基于拉曼光谱技术的变压器绝缘油老化研究

周宇含　刘庆珍
（福州大学电气工程与自动化学院，福州　350108）

摘要　针对变压器绝缘油的实测拉曼光谱数据，以拉曼光谱技术为手段，以准确判别变压器绝缘老化状态为目的，开展基于拉曼光谱数据处理、特征提取和老化诊断的研究。首先，结合拉曼光谱噪声的变化规律，基于小波变换理论提出全局阈值小波变换滤波法，有效地去除拉曼光谱中的噪声信号。然后，基于自适应迭代重加权惩罚最小二乘法（airPLS）提出一种改进的基线校正方法，准确地去除拉曼光谱的荧光背景。其次，运用连续投影算法（SPA）提取拉曼光谱中的老化特征信息，并分析其与变压器绝缘油老化程度之间的关系。最后，基于轻量级梯度提升机（LightGBM）分类模型实现对变压器绝缘老化状态的准确判别，并以极限梯度提升（XGBoost）模型作为对照，比较二者的诊断精度。实验结果表明，轻量级梯度提升机模型的诊断精度具有明显优势，验证了所提取老化特征信息的有效性。

关键词：变压器绝缘油；拉曼光谱；噪声；荧光背景；特征提取；老化判别

Research on transformer insulating oil aging based on Raman spectra

ZHOU Yuhan　LIU Qingzhen
(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou　350108)

Abstract　With regard to the measured Raman spectral data of transformer insulating oil, the relevant research is carried out based on Raman spectral data processing, feature extraction and aging diagnosis with Raman spectral technology as a means to accurately identify the aging state of transformer insulation. First of all, combined with the law of change of Raman spectral noise, based on the wavelet transform theory, the global threshold wavelet transform filtering method is proposed, which effectively removes the noise signals in Raman spectra. Further, an improved baseline correction method is proposed based on the adaptive iterative reweighting penalized least squares (airPLS) method, which accurately removes the fluorescence background of Raman spectra. Secondly, the aging feature information in Raman spectra is extracted using successive projections algorithm (SPA), and its relationship with the aging degree of transformer insulating oil is analyzed. Finally, the light gradient boosting machine (LightGBM) classification model is used to realize accurate discrimination of transformer insulation aging state, and the extreme gradient boosting (XGBoost) model is used as a control group to compare the diagnostic accuracy of the two. The experimental results show that the LightGBM model possesses obvious advantages in diagnostic accuracy, and also further verifies the validity of the extracted aging feature information.

Keywords：transformer insulating oil; Raman spectra; noise; fluorescence background; feature extraction; determination of aging

0　引言

变压器作为电力系统的核心设备，其绝缘状态直接关系着电力系统的稳定运行[1-2]。变压器的绝缘系统会随投运年限增加而逐渐老化，可能导致设备故障并引发重大的电力事故。因此，准确判别变压器绝缘老化状态具有重要意义[3-4]。

传统的变压器绝缘老化诊断方法的检测对象有

国家自然科学基金资助项目（No. 51807030）

油中溶解气体、糠醛含量、微水含量和绝缘纸聚合度等[5]，这类指标可以直接、准确地反映绝缘系统的老化程度，然而复杂的检测步骤与较高的环境技术要求限制了这些方法在实际工程中的应用。因此，为了实现准确的变压器绝缘老化状态评估，探索新的检测诊断技术很有必要。

近年来，拉曼光谱检测技术因无损、简单、快速等优点而被广泛应用于物质检测。其基本原理是通过激光照射不同物质所产生的拉曼散射光的波长不同，且散射光的强度随被检测物质的浓度大小而变化[6]。相关研究[7-10]表明，拉曼光谱中蕴含丰富的绝缘油老化特征信息，可作为变压器绝缘老化评估的重要依据。然而，拉曼光谱信号在采集的过程中通常会受到噪声、荧光背景等因素的干扰，导致拉曼信号质量降低，甚至淹没特征信息，造成老化状态判别的准确率大幅下降。因此，基于拉曼光谱的变压器绝缘油老化研究必须针对光谱信号进行数据处理。文献[7]运用小波模极大值法和包络线迭代法有效地去除了拉曼光谱的噪声信号和荧光背景。文献[8]结合五点三次平滑法、迭代自适应加权惩罚最小二乘法和主成分分析对原始拉曼光谱进行预处理与特征提取，并采用线性回归分类和支持向量机实现老化诊断。文献[9]基于小波包能量熵提取特征信息，并结合 Fisher 判别法建立了老化诊断模型。文献[10]采用竞争自适应重加权算法从多种角度提取了拉曼光谱的特征信息，并基于小波包能量熵分析了特征信息与老化程度之间的关系。这些研究在一定程度上实现了拉曼光谱数据处理、特征提取与老化判别，但没有充分结合拉曼光谱干扰信号的特点，诊断模型的准确率也有待提高。因此，在拉曼光谱的数据处理与特征提取方面仍需要进一步的深入研究。

本文以拉曼光谱技术为手段，以准确判别变压器绝缘老化状态为目的，开展基于拉曼光谱数据处理、特征提取与老化判别的相关研究。首先，通过加速老化实验制备老化样本，获取变压器绝缘油老化全寿命周期的拉曼光谱数据。并且，基于小波变换理论提出全局阈值小波变换滤波法，以有效去除拉曼光谱中的噪声信号。以方均根误差与信噪比作为评价指标，比较全局阈值小波变换滤波法与 Savitzky-Golay 滤波法对拉曼光谱的去噪效果，进一步验证其去噪性能。同时，基于自适应迭代重加权惩罚最小二乘法（adaptive iterative reweighting penalized least squares method, airPLS）提出改进 airPLS 基线校正法，并将其用于去除拉曼光谱荧光背景。进一步地，运用连续投影算法（successive projections algorithm, SPA）提取拉曼光谱中的老化特征信息，分析特征信息与绝缘油老化程度之间的关系。最后，基于轻量级梯度提升机（light gradient boosting machine, LightGBM）分类模型实现对变压器绝缘老化状态的准确判别，并以极限梯度提升（extreme gradient boosting, XGBoost）模型作为对照组，比较二者的预测准确率，以验证 LightGBM 模型在诊断精度上的优越性。

1　拉曼散射原理与光谱数据获取

1.1　拉曼散射原理

拉曼散射又称拉曼效应，是光通过介质时与分子运动相互作用而引起光波频率变化的非弹性散射，具体原理如图 1 所示：发生拉曼散射的过程中，处于基态的散射分子吸收入射光子的能量跃迁至高能态或振动终态，从而出现一系列的散射现象，包括瑞利散射、斯托克斯拉曼散射和反斯托克斯拉曼散射[11]。

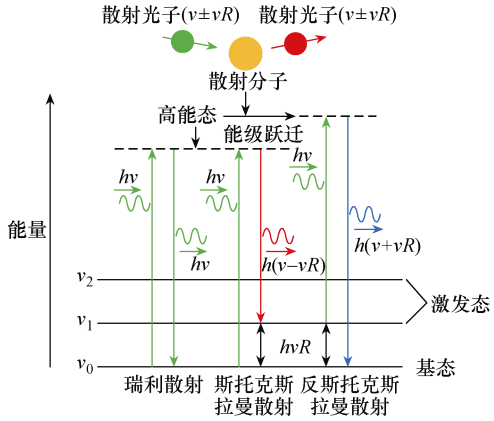

图 1　拉曼散射原理

1.2　拉曼光谱数据及其特征

本文以变压器绝缘油作为研究对象，对其进行加速老化实验，获得加热时长 0h、48h、72h、96h、120h、144h、168h、192h、240h、288h、307h、409h 共 12 组老化样本，采用 RA200 手持式拉曼光谱检测仪分别采集各老化样本对应的拉曼光谱信号。实验中，加热箱温度为 130℃，加热周期 17 天。为了消除不同条件下光谱强度差异较大对后续研究的影响，对采集到的拉曼光谱信号统一做归一化处理，使数据具备可比性。同时，为了使拉曼光谱信号特征更加明显、突出，选取波长 1 250～3 500cm^{-1} 的

拉曼光谱进行研究与分析。

归一化后的拉曼光谱曲线如图 2 所示。仔细观察拉曼光谱曲线可以发现，波长区间 1 400～1 500cm^{-1} 和 2 800～3 000cm^{-1} 内存在明显的高强度特征峰，波长 1 300cm^{-1} 附近也存在强度较低的特征峰。并且，注意到特征峰的高度随加热时长变化，由此推测其峰值强度与绝缘油老化程度之间可能存在关系。同时，从图 2 可以看出，拉曼光谱数据中存在大量毛刺状的噪声信号，导致无法确定各特征峰所处的具体波段。因此，需要对拉曼光谱信号做去噪处理，以消除噪声对后续研究的影响。

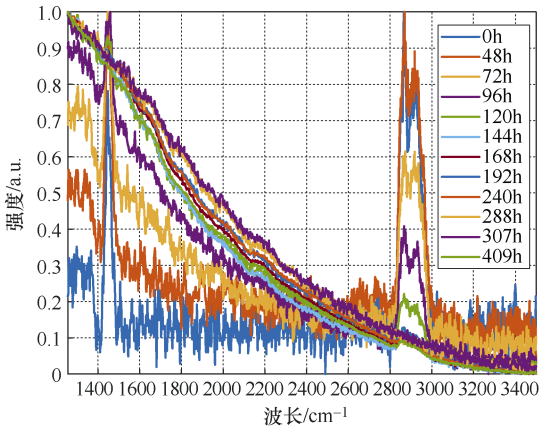

图 2 归一化后的拉曼光谱曲线

2 拉曼光谱噪声去除

拉曼散射信号在检测过程中会不可避免地受各种噪声信号的干扰，如高斯噪声、散粒噪声、环境噪声等。过多的噪声会导致实际拉曼散射信号被覆盖，造成特征信息难以识别。因此，为了消除噪声信号的干扰，对光谱信号进行去噪处理以提高其信噪比[12]。本文结合拉曼光谱噪声信号的变化规律，基于小波变换理论提出一种全局阈值小波变换滤波法，将该方法用于去除拉曼光谱中的噪声。

2.1 Savitzky-Golay 滤波法

Savitzky-Golay 滤波法是一种基于多项式拟合的滤波技术，其核心思想是在一定长度的窗口内对数据点进行高阶多项式拟合来平滑信号，目前被广泛应用于数据流平滑去噪[13]。

Savitzky-Golay 滤波的工作原理如下：

1）对每一个数据点，创建一个局部逼近的滑动窗口，并确定滑动窗口大小与多项式拟合阶数。

2）运用最小化二乘法对每个滑动窗口内的数据点进行多项式拟合，得到拟合曲线。

3）在窗口滑动的过程中，将拟合曲线在窗口中心的值作为平滑后的数据点，每个数据点至少参与一次多项式拟合，从而得到包含全部数据的拟合曲线。

运用 Savitzky-Golay 滤波法进行去噪，需要先确定滑动窗口大小与多项式拟合阶数。其中，滑动窗口大小确定平滑去噪的区域，多项式拟合阶数确定拟合曲线的复杂程度。然而，由于拉曼光谱不同波长区间内的信号强度与噪声水平差异较大，滑动窗口大小与多项式阶数不易调节。当滑动窗口过大时可能会平滑掉光谱中的细节，多项式阶数过高则可能导致过度拟合。因此，运用 Savitzky-Golay 滤波法去除拉曼光谱中的噪声存在一定的局限性。

2.2 全局阈值小波变换滤波法

小波变换[14]理论最初由法国工程师 Morlet J.于 1974 年提出，后经众多科学家接续研究与发展而趋于成熟。小波变换具备自适应时频信号分析要求的能力，能够提供一个随频率改变的“时间-频率”窗口，克服了传统滤波技术窗口大小不随频率变化的缺点，目前被广泛应用于信号的分析与处理。

小波变换滤波法的基本思想是将信号分解成多个小波系数，采用软阈值和硬阈值相结合的方法，通过阈值参数与阈值函数对小波系数进行处理，依据处理后的小波系数对信号进行重构。其中，阈值参数和阈值函数的选择是小波变换去噪的关键，直接影响重构信号的质量[15]。

传统的阈值函数有硬阈值和软阈值两种，软阈值处理可以有效地保留信号的细节，而硬阈值处理可以更彻底地滤除噪声，具体函数如下。

1）硬阈值函数

$$\hat{d}=\begin{cases}0 & |d|<\lambda\\ d & |d|\geqslant\lambda\end{cases} \tag{1}$$

2）软阈值函数

$$\hat{d}=\begin{cases}0 & |d|<\lambda\\ \mathrm{sign}(d)\left(|d|-\lambda\right) & |d|\geqslant\lambda\end{cases} \tag{2}$$

式中：d 为小波系数；sign(·)为符号函数；λ 为阈值，$\lambda=\sigma\sqrt{2\ln w}$，$\sigma=M_{\mathrm{mid}}/0.674\,5$，$M_{\mathrm{mid}}$ 为最低层小波系数的中位数，w 为信号尺度；$\hat{d}$ 为经阈值处理后的小波系数。

从硬、软阈值函数表达式可以看出，硬阈值将小于阈值的小波系数置零，软阈值将大于阈值的小

波系数减去阈值，二者配合使曲线连续且平滑。传统的阈值函数在对小波系数作阈值处理时，阈值不随窗口内噪声水平的变化而改变，从而限制了其滤波效果。因此，本文结合拉曼光谱噪声的变化规律，对传统阈值进行修正，提出一种自适应全局阈值参数 δ，其表达式为

$$\begin{cases}\delta=(1+F)\ \sigma\sqrt{2\ln w}\\ F=\dfrac{1}{1-\left(\max|d|-\min|d|\right)}\end{cases}\tag{3}$$

式中：F 为修正系数；$\max|d|$与 $\min|d|$分别为当前窗口内小波系数的最大值与最小值。

小波系数较大通常表示小波与原始信号的相似程度较高，反之较小。因此，当小波系数最大值与最小值差异较大时，说明当前窗口内的噪声信号强度较高，则应当提高阈值来增强噪声的滤除水平。由式（3）可知，当窗口内小波系数最大值与最小值差异较大时，修正系数 F 增大，相应地，全局阈值 δ 提高，符合拉曼光谱噪声分布的一般规律。

2.3　去噪效果对比

以加热时长 72h 的拉曼光谱信号为例，运用全局阈值小波变换滤波法对其进行去噪处理，去噪后的光谱曲线如图 3 所示。图中蓝色实线为原始拉曼光谱，红色虚线为去噪后的光谱，黄色点划线为噪声信号。

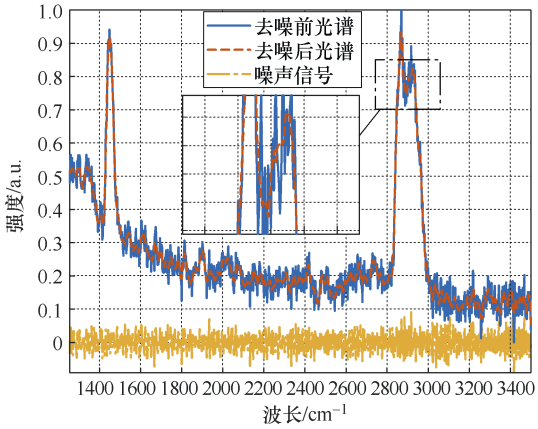

图 3　全局阈值小波变换滤波法去噪后的光谱曲线

从图 3 可以看出，原始拉曼光谱中毛刺状的噪声被消除，同时极大程度地保留了曲线的特征信息。仔细观察光谱曲线与噪声信号可以发现，噪声信号随波长的增加而不断增强，平均强度也逐渐提高，符合拉曼光谱噪声分布的变化规律，与本文提出的修正系数与全局阈值的设计目标相符。

为了进一步验证全局阈值小波变换滤波法的去噪性能，选取加热时长 0～288h 的拉曼光谱作为研究对象，分别运用 Savitzky-Golay 滤波法和全局阈值小波变换滤波法对其进行去噪处理，以方均根误差[16]（root mean square error, RMSE）与信号噪声比[17]（signal-noise ratio, SNR）作为评价指标，对比分析二者的去噪性能。

RMSE 是衡量预测值和观测值之间差异的指标，反映数据噪声去除的效果，具体定义为

$$R_{\mathrm{MSE}}=\frac{1}{N_{\mathrm{T}}}\sqrt{\sum_{h=1}^{N_{\mathrm{T}}}\left(y_{rh}-y_{sh}\right)^2}\tag{4}$$

式中：N_{T} 为波段中的波长个数；y_{rh} 与 y_{sh} 分别为预测值与观测值。

SNR 是衡量信号与噪声比例的重要指标，表征光谱有效信息的保留程度，具体定义为

$$S_{\mathrm{NR}}=10\lg\frac{E_{\mathrm{s}}}{E_{\mathrm{v}}}\tag{5}$$

其中

$$E_{\mathrm{s}}=\sum_{h=1}^{N_{\mathrm{T}}}(S(h)-S_{\mathrm{p}})^2$$

$$E_{\mathrm{v}}=\sum_{h=1}^{N_{\mathrm{T}}}(S(h)-S_{\mathrm{c}})^2$$

式中：E_{s} 为信号强度；E_{v} 为噪声强度；S 为原始拉曼光谱；S_{p} 为原始光谱均值；S_{c} 为去噪后的光谱。

Savitzky-Golay 滤波法中滑动窗口大小设置为 11，多项式拟合阶数为 3。经两种滤波法去噪后的各加热时长拉曼光谱的 RMSE 见表 1。

表 1　两种滤波法去噪后拉曼光谱的 RMSE

加热时长/h	RMSE	
	Savitzky-Golay 滤波	全局阈值小波变换滤波
0	3.612 6	2.452 7
48	2.938 4	2.120 5
72	2.641 7	1.201 8
96	3.235 1	1.312 2
120	3.362 8	1.387 9
144	3.510 2	1.488 1
168	3.771 9	1.570 3
192	4.065 2	1.681 6
240	4.395 0	1.871 2
288	4.837 4	2.121 3

RMSE 越小，表示噪声滤除的效果越好。从表 1 所示结果可以发现，Savitzky-Golay 滤波法的 RMSE 保持在 2.5～5.0，而全局阈值小波变换滤波法的 RMSE 稳定在 1.2～2.5，去噪效果明显优于 Savitzky-Golay 滤波法。

Savitzky-Golay 滤波法参数设置不变，经两种滤波法去噪后的各加热时长拉曼光谱的 SNR 见表 2。

表 2 两种滤波法去噪后拉曼光谱的 SNR

加热时长/h	SNR	
	Savitzky-Golay 滤波	全局阈值小波变换滤波
0	21.284 3	25.650 9
48	20.875 1	24.976 0
72	19.963 4	24.121 8
96	18.715 6	22.069 1
120	17.481 5	20.812 6
144	16.029 7	19.787 5
168	14.548 2	18.503 7
192	12.907 8	17.438 3
240	11.152 0	16.145 2
288	9.036 9	14.954 4

SNR 越高，表明信号中噪声的比例越低，即噪声滤除的效果越好。从表 2 所示结果不难看出，对于全部加热时长的拉曼光谱曲线，全局阈值小波变换滤波法的 SNR 均明显大于 Savitzky-Golay 滤波法。并且，在噪声强度随老化程度加深而不断提高的情况下，全局阈值小波变换滤波法的 SNR 下降速率明显低于 Savitzky-Golay 滤波法。

结合两种滤波法的方均根误差和信噪比的对比结果可以得出结论：对于拉曼光谱数据，全局阈值小波变换滤波法比传统的 Savitzky-Golay 滤波法具有更强的去噪性能。

3 拉曼光谱基线校正

原始拉曼光谱中不仅包含各类噪声信号，还存在荧光背景的干扰。荧光背景又称作基线干扰，一般是由绝缘油老化产生的荧光物质及油中杂质的荧光性所引起，其存在不利于特征信息的提取。因此，对采集到的拉曼光谱信号进行基线校正很有必要。

3.1 改进 airPLS 基线校正法

airPLS 是目前使用最广泛的基线校正方法，具备良好的自适应性和迭代加权惩罚特性。其原理是将惩罚最小二乘算法与自适应迭代重加权算法相结合，通过不断调整权重来拟合基线。然而，在拉曼光谱的谱峰区域，airPLS 拟合基线可能会由于谱峰强度的剧烈变化而明显抬升，从而导致校正后的光谱谱峰强度减弱，这不利于拉曼特征峰的研究与分析[18]。

针对 airPLS 拟合拉曼光谱基线存在的缺陷，本文结合双尺度高斯判别法[19]与多项式拟合法提出一种改进 airPLS 基线校正法，并通过实例验证其去除拉曼光谱荧光背景的应用效果。

双尺度高斯判别法基于高斯峰型计算滑动位置的相关系数来确定潜在的拉曼谱峰。每个滑动位置都对应一个相关系数，该系数反映信号与高斯峰型之间的相似程度。相关系数越大，表明该位置存在拉曼谱峰的可能性越高。为了确认某处的拉曼谱峰是否存在，将该位置的信噪比作为阈值，当相关系数大于阈值时，将该位置识别为拉曼谱峰。相关系数的计算公式为

$$R=\frac{\sum_{i=1}^{N}\left(x_i-x_{\mathrm{p}}\right)\left(y_i-y_{\mathrm{p}}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i-x_{\mathrm{p}}\right)^2\sum_{i=1}^{N}\left(y_i-y_{\mathrm{p}}\right)^2}} \tag{6}$$

式中：x_i、y_i 分别为拉曼光谱和高斯峰型在滑动位置 i 处的强度值；x_{p}、y_{p} 为对应信号的均值；N 为滑动位置总个数。

高斯拟合峰型的表达式为

$$f=\sum_{i=1}^{N}f(i)=\sum_{i=1}^{N}\frac{A_i}{\sqrt{2\pi}\mu_i}\exp\left[\frac{\left(i-u_i\right)^2}{-2\mu_i^2}\right] \tag{7}$$

式中：A_i 为峰的振幅；u_i 为平移参数；μ_i 为标准差。

多项式拟合法也是目前常用的基线校正方法，其原理是通过一个多项式函数对光谱背景信号进行建模，多项式函数为

$$Q(v)=a_0+a_1v+a_2v^2+\cdots+a_bv^b \tag{8}$$

式中：v 为波长；$a_0, a_1, a_2,\cdots, a_b$ 为多项式系数；b 为多项式阶数。

改进 airPLS 基线校正的具体步骤如下：

1）运用双尺度高斯判别法对拉曼光谱寻峰，以谱峰横坐标为中心向两侧寻找一阶导数极小值。

2）截断两极小值点之间的拉曼谱峰区域，保留其余区域的信号对其作 airPLS 基线校正。

3）对截断的拉曼谱峰区域运用多项式拟合法进行基线校正，得到多项式拟合基线。

4）将 airPLS 拟合基线与多项式拟合基线拼接得到组合基线，再作平滑处理得到最终的拟合基线。

3.2　基线校正效果对比

为了验证改进 airPLS 去除拉曼光谱荧光背景的应用效果，以去噪后的加热时长 48h 的拉曼光谱为例，分别运用 airPLS 与改进 airPLS 对其进行基线校正。同时，为了尽可能消除拟合基线在谱峰区域及其附近出现的起伏现象，适当降低双尺度高斯判别法中的谱峰识别阈值。

运用 airPLS 基线校正前后的光谱曲线与拟合基线如图 4 所示。图中蓝色实线为校正前的拉曼光谱，红色虚线为 airPLS 拟合基线，黄色点划线为校正后的拉曼光谱。

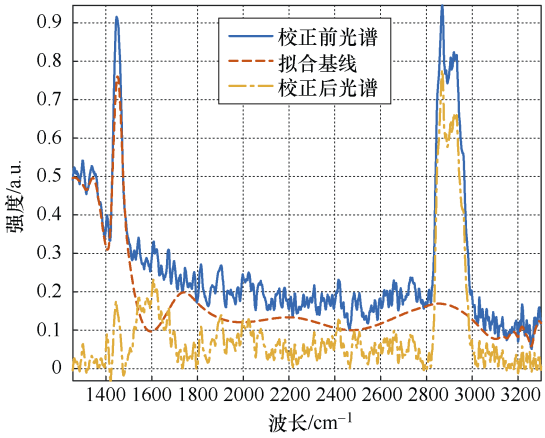

图 4　加热 48h 拉曼光谱的 airPLS 基线校正结果

从图 4 所示曲线可以看出，airPLS 拟合基线在位于 1 400～1 500cm^{-1} 波长的谱峰区域明显抬升，导致校正后的谱峰强度大幅减弱。同时，拟合基线在谱峰附近区域也出现了较为明显的起伏，不符合实际拉曼光谱基线的变化规律。

运用改进 airPLS 基线校正前后的拉曼光谱曲线与拟合基线如图 5 所示。图中蓝色实线为基线校正前的光谱，红色虚线为改进 airPLS 拟合基线，黄色点划线为基线校正后的光谱。

结合图 4 和图 5 可以看出，改进 airPLS 拟合基线消除了 airPLS 拟合基线在谱峰区域及其附近的抬升与起伏现象，在去除荧光背景的同时极大程度地保留了拉曼光谱的特征信息。实际应用结果表明，本文提出的改进 airPLS 基线校正法对于准确去除拉曼光谱的荧光背景具有良好的应用效果。

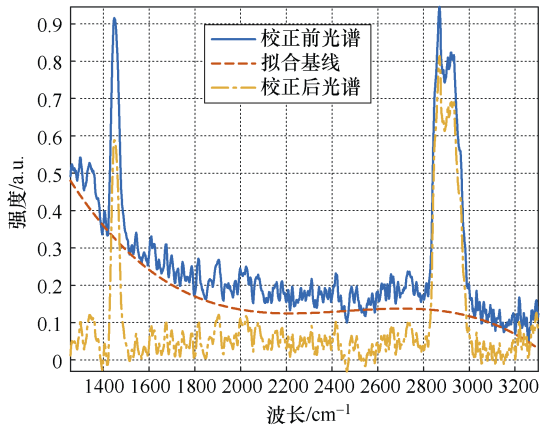

图 5　加热 48h 拉曼光谱的改进 airPLS 基线校正结果

4　拉曼光谱特征提取与分析

拉曼光谱中蕴含丰富的绝缘老化特征信息，可作为绝缘老化评估的重要依据。为了防止不重要的特征信息对实验结果造成影响，必须从拉曼光谱众多的老化特征信息中提取出最具代表性的特征量，以达到提高老化判别准确性的目的。

4.1　连续投影算法

连续投影算法[20-21]是一种使矢量空间共线性最小化的前向变量选择算法，利用向量的投影分析筛选出共线性最小的特征波长组合。SPA 能够消除原始光谱中冗余的信息和光谱波长共线性的影响，从而有效降低模型的复杂度。

作为一种前向迭代搜索算法，SPA 在迭代过程中将波长投影到其他波长上，选择投影向量最大的波长构成共线性最小的特征波长组合[22-23]。

SPA 具体步骤如下：

1）迭代开始前，建立光谱矩阵 $\boldsymbol{X}_{n\times m}$（$n$ 为样本数，m 为波长数），并确定待选特征波长的个数 H。

2）初始迭代周期 t=1 时，在光谱矩阵中任选一列向量 $\boldsymbol{x}_j$ 记为 $\boldsymbol{x}_{k(0)}$，$k(0)$为所选变量的最初位置（$j=k(0)$，$1\leqslant j\leqslant m$），其余变量的位置集合 s 定义为

$$s=\left\{j,1\leqslant j\leqslant m,j\notin\left\{k(0),\cdots,k(H-1)\right\}\right\} \tag{9}$$

3）计算其余列向量 $\boldsymbol{x}_j$（$j\in s$）在 $\boldsymbol{x}_{k(t-1)}$构成的正交向量空间的投影，即

$$\begin{cases}\boldsymbol{x}_j=\boldsymbol{P}\boldsymbol{x}_j\\ \boldsymbol{P}=\boldsymbol{I}-\dfrac{\boldsymbol{x}_{k(t-1)}\cdot\left(\boldsymbol{x}_{k(t-1)}\right)^{\mathrm{T}}}{\left(\boldsymbol{x}_{k(t-1)}\right)^{\mathrm{T}}\cdot\boldsymbol{x}_{k(t-1)}}\end{cases} \tag{10}$$

式中：$\boldsymbol{I}$ 为单位矩阵；$\boldsymbol{P}$ 为投影算子。

4）选择投影向量最大的变量 arg[max(||$\boldsymbol{Px}_j$||)]构成变量集合 S_x。

5）$t=t+1$，若 $t<H$，则返回步骤 3）循环计算。

当循环结束时，得到变量集合 $S_x=\{\boldsymbol{x}_{k(0)}, \boldsymbol{x}_{k(1)},\cdots, \boldsymbol{x}_{k(H)}\}$。对光谱矩阵 $\boldsymbol{X}_{n\times m}$ 中的每个波长都进行上述迭代得到若干个变量集合 S_x，计算每个变量集的交叉验证方均根误差，选择交叉验证方均根误差最小值对应的 $k(0)$和变量集合 S_x，即为筛选出的最优特征变量组合。

4.2 特征提取与分析

选取加热时长 0～409h 的全部拉曼光谱数据（波长区间 1 250～3 500cm^{-1}）作为研究对象，运用 SPA 对其进行特征提取。设置 SPA 待选特征波长个数 H=4，特征波长选择结果如图 6 所示。

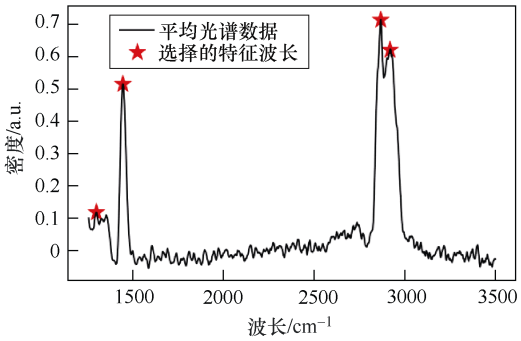

图 6 SPA 特征波长选择结果

从图 6 可以看出，SPA 提取的 4 个特征波长与光谱中强度前 4 的拉曼峰值点所处位置相对应，分别为波长 1 284.36cm^{-1}、1 424.80cm^{-1}、2 895.42cm^{-1}、2 946.34cm^{-1}。此结果初步表明，拉曼峰值点所处的波长及其峰值强度与变压器绝缘油老化程度之间存在一定关系。

为了进一步分析特征波长对应的拉曼强度与变压器绝缘油老化程度之间的相关性，选取加热时长 0～288h 的变压器油样作为研究对象，绘制各特征波长对应的拉曼强度随加热时长（即老化程度）变化的曲线，具体如图 7 所示。

从图 7 可以看出，各特征波长对应的拉曼强度随绝缘老化程度的加深而减弱。查阅相关文献[24]可知，上述 4 个特征波长与变压器油中溶解的特征化学物质相对应。其中，特征波长 1 284.36cm^{-1} 与 1 424.80cm^{-1} 为故障特征气体 CO_2 的拉曼特征峰，特征波长 2 895.42cm^{-1} 与 2 946.34cm^{-1} 为故障特征气体 C_2H_6 的拉曼特征峰。此结果充分说明了 SPA 提取的特征波长的有效性，同时也表明拉曼特征峰

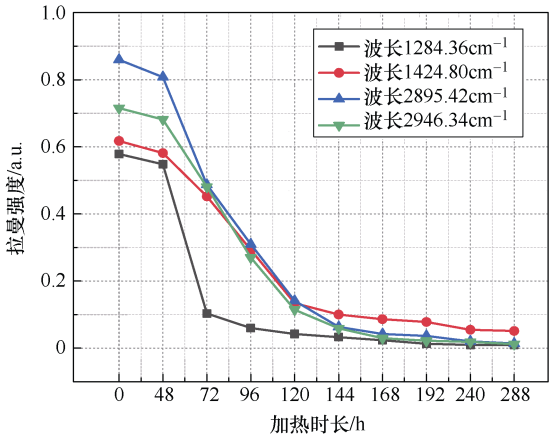

图 7 各特征波长对应的强度随加热时长变化的曲线

的强度可以作为初步判别老化程度的依据。

5 绝缘油老化状态评估

目前，XGBoost[25]分类模型因其诊断精度的优势而被广泛应用于变压器绝缘老化状态判别。由于基于拉曼光谱判别变压器老化状态所需要的光谱数据样本量级较大、噪声信号占比较高，因此为了尽可能提高老化判别的准确率与效率，本文提出运用 LightGBM[26]分类模型实现对变压器绝缘老化的准确判别。

5.1 XGBoost 分类模型

XGBoost 是一种基于梯度提升决策树（gradient boosting decision tree, GBDT）的改进提升（boosting）集成算法，支持分类、回归、排序等功能，具备高效、灵活等特点。XGBoost 以分类回归决策树（classification and regression tree, CART）为基础模型[27]，其实现分类的原理是在每次迭代过程中都增加一棵分类回归决策树，通过不断拟合残差建立新的树，逐渐形成由众多决策树集成的强评估器，最后将所有决策树对应的分数相加，得到最终的预测结果[28]。

GBDT 算法的基本思想是将上一轮的训练残差作为下一轮学习器训练的输入，这使 GBDT 在计算每一维特征所有可能分割点的增益时都必须遍历整个数据集。因此，当面对大规模数据样本及高维特征空间时，GBDT 的运算时间成本和内存资源消耗将大大增加[29]。XGBoost 作为 GBDT 的一种改进算法，在面对高维特征空间及数据量级较大的情况时同样会带来极大的计算代价。

5.2 LightGBM 分类模型

LightGBM 是一种基于 GBDT 的改进 Boosting

集成算法，但相较于 XGBoost，能有效地解决算法处理大量数据时存在的问题，更适用于基于拉曼光谱判别变压器绝缘油老化状态这一具体的应用领域。

LightGBM 算法相较于 XGBoost 做了较多的改进与优化，包括基于直方图（histogram）的数据处理方法、带深度限制的按叶（leaf-wise）生长策略和基于单边梯度采样（gradient-based one-side sampling, GOSS）的训练处理策略等，增强了模型对噪声的鲁棒性和处理大规模数据的能力，同时具有更高的预测精度、更快的模型训练速度和更低的内存占有率等优势，可有效满足构建准确、高效的变压器绝缘老化状态判别模型的需求。

5.3　老化状态判别

为了实现准确的变压器绝缘老化状态判别，通过测定 12 组绝缘油样中的糠醛含量来划分样本的老化阶段。依据我国 DL/T 596—2021《电力设备预防性试验规程》中详细规定的油纸绝缘变压器不同绝缘寿命下的糠醛含量限值划分样本老化阶段：当油中糠醛含量达到 0.5mg/L 时，变压器整体进入绝缘寿命中期；当油中糠醛含量达到 1～2mg/L 时，变压器整体绝缘劣化严重；当糠醛含量大于 4mg/L 时，变压器整体进入绝缘寿命晚期。需要注意的是，测定的加热时长为 409h 的变压器绝缘油样的油中糠醛含量略超出 2mg/L，考虑到从加热实验到送检的时间差和检测过程中可能存在的误差，将加热时长为 409h 的变压器绝缘油样归入绝缘劣化严重这一分类区间。因此，各老化样本的油中糠醛含量测定结果见表 3。

表 3　各老化样本油中糠醛含量测定结果

老化样本加热时长/h	油中糠醛含量/(mg/L)	所处老化阶段
0	0	绝缘寿命初期
48	0.451	
72	0.669	绝缘寿命中期
96	0.941	
120	1.217	绝缘劣化严重
144	1.401	
168	1.413	
192	1.433	
240	1.626	
288	1.878	
307	1.964	
409	2.095	

为了进一步验证本文采用的 LightGBM 模型在诊断精度方面的优越性，运用 LightGBM 模型与 XGBoost 模型对同一老化数据样本进行状态判别，对比二者的判别准确率。

验证实验以变压器绝缘油作为研究对象，依据表 3 所示划分结果对这 12 组老化样本进行类别划分，整理 4.2 节中提取到的特征波长对应的拉曼强度数据并建立样本数据集。样本共包含 4 508 条数据，将数据集按 7:2:1 的比例划分成训练集、测试集和验证集，对 LightGBM 模型和 XGBoost 模型进行训练，比较训练后的两种模型前 10 次预测的准确率，如图 8 所示。

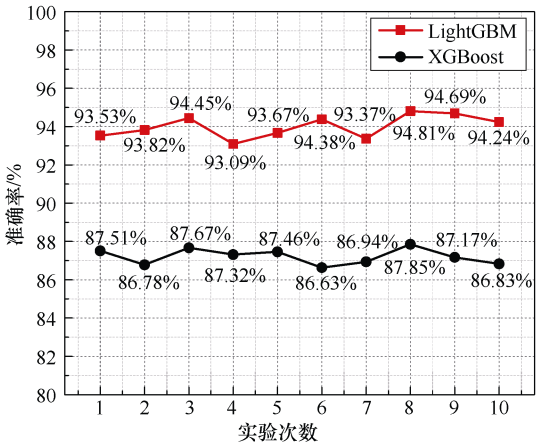

图 8　LightGBM 与 XGBoost 前 10 次预测的准确率对比

从图 8 可以看出，XGBoost 模型前 10 次预测结果的准确率保持在 86%～88%，而 LightGBM 模型前 10 次预测的准确率稳定在 93%～95%，平均准确率为 94.11%，诊断精度明显优于 XGBoost 模型。同时，在实验过程中注意到，相比 XGBoost 模型，LightGBM 模型的平均运行时间大幅缩减，诊断效率更高。此实验结果说明，LightGBM 模型对变压器绝缘老化状态的判别具有良好的应用效果，同时也进一步验证了 4.2 节中所提取特征信息的有效性及其与绝缘油老化程度之间的相关性。

6　结论

本文以拉曼光谱技术为手段，以准确判别变压器绝缘老化状态为目的，制备了绝缘油老化样本，开展了基于拉曼光谱数据处理、特征提取与老化诊断的相关研究。

1）结合拉曼光谱噪声变化规律，基于小波变换理论提出全局阈值小波变换滤波法，有效地去除了拉曼光谱中的噪声信号。以 RMSE 与 SNR 作为评价

指标，比较了全局阈值小波变换滤波法与 Savitzky-Golay 滤波法的去噪效果，进一步验证了其去噪性能。

2）充分考虑拉曼光谱信号特征，基于 airPLS 提出改进 airPLS 基线校正法。实际应用结果表明，该改进方法不仅能够准确地去除拉曼光谱的荧光背景，还可以极大程度地保留拉曼光谱的特征信息。

3）运用 SPA 提取拉曼光谱中的老化特征信息，通过查找油中特征化学物质所处的拉曼波长验证了所提取特征信息的有效性，并深入分析了其与绝缘油老化程度之间的关系。

4）运用 LightGBM 分类模型实现了对变压器绝缘油老化状态的高效、准确判别，进一步验证了所提取特征信息的有效性及其与绝缘老化程度之间的相关性。

参考文献

[1] 邹阳, 林锦茄, 李安娜, 等. 基于灰色关联分析和聚类云模型的变压器油纸绝缘状态评估[J]. 电力系统保护与控制, 2023, 51(21): 35-43.

[2] 苏凯强, 刘庆珍. 基于 TLS-ESPRIT 算法的变压器油纸绝缘等效电路参数辨识及新特征量提取[J]. 电气技术, 2022, 23(7): 89-96.

[3] 夏源, 杨丽君, 吕晓露, 等. 基于介质响应电流频变特性的油纸绝缘受潮状态评估方法[J]. 电工技术学报, 2024, 39(11): 3444-3456.

[4] 林智勇, 李荣华, 黄国泰, 等. 基于频域介电谱二次微分解谱法的变压器油纸绝缘介电响应等效电路参数辨识[J/OL]. 电工技术学报, 1-10 [2025-02-18]. https://doi.org/10.19595/j.cnki.1000-6753.tces.241347.

[5] 陈新岗, 陈姝婷, 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.

[6] 张振宇, 李永祥, 阎寒冰, 等. SF_6 分解特征组分拉曼光谱检测分析系统设计[J]. 光学与光电技术, 2021, 19(6): 11-18.

[7] 宋睿敏, 王建新, 陈伟根, 等. 矿物油中糠醛液芯光纤增强拉曼光谱原位检测方法[J]. 电工技术学报, 2023, 38(24): 6828-6838.

[8] 范元超, 陈孝敬, 黄光造, 等. 基于拉曼光谱的电线绝缘材料老化状态评估[J]. 光谱学与光谱分析, 2022, 42(10): 3161-3167.

[9] 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.

[10] 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.

[11] 王建新, 陈伟根, 王品一, 等. 变压器故障特征气体空芯反谐振光纤增强拉曼光谱检测[J]. 中国电机工程学报, 2022, 42(16): 6136-6144, 6187.

[12] 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.

[13] WANG Ruiqi, QI Ying, ZHANG Qiang, et al. A multi-step water quality prediction model based on the Savitzky-Golay filter and Transformer optimized network[J]. Environmental Science and Pollution Research, 2023, 30(50): 109299-109314.

[14] 胡睿喆, 杨晓峰. 基于小波散射网络-贝叶斯优化门控循环单元的电力变压器声纹识别方法[J]. 电气技术, 2024, 25(8): 35-40, 46.

[15] 方松琼, 邵荣君, 邱丽荣, 等. 基于小波变换的共焦拉曼图像去噪方法[J]. 光学技术, 2019, 45(3): 330-335.

[16] 刘宗溢, 张彩虹, 蒋健康, 等. 基于拉曼光谱结合 CNN-LSTM 深度学习方法的铁皮石斛总黄酮含量快速检测研究[J]. 光谱学与光谱分析, 2024, 44(4): 1018-1024.

[17] 王忠, 万冬冬, 单闯, 等. 基于反向传播神经网络的拉曼光谱去噪方法[J]. 光谱学与光谱分析, 2022, 42(5): 1553-1560.

[18] 白云飞, 罗海燕, 李志伟, 等. 基于改进的自适应迭代重加权惩罚最小二乘的空间外差拉曼光谱基线校正方法[J]. 光学学报, 2024, 44(7): 251-259.

[19] 刘庆珍, 张溢, 鄢仁武. 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[J]. 电力系统保护与控制, 2024, 52(8): 158-166.

[20] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.

[21] CHEN Yiyang, CHU Bing, FREEMAN C T. Generalized iterative learning control using successive projection: algorithm, convergence, and experimental verification[J]. IEEE Transactions on Control Systems Technology, 2020, 28(6): 2079-2091.

[22] 贾宗潮, 王子鉴, 李雪莹, 等. 主成分分析和连续投影融合的海洋沉积物粒度分类研究[J]. 光谱学与光

（下转第 28 页）

（续表 A.2）

始端	末端	长度/m	始端	末端	长度/m
3	6	225	17	18	525
4	7	225	18	19	525
5	8	225	17	20	1 050
5	9	1 050	18	21	225
2	10	525	19	22	225
10	11	525	19	23	225
11	12	525	4	11	630
10	13	225	12	19	630

附录 B

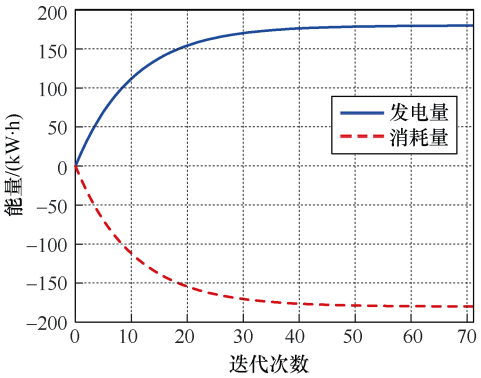

图 B.1　网络 2 总发电量和消耗量

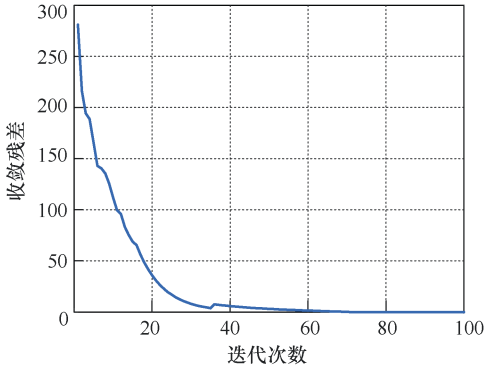

图 B.2　ADMM 收敛曲线（网络 2）

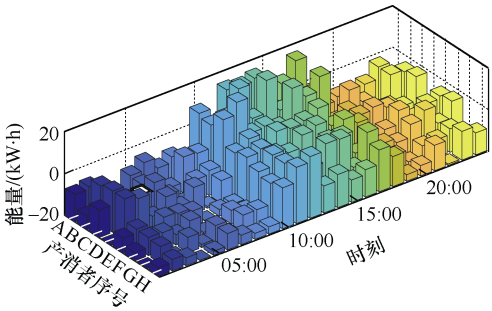

图 B.3　P2P 总交易量（网络 2）

收稿日期：2025-01-15

修回日期：2025-02-25

作者简介

陈俞邑（1999—），男，硕士研究生，研究方向为电力市场、分布式能源系统优化。

（上接第 16 页）

谱分析, 2023, 43(10): 3075-3080.

[23] 牛芳鹏, 李新国, 白云岗, 等. 遗传算法和连续投影算法结合的土壤有机碳含量高光谱估算模型[J]. 光谱学与光谱分析, 2023, 43(7): 2232-2237.

[24] 钱国超, 彭庆军, 彭惠. 变压器油中气体拉曼光谱光纤增强检测实验研究[J]. 云南电力技术, 2021, 49(6): 21-25.

[25] PUNITHA N, DIVYA BHARATHI K, MANUSKANDAN S R, et al. Analysis of muscle fatigue progression using geometric features of surface electromyography signals and explainable XGBoost classifier[J]. Journal of Medical and Biological Engineering, 2024, 44(2): 191-197.

[26] 韩璟琳, 冯喜春, 胡平, 等. 基于 Hyperopt-LightGBM 的直流配电网短期负荷抗噪声预测[J]. 高电压技术, 2024, 50(11): 4902-4911.

[27] 王俊杰, 毕利, 张凯, 等. 基于多特征融合和 XGBoost-LightGBM-ConvLSTM 的短期光伏发电量预测[J]. 太阳能学报, 2023, 44(7): 168-174.

[28] FRIFRA A, MAANAN M, MAANAN M, et al. Harnessing LSTM and XGBoost algorithms for storm prediction[J]. Scientific Reports, 2024, 14(1): 11381.

[29] 赖汶鸿, 刘庆珍, 鄢仁武. 基于关联信息熵和轻量级梯度提升机的油纸绝缘特征优选策略[J]. 电气技术, 2024, 25(1): 34-41, 47.

收稿日期：2025-02-18

修回日期：2025-03-01

作者简介

周宇含（2000—），男，福建省福州市人，硕士研究生，研究方向为电力设备绝缘老化评估与智能故障诊断。