

# 基于可见-近红外光谱及增强回归树算法的鸡蛋种类鉴别

王彬<sup>1</sup> 王巧华<sup>1,2,3</sup> 肖壮<sup>1</sup> 李理<sup>1</sup> 马逸霄<sup>1</sup> 杨朋<sup>1</sup>

1. 华中农业大学工学院, 武汉 430070; 2. 华中农业大学农业部长江中下游农业装备重点实验室, 武汉 430070;  
3. 华中农业大学国家蛋品加工技术研发分中心, 武汉 430070

**摘要** 利用可见-近红外光谱技术, 选取湖北地区同一品种不同饲养环境下的鸡蛋, 提取鸡蛋的光谱透射率(500~900 nm), 利用标准正态变量变换对光谱数据进行预处理, 结合竞争性自适应重加权与主成分分析对光谱数据进行二次降维, 并将提取的特征信息输入增强回归树算法, 建立鸡蛋土洋种类鉴别模型, 模型的训练集和测试集判别正确率分别为 98.33% 和 97.00%。结果表明, 应用基于可见-近红外光谱及增强回归树方法, 针对同一母鸡品种但不同饲料产出的土洋鸡蛋的种类鉴别是可行的。

**关键词** 鸡蛋; 种类鉴别; 可见-近红外光谱; 增强回归树; 竞争性自适应重加权

**中图分类号** TS 253.7 **文献标识码** A **文章编号** 1000-2421(2018)01-0095-06

鸡蛋以含有丰富的营养物质而广受大众的喜爱。市售鸡蛋大致分为两类: 一般自然放养的或较大范围圈养的鸡产蛋(土鸡蛋)和人工笼养的饲料鸡产蛋(洋鸡蛋)。刘艳芳等<sup>[1]</sup>对土鸡蛋的各种成分进行研究, 发现土鸡蛋相比洋鸡蛋的蛋黄比例高, 蛋清中水分含量少, 胆固醇含量较低, 蛋黄中类胡萝卜素含量明显较高。郭春燕等<sup>[2]</sup>研究不同种类鸡蛋的哈氏单位、蛋重、蛋黄比率等基本指标, 结果发现一些指标上有明显差异。左瑞华等<sup>[3]</sup>研究土鸡蛋和洋鸡蛋的化学指标, 包括内脂肪、蛋白质、胆固醇等, 结果表明土鸡蛋比洋鸡蛋营养价值高。由于土鸡蛋价格高, 在利益的驱使下, 不法商人将饲料鸡蛋充当土鸡蛋, 不仅损害了消费者利益, 也破坏了市场的公平。目前鸡蛋种类鉴别主要通过人工感官, 不仅浪费人力物力, 而且准确率低。因此, 市场迫切需要一种准确快速的鸡蛋种类鉴别无损检测方法。

可见-近红外光谱分析技术广泛用于农产品品种的识别与检测分级<sup>[4]</sup>。刘心如等<sup>[5]</sup>利用可见-近红外漫反射光谱结合主成分-马氏距离聚类判别法对羊毛和羊绒进行鉴别研究; 温珍才等<sup>[6]</sup>利用可见-近红外联合 UVE-PLS-LDA 鉴别压榨和浸出山茶油, 校正集和预测集样本的鉴别正确率均为 100%;

李晓丽等<sup>[7]</sup>利用可见-近红外光谱分析方法结合小波变换、主成分分析和 BP 神经网络的组合方法对稻谷品种进行了模式识别; 徐文杰等<sup>[8]</sup>利用近红外光谱分析方法结合主成分分析和 BP 神经网络对淡水鱼品种进行鉴别; 王梦东等<sup>[9]</sup>主要利用近红外光谱分析方法结合偏最小二乘法及主成分分析建立茶叶中含水量、茶多酚和咖啡碱的定量校正模型, 实现对 3 类茶的定性鉴别和主要内含物含量的预测; 付苗苗等<sup>[10]</sup>利用近红外光谱结合逐步多元线性回归、偏最小二乘回归和主成分回归定量分析水稻秸秆可溶性糖。另外, 国内外学者利用光谱技术在蛋品的检测上也进行了相关研究。段宇飞等<sup>[11]</sup>利用可见-近红外光谱无损检测方法结合 LLE\_SVR 实现对鸡蛋新鲜度精确预测; Abdel-Nour 等<sup>[12]</sup>利用可见-近红外光谱漫透射方法对鸡蛋的新鲜度实现预测; Bamelis 等<sup>[13]</sup>利用可见-近红外光谱漫透射的方法也对鸡蛋的哈夫单位与 PH 值实现预测。上述研究结果表明, 可见-近红外光可以穿透整个鸡蛋, 所得到的可见-近红外光谱带有大量的鸡蛋信息。上述研究中大多集中于鸡蛋内部品质及其他农产品种类鉴别, 对鸡蛋土洋种类鉴别的研究较少, 而消费者及蛋品行业又有鉴别两类鸡蛋的需求, 因此,

收稿日期: 2017-06-28

基金项目: 国家自然科学基金项目(31371771); 湖北省科技支撑计划项目(2015BBA172); 国家科技支撑计划项目(2015BAD19B05); 公益性行业(农业)科研专项(201303084)

王彬, 硕士. 研究方向: 农产品无损检测. E-mail: 399212883@qq.com

通信作者: 王巧华, 博士, 教授. 研究方向: 农畜禽产品无损检测及机电一体化. E-mail: wqh@mail.hzau.edu.cn

本研究通过自行搭建的可见-近红外光谱检测装置,利用可见-近红外透射光谱,对外形相近的肉眼难以分辨的土鸡蛋与洋鸡蛋进行区分,验证了可见-近红外光谱鉴别鸡蛋种类的可行性,为鸡蛋种类无损鉴别提供新的参考方法。

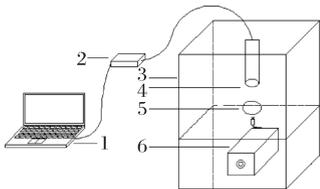
## 1 材料与方 法

### 1.1 试验样本

试验样本选用罗曼粉鸡种,分别选取武汉地区和孝感地区的洋鸡蛋和土鸡蛋,两个地区产洋鸡蛋的母鸡喂养的是同一鸡饲料,产土鸡蛋的母鸡觅食品种均为青草、虫子及谷粒等。4 种鸡蛋样本外形类似,通过外形指标难以区分,蛋质量均在 45~60 g 之间。鸡蛋生产日期相近,均为产后 1~2 d,经随机抽样检测其新鲜度均达到 AA 级。试验前,清除蛋壳表面污渍,剔除破蛋、畸形蛋和裂纹蛋,随机挑选 4 种鸡蛋样本各 100 枚(共 400 个样本)。所有样本按照每种鸡蛋 3:1 的比例选取训练集、预测集,并对所有的试验样本进行编号。

### 1.2 光谱数据采集

试验采用海洋光学公司的 USB2000+ 光纤光谱仪,采集软件为 Spectrasuite 光谱采集软件。采集参数为积分时间 60 ms、平均扫描 5 次、平滑宽度为 3。光源选用可调 50 W 卤素灯,型号为 LS-3000,且内置制冷风扇,可以调节色温稳定,不会使鸡蛋蛋白达到变性的温度,保证光谱采集后鸡蛋的品质安全。先预热光谱仪 30 min,每枚鸡蛋水平放置在光源透过孔上,长轴中心对准光纤探测仪,样本重复扫描 5 次,取 5 次平均后的光谱曲线。光谱采集装置如图 1 所示。



1.计算机 Computer; 2.光纤光谱仪 Fiber optic spectrometer;  
3.光谱采集支架 Spectral acquisition cradle; 4.光纤探测仪 Optical fiber probe; 5.鸡蛋 Egg; 6.光源 Light source.

图 1 鸡蛋光谱采集装置

Fig.1 Egg spectrum collection device

### 1.3 光谱数据预处理

光谱采集过程中不可避免受外界及实验设备影响,需对剔除后的光谱数据进行校正处理,从而消除

随机噪音、基线漂移等影响。分别对移动平均法(moving average method, MA)、标准正态变量(SNV)、Savitzky-Golay 平滑滤波(Savitzky-Golay smoothing filter, SG)、直接差分一阶导(direct difference first order derivative, 1-D)和直接差分二阶导(direct difference first order derivative, 2-D)处理后的光谱数据进行建模分析,选择最优校正处理方法。

### 1.4 建模方法

将提取的特征信息输入增强回归树(boosting regression trees, BRT)算法,建立鸡蛋土洋种类鉴别模型,算法实现过程如下:

1)初始化:训练集为 $(x_i, y_i)$ ,其中 $x$ 为鸡蛋种类特征值, $y$ 对应鸡蛋种类标签,本试验设置土鸡蛋为 1,洋鸡蛋为 2, $i$ 为训练集样本数量。

2)for 所有的回归树  $m=1$  to  $N$  do, 其中  $N$  为设置迭代次数,本试验选用  $N=500$ 。

3)计算当前模型的残差: $f_{im} = y_i - f_{m-1}(x_i)$ , 其中  $x_i$  表示第  $i$  个训练集样本值, $f_{m-1}$  是指所有之前的回归树的总和。

4)用一个指定终端节点数的回归树  $H_m$  拟合残差。

5)对于新的回归树的每个终端节点,计算平均残差。这个平均残差作为所有落在相应终端节点的残差的预测值。

6)将新的回归树添加进当前增强回归树  $f_m(x) = f_{m-1} + \lambda \times H_m$ 。

7)end for。

还需要指定 2 个参数:分割数和收缩参数( $\lambda$ )。本试验选用分割数为 5,收缩参数为 0.01。收缩参数减少或收缩每棵新添加树的影响,以避免模型过拟合<sup>[14]</sup>。

## 2 结果与分析

### 2.1 鸡蛋样本原始数据

图 2 为所有鸡蛋样本的原始透射率光谱曲线,光谱采集波长范围为 500~900 nm。

### 2.2 鸡蛋基本参数

光谱数据采集完后测量鸡蛋的基本参数,包括长轴、短轴、蛋壳厚度和哈夫值。使用精度 0.01 mm 的游标卡尺测量蛋的长轴、短轴和蛋壳厚度,测量 3 次,取平均值。根据农业部标准测量哈夫值( $H_a$ )。利用 SPSS 软件对 4 种鸡蛋样本长轴、短轴、蛋壳厚度进行显著性分析(表 1)。结果表明,3 个基本参数

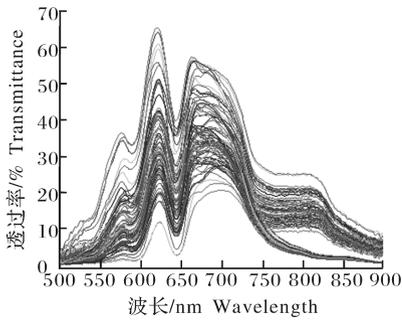


图 2 鸡蛋透射率光谱

Fig.2 Transmittance spectrum of eggs

表 1 4 种鸡蛋样本之间基本参数显著性分析

Table 1 Analysis of the parameters of four egg samples

参数 Parameters	武汉洋鸡蛋 Egg-hen egg of Wuhan	武汉土鸡蛋 Native egg of Wuhan	孝感土鸡蛋 Native egg of Xiaogan	孝感洋鸡蛋 Egg-hen egg of Xiaogan
长轴/mm Long axis	52.581±1.330a	52.722±1.211a	52.664±1.132a	52.783±1.232a
短轴/mm Short axis	40.312±0.983a	40.435±1.129a	40.336±1.147a	40.412±1.136a
蛋壳厚度/mm Eggshell thickness	0.356±0.142a	0.375±0.053a	0.364±0.067a	0.368±0.115a

注：同一行数据字母相同表示相互间无显著差异 ( $P > 0.05$ )，字母不同且小写表示相互间差异显著 ( $P < 0.05$ )。Note: There is no significant difference between the two letters in the same line, and the letters are different and the lower case shows significant difference ( $P < 0.05$ ) between the two letters ( $P > 0.05$ ).

表 2 4 种鸡蛋样本哈夫值数据

Table 2 Haugh unit data of four egg samples

样本 Samples	最大值 Max	最小值 Min	均值 Mean	标准差 SD
武汉洋鸡蛋 Egg-hen egg of Wuhan	90	84	85.24	1.371
武汉土鸡蛋 Native egg of Wuhan	91	86	87.13	1.547
孝感土鸡蛋 Native egg of Xiaogan	90	86	88.19	1.695
孝感洋鸡蛋 Egg-hen egg of Xiaogan	90	85	87.67	1.463

### 2.3 光谱数据初步分析

图 3 为 4 种鸡蛋样本平均光谱图。由于不同样本蛋壳与蛋黄颜色存在差异，在 590~670 nm 之间形成多个波峰波谷<sup>[15]</sup>，为了消除壳色及蛋黄颜色的影响，对比多次数据处理的结果，将此波段剔除。由图 3 可知，在 530~590 nm 波段，洋鸡蛋与土鸡蛋区别较大。在 670~730 nm 波段，土鸡蛋透射率有一个上升波峰，洋鸡蛋则为下降趋势。在 780~830 nm 波段，土鸡蛋透过率均高于洋鸡蛋。特别在近红外波段，土鸡蛋与洋鸡蛋透过率差异性显著。这可能是土鸡蛋与洋鸡蛋内部含氢基团物质含量的不

同所引起的<sup>[16]</sup>。综上所述，区别土洋鸡蛋的敏感波段是 530~590、670~730、780~830 nm。

### 2.4 数据处理

1) 预处理方法选择。为了比较预处理方法对模型的预测效果，通过反向传播神经网络(back propagation neural network, BPNN)建模方法，利用反向传播神经网络对经过 5 种预处理方法处理的数据进行分析。其中 BPNN 将数据分为两类，洋鸡蛋标签为 1，土鸡蛋标签为 2。BPNN 核函数采用 sigmoid 函数，并采用训练集预测集留一法交叉验证的准确率(cross validation accuracy, CVA)、训练集准确率(train accuracy, TRA)、测试集准确率(test accuracy, TEA)来评价模型的好坏。准确率代表模型的精度，其结果如表 3 所示。通过表 3 可以得出，利用标准正态变量变换(SNV)方法建立的模型效果最好。

2) 数据降维。首先通过 CARS 提取光谱特征波长(图 4)。图 4A 显示的是选取变量数与运行次数之间的关系，符合递减规律<sup>[17]</sup>。图 4B 显示的是

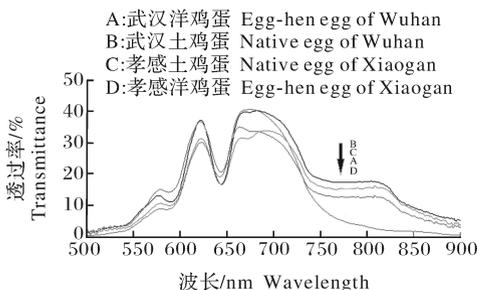


图 3 4 种鸡蛋样本平均光谱

Fig.3 The average spectrum of four egg samples

表 3 不同预处理方法的预测结果

Table 3 Prediction results based on different pretreatment methods

%

预处理 Pretreatment	训练集准确率 TRA	测试集准确率 TEA	交叉验证的准确率 CVA
MA	93.33(280/300)	91.00(91/100)	91.65
SNV	95.00(285/300)	93.00(93/100)	93.38
SG	92.00(276/300)	88.00(88/100)	89.37
1-D	93.67(281/300)	92.00(92/100)	92.12
2-D	93.00(279/300)	86.00(86/100)	88.63

注 Note: TRA: Train accuracy; TEA: Test accuracy; CVA: Cross validation accuracy.

RMSECV 的变化趋势,当 RMSECV 变小时,剔除了无效信息;当 RMSECV 变大时,剔除了有效信息。图 4C 表示的是当运行次数为 21 时(中线位置),得到的 RMSECV 最小。反复进行相关参数的选取调试,最终 CARS 选择的特征波长有 35 个,分别为 535、536、551、552、577、670、690、703、704、743、787、792、794、795、798、806、824、830、831、835、837、840、847、848、851、852、855、858、864、865、871、874、876、886、894。对于 CARS 选取的特征波长,可见光波段数为 10 个,近红外波段数为 25 个,选取的近红外波段占大部分,这与光谱数据初步分析的结果吻合。通过 CARS 方法,实现对数据的降维,前面分析了蛋壳主要是对可见光数据有影响,故在此处理过程中可以较好地剔除蛋壳对光谱数据的影响,提高数据的可靠性。

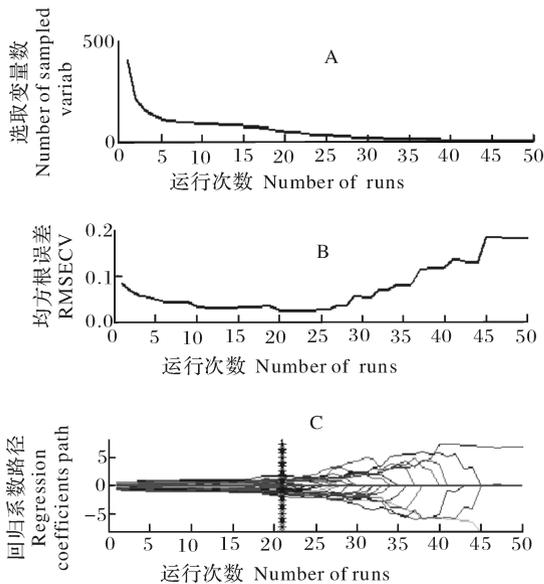


图 4 CARS 工作过程的可视化

Fig.4 Visualization of CARS work process

尽管运用 CARS 已经实现对高维光谱数据的降维,但是选取的特征波长还是较多,所以再运用 PCA 继续进行降维。将 CARS 提取的特征波长信

息导入进 PCA 算法中进行再降维。图 5 为前 2 个主成分对应的贡献率。CARS 数据经过 PCA 处理后维数进一步减小,前 2 个成分的累计贡献率已达到 97.22%,选取前 2 个主成分进行建模,进一步提取有效信息,使最终的模型进一步得到优化。

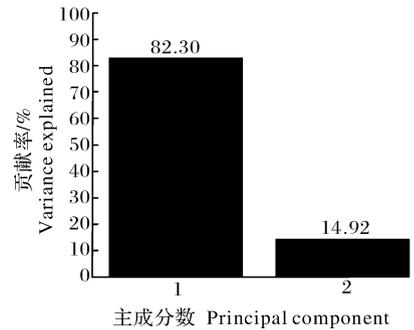


图 5 前 2 个成分对应的贡献率

Fig.5 Cumulative contribution rate of the first two principal components

## 2.5 建模分析

运用 Matlab 编写 BRT 程序,进行 BRT 分析,图 6 为 BRT 迭代 500 次的训练集准确率。当迭代 326 次时,模型已达到最优。最终模型的训练集正确率为 98.33%,交叉验证正确率为 96.54%。再对预测集进行分析,得到预测集正确率为 97%。训练集建模时间为 20.07 s,预测集判别时间为 1.8 s。训练集和预测集样本的结果如表 4 所示。由表 4 可

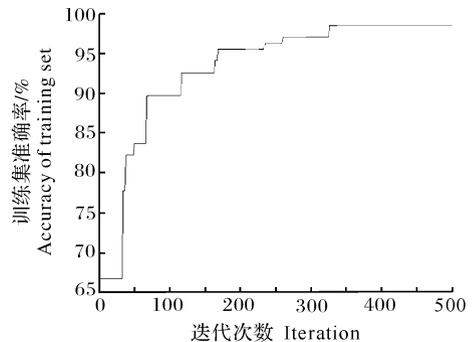


图 6 训练集准确率

Fig.6 Accuracy of training set

知, 样本 A、C、D 的训练集准确率均高于预测集, 但是样本 B 的训练集准确率低于预测集。由训练集准确率、预测集准确率和交叉验证准确率可知模型

既没有欠拟合, 也未过拟合<sup>[18]</sup>, 导致样本 B 的准确率的原因可能是随机取样时预测集样本随机分布, 以及模型需要进一步优化。

表 4 训练集和测试集样本的结果

Table 4 Results of training samples and prediction samples

样本 Samples	训练集 Training samples			预测集 Prediction samples		
	个数 Numbers	判错数 Error numbers	准确率/% Accuracy rate	个数 Numbers	判错数 Error numbers	准确率/% Accuracy rate
武汉洋鸡蛋 Egg-hen egg of Wuhan	75	2	97.33	25	1	96.00
武汉土鸡蛋 Native egg of Wuhan	75	2	97.33	25	0	100.00
孝感土鸡蛋 Native egg of Xiaogan	75	1	98.67	25	1	96.00
孝感洋鸡蛋 Egg-hen egg of Xiaogan	75	0	100.00	25	1	96.00
总计 Total	300	5	98.33	100	3	97.00

### 3 讨 论

光谱数据是高维数据, 直接利用高维数据进行分析, 往往耗时长, 并且部分光谱对结果而言是无效信息。为了进一步优化数据, 需要对光谱数据进行降维, 提取特征信息。对全波段进行特征信息提取时, 主要用到两种方法, 一种是对信息高低空间映射, 一种是选取特征波长。竞争性自适应重加权算法, 其思想是将输入的每个变量都看成是 1 个独立的个体, 通过自适应加权采样技术筛选出模型中回归系数绝对值较大的波长, 去掉权重较小的波长并且通过交互验证优选出模型交互验证均方根误差最小时所对应的波长组合<sup>[19-20]</sup>。因此, 对于光谱数据, 可以通过 CARS 选取特征波长实现降维。主成分分析算法, 其思想是把数据变换到一个新的坐标系中, 使得数据投影的第一大方差在第一主成分上, 第二大方差在第二主成分上, 依次类推。数据集的维数的减少是通过保留低阶主成分, 忽略高阶主成分。低阶成分往往能够保留住数据的最重要方面。增强回归树是基于分类回归树算法的一种自学习方法, 该方法通过自学习以及随机选择方法产生多重回归树, 能够提高模型的稳定性和预测精度<sup>[21]</sup>。在运算过程中, 多次随机取一定量的数据, 分析自变量对因变量的影响程度, 剩余数据用来对拟合结果进行实验, 最后对生成的多重回归取均值并输出。BRT 方法可以使结果的稳定性和精度得到提高, 得出自变量对因变量的相关关系<sup>[22-23]</sup>。

本研究测定了 4 种鸡蛋样本的可见-近红外光谱, 剔除其他影响因素, 提取敏感波段, 利用标准正态变量变换 (SNV) 较好地消除随机噪音、基线漂移等干扰。鸡蛋的光谱信息是高维空间的大量数据, 本研究采用了竞争性自适应重加权 (CARS) 与主成

分分析 (PCA) 方法进行二次降维, 利用 CARS 将全波段降为 35 个特征波长, PCA 将特征波长降为二维特征信息, 得到鸡蛋光谱低维有效特征信息, 剔除无效信息。利用增强回归树 (BRT) 方法建立鉴别模型, 可以避免模型过拟合, 训练集和预测集的判别正确率分别为 98.33%、97.00%, 训练集交叉验证正确率为 96.54%。本研究提出的基于可见-近红外光谱及增强回归树的鸡蛋种类鉴别方法的预测精度高, 可以鉴别出同一品种但不同饲料产出的土洋鸡蛋的种类。

### 参 考 文 献

- [1] 刘艳芳, 孙智达. 土鸡蛋中类胡萝卜素的提取、分离鉴定及其稳定性研究 [D]. 武汉: 华中农业大学, 2010.
- [2] 郭春燕, 杨海明, 王志跃, 等. 不同品种鸡蛋品质的比较研究 [J]. 家禽科学, 2007, 7(2): 14-16.
- [3] 左瑞华, 胡恒龙, 雷翠萍, 等. 六安市售品种鸡蛋与土鸡蛋营养成分的比较分析 [J]. 皖西学院学报, 2007, 23(5): 90-92.
- [4] 付苗苗, 刘梅英, 牛智有. 基于高光谱图像技术的配合饲料主要营养成分的检测方法 [J]. 华中农业大学学报, 2017, 36(2): 123-129.
- [5] 刘心如, 张利平, 王建福, 等. 可见-近红外漫反射光谱技术对羊毛和羊绒的鉴别研究 [J]. 光谱学与光谱分析, 2013, 33(8): 2092-2095.
- [6] 温珍才, 孙通, 耿响, 等. 可见/近红外联合 UVE-PLS-LDA 鉴别压榨和浸出山茶油 [J]. 光谱学与光谱分析, 2013, 33(9): 2354-2358.
- [7] 李晓丽, 唐月明, 何勇, 等. 基于可见/近红外光谱的水稻品种快速鉴别研究 [J]. 光谱学与光谱分析, 2008, 28(3): 578-581.
- [8] 徐文杰, 刘茹, 洪响声, 等. 基于近红外光谱技术的淡水鱼品种快速鉴别 [J]. 农业工程学报, 2014, 30(1): 253-261.
- [9] 王梦东, 王胜鹏. 适用于 3 类茶的定性分类及主要内含成分定量分析的近红外预测模型的建立 [J]. 华中农业大学学报, 2015, 34(1): 123-127.
- [10] 付苗苗, 刘梅英, 牛智有, 等. 基于近红外光谱法的水稻秸秆可溶性糖快速检测 [J]. 华中农业大学学报, 2016, 35(2): 115-121.

- [11] 段宇飞,王巧华,马美湖,等.基于 LLE\_SVR 的鸡蛋新鲜度可见-近红外光谱无损检测方法[J].光谱学与光谱分析,2016,36(4):981-985.
- [12] ABDEL-NOUR N, NQADI M, PRASHER S, et al. Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy[J]. Food and bioprocess technology, 2011, 4(5): 731-736.
- [13] BAMELIS F R, TONA K, DE BAERDEMAEKER J G, et al. Detection of early embryonic development in chicken eggs using visible light transmission[J]. British poultry science, 2002, 43(2): 204-212.
- [14] 于旭华, 冯定远, 王芳. 通过营养调控提高蛋壳质量[J]. 中国家禽, 2003, 25(5): 32-34.
- [15] 秦五昌, 汤修映, 彭彦昆, 等. 基于可见/近红外透射光谱的孵化早期受精鸡蛋的判别[J]. 光谱学与光谱分析, 2017, 37(1): 200-204.
- [16] 王立克, 戴四发, 汪金菊, 等. 不同品种鸡蛋品质及营养成分比较研究[J]. 畜牧与兽医, 2005, 37(7): 33-34.
- [17] 李洪波, 郭志明, 黄文倩, 等. 应用 CARS 和 SPA 算法对草莓 SSC 含量 NIR 光谱预测模型中变量及样本筛选[J]. 光谱学与光谱分析, 2015, 35(2): 372-378.
- [18] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 23-27.
- [19] 李洪波, 郭志明, 黄文倩, 等. 应用 CARS 和 SPA 算法对草莓 SSC 含量 NIR 光谱预测模型中变量及样本筛选[J]. 光谱学与光谱分析, 2015, 35(2): 372-378.
- [20] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica chimica acta, 2009, 648(1): 77-84.
- [21] ELITH J, LEATHWICK J R, HASTIE T. A working guide to boosted regression trees[J]. Journal of animal ecology, 2008, 77(4): 802-813.
- [22] HARA K, CHELLAPPA R. Computationally efficient regression on a dependency graph for human pose estimation[J]. Computer vision and pattern recognition, 2013, 9(4): 3390-3397.
- [23] 曹铭昌, 周广胜, 温恩生. 广义模型及分类回归树在物种分布模拟中的应用与比较[J]. 生态学报, 2005, 25(8): 2031-2040.

## Identifying egg varieties based on boosting regression trees algorithm and visible-near-infrared spectrum

WANG Bin<sup>1</sup> WANG Qiaohua<sup>1,2,3</sup> XIAO Zhuang<sup>1</sup> LI Li<sup>1</sup> MA Yixiao<sup>1</sup> YANG Peng<sup>1</sup>

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;

2. Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River, Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China;

3. National Research and Development Center for Egg Processing, Huazhong Agricultural University, Wuhan 430070, China

**Abstract** The varieties of soil and native eggs relate to its internal quality and sales price. Identifying the egg types quickly and nondestructively will be of great significance to regulate the market of agricultural products. The visible/near-infrared spectrum technology was used to extract the spectral transmittance (500-900 nm) of free-range and ordinary of the same egg variety collected from different breeding environment of Hubei Province. The spectral data were pretreated by the standard normal variate (SNV). The competitive adaptive reweighted sampling (CARS) combined with the principal components analysis (PCA) method was used to perform two times dimensionality reduction of spectral data. The processed data were transmitted as the input of boosting regression trees (BRT) and established the model for identifying egg varieties. The correct rate of the model set and the prediction set are 98.33% and 97.00%. The results showed that applying visible-near infrared spectrum based on boosting regression trees to identify eggs with the same hen breeds but different feeds is feasible.

**Keywords** egg; variety identification; visible-near-infrared spectrum; boosting regression trees; competitive adaptive reweighted sampling

(责任编辑:陆文昌)