

DOI: 10.12158/j.2096-3203.2024.03.022

# 基于声纹压缩和代价敏感的变压器状态检测评估方法

胡赵宇<sup>1</sup>, 李喆<sup>1</sup>, 陈海威<sup>2</sup>, 陆忻<sup>2</sup>

(1. 上海交通大学电气工程系, 上海 200240)

(2. 中国能源建设集团广西电力设计研究院有限公司, 广西 南宁 530007)

**摘要:** 声纹检测技术可以助力巡检人员对变压器状态进行检测和评估。文中提出一种基于声纹压缩和代价敏感的变压器状态检测和评估方法。该方法首先提取变压器音频的声纹特征, 然后在频率维度上对声纹特征进行筛选和压缩, 最后使用卷积神经网络评估变压器状态, 并引入代价敏感损失函数以提高对难检出样本的关注度。以某 35 kV 变压器为研究对象, 通过收集现场音频、模拟实验和样本扩充得到变压器音频数据集。测试结果表明, 文中所提方法将声纹维度从 1 025 维降低到 80 维, 计算量和显存分别降低到 1 025 维的 8.1% 和 7.7%。同时, 所提方法的声纹识别准确率高达 83.5%, 并将最难检出的短路电流异常状态的召回率从 48.2% 提升至 63.6%。

**关键词:** 变压器检测; 声纹识别; 声纹压缩; 代价敏感; 卷积神经网络; 模式识别

中图分类号: TM73

文献标志码: A

文章编号: 2096-3203(2024)03-0209-08

## 0 引言

随着我国城市化的推进和电力需求的增长, 电力变压器的应用越来越广泛。电力变压器的正常稳定运行对人们正常的生产生活具有重要意义, 因此, 需要加强对变压器的在线检测和状态评估工作, 为巡检和维修提供参考。

根据吴晓文<sup>[1]</sup>、高沛<sup>[2]</sup>、王宏升<sup>[3]</sup>等对变压器振动原理的分析和实验可知, 变压器处于不同工作状态时, 铁心、绕组等结构的振动特性不同, 振动产生的声音也不同。因此, 可以使用声纹识别技术<sup>[4]</sup>对变压器的工作状态进行在线检测。声纹识别技术包括声纹特征提取和声纹识别 2 个环节, 首先从音频中提取声纹特征作为音频的低维表征, 然后使用声纹识别算法实现音频的分类识别。

声纹特征方面, 除了梅尔时频谱图、梅尔倒谱系数(Mel frequency cepstrum coefficient, MFCC)、线性预测倒谱系数<sup>[5-7]</sup>等语音领域常用的特征, 不少学者还提出了新的变压器声纹特征。王丰华等使用加权处理法和主成分分析法对 MFCC 特征进行了改进<sup>[8]</sup>, 该方法主要从统计的角度实现了对音频信息的高效压缩, 但对变压器的振动特性考虑不足。刘云鹏等<sup>[9]</sup>和李楠等<sup>[10]</sup>根据变压器音频主要分量集中在 50 Hz 倍频这一特点, 分别设计了 50 Hz 倍频倒谱系数和 50 Hz 倍频小波时频熵。这 2 种方法结合了变压器振动特性和常用的音频处理方式, 提高了识别准确率, 但人为指定 50 Hz 倍频也存在可能

遗漏或者忽视部分频率信息的问题。

声纹识别算法方面, 神经网络算法因其强大的模式识别能力而得到广泛应用。刘云鹏等<sup>[9]</sup>和崔佳嘉等<sup>[11]</sup>分别将门控循环单元和卷积神经网络应用于变压器声纹识别, 大幅提高了声纹识别的准确率。但这 2 种方法都将变压器的不同状态简单地划分为不同类别, 没有考虑不同状态之间的联系, 也缺少对音频数量和难度不平衡的处理方法, 制约了模型效果的进一步提高。

综合上述研究现状, 文中提出基于声纹压缩和代价敏感的变压器状态检测评估方法。该方法使用递归消除法和梯度下降微调法对声纹特征进行压缩和优化, 使用代价敏感的损失函数训练卷积神经网络, 提高神经网络对难检出样本的召回率, 训练后的神经网络从电压电流工况和故障情况 2 个维度对变压器状态进行检测和评估。为验证所提方法的有效性, 以某 35 kV 变压器为对象, 通过收集现场音频、模拟实验和样本扩充得到数据集作为算例, 以供分析验证。

## 1 变压器音频数据集制作

使用模型对变压器的状态进行识别, 首先要收集变压器在不同状态下的音频。由于在真实工作环境下录制的变压器音频大多为变压器正常状态下的音频, 因此需要通过模拟实验来补充比较少见的变压器状态音频。首先介绍变压器状态模拟实验的方案, 然后介绍用于丰富变压器音频数据集的样本扩充方法。

### 1.1 变压器状态模拟实验

变压器状态模拟实验如表 1 所示。

收稿日期: 2023-11-18; 修回日期: 2024-01-25

基金项目: 国家自然科学基金资助项目(52077133)

表 1 变压器状态分类和模拟实验内容  
Table 1 Transformer condition classification and simulation experiment content

电压/电流状态 模拟实验	故障状态模拟实验		
	无故障	铁心松动	线圈松动
无异常	状态一	状态二	状态三
开路励磁	状态四	状态五	状态六
短路电流	状态七	状态八	状态九

模拟实验分为电压/电流状态模拟实验和故障状态模拟实验。其中,电压/电流状态模拟实验包括无异常实验、开路励磁实验和短路电流实验,故障状态模拟实验包括无故障实验、铁心松动实验和线圈松动实验。将电压/电流状态模拟实验和故障状态模拟实验两两组合,可以得到9类变压器状态。具体的模拟实验内容如下。

无故障实验:变压器铁心、线圈等结构无故障。

铁心松动实验:拧松变压器铁心水平方向的紧固螺丝,以完全拧松螺丝需要的圈数为基准,模拟松动程度为50%的情况。

线圈松动实验:拧松变压器线圈绕组垂直方向的紧固螺丝,以完全拧松螺丝需要的圈数为基准,模拟松动程度为50%和100%的情况。

无异常实验:变压器电压和电流为额定值。

开路励磁实验:将变压器高压侧开路,低压侧以额定电压为基准按40%、70%、90%、100%、105%、110%、115%梯度施加电压。

短路电流实验:变压器低压侧短路,在高压侧以额定电流为基准按40%、70%、90%、100%、110%梯度施加电流。

图1展示了模拟实验的场景和对应的俯视图。模拟实验用3只高精度的枪指型麦克风从正面、斜侧面45°和侧面录制音频,麦克风高度与变压器中心高度平齐,距离1 m。设置麦克风高度与中心高度平齐是为了模拟巡检人员巡检时人脑高度与变压器高度一致的场景,不仅能提高收音的效果,而且便于后续检测变压器状态的传感器的安装定位。

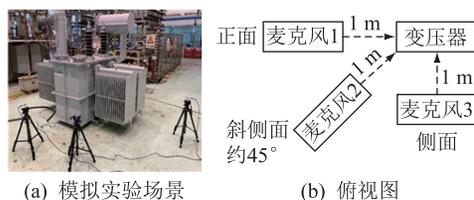


图 1 模拟实验场景和俯视图

Fig.1 Simulated experimental site and top view

## 1.2 变压器音频样本扩充

构建数据集的音频来自模拟实验得到的音频

和现场采集的音频。变压器现场环境和模拟实验环境存在一定的差异,因此,为了让模拟实验得到的音频更贴近现场音频,文中根据现场音频的特点和模拟实验音频的特点,对模拟实验得到的音频作表2所示的样本扩充处理。样本扩充处理使得模拟实验的音频与现场音频尽可能接近,同时增加了样本数量,能提高神经网络的泛化效果。

表 2 现场和模拟实验音频特点对比和后期处理方法

Table 2 Comparison of features and post-processing methods for field audio and simulated experiments audio

现场特点	模拟实验特点	后期样本扩充处理
角度不固定, 1支麦克风	角度固定, 3支麦克风	3支麦克风采集得到的音频按随机比例混合,模拟1支麦克风不同角度录制音频
距离不固定, 音量有大有小	距离固定为1 m	实验音频的幅值随机放大或缩小,模拟不同距离
存在各种各样的非稳态环境噪声, 如风雨声, 降噪困难	周围相对安静	调查现场的噪声水平, 根据-20 dB(模拟嘈杂环境)~20 dB(模拟安静环境)的信噪比将噪声与变压器音频混合, 噪声来自现场变压器周边和网络公开数据集

模拟实验音频经过后期处理后,与现场音频一起构成了文中的音频数据集。数据集一共有9 000条音频,1 000条来自现场音频,8 000条来自模拟实验。每条音频时长5 s,按照8:1:1的比例划分为训练集、验证集和测试集。训练集用于模型的训练,验证集用于模型结构和超参数效果的验证,测试集用于结果的测试和评估。

## 2 变压器声纹特征提取和压缩

变压器的音频信号包含大量数据点,提取音频的声纹特征可以降低数据量,便于模型识别。下面介绍变压器的声纹特征提取流程和对声纹特征进一步压缩的方法。

### 2.1 变压器声纹特征提取

时频谱图是常用的声纹特征,提取过程如图2所示。图2(a)为输入的音频波形,横轴为时间,纵轴为归一化处理后的幅值,由麦克风录得的幅值除以麦克风量程得到。为更清晰地展示图中细节,图中展示的是时长为1 s的音频的声纹特征提取流程(实际使用的音频为5 s)。图2(b)为从变压器音频中提取得到的时频谱图特征。首先,对音频进行预加重、分帧和加窗处理;然后,对每帧音频进行快速傅里叶变换(fast Fourier transform, FFT),并计算功率谱密度;最后,将每帧音频的功率谱拼接,得到时频谱图,为一个二维矩阵,如图2(b)所示。图2(b)

中横轴为时间,纵轴为音频的频率,不同的颜色代表不同的功率谱密度大小。

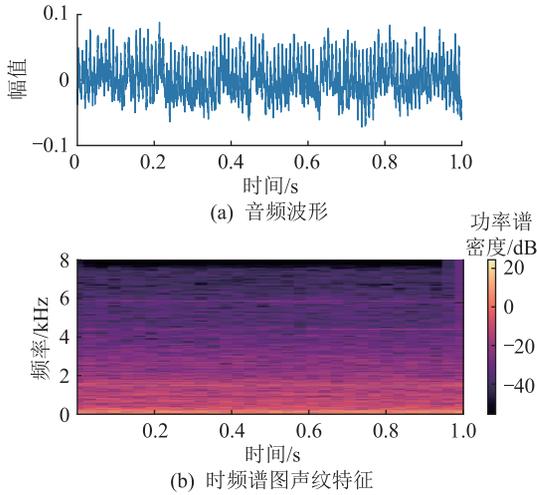


图2 变压器时频谱图声纹特征提取

Fig.2 Transformer time-frequency spectrogram acoustic feature extraction

## 2.2 变压器声纹特征压缩

2.1 节中提取到的变压器时频谱图声纹特征能较好地描述变压器的时频域特征,但数据量仍然很大。目前对变压器振动和音频特性的分析表明,变压器音频分量主要集中在0~4 kHz 区间内的50 Hz 倍频分量上<sup>[1-3,9]</sup>,而时频谱图的频率在0~8 kHz 区间内均匀分布,因此可以在频率维度上对时频谱图进行压缩。

文中使用的声纹压缩方法基于三角滤波器组,首先介绍三角滤波器组对声纹的压缩原理,然后介绍如何使用递归消除法和梯度下降法确定三角滤波器组的参数。

### 2.2.1 三角滤波器组压缩声纹

三角滤波器组是常见的声纹压缩方式,其数学表达如式(1)所示。其中, $H_m(k)$ 为第 $m$ 个滤波器在第 $k$ 个频率分量上的值; $f(m)$ 为第 $m$ 个滤波器的中心频率点。 $M$ 个滤波器的中心频率可以用一个向量 $\mathbf{f}=(f(1), f(2), \dots, f(M))$ 表示。

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k < f(m) \\ 1 & k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (1)$$

由式(1)可知,滤波器组对声纹的压缩效果由

滤波器数量 $M$ 和中心频率 $\mathbf{f}$ 决定。 $M$ 决定了压缩后声纹的维数, $\mathbf{f}$ 决定了滤波器组关注的重点频率,不同三角滤波器组的差异就在于 $M$ 和 $\mathbf{f}$ 不同。为确定合适的三角滤波器数量 $M$ 和中心频率 $\mathbf{f}$ ,文中提出“递归消除法粗调参数+梯度下降法微调参数”(后文简称递归消除+微调)的声纹压缩方法。

### 2.2.2 递归消除法粗调参数

递归消除法<sup>[12]</sup>使用迭代更新的方式来初步确定滤波器数量 $M$ 和中心频率 $\mathbf{f}$ ,其流程如图3所示。

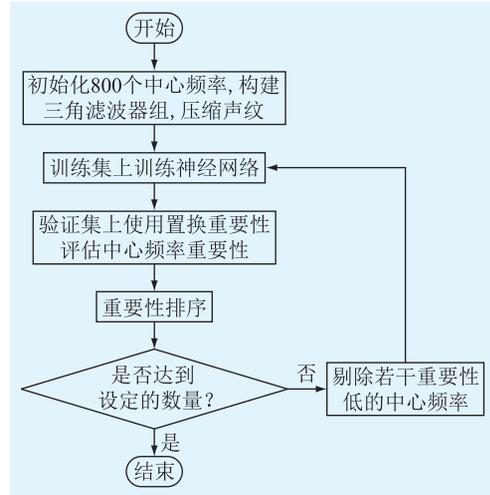


图3 递归消除法流程

Fig.3 Flow chart of recursive elimination method

首先,以10 Hz为间隔初始化800个三角滤波器,压缩声纹。然后,训练神经网络,并使用置换重要性<sup>[13]</sup>来评估不同中心频率对于神经网络识别声纹的重要性。该方法通过将特征值随机打乱后观察模型的性能变化来对特征的重要性进行评估。最后,剔除重要性较低的若干中心频率,重新训练神经网络。如此不断递归消除重要性低的中心频率,直到保留的中心频率数量达到设定值。考虑到文中的数据规模和模型规模都比较小,训练网络的速度较快,递归消除法的计算成本可以接受。

需要说明的是,训练模型、置换重要性评估和最终的模型评价分别在训练集、验证集和测试集上进行,3个数据集之间数据独立,因此递归消除法不存在信息泄露问题。

### 2.2.3 梯度下降法微调参数

使用递归消除法得到一组重要性较高的中心频率 $\mathbf{f}$ 后,文中采用了梯度下降法<sup>[14]</sup>对中心频率 $\mathbf{f}$ 的数值进行微调,与后续的声纹识别网络合并训练,以进一步提高声纹特征提取效果。梯度下降法的过程如下。

首先,将中心频率 $\mathbf{f}$ 视作一个可变量,则时频谱图 $\mathbf{S}$ 压缩得到 $\mathbf{S}'$ 的过程可以视作一个函数 $g_c$ ,如式

(2) 所示。

$$S' = g_c(f, S) \quad (2)$$

然后,将压缩后时频谱图  $S'$  输入神经网络  $g_n$ , 得到预测的状态分类  $y_p$ , 如式(3)所示,并根据损失函数  $g_l$  计算预测分类  $y_p$  和实际分类  $y$  之间的误差损失  $l$ 。

$$\begin{cases} y_p = g_n(S') = g_n(g_c(f, S)) \\ l = g_l(y_p, y) \end{cases} \quad (3)$$

最后,使用梯度下降法更新中心频率  $f$ , 如式(4)所示。

$$f' = f - \eta \frac{\partial l}{\partial f} = f - \eta \frac{\partial l}{\partial g_n} \times \frac{\partial g_n}{\partial g_c} \times \frac{\partial g_c}{\partial f} \quad (4)$$

式中: $f'$  为更新后的中心频率; $\eta$  为梯度下降的学习率。

### 3 变压器声纹识别神经网络

卷积神经网络是当今最成功的深度学习算法之一,在图像分类、目标检测等计算机视觉领域得到了广泛应用<sup>[15]</sup>,也是声纹识别领域常用的网络。首先介绍声纹识别使用的卷积神经网络的结构,然后介绍用于提高变压器难检出样本识别效果的代价敏感损失函数。

#### 3.1 卷积神经网络结构

文中使用残差网络<sup>[16]</sup>的残差块结构组成文中使用的卷积神经网络,如图4所示。图4中最左侧为输入的压缩后的声纹特征,形状为  $1 \times 80 \times 157$ , 3个数字分别代表声纹的通道数、长和宽。声纹特征首先通过卷积层、批归一化层<sup>[17]</sup>、线性整流(rectified linear unit, ReLU)函数<sup>[18]</sup>和最大池化层<sup>[19]</sup>进行初步的特征提取。然后,输入到6个残差块级联的网络,进行深层特征提取。随着网络深度的增加,声纹的通道数不断增加,长、宽不断降低,特征的层次从局部逐渐提升至整体。最后,进行全局平均池化<sup>[21]</sup>,得到一个  $256 \times 1 \times 1$  的向量,用2个不同的全连接层将向量与音频分类进行连接,分别从电压电流状态和故障状态2个维度对音频进行分类,得到变压器状态评估结果。

#### 3.2 卷积神经网络代价敏感损失函数

神经网络前向传播输出预测值,然后通过损失函数计算预测值和真实值之间的损失(误差),最后通过反向传播进行参数更新。

变压器音频通常存在分类难易程度不一致的特点,有些变压器状态的声音特征明显,容易检出;而有些状态没有明显的声音特征,检出比较困难。针对该特点,文中使用聚焦损失(Focal Loss)<sup>[20]</sup>作

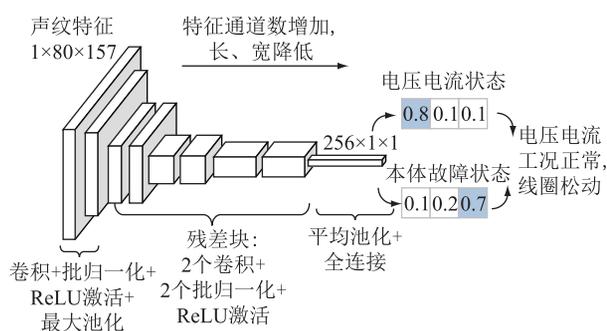


图4 卷积神经网络结构

Fig.4 Convolutional neural network structure

为损失函数,提高难检出样本在训练过程中的权重,使得模型对难检出样本的关注度提高。

Focal Loss 如式(5)所示, $y_i$  为模型对于真实分类的预测概率值, $y_i$  越高,损失值越小,说明模型效果越好。式中  $\gamma$  用于调节样本难度, $\gamma$  越大,对于分类错误的难检出样本的关注度就越高。根据文献[20]中的实验,将  $\gamma$  设置为 2 可以适当提高模型对难检出样本的关注度,但又防止模型过于关注难检出样本而降低整体的预测效果。此外,式(5)中还可以通过  $\alpha_i$  调节不同分类样本损失的权重,对于数量较少的分类,可以提高权重,使得模型对其更加关注,这也有助于缓解变压器正常和故障样本不平衡的问题。

$$g_l(y_i) = -\alpha_i (1 - y_i)^\gamma \log_2 y_i \quad (5)$$

### 4 变压器状态检测结果和分析

文中提出的变压器状态识别方法在训练集和验证集上进行训练和超参数验证,在测试集上测试有效性。下面对测试集上的结果进行展示和分析。

#### 4.1 声纹压缩对变压器状态检测的影响

在实际应用场景中,部署在现场的传感器设备通常为算力和存储空间较低的嵌入式设备,因此有必要压缩声纹的维度,从而降低模型的算力和存储要求。通过实验,得到一段 5 s 的音频在不同的声纹维度下的算力、显存占用情况和准确率情况,如图5和图6所示。图5和图6中从右到左声纹维度的下降过程对应使用递归消除法不断筛选特征的压缩过程。当特征数量较多时,筛选掉较多的特征对模型效果影响不大,因此采用步长为 100 筛选特征。而数量比较少的时候,特征数量变化对模型效果有明显的影 响,因此采用步长为 10 筛选特征。

图5展示了不同声纹维度对应的神经网络模型的计算量和显存,横轴为不同的声纹维度,左侧和右侧的纵轴分别对应计算量(单位为 MFLOPs,即百万浮点运算次数)和显存(单位为 MB,即百万字

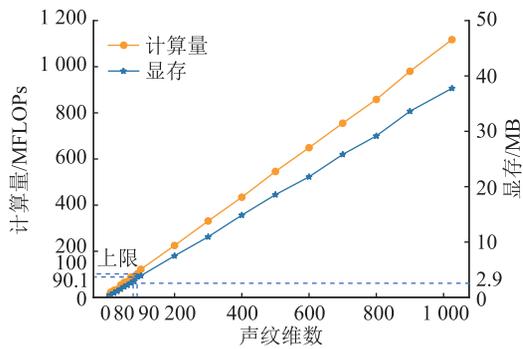


图5 声纹维数和计算量、显存曲线

Fig.5 Curves of voiceprint dimensions and computation and GPU memory

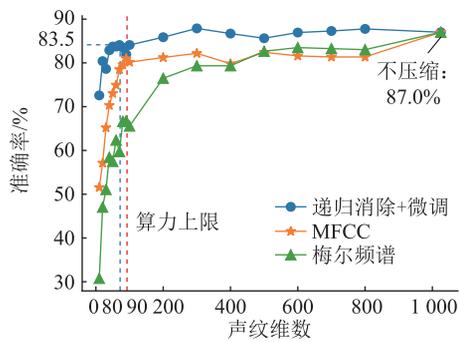


图6 声纹维数和准确率关系曲线

Fig.6 The relationship between voiceprint dimension and accuracy

节),分别用黄色和蓝色曲线表示。图5中,模型的计算量和显存与声纹维度呈现一种近似的线性关系,随着声纹维度的下降,计算量和显存也都线性下降。

考虑应用场景,当前应用最广的嵌入式硬件STM32系列和ESP32系列的每秒浮点计算次数峰值为100~200 MFLOPs<sup>[24-25]</sup>。对于一段时长为5s的音频,如果要实现实时监测,则处理时间需要小于等于5s。由于音频录制、声纹特征提取、通信等任务会占据较多的时间,并考虑一定的裕度,神经网络推理时间应在2s以下,对应的计算量为200~400 MFLOPs。此外,大多数嵌入式设备没有对神经网络的矩阵并行计算进行特别优化,因此很难发挥硬件设备的峰值计算能力,且由于嵌入式设备内存较小,来回加载数据到内存中也会占用一定时间,因此设定神经网络计算量上限为100 MFLOPs,对应的声纹维度约为90维,图5中蓝色的虚线标明了计算量上限。

图6的蓝色曲线展示了文中提出的“递归消除+微调”法在不同声纹维度下的准确率,在从右往左的压缩过程中,准确率先是缓慢下降,然后到达某一拐点时快速下降。准确率的部分波动可能是

因为测试的样本数量较少以及神经网络本身具有一定随机性。

图6还将文中方法与梅尔频谱、MFCC这2种常见的声纹压缩方法进行对比。可以看到,文中提出的递归消除+微调的蓝色曲线一直位于最上方,表明压缩相同的维度,文中方法能保留尽可能多的信息,得到最高的识别准确率。

图6中的红色虚线对应的90维为算力上限(对应图5中的100 MFLOPs计算量)。考虑到图中90维时准确率出现了下降,因此文中选择接近的80维作为声纹压缩维度。压缩到80维后,计算量和显存(90.1 MFLOPs和2.9 MB)分别为压缩前(1117.4 MFLOPs和37.8 MB)的8.1%和7.7%,便于实际的应用和普及推广。同时,83.5%的识别准确率,对实际使用的影响较小。

#### 4.2 代价敏感函数对变压器难检出样本的影响

文中选择具有代价敏感的Focal Loss损失函数用于网络的训练来提高对难检出样本的关注度。为验证Focal Loss的有效性,将其与交叉熵这一最常用的分类损失函数进行对比。

对于比较少见的变压器电压/电流异常状态或故障状态,通常希望能尽可能地检出这些状态,因此文中选择了召回率(检出率)<sup>[22]</sup>作为识别效果的评价指标。召回率是指检查出的目标状态的数量占该目标状态总数的比例,计算如式(6)所示。式中 $T_p$ 为判断正确的正类数量,即检出的数量; $F_N$ 为判断错误的负类数量, $F_N+T_p$ 即为该类的总量。

$$R = \frac{T_p}{T_p + F_N} \quad (6)$$

图7展示了分别使用交叉熵损失函数和Focal Loss作为损失函数训练神经网络时,在测试集上变压器不同状态的召回率。图7中,在大多数本身召回率较高的类别下,二者差别不大,但是对于召回率较低的状态七(短路+无故障)而言,使用Focal Loss作为损失函数进行训练后,召回率为63.6%,相较于使用交叉熵作为损失函数时的48.2%有了较大的提升,说明Focal Loss确实能改善对难检出样本的检出效果。

提高难检出样本的检出率对于实际应用而言有一定的意义。对于容易识别的状态,巡检人员平时也可以凭借经验发现;而对于难识别的状态,容易被巡检人员忽视或者误判,提高检出率可以帮助工作人员找到容易被忽视或误判的状态,避免问题扩大造成更大的损失。

#### 4.3 变压器状态识别结果混淆矩阵

为了更好地评估模型的表现,文中绘制了变压

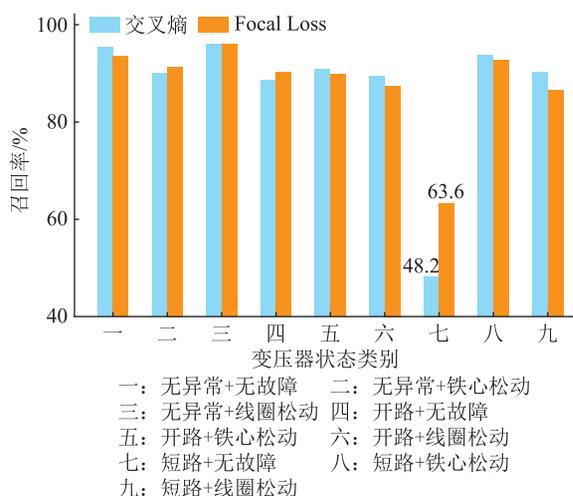


图7 不同变压器状态的召回率

Fig.7 Recall for different transformer states

器不同状态的混淆矩阵<sup>[23]</sup>,如图8所示,其中,数字表示预测的状态占实际状态的百分比,颜色的深浅对应数字的大小,数字越大,颜色越深。

实际状态	一	二	三	四	五	六	七	八	九
一	91.7%	5.0%	0	0	0	0	0	0	3.3%
二	9.5%	88.1%	2.4%	0	0	0	0	0	2.0%
三	2.0%	2.0%	95.9%	0	0	0	0	0	0
四	0	0	0	86.2%	0	4.6%	4.6%	1.8%	2.8%
五	0	0	0	2.8%	85.3%	5.5%	1.8%	4.6%	0
六	1.4%	1.4%	0.5%	5.1%	0.9%	81.4%	2.3%	0.9%	6.0%
七	0	1.5%	0	12.1%	0	7.6%	63.6%	0	15.2%
八	3.2%	0	0	0	1.6%	0	0	90.5%	4.8%
九	0.8%	0.8%	0.8%	3.9%	0	3.9%	9.3%	0.8%	79.8%
	一	二	三	四	五	六	七	八	九

预测状态

图8 变压器不同状态识别结果混淆矩阵

Fig.8 Confusion matrix of transformer identification results for different conditions

图8 从左上到右下的对角线上的数字即为召回率(检出率),对角线以外的数字代表不同状态被误判的比例,这有助于找到不同变压器状态之间的相似关系,方便巡检人员有针对性地排查问题。例如,对于召回率最低的状态七(短路+无故障)而言,其最容易被误判为状态九(短路+线圈松动)和状态四(开路+无故障),误判比例分别达到了15.2%和12.1%。这说明状态七与状态九、状态四之间的相似性较高,因此在巡检时要更加注意区分。

从电压/电流状态和故障状态2个维度分别对图8的变压器状态混淆矩阵做进一步细分,得到电压/电流状态和变压器故障状态的混淆矩阵如图9所示。

图9(a)为电压电流状态的混淆矩阵。可以看

实际状态	电压/电流状态识别结果			实际状态	故障状态识别结果		
	无异常	开路	短路		无故障	铁心松动	线圈松动
	无异常	开路	短路		无故障	铁心	线圈
无异常	98.7%	0	1.3%	无故障	86.8%	2.6%	10.6%
开路	1.6%	89.8%	8.5%	铁心松动	5.1%	90.2%	4.7%
短路	2.3%	9.3%	88.4%	线圈松动	9.7%	2.5%	87.8%

图9 变压器电压/电流和故障状态识别结果混淆矩阵

Fig.9 Confusion matrix for transformer voltage-current and faultcondition identification results

到,无异常的召回率有98.7%,这说明模型能准确分辨出是否存在电压电流异常状态。而对于具体的异常电压电流状态而言,有8.5%的开路被误判为短路,有9.3%的短路被误判为开路,说明开路和短路音频之间还是存在较高的相似度,存在误判的可能。

图9(b)为变压器故障状态的混淆矩阵。可以看到,相比电压电流的异常状态,变压器故障和无故障状态下的音频之间的差异更小,分别有5.1%的铁心松动和9.7%的线圈松动被误判为正常状态,说明变压器的故障状态更加难以被发现检出,因此巡检人员需要对变压器的故障状态投入更多的关注。

#### 4.4 环境噪声对变压器状态识别影响

在构建音频数据集时,为了模拟不同现场环境从安静到嘈杂的情况,往模拟实验得到的变压器音频中根据五档信噪比(-20 dB、-10 dB、0 dB、10 dB、20 dB)添加了不同强度的环境噪声。为探究不同强度的环境噪声对声纹识别的影响,分别计算每档信噪比下的识别准确率,得到结果如图10所示。

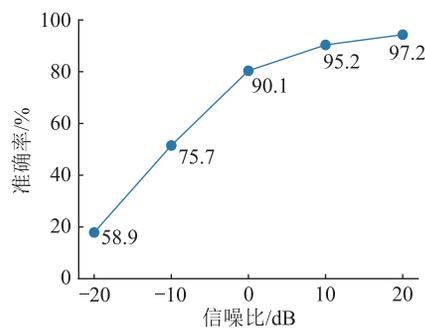


图10 不同信噪比下的识别准确率

Fig.10 Recognition accuracy at different signal noise ratios

图10中,准确率与信噪比成正相关,信噪比越高,声纹识别的准确率越高。当信噪比为-20 dB和-10 dB时,环境噪声的强度大于变压器音频的强度,对识别准确率有较大影响,准确率处于较低的水平,只有58.9%和75.7%;当信噪比大于等于0 dB时,变压器音频的强度大于等于环境噪声,此时识

别准确率能超过 90%,处于比较高的水平,特别是当信噪比达到 20 dB 时,准确率能够达到 97.2%。

信噪比和准确率的关系曲线说明,变压器的声纹识别更适用于环境噪声较小的场景,例如室内的场景,此时能取到比较好的效果;而对于室外等高噪声的场景,则需要使用指向性更好的录音设备或进行降噪处理,提高声纹识别的准确率。

## 5 结论

文中提出基于声纹压缩和代价敏感的变压器状态检测评估方法,通过模拟实验、现场采集和样本扩充得到变压器音频数据集,并在数据集上使用所提方法对变压器状态进行检测识别,分析检测结果得到结论如下:

(1) 文中所提的递归消除+微调的声纹压缩方法能有效压缩声纹的维度,且相比其他方法对声纹识别的影响更小。递归消除+微调能将声纹的维度从 1 025 维压缩至 80 维,计算量和显存分别降低到 1 025 维的 8.1% 和 7.7%,准确率达到 83.5%,有效地降低了对设备算力和存储的要求。

(2) 使用代价敏感的 Focal Loss 作为损失函数训练神经网络,能有效检出较难检出的样本,将最难检出的短路电流异常状态的召回率从 48.2% 提升至 63.6%,可以帮助巡检人员找到容易被忽视或误判的状态。

(3) 根据变压器状态识别结果混淆矩阵的分析,变压器电压电流工况的无异常样本和短路电流异常、开路电压异常这 2 种异常样本的差别较大,容易分辨。而变压器结构无故障和铁心松动故障、线圈松动故障这 2 种故障样本的差异较小,因此可以将巡检的重点更多地分配在比较难检出的故障的检查上。

(4) 声纹检测方法的准确率受环境噪声的影响,在信噪比为 -20 dB 的高噪声环境,识别准确率仅有 58.9%,信噪比为 20 dB 的低噪声环境,识别准确率可以达到 97.2%。因此,文中提出的声纹检测方法更适用于环境噪声较小的环境。

### 参考文献:

[1] 吴晓文,周年光,彭继文,等. 电力变压器噪声特性与相关因素分析[J]. 电力科学与技术学报,2018,33(3):81-85,146.  
WU Xiaowen,ZHOU Nianguang,PENG Jiwen,et al. Noise characteristic and relevant factors analysis of power transformers[J]. Journal of Electric Power Science and Technology,2018,33(3):81-85,146.

[2] 高沛,王丰华,苏磊,等. 直流偏磁下电力变压器的振动特性[J]. 电网技术,2014,38(6):1536-1541.

GAO Pei,WANG Fenghua,SU Lei,et al. Analysis on vibration characteristics of power transformer under DC bias[J]. Power System Technology,2014,38(6):1536-1541.

[3] 王宏升. 电力变压器铁芯及绕组振动特性研究与仿真分析[D]. 沈阳:沈阳工业大学,2022.  
WANG Hongsheng. Research and simulation analysis on vibration characteristics of iron cores and winding of power transformer[D]. Shenyang: Shenyang University of Technology, 2022.

[4] 曾桂南,吴恋,何燕琴,等. 基于声纹识别技术的常见模型与发展应用[J]. 现代计算机,2021(21):72-75,80.  
ZENG Guinan,WU Lian,HE Yanqin,et al. Common models and development applications based on voiceprint recognition technology[J]. Modern Computer,2021(21):72-75,80.

[5] ZHANG T,FENG G Q,LIANG J H,et al. Acoustic scene classification based on Mel spectrogram decomposition and model merging[J]. Applied Acoustics,2021,182:108258.

[6] MOHD HANIFA R,ISA K,MOHAMAD S. Speaker ethnic identification for continuous speech in Malay language using pitch and MFCC[J]. Indonesian Journal of Electrical Engineering and Computer Science,2020,19(1):207.

[7] PASEDDULA C,GANGASHETTY S V. Late fusion framework for acoustic scene classification using LPCC,SCMC, and log-Mel band energies with Deep Neural Networks[J]. Applied Acoustics,2021,172:107568.

[8] 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):1535-1542.  
WANG Fenghua,WANG Shaojing,CHEN Song,et al. Voiceprint recognition model of power transformers based on improved MFCC and VQ[J]. Proceedings of the CSEE,2017,37(5):1535-1542.

[9] 刘云鹏,王博闻,岳浩天,等. 基于 50 Hz 倍频倒谱系数与门控循环单元的变压器偏磁声纹识别[J]. 中国电机工程学报,2020,40(14):4681-4694,4746.  
LIU Yunpeng,WANG Bowen,YUE Haotian,et al. Identification of transformer bias voiceprint based on 50 Hz frequency multiplication cepstrum coefficients and gated recurrent unit[J]. Proceedings of the CSEE,2020,40(14):4681-4694,4746.

[10] 李楠,马宏忠,朱昊,等. 基于 50 Hz 倍频小波时频熵和 RUSBoost 的变压器绕组松动声纹识别[J]. 电机与控制应用,2022,49(5):87-93,102.  
LI Nan,MA Hongzhong,ZHU Hao,et al. Transformer winding looseness voiceprint recognition based on 50 Hz frequency multiplying wavelet time-frequency entropy and RUSBoost[J]. Electric Machines & Control Application,2022,49(5):87-93,102.

[11] 崔佳嘉,马宏忠. 基于改进 MFCC 和 3D-CNN 的变压器铁心松动故障声纹识别模型[J]. 电机与控制学报,2022,26(12):150-160.  
CUI Jiajia,MA Hongzhong. Voiceprint recognition model of transformer core looseness fault based on improved MFCC and

- 3D-CNN[J]. Electric Machines and Control, 2022, 26(12): 150-160.
- [12] JEON H, OH S. Hybrid-recursive feature elimination for efficient feature selection[J]. Applied Sciences, 2020, 10(9): 3211.
- [13] FISHER A, RUDIN C, DOMINICI F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously[J]. Journal of Machine Learning Research; JMLR, 2019, 20: 177.
- [14] AMARI S. Backpropagation and stochastic gradient descent method[J]. Neurocomputing, 1993, 5(4-5): 185-196.
- [15] ALZUBAIDI L, ZHANG J L, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. Journal of Big Data, 2021, 8(1): 53.
- [16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. New York. PMIR, 2015: 448-456.
- [18] ECKLE K, SCHMIDT-HIEBER J. A comparison of deep networks with ReLU activation function and linear spline-type methods[J]. Neural Networks, 2019, 110: 232-242.
- [19] AKHTAR N, RAGAVENDRAN U. Interpretation of intelligence in CNN-pooling processes: a methodological survey[J]. Neural Computing and Applications, 2020, 32(3): 879-898.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2999-3007.
- [21] YU D J, WANG H L, CHEN P Q, et al. Mixed pooling for convolutional neural networks[C]//MIAO D, PEDRYCZ W, SLEZAK D, et al. International Conference on Rough Sets and Knowledge Technology. Cham; Springer, 2014: 364-375.
- [22] KAYA, BILGE. Deep metric learning: a survey[J]. Symmetry, 2019, 11(9): 1066.
- [23] SALIH A A, ABDULAZEEZ A M. Evaluation of classification algorithms for intrusion detection system: a review[J]. Journal of Soft Computing and Data Mining, 2021, 2(1): 31-40.
- [24] ST. STM32 主流 MCU [EB/OL]. [2023-11-15]. <https://www.st.com/zh/microcontrollers-microprocessors/stm32-mainstream-mcus.html>.
- ST. STM32 mainstream MCU [EB/OL]. [2023-11-15]. <https://www.st.com/zh/microcontrollers-microprocessors/stm32-mainstream-mcus.html>.
- [25] Espressif. Esp32 echnical reference manual[EB/OL]. [2023-11-15]. [https://www.espressif.com/sites/default/files/documentation/esp32\\_technical\\_reference\\_manual\\_en.pdf](https://www.espressif.com/sites/default/files/documentation/esp32_technical_reference_manual_en.pdf).

作者简介:



胡赵宇

胡赵宇(1999),男,硕士在读,研究方向为高电压与绝缘技术(E-mail: 1755475455@qq.com);

李喆(1978),男,博士,教授,博士生导师,研究方向为高电压与绝缘技术、泛在电力物联网、人工智能等;

陈海威(1971),男,博士,工程师,从事高电压与绝缘技术相关工作。

## Transformer state detection and assessment method based on voiceprint compression and cost-sensitive techniques

HU Zhaoyu<sup>1</sup>, LI Zhe<sup>1</sup>, CHEN Haiwei<sup>2</sup>, LU Xin<sup>2</sup>

(1. Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

2. China Energy Engineering Group Guangxi Electric Power Design Institute Co., Ltd., Nanning 530007, China)

**Abstract:** Voiceprint detection technology can assist inspectors in assessing the state of transformers. A method for detecting and assessing transformer states based on voiceprint compression and cost-sensitive techniques is proposed. The method first extracts voiceprint features from transformer audio, then filters and compresses these features in the frequency domain. Subsequently, a convolutional neural network is employed to evaluate the transformer's state, incorporating a cost-sensitive loss function to enhance attention towards difficult samples. Using a 35 kV transformer as the experimental subject, transformer audio data is collected through on-site recordings, simulated experiments and sample augmentation. Test results demonstrate that the proposed method reduces the voiceprint dimensionality from 1 025 to 80, decreasing computational complexity and video memory usage to 8.1% and 7.7% of the original 1 025 dimensions, respectively. Simultaneously, the proposed method achieves a voiceprint recognition accuracy of 83.5% and improves the recall rate of the most challenging short-circuit current anomaly from 48.2% to 63.6%.

**Keywords:** transformer detection; acoustic pattern recognition; acoustic pattern compression; cost sensitivity; convolutional neural networks; pattern recognition

(编辑 钱悦)