近红外光谱-BP 神经网络-PLS 法用于橄榄油掺杂分析

翁欣欣[1]，陆　峰[1]，王传现[2]，亓云鹏[1*]

1. 第二军医大学，上海　200433
2. 上海出入境检验检疫局，上海　200135

摘　要　橄榄油兼有食用和保健的作用，价值与价格远远高于其他食用油，所以橄榄油中以劣充好的现象十分普遍。可采用近红外光谱法测定初榨橄榄油中掺杂芝麻油、大豆油和葵花籽油的光谱数据，运用改进的 BP 算法——Levenberg-Marquardt 方法，建立 PCA-BP 人工神经网络方法对其进行定性判别。同时采用偏最小二乘法(PLS)建立了初榨橄榄油中芝麻油、大豆油、葵花籽油含量的近红外光谱定标模型，用交互验证法进行验证。结果表明，BP 人工神经网络有很好的定性鉴别能力，PLS 建立的芝麻油、大豆油、葵花籽油定标模型的相关系数分别为 98.77，99.37，99.44，交叉验证的均方根误差分别为 1.3，1.1，1.04。该方法无损、快速、简便，为橄榄油掺杂的检测提供了一种新的途径。

关键词　近红外光谱；橄榄油；鉴别和定量；BP 人工神经网络；偏最小二乘法(PLS)

中图分类号：O657.3　　**文献标识码**：A　　**DOI**：10.3964/j.issn.1000-0593(2009)12-3283-05

引　言

食用油是膳食的主要成分之一，目前市场上销售的食用油种类繁多，主要有大豆油、葵花籽油、芝麻油以及各种调和油等[1]，其中橄榄油兼有食用和保健作用，是医学界公认为的最有益健康的食用油之一，在市场上长期以来价格都高于其他的食用油[2]。目前，橄榄油中以劣充好的现象十分普遍，虽然国家已发布了食用油市场准入制度和食用油标准，但目前对食用油的掺伪判别还缺乏快速、方便、准确的检测手段[3, 4]。

油脂掺伪检验十分复杂，一直是油脂安全监控领域的一项技术性难题。常采用气相色谱(GC、GC-MS)[5, 6]和液相色谱(HPLC-DAD、HPLC-MS)[7, 8]技术分析植物油中脂肪酸、甾醇、甘油三酯等成分，对植物油进行分类和质量鉴别。但这些方法往往费时、操作繁琐、费用高等。现已获得公认可被采用的检测方法较少，而且也很不完善。植物油是十分复杂的多组分体系，像其他油品一样，橄榄油也含有饱和脂肪酸、多双键不饱和脂肪酸和单双键不饱和脂肪酸[9]，这使得植物油掺伪的检测工作更加困难。但是橄榄油与其他油品的脂肪酸含量存在着一些差异，其中最重要的是橄榄油中含有 60%～80%的单双键不饱和脂肪酸——油酸，其含量明显地高于其他油品。

光谱分析法已经在食品(药品)掺伪检测的研究领域发挥了非常重要的作用[10, 11]，其中近红外光谱(NIR)分析技术以其速度快、不破坏样品、操作简单、稳定性好、效率高等特点，已经广泛地应用于食品工业、石油化工、制药工业等领域，包括油品掺伪检测的研究等[12-14]。

1　材料与方法

1.1　实验仪器

VECTOR 22/N 傅里叶变换红外光谱仪(德国 Bruker 公司)。分析软件 SPSS 15.0，OPUS 5.0，MATLAB 7.0。

1.2　实验样品

28 种不同品牌的特级初榨橄榄油、芝麻油、大豆油及葵花籽油。橄榄油和芝麻油由上海出入境检验检疫局提供，大豆油和葵花籽油为超市购得。

将一种芝麻油掺杂到 10 种橄榄油中，按质量比(ω)5%～50%掺杂油样品 26 份。将一种大豆油掺杂到 10 种橄榄油中，按质量比(ω)5%～50%掺杂油样品 45 份。将一种葵花籽油掺杂到 10 种橄榄油中，按质量比(ω)5%～50%掺杂油样品 48 份。将一种大豆油和一种芝麻油掺杂到五种橄榄油中，按质量比(ω)5%～30%掺杂油样品 5 份。

Table 1　Ten PCs and reliabilities

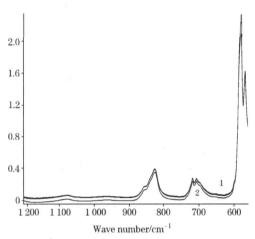| 主成分	PC1	PC2	PC3	PC4	PC5
累积贡献率/%	99.789	99.947	99.974	99.988	99.996
主成分	PC6	PC7	PC8	PC9	PC10
累积贡献率/%	99.999	99.999	99.999	100.000	100.000

Fig. 1　Near infrared spectra of pure and adulterated virgin olive oil

1: Adulterated virgin olive oil; 2: Pure virgin olive oil

1.3　光谱采集

取上述样品于 5 mm 的检测池中，用近红外透射光谱法进行光谱采集，扫描范围 3 700～12 000 cm^{-1}，分辨率 4 cm^{-1}，扫描次数 32 次，每个样品重复测定 5 次后取平均值。实验温度 25 ℃，空气湿度 70%。样品的近红外光谱如图 1 所示。从图 1 可以看出，纯的初榨橄榄油和掺杂的初榨橄榄油的原始光谱非常相似。

2　结果与讨论

2.1　BP 神经网络的建立与定性分析

2.1.1　光谱预处理

采集的光谱受到高频随机噪声、基线漂移、样品不均匀、光散射等影响，需要进行光谱预处理来消除噪声。采用基线校正，向量标准化的预处理方法。由于系统样品光谱曲线在尾部有较大的噪声，所以选用 12 000～5 390 cm^{-1} 波段的光谱用于分析。

2.1.2　主成分分析得到新的特征变量

主成分分析是把多个指标化为少数几个综合指标的一种统计分析方法，其主要目的是用来降维。经主成分分析光谱数据后，得到前 10 个主成分累积贡献率如表 1 所示，前 6 个主成分的累积贡献率已经达到 99.999%，所以每个样品的光谱数据可以用 6 个主成分代替。

2.1.3　BP 神经网络的建立与定性分析

本文建立的神经网络为误差反向传播算法，也称为 BP (Back-Propagation Network) 算法[15]。在利用人工神经网络进行建模之前，首先采用主成分分析对近红外光谱数据进行压缩和降维，得到的主成分作为 BP 神经网络输入。

BP 神经网络分 3 层，即输入层、隐含层和输出层，采用 Sigmoid 激发函数，并运用改进的 BP 算法——Levenberg-Marquardt 方法。主成分分析将所有光谱信息提取出 6 个主成分，将其作为神经网络的输入来建立一个三层的网络模型。经多次实验确定隐含层单元数为 12，网络的输出为 2，纯的橄榄油的代码为[0 1]，掺杂的橄榄油的代码为[1 0]。选择 20 个纯橄榄油，20 个掺杂芝麻油的橄榄油，30 个掺杂大豆油的橄榄油和 30 个掺杂葵花籽油的橄榄油，共 100 个

Table 2　Prediction result for unknown samples by BP model

预测样本序号	真实值		预测值		预测样本序号	真实值		预测值	
1	0	1	−0.003 7	1.003 7	27	1	0	0.999 8	0.000 1
2	0	1	0.005 2	0.994 9	28	1	0	0.999 7	0.000 2
3	0	1	−0.000 1	1.000 1	29	1	0	0.999 5	0.000 4
4	0	1	0.004 0	0.996 0	30	1	0	1.000 1	−0.000 1
5	0	1	−0.000 2	1.000 2	31	1	0	1.000 2	−0.000 2
6	0	1	0.000 2	0.999 9	32	1	0	1.000 0	0.000 0
7	0	1	0.000 0	1.000 0	33	1	0	0.997 8	0.002 1
8	0	1	−0.027 9	1.027 8	34	1	0	1.000 0	−0.000 0
9	1	0	1.000 1	−0.000 0	35	1	0	1.000 1	−0.000 1
10	1	0	1.000 0	0.000 0	36	1	0	0.999 7	0.000 2
11	1	0	1.000 1	−0.000 0	37	1	0	0.998 6	0.001 5
12	1	0	0.999 9	0.000 1	38	1	0	0.996 0	0.004 0
13	1	0	1.004 8	−0.004 8	39	1	0	1.000 0	−0.000 1
14	1	0	1.000 0	0.000 0	40	1	0	1.000 2	−0.000 2
15	1	0	0.999 7	0.000 3	41	1	0	0.999 9	−0.000 0
16	1	0	0.957 9	0.042 1	42	1	0	1.000 1	−0.000 0
17	1	0	0.997 7	0.002 1	43	1	0	1.000 0	−0.000 1
18	1	0	0.999 8	0.000 1	44	1	0	0.999 9	0.000 0
19	1	0	1.000 0	−0.000 0	45	1	0	1.040 3	−0.040 3
20	1	0	1.000 0	0.000 0	46	1	0	0.986 9	0.013 1

续表 2

21	1	0	0.999 8	0.000 2	47	1	0	1.000 3	−0.000 3
22	1	0	0.998 2	0.001 8	48	1	0	1.000 0	0.000 0
23	1	0	0.999 0	0.001 0	49	1	0	0.999 7	0.000 3
24	1	0	0.999 9	0.000 2	50	1	0	1.000 0	−0.000 0
25	1	0	0.999 9	0.000 0	51	1	0	0.999 9	0.000 0
26	1	0	0.999 6	0.000 3	52	1	0	0.999 9	0.000 0

样品作为建模集，剩余的 52 个样品作为预测集。100 个建模集的拟合残差为 $6.762\ 31\times10^{-7}$，训练迭代次数为 1 000 次。对未知的 52 个样品进行预测，结果准确率为 100%（见表 2）。

2.2 定标模型的建立

2.2.1 初榨橄榄油中掺杂芝麻油定标模型建立

光谱经基线校正、平均值中心化(mean centering)和直线扣除(straight line subtraction)处理以及消除异常点后，采用偏最小二乘法(PLS)建立初榨橄榄油中芝麻油含量的定标模型，并经内部交互验证[16]。在用偏最小二乘法建立模型时，最重要的是确定建模所需要的主因子数，采用的主因子数过多或过少，都会使拟合的精度降低，预测效果变差。应用交互验证方法，通过 F-检验来确定当预测残差平方和(PRESS)达到最小时的主因子数[17]，即最优主因子数。优化模型是一个检测奇异点、强响应点、确定最优主因子和相应交互验证均方残差 RMSECV 的协同过程，以主因子数与内
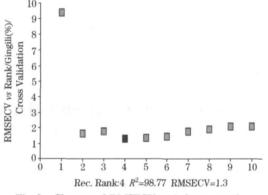
部交互验证的均方残差(RMSECV)作图(图 2)。由图可以看出，当主因子数为 4 时，RMSECV 值达到最小，故取主因子数 4 建立 PLS 定标模型。图 3 为初榨橄榄油中芝麻油含量从 0%～50% 的实际值与近红外光谱定标模型预测值的相关曲线，横轴为芝麻油含量的实际值，纵轴为预测值。由图 3 知模型的相关系数为 98.77，RMSECV＝1.3。
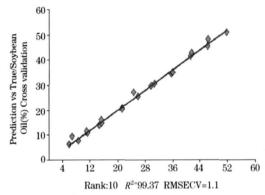
Fig. 4 Scattering plots of the prediction values and the true values of the content of soybean oil in virgin olive oil

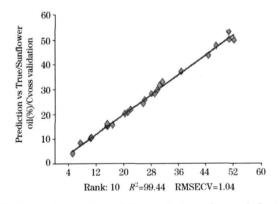**Fig. 2 Changes of RMSECV with factor number**

Fig. 5 Scattering plots of the prediction values and the true values of the content of sunflower oil in virgin olive oil

2.2.2 初榨橄榄油中掺杂大豆油定标模型的建立

光谱经基线校正、平滑(平滑点数 9)、向量标准化和一阶导数处理后，经过消除异常点，选择主因子数为 10，建立初榨橄榄油中大豆油含量的 PLS 定标模型，并经内部交互验证。由图 4 知模型的相关系数为 99.37，RMSECV＝1.1。

2.2.3 初榨橄榄油中掺杂葵花籽油定标模型的建立

光谱经基线校正、平均值中心化和直线扣除处理后，经过消除异常点，选择主因子数为 10，建立初榨橄榄油中葵花

Fig. 3 Scattering plots of the prediction values and the true values of the content of sesame oil in virgin olive oil

籽油含量的 PLS 定标模型，并经内部交互验证。由图 5 知模型的相关系数为 99.44，RMSECV＝1.04。

3 结 论

通过近红外光谱仪测得吸收光谱，结合主成分分析和人工神经网络建立了掺杂的初榨橄榄油的鉴别模型。结果表明，初榨橄榄油的识别率达到了 100%。说明运用近红外光谱可以准确、快速地对初榨橄榄油的掺杂进行定性鉴别。而偏最小二乘法建立的定标模型，可以检验初榨橄榄油中掺杂芝麻油、大豆油及葵花籽油的含量，检测结果理想。本文所建的研究方法快速、简便，为食用油的快速检验提供了一种新的技术手段。

参 考 文 献

[1] XIAO Chang-zhen，WU Yu，GAN Dong-sheng，et al(肖昌珍，吴　渝，甘东生，等)· Chinese Journal of Oil Crop Sciences(中国油料作物学报)，1999，21：63.
[2] ZHAO Dan，XIAO Ping，REN Yan(赵　丹，肖　平，任　雁)· South China Agriculture(南方农业)，2007，1(2)：49.
[3] WANG Han-zhong(王汉中)· Chinese Journal of Oil Crop Sciences(中国油料作物学报)，2007，29(3)：347.
[4] LU Feng，WU Yu-tian，CHAI Yi-feng(陆　峰，吴玉田，柴逸峰)· Life Science Instruments(生命科学仪器)，2007，5(11)：6.
[5] Gema Flores，Maria Luisa Ruiz Del Castillo，Gracia Planch，et al· Food Chemistry，2006，97：336.
[6] Luisito Cercaci，Maria Teresa Rodriguez-Estrada，Giovanni Lercker· Journal of Chromatography A，2003，985：211.
[7] Zabaras D，Gordon M H· Food Chemistry，2004，84：475.
[8] LU Yong-xin，LI Yao-hui，CUI Zhou-ping(卢永欣，李耀辉，崔周平)· Lit. & Information Med.(预防医学文献信息)，2002，2(8)：48.
[9] HAN Yu-zhi，QIAO Hong-xia(韩玉之，乔洪霞)· Heilongjiang Information of Science and Technology(黑龙江科技信息)，2007 4：16.
[10] LU Feng，LE Jian，CHEN Gui-liang，et al(陆　峰，乐　建，陈桂良，等)· China Journal of Chinese Materia Medica(中国中药杂志)，2007，32(1)：6.
[11] LU Feng，LI Shu，LE Jian，et al· Analytica Chimica Acta，2007，589(2)：200.
[12] Alfred A Christy，Sumaporn Kasemsumran，Yiping Du· Analytical Science，2004，20：935.
[13] LI Wang，Frank S C Lee，Wang Xiaoru· Food Chemistry，2006，95：529.
[14] Galtier O，Dupuy N，Ollivier D· Analytica Chimica Acta，2007，595：136.
[15] YANG Nan-lin，CHENG Yi-yu，QU Hai-bin(杨南林，程翼宇，瞿海斌)· Chinese Journal of Analytical Chemistry(分析化学)，2003，31(6)：664.
[16] ZHANG Ling，YU Jian(张　玲，于　健)· The Food Industry(食品工业)，2006，27(4)：27.
[17] WANG Jia-jun，LIANG Yi-zeng(王家俊，梁逸曾)· Chinese Journal of Analytical Chemistry(分析化学)，2005，33(6)：793.

Discriminating and Quantifying Potential Adulteration in Virgin Olive Oil by Near Infrared Spectroscopy with BP-ANN and PLS

WENG Xin-xin[1]，LU Feng[1]，WANG Chuan-xian[2]，QI Yun-peng[1]*
1. Second Military Medical University，Shanghai　200433，China
2. Shanghai Exit-Entry Inspection and Quarantine Bureau，Shanghai　200135，China

Abstract In the present paper，the use of near infrared spectroscopy (NIR) as a rapid and cost-effective classification and quantification techniques for the authentication of virgin olive oil were preliminarily investigated· NIR spectra in the range of $12\ 000$-$3\ 700\ cm^{-1}$ were recorded for pure virgin olive oil and virgin olive oil samples adulterated with varying concentrations of sesame oil，soybean oil and sunflower oil (5%-50% adulterations in the weight of virgin olive oil)· The spectral range from $12\ 000$ to $5\ 390\ cm^{-1}$ was adopted to set up an analysis model· In order to handle these data efficiently，after pretreatment，firstly，principal component analysis (PCA) was used to compress thousands of spectral data into several variables and to describe the body of the spectra，and the analysis suggested that the cumulate reliabilities of the first six components was more than 99.999%· Then ANN-BP was chosen as further research method· The six components were secondly applied as ANN-BP inputs· The experiment took a total of 100 samples as original model examples and left 52 samples as unknown samples to predict· Finally，the results showed that the 52 test samples were discriminated accurately· And the calibration models of quantitative analysis were built using partial-least-square (PLS)· The R values for PLS model are 98.77，99.37 and 99.44 for sesame oil，soybean oil and sunflower oil respectively；the root mean standard errors of cross validation (RMSECV) are 1.3，1.1 and 1.04 respectively· Over-

all, the near infrared spectroscopic method in the present paper played a good role in the discrimination and quantification, and offered a new approach to the rapid discrimination of pure and adulterated virgin olive oil.

* Corresponding author

2010 Winter Conference on Plasma Spectrochemistry
Fort Myers, Florida, January 4-9, 2010
Plasma Spectrochemistry for Trace Element, Stable Isotope, and Elemental Speciation Analyses

The 2010 Winter Conference on Plasma Spectrochemistry, sixteenth in a series of biennial meetings sponsored by the ICP Information Newsletter, features developments in plasma spectrochemical analysis by inductively coupled plasma (ICP), dc plasma (DCP), microwave plasma (MIP), glow discharge (GDL, HCL), and laser sources. The meeting will be held Monday, January 4 through Saturday, January 9, 2010, in Fort Myers, Florida (www.fortmyers-sanibel.com) at the Sanibel Harbour Resort and Spa (www.sanibel-resort.com). Continuing education short courses at introductory and advanced levels and manufacturers' seminars will be offered Saturday through Monday, January 2 to 4. Spectroscopic instrumentation and accessories will be shown during a three-day exhibition from January 5 to 8, and a Workshop on New Plasma Instrumentation will be presented on Tuesday, Wednesday, and Thursday afternoons. A golf tournament is planned for Sunday, January 3.

Objectives and Program

The continued growth in popularity of plasma sources for atomization and excitation in atomic spectroscopy and ionization in mass spectrometry and the need to discuss recent developments of these discharges in spectrochemical analysis stimulated the organization of this meeting. The Conference will bring together international scientists experienced in applications, instrumentation, and theory in an informal setting to examine recent progress in the field. Approximately 600 participants from 30 countries are expected to attend. Over 300 papers describing applications, fundamentals, and instrumental developments with plasma sources will be presented in lecture and poster sessions by more than 200 authors. Symposia organized and chaired by recognized experts will include the following topics: 1) Sample introduction and transport phenomena; 2) Micronebulization and flow processing spectrochemical analysis; 3) Elemental speciation and sample preparation for speciation; 4) Plasma instrumentation, including chemometrics, expert systems, on-line analysis, microplasmas, software, and remote-system automation; 5) Sample preparation, treatment and automation, 6) Excitation mechanisms, plasma phenomena and modeling; 7) Spectroscopic standards and reference materials, and high-purity materials; 8) Plasma source mass spectrometry, 9) Glow discharge atomic and mass spectrometry, 10) Stable isotope analyses, and 11) Laser-assisted plasma spectrometry. Six plenary and 26 invited lectures will highlight advances in these areas. Four afternoon poster sessions will feature applications, automation, and new instrumentation. Six Heritage Lectures will be presented by distinguished scientists and investigators, who have contributed significantly to the development of plasma spectrochemistry and will address critical development areas in sample introduction, instrumentation, elemental speciation, plasma source mass spectrometry, and novel software and hardware. Plenary, invited, and submitted papers will be published in September 2010 after peer review as the official Conference proceedings.

Schedule of Activities, Deadlines

Call for Papers, Abstracts Due, Early Bird Registration   Friday, September 18, 2009
Exhibitor Booth Reservation and Pre-Registration Deadlines   Friday, September 18, 2009
Final Abstracts for All Paper Deadline   Friday, October 16, 2009
Exhibitor Reservation Deadline   Friday, October 16, 2009
Conference Pre-Registration Deadline   Friday, October 16, 2009
Hotel Reservation Deadline   Monday, December 7, 2009
Late Pre-Registration Deadline   Friday, December 11, 2009
2010 Winter Conference Short Courses   Saturday - Monday, January 2-4, 2010
2010 Manufacturers' Seminars   Saturday-Sunday, January 2-3, 2010