

Discrimination of Chinese Herbal Medicines from Different Growing Areas and Harvesting Dates Based on E-Nose*

WU Shiyuan, LUO Dehan*, DENG Bingrong, ZHAO Qinghui

(Faculty of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: It is difficult to distinguish the same kind of Chinese herbal medicines (CHMs) with different growing areas and harvesting dates. An electronic nose (E-nose) PEN3 was used to discriminate the CHMs with different growing areas and harvesting dates under natural state. Firstly, odor information of samples was obtained by E-nose, and feature set was constituted by all aspects of the representative characteristics of odor information. Then, negative feedback was applied to extract features, and the most effective feature subset was determined. PCA+LDA was used to classify samples, finally, unknown samples were identified by using Euclidean distance, Mahalanobis distance. Identification of unknown samples was up to 100% accuracy, showing that it is feasible to evaluate the quality of the CHMs used e-nose. The paper provides a new method for quality control of CHMs.

Key words: electronic nose; Chinese herbal medicines; feature extraction and selection; PCA + LDA algorithm; quality control

EEACC:7230J

doi:10.3969/j.issn.1004-1699.2011.01.003

不同产地和采收期的中药材电子鼻鉴别研究*

伍世元, 骆德汉*, 邓炳荣, 赵庆辉

(广东工业大学信息工程学院, 广州 510006)

摘要: 不同产地和采收期的同种中药材在感官上很难区分。提出一种在自然态下采用 PEN3 电子鼻鉴别不同产地和采收期中药材的方法。首先通过电子鼻获取样本的气味信息, 综合气味信息各方面的特征组成一个特征集合, 采用逆向反馈方法提取特征, 获得最能区别样品的特征子集, 利用 PCA+LDA 算法对样本进行分类, 最后运用欧式距离、马氏距离对未知样本进行鉴别。对未知样本的识别正确率达 100%, 证明采用电子鼻对中药材质量进行评判是可行的。为中药材质量控制提供了一种新的手段。

关键词: 电子鼻; 中药材; 特征提取与选择; PCA+LDA 算法; 质量控制

中图分类号: TP212; R282.5

文献标识码: A

文章编号: 1004-1699(2011)01-0010-04

中药材存在产地、批次、品种、采收季节等因素的影响^[1], 致使中药材很难达到质量标准化, 中药材质量可控性成为中医药界关注的热门话题。基于感官的鉴别方法不可避免地要受生理、经验、情绪、环境等因素的影响, 其主观性强、重复性差, 难以形成标准。高效液相色谱、气相色谱等理化方法需要较长的前处理时间, 很难将获得的数据和样本的气味直接联系起来^[1-2], 而且是典型的有损检测。

电子鼻是由具有交叉敏感特性的传感器阵列和适当的模式识别算法组成的仪器, 能识别简单和复

杂的气味或气体^[3]。它不但具有客观性强、重现性好、操作简单等特点, 更重要的是对样品的测定可以做到不失原本性、无损性, 能像人类鼻子一样获得样品气味的整体信息^[5], 即“气味指纹图谱”。目前电子鼻在食品品质评价、环境检测、疾病诊断等方面获得广泛应用^[4-7]。

文献[2]采用电子鼻识别不同种类的中药材, 而不同种类的中药材外在特性有很大的差异, 用电子鼻来鉴别没有充分发挥其在中药材鉴别中的优势。本文提出在自然态下利用电子鼻对不同产地和

项目来源: 2009 年国家自然科学基金资助“基于仿生嗅觉的辛味中药材气味指纹图谱研究”(60971105)

收稿日期: 2010-05-20 修改日期: 2010-09-07

采收期的同种中药材进行分类鉴别,为中药材质量控制提供一种新的有效方法。

1 实验装置、样品及方法

1.1 实验装置

采用德国 AIRSENSE 公司制造的 PEN3 便携式电子鼻, PEN3 电子鼻是一种用来检测气体和蒸汽的小巧、快捷、高效的检测系统,经过训练后可以很快辨别单一化合物或者混合气体,其测量气室内置 10 个金属氧化物气体传感器组成传感器阵列,分别以“S1”到“S10”命名。

1.2 检测参数的设置

确定检测参数如下,工作环境温度:25℃;顶空生成时间:60 min(静态顶空);进气流量:200 mL/min;采样时间:80 s;相对湿度:58%~62%;清洗时间:100 s;采样间隔时间:1.0 s;零点漂移时间:10.0 s;预抽样时间:6.0 s;顶空空间:250 mL。

1.3 检测样品和测量方法

实验样品由广东药学院提供,测量数据一半用做训练样本,一半用做测试样本。

(1) 不同采收期中药材对象和检测方法

选择 2 个批次不同采收期的阳春砂为研究对象(批次 1 为成熟的阳春砂,批次 2 为成熟前 1 个月采收的)。由于要求在自然态下对样品进行测量并分类,对中药材无需经过研磨粉碎,而阳春砂颗粒有大有小,致使两批次阳春砂样本占住的体积、外观状态、散发气味浓度不一致,在其他检测参数基本确定的情况下,质量因素将是影响测量结果的主要因素。所以对 2 个不同采收期的阳春砂各取 5 个质量水平(5 g、10 g、15 g、20 g、25 g),分别以“阳春砂 01”到“阳春砂 10”命名,其中“阳春砂 01”到“阳春砂 05”代表批次 1 各样本,“阳春砂 06”到“阳春砂 10”代表批次 2 各样本,每个水平重复测量 12 次,共得到 120 个样本。

(2) 不同产地中药材对象和检测方法

选择安徽太和、广东揭阳、江苏太仓产的薄荷为研究对象,根据不同采收期的阳春砂分析结果,样本质量水平统一取 15 g,每个水平重复测量 16 次,共得到 48 个样本。

2 特征提取

电子鼻面临的一个问题是对差别微小,浓度甚微的气体进行检测,类别之间的测量空间间隔很近。仅选择一个特征只能代表传感器响应数据的某一方面的信息,所获信息有限^[6];采用多个具有代表性万方数据

的特征增加了信息量,但是会增加特征阵列的维数,特征之间也会出现相互抵消的现象^[7],使得分类判别处理数据量大而且分类鉴别结果不一定理想。鉴于上述讨论,选择代表传感器响应曲线不同方面的各个特征组成原始特征集合,通过鉴别结果的反馈对特征组合进行优化,得到一个使判别分类最优的特征子集,并用于对待测样本进行鉴别。

特征集合包括以下特征:各传感器第 15 秒数据、各传感器第 30 秒数据、各传感器第 40 秒数据、各传感器第 50 秒数据、各传感器第 60 秒数据、各传感器第 70 秒数据、各传感器第 80 秒数据、各传感器的均值、各传感器的峰值、各传感器的方差、各传感器的标准差、各曲线的微分值、各曲线的积分值、各传感器四阶拟合系数(5 个拟合参数)、各传感器方差最大时候的响应值。

3 分类判别

主成分分析(PCA)虽能最大限度地保持原有测量数据集的信息,但经过降维后保存下来的信息却不一定对分类最有用,这是因为被 PCA 算法抛弃的那些分布方向有可能正是能够把不同类别区分开来的方向。线性判别分析(LDA)由原始数据经线性组合构造判别函数,将多维空间分成一些子空间,能够最大限度地区分不同的样品集,分类效果好且易实现,但是,当样本总数较少或选取的特征数较多时,直接采用 LDA 算法会出现小样本问题,即导致样本类内散布矩阵奇异, LDA 算法将无法进行下去,而解决小样本问题可通过降维的方法使类内散布矩阵非奇异或者避免对其求逆^[8-9]。所以,本文综合 PCA 和 LDA 的优缺点,采用由 PCA 算法得到的特征矩阵作为 LDA 算法的输入矩阵,即 PCA+LDA 分析法^[10]。

在特征集合里选择特征子集,通过 PCA 对特征子集进行降维,减少数据计算量,进而优化特征向量,得到能够代表原始特征主要方面的少量特征作为 LDA 的输入,即得到足够的信息,又减少数据的处理量,使类内散布最小化、类间散布最大化。

3.1 不同采收期中药材分类结果

在选择某一个特征或特征集合全部特征的情况下,分类结果都不理想。如图 1 所示,在特征集合里选取单一特征的情况下,得到区分度最好的特征是“各传感器的方差”,但区分效果还是不太理想,在相同质量水平的“阳春砂 04”和“阳春砂 09”之间决策边界靠的很近;不同质量水平的同一批次样本聚合度不理想,质量因素对分类影响很大。图 2 是选

择特征集合的全部特征,对其进行 PCA+LDA 分析,结果各个样本数据混杂在一起,分类效果反而更差。可见,单一的特征不能代表样本的整体信息,对样本的反映不全面,而太多的特征往往使得特征之间相互抵消,分类反而不理想。

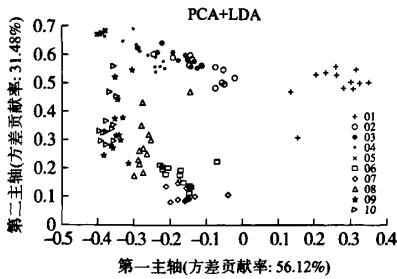


图 1 单一特征下的不同采收期阳春砂分类结果

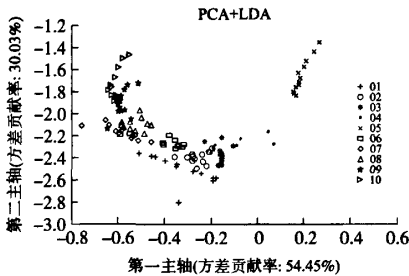


图 2 特征集合下的不同采收期阳春砂分类结果

通过选择特征集合里的特征反复试验,当选择“各传感器的方差”、“各传感器的标准差”、“各曲线的微分值”这一特征子集时,能够将两个不同采收期的阳春砂很好的区分开来,如图 3 所示。通过一条直线可以清楚地把 2 个批次不同采收期的阳春砂区别开来,决策边界清晰。尽管在不同质量的情况下,两个批次的阳春砂数据不重合,说明质量会影响分类的结果,但是质量引起的差别比单一特征和特征集合下的要紧凑得多。“阳春砂 06”到“阳春砂 10”靠得比较紧凑,说明质量对这一批次的影响较小,是采收时间较早的批次,散发的气味比较淡;“阳春砂 02”、“阳春砂 03”、“阳春砂 04”重叠在一起,质量因素对这一区段的样本影响较小;总体上讲,“阳春砂 01”到“阳春砂 05”质量因素对它的影响很大,尤其是在“阳春砂 05”有强烈离群现象,说明在 20 g、25 g 质量之间,质量因素对测量影响很大,所以应该避免选择位于此区间的质量的样本作为检测对象。通过选择一个特征子集,收到了很好的分类结果,说明“各传感器的方差”、“各传感器的标准差”、“各曲线的微分值”这一特征子集能够充分代表不同采收期阳春砂的特点。

万方数据

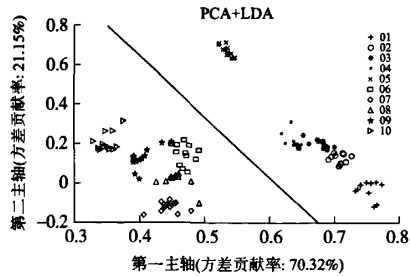
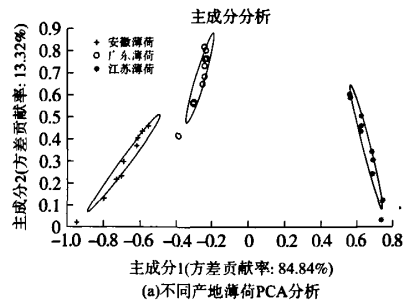


图 3 特征子集下的不同采收期阳春砂分类结果

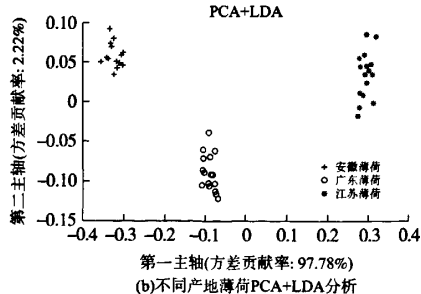
3.2 不同产地中药材分类结果

通过采用与不同采收期阳春砂分类类似的方法反复验证,在特征集合里面选择适当的特征子集,对不同产地的薄荷进行分类。

通过试验,最终选择“各传感器的均值”、“各传感器的峰值”、“各曲线的标准差”、“各曲线的微分值”、“四阶拟合系数”这一特征子集,获得较好的分类结果,说明这一特征子集能够充分代表不同产地薄荷的特点,如图 4(a)、4(b)所示。但是,在相同的特征子集情况下,PCA 分析方法中广东薄荷和安徽薄荷的分类界面靠得很近,PCA+LDA 分析方法内更加紧凑、类间分界面更加明显,证实 PCA+LDA 分析方法的优越性。



(a)不同产地薄荷PCA分析



(b)不同产地薄荷PCA+LDA分析

图 4 不同产地中药材分类结果

3.3 未知样本识别结果

选择和训练样本相一致的特征子集,把待测样本数据通过 PCA+LDA 映射到特征空间,分别采用欧式距离、马氏距离分析法,计算未知样本和各个训

练类别之间的距离,把未知样本归类到距离最小的训练类别,得出未知样本的类别属性。由于采用反馈方法获得最能区别样本的特征子集,通过分别采用欧式距离、马氏距离分析法,都能得到正确的识别结果,正确率为100%。

4 结束语

事先确定特征,通过模式识别技术做分类判别的方法有很大的劣势,因为无法确定事先选择的特征就能确切的代表样本类别之间的最大差别。况且选择某一个特征得到的信息有限,而选择过多的特征使得特征相互抵消,并不一定能获得较好的分类判别结果。应该采取逆向思维的方法,通过一个反馈的过程,根据测量结果不断优化特征子集,最终达到较好的分类判别结果。同时,采用PCA+LDA分析方法,能够解决PCA算法对不同的训练样本数据不敏感的问题和LDA算法可能出现的小样本问题,又能够实现PCA算法的特征压缩,使得处理变得简单有效。

通过此种方法,对不同产地的薄荷以及不同采收期的阳春砂进行分类鉴别,获得较理想的结果。通过电子鼻构建中药材气味指纹图谱的方法具有可

观的前景,将为中药材质量控制以及中药材的质量标准化提供一种行之有效的方法。

参考文献:

- [1] 孟岩,郑旭光,郑燕,等. 色谱指纹图谱在中药与天然产物研究中的应用[J]. 河北医药,2009,31(18):2473-2474.
- [2] 刘红秀,姬生国,庄家俊,等. 基于仿生嗅觉的中药材鉴别的实现[J]. 广东药学院学报,2009,25(4):356-359.
- [3] Gardner J W, Bartlett N. A Brief History of Electronic Nose[J]. Sensors and Actuators B,1994,18-19:211-220.
- [4] Huichun Yu, Jun Wang. Discrimination of LongJing Green-Tea Grade by Electronic Nose [J]. Sensors and Actuators B: Chemical,2007,122(1):134-140.
- [5] Ghasemi-Varnamkhashti M, Mohtasebi S S, Siadat M, et al. Meat Quality Assessment by Electronic Nose (Machine Olfaction Technology)[J]. Sensors,2009,9(8):6058-6083.
- [6] 周海涛,殷勇,于慧春. 劲酒电子鼻鉴别分析中传感器阵列优化方法研究[J]. 传感技术学报,2009,22(2):175-177.
- [7] 海铮,王俊. 电子鼻信号特征提取与传感器优化的研究[J]. 传感技术学报,2006,19(3):606-609.
- [8] 宋枫溪,程科,杨静宇. 最大散度差和大间距线性投影与支持向量机[J]. 自动化学报,2004,30(6):890-896.
- [9] 陈伏兵,张生亮,高秀梅. 小样本情况下 Fisher 线性鉴别分析的理论及其验证[J]. 中国图象图形学报,2005,10(8):984-991.
- [10] 邹宇华. 岭南中草药分类鉴别的机器嗅觉实现方法研究[D]. [硕士学位论文]. 广州:广东工业大学,2009.



伍世元(1985-),男,湖南邵阳人,硕士生,主要研究方向为嵌入式电子鼻及传感信号处理,仿生嗅觉技术在中药材、食品等质量检测中的应用,wsy-2050@163.com;



骆德汉(1958-),男,安徽人,博士,教授,主要研究方向为仿生嗅觉与智能信号处理,绿色电子产品的研发,dehanluo@gdut.edu.cn。