

基于小波包和改进核最近邻算法的风机齿轮箱故障诊断方法^{*}

王栋瑾¹, 丁云飞¹, 朱晨烜¹, 孙佳林²

(1. 上海电机学院 电气学院, 上海 200240;

2. 上海电气风电公司, 上海 200241)

摘 要: 齿轮箱作为风力机组的核心部件, 故障频发, 研究风机齿轮箱的故障诊断方法意义重大。针对最近邻(KNN)诊断方法对离群噪声不敏感和诊断精度较低的缺陷, 提出了基于小波包和改进核最近邻算法的风机齿轮箱故障诊断方法。该方法应用小波包分析技术对故障特征进行提取, 利用互近邻准则将故障数据集中的离群噪声点剔除, 构建出基于核空间的改进型最近邻分类决策规则来识别齿轮箱的故障类型。试验表明: 该方法可以有效地提升故障诊断精度和鲁棒性, 为智能诊断技术的研究提供新思路。

关键词: 风机齿轮箱; 故障诊断; 小波包分析; 最近邻; 互近邻

中图分类号: TM 315; TP 29 **文献标志码:** A **文章编号:** 1673-6540(2019)01-0108-06

Fault Diagnosis for Wind Turbine Gearbox Based on Wavelet Packet and Improved Kernel K-Nearest Neighbors Algorithm^{*}

WANG Dongcui¹, DING Yunfei¹, ZHU Chenxuan¹, SUN Jialin²

(1. School of Electrical Engineering, Shanghai Dianji University, Shanghai 200240, China;

2. Shanghai Electric Wind Power Group, Shanghai 200241, China)

Abstract: As the core component of wind turbines, gearboxes frequently fail. It is significant to study the fault diagnosis methods of the wind turbine gearboxes. Considering that the K -nearest neighbors (KNN) diagnosis method was insensitive to noise and the accuracy of fault diagnosis was low, a fault diagnosis method based on wavelet packet and improved kernel K -nearest neighbors algorithm was proposed. This method used wavelet packet analysis technology to extract the fault features, and eliminated the noise by mutual nearest neighbor criterion. Then, an improved K -nearest neighbors classification decision rule based on kernel method was established. Experiments showed that this method could effectively improve fault diagnosis accuracy and robustness, and provide new ideas for the research of intelligent diagnosis technology.

Key words: wind turbine gearbox; fault diagnosis; wavelet packet analysis; K -nearest neighbors; mutual nearest neighbor

0 引 言

随着风能在清洁能源中的广泛使用, 大型风力发电机组的装机容量日趋增长^[1-2], 运行故障风险与维护成本也随之增加。据统计, 齿轮箱作

为双馈型风力发电机的核心部件, 由于长期工作在恶劣的工况条件下, 最易发生机械故障, 占总故障率的 65%^[3], 故风机齿轮箱的故障诊断成为热点课题, 国内外学者对此做了大量研究。以傅里叶变换为基础的时频分析^[4]早期在实际工程中得

^{*} 基金项目: 国家自然科学基金项目(11302123); 上海市浦江人才计划(15PJ1402500)

作者简介: 王栋瑾(1989—), 男, 硕士, 研究方向为故障诊断与智能控制。

丁云飞(1976—), 女, 教授, 研究方向为智能控制与模式识别。

到了应用,但其无法满足非平稳、非线性信号的精确分析要求,而小波分析^[5]以独有的多尺度特性和局部化特性突破了这一局限,在故障特征提取方面取得了丰硕的研究成果。卢锦玲等^[6]利用小波包分解方法提取风机齿轮箱故障的能量向量,并结合相关向量机成功地实现了风机齿轮箱的多分类故障诊断。杨明莉等^[7]针对风力机齿轮箱故障类型,提出基于多尺度小波分析和希尔伯特变换的实时齿轮箱故障诊断方法,通过分析能量包络谱相应的波形参数值来判定预测故障的类型。朱俊等^[8]采用小波包降噪技术提取出振动信号的特征值,与优化的神经网络相结合,诊断出风机齿轮箱轴承的故障。

随着人工智能技术的发展,基于智能算法的模式识别技术被引入故障诊断领域。其中,最近邻(KNN)分类算法^[9]以其结构简单、无需训练和估计参数等优点,广泛应用于故障模式识别。孙斌等^[10]提出一种基于局部切空间排列(LTSA)和KNN分类器的转子故障诊断模型,利用LTSA提取到低维特征向量作为KNN分类器的输入进行故障模式识别,在汽轮机转子的故障诊断中得到了验证。侯平智等^[11]将证据理论与KNN算法相结合,提出了基于KNN证据融合的故障诊断方法,在电机转子故障诊断中取得了显著的成果。陈法法等^[12]利用等距映射非线性流形学习提取故障特征,并与KNN算法相结合,对齿轮箱进行了有效的故障诊断。以上文献结合KNN算法均可实现对故障的分类识别,却忽视了近邻算法本身无法消除离群噪声点干扰的问题,没有考虑到样本权重的分配对分类贡献率的影响,从而制约了诊断精度的进一步提高,限制了KNN算法在工程领域中的应用与推广。

针对以上问题,本文提出基于小波包和改进核最近邻(1KNN)算法风机齿轮箱故障诊断方法。首先采用小波包分解^[13]把振动信号分解到相邻的不同频段上,提取感兴趣的频段成分进行重构,并计算出重构信号的能量作为特征向量。随后利用互近邻准则消除离群噪声点对故障分类的影响,在训练样本集中引入核空间和类隶属度的概念,细化样本的类别信息。最后建立新的分类决策机制,实现齿轮箱的故障诊断。

1 基本算法理论

1.1 KNN 分类算法

KNN分类算法属于基于先验知识的分类学习算法。现以待分类样本 $\mathbf{x}_s$ 为例简述KNN分类算法的计算步骤。首先设训练样本数据集 $U = \{(\mathbf{x}_i, c_i) | i=1, \dots, N\}$,包含 M 类训练样本,记类别集合为 $C = \{c_l | l=1, \dots, M\}$,其中训练样本 $\mathbf{x}_i$ 为 p 维的列向量, c_i 为 $\mathbf{x}_i$ 所对应的类别标签。

(1) 计算待分类样本 $\mathbf{x}_s$ 与所有训练样本的距离,通常采用欧氏距离作为衡量标准。其计算公式为:

$$d(\mathbf{x}_s, \mathbf{x}_i) = \sqrt{(\mathbf{x}_s - \mathbf{x}_i)^T (\mathbf{x}_s - \mathbf{x}_i)} \quad (1)$$

(2) 选择最近邻参数 k ,依据计算出的欧氏距离寻找出 $\mathbf{x}_s$ 的 k 个最近邻样本,记为 $Q_s = \{(\mathbf{x}_{s_j}, c_{s_j}) | j=1, \dots, k\}$ 。 $\mathbf{x}_{s_j}$ 表示待分类样本 $\mathbf{x}_s$ 的第 j ($1 \leq j \leq k$)个最近邻样本, c_{s_j} 为 $\mathbf{x}_{s_j}$ 所对应的类别。

(3) 根据 Q_s 中的 k 个最近邻样本的类别信息进行投票,投票结果记为 $v = [v_1, \dots, v_l, \dots, v_M]$,并按投票结果进行决策,决策规则为

$$c(\mathbf{x}_s) = \operatorname{argmax}_l (v_l) \quad (2)$$

1.2 互近邻去噪准则

为了去除离群噪声点的干扰,提高数据分类质量,本文引入互近邻准则来剔除离群点。所谓的互近邻就是互相为对方的近邻。在选定近邻参数 k 的前提下,若 y 是 x 的 K 近邻之一, x 也是 y 的 K 近邻之一,那么就称 x 与 y 互为彼此的近邻。例如图1所示为9个样本数据点的分布图,取近邻参数为3,可知1、2、3、4号样本点互为彼此的近邻,同理,5、6、7、8号样本点也互为彼此的近邻。对于9号样本点,其3个近邻样本点分别为5、6、8号,然而这3个样本点的近邻点却都不包含9号样本点,即9号样本点与5、6、8号样本点均不构成互近邻关系。事实上9号样本点不存在互近邻点,这是因为9号样本点距其他样本点较远,在一定程度上属于伪近邻,是一个孤立的离群点,即噪声数据点。用 $C_k(x)$ 来表示 x 的 K -互近邻的集合,则图1所反映的互近邻去噪准则可以简述为:设在数据集 S 中,测试样本为 x_j ,近邻个数为 k ,若 $C_k(x_j) = \emptyset$,则将 x_j 从数据集 S 中删除,

否则继续保留在数据集中。

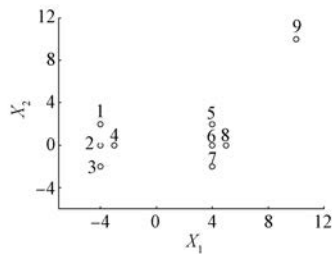


图1 9个样本数据点的分布图

1.3 IKKNN 分类算法

由于传统的 KNN 诊断方法依据近邻投票法则对待测样本进行决策分类,忽视了样本不同特征的差异,进而导致在样本数据不平衡时容易出现误分类的情况,降低分类精度。将隶属度的概念植入到 KNN 分类算法中,考察待分类样本与近邻样本的相似度,并将相似度与近邻样本点的类隶属度进行加权,综合考虑近邻样本对分类的贡献率,同时采用核方法,将低维空间线性不可分的样本映射到高维可分的特征空间中,提高算法的分类准确率。具体的算法流程如下。

(1) 确定训练集中每类训练样本的类重心 c_j 。

(2) 设数据集中包含 M 类训练样本,在核空间中计算训练样本 x_i 的类隶属度 T_{ij} 。具体的计算公式如下:

$$T_{ij} = \frac{\|\Phi(x_i) - \Phi(c_j)\|^{-\left(\frac{2}{m-1}\right)}}{\sum_{j=1}^M \|\Phi(x_i) - \Phi(c_j)\|^{-\left(\frac{2}{m-1}\right)}} \quad (3)$$

$$\|\Phi(x_i) - \Phi(c_j)\| = \sqrt{K(x_i, x_i) - 2K(x_i, c_j) + K(c_j, c_j)} \quad (4)$$

$$K(x_i, c_j) = \exp\left(\frac{-\|x_i - c_j\|^2}{2\delta^2}\right) \quad (5)$$

式中: $K(\cdot)$ ——高斯核函数;

δ ——核函数参数,一般取 1;

m ——权重系数。

(3) 选定近邻参数 k ,找出待分类样本 x_s 的 k 个最近邻训练样本,分别计算待分类样本 x_s 与各个最近邻样本间的相似度。相似度计算公式为

$$d_i = \frac{1}{\|\Phi(x_s) - \Phi(x_i)\|^{\frac{2}{q-1}}} \quad (6)$$

其中: q 为常数,一般取 2。

(4) 计算待分类样本 x_s 的类隶属度,进行分类决策。 x_s 的类隶属度及决策计算公式为

$$\mu_j(x_s) = \frac{\sum_{i=1}^k d_i T_{ij}}{\sum_{i=1}^k d_i} \quad (7)$$

$$c(x_s) = \operatorname{argmax}_j [\mu_j(x_s)] \quad (8)$$

2 基于 IKKNN 算法的故障诊断方法

齿轮箱的故障诊断大致可以分为试验数据的采集、故障特征的提取和故障模式的识别 3 个阶段。具体的诊断流程如图 2 所示。其中,试验数据采集部分将在第 3 节详细给出。对采集到的试验数据,首先利用小波阈值降噪^[14]对振动信号进行处理,达到衰减噪声的目的,本文选用 SYM8 小波降噪技术对原始信号进行降噪。考虑到振动能量可以有效地反映出故障信息,因而采用小波包分解和重构技术实现对故障特征的提取,具体的提取步骤在第 3.2 节给出。

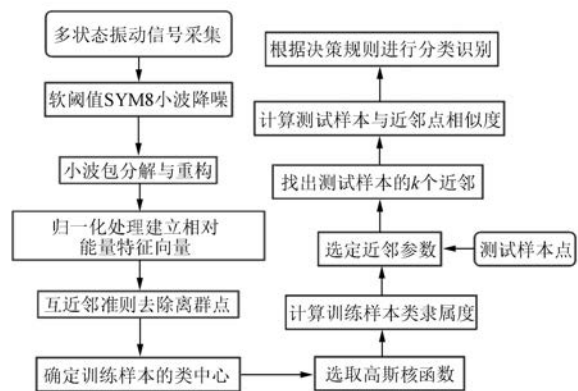


图2 基于 IKKNN 算法的故障诊断流程图

为了方便进行数据处理,将提取到的小波包特征能量进行数据的归一化处理,并以此作为故障特征向量。最后,依据 IKKNN 分类算法对特征向量进行模式识别。具体步骤如下。

(1) 根据互近邻准则,剔除故障样本集中的离群点,消除离群噪声点对故障模式识别的影响。

(2) 从剔除离群点后的样本数据中选取若干数据,作为训练样本和测试样本。

(3) 确定训练样本的类中心,选取高斯核函数,在核空间中计算训练样本的类隶属度。

(4) 选定近邻参数 k , 找出测试样本的 k 个近邻。

(5) 分别计算测试样本与 k 个近邻训练样本的相似度。

(6) 计算测试样本的类隶属度, 并根据分类决策规则, 对测试样本的所有类隶属度进行比较, 得出分类结果。

基于以上说明, 构建出基于 IKNN 算法的风机齿轮箱故障诊断模型, 并传统的 KNN、加权 KNN (WKNN)、模糊 KNN (FKNN) 以及支持向量机 (SVM) 等其他类型的诊断方法进行对比, 验证该方法的有效性。

3 试验与结果分析

3.1 风机齿轮箱故障模拟试验

试验所采集的风机齿轮箱振动信号来自上海电气风电集团的故障诊断试验平台系统, 试验平台由风机齿轮箱、变速驱动电机、轴承、主轴、联轴器、制动器、传感器和数据采集系统组成。选择振动能量较为集中和突出的齿轮箱输出端的轴承压盖作为测点位置, 并布置压电式加速度传感器, 测取振动加速度信号, 如图 3 所示。

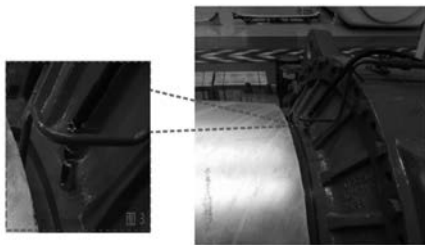


图 3 风机齿轮箱测点布置图

试验根据风机齿轮箱常见的几种故障模式, 分别模拟了齿轮正常状态以及由断齿和磨损失效形式所引起的 3 种故障状态 (大齿轮断齿故障、小齿轮磨损故障、大齿轮断齿+小齿轮磨损的混合故障)。为保证模拟的故障振动信号与实际所产生的齿轮故障信号相似, 试验采用电火花加工技术对齿轮进行了处理。通过在不同的齿轮上分别植入损伤点来模拟不同类型的齿轮失效形式。数据采集系统在采样频率为 5 120 Hz, 主轴转速为 1 500 r/min 的情况下, 采集到每种状态数据各 25 组, 共 100 组试验样本。试验所采集的 4 种振动

信号的时域波形如图 4 所示。

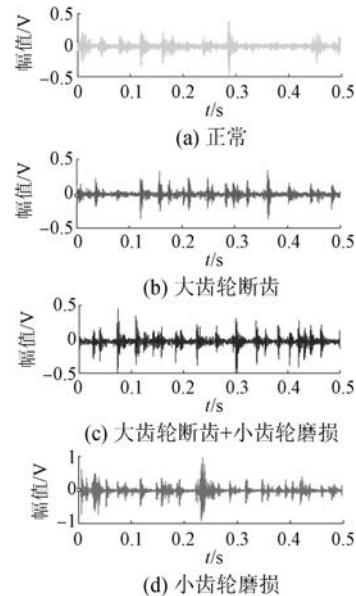


图 4 4 种状态下的时域波形图

3.2 故障特征提取

合理地提取到信号的特征是提高故障诊断精度的前提。考虑到振动信号中的频带能量包含了丰富的故障信息, 以能量为元素构造出的特征向量可以区分出风机齿轮箱的故障类型。因此, 采用能够对非平稳信号进行有效分析的小波包变换对信号特征进行提取, 提取步骤如下。

(1) 利用 db3 小波对信号进行 3 层小波包分解, 分别提取第 3 层从低频到高频 8 个频带成分的小波包分解系数。

(2) 对第 3 层小波包分解系数进行重构, 提取各频带信号。用 $S_{3j} (j=0, 1, \dots, 7)$ 表示第 3 层小波包分解系数的单支重构信号。

(3) 计算各频带信号的能量, 构造能量特征向量。设 S_{3j} 对应的频带能量为 E_{3j} , 则:

$$E_{3j} = \int |S_{3j}|^2 dt = \sum_{k=1}^n |x_{jk}|^2 \quad (12)$$

其中: x_{jk} 为重构信号 S_{3j} 的离散点的幅值 ($j=0, 1, \dots, 7; k=1, 2, \dots, n$); n 为振动信号的采样点个数。

(4) 能量所反映的故障信息通常是一个比较大的数值, 为方便进行数据处理, 需要对提取到的小波包能量进行归一化处理。最终可得能量特征向量表示为

$$T = [E_{30}, E_{31}, E_{32}, E_{33}, E_{34}, E_{35}, E_{36}, E_{37}] \quad (13)$$

3.3 结果分析

为了验证所提方法的有效性,分别利用传统的 KNN 诊断方法和所提方法对故障类型进行分类识别并比较试验结果。设置近邻参数为 3, 顺次选取 4 种状态下的训练样本各 10 组作为训练样本集, 将剩余的 60 组作为待测样本进行分类识别。图 5 所示为基于 KNN 的故障分类试验结果。传统的 KNN 的诊断算法分类性能较差, 有 4 个测试样本被错误分类, 并且由于无法识别离群点, 使离群样本被错分为大齿轮断齿故障, 其故障识别精度仅为 86.67%; 而改进的诊断算法利用了互近邻准则可以有效地甄别出这 4 个离群点样本, 并将其剔除。如图 6 所示, 改进方法所得到的故障识别精度达到了 100%, 良好地克服了传统的 KNN 算法受离群点干扰的缺陷, 准确地区分出了故障类型。图 7 所示为试验所用的 40 组训练样本和 4 个离群点(黑色小方块)样本的三维散点分布。

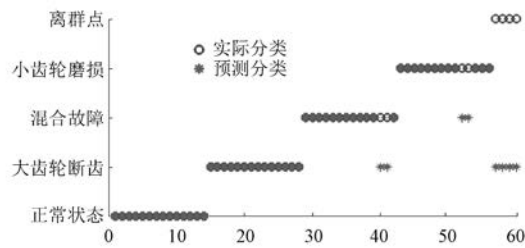


图 5 基于 KNN 的故障分类结果($k=3$)

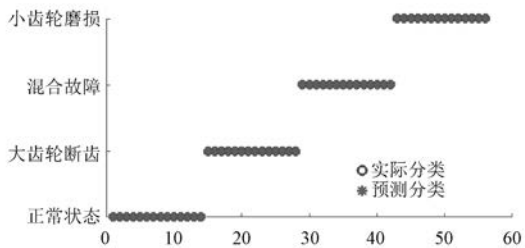


图 6 基于 IKKNN 的故障分类结果($k=3$)

近邻学习依赖于先验信息, 训练样本越多, 先验知识就越丰富, 反之亦然。因此本文增加了不同的分类方法随样本数目变化的对比试验分析。将已知的故障样本按比例划分, 分别随机选取 80%、60%、40% 和 20% 的样本作为训练样本集, 相应地将剩余样本作为测试样本集。为保证试验结果的客观性, 采用交叉验证^[15]的方法, 在近邻

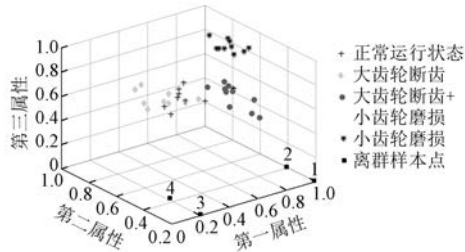


图 7 训练样本点和离群点的三维散点分布图

参数为 3 的情况下, 分别利用不同的近邻分类方法对测试样本进行诊断识别, 每组试验均进行 N ($N=10$) 次取平均值, 所得试验结果如表 1 所示。

表 1 不同分类器的故障识别率($k=3$)

分类器 类型	训练样本数占比			
	80%	60%	40%	20%
KNN	93.25	91.68	88.84	83.82
FKNN	97.95	95.33	92.65	89.68
WKNN	97.83	95.60	93.03	91.25
IKKNN	99.75	99.01	96.53	95.28
SVM	99.25	98.87	95.33	93.75

结果表明, 在不同的训练样本数目下, IKKNN 诊断方法所获得的识别率均明显高于其他 3 种近邻诊断方法。这主要是由于 IKKNN 方法所构建的隶属度量化了不同特征对样本分类的贡献, 使得样本分类更加灵活, 同时将低维空间中线性不可分的样本映射到高维可分的特征空间中, 整体优化了分类性能。就 SVM 分类方法而言, 其故障识别率接近 IKKNN, 但由于需要训练样本模型会增加计算开销进而影响诊断效率。

3.4 鲁棒性分析

近邻信息的不确定性导致近邻参数的选择势必会对诊断结果产生影响。所以本文考察了不同的近邻分类诊断算法随 k 值变化的鲁棒性。取训练样本 60 组, 测试样本 40 组, 进行试验, 并分析比较试验结果, 如图 8 所示。

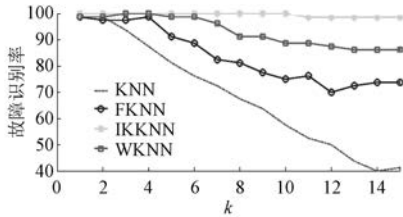


图 8 不同分类器随 k 值变化的故障识别率

随着参数 k 的变化, 4 种不同的近邻诊断方法所得到的故障识别率均有所变化。其中传统的 KNN 方法的分类性能随着近邻参数的增加而逐渐恶化, 故障识别率大幅降低, 无法准确地识别故障类型; FKNN 诊断方法的识别精度虽高于 KNN, 但也有所降低, 特别是在 $k=4$ 以后, 其故障识别率出现波动并开始下降, 分类性能逐步退化; WKNN 的故障识别精度整体上要高于上述两种方法, 特别是在近邻参数 $k=6$ 之前, 表现出一定的鲁棒性, 但随着 k 值的增大 ($k=6$ 以后), 故障识别率呈现下降趋势, 鲁棒性开始受到影响。相比上述几种方法, 本文所提的 IKKNN 诊断方法体现出了明显的优势。随着近邻参数的变化, 故障识别率没有出现大的波动, 始终保持在 95% 以上的较高水平, 表现出了良好的鲁棒性和适应性。

4 结 语

本文提出的诊断方法, 借助互近邻准则将故障数据中的离群噪声点剔除, 有效地克服了基于近邻决策的诊断方法对噪声点不敏感的缺陷, 体现出融合互近邻准则的改进方法在故障诊断中的优越性。

利用核方法和训练样本的类重心重新定义了样本的隶属度和样本间的相似度, 建立起新的分类决策机制, 弥补了传统的近邻算法在贡献权重分配上的缺陷, 充分考虑了不同特征属性对分类的影响, 提升了故障类型的识别精度。

通过对比试验, 综合评价了不同方法的故障诊断能力。就分类性能而言, 本文所提故障诊断方法不但具有较高的故障识别率, 而且随近邻数参数 k 的变化, 还表现出了较强的鲁棒性, 体现了该方法在故障诊断中的优越性。该方法可为风机齿轮箱智能诊断技术的研究提供参考。

【参考文献】

[1] 曾军, 陈艳峰, 杨苹, 等. 大型风力发电机组故障诊断综述[J]. 电网技术, 2018, 42(3): 849.
[2] 张文秀, 武新芳. 风电机组状态监测与故障诊断相关技术研究[J]. 电机与控制应用, 2014, 41(2):

50.
[3] 徐志翔. 基于改进 PSO 和参数优化的 LSSVM 的风力发电机齿轮箱故障诊断[D]. 上海: 华东理工大学, 2012.
[4] RICHARD C, LENGELLÉ R. Joint recursive implementation of time-frequency representations and their modified version by the reassignment method [J]. Signal Processing, 1997, 60(2): 163.
[5] MELTZER G, DIENB N P. Fault diagnosis in gears operating under non-stationary rotational speed using polar wavelet amplitude maps[J]. Mechanical Systems & Signal Processing, 2004, 18(5): 985.
[6] 卢锦玲, 绳菲菲, 赵洪山. 基于相关向量机的风机齿轮箱故障诊断方法[J]. 华北电力大学学报(自然科学版), 2017, 44(2): 91.
[7] 杨明莉, 刘三明, 张卫, 等. 基于振动分析的风力机齿轮箱故障诊断[J]. 电机与控制应用, 2015, 42(1): 66.
[8] 朱俊, 刘天羽, 王致杰, 等. 基于蜂群算法的选择性神经网络集成的风机齿轮箱轴承故障诊断[J]. 电机与控制应用, 2017, 44(1): 6.
[9] 刘凡, 张昀, 姚晓, 等. 基于 K 近邻算法的换流变压器局部放电模式识别[J]. 电力自动化设备, 2013, 33(5): 89.
[10] 孙斌, 刘立远, 牛翀. 基于局部切空间排列和 K -最近邻分类器的转子故障诊断方法[J]. 中国机械工程, 2015, 26(1): 74.
[11] 侯平智, 张明, 徐晓滨, 等. 基于 K 近邻证据融合的故障诊断方法[J]. 控制与决策, 2017, 32(10): 1767.
[12] 陈法法, 汤宝平, 苏祖强. 基于等距映射与加权 KNN 的旋转机械故障诊断[J]. 仪器仪表学报, 2013, 34(1): 215.
[13] 于志伟, 苏宝库, 曾鸣. 小波包分析技术在大型电机转子故障诊断系统中的应用[J]. 中国电机工程学报, 2005, 25(22): 158.
[14] 武瑞兵. 基于 8 层小波包分解的电机定子电流故障诊断新方法[J]. 电机与控制应用, 2015, 42(4): 32.
[15] 汪庆华, 刘江伟, 张兰兰. 交叉验证 K 近邻算法分类研究[J]. 西安工业大学学报, 2015, 35(2): 119.

收稿日期: 2018-07-05