

# 基于激光诱导击穿光谱的茶叶品种快速分类

徐向君<sup>1</sup>, 王宪双<sup>1</sup>, 李昂泽<sup>1</sup>, 何雅格<sup>1</sup>, 柳宇飞<sup>2</sup>, 何锋<sup>1</sup>, 郭伟<sup>1</sup>, 刘瑞斌<sup>1\*</sup>

<sup>1</sup>北京理工大学物理学院, 北京 100081;

<sup>2</sup>宝瑞激光科技(常州)有限公司, 江苏 常州 213000

**摘要** 在提取激光诱导击穿光谱(LIBS)全部特征峰的基础上,利用支持向量机建立了有效的茶叶分类模型。采集了 15 种茶叶样品的有效 LIBS 光谱数据(190~720 nm),运用窗口平移平滑和峰位漂移函数修正对光谱进行了预处理,再结合主成分分析降维,对绿茶、红茶、白茶实现了 98.3% 的识别率;对同一种类中不同品种的茶叶也实现了较好的识别。研究结果表明,LIBS 在茶叶品种快速识别应用中具有较好的前景。

**关键词** 光谱学;激光诱导击穿光谱;茶叶品种;快速分类;光谱预处理;主成分分析;支持向量机

中图分类号 O433.4

文献标识码 A

doi: 10.3788/CJL201946.0311003

## Fast Classification of Tea Varieties Based on Laser-Induced Breakdown Spectroscopy

Xu Xiangjun<sup>1</sup>, Wang Xianshuang<sup>1</sup>, Li Angze<sup>1</sup>, He Yage<sup>1</sup>, Liu Yufei<sup>2</sup>,  
He Feng<sup>1</sup>, Guo Wei<sup>1</sup>, Liu Ruibin<sup>1\*</sup>

<sup>1</sup>School of Physics, Beijing Institute of Technology, Beijing 100081, China;

<sup>2</sup>Bao Rui Laser Technology Co., Ltd., Changzhou, Jiangsu 213000, China

**Abstract** On the basis of extracting all the characteristic peaks of laser-induced breakdown spectroscopy (LIBS), an effective tea classification model is established based on support vector machine. The effective LIBS spectral data (190-720 nm) of fifteen tea samples are collected, and the spectra are preprocessed by window translation smoothing and peak shift function correction. Combined with principal component analysis for dimensionality reduction, the recognition rate of green tea, black tea and white tea is 98.3%. Different varieties of tea in the same species also achieve good recognition. The research results show that LIBS has a good prospect in the rapid identification of tea varieties.

**Key words** spectroscopy; laser-induced breakdown spectroscopy; tea variety; rapid identification; spectral pretreatment; principal component analysis; support vector machine

**OCIS codes** 300.2530; 300.6210; 300.6365

## 1 引 言

作为我国的传统饮品,茶叶越来越受到人们的喜爱,已成为世界三大饮品之一。中国是茶的故乡<sup>[1]</sup>,地理环境非常适宜茶树的生长,茶树品种繁多,采制经验丰富。我国人民在长期的制茶过程中积累了丰富的经验,创造出了精湛的制茶工艺,产生了众多茶叶品种,一般根据茶多酚的氧化聚合程度由浅入深地将茶叶归纳为 6 大类,即绿茶、黄茶、白

茶、青茶、黑茶和红茶。传统对茶叶品种的识别以专家感官评审为主,人为因素对茶叶品种的精确判定有较大的影响。因此有必要研究出一种快速、精确、可靠的茶叶种类识别方法。

激光诱导击穿光谱(LIBS)具有检测速度快、无需样品处理、对样品损伤小等特点,近年来已经越来越多地被应用到物质检测领域,如水体污染<sup>[2]</sup>、土壤分析<sup>[3]</sup>、工业评估<sup>[4]</sup>、食品安全<sup>[5]</sup>、环境监测<sup>[6]</sup>、考古文物<sup>[7]</sup>、医药分析<sup>[8]</sup>等。目前,国内外利用 LIBS 检

收稿日期: 2018-09-19; 修回日期: 2018-10-23; 录用日期: 2018-12-12

基金项目: 国家自然科学基金(61574017)、北京理工大学火炸药全链条创新专项(2017CX10007)

\* E-mail: liuruibin8@gmail.com

测技术快速评价产品质量和快速分类待测样品(如茶叶、药品、塑料产品<sup>[9]</sup>)已经成为 LIBS 领域实际应用的热点。Wang 等<sup>[10]</sup>利用 LIBS 技术结合判别分析(DA)法对铁观音、龙井、普洱等 6 种茶叶进行识别,训练集识别准确率达 98%,测试集识别准确率达 95.33%。饶刚福等<sup>[11]</sup>基于 LIBS 技术,采用不同光谱预处理方法结合 DA、径向基函数网络(RBF)和反向传播网络(MLP)进行建模,实现了对铁观音、碧螺春、普洱等 7 种茶叶的高准确率识别。Zhang 等<sup>[12]</sup>将 LIBS 技术与偏最小二乘判别分析(PLS-DA)法相结合,对龙井、铁观音、金豪等 6 种茶叶进行识别,训练集识别准确率达 100%,测试集识别准确率达 99.77%。蒋帆等<sup>[13]</sup>采用高光谱分析技术结合支持向量机(SVM)对机炒龙井茶等级进行识别,训练集识别准确率达 98%,测试集识别准确率达 95.33%。

以上研究表明,茶叶品种的快速识别具有一定的可行性。目前,LIBS 技术在茶叶种类识别中的研究主要分为 3 个方面:1)针对我国六大茶类不同类别的茶进行识别;2)采用不同的光谱预处理方法提高分类准确率;3)将机器学习中的不同算法应用于模型建立,实现高准确率识别。上述已报道研究可实现对不同种类茶叶的较高识别率的区分,但对同一种类中不同品种茶叶的区分还不全面,识别率也较低。本文在传统光谱预处理中加入了峰位漂移函数修正,结合主成分分析提取特征量并利用支持向量机进行建模,尝试对不同类型和同一种类中不同品种的茶进行同时快速分类识别,获得了准确率更高的判别模型,具有更广的应用范畴。三大品类

包括典型的绿茶、白茶和红茶,交叉检验准确率为 98.3%,分别对红茶、绿茶和白茶每个品类中的 5 个不同品种进行识别,交叉检验准确率分别为 100%、97.6%和 95.2%。

## 2 实验过程

### 2.1 实验装置

实验中,采用的空间光路如图 1 所示,激光器为主动调 Q 的 Nd:YAG 1064 nm (Ultra 100, Quantel, 法国)固体激光器,脉宽为 7 ns,重复频率为 20 Hz,激光器的最大输出能量为 80 mJ。光谱仪为三通道光纤光谱仪(AvaSpec-Dual, Avantes, 荷兰),CCD 分辨率为 2048 pixel×2048 pixel,探测波长范围为 190~720 nm,光学分辨率为 0.1 nm。数字脉冲发生器(DG535, Stanford Research Systems, 美国)为激光器和光谱仪提供精准的外触发信号。

图 1 中激光经过反射镜 M 反射后,再经透镜(焦距  $f=75$  mm)聚焦在放置于三维电动平台的样本表面,产生等离子体。收集透镜( $f=35$  mm)在与激光聚焦方向成  $\theta=45^\circ$  的方向进行光信号收集,并将收集到的光耦合至光纤,使光信号传输至三通道光纤光谱仪,完成光谱的分光与光电转换<sup>[14]</sup>。实验中 DG535 精确控制光谱仪三个通道的延迟时间,从而控制连续背景辐射的强度,进而得到连续背景较小的光谱图。本实验采用的 LIBS 参数为:激光能量为 30 mJ,光谱仪三个通道的延迟分别为 764.04, 752.01, 763.78  $\mu$ s,积分时间为 1 ms,每个样品分别进行 250 次脉冲打样并将其作为待处理的数据。

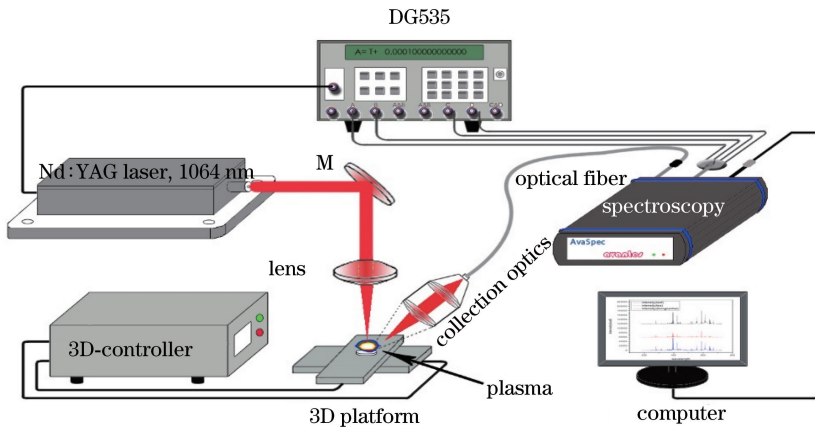


图 1 实验示意图

Fig. 1 Experimental diagram

## 2.2 样品制备

实验中选取市面上常见且具有代表性的绿茶[西湖龙井(XHLJ)、碧螺春(BLC)、日照绿茶(RZGT)、大佛龙井(DFLJ)、山东绿茶(SDGT)]、红茶[祁门红茶(QM)、暖阳红茶(NY)、坦洋工夫红茶(TYGF)、阿萨姆红茶(Assam)、金骏眉(JJM)]、白茶[白牡丹(BMD)、寿眉(SM)、野茶(Wild)、白露茶(BL)、贡眉(GM)]三大类、15种茶叶制成样品。根据以往的实验经验,样品的干燥性、研磨均匀性和密度等物理特性对光谱信号有很大的影响。为了减小其他因素对光谱的影响,需要对茶叶样品进行必要的预处理,然后再进行测定。这15种样品首先需要用吹风机烘干,然后置于80℃的恒温干燥箱5h,再用玛瑙研钵研磨1h,获得质地均匀的茶叶粉末,最后用15MPa的机械粉末压片机将样品压制厚度为2mm、直径为10mm的小圆饼。

## 2.3 茶叶识别模型的构建

实验中采用支持向量机的方法构建分类模型<sup>[15]</sup>。模型建立过程中通过选取不同的核函数,调整支持向量个数,优化惩罚参数 $c$ 和核参数 $g$ ,从而确定最优的

分类超平面,建立预测精度较高的分类模型。

## 3 结果与分析

### 3.1 光谱数据的预处理

本实验中通过多次脉冲打样进行光谱采集并对光谱进行平均的方法来降低激光能量波动或仪器误差带来的光谱信号的波动。对每种茶叶样品分别进行多次打点并采集数据,每种样品收集250组光谱数据,通过对每5组数据求平均获得1幅光谱,每种茶叶样品对应50幅光谱,15种茶叶共对应750幅光谱。图2(a)为白牡丹的典型光谱图,由于其组成成分复杂,光谱中有多条特征峰,此外,在经过一定的延迟后第二通道收集的光谱仍然存在轻微的背景,采用特别的算法来消除光谱中不必要的背景辐射。消除光谱背景常用插值法和窗口平移平滑法<sup>[16]</sup>。但是,采用插值法去除背景得到的光谱存在信息丢失,如特征峰的相对强度改变、半峰全宽信息丢失、峰与峰之间的差异性消失、个别带状分子峰丢失等。本实验中采用窗口平移平滑法,图2(b)为去除背景后的光谱图。

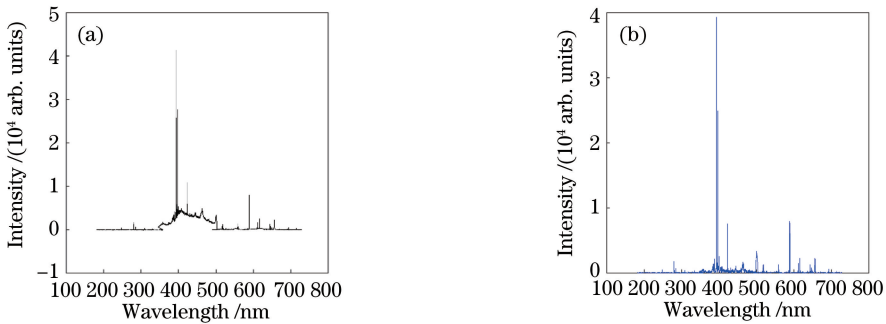


图2 白牡丹茶的 LIBS 光谱。(a)原始光谱;(b)去除背景后的光谱

Fig. 2 LIBS spectra of BMD tea. (a) Original spectrum; (b) spectrum after removing background

由于基体效应和实验误差的存在,同一样品在不同测试点得到的光谱会出现峰位漂移的情况。峰位最大值在两次光谱测量中对应的像素点出现差别,出现约为1 pixel的漂移,对应的波长漂移量为光谱分辨率单位,即0.1 nm<sup>[17]</sup>。为方便处理,先将背景消除的光谱数据进行寻峰处理,找出每组光谱中的特征峰的峰值强度和对应波长并记录,然后用峰位漂移函数对漂移峰进行修正<sup>[17]</sup>。图3为白牡丹样品峰位漂移修正后的光谱数据。

本实验中光谱仪探测波长范围为190~720 nm,图4为白牡丹样品的 LIBS 光谱图。根据美国国家标准与技术研究院(NIST)的标准原子光谱数据库,对其元素组成进行了鉴定和标记。如

LIBS 光谱所示,光谱中含有 Mg、Ca、Fe、Mn 等矿物质营养元素以及 C、H、Si 等非金属元素的原子发射谱线。这些特征谱线是通过仔细观察线型和强度确定的,以确保这些谱线不受其他谱线自吸收效应的影响并提供较强的光谱强度。

### 3.2 主成分分析降维并提取特征量

主成分分析是把多个特征映射为少数几个综合特征的一种统计分析方法,以快速处理数据。在对绿茶、红茶和白茶三种品类的分类研究中,不同品类的样本集是由该品类下5种不同茶叶各随机抽取20幅光谱数据组成,对获得的光谱数据峰位漂移函数进行修正和归一化后,再进行主成分分析降维<sup>[18]</sup>,将全谱变换为多个主成分<sup>[19]</sup>。图5(a)为前

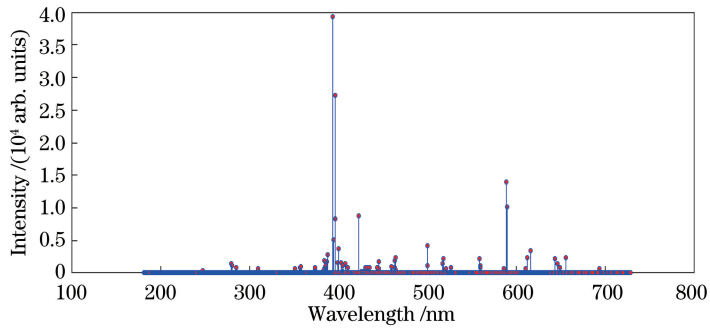


图 3 峰位漂移修正后白牡丹茶的光谱

Fig. 3 Spectra of BMD tea after peak drift correction

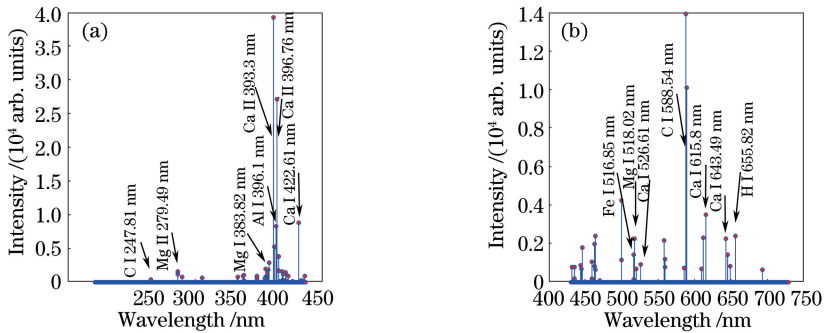


图 4 白牡丹茶的 LIBS 光谱图。(a) 波长范围为 200~430 nm; (b) 波长范围为 430~720 nm

Fig. 4 LIBS spectra of BMD tea. (a) Wavelength range of 200-430 nm; (b) wavelength range of 430-720 nm

10 个主成分中每个主成分对原始数据的解释率, 取 10 个主成分作为特征向量, 累积解释率为图 5(b) 为主成分对原始数据的累积解释率, 这里选 90.13%。

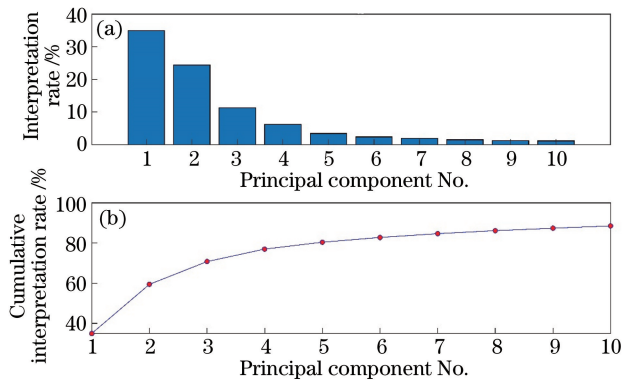


图 5 主成分对原始数据的解释率。(a) 每个主成分的解释率; (b) 主成分的累积解释率

Fig. 5 Rate of interpretation of principal component to original data. (a) Rate of interpretation of each principal component; (b) cumulative interpretation rate of principal components

前 3 个主成分的累积解释率为 76.98%, 图 6 给出前 3 个主成分的三维散点图, 每个散点代表一个样本, 显示出较好的聚类效果。从图中可以看出, 每种品类的不同样品都有其特定的聚集区域, 并与其他类别的样品有一定的间隔。聚类结果是由物质的基本成分含量和分子间关系决定的, 因此, 对于不同类别的茶叶, PCA 是一种有效提取特征量的方法。

对绿茶(西湖龙井、碧螺春、日照绿茶、大佛龙井、山东绿茶)、红茶(祁门、暖阳、坦洋工夫、金骏眉、阿萨姆)和白茶(寿眉、野茶、白露茶、贡眉、白牡丹)各 5 个不同品种分别进行 PCA 降维以提取特征量, 3 种情况下提取的 10 个主成分累积解释率分别为 91.75%、90.79% 和 92.48%, 5 种绿茶和白茶的三维散点图分别有少部分散点重叠, 5 种红茶的三维散点图显示出很好的聚类效果。

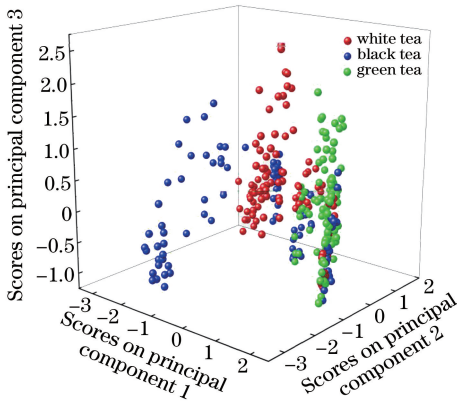


图 6 前 3 个主成分的三维散点图

Fig. 6 Three-dimensional scatter plot of first three principal components

### 3.3 分类结果

支持向量机在解决小样本、非线性及高维模式识别中表现出许多特有的优势,非常适合文中判别

模型的建立。在 PCA 降维之后,提取前 10 个主成分构建特征空间,在特征空间下需要选取训练集和测试集,文中采用小样本的 5-折交叉验证法,即将样本随机分为 5 个包,每次将其中一个包作为测试集、剩下 4 个包作为训练集进行训练,将 5 次结果的正确率(或差错率)平均值作为对算法精度的估计。

白茶、红茶、绿茶的分类结果如表 1 所示,识别率分别为 100%、99%和 96%;5 种绿茶(西湖龙井、碧螺春、日照绿茶、大佛龙井、山东绿茶)的分类结果如表 2 所示,识别率分别为 94%、100%、98%、100%和 96%;5 种红茶(暖阳、坦洋工夫、阿萨姆、金骏眉、祁门)的分类结果如表 3 所示,识别率均为 100%;5 种白茶(寿眉、野茶、白露茶、贡眉、白牡丹)的分类结果如表 4 所示,识别率分别为 100%、100%、100%、80%和 96%。

表 1 基于 SVM 模型的 3 种茶的分类结果

Table 1 Classification results of three kinds of tea based on SVM model

Variety	Number	Predicted variety			True positive rate / %
		white tea	black tea	green tea	
white tea	100	100			100
black tea	100	1	99		99
green tea	100		4	96	96

表 2 基于 SVM 模型的 5 种绿茶的分类结果

Table 2 Classification results of five kinds of green tea based on SVM model

Variety	Number	Predicted variety					True positive rate / %
		XHLJ	BLC	RZGT	DFLJ	SDGT	
XHLJ	50	47		1	2		94
BLC	50		50				100
RZGT	50		1	49			98
DFLJ	50				50		100
SDGT	50	2				48	96

表 3 基于 SVM 模型的 5 种红茶的分类结果

Table 3 Classification results of five kinds of black tea based on SVM model

Variety	Number	Predicted variety					True positive rate / %
		NY	TYGF	Assam	JJM	QM	
NY	50	50					100
TYGF	50		50				100
Assam	50			50			100
JJM	50				50		100
QM	50					50	100

表 4 基于 SVM 模型的 5 种白茶的分类结果

Table 4 Classification results of five kinds of white tea based on SVM model

Variety	Number	Predicted variety					True positive rate / %
		SM	Wild	BL	GM	BMD	
SM	50	50					100
Wild	50		50				100
BL	50			50			100
GM	50				40	10	80
BMD	50				2	48	96



## 4 结 论

通过采集 15 种茶叶在 190~720 nm 波长范围的 LIBS 光谱信息,采用窗口平移平滑法去除背景和峰位漂移函数修正的光谱预处理结合主成分分析,运用 SVM 对茶叶进行建模,红茶、绿茶、白茶的识别率分别为 100%、99% 和 96%,平均识别率为 98.3%;5 种红茶的识别率均为 100%;5 种绿茶的识别率分别为 94%、100%、98%、100% 和 96%,平均识别率为 97.6%;5 种白茶的识别率分别为 100%、100%、100%、80% 和 96%,平均识别率为 95.2%;这说明采用一定的光谱预处理方法结合支持向量机建模,在以上情况下可以实现对茶叶品类和品种较高识别率的快速分类。

## 参 考 文 献

- [1] Yu Z Y, Shu Y. Study on the establishment of tea culture and Taoist culture experience museum in Taishan Scenic area [J]. *China Market*, 2016(4): 106, 108.  
于振宇, 舒源. 泰山景区建立茶文化与道教文化体验馆的研究[J]. *中国市场*, 2016(4): 106, 108.
- [2] Wu J, Wu Y J, Fan S, *et al.* Signal stability of collinear double pulse laser induced breakdown spectroscopy combining with auto-focusing system [J]. *Chinese Journal of Lasers*, 2018, 45(7): 0711003.  
武进, 吴跃进, 范爽, 等. 结合自动聚焦系统的共线双脉冲激光诱导击穿光谱信号稳定性研究[J]. *中国激光*, 2018, 45(7): 0711003.
- [3] Wang J M, Yan H Y, Zheng P C, *et al.* Quantitative detection of nutrient elements in soil based on laser induced breakdown spectroscopy [J]. *Chinese Journal of Lasers*, 2017, 44(11): 1111002.  
王金梅, 颜海英, 郑培超, 等. 基于激光诱导击穿光谱定量检测土壤中营养元素的研究[J]. *中国激光*, 2017, 44(11): 1111002.
- [4] Gu Y H, Li Y, Tian Y, *et al.* Study on the multivariate quantitative analysis method for steel alloy elements using LIBS [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(8): 2244-2249.  
谷艳红, 李颖, 田野, 等. 基于 LIBS 技术的钢铁合金中元素多变量定量分析方法研究[J]. *光谱学与光谱分析*, 2014, 34(8): 2244-2249.
- [5] He X W, Huang L, Liu M H, *et al.* Determination of Cd in rice by laser-induced breakdown spectroscopy [J]. *Applied Laser*, 2014, 34(1): 72-75.  
何秀文, 黄林, 刘木华, 等. 激光诱导击穿光谱对大

米中镉元素的检测分析[J]. *应用激光*, 2014, 34(1): 72-75.

- [6] Wainner R T, Harmon R S, Miziolek A W, *et al.* Analysis of environmental lead contamination: comparison of LIBS field and laboratory instruments [J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2001, 56(6): 777-793.
- [7] Yan H T, Chang Z. The laser-induced breakdown spectroscopy and its application in the polychrome cultural relics [J]. *Relics and Museology*, 2009(6): 229-233.  
阎宏涛, 昌征. 激光诱导击穿光谱分析法及其在彩绘文物分析与表征中的应用[J]. *文博*, 2009(6): 229-233.
- [8] Liu X N, Wu Z S, Qiao Y J. Review on rapid evaluation of product quality attributes and application prospects in Chinese material medica [J]. *World Chinese Medicine*, 2013, 8(11): 1269-1272.  
刘晓娜, 吴志生, 乔延江. LIBS 快速评价产品质量属性的研究进展及在中药的应用前景[J]. *世界中医药*, 2013, 8(11): 1269-1272.
- [9] Sun Q Q, Du M, Guo L B, *et al.* Fast identification of plastics with laser-induced breakdown spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2017, 37(7): 2205-2209.  
孙倩倩, 杜敏, 郭连波, 等. 塑料激光诱导击穿光谱技术快速分类应用研究[J]. *光谱学与光谱分析*, 2017, 37(7): 2205-2209.
- [10] Wang J M, Zheng P C, Liu H D, *et al.* Classification of Chinese tea leaves using laser-induced breakdown spectroscopy combined with the discriminant analysis method [J]. *Analytical Methods*, 2016, 8(15): 3204-3209.
- [11] Rao G F, Huang L, He X W, *et al.* Comparison of tea variety discriminating models based on laser induced breakdown spectroscopy [J]. *Laser & Optoelectronics Progress*, 2018, 55(2): 023002.  
饶刚福, 黄林, 何秀文, 等. 基于激光诱导击穿光谱的茶叶品种识别模型对比[J]. *激光与光电子学进展*, 2018, 55(2): 023002.
- [12] Zhang H Y, Zhu Q B, Huang M, *et al.* Automatic determination of optimal spectral peaks for classification of Chinese tea leaves using laser-induced breakdown spectroscopy [J]. *International Journal of Agricultural and Biological Engineering*, 2018, 11(3): 154-158.
- [13] Jiang F, Qiao X, Zheng H J, *et al.* Grade discrimination of machine-fried Longjing tea based on hyperspectral technology [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(7): 343-348.

- 蒋帆, 乔欣, 郑华军, 等. 基于高光谱分析技术的机炒龙井茶等级识别方法[J]. 农业工程学报, 2011, 27(7): 343-348.
- [14] Li A, Guo S, Wazir N, *et al.* Accuracy enhancement of laser induced breakdown spectra using permittivity and size optimized plasma confinement rings [J]. *Optics Express*, 2017, 25(22): 27559-27569.
- [15] Yang Z H. Research on application of machine learning algorithm in data classification [D]. Taiyuan: North University of China, 2017: 14-15.  
杨志辉. 基于机器学习算法在数据分类中的应用研究[D]. 太原: 中北大学, 2017: 14-15.
- [16] Yang C R. Study of the spectral data processing in laser induced breakdown spectroscopy technology [D]. Beijing: Beijing Jiaotong University, 2014: 20-21.
- 杨崇瑞. 激光诱导击穿光谱数据处理方法研究[D]. 北京: 北京交通大学, 2014: 20-21.
- [17] Wang X S, Li A, Wazir N, *et al.* Accuracy enhancement of laser induced breakdown spectroscopy by safely low-power discharge [J]. *Optics Express*, 2018, 26(11): 13973-13984.
- [18] Chen P. Study on principal component analysis and its application in feature extraction [D]. Xi'an: Shaanxi Normal University, 2014: 7-10.  
陈佩. 主成分分析法研究及其在特征提取中的应用[D]. 西安: 陕西师范大学, 2014: 7-10.
- [19] Wang J M, Liao X Y, Zheng P C, *et al.* Classification of Chinese herbal medicine by laser-induced breakdown spectroscopy with principal component analysis and artificial neural network[J]. *Analytical Letters*, 2018, 51(4): 575-586.