

基于激光诱导击穿光谱的茶叶品种识别模型对比

饶刚福, 黄林, 何秀文, 林金龙, 杨晖, 刘木华, 陈添兵, 陈金印, 姚明印

江西省果蔬采后处理关键技术及质量安全协同创新中心, 江西省高校生物光电及应用重点实验室, 江西 南昌 330045

摘要 为了快速识别茶叶品种,提出了激光诱导击穿光谱全光学诊断方法。采集7种茶叶样品在200~480 nm波长范围的激光诱导击穿光谱的全谱数据,分别运用九点平滑和九点平滑/一阶导数方法对光谱进行降噪、消除干扰预处理,再结合主成分分析对预处理后的光谱进行降维。选择判别分析(DA)、径向基函数网络(RBF)和B-P反向传播网络(又称MLP)三种模型对7种茶叶进行品种识别。结果显示:综合九点平滑和一阶导数预处理后,再结合主成分分析降维,可使三种模型对茶叶品种的识别准确率均有一定程度的提高,MLP的识别准确率高于DA和RBF,其训练集识别准确率为99.6%,测试集识别准确率为99.1%。选择合适的激光诱导击穿光谱预处理及模型构建方法,对快速准确识别茶叶品种具有可行性。

关键词 光谱学; 茶叶品种; 快速识别; 激光诱导击穿光谱; 主成分分析; 识别模型

中图分类号 TN249

文献标识码 A

doi: 10.3788/LOP55.023002

Comparison of Tea Variety Discriminating Models Based on Laser Induced Breakdown Spectroscopy

Rao Gangfu, Huang Lin, He Xiuwen, Lin Jinlong, Yang Hui, Liu Muhua,
Chen Tianbing, Chen Jinyin, Yao Mingyin

*Collaborative Innovation Center of Postharvest Key Technology and Quality Safety of Fruits
and Vegetables in Jiangxi Province, Key Laboratory of Optics-Electrics Application of Biomaterials of
Jiangxi Province Higher Education, Nanchang, Jiangxi 330045, China*

Abstract Laser induced breakdown spectroscopy (LIBS) is proposed to identify tea variety rapidly. LIBS spectra of seven kinds of teas are collected at 200-480 nm wavelength. Two approaches in preprocessing spectra are applied to decrease noise and eliminate disturbance. One is nine-point smoothing (NPS), the other is NPS combined with first derivative (FD). Principal component analysis (PCA) is adopted to reduce the dimensions of processed spectra. Three models like discriminant analysis (DA), radical basic function (RBF) and BP-ANN (multi layer perception, MLP) are selected to discriminate the tea variety. The results demonstrate that the recognition accuracy of tea variety is improved while NPS, FD and PCA are utilized according to priority. And the recognition accuracy of MLP is higher than that of DA and RBF. The recognition accuracy of MLP is 99.6% in training set and 99.1% in test set. It is feasible to select suitable LIBS spectra preprocessing and model construction method to identify tea variety.

Key words spectroscopy; tea variety; rapid identification; laser induced breakdown spectroscopy; principal component analysis; discriminating model

OCIS codes 300.6365; 140.3440

收稿日期: 2017-07-28; 收到修改稿日期: 2017-09-03

基金项目: 国家自然科学基金(31460419, 31560482)、江西省科技支撑计划(20151BBG70063)、江西省远航工程计划(20140142)、江西省研究生创新专项资金(YC2017-S177)

作者简介: 饶刚福(1993—),男,硕士研究生,主要从事农产品/食品及其源头质量安全光学绿色检测方面的研究。

E-mail: 1158688677@qq.com

导师简介: 姚明印(1980—),女,博士,副教授,硕士生导师,主要从事农产品/食品及其源头质量安全光学绿色检测方面的研究。E-mail: mingyin800@126.com(通信联系人)

1 引言

中国是茶的故乡和茶文化的发源地^[1],茶叶资源丰富,种类繁多,各地的茶叶具有很强的多样性,导致市面上售卖的茶叶品种难以分辨。一些不法商家为了牟取利益,更是以次充好。因此,对茶叶品种的识别意义重大。

目前,茶叶品种及品质的识别主要是凭借人工从色、香、味、形四个方面进行鉴别,具有较强的个人主观意识,需要丰富的经验,普通饮茶者难以正确识别。因此,对茶叶品种的通用化快速识别具有重要的现实意义。

近年来,越来越多的学者在茶叶品种和品质的快速辨别方面进行了许多有价值的研究。王丽鸯等^[2]采用多元化学指纹图谱结合判别分析(DA)对武夷茶进行了识别研究,回判准确率达100%,交叉验证准确率达97%。蒋帆等^[3]采用高光谱结合支持向量机(SVM)对机炒龙井茶的等级进行了识别,识别准确率达98.3%。刘红弟^[4]采用激光诱导击穿光谱(LIBS)结合DA对龙井绿茶、蒙顶黄芽等6类茶叶进行了判别,测试集识别准确率达98%,预测集识别准确率达95.33%。

以上研究表明茶叶品种的快速识别具有一定可行性,但多数研究只采用单一模型对茶叶进行识别,缺乏多种分类模型对茶叶鉴别精度的比较。为此,本试验采集7种茶叶品种的LIBS^[5-8]全谱数据,在两种光谱数据预处理条件下,研究了DA^[9-13]、径向基函数网络^[14-16](RBF)、B-P反向传播网络(又称为多层感知机MLP)^[17-18]三种模型对茶叶品种的识别准确度。

2 材料与方法

2.1 LIBS 试验装置

LIBS 试验装置示意图如图1所示,采用Vlite-200型Nd:YAG共轴双脉冲激光器作为激光发射光源,中心工作波长1064 nm,脉冲宽度6~8 ns,产生的激光经过45°反射镜垂直透过穿孔反射镜、平凸透镜(焦距 $f=100$ mm)后聚焦在放置于二维旋转平台的样本表面,产生等离子体。等离子体透过穿孔反射镜反射至平凸透镜($f=100$ mm),然后聚焦耦合至光纤,再传输至双通道光谱仪完成光谱的分光与光电转换。DG645数字脉冲发生器(Stanford Research Systems, 美国)为激光器和光谱仪提供精准的外触发信号。本试验采用的LIBS参数为:A路激光能量151 mJ, B路激光能量122 mJ, 两路激光脉

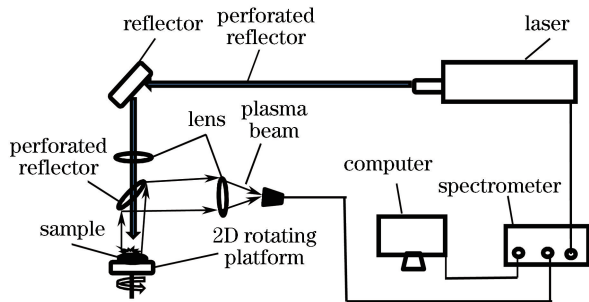


图1 LIBS装置示意图

Fig. 1 Schematic of LIBS setup

冲间延迟55 ns,采集延时 $2 \mu\text{s}$,积分时间2 m。

试验用的7种茶叶均购于南昌市沃尔玛超市,这7种茶叶分别为铁观音、宁红工夫茶(红茶)、普洱茶、峨眉山毛峰、邓村绿茶、信阳毛尖和碧螺春。为了尽可能保证离焦量、光谱稳定性等方面的一致性,在相同的试验条件下,对每种茶叶分别进行粉碎,然后压制成10个厚为3 mm、直径为30 mm的小圆饼,每个小圆饼采集5副光谱,每副光谱均是三次脉冲的平均结果,每种茶叶采集50副光谱,7种茶叶共采集350副光谱。

2.2 茶叶判别模型的构建

试验分别采用DA、RBF、MLP三种模型对7种茶叶进行对比分析识别。

DA采用Fisher判别法。Fisher判别法的基本思路就是投影,针对 P 维空间中的某点寻找一个能使它降为一维数值的线性函数,然后应用这个线性函数把 P 维空间中已知类别归属的样本都变换为一维数据,再根据它们的亲疏程度判定未知归属的样本。

RBF是在借鉴生物局部调节和交叠接受区域知识的基础上提出的一种采用局部接受域来执行函数映射的人工神经网络,具有最优逼近和全局逼近的特性。RBF由一个隐含层和一个输出层组成,隐含层最常用的是高斯径向基函数;输出层最常用的是线性激活函数,其网络规模大,学习速度较快。

MLP是一种最常见的且较为复杂的神经网络模型。该神经网络一般包含组成网络输入层的一组输入,一个或多个具有计算节点的隐层和一个具有计算节点的输出层。MLP采用Sigmoid函数作为激活函数,Sigmoid函数不但具有非线性和单调的特点,还具有无限可微的特点。

3 结果与分析

3.1 光谱数据的提取和处理

噪音信号的干扰会影响LIBS光谱的质量,九

点平滑预处理具有良好的降噪效果。图 2 为铁观音原始 LIBS 光谱和九点平滑处理后的 LIBS 光谱,可以看出,经过九点平滑处理后的噪声信号得到了一定的消除,达到了降噪的目的。此外,从图 2 可知茶叶中含有丰富的 Mg、Fe、Ca 等矿质营养元素以及 C 等非金属元素。通过分析 LIBS 光谱可知,七种茶叶中所含的元素种类基本相同,但各种元素的光谱强度有差异,说明针对茶叶 LIBS 光谱进行分类具有可行性依据。此外,光谱导数校正可以消除光谱中基线的平移、漂移及其他背景的干扰,分辨重叠

峰,提高分辨率和灵敏度。

对采集的茶叶光谱数据进行如下两种预处理: 1) 在九点平滑的基础上结合主成分分析对波长范围为 200~480 nm 的 LIBS 全谱数据进行处理,提取前 9 个主成分,其累积贡献率为 96.5%; 2) 在九点平滑基础上增加一阶导数处理,再结合主成分分析对 LIBS 全谱进行处理,提取前 3 个主成分,其累积贡献率为 96.6%。在累积贡献率相近的情况下,九点平滑/一阶导数光谱预处理方法减少了主成分数,提高了建模效率。

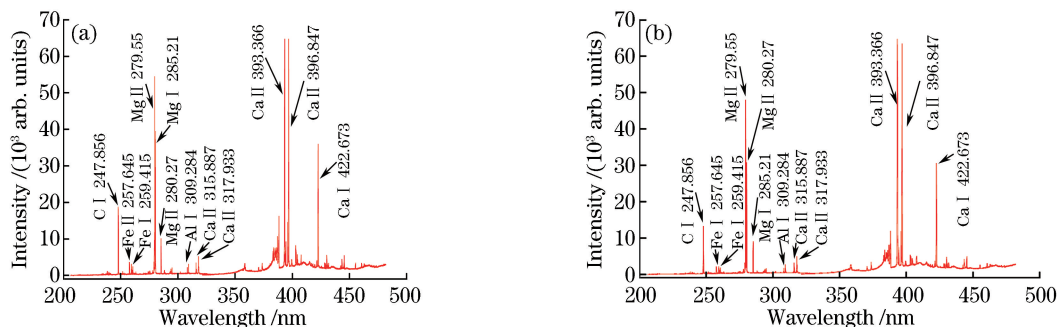


图 2 铁观音样品的 LIBS 光谱。(a)原始光谱;(b)九点平滑预处理后的光谱

Fig. 2 LIBS spectra of Tieguanyin tea sample. (a) Original spectrum; (b) spectrum after nine-point smoothing pretreatment

3.2 判别模型的测定结果

针对以上两种数据预处理方法提取的主成分数,分别采用 DA、RBF 和 MLP 三种方法构建茶叶品种的

判别模型,其中每种茶叶采集 50 副光谱,以第 2 副光谱开始,步长为 3,共 15 副光谱数据作为测试集,其他 35 副光谱数据作为训练集,判别结果如表 1、2 所示

表 1 九点平滑结合主成分分析构建三种模型的识别准确率

Table 1 Classification accuracy of model built by nine-point smoothing combined with principal component analysis

Sample	DA				RBF				MLP			
	Training		Test		Training		Test		Training		Test	
	C	E	C	E	C	E	C	E	C	E	C	E
T-G-Y	35	0	12	3	26	9	9	6	34	1	15	0
N-H-G-F-T	33	2	15	0	30	5	11	4	34	1	14	1
P-E	35	0	15	0	35	0	15	0	35	0	15	0
E-M-S-M-F	35	0	12	3	24	11	10	5	33	2	13	2
D-C-G-T	32	3	12	3	25	10	10	5	32	3	13	2
X-Y-M-J	30	5	10	5	24	11	8	7	31	4	14	1
B-L-C	29	6	12	3	25	10	12	3	31	4	14	1
Accuracy / %	93.5		82.9		77.0		73.0		93.7		93.9	

Notes: C is the number of correct; E is the number of error; T-G-Y is the abbreviation of Tieguanyin; N-H-G-F-T is the abbreviation of Ninghong Gongfu tea; P-E is the abbreviation of Pu'er; E-M-S-M-F is the abbreviation of Emeishan Maofeng; D-C-G-T is the abbreviation of Deng Village Green Tea; X-Y-M-J is the abbreviation of Xinyang Maojian; B-L-C is the abbreviation of Biluochun.

综合表 1 和表 2 可知,LIBS 全谱数据经九点平滑/一阶导数预处理后,相较于单独经九点平滑处理,在结合主成分分析的情况下,可使三种模型对 7 种茶叶品种的识别准确率均有一定程度的提高。在结合主成分分析的情况下,增加一阶导数预处理后,

DA 对 7 种茶叶品种的识别准确率有单一平滑处理的 93.5%(训练集)和 82.9%(测试集)提高到 99.6%和 94.3%;RBF 对 7 种茶叶品种的识别准确率由 77.0%(训练集)和 73.0%(测试集)提高到 86.4%和 80.4%;MLP 对 7 种茶叶品种的识别准确率由

93.7%(训练集)和93.9%(测试集)提高到99.6%和99.1%。这表明,光谱数据经平滑处理后再进行一阶导数预处理,在结合主成分分析的情况下能提高茶叶品种的识别准确率,且MLP的识别准确率最

高。这可能是因为DA只对线性函数具有较高的识别率,除此之外MLP相比较于RBF还具有多个隐藏层,对于数据量大、复杂的多分类问题具有更好的识别效果。

表2 九点平滑/一阶导数处理结合主成分分析构建三种模型的识别准确率

Table 2 Classification accuracy of model built by nine-point smoothing and first derivative combined with principal component analysis

Sample	DA				RBF				MLP			
	Training		Test		Training		Test		Training		Test	
	C	E	C	E	C	E	C	E	C	E	C	E
T-G-Y	35	0	15	0	33	2	12	3	35	0	15	0
N-H-G-F-T	34	1	13	2	33	2	14	1	35	0	14	1
P-E	35	0	15	0	35	0	15	0	35	0	15	0
E-M-S-M-F	35	0	15	0	26	9	9	6	34	1	15	0
D-C-G-T	35	0	15	0	31	4	12	3	35	0	15	0
X-Y-M-J	35	0	11	4	25	10	11	4	35	0	15	0
B-L-C	35	0	15	0	31	4	9	6	35	0	15	0
Accuracy / %	99.6		94.3		86.4		80.4		99.6		99.1	

Notes: C is the number of correct; E is the number of error; T-G-Y is the abbreviation of Tieguanyin; N-H-G-F-T is the abbreviation of Ninghong Gongfu tea; P-E is the abbreviation of Pu'er; E-M-S-M-F is the abbreviation of Emeishan Maofeng; D-C-G-T is the abbreviation of Deng Village Green Tea; X-Y-M-J is the abbreviation of Xinyang Maojian; B-L-C is the abbreviation of Biluochun.

4 结 论

通过采集7种茶叶在200~480 nm波长范围的LIBS全谱信息,采用九点平滑/一阶导数预处理结合主成分分析,运用DA、RBF、MLP三种模型对茶叶品种识别进行建模,结果发现,DA对7种茶叶品种识别的准确率为99.6%(训练集)和94.3%(测试集);RBF的准确率为86.4%(训练集)和80.4%(测试集);MLP的准确率为99.6%(训练集)和99.1%(测试集)。这说明采用一定的数据预处理方法对LIBS全谱数据进行处理后,可以结合合适的模型构建方法实现茶叶品种的快速识别。

参 考 文 献

- [1] 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.
- [2] Wang L Y, Cheng H, Zhou J, *et al.* Discriminant classification of Wuyi Yan Tea based on multiple chemical fingerprint [J]. Journal of Tea Science, 2010, 30(2): 83-88.
王丽鸳, 成浩, 周健, 等. 基于多元化学指纹图谱的武夷岩茶身份判别研究[J]. 茶叶科学, 2010, 30(2): 83-88.
- [3] Jiang F, Qiao X, Zheng H J, *et al.* Grade discrimination of machine-fried Longjing tea based on

hyperspectral technology [J]. Transactions of the CSAE, 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.

- [4] Liu H D. Research on the characteristics and application of laser-induced plasma[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2015.
刘红弟. 激光诱导等离子体的特性与应用研究[D]. 重庆: 重庆邮电大学, 2015.
- [5] Hu L, Zhao N J, Liu W Q, *et al.* Study on removing method of continuous background spectrum in LIBS of multi-element heavy metals in water [J]. Chinese Journal of Lasers, 2014, 41(7): 0715003.
胡丽, 赵南京, 刘文清, 等. 水体重金属多元素LIBS测量连续背景光谱去除方法研究[J]. 中国激光, 2014, 41(7): 0715003.
- [6] Duan W Z, Xu S N, Ning R B, *et al.* Effect of spectral characteristics of Cr and Cu in soil on laser induced breakdown spectroscopy [J]. Laser & Optoelectronics Progress, 2016, 53(2): 023003.
段文钊, 徐送宁, 宁日波, 等. 土壤中Cr、Cu元素谱线特性对LIBS的影响[J]. 激光与光电子学进展, 2016, 53(2): 023003.
- [7] Hu L, Zhao N J, Liu W Q, *et al.* Quantitative

- analysis of Pb in water based on multivariate calibration with LIBS[J]. *Acta Optica Sinica*, 2015, 35(6): 0630001.
- 胡丽, 赵南京, 刘文清, 等. 基于多元校正的水体 Pb 元素 LIBS 定量分析[J]. *光学学报*, 2015, 35(6): 0630001.
- [8] Ghezelbash M, Majd A E, Darbani S M R, *et al.* Experimental investigation of atomic and ionic titanium lines, diatomic TiO_2 , transition and continuum background radiation via magnetically confined LIBS[J]. *Ceramics International*, 2017, 43(11): 8356-8363.
- [9] 张初. 基于光谱与光谱成像技术的油菜病害检测机理与方法研究[D]. 杭州: 浙江大学, 2016.
- [10] Wang P P. Study on the breed discrimination and quality determination of lamb by near infrared spectroscopy [D]. Beijing: Chinese Academy of Agricultural Sciences, 2012.
- 王培培. 基于近红外光谱的不同品种绵羊肉鉴别及品质检测技术研究[D]. 北京: 中国农业科学院, 2012.
- [11] Lu J Z. Plant leaves diseases detection using spectral imaging technology [D]. Hangzhou: Zhejiang University, 2016.
- 卢劲竹. 基于光谱成像技术的植物叶部病害检测研究[D]. 杭州: 浙江大学, 2016.
- [12] Chu X, Wang W, Zhang L D, *et al.* Hyperspectral optimum wavelengths and Fisher discrimination analysis to distinguish different concentrations of aflatoxin on corn kernel surface[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(7): 1811-1815.
- 褚璇, 王伟, 张录达, 等. 高光谱最优波长选择及 Fisher 判别分析法判别玉米颗粒表面黄曲霉毒素[J]. *光谱学与光谱分析*, 2014, 34(7): 1811-1815.
- [13] Ye Y H, Chen Y, Li Y Z, *et al.* Discrimination of nasopharyngeal carcinoma and normal nasopharyngeal cell lines based on confocal Raman microspectroscopy [J]. *Chinese Journal of Lasers*, 2012, 39(5): 0504003.
- 叶宇煌, 陈阳, 李永增, 等. 基于拉曼光谱的鼻咽癌与正常鼻咽细胞株的分类研究[J]. *中国激光*, 2012, 39(5): 0504003.
- [14] Teng W Z, Song J, Meng F X, *et al.* Study on the detection of active ingredient contents of *Paecilomyces hepialid* mycelium via near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2014, 34(10): 2645-2651.
- 滕伟卓, 宋佳, 孟凡欣, 等. 采用近红外漫反射光谱进行蝙蝠蛾拟青霉菌丝体组分检测的研究[J]. *光谱学与光谱分析*, 2014, 34(10): 2645-2651.
- [15] Liao D J. Studies on the test method of camellia oleifera seed quality inspection based on hyperspectral [D]. Changsha: Hunan Agricultural University, 2013.
- 廖敦军. 基于高光谱的油茶籽内部品质检测方法研究[D]. 长沙: 湖南农业大学, 2013.
- [16] Liu Y. Study on new fast methods of Raman spectroscopy for testing counterfeit drugs [D]. Shanghai: The Second Military Medical University, 2012.
- 柳艳. 基于拉曼光谱法的假药快速检测新方法的研究[D]. 上海: 第二军医大学, 2012.
- [17] Hu Z T, Yuan G Y, Hu Y M, *et al.* Training method of neural network based on cubature Kalman filter[J]. *Control and Decision*, 2016, 31(2): 355-360.
- 胡振涛, 袁光耀, 胡玉梅, 等. 基于容积卡尔曼滤波的神经网络训练算法[J]. *控制与决策*, 2016, 31(2): 355-360.
- [18] Li X G, Huang X Q. Deep neural networks based on hyperspectral image classification [J]. *Electronic Measurement Technology*, 2016, 39(7): 81-86.
- 李新国, 黄晓晴. 一种基于 DBN 的高光谱遥感图像分类方法[J]. *电子测量技术*, 2016, 39(7): 81-86.