

文章编号: 1006-1355(2014)05-0187-05

## 基于谱质心直方图—SVM的滚动轴承故障诊断

李军杰, 马建仓, 柏会宁, 孟宏伟

(西北工业大学 电子信息学院, 西安 710129)

**摘要:** 轴承早期故障信号的信噪比较低, 较难识别。为提高轴承早期故障诊断的准确率, 分析了现有方法和视觉信息处理方式的相似点, 从视觉信息处理的角度进行研究, 提出一种基于谱质心直方图的滚动轴承故障特征提取方法, 将所提出方法用于滚动轴承不同运行状态下的故障特征提取, 将所提取特征作为支持向量机的输入特征向量从而实现滚动轴承不同运行状态的智能分类。实验证明, 谱质心直方图提高了轴承早期故障诊断准确率, 准确率平均提高了2%, 优于Mel频率倒谱系数特征。

**关键词:** 振动与波; 谱质心直方图; 谱峭度; 支持向量机; 轴承早期故障诊断; 视觉信息处理

中图分类号: TN911.7

文献标识码: A

DOI编码: 10.3969/j.issn.1006-1355.2014.05.041

## Fault Diagnosis of Rolling Bearings Based on Spectral Centroid Histograms-SVM

LI Jun-jie, MA Jian-cang, BAI Hui-ning, MENG Hong-wei

(School of Electronics & Information, Northwestern Polytechnical University, Xi'an, 710129, China)

**Abstract:** Early fault features of rolling bearings are usually immersed in heavy noise background. In other words, its signal-to-noise ratio (SNR) is too low to diagnose the fault. In this paper, the similarity of the existing methods and the way of visual information processing is analyzed in order to improve the diagnosis accuracy of the early fault diagnosis of the rolling bearings. From the viewpoint of visual information processing, a method of fault feature extraction of rolling bearings under different operation conditions based on spectral centroid histogram is proposed. Then, the extracted features are used as the input feature vectors of SVM to realize the intelligence diagnosis in different operation status. Experiments show that using the spectral centroid histogram can raise the diagnosis accuracy of the early fault diagnosis of the rolling bearings, and the accuracy is increased by 2%, better than that of the MFCC features.

**Key words:** vibration and wave; spectral centroid histograms; spectral kurtosis; SVM; early fault diagnosis; visual information processing

滚动轴承广泛应用于旋转机械设备, 轴承过大的振动降低生产线的质量, 严重的轴承振动甚至可以使整个系统无法正常工作<sup>[1]</sup>。因此对轴承故障诊断技术进行研究具有较大经济效益。据统计, 滚动轴承90%的故障发生在内环、外环和滚动体上<sup>[2]</sup>。

因此, 本文只对这3种故障进行诊断实验。在实际应用中, 由于设备环境复杂, 各种振动和噪声混合在一起, 故障信号往往淹没于噪声之中。当故障特征明显时, 已经十分严重, 仍有可能带来巨大损失, 因此, 在故障特征微弱时对故障进行诊断显得十分重要<sup>[3]</sup>。

滚动轴承的故障诊断本质上可以看成模式识别问题, 诊断步骤主要包括<sup>[4]</sup>: 数据获取、特征提取、特征选择和识别几个步骤。其中特征提取和特征选择是故障诊断的关键, 识别是故障诊断的核心。文献[5]提出一种基于谱峭度的早期故障检测方法, 该方法采用短时傅里叶变换、FIR滤波器组对信号进行分解, 获得信号谱峭度特征进行故障诊断。文献[6]运用基于多小波的谱峭度对轴承故障进行诊断, 提

收稿日期: 2013-12-15

基金项目: 西北工业大学研究生创业种子基金项目(Z2014132)

作者简介: 李军杰(1990-), 男, 河南开封人, 硕士生。主要研究方向: 振动故障诊断、神经网络、深度学习。

E-mail: lijunjiehnkf@hotmail.com

通讯作者: 马建仓(1956-), 男, 陕西宝鸡人, 教授, 硕士生导师, 主要研究方向: 信息获取处理与传输、故障诊断、盲信号处理、仿生信号处理。

E-mail: majc@nwpu.edu.cn

高了频带选择的准确性与滤出信号的信噪比,获得较好的诊断效果。文献[7]提出一种基于谱质心直方图的语音特征提取方法,并应用于强噪声环境中的语音特征提取,最终识别结果比Mel倒谱系数(MFCC)特征还要好。上述两种方法都是基于频率的多通道模型,这和视觉信息处理机制(如图1)惊人地相似,人类视觉是通过不同细胞去感知不同的频率成分,先进行初级的信息处理,然后传输到下一层(分类器)进行高级处理。由文献[7]可知谱质心直方图在强噪声环境下的语音信号特征提取方面表现很好。因此,本文对谱质心直方图特征在轴承早期故障诊断方向的应用进行了研究。

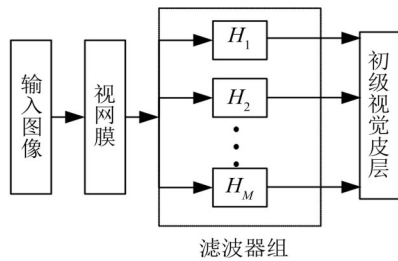


图1 视觉信息处理模型

设计了两组实验对谱质心直方图特征的轴承故障诊断能力进行研究,在不同信噪比环境下进行对比研究,特征选择采用主成分分析方法<sup>[8]</sup>(PCA),分类器采用支持向量机分类器。实验1对谱质心直方图特征的轴承故障诊断能力进行了仿真实验,与具有较好诊断能力的MFCC特征进行对比分析。实验2对基于小波包包络的谱质心直方图特征和谱峭度直方图特征的诊断结果进行了对比。实验证明谱质心直方图可以较好地提高轴承早期故障诊断精度。

## 1 特征提取

### 1.1 谱峭度直方图

谱峭度(Spectral Kurtosis, SK)基本思想是:对信号进行包络分解,得到各个频段的包络信号的峭度,最终得到一个谱峭度直方图,峭度值较大的位置就是非高斯信号所在的频段,可以对故障信号所在频段进行定位。该方法完全是启发式的,对附加噪声相对鲁棒,文献[5]根据Wold-Cramer分解给出了非平稳信号SK的正式定义。SK定义为归一化4阶累积量

$$K_x(f) = \frac{\left\langle |H(n, f)|^4 \right\rangle}{\left\langle |H(n, f)|^2 \right\rangle^2} - 2 \quad (1)$$

其中  $H(n, f)$  是信号  $x(n)$  在频率  $f$  处的复包络,

$\langle \cdot \rangle$  是一个时间平均算子,这里  $-2$  而不是  $-3$  是因为  $H(n, f)$  是复数。如果令  $y(n) = x(n) + N(n)$ ,  $N(n)$  为高斯噪声,则对应  $y(n)$  的谱峭度定义为

$$K_y(f) = \frac{K_x(f)}{[1 + \rho(f)]^2}, \quad f \neq 0 \quad (2)$$

其中  $\rho(f)$  是在频率  $f$  处的信噪比。

谱峭度直方图特征有许多好的性质,它可以描述信号高斯性,对于高斯过程SK值为零,非高斯过程SK非零,而且信号偏离高斯过程越远,SK值越大。SK还可以用来度量随机过程概率密度函数在频率  $f$  处的峰值性,对于平稳过程SK值是频率  $f$  的常量函数。因此谱峭度能够检测定位信号的非平稳成分,能够较好地对轴承故障信号进行识别。进行谱峭度分析的较好的工具是时间-频率分析,例如短时傅里叶变换等。上述的时间-频率平面也可以推广到时间-尺度平面,通过小波包滤波进行分析。

这里选择小波包对轴承故障信号进行谱峭度直方图特征提取。具体步骤是:首先利用小波包分解得到各层小波系数,然后利用Hilbert变换求各层小波系数的包络信号,即得到(1)式中的  $H$ ,最后求得谱峭度直方图特征。虽然小波包变换具有粗糙的频率分辨率,但是在一定程度上提高了轴承早期故障的诊断能力。

### 1.2 谱质心直方图

谱质心直方图主要反映了信号频谱各个子带主要能量集中的频率范围,而子带能量特征只有能量信息,能量信息对信号微弱变化不敏感,这就限制了能量特征的早期故障诊断准确率。所以谱质心直方图特征是谱峭度直方图特征、小波熵特征的完美补充。

谱质心直方图计算步骤<sup>[7]</sup>如下:第1步,对信号进行功率谱估计,这里选择经典的周期图法进行谱估计,窗函数利用Kaiser窗。第2步,选择滤波器组进行滤波,把采样信号的功率谱通过滤波器组在频域划分为一系列子带,这里选用均匀的矩形窗进行处理,也可以选用其他特征的滤波器组,只要达到较好的诊断效果即可。第3步,谱质心特征提取,在每个频带上计算谱均值、谱质心特征。令  $SC_m$  表示第  $m$  个频带上的谱质心,则

$$SC_m = \frac{\sum_{k=1}^N f_k H_m(f_k) P^\gamma(f_k)}{\sum_{k=1}^N H_m(f_k) P^\gamma(f_k)} \quad (3)$$

其中  $P$  表示功率谱估计,  $H_m(f_k)$  表示第  $m$  个带通滤波器的频率响应,  $N$  为数据采样点数,  $\gamma$  决定频谱

的动态范围,  $\gamma$  如果太小则计算得到的谱质心为子带中心, 如果太大则得到的谱质心为子带内的峰值。一般取  $\gamma=1$ , 此时功率谱质心指标反映了每个子带内功率谱重心位置, 当故障出现时振动信号功率谱值将发生变化, 会在很大程度上影响谱质心值。

第四步, 围绕谱质心计算每个频带内的谱能量、谱方差, 谱方差计算公式如下

$$P_{\sigma} = \frac{\sum_{k=1}^N (f_k - SC_m)^2 H_m(f_k) P^{\gamma}(f_k)}{\sum_{k=1}^N H_m(f_k) P^{\gamma}(f_k)} \quad (4)$$

这里不像 MFCC 那样计算整个子带的能量, 而是以谱质心为中心的一个更小的子带内的能量。因为谱质心周围的谱线具有相对有限的附加噪声, 提高了特征的抗噪声能力。

由于谱质心直方图携带了每个子带上主要的频率信息, 弥补了能量等特征的不足, 结合以谱质心为中心的谱方差特征, 能够较好地地区分微弱的轴承故障信号, 实现轴承的早期故障诊断。

## 2 支持向量机分类器

支持向量机是一种统计学习方法, 是基于结构风险最小化的方法, 具有较好的泛化能力。其基本思想<sup>[9]</sup>是: 给定训练样本, 建立一个超平面作为决策平面, 使得正例和反例之间的隔离边缘最大化。其中支持向量是由算法从训练数据中抽取的小的子集构成, 这些支持向量组成训练数据的分类曲面, 对于复杂的模式分类问题, 支持向量机具有较好的分类性能, 在某些公共数据集上的效果甚至已经超越 BP 神经网络。

支持向量机的原问题数学上可以描述为: 给定训练样本  $\{(x_i, d_i)\}_{i=1}^N$ , 找到权值向量  $w$  和偏置  $b$  的最优值使得他们满足下面的约束条件

$$d_i(w^T \phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, N \quad (5)$$

其中  $\phi(\cdot)$  代表核函数, 并且权值向量  $w$  最小化代价函数  $\Phi(w) = 1/2 w^T w$ , 这个约束优化问题称为原问题。然后使用拉格朗日乘子法解决该约束优化问题, 建立拉格朗日函数为

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i(w^T \phi(x_i) + b) - 1] \quad (6)$$

对  $w$  和  $b$  求偏导并置零, 得到最优化条件如下

$$w = \sum_{i=1}^N \alpha_i d_i \phi(x_i) \quad (7)$$

把(11)式代入决策面  $w^T \phi(x) + b = 0$  得到

$$\sum_{i=1}^N \alpha_i d_i k(x, x_i) + b = 0 \quad (8)$$

其中  $k(x, x_i) = \phi^T(x_i) \phi(x)$ , 称为内积核, 正是因为内积核, 把一个非线性可分的输入空间映射到一个线性可分的输出空间, 从而只需要解决一个线性可分问题即可。因此, 支持向量机分类的对偶问题可以描述为: 寻找拉格朗日乘子  $\{\alpha_i\}_{i=1}^N$  最大化目标函数

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j k(x_i, x_j) \quad (9)$$

并满足约束条件

$$1) \sum_i \alpha_i d_i = 0;$$

$$2) 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N$$

其中  $C$  为用户选定的惩罚参数。求解(9)式的线性约束二次规划问题采用序列最小优化算法<sup>[10]</sup> (SMO), 把全局的 QP 问题分解为一系列的 QP 子问题, 解析地进行优化。

SVM 分类器的设计需要对核函数和模型参数行设置, 对于线性不可分问题, RBF 核函数是第一选择<sup>[8]</sup>, 因为多项式核会含有过多的参数, 过多的参数影响模型的复杂度; 线性核是一种特殊的 RBF 核函数; 而 Sigmoid 核和 RBF 核的作用相似。参数  $C$  和核参数  $\sigma$  的选择一般采用交叉验证法, 先设置较大范围的  $C$  和  $\sigma$  进行网格搜索, 获得准确率较大的参数值, 然后进行小范围搜索, 从而获得最优的模型参数  $C$  和  $\sigma$ 。

## 3 实验结果与分析

实验采用美国凯斯西储大学轴承数据中心的轴承振动故障数据进行诊断分析。信号采样频率为  $f_s = 12$  kHz, 采样长度为  $N = 4096$ 。该数据的故障特征比较明显, 实验通过加入高斯白噪声来模拟轴承早期故障信号, 本文共选取 3900 组样本进行实验, 其中正常样本 826 个, 内圈故障 812 个, 滚动体故障 812 个, 外圈故障 1450 个, 对所有实验样本加入高斯白噪声, 然后进行故障诊断。首先, 求得滚动轴承 4 种不同状态下信号的时域波形和对应的包络谱, 如图 2 所示。由图可知仅仅依据包络谱较难判断轴承的故障类型, 所以引入本文算法, 如下实验。

实验 1: 谱质心直方图特征计算时采用周期图法进行功率谱估计, 窗函数选择 Kaiser 窗, 窗参数取 4。由于谱对称, 所以只需要选择 2048 个点进行计算即可, 对频谱进行多通道滤波, 这里可以选用不同的滤波器组, 下面实验均选择矩形滤波器组进行均匀划分为 2 的整数次方段, 选择滤波器个数为 32。然后计算谱质心直方图特征, 获得原始数据的特征向量。第 1 组实验只选择频域统计特征进行诊断; 第 2 组实验在第 1 组实验基础上加上谱质心直方



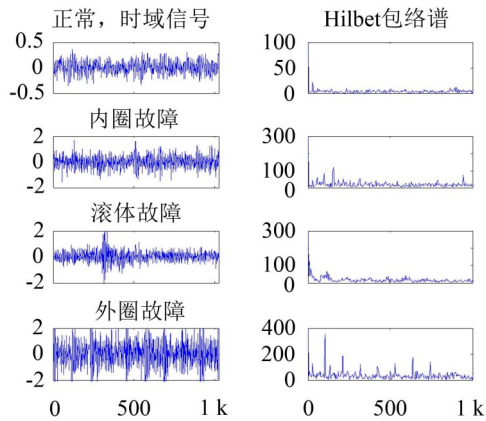


图 2 不同故障的信号波形和包络谱

图特征进行对比;第3组实验采用的是现有的故障诊断效果比较好的MFCC特征进行对比。

由于特征提取过程中,特征向量往往带有很大冗余,较高的特征维数会加重分类器的负担,甚至产生维数灾难,因此需要对提取特征进行降维。这里采用PCA方法进行降维,然后运用PCA处理后的训练集特征数据对SVM分类器模型进行训练,模型参数通过10重交叉验证和 $(C, \sigma)$ 网格搜索获得最优模型参数。获得SVM模型参数后,运用PCA处理后的测试集特征数据对获得的SVM模型进行测试,选取2 680个样本作为训练样本,1 220个样本作为测试样本,最终获得诊断结果如表1所示。

由表1可知,频域统计特征加上谱质心直方图后,准确率有较大的提高,特别是在信噪比为-1 dB时,提高了将近6个百分点,其它信噪比情况下提高3~5个百分点,而且和经典的MFCC特征相比也有一定提高。说明谱质心特征具有一定的抗噪声能力,使得SVM分类精度在低信噪比时具有较大提高,可以较好地发现轴承的早期故障(信号淹没于噪声中)。SVM分类器在测试集上的故障诊断准确率

随信噪比变化的趋势如图3,由图可知,谱质心直方图特征明显改善频域统计特征诊断精度,而且精度随信噪比变化较为稳定,稍微优于MFCC特征。图4是SVM分类器在测试集上的均方误差随信噪比变化的趋势图,第1组特征的分类器误差较大,特别是在低信噪比情况下,具有较高的均方误差;第2组实验的分类器分类误差明显减小,验证了谱质心特征对轴承内圈故障、外圈故障、滚动体故障的较好的区分能力,对噪声有相对好的鲁棒性。

实验2:运用小波包对信号进行3层小波包分解,小波基选择db 4小波,因为选取db 5以上的小波诊断效果得不到明显改善。然后对每一层小波包系数进行特征提取,包括谱峭度直方图、谱质心直方图,获得原始信号的特征向量。第1组实验特征向量包括统计特征和谱峭度直方图特征;第2组特征向量包括统计特征和谱质心直方图特征。分类器仍然选择SVM分类器,运用实验1中的方法对SVM进行参数设置和优化,仍然选取2 680个样本作为训练样本,1 220个样本作为测试样本,最终获得故障诊断结果如表2所示。

由表2可知谱质心直方图特征的诊断准确率相对谱峭度特征得到明显改善,信噪比为-1 dB时提高了4个百分点,其它情况下平均提高2个百分点,说明谱质心特征具有较好的抗噪声能力,在强噪声环境下对轴承故障具有较好的诊断效果。图5中为测试集上分类器正确率趋势图,可知谱质心直方图特征的诊断结果相对谱峭度特征得到很大提高,而且准确率随信噪比的增加变化较为稳定,说明谱质心直方图具有更好的轴承早期故障诊断能力。由图6可知基于谱质心直方图特征的分类器误差几乎呈直线下降,说明算法具有较好的鲁棒性,适合对轴承早期故障进行诊断。

表 1 实验1轴承故障诊断准确率 单位:%

信噪比 方法	- 1 dB		3 dB		5 dB		8 dB	
	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集
频域统计特征	78.9	71.7	91.0	86.2	95.6	90.8	98.4	96.7
同上+谱质心	82.8	77.1	94.0	91.3	98.0	94.9	99.8	99.3
Mel倒谱系数	83.7	75.5	93.2	89.8	96.5	94.0	98.9	98.0

表 2 实验2轴承故障诊断准确率 单位:%

信噪比 方法	- 1 dB		3 dB		5 dB		8 dB	
	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集
谱峭度直方图	79.2	75.1	87.9	86.5	91.6	88.4	96.2	94.1
谱质心直方图	82.0	79.6	90.6	87.4	94.4	91.6	98.3	96.9

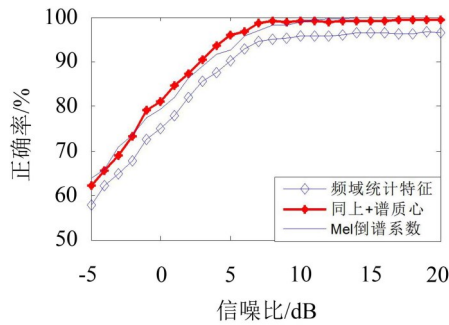


图3 分类器正确率趋势图

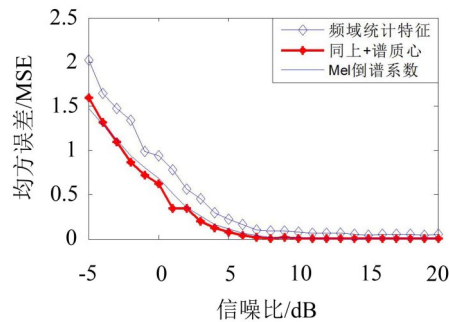


图4 分类器均方误差趋势图

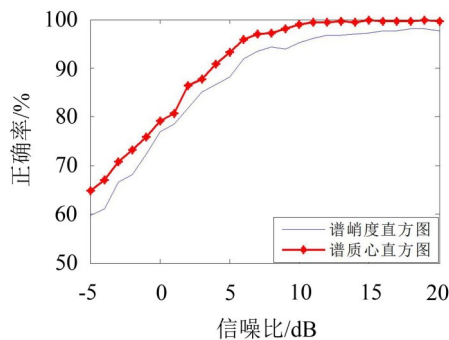


图5 分类器正确率趋势图

## 4 结语

本文对基于谱质心直方图的轴承早期故障特征提取方法进行了研究。与采用统计特征提取轴承故障特征相比较，谱质心直方图提供了不同子带的频率信息，说明了不同频域子带能量集中的频率范围。实验证明，谱质心能够改善强噪声环境下轴承故障诊断准确率，加入谱质心直方图特征后频域特征比MFCC特征的识别精度要高，弥补了基于小波包谱峭度直方图特征的不足，达到更高的识别精度。因此，谱质心直方图特征对于轴承早期故障诊断具有较高的工程应用价值。

谱质心直方图特征也有缺陷，算法第二步中需要对多通道滤波器组设计，如何设计适合当前问题的滤波器组，针对具体问题设计具体滤波器组，这是个值得研究的问题。除此之外，滤波器、滤波器个

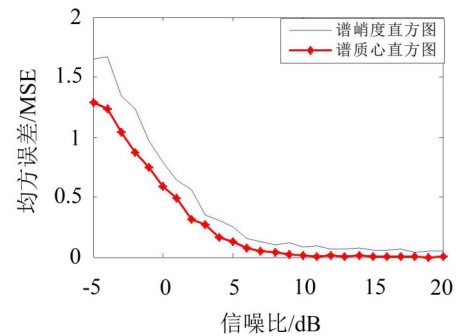


图6 分类器均方误差趋势图

数、谱质心参数 $\gamma$ 等的选择没有统一的理论方法。采用贪婪搜索的方法进行搜索，通过观察设置不同参数时的分类结果选择最优参数。模型参数选择是前期设计阶段的任务，一旦确定，系统设计完成，所以不影响最后的工程应用。

### 参考文献:

- [1] Kankar P K, Sharma S C, Harsha S P. Fault diagnosis of ball bearings using machine learning methods[J]. **Expert Systems with Applications**, 2011, 38(3): 1876-1886.
- [2] Jardine A K S, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J]. **Mechanical Systems and Signal Processing**, 2006, 20(7): 1483-1510.
- [3] 夏均忠, 刘远宏, 冷永刚, 等. 微弱信号检测方法的现状分析[J]. **噪声与振动控制**, 2011(03): 156-161.
- [4] 张新海, 雷勇. BP神经网络在机械故障诊断中的应用[J]. **噪声与振动控制**, 2008(05): 95-97.
- [5] Antoni J. The spectral kurtosis: a useful tool for characterising nonstationary signals[J]. **Mechanical Systems and Signal Processing**, 2006, 20(2): 282-307.
- [6] 王晓冬, 何正嘉, 訾艳阳. 滚动轴承故障诊断的多小波谱峭度方法[J]. **西安交通大学学报**, 2010, 44(03): 77-81.
- [7] Gajic B, Paliwal K K. Robust speech recognition in noisy environments based on subband spectral centroid histograms[J]. **Audio, Speech, and Language Processing, IEEE Transactions on**, 2006, 14(2): 600-608.
- [8] Widodo A, Yang B. Support vector machine in machine condition monitoring and fault diagnosis[J]. **Mechanical Systems and Signal Processing**, 2007, 21(6): 2560-2574.
- [9] Haykin S. Neural Networks and learning machines[M]. Upper Saddle River, New Jersey: PEARSON Prentice Hall, 2009.
- [10] FAN R, CHEN P, LIN C. Working set selection using second order information for training support vector machines[J]. **The Journal of Machine Learning Research**, 2005, 6: 1889-1918.