激光写光电子学进展

研究论文

基于集成增强 KNN 的油纸绝缘原始拉曼光谱 老化状态判别

陈新岗^{1,2},范益杰¹,马志鹏^{1*},谭世耀¹,李宁一¹,宋成¹,黄宇杨¹,张金京¹,张文轩¹ ¹重庆理工大学电气与电子工程学院,重庆 400054; ²重庆市能源互联网工程技术研究中心,重庆 400054

摘要 油纸绝缘老化状态的快速准确检测一直备受关注,本研究对未进行特征提取的油纸绝缘原始拉曼光谱老化状态 类别进行判定。首先,根据所测得的绝缘纸聚合度将其老化状态划分为10个类别。同时,对不同老化状态的油纸绝缘样 本进行拉曼光谱检测。最后,通过K最近邻(KNN)算法、集成增强KNN算法分别对169组拉曼光谱样本进行老化状态 类别判定。结果表明:经过集成增强后的KNN算法对原始拉曼光谱具有更强的识别能力,其判别准确率为98.32%,且 具有更好的稳定性。证明了由集成增强KNN算法构建的判别模型能够较为准确地判别油纸绝缘原始拉曼光谱,该模型 简化了变压器油纸绝缘拉曼光谱老化状态的诊断步骤,对油纸绝缘拉曼光谱检测方面的研究具有重要意义。 关键词 拉曼光谱;油纸绝缘;集成增强;状态判别 中图分类号 O433.4 文献标志码 A DOI: 10.3788/LOP231050

DOI: 10.3788/LOI 231030

Aging State Discrimination of Oil-Paper Insulation Using Raman Spectroscopy and Integrated Enhanced KNN

Chen Xingang^{1,2}, Fan Yijie¹, Ma Zhipeng^{1*}, Tan Shiyao¹, Li Ningyi¹, Song Xin¹, Huang Yuyang¹, Zhang Jinjing¹, Zhang Wenxuan¹

¹School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China; ²Chongqing Engineering Research Center of Energy Interconnection, Chongqing 400054, China

Abstract The rapid and accurate detection of the oil-paper insulation aging state has attracted considerable attention. In this study, classification of the original Raman spectral aging state of oil-paper insulation is performed without feature extraction. First, the aging state of the insulation paper is divided into 10 categories according to the measured polymerization degree. Raman spectroscopy is performed on the oil-paper insulation samples in each aging state. Finally, 169 groups of Raman spectra are classified by the *K*-nearest neighbour(KNN) algorithm and integrated enhanced KNN algorithm. The results indicate that the KNN algorithm after integration enhancement has a stronger recognition ability for the original Raman spectrum, its discriminant accuracy is 98.32%, and it has better stability. It is proved that the discriminant model based on the integrated enhanced KNN algorithm accurately discriminates the original Raman spectra of oil-paper insulation. The proposed model simplifies the diagnosis of the aging state of transformer oil-paper insulation using Raman spectra and is of considerable significance for research on this topic.

Key words Raman spectroscopy; oil-paper insulation; integrated enhancement; state discrimination

1引言

变压器作为电力系统的重要组成部分,其绝缘性 能关乎电网的稳定运行^[1]。油纸绝缘为电力变压器常 用的绝缘方式^[2]。因此,准确预测油纸绝缘老化状态 能为电力系统的安全运行提供可靠保障。针对油纸绝 缘的检测,常用的检测方法有:聚合度检测法、油中糠 醛含量检测法、碳氧气体检测法等^[34]。聚合度检测法

收稿日期: 2023-04-07;修回日期: 2023-05-08;录用日期: 2023-05-19;网络首发日期: 2023-06-09

基金项目:重庆市教育委员会科学技术研究项目(KJZD-K202101103)、重庆理工大学科研启动基金(2021ZDZ016)、重庆理工 大学研究生教育高质量发展行动计划(gzlcx20233098)

第 60 卷第 21 期/2023 年 11 月/激光与光电子学进展

研究论文

具有破坏性且检测步骤复杂;油中糠醛含量检测法需 要复杂的实验预处理过程且须在实验室进行;碳氧气 体检测法易发生色谱柱老化且所需的检测时间过 长^[5-6],上述方法未能满足快速、准确地检测实际运行 变压器的老化状态要求。

拉曼光谱具有检测速度快、与待测物无直接接触、 检测精度高等优点,近年来,其已在石油、化工及医疗 方面的物质检测中得到广泛应用^[7-9]。将拉曼光谱检 测技术引入变压器油纸绝缘的老化状态检测中,可满 足变压器实际运行中快速、准确、无损耗等需求^[10]。

变压器绝缘油中包含大量老化特征物,其拉曼光 谱含有所有老化特征物信息。针对多维拉曼光谱数据 集,已有的单一机器学习算法构建的判别模型不能满 足绝缘油检测的稳定性及精度要求^[11]。相关研究学者 通过多维尺度变换等方式降低光谱数据量,从而提高 后续判别模型的准确率。但过于复杂的特征提取方 法,使得老化特征物的原始拉曼光谱信息丢失,且增加 了数据处理的程序^[12]。因此,研究一种针对油纸绝缘 原始拉曼光谱数据的判别模型具有重要意义。

集成增强算法可充分利用已有的老化样本标签, 通过对多个单一分类模型的集成优化构建出高精确的 判别模型。K-最近邻(KNN)算法可对拉曼光谱样本 进行状态识别,但其准确率、稳定性未能满足油纸绝缘 老化程度判定要求。因此,引入自适应增强算法 (Adaboost)整合多个KNN从而构建出集成增强KNN (AA-KNN)模型,从而提高判别模型的准确率、稳定 性、可信度^[13-15]。

本研究中,采用加速热老化实验模拟变压器油纸 绝缘老化过程,通过绝缘纸聚合度检测制定不同老化 时间样本的老化标签;运用自主搭建的实验平台对绝 缘油进行拉曼光谱信号采集;采用 Savitzky-Golay 法 和3次样条平滑法对拉曼光谱信号进行降噪与基线校 正等数据预处理;分别采用 KNN 算法与 AA-KNN 算 法构建判别模型对原始拉曼光谱进行老化状态判别; 从准确率、稳定性、可信度等方面对比分析集成增强所 带来判别能力的提升效果。最后,引入极度梯度提升 决策树(XGBoost)、随机森林(RF)算法来构建判别模 型并与 AA-KNN 进行对比研究。

2 实验部分

2.1 变压器油纸绝缘热老化实验

根据IEEE标准,选择130℃为加速热老化实验温度,油纸比定为10:1,老化前10d,每2d采样一次;老化10~20d,每4d采样一次;老化20~30d,每5d采样 一次^[15],热老化试验流程如图1所示。

采样之后分别测量绝缘纸在铜乙二胺溶液与铜乙 二胺溶剂中的黏度,计算出比黏度,再通过马丁经验方 程由比黏度与浓度计算出溶液的本征黏度[v],最后计



图 1 热老化实验流程 Fig. 1 Thermal aging experimental process

算绝缘纸聚合度 fue 可表示为

$$f_{\rm DP}^{\alpha} = \lceil \nu \rceil / R, \qquad (1)$$

式中:*R*为绝缘纸与铜乙二胺溶剂比值;α为马克-豪温 克特征系数。

不同老化天数下绝缘纸聚合度变化情况如图2所示。由图2可知,在热老化前期,绝缘纸聚合度下降速 率较快,中后期下降速率逐渐减慢,这也符合实际运行 变压器的绝缘老化规律。





根据DL/T 984-2018 导则,并基于老化状态类别 划分越多、老化状态判别越准确的原理,以绝缘纸聚合 度为依据对老化样本进行类别划分^[16],如表1所示。

2.2 绝缘油拉曼光谱信号检测与预处理

拉曼光谱检测平台如图 3 所示,激光器的功率设置为 300 mW,积分时间为 0.25 s,积分次数为 10次。 检测前设置 CCD 内部温度为-20 ℃以提高探测器灵

研究论文

表1 老化状态类别划分 Table 1 Classification of aging status categories

1 4610 1	Chabbiliteation of aging status categories		
Aging state	Degree of Sample s		
category	polymerization range	Sample Size	
Phase 1	>1000	30	
Phase 2	900-1000	16	
Phase 3	800-900	19	
Phase 4	700-800	17	
Phase 5	600-700	17	
Phase 6	500-600	18	
Phase 7	350-500	15	
Phase 8	250-350	13	
Phase 9	150-250	12	
Phase 10	0-150	12	

敏度。为避免外部光线干扰,实验在黑暗环境中进行。

为降低样本拉曼光谱噪声,并抑制不同老化阶段的基线漂移,引入Savitzky-Golay法和3次样条平滑法分别对拉曼光谱样本进行降噪与基线扣除等预处理,如图4所示。







第 60 卷第 21 期/2023 年 11 月/激光与光电子学进展





3 油纸绝缘老化状态判别模型

判別模型有两种类型,单一机器学习分类器与多 个机器学习通过串行或并行方式集成获得的强分类 器。Adaboost采用串行的方式整合单一弱分类器,在 每次对样本进行类别判定前利用已有的判别结果调整 样本权重,使后续更注重对错误判别样本的分类。因 此,Adaboost可提升单一分类器的判别准确率与稳定 性。KNN算法作为最常见的分类算法,其原理简单, 利用训练集样本已有老化标签对测试集样本进行老化 状态判别,具有监督特性,但随机性较强,难以满足针 对样本的重复性判别中稳定性与准确率的要求。

引入 KNN 作为 Adaboost 的基分类器构成 AA-KNN,可弥补油纸绝缘拉曼光谱样本应用传统 KNN 判别模型时的准确率与稳定性不足, AA-KNN 的流程 如图 5 所示。

首先,定义第i个拉曼光谱样本为

$$\{(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, 2, \cdots, n\},$$
(2)

式中:n为拉曼光谱样本总数;x_i为样本拉曼光谱向量, 其向量维度为*d*,即每一拉曼光谱样本数据点数;y_i为 拉曼光谱样本老化标签,如表1所示。

在已有研究基础上^[17],将KNN判别模型用于油纸 绝缘拉曼光谱样本分类时,对其距离度量方式进行改 进,定义欧氏距离与余弦距离的比值作为新度量方式 D_{ab},D_{ab}可表示为

$$D_{ab} = L_{ab} / \sin\left(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}\right), \qquad (3)$$

式中:Lab为两拉曼光谱样本欧氏距离,Lab可表示为

$$L_{ab} = \sqrt{\sum_{l=1}^{d} \left| \mathbf{x}_{i}^{l} - \mathbf{x}_{j}^{l} \right|^{2}}, \qquad (4)$$

 $sin(\mathbf{x}_i, \mathbf{x}_j)$ 为拉曼光谱样本间的余弦距离, $sin(\mathbf{x}_i, \mathbf{x}_j)$ 可表示为

$$\sin(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|^{\circ}}$$
(5)

改进KNN距离度量方式后,对拉曼光谱样本进行





老化状态判别,KNN近邻参数参照下式设置为12。

$$N = \min\{n_1, n_2, \cdots, n_m\},\tag{6}$$

式中:n_x为第x类别的拉曼光谱样本数量。

KNN判别模型对拉曼光谱判别过程为

$$\bar{y}_i = \arg\max_c \sum_{i=1}^N \lfloor y_i = c \rfloor, \qquad (7)$$

式中: $[y_i = c]$ 为指示函数,若该等式成立,则函数值为 1,否则为0,遍历样本周边距离最近的N个样本,其出 现次数最多的老化标签即为样本的判别结果 \bar{y}_i 。得出 每个KNN判别结果误差率 e_k , e_k 可表示为

$$e_{k} = \sum_{i=1}^{n} (y_{i} - \bar{y}_{i})_{\circ}$$
(8)

第k个KNN判别结果权重系数 ∂_k 为

$$\partial_k = \frac{1}{2} \ln \frac{1 - e_k}{e_k} + \ln(m - 1),$$
 (9)

式中:m为样本总类别数,第k+1次判别的样本权重系数w_{k+1}由第k次判别结果更新得到,w_{k+1}可表示为

$$w_{k+1} = \frac{w_k}{Z_k} \exp\left(-\partial_k y_i \bar{y}_i\right), \qquad (10)$$

式中:Z_k为规范化因子,Z_k可表示为

$$Z_{k} = \sum_{i=1}^{n} w_{k} \exp\left(-\partial_{k} y_{i} \bar{y}_{i}\right)_{\circ}$$
(11)

设置AA-KNN的基础分类器KNN个数为3,通过 多个KNN模型判别结果与权重系数采用加权平均求 和法得到拉曼光谱判别结果:

$$\hat{y}_i = \operatorname{sign}\left(\sum_{k=1}^{K} \partial_k \bar{y}_i\right), \qquad (12)$$

式中:K为基分类器 KNN的数量; \hat{y}_i 为拉曼光谱经 AA-KNN老化状态判别后的最终结果。

4 结果与讨论

4.1 拉曼光谱检测结果

不同老化天数油纸绝缘老化样本经预处理后的拉 曼光谱如图6所示。不同老化阶段样本在拉曼频移 2800~3000 cm⁻¹范围内及1450 cm⁻¹附近拉曼谱峰强 度具有明显差异,1450 cm⁻¹附近的拉曼谱峰主要来自 于有机分子基团中C—H键的弯曲振动与C=O键的 伸缩振动,2800~3000 cm⁻¹的拉曼谱峰主要由C—H 键伸缩振动产生。在老化过程中,绝缘油氧化反应带 来的酮类物质、绝缘纸中纤维素裂解产生的呋喃类物 质与醇类物质以及油纸共同老化带来的酸类物质中含 有大量的C—H键与C=O键,使得油纸绝缘不同老化 阶段物质含量及物质种类的差别可通过拉曼谱峰强度 差异体现。





4.2 两模型判别结果

将169组样本按7:3划分为训练集与测试集,并 参照表1为样本添加老化标签。分别引入传统KNN 算法与AA-KNN算法构建老化状态判别模型,对油 纸绝缘拉曼光谱样本进行老化状态预测,设置随机数 种子为100,两种判别模型预测结果如图7所示。

由图 7 可知,51 组测试集样本的 AA-KNN 算法判别结果仅有处在第6阶段的一个样本被错误划分在第

第 60 卷第 21 期/2023 年 11 月/激光与光电子学进展



图 7 油纸绝缘样本老化状态判别结果。(a) AA-KNN;(b) KNN Fig. 7 Discrimination results of aging state of oil-paper insulation samples. (a) AA-KNN; (b) KNN

5阶段,其余50组样本老化类别均被正确预测,KNN 算法判别结果分别在老化阶段1、3、4、6、9的一个样本 出现预测类别错误,被错误划分至相邻类别。

为保证判别模型的稳定性,以及验证模型的判别 准确率,随机数种子数量依次增加100,从51组样本中 随机挑选测试集与训练集重复进行50次判别,并计算 判别准确率fauc可表示为

$$f_{\rm AUC} = \frac{f_{\rm TP}}{f_{\rm TP} + f_{\rm FP}},\tag{13}$$

式中:f_{TP}为测试集中正确预测样本数;f_{FP}为测试集中 错误预测样本数。针对同一批次样本多次判别准确率 如图8所示。



图 8 多次实验准确率变化图 Fig. 8 Variation of accuracy of multiple experiments

由图 8 可知, AA-KNN 准确率范围为 94%~ 100%, 传统 KNN 准确率范围为 82%~98%。此外, AA-KNN 判别模型每次准确率均大于或等于传统 KNN 判别模型准确率。

4.3 多种判别模型的结果对比

为进一步评估AA-KNN判别模型的分类性能,采

用同为串行集成的XGBoost分类判别模型、并行集成的RF分类判别模型对169组拉曼光谱样本进行老化状态判别。

分别引入Kappa指数、平均判别准确率、准确率标 准差作为评价指标对4种判别模型预测结果进行对比 分析,如表2所示。

	表 2	判别模型效果对比图
Table 2	Comparis	son chart of discriminant model effect

	1		
Discriminant model	Average discriminant accuracy	Kappa index	Accuracy standard deviation / 10^{-2}
AA-KNN	0.9832	0.9780	2.11
KNN	0.9136	0.8896	3.20
RF	0.8934	0.8788	3.01
XGBoost	0.9013	0.8911	4.21

由表2可知,AA-KNN模型判别平均准确率在4种判别模型中最高(98.32%);Kappa指数可用于衡量多类别判别结果的可信度及表达分类时各个类别样本的平衡性,由4种判别模型的Kappa指数可知,AA-KNN判别模型在可信度及平衡性上均高于其余3种判别模型;准确率标准差用于衡量判别模型的稳定性,由4种判别模型的准确率标准差可知,AA-KNN的标准差最低,这体现出AA-KNN具有更好的稳定性。

5 结 论

通过油纸绝缘热老化实验获取169组不同老化阶段的油纸绝缘老化样本,并进行绝缘纸聚合度检测与 老化状态类别划分,使用拉曼光谱检测平台对绝缘油 样本进行检测。

将样本按照7:3划分训练集与测试集,分别采用 AA-KNN、KNN、RF、XGBoost构建针对原始高维拉 曼光谱样本的判别模型,通过调整随机数种子对样本 进行重复性判别,引入多种评价指标对重复性判别结

研究论文

果进行分析。结果显示,AA-KNN的平均判别准确率 达到 98.32%,高于 KNN 的 91.36%、RF 的 89.34%、 XGBoost 的 90.13%,AA-KNN 判别模型在平均判别 准确率以及 Kappa 指数上均高于其余 3 种判别模型, 而 AA-KNN 的准确率标准差低于其余 3 种判别模型。 这表明 AA-KNN 判别模型具有更高的准确率、更好的 稳定性、更大的可信度。

因此,AA-KNN判别模型可直接对高维原始绝缘 油拉曼光谱样本进行状态判别,无须通过特征提取减少 数据维度,在简化油纸绝缘拉曼光谱样本老化状态判别 过程的同时,保证判别结果的准确率、稳定性与可信度, 有利于变压器油纸绝缘老化评估技术的发展,为变压器 在线老化状态监测奠定基础。

参考文献

- [1] Yang D K, Chen W G, Wan F, et al. Identification of the aging stage of transformer oil-paper insulation via Raman spectroscopic characteristics[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2020, 27(6): 1770-1777.
- [2] Zhang M Z, Lei S J, Liu H Q, et al. Research on nonlinear characteristics for frequency domain dielectric response of transformer oil-paper insulation[J]. Measurement, 2022, 204: 112103.
- [3] 陈新岗,冯煜轩,李昌鑫,等.基于多参量的变压器油 中特征气体拉曼光谱分析[J].光谱学与光谱分析, 2020,40(6):1916-1922.

Chen X G, Feng Y X, Li C X, et al. Quantitative analysis of dissolved gases in transformer oil based on multi-parameter[J]. Spectroscopy and Spectral Analysis, 2020, 40(6): 1916-1922.

- [4] Feng D W, Yang L J, Zhou L W, et al. Effect of oilpaper-pressboard mass ratio on furfural content in transformer oil[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2019, 26(4): 1308-1315.
- [5] 余银辉,朱文江,吴苏敏,等.基于CNTs-FAgNPs基底的油中溶解糠醛SERS原位检测研究[J].光子学报,2022,51(9):0930001.
 Yu Y H, Zhu W J, Wu S M, et al. Study on SERS *in situ* detection of dissolved furfural in oil based on CNTs-

FAgNPs substrate[J]. Acta Photonica Sinica, 2022, 51 (9): 0930001.

 [6] 邓先钦,彭伟, 厉敏宪,等.应用氦离子化气相色谱法 测定变压器油中溶解气体[J].中国电力,2017,50(10): 148-152.

Deng X Q, Peng W, Li M X, et al. Determination of dissolved gases in transformer oil with gas chromatographic helium ionization detector[J]. Electric Power, 2017, 50(10): 148-152.

 [7] 刘金阳,于明鑫,冀胜楠,等.基于深度学习的舌鳞癌 拉曼光谱分割方法[J].激光与光电子学进展,2023,60
 (2):0230001.

第 60 卷第 21 期/2023 年 11 月/激光与光电子学进展

Liu J Y, Yu M X, Ji S N, et al. Raman spectral segmentation method for tongue squamous cell carcinoma using deep learning[J]. Laser & Optoelectronics Progress, 2023, 60(2): 0230001.

[8] 杨勇,董浩,桑瑶烁,等.基于密集连接网络模型的致病菌拉曼光谱分类[J].激光与光电子学进展,2023,60 (1):0130003.

Yang Y, Dong H, Sang Y S, et al. Raman spectrum classification of pathogenic bacteria based on dense connection network model[J]. Laser & Optoelectronics Progress, 2023, 60(1): 0130003.

[9] 杨勇,董浩,王澍,等.正电纳米银与卷积神经网络用 于4种食源性致病菌的SERS鉴定[J].中国激光,2022, 49(15):1507405.

Yang Y, Dong H, Wang S, et al. SERS identification of four food-borne pathogens by positively charged nanosilver and convolutional neural network[J]. Chinese Journal of Lasers, 2022, 49(15): 1507405.

- [10] Yang D K, Chen W G, Zhou W R, et al. Recognition of aging stage of oil-paper insulation based on surface enhanced Raman scattering and kernel entropy component analysis[J]. IEEE Access, 2019, 7: 127862-127873.
- [11] 杨定坤,陈伟根,万福,等.基于绝缘油拉曼光谱的聚合度定量评估方法[J].中国电机工程学报,2021,41 (13):4710-4720.
 Yang D K, Chen W G, Wan F, et al. Quantitative

evaluation method of polymerization degree based on Raman spectrum of insulating oil[J]. Proceedings of the CSEE, 2021, 41(13): 4710-4720.

- [12] 范舟,陈伟根,万福,等.基于小波包能量熵和Fisher 判别的油纸绝缘老化拉曼光谱诊断[J].光谱学与光谱分析,2018,38(10):3117-3123.
 Fan Z, Chen W G, Wan F, et al. Raman spectroscopy diagnosis of oil-paper insulation ageing based on wavelet packet energy entropy and fisher discrimination[J].
 Spectroscopy and Spectral Analysis, 2018, 38(10): 3117-3123.
- [13] 邹经鑫,刘彦琴,袁明哲,等.油纸绝缘老化状态判别的拉曼光谱特征研究[J].光谱学与光谱分析,2021,41
 (10):3159-3165.

Zou J X, Liu Y Q, Yuan M Z, et al. Study on the Raman spectral characteristics in ageing condition discrimination of oil-paper insulation[J]. Spectroscopy and Spectral Analysis, 2021, 41(10): 3159-3165.

- [14] Chen X G, Chen S T, Yang D K, et al. Study on the evaluation method of oil-paper insulation aging in transformer based on high dimensional Raman spectral data[J]. Spectroscopy and Spectral Analysis, 2021, 41 (5): 1463-1469.
- [15] Song R M, Chen W G, Yang D K, et al. Aging assessment of oil-paper insulation based on visional recognition of the dimensional expanded Raman spectra[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 6007110.

研究论文

第 60 卷第 21 期/2023 年 11 月/激光与光电子学进展

- [16] 国家能源局.油浸式变压器绝缘老化判断导则:DL/T 984—2018[S].北京:中国电力出版社,2018.
 National Energy Bureau of the People's Republic of China. Guide for the diagnosis of insulation aging in oilimmersed power transformer: DL/T 984—2018[S].
 Beijing: China Electric Power Press, 2018.
- [17] 陈新岗, 崔炜康, 朱莹月, 等. 适用于不同油纸比的改

进 KNN 变压器拉曼光谱老化评估模型[J]. 重庆理工大 学学报(自然科学), 2021, 35(7): 176-182.

Chen X G, Cui W K, Zhu Y Y, et al. A transformer Raman spectrum aging evaluation model based on improved KNN algorithm for different transformer's oilpaper ratio[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(7): 176-182.