

变压器油中铁颗粒检测的激光诱导击穿光谱处理方法研究

叶哲, 袁欢*, 刘定新**, 王小华, 杨爱军, 荣命哲

西安交通大学电气工程学院电力设备电气绝缘国家重点实验室, 陕西 西安 710049

摘要 为实现激光诱导击穿光谱技术对变压器油中铁颗粒含量的准确定量检测, 本文提出了一套光谱处理算法。该算法基于特征峰的稀疏性建模并结合凸优化规则改进数值迭代过程, 以实现低通基线估计和特征峰分离; 进一步, 用目标分析线对光谱进行降维, 并依靠基于袋装决策树的集成分类器完成烧蚀点的二分类。滤波器截止频率、滤波阶数、不对称比例、正则化参数分别取 0.15、1.6、0.8 时, 基线估计及 Fe 特征谱线强度的准确性高; 袋装决策树集成分类器对降维后光谱的识别准确率高达 98.33%; 有效点的光谱校正强度与 Fe 颗粒质量比的线性相关系数为 0.9983。所提方法能精确实现特征谱线的提取和有效点筛选, 保证油中 Fe 颗粒质量比定量检测的准确性。

关键词 光谱学; 激光诱导击穿光谱; 光谱处理算法; 变压器油; 铁颗粒

中图分类号 O657.38; O657.319; O433.5+4

文献标志码 A

DOI: 10.3788/CJL220852

1 引言

在油浸式变压器运行过程中, 绝缘纸板老化^[1]、潜油泵机械故障^[2-3]以及分接开关动作电弧都会在变压器油内引入不同的颗粒杂质。若颗粒杂质悬浮或附着在变压器绕组、部件表面时, 将危及变压器的运行安全^[4-8]。近年来, 激光诱导击穿光谱(LIBS)被广泛应用于非金属、金属含量检测领域^[9-11], 在变压器油中颗粒物检测方面也有相关研究, 而且滤纸辅助 LIBS 面扫描方法已被证明可以有效完成颗粒定量分析任务^[12-13]。

激光烧蚀等离子体冷却时, 处于激发态的原子或离子的外层电子跃迁产生特征谱线, 为颗粒物检测提供了依据^[14]。除了特征谱线外, 激光烧蚀等离子体光谱中还包含分子谱带以及等离子体初期的韧致辐射和复合辐射, 这些辐射会影响特征谱线的提取。此外, 激光对颗粒沉积面积进行面扫描时, 需要对大量烧蚀点的光谱数据进行二分类。人工或半人工的方法无法满足大量高维数据的处理需求, 因此, 迫切需要一套数据处理算法对原始数据进行基线校正、降维和分类。

近年来, 稀疏化模型在谱线分析领域取得了一定进展, 利用光谱特征谱线的稀疏性建模可使基线拟合具有更高的可解释性和准确性。Ning 等^[15]将稀疏性去噪与凸优化技术结合的高效迭代数值算法应用于色

谱分析, 得出了优于多项式拟合的基线校正结果。Yi 等^[16]提出了一种基于凸优化框架的基线校正算法, 该算法采用不对称罚函数和迭代算法保证了基线估计结果的精度和收敛性, 提高了 LIBS 信号的信噪比和定量精度。

典型的决策树^[17]、支持向量机(SVM)^[18]、K-最近邻分类法(KNN)^[19]等算法常被用来进行数据分类, 但实际数据的复杂性常常导致单一算法的分类效果较差。将多种分类器结合成集成学习分类器能解决单一分类器的不稳定问题, 提高分类的准确性。集成学习方式主要有 Bagging^[20]、Boosting^[21]及 Stacking^[22]。Bagging 方式对有放回抽取的数据子集训练所得模型进行投票, Boosting 方式将弱分类器组合后通过迭代学习完善模型, Stacking 方式以预先训练的多个模型的输出作为输入来训练目标模型。目前的分类算法各有所长, 需要结合实际数据判断其分类效果。此外, 机器学习算法在处理高维数据时通常会受到冗余信息干扰而产生“维度灾难”, 影响分类准确性。鉴于此, 本课题组采用基于稀疏性的基线估计和去噪算法(BEADS)对 LIBS 光谱进行基线校正, 采用目标分析线对数据进行降维, 并结合机器学习分类算法对 LIBS 烧蚀点进行二分类筛选, 以期提高变压器油中颗粒杂质的 LIBS 定量分析方法的数据处理效率和准确性。

收稿日期: 2022-05-12; 修回日期: 2022-06-30; 录用日期: 2022-07-11; 网络首发日期: 2022-07-25

基金项目: 国家自然科学基金(51777154)、中国博士后科学基金(2020M683481)、陕西省“三秦学者”创新团队项目(西安交通大学先进直流电力装备关键技术及其产业化创新团队)

通信作者: huanyuan@xjtu.edu.cn; liudingxin@mail.xjtu.edu.cn

2 实验条件

LIBS 实验装置示意图如图 1 所示。光学系统主要由激光器、光路、延迟发生器、光谱仪、计算机组成。选用波长为 1064 nm 的 Nd:YAG 激光器,激光脉宽为 8 ns,脉冲能量为 60 mJ,激光焦点与试样表面相距 2 mm。多通道光谱仪的光谱窗口范围为 200~1015 nm,通过 DG645 控制光谱仪相对激光发射的延迟时间为 600 ns,光谱仪采集门宽设置为 1.05 ms。位移台用于控制试样移动,使激光对颗粒沉积面进行面扫描。

对含 Fe 颗粒的变压器油进行 LIBS 检测。油样由 KI25X 矿物油与 Fe 颗粒配制而成,油液中 Fe 颗粒的

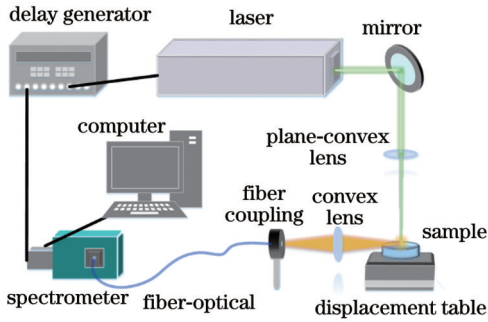


图 1 LIBS 实验装置示意图

Fig. 1 Schematic diagram of LIBS experiment

质量比分别 0.271、0.191、0.135、0.096、0.048、0.016、0.006 μg/g。对油样进行超声波振荡处理后,Fe 颗粒悬浮于油中,随后采用真空负压砂芯过滤装置将油样中的 Fe 颗粒富集到滤纸上制得 LIBS 试样。

3 算法原理

3.1 基于稀疏性的基线估计和去噪算法

将实测 LIBS 光谱图建模为特征谱线、基线、噪声的线性叠加,按式(1)对基线和特征谱线进行分离。

$$y = x + f + w, \quad (1)$$

式中: y 为实测光谱数据向量;向量 x 为一系列特征谱线的稀疏导数;向量 f 为低频背景信号; w 是方差为 σ^2 的高斯白噪声向量。

Selesnick 等^[23]假设滤波器为零向量,且不改变特征峰位置,将高通滤波器描述为

$$H = BA^{-1}, \quad (2)$$

式中: A 和 B 是带状卷积(Toeplitz)矩阵。用带状卷积矩阵表示 H 可利用带状线性系统的快速求解器,采用式(3)所示高通滤波器的二次数据保真度和非对称罚函数的凸优化求解特征峰估计值 \hat{x} 的最优化问题。该式本质上是对残差进行考察。

$$\|y - \hat{s}\|_2^2 = \|y - Ly - H\hat{x}\|_2^2 = \|H(y - \hat{x})\|_2^2, \quad (3)$$

其中,

$$\hat{x} = \arg \min_x \left\{ F(x) = \frac{1}{2} \|H(y - x)\|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_\epsilon(x_n; r) + \sum_{i=0}^M \lambda_i \sum_{n=0}^{N_i-1} \phi([D_i x]_n) \right\}, \quad (4)$$

式中: s 为 x 与 f 之和; L 为 H 对应的低通滤波器; D_i 为 i 阶差分算子; $\theta_\epsilon(x_n; r)$ 是非对称罚函数,其中 r 为不对称比例; ϕ 为可微的对称罚函数,用于稀疏性建模; λ 为正则化参数,与稀疏性大小紧密相关; N_i 是 $D_i x$ 的长度; ϵ 是与罚函数相关的正数,其值无限接近 0,但不能过小,一方面要保证罚函数的有效稀疏性,另一方面要避免算法在优化过程中出现问题;参数 r 与非对称罚函数的峰值正性有关。式(4)本质上是对残差进行考察。

适用于 LIBS 光谱基线校正的 $\theta_\epsilon(x_n; r)$ 见式(5)。BEADS 算法迭代收敛便可完成 LIBS 光谱的基线校正和特征峰提取。

$$\theta_\epsilon(x_n; r) = \begin{cases} \frac{1+r}{4|v|} x^2 + \frac{1-r}{2} x + \frac{(1+r)|v|}{4}, & |v| > \epsilon \\ \frac{1+r}{4\epsilon} x^2 + \frac{1-r}{2} x + \frac{(1+r)}{4}, & |v| \leq \epsilon \end{cases}, \quad (5)$$

式中: $|v|$ 表示非对称罚函数值优化器。

3.2 机器分类算法

3.2.1 决策树

将样本集按照一定的特征顺序进行分割,若按特

征构建的子集的信息熵或基尼指数最小,则将该特征确定为节点。通过一系列的特征选择来确定决策树。信息熵用来度量特征的纯度,定义为

$$H(A) = - \sum_{K=1}^k P(A_K) \log_2 [P(A_K)], \quad (6)$$

式中: $H(A)$ 为样本集 A 的信息熵; $P(A_K)$ 为样本 A 分为 k 类后的第 K 类样本所占比例。

如果用特征 a 对样本 A 分类产生了 n 个节点,则特征 a 对样本 A 划分的信息增益定义为

$$G(A, a) = H(A) - \sum_{N=1}^n \frac{|A^N|}{|A|} H(A^N), \quad (7)$$

式中: $G(A, a)$ 为信息增益; $|A^N|$ 为 A 在特征 a 上取值 A^N 的样本数。信息增益越大,特征划分样本的不确定度就越小。

3.2.2 支持向量机

支持向量机是常用的二分类算法,该算法采用超平面作为决策边界对样本进行分类。两类样本中与超平面最近的点到超平面的距离和最大时,该超平面即为决策边界。本质上,决策边界仅由支持向量决定,支持向量就是距离超平面最近的几个点。样本线性不可分时,通过核函数将数据映射到高维空间,以使其在高

维空间线性可分。

参考直线方程,定义超平面方程为

$$\mathbf{W}^T \mathbf{X} + b = 0, \quad (8)$$

式中: \mathbf{W} 、 \mathbf{X} 为 n 维向量; b 为常数。 n 维空间内点 X_0 到超平面的距离 d_{is} 为

$$d_{is} = \frac{|\mathbf{W}^T \mathbf{X}_0 + b|}{\|\mathbf{W}\|}, \quad (9)$$

式中: $\|\mathbf{W}\|$ 为范数,即 \mathbf{W} 所有元素平方和的开方; d_{is} 为分类间隔值。

支持向量机将所有样本点 X_i 中最小的分类间隔定为 1,则所有样本点的分类间隔都不会小于 1。对每个样本点添加分类标签 y_i (y_i 为 +1 或 -1),则可正确分类的样本点满足约束条件

$$y_i(\mathbf{W}^T \mathbf{X}_i + b) \geq 1. \quad (10)$$

上述为支持向量机的核心思想。接下来通过构造拉格朗日函数求解超平面。

3.2.3 K 最近邻分类法(KNN)

K 最近邻分类法的分类思想是:对于输入向量,从训练数据集中寻找 K 个与之接近的向量的集合,将 K 个向量中数目最多的分类标签赋予该输入向量。向量空间内两样本之间的距离表征它们之间的相似度,距离越小,相似度就越高。常用的距离计算方法有闵可夫斯基距离、欧氏距离、曼哈顿距离、切比雪夫距离等。

对于 j 维空间内的两点 x 和 y ,闵可夫斯基距离的计算公式为

$$d_{xy} = \sqrt[p]{\sum_{j=1}^p [(x_j - y_j)]^p}, \quad (11)$$

式中: d_{xy} 为 x 、 y 之间的距离; x_j 、 y_j 为 x 、 y 点对应的元素值; p 为可变参数。当 $p=1$ 时, d_{xy} 为曼哈顿距离;当 $p=2$ 时, d_{xy} 为欧氏距离;当 $p=3$ 时, d_{xy} 为切比雪夫距离。

4 结果与讨论

4.1 基线估计与扣除

采用 BEADS 算法对 Fe 颗粒的光谱图进行基线估计与扣除,将光谱数据按稀疏峰、低通基线和随机残差三部分建模。非对称惩罚函数和凸优化技术的应用有利于降低过拟合程度,优化基线估计。BEADS 算法的可调参数包括滤波器截止频率 f_c ($0 < f_c < 0.5$)、滤波阶数 d 、不对称比例 r 、正则化参数 λ 。图 2 为调参实验的基线估计结果。 f_c 增大到 0.15 时,基线估计与原始基线的重合度极高,而且 Fe 特征谱线的极小值落在基线上;随着 f_c 继续增大,特征峰发生过拟合。图示特征谱线如 Fe I 360.89 nm、Fe I 361.88 nm、Fe I 363.15 nm、Fe I 364.78 nm、Fe I 371.99 nm 的稀疏性优秀,而 Fe I 373.49 nm 与 Fe I 373.71 nm 之间、Fe I 374.83 nm 与 Fe I 374.95 nm 之间、Fe I 376.38 nm 与周围噪声信号之间有重叠,稀疏性相对较差,可能需要更高的滤波阶数。当 $d=1$ 时,特征峰的基线拟合良好,增大 d 反而会发生过拟合。不对称比例 r 用于对谱峰的对称性进行规范。该值过大时基线

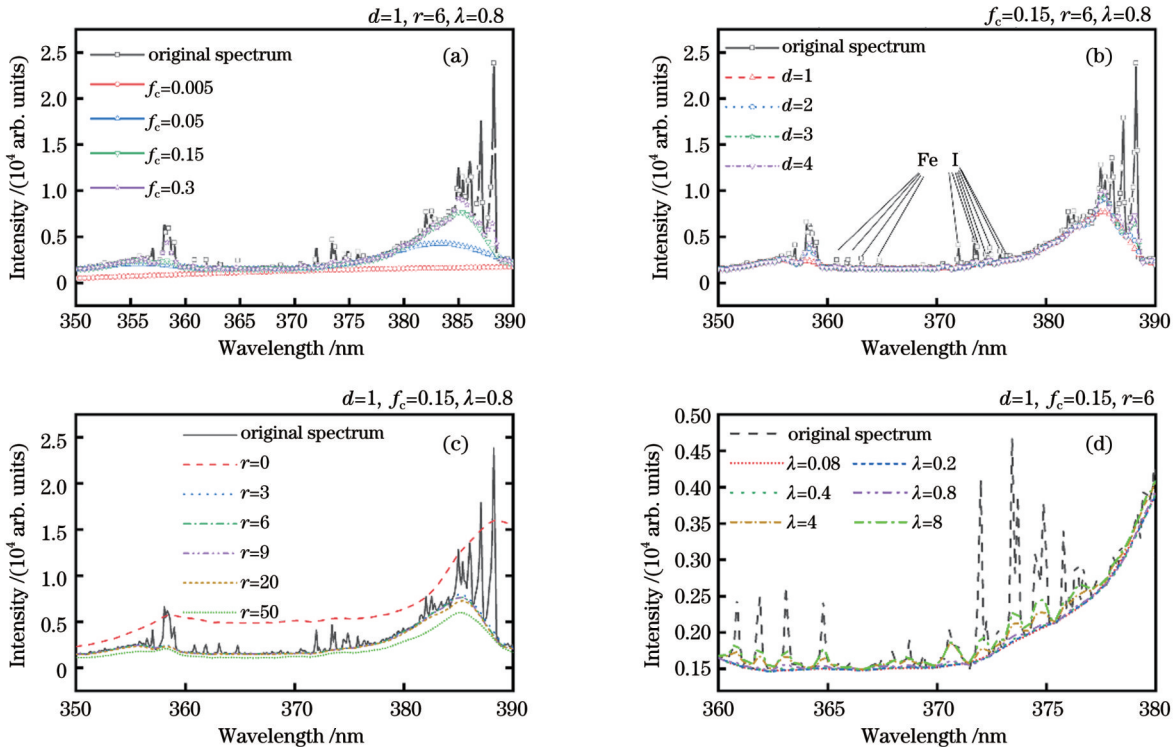


图 2 基线估计曲线的参数化研究。(a) f_c ; (b) d ; (c) r ; (d) λ

Fig. 2 Parameterization of baseline estimation curve. (a) f_c ; (b) d ; (c) r ; (d) λ

欠拟合,过小时会发生基线过拟合,在 3~9 之间取值时基线估计性能良好。因此, r 取 6 是较为合适的选择。正则化参数 λ 在 0.08~0.4 内取值时,基线轻微欠拟合; $\lambda=0.8$ 能使欠拟合得到校正,从而提升特征谱线强度提取的准确性;当 λ 取 4 或 8 时,基线发生过拟合。

选择 $f_c=0.15, d=1, r=6, \lambda=0.8$ 时,基线估计的准确度高,基于此得到的光谱原始数据、基线估计、基线扣除校正光谱、残差如图 3 所示。对于含 Fe 颗粒的油样的 LIBS 光谱, BEADS 算法出色地完成了连续背景的扣除,使 Fe I 稀疏峰强度得到校正,同时基线拟合所得的残差值很小,充分说明了基线估计的准确性。

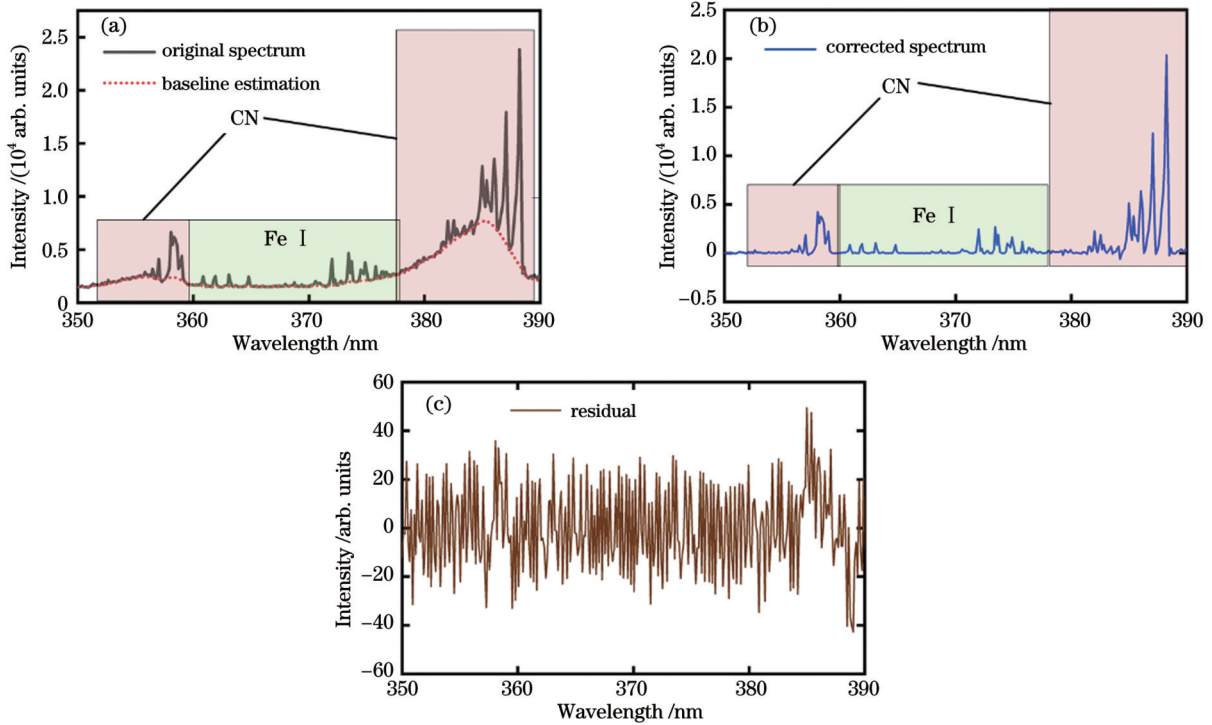


图 3 基于最优参数的光谱基线估计与扣除。(a)原始光谱与基线估计;(b)校正光谱;(c)残差

Fig. 3 Baseline estimation and deduction of spectrum based on the optimal parameters. (a) Original spectrum and baseline estimation; (b) corrected spectrum; (c) residual

4.2 光谱降维与分类

LIBS 全谱为 5610 维数据,对海量烧蚀点进行分类时难免会发生“维数灾难”。通过提取定量分析的特征参量对数据进行降维处理能去除无关光谱信息,提高分类器的准确性。LIBS 定量分析基于目标元素特征谱线的峰值强度展开,不同烧蚀点光谱特征谱线强度的差异归因于 Fe 颗粒的烧蚀量。因此,通过提取 Fe 元素特征谱线附近的光谱数据可以实现降维。结合美国国家标准与技术研究院(NIST)元素发射谱线数据库可知,Fe 的原子发射光谱有 103 条之多,而由于跃迁概率等原因,其中理论强度较高的谱线有 27 条,如表 1 所示。

Fe 颗粒在滤纸衬底表面分布极不均匀,随着 Fe 颗粒含量降低,仅部分烧蚀点的光谱包含 Fe 谱线。本文将光谱中包含 Fe 谱线的烧蚀点称为有效点,不包含 Fe 谱线的烧蚀点称为无效点。Fe 颗粒质量比低至 0.016 $\mu\text{g/g}$ 以下时,表 1 中理论强度较低的特征谱线难以检测到,此时试样有效点的校正光谱如图 4 所示,图中每个颜色代表一个有效点的光谱数据。结合表 1 可以发现,仅有 12 条特征谱线在所有有效点光谱中均被

表 1 理论强度较高的 Fe 原子发射谱线

Table 1 Fe atomic emission lines with high theoretical intensity

Wavelength / nm	Intensity / arb. units	Wavelength / nm	Intensity / arb. units
344.06	400	374.83	300
346.59	120	374.95	400
347.55	200	375.82	300
349.06	250	376.38	150
356.54	100	382.04	500
357.01	120	382.44	250
358.12	600	404.58	300
360.89	150	406.36	150
361.88	150	407.17	120
363.15	120	428.24	120
364.78	150	430.79	120
371.99	600	432.58	150
373.49	700	438.35	200
373.71	600		

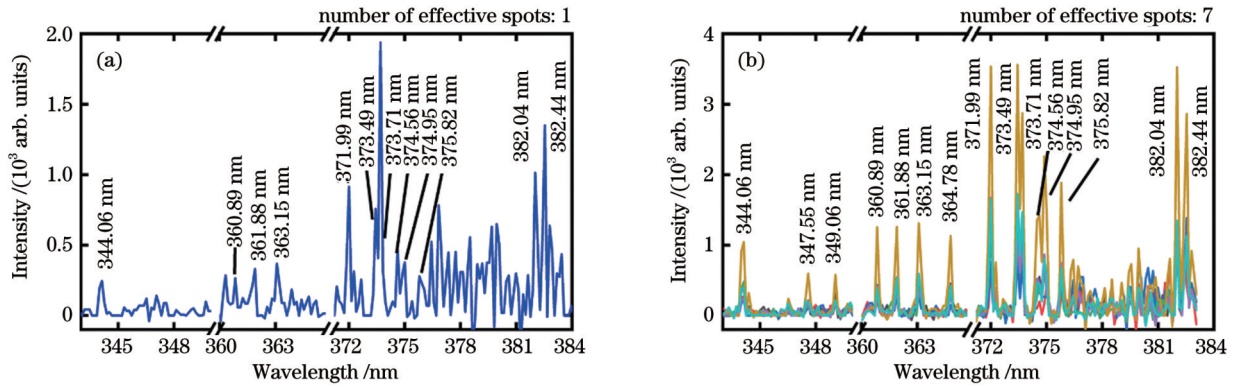


图 4 低质量比 Fe 颗粒有效点的校正光谱图。(a)0.006 μg/g; (b)0.016 μg/g

Fig. 4 Corrected spectra of effective spots of Fe particles with low mass ratio. (a) 0.006 μg/g; (b) 0.016 μg/g

检测到。因此,本文光窗取值范围主要包含这 12 条特征谱线。用这些分析线附近的数据可将 5610 维原始光谱降维至 68 维,同时保证了有效信息的完整性,能大幅提高烧蚀点的分类效率。

采用决策树、支持向量机、K 最近邻分类法以及基于它们的集成分类器进行烧蚀点识别,结合 10 折交叉验证法对光谱分类效果进行验证,校正后的高维数据、经主成分分析法(PCA)降维的数据、目标分析线降维数据的分类准确率如图 5 所示。相较于校正后的高维

数据和 PCA 降维后的数据,采用目标分析线进行降维后,烧蚀点的分类准确率大幅提高,说明该降维处理方式相较于 PCA 对光谱有效信息保存的完整性和科学性更高。在所有分类算法中,袋装决策树(Bagged Trees)的集成分类器的性能最优异,其对高维光谱和目标分析线降维处理后的光谱识别准确率分别高达 95% 和 98.33%。该算法在决策树的基础上,有放回地抽取 n 个样本,构建 n 个决策树模型,可以降低方差,实现分类准确率的提升。

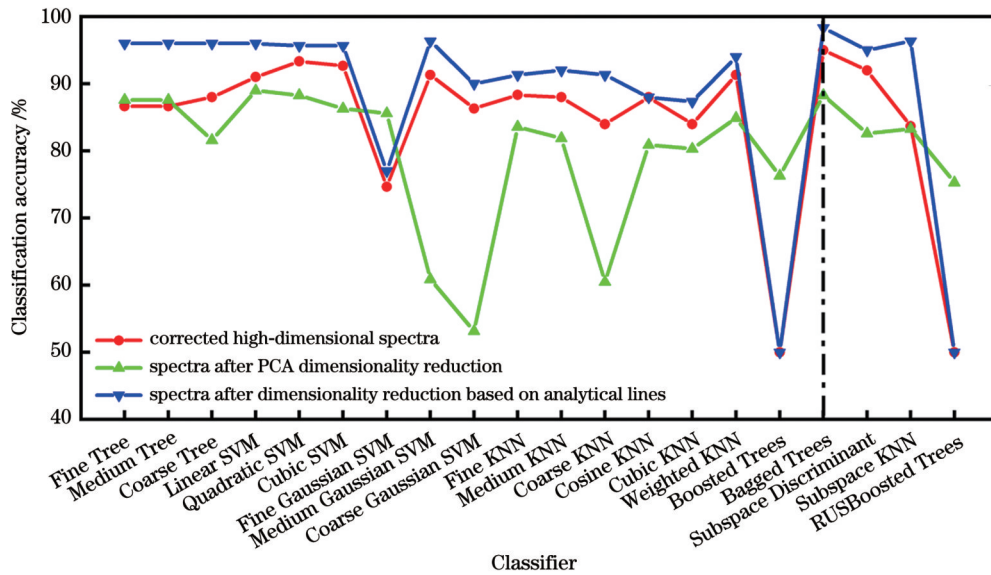


图 5 光谱分类准确性

Fig. 5 Spectral classification accuracy

袋装决策树分类混淆矩阵如图 6 所示。该分类器对绝大部分烧蚀点分类正确,而产生极少数错误分类的主要原因是原始光谱数据特征谱线极低的信噪比。这些信噪比极低的信号可能会导致分类误差,同时其对定量分析的影响极小,甚至可以忽略。因此,该分类器保证了绝大多数烧蚀点的正确分类,甚至可能还会降低分类误差。

4.3 Fe 颗粒定量分析

采用 BEADS 算法对烧蚀点光谱进行基线扣除,

并依据目标分析线对其降维,然后采用袋装决策树集成分类器筛选有效点完成定量分析。以 4.2 提到的 12 条 Fe I 特征谱线为分析线,选用 CN 358.6 nm 为内标线,以分析线校正强度之和作为定量分析的特征参量,得到了图 7 所示的结果。图中各数据点为 5 次重复实验的平均值,误差棒为重复实验的标准偏差。LIBS 有效点光谱校正强度与 Fe 颗粒质量比的线性相关系数达到了 0.9983,且实验重复性较高,说明了该光谱数据处理流程的合理性。

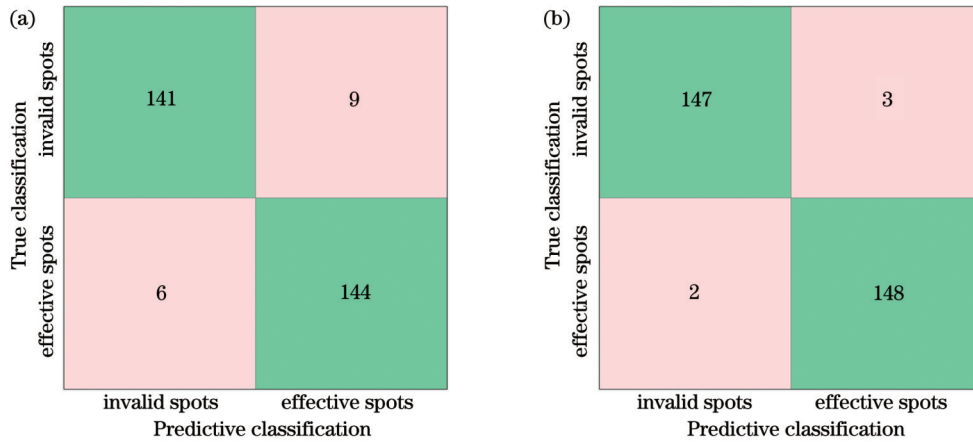


图 6 袋装决策树分类混淆矩阵。(a)高维光谱;(b)目标分析线降维后的光谱

Fig. 6 Confusion matrix for Bagged Trees classification results. (a) High-dimensional spectra; (b) spectra after dimensionality reduction based on analytical lines

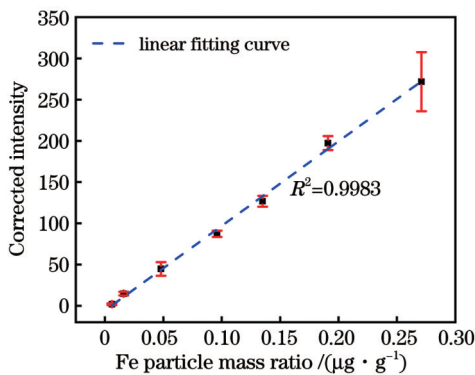


图 7 Fe 颗粒质量比检测的定标曲线

Fig. 7 Calibration curve of Fe particles mass ratio detection

5 结 论

本文提出了一套结合光谱基线扣除、数据降维和分类的光谱处理算法,有效提高了光谱数据的处理效率。BEADS 算法的校正基线与原始光谱背景信号具有极高的重合度,能准确提取 Fe I 特征谱线的强度。以分析线光谱进行降维相较于 PCA 降维更能保证数据的完整性,基于袋装决策树的集成分类器对烧蚀点的自动分类准确率达 98.33%。该光谱处理算法的 LIBS 光谱校正强度与 Fe 颗粒质量比之间的线性相关系数达到了 0.9983。本文所提光谱处理算法不仅有效简化了变压器油中金属颗粒 LIBS 检测的光谱处理过程,还保证了 LIBS 技术的定量精度。

参 考 文 献

- [1] 严家明. 油浸绝缘纸局部放电损伤特性研究[D]. 重庆: 重庆大学, 2011.
Yan J M. Study on characteristics of partial discharge damage to oil-impregnated insulation paper[D]. Chongqing: Chongqing University, 2011.
- [2] Oommen T V, Petrie E M. Particle contamination levels in oil-filled large power transformers[J]. IEEE Transactions on Power Apparatus and Systems, 1983, 102(5): 1459-1465.
- [3] 袁道君. 油泵故障引起的变压器油金属污染及现场再生处理[J]. 变压器, 1999, 36(1): 35-38.
Yuan D J. Metal pollution of transformer oil caused by oil pump fault and its on-site treatment[G]. Transformer, 1999, 36(1): 35-38.
- [4] Guo R C, Wang Z, He C, et al. Breakdown characteristics of transformer oil with cellulose particles in a non-uniform field under lightning impulse voltage[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2020, 27(5): 1627-1635.
- [5] Krins M, Borsi H, Gockenbach E. Influence of carbon particles on the breakdown voltage of transformer oil[C] // ICDL'96.12th International Conference on Conduction and Breakdown in Dielectric Liquids, July 15-19, 1996, Roma, Italy. New York: IEEE Press, 1996: 296-299.
- [6] Mahmud S, Chen G, Golosnoy I O, et al. Bridging in contaminated transformer oil under DC and AC electric field[J]. Journal of Physics Conference Series, 2013, 472(1): 012007.
- [7] Wang X, Wang Z D. Particle effect on breakdown voltage of mineral and ester based transformer oils[C]// 2008 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, October 26-29, 2008, Quebec, QC, Canada. New York: IEEE Press, 2008: 598-602.
- [8] Mathes K N, Atkins J M. Influence of particles on partial discharges and breakdown in oil[C] // 1978 IEEE International Conference on Electrical Insulation, June 12-14, 1978, Philadelphia, PA, USA. New York: IEEE Press, 1978: 226-231.
- [9] 傅院霞, 贾韧, 徐鹏, 等. 金属基质激光诱导击穿光谱最佳离焦量的基质效应研究[J]. 中国激光, 2022, 49(13): 1311003.
Fu Y X, Jia R, Xu P, et al. An exploration of matrix effect on optimal defocus distance of metal matrix in laser-induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2022, 49(13): 1311003.
- [10] 许铖, 李芳, 陈锋, 等. 钛合金的激光诱导击穿光谱快速分类[J]. 光子学报, 2022, 51(4): 0430001.
Xu C, Li F, Chen F, et al. Rapid classification of laser induced breakdown spectroscopy of titanium alloys[J]. Acta Photonica Sinica, 2022, 51(4): 0430001.
- [11] 杨明磊, 刘玉柱. 基于激光诱导击穿光谱技术的条纹紫菜元素探测研究[J]. 激光与光电子学进展, 2022, 59(10): 1030001.
Yang M L, Liu Y Z. Element detection in porphyra yezoensis via laser-induced breakdown spectroscopy[J]. Laser & Optoelectronics Progress, 2022, 59(10): 1030001.
- [12] Ye Z, Zhu Z H, Ma N, et al. Copper particle contamination detection of oil-immersed transformer using laser-induced breakdown spectroscopy[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2020, 167: 105820.

- [13] Yuan H, Ye Z, Wang X H, et al. Study on spark discharge enhanced laser-induced breakdown spectroscopy of Fe particles in transformer oil[J]. *Journal of Analytical Atomic Spectrometry*, 2022, 37(2): 381-389.
- [14] Guo L B, Zhang D, Sun L X, et al. Development in the application of laser-induced breakdown spectroscopy in recent years: a review[J]. *Frontiers of Physics*, 2021, 16(2): 22500.
- [15] Ning X R, Selesnick I W, Duval L. Chromatogram baseline estimation and denoising using sparsity (BEADS) [J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, 139: 156-167.
- [16] Yi C C, Lü Y, Xiao H, et al. A novel baseline correction method using convex optimization framework in laser-induced breakdown spectroscopy quantitative analysis[J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2017, 138: 72-80.
- [17] Gordon A D, Breiman L, Friedman J H, et al. Classification and regression trees[J]. *Biometrics*, 1984, 40(3): 874.
- [18] Leslie C, Eskin E, Noble W S. The spectrum kernel: a string kernel for SVM protein classification[J]. *Pacific Symposium on Biocomputing*, 2002: 564-575.
- [19] Lah M F B C. Pattern recognition using K-nearest neighbors (KNN) technique[D]. Batu Gajah Perak: Universiti Teknologi Petronas, 2010.
- [20] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123-140.
- [21] Valiant L G. A theory of the learnable[J]. *Communications of the ACM*, 1984, 27(11): 1134-1142.
- [22] Breiman L. Stacked regressions[J]. *Machine Learning*, 1996, 24(1): 49-64.
- [23] Selesnick I W, Graber H L, Pfeil D S, et al. Simultaneous low-pass filtering and total variation denoising[J]. *IEEE Transactions on Signal Processing*, 2014, 62(5): 1109-1124.

Spectral Processing for Detection of Iron Particles in Transformer Oil by Laser-Induced Breakdown Spectroscopy

Ye Zhe, Yuan Huan*, Liu Dingxin**, Wang Xiaohua, Yang Aijun, Rong Mingzhe

State Key Lab of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

Abstract

Objective During the operation of an oil-immersed transformer, aging of the insulating cardboard, mechanical failure of the submersible pump, and action arc of the tap change introduce different particulate impurities into the transformer oil. If the particulate impurities suspend in the oil flow or adhere to the surface of the transformer windings and components, the safety of the transformer will be compromised. In recent years, laser-induced breakdown spectroscopy (LIBS) has been widely used for the detection of nonmetals and metals. There are also reports on the detection of particulate matter in transformer oil, and the filter-paper-assisted LIBS map-scanning method has proven to be effective for the quantitative analysis of particles. In this study, in addition to the target spectral lines, molecular bands and bremsstrahlung and recombination radiation in the plasma generated by laser ablation are present, and the resulting continuous background spectrum cannot be shielded during spectral analysis, which adversely affects the spectral intensity of the characteristic spectral lines of the target element. However, the ablation spots obtained by laser scanning the deposition area of the sample do not contain characteristic spectral lines of the target element, hence necessitating binary classifications to screen out the spectrally effective spots for quantitative analysis. Therefore, there is an urgent need for a set of data-processing algorithms for baseline correction, dimensionality reduction, and classification of the original data to meet the processing requirements of a large number of high-dimensional spectral data.

Methods Based on the sparsity of characteristic spectral lines, this study investigates the application of the baseline estimation and denoising algorithm (BEADS) in LIBS spectral baseline correction and subtraction. The results of the dimensionality reduction using the target analysis line and principal component analysis (PCA) are compared. In addition, the binary classification effects of different machine learning algorithms (e.g., decision tree, support vector machine, K-nearest neighbor classification, and ensemble classifier) on laser ablation spots are studied, in which the ensemble classifier perfects the classifier model through different optimization methods. Finally, based on the above spectrum-processing algorithm, a quantitative analysis and calibration of Fe particle detection is completed.

Results and Discussions Spectral analysis reveals that the characteristic spectral lines such as Fe I 360.89 nm, Fe I 361.88 nm, Fe I 363.15 nm, Fe I 364.78 nm, and Fe I 371.99 nm have excellent sparsity, whereas Fe I 373.49 nm, Fe I 373.71 nm, Fe I 374.83 nm, Fe I 374.95 nm, and Fe I 376.38 nm exhibit relatively poor sparsity owing to interference with each other or overlap with the background spectrum. The experimental results show that, as shown in Fig. 3, the application of asymmetric penalty functions and convex optimization techniques is beneficial for reducing overfitting. When the parameters are adjusted to $f_c=0.15$, $d=1$, $r=6$, and $\lambda=0.8$, the accuracy of the baseline estimation is very high, and the residual value obtained by the baseline fitting is small. For the detection of Fe particles in transformer oil, the BEADS algorithm can excellently deduct the continuous background, so that the intensity of the Fe characteristic spectral lines can be accurately corrected. As shown in Fig. 5, in terms of the classification accuracy of ablation spots by the classifier, the spectral data after dimensionality reduction using the target analysis line is better than the original spectral data and the spectral data after PCA is performed. This demonstrates that the dimensionality reduction processing

method using the target analysis line is scientific and reasonable. Based on the decision tree algorithm, the Bagged Trees ensemble classifier constructs multiple decision tree models through the extraction of different samples, thereby reducing the variance, optimizing the classification model, and improving classification accuracy. The spectral classification accuracy for data after dimensionality reduction is as high as 98.33%.

Conclusions Based on the spectral processing method, the linear correlation coefficient between the spectral correction intensity and particle mass ratio is 0.9983, and the relative standard deviation of repeated experiments is small, which proves the scientificity and robustness of the method. The method can realize batch processing of a large number of ablation spot data generated by laser map scanning, which greatly improves data-processing efficiency while reducing errors introduced by manual processing. It provides convenience for LIBS automatic acquisition and data processing and lays a theoretical foundation for LIBS detection of particulate matter in transformer oil.

Key words spectroscopy; laser-induced breakdown spectroscopy; spectral processing algorithm; transformer oil; iron particle