

基于特征筛选和改进深度森林的变压器内部机械状态声纹识别^{*}

李楠¹, 马宏忠¹, 张玉良¹, 段大卫¹, 崔佳嘉¹, 何萍²

(1. 河海大学 能源与电气学院, 江苏 南京 211100; 2. 国网南京供电公司, 江苏 南京 210008)

摘要: 变压器声纹信号包含大量反映内部机械状态的有效信息。为实现变压器内部机械状态不停电检测, 提出一种基于特征筛选和改进深度森林的变压器机械状态声纹识别方法。首先, 利用自适应噪声完备集合经验模态分解(CEEMDAN)声纹信号得到本征模态函数(IMF), 并通过频谱分析和皮尔逊相关系数对IMF分量进行筛选, 得到包含故障信息的IMF分量。其次, 利用各IMF分量在频段上的分布情况进行高、低频段划分, 依据高、低频段IMF分量的差异性, 将高频段IMF分量的时频能量和低频段IMF分量的幅值特性作为特征指标, 构成特征向量, 输入改进后的深度森林模型, 得到10种机械松动状态的声纹识别结果。最后, 通过现场试验验证了该方法的有效性。研究表明: 所提方法对10种机械松动状态的平均识别准确率达99.2%。与传统变压器声纹特征相比, 所提声纹特征区分度更高; 与传统识别模型相比, 所提改进深度森林识别模型复杂度更低, 训练速度更快, 识别准确率更高。

关键词: 电力变压器; 特征筛选; 深度森林; 模态分解; 声纹识别

中图分类号: TM407 **文献标志码:** A **文章编号:** 1673-6540(2022)09-0057-09

doi: 10.12177/emca.2022.099

Voiceprint Recognition of Transformer Internal Mechanical State Based on Feature Screening and Improved Deep Forest^{*}

LI Nan¹, MA Hongzhong¹, ZHANG Yuliang¹, DUAN Dawei¹, CUI Jiajia¹, HE Ping²

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China;
2. State Grid Nanjing Power Supply Company, Nanjing 210008, China)

Abstract: Transformer voiceprint signal contains a lot of effective information reflecting the internal mechanical state. In order to realize uninterrupted detection of internal mechanical state of transformer, a voiceprint recognition method of transformer mechanical state based on feature screening and improved deep forest is proposed. Firstly, the intrinsic mode function(IMF) is obtained by decomposing the voiceprint signal with the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and the IMF component containing the fault information is obtained by filtering the IMF components through spectrum analysis and Pearson correlation coefficient. Secondly, the distribution of each IMF component in the frequency band is used to divide the high and low frequency bands. According to the difference of the IMF components in the high and low frequency bands, the time-frequency energy of the IMF component in the high frequency band and the amplitude characteristic of the IMF component in the low frequency band are used as characteristic indicators to form a feature vector, which is input into the improved deep forest model, and the voiceprint recognition results of 10 mechanical loose states are obtained. Finally, the

收稿日期: 2022-07-04; 收到修改稿日期: 2022-08-23

^{*} 基金项目: 国家自然科学基金项目(51577050); 国网江苏省电力公司科技项目(J2021053)

作者简介: 李楠(1997—), 男, 硕士研究生, 研究方向为电力设备故障诊断。

马宏忠(1962—), 男, 博士, 教授, 研究方向为电力设备故障诊断。

effectiveness of the method is verified by field experiments. The research results show that the average recognition accuracy of the proposed method is 99.2% for 10 mechanical loose states. Compared with the traditional transformer voiceprint feature, the proposed voiceprint feature has higher discrimination; Compared with the traditional recognition model, the proposed improved deep forest recognition model has lower complexity, faster training speed and higher recognition accuracy.

Key words: power transformer; feature screening; deep forest; mode decomposition; voiceprint recognition

0 引言

电力变压器是电力系统的核心设备之一,其可靠运行与电力系统的安全稳定密切相关^[1-2]。机械故障是变压器主要故障类型,也是部分电气故障、发热等问题的主要诱因,具有累计效应,易造成绝缘老化加速、抗短路能力降低等隐患^[3-4]。国家电网统计表明,由绕组故障引起的变压器事故占变压器总事故的 69.8%^[5-6]。因此,开展变压器绕组及铁心机械状态监测与识别研究,对电力系统的安全稳定运行具有重要意义。

目前,现有的变压器机械故障诊断方法包括频率响应法^[7]、低压脉冲法^[8]、短路电抗法^[9]、油色谱分析法^[10]以及振动分析法^[11]等。频率响应法、低压脉冲法以及短路电抗法需要离线检测并且一定程度地影响变压器寿命。油色谱分析法可以实现在线监测,但难以准确地识别故障类型及确定程度。振动分析法具有无电气联系、可以在线检测等优点,但对振动传感器安装位置、安装方式要求严苛,尤其是测点位置的改变极易造成测量结果相差较大,因此该方法难以得到推广。

变压器运行时产生的声纹信号包含大量反映设备内部机械状态的有效信息。基于声纹的变压器故障诊断方法具有非接触、测量方便快捷以及可以在线监测等优点,且能够有效解决振动分析法中空间敏感度要求高的问题,受到广泛关注。文献[12]提出了基于 Mel 频率倒谱系数(MFCC)和矢量量化算法的变压器声纹识别模型,对不同铁心压紧状态下的变压器声纹信号进行识别。文献[13]使用 MFCC 进行特征提取,建立了基于快速增量式以及门控循环单元的变压器机械故障声纹识别模型。文献[14]提取 MFCC 特征量,并建立两级电气故障声纹识别算法,对 4 种不同工况进行识别。MFCC 利用人类听觉感知试验得到的 Mel 非线性谱,进行声纹特征提取,目前已成为主

流的声纹识别特征向量提取方法。然而,MFCC 特征参数主要针对人对声音的感知特性,考虑到变压器声纹信号与人声之间存在较大差异,MFCC 难以用于变压器声纹识别。因此,亟需一种切实有效的变压器声纹特征提取方法,以提高声纹识别准确率。

以深度神经网络为代表的深度学习算法因其强大的数据处理与挖掘能力受到广泛关注。文献[15]利用卷积神经网络在图像深度数据挖掘方面的优势,提出了基于 Mel 时频谱-卷积神经网络的变压器铁心声纹识别模型,对 3 种不同工况下的声纹信号进行识别。文献[16]利用深度神经网络中的门控循环单元强大的特征学习能力,提出一种改进的 MFCC 和门控循环单元的变压器偏磁声纹识别方法。尽管上述方法证明了深度学习用于声纹识别的有效性,但所提模型均存在网络结构复杂、对样本数量要求高、训练时间长以及过拟合等问题。

为了利用深度神经网络挖掘故障特征的优势,同时避免建立模型,深度森林应运而生。文献[17]提出基于稀疏深度森林的调相机轻微定子匝间短路故障方法,取得了较高的诊断精度和效率。文献[18]提出一种改进的新型深度森林算法,有效解决了滚动轴承剩余寿命预测精度差、运算效率低等问题。然而,目前深度森林还未应用于变压器机械状态声纹识别^[19-21]。

为此,本文提出一种基于特征筛选和改进深度森林的变压器机械状态声纹识别方法。先利用自适应噪声完备集合经验模态分解(CEEMDAN)将声音传感器采集到的声纹数据分解成多个本征模态函数(IMF)分量,并根据频谱分析和皮尔逊相关系数进行 IMF 分量优化选择,筛选出有效的 IMF 分量。再利用各 IMF 分量在频段上的分布情况划分高、低频段,依据高、低频段 IMF 分量的差异性,提取高频段 IMF 分量时频能量和低频段

IMF 分量幅值特征构成特征向量。为平衡模型的级联层数和精度,以提高运算效率,对深度森林进行改进,并将特征向量输入改进后的深度森林模型,输出 10 种机械松动状态的声纹识别结果。最后,在现场实测的基础上对所提方法进行验证。

1 模态分解与筛选

1.1 模态分解

经验模态分解(EMD)是一种自适应时频分析方法,通过多次迭代得到多个单 IMF 分量,但由于极值点的选择问题会出现模态混叠现象。为解决极值点的选择问题,集合经验模态分解(EEMD)引入高斯白噪声,利用高斯白噪声频谱均匀分布特性,使信号自动分布到合适的参考尺度上,有效抑制 EMD 的模态混叠现象,但存在计算速度慢和模态分解结果中存在多个无效低频 IMF 分量等问题。

CEEMDAN 在 EEMD 方法基础上,通过加入残值、高斯白噪声和多次叠加求平均以抵消噪声等操作,进一步减少模态混叠,提高了分解的完备性以及对噪声干扰的鲁棒性。具体算法原理如下所示。

1.1.1 第一阶段模态分量

对于加入高斯白噪声的原始信号:

$$X(i) = y(t) + k_0 n^i(t) \quad (1)$$

式中: $y(t)$ 为原始信号; $n^i(t)$ 为满足 $N(0,1)$ 分布的高斯白噪声; k_0 为第一阶段高斯白噪声的幅值。

将带噪原始信号 EMD 分解,并将分解得到的 IMF_{1,i} 分量平均得到:

$$X_{\text{IMF}_1} = \frac{1}{N} \sum_{i=1}^N X_{\text{IMF}_{1,i}} \quad (2)$$

式中: N 为分解后的模态分量个数。

1.1.2 第二阶段模态分量

在第一阶段模态分量的基础上计算第一阶段残余分量:

$$x_1(t) = X(t) - X_{\text{IMF}_1} \quad (3)$$

分解加入白噪声的第一阶段残余分量:

$$r_1(t) = x_1(t) + k_1 \text{EMD}_1[n^i(t)] \quad (4)$$

式中: $\text{EMD}_i[n^i(t)]$ 为分解信号得到的第 i 个分量; k_1 为第二阶段高斯白噪声幅值。

将带噪第一阶段残余分量 EMD 分解,并将得到的 IMF_{2,i} 分量平均得到:

$$X_{\text{IMF}_2} = \frac{1}{N} \sum_{i=1}^N X_{\text{IMF}_{2,i}} \quad (5)$$

1.1.3 第 k 阶段模态分量

对于 k 取 $3, 4, \dots, k$, 同样重复上述步骤计算 $x_k(t)$ 和 $r_k(t)$, 并通过 EMD 分解及平均操作,直到剩余信号不能被分解,得到第 k 阶段模态分量为

$$X_{\text{IMF}_k} = \frac{1}{N} \sum_{i=1}^N X_{\text{IMF}_{k,i}} \quad (6)$$

1.2 模态筛选

1.2.1 频谱分析

现有研究表明,变压器声纹信号分布于 $0 \sim 2\,000$ Hz 的频率范围内,且以 50 Hz 及其倍频分量为主。本文试验中的声纹信号主要由 3 种频率成分构成:(1) 频点为 0 的直流分量,属于传感器测量所造成的直流偏移;(2) $0 \sim 50$ Hz 之间的频率范围,属于测试环境中的背景噪声;(3) 50 Hz 及其倍频分量,是变压器本体产生的声音成分。

1.2.2 皮尔逊相关系数

皮尔逊相关系数用以度量两组向量之间的关联程度,本文选择皮尔逊相关系数量化各个 IMF 分量与原始信号之间的关联程度,进而刻画 IMF 分量中包含特征信息的丰富程度,其值在 $-1 \sim 1$ 之间,且越接近 1,表示两者之间的相关性越强。皮尔逊相关系数 ρ_{XY} 计算公式如下所示:

$$\rho_{XY} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (7)$$

式中: $X = \{x_1, x_2, x_3, \dots, x_n\}$; $Y = \{y_1, y_2, y_3, \dots, y_n\}$ 。

2 特征提取

不同 IMF 分量在频谱中表现出较大的差异性。为依据不同频段 IMF 分量的特点,有针对性地提取故障特征,可以将 IMF 分量分为能量大多集中于 500 Hz 以下的低频段 IMF 分量和 500 Hz 以上的高频段 IMF。低频段 IMF 分量频谱简单,其主频集中于 100 Hz、200 Hz 以及 300 Hz,且能量集中、幅值较大。高频段 IMF 分量频谱复杂,表现出非线性、非平稳特征,且能量分散、幅值较小。本文针对高、低频段 IMF 分量差异性提取不

同特征,并进行归一化构成故障特征量。

2.1 低频段 IMF 分量幅值

低频段 IMF 分量相对平稳,频谱简单且集中。针对此特点,本文提取低频段中各 50 Hz 及其倍频信号幅值之和作为低频段 IMF 分量特征(简称为 LFA)。其计算方法为先利用短时傅里叶变换,计算各 50 Hz 及其倍频信号幅值对应的幅值 $f_{(50 \times i)}$ 并将其叠加,计算式如下所示:

$$f = \sum_{i=1}^{10} f_{(50 \times i)} \quad (8)$$

式中: i 为整数, $i=1,2,3,\dots,10$ 。

2.2 高频段 IMF 分量时频能量

高频段 IMF 分量频谱复杂,且表现出非线性、非平稳的特征。小波能量熵针对高频段 IMF 特征,融合小波变化和能量熵优势,能够从时间和频率两方面监测能量分布变化情况,因此,本文选择小波能量熵(简称为 WEE)对高频段 IMF 分量进行特征提取,利用特征向量构成识别参数,输入识别模型,计算过程如下所示。

2.2.1 小波时频分析

对声音信号进行分帧以及加窗处理后,对声音信号进行连续小波变换,先确定小波基与尺度,而后求出小波系数,其计算公式如下所示:

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \psi\left(\frac{t - \tau}{a}\right) dt \quad (9)$$

式中: $x(t)$ 为输入声音信号序列; ψ 为母小波,采用复数小波 Complex Morlet,其在时频两域具有很好的分辨率,适合处理非平稳的声音信号。

2.2.2 IMF 分量能量计算

从小波系数序列中提取小波系数 $coefs_{(50 \times i)}(n)$,对于给定小波系数 $coefs_{(50 \times i)}(n)$:

$$coefs_{(50 \times i)}(n) = [coefs_{(50 \times i)}(1), coefs_{(50 \times i)}(2), coefs_{(50 \times i)}(3), \dots, coefs_{(50 \times i)}(N)] \quad (10)$$

定义时间序列数据平方和 E_j ,其为该时间序列数据能量:

$$E_j = \sum_{i=1}^N |coefs_{(50 \times i)}(i)|^2 \quad (11)$$

2.2.3 总能量值计算 E

总能量值 E 的计算式为

$$E = \sum_{j=1}^A E_j \quad (12)$$

式中: A 为 CEEMDAN 分解的 IMF 分量个数。

2.2.4 IMF 能量熵计算

IMF 能量熵计算式为

$$E_n = - \sum_{j=1}^A \frac{E_j}{E} \ln \frac{E_j}{E} \quad (13)$$

3 深度森林

深度森林主要包括多粒度扫描和级联森林。多粒度扫描是通过扫描窗口遍历以提高子样本数量的方法。级联森林采用级联方式连接各层,且每一层均由随机森林和完全随机森林组成。本文对多粒度扫描和级联森林进行参数调整和模型改进,提高了运算效率和识别准确率。

3.1 多粒度扫描

多粒度扫描利用滑动扫描窗口遍历样本,提取时间序列数据空间位置信息。扫描窗口为 L 维向量,扫描滑动步长为 1,对于 N 维时间序列数据样本,多粒度扫描后将获得 $N-L+1$ 个 L 维子样本。利用随机森林对子样本进行训练,产生 $N-L+1$ 个 K 维概率特征向量,将上述 $N-L+1$ 个概率特征向量利用张量压缩拼接成 $K(N-L+1)$ 维向量,作为后续级联森林输入特征。

3.2 级联森林

级联森林是一种借鉴决策树混合堆叠思想改进的多层级结构,每一级联层中包含随机森林和完全随机森林,其训练后生成的特征向量为类向量,生成的类向量和原始数据向量拼接构成新的向量,作为下一层的输入特征向量。当训练达到较好效果或训练至最后一层时,对各层输出的概率特征向量取平均值,则最大的类概率值所对应标签为最终输出结果。本文在原始模型的基础上,增加 XGBoost 基学习器以调用 GPU 参与运算,提高计算速度和效率。此外,增加逻辑回归基学习器以降低数据资源存储要求,大幅度提高级联森林网络模型的性能。

级联森林是多层堆叠,随着级联层数增加,深度森林复杂度线性增加,但是随着层级增加,训练准确率提升逐渐变缓。为避免无效级联,提高模型训练效率,引入神经网络中的识别准确率和交叉熵损失值作为级联结束判定指标。在模型训练以及层级增加过程中,当识别准确率连续 5 次未提高或损失值连续 5 次未降低,则层级增加终止,避免无效的级联层数增加,以提高模型运算效率。

模型训练准确率计算公式如下所示:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

式中: TP 为将正类预测为正类的数量; TN 为将负类预测为负类的数量; FN 为将正类预测为负类的数量; FP 为将负类预测为正类的数量。

模型训练损失值计算公式如下所示:

$$L = -\frac{1}{s} \sum_{i=1}^s \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (15)$$

式中: s 为样本个数; M 为多分类类别数; y_{ij} 为样本标签, 当样本 i 的真实预测类别标签为 j 时取 1, 否则取 0; p_{ij} 为观测样本 i 属于 j 的概率。

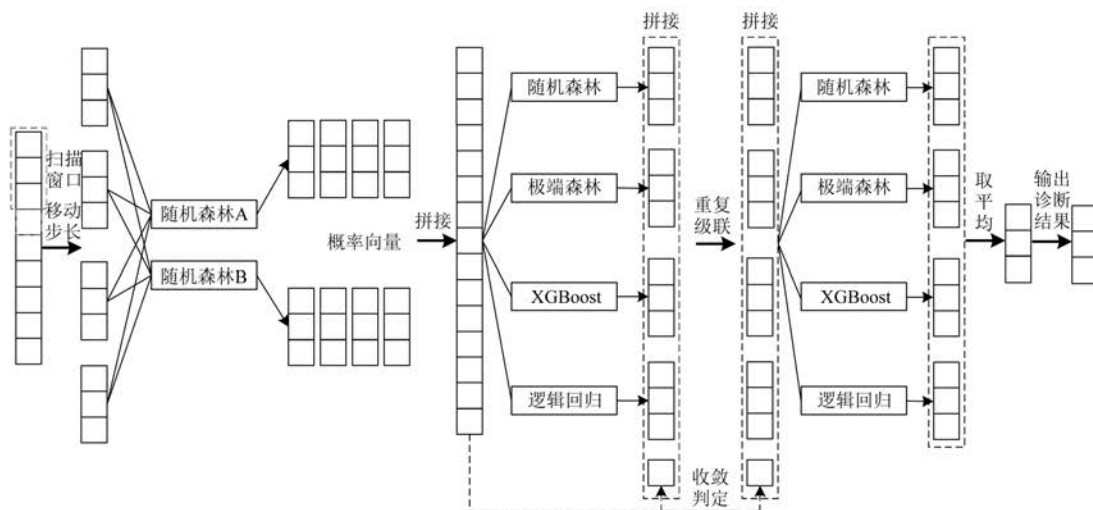


图1 改进深度森林模型

4 试验与分析

4.1 试验描述与频谱分析

为验证本文方法的有效性,在句容变压器厂搭建变压器声纹数据采集平台,以一台型号为 S-13-M-200/10 的 10 kV 油浸式配电变压器为研究对象,通过试验模拟其在不同工况下发生机械故障,并通过采集平台采集声纹信号。试验平台包括电脑、DHDAS 动态信号采集仪、信号传输线、前置放大器 HS14618 以及电容式声传感器 HS14401 等如图 2 所示。本文设置采集仪的采样频率为 50 kHz 以提高试验精度,按照国标 GB/T

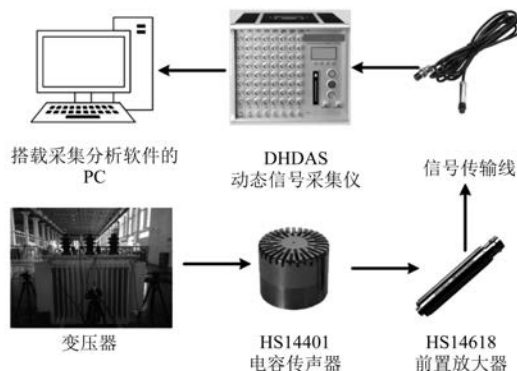


图2 变压器声纹数据采集平台

1094.10—2003 规范,传感器布置在变压器四面,距离变压器油箱 30 cm,距离地面高度 35 cm,利用龙门吊和扭力扳手设置不同类型的机械故障。

变压器空载试验下,绕组中仅流过励磁电流,不足以引起绕组的振动,此时声音信号主要来源于铁心振动。变压器短路试验下,绕组振动引起的声音远大于铁心振动引起的声音。本文开展空载试验和短路试验,如图 3 所示。为了尽可能采

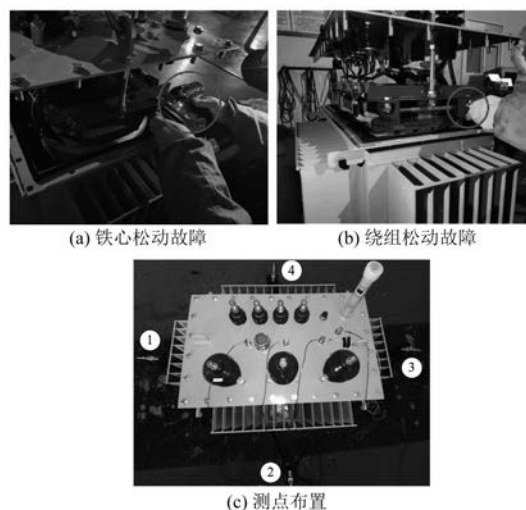


图3 变压器机械松动故障试验

集到丰富的声音信号,本文将4个相同型号的声音传感器均匀地放置在变压器箱体的4个侧面。1号和3号传感器对应的箱体侧面记为“左右侧面”,2号和4号传感器对应的箱体侧面记为“上下侧面”,以研究变压器铁心、绕组的松动机械故障,具体试验流程如下。

(1) 空载试验下的铁心松动故障。设置变压器高压侧开路,低压侧保持额定电压,利用龙门吊吊心,并使用扭力扳手调整铁心水平方向紧固螺丝的松紧程度,调整预紧力为 0 、 $0.25F_N$ 、 $0.5F_N$ 、 $0.75F_N$ 和 F_N ,此处 F_N 为额定预紧力,最后装机静置采集声纹数据。

(2) 短路试验下的绕组松动故障。设置变压器低压侧短路,高压侧逐渐增加电压至绕组中电流达到额定电流时停止,通过重复变压器吊心,并利用扭力扳手调整变压器绕组垂直方向上的紧固螺丝松紧程度,分别调整预紧力为 0 、 $0.25F_N$ 、 $0.5F_N$ 、 $0.75F_N$ 和 F_N ,静置测量变压器声纹数据。

试验采集并切片构成11 000个样本,对样本进行分类、附标签以及频谱分析,将其分为铁心松动和绕组松动2种松动故障类型。每种松动故障又被分为 0 、 $0.25F_N$ 、 $0.5F_N$ 、 $0.75F_N$ 和 F_N 5种不同预紧力,对应100%、75%、50%、25%以及正常5种松动程度,并按照9:1划分训练集与测试集,各类别标签及样本个数如表1所示。

表1 样本标签和样本个数

故障类型	松动程度	标签	样本数量/个
铁心松动	100%	1	1 120
	75%	2	1 320
	50%	3	1 400
	25%	4	950
	正常	5	1 310
绕组松动	100%	6	870
	75%	7	1 130
	50%	8	870
	25%	9	1 040
	正常	10	900

如图4所示,变压器声纹信号主频为100 Hz,主要成分为50 Hz及其偶次倍频分量,50 Hz奇次倍频分量含量较少,且多集中于低频段。变压器铁心及绕组松动程度的变化,只改变特征频点的幅值,未改变频谱分布。此外,在低于50 Hz的频段也存在少量分布,属于测试环境中的背景噪声。

此外,2号测点频谱稳定性高且信号相对简单,以其为代表进行分析处理。

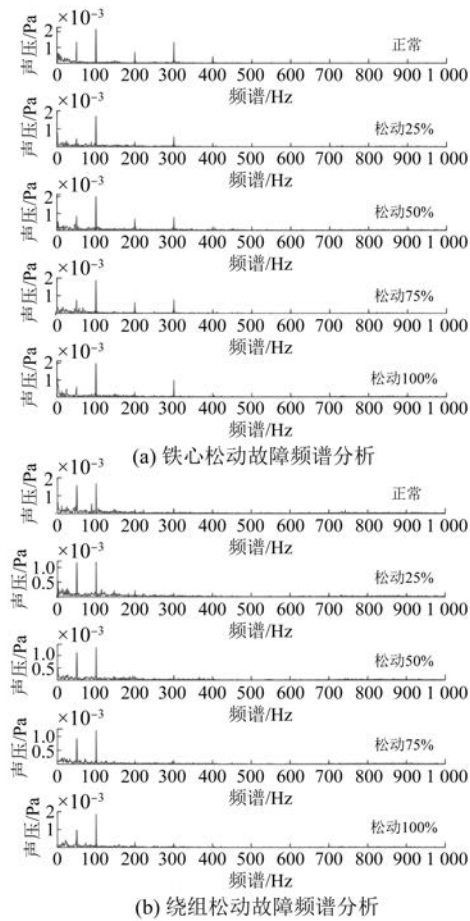


图4 声纹信号频谱分析

在不同预紧力作用下,声纹信号频谱变化各不相同,但未表现出明显的规律性,且故障特征提取较困难。因此,迫切需要切实可行的故障特征提取方法。

4.2 变量筛选

为验证本文所提方法的有效性,对真实声纹数据进行分解,摒弃由于传感器自身及测量所造成的直流偏移和外界环境噪声干扰等无关分量,保留包含故障特征的IMF分量。如图5所示,IMF₁~IMF₃为全频段的均匀噪声,且其幅值远低于其他IMF分量。IMF₁₂~IMF₁₃所包含的特征分量均为低于50 Hz的低频段,其属于测试环境中的背景噪声。IMF₁₄为传感器及测量造成的直流偏移。因此,本文舍弃无关分量,保留IMF₄~IMF₁₁等包含故障特征的分量。

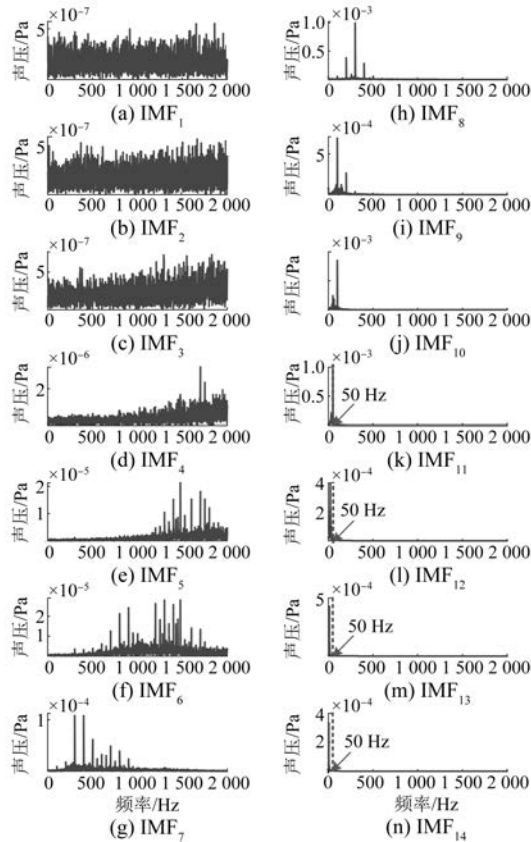


图5 IMF分量频谱分析

在去除背景噪声分量的基础上,通过皮尔逊相关系数刻画原始信号与各 IMF 分量之间及各 IMF 分量之间的关联程度,刻画各 IMF 分量包含信息的丰富程度和独立性。各 IMF 分量相关系数,如图 6 所示。IMF₄ 与原始信号之间的皮尔逊相关系数仅为 0.008 4,相关程度较低。各 IMF 分量之间的相关系数均小于 0.2,说明各分量之间相对独立。

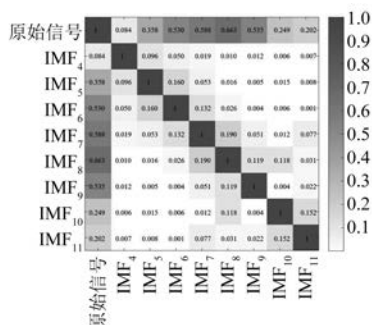


图6 IMF分量相关系数图

4.3 特征量提取

变压器声纹信号经过 CEEMDAN 分解后,对各

IMF 分量进行频谱分析和相关系数筛选,得到包含故障特征信息的 IMF 分量。针对不同 IMF 分量特点,利用短时傅里叶变换和小波时频熵分别提取不同频段的 IMF 分量特征,同时随机选择提取 MFCC 一帧进行对比。以 5 种不同程度铁心松动故障为例,特征提取结果如图 7 所示。MFCC 对于不同程度铁心松动故障区分度较差,存在大量交叠,且特征维度远超其他特征。短时傅里叶变换幅值特征提取,针对简单平稳的 IMF₈~IMF₁₁ 具有良好的区分度,但对复杂非平稳的 IMF₅~IMF₇ 存在交叠现象,难以有效区分。小波能量熵针对复杂非平稳的 IMF₅~IMF₇ 具有较好的区分效果,但不适用于简单平稳的 IMF₈~IMF₁₁。本文针对不同分量特征,融合不同特征提取方法优点,提出筛选特征方法,所提取的特征对有效的 IMF₅~IMF₁₁ 均具有较高区分度。

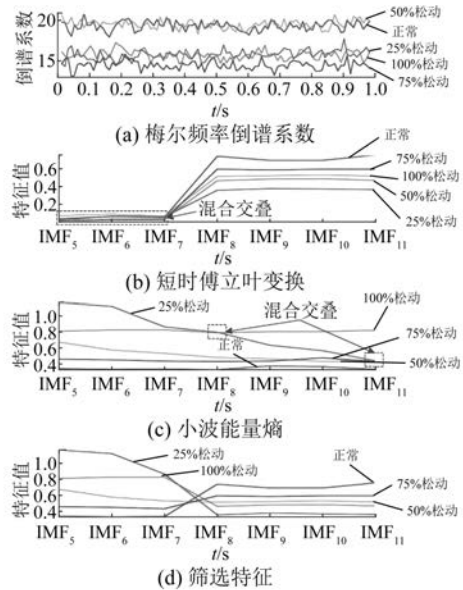


图7 特征值区分度折线图

4.4 识别模型搭建与训练

为避免变电站其他变压器声纹信号对测试变压器产生干扰,在识别模型前增加辅助判据。当变压器处于空载状态且 50 Hz 频点幅值均小于 0.001 Pa,启动铁心松动程度识别程序。当变压器负载状态下且 50 Hz 频点幅值均小于 0.001 5 Pa,启动绕组松动程度识别程序。

本文针对 10 种类别、共计 6 000 组样本数据,随机提取 80% 样本作为训练集、20% 样本作为测试集。在经过 CEEMDAN 分解后,选取 IMF₅~

IMF₁₁ 的样本熵和能量熵所构成的特征向量,作为深度森林的输入进行分类识别。

本文对改进深度森林中的 3 个重要超参数进行优化,包括学习率、基学习器中树的数量以及 k 折交叉验证中的 k 值。其中 k 折交叉验证的 k 值分别取 3,4,5,⋯,15,基学习器中树的数量分别取 100,200,300,⋯,1 000,学习率分别取 0.1,0.01,0.001,0.000 1。为防止偶然性出现,每种超参数取值后进行 5 次重复试验,将 5 次试验结果取平均值进行对比,如图 8 所示。本文设置 k 折交叉验证中的 k 值取 12,基学习器中树的数量取 700,学习率为 0.001。

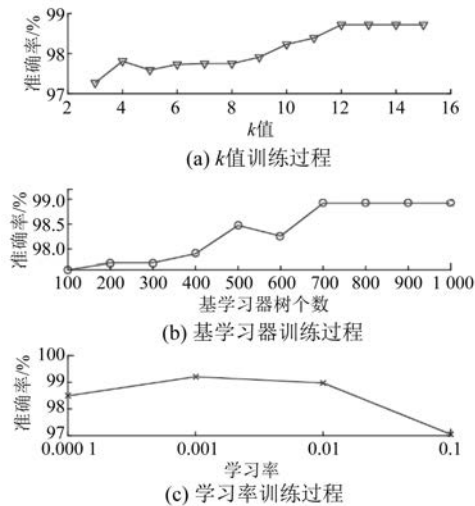


图 8 识别模型超参数训练过程

按上文重新设置模型超参数,并导入样本集对模型进行重新训练,本文采用混淆矩阵来衡量模型对 1 100 个测试样本的识别分类结果。如图 9 所示,本文所提模型对绕组及铁心不同松动程度故障均能有效识别,平均识别率达 98.7%。

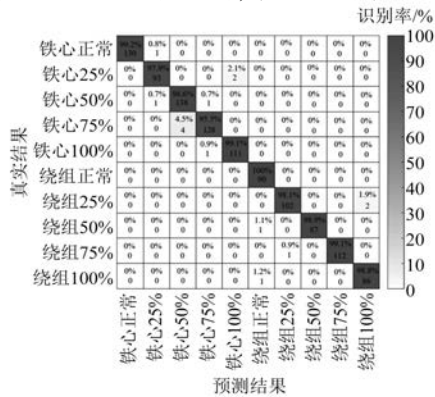


图 9 识别结果混淆矩阵

4.5 模型优越性验证

为验证本文所提特征筛选方法的有效性,本文对比 MFCC、IMF₁ ~ IMF₁₄ 中提取的倍频幅值、IMF₁ ~ IMF₁₄ 中提取小波能量熵、IMF₅ ~ IMF₁₁ 中提取的幅值以及 IMF₅ ~ IMF₁₁ 中提取的小波能量熵 5 种特征值,分别记为 MFCC、LFA₁₋₁₄、WEE₁₋₁₄、LFA₅₋₁₁、WEE₅₋₁₁。对 1 100 组测试样本进行测试,测试结果如图 10 和表 2 所示。MFCC 特征识别准确率较低、训练时间较长。对比 IMF₁ ~ IMF₁₄ 所提特征和 IMF₅ ~ IMF₁₁ 所提特征,IMF₁ ~ IMF₁₄ 所提特征的识别准确率均低于 IMF₅ ~ IMF₁₁ 所提特征的识别准确率,IMF₁ ~ IMF₁₄ 所提特征的训练时间均高于 IMF₅ ~ IMF₁₁ 所提特征的训练时间。可知无效的 IMF 分量所提特征不仅降低识别准确率且大幅度增加训练时间。本文所提筛选的特征识别准确率较 LFA₁₋₁₄、LFA₅₋₁₁、WEE₁₋₁₄、WEE₅₋₁₁ 以及 MFCC 5 种特征,至少提高了 0.8%,且针对 10 类松动故障均有明显提升。

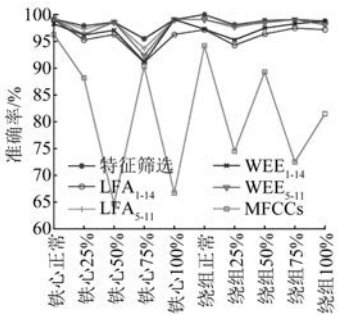


图 10 不同特征识别结果对比图

表 2 各故障特征训练时间和识别准确率

特征名称	训练时间/s	测试时间/s	平均准确率/%
特征筛选	77.20	0.016	99.2
LFA ₁₋₁₄	121.61	0.124	93.7
LFA ₅₋₁₁	97.10	0.021	98.4
WEE ₁₋₁₄	115.71	0.138	94.4
WEE ₅₋₁₁	91.15	0.133	98.0
MFCC	180.82	0.019	81.8

为验证本文所提识别算法的优越性,将常见的机器学习算法如决策树(DT)、朴素贝叶斯(NB)、随机森林(RF)、K近邻(KNN)、支持向量机(SVM)以及 RUSBoost 等与本文所提的改进深度森林模型进行对比。对 1 100 组样本进行测试,测试结果如图 11 所示。对比其他常见机器学习模型的平均识别准确率以及单一故障识别准确

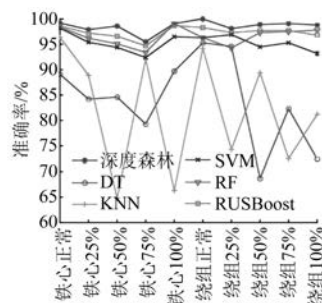


图 11 传统机器学习算法识别结果对比图

率,所提方法均有所提高。

为验证改进深度森林算法比深度学习算法拥有更强的泛化能力,与深度学习算法中的一维卷积神经网络(ID-CNN)的训练过程和识别结果进行对比,如图 12 所示。CNN 模型在训练过程中能够快速收敛,训练集准确率最终稳定在 100%。虽然改进的深度森林算法识别准确率存在一定程度的波动性,但卷积神经网络测试集识别准确率仅为 62.1%,改进深度森林测试集准确率达到 99.2%,可知改进深度森林对变压器机械状态的识别具有更高的泛化能力。

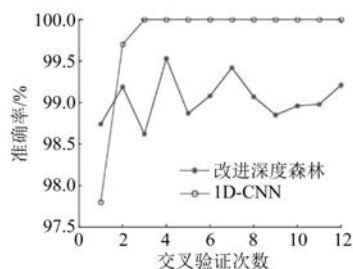


图 12 卷积神经网络识别结果对比图

5 结 语

变压器运行过程中发出的声音信号包含大量能够反映设备内部机械状态的有效信息。本文搭建变压器声纹信号采集平台,对采集到的声纹信号进行分析与处理,并对变压器不同机械状态的声纹信号进行识别。在后续的研究中,会对现场实际运行的变压器正常及故障数据进行收集处理,以进一步验证本文所提方法的有效性。本文得到如下结论。

(1) 利用 CEEMDAN 对声纹信号进行分解,通过频谱分析和皮尔逊相关系数对 IMF 分量进行筛选,有效剔除无效的 IMF 分量,保留了包含

变压器本体声纹信号的 IMF 分量。

(2) 通过对不同 IMF 分量在高低频段上表征出的差异性,提取高频段 IMF 分量的时频能量和低频段 IMF 分量的幅值特性作为特征指标,具有针对性的提取特征量,有效提高了特征向量的区分度,降低了特征向量的冗余,在一定程度上减少了识别模型复杂程度。

(3) 利用改进的深度森林算法对变压器机械状态进行识别,取得了较好的识别效果。

【参 考 文 献】

- [1] 李鹏,毕建刚,于浩,等. 变电设备智能传感与状态感知技术及应用[J]. 高电压技术, 2020, 46(9): 3097.
- [2] 谢国民,倪乐水. 基于 IABC 优化 SVM 的变压器故障诊断[J]. 电力系统保护与控制, 2020, 48(15): 156.
- [3] 朱保军,咸日常,范慧芳,等. WRSR 与改进朴素贝叶斯融合的变压器故障诊断技术研究[J]. 电力系统保护与控制, 2021, 49(20): 120.
- [4] 方涛,钱晔,郭灿杰,等. 基于天牛须搜索优化支持向量机的变压器故障诊断研究[J]. 电力系统保护与控制, 2020, 48(20): 90.
- [5] 王梦云. 110 kV 及以上变压器事故与缺陷统计分析[J]. 供用电, 2007, 24(1): 1.
- [6] 王梦云. 2004 年度 110 kV 及以上变压器事故统计分析[J]. 电力设备, 2005, 6(11): 31.
- [7] 甘锡淞,李云,傅成华,等. 基于信息融合和 CS-SVM 的变压器绕组变形故障诊断方法研究[J]. 电力系统保护与控制, 2018, 46(1): 156.
- [8] 李宏达,黄鼎琨,张彬,等. 改进的低压脉冲法对变压器绕组变形的探测研究[J]. 南京理工大学学报, 2020, 44(1): 15.
- [9] 徐建源,陈彦文,李辉,等. 基于短路电抗与振动信号联合分析的变压器绕组变形诊断[J]. 高电压技术, 2017, 43(6): 2001.
- [10] 明菊兰,钱洲亥,牛莉,等. 基于最小二乘原理的电力变压器绝缘油色谱在线监测系统校准方法[J]. 武汉大学学报(理学版), 2021, 67(3): 281.
- [11] 曹辰,徐博文,李辉. 基于振动与电抗信息的变压器绕组形变状态综合监测方法[J]. 高电压技术, 2022, 48(4): 1518.
- [12] 王丰华,王邵菁,陈颂,等. 基于改进 MFCC 和 VQ 的变压器声纹识别模型[J]. 中国电机工程学报, 2017, 37(5): 1535.

(下转第 74 页)