第19卷 第3期
2006年6月

传感技术学报
CHINESE JOURNAL OF SENSORS AND ACTUATORS

Vol.19 No.3
Jun.2006

Studies on Signal Feature Extraction and Sensor Optimization of an Electronic Nose

HAI Zheng, *WANG Jun*

(*College of Biosystems Engineering and Food Science*, *Zhejiang University*, *Hangzhou* 310029, *China*)

Abstract: An electronic nose was used to detect the adulteration of sesame oil with corn oil. In order to reduce the dimension of the data matrix, three different variable selection techniques were employed: PCA, step-LDA and Fisher linear transformation. And then the pattern recognition technique of linear discriminant analysis (LDA) and artificial neural network (ANN) were used to check the effect of the three dimension reduction methods mentioned above. In the process of LDA, the Fisher linear transformation is most effective, the error rate was 0.61%, only 1 sample was misclassified; in the process of ANN, the results obtained using the ten variables selected by step-LDA were more acceptable than others, the 95% individual confidence interval is (−4.71%, 3.38%), the mean square of error was 4.75, and the correlation between predicted concentration and genuine concentration is 0.998 08.

Key words: electronic nose; principal component analysis; stepwise linear discriminant analysis; BP-neural network.

EEACC: 7230L

电子鼻信号特征提取与传感器优化的研究

海 铮，王 俊

（浙江大学生物系统工程与食品科学学院，杭州 310029）

摘 要：采用PEN2型电子鼻系统对芝麻油的玉米油掺假进行定性鉴别和定量预测，运用主成分分析，逐步判别分析和Fisher线性判别函数变换对原始数据进行预处理，从而降低原始数据空间的维数，并用判别分析与人工神经网络对数据进行进一步分析，考察了不同的数据预处理方法的效果。判别分析结果表明，采用Fisher线性判别函数变换所得到的十个变量判别能力最强，误判率为0.61%，仅有1个样品出现误判。在BP神经网络的定量预测中，采用逐步判别分析所筛选出的十个变量作为网络输入，所得的预测结果最为理想，绝对误差个体值的95%置信区间最小，为(−4.71%，3.38%)，均方误差为4.75，预测值与实际值之间有极显著的相关性，相关系数 $R=0.998\ 08$。

关键词：电子鼻；主成分分析；逐步判别分析；BP神经网络

中图分类号：TP212.6　　**文献标识码**：A　　**文章编号**：1004-1699(2006)03-0606-05

电子鼻是一种由具有部分选择性的气敏传感器阵列和适当的模式识别系统组成，能识别简单或复杂气体的仪器。1982年，英国Warwick大学的学者Persuad和Dodd教授用3个商品化的SnO_2气体传感器模拟哺乳动物嗅觉系统中的多个嗅觉感受器细胞对多种复杂的有机挥发气进行了类别分析，开辟电子鼻研究之先河。此后，随着相关学科的发展，电子鼻的研究迅速发展。目前，在食品品质评价[1-2]、环境监测[3]以及疾病诊断[4]等方面，国内外都作了不少的研究。

收稿日期：2005-07-27

基金项目：国家教育部新世纪人才支持计划资助(NCET-04-0544)；国家自然科学基金资助(30571706)

作者简介：海 铮(1980-)，男，硕士研究生，研究方向电子鼻的应用技术，sea333355@yahoo.com；

王 俊(1965-)，男，博士，教授，博士生导师，从事农产品品质电子鼻检测技术，jwang@zju.edu.cn

电子鼻系统中的气敏传感器阵列一般由 8-32 个传感器组成，并在多个时刻读入数据，所以，对于每个样品所获得的数据都在几百个以上。在这些数据中，有很多数据之间存在着极其显著的相关性，因而提供的信息也相互重叠；此外，由于某些传感器对样品气体不敏感，或对环境因子过于敏感，造成信号响应紊乱，不能提供有用信息。大量冗余信息的引入，会导致模式识别过程耗费时间长，计算精度低，系统稳定性差。对传感器所获得的原始数据进行适当的特征选择与特征提取是极为重要的，不但有利于去除数据中的冗余信息，减少进行模式识别的计算时间，而且可以省去那些对模式识别效果没有显著影响甚至有负面影响的传感器，从而对降低电子鼻的制造成本，提高系统稳定性都有一定的积极意义。在食用油的品质[5,6]检测方面，国外已有一些成功的范例。本文用电子鼻对芝麻油、大豆油及两者不同比例的混合物进行检测，采用主成分分析、逐步判别和 Fisher 线性函数变换三种方法对原始数据进行预处理，降低数据空间的维数，并用判别分析和人工神经网络对数据进行进一步分析，以考察和比较数据经不同的预处理后模式识别的效果。

1　实验材料、仪器与方法

1.1　实验材料

本实验采用超市选购的金龙鱼芝麻油和玉米油，并在实验前密封避光贮藏，防止其氧化变质。

1.2　实验仪器

在本实验中，采用的电子鼻系统是德国 AIRSENSE 公司的 PEN2 便携式电子鼻(Portable Electronic nose)。PEN2 电子鼻包含 10 个金属氧化物传感器阵列，各个传感器的名称及性能描述见表 1。本电子鼻系统所获取的数据是样品气体经过传感器时其电导率 G 与经活性炭过滤后的标准气体经过传感器时其电导率 G_0 的比值，即 G/G_0。系统组成主要包含下面几个部分：传感器阵列、采样通道，内置泵，控制单元和计算机。传感器工作温度为 300℃。

表 1　PEN2 的各传感器名称及性能特点

阵列序号	传感器名称	性能描述	备注
1	W1C	对芳香成分灵敏	甲苯，10 mL/m³
2	W5S	灵敏度大，对氮氧化合物很灵敏	NO_2，1 mL/m³
3	W3C	对氨水、芳香成分灵敏	苯，10 mL/m³
4	W6S	主要对氢气有选择性	H_2，100 mL/m³
5	W5C	对烷烃，芳香成分灵敏	丙烷，1 mL/m³
6	W1S	对甲烷灵敏	CH_3，100 mL/m³
7	W1W	对硫化物灵敏	H_2S，1 mL/m³
8	W2S	对乙醇灵敏	CO，100 mL/m³
9	W2W	对芳香成分、有机硫化物灵敏	H_2S，1 mL/m³
10	W3S	对烷烃灵敏	CH_3，10 mL/m³

1.3　实验方法及数据的获取

芝麻油与玉米油按不同比例混合，并用搅拌器搅拌均匀，其中玉米油所占比例(体积)分别为 10%、20%、30%、40%、50%、60%、70%、80%、90%。九类混合样品与芝麻油和玉米油纯样品共 11 类样品，每类样品分装于 15 支相同的带胶塞的玻璃瓶中，每瓶 15 mL。盖紧瓶盖，静置 1 h，使其顶部气体成分稳定后，采用顶空抽样进行测量。图 1 所示为芝麻油与玉米油的传感器电导率比值随时间变化的曲线图。

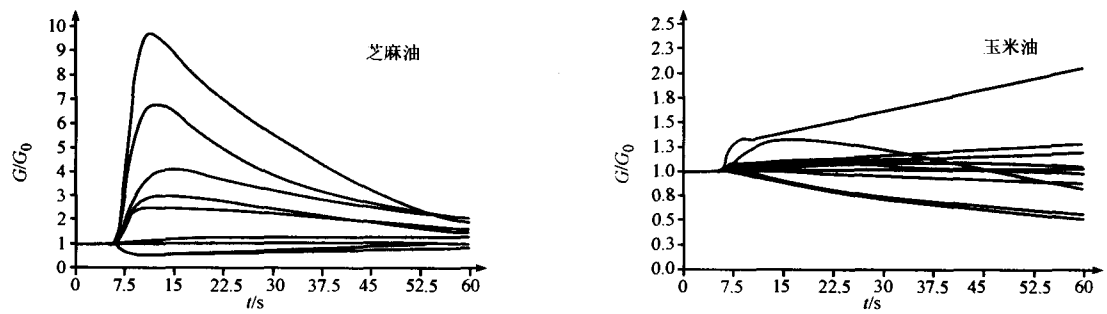

图 1　芝麻油和玉米油的传感器响应曲线图

1.4 数据获取与特征提取

电子鼻抽样检测时间为 60 s,但多数数据之间高度相关,所以只选取较有代表性的 15、30 和 60 s 的数据作为原始数据进行分析。

1.4.1 主成分分析(PCA)

主成分分析是将多个指标化为较少的几个综合指标的一种统计方法,在回归分析、聚类分析等过程中是一种简化数据结构的有力工具。综合指标,即主成分,是原来多个指标的线性组合。虽然这些综合指标不能直接观察,但它们之间互不相关,能反映原来多指标的信息。通常,对原始的 n 维数据进行主成分分析,可以得到 n 个主成分,然而在这 n 个主成分中,只用特征值较大的几个主成分即可反映原始数据中的大部分信息。将特征向量较小,即方差贡献率接近于零的主成分删除,对余下的主成分(通常累计方差贡献率大于 85%),进行进一步的分析或处理,便可以在不丢失或很少丢失原始信息的前提下减少分析变量的数目[7]。

1.4.2 逐步判别分析(Step-LDA)

逐步判别与逐步回归的基本思想相似,都采用有进有出的算法,即每一步都进行检验,把一个最重要的变量选入判别式,同时也考虑较早进入判别式的某些变量,如果其重要性也随着其后一些变量的选入而变化,已失去原有的重要性时,应把它及时地从判别式中剔除出去,使判别式中仅仅保留重要的变量,以达到对原有数据向量降维的目的。

1.4.3 Fisher 线性函数变换

Fisher 判别分析是一种基于线性映射的判别方法。这种方法实际上是致力于寻找一个最能反映组与组之间差异的投影方向,即寻找线性判别函数:

$$Y = C_1X_1 + C_2X_2 + \cdots + C_pX_p$$

使得所有的样品经线性变换后的函数值 Y 应满足组内离差平方和最小,而组间离差平方和最大。通常,满足上述条件的线性函数有 $k(k \leqslant p)$ 个,当组数太大,讨论的指标太多,则一个判别函数是不够的,这时需要寻找第二个,甚至第三个线性判别函数。从而,选取判别效率最大的前 $n(\leqslant k)$ 各判别函数,并使累计判别效率 $\geqslant 85\%$,即可在最大限度地降低映射过程中的数据信息丢失的同时,使组间离差与组内离差的比值达最大,也就使得不同类别的样品更易于区分[8]。本文即借鉴 Fisher 判别分析中这种线性映射方法对原始数据进行线性变换,达到特征提取的目的。

1.5 模式识别方法

用于电子鼻系统的模式识别方法主要有两类,一类是基于多元统计分析的方法,另一类是基于网络的分析方法。本文分别采用多元统计中线性判别函数分析对样品进行定性判别,采用网络分析中 BP 神经网络对样品进行定量预测,从而考察和比较数据经不同的预处理后的判别能力和预测效果。

1.5.1 线性判别函数分析(LDA)

线性判别函数分析是一种常规的模式识别和样品分类方法,在医学诊断、指纹识别等领域都有广泛的应用,同时也是在电子鼻系统的一种重要的模式识别方法,在很多实验中已取得了良好的效果。线性判别函数分析假定数据的条件概率呈正态分布,并且各组样品有相同的协方差矩阵。此方法通过计算未知样品与各类已知样品的马氏距离,并考虑到样品归属的先验概率,针对每一组样品计算出一个以原有变量为基础的线性判别函数,用以计算未知样品归属的后验概率。线性判别函数分析也因此而得名[9]。

1.5.2 人工神经网络(ANN)

人工神经网络是在人类对其大脑神经网络认识理解的基础上人工构造的能够实现某种功能的神经网络。它是理论化的人脑神经网络的数学模型,是基于模仿大脑神经网络结构和功能而建立的一种信息处理系统。它实际上是由大量简单元件相互连接而成的复杂网络,具有高度的非线性,能够进行复杂的逻辑操作和非线性关系实现的系统。基于误差反向传播算法(BP 算法)的 BP 神经网络是一种多层前向网络,采用 Sigmoid 型函数作为传递函数,可以实现输入和输出间的任意非线性映射,是目前应用最为广泛的一种神经网络,在诸如函数逼近、模式识别、数据压缩等领域有着广泛的应用[10]。在电子鼻系统中,BP 神经网络也是一种很好的模式识别方法,有着广泛的应用并取得了很好的效果。

2 结果与分析

2.1 特征提取

采用 Matlab 7.0 软件对原始数据进行主成分分析,获得样品的各主成分得分值以及各主成分的方差贡献率和累计方差贡献率,对方差贡献率为零或接近于零的主成分予以删除,所余前十个主成分,其累计方差贡献率为 99.9%。采用 SAS8.0 软件进行逐步判别分析筛选变量,选取判别能力较大的前十个变量,分别为 15 s 时的传感器 1、3、4、5、6、8、9,30 s 时的传感器 1、4、6,共涉及到七个传感器。对原始数据进行 Fisher 线性函数变换,取前十个线性函数,使数据空间降至十维,累计判别效率为

100%。

2.2　线性判别分析

交互验证法(cross-validation)是 SAS 软件中提供的一种判别方法，其具体做法是：设资料中共有 n 个样品，每次留下一个样品作为未知样品，用(n－1)个样品建立判别函数，将留下的样品代入判别函数，判别其归属，这样就有利于消除用全部数据建立的判别函数再对样品进行回代时所产生的偏性。因此，在本次实验中，采用交互验证法对 165 个样品进行判别分析。进行分析的变量分别为主成分分析所选入的十个变量(PCA)，逐步判别分析所选入的十个变量(SLDA)，Fisher 线性判别函数变换后的十个变量(FLT)，以及 15 s、30 s、60 s 时的数据和原始数据。各组数据的判别结果见表 2。

表 2　各类样品的误判率　　单位：(%)

变量	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	total
PCA	0	0	0	0	0	6.67	0	26.67	6.67	6.67	0	4.24
SLDA	0	0	0	0	0	6.67	0	6.67	6.67	6.67	0	2.42
Fisher	0	0	0	0	0	6.67	0	0	0	0	0	0.61
15 s	0	0	0	0	13.33	6.67	0	13.33	6.67	6.67	0	4.24
30 s	6.67	0	0	0	6.67	6.67	0	40	13.33	0	0	6.67
60 s	0	0	0	0	13.33	6.67	6.67	26.67	13.33	0	0	6.06
原始数据	6.67	0	0	0	0	6.67	0	6.67	0	0	0	1.82

由表 2 中各类样品的误判率可以看出，选用逐步判别法，Fisher 线性判别函数变换以及初始的 30 个变量进行判别分析所得的结果较为理想。选用初始的 30 个变量进行判别分析，误判率为 1.82%，即有 3 个样品被判错；选用逐步判别函数法筛选出的 10 个变量进行判别误判率为 2.42%，即有 4 个样品被判错，与选用全部变量进行判别分析所得的结果相差不多；选用其他的变量组合进行判别分析，误判率均大于 4%，被误判的样品个数均在 7 个以上。而选用 Fisher 线性判别函数变换后的数据进行判别分析，误判率为 0.61%，仅有一个样品被判错，效果明显优于其它特征提取方法。这表明采用 Fisher 线性判别函数对原始数据进行变换，使得组间离差与组内离差的比值变大，也就使得不同类别的样品之间的马氏距离变大，样品也就更易于区分。

2.3　BP 神经网络

采用 MATLAB 软件中的神经网络工具箱构建 BP神经网络，网络的基本属性见表3。首先对以各种方法所选入的变量进行归一化处理，而后作为网络输入，以被测样品中芝麻油的百分含量作为输出。由每类样品中随机选取 10 个，共 110 个样品作为训练样本，以每类样品中剩下的 5 个样品，共 55 个样品作为测试样本。网络经训练样本训练之后，对测试样本进行仿真，得仿真结果，即被测样品的预测值(原始数据此处从略)。由样品的预测值与实际值之差，计算预测的绝对误差，并采用 SAS 软件计算绝对误差个体值的 95%置信区间及均方误差，结果见表 4。

表 3　神经网络的基本属性

最大迭代次数	5000
网络性能目标	0.001
最长运算时间	Inf
性能函数的最小梯度	1.00E-06
输入层传递函数	TANSIG
输出层传递函数	PURELIN
训练函数	TRAINLM
性能函数	MSE
自适应学习函数	LEARNGDM
隐含层神经元数	15
输出层神经元数	1

表 4　绝对误差个体值的容许区间　单位：(%)

变量	区间上限	区间下限	置信度	均方误差
PCA	－6.48	4.30	95	8.48
SLDA	－4.71	3.38	95	4.75
FLT	－3.95	7.56	95	11.20
15 s	－9.72	21.05	95	88.89
30 s	－5.55	16.24	95	57.00
60 s	－6.98	17.95	95	67.33
原始数据	－6.89	8.48	95	25.54

由表 3 可见，采用 15 s、30 s、60 s 或全部原始数据进行网络预测时，由于蕴含了较多的冗余和干扰信息，预测效果较差。采用主成分分析，逐步判别分析和 Fisher 线性判别函数变换所得到的三组变量组合，作为网络输入时，预测效果较为理想。采用逐步判别函数法筛选出的十个变量进行网络预测，绝对分析和均方误差均最小，预测精度最高。

选用逐步判别分析所筛选的一组变量的预测结果，以被测样品中芝麻油的实际含量作为 X 轴，以网络对 55 个预测样本的预测输出作为 Y 轴，利用 Origin 软件作图(见图 2)并进行回归分析。结果表明，实际值与预测值之间存在极显著的相关性($P<0.0001$)，得回归方程为 $y=-0.00823+0.98681x$，相关系数 $R=0.998\ 08$。可见，预测值与实际值较

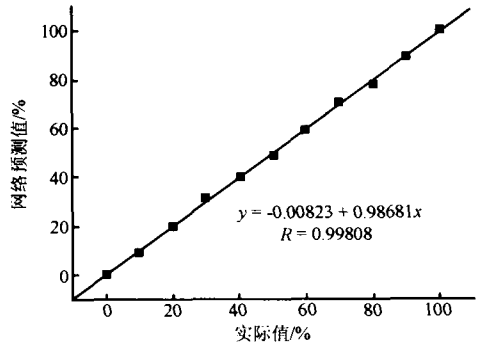

图 2 网络预测值与实际值之间的关系

为接近，说明以神经网络作为模式识别方法，用电子鼻对芝麻油的掺假进行检测，结果比较可靠。

3 总 结

(1) 采用主成分分析，逐步判别分析，Fisher 线性函数变换将原始数据降至 10 维，与采用 15 s、30 s、60 s 以及原始数据的判别分析的结果进行比较，经 Fisher 线性判别函数变换所得的十个变量具有较好的辨别能力，在线性判别函数分析中误判率为 0.61%，仅有一个样品出现误判。

(2) 采用 BP 神经网络的预测结果表明，逐步判别分析所筛选出的变量组合预测精度明显高于其他的变量组合，预测值的绝对误差的 95%置信区间为 (−4.71%, 3.38%)，均方误差为 4.75。

(3) 线性判别函数分析和 BP 神经网络分析均取得了较好的效果，说明通过适当的特征提取，基于这两种模式识别方法，采用电子鼻对芝麻油中玉米油的掺假进行检测是十分可行的。

参考文献：

[1] Corrado Di Natale, Antonella Macagnano, Eugenio Martinelli. The Evaluation of Quality of Post-Harvest Oranges and Apples By Means of an Electronic Nose[J]. Sensors and Actuators B, 2001, 78: 26-31.

[2] 皱小波，赵杰文。电子鼻快速检测谷物霉变的研究[J]. 农业工程学报，2004，20 (4)：121-124.

[3] 张斌，宋维，符若文，章明秋，董先明，赵斌。炭黑/聚苯乙烯复合材料在环境监测方面的应用[J]. 环境技术，2004，22(5)：28-30

[4] 王平. 仿生传感技术的研究进展[J]. 中国医疗器械杂志，2004，28(4)：235-238.

[5] Rita Stella, Joseph N. Barisci, Giorgio Serra, Gordon G. Wallace, Danilo De Rossi. Characterization of Olive oil by an Electronic Nose Based on Conducting Polymer Sensors[J]. Sensors and Actuators B, 2000, 63: 1-9.

[6] Gan H L, Che Man Y B, Tan C P, NorAini I, Nazimah S A H. Characterisation of Vegetable Oils by Surface Acoustic Wave Sensing Electronic Nose[J]. Food Chemistry, 2005, 89: 507-518.

[7] 袁志发，周静芋. 多元统计分析[M]. 第一版，科学出版社，2002，188-201.

[8] Ampuero S, Bosset J O. The Electronic Nose Applied to Dairy Products: a Review[J]. Sensors and Actuators B, 2003, 94: 1-12.

[9] Marini F, Balestrieri F, Bucci R, Magri A D, Magri A L, Marini D. Supervised Pattern Recognition to Authenticate Italian Extra Virgin Olive Oil Varieties[J]. Chemometrics and Intelligent Laboratory Systems, 2004, 73: 85-93.

[10] 许东，吴铮. 基于 MATLAB6.X 的系统分析与设计—神经网络[M]. 第二版，西安电子科技大学出版社，2002 年，19-25.

(上接第 605 页)

[3] Frank J, Meixner H. Sensor System for Indoor Air Monitoring Using Semiconducting Metal Oxides and IR-absorption [J]. Sensor and Actuators B. 2001, 78: 298-302.

[4] Meléndez J, J de Castro A, Lopez F, Meneses J. Spectrally Selective Gas Cell for Electrooptical Infrared Compact Multigas Sensor[J]. Sensor and Actuators A. 1995, 47: 417-421.

[5] Dirk Rossberg, Optical Properties of the Integrated Infrared Sensor[J]. Sensor and Actuators A. 1996, 54: 793-797.

[6] Rothman L S, et al. The HITRAN Database[J]. Appl Opt, 1987, 26: 1051-1097.

[7] Silveira J P, Anguita J, Briones F, et al. Micromachined Methane Sensor Based on Low Resolution Spectral Modulation of IR Absorption Radiation[J]. Sensor and Actuators B. 1998, 48: 305-307.

[8] Bauer D, Heeger M, Gebhard M, Benecke W. Design and Fabrication of a Thermal Infrared Emitter[J]. Sensor and Actuators A, 1996, 55: 57-63.

[9] Thierry Corman, Edvard Kalvesten, Matti Huiku, et al. New CO_2 Filters Fabricated by Anodic Bonding at Overpressure in CO_2 Atmosphere[J]. Sensor and Actuators A. 1998, 69: 166-171.