

基于 Transformer 神经网络的变压器状态监测

冯治国¹ 金日¹ 罗冲² 李昂¹

(1. 国能大渡河检修安装有限公司 成都 610041; 2. 四川大学电气工程学院 成都 610065)

摘要: 电力变压器是电力系统的关键设备,为保障电力系统健康稳定运行,对电力变压器开展状态监测十分必要。提出基于 Transformer 神经网络的变压器状态监测方法,Transformer 神经网络具有自注意力机制,能够挖掘不同特征维度之间的关联性,为变压器状态监测提供更可靠的决策能力。在进行变压器数据收集时,将采集到的数据集分为健康、亚健康、病态 3 个类别;之后采用原始数据、小波特征以及傅里叶特征融合的方式对数据进行预处理,增加特征维度;通过数据生成和 Focal Loss 的方法降低模型训练时数据不平衡带来影响,再将处理后的数据输入 Transformer 神经网络进行模型训练,最终利用训练好的模型预测变压器健康状态。与传统机器学习方法、卷积神经网络、长短时记忆网络相对比,所提方法预测精度有明显提升,能够准确的监测变压器设备状态,预测准确率能达到 90%,是一种有效的变压器状态监测方法。

关键词: 电力系统;电力变压器;状态监测;Transformer 神经网络;特征融合

中图分类号: TM41 **文献标识码:** A **国家标准学科分类代码:** 470.4099

Power transformer state monitoring based on transformer deep neural network

Feng Zhiguo¹ Jin Ri¹ Luo Chong² Li Ang¹

(1. CHN Energy Dadu River Maintenance and Installation Co., Ltd., Chengdu 610041, China; 2. College of Electrical Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Power Transformer is the critical equipment in electric power system. Transformer status monitor is very important for electric power system stability. This paper proposed a Transformer neural network based power Transformer monitoring system. Transformer neural network deeply mined the relationship between different input feature dimension, which offers a more stable and high accuracy power for Transformer decision system. Training data set was labeled as health status, sub health status and failure status. In feature extraction step, this paper used multiple feature fusion method which included origin data, wavelet transform feature and Fourier transform feature. For training, we used data generator and Focal loss method to avoid data imbalance influence. Then the processed data was used for model training, we used the trained model for current power transformer status prediction. The accuracy of the proposed method is better than traditional machine learning methods and other deep learning model, which achieve almost ninety percent. The experiment result shows this method can be used for power transformer monitor system.

Keywords: power system; power transformer; status monitor; transformer neural network; feature fusion

0 引言

电力变压器在电力系统中是关键枢纽设备,在发电侧,电力变压器提高发电厂输出的电压,使之满足输送电网额定高电压的要求,将发电厂所发的电能变送到电网中去。如果变压器出现故障,则电能无法正常上网,将造成不同程度的经济损失。变压器故障是导致电网事故的

重要原因,如果变压器出现故障而未及时发现,可能造成大停电对人民的生产生活造成重大影响。由此可见电力变压器的安全稳定与电网稳定运行密切相关,及早发现变压器故障有利于降低事故的严重性。为保障变压器安全稳定的运行,在选择技术可靠、质量优良的变压器以外,还应提高变压器的运行维护、检修技术水平,利用科学手段做到防患于未然。

传统的电力变压器健康管理主要依赖于定期检修和专家经验,定期检修不可避免存在“过维修”和“欠维修”欠维修的情况,造成资源浪费且效率低下,难以满足智能电网背景下变压器安全稳定运行的要求;而专家经验对变压器的状态评估存在主观因素,且由于不同生产厂家、不同运行环境的变压器在同种条件下其特性存在差异,依赖专家经验难以有效判断变压器的复杂故障。随着通信技术、计算机技术、大数据处理技术等先进技术在电力系统中的逐步应用。文献[1]分析了电力变压器检测技术的研究现状以及在发展过程中存在的问题,同时结合当前的技术发展,概述了电力变压器故障检测技术,包括油气分析技术、红外光谱技术、气相色谱仪技术以及电路故障分析技术等相关应用。文献[2-3]收集变压器的振动信号,将振动信号通过傅里叶变换和小波变换,在频率域将故障信号划分不同的频段进行分析。文献[4-5]采用传统的机器学习算法支持向量机(support vector machine, SVM)和集成算法(Boost)方法作为决策单元对变压器故障进行检测。文献[6]采用基于随机森林的方法对变压器故障进行建模,以此实现变压器故障诊断。文献[7-8]在随机森林的基础上分别使用麻雀搜索优化和粒子群优化的方法来提高监测模型的准确度。诸如此类传统机器学习方法高度依赖特征工程,决策模型的精度受限于特征提取算法的有效性,以及传统决策算法的非线性映射能力。近年来,随着深度学习方法的研究和推广,传统的机器学习方法逐步被替代,相对于传统机器学习方法,深度学习方法大大降低了特征提取的重要性,通过复杂的网络结构,将特征提取的步骤融入到网络学习的过程中,让网络自动去学习有效特征。文献[9-10]训练一个反向传播(back propagation, BP)神经网络方法得到一个变压器故障诊断的分类器,文献[11-12]在深度学习的基础上,利用射频识别(radio frequency identification, RFID)传感器标签技术采集传感器震动信号,对变压器状态进行监测和故障诊断。文献[13]采用深度循环神经网络(recurrent neural network, RNN)结构作为骨干网络来预测变压器的欠励、过励以及匝间故障。文献[14]将长短期记忆网络(long short-term memory, LSTM)和多策略改进麻雀算法相结合,训练神经网络,提升模型的训练速度和检测精度。

本文提出基于 Transformer^[15]神经网络模型的变压器状态监测算法,Transformer神经网络模型相对于卷积神经网络(convolutional neural network, CNN)或者 LSTM,由于其独特的自相关结构,能够对不同特征维度之间的关系进行充分地挖掘,使得 Transformer 网络模型相对于其他深度网络,其对特征维度之间的数据关系更为敏感;而由于其较为复杂的多层网络模型结构,也能够为决策层提供优秀的非线性映射能力,更适用于变压器监测方法的开发。本文提出的方法通过传感器采集的大量的

多维度的变压器运行历史数据,对设计好的模型进行训练;在数据预处理阶段,本文采用多维度数据融合的方式,减低模型训练对数据体量的要求,加快模型的收敛。实验证明,基于 Transformer 模型的神经网络对变压器状态监测能够提供更好的精度判断。

1 变压器状态监测方法框架

本文提出基于 Transformer 深度神经的变压器状态监测方法,整体框架如图 1 所示。

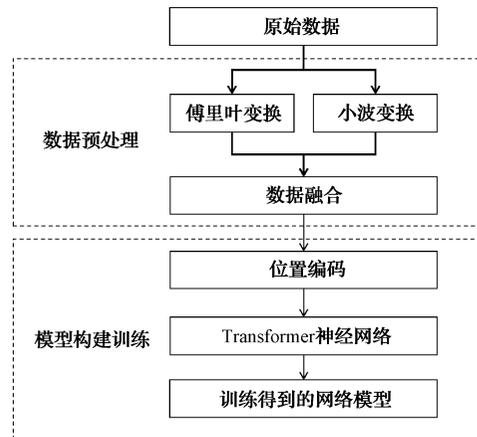


图 1 变压器状态监测方法框架

变压器状态监测方法主要分为数据预处理和 Transformer 模型训练两个部分。数据处理主要分为傅里叶变换、小波变换^[16]、数据融合 3 个模块。傅里叶变换得到原始数据的频谱特性,小波变换得到原始数据的频率能量特性,原始数据中包含了数据本身的时域特性。数据融合模块将获得的数据进行数据对齐和拼接,整理好的所有数据将随机分为训练数据集和测试数据集送入 Transformer 模型中进行训练。训练过程中当测试集的精确度能达到预设值时,终止训练,保存模型。在模型推理阶段,方法流程如上图 2 所示,在获得传感器上传的实时数据后,用同样的方式进行数据预处理,将处理后的数据送入训练好的模型中进行推理,最终得到当前数据对应的状态值。

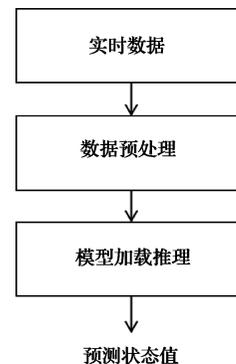


图 2 模型推理流程

2 数据预处理

2.1 傅里叶变换

傅里叶变换是一种线性积分变换,用于信号在时域和频域之间的变换,是一种全局变换,时域信号经过傅里叶变换后,得到信号的频域表达,能观测到时域观测不到的特征,对于一个连续信号 $f(t)$,其变换公式如下:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (1)$$

在实际应用中,计算机只能处理离散信号,因此,对连续信号 $x(t)$ 进行时域采样,得到离散样本 $x(n)$,对该样本进行离散傅里叶变换,其公式如下:

$$X(\omega) = \sum_{n=-\infty}^{+\infty} x(n)e^{-j\omega n} \quad (2)$$

变换后,在频域进行采样得到:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}} \quad (3)$$

傅里叶变换对于平稳信号的频率特性有很好的表达,但对非平稳信号处理欠佳。非平稳信号时域信息也非常重要,因此,在数据预处理阶段,使用小波变化对非平稳信号频域表达进行补充。

2.2 小波变换

小波变换本质上是采用小波基替换傅里叶变换中的正弦基对信号进行频率分解。相对于正弦基,小波是一种在时域非常集中的波,其能量是有限的,且集中在某一点附近。对于分析瞬时时变信号非常有效,通过伸缩和平移等运算功能对信号进行多尺度细化分析,解决了傅里叶变换的时域信息丢失等问题。对于信号 f 其连续小波变换定义为:

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\varphi\left(\frac{t-b}{a}\right) dt \quad (4)$$

式中: $\varphi(t)$ 是小波基,满足 $\varphi(\pm\infty) = 0$, $\varphi(0) = 0$, $\int_{-\infty}^{+\infty} \varphi(t) dt = 0$ 。该方法使用的 Symlets 小波基族进行小波变换。

2.3 数据融合

本文首先对原始数据进行傅里叶变换和小波变换之后,构建 Transformer 网络模型的输入序列和目标序列。本文将原始数据和傅里叶特征以及小波特征进行统一维度拼接,构建 Transformer 网络的输入序列,如图 3 所示。目标序列采用输入序列右移(shifted right)方法滑窗处理得到。

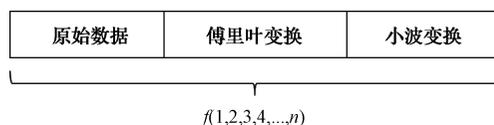


图 3 数据融合

3 Transformer 网络模型

3.1 模型架构

Transformer 神经网络结构由谷歌于 2017 年发布,其最初主要是应用在自然语言处理领域,近年来逐步的扩散到了其他领域,应用非常广泛。该网络的创新点主要是自注意力机制(self-attention),而自注意力机制其本质主要是寻找输入特征维度之间的关联性,并做出预测。Transformer 网络模型如图 4 所示。

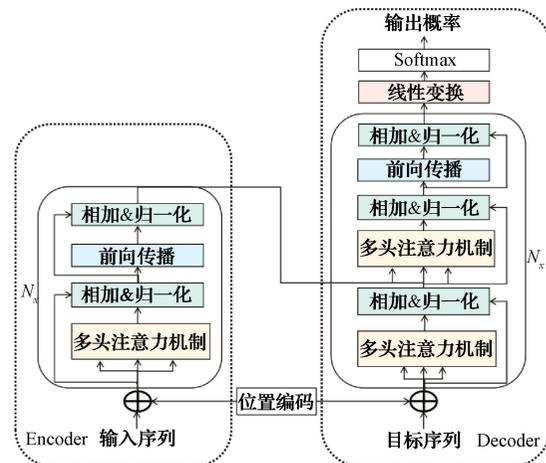


图 4 Transformer 网络模型结构

Transformer 网络模型可分为编码(encode)和解码(decode)两部分。编码有两个子层,一个是多头注意力机制(multi-head attention)层,利用自注意力机制学习源数据不同维度之间的关系。另一个是前向传播(feed forward)层,即简单的全连接网络,对每个位置的向量分别进行相同的操作,包括两个线性变换和一个 ReLU 激活函数,之后产生编码,将输出数据传给解码层。解码层中有 3 个子层,其中两个多头注意力机制层。下面的注意力机制层是利用自注意力机制学习数据内部的关系。之后该层输出与编码传过来的结果一起输入到上面的注意力层,注意力层是编码与解码之间的相关性计算,用来挖掘输入序列与目标序列之间的关系。

3.2 位置编码

在神经网络模型中,通过优化器迭代训练,模型会自动挖掘不同输入数据之间的关系,构建有效的输入特征矩阵。在循环卷积网络模型中,模型天生就包含了数据序列的位置信息,而在卷积神经网络中,卷积核其本质就是挖掘数据局部位置之间的相对关系,而 Transformer 网络结构中的自相关结构无法获取位置信息,因此,需要对输入的数据序列进行位置编码,加入位置信息。

Transformer 网络模型中的位置编码大致可分为两大类,基于表格的归一化方法和基于函数的相对位置关系方法。基于表格的归一化方法更适用于固定长度的输入序列,而本文提出的方法希望网络模型可以更为灵活的去

适应不同长度的特征序列,方便进行扩展,因此,本文选用基于正弦和余弦函数的位置编码方法:

$$\vec{P}_i^{(t)} = f(t)^i = \begin{cases} \sin(\omega_k \cdot t), & i = 2k \\ \cos(\omega_k \cdot t), & i = 2k + 1 \end{cases} \quad (5)$$

式中: \vec{P}_i 表示位置 t 对应的向量; i 为输入数据位置; $f(t)$ 为生成位置向量 \vec{P}_i 的函数; ω_k 为频率。

$$\omega_k = \frac{1}{10\,000^{2k/d}} \quad (6)$$

式中: d 为输入向量长度,频率沿着向量维度而减小,而得到的位置编码 P_i 是一个包含了频率的正弦余弦对,根据式(7)进行位置信息嵌入。

$$\psi'(\omega_i) = \psi(\omega_i) + \vec{P}_i \quad (7)$$

3.3 模型训练

模型训练阶段,由于 Transformer 模型训练需要大量的数据,本文采用 K 折交叉验证的方式对数据集进行划分,弥补数据集不足的同时,保证模型的泛化能力,防止过拟合。而由于在数据集的采集过程中,变压器异常状态数据非常少,会导致数据集不平衡,本文采用两种方式来弱化数据集不平衡带来的影响。1)数据生成,假定数据是连续的,基于现有数据集中异常样本数据,通过随机插值法,生成新的异常数据样本;2)采用 Focal Loss 的方式,对异常数据样本损失值赋予更大权重,让模型在训练过程中更关注异常数据,提高检测率,抑制由于数据不平衡带来的影响。

在模型训练过程中,将解码器的输出和当前训练数据的标签一起送入损失函数计算损失值,本文采用交叉熵损失函数。为了得到最小损失值,本文采用 ADAMAX 优化器训练模型,当损失值小于预设值的阈值 T 或者迭代次数大于最大迭代次数时,训练终止。阈值 T 和最大迭代次数为模型超参数。

4 实验结果

4.1 数据采集与标注

本文采用 Transformer 深度神经网络来对变压器健康状态进行实时监测,在实验中,对每个变压器分别采油气数据和振动数据,其中油气数据主要包含 H_2O 、 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 、 C_2H_6 、 CO 、 CO_2 、 O_2 、 N_2 等数据特征,振动数据通过装在变压器上的振动传感器进行数据采集。数据收集后,由专业人员对数据进行分类标记。之后通过数据生成的方式进行数据增广。实验采用的变压器型号为 SSP-723000/500,3 相 50 Hz,额定容量 723 000 kVA,额定电压 550/18 kV,冷却方式为强迫油循环冷却(ODWF),变压器参数如表 1 所示。

将变压器状态分为健康、亚健康、病态 3 种状态,病态表示当前变压器状态很差,需要立刻进行检测和维修;健康表示当前变压器状态良好,亚健康状态介于病态和健康状态之间,表示当前变压器可运行,但是需要特别关注其

表 1 电力变压器详细参数

型号	SSP-723000/500	短路阻抗/%	16.24
3 相/Hz	50	空载电流/%	0.04
额定容量/kVA	723 000	空载损耗/kW	244.4
额定电压/kV	550/18	联结组标号	YNd11
冷却方式	ODWF	负载损耗/kW	1 151
器身重/t	332	油重/t	85.8
总重/t	502.5	上节油箱重/t	31.7
绝缘水平			
h. v. :	线路端子 SI/LI/AC	1 175/1 550/680 kV	
h. v. :	中性点端子 LI/AC	185/85 kV	
l. v. :	线路端子 LI/AC	125/55 kV	

状态变化。整个数据集分布如图 5 所示。

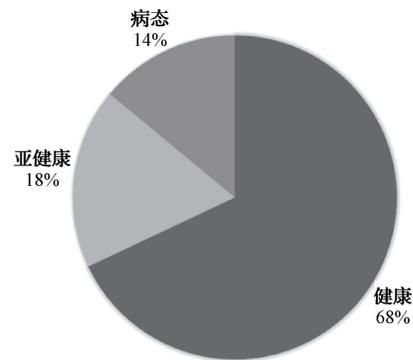


图 5 数据集分布状态

4.2 输入数据组合消融验证

数据标注完后,本文采用 10 折交叉验证的方式,对数据集进行平均划分,构建训练集和测试集。在预处理阶段,采用小波变换、傅里叶变换以及原始数据的方式,进行数据融合后对 Transformer 模型进行训练。

针对如何进行输入数据组合,本文分别对上述特征不同数据组合的方式进行消融验证,如图 6 所示。

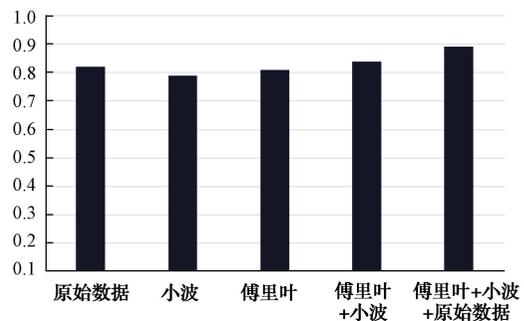


图 6 不同输入消融验证

实验结果表明,融合了原始数据加小波特征以及傅里叶特征的方式表现最好,其原因本质在于通过数据预处理扩大了特征维度,模型可以更有效地进行特征挖掘。

4.3 决策模型对比

机器学习模型分为传统机器学习方法和基于深度学习的方法,在传统机器学习方法中,实验筛选了朴素贝叶斯(naive Bayes)、KNN(K-nearest neighbor)、线性判别分析(LDA)、SVM、决策树(tree bagging)等方法作为决策模型和 Transformer 方法进行对比,如图 7 所示。

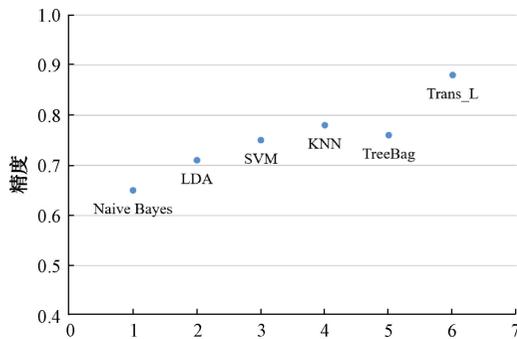


图 7 Transformer 与传统机器学习方式

实验结果表明,基于 Transformer 网络模型的方法效果远远领先于传统机器学习方法,精度表现优于超过传统方法 15%。

在基于深度学习的方法中,实验选取了常用的 CNN 模型和 LSTM 模型进行验证比对。CNN 模型由于其卷积结构,需要对输入数据组合成一个二维的输入向量,如图 8 所示。

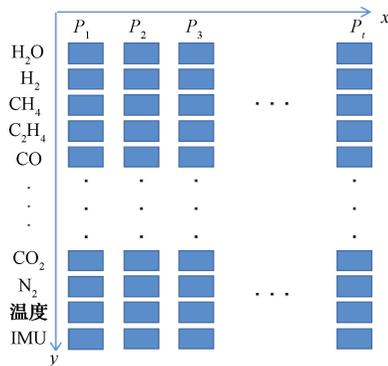


图 8 CNN 模型输入构建

在 y 方向上,为某一时刻 t 各类型传感器采集的不同维度的数据, x 方向是时间维度,连续不同的时刻。因此,对于 CNN 模型来说,当前时刻 P_t 的映射关系可以等价为:

$$f(P_t) \rightarrow f(P_t, P_{t-1}, P_{t-2} \dots P_{t-n}) \quad (8)$$

其中, f 为 CNN 非线性映射模型, n 为超参数,在实验中 n 取 50。CNN 模型采用 3 个卷积层,3 个池化层,1 个全连接层,卷积核采用 5×5 尺寸大小卷积核,构建 CNN 网络模型。而 LSTM 网络模型采用双层 LSTM 层构建,单层由 128 个神经元构成^[17]。数据采用同样的健康、亚健康、病态 3 类划分的数据集进行训练和验证,3 类

模型的表现如表 2 所示。

表 2 不同网络模型表现对比

模型对比	病态		亚健康		健康		平均精度
	精度	误报	精度	误报	精度	误报	
CNN	0.779	0.054	0.819	0.065	0.819	0.067	0.805
LSTM	0.803	0.044	0.872	0.049	0.865	0.049	0.846
Transformer	0.83	0.032	0.91	0.034	0.903	0.04	0.881

对训练好的模型,按照类别分别统计其模型推理的准确率和误报率,准确率 P_i 计算公式如下:

$$P_i = \frac{T_i}{N_i} \quad (9)$$

式中: N_i 为样本中标注为 i 类的样本集, $i \in$ (病态、亚健康、健康); T_i 为标注为 i 类且模型推理为 i 类的样本集。误报率 E_i 的计算公式如下:

$$E_i = \frac{F_i}{N_{att}} \quad (10)$$

式中: N_{att} 为所有样本集的数量; F_i 为标注为非 i 类且模型推理为 i 类的样本集数量。可以看出,相对于 CNN 和 LSTM 模型,Transformer 模型在相同的数据下,其每一类的准确率表现都更好,符合实验预期。

Transformer 模型表现更好的原因是由于其模型自身的 self-attention 机制能够更好的挖掘不同维度输入的相关关系,在增加准确性的同时,相应的也提升了模型的复杂度。为了验证不同深度的 Transformer 神经网络对模型性能的影响,实验构建了 3 个不同深度的 Transformer 神经网络模型,将 Transformer 神经网络为 8 层、12 层、16 层,分别标记为 Trans_8、Trans_12、Trans_16,3 个不同深度模型的表现如表 3 所示。

表 3 不同深度网络模型表现对比

模型对比	病态		亚健康		健康		平均精度
	精度	误报	精度	误报	精度	误报	
Trans_8	0.766	0.059	0.833	0.063	0.824	0.061	0.808
Trans_12	0.830	0.032	0.910	0.034	0.903	0.040	0.881
Trans_16	0.826	0.031	0.908	0.034	0.908	0.039	0.881

实验结果表明,随着模型深度的加深,模型复杂度的提高,模型的决策精度能够进一步提升,然而当模型深度过度加深后,其带来的精确度的提升度在下降,可能对某一细分分类其精度在上升,但是提高并不明显,反而模型复杂度却在上升,计算代价大大增加。综上可知,Transformer 网络模型在训练过程中,可以自动构建用于做决策的特征,相对于其他决策模型表现更好,符合实验预期。

5 结论

本文提出了基于 Transformer 深度神经网络的变压

器健康状态监测方法,在数据输入阶段,采用原始数据加小波特征加傅里叶变换特征融合的方式进行数据预处理,数据预处理在加快模型训练收敛速度的同时,也加宽了数据维度,提高了模型的决策精度。相对于传统基于人工特征的机器学习方法和 CNN 和 LSTM 深度神经网络,Transformer 模型基于其自注意力机制,会主动挖掘不同特征之间的关联关系,尤其是加深网络层级后,Transformer 网络模型由于自身复杂的网络结构带来的更强的非线性映射的能力,因此在电力变压器状态监测方法中,能够提供更高的决策精度。

利用 Transformer 网络的特性以及自适应位置编码方法,本文提出的方法可扩展数据输入的维度。在未来的研究中,可以尝试在输入端添加更多维度的特征,例如红外光谱特征、变压器运行时长等,在输入端构成更大的特征矩阵,获得更高的决策精度。另外在工程化应用中,未来可以对模型进行边缘端部署,完成实时监测运行。

参考文献

- [1] 张国栋,苗青,滕伟,等.基于振动特征主分量分析的电力变压器故障诊断[J].变压器,2019,56(5):62-65.
- [2] 刘彪,于一三,白祥宇,等.基于振动信号分析的变压器故障检测系统的研究[J].电气自动化,2020,42(5):80-82,86.
- [3] 吴琼,崔昊杨,周坤,等.多源检测数据融合的变压器故障诊断模型[J].现代电子技术,2022,45(1):181-186.
- [4] 张琳,马宏忠,王涛云,等.基于振动-SVM的变压器绕组缺陷诊断方法[J].陕西电力,2016,44(11):14-18.
- [5] 郭林,唐晶,唐黎哲,等.一种基于改进BP神经网络的变压器故障诊断方法[J].大功率变流技术,2021(5):71-77.
- [6] 殷作洋,吴肖锋,仲伟坤.基于随机森林的变压器故障识别与实例分析[J].电子测量技术,2020,43(23):63-67.
- [7] 曹伟嘉,杨留方,徐天奇,等.基于精英反策略麻雀搜索优化随机森林的变压器故障诊断[J].国外电子测量技术,2022,41(2):138-143.
- [8] 杨毅,徐晓冉,刘雅琳,等.基于离散二进制粒子群优化随机森林的平面变压器寄生参数预测模型[J].国外电子测量技术,2021,40(4):37-41.
- [9] 张景明,刘建国.粗糙集和BP神经网络在变压器故障诊断中的应用[J].变压器,2009,46(4):18-21.
- [10] 黄新波,王娅娜,刘林,等.变压器油色谱在线监测中BP神经网络算法分析[J].陕西电力,2013,41(6):56-60.
- [11] 何怡刚,汪涛,施天成,等.基于RFID传感器标签与深度学习的变压器状态监测方法研究[J].电子测量与仪器学报,2018,32(9):72-79.
- [12] 张朝龙,何怡刚,杜博伦,等.基于深度学习的电力变压器智能故障诊断方法[J].电子测量与仪器学报,2020,34(1):81-89.
- [13] 李辉,黄祖源,田园,等.基于深度神经网络的变压器故障诊断方法研究[J].变压器,2022,59(4):35-40.
- [14] 王雨虹,王志中,付华,等.多策略改进麻雀算法与BiLSTM的变压器故障诊断研究[J].仪器仪表学报,2022,43(3):87-97.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [16] 马学娟.基于快速傅里叶变换(FFT)和小波变换的大型风机机械振动故障的分析[J].科技与创新,2016(11):121-121,125.
- [17] 邱大伟,刘子辰,周一青,等.基于Transformer神经网络的滚动轴承故障类型识别[J].高技术通信,2021,31(1):1-11.

作者简介

冯治国,硕士,高级工程师,主要研究方向为水电站机电设备检修管理及科技创新等。

罗冲(通信作者),博士研究生,主要从事装备故障诊断相关研究。

E-mail:luochong33@foxmail.com