, KUMAR N. Combined rotor fault diagnosis in rotating machinery using empirical mode decomposition[J]. Journal of Mechanical Science & Technology, 2014, 28(12): 4869-4876.

作者简介



刘浩炜, 2016 年于昆明理工大学获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为设备状态监测与故障诊断。
E-mail: 1315650794@qq.com

Liu Haowei received his B. Sc. degree from Kunming University of Science and Technology in 2016. Now he is a M. Sc. candidate at Kunming University of Science and Technology. His main research interest includes equipment condition monitoring and fault diagnosis.



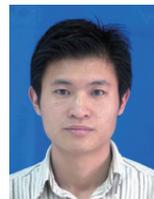
刘韬(通信作者), 2005 年于西安理工大学获得硕士学位, 2014 年于上海交通大学获得博士学位, 现为昆明理工大学副教授、硕士生导师, 主要研究方向为设备状态监测及故障诊断。
E-mail: kmlitao@aliyun.com

Liu Tao (Corresponding author) received his M. Sc. degree from Xi'an University of Technology in 2005, Ph. D. degree from Shanghai Jiaotong University in 2014. now he is an associate professor and M. Sc. supervisor at Kunming University of Science and Technology. His main research interest includes equipment condition monitoring and fault diagnosis.



涂文涛, 2017 年于昆明理工大学获得学士学位, 现为昆明理工大学硕士研究生, 主要研究方向为设备性能评估和寿命预测。
E-mail: 1315650794@qq.com

Tu Wentao received his B. Sc. degree from Kunming University of Science and Technology in 2017. Now he is a M. Sc. candidate at Kunming University of Science and Technology. His main research interest includes equipment performance assessment and life prediction.



陈庆, 现为昆明理工大学博士研究生, 主要研究方向为机电系统故障诊断。
E-mail: qingchen@kmust.edu.cn

Chen Qing is a Ph. D. candidate at Kunming University of Science and Technology. His main research interest includes electromechanical system fault diagnosis.