

人工神经网络用于直接化学电离质谱分析食用油品质的研究

周志权¹ 张婷婷¹ 贾滨² 欧阳永中² 方小伟² 陈焕文^{* 2}

¹(哈尔滨工业大学(威海)信息工程研究所,威海 264209)

²(江西省质谱科学与仪器重点实验室,东华理工大学,南昌 330013)

摘 要 无需任何样品预处理,采用表面解吸常压化学电离质谱(DAPCI-MS)技术直接对涂覆在载玻片表面的食用油样品和地沟油样品进行检测,快速获得了不同油类样品的质谱信号;并运用改进的反向传输(BP)人工神经网络对 DAPCI-MS 所得到的油类样品质谱数据进行有监督的分类识别,建立多分组预测模型。结果表明: DAPCI-MS 能够承受食用油中复杂基体的影响,可对油类样品进行直接快速质谱分析;误差反转(BP)神经网络具有良好的分类判别能力,对食用油品质谱数据识别效果比较理想,能够在地沟油和非地沟油样品进行有效区分的同时,实现对不同品种的食用油的分离及分类判别。本方法分析速度快,信息提取准确,识别精度高,对快速质谱技术结合神经网络在该领域的应用以及食用油品质的快速鉴定具有重要的借鉴意义。

关键词 表面解吸常压化学电离;质谱;反向传输人工神经网络;食用油;地沟油

1 引言

近年来,“地沟油”成为了人们关注的热点之一。“地沟油”因受污染、变质等原因产生的黄曲霉素等有害物质会导致腹泻、腹痛等疾病,甚至诱发癌变^[1]。因此,对地沟油进行快速检测和识别具有重要意义。常规地沟油检测主要有水分含量测定法、酸价测定法、胆固醇含量测定法、近红外光谱法、电导率与极性物质测定法、重金属含量测定法^[1,2]。上述方法操作较为复杂,费时费力,并且由于主要针对地沟油中的特殊成分进行检测,无法对合法食用油类(如大豆油、花生油等)进行识别。

表面解吸常压化学电离质谱(DAPCI-MS)技术无需有毒化学试剂,即可对样品完成无创、无污染的检测,在食品安全领域具有独特优势^[3,4]。人工神经网络现已成为解决化学问题的重要化学计量学手段,其中反向传输(BP)人工神经网络是目前应用最广泛的神经网络模型之一^[5,6]。目前已有大量近红外光谱与神经网络结合的分组预测模型的成功应用的报道^[7,8],但质谱与神经网络结合的分析识别应用的报道鲜见。

本研究采用 DAPCI-MS 技术对地沟油和 3 种主要食用油类(大豆油、花生油和油菜籽油)进行快速质谱分析,同时利用改进的 BP 神经网络作为识别模型,实现了对 DAPCI-MS 所得到的油类样品质谱数据的自动辨识。本方法分析速度快、信息提取准确、识别精度高,对油类样品质谱数据具有较好的识别效果。

2 基本理论

标准的 BP 算法使用最速梯度下降法进行误差调整,其权值的修正是沿着误差性能函数梯度的反方向进行的:

$$x(k+1)=x(k)-\alpha g(k) \tag{1}$$

其中, $x(k)$ 为第 k 次迭代各层之间的连接权向量或阈值向量; α 为学习速率; $g(k)=\frac{\partial E(K)}{\partial x(k)}$ 为第 k 次迭代的神经网络输出误差对各个权值或阈值的梯度向量。由于 BP 神经网络的隐层采用 Sigmoid 型传输函数,在输入量很大或很小时,输出函数的斜率趋近于零。因此在权值和阈值距离最佳值甚远时,权值和阈值的修正量也很小。所以标准的 BP 算法存在收敛速度慢,训练时间长等不足^[9~11]。

2010-12-12 收稿;2011-04-18 接受

本文系科技部创新方法专项基金(No. 2008IM 040400)和哈尔滨工业大学研究基金(No. HIT WHXB200803)资助

* E-mail: chw8868@gmail.com

BP 算法改进的主要目标是加快训练速度,消除梯度幅度的不利影响。算法的改进基于对表现函数梯度加以分析:权值的修正值取决于 $g(k)$ 的正负号,而 $g(k)$ 的幅度与权值的修正无关^[14]。权值或阈值 $x(k)$ 的第 $k+1$ 次迭代的幅度修正值 $\Delta x(k+1)$ 的算法描述如下^[12]:

$$\Delta x(k+1)=\begin{cases}\Delta x(k)\cdot k_{inc}\cdot \text{Sign}(g(k)) & (\text{当连续两次迭代的梯度方向相同}) \\ \Delta x(k)\cdot k_{dec}\cdot \text{Sign}(g(k)) & (\text{当连续两次迭代的梯度方向相反}) \\ \Delta x(k) & (\text{当 } g(k)=0)\end{cases}\tag{2}$$

当连续两次迭代的梯度方向相同时,修正值随增量因子 k_{inc} 增加;当连续两次迭代的梯度方向相反时,修正值随减量因子 k_{dec} 减小;当梯度为零时,修正值保持不变。改进以后,如果权值在相同的梯度上被连续修正,则增加权值变化的量级,克服了幅度偏导带来的不利影响,加快了训练速度。

3 实验部分

3.1 仪器与试剂

DAPCI 离子源(东华理工大学研制^[15,16]); LTQ-XL 增强型线性离子阱质谱仪(美国 Finnigan 公司),配有 Xcalibur 数据系统。甲醇(色谱纯);食用油样品(采购自当地超市);地沟油样品(当地工商部门提供)。

3.2 质谱分析

将未经任何样品预处理的油样品直接涂覆在干净的载玻片上,待其扩散成较薄的一层油膜后,放置在 DAPCI 离子源下,直接进行质谱分析。设置 DAPCI 离子源为负离子检测模式,质量范围为 50~300 Da,电离电压 3.6 kV,离子传输管温度 150 ℃,放电针与水平面夹角约 43°。通过针尖电晕放电产生的大量初级离子,将承载在载玻片上的食用油样品解吸和电离,形成的离子引入质谱进行分析。其它实验参数由系统自动优化。

3.3 样本的使用

为了减少样本分布带来的偶然性,每次从每类油品的 50 个样本中按次序取出 30 个(4 类共 120 个)样本参与训练,其余 20 个(4 类共 80 个)样本参与测试,共进行 6 次训练和测试。采用基于类内类间距离的可分离性判据,对质谱仪记录的样品数据进行特征选择与提取^[17]。将选定的 26 个特征向量归一化转换得到的相对丰度输入到改进的 BP 神经网络中,建立两层的网络模型。

4 结果与讨论

4.1 油类样品的 DAPCI-MS 分析

DAPCI-MS 实验记录了同种品牌的大豆油、花生油、油菜籽油样品和地沟油样品的质谱图(图 1)。由于地沟油在其变质、非法处理等过程中形成了较多小分子有毒物质,在 m/z 50~300 质量范围内,质谱信号较为丰富,信号强度较高,3 种食用油的谱图相似度较高,差别较小。由于 3 种食用油的油料不同,其主要营养成分及微量物质的含量存在差异,其质谱图中信号间的相对丰度存在细微差别,并具有独自的特征。

采用主成分分析(PCA)法处理质谱数据,只能区分地沟油和非地沟油,不能有效区分开 3 种食用油。

如图 2 所示,本实验将食用油样品连续进样 11 次,相对标准偏差(RSD)为 7.2%,平均一个样品的分析时间约 16 s。表明 DAPCI-MS 能够实现对食用油样品的高通量分析,检测效果较为稳定。

4.2 BP 神经网络分析

4.2.1 BP 神经网络的设计 在本实验中,网络的输入节点为 26 个,在地沟油与正常食用油的分类识别中,输出量的表示采用 n 中取 1 表示法(10 和 01),输出节点为 2 个;在 4 种油品的分类识别中,输出量的表示采用 $n-1$ 表示法(000,001,010 和 100),输出节点为 3 个。网络隐层节点数的确立根据公式(3):

$$m=\sqrt{n+l}+\alpha\tag{3}$$

其中, m 为隐层节点数, n 为输入层节点数, l 为输出层节点数, α 取值范围 1~10,则隐层节点数 m 的取值范围 6~15。通过多次实验并选取最优分类结果,最终确定网络隐层节点数分别为 7 和 13, $\Delta x(0)$ 为 0.07;增量因子 k_{inc} 为 1.3;减量因子 k_{dec} 为 0.4。上述数据处理方法均在 Matlab 7.8 中编程实现。

4.2.2 地沟油与正常食用油的分类识别 为实现地沟油样品与非地沟油样品的快速鉴别,将样品分为

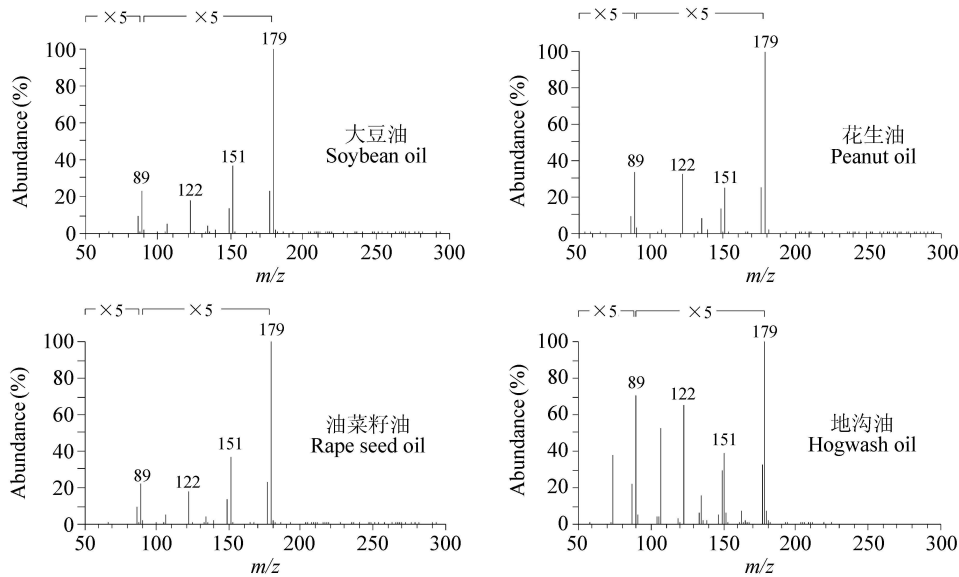


图 1 地沟油与食用油样品的 DAPCI-MS 谱图

Fig. 1 Mass spectra of oil samples recorded using surface desorption atmospheric pressure chemical ionization (DAPCI)-MS

两类供学习。设置地沟油的目标输出为(10), 非地沟油(大豆油、花生油、油菜籽油)的目标输出为(01), 上述油品的质谱数据对应图 1 所示谱图。网络分类以及预测所得结果映射到二维平面如图 3 所示, 每类结果的判决区间由相同类型的判决线决定。网络的分类识别结果列于表 1。经计算, 训练样本预测结果准确率为 100. 0%; 测试样本预测结果准确率为 98. 5%, 地沟油样品与非地沟油样品被明显区分, 表明本方法能够有效地对地沟油样品进行分类和鉴定。

4. 2. 3 4 种油品的分类识别 为了进一步考察 3 种食用油(大豆油、花生油和油菜籽油)之间的差异, 将样品分为 4 类供网络学习。设置大豆油、花生油、油菜籽油和地沟油的目标输出分别为(000), (100), (010) 和

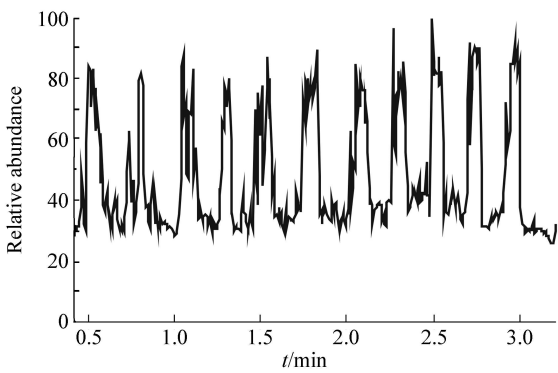


图 2 检测食用油样品的 DAPCI-MS 离子流图

Fig. 2 Total ion chromatogram of ions recorded from oil samples

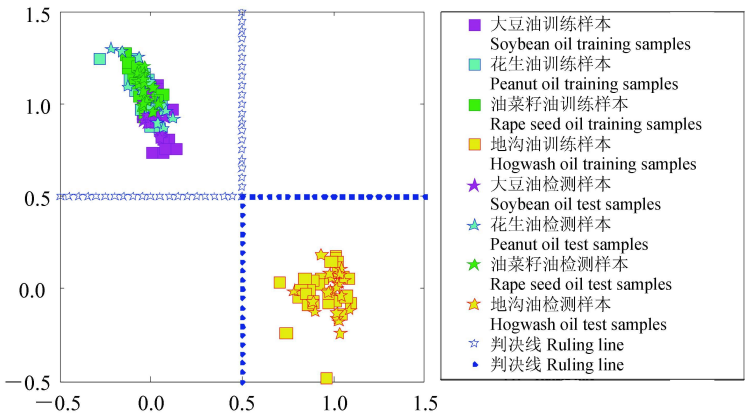


图 3 BP 神经网络地沟油与非地沟油的分类识别结果二维平面图

Fig. 3 2-D Classification result of hogwash oil and edible oil based on BP neural network

表 1 BP 神经网络地沟油与非地沟油的分类识别结果
Table 1 Classification result of hogwash oil and edible oil based on BP neural network

样本集 Data set	Group 1 (%)	Group 2 (%)	Group 3 (%)	Group 4 (%)	Group 5 (%)	Group 6 (%)	Total (%)
训练样本 Trainingset	100.0	100.0	100.0	100.0	100.0	100.0	100.0
测试样本 Test set	97.5	100.0	96.3	98.8	100.0	98.8	98.5

(001), 上述油品的质谱数据对应图 1 所示谱图。4 种油品的网络分类及预测结果映射到二维平面如图 4 所示, 每类结果的判决区间由相邻两条判决线决定。网络的分类识别结果列于表 2。经计算, 训练样本预测结果准确率为 98.0%, 测试样本预测结果准确率为 85.8%, 表明本方法能够有效地对不同种类的食用油进行分类和鉴定。

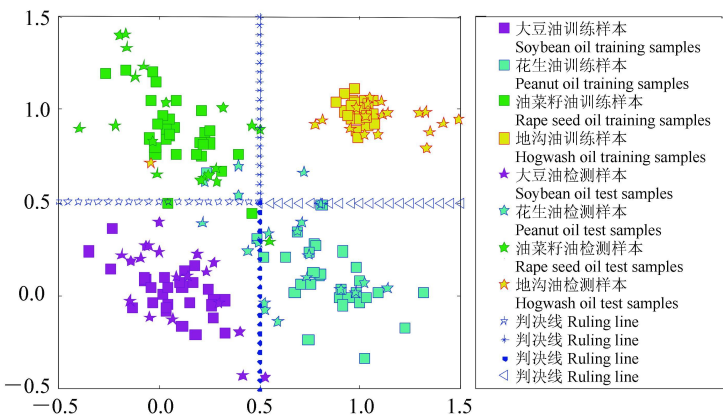


图 4 BP 神经网络 4 种油品的分类识别结果二维平面图
Fig. 4 2-D Classification result of oil samples based on BP neural network

表 2 BP 神经网络 4 种油品的的分类识别结果
Table 2 Classification result of oil samples based on BP neural network

样本集 Data set	Group 1 (%)	Group 2 (%)	Group 3 (%)	Group 4 (%)	Group 5 (%)	Group 6 (%)	Total (%)
训练样本 Trainingset	99.0	99.0	100.0	100.0	92.5	97.5	98.0
测试样本 Test set	85.0	87.5	86.3	83.8	82.5	90.0	85.8

改进的 BP 神经网络方法能够利用 DAPCI-MS 数据对大豆油、花生油和油菜籽油进行有效分类和鉴别。不同种类的食用油,因其油料种类等的不同,所含微量化合物的种类及相对浓度也必然存在差别。该细微差别所导致的质谱信号差异,能够被 DAPCI-MS 检测和记录,并通过神经网络的计算展现出来。因此 DAPCI-MS 能够抵抗油类基体的干扰,并对其中的微量物质进行分析。

研究结果表明, 将改进的 BP 神经网络与 DAPCI-MS 相结合, 能够在无需任何样品预处理的条件下, 直接对食用油类样品进行高通量分析, 并实现了对地沟油的快速准确鉴别。同时 DAPCI-MS 数据的神经网络计算结果可以实现对大豆油、花生油和油菜籽油的分类识别, 检测结果理想。本方法分析速度快、信息提取准确、识别精度高。

References

1 WANG Le, LIU Yao-Gang, CHEN Feng-Fei, HU Jian-Hua(王 乐, 刘尧刚, 陈凤飞, 胡建华). *J. Wuhan Polytech. Univ.* (武汉工业学院学报), **2007**, 26(4): 1~ 12
2 WANG YAO, YIN Ping-He(王 耀, 尹平河). *Chinese J. Anal. Lab* (分析试验室), **2006**, 25(3): 92~ 94
3 Wu Z C, Chen H W, Wang W L, Jia B, Yang T L, Zhao Z F, Ding J H, Xiao X X. *J. Agric. Food Chem.*, **2009**, 57(20): 9356~ 9364

- 4 YANG Shui-Ping, CHEN Hua-Wen, YANG Yi-Ling, HU Bin, ZHANG Xie, ZHOU Yi-Fen, Zang Li-Li, Gu Hai-Wei(杨水平, 陈焕文, 杨宇玲, 胡斌, 张燮, 周瑜芬, 张丽丽, 顾海威). *Chinese J. Anal. Chem.* (分析化学), **2009**, 37(3): 315~318
- 5 XU Lu, HU Chang-Yu(许禄, 胡昌玉). *Comput. Appl. Chem.* (计算机与应用化学), **2000**, 17(2): 145~147
- 6 Marengo E, Bobba M, Robotti E, Lenti M. *Anal. Chim. Acta*, **2004**, 511(2): 313~322
- 7 WENG Xin-Xin, LU Feng, WANG Chuan-Xian, QI Yun-Peng(翁欣欣, 陆峰, 王传现, 亓云鹏). *Spectrosc. Spect. Anal.* (光谱学与光谱分析), **2009**, 29(12): 3283~3287
- 8 TANG Song-Peng, YAO Xin-Feng, YAO Xia, TIAN Yong-Chao, CAO Wei-Xing, ZHU Yan(汤守鹏, 姚鑫锋, 姚霞, 田永超, 曹卫星, 朱艳). *Chinese J. Anal. Chem.* (分析化学), **2009**, 37(10): 1145~1450
- 9 Sun S P, Yi D Q, Jiang Y. *Mater. Chem. Phys.* **2011**, 126(3): 632~641
- 10 Oh S H. *Neurocomputing*, **2011**, 74(6): 1058~1061
- 11 Phua P K H, Ming D H. *IEEE Trans. Neural Networks*, **2003**, 14(6): 1460~1467
- 12 ZHOU Kai-Li, KANG Yao-Hong(周开立, 康耀红). *Neural network model and its MATLAB simulation program*(神经网络模型及其 MATLAB 仿真程序). Beijing(北京): Tsinghua University Press(清华大学出版社), **2005**: 69~100
- 13 Gunturkun R J. *Med. Sys.*, **2010**, 34(4): 493~497
- 14 Mastorocostas P A. *Electron. Lett.*, **2004**, 40(1): 57~58
- 15 YANG Shui-Ping, HU Bin, LI Jian-Qiang, HAN Jing, ZHANG Xie, CHEN Hua-Wen(杨水平, 胡斌, 李建强, 韩京, 张燮, 陈焕文). *Chinese J. Anal. Chem.* (分析化学), **2009**, 37(5): 691~694
- 16 Yang S P, Ding J H, Zheng J, Hu B, Li J Q, Chen H W, Zhou Z Q, Qiao X L. *Anal. Chem.*, **2009**, 81(7): 2426~2436
- 17 BIAN Zhao-Qi, ZHANG Xue Gong(边肇祺, 张学工). *Pattern Recognition*(模式识别). Beijing(北京): Tsinghua University Press(清华大学出版社), **2000**: 176~208

Surface Desorption Atmospheric Pressure Chemical Ionization Mass Spectrometry for Edible Oil Analysis Based on Back Propagation Neural Networks

ZHOU Zhi-Quan¹, ZHANG Ting-Ting¹, JIA Bin², OUYANG Yong-Zhong²,
FANG Xiao-Wei², CHEN Hua-Wen^{*2}

¹(Institute of Information Engineering, Harbin Institute of Technology at Weihai, Weihai 264209)

²(Jiangxi Key Laboratory for Mass Spectrometry and Instrumentation,
East China Institute of Technology, Nanchang 330013)

Abstract Without any sample pretreatment, oil samples smeared on slide were directly detected by surface desorption atmospheric pressure chemical ionization mass spectrometry (DAPCI-MS), the mass spectra rapidly recorded by DAPCI-MS were subjected to data processing for classification using improved BP(Back Propagation) neural networks. The results showed that DAPCI-MS was a practically convenient tool for edible and hogwash oil detection without much matrix effect. The improved BP neural network can be applied to the simultaneous determination of hogwash standard oil and different kinds of edible oil samples. The data demonstrated that the DAPCI-MS combined improved BP neural network methods was a promising technique for edible oil rapid identification with expedite convergence pace and superior prediction precision.

Keywords Surface desorption atmospheric pressure chemical ionization; Mass spectrometry; Back propagation neural network; Edible oil; Hogwash oil

(Received 12 December 2010; accepted 18 April 2011)