

基于拉曼光谱数据处理和谱峰识别的变压器油绝缘老化研究

刘庆珍¹, 张溢¹, 鄢仁武²

(1. 福州大学电气工程与自动化学院, 福建省新能源发电与电能变换重点实验室, 福建 福州 350108;

2. 福建理工大学电子电气与物理学院, 福建 福州 350118)

摘要: 针对变压器油的拉曼光谱分析通常受到噪声和荧光背景等的干扰以及谱峰位置难以识别的问题, 提出了一种改进的数据处理和谱峰识别算法, 用于变压器油老化评估时的拉曼光谱分析。提出一种自适应 Savitzky-Golay 滤波法, 引入自适应窗口大小拉曼光谱数据进行去噪处理。采用改进的多项式拟合算法对去噪后的数据进行去除荧光背景处理, 减小荧光背景对拟合结果的影响。通过数据点与期望的拉曼信号的接近程度为每个数据点赋予权重, 以实现更准确的去荧光背景处理。利用谱峰识别技术判别变压器油的老化程度, 采用大小两种尺度高斯窗口判别法来识别谱峰, 并结合局部加权信噪比(local weighted signal-to-noise ratio, LW_SNR)来判断疑似拉曼谱峰的真实性。最后通过实验验证了所提算法在变压器油老化评估中的有效性。

关键词: 去噪; 荧光背景; 谱峰识别; 局部窗口加权信噪比; 变压器油老化评估

Transformer oil insulation aging based on Raman spectral data processing and peak identification

LIU Qingzhen¹, ZHANG Yi¹, YAN Renwu²

(1. Fujian Key Laboratory of New Energy Generation and Power Conversion, College of Electrical Engineering and

Automation, Fuzhou University, Fuzhou 350108, China; 2. School of Electronic, Electrical Engineering

and Physics, Fujian University of Technology, Fuzhou 350118, China)

Abstract: There are problems in that the Raman analysis of transformer oil is usually interfered with by noise and fluorescent background, and it is difficult to identify the position of the spectral peak. Thus this paper proposes an improved data processing and spectral peak recognition algorithm for the Raman analysis of transformer oil aging evaluation. An adaptive Savitzky-Golay filtering method is proposed, and adaptive window-size Raman spectral data is introduced for denoising. An improved polynomial fitting algorithm is used to remove the fluorescence background processing of the de-noised data to reduce its influence on the fitting results. Each data point is weighted according to the distance between the data point and the expected Raman signal, so as to achieve more accurate de-fluorescence background processing. The aging degree of transformer oil is identified by spectral peak recognition technology, and the spectral peak is identified by the Gaussian window discrimination method with two scales, and the authenticity of the suspected Raman spectral peak is judged by the local weighted signal-to-noise ratio (LW_SNR). Finally, the effectiveness of the proposed algorithm in transformer oil aging evaluation is proved by experiment.

This work is supported by the National Natural Science Foundation of China (No. 51807030).

Key words: denoising; fluorescence background; spectral peak identification; local weighted signal-to-noise ratio; transformer oil aging evaluation

0 引言

变压器油是电力系统中至关重要的电气设备, 其绝缘性能与状态对系统的安全运行有至关重要的

影响, 因此, 变压器的绝缘检测与老化评估一直是一项重要的研究课题^[1-3]。在变压器绝缘老化的诊断评估研究领域, 油色谱分析和介电响应分析是两种比较经典和成熟的诊断方法, 取得了很多理论成果^[4-8], 部分成果在生产中得到了广泛应用。

近年来, 拉曼光谱技术在变压器油老化评估中

基金项目: 国家自然科学基金项目资助(51807030)

引起了广泛关注。拉曼光谱具有非侵入性、快速、灵敏和特异性的优点, 可以提供关于样品化学成分和分子结构的丰富信息^[5]。同时, 拉曼光谱中蕴含着丰富的绝缘老化特征信息, 从光谱中提取绝缘老化特征量对老化状态直接进行诊断的研究成果积累不多, 但已经逐渐引起重视。文献[6]基于小波包能量熵和 Fisher 判别的方法, 提取了拉曼光谱中的特征信息, 对油纸绝缘老化拉曼光谱进行诊断。文献[7]基于拉曼光谱的相似度比对, 实现了油纸绝缘老化阶段的诊断, 此方法通过比对样本的拉曼光谱与已知老化阶段的样本进行相似性分析。但该方法对于油纸绝缘老化评估存在主观性太强的弊端。文献[8]结合机器学习的方法, 利用高维数据处理技术, 可以深入挖掘拉曼光谱中的潜在特征, 为绝缘老化评估提供了新的视角和方法。但是, 采用算法进行参数求解, 实质上是一种数学寻优过程, 存在一定的局限性, 这可能会导致部分参数在优化过程中出现发散、收敛缓慢的问题。从现有的研究成果看, 提取光谱特征量对绝缘油老化的评估研究还处在初期阶段, 需要进一步深入研究探索。

由于仪器和环境因素的影响, 变压器油的拉曼光谱时常受激光噪声、电荷耦合器件(charge-coupled device, CCD)噪声等造成的毛刺状的噪声信号和荧光背景等的影响, 导致信号质量下降和谱峰模糊不清^[9], 使老化特征的识别和分析变得更加困难, 这是采用拉曼光谱进行绝缘老化评估首先要处理的问题。文献[10]基于三点循环快速傅里叶变换中值的平滑去噪方法, 尽可能地保留了油纸绝缘老化的拉曼光谱信息, 用于后续老化诊断。文献[11]运用小波模极大值法对特定气体的拉曼光谱检测信号进行了去噪处理。

对光谱曲线进行去噪时, 需要注意去噪后的波形将产生一定程度的失真, 如何尽可能地减少失真是多数学者需要研究的问题, 同时应注意不同阶段的油样所对应的窗口数量是不同、多项式数量的要求也比较严格等问题。

拉曼光谱的谱峰识别是绝缘老化评估研究的重要内容。针对各类化学图谱的谱峰识别问题, 有几种方法已经获得认可, 如: k 邻近距离法、模仿人眼识别色谱谱峰法、导数法、改进小波变换法^[12]、离子迁移谱-飞行时间质谱(ion mobility spectrometry-time of flight mass spectrometry, IMS-TOFMS)自适应算法寻峰^[13]等。虽然有些方法并不是用于拉曼光谱的谱峰识别, 但是其方法原理及优点都值得借鉴。

本文基于拉曼光谱分析获取光谱中的绝缘老化

特征量并实现变压器油的绝缘老化诊断。通过对原始数据进行去噪处理和去除荧光背景处理, 获取纯净的光谱数据, 引入一种改进的谱峰识别方法, 准确评估变压器油的老化程度。

本文首先利用改进 Savitzky-Golay 滤波法对加热过程中获得的拉曼光谱数据进行去噪处理。在传统的 Savitzky-Golay 滤波法基础上, 将窗口大小依据频率自适应调整, 提高去噪效果。然后, 利用加权多项式拟合算法对去噪后的数据进行去除荧光背景处理。与传统的多项式拟合算法不同, 该算法提出使用最小二乘法为每个数据点分配不同的权重, 以减小荧光背景对拟合结果的影响。最后, 采用两种大小尺度高斯窗口判别法来进行谱峰识别, 实现变压器油老化程度的判别。参考高斯峰型, 与被检测光谱进行滑动相关运算, 得到两个尺度下的相关系数。将相关系数大于某一阈值的局部极大值处定义为疑似拉曼谱峰^[14], 并结合局部加权信噪比(local weighted signal-to-noise ratio, LW_SNR)来判断该处是否为真正的拉曼谱峰。

1 拉曼光谱数据去噪处理

变压器油的拉曼光谱检测是评估变压器油老化程度的重要手段。拉曼光谱是一种散射光谱, 其基本原理是: 激发光频率的光子入射到介质内部, 光子与介质相互作用后发生能量交换, 并发生弹性散射和非弹性散射^[15]。

然而, 仪器噪声、环境噪声、样品制备噪声干扰等影响会造成信号质量下降和谱峰模糊不清。因此, 在进行谱峰识别和老化评估之前, 必须对原始拉曼光谱进行有效的去噪处理。

1.1 数据来源

本文研究对象选取的是加热周期为 17 天的变压器油, 加热箱的温度设置为 120℃, 并通过 RA200 手持式拉曼光谱仪器, 检测出拉曼光谱数据。随着油样加热时间的增加、老化程度的加深, 波峰处的特征也在相对减弱, 为了能更加清晰地观察到波峰, 对所有数据进行归一化处理, 得到的曲线如图 1 所示。由图 1 曲线可以发现, 变压器油的拉曼光谱有两个比较明显的波峰, 分别处于 1442 cm^{-1} 与 2866 cm^{-1} 附近。可以通过波峰的变化初步判别油样的老化情况。同时, 各组波谱均具有明显的毛刺, 影响对图谱的精确分析。

1.2 自适应 Savitzky-Golay 滤波

传统的 Savitzky-Golay 滤波法是通过数据点

周围的多项式拟合来平滑信号,从而减少噪声的影响^[16]。此滤波法固定了窗口大小,但是,不同的光谱区域或者具备不同声波水平的光谱数据,需要有不同的窗口大小来适应其特征变化,这就可能导致在不同频率范围内信号的平滑程度不一致,从而影响谱峰的清晰度和峰值强度的识别准确性。为了改善传统 Savitzky-Golay 滤波法的不足,本文提出了一种改进 Savitzky-Golay 滤波方法,采用自适应窗口大小对变压器油拉曼光谱进行去噪处理。改进的 Savitzky-Golay 滤波法根据信号频率的变化自动调整窗口大小,以确保在不同频率范围内信号平滑程度的一致性。该方法通过在每个数据点周围进行多项式拟合来平滑信号,其特点是窗口的大小可以自适应调整。根据信号的频率变化,窗口大小会自动增大或缩小,以适应不同频率范围内信号的特征。

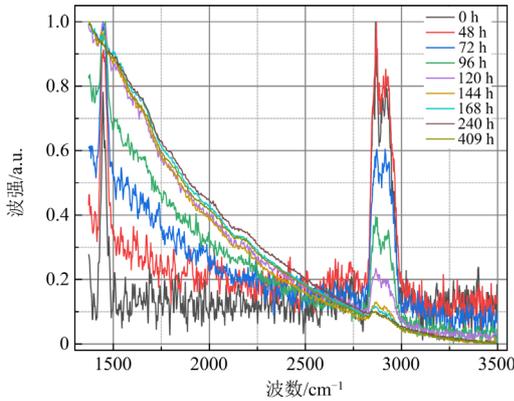


图 1 归一化后的原始拉曼光谱曲线

Fig. 1 Original Raman spectral curve after normalization

自适应 Savitzky-Golay 滤波法的详细步骤如下。

1) 对于每个数据点,确定一个窗口,该窗口包含该数据点及其周围的数据点。

2) 计算窗口内每个数据点相对于当前数据点的权重,计算公式为

$$\theta = \frac{1}{1 + (S'(i) - S(i))} \quad (1)$$

式中: $S'(i)$ 为窗口内数据点的波强; $S(i)$ 为当前数据点的波强。窗口内的数据点的波强与当前数据点的波强差异越小,即它们越接近,权重值就越大,表示该数据点在滤波过程中的重要性较高。相反,差异越大,权重值就越小,表示该数据点在滤波过程中的贡献较小。

3) 对窗口内的数据点应用 Savitzky-Golay 滤波方法来平滑数据。

4) 根据权重对滤波后的数据点进行加权平均,得到当前数据点的滤波结果。

5) 重复以上步骤,对所有数据点进行滤波处理。

改进的 Savitzky-Golay 滤波方法中的自适应窗口大小是基于对局部噪声水平的估计来确定的。自适应窗口大小的计算公式为

$$N = C(x) \quad (2)$$

$$x = Y\sigma \quad (3)$$

式中: N 表示计算出的自适应窗口大小;函数 $C(x)$ 表示将结果四舍五入为最接近的整数,以确保窗口大小为整数值; Y 表示原始固定窗口大小; σ 表示局部噪声水平的估计。

1.3 拟合结果对比

对加热了 0~409 h 的变压器油样分别使用传统的 Savitzky-Golay 滤波法和自适应 Savitzky-Golay 滤波法进行去噪处理。以均方根误差(root mean square error, RMSE)值 e_{RMSE} 作为最终的评判标准。RMSE 反映了每个数据点的预测值与观测值之间的差异,可用来衡量方法的数据处理性能。

$$e_{\text{RMSE}} = \frac{1}{N_T} \sqrt{\sum_{n=1}^{N_T} (y_a - y_b)^2} \quad (4)$$

式中: N_T 为波数; y_a 为第 n 个样本原始数据值; y_b 为第 n 个样本预测数据值。

以加热 48 h 的拉曼光谱曲线为例,使用自适应 Savitzky-Golay 滤波法进行去噪后的数据曲线如图 2 所示,红色为加热 48 h 后的原始拉曼光谱曲线,黑色为经过自适应 Savitzky-Golay 滤波法去噪后的拉曼光谱曲线,绿色为原始曲线中的噪声曲线。通过图 2 中局部放大图可以看出,去噪后的曲线中毛刺状的噪声信号被消除,并且曲线呈现出更加平滑的状态,更有利于谱峰的识别。

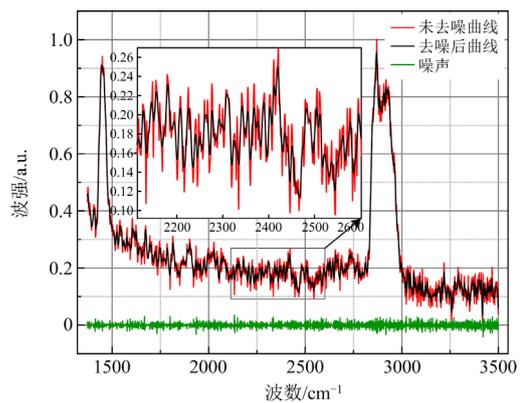


图 2 自适应 Savitzky-Golay 滤波法去噪后的数据曲线

Fig. 2 Data curve after denoising by adaptive Savitzky-Golay filtering method

各个时段的拉曼光谱曲线,经自适应 Savitzky-Golay 滤波法去噪后的 RMSE 值,如表 1 所示。

表 1 自适应 Savitzky-Golay 滤波法去噪后的 RMSE 值
Table 1 RMSE value after noise removal by adaptive Savitzky-Golay filtering method

加热 时长/h	RMSE 值/a.u.	
	传统 Savitzky-Golay 滤波	自适应 Savitzky-Golay 滤波
0	3.7540	2.8050
48	2.9706	2.3243
72	2.6796	1.3157
96	3.1048	1.4013
120	3.3206	1.5222
144	3.4062	1.5401
168	3.8611	1.6431
240	4.4467	1.9074
409	5.5570	2.3687

去除噪声的效果越好, RMSE 的值越小。由表 1 可以看出, 经改进的 Savitzky-Golay 滤波后, RMSE 值保持在 1~3 之间, 而传统方法滤波后的 RMSE 值在 2~6 之间。这说明加入自适应窗口的方法后, 不同阶段的油样所对应的窗口的数量是不一样的问题得到解决; 对于持续加热的变压器油样的拉曼光谱曲线, 改进后的方法能够有效提升其去噪能力。

2 拉曼光谱去荧光背景处理

变压器油的拉曼光谱技术中, 荧光背景的干扰是影响图谱特征识别精度的重要因素。因此, 在进行谱峰识别和老化程度评估之前, 必须对拉曼光谱数据进行荧光背景去除处理^[17]。

2.1 多项式拟合算法

传统的多项式拟合算法是一种常用的拉曼光谱的荧光背景处理方法。该方法通过拟合多项式函数来对背景信号进行建模和去除。然而, 传统的多项式拟合算法存在一些缺陷^[18], 主要体现在以下几个方面。

1) 传统算法没有考虑到不同数据点对拟合结果的贡献度不同。它在拟合过程中将每个数据点视为同等重要, 忽略了荧光背景与期望的拉曼信号之间的接近程度^[19]。在拟合过程中, 荧光背景信号的影响可能会被过度估计或低估, 从而导致拟合结果失真。

2) 传统算法未能充分考虑到荧光背景信号的变化性^[20]。传统的多项式拟合算法通常选择一个固定的多项式函数来对背景信号进行建模, 其公式为

$$P(\nu) = a_0 + a_1\nu + a_2\nu^2 + \dots + a_n\nu^n \quad (5)$$

式中: ν 表示波数, 代表波传播方向上单位长度内的波周数; a_0, a_1, \dots, a_n 为多项式系数; n 为多项式的阶数。当 RMSE 最小时, 可以通过拟合原始拉曼

光谱 $R(\nu)$ 和多项式函数 $P(\nu)$, 得到最优的多项式系数 a_0, a_1, \dots, a_n 等, 便可实现多项式拟合。去除背景后的拉曼光谱为 $R'(\nu)$, 计算公式为

$$R'(\nu) = R(\nu) - P(\nu) \quad (6)$$

然而, 拉曼光谱中的荧光背景通常具有复杂的特征, 可能在不同波数的位置和不同样品之间表现出明显的变化。传统的多项式拟合算法无法捕捉到这种变化性, 因而可能无法准确地估计和去除荧光背景信号。

3) 传统算法中所选取的多项式阶数也会对拟合结果产生一定影响。选择过低的阶数可能无法有效地拟合复杂的荧光背景信号, 而选择过高的阶数可能会导致过度拟合, 将背景信号误认为拉曼信号的一部分。

综上所述, 传统的多项式拟合算法在拉曼光谱荧光背景去除处理中存在一些局限性。

2.2 改进多项式拟合算法

为了克服传统多项式拟合算法在荧光背景去除处理中存在的不足, 本文提出改进的多项式拟合算法。该算法通过考虑权重分配的方式, 对变压器油拉曼光谱进行荧光背景去除处理, 以提高拟合结果的准确性和可靠性。在改进的多项式拟合算法中, 为每个数据点分配不同的权重是关键。权重的分配是基于数据点与期望的拉曼信号之间的相似程度, 即数据点与理想的拉曼信号之间的差异。根据差异的大小, 赋予每个数据点一个适当的权重, 使其对拟合结果的贡献度不同^[21]。

进行加权多项式拟合时, 本文使用了加权最小二乘法来分配每个数据点的权重。加权最小二乘法的目标是最小化加权残差的平方和, 其中每个数据点的权重表示其对拟合结果的贡献程度^[22]。

假设有 n 个数据点, 表示为 (x_i, y_i) , 其中 $i = 1, 2, 3, \dots, n$ 。拟合一个阶数为 m 的多项式函数, 如式(7)所示。

$$f(x) = c_0 + c_1x + c_2x^2 + \dots + c_mx^m \quad (7)$$

式中, c_0, c_1, \dots, c_m 为加权多项式系数, 该系数决定多项式函数的形状、斜率和曲率。

使用加权最小二乘法, 将每个数据点的加权残差的平方和最小化, 即最小化目标函数式(8)。

$$F(c_0, c_1, \dots, c_m) = \sum [w_i(y_i - f(x_i))^2] \quad (8)$$

式中: w_i 为第 i 个数据点的权重, 表示该数据点的重要性或接近程度^[23]; y_i 表示第 i 个数据点的实际观测值。

对于加权最小二乘法, 使用线性代数的方法来求解最佳的多项式系数。通过构建一个带有权重的

设计矩阵和响应向量,可以得到式(9)所示的线性方程组。

$$A^T W A c = A^T W y \quad (9)$$

式中: A 为设计矩阵,其第 i 行元素表示第 i 个数据点对应的多项式函数的各个次幂的值; W 为一个对角矩阵,其对角线上的元素是权重向量, $W = [w_1 \ w_2 \ \dots \ w_n]$; c 为多项式的系数向量; y 为响应向量,即数据点的观测值。

通过解这个线性方程组,可以得到最佳的多项式系数 c_0, c_1, \dots, c_n , 从而实现加权多项式拟合。

2.3 拟合结果

以加热 48 h 与 192 h 油样的拉曼光谱曲线为例,使用加权多项式拟合算法去除荧光背景后的曲线如图 3 和图 4 所示。其中蓝色曲线为加热油样原始拉曼光谱,橙色曲线为荧光背景,绿色曲线为去除荧光背景后的拉曼光谱。

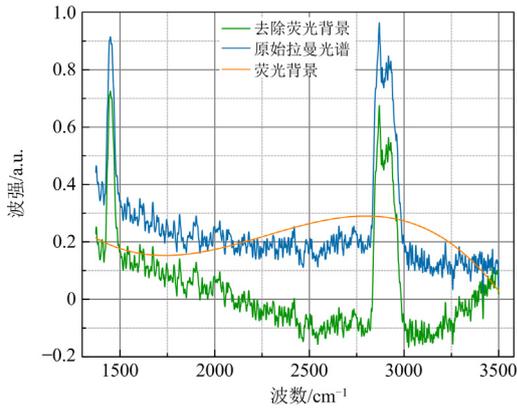


图 3 加热 48 h 的加权多项式拟合算法去除荧光背景结果
Fig. 3 Results of fluorescence background removal by weighted polynomial fitting method after heating for 48 h

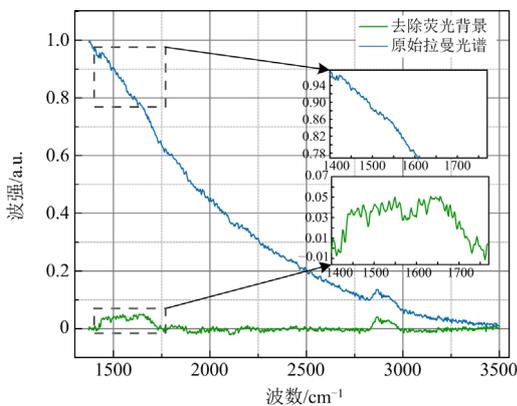


图 4 加热 192 h 的加权多项式拟合算法去除荧光背景结果
Fig. 4 Results of fluorescence background removal by weighted polynomial fitting method after heating for 192 h

图 4 结果体现了改进的多项式拟合算法在谱峰保留方面的显著成果。由于油样加热 48 h 时仍然处于老化前期,波峰特征明显,故去除荧光背景在图中效果表现的不明显,然而当加热到 192 h 之后,油样已进入老化中期,波数在 1442 cm^{-1} 附近的特征峰已经十分微弱,此时对数据进行去除荧光背景的效果相对明显。且随着加热时长的增加, $1400\sim 1750 \text{ cm}^{-1}$ 波数范围内的特征峰强度已经开始超过第一个特征峰的强度。

荧光背景去除后谱峰更加清晰可辨,荧光背景对于谱峰的干扰程度减小。这对于变压器油老化程度的判别十分关键,因为谱峰的位置和强度变化能够提供有关油品质量和老化程度的重要信息。

3 拉曼光谱谱峰识别与老化判别

3.1 双尺度高斯窗口判别法原理

由于拉曼光谱的特性以及变压器油中成分多样的缘故,单尺度参考峰形的形状和大小是固定的。这意味着对光谱数据的分析会集中在特定尺度或特征大小上。这种方法通常适用于已知谱峰大小的情况。针对这个问题,本文采用大小双尺度的相关算法来分析变压器油的拉曼光谱中的特征峰值^[14]。该方法可以全面地分析光谱数据,捕获不同宽度的特征,并具有更高的分析灵活性。这种方法对于处理具有多个不同尺寸峰值或未知尺寸的数据更加有效。

双尺度图谱分析法是基于高斯峰型进行滑动相关运算^[12],它通过计算相关系数来确定潜在的拉曼谱峰。滑动相关运算是一种用于测量信号之间相似度的方法,它通过在不同位置上滑动参考峰型,计算其与每个位置处的信号的相关系数。对于每个滑动位置,通过计算得到一个相关系数,该系数反映了信号与参考峰型之间的相似程度^[24]。较大的相关系数表示信号在该位置上可能存在拉曼谱峰。

为了确定真实的拉曼谱峰,选择一个适当的相关系数阈值。当相关系数超过该阈值时,将其识别为具有较高相关性的局部极大值,作为疑似拉曼谱峰^[25]。

利用双尺度的谱峰识别方法,能够准确地提取出变压器油拉曼光谱中的特征峰值。

相关系数采用式(10)计算。

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (10)$$

式中: x_i 和 y_i 分别表示原始拉曼光谱和高斯峰型

在滑动位置 i 处的强度值; \bar{x} 和 \bar{y} 表示对应信号的均值。

通过计算相关系数,能够确定潜在的拉曼谱峰,并进一步分析其强度和位置等特征,以评估变压器油的老化程度和化学变化情况。

基于高斯拟合峰形的大小尺度谱峰识别方法提高了变压器油拉曼光谱分析的准确性和可靠性,为变压器油的老化程度判别提供了重要的依据。高斯拟合峰形表达式为

$$f(t) = \sum_{i=1}^N f_i(t) = \sum_{i=1}^N \frac{A_i}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(t-u_i)^2}{2\sigma_i^2}\right] \quad (11)$$

式中: A_i 为峰的振幅; u_i 为平移参数; σ_i 为标准差。

3.2 局部加权信噪比

尽管通过双尺度谱峰识别方法可以初步确定谱峰,但仅仅识别疑似谱峰并不能确保其真实性。为了进一步评估谱峰的可靠性,引入 LW_SNR, 加权指的是使用高斯窗口函数对局部区域内的信号强度进行加权处理。

具体计算步骤如下。

1) 计算窗口加权信号。对于给定的位置 i , 将窗口函数应用于其周围的信号强度值, 根据窗口函数的权重分布对这些值进行加权。这意味着在使用高斯窗口函数计算局部加权信噪比时, 当前位置 i 的信号强度会被窗口函数在 i 附近的值所加权。窗口加权信号使用式(12)计算。

$$W(i) = \frac{\sum[w(n)S(n)]}{\sum w(n)} \quad (12)$$

式中: $W(i)$ 为位置 i 处的加权平均值; $w(n)$ 为高斯窗口函数在位置 n 处的值; $S(n)$ 为谱峰位于位置 n 处的光谱强度。

2) 计算信噪比。使用加权的信号强度值, 通过与噪声的比较计算局部高斯加权信噪比。通过设置合适的信噪比阈值作为最终评判标准, 判断 3.1 节中的疑似拉曼谱峰是否为真正的拉曼谱峰。

使用噪声的标准差作为归一化因子, 以将信号强度与噪声水平进行比较。加权信噪比计算公式为

$$L_{W_SNR} = \frac{2W(i) - W(i+m) - W(i-m)}{\sigma} \quad (13)$$

式中, m 为计算局部信噪比的高斯窗口半宽度。

LW_SNR 方法考虑了局部高斯窗口内信号与噪声的相对比较。传统 SNR 方法仅仅通过计算信号的能量与噪声的方差之比来评估信噪比的大小, 而没有考虑到信号和噪声在局部窗口内的分布情况。相比之下, LW_SNR 方法考虑了局部高斯窗口内信号和噪声的相对位置关系, 通过引入高斯窗口函数

的权重, 能够更加准确地估计谱峰的强度和噪声水平, 提高了对干扰的鲁棒性, 从而有助于凸显噪声较低的真实峰值, 使其更加明显, 容易被识别。具体来说, LW_SNR 方法利用高斯窗口函数在不同位置的权重来调整信号和噪声的相对贡献, 使得窗口内的谱峰信号更加突出, 同时抑制了窗口内的噪声干扰, 避免了伪峰的出现, 有效地降低了错误识别的风险, 从而提高了信号处理的精确性和可靠性。

通过结合大小尺度高斯窗口判别法和 LW_SNR 方法, 能够更准确地判别变压器油的老化程度, 有效地识别和验证谱峰的真实性, 可以提供关键的信息用于评估油品的质量和老化程度。

3.3 谱峰识别的结果分析

3.3.1 小波分析法

本文选用小波分析法作为本文研究的对比方法。小波分析法利用在时域和频域都具有紧支撑的小波对信号进行小波分解, 可获得信号局部的时频信息, 通过小波尺度的伸缩可以实现对信号的多分辨率分析^[26]。

由于高斯函数形式与拉曼谱峰十分相近, 其二阶导数后的谱峰位置必定为局部极大值。两种尺度下测出的同一谱峰局部极大值位置相差不多, 而小波系数从三维图上来看, 小波系数是个山脊的轮廓。将这些位置相近的局部极大值相连便形成了脊线, 由这些脊线组成的矩阵中可以识别出谱峰^[27]。但是脊线的长度和信噪比的关系目前并不明确, 由于这些参数是谱峰识别的关键参数, 所以会导致谱峰的识别准确率不高及峰位定位不准确。

3.3.2 双尺度高斯加权信噪比法的绝缘老化判别

在变压器油样的拉曼光谱谱峰识别中, 油样的加热时长是一个关键的特征量, 当加热时长增加时, 数据的局部信噪比会降低, 传统的谱峰识别技术或多或少都存在识别不准确的问题。故本文研究选取加热 48 h、409 h 作为变压器油样的两个阶段, 作为寻峰算法的示例样本。表 2 为通过式(13)所得出的不同加热时长所对应的信噪比阈值。

表 2 不同加热时长的油样信噪比阈值

Table 2 SNR thresholds of oil samples with different heating time

加热时长/h	0	48	96	120	168	240	409
信噪比阈值	8	6	6	6	5	5	4

大小双尺度高斯窗口判别结合局部加权信噪比的谱峰识别结果如图 5 和图 6 所示, 其中红色星号代表大尺度算法判别出的峰值点, 蓝色星号代表小尺度算法判别出的峰值点。由图 5 可以发现, 由

于加入了双尺度判别方法，峰值识别的准确率明显提高了。

由图 5 可知，在加热前期，谱峰的位置相对明显，可以看到本研究成功识别出 48 h 的拉曼光谱中 1442 cm⁻¹ 和 2866 cm⁻¹ 两处特征峰的 3 个峰位点。随着加热时长的增加，加热时长为 409 h 时，如图 6 所示，1442 cm⁻¹ 处的谱峰逐渐削弱，处于识别的最低限，本研究的寻峰方法，对比周围的噪声点，舍弃了噪声的峰值点，成功标记出了 1442 cm⁻¹ 处的峰值点，并且标记出 1400~1500 cm⁻¹ 波数范围内最明显的峰值位置。

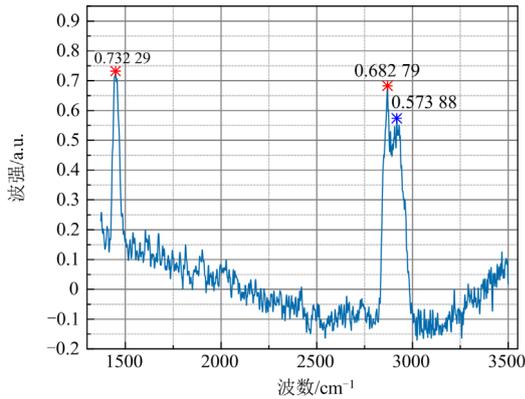


图 5 加热 48 h 的谱峰识别结果

Fig. 5 Results of spectrum peak identification after heating for 48 h

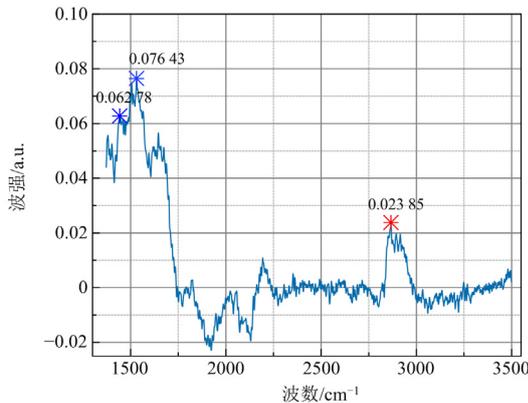


图 6 加热 409 h 的谱峰识别结果

Fig. 6 Results of spectrum peak identification after heating for 409 h

3.3.3 谱峰识别方法性能对比

本文采用前文预处理后的数据，对处理后的光谱分别使用小波分析法、双尺度窗口加权信噪比法进行寻峰操作，并统计两种算法在谱峰识别中的准确率，对比结果如图 7 所示。

由图 7 可知，随着加热时长的不断增加，数据

的峰值信噪比逐渐降低，故两种方法的谱峰识别准确率都有一定程度的降低。小波分析法在检测油样老化后期的峰值点时存在着误判、漏判等问题，这是由于加热 240 h 后的噪声已经接近或者覆盖两个特征峰，导致识别的难度逐渐变大。而本文提出方法的识别准确率都达到了 93%以上，比小波分析法在识别准确率上更加具有优势。

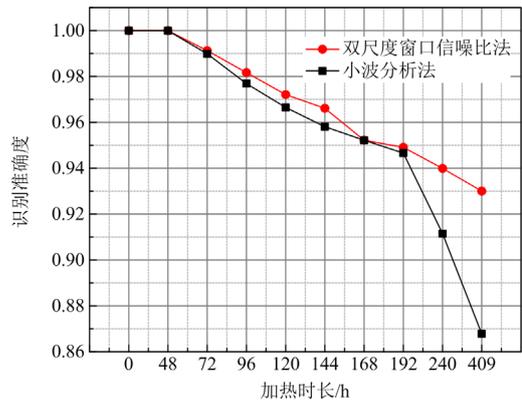


图 7 加热时长与峰值识别准确率的关系

Fig. 7 Relationship between heating time and peak recognition accuracy

4 结论

本研究的数据样本采用的是加热时长为 0~409 h 的变压器油样，并通过 RA200 手持式拉曼光谱仪测试其拉曼光谱曲线和数据。本文的主要研究结论如下。

1) 采用自适应 Savitzky-Golay 滤波法对数据进行去噪声处理，该方法可以根据信号频率的变化自动改变窗口大小，比传统的 Savitzky-Golay 滤波法更准确地识别出噪声。

2) 采用改进多项式拟合法对拉曼光谱进行去除荧光背景处理。利用加权最小二乘法，根据每个数据点中荧光背景与理想的拉曼光谱之间的差值来分配权重，准确去除荧光背景数据。

3) 提出双尺度高斯窗口判别法结合局部加权信噪比的方法来对拉曼光谱进行峰位识别。随着油样老化的加剧，数据的整体信噪比会降低，谱峰识别的难度增大^[28]，但是本方法的谱峰识别准确率仍然能够达到 93%以上。而小波分析法在油样老化后期的准确率只有 86.78%，说明大小尺度窗口加权信噪比法在变压器油样老化后期也能进行准确识别，是一种有效的拉曼谱峰识别方法。

综上所述，本文提出的拉曼光谱数据处理和谱峰识别方法在去噪、去除荧光背景和准确识别谱峰方面是有效的和可靠的，与小波分析法相比，在诊

断准确率上更加具有优势。本文方法可以有效评估变压器油的老化程度, 为变压器油老化监测提供了一种新颖的实用方法, 在电力系统的可靠性和安全性方面具有重要的应用前景。

参考文献

- [1] 刘庆珍, 黄昌硕. 基于 FCBF 特征选择和 XGBoost 原则的油纸绝缘介电响应特征量优选研究[J]. 电力系统保护与控制, 2022, 50(15): 50-59.
LIU Qingzhen, HUANG Changshuo. Optimization of dielectric response characteristics of oil paper insulation based on FCBF feature selection and the XGBoost principle[J]. Power System Protection and Control, 2022, 50(15): 50-59.
- [2] 邹阳, 俞豪奕, 金涛. 融合模糊 K 近邻及证据理论的变压器油纸绝缘状态评估方法[J]. 电力系统保护与控制, 2023, 51(14): 55-63.
ZOU Yang, YU Haoyi, JIN Tao. Evaluation method of the oil-paper insulation condition of a transformer based on fuzzy K nearest neighbor and evidence theory[J]. Power System Protection and Control, 2023, 51(14): 55-63.
- [3] 石宜金, 谭贵生, 赵波, 等. 基于模糊综合评估模型与信息融合的电力变压器状态评估方法[J]. 电力系统保护与控制, 2022, 50(21): 167-176.
SHI Yijin, TAN Guisheng, ZHAO Bo, et al. Condition assessment method for power transformers based on fuzzy comprehensive evaluation and information fusion[J]. Power System Protection and Control, 2022, 50(21): 167-176.
- [4] 刘焕, 薛守洪, 孙利强. 变压器油糠醛含量检测与固体绝缘老化的研究[J]. 内蒙古科技与经济, 2021(24): 86-87.
- [5] 张腾翼, 古亮, 陈新岗, 等. 基于拉曼光谱相似度比对的油纸绝缘老化阶段诊断[J]. 光散射学报, 2020, 32(3): 202-209.
ZHANG Tengyi, GU Liang, CHEN Xingang, et al. Aging phase diagnosis of oil paper insulation based on Raman spectral similarity ratio[J]. The Journal of Light Scattering, 2020, 32(3): 202-209.
- [6] 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 Zhou, CHEN Weigen, WAN Fu, 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.
- [7] 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 Dingkun, CHEN Weigen, WAN Fu, et al. Quantitative evaluation method of degree of polymerization based on Raman spectra of insulating oil[J]. Proceedings of the CSEE, 2021, 41(13): 4710-4720.
- [8] 陈新岗, 陈妹婷, 杨定坤, 等. 基于高维拉曼光谱数据的变压器油纸绝缘老化评估方法研究[J]. 光谱学与光谱分析, 2021, 41(5): 1463-1469.
CHEN Xingang, CHEN Shuting, YANG Dingkun, 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.
- [9] 郭永恒. 拉曼光谱分析算法研究与系统实现[D]. 上海: 华东师范大学, 2021.
GUO Yongheng. Raman spectrum analysis algorithm research and system realization[D]. Shanghai: East China Normal University, 2021.
- [10] 杨定坤. 油纸绝缘老化拉曼光谱多尺度特征提取及集成增强神经网络诊断研究[D]. 重庆: 重庆大学, 2022.
YANG Dingkun. Study on multi scale feature extraction and integrated enhanced neural network diagnosis of oil-paper insulation aging by Raman spectroscopy[D]. Chongqing: Chongqing University, 2021.
- [11] 万福, 杨曼琳, 贺鹏, 等. 变压器油中气体拉曼光谱检测及信号处理方法[J]. 仪器仪表学报, 2016, 37(11): 2482-2488.
WAN Fu, YANG Manlin, HE Peng, et al. Raman spectroscopy detection and signal processing method for the gases in transformer oil[J]. Chinese Journal of Scientific Instrument, 2016, 37(11): 2482-2488.
- [12] 吕罗庚, 史晋芳, 邱荣, 等. 基于尺度自适应的小波变换 LIBS 谱峰识别算法[J]. 制造业自动化, 2020, 42(11): 51-55.
LÜ Luogeng, SHI Jinfang, QIU Rong, et al. Peak detection algorithm for laser-induced breakdown spectroscopy based on scale adaptation of wavelet transform[J]. Manufacturing Automation, 2020, 42(11): 51-55.
- [13] 刘薇. IMS-TOFMS 自适应去噪与谱峰识别算法研究及分析软件实现[D]. 西安: 西北大学, 2022.
LIU Wei. Research on IMS-TOFMS adaptive denoising and spectral peak identification algorithm and implementation of analysis software[D]. Xi'an: Northwest University, 2022.
- [14] 姜承志, 孙强, 刘英, 等. 基于双尺度相关运算的拉曼谱峰识别方法[J]. 光谱学与光谱分析, 2014, 34(1): 103-107.
JIANG Chengzhi, SUN Qiang, LIU Ying, et al. A new peak detection algorithm of Raman spectra[J]. Spectroscopy

- and Spectral Analysis, 2014, 34(1): 103-107.
- [15] 李奇辰, 李民赞, 杨玮, 等. 拉曼光谱在精细农业土壤成分快速检测中的研究进展[J]. 农业工程学报, 2023, 39(7): 1-9.
LI Qichen, LI Minzan, YANG Wei, et al. Research progress on the rapid detection of soil components using Raman spectroscopy: a review[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(7): 1-9.
- [16] 雷林平. 基于 Savitzky-Golay 算法的曲线平滑去噪[J]. 电脑与信息技术, 2014, 22(5): 30-31.
LEI Linping. Curve smooth denoising based on Savitzky-Golay algorithm[J]. Computer and Information Technology, 2014, 22(5): 30-31.
- [17] 王海朋, 褚小立, 陈瀑, 等. 光谱基线校正算法研究与应用进展[J]. 分析化学, 2021, 49(8): 1270-1281.
WANG Haipeng, CHU Xiaoli, CHEN Pu, et al. Research and application progress of algorithms for spectral baseline correction[J]. Chinese Journal of Analytical Chemistry, 2021, 49(8): 1270-1281.
- [18] 司赶上, 刘家祥, 李振钢, 等. 基于形态学与多项式拟合的紫外拉曼荧光背景扣除算法[J]. 光学学报, 2022, 42(22): 200-208.
SI Ganshang, LIU Jiexiang, LI Zhengang, et al. Fluorescence background subtraction algorithm of UV Raman based on morphology and polynomial fitting[J]. Acta Optica Sinica, 2022, 42(22): 200-208.
- [19] 谢佳文. 基于 Ipro-Poly 算法的拉曼光谱去背景模型及系统设计[D]. 福州: 福建师范大学, 2020.
XIE Jiawen. Background model and system design of Raman spectroscopy based on Ipro-Poly algorithm[D]. Fuzhou: Fujian Normal University, 2020.
- [20] HU Haibing, BAI Jing, XIA Guo, et al. Improved baseline correction method based on polynomial fitting for Raman spectroscopy[J]. Photonic Sensors, 2018, 8(4): 332-340.
- [21] 王玲, 邓志, 马明, 等. 基于状态估计残差比较的配电网故障区段定位方法[J]. 电力系统保护与控制, 2021, 49(14): 132-139.
WANG Ling, DENG Zhi, MA Ming, et al. A method for locating fault sections in distribution networks based on the comparison of state estimation residual errors[J]. Power System Protection and Control, 2021, 49(14): 132-139.
- [22] YAN Yuguo, XU Tianhe, SUN Zhangzhen, et al. Middle- and long-term UT1-UTC prediction based on constrained polynomial curve fitting, weighted least squares and autoregressive combination model[J]. Remote Sensing, 2022, 14(14): 3252.
- [23] HABERSTICH C, NOUY A, PERRIN G. Boosted optimal weighted least-squares[J]. Mathematics of Computation, 2022, 91(335): 1281-1315.
- [24] 侯岩. 基于相关系数与局部信噪比的拉曼谱峰识别技术研究[D]. 成都: 电子科技大学, 2018.
HOU Yan. Research on Raman spectra recognition technology based on correlation coefficient and local SNR[D]. Chengdu: University of Electronic Science and Technology of China, 2018.
- [25] 姜承志. 拉曼光谱数据处理与定性分析技术研究[D]. 长春: 中国科学院研究生院(长春光学精密机械与物理研究所), 2014.
JIANG Chengzhi. Research on data processing and qualitative analysis of Raman spectrum[D]. Changchun: Graduate School of Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics), 2014.
- [26] 吴童, 孙抗, 师文文. 基于 AVMD-自适应小波包法的电缆局部放电去噪研究[J]. 电力系统保护与控制, 2020, 48(14): 95-103.
WU Tong, SUN Kang, SHI Wenwen. Research on cable partial discharge denoising based on an AVMD-adaptive wavelet packet method[J]. Power System Protection and Control, 2020, 48(14): 95-103.
- [27] DU P, KIBBE W A, LIN S M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching[J]. Bioinformatics, 2006, 22(17): 2059-2065.
- [28] 颜凡, 朱启兵, 黄敏, 等. 基于拉曼光谱的已知混合物组分定量分析方法[J]. 光谱学与光谱分析, 2020, 40(11): 3599-3605.
YAN Fan, ZHU Qibing, HUANG Min, et al. Quantitative analysis method for mixture with known components based on Raman spectroscopy[J]. Spectroscopy and Spectral Analysis, 2020, 40(11): 3599-3605.

收稿日期: 2023-08-01; 修回日期: 2023-11-26

作者简介:

刘庆珍(1971—), 女, 博士, 副教授, 研究方向为电力设备绝缘老化评估; E-mail: lqz515@126.com

张溢(1998—), 男, 通信作者, 硕士研究生, 研究方向为变压器绝缘老化评估。E-mail: 291997431@qq.com

(编辑 周金梅)