

近红外光谱和特征光谱的山茶油掺假鉴别方法研究

褚璇¹, 王伟², 赵昕¹, 姜洪喆¹, 王伟^{1*}, 刘声泉¹

1. 中国农业大学工学院, 现代农业装备优化设计北京市重点实验室, 北京 100083
2. 塔里木大学机械电气化工程学院, 新疆 阿拉尔 843300

摘要 山茶油素有“东方橄榄油”美誉, 实现掺假山茶油的鉴别具有重要实用价值, 采用近红外光谱技术对掺有葵花油的山茶油进行检测。分别以1%, 5%, 10%为梯度制备掺假比例不同的山茶油样品, 并根据掺假比例将其分为A组(0%~5%)和B组(6%~10%)共11个样品, C组(15%~40%)6个和D组(50%~100%)6个样品。将每个掺假样品充分混匀后再分为9份, 依次采集其1 000~2 500 nm范围的吸收光谱, 共获得207条光谱曲线。每组样品的光谱数据按2:1随机分为训练集与验证集。经去除首尾噪声后, 通过主成分分析法(principal component analysis, PCA)降维, 并利用前四个主成分建立了鉴别山茶油不同掺假等级的主成分-支持向量机判别模型, 训练集与验证集的总体判别准确率分别达96.38%和94.20%;进一步, 通过对前四个主成分的载荷系数的分析, 并结合原始光谱, 提取建模过程中权重较大的波长并解析其化学含义, 最终确定出五个特征波长: 1 212, 1 705, 1 826, 1 905及2 148 nm, 以此波长重新建立近红外特征光谱山茶油掺假等级判别模型, 对训练集与验证集的总体判别准确率也达到了94.20%和92.75%。研究结果表明, 利用近红外光谱和特征光谱均能够较好实现山茶油掺假等级的鉴别, 同时所建立的近红外特征光谱模型也为设计相应的掺假山茶油实用便携式检测仪器提供了理论基础。

关键词 山茶油; 掺假检测; 近红外光谱技术; 特征光谱; 支持向量机

中图分类号: S513 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2017)01-0075-05

引言

山茶油是由油茶籽制得的食用植物油脂, 富含油酸、亚油酸、亚麻酸等不饱和脂肪酸, 营养丰富, 不仅气味清香, 还具有保健功能, 有着十分广阔的市场前景, 售价也远远高于普通植物油^[1]。因此, 屡见一些不法商家在山茶油中掺入其他低价食用油以获取高利润的现象。为了保证山茶油品质, 维护消费者的权益, 实现掺假山茶油的快速鉴别就显得尤为重要。

在理化方法检测山茶油品质中, 可以通过添加相应的化学试剂观察其变色反应, 以实现掺假的定性检测; 也可以利用气相色谱、液相色谱等仪器观察其脂肪酸等成分含量与构成比, 以实现山茶品质的定量检测^[2-3], 这些方法均对样品具有污染性或需要对样品进行稀释、配置等预处理, 难以满足现场检测中快速无损等要求。除这些理化方法外, 许多无

损检测方法也应用在食用油的检测中, 其中振动光谱技术以其可现场直接检测、快速、无损等特点在检测应用中尤为突出。例如, 利用近红外光谱技术鉴别橄榄油产地^[4]、区分混合油组分; 检测菜籽油、大豆油等各类植物油品中的脂肪酸、芥酸等成分含量^[5-7]; 鉴别植物油的优劣等级^[8]等等, 均取得了较好的检测结果。另外, 虽然近红外光谱技术在山茶油掺假方面的检测较少, 但其在定性定量鉴定橄榄油、芝麻油等油品的掺假中已有一定应用^[9-11], 这些均为实现检出掺假山茶油奠定了基础。

本工作旨在利用近红外光谱技术结合主成分分析、支持向量机等化学计量学方法实现对掺入葵花油的山茶油进行判别并判定其掺假程度。

1 实验部分

1.1 样品

收稿日期: 2016-01-13, 修订日期: 2016-05-25

基金项目: 国家科技支撑计划项目(2012BAK08B04)资助

作者简介: 褚璇, 1990年生, 中国农业大学工学院博士研究生 e-mail: cauchx0105@163.com

王伟, 1979年生, 塔里木大学机械电气化工程学院副教授 e-mail: syww77@163.com

褚璇, 王伟: 并列第一作者 *通讯联系人 e-mail: playerwxw@cau.edu.cn

试验所用山茶油与葵花油,均为购于北京某大型超市的品牌纯油。将葵花油以一定质量分数掺入山茶油中,并充分混匀。首先以 1% 为梯度,配置质量分数范围为 0%~10% 的样品 11 个;然后以 5% 为梯度,配置质量分数范围为 15%~40% 的样品 6 个;最后以 10% 为梯度,配置质量分数范围为 50%~100% 的样品 6 个,共获得 23 个不同掺假比例的山茶油样品。然后按照掺入葵花籽油的质量分数将样品分为四个组: A 组(0%~5%), B 组(6%~10%), C 组(15%~40%) 和 D 组(50%~100%)。

1.2 仪器与光谱采集

利用北京爱万提斯科技有限公司提供的 AvaSpec-NIR256 型近红外光谱仪(AvaSpec-NIR256, 荷兰)采集光谱数据。光谱范围为 1 000~2 500 nm,分辨率为 6.4 nm。在采集光谱前,设置光谱仪积分时间为 0.031 9 ms,扫描次数为 100 次;并采集专用白板信息与光谱仪的暗电流以实现校正;最后将每个掺假样品再分为 9 份,置于石英比色皿中并放置在专用比色皿支架上,采集其吸收光谱,光谱数据通过仪器配套软件 AvaSoft 获得。由于光谱仪首尾噪声较大,截取 1 000~2 300 nm 波长范围的光谱数据以初步地去除噪声,图 1 为消除首尾噪声后样品的 207 条原始光谱曲线,可以看出不同掺假比例的山茶油样品光谱曲线十分相似,无法直接从原始光谱曲线上判断其掺假程度,因此需要借助化学计量学方法作进一步分析。

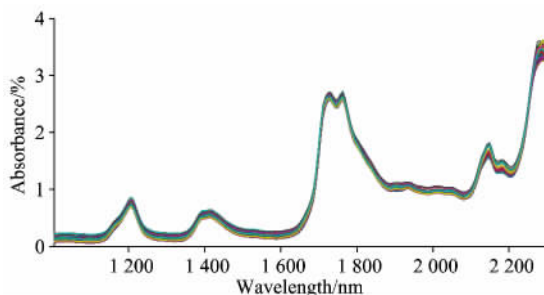


图 1 原始光谱

Fig 1 Original spectra

1.3 支持向量机

采用 Vapnik 提出的支持向量机(support vector machine, SVM)实现不同掺假等级样品的分类,其主要思想是建立分类线或在高维空间建立分类平面,使各类数据之间分类间隔最大。对于线性可分问题,可在原始空间中直接求得简单超平面;而对于线性不可分问题,可通过非线性函数将低维的输入空间数据映射到某个高维空间中,再在此变换空间求最优分类面^[12]。另外,在寻找最优分类面过程中,引入核函数,使其对应某一变换空间中的内积,以避免非线性变换后的“维数灾难”^[13]。

常用的核函数有线性核函数、多项式核函数(polynomial function)、两层感知器函数(sigmoid function)以及径向基核函数(radial basis function)。

2 结果与讨论

对采集的四组样品的光谱数据,每组分别按照 2:1 随机划分,共 138 个样品组成训练集,69 个样品组成验证集。利用 MATLAB 2013b 平台结合 LIBSVM 工具箱对光谱数据进行分析,建立掺假山茶油判别模型,模型通过对不同掺假程度样品的判别准确率来评价。

2.1 主成分选择

在消除原始光谱首尾噪声后,每个样品的光谱数据包含 201 个数据点,数据量大且容易造成信息冗余,因此采用主成分分析法(principal component analysis, PCA)对数据降维,同时利用关键主成分得分来表征样品内部特征和聚类信息^[14-15]。如图 2 为前三个主成分得分图下 138 个训练集样本分布,可以看出,PC₂ 在掺假样品聚类中作用最为显著,能较好区分出掺假质量分数在 50% 以上的样品,而且也能够大致区分开其余三组样品,但是其对边缘区域的区分不是很明显。

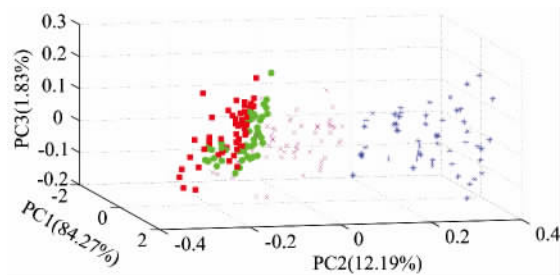


图 2 前三个主成分的得分图

Fig. 2 Score of the first three PCs

将主成分作为支持向量机的输入进一步建立掺假鉴别模型,逐个增加输入主成分的个数,当选取前 4 个主成分作为输入时所建模型有较高准确率,且对照表 1,这些主成分已能够表达 99% 以上的原始光谱信息。因此在后续模型建立过程中,选择前 4 个主成分作为输入。

表 1 主成分累计贡献率

Table 1 Cumulative contribution rate of PCs

主成分	累计贡献率/%	主成分	累计贡献率/%
PC ₁	84.27	PC ₅	99.53
PC ₂	96.46	PC ₆	99.78
PC ₃	98.29	PC ₇	99.87
PC ₄	99.07	PC ₈	99.91

2.2 主成分-支持向量机判别模型建立

在主成分分析的基础上,利用支持向量机建立掺假山茶油的鉴别模型。采用径向基核函数作为支持向量机核函数,利用网格寻优法确定惩罚参数与核函数参数。所建模型对掺假山茶油有较好的鉴别效果,鉴别结果如表 2 所示,训练集与验证集总体判别准确率分别为 96.38% 与 94.20%。其中模型对掺入杂质质量分数较高的(>10%)C 组、D 组识别

准确率达到 100%；对于掺入杂质质量分数较低的($\leq 10\%$) A 组、B 组，虽然对部分样品发生误判，但其单组判别准确率也在 86.67% 以上，并且对照误判样品的掺假质量分数可发现，这些发生误判的样品均是在分组阈值边缘的样品，例如，A 组中误判至 B 组的样品掺假质量分数为 4% 和 5%，B

组中误判至 A 组的样品掺假质量分数为 6%，B 组中误判至 C 组的样品掺假质量分数为 10%。因此，利用前 4 个主成分结合 SVM 建立的鉴别模型能够实现掺假山茶油的鉴别，并且能够较好地地区分出掺假等级。

表 2 主成分-支持向量机判别模型判别准确率
Table 2 Discriminant accuracy of PCA-SVM identification model

实际分组	训练集					验证集				
	A 组	B 组	C 组	D 组	准确率/%	A 组	B 组	C 组	D 组	准确率/%
A 组	34	2	0	0	94.44	16	2	0	0	88.89
B 组	2	27	1	0	90.00	0	13	2	0	86.67
C 组	0	0	36	0	100	0	0	18	0	100
D 组	0	0	0	36	100	0	0	0	18	100
总体准确率/%	96.38					94.20				

2.3 特征波长分析

不同植物油的脂肪酸组成和含量不同，当把其他油品掺入山茶油中时，必然会改变其脂肪酸的组成和含量^[2]。因此可从这些主要成分的分析入手实现掺假山茶油的鉴别。而近红外光谱属于分子振动光谱，许多波段、光谱波峰或波谷等都与样品内部化学成分的官能团有关^[16]，因此找到与这些化学成分相对应的特征波长，对鉴别山茶油是否掺假有重要作用。

首先对所建模型中 4 个主成分的载荷图进行分析，如图 3 所示，每条曲线表示每个主成分中相应波长原变量的系数，系数的绝对值越大，表示对应波长在判别模型中发挥的作用越显著。对这些特征波长与特征波段进行归属分析发现，其均与山茶油中主要化学成分或官能团的结构伸展或振动有关。1 211 nm 波长附近的波峰、1 828 nm 波长附近的波峰分别与甲基或亚甲基中—C—H 结构振动的二阶泛音与—CH₂ 结构振动的一阶泛音有关^[5, 17]；1 903 nm 波长附近的波峰与 C=O 的伸展以及—CO₂H 结构振动的二阶泛音相关；1 705 nm 所在的 1 666~1 818 nm 波段与油或脂肪类物质官能团中结构振动的一阶泛音有关^[18]；2 148 nm 所在的 2 145~2 176 nm 波段与 C—H 和 C=O 结构的振动有关^[19]。

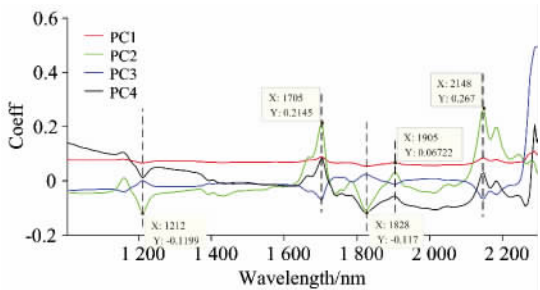


图 3 前四个主成分的载荷系数图
Fig 3 Loading plot of the first four PCs

观察这些波长在原始光谱的位置，如图 4(a)所示，1 212

与 2 148 nm 处于原始光谱中波峰位置；而 1 705, 1 826 以及 1 905 nm 在原始光谱中处于光谱上升沿或下降沿的位置，进一步计算原始光谱一阶导与二阶导，如图 4(b)和图 4(c)所示，1 705 与 1 905 nm 处于一阶导光谱的波峰位置，1 828 nm 处于二阶导光谱的波峰位置。因此为减少变量简化鉴别模型，选择这 5 个波长作为山茶油掺假鉴别的特征波长。

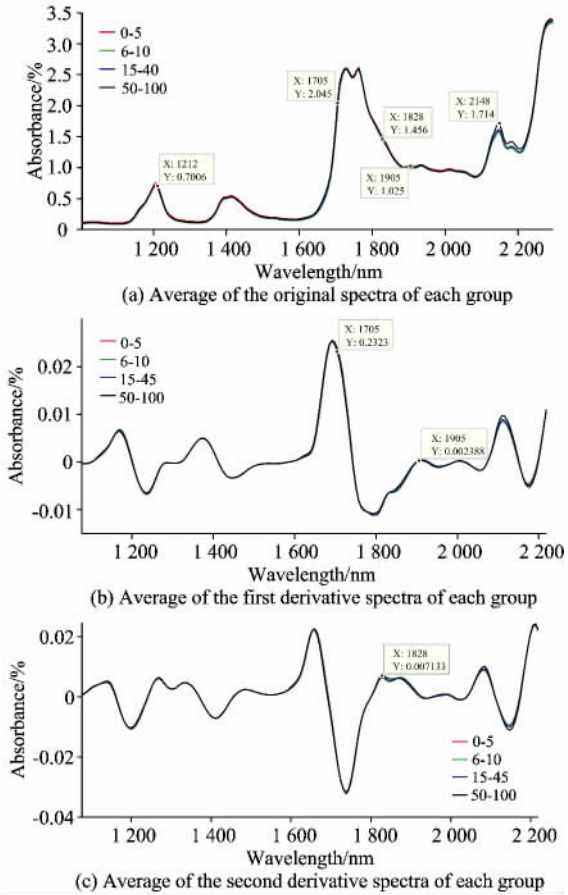


图 4 每组样品平均光谱图
Fig 4 Average spectra of each group

2.4 近红外特征光谱判别模型的建立

为建立近红外特征光谱山茶油掺假判别模型,将所分析出的 5 个特征波长作为支持向量机的输入,与主成分-支持向量机判别模型的建立过程相似,以径向基核函数作为核函数,以网格寻优法确定惩罚参数与核函数参数。所得模型判别效果如表 3 所示。相比于主成分-支持向量机判别模型,特征光谱判别模型所用到的光谱变量大大减少,建模过程得到

简化。其总体鉴别准确率虽有所下降,但也达到了 94.20%与 92.75%,对各组判别情况进行分析,其对掺入杂质质量分数较高的($>10\%$)C 组、D 组识别准确率也能够达到 100%;而在 A, B 两组的误判中,与主成分-支持向量机模型的判别效果相似,均是掺假含量为 4%, 5%, 6%与 10%这些处于阈值附近的样品被误判至相邻组。因此所建立的特征光谱判别模型也能够较好的实现山茶油的掺假判别。

表 3 多光谱判别模型判别准确率

Table 3 Discriminant accuracy of NIR characteristic spectra identification model

实际分组	判别分组									
	训练集					验证集				
	A 组	B 组	C 组	D 组	准确率/%	A 组	B 组	C 组	D 组	准确率/%
A 组	34	2	0	0	94.44	17	1	0	0	94.44
B 组	5	24	1	0	80.00	2	11	2	0	73.33
C 组	0	0	36	0	100	0	0	18	0	100
D 组	0	0	0	36	100	0	0	0	18	100
总体准确率/%	94.20					92.75				

3 结 论

以不同掺假比例的山茶油为研究对象,通过对其 1 000~2 500 nm 下的近红外光谱数据进行分析,结合主成分、特征波长和支持向量机建立判别模型,均实现了山茶油掺假的鉴别。

首先利用光谱数据的前四个主成分结合支持向量机建立了能够鉴别不同掺假等级的模型,对不同等级掺假山茶油的总体判别准确率为 96.38%与 94.20%。另外,对所用前四个主成分的载荷系数分析,主成分中发挥重要作用的波长恰好

对应脂肪、油类等掺假后含量会发生改变的物质,并且这些波长与原始光谱及其差分光谱的吸收峰谷存在一致性。因此,以挑选出的五个特征波长(1 212, 1 705, 1 826, 1 905 及 2 148 nm)作为支持向量机的输入,建立了具有实用意义的近红外特征光谱判别模型,其对掺假等级的总体判别准确率也达到了 94.20%和 92.75%。

因此主成分-支持向量机模型与近红外特征光谱判别模型均对掺假的山茶油有较好的鉴别效果。仅利用五个特征波长,便可实现山茶油掺假的鉴别,这也为无损检测仪器的设计提供了理论基础。

References

- [1] 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.
- [2] XIE Ting-ting, HUANG Li(谢婷婷, 黄 丽). Science and Technology & Innovation(科技与创新), 2014, (21): 9.
- [3] Xie J, Liu T, Yu Y, et al. Journal of the American Oil Chemists' Society, 2013, 90(5): 641.
- [4] Galtier O, Abbas O, Le Dréau Y, et al. Vibrational Spectroscopy, 2011, 55(1): 132.
- [5] Yuan J, Wang C, Chen H, et al. Food Chemistry, 2013, 138(2): 1657.
- [6] Patil A G, Oak M D, Taware S P, et al. Food Chemistry, 2010, 120(4): 1210.
- [7] CHEN Dan, CHEN Bin, LU Dao-li, et al(陈 蛋, 陈 斌, 陆道礼, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2007, 23(1): 234.
- [8] Sinelli N, Cerretani L, Di Egidio V, et al. Food Research International, 2010, 43(1): 369.
- [9] Mendes T O, da Rocha R A, Porto B L S, et al. Food Analytical Methods, 2015, 1.
- [10] Azizian H, Mossoba M M, Fardin-Kia A R, et al. Lipids, 2015, 50(7): 705.
- [11] LIU Yan-de, WAN Chang-lan(刘燕德, 万常澜). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2012, 43(7): 136.
- [12] Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
- [13] Devos O, Downey G, Duponchel L. Food Chemistry, 2014, 148: 124.
- [14] Juneja J. International Review of Financial Analysis, 2012, 24: 48.
- [15] Xing J, Symons S, Shahin M, et al. Biosystem Engineering, 2010, 106(2): 188.
- [16] Yang H, Irudayaraj J, Paradkar M M. Food Chemistry, 2005, 93(1): 25.
- [17] Sinelli N, Cerretani L, Di Egidio V, et al. Food Research International, 2010, 43(1): 369.

- [18] Hourant P, Baeten V, Morales M T, et al. *Applied Spectroscopy*, 2000, 54(8): 1168.
- [19] Inarejos-García A M, Gómez-Alonso S, Fregapane G, et al. *Food Research International*, 2013, 50(1): 250.

Detection of Camellia Oleifera Oil Adulterated with Sunflower Oil with Near Infrared (NIR) Spectroscopy and Characteristic Spectra

CHU Xuan¹, WANG Wei², ZHAO Xin¹, JIANG Hong-zhe¹, WANG Wei^{1*}, LIU Sheng-quan¹

1. Beijing Key Laboratory of Optimized Design for Modern Agricultural Equipment, College of Engineering, China Agricultural University, Beijing 100083, China
2. College of Mechanical and Electronic Engineering, Tarim University, Alar 843300, China

Abstract Camellia oleifera oil has the reputation of “oriental olive oil”; it is important to detect the adulterated camellia oleifera oil. In this paper, NIR spectra were used to detect camellia oleifera oil adulterated with sunflower oil. Camellia oleifera oil adulterated with varying mass fraction of sunflower oil were prepared, i. e., 11 samples in 0%~10% with the gradient of 1%, 6 samples in 15%~40% with the gradient of 5%, 6 samples in 50%~100% with the gradient of 10%, and all the samples were divided into four groups such as A(0%~5%), B(6%~10%), C(15%~40%) and D(50%~100%). A total of 207 absorbance spectra(1 000~2 500 nm) were acquired by sampling 9 times in each adulteration. Calibration set was consist of two-thirds of the spectra data in each group selected randomly, and the validation set was made up of the last spectral data. After removing the noise in both ends of the spectra, principal component analysis(PCA) was used to reduce the dimensionality, then the first four PCs were used to build the support vector machine (SVM) identification model, and the identification accuracies of 96.38% and 94.20% in calibration and validation set were obtained. Furthermore, five characteristic wavelengths (1 212, 1 705, 1 826, 1 905 and 2 148 nm) were selected based on the loading of the PCs, the peaks or troughs of the original spectra and the chemical functional groups they were corresponding to. A NIR simplified SVM identification model was built by them, and the identification accuracies were 94.20% and 92.75%. Overall, both NIR spectroscopy and NIR characteristic spectra can realize the identification of camellia oleifera oil adulterated with sunflower oil, and the characteristic wavelengths, selected in this study, provide a basis for the design of corresponding instrument.

Keywords Camellia oleifera oil; Detection of adulterations; NIR spectroscopy; Characteristic wavelengths; SVM

(Received Jan. 13, 2016; accepted May 25, 2016)

CHU Xuan and WANG Wei: joint first authors

* Corresponding author