

文章编号:1006-1355(2022)01-0144-06

基于WELCH算法集成学习模型的 滚动轴承故障诊断

张 龙, 周 俊

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘 要: 针对滚动轴承振动信号具有非平稳性、非线性且易受背景噪声干扰,故障特征难以提取等问题,提出一种基于WELCH功率谱算法的集成学习模型的故障诊断方法。首先使用WELCH算法对轴承的原始振动信号进行预处理,从中提取峭度、偏度、波形因子、峰值因子、脉冲因子和裕度因子6个参数,作为支持向量机的特征向量;然后结合集成学习算法构造Bagging-SVM集成学习模型。实验结果表明,与单一的SVM分类器相比较,Bagging-SVM集成模型对于轴承的故障诊断性能更优;在不同电机转速下的轴承故障诊断中,诊断率分别为97%,98%,98%和99.5%;说明了该集成模型在不同工况下的适用性强,诊断性能优秀。

关键词: 故障诊断;WELCH算法;集成学习;支持向量机;滚动轴承

中图分类号:TH133.33

文献标志码:A

DOI编码:10.3969/j.issn.1006-1355.2022.01.023

Rolling Bearing Fault Diagnosis Based on Ensemble Learning Model with WELCH Algorithm

ZHANG Long, ZHOU Jun

(School of Mechanical and Automobile Engineering, Shanghai University of Engineering Science,
Shanghai 201620, China)

Abstract: Aiming at the problem that the vibration signal of rolling bearings is non-stationary, non-linear and can be easily disturbed by background noise, and its fault characteristics are hardly to be extracted, this paper presents a fault diagnosis method based on the ensemble learning model with WELCH power spectrum algorithm. Firstly, the original vibration signal of the bearing is preprocessed by using the WELCH algorithm, and six parameters such as kurtosis, skewness, form factor, crest factor, impulse factor and margin factor are extracted from the signal and used as the feature vector of the support vector machine. Then, the ensemble learning algorithm is used to construct a Bagging-SVM ensemble learning model. Experimental results show that compared with the single SVM classifier, the Bagging-SVM ensemble model has a better bearing fault diagnosis performance. In the bearing fault diagnosis at different motor speeds, the diagnosis rates are 97%, 98%, 98%, and 99.5%. It shows that the ensemble model has strong applicability and excellent diagnostic performance under different working conditions.

Key words: fault diagnosis; WELCH algorithm; ensemble learning; support vector machine; rolling bearing

滚动轴承是最常见、最易受损的机械部件之一,作为机电设备中的核心零部件,它也是保证旋转机械高效稳定运行的重要部件。据相关统计显示,滚动轴承每年的故障率在35%左右,有超过一半的滚动轴承需要进行检查^[1-2]。滚动轴承的优劣对机

械设备能否高效稳定运行有很大的影响,因此对滚动轴承的质量检测、状态监测和故障诊断具有十分重大的意义。

现代化生产中,大型化、复杂化、高速化、自动化、智能化是机械设备未来发展的方向,旧的依赖于人的传统诊断方法已远远不能满足当前各式各样复杂化的系统需要,工业生产迫切需要融合智能传感网络、智能诊断算法和智能决策预示的智能诊断系统等。发展智能化的故障诊断方法是一条全新的途径^[3-4]。夏田等^[5]采用小波包分解方法对轴承信号进行分解,再计算每一频带的小波包能量作为轴承故障特征,结合梯度提升决策树构造轴承故障诊断模型。刘长良等^[6]先将原始轴承故障信号进行变分模

收稿日期:2021-03-18

作者简介:张龙(1995-),男,安徽省马鞍山市人,硕士研究生,专业方向为故障诊断。

E-mail: 1203391325@qq.com

通信作者:周俊(1967-),女,博士,副教授,研究生导师,专业方向为制造系统的建模、仿真及优化设计,机器学习,服务型制造。

E-mail: shuu_shun@163.com

态分解 (Variational mode decomposition, VMD), 再利用奇异值分解技术进一步提取各模态的特征, 最后采用标准模糊 C 均值聚类 (Fuzzy cmeans clustering, FCM) 进行轴承故障识别; 刘尚坤等^[7]先利用自适应 MED (Minimum entropy deconvolution) 降噪方法对轴承信号进行最优降噪处理, 再通过 EMD (Empirical mode decomposition) 分解出若干个 IMF (Intrinsic mode function) 分量, 选取峭度值最大的 IMF 进行包络谱分析, 通过其故障特征的频率实现故障诊断。全兆景等^[8]通过变步长粒子群算法优化的变分模态分解与 Hilbert 变换提取故障特征并将其作离散化处理, 结合贝叶斯网络构造故障诊断模型。周建民等^[9]对原始轴承信号采用时域方法和集成经验模态分解 (Ensemble empirical mode decomposition, EEMD) 能量熵提取轴承特征, 通过遗传算法 (Genetic Algorithm, GA) 优化 SVM (Support vector machine) 相关参数, 建立故障诊断模型。

上述方法存在两个主要问题: (1) 对于轴承信号预处理的方法处理过程复杂且计算量大; (2) 在对轴承故障进行分类时往往只针对单一分类器进行故障诊断研究。本文提出将 WELCH 功率谱算法与集成学习相结合的集成故障诊断模型。其中 WELCH 算法计算简单, 有利于滚动轴承故障特征的提取; 然后将单一的 SVM 分类器与 Bagging 算法相结合构造 Bagging-SVM 集成模型; 最后验证该集成模型在不同电机转速下的诊断性能以及抗噪性测试。

1 理论基础

1.1 WELCH 功率谱算法

WELCH 功率谱是一种有效且实用的经典的谱估计运算, 具有很好的信号分析能力。主要通过窗函数在被分析数据串上滑动截取数据并进行交叠, 使数据中被截取的段数增加, 并且对每一段被截取的数据都进行求取功率谱转换并平均后叠加, 这使得转换后的信号数据更加平滑。这样不仅降低了谱估计方差, 还减少了对信号分辨率破坏, 对信号所蕴藏的信息进行有效的保留, 且计算量相对较小, 对故障特征的提取相对平稳^[10-11]。

WELCH 算法进行功率谱转换主要有参数及窗函数的确定、数据分段、部分重叠几个步骤组成^[12]。

如果用 $\tilde{P}_{\text{PER}}(e^{j\omega})$ 表示 WELCH 算法计算得到的信号 $x(n)$ 功率谱估计, 则有:

$$\tilde{P}_{\text{PER}}(e^{j\omega}) = \frac{1}{\text{MUL}} \sum_{i=1}^L \left| \sum_{n=0}^{M-1} x_N^i(n) \omega(n) e^{-j\omega n} \right|^2 \quad (1)$$

其中 $x(n)$ 的长度为 N , 被分成了 L 段, 每一段包含 M

个数据, $\omega(n)$ 为窗函数, $U = \frac{1}{M} \sum_{n=0}^{M-1} \omega^2(n)$ 为归一化因子, 它的存在使得功率谱估计是渐进无偏估计。

1.2 集成学习——Bagging 算法

Bagging 算法是一种经典的并行式集成学习方法, 可以提高分类精度, 推广数据模式^[13]。首先给定一个数据集, 其中包含有 m 个样本, 从中有放回的随机抽取一部分样本作为训练集, 通过 T 次随机采样操作后, 产生了 T 个基学习器, 将所有的基学习器按照某种集成策略相互结合, 就构成了集成学习模型。具体步骤如下。

输入: 训练集

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\};$$

基学习算法 \mathfrak{S} ; 训练轮数 T 。

过程:

1: for $i=1$ to T do

2: $h_i = \mathfrak{S}(D, D_{bs})$

3: endfor

输出:

$$H(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{i=1}^T \mathbb{I}(h_i(x) = y)。$$

2 集成学习模型

2.1 基于 WELCH 算法的 Bagging-SVM 集成模型

滚动轴承的故障诊断主要有 3 个步骤: 原始信号预处理、故障特征提取和模式识别^[14]。将已知的故障状态的滚动轴承的原始振动信号先通过 WELCH 算法进行预处理, 然后从功率谱中提取相关特征参数, 输入到 SVM 分类器中, 进行滚动轴承的故障诊断。图 1 为 Bagging-SVM 集成模型。

2.2 集成策略

将所有基分类器的输出结果进行整合时, 通常采用以下两种集成策略:

(1) 多数投票法。假设基分类器有 T 个, 多数投票法就是先将 T 个分类器输出相同结果的个数统计起来, 最终的输出结果为基分类器相同个数最多的输出结果。当个数相同时, 则随机选择一个结果作为最终输出结果。

(2) 简单平均法。假设基分类器有 T 个, 简单平均法就是先累加所有分类器的输出结果, 通过平均后再进行输出, 即: $H(x) = \frac{1}{T} \sum_{i=1}^T h_i(x)$ 。由于最终的输出结果为对应的故障标签均为整数, 故在用简单平均法进行集成后, 再对结果采取了就近取整的原则。

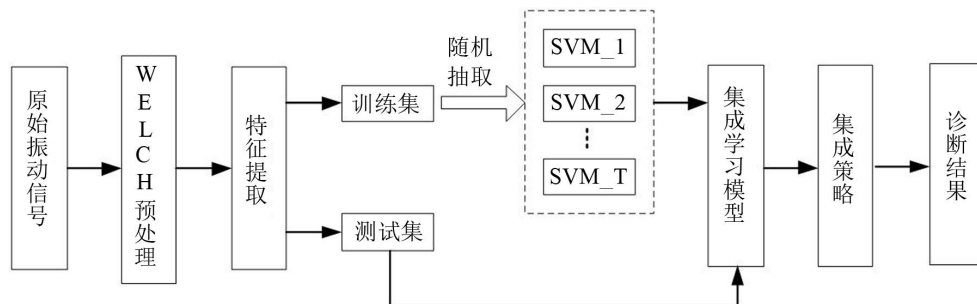


图1 基于WELCH算法的Bagging-SVM集成模型

3 实验结果与分析

本文实验的软件环境是 Windows1064 位操作系统，硬件配置为 Intel(R)Core(TM)i7-9750H CPU@2.60GHz，GPU 为 NVIDIA GeForce GTX16504 GGDDR5 独立显卡，内存为 16G，使用的 MATLAB 软件是 R2018a 版本。

3.1 数据介绍

数据集来自 Case Western Reserve University 轴承数据中心。选取了电机转速为 1 730 r/min、1 750 r/min、1 772 r/min、1 797 r/min，采样频率为 12 kHz 下的驱动端滚动轴承数据，分成 10 种故障类型，分别为正常状态以及内圈、外圈和滚动体故障直径分别为 0.18 mm、0.36 mm、0.53 mm 下的故障。每个样本包含 1 200 个采样点，其中每一种故障状态都选取了 100 个样本，总共 1 000 个样本数据。每种故障类型选取 80 个样本作为训练集，共 800 个，剩余的作为测试集，共 200 个。表 1 为转速为 1 772 r/min 下的实验数据集。

图 2 为电机转速 1 772 r/min 下的滚动轴承部分故障类型振动波形图，分别是正常状态以及内圈、滚

表 1 实验数据集

类别	故障类型	故障直径/mm
1	正常	—
2		0.18
3	内圈故障	0.36
4		0.53
5		0.18
6	滚动体故障	0.36
7		0.53
8		0.18
9	外圈故障	0.36
10		0.53

动体和外圈故障直径在 0.18 mm 下的故障状态。

3.2 特征提取

将原始的滚动轴承的原始振动信号通过 WELCH 算法进行预处理，选取采样点数为 1 200，FFT 点数为 600，采用矩形窗，窗函数长度设置为 100，重叠样本数为 25。图 3 是经过 WELCH 算法预处理后的功率谱。

从功率谱中获得峭度 X_{ku} 、偏度 X_{sk} 、波形因子 X_s 、峰值因子 X_c 、脉冲因子 X_z 和裕度因子 X_L 6 个参数，构成 SVM 的特征矢量，作为输入。部分相关数据见

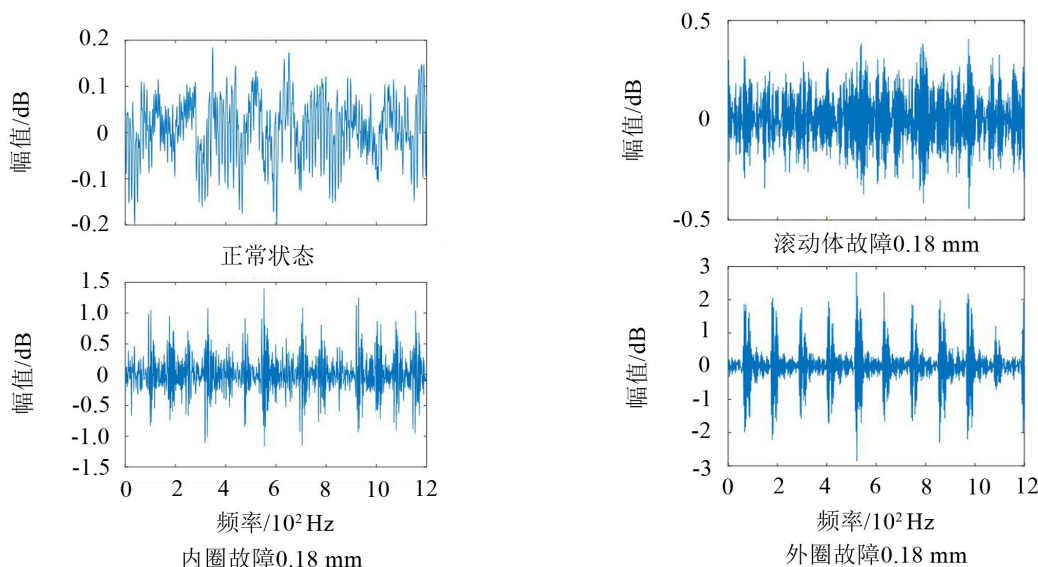


图2 不同故障类型的振动波形图

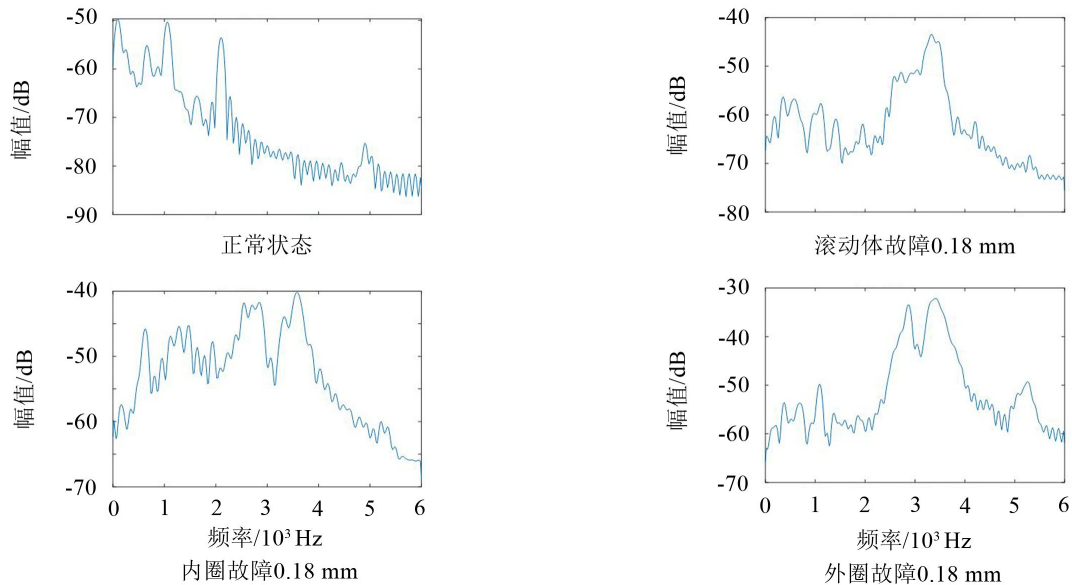


图3 WELCH法预处理后的功率谱

表2。

表2 部分SVM特征向量数据

X_{ku}	X_{dk}	X_s	X_c	X_z	X_l
2.226 9	0.699 0	1.009 4	-0.676 0	0.503 2	4.304 8
2.168 4	0.700 0	1.009 6	-0.668 5	0.526 4	4.496 0
2.079 6	0.684 1	1.009 7	-0.680 4	0.496 4	4.247 8
2.275 4	0.730 9	1.009 3	-0.676 4	0.515 8	4.401 2
.....

3.3 故障识别

首先通过对单个SVM分类器进行实验研究,采用RBF核函数,利用C-V交叉验证法求得最优的惩罚因子 $C=16$,核函数参数 $\gamma=2$ 。未进行预处理的故障诊断率为90.5%,经过WELCH预处理后的故障诊断率为96.5%。说明WELCH算法有利于故障特征的提取,SVM分类器对于轴承的故障诊断有一定的识别率,但仍有上升空间。本文则是在单个SVM分类器的基础上,结合Bagging算法,建立Bagging-SVM集成模型,将原始数据与经WELCH算法预处理过的数据进行对比实验,并在最终的集成策略上分别采用多数投票法和简单平均法进行对比实验。

实验一:设定分类器个数为20个,从数据集中有放回的随机抽取不同比例的样本数量构成训练子集进行实验,每次实验均重复进行10次,取其平均值。诊断结果如图4所示。两种集成方式下的诊断率随着样本数量的增加而不断上升;说明训练样本越多就越有利于诊断模型更好的将这些故障类型区分开来。

实验二:从数据集中有放回的随机抽取900个训练样本构成训练子集,设定不同数量的分类器进行实验,每次实验均重复进行10次,取其平均值。

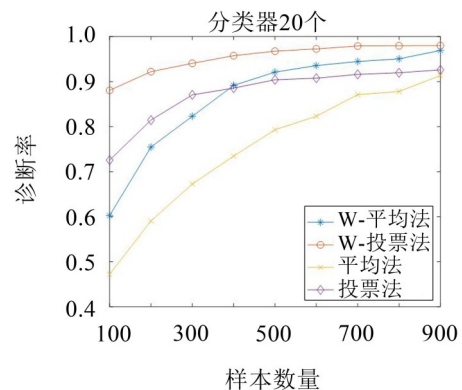


图4 不同样本数量下的诊断率

诊断结果如图5所示。当分类器个数为20个时,两种集成策略下的诊断率均达到最高,分别为98%和97%。在此之后,多数投票法下的诊断率趋于稳定,而简单平均法下的诊断率在97%左右波动。可见,当分类器超过一定数量时,诊断率便不再增加,甚至呈下降趋势;这表明产生了分类效果较差的弱分类器,影响了集成模型的性能,降低了故障诊断率。

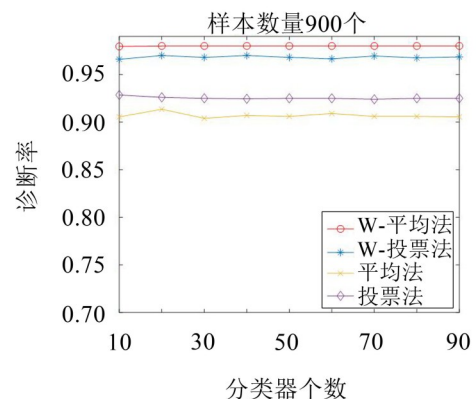


图5 不同分类器个数下的诊断率

WELCH算法预处理后的故障诊断率明显高于原始数据的故障诊断率;并且多数投票法的集成策略也优于简单平均法。因此设定分类器个数为20个,训练样本数量为900个,采用多数投票法进行实验,构造该数据下最优的集成模型,每次实验均重复进行10次,取其平均值。将其与单一SVM分类器的实验结果相对比,实验结果见表3。

表 3 诊断模型性能比较

故障类别	准确率/(%)	
	单一SVM	Bagging-SVM
1	100	100
2	100	100
3	100	100
4	85	90
5	95	100
6	95	95
7	100	100
8	95	100
9	100	100
10	95	95
平均准确率	96.5	98

3.4 不同工况下实验测试

本文提出的集成模型是针对电机转速为1 772 r/min下的滚动轴承的故障诊断,为验证该模型是否在其他电机转速下一样具有良好的诊断性能,故设定以下实验。将电机转速为1 730 r/min,1 750 r/min,1 772 r/min和1 797 r/min下的轴承数据分别用该集成模型进行实验,诊断结果如图6所示。相关数据见表4。

在图6中可以清楚地了解到本文提出的模型对于不同电机转速下的故障诊断效果相对均衡,其结果并无明显的波动,诊断率分别为99.5%,98%,98%和97%。这很好地证明了本文提出的模型同样适用于不同工况下的轴承故障诊断。

由表4可知,对于滚动体故障直径在0.36 mm的情况下,平均诊断率较低,仅为88.75%,甚至在电机转速为1 797 r/min下仅为75%。分析其原因,可能与滚动体本身的工作原理有关,导致所采集的振动数据不规则且具有随机性,从而导致诊断率偏低。但是在大部分情况下,对于不同转速下同种故障类

表 4 4 种不同电机转速下的诊断准确率/(%)

不同转速/(r·min ⁻¹)	1 730	1 750	1 772	1 797	平均准确率/(%)
正常	100	100	100	100	100
内圈故障0.18 mm	100	100	100	100	100
内圈故障0.36 mm	100	100	100	100	100
内圈故障0.53 mm	100	100	90	100	97.50
滚动体故障0.18 mm	100	100	100	100	100
滚动体故障0.36 mm	95	90	95	75	88.75
滚动体故障0.53 mm	100	100	100	100	100
外圈故障0.18 mm	100	100	100	100	100
外圈故障0.36 mm	100	100	100	100	100
外圈故障0.53 mm	100	90	95	95	95
准确率	99.50	98	98	97	-

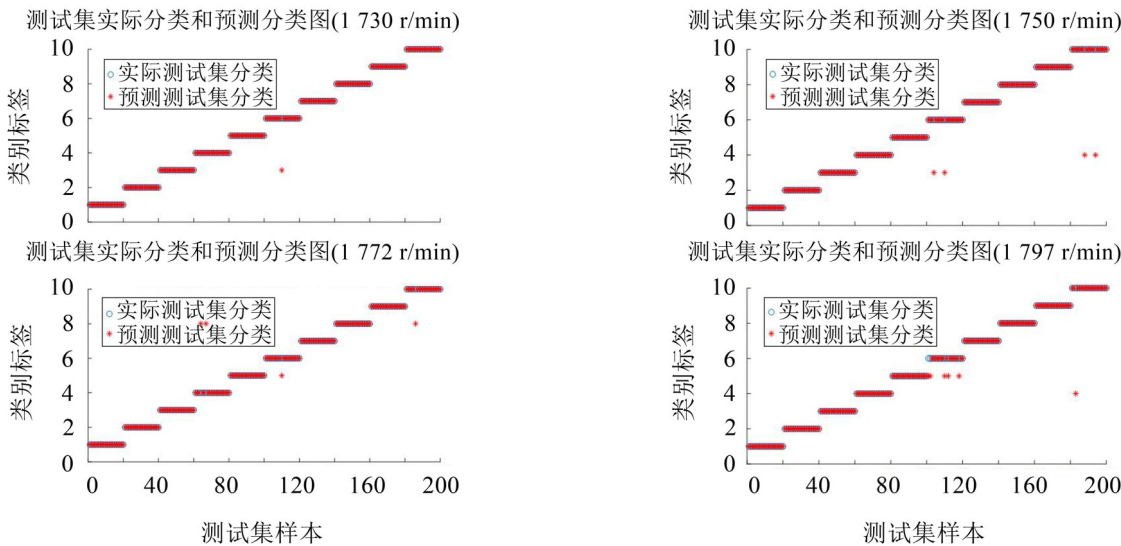


图 6 不同电机转速下的诊断结果

型,该集成模型的故障诊断性能稳定,平均诊断率为95%,97.5%,100%。

3.5 噪声实验测试

为验证该集成模型在噪声条件下的诊断性能,在电机转速为1 772 r/min的工况下,给每类故障的原始信号中分别加入信噪比(Signal-to-noise ratio, SNR)为-8 dB~8 dB的高斯白噪声构成带有噪声的数据集。每组实验均做10次,取平均值。实验结果如表5所示。

表5 不同信噪比下的模型诊断结果

序号	SNR/dB	平均准确率/(%)
1	-	98
2	8	96.45
3	6	92.45
4	4	92.3
5	2	90.25
6	0	88.4
7	-2	86.9
8	-4	86.85
9	-6	84
10	-8	81.15

SNR的定义如下:

$$SNR = 10 \lg \left(\frac{P_s}{P_n} \right) \quad (2)$$

式中: P_s ——原始信号能量大小;

P_n ——噪声信号能量大小。

由表5可知,本文提出的集成模型具有一定的抗噪性能,随着SNR值的减小,也就是说随着噪声干扰的增强,故障诊断准确率也在不断下降,但下降幅度相对平缓。在SNR达到-8 dB的情况下也能达到80%以上的准确率。

4 结 语

(1) 通过WELCH功率谱算法预处理后的轴承数据,能够很好地将各种故障类型区分开来,有利于提取故障特征,提高故障诊断率;相比于原始数据而言,两者的诊断误差最高能达到15%左右。

(2) 单一的SVM分类器,其诊断率为96.5%;而采用集成学习Bagging算法构造的Bagging-SVM集成模型诊断率在同种电机转速下能达到98%;在电机转速为1 730 r/min下甚至能达到99.5%;说明了集成模型能弥补单个分类器的不足,提高分类系统的分类能力。

(3) 当分类器数量不断增加时,故障的诊断率并不会一直上升,甚至可能出现下降的情况。由此可知,分类器个数并不是越多越好,对于分类器数量的确定以及适当的筛选一部分分类器进行集成,需要进一步的深入研究。并且集成策略的选择很大程度上会直接影响到最终的诊断结果,对于本文的集

成模型而言,多数投票法明显优于简单平均法,两者之间的诊断误差最高能达到20%左右。

(4) 本文是在电机转速为1 772 r/min下构造的集成模型,在变转速故障诊断中,除了转速为1 797 r/min下滚动体故障0.36 mm之外,其余故障类型均取得了很高的故障诊断率。

(5) 本文提出的集成模型具有一定的抗噪性,在SNR为-8 dB的情况下也能达到80%以上的诊断准确率。

参考文献:

- [1] 吕楠,姚平喜. 基于BP神经网络的滚动轴承故障诊断[J]. 煤矿机械, 2020, 41(8): 172-173.
- [2] 葛新峰,王志豪,顾青,等. 基于信号分析的滚动轴承故障诊断[J]. 数据采集与处理, 2010, 25(S1): 160-163.
- [3] 陈雪峰. 智能运维与健康管理[M]. 北京: 机械工业出版社, 2020.
- [4] 任世锦,潘剑寒,李新玉,等. 基于ELMD与改进SMS-VM的机械故障诊断方法[J]. 南京航空航天大学学报, 2019, 51(5): 693-703.
- [5] 夏田,詹瑶,郭建斌. 基于小波包和梯度提升决策树的轴承故障诊断[J]. 陕西科技大学学报, 2020, 38(5): 144-149.
- [6] 刘长良,武英杰,甄成刚. 基于变分模态分解和模糊C均值聚类的滚动轴承故障诊断[J]. 中国电机工程学报, 2015, 35(13): 3358-3365.
- [7] 刘尚坤,唐贵基. 自适应MED结合EMD诊断滚动轴承早期故障[J]. 噪声与振动控制, 2015, 35(6): 159-162.
- [8] 全兆景,芦彤,秦紫霓. 基于PSO-VMD与贝叶斯网络的滚动轴承故障诊断[J]. 河南理工大学学报(自然科学版), 2021, 40(1): 95-104.
- [9] 周建民,王发令,张臣臣,等. 基于特征优选和GA-SVM的滚动轴承智能评估方法[J]. 振动与冲击, 2021, 40(4): 227-234.
- [10] JIN ZH, HAN QC, ZHANG K, et al. An intelligent fault diagnosis method of rolling bearings based on Welch power spectrum transformation with radial basis function neural network[J]. Journal of Vibration and Control, 2020, 26 (9-10): 629-642.
- [11] 薛征宇,郑新潮,邱翔,等. 基于支持向量机的船舶感应电机轴承故障在线诊断方法[J]. 船海工程, 2020, 49(5): 1-5+9.
- [12] 杨婧,程乃平,倪淑燕. Welch算法在弱信号检测中的性能分析[J]. 计算机仿真, 2020, 37(5): 235-240.
- [13] WANG SM, ZHOU J, LI CQ, et al. Rockburst prediction in hard rock mines developing bagging and boosting tree-based ensemble techniques[J]. Journal of Central South University, 2021, 28(2): 527-542.
- [14] 涂志松. 基于小波包和支持向量机的滚动轴承故障诊断[J]. 机电工程技术, 2020, 49(12): 208-211.