

基于优化 Focal-XGBoost 的变压器状态声振识别模型*

许洪华, 尹来宾, 李 勇

(国网江苏省电力有限公司南京供电分公司, 江苏 南京 210000)

摘要: 受数据样本难以区分和数据平衡性不佳影响, 采用声振信号的变压器状态识别模型往往准确率低下。针对这一问题, 引入了 Focal 损失, 根据样本训练过程的准确度动态反馈权重, 从而构成了 Focal-XGBoost 优化模型。先通过一组贴合变压器频谱的滤波器充分提取声振信号有效信息, 再作 XGBoost-PCA 筛选降低样本维度。然后采用 Focal 损失优化原模型中的 Softmax 目标函数形成 Focal-XGBoost 模型, 并在输入上述样本后根据准确率波动作 Focal 的超参数优化, 进而输出变压器状态识别结果。10 kV 和 110 kV 变压器的试验结果表明, 相较传统 SVM、KNN 等学习模型, Focal-XGBoost 减少了 XGBoost 测试样本中难分样本的误分量 44.7%, 从而使模型识别准确率更高; 此外, 非均匀提取在平均精度损失低于 0.5% 的基础上压缩 50% 样本空间, 进一步降低了模型训练成本。

关键词: 非均匀滤波; Focal 损失; Softmax 分类; 变压器状态识别; XGBoost 算法

中图分类号: TM41 文献标志码: A 文章编号: 1673-6540(2023)08-0038-08

doi: 10.12177/emca.2023.087

Vibro-Acoustic Identification Model for Transformer State Based on Optimized Focal-XGBoost*

XU Honghua, YIN Laibin, LI Yong

(State Grid Jiangsu Electric Power Co., Ltd., Nanjing Power Supply Branch, Nanjing 210000 China)

Abstract: Influenced by indistinguishable data samples and poor data balance, transformer state identification models using vibro-acoustic signals often have low accuracy. To address this problem, Focal loss is introduced to dynamic feedback weights according to the accuracy of the sample training process, thus constituting a Focal-XGBoost optimization model. Firstly, a set of filters that fit the transformer spectrum are used to fully extract the effective information of the vibro-acoustic signal, and then XGBoost-PCA is used to reduce the dimensionality of the samples. Then, the Softmax objective function in the original model is optimized using Focal loss to form the Focal-XGBoost model. After inputting the above samples, the hyperparameters of Focal are optimized based on the accuracy wave action, and the transformer state recognition results are output. The experimental results of 10 kV and 110 kV transformers show that Focal-XGBoost can reduce the number of samples compared with traditional SVM and KNN models. Focal-XGBoost reduces the misspecification of difficult samples in XGBoost test samples by 44.7%, which results in higher model recognition accuracy. In addition, non-uniform extraction compresses the sample space by 50% on the basis of average accuracy loss below 0.5%, which further reduces the model training cost.

Key words: non-uniform filtering; Focal loss; Softmax classification; transformer state identification; XGBoost algorithm

收稿日期: 2023-03-16; 收到修改稿日期: 2023-04-21

* 基金项目: 江苏省电力有限公司重点科技项目(J2021053)

作者简介: 许洪华(1981—), 男, 博士, 高级工程师, 研究方向为电力设备故障诊断技术。

0 引言

电力系统的运行稳定性与电力变压器的健康状态息息相关。电力变压器在包括发电厂、变电站、配电所等电网重要场所承担着电压变换、电能传输与调节、用户侧电能质量维持等任务。因此,及时发现并排除变压器异常状态,确保其健康稳定工作具有重要意义^[1-2]。

对于变压器,常采用在线监测实现状态辨识。现有的在线监测方式主要包括油气比、振动监测、局放监测和光纤传感法。相较传统非在线监测,在线监测无需变压器停机或进行非破坏性试验,因此适用于不方便随时停运的重要电力变压器,并具有数据实时更新的优势^[3]。然而,在线监测装置的后期安装对已投运变压器是一大难题,基于声纹特征的在线监测能够克服设备装设困难且无需与带电部分直接接触,此外声纹信号蕴含的信息量丰富多样,故应用前景良好^[4]。

目前有关变压器声振识别的研究主要分为局放超声信号和工况信号两类。文献[5]研究了变压器绕组对局放信号的定位影响,提高了基于超高频(UHF)的定位精度。文献[6]采用了一种光纤式传感器对超声信号进行测量,具有较高的灵敏度且增加了局放信号可测范围。文献[7]提出了一种时差筛选和ABC二次优化的变压器超声定位方法,结合蜂群算法和K-means进行局放定位校正。然而变压器超声信号的研究主要针对局放故障及其定位,无法对变压器运行状态进行识别也难以发现潜在隐患进行预防,故存在较大局限性。

在工况信号方面,文献[8]采用改进的梅尔频谱倒谱系数(MFCC)结合矢量化模型对变压器不同松紧铁心信号进行了测试。文献[9]在MFCC的基础上采用具有抗噪能力和鲁棒性的新混合参数(MGCC)对干式变压器进行诊断。文献[10]研究了有关振动和声压测量的变压器声功率谱计算。但现有研究主要关注声振信号特征的提取和优化对状态识别的影响,很少针对具体算法在处理声振样本时存在的缺陷进行分析,因此部分样本误分率高,进而易导致整体识别准确率低。

基于声振的变压器状态识别准确率不高,其

根本原因在于状态特征提取的准确度不够,而机器学习能充分挖掘状态和特征间的潜在联系,因此,近年来包括支持向量机(SVM)、K临近(KNN)、深度神经网络、模糊逻辑、随机森林等模型被应用于状态识别。实际工程中变压器工况的声振数据存在异常、样本难获取、正常样本量多的严重不平衡问题,且部分状态间界限模糊。有关学者针对这一情况从样本采集方式和处理方式上进行了研究,提出了如人工少数类过采样法(SMOTE)过采样^[11]、混合采样^[12]、聚类采样^[13]、样本预筛选^[14]等不平衡缺陷补偿方案。但大部分研究把重心放在少量样本生成、扩充或过量样本筛选方面,若相对不平衡程度较高则容易造成理想样本库过大,有效信息量缺失等问题,进而导致模型训练效率降低,且很少关注样本自身区分难易性。

鉴于此,本文基于非均匀声振特征提取及优化降维,提出一种采用Focal损失的XGBoost优化模型。该种新型交叉熵损失函数(CE)能有效针对变压器正常和异常运行样本数量差较大的不平衡问题进行损失重新分配,且无需改变原样本大小。此外,对于变压器声振样本存在的区分差异,Focal损失能够让训练过程聚焦更难分样本,从而提高模型整体运行准确率。最后,本文对构建的Focal-XGBoost模型进行10 kV和110 kV变压器状态的识别测试与分析,以验证该模型的有效性和可靠性。

1 特征提取和降维模型

目前,声学信号的识别依赖语音信号模型,即通过MFCC提取特征,但反应电气设备运行能力较差^[15]。因此,本文从MFCC出发,针对变压器声纹特点提出一套更贴合其实际运行的特征提取方法。

1.1 非均匀声振特征提取

工况下变压器声纹高频有效成分占比不高,预加重容易导致频谱失真或无用特征凸显,进而引起特征簇发散,故不考虑对信号作预加重处理;常规MFCC基于人声提出,其特征频率 f_{Mel} 也围绕可听声非线性展开,具体如下式:

$$f_{\text{Mel}} = 2595 \lg \left(1 + \frac{f}{700} \right) \quad (1)$$

式中: f 为原始频率。

而变压器声纹特征在常规频率尺度上呈间断非线性分布,由 MFCC 基本原理可知,滤波器设置密度决定了运算后该频段区域的特征精度,语音信号在变压器频谱许多信息集中位置滤波器设置稀疏,为使中心频率尽可能还原信号特征,必须对滤波器各参数进行重设。

变压器运行过程中,交变电流于铁心激发出 100 Hz 基频的倍频量,并产生以 50 Hz(由变压器剩磁和非对称产生)及 100 Hz 倍频为主的声振量,为充分捕捉特征信息,本文在 400 Hz 以下每 50 Hz、400 Hz 以上每 100 Hz 设一组滤波器,形成如式(2)的局部均匀总体非均匀模式:

$$W_i = \begin{cases} 0 & f \leq f_{i-1}, f \geq f_{i+1} \\ \mu_i \frac{f - f_{i-1}}{f_i - f_{i-1}} & f_{i-1} \leq f \leq f_i \\ \mu_i \frac{f_i - f}{f_i - f_{i-1}} & f_i \leq f \leq f_{i+1} \end{cases} \quad (2)$$

式中: W_i 为滤波参数; μ_i 为三角中心滤波器权参; f_n 为第 n 个滤波器的中心频率。

非均匀三角滤波器组如图 1 所示。

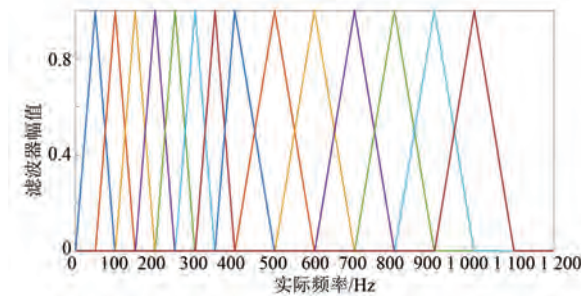


图 1 非均匀三角滤波器组

采用高斯滤波器代替部分三角滤波器进行优化,带通的高斯滤波能有效滤除变压器声学信号中的白噪成分,并较三角滤波减少有效信息损失,如下式所示:

$$W_i = \frac{\mu_i}{\sqrt{2\pi}\sigma} e^{-\frac{(f-f_i)^2}{2\sigma^2}} \quad (3)$$

指数形式滤波具有快速衰减特性,且能根据需要调节 σ 参数控制特征提取密度,例如 50、100、200 Hz 附近频谱特征密集,放大 σ 参数可以扩大特征提取时该频段成分有效占比。本文认为

0.8 以上的部分为有效带宽,当干扰强度较大而特征量少时,选择窄带高斯滤波器;而 400 Hz 以内尤其 100 Hz 附近干扰较低且邻近特征信息密集时,选择宽带高斯滤波器。

1.2 基于 XGBoost-PCA 的优化降维

在分析变压器的声振时,为提高计算效率,除离散余弦(DCT)变换外,还需要对数据进行横向压缩^[16]。

PCA 降维的基本思想是将原特征的分量进行线性组合生成新特征向量,且相关度应尽可能低、方差尽可能大,本文对压缩后特征对应特征值的选取要求其总贡献不低于全体的 80%,且单个贡献不低于全体的 5%。但常规 PCA 只关注特征本身的差异程度,并不关心同类样本特征差异是否过大,或不同样本间特征相似度是否过高,因此,直接进行降维可能造成大量信息损失,进而导致识别准确率严重降低。

基于 XGBoost-PCA 的优化降维先对不同类样本的特征值进行简单聚类,使协方差矩阵降维时有效特征更加突出。本文从去中心化后的样本矩阵出发设定分类阈值,不足阈值部分保持不变或降权、剔除,超出阈值部分则放大该部分作用,起到特征人为预筛选、修正的作用。具体来说,取一组样本类型中的某特征量上下限,记 $H_{ij} = [h_1, h_2, \dots, h_m]$ 为第 i 个样本全体的第 j 个特征值集合,如式(4):

$$\mu_1 = \max H_{ij} - \min H_{ij} \quad (4)$$

式中: μ_1 为该特征量在判别中的稳定性。

取同样本不同特征量间的重叠限,记 A 和 B 分别为两个样本值域内的重心附近样本数,如式(5):

$$\mu_2 = \frac{A \cap B}{\max(A, B)} * 100\% \quad (5)$$

式中: μ_2 为重叠阈值,重叠阈值越高可大概理解为两种特征越相似。

取第 k 个特征量在全样本中的上下限,如式(6):

$$\mu_3 = \frac{\max H_{ik} - \min H_{ik}}{\mu_1} * 100\% \quad (6)$$

式中: μ_3 为特征有无作为辨别量的潜力。

表 1 给出本文预筛选分类的评估标准,并以图 5 中特征标量大小为例。

表 1 样本预筛选评估标准

辨别值	取值大小	状态描述	进行的操作
	<5	完全稳定	保持不变
μ_1	>5, <20	存在不稳定因素	降低权重
	>20	不稳定	根据重要性合并或删除
μ_2	<80%	分辨率高	提高权重
	>80%, <95%	分辨率模糊	保持不变
	>95%	无可靠分辨率	根据重要性合并或删除
μ_3	>120%	辨别潜力大	保持不变
	接近 100%	无法辨别潜力	根据重要性保持 不变或降低权重

重要性根据 XGBoost 进行划分,如图 2 所示为声振 12 维 DCT 特征在 XGBoost 中训练 100 次后的贡献度。本文认为当某一特征贡献小于 800 时重要性一般,当贡献低于 500 时重要性很低。

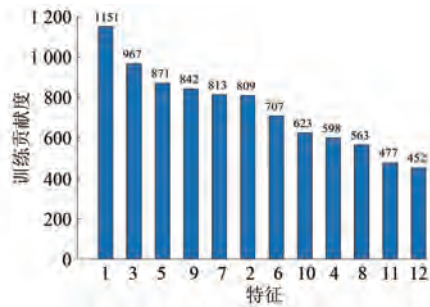


图 2 基于 XGBoost 的特征贡献评分

2 Focal-XGBoost 优化模型

本文所研究的 Focal-XGBoost 优化模型结合了变压器实际工况,包括运行过程中正常样本和异常样本存在数量大小差异,以及变压器非突变状态间存在渐变性导致的声振特征难分和易分差异。而 Focal Loss 能动态针对模糊样本加强训练,同时根据样本尺度修正损失权重,从而全面提升原训练模型的效率和准确率。

2.1 XGBoost 训练模型

XGBoost 算法是一种集成树的提升训练模型,当一种模型在训练结果上表现效果不好时,算法将从原树模型中提取训练不良部分,并针对该部分数据使用第二个模型进行训练,以此类推直到训练全局都表现出较好的结果^[17]。

应用 XGBoost 模型首先考虑目标函数,如下所示:

$$\hat{y}_i^{(l)} = \sum_{k=1}^l f_k(x_i) = \hat{y}_i^{(l-1)} + f_i(x_i) \quad (7)$$

式中: $\hat{y}_i^{(l)}$ 为迭代格式,表征了每个树对 i 状态的预测值总和。

由于 XGBoost 的特点是正则化提升,用于控制复杂度,故须在目标函数中引入复杂度参数,如下所示:

$$O_{bj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{j=1}^l \Omega(f_j) \quad (8)$$

式中: l 为损失函数; Ω 为算法复杂度。

进一步考虑损失函数二阶泰勒展开形式:

$$l(y_i, \hat{y}_i^{(l)}) = l[y_i, \hat{y}_i^{(l-1)} + f_i(x_i)] = l(y_i, \hat{y}_i^{(l-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \quad (9)$$

最后对模型参数及复杂度作两方面的优化,子节点 j 所包含的样本的一阶和二阶导数累加之和分别为 G, H ,如下所示:

$$\begin{cases} G_j = \sum_{i \in I_j} g_i \\ H_j = \sum_{i \in I_j} h_i \end{cases} \quad (10)$$

进一步对目标函数求导,得到范数 w_j 的极值作为树的模型参数参考, λ 为复杂度系数,如下所示:

$$\begin{cases} O_{bj}^{(l)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \\ w_j = - \frac{G_j}{H_j + \lambda} \end{cases} \quad (11)$$

但采用 Softmax 函数作为 XGBoost 多分类的 CE 损失目标时,同样对样本本身监督不足,仅考虑预测值与真实值差异,并不考虑样本本身存在的不平衡问题,因此训练重心偏向大样本识别,而对少量难分样本识别能力较差^[18]。

2.2 基于 Focal 的 XGBoost 训练模型

XGBoost 模型具有高度可拓性,除训练参数外允许使用不同损失函数进行算法优化,因此能根据变压器实际工程需求形成基于 Focal 的 XGBoost 模型,针对训练过程中难分样本和样本量差进行平衡调整和训练侧重调整^[19]。

Focal 作为损失函数的一种,随训练过程动态调节特征的输出准确度 p_i ,与常规 CE 交叉熵损失不同,Focal 引入了超参数 γ 和 α_i 平衡样本数

量和难易区分重心问题,二类的 Focal 损失如下:

$$L_{\text{Focal}} = -\alpha_i(1 - p_i)^\gamma y \log(p_i) - (1 - \alpha_i)p_i^\gamma(1 - y) \log(1 - p_i) \quad (12)$$

式中: γ 参数常规大于 1,能使易分样本较难分样本的损失衰减增加,变相提高训练中难分样本的损失权重; α_i 参数则用于平衡数据量差异,其值越大权重越高,并随 γ 增加会适当降低一些,这是由于普遍难分样本数据较少而易分样本数据多; γ 已对易分样本损失进行过限制; y 为理论输出值,在二分类中仅以 0、1 表示。

将 Focal 损失作为优化加入 XGBoost 多分类模型时,需根据式(9)考虑其一阶和二阶导数作为训练目标,以 Softmax 为激活函数的多分类 Focal 损失表达,如下所示:

$$L_{\text{Focal}} = -\sum_{i=0}^n \alpha_i(1 - p_i)^\gamma y_i \log(p_i) \quad (13)$$

式中: α_i 和 p_i 分别为第 i 个状态平衡样本大小与输出准确度的参数; y_i 由 Softmax 函数的 one-hot 编码决定,仅在对第 i 个位置上升为 1,其余为 0,故在计算第 i 类样本损失时简化为下式(14),并将 Softmax 对应的目标代入 p_i 即可。

$$L_{\text{Focal},i} = -\alpha_i(1 - p_i)^\gamma \log(p_i) \quad (14)$$

这样针对变压器声振样本中的难分对象,如不同层次松动,或不同负载情况等,通过 Focal 损失函数能够有效提高该类样本产生的损失值贡献;而对于不同类样本数量差较大问题,通过 α_i 与 γ 共同调整其损失权重,能够避免易区分大样本在损失计算时占据主导地位^[20]。

3 试验测试

为验证本文提出的 Focal-XGBoost 模型可行性,本文以一台 10 kV 油浸式变压器为试验对象,在利用龙门吊将机箱、铁心绕组分离后,分别对绕组、铁心的正常及不同松动程度下的声振量进行识别测试。其中,铁心和绕组松动程度通过改变安装螺栓的压紧程度进行,分别在原始安装、轻微松动、严重松动及完全松动情况下进行上电试验。

试验在变压器 4 面分别布置麦克风,采样频率取 50 kHz,每次采样时间在 15~20 s。以 2 号声传感器信号为例,图 3 为实地安置测量图。

3.1 特征提取及压缩结果

图 4 给出了不同运行状态的频谱分布,其中

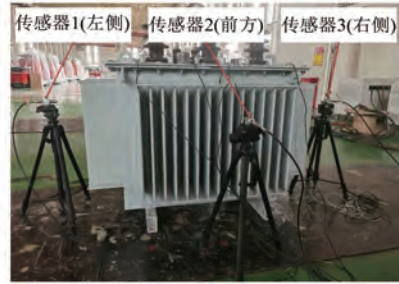


图 3 传感器实测安装图

声压值为避免由于仪器校准或距离产生偏差,以同一标准下的归一值代替。

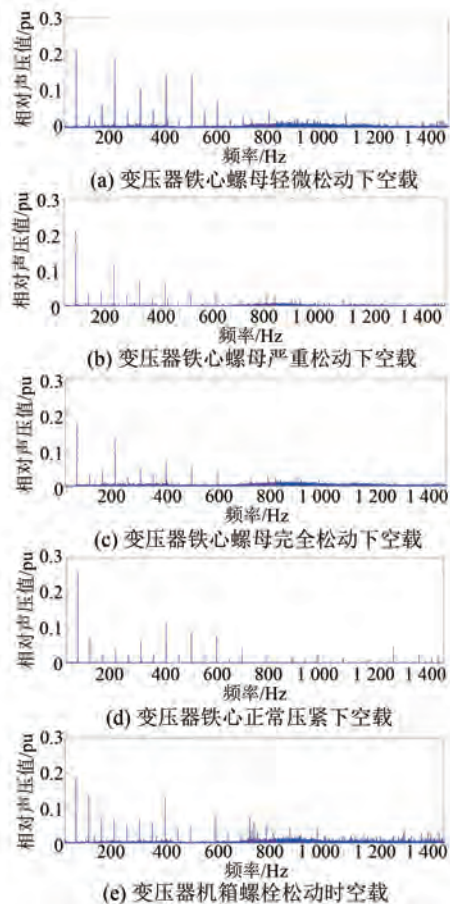


图 4 变压器不同状态频谱图

由图 4 可知,变压器在铁心松动时干扰更大且 100 Hz 成分减少,100 Hz 倍频尤其以 200 Hz 为主的谐波分量上升;在机箱螺栓松动时 100 Hz 及其倍频分量,包括 50 Hz 奇频分量都有所增加;但详细考虑不同松动类型、松动程度时,对应的声纹特征较难直接区分,基于 Focal-XGBoost 训练该样本数据的结果如下。经本文上述非均匀特征提

取流程,对试验变压器不同运行状态进行特征提取结果如图 5 所示。

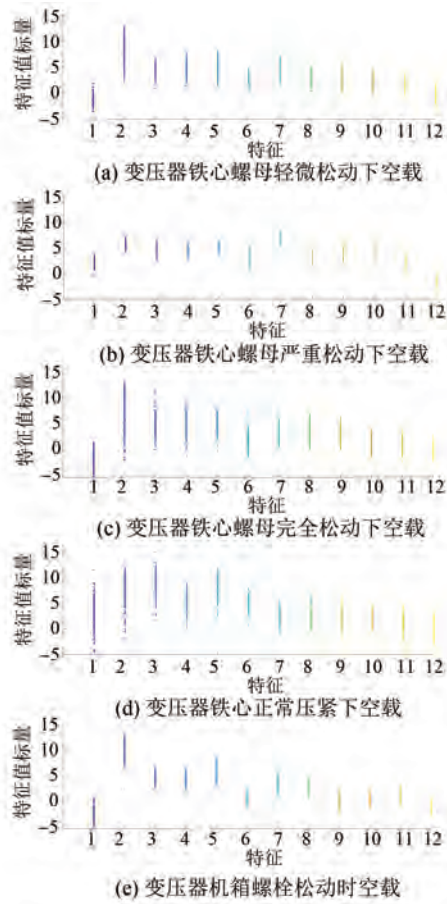


图 5 变压器不同状态特征图

采用 XGBoost-PCA 的部分类型特征降维结果将信息量主要集中在前 4 维^[21],并最终得到表 2 的样本库,其中 α_i 初步考虑为占比的反比, γ 常规取 2,单类别样本最多占 36.77% 最少占 12.24%,具有样本大小的不平衡性,且铁心不同程度松动区分较难。

表 2 五类状态训练样本库

样本类型	样本数量	占比%	反比	标准 α_i
正常	2 981	36.77	2.72	0.095
轻微松动	1 433	17.68	5.66	0.197
严重松动	1 482	18.28	5.48	0.191
完全松动	992	12.24	8.17	0.285
螺栓松动	1 218	15.03	6.65	0.232

3.2 基于 Focal-XGBoost 的训练结果

表 3 列出了基于 Focal-XGBoost 与其他训练

模型的识别结果,并比较了非均匀特征提取与 XGBoost-PCA 降维下的准确率。

表 3 不同模型识别准确率 %

使用模型	五类松动状态识别准确率		
	常规提取	非均匀提取	优化降维
树模型	82.0	93.9	94.9
KNN 模型	94.2	96.1	95.7
XGBoost	92.5	94.9	94.8
Focal-XGBoost	94.8	97.8	97.6
SVM 模型	93.4	96.6	96.4
线性模型	89.2	95.8	95.5

由表 3 可知 Focal-XGBoost 训练模型相较常规 XGBoost 准确率提高了近 2.5%,且在不同条件下均优于 SVM 和 KNN 模型;非均匀特征提取提高了 XGboost 识别准确率 2% 以上,尤其适用于线性模型和常规树模型;基于 XGBoost-PCA 的优化降维在压缩一半特征量的前提下,将平均准确率波动控制在 0.5% 以内,提高了模型运行准确率。

具体地,以混淆矩阵图 6 的形式描述 XGBoost 和 Focal-XGBoost 对五类状态识别的误分情况,测试集占总数的 15%,显著减轻了轻微和完全松动的样本误分量。

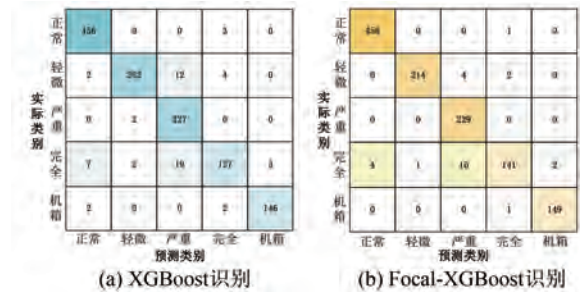


图 6 松动状态识别混淆矩阵

此外,Focal-XGBoost 对类别少的特征训练效果更优,当类别总数为 2 时识别准确率达 98.6%,即使针对较难识别的松动程度其也达到 96.2%。且压缩后总运行内存更小,总样本数在 4029 时平均运行时间为 3.2 s 左右,在 8106 时平均运行时间为 4.0 s 左右,都低于 KNN 和 SVM 的训练时间。

4 现场实测分析

110 kV 以上在运大型变压器除大型年检或

重大事故外,无法进行拆装也不允许进行故障检测设置,为验证本文提出的变压器声振识别模型在现场环境的有效性,选择以 110 kV 大型变压器实际运行过程中,不同时段对应不同负载状态下产生的声振实测信号作为测试样本,并以峰时负载、谷时负载和常规负载作为样本类型划分进行训练。

由于不同负载程度的声振样本同样具有辨识难易区分性及尺寸差异性,且也属变压器运行状态的一种,能有效对本文 Focal-XGBoost 优化模型的状态识别能力进行验证。图 7 为现场实测图选择在风机背面进行声振信号采集。



图 7 现场实测安装图

在相同数据样本中,本文选择采用神经网络模型、KNN 和 SVM 训练模型与 Focal-XGBoost 进行对比。

4.1 Focal-XGBoost 训练结果分析

表 4 为基于 Focal-XGBoost 和其他模型对三种负载情况的多次训练结果,共取峰时负载样本 1 982 个,谷时负载 3 044 个,常规状态负载 5 817 个,对应测试集的分部及整体识别准确率如下。

表 4 负载状态识别结果 %

使用模型	三类负载状态识别准确率			综合
	谷时	峰时	常规	
XGBoost	93.4	97.6	99.8	97.9
Focal-XGBoost	95.5	99.0	99.8	98.9
KNN 模型	92.7	98.8	99.9	98.2
SVM 模型	90.7	99.8	99.7	98.0
神经网络	95.8	99.8	96.8	97.5

其中神经网络的准确率随样本划分和结构变化在 96.7% 到 98.3% 之间波动。为直观展现误分情况,基于 XGBoost 和 Focal-XGBoost 的负载状态

混淆矩阵如图 8 所示,测试样本为总数的 20%。

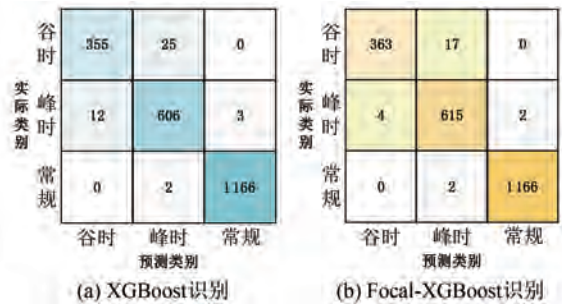


图 8 负载识别混淆矩阵

结果表明 Focal-XGBoost 降低了实测谷时和峰时样本误分至常规样本的概率,整体减少了变压器状态误判样本量 44.7%,并在难分样本的识别中明显优于其他训练模型,且稳定性更优秀。

4.2 Focal 损失对比结果

对实测样本进行 Focal 超参数优化,不同 γ 参数下 Focal-XGBoost 对该不平衡负载状态测试样本的误分波动如图 9 所示。

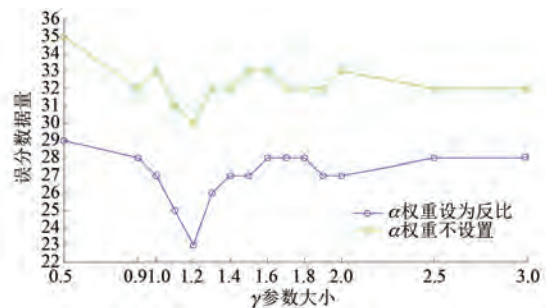


图 9 γ 与负载样本误分关系

故本文在 α_i 取样本量反比的情况下,在实测中选取 γ 为 1.2 以获得最佳训练结果;而对于更一般的变压器声振数据不平衡情况,可根据不平衡程度初步取 γ 在 1~1.6 之间进行 Focal 损失参数优化,但不宜超过 2.5。

5 结 语

本文针对变压器声振数据存在的样本数量及区分难度不平衡问题,提出一种采用 Focal 损失优化的 XGBoost 识别模型,据仿真分析和现场测试中不同样本的训练结果,有下述结论:

(1) 采用贴合变压器实际频谱的非均匀特征提取压缩方式能有效减少环境白噪影响和信号信息泄露,使关键特征值围绕频谱突出值展开,从而

提高各类模型状态识别准确率;

(2) 将 Focal 损失与 XGBoost 多分类中 Softmax 目标函数相结合构成的 Focal-XGBoost 模型,能有效针对声振数据样本的不平衡问题进行动态优化,显著减少了原 XGBoost 测试中,少量、难分异常样本的误分量 44.7%,进而提高了难分样本识别准确率 2.5% 左右;

(3) 据仿真分析和实测样本训练,本文采用 Focal-XGBoost 的识别模型较 SVM、KNN 等常规训练算法运行时间更短,更适合在线监测应用,且难分样本识别准确率提高 3% 以上,进而使整体识别准确率平均提升 1%;较神经网络模型在结构上更加稳定,不受网络结构影响。因而 Focal-XGBoost 模型能够更加精确地对变压器运行状态进行识别,具有一定实际工程价值。

【参考文献】

- [1] 张又文,冯斌,陈页,等. 基于遗传算法优化 XGBoost 的油浸式变压器故障诊断方法[J]. 电力自动化设备, 2021,41(2): 200.
- [2] 狄晓栋,李震梅,李宗哲,等. 基于混合特征 MGCC 的干式变压器故障诊断[J]. 电子测量技术, 2021,44(12): 57.
- [3] 刘晨辉,温广瑞,苏宇,等. 基于混合域相对特征和 FOA-XGBoost 滚动轴承退化评估[J]. 振动、测试与诊断, 2021,41(5): 880.
- [4] 武天府,刘征,王志强,等. 基于 Focal 损失 SSDAE 的变压器故障诊断方法[J]. 电力工程技术, 2021,40(6): 18.
- [5] 郑书生,陈金祥,李成榕,等. 变压器绕组对局部放电特高频定位方法的影响[J]. 高电压技术, 2013,39(2): 324.
- [6] 周宏扬,马国明,张猛,等. 基于迈克尔逊光纤干涉的变压器局部放电超声信号检测技术[J]. 中国电机工程学报,2021,42(21): 8016.
- [7] 余昌佳,郑建勇,何嘉弘,等. 时差筛选和 ABC 二次寻优的变压器局放超声定位方法[J]. 高电压技术,2021,47(8):2820.
- [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.
- [9] 狄晓栋,李震梅,李宗哲,等. 基于混合特征 MGCC 的干式变压器故障诊断[J]. 电子测量技术,2021,44(12):57.
- [10] 张建功,王延召,陈豫朝,等. 基于振动和声压测量的特高压变压器声功率估算方法[J]. 高电压技术,2019,45(6): 1843.
- [11] 王文博,曾小梅,赵引川. 基于 SMOTE-XGBoost 的变压器缺陷预测[J]. 华北电力大学学报, 2021,48(5): 54.
- [12] 安国庆,史哲文,马世峰,等. 基于 RF 特征优选的 WOA-SVM 变压器故障诊断[J]. 高压电器, 2022,58(2): 171.
- [13] 赵文清,严海,周震东,等. 基于残差 BP 神经网络的变压器故障诊断[J]. 电力自动化设备, 2020,40(2):143.
- [14] WANG Y X, TAN J X, QIN X L, et al. Python language training system based on MFCC, VQ, variational coefficient and KNTM algorithm [J]. Mathematics and Computer Science, 2021, 6 (2): 38.
- [15] PRABAKARAN D, SRIUPPILI S. Speech processing: MFCC based feature extraction techniques-an investigation [J]. Journal of Physics: Conference Series, 2021,1717: 012009.
- [16] 王荣昊,李喆,孙正. 基于 FISVDD 与 GRU 的变压器声纹识别技术[J]. 高电压技术, 2022, 48 (11): 4546.
- [17] LI B B, CHENG D F, TIAN Y, et al. GIS insulation defect diagnosis method based on improved MFCC and PCA-SVM model [C] // Proceedings of the 5th International Conference on Communication, Image and Signal Processing (CCISP 2020), 2020.
- [18] ZHU K J, MA H Y, WANG J N, et al. Optimization research on abnormal diagnosis of transformer voiceprint recognition based on improved wasserstein GAN [J]. Journal of Physics: Conference Series, 2020,1746: 012067.
- [19] 张重远,罗世豪,岳浩天,等. 基于 Mel 时频谱-卷积神经网络的变压器铁芯声纹模式识别方法 [J]. 高电压技术, 2020,46(2): 413.
- [20] 吴国鑫,詹花茂,李敏. 声纹的变压器放电与机械故障诊断研究 [J]. 应用声学, 2021,40(4): 602.
- [21] XI K, CAI C. Feature selected based on PCA and optimized LMC [J]. MATEC Web of Conferences, 2021,336: 06034.