

基于 RBF-RF 级联分类器电子鼻对中药的快速鉴别*

邹慧琴¹, 李 硕², 闫永红^{3**}, 刘 勇³, 赵 婷³, 韩 玉⁴,
苏玉贞⁵, 彭 莲³

(1. 北京中医药大学图书馆 北京 100102; 2. 北京交通大学理学院 北京 100044;
3. 北京中医药大学中药学院 北京 100102; 4. 北京中医药大学中医药博物馆 北京 100029;
5. 天津医科大学附属肿瘤医院 天津 300060)

摘 要:目的:将电子鼻引入中药研究领域,探讨其在实际应用中的难点并提出解决方案,建立优化判别模型,为中药鉴别提供一种简便、快速、有效的分析方法,同时为气敏传感器的研发及应用提供新思路。方法:采用电子鼻提取中药气味特征,基于 MOS 传感器的离子迁移谱,建立中药气味指纹图谱。以传感器最大响应值为分析指标,针对鉴别难点,提出两种解决方案:尝试不同检测器,即扩充传感器数量,尽量缩小“嗅觉盲区”;采用“级联分类器”构建法,即采用径向基函数(RBF)与随机森林(RF)二级级联分类器构建判别模型。通过十折交叉验证和外部测试集验证对所建模型进行系统性能的评估。结果:两种方案准确、可行,具有较高的正判率和较好的泛化能力(所得最高正判率分别为 95%和 100%、96%和 80%)。结论:本研究首次采用“级联分类器”模式构建中药电子鼻鉴别的判别模型,在传感器数量有限的情况下,从所得数据中挖掘最大信息量,以“拆分任务、剥离难点、由易到难、分级递进”为原则,实现电子鼻对中药的快速、准确鉴别。所建模式识别法在可操作性、鉴别准确率和稳定性上均优于传统嗅觉识别法,为中药鉴别提供一种简便、快速的分析方法。

关键词:电子鼻 级联分类器 中药鉴别 径向基函数 随机森林

doi: 10.11842/wst.2013.09.003 中图分类号:R282.5 文献标识码:A

中药鉴定是中药生产、研究及应用领域至关重要的一步^[1]。传统性状鉴别法基于“辨状论质”,建立了一整套简便、有效的中药质量评价标准。现行《中国药典》中^[2]“性状”标准始终占据重要地位,具有独特的优越性。其中,气味特征与中药所含化学成分密切相关,是评价质量的重要指标之一^[3]。然而,传统嗅觉识别法主观经验性较强,不利于推广应用。目前常用的气味分析技术,如气相色谱法,前处理繁琐、溶剂消耗多,且不能反映气味整体特征^[4]。因此,在传统性状鉴别基础上,早日引入能数字化、量化气味特征的仪器迫在眉睫。

电子鼻(Electronic nose, E-nose)是近年发展迅速的高新技术,具有快速、灵敏、准确、无损和无污染等特点^[5]。但其检测系统——传感器阵列具有广谱响应性、交叉敏感特性等特点,导致信息冗余、模式识别难等局限性^[6]。在实际应用中,需寻找更适合的传感器、更优化的判别模型,从而获取最简便、准确的识别方法,实现不同样品的快速分类及鉴别。

本研究针对中药气味特征缺乏快速、有效鉴别手段的难题,以姜科常用 10 味中药为研究载体,从不同检测系统、不同分类器的筛选展开研究,首次将电子鼻结合级联分类器引入中药研究领域。针对鉴别难点,提出两种解决方案:扩充传感器数量,尽

收稿日期:2013-03-13

修回日期:2013-04-03

* 北京中医药大学自主选题项目(JYB22-XS041):基于仿生嗅觉系统的砂仁气味与药材品质相关性研究,负责人:邹慧琴。

** 通讯作者:闫永红,教授,博士生导师,主要研究方向:中药材鉴定,中药质量评价。

量缩小“嗅觉盲区”;建立径向基函数(Radial Basis Function, RBF)与随机森林(Random Forest, RF)的“二级分类器”,构建判别模型。所建模型实现了基于气味特征电子鼻对不同中药的准确、快速鉴别。结果显示,本研究针对鉴别难点提出的两种方案均可行,相关研究尚未报道。

1 仪器与试剂

1.1 实验材料

姜科常用 10 味中药饮片干姜、姜黄、高良姜、莪术、郁金、白豆蔻、草豆蔻、草果、益智和砂仁,购于北京同仁堂股份有限公司,经北京中医药大学闫永红教授鉴定分别来源于姜科植物姜 *Zingiber officinale* Rosc. 的干燥根茎、姜黄 *Curcuma longa* L. 的干燥根茎、高良姜 *Alpinia officinarum* Hance. 的干燥根茎、蓬莪术 *Curcuma phaeocaulis* Val. 的干燥根茎、广西莪术 *Curcuma kwangsiensis* S. G. Lee et C. F. Liang 的干燥块根、白豆蔻 *Amomum kravanh* Pierre ex Gagnep. 的干燥成熟果实、草豆蔻 *Alpinia katsumadai* Hayata. 的干燥近成熟种子、草果 *Amomum tsao-ko* Crevost et Lemaire. 的干燥成熟果实、益智 *Alpinia oxyphylla* Miq. 的干燥成熟果实以及阳春砂 *Amomum villosum* Lour. 的干燥成熟果实。样品的生产年份均为 2011 年。

1.2 电子鼻系统

采用法国 Alpha M.O.S. 公司生产的 α -FOX3000、 α -FOX4000 电子鼻系统。 α -FOX3000 包含 12 根金属氧化物传感器阵列,型号分别为 LY2/LG、LY2 /G、LY2/AA、LY2/GH、LY2/gCTL、LY2/gCT、T30/1、P10/1、P10/2、P40/1、T70/2、PA/2; α -FOX4000,比 α -FOX3000 多 6 根传感器:P30/1、P40/2、P30/2、T40/2、T40/1、TA/2。

2 方法

2.1 气味提取法

药材粉碎,过 2 号筛,精密称取 0.4 g 样品装入 10 mL 顶空瓶中,压盖密封。电子鼻对中药气味特征采集分为 3 个流程:顶空平衡、自动进样和信号采集。具体实验参数见表 1。

2.2 分类器的设计及判别模型的构建

各测试中药分别为一类,进行分类器的筛选,建立判别模型。在此判别模型中,可能存在 2 种(或以上)中药,彼此之间分类效果不佳。基于检测系统(传感器阵列)和后期数据处理(判别模型)是影响电子鼻工作效能的主要因素,原因可能为:一是传感器数量不足,对中药气味存在“嗅觉盲区”,从而影响判别灵敏度;二是气味相近的中药难以与其它几种药材在单一判别模型中实现分类。因此,结合实际分类情况,有必要进一步分析:尝试采用数量更多的传感器,旨在提取更完整、更全面的中药气味信息,尽量缩小由于检测器局限性带来的“嗅觉盲区”;提取鉴别难点,以气味相近、区分难度较大的 2 种(或以上)中药为各类,采取“级联分类器”模式再筛选二级分类器,在传感器数量有限的条件下,对所得信息进行充分数据挖掘,综合建立判别模型。

前期实验中,分别采用朴素贝叶斯(Naive Bayes Net, NBN)、多层感知器、RBF 网络、RF 等 12 种算法建立中药鉴别分类器,并进行比对分析。结合各算法的原理、判别效果以及对中药的适用性,较适宜算法为朴素贝叶斯、RBF 网络和 RF。

朴素贝叶斯是一个包含一个根节点、多个叶节点的树状贝叶斯网络。根节点是类别变量,描述对象的类别;叶节点是属性变量,描述待分类对象的

表 1 电子鼻实验参数表

顶空平衡		自动进样		信号采集	
类型	取值	类型	取值	类型	取值
顶空时间(s)	600	进样量(μ L)	500	采集时间(s)	200
顶空温度($^{\circ}$ C)	45	注射速度(μ L \cdot s $^{-1}$)	500	采集周期(s)	1.0
搅动速度(rpm)	250	注射器温度($^{\circ}$ C)	55	采集延迟(s)	600

属性。它是一种简单高效的分类器 ,在数据挖掘和模式识别中应用广泛^[7]。

RBF 网络是在借鉴生物局部调节和交叠接受区域知识的基础上提出的一种采用局部接受域来执行函数映射的人工神经网络，其在逼近能力、分类能力和学习速度等方面均优于 BP 网络^[8]。由于 RBF 网络能够逼近任意的非线性函数 ,学习收敛速度快等 ,目前已成功应用于非线性函数逼近、模式识别^[9]、信息处理^[10]等。

RF 是一个包含多个决策树的分类器，其输出类别是由个别树输出类别的众数而定 ,其中每棵决策树都会完整成长而不会剪枝。它有准确度高、学习速度快、能容忍内部噪声且不容易出现过拟合等优点 ,可用于数据挖掘、侦测偏离值和资料视觉化等 ,在医学、生物信息、管理学等领域广泛应用^[11]。

本研究所有计算均基于 Weka 软件，相关参数为默认值。

3 结果与讨论

3.1 中药气味分析及训练集、外部测试集的建立

α -FOX3000 电子鼻测定中药气味后 12 根传感器的阻值变化响应见图 1，12 条曲线分别代表 12 根传感器响应。本研究采用传感器最大响应值进行数据分析。

上述 10 味中药各取 10 个样本，采用循环交叉排列的方式组成序列，以减小实验的系统误差^[12]，组成训练集。另外每味再各取 2 个样本 ,组成外部测试集。

3.2 判别模型的初步构建及分类结果的比较

首先，以每味中药各为一类建立训练集($\varphi 1$)。采用 NBN、RBF、RF 3 种方法分别建立分类器 ,系统性能评估采用十折交叉验证法 (Ten-folds Cross-Validation ,TFCV) 和外部测试集验证法 (External Test set Validation ,ETSV)。其分类结果如表 2 ,RBF 分类器所建模型的正判率最高 ,分别为 96%、80%。因此，确定采用 RBF 网络分类器建立姜科 10 味药材的判别模型。3 种分类器的判别结果均显示莪术与益智相互错判的概率最大，由此确定此 10 味中药的鉴别难点为莪术与益智。鉴于传感器种类及数

表 2 3 种分类器对训练集($\varphi 1$)正判率的比较

分类器	TFCV/%	ETSV/%
NBN	91	80
RBF	96	80
RF	94	80

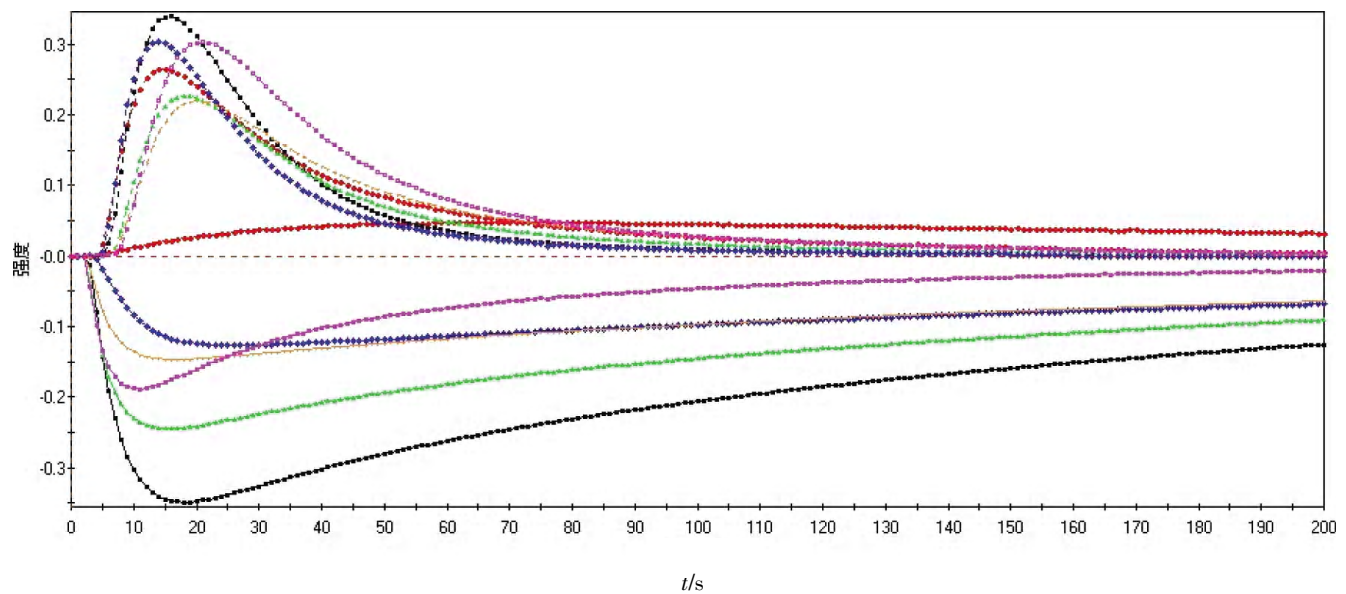


图 1 α -FOX3000 电子鼻 12 根传感器阻值变化响应图

量有限，要解决此鉴别难点可以通过两条途径：一是采用数量更多的传感器，提取更完整、更全面的中药气味信息，尽量缩小由于检测器的“嗅觉盲区”；二是利用多级分类器相结合建立判别模型，即针对分类结果相近的莪术、益智作二级分类器的筛选，从有限的数据中挖掘出最大的有效信息量。

3.3 相同样品、不同检测器的分类结果

α -FOX4000 电子鼻测定中药气味后 18 根传感器的阻值变化响应见图 2。同样，以每味中药的传感器最大相应数值各为一类建立训练集(φ_2)。

同样采用 NBN、RBF、RF 3 种方法分别建立分类器，系统性能评估采用 TFCV 和 ETSV。其分类结果如表 3，NBN 分类器所建模型的正判率最高，分别为 99%、100%。综合比较表 2、表 3 的分类结果，不难看出： α -FOX4000 所含的 18 根传感器对姜科 10 味药材的区分能力更佳，所得正判率均在 85% 以上。

3.4 级联分类器的构建

基于表 2，在 RBF 分类器为一级分类器的基础上，针对鉴别难点-莪术、益智作二级分类器筛选。以莪术、益智各为一类、其它 8 味中药为一类建立

训练集(φ_3)。结果显示：RF 所建分类器模型的正判率最高，分别为 97%、95%，见表 4。因此，确定采用 RF 法建立莪术、益智的二级分类器模型。

综上所述，在原始观测值基础上，采用更多、更全面的传感器阵列提取中药气味特征；或采用 RBF、RF 分别为一级、二级分类器，建立级联分类

表 3 3 种分类器对训练集(φ_2)正判率的比较

分类器	TFCV/%	ETSV/%
NBN	99	100
RBF	95	100
RF	95	85

表 4 3 种分类器对训练集(φ_3)正判率的比较

分类器	TFCV/%	ETSV/%
NBN	83	85
RBF	93	90
RF	97	95

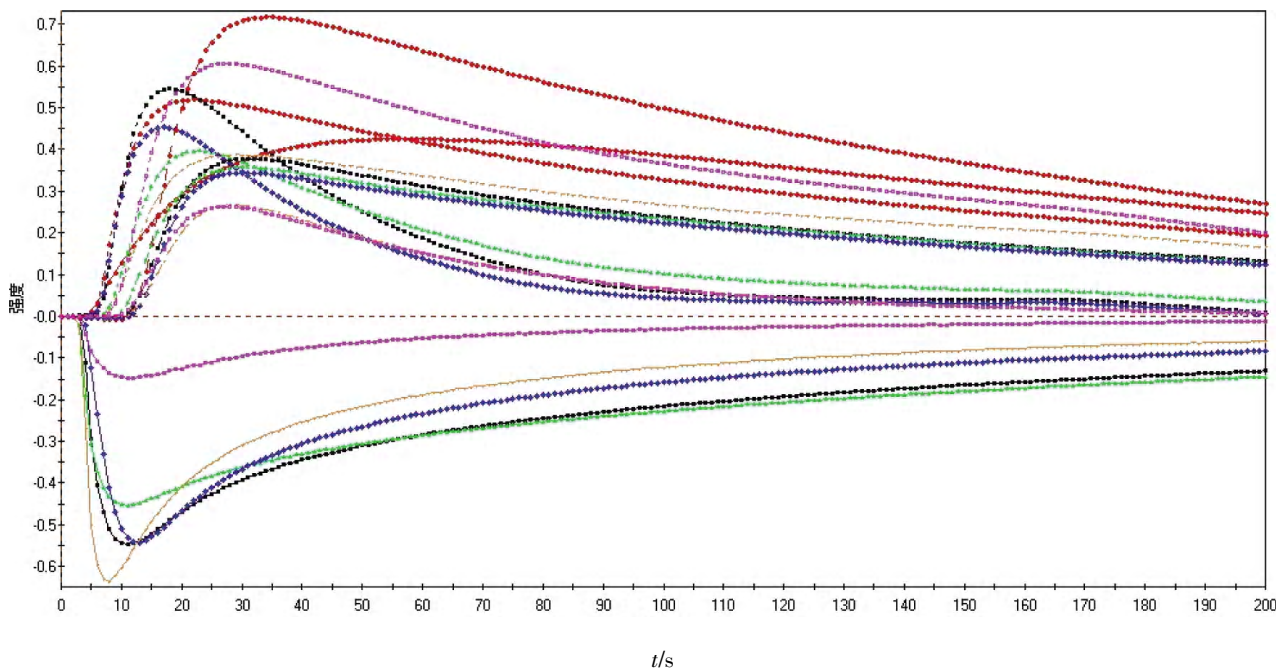


图 2 α -FOX4000 电子鼻 12 根传感器阻值变化响应图

器,综合构建判别模型,均能实现电子鼻对 10 味中药的快速鉴别。

4 结语

相较于 α -FOX3000 而言, α -FOX4000 增加的传感器数量为 6 根,均为 T 型和 P 型,此次实验结果表明,这两种类型的传感器对中药气味的提取及识别专属性更强,为今后电子鼻应用于中药研究领域的传感器筛选与优化提供思路。

本实验样品仅为 10 味中药,在实际应用中,有可能面对更多、更复杂的中药样品,笔者建议不必拘泥于级联分类器的级数,可根据具体分析对象和研究目的,再结合本课题组提出的传感器优化法^[13],多法交叉、择优建模。本研究在“拆分任务、由易到难,分级递进”的原则下,基于级联分类器构建的电子鼻判别模型,具有学习速度快、准确率高及无过拟合现象等优点,在可操作性、鉴别准确率和稳定性上均优于传统嗅觉识别法,可望为中药鉴别提供一种简便、快速的分析方法。

参考文献

1 陶燕蓉,陈曦.中药质量评价技术的国内外研究现状及分析.中药与临床,2011,2(2):59~62.

- 2 国家药典委员会.中华人民共和国药典(一部).北京:中国医药科技出版社,2010.
- 3 杨锡仓,马炎铭.5 种中药材的伪品识别经验介绍.甘肃中医,1994,7(24):43~45.
- 4 丁平,方琴,徐鸿华.砂仁及其近缘植物化学成分的气相色谱指纹图谱研究.华西药理学杂志,2004,19(5):330~332.
- 5 Tang X W, He H J, Geng L H, *et al.* Evaluation of maturity and flavour of melons using an electronic noses. *Agricultural Science & Technology*, 2011, 12(3):447~450.
- 6 邹慧琴,刘勇,闫永红,等.电子鼻技术及应用研究进展.传感器世界,2011,17(11):6~11.
- 7 张连文.贝叶斯网引论.北京:科学出版社,2006.
- 8 侯媛彬.神经网络.西安:西安电子科技大学出版社,2007.
- 9 Evans P, Persaud K C, McNeish A S, *et al.* Evaluation of a radial basis function neural network for the determination of wheat quality from electronic nose data. *Sens Actuators B Chem*, 2000, 69(3):348~358.
- 10 Tudu B, Jana A, Metla A, *et al.* Electronic nose for black tea quality evaluation by an incremental RBF network. *Sens Actuators B Chem*, 2009, 138(1):90~95.
- 11 方匡南,吴建彬,朱建平,等.随机森林方法研究综述.统计与信息论坛,2011,26(3):32~38.
- 12 赵镭,史波林,汪厚银,等.电子鼻传感器筛选的组合优化法研究.食品科学,2009,30(20):367~370.
- 13 邹慧琴,刘勇,闫永红,等.电子鼻 MOS 传感器阵列优化及其在中药材快速鉴别中的应用.中国中药杂志,2013,38(2):161~165.

Rapid Identification of Traditional Chinese Medicine Using Electronic Nose Based on RBF-RF Cascade Classifier

Zou Huiqin¹, Li Shuo², Yan Yonghong³, Liu Yong³, Zhao Ting³, Han Yu⁴, Su Yuzhen⁵, Peng Lian³

(1. Library, Beijing University of Chinese Medicine, Beijing 100102, China;

2. College of Science, Beijing Jiaotong University Beijing 100044, China;

3. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100102, China;

4. Museum of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China;

5. Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China)

Abstract: This study was aimed to apply the electronic nose (E-nose) in the research of traditional Chinese medicine (TCM). The discussion was made on difficulties of using E-nose. The solution plan was proposed and the discrimination model was established. It provided a simple, rapid and effective analysis method in the identification of TCM. It also provided new ideas for the research and application of gas sensor arrays. E-nose was used in the extraction of TCM scent characteristics. Based on ion mobility spectrometry of MOS sensor, the fingerprint of TCM scent was established. The maximum response value of the sensor was used as analysis index. According to the difficulties of identification, two solution plans were proposed. Firstly, different detectors were employed to complete the classification. Secondly, radial basis function (RBF) and random forests (RF) were combined and then a cascade

classifier was constructed in order to achieve the maximum of information obtained in conditions where the number of measurements, metal oxide semiconductor sensors in E-nose was limited. The results showed that both plans were accurate and practical with relatively high upper correct judge rate and better cross-validation (The highest upper correct judge rates were 95% and 100%, 96% and 80%, respectively). It was concluded that this study firstly applied cascade classifier in the establishment of TCM identification by E-nose. With limited amount of sensors, the maximum information was received through data mining. Using E-nose in the identification of TCM was rapid and accurate. The established pattern recognition method was maneuverable with accurate identification rate and stability compared to conventional sensory identification method. It provided a simple and rapid analysis method for the identification of TCM.

Keywords: Electronic nose, cascade classifier, traditional Chinese medicine identification, radial basis function, random forests

(责任编辑 叶丽萍 张志华, 责任译审 王 晶)