

文章编号: 1000-7032(2018)04-0568-05

三维荧光光谱结合 Tchebichef 矩快速鉴别掺伪芝麻油

潘 钊¹, 崔耀耀^{1,2}, 吴希军^{1*}, 刘婷婷¹, 苑媛媛¹

(1. 燕山大学 测试计量技术及仪器河北省重点实验室, 河北 秦皇岛 066004;

2. 燕山大学 信息科学与工程学院, 河北 秦皇岛 066004)

摘要: 应用 FS920 荧光光谱仪测定样品的三维荧光光谱数据, 直接利用 Tchebichef 矩提取三维光谱灰度图的特征信息, 然后对其进行聚类分析, 最后通过逐步回归建立样本中各成分的线性模型。聚类分析能够准确识别掺伪芝麻油, 并正确解析其组成成分, 得到的线性模型相关系数 $R > 0.99$ 。研究表明, Tchebichef 矩能够有效提取光谱的特征信息, 应用于掺伪芝麻油鉴别可获得良好的定性和定量分析结果。

关键词: 三维荧光光谱; Tchebichef 矩; 聚类分析; 定量分析; 掺伪鉴别

中图分类号: O433.4

文献标识码: A

DOI: 10.3788/fgxb20183904.0568

3D Fluorescence Spectra Combined with Tchebichef Moments for Rapid Identification of Doping Sesame Oil

PAN Zhao¹, CUI Yao-yao^{1,2}, WU Xi-jun^{1*}, LIU Ting-ting¹, YUAN Yuan-yuan¹

(1. Key Lab of Measurement Technology and Instrumentation of Hebei Province, Yanshan University, Qinhuangdao 066004, China;

2. College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)

* Corresponding Author, E-mail: wuxijun@ysu.edu.cn

Abstract: The three-dimensional fluorescence spectra of the samples were measured by FS920 fluorescence spectrometer, and the characteristic information of three-dimensional spectral grayscale was extracted directly by Tchebichef moments. And then, the cluster analysis was carried out. Finally, a linear model of each component in the sample was established by the stepwise regression. Clustering analysis can identify doping sesame oil with a high recognition rate and can correctly analyze its constituent components. R -squared of the obtained linear model is greater than 0.99. The results show that Tchebichef moments can effectively extract the characteristic information of the spectrum and can be used to identify the doping sesame oil and obtain good qualitative and quantitative analysis results.

Key words: three-dimensional fluorescence spectroscopy; Tchebichef moments; clustering analysis; quantitative analysis; adulteration identification

1 引 言

“民以食为天, 食以安为先”, 食品安全始终

是人们关注的热点问题。植物油掺杂掺假现象普遍存在, 严重损害了消费者的健康, 然而由于植物油所含成分类似且比例相近, 因此不易鉴别其

收稿日期: 2017-07-24; 修订日期: 2017-10-19

基金项目: 国家自然科学基金(61471312, 11674275, 11601469); 河北省自然科学基金(F2015203072, F2016203282, C2014203212); 燕山大学基础研究专项课题(16LGA008)资助项目

Supported by National Natural Science Foundation of China(61471312, 11674275, 11601469); Natural Science Foundation of Hebei Province(F2015203072, F2016203282, C2014203212); Yanshan University Basic Research Project(16LGA008)

真伪。

荧光光谱分析技术由于其灵敏度和准确度日益提高,是一种重要而有效的检测分析手段。三维荧光光谱具有灵敏度高、选择性好、信息含量丰富、不破坏物质结构等优点,是诸多专家学者在植物油检测研究中的热点选择^[1-2]。荧光光谱的解析和特征信息的有效提取始终是荧光分析中的难点与关键问题。近期有研究者将数字图像处理技术中几何图像矩算法与光谱方法结合,用于对荧光物质的直接定量分析^[3-6],这些方法无需预处理,直接提取目标分析物的重要特征,用于进一步的定性和定量分析,通过图像矩的多分辨能力解决诸如重叠峰、噪声和散射等干扰问题。

形状是最直观的图像信息的表现方式,几何矩中的 Tchebichef 矩对形状信息的描述没有冗余,并且没有传统几何矩中的坐标空间转化以及连续积分近似所带来的误差^[7-8]。本文在没有任何数据预处理的前提下,直接应用 Tchebichef 矩提取荧光光谱数据的特征,结合聚类分析算法实现了掺伪芝麻油的鉴别并通过逐步回归建立了定量模型。实验结果表明 Tchebichef 矩能够有效地提取光谱的特征信息,验证了该方法用于定性和定量分析的可行性。

2 实 验

2.1 材料与仪器

从市场购得不同品牌、同品牌不同批次的芝麻油、玉米油、大豆油、菜籽油和芝麻香精,其中包括芝麻油样本 14 个(Z1 ~ Z14),大豆油样本 4 个(D1 ~ D4),玉米油样本 7 个(Y1 ~ Y7),菜籽油样本 1 个(C1),芝麻香精样本 2 个(J1, J2);在实验室配置芝麻油掺伪样本:菜籽油和芝麻香精按不同体积比配置混合样本 9 个(JC1 ~ JC9),样本分布见表 1。

采用 FS920 荧光光谱仪(英国 Edinburgh Instruments 公司)测量各植物油样本的荧光光谱,激发波长范围设置为 250:10:550 nm,发射波长范围设置为 260:2:750 nm;激发和发射端狭缝宽度为 1.11 mm,对应光谱分辨率为 2 nm。

表 1 植物油样本

Tab.1 Vegetable oil samples

Sample	Brand	Sample	Brand
Z1	Damingyongzhen	Y1	Wannian
Z2	Fulinmen	Y2	Jinlongyu
Z3	Huarui	Y3	Luhua
Z4	Lihong	Y4	Fulinmen
Z5	Mingxuan	Y5	Xiwang
Z6	Fulaiwei	Y6	Duoli
Z7	Jinlongyu	Y7	Fulinmen
Z8	Damingyongzhen	C1	Jinding
Z9	Fulaiwei	JC1	J2: C1 = 1: 9
Z10	Fulinmen	JC2	J2: C1 = 2: 8
Z11	Jinlongyu	JC3	J2: C1 = 3: 7
Z12	Luhua	JC4	J2: C1 = 4: 6
Z13	Mingxuan	JC5	J2: C1 = 5: 5
Z14	Damingyongzhen	JC6	J2: C1 = 6: 4
D1	Fulinmen	JC7	J2: C1 = 7: 3
D2	Shiweitian	JC8	J2: C1 = 8: 2
D3	Fulinmen	JC9	J2: C1 = 9: 1
D4	Huading		

2.2 荧光光谱数据采集

对表 1 中各样本进行荧光光谱扫描,得到相应的荧光光谱数据,样本 JC5 的三维荧光光谱图和等高线图如图 1 所示。荧光光谱中荧光强度记为 $n \times m$ 维的矩阵(n 对应发射波长, m 对应激发

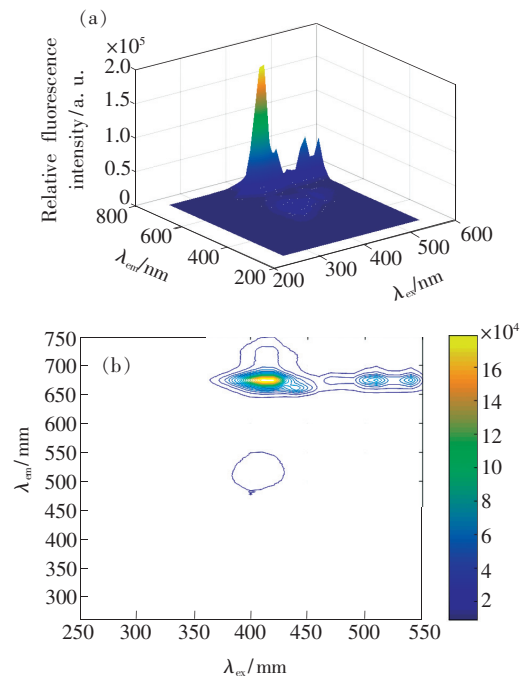


图 1 JC5 的三维荧光光谱图(a)和等高线图(b)

Fig.1 3D fluorescence spectra(a) and contours(b) of JC5

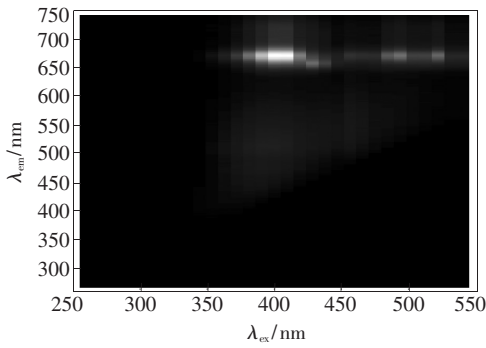


图 2 JC5 的灰度图

Fig. 2 Grayscale image of JC5

波长),将矩阵中的荧光强度值映射到 $[0, 255]$ 灰度范围内,则荧光强度矩阵可以被视为“灰度图像”。如图 2 所示,该图像可以完全地表示三维荧光光谱中的强度。

2.3 数据处理

2.3.1 Tchebichef 矩

对于大小 $N \times N$ 像素的图像,用 $f(x, y)$ 表示图像坐标 (x, y) 处的灰度值,则该图像的 $p + q$ 阶 Tchebichef 矩定义为:

$$T_{pq} = \frac{1}{\rho(p, N)\rho(q, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t_{p, N}(x)t_{q, N}(y)f(x, y),$$

$$p, q = 0, 1, \dots, N - 1, \quad (1)$$

其中, $t_{p, N}(x)$ 为 n 阶 Tchebichef 多项式,定义为:

$$t_{p, N}(x) = \frac{p!}{N^p} \sum_{k=0}^p (-1)^{p-k} \binom{N-1-k}{p-k} \binom{p+k}{p} \binom{x}{k}, \quad (2)$$

而 $\rho(n, N)$ 是归一化系数,定义为:

$$\rho(n, N) = \frac{N \left(1 - \frac{1}{N^2}\right) \left(1 - \frac{2^2}{N^2}\right) \cdots \left(1 - \frac{n^2}{N^2}\right)}{2n + 1}, \quad (3)$$

Tchebichef 矩是一种以离散的正交多项式为基函数的正交矩,因而在图像信息的表征上没有信息冗余^[7-8]。考虑到矩的计算复杂度以及对图像描述能力的完整性,本文初步选择前 11 + 11 阶($p + q$ 阶) Tchebichef 矩(共 121 个矩特征)提取样本特征,用于对样本的定性和定量分析。

2.3.2 聚类分析

为最大化类间对象的异质性和类内对象的同质性,聚类分析把对象集合进行分组,使同一类中的个体之间相似性强于该个体与其他类对象的相似性^[9-10],样品间的距离采用欧氏距离,利用类平

均法计算聚类树形图。

设样品 X_1, X_2, \dots, X_n 为总体样本,记第 i 个样品 $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ($i = 1, 2, \dots, n$),其中第 i 个样品 X_i 和第 j 个样品 X_j 之间的欧式距离定义为:

$$d_{ij} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}}, \quad (4)$$

类与类之间的平均距离定义为样品对之间的平方距离的平均值。 X_K 和 X_L 之间的平方距离定义为:

$$D_{KL}^2 = \frac{1}{n_K n_L} \sum_{x_i \in X_K, x_j \in X_L} d_{ij}^2. \quad (5)$$

2.3.3 逐步回归

从低阶到高阶的图像矩描述了图像的整体形状,其中包括目标化合物、噪声和不相关分量的完整信息^[2]。建立线性模型时,如果把所有特征全部考虑进去,会引入噪声和不相关分量等无意义的矩特征,从而影响定量的精度。因此,为了建立一个最优的回归方程,本文选用逐步回归作为对矩特征的优选方法。逐步回归根据自变量(Tchebichef 矩)对因变量(相对体积)的影响大小,将对因变量影响显著的自变量引入方程,并剔除影响不显著的自变量,最终得到最优回归方程。

3 结果与讨论

离散正交矩可以有效地提取不同层次上灰度图像的信息,不同阶的矩或其组合可以描述图像中不同的信息^[6](如目标化合物、噪声、不相关分量等信息),因此选择特定阶的矩或其组合就可以实现对样本的定性和定量分析,并且可以去除重叠峰、噪声和散射等干扰问题。

3.1 Tchebichef 矩结合聚类分析鉴别掺伪样本及其成分

鉴别样本是否为掺伪芝麻油,首先应观察样本是否和纯芝麻油属于一类。利用类平均法对纯芝麻油样本(Z1 ~ Z15)、芝麻香精样本(J1、J2)与掺伪样本(JC1 ~ JC9)进行聚类,得到如图 3 的聚类树形图,其中横坐标表示样本间的欧式距离,纵坐标表示样本。当选取前 1 + 2 阶 Tchebichef 矩作为样本的特征信息组合时,即可完全区分掺伪样本。由图 3 可知,在欧式距离 $L = 1.68$ 以上水平被分为 2 类,纯芝麻油样本(Z1 ~ Z14)被分为一类,芝麻香精样本(J1、J2)与掺伪样本(JC1 ~ JC9)被分为一类。通过以上分类结果可知,部分

Tchebichef 矩特征即可有效区分掺伪样本, 且能确定掺伪样本中含有芝麻香精成分。

掺伪样本被正确区分之后, 仍需对其所含的成分进行鉴别, 以便对其进行定量分析。由大豆油样本(D1 ~ D4)、玉米油样本(Y1 ~ Y7)、菜籽油样本(C1) 构造样本库。并把掺伪样本(JC1 ~ JC9) 与样本库进行聚类, 选择 1 + 2 阶 Tchebichef 矩可得如图 4 所示的聚类树形图。由图 4 可知, 在欧式距离 $L = 2.13$ 以上即可将菜籽油样本(C1) 与掺伪样本(JC1 ~ JC9) 分为一类, 而大豆油样本(D1 ~ D4) 和玉米油样本(Y1 ~ Y7) 在 $L = 1.49$ 以上也被完全区分, 从而可知掺伪样本含有菜籽油。因此少数 Tchebichef 矩特征即可鉴别掺伪样本的成分。

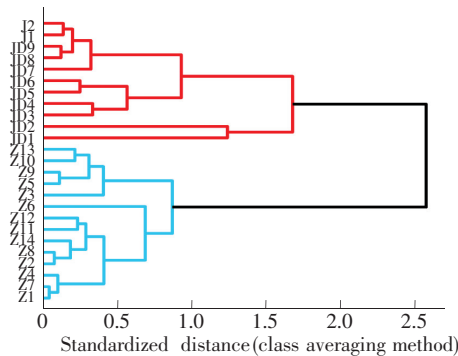


图 3 掺伪芝麻油样本聚类树形图

Fig. 3 Dendrogram of doping sesame oil

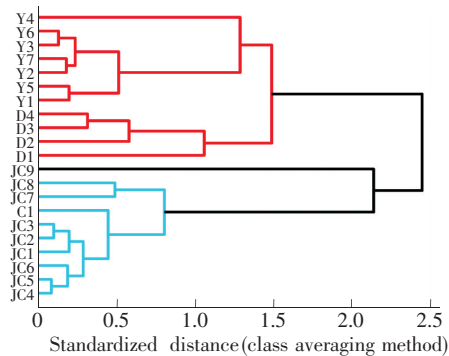


图 4 掺伪芝麻油成分聚类树形图

Fig. 4 Dendrogram of doping sesame oil's ingredients

3.2 逐步回归线性模型

应用掺伪样本(JC1 ~ JC9)、菜籽油样本(C1)、芝麻香精样本(J1) 建立定量线性模型。对样本的 121 个 Tchebichef 矩特征向量 $\{T_{1,1}, T_{1,2}, \dots, T_{11,11}\}$ 进行逐步回归, 相关系数最高的项 $T_{9,9}$ 首先进入方程, 其次 $T_{11,6}$ 和 $T_{5,10}$ 进入方程, 当回归方程中有 3 个变量时, 逐步回归方程线性相关度

最高 ($R = 0.999$), 剔除剩余变量, 相应的回归方程如下, 其中 A 是掺伪样本中一种目标化合物的相对体积, T 的下标表示阶数 (p, q):

(1) 芝麻香精模型:

$$A_J = 4.1159 - 18.681T_{5,10} - 58.76T_{9,9} - 113.24T_{11,6},$$

RMSE: 0.125, $R: 0.999, R - A_{dj} = 0.999, F - K_{\text{test value}} = 2.34e + 03, p - K_{\text{value}} = 7.21e - 11$ 。

(2) 菜籽油模型:

$$A_C = 5.8841 + 18.681T_{5,10} + 58.76T_{9,9} + 113.24T_{11,6},$$

RMSE: 0.125, $R: 0.999, R - A_{dj} = 0.999, F - K_{\text{test value}} = 2.34e + 03, p - K_{\text{value}} = 7.21e - 11$ 。

其中对回归曲线检验的 P 值 ($p - K_{\text{value}} = 7.21e - 11$) 均小于 0.05, 说明相对体积 A 和 Tchebichef 矩的线性关系是显著的, 模型拟合图如图 5 所示。

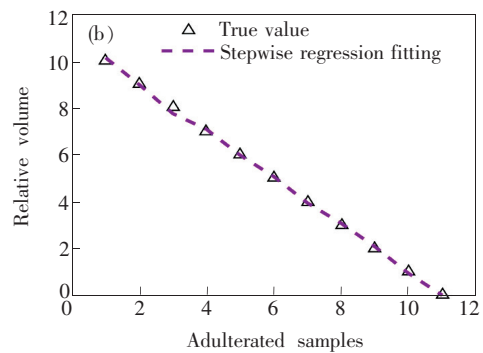
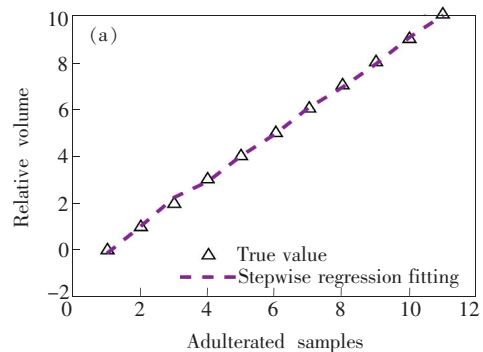


图 5 逐步回归结果

Fig. 5 Results of stepwise regression

4 结 论

本文利用图像处理技术中 Tchebichef 矩结合聚类分析和逐步回归建立了对掺伪芝麻油快速、准确的定量和定性分析方法。未作数据预处理, 直接用 Tchebichef 矩提取三维光谱数据的特征, 在此基础上通过聚类分析和逐步回归实现对掺伪芝麻油的鉴别, 并准确解析其组成成分及含量。实验结果表明该方法能够有效提取光谱的特征,

并能够得到良好的定性和定量结果,为质监部门及生产企业提供了新的油品鉴别手段。在此基础上,未来的研究会朝着增加样本数量以进一步完

善光谱数据库的方向进行,并分析讨论该算法对不同油种之间的适用性,以进一步提高检测的效率和实用性。

参 考 文 献:

- [1] 吴希军, 潘钊, 赵彦鹏, 等. 荧光光谱及平行因子分析法在植物油鉴别中的应用 [J]. 光谱学与光谱分析, 2014, 34(8):2137-2142.
WU X J, PAN Z, ZHAO Y P, *et al.*. Application of fluorescence spectroscopy and parallel factor analysis in the identification of vegetable oils [J]. *Spectrosc. Spect. Anal.*, 2014, 34(8):2137-2142. (in Chinese)
- [2] 吴希军, 田瑞玲, 孙梦菲, 等. 基于荧光光谱及矩阵分析的植物油鉴别技术 [J]. 光谱学与光谱分析, 2016, 36(7):2155-2161.
WU X J, TIAN R L, SUI M F, *et al.*. Identification technology of vegetable oil based on fluorescence spectroscopy and matrix analysis [J]. *Spectrosc. Spect. Anal.*, 2016, 36(7):2155-2161. (in Chinese)
- [3] ZHAI H L, ZHAI Y Y, LI P Z, *et al.*. A simple approach to quantitative analysis using three-dimensional spectra based on selected Zernike moments [J]. *Analyst*, 2012, 138(2):683-687.
- [4] JING C, BAO Q L, HONG L Z, *et al.*. A practical application of wavelet moment method on the quantitative analysis of Shuanghuanglian, oral liquid based on three-dimensional fingerprint spectra [J]. *J. Chromat. A*, 2014, 1352:55-61.
- [5] CHEN J, LI B Q, XU M L, *et al.*. Krawtchouk image moment method for the simultaneous determination of three drugs in human plasma based on fluorescence three-dimensional spectra [J]. *Talanta*, 2016, 161:99-104.
- [6] LI B Q, CHEN J, XU M L, *et al.*. The determination of multi-components utilizing 1 H NMR three-dimensional spectra combined Tchebichef moments [J]. *Chemomet. Intell. Lab. Syst.*, 2016, 156:128-136.
- [7] MUKUNDAN R, ONG S H, LEE P A. Image analysis by Tchebichef moments [J]. *IEEE Trans. Image Proc.*, 2001, 10(9):1357-1364.
- [8] 张辉, 周健, TOUMOULIN C, 等. Tchebichef 矩的快速算法 [J]. 东南大学学报(自然科学版), 2006, 36(5):857-862.
ZHANG H, ZHOU J, TOUMOULIN C, *et al.*. Fast algorithm for Tchebichef moments [J]. *J. Southeast Univ. (Nat. Sci. Ed.)*, 2006, 36(5):857-862. (in Chinese)
- [9] 梁曼, 黄富荣, 何学佳, 等. 荧光光谱成像技术结合聚类分析及主成分分析的藻类鉴别研究 [J]. 光谱学与光谱分析, 2014, 34(8):2132-2136.
LIANG M, HUANG F R, HE X J, *et al.*. Identification of algae by fluorescence spectroscopy combined with cluster analysis and principal component analysis [J]. *Spectrosc. Spect. Anal.*, 2014, 34(8):2132-2136. (in Chinese)
- [10] 刘丙新, 李颖, 韩亮, 等. 基于光谱反射率数据的水面油种鉴别研究 [J]. 光谱学与光谱分析, 2016, 36(4):1100-1103.
LIU BI X, LI Y, HAN L, *et al.*. Identification of surface oil based on spectral reflectance data [J]. *Spectrosc. Spect. Anal.*, 2016, 36(4):1100-1103. (in Chinese)



潘钊(1982 -),男,河北石家庄人,博士,副教授,2012年于燕山大学获得博士学位,主要从事荧光检测技术、光电检测技术和光谱分析方面的研究。
E-mail: panzh_zach@hotmail.com



吴希军(1979 -),男,河北邯郸人,博士,副教授,2010年于燕山大学获得博士学位,主要从事光纤传感与光谱测试系统设计等方面的研究。
E-mail: wuxijun@ysu.edu.cn