

# KNN—朴素贝叶斯算法的滚动轴承故障诊断

路敦利<sup>1</sup>, 宁芊<sup>1,2</sup>, 杨晓敏<sup>1</sup>

(1. 四川大学 电子信息学院, 成都 610065; 2. 电子信息控制重点实验室, 成都 610036)

**摘要:** 滚动轴承的故障诊断对于提高工业生产效率, 保障工业生产的稳定安全地运行具有重要意义; 为了提高滚动轴承故障识别的正确率, 提出一种使用 KNN—朴素贝叶斯决策组合算法对滚动轴承故障诊断; 组合算法利用朴素贝叶斯算法对使用不同  $K$  值的 KNN 算法初步分类结果进行再分类以达到提高滚动轴承故障识别的目的; 首先, 使用小波包能量法对滚动轴承振动信号进行能量特征提取, 然后使用多个参数  $K$  值不同的 KNN 算法对能量特征数据预分类, 得到多个 KNN 算法分类结果集, 将分类结果集进行处理得到预分类结果集, 将预分类结果集作为朴素贝叶斯算法的输入, 使用朴素贝叶斯算法对数据再分类; 实验结果表明, 组合算法相较于传统 KNN 算法及贝叶斯算法在滚动轴承的故障诊断率得到了有效提高, 实现了对滚动轴承故障的有效诊断。

**关键词:** KNN; 贝叶斯算法; 故障诊断; 滚动轴承; 小波包

## Fault Diagnosis of Rolling Bearing Based on KNN—Naive Bayesian Algorithm

Lu Dunli<sup>1</sup>, Ning Qian<sup>1,2</sup>, Yang Xiaomin<sup>1</sup>

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

2. Science and Technology on Electronic Information Control Laboratory, Chengdu 610036, China)

**Abstract:** Fault diagnosis of rolling bearings is of great importance to improve the efficiency of industrial production and ensure the stable and safe operation of industrial production. In order to improve the accuracy of rolling bearing fault recognition, a fault diagnosis method based on KNN—naive Bayesian decision combination algorithm is proposed. The combined algorithm uses the naive Bayesian algorithm to reclassify the KNN algorithm using different  $K$  values to achieve the purpose of improving the fault identification of rolling bearings. Firstly, the energy feature extraction of the rolling bearing vibration signal is carried out by using the wavelet packet energy method. Then, the KNN algorithm is used to pre-classify the energy characteristic data, and the KNN algorithm is used to classify the result group. The classification result set is used as the input of the naive Bayesian algorithm, and the data is reclassified using the naive Bayesian algorithm. The experimental results show that the combined algorithm is effective compared with the traditional KNN algorithm and the Bayesian algorithm in the fault diagnosis of rolling bearings. The combination algorithm realizes the effective diagnosis of rolling bearing faults.

**Keywords:** KNN; Bayesian algorithm; troubleshooting; rolling bearings; wavelet package

## 0 引言

滚动轴承是旋转机械中较为常见的零部件, 滚动轴承故障往往会导致其他零部件发生故障, 进而影响设备的正常运行。滚动轴承在使用过程中一般会经历正常运转、故障、失效几个阶段。若检测到滚动轴承发生故障, 及时停机更换部件, 以防轴承失效, 造成更加严重的后果, 同时可以有效地降低设备故障带来经济损失。因此, 对滚动轴承的故障诊断有着重要的研究意义<sup>[1]</sup>。

对于旋转设施关键的滚动轴承故障诊断, 获取故障信息的方法有温度检测法, 振动监测法以及光纤检测法等<sup>[1]</sup>。目前一般采用振动监测法对滚动轴承故障诊断。传统的诊断技术一般是针对滚动轴承振动信号的时频域特征来提取特征向量进行故

障识别。常用的识别算法有支持向量机 (SVM)、人工神经网络 (ANN)、决策树、随机森林等。使用支持向量机进行故障诊断获取的诊断结果置信度偏低; 人工神经网络对训练样本要求较高, 容易陷入局部极小值, 并且运算速度慢。使用小波变化与小波分析对滚动轴承信号处理可以将信号的低频特征与高频特征很好的表征出来, 小波包分析是在小波分析的基础上对信号更进一步的分析与重构。滚动轴承振动信号的时频信号经小波包能量法提取特征后, 各子频带能量分布存在较大差异, 相较于时频域特征更适合做故障识别。小波变换由于其能够对不同频域的信号进行特征分析, 抗干扰能力强, 并且对输入信号要求不高, 成为近些年研究的热点<sup>[2-9]</sup>。目前, 越来越多的研究将小波包能量法与机器学习相结合, 实现滚动轴承的故障诊断。文献 [10] 中提出使用小波包能量法与 BP 神经网络算法相结合, 对滚动轴承故障诊断, 该方法在神经网络设计合理的情况下, 效果较好, 但算法容易陷入局部极值。文献 [11] 提出 C4.5 决策树与小波包能量法相结合, 总体正确率达到 92%。传统分类算法 KNN 算法与朴素贝叶斯算法在模式识别中的到广泛应用<sup>[12-15]</sup>。KNN 算法与朴素贝叶斯算法在滚动轴承故障诊断中单独使用分类效果比较一般, 无法达到工业要求。

本文提出一种 KNN 算法与朴素贝叶斯算法结合的改进算

收稿日期: 2017-10-06; 修回日期: 2017-11-06。

**作者简介:** 路敦利(1991-), 男, 山东济宁人, 硕士研究生, 主要从事模式识别方向的研究。

宁芊(1969-), 女, 四川成都人, 副教授, 硕士研究生导师, 主要从事模式识别方向的研究。

杨晓敏(1980-), 女, 四川成都人, 副教授, 硕士研究生导师, 主要从事图像处理方向的研究。

法, 可以有效地提高滚动轴承故障诊断的识别率。首先使用  $K$  值为 1、3、5、7 的 KNN 算法对滚动轴承振动信号特征数据进行分类, 得到初步分类结果集, 将结果集作为朴素贝叶斯算法的输入, 再使用朴素贝叶斯算法进行分类, 得到最终分类结果。经实验验证表明, 改进后算法的准确率得到了有效提高。

## 1 小波包的滚动轴承故障特征向量提取

小波包变换是在多分辨率分析基础上构成的一种更为精细的正交分解方法, 它在全频带内对信号进行多层次的频带划分, 不仅继承了小波变换所具有的良好时频局域化优点, 而且继续对没再分解的高频频带进一步分解, 并能够根据被分析信号的特征, 自适应地选择相应频带, 使之与信号频谱相匹配, 从而提高了时频分辨率, 因此小波包在信号特征提取方面具有更广泛的应用价值<sup>[16]</sup>。

### 1.1 小波包定义

给定正交尺度函数  $\varphi(t)$  与小波函数  $\varphi(t)$ , 其尺度关系为

$$\varphi(t) = \sqrt{2} \sum_k h_{0k} \varphi(2t - k) \quad (1)$$

$$\varphi(t) = \sqrt{2} \sum_k h_{1k} \varphi(2t - k) \quad (2)$$

其中: 式 (1)、(2) 中  $h_{0k}$ 、 $h_{1k}$  为多分辨率分析中的滤波器系数。

进一步推广至二尺度方程, 可以得到如下递推关系

$$\omega_{2n}(t) = \sqrt{2} \sum_{k \in \mathbf{Z}} h_{0k} \omega_n(2t - k) \quad (3)$$

$$\omega_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbf{Z}} h_{1k} \omega_n(2t - k) \quad (4)$$

当  $n=0$  时,  $\omega_0(t) = \varphi(t)$ ,  $\omega_1(t) = \varphi(t)$ , 以上定义的函数集合  $\{\omega_n(t)\}_{n \in \mathbf{Z}}$  为由  $\omega_0(t) = \varphi(t)$  确定的小波包。

小波包分解:

$$d_k^{i, 2n} = \sum_k h_{0k} - 2ld_k^{i+1}, d_k^{i, 2n+1} = \sum_k h_{1k} - 2ld_k^{i+1} \quad (5)$$

小波包重构:

$$d_i^n = \sum_k g_{0, l-2kd_k^{i+1, 2n}} + \sum_k g_{1, l-2kd_k^{i+1, 2n+1}} \quad (6)$$

### 1.2 小波包能量提取特征

滚动轴承的振动信号共分为正常信号、外圈故障信号、内圈故障信号、滚动体故障信号 4 种状态信号。对滚动轴承的时域振动信号采用 db5 作为小波包函数对振动信号进行三层小波包分解, 这样可以得到第 3 层上 8 个频带的小波包分解系数, 然后对第 3 层上每个节点进行小波包重构, 对重构信号进行能量统计分析, 对第 3 层上各个子频带上的能量值做归一化处理, 得到 8 个归一化的频带能量特征, 作为组合算法的输入。

## 2 KNN-朴素贝叶斯组合算法及流程

$K$  近邻 ( $k$  nearest neighbor, KNN) 算法是解决分类问题中最为常用的算法之一, 其算法具有原理简单易于理解、实现方便、分类有效等特点。KNN 分类算法的主要思想是: 先计算待分类样本与已知类别的训练样本之间的距离, 找到距离与待分类样本数据最近的  $K$  个样本; 再根据这些样本所属的类别来判断待分类样本数据的类别。由于同时使用多个参数  $K$  值不同的 KNN 算法对同一特征数据集分类得到的结果是相互独立的, 符合朴素贝叶斯算法中假设所有属性相互独立的假设, 使用朴素贝叶斯算法可以进一步提高 KNN 分类器性能。朴素贝叶斯分类的原理是求解向量  $D(d_1, d_2, \dots, d_m)$  属于类

别  $C(C_1, C_2, \dots, C_n)$  的概率值  $(P_1, P_2, \dots, P_n)$ , 其中  $P_n$  是指向量  $D(d_1, d_2, \dots, d_m)$  属于类别  $C_n$  的概率,  $\max(P_1, P_2, \dots, P_n)$  为向量  $D$  对应的类别。组合算法构造算法步骤如下。

Step 1: 设  $X(x_1, x_2, \dots, x_m)$  作为小波包能量法提取的滚动轴承振动信号特征向量集合。

Step 2: 将  $X$  分为训练集  $X_{train}$  与测试集  $X_{test}$  两部分。

Step 3: 设置参数  $K$  值:  $K$  值的一般选择 1~10, 本文  $K$  值选取 1、3、5、7, 选择的  $K$  值为奇数, 尽量避免了投票过程中可能产生不同类别的最邻近数目相同的情况。

Step 4: 计算训练集样本数据和测试集样本数据的距离。

Step 5: 为测试集样本数据选择  $K$  个与其距离最小的样本。

Step 6: 得到测试集分类结果。对每个测试样本统计出最临近  $K$  个样本中大多数样本所属的分类, 使用投票法, 确定待分类数据的类别, 得到 4 个参数  $K$  值不同的 KNN 算法的初步分类结果集  $D(d_1, d_2, \dots, d_m)$ 。

Step 7: 得到  $C(C_1, C_2, \dots, C_n)$ , 对每个特征属性计算所有划分的条件概率  $(P_1, P_2, \dots, P_n)$ 。

Step 8: 对于一条待分类数据如果  $P_i = \max(P_1, P_2, \dots, P_n)$ , 那么该数据的类别为  $i$ 。以此训练分类器, 得到贝叶斯分类模型。

Step 9: 使用训练模型对测试数据集分类, 得到最终分类结果。

## 3 故障诊断实验

### 3.1 实验数据

本文实验数据使用美国 Case Western Reserve University 电气工程实验室轴承中心实验台所采集的数据。滚动轴承实验平台如图 1 所示, 实验平台包括左侧一个 2 马力的电机 (1 hp = 746 W), 中间一个转矩传感器, 右侧一个功率计和电子控制设备。

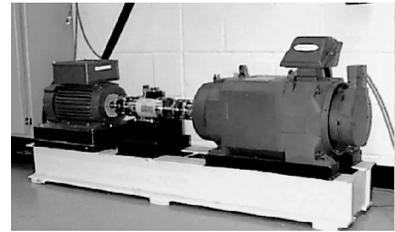


图 1 凯斯西储大学轴承故障实验台

实验时, 采用数据为电机故障直径为 0.007 英寸, 转速为 1 750 rmp, 采样频率为 12 kHz 部分数据, 其中外圈故障信号取采集点为 6 点钟方向作为实验数据。原始实验数据包含滚动轴承的 4 种运行状态, 分别是正常状态、内圈故障状态、滚动体故障状态、外圈故障状态。如下图 2 所示为采集的滚动轴承时域振动信号处于正常状态及发生不同故障时的时域振动信号波形。

通过观察图 2, 滚动轴承的时域振动波形区分度较低, 很难区分出正常信号与故障信号, 为了获取区分度更好的特征, 对滚动轴承实现有效的故障诊断, 对 4 种运行状态下的滚动轴承振动信号进行小波包能量特征提取。对滚动轴承 4 种不同状态下的能

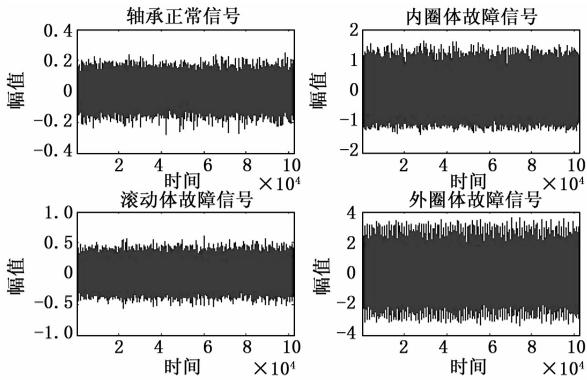


图 2 轴承各个状态下时域振动信号

量特征数据各取 80 组, 一共 320 (4 \* 80) 组实验数据。每组数据时域振动信号包含 1 024 个点, 对时域振动信号数据进行小波包能量特征提取, 首先使用 db5 作为小波包函数对振动信号进行三层小波包分解, 并在每个节点上进行小波包重构, 对重构信号进行能量统计分析, 得到一个 8 维小波包能量频带特征向量, 然后将获取的特征向量进行归一化处理。滚动轴承正常状态以及不同故障状态下的能量分布情况如图 3 所示。

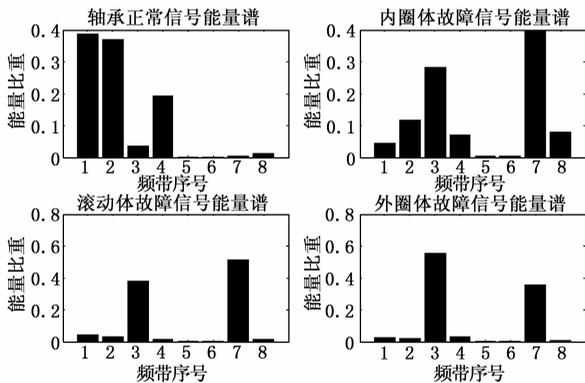


图 3 轴承各个状态下能量谱

通过图 3 我们可以比较直观的观察滚动轴承处于不同状态时, 其振动信号各子频带的能量分布存在的差异较大, 比较易于识别。轴承处于正常状态时, 能量主要集中在前 2 个子频带, 其中前两个子频带之和所占比重超过了总能量的 80%, 与其他故障状态下能量分布有很大不同, 易于区分。内圈故障、滚动体故障和外圈故障 3 种状态的能量主要分布在子频带 3 和子频带 7 上, 其余频带所占能量比重相对较少, 但 3 种不同状态下能量谱分布还是有所不同。对比滚动体故障信号能量谱与外圈故障信号能量谱, 滚动体故障信号能量谱子频带 3 能量所占比重小于子频带 7, 而外圈故障与之相反。相比这两种故障能量谱能量分布, 内圈故障信号能量谱分布相对分散, 子频带 3 和子频带 7 的能量和平均只占总能量的 65% 左右。

### 3.2 实验方案及分析

对上述滚动轴承 4 种状态下 4 \* 80 (320) 组数据中每种状态下各取 60 组数据作为训练数据集, 其余 20 组作为测试数据集验证组合算法性能。实验中分别使用参数  $K$  值取 1、3、5、7 的 KNN 算法以及朴素贝叶斯算法与组合算法做比较。

对于 KNN 算法使用训练集作为样本数据集, 直接对测试

数据集进行预测, 得到分类结果。实验结果如表 1 所示。

表 1 KNN 算法分类结果比较

参数 $K$ 值不同的 KNN 算法分类精度				
	$K=1$	$K=3$	$K=5$	$K=7$
正常状态	1	1	0.95	0.95
内圈故障	1	0.95	0.80	0.85
滚动体故障	0.95	0.85	0.85	0.85
外圈故障	0.9	0.85	0.9	0.8
分类精度	0.937 5	0.917 5	0.877 5	0.862 5

从表 1 中可以看出使用测试数据集使用  $K$  值不同的 KNN 算法分类效果也不同, 从分类结果中可以看出当  $K$  值为 1 时 KNN 算法对滚动轴承的故障类型分类效果最好, 但其他 KNN 算法在某种故障类别上诊断率同样可以达到较好的诊断效果, 例如参数  $K$  为 5 的 KNN 算法在外圈故障诊断中表现出较好的精度。当使用参数  $K=1$  的 KNN 分类器时分类性能最好, 能达到 93.75% 的正确率。

使用朴素贝叶斯算法与组合算法对滚动轴承进行故障诊断, 同样取滚动轴承 4 种状态下 4 \* 80 (320) 组数据中 4 \* 60 组数据 (每种状态下各取 60 组) 作为训练数据集, 剩余数据 (每种状态各 20 组) 作为测试数据集。得到分类结果如下表 2。

表 2 分类结果比较

分类算法总体分类效果比较			
算法名称	KNN 算法	朴素贝叶斯算法	组合算法
算法精度	0.937 5	0.875	0.987 5

由上表 2 中可以看出组合算法的分类精度较  $K=1$  的 KNN 算法提高了 5%, 比朴素贝叶斯算法精度提高了 11.25%, 由此可见组合算法的分类效果相较于传统 KNN 算法及朴素贝叶斯算法有所提高, 且算法精度达到 98.75%, 能够有效的识别滚动轴承的故障状态, 达到工业要求水准, 可以在实际生产中应用推广。

## 4 结语

提出了一种 KNN-朴素贝叶斯组合算法用于提高滚动轴承的故障识别率, 实现滚动轴承故障的有效识别。单一使用传统 KNN 算法或朴素贝叶斯算法在滚动轴承的故障识别中分类效果不佳, 不能有效的识别滚动轴承的故障状态, 达到工业应用标准。针对时域特征不明显的滚动轴承振动信号使用小波包能量法提取特征, 可以得到差异较大的一组特征量, 使用多个参数  $K$  值不同的 KNN 算法对数据进行初步分类, 得到一个初步分类结果集, 再使用朴素贝叶斯算法对初步分类结果集进行优化再分类, 可有效的提高分类精度, 实现滚动轴承故障的有效诊断, 在实际应用中有较好的推广意义。

### 参考文献:

[1] 王泽文. 基于振动信号的滚动轴承故障诊断与预测系统研究 [D]. 中国矿业大学, 2014.  
 [2] Kankar P K, Sharma S C, Harsha S P. Fault diagnosis of ball bearings using continuous wavelet transform [J]. Applied Soft Computing, 2011, 11 (2): 2300 - 2312.