

专题论述

化学计量学方法及其在油脂识别分析中的应用

张 骊¹, 王亚瑞², 范 璐³, 李 娟³

(1. 国家粮食储备局 西安油脂科学研究设计院, 西安 710082; 2. 西安职业技术学院, 西安 710077;

3. 河南工业大学 化学化工学院, 郑州 450001)

摘要:综述了一些常见的化学计量学方法(主成分分析、聚类分析、判别分析和SIMCA等)在油脂识别分析中的应用。化学计量学方法主要用于光谱分析和色谱分析中的数据处理,从而最大限度地从中获取油脂的成分、结构及其他相关信息,解决一些常规数据处理方法无法解决的问题,使得同时识别不同的油脂成为可能。

关键词:化学计量学;油脂;识别;数据处理

中图分类号:TQ646

文献标志码:A

文章编号:1003-7969(2010)01-0001-05

Application of chemometrics methods in the discrimination of oils and fats

ZHANG Li¹, WANG Yarui², FAN Lu³, LI Juan³

(1. Xi'an Oils & Fats Research & Design Institute, State Administration of Grain Reserve of P. R. C.,

Xi'an 710082, China; 2. Xi'an College of Career and Technology, Xi'an 710077, China;

3. School of Chemistry & Chemical Engineering, Henan University of Technology,

Zhengzhou 450001, China)

Abstract: The applications of chemometrics methods (principal components analysis, cluster analysis, discriminant analysis and soft independent modelling class analogy) in the discrimination of different kinds of oils and fats were described. The chemometrics methods could solve problems that could not be solved by conventional data processing methods. The chemometrics methods were mainly used for the spectra analysis and chromatogram analysis, so as to maximumly obtain the composition, structures and other correlations information of the different oils and fats, which made it possible to discriminate different oils and fats at the same time.

Key words: chemometrics; oils and fats; discrimination; data processing

化学计量学是一个化学分支,它利用数学和统计学方法进行设计和选择最优的测量程序和实验方案,并通过对化学数据的分析提供最大限度的化学信息^[1]。化学计量学是分析化学的前沿研究方向之一^[2],是化学量测的基础理论和方法学。

化学计量学的研究范围极为广泛,内容丰富。化学实验设计与优化、定量校正理论、分析信号处理、化学模式识别、模型与参数估计、数据解析、过程

模拟、人工智能、情报检索、实验室自动化等等都是化学计量学的研究范围。其中,化学模式识别技术是粮油食品类别分析中应用较广而且较有效的化学计量学方法。化学模式识别就是借助计算机来揭示隐含于化学量测数据内部规律的一种多元分析技术,它的任务就是要找出相同样品具有某些共同性的特征,进而根据这些特征去预报一些未知样本的性质,来判别未知样品属于哪一个模式,也就是属于哪一类或哪一种。

随着计算机科学、应用数学和统计学方法在化学中应用的日益广泛和深入,化学计量学理论和方法已渗透到化学中的各个领域。不像以往的分析工

收稿日期:2009-07-15;修回日期:2009-10-22

作者简介:张 骊(1964),女,高级工程师,主要从事油脂工程设计及新产品开发工作。

作者对研究过程中的预处理及实验过程研究较多,而对方法选择、数据处理及结果解释方面研究较少,往往不能以最优的实验设计去分析和处理问题,并尽可能多地提取有用的化学信息。而化学计量学正是以研究过程中的方法选择、数据处理、信息提取及结果解释为主要研究内容,它在实验设计、数据处理、信号解析、化学分类决策及预报方面具有巨大的优势,能够解决传统的化学研究方法难以解决的复杂问题,为此,它有着强大的生命力,得到了广泛的应用。

1 常用的化学计量学方法

随着化学计量学的发展人们越来越多地倾向于用模式识别技术处理食品中的有机成分实验数据,以解决植物油分析和鉴别的难题,主成分分析法、判别分析、聚类分析^[3]、偏最小二乘法等受到普遍重视,还有其他化学计量学方法如人工神经网络、多元线性回归法^[4]、SIMCA 法^[5]、校正变换矩阵法在油脂识别分析中也有应用。

1.1 主成分分析 (principal components analysis, PCA)^[6]

主成分分析是模式识别分析中最常用的一种线性映射方法,这种方法是根据样本点在多维模式空间的位置分布,以样本点在空间中变化最大方向,即方差最大方向作为判别矢量。从概率统计观点可知,一个随机变量的方差越大,该随机变量所包含的信息越多。所谓主成分就是原始数据的 p 个变量经线性组合(或映射)得到的变量,其方差为最大(第一主成分)。主成分之间是相互线性无关的,一般前几个主成分可包括原变量的大部分信息。主成分分析是多元统计中的一种数据压缩技术,在不丢失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量,解决了由于谱带重叠而无法分析的困难^[7]。

1.2 聚类分析 (cluster analysis, CA)

聚类分析是按“物以类聚”的原则将特性相近的变量或观察单位进行归类。聚类分析是数理统计的一种方法,适用于对于样本没有类的先验知识的情况^[8]。聚类分析可分为变量聚类和样品聚类两类,前者又称指标聚类或 R 型聚类;后者又称 Q 型聚类。用于近红外光谱定性分析常用的方法有系统聚类分析法、 K -均值聚类方法、模糊 K -均值聚类方法、以及自组织 (Kohonen) 神经网络。系统聚类法和 K -均值聚类方法是目前聚类分析中应用最多的两种方法,许多著名的统计软件,如 SAS、SPSS 等都包含相应模块。模糊 K -均值聚类方法和自组织

(kohonen) 神经网络方法是近几年发展最快的聚类方法。聚类分析的重要组件为样品间的距离、类间的距离、并类的方式和聚类数目的判定。对已知种类的样品进行聚类分析的基本原理是近红外光谱或其压缩的变量(如主成分)组成一个多维的变量空间,通过化学计量学计算,反映出不同种类样品间的差别,由于同类物质在该多维空间聚集于相近的位置,从而对样品进行分类。

1.3 判别分析 (discriminant analysis, DA)^[9]

判别分析是判别样品所属类型的一种统计方法,是在已知研究对象分成若干类型(或组别)并已取得各种类型的一批已知样品的观测量数据的基础上,根据某些准则建立判别式,然后对未知类型的样品进行判别分析。

1.4 人工神经网络 (artificial neural network, ANN)^[10]

人工神经网络又称神经网络,是通过已知类别的训练集的拟合,来建立识别模型对未知样本进行分类和预测。人工神经网络的算法^[10]就是给网络各种训练样本,把网络的实际输出和正确的目标值相比较,然后根据偏差的情况修改各节点的连接权重,网络不断朝误差减小的方向进行变化,直到输出值与正确的目标值的偏差满足工作所需要的精度。经过训练的网络把系统规则、预测能力、变量转化等都以权值的形式隐含在网络之上,然后向输入层输入信息,即可给出结果。它的学习算法称为误差反传学习算法 (back propagation of error, 简称 BP 算法),这种网络也称为 BP 网络。植物油脂的组成和含量信息复杂,难以用解析式表达,选用 BP 法校正则较为有效。

1.5 SIMCA (soft independent modeling of class analogy)

SIMCA 是一种有监督模式的识别方法,是一种二值判定方法。先对训练集中每一类样本的量测数据矩阵分别进行主成分分析,建立每一类主成分分析数学模型,然后在此基础上对未知样本进行分类。基于每一个类别训练集的主成分分析建立的模型,计算未知样本点与培训集的主成分分析模型的距离,根据距离判别方法判别未知样本的类别。

随着粮油及其制品中研究因素的不断增多,种类识别和掺伪分析趋于复杂,而化学计量学的应用为粮油种类的识别和掺伪分析提供了更简单、更科学的数据处理方法。

2 计量学方法在油脂识别分析中的应用

油脂是一种复杂的混合物,其中含有的甾醇、维生素、胡萝卜素等物质,可能对红外光谱产生干扰,

给解析谱图增加了难度;甘三酯的立体结构的差异也会给红外光谱的分析带来一定的难度。用一种光谱分析法并不能完全地对各类油脂进行识别,需要将多种分析方法相结合^[11]。因此,不仅仅只根据单一的红外图谱来对油脂进行识别,而是将光谱法和计量学的方法相结合来对不同种类的油脂进行分析。

2.1 计量学方法与红外光谱法结合识别分析油脂

Jefferson等^[12]人将偏最小二乘法(PLS)与人工神经网络技术相结合,建立分析模型,对FTIR-ATR转换及FT-NIR红外光谱信息进行了分析处理,用于生物柴油混合物中甲酯含量的测定,研究结果表明所建模型的精确度高,分析结果理想。Vlachos等^[13]人将红外光谱法与计算机软件技术相结合,对掺有玉米油及芝麻油的橄榄油样品进行了测定,通过红外光谱 $3\ 009\text{ cm}^{-1}$ 处吸收峰的变化情况,分析出了理想的识别结果。Mossoba等^[14]人采用红外光谱法,将植物油的红外光谱进行二阶导数运算后,选取导数光谱中 966 cm^{-1} 处的特征峰进行分析,检测出了用气相色谱没能检测出的饱和脂肪酸中的微量反式酸。Woodcock等^[15]人采集了913个样品的近红外光谱图,对光谱进行一阶导数处理,结合主成分分析和偏最小二乘法建立了最好的识别模型,可以92.8%和81.5%识别利古里亚和非利古里亚橄榄油。Galtier等^[16]人应用偏最小二乘法分析125个橄榄油样品的NIR光谱,区分法国橄榄油的产地及组成,得到满意的结果,并表明此方法优于传统的GC或HPLC方法。Özdemir等^[17]人用NIR光谱法检测26个掺有葵花籽油和玉米油的橄榄油(掺伪体积分数4%~96%),多元校正模型通过遗传反转最小二乘法(GILS)建立,结果显示,能够识别橄榄油和葵花籽油的二重混合油预测值的平均标准误差在2.49%~2.88%,橄榄油、葵花籽油和玉米油的三重混合油在1.42%~6.38%。Gurdeniz等^[18]人用主成分分析和偏最小二乘法分析红外光谱检测掺伪的橄榄油,主成分分析是通过小波压缩进行降维,掺有玉米油和葵花籽油的橄榄油的检出限为5%;通过直角信号正交(OSC)和小波压缩处理光谱数据后进行偏最小二乘法分析,预测掺有玉米油-葵花籽油、棉籽油和菜籽油的橄榄油的误差限分别为1.04、1.4和1.32;此外,经过建立的主成分分析模型和偏最小二乘法判别分析(PLS-DA)对不同类型的掺伪油脂,得到最低检出限为10%。Marini等^[19]人运用NIR结合化学计量学方法识别

不同种类的橄榄油,最理想的模型中验证 Q^2 达到0.91~0.96,且平均预测误差为5%~7.5%。Chen等^[20]人应用近红外光谱技术结合遗传算法完全正确分类不同来源的橄榄油,建立三层神经网络模型可以100%识别30个未知橄榄油,提供了一个新的识别油脂的方法。意大利Monica等^[21]人采集了195个利古里亚特级初榨橄榄油的近红外光谱法,采用了位函数技术、SIMCA、二次不等判别分析和多元归类建模技术对样品进行识别分析,结果显示,利古里亚橄榄油可以与其他油脂清晰的区分,并且应用多元归类建模技术建立的橄榄油分类模型有很好的预测能力、高灵敏度和充足的特异性,结果指出NIR结合化学计量学方法对于检测橄榄油是很好的方法。

2.2 计量学方法与色谱法结合识别分析油脂

目前已经有化学计量学方法结合气相色谱法在油脂识别分析中的报道。Rezanka等^[22]人用毛细管气相色谱法测定了摩洛哥坚果油和其他植物油的脂肪酸含量和植物油中甘三酯的种类及含量,运用SPSS8.0.1统计学分析软件的欧式距离和平方欧式距离的方法识别摩洛哥坚果油的掺伪,表明系统聚类分析能清楚地区别不同的植物油。Doutoglou等^[23]人用气相色谱法提取了橄榄油和其他植物油的全部脂肪酸信息和1,3位脂肪酸信息,采用主成分分析的方法,建立了一种用于识别橄榄油掺伪的方法。研究表明,主成分分析三维图用于识别橄榄油和其他植物油是可行的。Lee等^[24]人用气相色谱法测定了芝麻油、紫苏油、大豆油、玉米油、卡诺拉油、菜籽油、橄榄油和椰子油中的脂肪酸和三酰甘油,并测定了掺有不同比例大豆油的芝麻油的脂肪酸组成和含量,以棕榈酸、硬脂酸、油酸、亚油酸和亚麻酸的含量作为特征参数,运用主成分分析和判别分析的方法,完成了芝麻油真伪识别。Voncina等^[25]人用气相色谱测定植物油中的脂肪酸种类和含量,以棕榈酸、硬脂酸、油酸、亚油酸、亚麻酸、花生酸和花生烯酸的含量作为特征参数,用主成分分析和基于神经网络的聚类分析实现了植物油的分类,交互验证准确率为95%,线性判别分析的判别准确率为96.84%,表明了气相色谱技术结合化学计量学方法分析识别油脂是可行的。Hajimahmoodi等^[26]人用气相色谱脂肪酸指纹图谱法结合统计学中的最小二乘法模型,对含有棉籽油、橄榄油、大豆油、葵花籽油这4种植物油的混合油进行了同时测定,建立校正模型,测定结果的相对标准偏差低于10%。结果表明,这种方法能用于识别这4种油的

可能的掺伪。国内也有用气相色谱法结合化学计量学方法的报道,马亨等^[27]人根据食用油中脂肪酸、甾醇以及生育酚含量,应用校正转换矩阵法对花生油掺伪进行了定量检测,结果令人满意。何小青等^[28]人建立了微波辅助衍生化 GC-MS 测定植物油中的脂肪酸含量,使用校正变换矩阵法对食用植物油的成分进行测定,可用于快速、准确测定食用油中各成分含量或定性、定量识别掺伪成分。化学计量学方法结合气相色谱的方法在油脂识别分析方面有很大的优势,是目前油脂识别分析方面发展的趋势。但这方面的研究目前还处在起步阶段,只是针对一种植物油脂的识别分析,还需要加大可靠样品的分析,增加油脂样品的品种,对其进行进一步深入的研究。化学计量学方法结合高效液相色谱对植物油脂进行识别分析的报道比较少,主要是测定各种甘三酯含量^[29,30],然后结合化学计量学的方法对油脂进行识别分析。

3 结束语

由于天然植物油脂是复杂的有机化合物,而不同的植物油脂组成的差异是多变量,传统的数据分析方法很难用系统的方法表征多种油脂之间的差异,而化学计量学与计算机结合,通过数学计算达到了识别的目的,使得长期困扰学者的混合物波谱同时识别成为可能,不仅体现了绿色分析的优势,而且拓宽了油脂识别分析思路。随着各学科的发展,可以预期,化学计量学在油脂分析中的应用也将继续得到更蓬勃的发展。

参考文献:

- [1] 奥托 M. 化学计量学——统计学与计算机在分析化学中的应用[M]. 邵学广,蔡文生,徐筱杰,译. 北京:科学出版社,2003:1-2.
- [2] 刘树深. 基础化学计量学[M]. 北京:科学出版社,1999:1-2.
- [3] 张国文,倪永年,涂北平. 食用植物油的分类和质量鉴别模式识别研究[J]. 食品科学,2005,26(1):71-74.
- [4] PEN F, CARDENAS S, GALLEGO M, et al. Direct olive oil authentication; detection of adulteration of olive oil with hazelnut oil by direct coupling of headspace and mass spectrometry, and multivariate regression techniques [J]. J Chromatogr A, 2005, 1074: 215-221.
- [5] WANG L, LEE F S C, WANG X R, et al. Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR [J]. Food Chem, 2006, 95: 529-536.
- [6] 朱尔一,杨芑元. 化学计量学技术及应用[M]. 北京:科学出版社,2001:87,149.
- [7] REN Y L, BING C T, LU J H, et al. Principal component analysis of near-infrared diffuse reflectance spectra [J]. Spectroscopy and Spectral Analysis, 1996, 16(6): 31-35.
- [8] LI Y Z, MIN S G, LIU X. Study on the methods and applications of near-infrared spectroscopy chemical pattern recognition [J]. Spectroscopy and Spectral Analysis, 2007, 27(7): 1299-1303.
- [9] 罗应婷,杨钰娟. SPSS 统计分析——从基础到实践[M]. 北京:电子工业出版社,2007:252.
- [10] 李燕,王俊德,顾炳和,等. 人工神经网络及其在光谱分析中的应用[J]. 光谱学与光谱分析, 1999, 19(6): 844-849.
- [11] YANG H, IRUDAVARA J, PARDKAR M M, et al. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy [J]. Food Chem, 2005, 93: 25-32.
- [12] JEFFERSON S O, MONTALVAO R, DAHER L, et al. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies [J]. Talanta, 2006, 69: 1278-1284.
- [13] VLACHOS N, SKOPELITIS Y, PSARONDAKI M, et al. Applications of Fourier transform-infrared spectroscopy to edible oils [J]. Analytica Chimica Acta, 2006, 573/574: 459-465.
- [14] MOSSOBA M M, KRAMER J K G, MILOSEVIC V, et al. Interference of saturated fats in the determination of low levels of trans fats (below 0.5%) by infrared spectroscopy [J]. J Am Oil Chem Soc, 2007, 84(4): 339-342.
- [15] WOODCOCK T, DOWNEV G, O'DONNELL C. Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy [J]. J Agri Food Chem, 2008, 56(23): 11520-11525.
- [16] GALTIER O, DUPUV N, LE DREAU Y, et al. Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra [J]. Analytica Chimica Acta, 2007, 595(1/2): 136-144.
- [17] ÖZDEMİR D, ÖZTÜRK B. Near infrared spectroscopic determination of olive oil adulteration with sunflower and corn oil [J]. Betul Food and Drug Analysis, 2007, 15(1): 40-47.
- [18] GURDENİZ G, OZEN B. Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data [J]. Food Chem, 2009, 116(2): 519-525.
- [19] MARINI F, MAGRI A L, BUCCI R. Use of different artificial neural networks to resolve binary blends of monocultivar Italian olive oils [J]. Analytica Chimica Acta, 2007, 599(2): 232-240.
- [20] CHEN Y M, LIN P, HE Y. Study on discrimination of pro-

- ducing area of olive oil using near infrared spectra based on genetic algorithms[J]. Spectroscopy and Spectral Analysis, 2009, 29(3): 671-674.
- [21] MONICA C, CHIARA C, GIUSEPPE F, et al. Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil [J]. J Near Infrared Spectrosc, 2008, 16(1): 39-47.
- [22] REZANKA T, REZANKOVA H. Characterization of fatty acids and triacylglycerols in vegetable oils by gas chromatography and statistical analysis [J]. Anal Chim Acta, 1999, 398: 253-261.
- [23] DOUTOGLOU V G, DOUTOGLOU T, ANTONOPOULOS A, et al. Detection of olive oil adulteration using principal component analysis applied on total and regio FA content [J]. J Am Oil Chem Soc, 2003, 80: 203-208.
- [24] LEE D S, NOH B S, BAE S Y, et al. Characterization of fatty acid composition in vegetable oils by gas chromatography and chemometrics [J]. Anal Chim Acta, 1998, 358: 163-175.
- [25] VONCINA D B, KODBA Z C, NOVIC M. Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids [J]. Chemom Intell Lab Syst, 2005, 75: 31-43.
- [26] HAJIMAHMOODI M, HEYDEN Y V, SADEGHII N, et al. Gas - chromatographic fatty - acid fingerprints and partial least squares modeling as a basis for the simultaneous determination of edible oil mixtures [J]. Talanta, 2005, 66: 1108-1116.
- [27] 马亭, 邹世春, 陈文锐, 等. 应用校正变换矩阵法识别掺伪食用油 [J]. 分析测试学报, 2000, 19(2): 21-24.
- [28] 何小青, 许德英, 罗美中, 等. 微波辅助衍生 GC-MS 测定脂肪酸 - 校正变换矩阵法鉴别食用植物油的研究 [J]. 现代科学仪器, 2004(3): 36-39.
- [29] HAMDY A H E, FIZGA N K E. Detection of olive oil adulteration by measuring its authenticity factor using reversed - phased high - performance liquid chromatography [J]. J Chromatogr A, 1995, 708: 351-355.
- [30] LEE D S, LEE E S, KIM H J, et al. Reversed phase liquid chromatographic determination of triacylglycerol composition in sesame oils and the chemometric detection of adulteration [J]. Anal Chim Acta, 2001, 429: 321-330.

· 信息 ·

花生产业或列入农业补贴名单

北京大学中国地方政府研究院院长彭真怀一年多时间一直在陕西、河南、山西等花生主产区调研,他完成的一份《振兴我国花生产业的政策建议》报告已引起国务院政研室、发改委、农业部等部委的重视,并已呈送国务院高层。目前国务院中农办、发改委、农业部、财政部等部委正在密集调研会签,部门之间也在交换意见,花生产业已被高度关注,或列入农业补贴名单。

一些农业专家担心扩大花生种植会不会“与粮争地”?但有关部门调研发现,作为一种耐瘠薄、抗干旱作物,花生可以最大程度地开发贫瘠和沙化土地,提高非耕地的有效利用率,是贫困山区致富的有效途径。

有关专家将花生与大豆进行比较后发现,每亩花生产量为250公斤左右,而大豆在120公斤左右;花生含油率为50%左右,大豆为20%左右,花生亩产和含油率都在大豆的2倍以上。也就是说,生产同样数量的食用油,种大豆需要4亩地,而种花生仅要1亩地。

由于花生收获季节相对集中,一些企业考虑到资金占用周期长等因素,在得不到国家政策支持的情况下,很难用自有资金进行收储。2009年3月,国内花生价格跌到了每吨4000元,导致种植面积比2008年减少约650万亩。

另外,花生科研投资多、见效慢,没有中央财政专项支持,有限的科研力量大多分散在相关科研院所及个别企业内,长期不能有效整合资源。以追求粮食产量增加为目标的科研经费,2/3以上投向了高产粮食作物,而花生主产区农民仍沿用20年前的老品种,产量低,出油率低,混种混收混加工,品种退化已经到了濒临绝收的严重地步。

因此,报告建议,把花生种植作为开发革命老区和贫困山区的重要措施,把花生深加工作为振兴我国农产品加工的支柱产业给予优先扶持。可在国务院直接领导下,由国家发改委牵头,会同财政部、农业部、科技部、商务部、卫生部和质检总局等有关涉农部门,共同研究制定《花生产业振兴规划》、《花生种植布局规划》和《花生科技支撑规划》,形成系统完整、高效运转的制度保障体系。 (SPSWW)