

文章编号:1000-582X(2012)09-157-06

## 可视化仿生鼻对五小香型白酒的识别

罗小刚<sup>1</sup>, 张 亚<sup>1</sup>, 侯长军<sup>1</sup>, 易 彬<sup>2,3</sup>, 霍丹群<sup>1</sup>, 赵飞翔<sup>1</sup>

(1. 重庆大学 a. 生物流变科学与技术教育部重点实验室; b. 生物工程学院, 重庆 400044;  
2. 泸州老窖股份有限公司, 四川 泸州 646000; 3. 国家固态酿造工程技术中心, 四川 泸州 646000)

**摘 要:**建立了一种有潜力的人工仿生鼻方法,用于分析典型的中国五小香型白酒。为减少误差、去除冗余数据,采用阈值限定结合平方和均值方根的方法对原始 RGB 数据进行预处理;在可视化区分的基础上采用分层聚类分析(HCA)、主成分分析(PCA)以及支持向量机(SVM)的分析方法,对预处理后的数据进行分析。聚类分析方法依据香型衍生的不同可以实现正确的归类;利用主成分分析得到的前 3 个主成分包含了白酒 80.79% 信息量,可以将不同香型白酒正确区分;支持向量机的可视化仿生鼻能对白酒香型进行有效区分,其识别的准确率达到 100%。研究结果表明,基于可视化传感技术的可视化仿生鼻可以用于五小香型白酒的识别。

**关键词:**可视化仿生鼻;香型;支持向量机;白酒识别

**中图分类号:**TS262.3; TS261.7; TP29 **文献标志码:**A

## Identification of different aromatic Chinese liquors by colorimetric artificial nose

Luo Xiaogang<sup>1</sup>, Zhang Ya<sup>1</sup>, Hou Changjun<sup>1</sup>, Yi Bin<sup>2,3</sup>, Huo Danqun<sup>1</sup>, Zhao Feixiang<sup>1</sup>

(1a. Key Laboratory of Biorheological Science and Technology; 1b. Ministry of Education, Chongqing University, Chongqing 400030, China; 2. Luzhou Laojiao Co. Ltd., Luzhou, Sichuan 646000, China; 3. Nation Engineering Research Center of Solid-State Brewing, Luzhou, Sichuan 646000, China)

**Abstract:** A potential instrument, the colorimetric nose, is developed to identify the fragrances of five different Chinese liquors. Firstly, in order to minimize error, the RGB values of the raw data are preprocessed using a threshold before further analysis. The output of the artificial nose is then analyzed by hierarchical routing cluster analysis (HCA), principal component analysis (PCA) and support vector machine (SVM). It is found that HCA can perform correctly classify fragrances into five different classes. However, using the first three components identified by PCA analysis, representing 80.79% of the variance, the five individual fragrances can be reliably distinguished. Finally, it also shows that the five constituent fragrant liquors can also be reliably classified with 100% accuracy by SVM. These results show that the colorimetric artificial nose, a simple and efficient detection and identification tool, has great potential to identify different constituent fragrant liquors reliably well.

**Key words:** colorimetric artificial nose; fragrances; support vector machine (SVM); liquor identification

收稿日期:2012-04-07

基金项目:国家自然科学基金资助项目(30770568);中央高校基本科研业务费资助项目(CDJXS10231179, CDJSX102300);四川省重点实验室开放基金资助项目(NJ20094);重庆市科委攻关项目(2008AC7037)

作者简介:罗小刚(1974-),男,重庆大学副教授,博士后,主要从事生物医学传感器、生物医学仪器方面的研究,  
(E-mail)luosteel@cqu.edu.cn。

白酒的香型是指具有悠久历史的独特酿造工艺形成的以某些香味物质成分为主的产品。五小香型是在五大香型的基础上衍生出来的自成一体香型,是继五大香型白酒之后又划分出来的 5 个典型的香型代表,它包括:药香型、芝麻香型、豉香型、特香型、兼香型<sup>[1]</sup>。目前,在白酒的成分检测中主要采用质谱、气相色谱-质谱联用技术等常规方法。文献[2]采用气相色谱技术对不同白酒中的 34 种主要挥发性香味化学物质进行分析;文献[3]采用气相色谱-质谱联用技术对浓香型白酒的微量香型组分进行了分析研究;文献[4]利用气相色谱技术对 8 种酱香型白酒的指纹图谱进行了研究。虽然这些技术能够定性和定量检测出白酒中的几百种化合物,但是通常需要昂贵的大型仪器,并需要专门人员操作。另外,由于白酒中含有多种微量挥发性香味物质,仅依靠大型设备检测,并不能完全展现白酒的成分特征。基于交叉响应的电子鼻/电子舌技术很好地弥补了上述技术的不足,该技术是由具有广谱响应特性、较大交叉灵敏度以及对不同气体有不同灵敏度的敏感元件组成,具有响应时间短、检测速度快等优点。但是,由于电子鼻/电子舌中的传感器元件呈现出非特异性,对成分或化学结构相似的复杂混合物识别能力有限<sup>[5]</sup>。同时白酒中的乙醇和水也严重影响电子鼻/电子舌对酒中微量香味成分识别<sup>[6]</sup>。

作为一种新出现的基于可视化传感器阵列技术的可视化仿生鼻技术,与传统的电子鼻/电子舌相比更加简便直观,特异识别能力更强,在区分复杂的混合物甚至高度相似的复杂混合物方面具有广阔的前景<sup>[5,7-8]</sup>。笔者利用可视化仿生鼻已成功实现了对不同种类的氨基酸、蛋白质<sup>[9,10]</sup>、有毒气体<sup>[11]</sup>以及酒<sup>[12]</sup>的识别。五小香型是最近才划分出来的 5 个典型的新香型代表,目前对五小香型的不同香型识别以及影响香型的因素还没有系统的研究。笔者利用可视化仿生鼻对五小香型白酒的香型进行识别,探讨了该仿生鼻用于白酒检测的方法,以便于生产者对形成不同香型的因素进行深入探究。

## 1 实验部分

### 1.1 检测对象

五小香型的典型代表酒样:兼香型——湘泉的酒精精度为 54% vol、药香型——董酒的酒精精度为 45% vol、芝麻香型——景芝白干的酒精精度为 49%

vol、豉香型——玉冰烧的酒精精度为 29% vol、特香型——四特老窖的酒精精度为 45% vol。

### 1.2 实验装置及方法

如图 1 所示,实验装置由 3 部分构成:气体挥发装置、气体检测装置以及控制和数据采集处理装置。其中,气体检测装置及控制和数据采集处理装置构成可视化电子鼻。气体挥发装置主要由挥发室、真空泵以及阀门、导管组成。气体检测装置主要由 LED 光源,可视化阵列传感器以及 CCD 传感器构成。其中可视化阵列传感器采用卟啉及其衍生物(主要为金属卟啉)、酸碱指示剂以及极性指示剂,构建了 6×6 的阵列传感器;疏水性的反相硅胶板作为阵列传感器的基底材料,有效阻止了空气湿度对检测造成的干扰。控制和数据采集处理系统主要是实现检测过程的自动化以及数据采集处理的自动化。实验过程中为了减少乙醇和水蒸气对实验结果的影响,气体挥发和检测装置始终处在恒温的环境下;同时为了实时在线监控温度,气体挥发和检测装置内均安装有温度传感器。

实验开始前,将待测酒样通过进样口装入挥发。启动可视化电子鼻系统,系统会自动检测环境温度,同时打开温度加热装置。当温度到达恒温后,继续保持 10 min 后系统自动打开真空泵,气体通过导管进入气体检测装置内,反应开始;反应 5 min 后,系统通过 CCD 传感器采集图像即为反应后的图像。反应前后光谱图通过控制和数据采集系统的一系列的图像处理最后获取反应前和反应后的光谱差图(即差谱图),同时系统保存每一个颜色分量的变化数值[即反应前后图像的红(R)、绿(G)、蓝(B)分量的差值(以阵列中点(2,4)为例,如式(1)所示)]。差谱图反应结果的直观表达,RGB 值则代表了反应过程的详细变化。实验过程中分别对每种待测酒样检测 10 次,即每个酒样有 10 组平行样,最后得到 50 组实验数据向量。

$$\left. \begin{aligned} \Delta R_{i(2,4)} &= |R_{i(2,4)} - R_{0(2,4)}|, \\ \Delta G_{i(2,4)} &= |G_{i(2,4)} - G_{0(2,4)}|, \\ \Delta B_{i(2,4)} &= |B_{i(2,4)} - B_{0(2,4)}|. \end{aligned} \right\} \quad (1)$$

式中: $R_{0(2,4)}$ 、 $G_{0(2,4)}$ 、 $B_{0(2,4)}$  分别表示传感器阵列中第 2 行第 4 列阵列点在反应前的红色、绿色、蓝色反应分量; $R_{i(2,4)}$ 、 $G_{i(2,4)}$ 、 $B_{i(2,4)}$  分别为阵列点反应后的红色、绿色、蓝色分量; $\Delta R_{i(2,4)}$ 、 $\Delta G_{i(2,4)}$ 、 $\Delta B_{i(2,4)}$  分别为该阵列点反应前后红色、绿色和蓝色分量差值。

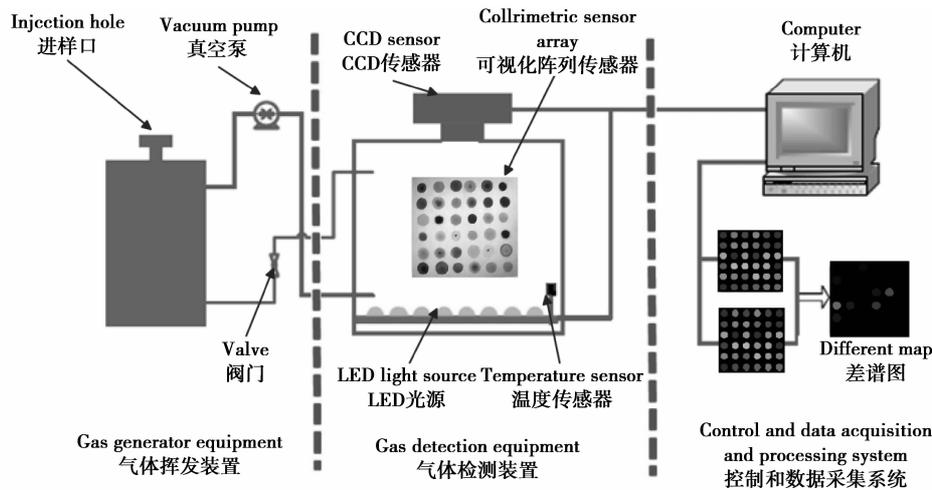


图1 可视化仿生鼻检测系统示意图

### 1.3 数据预处理

本实验中采用 $6 \times 6$ 的传感阵列,每个颜色响应点又包含 $R$ 、 $G$ 、 $B$ 3种颜色分量。因此响应结果具有 $6 \times 6 \times 3$ 即108维数据。每一维数据理论上的变化范围为 $0 \sim 255$ ,因此可视化阵列传感器在理论上的响应模式可达 $256^{108}$ 种,保证了阵列传感器的巨大的识别能力。但是实际上由于传感器制备过程中以及实验设备和操作,实验结果都会存在一定程度的误差。因此为了减少误差需要对原始数据进行预处理。通过设定阈值,将同一传感元件 $RGB$ 之和小于25的点的 $RGB$ 变化值归零,从而将5种酒样的光谱变化值显著的阵列点的光谱变化数据作为指纹数据。由于颜色变化是一个综合的变化量,而单独的光谱变化并不能充分反应颜色变化,因此为了充分反应颜色的变化特性和减少误差值在实验结果中的比重,利用公式 $\sqrt{(R^2+G^2+B^2)}/3$ 进行降维,最后获取每个显著变化点的综合变化特征值。

## 2 结果与讨论

### 2.1 不同香型白酒的传感器指纹图谱

利用可视化仿生鼻对五小香型的酒样进行检测,把反应5 min后的同种酒样的平行样差谱图进行平均作为该种酒样最终的指纹图谱。图2所示为5种小香型白酒的反应差谱图,也可以称作5种小香型的指纹图谱。不同香型的差谱图中响应点的数量和光谱变化强度是不同的,每种香型的酒样都有其独有的反应差谱图。差谱图上有些阵列点在5种小香型白酒中都是敏感点,都会发生光谱变化。酒样中对应这些敏感点所发生反应的物质在各种香型中都存在,但是由于含量的不同,与敏感点反应的

强度不同,导致光谱变化强度也不同。同时,酒样中的各种成分与传感器阵列上的材料发生反应,由于阵列点并非绝对特异性,可以与多种物质发生反应,从而产生交叉响应。最终的反应差谱图既能反映不同成分,也能反映相同成分的不同含量。反应差谱图虽然能给出不同香型酒样的综合特征直观的描述,可以直观地区分不同香型的酒样,但却无法给出具体的区分依据。

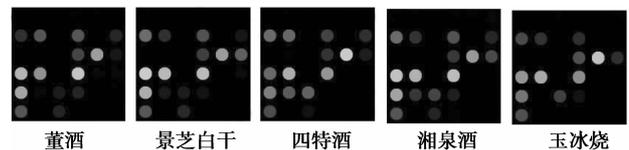


图2 五小香型指纹图谱

### 2.2 不同香型白酒的聚类分析

分层聚类分析(hierarchical cluster analysis, HCA)是将物理或者抽象对象的集合分组成为由类似的对象组成的多个类的分析过程,它可以衡量不同数据源间的相似性,并把数据源分类到不同的簇中,是目前最常用的模式识别方法。聚类图表示的是不同香型白酒之间的相似度,计算过程中所用的数据是阵列上显著变化特征点与不同香型白酒之间反应前后的 $RGB$ 综合变化特征值,同时聚类图还反映平行样实验之间的差异。笔者采用Spss软件对经过预处理后的25次试验所获得的数据进行聚类分析,结果如图3所示。在图3中可以看出,同种香型的酒样可以完全聚在一起,而不同香型的酒样可以完全正确分离;5种小香型的酒样距离由近到远分别为药香型—兼香型—芝麻香型—豉香型—特香

型。四特老窖的酒精度为 45%vol,玉冰烧的酒精度为 29%vol,湘泉的酒精度为 54%vol,景芝白干的酒精度为 49%vol,董酒的酒精度为 45%vol。四特老窖和董酒的酒精度虽然相似,但在聚类图上却是距离最远的,这表明可视化阵列传感器识别的只是白酒中的特殊的微量香味成分而非乙醇的浓度。由图 3 可知,5 种不同的香型大致可以聚为两大类:一类是玉冰烧和四特,另一类是董酒、湘泉和景芝白干。芝麻香型的景芝白干和兼香型的湘泉酒都是以优质高粱作为主要原料,四特酒和玉冰烧则采用大米作为酿酒原料,董酒是以优质高粱为主要原料,配以小麦和小米以及中草药作为原料。由此可见,不同香型之间的关系与他们使用的主要原料有着一定的关系。

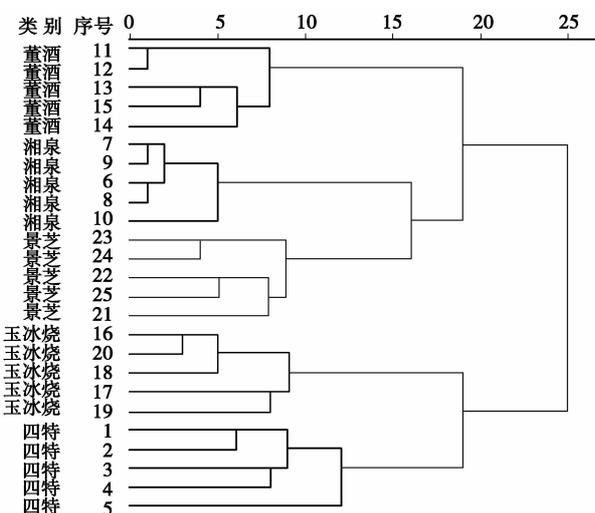


图 3 五小香型聚类图

五大香型中的浓、酱、清、米香型是基本香型,它们是构成其他香型的基础,其他香型是在这 4 种基本香型的基础上,以其中一种、两种或者两种以上的香型,在不同酿造工艺的揉和下衍生而成的。药香型是由浓香型、酱香型、米香型 3 种香型结合演变而来的香型,芝麻香型则是以酱香为基础衍生出的香型,兼香型白酒中的微量成分的种类和含量介于浓香型和酱香型之间,既具有浓香型的窖香浓郁也具有酱香型的留香持久等特点。药香型、芝麻香型、兼香型 3 种香型都跟酱香有很大关联,主体香味中酱香占很大比例。豉香型是以米香为基础演变而来的,特香型是浓香型、清香型辅以酱香型衍生而成的。由此可见,聚类分析是根据不同香型的演变关系来区分不同香型的。与五大香型相比,五小香型没有五大香型中那么明显的主体香味物质<sup>[12]</sup>,香味

本身就是由四大香型结合而成的,所以无法直接根据主体香味来分类,而是根据香型的演变关系或者是香型的远近来区分,本质上还是依据不同香型之间的香味差别来区分的,与五大香型白酒的区分还是有一定的相似性。

### 2.3 不同香型白酒的主成分分析

主成分分析方法主要是将原来多个特征值化为少数几个互不相关或者相互独立的综合特征值的一种统计方法,可以简化数据,揭示变量之间的关系,为进一步分析总体的行政和数据的统计特性提供一些重要的信息。对数据预处理后的 25 组实验数据向量进行主成分分析,第 1 主成分的贡献率为 44.12%,第 2 主成分的贡献率为 19.88%,第 3 主成分的贡献率为 16.79%。通常电子鼻/电子舌技术前两个主成分可以占到贡献率的 95%~99% 的信息量,代表相同信息量所需的主成分越多,传感器对不同物质的识别能力越强。而可视化仿生鼻前 3 个主成分仅仅贡献信息量的 80.79%,可视化传感阵列信号的这种高离散性是因为金属卟啉的分子识别功能,即可视化阵列传感器比一般的化学传感器更适合用于相似样本的区分<sup>[8]</sup>。图 4(a)是第 1 主成分和第 2 主成分散点图,5 种小香型在该图上可以完全区分开来。图 4(b)是第 1 主成分和第 3 主成分散点图,在这上面五小香型的白酒可以完全正确区分开来。第 1 主成分所在的横坐标上,自左向右分别为玉冰烧、四特、景芝、湘泉、董酒。第 2 主成分所在的纵坐标上,自上向下分别为景芝、湘泉、玉冰烧、四特、董酒,其中董酒和湘泉、四特和景芝在第 2 主成分上的值相似。第 3 主成分所在的纵坐标上,自上到下分别为四特、湘泉、玉冰烧、董酒、景芝,其中四特和湘泉、董酒和景芝在第 3 主成分上的值相似。

芝麻香型的景芝白干和兼香型的湘泉酒都是以优质高粱作为主要原料,四特酒和玉冰烧则采用大米作为酿酒原料,董酒是以优质高粱为主要原料,配以小麦和小米以及中草药作为原料。不同的酿酒原料带来了不同的分类结果。从图 4(b)中可以看出,高粱为主要原料的景芝白干、湘泉以及董酒分布在图中的右下角,而以大米作为主要原料的四特酒和玉冰烧分布在图中的左上角。第 1 主成分所反映的正是不同香型白酒的酿酒原料的差异。这与聚类分析的结果相对应。湘泉酒的酒精度为 54%vol,董酒和四特酒均为 45%vol,景芝白干和玉冰烧分别为 49%vol 和 29%vol。酒精度的不同不仅与后期勾兑有关,而且还与发酵工艺有关,发酵工艺中的各个阶段均能对白酒的香味成分产生影响。第 2 主成分所

代表的是不同的发酵和勾兑工艺对白酒香型的影响。在第3主成分中,湘泉酒和景芝、四特和董酒比较相似。这与发酵过程中使用的发酵剂和制曲的原料有关。湘泉酒采用的是小麦制成的高温大曲作为糖化发酵剂,四特酒利用的也是大曲,制曲原料是由小麦面粉和麸皮混合而成。景芝酒的发酵使用的是

以高粱为原料的高温曲、以小麦制成的中温曲、强化菌曲组成的混合发酵剂,董酒使用的是以大麦为原料的大曲和以大米为原料的小曲混合。玉冰烧酒使用的糖化发酵剂为大酒饼,大酒饼是以大米和大豆为原料制成的,独特的配料和制曲工艺使得豉香型的玉冰烧具有独特的风味。

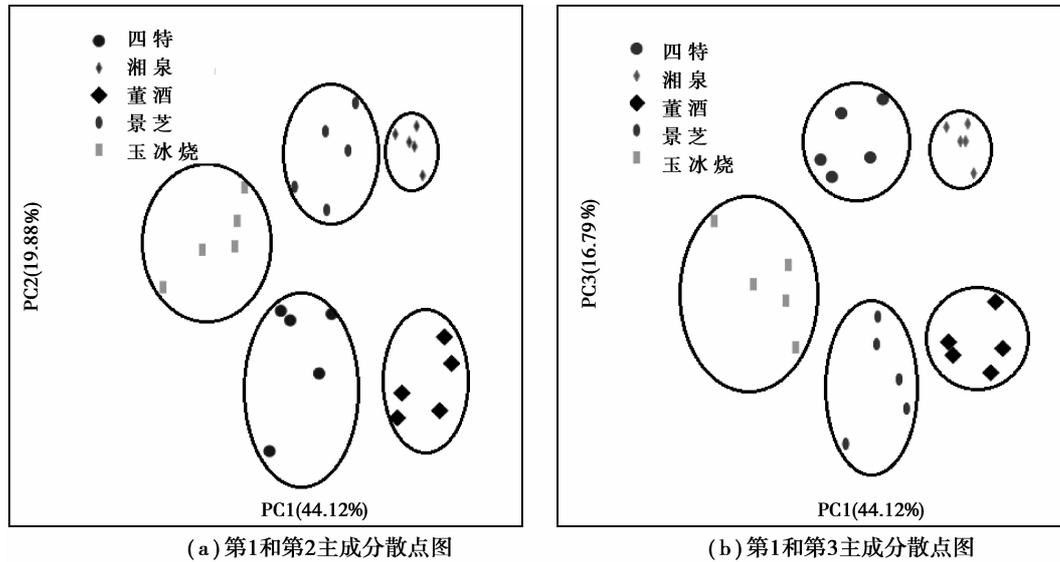


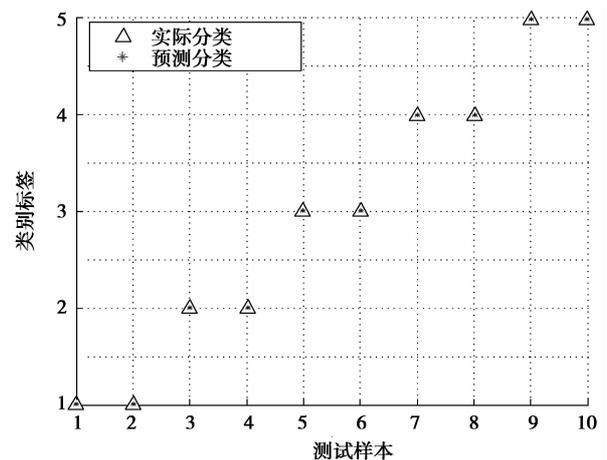
图4 主成分分析结果

五小香型白酒主成分分析的结果表明,主成分分析反映了白酒生产过程中各方面的差异。在香型区分过程中,第1主成分通常表示酿造原料的差异,第2主成分反映勾兑和发酵工艺的差异,第3主成分表示糖化发酵剂和制曲工艺的差异。

#### 2.4 支持向量机(SVM)对不同香型的识别

支持向量机(supporting vector machine, SVM)是基于统计学习理论上而发展起来的一种新型的通用学习方法,它遵循结构风险最小化原理,具有较好的泛化能力<sup>[13]</sup>。在解决了小样本、非线性、高纬度、局部极小点等实际问题中表现出许多特有的优势,并能够推广应用到函数拟合等其他机器学习问题中,已在模式识别、函数逼近、信号处理等领域得到了广泛的应用<sup>[14]</sup>。笔者利用MATLAB2009b及台湾大学林智仁(Lin Chih-Jen)等人开发的LibSVM工具箱,对全部50个样本建立SVM模型进行分析,从而对白酒的不同香型特征进行定性识别鉴定。从50个样本中随机抽取40个样本作为训练集,剩下10个样本作为测试集,即每个香型的样本有2个测试样本。识别结果如图5所示,图中横坐标表示测试的样本,纵坐标表示识别的类。从图5可以看出,实际测试的样本和预测样本

完全重合,SVM对测试集的预测准确率达到100%。



类别标签:1. 四特酒;2. 湘泉酒;3. 董酒;4. 景芝老白干;5. 玉冰烧酒

图5 SVM识别结果分类图

### 3 结论

笔者采用可视化仿生鼻对五小香型的典型代表酒样进行检测,得到了五小香型白酒的指纹图谱。HCA和PCA分析结果表明五小香型之间的差异来

自于原料的不同、勾兑和发酵工艺的差异以及糖化发酵剂和制曲工艺的差异。HCA 中依据香型衍生的不同可以实现正确的归类;PCA 结果前 3 主成分包含了每个香型 80.78% 的信息,不同香型之间在 PCA 下可准确进行分类;SVM 对香型的识别结果,表明基于可视化传感阵列技术的可视化仿生鼻可以根据香型的性质对不同的酒样进行显著的分辨,SVM 对五小香型分类识别准确率达到 100%,准确率优于传统的常规方法;同时检测时间仅需 5 min,优于各种大型仪器的检测时间。由此可见,基于本文的可视化仿生鼻系统,采用 HCA、PCA 以及 SVM 在五小香型白酒识别上都具有检测快速,操作简捷,分辨率高的特点,显示出其在白酒的在线质量检测 and 真伪鉴定上具有潜在的应用价值。

#### 参考文献:

- [1] 王传荣. 白酒的香型及其风味特征研究[J]. 酿酒科技, 2008(9): 49-52.  
WANG Chuanrong. Research on liquor flavor types & their flavoring characteristics [J]. Liquor-Making Science & Technology, 2008(9): 49-52.
- [2] 周围, 周小平, 赵国宏, 等. 名优白酒质量指纹专家鉴别系统[J]. 分析化学, 2004, 32(6): 735-740.  
ZHOU Wei, ZHOU Xiaoping, ZHAO Guohong, et al. Quality fingerprint expert discriminating system of Chinese famous white spirits [J]. Chinese Journal of Analytical Chemistry, 2004, 32(6): 735-740.
- [3] 黄艳梅, 卢建春, 李安军, 等. 采用气相色谱-质谱分析古井贡酒中的风味物质[J]. 酿酒科技, 2006(7): 91-94.  
HUANG Yanmei, LU Jianchun, LI Anjun, et al. Analysis of flavoring substances in Gujing Gongjiu liquor by GC/MS [J]. Liquor-Making Science & Technology, 2006(7): 91-94.
- [4] 吴天祥, 郑岩, 汤庆莉. 酱香型白酒 GC 指纹图谱的研究[J]. 酿酒科技, 2008(10): 30-36.  
WU Tianxiang, ZHENG Yan, TANG Qingli. Study on gas chromatography fingerprint map of maotai-flavor liquor [J]. Liquor-Making Science & Technology, 2008(10): 30-36.
- [5] Zhang C, Bailey D P, Suslick K S. Colorimetric sensor arrays for the analysis of beers: a feasibility study [J]. Journal of Agricultural and Food Chemistry, 2006, 54(14): 4925-4931.
- [6] García M, Aleixandre M, Gutiérrez J, et al. Electronic nose for wine discrimination [J]. Sensors and Actuators B: Chemical, 2006, 113(2): 911-916.
- [7] Zhang C, Suslick K S. Colorimetric sensor array for soft drink analysis [J]. Journal of Agricultural and Food Chemistry, 2007, 55(2): 237-242.
- [8] Feng L, Musto C J, Kemling J W, et al. Colorimetric sensor array for determination and identification of toxic industrial chemicals [J]. Analytical Chemistry, 2010, 82(22): 9433-9440.
- [9] 霍丹群, 张国平, 侯长军, 等. 基于新型可视化学传感阵列的氨基酸快速识别[J]. 分析化学, 2010, 38(8): 1115-1120.  
HUO Danqun, ZHANG Guoping, HOU Changjun, et al. A colorimetric sensor array for identification of natural amino acids [J]. Chinese Journal of Analytical Chemistry, 2010, 38(8): 1115-1120.
- [10] Hou C J, Dong J L, Zhang G P, et al. Colorimetric artificial tongue for protein identification [J]. Biosensors and Bioelectronics, 2011, 26 (10): 3981-3986.
- [11] Luo X G, Hou C J, Huo D Q, et al. A novel chemical detector using colorimetric sensor array and pattern recognition methods for the concentration analysis of  $\text{NH}_3$  [J]. The Review of Scientific Instruments, 2010, 81(10): 105113-105116.
- [12] 霍丹群, 尹猛猛, 侯长军, 等. 可视化阵列传感器技术鉴别不同香型白酒[J]. 分析化学, 2011, 39(4): 516-520.  
HUO Danqun, YIN Mengmeng, HOU Changjun, et al. Identification of different aromatic Chinese liquors by colorimetric array sensor technology [J]. Chinese Journal of Analytical Chemistry, 2011, 39 (4): 516-520.
- [13] 彭光金, 俞集辉, 韦俊涛, 等. 特征提取和小样本学习的电力工程造价预测模型[J]. 重庆大学学报, 2009, 32(9): 1104-1110.  
PENG Guanjin, YU Jihui, WEI Juntao, et al. Cost forecast model for power engineering based on feature extraction and small-sample learning [J]. Journal of Chongqing University, 2009, 32(9): 1104-1110.
- [14] 邓乃扬, 田英杰. 数据挖掘中的新方法-支持向量机[M]. 北京: 科学出版社, 2004.

(编辑 张 苹)