

可见/近红外光谱的葡萄籽油掺伪检测系统

唐云峰^{1,2}, 柴琴琴^{1,2*}, 林双杰^{1,2}, 黄捷^{1,2}, 李玉榕^{1,2}, 王武^{1,2}

1. 福州大学电气工程与自动化学院, 福建 福州 350108

2. 福建省医疗器械和医药技术重点实验室, 福建 福州 350108

摘要 葡萄籽油掺假种类繁多, 手段隐蔽, 成为食品安全检测的重要难点之一, 为规范食用油市场, 提供一种方便、可靠的葡萄籽油品质鉴别方法尤为重要。针对色谱和质谱等传统品质分析方法的耗时、试剂消耗大、专业性强等不足, 以及实现无损分析的近红外光谱仪价格昂贵、操作环境要求高等缺点, 研究设计一套低成本、高准确度的可见/近红外光谱仪检测系统来实现葡萄籽油品质掺假鉴别。首先, 依托 USB6500-Pro 探测器搭建可见/近红外光谱仪硬件平台, 并基于 Qt 设计一套简洁的人机交互界面, 用以实现光谱数据的采集、处理以及葡萄籽油掺假鉴别结果的显示; 其次, 针对硬件和检测环境带来的光谱噪声, 系统采用小波变换滤除噪声, 减小光谱失真; 最后, 考虑到现有的基于机器学习的品质鉴别模型往往依赖已知的油类训练样本集来实现对不同掺假类别油类的预测, 而利益驱使下层出不穷的掺假手段使得新的、未出现在原训练集中的掺假类别样本不断涌现, 现有的品质鉴别方法将很难给出准确的判别结果。因此, 研发的检测系统中设计一种能实现已知和新的掺假油品光谱的鉴别方法, 该方法分为分类和校正两步: 先用建模数据库中的训练集建立极限学习机(ELM)分类器模型, 实现初步掺假类别的分类; 然后再利用自动聚类算法对分类结果进一步校正。若与校正数据集产生一个聚类中心, 则证明分类结果正确且属于建模数据库中的已知掺假类别; 若产生两个聚类中心, 则分类结果不正确, 样本为新掺假类别, 未出现在建模数据库中, 最终得到准确的掺假类别结果。为检验系统性能, 用搭建的可见/近红外光谱仪硬件平台采集了纯葡萄籽油和掺入不同比例的大豆油、玉米油、葵花籽油和调和油的葡萄籽油的 5 类光谱数据, 每一类 30 组共计 150 组数据, 将得到的可见/近红外光谱数据先进行小波阈值法去噪和多元散射校正(MSC)预处理后输入到设计的检测系统中。假定前 4 类作为建模数据库中的已知掺假类别以及第 5 类作为新掺假类别, 先利用 K-S 算法将已知掺假类别的每类样本划分成训练集 20 组和测试集 10 组, 用训练集共 80 组样本建立 ELM 分类模型, 将 40 组测试集输入到 ELM 实现初步判别, 判别结果再进一步聚类分析校正, 只有一个聚类中心, 说明了模型判别准确, 且对已知类别能够 100% 识别; 当 30 组新掺假类别样本输入到 ELM 模型时, 均判别成了纯葡萄籽油, 进一步聚类分析校正, 产生了两个聚类中心点, 说明 ELM 模型误判, 定性判定第 5 类为新掺假类别。实验结果表明, 研发的葡萄籽油掺伪检测系统操作简单、快速, 不仅对已知掺假类别能够 100% 识别, 而且对新掺假类别能够实现定性判别。

关键词 极限学习机; 可见/近红外光谱; 聚类; 葡萄籽油掺假

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)01-0202-07

引言

葡萄籽油是一种重要的食用、保健资源, 长期食用葡萄籽油具有抗衰老、增强免疫力、促进生长发育、消除血清胆固醇等作用^[1]。具有的营养价值和医药价值使其成为一种高

级保健食用油。由于葡萄籽油的市场价格普遍高于大宗食用油, 受利益驱动的不法分子将廉价食用油掺入葡萄籽油^[2]。传统方法进行品质检测时, 从采样到测量整个过程, 工作量较大, 例如采样过程需要包装、记录、运输, 测量前还需破坏性预处理等, 这样的分析过程效率低、周期长, 有时还会因化学试剂的使用引起环境污染^[3]。为了保护消费者的合法

收稿日期: 2018-11-08, 修订日期: 2019-03-09

基金项目: 国家自然科学基金项目(61773124)和福建省科技计划项目(2016Y1002)资助

作者简介: 唐云峰, 1992 年生, 福州大学电气工程与自动化学院硕士研究生 e-mail: 1121421877@qq.com

* 通讯联系人 e-mail: qq.chai@fzu.edu.cn

权益和身体健康,规范食用油市场,现场快速鉴别掺伪的葡萄籽油具有现实意义。

近红外光谱(near infrared spectroscopy, NIR)通过检测含氢基团($-\text{CH}$, $-\text{OH}$, $-\text{NH}$, $-\text{SH}$)的振动组合频与倍频的吸收来实现对目标物的鉴定,是一种高性能的、快速、绿色、高灵敏度的食品无损检测方法,近年来已经被广泛地应用在食品安全的检测中^[4-7]。目前,我国近红外光谱仪主要依赖进口,进口产品高昂的价格和脱节的售后服务阻碍了该技术在我国的普及,严重影响了我国应用技术的进步^[3],实验室通用的傅里叶近红外光谱仪价格昂贵、操作环境要求高,并不适合现场检测操作,研究一款低成本、准确度高的光谱检测系统具有重大的实用价值,因此本文依托 USB6500-Pro 探测器搭建一个可见/近红外光谱检测平台,可用于检测葡萄籽油及掺入其他油类的葡萄籽油光谱数据。由于探测器厂家自带的软件仅作为光谱采集、操作复杂和无法显示分类判别结果,本文基于 Qt 自行设计一套简便、有效的人机交互界面的在线光谱检测软件,用以实现对光谱数据的采集、处理以及葡萄籽油掺假鉴别结果的显示。

在高精度建模方法方面,极限学习机(extreme learning machine, ELM)算法训练速度快、计算量小、具有优异的泛化性能等,近年来广泛应用到近红外光谱的分析中^[8-9]。王玮等^[8]提出酵母菌生长过程 4 个阶段的 CARS-ELM 分类模型,其 10 次运行在训练集和测试集中的平均识别率分别为 98.68% 和 97.37%;郭文川等^[9]采用近红外漫反射光谱,提出的 UVE-SPA-ELM 模型对采收贮藏 10 天内预测集中损伤猕猴桃和无损猕猴桃的总正确识别率为 92.4%。但是 ELM 预测效果取决于训练集对模型训练的好坏,只能预测训练集中的已知的类别,上述应用都缺乏对新类别样本的分类。而现实生活中,受利益驱使下层出不穷的掺假手段使得新的、未出现在原训练集样本中的掺假类别不断涌现,因此亟需一种能够鉴别新掺假类别的方法。针对这个问题,以 ELM 算法与自动聚类算法为核心,本研究小组研制出一种快速真伪判别检测系统,不仅实现已知掺假类别的判别,而且对新掺假类别也能定性分析,具有便携、操作简便、扩展性好、时间短等特点,符合现场检测的要求。

1 实验部分

1.1 样品制备

实验用葵花籽油、玉米油、大豆油、调和油和葡萄籽油,购买于本地的永辉超市。以葡萄籽油为溶剂,每次掺入单一定量的其他食用油(大豆油、葵花籽油、玉米油和调和油)制备总体积均为 200 mL 掺杂葡萄籽油样品,其中每次掺入的某种油品的体积(mL)分别为 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 和 200,充分振荡摇匀,每组各有 30 个样品,得到 30 个纯葡萄籽油样品和 120 个葡萄籽油掺假样品。

1.2 硬件检测平台设计

研发的检测系统的硬件平台主要包含:光源、样品池、

光纤、探测器和计算机。光源采用进口卤钨灯光源 HL-2000,光纤为 600 μm 优质石英光纤,探测器的型号是 USB6500-Pro,具体参数见表 1。

表 1 USB6500-Pro 探测器参数

Table 1 USB6500-Pro detector parameters

性能指标	具体参数
探测器范围	180~1 100 nm
信噪比	500 : 1
线性度	99.8%
入射狭缝	5, 10, 25, 50, 100 或 200 μm 可选
积分时间	0.01 ms~100 s
外型尺寸	85 mm×40 mm×120 mm
质量	500 g

整个搭建好的硬件检测平台如图 1 所示。光谱采集过程如下:首先将葡萄籽油样品装入比色皿,然后给仪器光源和 USB6500-Pro 光谱仪上电,并进行 2 分钟的预热处理以保证光源的光强输出稳定和光谱仪稳定工作,光源通过光纤将光照射进比色皿中,光线透过葡萄籽油样品后通过光纤传递给探测器,最后探测器将采集到的光谱吸收率通过 USB 数据线传送给计算机,然后用设计的光谱软件进行采集。

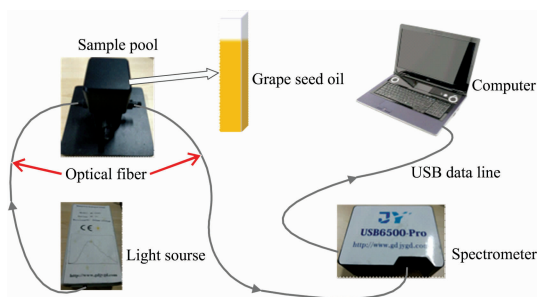


图 1 检测系统硬件平台

Fig. 1 Hardware platform of the detection system

1.3 检测系统软件设计

为了实现对光谱数据的采集,基于 Qt 开发环境设计了一套具有简便的人机交互界面在线光谱检测软件,软件功能有:光谱数据的采集、保存、数据处理及结果显示,如图 2 所示。该软件有光谱数据采集区,能有效采集 USB6500-Pro 光谱仪产生的光谱数据,并对采集到的数据进行显示和保存;此外还有光谱数据处理区,其中数据预处理则根据需要可以选择合适的预处理方法,本文采用 MSC 处理方法,模型在线分析则将采集的光谱数据通过程序中的光谱鉴别设计算法判别出来,最后在软件界面显示光谱分类判别结果。

1.4 光谱鉴别方法设计

光谱鉴别方法是光谱处理软件设计的关键部分。自动聚类算法采用快速搜索密度峰聚类算法^[10](clustering by fast search and find of density peaks, CFSFDP),CFSFDP 算法能够自动确定聚类中心个数,根据聚类中心个数就可以知道类别个数。因此在软件设计部分主要采用 ELM 结合 CFSFDP 算法,简称 ELM-CFSFDP 算法,对 ELM 的预测结果用

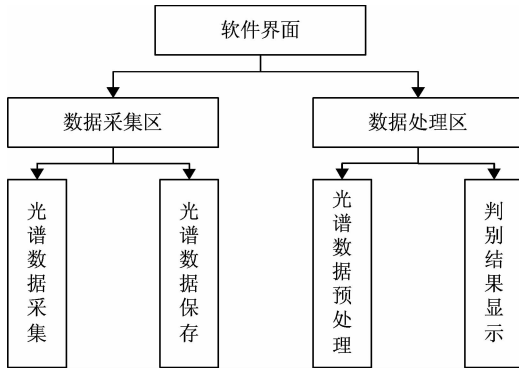


图 2 检测系统光谱软件设计

Fig. 2 Design of spectrum software for detection system

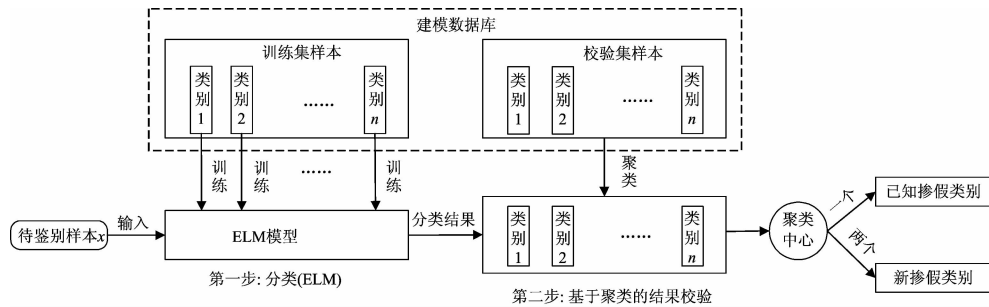


图 3 光谱鉴别方法设计

Fig. 3 Design of spectral identification methods

CFSFDP 算法进行聚类验证, 鉴别方法设计如图 3 所示, 分为两步:

第一步分类(ELM): 在建模数据库中, 利用训练集样本中的已知掺假类别(类别 1—类别 n)对 ELM 模型进行训练, 将待鉴别样本 x 输入到训练好的 ELM 中, 预测结果为类别 $m(m=1, 2, \dots, n)$ 。但是当样本 x 不属于建模数据库的已知掺假类别而是新掺假类别时, ELM 的分类结果为类别 1— n 中的近似或相似类别, 就会出现误判。而在现实生活中检验葡萄籽油掺假时, 事先并不知道待鉴别样本是否属于建模数据库中, 因此需要对分类结果进行校验。

第二步基于聚类(CFSFDP)的结果校验: 假设第一步 ELM 的分类结果为类别 m , 校验集样本中的有相同的类别 m 与之对应, 然后将校验集中类别 m 的样本与待鉴别样本 x 进行 CFSFDP 算法聚类, 当产生一个聚类中心时, 则待鉴别样本 x 属于类别 1— n , 为已知掺假类别, 而聚类中心为两个时, 则待鉴别样本 x 不属于类别 1— n , 为新掺假类别。

1.5 算法理论

1.5.1 极限学习机

极限学习机算法是 2006 年由 Huang 等^[11]提出的一种新型单隐层前馈神经网络, 设任意对于给定的 N 个不同的样本 $\{x_i, t_i | i=1, 2, \dots, N\}$, 其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^n$ 和 $t_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in R^m$, n 和 m 分别表示其维度, w_k 为输入节点与第 k 个隐含层节点的权值, b_k 为第 k 个隐含层节点的输出偏置。单隐层前馈网络的输入输出的数学模型的表示为

$$\sum_{k=1}^L \beta_k G(w_k^T x_i + b_k) = t_i, i = 1, 2, \dots, N \quad (1)$$

式中, L 为隐含层的个数, $G(\cdot)$ 为非线性连续激活函数, $\beta_k = [\beta_{k1}, \beta_{k2}, \dots, \beta_{km}]^T$ 为连接第 k 个隐含层节点与输出层的输出权值向量, 上述方程可以简写成如下形式

$$H\beta = T \quad (2)$$

式中, H 为隐含层输出矩阵, w_k 和 b_k 在模型的训练中是随机取值的, 因此, 方程(2)的最小范数最小二乘解 β 为

$$\hat{\beta} = H^+ T \quad (3)$$

式中, H^+ 为隐含层输出矩阵 H 的 Moore-Penrose Γ^+ 逆。

1.5.2 聚类算法

Alex Rodriguez 于 2014 年在 Science 上提出了一种聚类的新思路^[10], 简称为 CFSFDP 算法, 该算法实现聚类中心自

动确定是基于以下两个假设: (1) 聚类中心点具有较大的局部密度高于附近邻居点密度值, 聚类中心点被具有较低局部密度的邻居点包围, 并且不同的聚类中心点有较大的距离; (2) 噪声点具有较大的距离和相应较小的局部密度。

设聚类的数据集 $J = \{x_i | i=1, 2, \dots, N\}$, 对于要处理的数据 $x_i \in R^q$, q 表示数据点的维度, 每个数据点都需要计算相应的局部密度和最小距离

$$\rho_i = \sum_{j=1}^N e^{-\frac{d_{ij}}{d_c}} \quad (4)$$

其中, d_{ij} 表示不同两个数据点 x_i 与 x_j 间的距离值, 这里取欧氏距离值; 参数 $d_c > 0$ 为截断距离, 为一个超参数, 根据文献[10]描述采用可以使得每个数据点平均邻居点数约占整个数据点总数的 1%~2%。对于任意的数据 x_i , 到具有更高局部密度的其他数据对象的最小距离 δ_i 的公式计算如式(5)

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}) \quad (5)$$

CFSFDP 算法基于两个假设, 从而得出聚类中心点相比于非聚类中心点具有较高的局部密度 ρ 和最小距离 δ 值, 通过计算求得每个数据点的 ρ 和 δ 值, 以 ρ 和 δ 横纵坐标的决策图便能够定性识别聚类中心数, 相比较于其他传统的聚类算法, 它能根据聚类中心自动聚类, 因此也叫自动聚类算法。

2 结果与讨论

用自己搭建的可见/近红外光谱硬件平台采集了纯葡萄籽油和掺入不同比例的大豆油、玉米油、葵花籽油和调和油的葡萄籽油的共计 5 类油样品的光谱前 4 类作为建模数据库中的已知掺假类别, 而第 5 类掺入调和油的葡萄籽油为新掺假类别。K-S 算法是一种基于样本光谱间欧氏距离的定标集划分方法, 通过 K-S 算法将已知掺假类别的每类样本划分成训练集 20 组和测试集 10 组, 同时将已知类别打上类别标签 1~4, 新掺假类别全部作为测试集, 训练集用来实现软件中的鉴别算法, 测试集的数据用来测试光谱处理软件的效果。最后得到样本分配情况如表 2 所示。

表 2 样本分配表

Table 2 The sample allocation table

样本集	训练集/校验集	测试集	标签类别
纯葡萄籽油	20	10	1
掺入大豆油	20	10	2
掺入玉米油	20	10	3
掺入葵花籽油	20	10	4
掺入调和油	—	30	新(未知)

2.1 光谱去噪与预处理

采集的纯葡萄籽油和掺入不同比例大豆油、玉米油、葵花籽油的样品的可见/近红外原始光谱如图 3 所示。从图可以看出, 由于测量过程中存在杂散光、仪器不稳定等因素, 光谱存在大量的毛刺和噪声, 需要去噪处理。

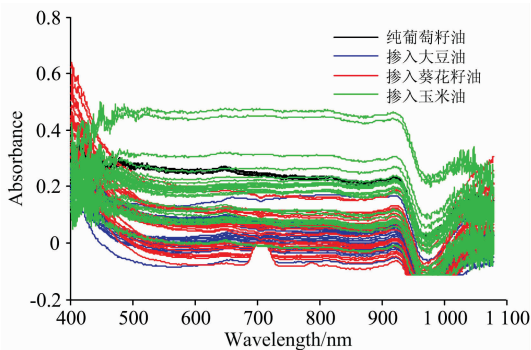


图 4 原始可见/近红外光谱图

Fig. 4 Original visible/near infrared spectra

小波阈值去噪已广泛应用于在光谱去噪中^[12-13], 小波变换去噪时, 参数的不当设定可能误将光谱部分特征信息当作噪音从光谱数据中剔除, 进而影响模型精度。采用信噪比(SNR)与均方根误差(RMSE)作为小波阈值法对光谱数据去噪的评价指标。在一定精度范围要求下, SNR 越大、RMSE 越小, 表明光谱数据中含有的噪声被剔除的越多, 特征信息保留的越充分^[13]。

$$SNR = 10 \ln \left(\frac{\sum_{t=1}^N s^2(t)}{\sum_{t=1}^N (s^2(t) - \hat{f}^2(t))} \right) \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (s^2(t) - \hat{f}^2(t))}{N}} \quad (7)$$

式中, N 为数据的长度, $s(t)$ 为去噪前的含噪声的光谱数据, $\hat{f}(t)$ 为去噪后的光谱数据, 是真实光谱数据 $f(t)$ 的估计值。

由于 dbN 系具有较好的正交紧支性^[14], 实验用掺假 75 mL 其他油类的葡萄油 and 一组纯葡萄籽油共计四组数据, 综合考虑去噪后的光谱平滑度, 最后小波基函数选择 db5 小波、分解层数 5 层, 阈值方案选为“rigsure”, 计算 SNR 和 RMSE, 如表 3 所示。

表 3 去噪后的 SNR 和 RMSE

Table 3 Signal to noise ratio and root mean square error after denoising

掺伪类型(掺伪 75 mL 其他油类)	SNR	RMSE/%
纯葡萄籽油	64.899 4	0.85
掺入大豆油的葡萄籽油	71.224 9	0.17
掺入葵花籽油的葡萄籽油	85.297 5	0.23
掺入玉米油的葡萄籽油	49.478 6	0.62

表中可以看出, 小波变换去噪后, 有较高的 SNR 和较低的 RMSE^[13], 符合后续分类要求。将其应用于所有的可见/近红外光谱数据, 得到去噪后的光谱如图 5 所示。

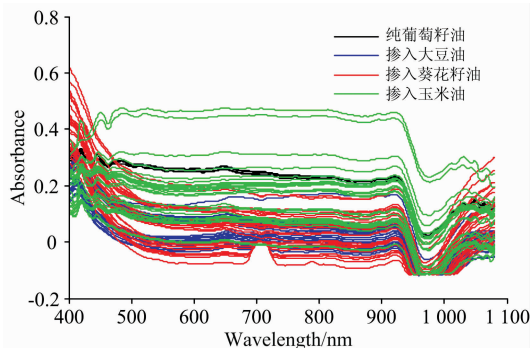


图 5 去噪后的可见/近红外光谱图

Fig. 5 Visible/near infrared spectra after denoising

为了提高数学预测模型的稳定性和鲁棒性, 分类器输入一般都需要进行预处理, MSC 算法能有效的消除散射的影响, 修正基线平移和偏移影响, 增强了与成分含量最相关光谱的吸收度信息, 因此选择 MSC 作为数据预处理。

2.2 已知掺假类别的葡萄籽油鉴别

去噪和预处理后的光谱数据, 按照表 2, 前 4 类(纯葡萄籽油和掺入不同比例的大豆油、玉米油、葵花籽油)作为已知掺假类别, 其中训练集共 80 组数据用来实现鉴别算法, 40 组用来测试系统效果。测试结果如图 6 所示, 可以看出, 利用开发的光谱处理软件可达到 100% 的识别率, 证明软件是

可用的。

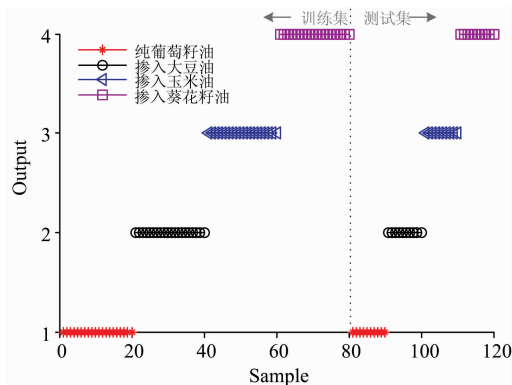


图 6 ELM 预测结果图

Fig. 6 The prediction results of ELM

对于测试用的 40 组数据，事先并不知道它的掺假类别，为检验软件的校正功能。以掺入大豆油的葡萄籽油的测试集 10 组数据进行验证，这些样本数据称为“新样本”。这 10 组“新样本”由图 5 可知，均判别成了“类别 2”，将“类别 2”对应的校验集 20 组数据（来源于类别 2 的训练集）与 10 组“新样本”进行软件中的第二步聚类验证。为了防止一组“新样本”与 20 组校验集样本在聚类过程中被覆盖，采用过采样技术将该组“新样本”扩充至与校验集一样（20 组）。结果每组聚类验证都只产生一个聚类中心点，其中一组聚类中心决策图如图 7 所示。

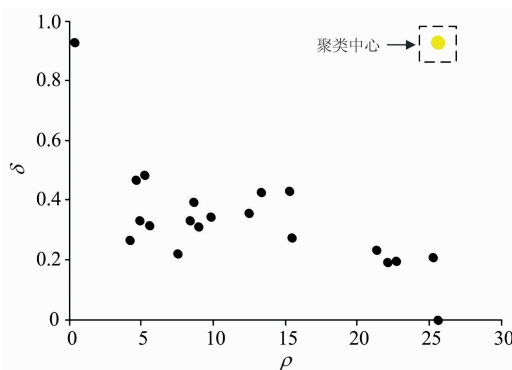


图 7 “新样本”与“类别 2”聚类中心决策图

Fig. 7 “New sample” and “category 2” cluster center decision graph

说明“新样本”与“类别 2”为同一属性的物质，因此最终确定该“新样本”确实为掺入大豆油的葡萄籽油，软件判定结果是一致的。“新样本”的光谱以及其判别结果在光谱处理软件中的显示如图 8。

2.3 新掺假类别的葡萄籽油鉴别

利益驱使下层出不穷的掺假手段使得新的、未出现在原训练集中的掺假类别样本不断涌现，因此往往有许多掺假是新（未知）的，并不在建模数据库中，为了验证算法的正确性，将掺入调和油的葡萄籽油的 30 组样本作为“新样本”进行验证，其原始光谱图如图 9 所示。

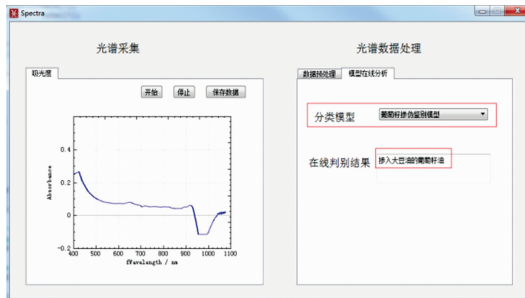


图 8 判别结果显示

Fig. 8 Discrimination result

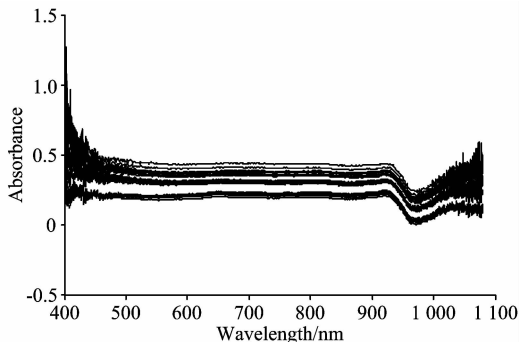


图 9 掺入调和油的葡萄籽油原始光谱图

Fig. 9 Original spectra of grape seed oil blended with blended oil

同样按 2.1 节的方法去噪和预处理后，进行第一步分类，ELM 将 30 组“新样本”均分类为“类别 1”。进一步验证 ELM 分类结果，和上述方法一样，将“新样本”与“类别 1”对应的校验集进行 CFSFDP 聚类得到聚类中心决策图，每组验证结果都产生了两个聚类中心，其中一组决策图如图 10 所示。

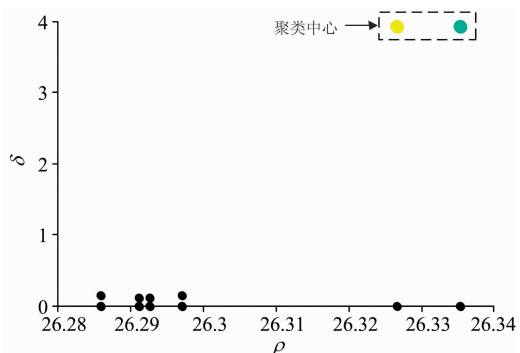


图 10 “新样本”与“类别 1”聚类中心决策图

Fig. 10 “New sample” and “category 1” clustering center decision graph

说明“新样本”与“类别 1”不是同一属性的物质，ELM 模型分类结果是有误，“新样本”不属于“类别 1”，确定了 30 组“新样本”都不属于建模数据库中的已知掺假类别，ELM 由于算法的局限性，只能把新掺假类别分类到相似或相近的已

知掺假类别(“类别 1”)中,但此时样本的聚类中心数为 2,说明样本不属于该类别,最终定性判别掺入调和油的葡萄籽油为新掺假类别。

3 结 论

设计的可见/近红外光谱仪硬件平台结合人机交互界面的光谱处理软件,能够有效地采集并处理葡萄籽油光谱数据。该仪器采集的光谱数据会有部分毛刺和噪声,但通过软件中的小波阈值法去噪和 MSC 预处理后,结合光谱处理算法就能够正确地鉴别葡萄籽油掺伪类别。特别是由 ELM 算

法和 CFSFDP 聚类分析相结合的葡萄籽油掺伪鉴别算法,不需要知道待鉴别样本是否在建模数据库中,就能够判别出葡萄籽油掺伪类别。实验用搭建的可见/近红外光谱采集平台采集了 5 类葡萄籽油掺假光谱数据进行算法验证,结果表明,研发的可见/近红外光谱检测系统的光谱处理软件不仅对已知掺假类别的识别率为 100%,而且可以定性判别出新掺假类别。综上所述,研究设计检测系统能够快速有效地鉴别葡萄籽油掺假,与实验室现有的通用傅里叶近红外光谱仪相比,该可见/近红外光谱检测系统成本低、针对性强、方便携带,可为其他同类光谱仪开发和油品鉴别提供借鉴。在后续研究中,将拓展应用面,以实现仪器的通用性。

References

- [1] Lutterodt H, Slavin M, Whent M, et al. Food Chemistry, 2011, 128(2): 391.
- [2] SHUAI Qian, ZHANG Liang-xiao, LI Pei-wu, et al(帅茜, 张良晓, 李培武, 等). Chinese Journal of Analytical Chemistry(分析化学), 2014, 42(10): 1530.
- [3] YE Hua-jun, XIA A-lin, ZHANG Xue-feng, et al(叶华俊, 夏阿林, 张学锋, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2012, 33(1): 85.
- [4] Zhou Yang, Liu Tiebing, Li Jinrong, et al. Analytical Methods, 2015, 7(6): 2367.
- [5] Wójcicki K, Khmelinskii I, Sikorski M, et al. Food Chemistry, 2015, 187: 416.
- [6] Zhang Liguang, Zhang Xin, Ni Lijun, et al. Food Chemistry, 2014, 145: 342.
- [7] SUN Tong, HU Tian, XU Wen-li, et al(孙通, 胡田, 许文丽, 等). China Oils and Fats(中国油脂), 2013, 38(10): 75.
- [8] WANG Wei, JIANG Hui, LIU Guo-hai, et al(王伟, 江辉, 刘国海, 等). Chinese Journal of Analytical Chemistry(分析化学), 2017, 45(8): 1137.
- [9] GUO Wen-chuan, WANG Ming-hai, GU Jing-si, et al(郭文川, 王铭海, 谷静思, 等). Optics & Precision Engineering(光学精密工程), 2013, 21(10): 2720.
- [10] Alex Rodriguez, Alessandro Laio. Science, 2014, 344(6191): 1492.
- [11] Huang Guangbin, Zhu Qinyu, SIEW Cheekheong. Neurocomputing, 2006, 70(1): 489.
- [12] ZHANG Li-guo, HU Yong-tao, ZHANG Shu-qing, et al(张立国, 胡永涛, 张淑清, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2016, 37(9): 2061.
- [13] LI Ying, LI Yao-xiang, LI Wen-bin, et al(李颖, 李耀翔, 李文彬, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(5): 1384.
- [14] FENG De-shan, YANG Bing-kun, WANG Xun, et al(冯德山, 杨炳坤, 王珣, 等). Chinese Journal of Geophysics(地球物理学报), 2016, 59(1): 342.

Study on Detection System of Grape Seed Oil Adulteration Based on Visible/Near Infrared Spectroscopy

TANG Yun-feng^{1,2}, CHAI Qin-qin^{1,2*}, LIN Shuang-jie^{1,2}, HUANG Jie^{1,2}, LI Yu-rong^{1,2}, WANG Wu^{1,2}

1. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

2. Ministry of Education Key Laboratory of Medical Instrument and Pharmaceutical Technology, Fuzhou University, Fuzhou 350108, China

Abstract Various kinds of adulterated grape seed oil and concealed adulterated means cause a severe problem in food safety detection. In order to regulate the edible oil market, it is especially important to provide a convenient and reliable method for identifying the quality of grape seed oil. However, traditional methods for chromatography and mass spectrometry are time consuming, reagent intensive, highly specialized, etc.; and the near infrared spectrometer that realizes non-destructive analysis is expensive and has high operating environment requirements. Thus, a visible/near infrared spectrometer with low cost and high accuracy was designed to discriminate grape seed oil adulteration. Firstly, a visible/near infrared spectrometer hardware platform

based on USB6500-Pro detector was built, and a simple human-computer interaction interface based on Qt was designed to realize the collection and processing of spectral data and the display of grape seed oil adulteration discrimination results. Secondly, for the spectral noise brought by hardware and detection environment, wavelet transform was used to filter out noise and reduce spectral distortion. Finally, considering that the existing quality discrimination models based on machine learning often rely on the known oil training sample set to predict the different adulterated categories; and driven by interest adulteration means will emerge in endlessly which will result in the emerging of new adulteration categories not in the original training set, the existing quality identification methods are difficult to give accurate results. Therefore, a discrimination method for known and new adulterated oil spectra was designed in the detection system. This method was realized by two steps: (1) classification: the extreme learning machine (ELM) classifier model was established by using the training set in the modeling database to realize the preliminary judgment of the preliminary adulteration category; (2) correction: the automatic clustering algorithm was then used to further correct the prediction result. If a clustering center is generated with the correction data set, it is proved that the prediction result is correct and belongs to the known adulteration category in the modeling database; if two cluster centers are generated, the prediction result is incorrect and the sample is a new adulteration category which does not appear in the modeling database. The result of the accurate adulterated category was eventually obtained. In order to test the performance of the system, five classes of oil, including pure grape seed oil, and grape seed oil blended with different proportions of soybean oil, corn oil, sunflower oil and blend oil were analyzed by the visible/near infrared hardware platform and their spectroscopy data were collected. It contains 30 sets of data for each class of oil, totals 150 sets. Before inputting the visible/near infrared spectroscopy data into the detection system, they were firstly de-noised by wavelet threshold method and pre-processed by multiple scattering correction. Assuming that the first four classes were known adulteration class in the modeling database and the fifth class was new adulteration class, samples from each of the four known adulteration classes were divided into 20 training sets and 10 test sets by using K-S algorithm. Then, ELM classification model was established by using 80 training sets, and 40 test sets were input into ELM for preliminary discrimination. The discrimination results were further analyzed and corrected by clustering. There was one clustering center, which meant that the ELM model discriminated accurately and could recognize 100% of the known classes. However, when 30 samples from the new adulterated class were put into the ELM model, all of them were discriminated as pure grape seed oil. The discrimination results were further clustered and corrected. There were two clustering centers, which showed that the model was misjudged and the fifth class was qualitatively determined as a new adulterated class. The experimental results showed that the designed visible/near infrared spectroscopy detection system was simple and fast, and can identify not only the known adulteration categories but also the new adulteration categories.

Keywords Extreme learning machine; Visible/near infrared spectra; Clustering; Grape seed oil adulteration

(Received Nov. 8, 2018; accepted Mar. 9, 2019)

* Corresponding author