

Study of Zanthoxylum Bungeanum Varieties Discriminating Method by Electronic Nose Technology*

WU Lili, ZHENG Dan, ZHENG Baozhou, LIN Aiyong, PAN Jianbin*

(College of Sciences, Henan Agricultural University, Zhengzhou He'nan 450002, China)

Abstract: A fast discriminating method of Zanthoxylum Bungeanum varieties was studied. In the experiment, the order information of six Zanthoxylum Bungeanum varieties were collected by the electronic nose system for investigation. The fifty-six groups training samples and thirty-two groups testing samples were obtained after feature extraction. The feature data were classified by the BP neural network, probabilistic neural network, and support vector machine respectively, and the rate of discrimination was 89.58%, 93.23%, 94.27%. Compared to the BP neural network and probabilistic neural networks, support vector machine had better classification results. The electronic nose system developed in this paper had achieved the purpose of identifying Zanthoxylum Bungeanum species non-destructively, rapidly and accurately, which provided a new way for agricultural products nondestructive testing research.

Key words: electronic nose; BP network; probabilistic neural network; support vector machine; Zanthoxylum Bungeanum
EEACC: 6140; 7230S **doi:** 10.3969/j.issn.1004-1699.2013.11.001

基于电子鼻技术的花椒品种鉴别方法研究*

吴莉莉, 郑丹, 郑宝周, 林爱英, 潘建斌*

(河南农业大学理学院, 郑州 450002)

摘要: 针对不同花椒品种的快速鉴别方法进行研究, 以花椒的气味信息检测为研究对象, 利用自行研制的电子鼻系统采集了6类花椒样品气味数据, 对这些数据样本进行特征提取, 得到了56组训练样本和32组测试样本。利用BP神经网络、概率神经网络和支持向量机对特征数据进行鉴别, 正确识别率分别为89.58%、93.23%、94.27%, 相对于BP神经网络和概率神经网络识别, 支持向量机具有更好的分类效果。本文研制的电子鼻系统能无损、快速、准确鉴别花椒的品种, 为农产品无损检测的研究提供了一种新的思路。

关键词: 电子鼻; BP网络; 概率神经网络; 支持向量机; 花椒

中图分类号: TP274; TN911.7

文献标识码: A

文章编号: 1004-1699(2013)11-1473-05

花椒是指芸香科植物花椒和青椒的干燥成熟果皮。我国是世界上花椒种植面积和产量的第一大国, 约有45种, 13变种, 分布于全国20多个省份, 尤以长江南及西南诸省居多。原产我国的花椒属植物主要有花椒、野花椒、川陕花椒、竹叶花椒、青花椒等。花椒作为我国的特色辛香料和中药材, 有着悠久的历史渊源。花椒中所含的丰富化学物质, 决定了其具有极高的药用价值和食用价值^[1-3]。花椒中的化学成分主要有挥发油、生物碱、酰胺、木脂素、香豆素、黄酮类物质和脂肪酸等。花椒的香气成分来自于其组织中所含的挥发油, 挥发油是由烯烃类等有机化合物及其含氧衍生物醇、醛、酮、酯等成分组成。由于品种、产

地的气候、雨量、日照和土质等诸多自然因素的不同, 造成其化学组成及含量差异很大。

花椒市场品种繁多, 由于各产地的种类和质量有所差别, 尤其是花椒所具有的特征性香味差异明显, 导致不同地方和不同品种的花椒价格差异很大, 常常出现品种混杂、真伪混淆等现象。目前对花椒的品种鉴别主要依靠感官和经验, 受主观影响较大, 而且对于具有强烈气味的花椒而言, 很难进行多次、连续的鉴别。而采用气相色谱法、色谱-质谱联用法、原子吸收光谱仪等仪器分析方法检测费用高、耗时长、样品前处理复杂, 无法满足现场快速检测的需求。

电子鼻是近年来发展非常迅速的现代分析技

项目来源: 河南省教育厅科学技术研究重点项目(12A510014); 郑州市科技攻关项目(121PPTGG465-3)

收稿日期: 2013-09-07 **修改日期:** 2013-10-10

术,它综合运用了仿生技术、电子技术、计算机技术、神经生理学和数学等多个学科的研究成果,具有成本低、稳定性好、分析速度快等优点,已逐渐成为农产品无损检测领域分析判定的重要手段。电子鼻模拟生物嗅觉系统,由传感器阵列结合模式识别系统构成^[4],虽然没有对气味的组分给出特定信息,也没有关于组分的识别,但是在合适的数学工具的帮助下,例如人工神经网络和统计方法,电子鼻能够从一个特定样品中识别气味模式并在区分其他样品中起到帮助作用^[5]。

目前利用电子鼻系统对花椒进行检测的研究已经展开了。阎红等^[6]利用法国 Alpha M. O. S. 公司生产的 FOX 4000 型电子鼻对四川汉源地区所产青花椒和红花椒样品进行了检测,运用主成分分析和判别因子分析方法进行分析,发现二者之间的气味存在明显的差异,与感官分析得出的结论一致。彭德川等^[7]利用电子鼻技术分析考察了四川部分地区花椒粉样品气味上的差异,并与人工品评结果相对比,结果表明电子鼻具有客观、快捷、重复性好的特点。

本文针对花椒品种的快速鉴别方法进行研究,以花椒的气味信息检测为研究对象,利用自行研制的电子鼻系统对不同品种的花椒样品进行分类识别,以期消除人为主观因素的影响,获得较高的识别精度。

1 实验方法

1.1 实验材料

实验用的花椒样品购于郑州某超市,品种分别为:四川汉源大红袍(标记为1号样品)、四川金阳青花椒(2号样品)、陕西韩城大红袍(3号样品)、河北涉县大红袍(4号样品)、云南鲁甸青花椒(5号样品)、甘肃陇南大红袍(6号样品)。每种样品取50克分别置于6个表面皿内。

1.2 实验装置

实验中花椒的气味信息采集装置采用自行研制的电子鼻系统,主要由硬件和软件两部分构成。硬件系统由传感器阵列、密闭气室、样品台、外围电路、数据采集卡以及计算机等组成,如图1所示^[8]。其中气体传感器阵列采用8只郑州炜盛电子科技有限公司

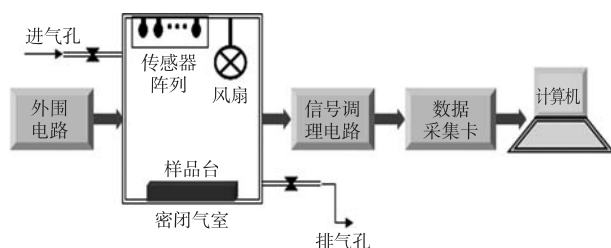


图1 电子鼻系统的硬件组成框图

公司生产的半导体气体传感器构成,各个传感器的性能描述如表1所示,它们对硫化物、烷烃类、乙醇类等都有不同的响应。外围电路的功能主要是对传感器阵列进行供电、加热等。传感器的加热电压为5V,工作电压为5V,负载电阻为4.7kΩ。传感器阵列的脱附采用洁净空气。

表1 气体传感器阵列

传感器	传感器型号	敏感性能描述
1	MQ-2	丙烷、氢气、液化气
2	MQ-3	乙醇
3	MQ-5	丁烷、丙烷、甲烷
4	MQ-7	一氧化碳
5	MQ-135	氨气、硫化物、苯系
6	MQ-136	硫化氢
7	MQ-137	氨气
8	MQ-138	甲苯、乙醇、丙酮、甲醛、氢气等

由于花椒的气味比较浓郁,易于挥发,因此在电子鼻进样系统中采取自动静态顶空取样设计。所谓静态顶空取样法就是将待测样本置于容器内,待其自然挥发至容器内气体饱和后,将容器顶空的混合气体作为样本气体进行测量。该方法可保证每次进样的容积相同,为后续分析提供精准的数据,而且操作简单、成本低,适合易挥发物的检测。

软件系统主要包括信号预处理算法、特征提取算法以及模式识别算法等构成。

1.3 传感器阵列对花椒样品的响应

为了方便后续数据处理和在一定程度上排除外部条件及传感器本身所引起的误差,对传感器阵列采集的原始数据进行去基准、平滑滤波和归一化处理。图2所示为8只传感器对1号花椒样品的响应曲线,该曲线是传感器信号经A/D转换采样和归一化处理后得到的一个个的离散点构成的。从图中可看出各传感器响应电压开始较低,随着花椒产生的气体富集在传感器表面,传感器输出电压不断增大,达到饱和后,输出电压为极大值;当撤去样品时,传感器的响应电压迅速下降,对于不同的传感器,其响应速度有所不同。

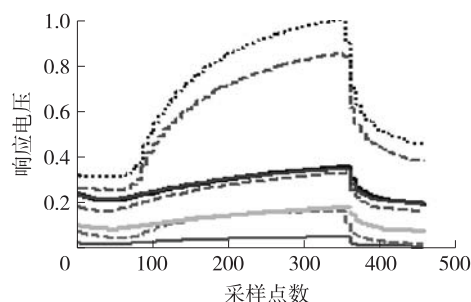


图2 传感器阵列对1号花椒样品的响应曲线

2 特征值的选择

对每种花椒样品, 8 只传感器阵列获得的响应曲线对应的数据量比较大, 实际做分析判别时, 不会利用整个曲线上的所有点, 而是提取其中最能代表本传感器对被测气味响应特性的一些点, 这些点的值就称为特征值。

以单只传感器 MQ-3 对 1 号样品的响应曲线为

$$\mathbf{R} = \begin{bmatrix} \text{Min}_1 & \text{Gramax}_1 & \text{Max}_1 & \text{Mean}_1 & \text{Gramin}_1 & 50s_1 & 100s_1 & 150s_1 & 200s_1 & 250s_1 & 300s_1 \\ \text{Min}_2 & \text{Gramax}_2 & \text{Max}_2 & \text{Mean}_2 & \text{Gramin}_2 & 50s_2 & 100s_2 & 150s_2 & 200s_2 & 250s_2 & 300s_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{Min}_8 & \text{Gramax}_8 & \text{Max}_8 & \text{Mean}_8 & \text{Gramin}_8 & 50s_8 & 100s_8 & 150s_8 & 200s_8 & 250s_8 & 300s_8 \end{bmatrix} \quad (1)$$

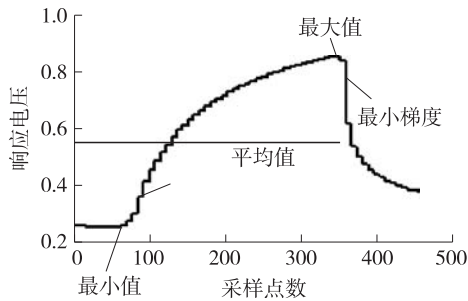


图 3 MQ-3 对 1 号花椒样品的响应曲线

3 模式识别方法

3.1 BP 网络

BP 网络是电子鼻领域应用较多的算法之一, 该算法功能强大, 训练简单。BP 网络是采用误差反向传播 BP (Back Propagation) 算法的多层前馈网络, 其学习规则采用梯度下降算法。典型的 BP 网络由输入层、隐含层和输出层组成, 其数学模型为:

$$\begin{cases} y_i^l = f(x_i^l) \\ x_i^l = \sum_{j=1}^{N_{l-1}} w_{ij}^l y_j^{l-1} + \theta_i^l \end{cases} \quad (2)$$

式中: y_i^l 为第 l 层第 i 个节点的输出值; x_i^l 为第 l 层第 i 个节点的激活值; w_{ij}^l 为第 $l-1$ 层第 j 个节点到第 l 层第 i 个节点的连接权值; θ_i^l 为第 l 层第 i 个节点的阈值; N_l 为第 l 层节点数; L 为总层数; f 为神经元激活函数。

BP 网络的学习过程分为 2 个过程: ①工作信号正向传播。输入信号从输入层经隐单元, 传向输出层, 在输出端产生输出信号, 这是工作信号的正向传播。在信号的向前传递过程中网络的权值是固定不变的, 每一层神经元的状态只影响下一层神经元的状态。如果在输出层不能得到期望的输出, 则转入误差信号反向传播。②误差信号反向传播。网络的实际输出与期望输出之间的均方误差即为误差信

例说明实验中对特征值的提取。图 3 中提取曲线的最小值、最大梯度值、最大值、平均值、最小梯度值、50 s、100 s、150 s、200 s、250 s、以及 300 s 采样点处的响应值作为特征值, 分别记为 Min、Gramax、Max、Mean、Gramin、50 s、100 s、150 s、200 s、250 s、300 s。8 只传感器均按此法选择特征值, 构成原始特征向量 \mathbf{R} 。即原始特征向量 \mathbf{R} 由 11 个特征子集组成, 由于有 8 只传感器组成阵列, 其维数为 $8 \times 11 = 88$ 维, 可表示为:

号, 逐层向输入层反向传播, 分配给各连接节点, 并计算出各连接节点的参考误差, 在此基础上调整各连接权值, 使得网络的期望输出与实际输出的均方误差达到最小。

3.2 概论神经网络

概率神经网络 PNN (Probabilistic Neural Network) 是由 Specht D F 在径向基神经网络的基础上提来的一种 4 层前馈神经网络, 其主要思想是利用贝叶斯决策准则, 在多维输入空间内分离决策空间。由于其采用了高斯核的 Parzen 窗来估计样本的后验概率实现贝叶斯分类, 因此具有结构简单、训练速度快等优点, 常用于解决模式分类问题^[9-12]。

PNN 为双隐层结构, 具有一个径向基网络层和一个竞争型网络层。径向基层计算该输入目标向量同输入样本向量之间的距离 $\| \text{dist} \|$, 用高斯核函数实现非线性映射:

$$p_j^k(\mathbf{x}) = \exp(-\| \mathbf{x} - \boldsymbol{\omega} \|^2 / 2\sigma^2) \quad (3)$$

其中 k 为分类类别数, $k = 1, 2, \dots, K$; j 为训练样本个数, $j = 1, 2, \dots, R$; $\boldsymbol{\omega}$ 为权向量。选择不同的方差可以得到不同的分类器。例如: $\sigma = \infty$ 时接近线性分类器; $\sigma = 0$ 是趋近近邻分类器。

3.3 支持向量机

支持向量机 SVM (Support Vector Machine) 是一种基于统计学习理论的机器学习算法, 它具有全局优化、适应性强、理论完备、泛化性能好等优点。

SVM 采用结构风险最小化准则, 不同于一些机器学习算法通常采用的经验风险最小化准则, 统计学习理论提出了一种新的策略: 将函数集构造为一个函数子集序列, 使各个子集按照 VC 维的大小排列; 在每个子集中寻找较小经验风险, 在子集间折衷考虑经验风险和置信范围, 取得实际风险的最小, 这种思想称作结构风险最小化。因此 SVM 能较好地解决小样本、非线性、高维数和存在局部极小值等实

际模式识别问题^[13]。

SVM 的实质在于:首先通过非线性变换将输入空间变换到一个高维空间,然后在这个高维空间中进行线性回归,求取最优线性分类面,而这种非线性变换是通过定义适当的内积函数实现的。

常用的内积核函数有以下 3 种^[14]:

(1)多项式函数

$$K(x,y)=[(x \cdot y)+c]^d \quad (c \geq 0) \quad (4)$$

当 $c > 0$, 为非齐次多项式核;当 $c = 0$, 为齐次多项式核。

(2)高斯径向基函数

$$K(x,y)=\exp(-\frac{1}{2\sigma^2} \|x-y\|^2) \quad (5)$$

(3)Sigmoid 函数

$$K(x,y)=\tanh[v(x \cdot y)+c] \quad (v > 0, c < 0) \quad (6)$$

SVM 是一种两分类器,实际应用时常常遇到多分类问题。对于多类 SVM 分类问题,目前主要有 2 种策略(假设类别数为 n):①1 对多。训练 n 个 SVM,每个 SVM 区分一类和其他类。②1 对 1。训练 $n(n-1)/2$ 个 SVM,每个 SVM 区分一对类别。

4 实验结果比较分析

4.1 BP 网络鉴别花椒品种

利用 BP 网络对花椒品种进行鉴别时,建立 8 维(传感器个数)输入和 6 维(待识别的种类)输出,理想输出为(0,0,0,0,0,1)类型,其中 1 表示属于该类,否则为 0。BP 网络采用 S 型激励函数的隐含层,隐层神经元数通过实验确定为 14(此时的正确识别率最高),网络训练函数采用 Levenberg-Marguardt 的优化算法,网络最大训练步长设置为 3 000,目标误差为 0.000 1。

BP 算法先对每种花椒样品进行网络训练,然后再进行鉴别,选用 56 组特征参数作为训练样本,32 组特征参数作为测试样本,识别结果如表 2 所示。

表 2 BP 网络的鉴别结果

花椒样品	训练样本		识别率 /%	测试样本		识别率 /%
	正确	错误		正确	错误	
1	55	1	98.18	28	4	87.50
2	54	2	96.43	29	3	90.63
3	54	2	96.43	28	4	87.50
4	55	1	98.18	29	3	90.63
5	56	0	100.00	30	2	93.75
6	55	1	98.18	28	4	87.50
总识别率/%	97.92			89.58		

对于 8 只传感器的输出响应,采用 BP 网络对训练样本进行模式识别时,分类效果尚可,但对测试样本进行分类时,无论如何修改隐层单元个数及和改进权值学习算法,都得不到较为理想的分类结果,说明 BP 神经网络在小样本的情况下泛化能力较低。

4.2 PNN 鉴别花椒品种

在设计 PNN 时,由于径向基函数的分布密度 Spread 对网络的分类性能影响较大,当 Spread 接近于 0 时,对应的 PNN 就成为一个最近邻分类器;当 Spread 增大后,对应的 PNN 就要考虑附近的设计向量。采用与 BP 网络相同的 56 组训练样本和 32 组测试样本,在训练网络时,通过实验确定 Spread 取值为 0.003 时,训练样本识别率最高。表 3 是 PNN 的识别结果,显然相比于 BP 网络识别率得到了提高。

表 3 PNN 的鉴别结果

花椒样品	训练样本		识别率 /%	测试样本		识别率 /%
	正确	错误		正确	错误	
1	56	0	100.00	30	2	93.75
2	55	1	98.18	29	3	90.63
3	56	0	100.00	30	2	93.75
4	55	1	98.18	29	3	90.63
5	56	0	100.00	31	1	96.88
6	56	0	100.00	30	2	93.75
总识别率/%	99.40			93.23		

4.3 SVM 鉴别花椒品种

实验中用 SVM 对花椒品种进行鉴别,是多分类问题,采用“1 对 1”策略。此法是基于两分类方法基础之上的,先在 6 类训练样本中构造所有可能的两类分类器,共应构建 15 个分类器,然后用这 15 个分类器对检验样本进行分类,对各分类结果进行投票,得票最多的类为检验样本所属的类别。

内积核函数的选择在 SVM 分类中非常重要,目前为止还没有一个公认比较好的选择标准。通过实验比较,选用高斯径向基函数的 SVM 取得了较好的识别结果,如表 4 所示。

表 4 SVM 的鉴别结果

花椒样品	训练样本		识别率 /%	测试样本		识别率 /%
	正确	错误		正确	错误	
1	56	0	100.00	31	1	96.88
2	56	0	100.00	29	3	90.63
3	56	0	100.00	30	2	93.75
4	56	0	100.00	29	3	90.63
5	56	0	100.00	32	0	100
6	56	0	100.00	30	2	93.75
总识别率/%	100.00			94.27		

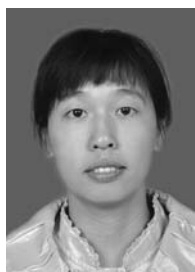
与传统的高斯径向基神经网络相比,支持向量充当了聚类中心的角色,当设定的方差较大时,支持向量数量减少;而当方差较小时,支持向量的数目反而增加。所不同的是,SVM 中的这些聚类中心及其权重都是算法自动寻找到的。

5 结束语

本文利用自行研制的电子鼻系统对 6 类不同品种的花椒样品在相同条件下进行了数据采集和预处理,分别采用了 BP 神经网络、PNN 和 SVM 3 种模式识别方法进行了对比分析。BP 网络由于在小样本的情况下泛化能力较低,因此对于 6 类不同品种的花椒正确识别率不高,仅为 89.58%。与 BP 算法相比,PNN 学习速度快,收敛性好,并且网络结构设计灵活方便,对花椒样品的正确识别相对较高,为 93.23%。基于小样本统计学习理论的 SVM 与神经网络方法相比结构简单、容易训练、收敛速度快,且有很高的分类精度,对花椒样品的正确识别率可达 94.27%,是一种有效地分类识别方法。如何优化 SVM,融合传感器阵列的其他信息诸如机器视觉、机器味觉等,来进一步提高识别率是将来要进行的研究。

参考文献:

- [1] 赵志峰. 汉源花椒风味物质研究及花椒油生产工艺优化[D]. 成都: 四川大学, 2005: 4.
- [2] 阎红, 辛松林, 彭德川. 花椒的气味指纹分析技术应用与探索[J]. 四川烹饪高等专科学校学报, 2010(2): 21-24.
- [3] 余晓琴. 花椒品质评价方法及其应用研究[D]. 重庆: 西南大学, 2010: 5.
- [4] 吴莉莉, 林爱英, 郑宝周, 等. 电子鼻检测技术在粮食霉变识别中的应用研究[J]. 安徽农业科学, 2009, 37(21): 10133-10135.
- [5] Balasubramanian S, Panigrahi S, Kottapalli B, et al. Evaluation of an Artificial of Lactroxy System for Grain Quality Discrimination[J]. LWT, 2007, 40: 1815-1825.
- [6] 阎红, 辛松林, 彭德川. 四川汉源地区青花椒和红花椒的气味指纹图谱分析[J]. 中国调味品, 2010, 35(4): 89-91.
- [7] 彭德川, 阎红, 辛松林. 电子鼻技术对红花椒粉感观品质的快速分析[J]. 现代食品科技, 2010, 26(9): 1018-1019.
- [8] 惠国华, 吴玉玲, 陈裕泉. 基于电子鼻的香蕉储存时间鉴别方法研究[J]. 传感技术学报, 2012, 25(5): 566-570.
- [9] Specht D F. Probabilistic Neural Networks for Classification Mapping or Associative Memory [C]//Neural Networks, IEEE International Conference on Digital Object Identifier, 1988, 1: 525-532.
- [10] Specht D F. Probabilistic Neural Networks [J]. Int J Neural Networks, 1990, 3: 109-118.
- [11] Roberto Miguez, Michael Georgiopoulos, Assem Kaylani. G-PNN: A Genetically Engineered Probabilistic Neural Network [J]. Nonlinear Analysis: Theory, Methods and Applications, 2010, 73(6): 1783-1791.
- [12] Michael R, Berthold Jay Diamond. Constructive Training of Probabilistic Neural Networks [J]. Neurocomputing, 1998, 19(1-3): 167-183.
- [13] 高剑, 罗志增. 支持向量机在肌电信号模式识别中的应用[J]. 传感技术学报, 2007, 20(2): 366-369.
- [14] 李弼程, 邵美珍, 黄洁. 模式识别原理与应用[M]. 西安: 西安电子科技大学出版社, 2008: 2.
- [15] Esposito Sonia, Montedoro Gian Francesco, Selvaggini Roberto, et al. Monitoring of Virgin Olive Oil Volatile Compounds Evolution during Olive Malaxation by an Array of Metal Oxide Sensors [J]. Food Chemistry, 2009, 113: 345-350.
- [16] Wu Lili, Yuan Chao, Lin Aiyong, et al. Identification of Early Moldy Rice Samples by PCA and PNN [J]. Communications in Computer and Information Science, 2012, 288(1): 506-514.



吴莉莉(1977-), 湖北荆州人, 2007 年获浙江大学生物医学工程专业博士学位, 师从陈裕泉教授。副教授, 主要从事传感器技术、信号处理、模式识别和农产品无损检测等方面的研究, 已发表论文四十余篇, willzju@126.com;



郑丹(1981-), 河南信阳人, 讲师, 主要从事大学物理教学研究、农业工程和物理农业等方面的研究, 已发表论文十余篇, masterzd@126.com;



潘建斌(1962-), 河南洛阳人, 教授, 主要从事智能仪器、农业工程及农产品加工等方面的研究, 已发表论文三十余篇, panjb01@126.com。