

基于近红外光谱与随机青蛙算法的褐变板栗识别

郑 剑^{1,2}, 周 竹^{3,4,5}, 仲山民^{1,2}, 曾松伟^{3,4,5}

(1. 浙江农林大学 农业与食品科学学院, 浙江 临安 311300; 2. 浙江农林大学 浙江省农产品品质改良技术研究重点实验室, 浙江 临安 311300; 3. 浙江农林大学 信息工程学院, 浙江 临安 311300; 4. 浙江农林大学 浙江省林业智能监测与信息技术研究重点实验室, 浙江 临安 311300; 5. 浙江农林大学 智慧农林业研究中心, 浙江 临安 311300)

摘要: 为了实现板栗 *Castanea mollissima* 的快速自动分选, 研究了基于近红外光谱技术的褐变板栗无损检测方法。首先采用 Antaris II 傅里叶变换近红外光谱仪获取 70 个正常板栗和 110 个褐变板栗的近红外光谱(1 000.00~2 500.00 nm), 比较了不同光谱预处理方法对褐变板栗识别的影响, 随后采用一种新的变量选择方法即随机青蛙算法(Random Frog)提取与板栗褐变相关的特征波长变量, 最后基于特征波长建立和比较了褐变板栗识别的偏最小二乘-线性判别分析模型(PLS-LDA)和最小二乘-支持向量机(LS-SVM)模型。结果显示: 经标准正态变量变换(SNV)预处理和随机青蛙算法优选的 23 个特征波长所建 LS-SVM 模型的性能最优, 该模型对测试集的敏感性、特异性和识别正确率分别为 0.92, 1.00 和 95.00%。随机青蛙算法可以有效筛选重要的特征变量, 不仅能简化模型, 而且可以提高识别准确率和识别速度。图 4 表 3 参 22

关键词: 经济林学; 近红外光谱; 褐变; 随机青蛙算法; 最小二乘-支持向量机; 偏最小二乘-线性判别分析; 板栗
中图分类号: S123; S664.2 **文献标志码:** A **文章编号:** 2095-0756(2016)02-0322-08

Chestnut browning detected with near-infrared spectroscopy and a random-frog algorithm

ZHENG Jian^{1,2}, ZHOU Zhu^{3,4,5}, ZHONG Shanmin^{1,2}, ZENG Songwei^{3,4,5}

(1. School of Agriculture and Food Science, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 2. The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 3. School of Information Engineering, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 4. Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 5. Research Center for Smart Agriculture and Forestry, Zhejiang A & F University, Lin'an 311300, Zhejiang, China)

Abstract: To develop a calibration model for rapid, accurate and nondestructive detection of chestnut browning with peeled chestnut *Castanea mollissima* by using near infrared spectroscopy technology. Seventy normal chestnuts and 110 browning chestnuts were prepared, and their diffuse reflectance spectrums were collected in the wavelength range from 1 000.00 to 2 500.00 nm. Spectral pretreatment methods, including standard normal variate (SNV), multiplication scattering correction (MSC), first derivative (FD), second derivative (SD), and detrend, were used and compared first. Then random-frog algorithm was applied to select effective wavelengths (EWs) from the SNV pretreated spectrum. Afterward, a partial least squares-linear discriminant analysis (PLS-LDA) and a least squares-support vector machine (LS-SVM) model were established to classify the browning

收稿日期: 2015-04-16; 修回日期: 2015-09-07

基金项目: 浙江省自然科学基金资助项目(Y3110450, LQ13F050006); 浙江农林大学科研发展基金资助项目(2008FR053, 2012FR085); 浙江省木本粮油产业科技创新团队项目(2011R50030-2)

作者简介: 郑剑, 讲师, 博士研究生, 从事农产品储藏加工与品质检测研究。E-mail: zhengjian622@126.com。

通信作者: 周竹, 讲师, 博士, 从事智能化检测与控制技术研究。E-mail: zhouzhu@zafu.edu.cn

chestnuts based on EWs, and the results were compared based on sensitivity, specificity and accuracy. For the validation set, the sensitivity, specificity and accuracy obtained by EWs-LS-SVM were 0.92, 1.00 and 95.00%, respectively. The results were better than those of full-PLS-LDA model, full-LS-SVM model and EWs-PLS-LDA model. Also, the random-frog algorithm effectively selected important wavelengths and simplified the discrimination model improving precision and recognition speed. The overall results demonstrate that random-frog algorithm is a powerful tool to select the efficient variables, and EWs-LS-SVM is excellent for the spectral calibration. [Ch, 4 fig. 3 tab. 22 ref.]

Key words: cash forestry; NIR spectroscopy; browning; random-frog algorithm; least squares-support vector machines; partial least squares-linear discriminant analysis; chestnut

板栗 *Castanea mollissima* 味道香甜, 营养丰富, 是良好的营养保健品。然而, 板栗在储藏和加工过程中常常会发生果肉褐变, 严重影响产品的品质、风味和营养成分, 降低了产品的经济价值。因此, 快速准确检测褐变板栗对于板栗的储藏与深加工具有重要的意义。近红外光谱技术(NIRS)可利用近红外光谱区域(780~2 500 nm)的全谱或部分波段光谱对被测物进行定量或定性分析。因其快速、无损等优点, 在肉类^[1-2]、果蔬^[3-6]等农产品品质检测中得到了广泛的研究。在板栗品质检测方面, 周竹等^[7-9]、Moscetti 等^[10-11]研究了带壳霉变、虫害板栗的近红外光谱定性检测方法; 刘洁等^[12-13]、傅谊等^[14]建立了板栗中水分、蛋白质等含量的近红外光谱模型。以上研究为板栗褐变的近红外光谱检测奠定了基础。然而, 近红外光谱通常包含 1 000.00~2 000.00 nm 多个波长点的光谱数据, 利用全波段数据建模分析时, 光谱中的大量冗余信息会使建模计算工作量大, 时间花费长。此外, 由于变量间的强相关性, 并非所有波长都能提供有用信息。周竹等^[7-8]前期采用傅里叶变换、主成分分析等方法提取少数特征量建立了霉变板栗识别模型, 但傅里叶系数、主成分得分均难以直观表征正常、褐变板栗光谱。已有的研究表明: 对波长变量进行优选, 一方面可以剔除不相关变量, 在简化模型的同时, 还可提高模型的预测精度以及稳健性^[15]; 另一方面可以根据获取的特征波长设计便携式在线检测仪器, 从而有效降低成本。本研究提出采用近红外光谱技术检测褐变板栗, 拟采用一种最新的变量提取算法——随机青蛙算法(Random-Frog, RF)寻找区分褐变板栗与正常板栗的有效光谱波长, 并采用线性建模方法偏最小二乘-线性判别分析(partial least squares-linear discriminant analysis, PLS-LDA)和非线性判别方法最小二乘-支持向量机(least squares-support vector machine, LS-SVM)建立褐变板栗识别模型, 且与全波段所建模型进行比较, 以期确定最优的板栗褐变检测方法, 为褐变板栗的快速准确识别提供方法与技术支持。

1 材料与方 法

1.1 试验材料

供试板栗品种为浙江省主栽品种毛板栗, 产自浙江省丽水市, 购于农贸批发市场。剔除虫害板栗, 选取大小均匀的板栗仁, 将其标号放置在 4~6 °C 冷库中 2 个月, 随后取出在室温下放置 10 d, 至部分板栗仁内部出现不同程度的褐变, 选取供实验板栗仁 180 个。

1.2 试验仪器

采用 Antaris II 傅里叶变换近红外光谱仪(Thermo Scientific Co., 美国)对板栗果仁进行光谱采集。该仪器使用 InGaAs 检测器, 内置参比背景, 光源为 25 W 卤素灯(仪器标配)。通过漫反射式积分球附件进行光谱采集。光谱扫描范围设置为 1 000.00~2 500.00 nm, 扫描次数为 32 次, 分辨率为 8.0 cm⁻¹。选择栗仁平整的面作为光谱采集面^[8,11], 平面中心以及 4 个方位上等距的区域作为光谱采样点。采集 5 个·样本⁻¹ 标记部位的光谱, 并将其平均光谱作为该样品的原始光谱。

1.3 栗仁褐变评价

光谱采集完成后, 在平行采集面 8.0 mm 的位置切开板栗, 以判断是否褐变及褐变程度。将板栗剖面置于 A4 打印纸上, 采用 Canon 5D3 数码相机(分辨率 5 184 × 3 456)对被切剖面拍照, 拍照时使剖面正对镜头。图片通过 ENVI 4.7(Research System Inc., 美国)选择感兴趣区域进行处理, 计算褐变面积比(褐变面积占整个剖面面积的比例)。

1.4 特征波段选择方法

采用随机青蛙算法(random-frog)进行特征波长选择。随机青蛙算法是一种最新的变量选择方法,由LI等^[16]提出并用于疾病的基因表达数据分析。该方法是一种类似于可逆跳转马尔科夫链蒙特卡洛(reversible jump Markov chain Monte Carlo, RJMCMC)的算法,通过在模型空间中模拟一条服从稳态分布的马尔科夫链,来计算每个变量的被选概率,从而用于评价变量的重要性^[16]。本研究中,将随机青蛙算法与PLS-LDA算法相结合,探寻褐变板栗识别的特征波段,其具体步骤为^[17]: 1)初始化参数: n 为迭代次数,设为10 000; q 为初始变量集 v_0 的变量个数,为1~1 557任意数,默认值为2; θ 为正态分布方差控制因子,默认值为0.3; ω 为比例因子,用于控制候选变量个数,默认值为3; η 是否接受建模结果差于 v_0 的新变量集 V^* 的概率上界,默认值为0.1。2)随机选择包含 q 个变量的初始变量集合为 v_0 , 并定义包含所有1 557个变量的集合为 v 。3)从均值为 q , 标准偏差为 θ_q 的正态分布中产生随机数并就近圆整为 q^* , 分情况构造一个含有 q^* 个变量的候选变量子集 v^* : ①若 $q^*=q$, 则 $v^*=v$; ②若 $q^*<q$, 先用变量集 v_0 建立 PLS-LDA 模型, 得到各变量回归系数, 然后将回归系数绝对值最小的 $q-q^*$ 个变量从 v_0 中去掉, 则剩余的 q^* 个变量构成 v^* ; ③若 $q^*>q$, 先从 $v-v_0$ 中随机选择 $\omega(q^*-q)$ 个变量, 构成新变量集 t , 然后将 t 与 v_0 组合建立 PLS-LDA 模型, 得到每个变量的回归系数, 随后将回归系数绝对值最大的 q^* 个变量挑选出来构成 v^* 。4) v^* 确定后, 要决定是否接受 v^* 。分别计算用 v_0 和 v^* 建立 PLS-LDA 模型的预测误差, 记为 e_0 和 e^* 。若 $e^*\leq e_0$, 接受 v^* 并令 $v_1=v^*$; 否则, 以 $\eta E_{e_0}/E_{e^*}$ 的概率接受 v^* 且令 $v_1=v^*$ 。用 v_1 代替 v_0 , 返回2)进行下一次迭代, 直到完成 n 次迭代。5) n 次迭代后, 得到 n 个变量子集。变量 i 的被选频次记为 n_i , 依公式计算该变量的选择概率 P_i :

$$P_i=n_i/n。 \quad (1)$$

式(1)中: $i=1, 2, \dots, 1 557$ 。变量对模型越重要, 其被选概率越大, 因此, 可将所有变量的被选概率排序, 选出概率较大的一些变量作为特征波长。

1.5 建模方法

采用 PLS-LDA 和 LSSVM 建立栗仁褐变识别模型。PLS-LDA 是偏最小二乘法与线性判别分析方法的有效结合, 该方法先利用 PLS 算法对光谱矩阵和类别属性进行分解, 得到光谱矩阵的得分向量, 然后将得分向量与类别属性做线性判别分析, 建立判别模型^[18]。如利用 n 个 PLS 得分向量建立 PLS-LDA 模型, 则其判别方程为:

$$\alpha_1 s_1 + \alpha_2 s_2 + \alpha_3 s_3 + \dots + \alpha_n s_n = c。 \quad (2)$$

式(2)中: $s=[s_1, s_2, \dots, s_n]$ 为 PLS 得分向量, $\alpha_1, \alpha_2, \dots, \alpha_n$ 是判别系数, c 为常数。

在 PLS-LDA 建模过程中, 采用蒙特卡洛交互检验^[19]来确定最优主成分数。蒙特卡洛交互检验的取样比率为0.75, 模拟次数为2 000。LS-SVM 是对经典支持向量机(SVM)的一种改进, 该方法采用最小二乘线性系统作为损失函数, 以求解一组线性方程代替经典 SVM 中较为复杂的2次规划问题, 从而降低了计算复杂性, 具有学习速度快, 泛化性好等特点^[20]。该方法用非线性映射将输入参数映射到高维特征空间, 在此空间构造最优决策函数, 并基于结构风险最小化原则确定决策函数参数^[21]。该方法中的核函数是高维特征空间的内积, 只要满足 Mercer 条件的函数均可作为核函数, 其常用的4个核函数中, 径向基核函数(RBF)应用最广泛, 故该文采用 RBF 核函数作为 LS-SVM 分类器的核函数, 并以留一交叉验证法结合网格搜索(grid search)策略来确定惩罚因子 c 以及 RBF 核函数参数 g 。

1.6 模型评价

模型的评价指标为敏感性、特异性、预测总正确率。其定义如式(3)~式(5)所示。

$$\text{敏感性} = t_p / (t_p + f_N); \quad (3)$$

$$\text{特异性} = t_N / (t_N + f_P); \quad (4)$$

$$\text{总正确率}(\%) = (t_p + t_N) \times 100\% / (t_p + t_N + f_P + f_N)。 \quad (5)$$

式(3)~式(5)中: t_p, t_N, f_P, f_N 分别表示真阳性、真阴性、假阳性和假阴性。试验中规定阳性和阴性分别代表褐变板栗和正常板栗。敏感性、特异性以及总正确率越高, 说明所建模型性能越好。

所有程序均在 Matlab 2010a (The MathWorks Inc., 美国)上运行实现。

2 结果与分析

2.1 栗仁褐变的评价结果

试验共使用板栗样本 180 个。表 1 给出了试样的形态学参数。从表 1 中可以看出：试样的各项形态学指标均在正常的分布范围内，具有一定的代表性。通过切割观察，最终确认 110 个样本褐变，褐变面积比例范围为 0.16%~49.43%。图 1 为典型样本的剖面图。图 2 为 110 个褐变样本的褐变程度分布。从图 1 和图 2 中可以看出：试验中包含了不同褐变程度的样本且褐变程度范围较广，因此，试样的光谱数据具有很强的代表性。将所有褐变样本依据褐变面积比梯度以 2:1 比例进行划分，而正常栗仁样本则以 2:1 比例随机划分，最终组成校正集的正常、褐变样本数分别为 47 个和 73 个，组成测试集的正常、褐变样本数分别为 23 个和 37 个。

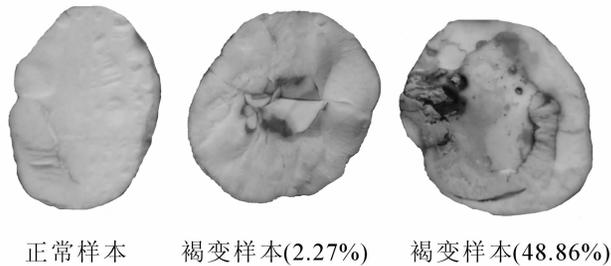


图 1 板栗样本
Figure 1 Chestnut samples

表 1 样本的形态学参数描述

属性	质量/g	长径/mm	短径/mm	厚/mm
范围	9.12~23.77	18.00~36.00	15.50~26.70	13.40~24.50
均值	15.86	28.30	21.32	18.20
标准差	3.34	2.93	2.30	1.78

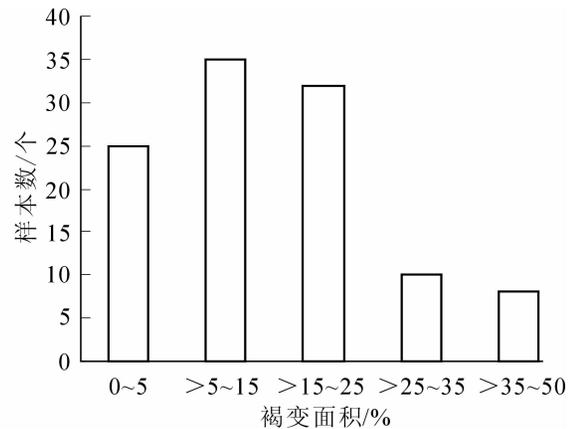


图 2 栗仁褐变样本的褐变程度分布图
Figure 2 Histogram of brown chestnut in various degrees

2.2 栗仁的光谱分析

图 3 为正常栗仁 and 不同程度褐变栗仁的原始光谱。从图 3 中可以看出：栗仁光谱在 1 200.00, 1 450.00 和 1 940.00 nm 处均有明显的吸收峰，在 1 780.00 nm 处略有波峰。在栗仁谱图中，1 200.00 nm 左右区域为甲基和亚甲基中 C—H 键二级倍频的特征吸收谱带，1 450.00 nm 附近区域包含了水分和糖分中 O—H 基团伸缩振动的一级倍频信息，1 940.00 nm 为水分中 O—H 键伸缩振动和弯曲振动的组合频带特征谱带，1 780.00 nm 则分别反映了糖分中 C—H 键的一级倍频和蛋白质中 N—H 键的组合频信息^[22]。从图 3 中还可以看出：不同褐变程度的板栗的原始光谱图趋势基本相似，没有太大的差别，即随着褐变程度的增加，吸收峰高没有呈现出规律性的变化，难以从图谱的表面提取褐变果仁的特征指纹。因此，需要进一步结合化学计量学方法对光谱进行分析处理，以便建立板栗果仁褐变的鉴别模型。

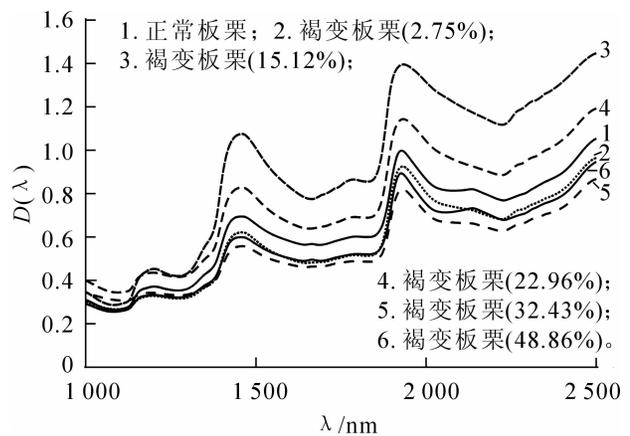


图 3 栗仁原始光谱
Figure 3 Spectra of peeled chestnut

2.3 光谱预处理方法的确定

受仪器、样品以及光谱采集环境等因素的影响，近红外光谱中常出现噪声、谱图基线漂移和平移等现象，为了消除上述因素的影响，应对原始光谱(none)进行预处理。采用标准正态变量变换(SNV)，去

趋势变换(detrend), 一阶导数(FD), 二阶导数(SD), 多元散射校正(MSC)等方法对原始光谱进行预处理并采用 PLS-LDA 建立判别模型, 结果如表 2 所示。由表 2 可知: SNV 预处理建立的 PLS-LDA 模型最优, 该模型对校正集、测试集的敏感性分别为 0.90 以及 0.92, 均明显高于一阶导数、二阶导数, 多元散射校正等光谱预处理方法的模型; 对校正集的特异性均为 1.00, 也高于一阶导数、二阶导数, 多元散射校正等光谱预处理方法的模型; 对校正集、测试集样本识别的准确率分别为 94.13%和 95.00%, 高于其他几种光谱预处理方法。这表明 SNV 预处理有利于减弱因受仪器、样品以及光谱采集环境等因素引起的噪声干扰, 可以增强模型的预测能力和稳健性。因此, 在下文均采用 SNV 处理后的光谱进行分析。

表 2 不同光谱预处理下全波段 PLS-LDA 模型的结果

Table 2 Classification results based on PLS-LDA with different preprocessing methods and full wavelengths

预处理方法	因子数	校正集			测试集		
		敏感性	特异性	正确率/%	敏感性	特异性	正确率/%
none	16	0.89	1.00	93.33	0.89	1.00	93.33
MSC	6	0.85	0.98	90.00	0.86	1.00	91.67
SNV	16	0.90	1.00	94.17	0.92	1.00	95.00
FD	6	0.86	0.98	90.83	0.86	1.00	91.67
SD	4	0.88	0.98	91.67	0.89	1.00	93.33
detrend	15	0.89	1.00	93.33	0.89	1.00	93.33

2.4 特征波长的选择

采用随机青蛙算法(random-frog)对全波段光谱进行特征变量选择。由于随机青蛙算法是基于蒙特卡罗(MC)思想, 每次运行的结果略有差异。为了减少随机因素的影响, 需多次运行, 并对结果进行统计。该研究共运行随机青蛙算法 100 次, 以 100 次运行结果的平均值作为最终特征波长选择的依据, 其结果如图 4 所示。由图 4 可知: 大部分变量的被选择概率都比较小, 只有小部分变量的概率比较大。将所有变量的被选概率进行排序, 以 0.1 为阈值, 选择出概率最大的前 23 个变量作为特征波长, 依次为 1 078.06, 1 028.45, 1 029.68, 1 081.21, 1 077.61, 1 183.90, 1 029.27, 1 020.36, 1 080.76, 1 145.20, 1 183.36, 1 124.34, 1 119.97, 1 032.96, 1 121.42, 1 032.96, 1 098.62, 1 028.05, 1 162.66, 1 120.94, 1 145.71 和 1 182.82 nm, 如图 3 所示。在 1 020.00~1 200.00 nm 波段范围内, 1 020.00 nm 附近为蛋白质中 N—H 键伸缩振动的一级倍频带, 1 030.00, 1 060.00 nm 附近为胺类中 N—H 键伸缩振动的二级倍频带, 1 080.00 nm 为芳香烃中 C—H, C—C 键伸缩振动的一级倍频带, 1 143.00

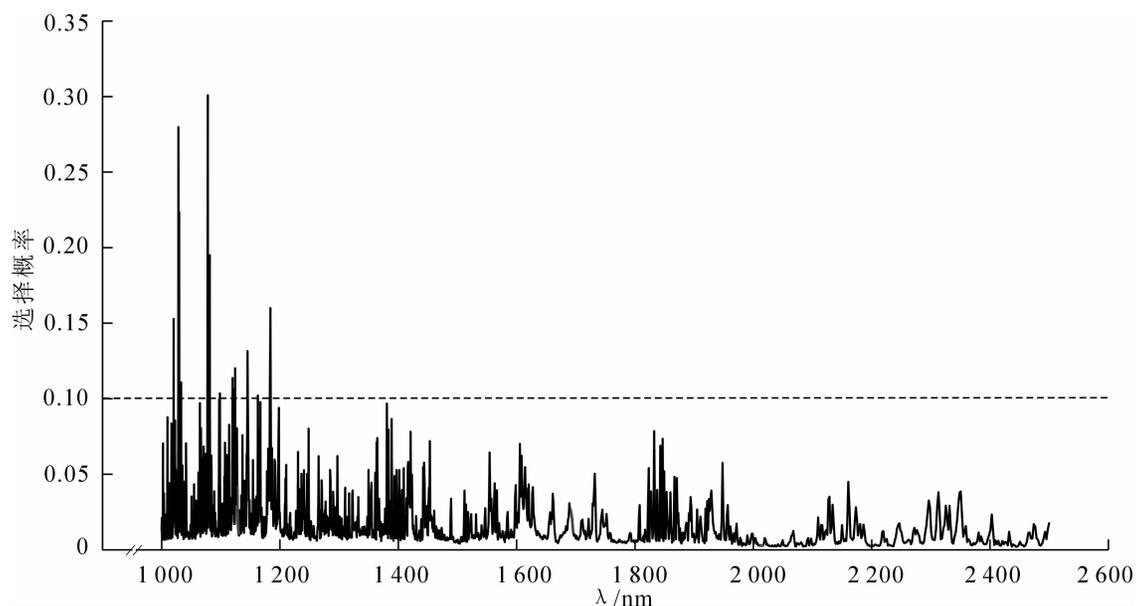


图 4 100 次随机青蛙算法的平均概率

Figure 4 Selection probability of each wavelength averaged over 100 runs of random-frog algorithm

nm 为芳烃中 C—H 键伸缩振动的二级倍频带, 1 152.00, 1 195.00 nm 为甲基中 C—H 键伸缩振动的二级倍频带^[22]。与板栗果仁褐变直接相关的物质主要是有机物酚、酸、醌、蛋白质、氨基酸等。这些物质都与 C—H 键、N—H 键以及 O—H 键相关。因此, 所选特征波长可以反映板栗栗仁褐变特性。从图 3 中还可以看出: 所选特征波段避开了水分的强吸收峰, 减少了水分因素对栗仁褐变识别的干扰。

2.5 模型的建立与比较

将得到的 23 个特征波长(effective wavelengths, EWs)下的光谱数据作为输入, 分别采用 PLS-LDA 以及 LS-SVM 建立褐变板栗的识别模型。表 3 给出了采用 23 个特征波长的光谱数据所建模型的结果。为了突出特征波长所建模型的优越性, 表 3 还给出了采用全波段(full wavelengths, FWs)光谱所建模型的结果。从表 3 中可以看出: 23 个特征波长数据所建 EWs-PLS-LDA 模型对校正集的敏感性(0.92)高于 FWs-PLS-LDA 模型的同类指标(0.90), 但特异性(0.98)低于 FWs-PLS-LDA 模型的同类指标(1.00), 识别正确率相同, 均为 94.17%; 当模型用于测试集数据时, EWs-PLS-LDA 模型的敏感性(0.89)低于 FWs-PLS-LDA 模型的同类指标(0.92), 识别正确率也降低为 93.33%。FWs-PLS-LDA 模型使用 1 557 个变量, 模型的主因子数为 16, 远多于特征波段模型的 23 个变量和 9 个主因子数, 模型较为复杂。23 个特征波长数据所建 EWs-LS-SVM 模型, 无论是校正集还是测试集, 其敏感性(0.92/0.92), 特异性(1.00/1.00), 识别正确率(95.00%/95.00%)均高于 FWs-LS-SVM 模型的同类指标。该模型与 FWs-PLS-LDA 模型相比, 对校正集的预测结果更优。相比于 EWs-PLS-LDA 模型, EWs-LS-SVM 模型结果略优, 原因可能是 PLS-LDA 只能处理栗仁光谱中的线性相关关系, 而 LS-SVM 则可以利用低维数据中的非线性特性, 从而获得较优的预测结果。

表 3 不同建模方法的结果

Table 3 Results for different methods

建模方法	因子数	校正集				测试集			
		敏感性	特异性	正确率/%	建模时间/s	敏感性	特异性	正确率/%	预测时间/ms
FWs-LS-LDA	16	0.90	1.00	94.17	67.53	0.92	1.00	95.00	18.08
FWs-LS-SVM		0.90	0.98	93.33	9.73×10 ⁻³	0.89	0.96	91.67	25.06
EWs-PLS-LDA	9	0.92	0.98	94.17	15.21	0.89	1.00	93.33	9.78
EWs-LS-SVM		0.92	1.00	95.00	3.09×10 ⁻³	0.92	1.00	95.00	9.55

表 3 还给出了不同建模方法建模和预测所用时间(模型建立环境: inter (R) core (TM) i5-3210 CPU@2.5 GHz RAM 4.00 GB Windows 7 旗舰版 64 Bit)。从表 3 中可以看出: 不管是 PLS-LDA 方法还是 LS-SVM 方法, 采用 23 个特征波长建模时间以及预测时间都要短于全波段模型的相应时间, 其中, 23 个波段光谱数据所建 EWs-LS-SVM 模型时间、预测时间最短, 仅为 3.09 和 9.55 ms。综上所述, 通过随机青蛙算法进行特征波长选择, 并且采用 LS-SVM 建立褐变板栗的识别模型, 不仅降低了模型的复杂度, 而且提高了识别准确率和识别速度。

上述结果表明: 应用近红外光谱技术结合化学计量学方法可以快速准确的对褐变板栗进行识别, 为板栗储藏、深加工等环节中板栗品质的快速无损检测提供了重要的参考, 可进一步研究不同品种, 不同产地等条件下的板栗样本, 从而提高模型的稳定性和适用性, 以便在实际应用中取得满意的结果。

3 结论

采用近红外光谱技术对褐变板栗栗仁进行快速无损检测。确定最优的光谱预处理方法为标准正态变量变换(SNV), 运用随机青蛙算法提取代表光谱有效信息的 23 个特征波长, 并作为输入变量建立褐变栗仁识别 PLS-LDA 和 LS-SVM 模型。23 个特征波长所建 LS-SVM 模型性能最优, 校正集和测试集的敏感性、特异性和识别正确率相同, 分别为 0.92, 1.00 和 95.00%。研究结果表明: 采用随机青蛙算法提取特征波长并采用非线性的 LS-SVM 方法建立栗仁褐变检测模型是可行的, 不仅能简化模型, 提高识别速度, 而且能够提高识别准确率。

4 参考文献

- [1] 黄涛, 李小昱, 彭毅, 等. 基于近红外光谱的淡水鱼新鲜度在线检测方法研究[J]. 光谱学与光谱分析, 2014, **34**(10): 2732 – 2736.
HUANG Tao, LI Xiaoyu, PENG Yi, *et al.* Freshwater fish freshness on-line detection method based on near-infrared spectroscopy [J]. *Spectrosc Spectral Anal*, 2014, **34**(10): 2732 – 2736.
- [2] 汤修映, 牛力钊, 徐杨, 等. 基于可见/近红外光谱技术的牛肉含水率无损检测[J]. 农业工程学报, 2013, **29**(11): 248 – 254.
TANG Xiuying, NIU Lizhao, XU Yang, *et al.* Nondestructive determination of water content in beef using visible/near-infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(11): 248 – 254.
- [3] LI Guanghui, REN Yamei, REN Xiaolin, *et al.* Non-destructive measurement of fracturability and chewiness of apple by FT-NIRS [J]. *J Food Sci Tech*, 2015, **52**(1): 258 – 266.
- [4] LIU Cong, YANG S X, DENG Lie. Determination of internal qualities of Newhall navel oranges based on NIR spectroscopy using machine learning [J]. *J Food Eng*, 2015, **161**: 16 – 23.
- [5] RADY A M, GUYER D E. Evaluation of sugar content in potatoes using NIR reflectance and wavelength selection techniques [J]. *Postharv Biol Technol*, 2015, **103**: 17 – 26.
- [6] 介邓飞, 谢丽娟, 饶秀勤, 等. 近红外光谱变量筛选提高西瓜糖度预测模型精度[J]. 农业工程学报, 2013, **29**(12): 264 – 270.
JIE Dengfei, XIE Lijuan, RAO Xiuqin, *et al.* Improving accuracy of prediction model for soluble solids content of watermelon by variable selection based on near-infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2013, **29**(12): 264 – 270.
- [7] 周竹, 李小昱, 李培武, 等. 基于 GA-LSSVM 和近红外傅里叶变换的霉变板栗识别[J]. 农业工程学报, 2011, **27**(3): 331 – 335.
ZHOU Zhu, LI Xiaoyu, LI Peiwu, *et al.* Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT [J]. *Trans Chin Soc Agric Eng*, 2009, **25**(8): 131 – 135.
- [8] 周竹, 刘洁, 李小昱, 等. 霉变板栗的近红外光谱和神经网络方法判别[J]. 农业机械学报, 2009, **40**(增刊1): 109 – 112.
ZHOU Zhu, LIU Jie, LI Xiaoyu, *et al.* Discrimination of moldy Chinese chestnut based on artificial neural network and near infrared spectra [J]. *Trans Chin Soc Agric Mach*, 2009, **40**(supp 1): 109 – 112.
- [9] LIU Jie, LI Xiaoyu, LI Peiwu, *et al.* Nondestructive detection of moldy chestnut based on near infrared spectroscopy [J]. *Afr J Agric Res*, 2010, **5**(23): 3213 – 3218.
- [10] MOSCETTI R, HAFF R P, SARANWONG S, *et al.* Nondestructive detection of insect infested chestnuts based on NIR Spectroscopy [J]. *Postharvest Biol Technol*, 2014, **87**: 88 – 94.
- [11] MOSCETTI R, MONARCA D, CECCHINI M, *et al.* Detection of mold-damaged chestnuts by near-infrared spectroscopy [J]. *Postharvest Biol Technol*, 2014, **93**: 83 – 90.
- [12] 刘洁, 李小昱, 李培武, 等. 基于近红外光谱的板栗水分检测方法[J]. 农业工程学报, 2010, **26**(2): 338 – 341.
LIU Jie, LI Xiaoyu, LI Peiwu, *et al.* Determination of moisture in chestnuts using near infrared spectroscopy [J]. *Trans Chin Soc Agric Eng*, 2010, **26**(2): 338 – 341.
- [13] LIU Jie, LI Xiaoyu, WANG Wei, *et al.* Measurement of protein content in chestnuts using near infrared spectroscopy [J]. *J Chem Pharm Res*, 2014, **6**(6): 938 – 941.
- [14] 傅谊, 张拥军, 陈华才, 等. 基于偏最小二乘法的板栗近红外光谱分析模型的建立[J]. 食品科技, 2012, **37**(5): 42 – 45.
FU Yi, ZHANG Yongjun, CHEN Huacai, *et al.* Establishment of NIR models for components determination in fresh chestnut [J]. *Food Sci Technol*, 2012, **37**(5): 42 – 45.
- [15] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, **16**(4): 528 – 542.
CHU Xiaoli, YUAN Hongfu, LU Wanzhen. Progress and application of spectral data pretreatment and wavelength se-

- lection methods in NIR analytical technique [J]. *Progr Chem*, 2004, **16**(4): 528 – 542.
- [16] LI Hongdong, XU Qingsong, LIANG Yizeng, *et al.* Random Frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification [J]. *Anal Chim Acta*, 2012, **740**: 20 – 26.
- [17] 李洪东. 广义灰色分析体系建模的基本问题及其模型集群分析研究[D]. 长沙: 中南大学, 2012.
LI Hongdong. *Fundamental Problems of Generalized Grey Analytical Systems and Their Research of Model Population Analysis* [D]. Changsha: Central South University, 2012.
- [18] 周竹, 李小昱, 高海龙, 等. 漫反射和透射光谱检测马铃薯黑心病的比较[J]. 农业工程学报, 2012, **28**(11): 237 – 242.
ZHOU Zhu, LI Xiaoyu, GAO Hailong, *et al.* Comparison of diffuse reflection and transmission mode of visible/near infrared spectroscopy for detecting black heart of potato [J]. *Trans Chin Soc Agric Eng*, 2012, **28**(11): 237 – 242.
- [19] LI Hongdong, LIANG Yizeng, XU Qingsong, *et al.* Model population analysis for variable selection [J]. *J Chemometr*, 2010, **24**(7/8): 418 – 423.
- [20] SUYKENS J A K, de BRABANTER J, LUKAS L, *et al.* Weighted least squares support vector machines: robustness and sparse approximation [J]. *Neurocomputing*, 2002, **48**(1/4): 85 – 105.
- [21] 顾燕萍, 赵文杰, 吴占松. 最小二乘支持向量机的算法研究[J]. 清华大学学报: 自然科学版, 2010, **50**(7): 1063 – 1066.
GU Yanping, ZHAO Wenjie, WU Zhansong. Least squares support vector machine algorithm [J]. *J Tsinghua Univ Sci & Tech*, 2010, **50**(7): 1063 – 1066.
- [22] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005: 31 – 39.