

三维荧光光谱结合 PARAFAC 和 GA 对中国白酒 品牌的鉴别

朱焯炜^{1,2} 阙立志¹ 吴亚敏¹ 陈国庆¹ 徐瑞煜¹ 朱 拓³

¹江南大学理学院, 江苏 无锡 214122

²江南大学物联网工程学院, 江苏 无锡 214122

³河海大学能源与电气学院, 江苏 南京 210098

摘要 为了实现对对中国白酒品牌的有效鉴别,在比较了不同品牌白酒三维荧光光谱特性的基础上,采用平行因子方法(PARAFAC)结合遗传算法(GA)获得训练样本和测试样本的浓度得分。同时利用支持向量机(SVM)方法建立中国白酒的鉴别模型,预测准确率为97.5%。PARAFAC分解得到的三个组分的浓度得分,在一定程度上反映了品牌之间的差异。PARAFAC与GA算法的有效结合为未知样本的鉴别提供了一种快速准确的方法。研究结果表明,PARAFAC-GA-SVM的鉴别模型具有更强的预测能力,该方法能够有效提取白酒的特征光谱信息,同时又大大降低了SVM输入变量的维数。研究结果为中国白酒的鉴别提供了一种新的思路。

关键词 光谱学; 平行因子分析法; 遗传算法; 中国白酒

中图分类号 O436

文献标识码 A

doi: 10.3788/CJL201542.0615002

Identification of Chinese Liquors by Three-Dimensional Fluorescence Spectra Combined with PARAFAC and Genetic Algorithm

Zhu Zhuowei^{1,2} Que Lizhi¹ Wu Yamin¹ Chen Guoqing¹ Xu Ruiyu¹ Zhu Tuo³

¹School of Science, Jiangnan University, Wuxi, Jiangsu 214122, China

²School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

³College of Energy and Electrical Engineering, Hohai University, Nanjing, Jiangsu 210098, China

Abstract In order to classify the brands of Chinese liquors effectively, the fluorescence characters of Chinese liquors are compared and analyzed. The concentration scores of training samples and testing samples are obtained by using parallel factor method (PARAFAC) combined with genetic algorithm (GA). Support vector machine (SVM) method is adopted to establish the identification model of Chinese liquors, and the accuracy rate of prediction is 97.5%. The experimental results show that the concentration scores of the three principal components reflect the difference between brands. The combination of PARAFAC and GA provides an accurate method for the rapid identification of unknown samples. The results indicate that PARAFAC-GA-SVM has higher prediction accuracy. The proposed method can effectively extract the spectral characteristics, and also reduce the dimension number of the input variables of SVM. The results can provide a new way for the identification of Chinese liquors.

Key words spectroscopy; parallel factor analysis; genetic algorithm; Chinese liquors

OCIS codes 300.2530; 300.6170; 300.6280

1 引 言

中国白酒有2000多年的历史,其独特的风味受到广大老百姓的喜爱。原料和酿造工艺等的不同,形成

收稿日期: 2014-12-09; 收到修改稿日期: 2015-03-02

基金项目: 国家自然科学基金(61378037)

作者简介: 朱焯炜(1981—),男,博士研究生,主要从事光谱检测技术方面的研究。E-mail: zhuzhuowei2004@163.com

导师简介: 陈国庆(1963—),男,博士,教授,主要从事光谱学和光谱技术等方面的研究。

E-mail: cgq2098@163.com(通信联系人)

了不同香型的白酒。主要香型有浓香、清香、酱香等。即使同种香型白酒,由于产地的环境和原料之间的差异,导致不同品牌白酒中一些微量香气成分及其含量的不同^[1],形成各自独特的酒体风格。知名品牌白酒企业已经逐渐重视对自己白酒品牌的保护,但是也有不法商家为了牟取巨额利益,仿冒知名白酒,不但侵犯了知名白酒企业的知识产权,也侵害了消费者权益。因此,对白酒品牌的鉴定研究,有利于监督白酒生产质量和保护消费者的利益。

目前白酒检测的方法主要有近红外光谱检测、电子鼻检测^[2]、阵列传感器检测^[3]、顶空固相微萃取与质谱联用技术等。这些方法都有一定的优势,但是也有许多不足之处。例如Chen等^[4-5]利用顶空固相微萃取与质谱联用技术进行白酒产地和等级的分类,取得较高的准确率,但是该方法操作复杂,成本较高。Zhao等^[6]利用近红外光谱技术对中国白酒品牌进行鉴别,该方法快速、低成本,但是模型预测精度有待进一步提高。荧光光谱技术由于检测成本低、灵敏度高、对样本无损,得到广大学者的认可^[7-8]。

三维荧光光谱是荧光光谱仪的一种采集方式,体现的是发射波长随着激发波长变化的荧光强度信息,获得的信息量比常规荧光光谱更多^[9]。平行因子分析法(PARAFAC)是利用交替最小二乘算法实现的一种多维数据分解法^[10],能够从多组分混合溶液中得到各个组分准确的定量结果,在三维荧光光谱分析中得到广泛的应用^[11-12]。但是大部分研究都是将训练样本和待测样本放在一起进行平行因子分解以及拟合分析,在实际快速检测应用中有一定的局限性。

本文将样本划分为训练样本和测试样本,采用平行因子获取训练样本的浓度得分,基于遗传算法(GA)得到测试样本的浓度得分。同时结合支持向量机(SVM)建立鉴别模型PARAFAC-GA-SVM,对白酒品牌进行鉴别,取得了较为满意的预测效果。

2 材料与方 法

2.1 仪器设备

实验所用仪器为英国爱丁堡公司的FLS920P全功能型荧光光谱仪。设置仪器的激发与发射狭缝宽度为5 nm,积分时间为0.1 s。激发波长为200~600 nm,步长为5 nm;发射波长为215~800 nm,步长为1 nm。发射波长始终滞后激发波长15 nm扫描的原因是为了消除瑞利散射对实验的影响。

2.2 实验材料

为了保证建立模型的可靠性,所有实验用样本均由厂家提供。包括5种常见香型和8个知名品牌。为了保护厂家的信息,没有列出具体白酒品牌,以香型加上字母表示(如浓香A)。主要品牌包括芝麻A、酱香A、浓香A、浓香B、浓香C、清香A、清香B、豉香A。将整个样品集分成训练集(120个酒样,每个品牌15个)和测试集(40个酒样,每个品牌5个)。

2.3 光谱预处理及数据分析

通过荧光光谱仪扫描得到160个样本的激发-发射矩阵(81×586)。对所有样本的荧光光谱数据进行预处理,采用插值拟合的方法消除水的拉曼峰影响,并将激发-发射矩阵中的缺失值置零。所有光谱数据进行中心化处理,以消除酒精度的影响。运用Matlab2013软件编程实现平行因子算法和遗传算法,支持向量机算法采用LIBSVM工具箱。

2.4 PARAFAC-GA-SVM方法

利用平行因子分析方法对训练样本光谱数据分解,获得载荷矩阵和浓度得分,将浓度得分作为支持向量机的输入,建立白酒品牌的鉴别模型。对于测试样本,保持载荷矩阵不变,利用遗传算法获得测试样本的浓度得分,并将浓度得分输入鉴别模型,输出预测的品牌值。具体建模和鉴别过程如下:

1) 实验测得其三维矩阵数据 $X(I \times J \times M)$ 进行三线性分解。 I 和 J 分别为样品的激发和发射波长扫描个数, M 为训练样本数。将 X 模型表示为^[13]

$$X = \sum_{f=1}^F a_f \otimes b_f \otimes c_f + E, \quad (1)$$

式中 a_f , b_f 和 c_f 分别是平行因子分解得到的载荷矩阵 A 、 B 以及浓度矩阵 C 在第 f 个组分上的元素, \otimes 是克罗内克积, E 是三维矩阵数据 X 的残差矩阵。

2) 对于 N 个测试样品,通过实验获得三维矩阵 \mathbf{X}^{new} 。根据三线性模型,保持载荷矩阵 \mathbf{A} 和 \mathbf{B} 不变;设置遗传算法参数,包括初始种群大小、最大进化代数、交叉比例等。以遗传算法迭代次数达到设定的均方根误差(RMSE)作为收敛终止条件。

$$f_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^N (\sum_{f=1}^F a_{if} b_{jf} c_{nf}^{\text{new}} - x_{ijn})^2}{I \times J \times N}}, \quad (2)$$

式中 x_{ijn} 为矩阵 \mathbf{X}^{new} 中的一个元素, c_{nf}^{new} 为测试样本集中第 n 个样本第 f 个组分对应的预测浓度。

3) 将训练样本浓度阵 \mathbf{C} 作为 SVM 的输入,进行网络模型训练,并达到预计目标。将预测浓度矩阵 \mathbf{C}^{new} 作为支持向量机的输入,得到预测的品牌值。支持向量机的模型表达式为^[14]

$$y(x) = \sum_{i=1}^n a_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) + b, \quad (3)$$

式中 a_i 为拉格朗日乘子, b 为偏差, x_i 为输入向量。选用径向基函数(RBF)作为其核函数, σ 和 c 称为核参数和惩罚因子。

3 结果与讨论

3.1 光谱特性分析

图 1(a)~(h)分别为 8 种品牌白酒的三维荧光等高线图。不同品牌白酒光谱特性既有共性又有个性。从图中可以看出中国白酒都具有较宽的荧光峰位,有效激发波长范围为 300~550 nm,荧光发射波长范围为 300~650 nm,这与其他蒸馏酒相比有较大的差别。同种香型品牌白酒光谱差异较小,不同香型品牌之间差异较大。例如清香 A 和清香 B 两种白酒的荧光光谱等高图较为类似,具有共同的荧光峰位 290/320。三种浓香型白酒有相似的激发和发射范围,具有相似的荧光峰位 355/430。豉香白酒有两个峰位,为 350/450 和 265/284。酱香型白酒与芝麻香型白酒虽然不属于同一香型,但也有类似的等高图谱,具有共同的荧光峰位 350/438。文献[15]报道芝麻香型白酒有一种轻微的酱香和近似焙炒芝麻的香气。因此荧光光谱在一定程度上反映了白酒的香型风格。

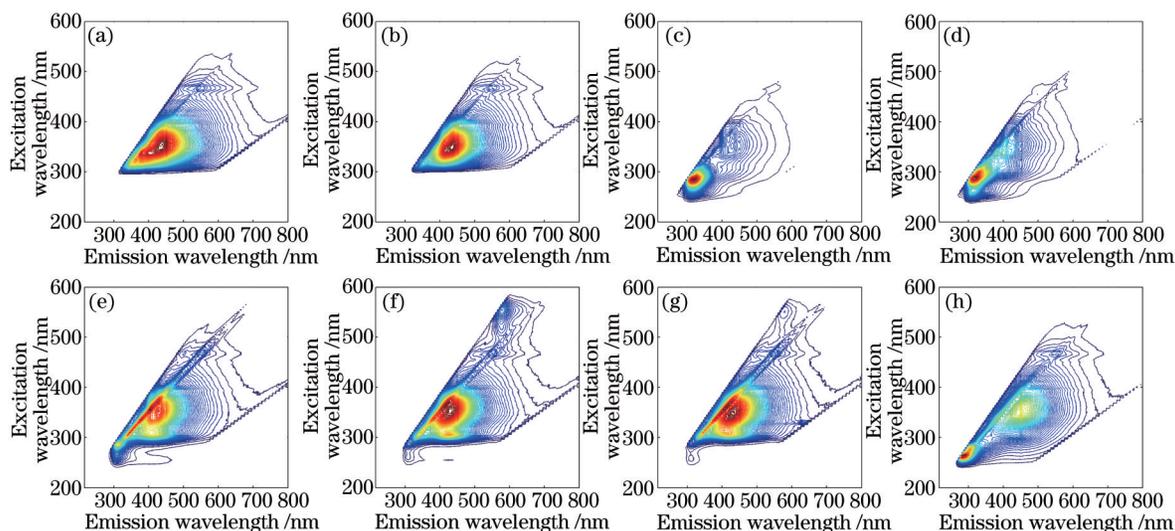


图 1 8 种白酒的三维荧光等高图。(a) 芝麻 A; (b) 酱香 A; (c) 清香 A; (d) 清香 B; (e) 浓香 A; (f) 浓香 B; (g) 浓香 C; (h) 豉香 A
Fig.1 Contour maps of eight groups of liquors with different brands. (a) Zhima A; (b) jiangxiang A; (c) qingxiang A; (d) qingxiang B; (e) nongxiang A; (f) nongxiang B; (g) nongxiang C; (h) chixiang A

通过三维荧光光谱图的分析可以看出每一种白酒都具有其独特的三维荧光光谱特性,这是由于白酒中除了酒精和水外,还有大约 2% 的其他微量成分^[16],不同品牌白酒微量成分及其含量的不同,表现出了光谱特性的差异。因此三维荧光光谱从整体上宏观地反映出每一种白酒的独特品质。从肉眼上判定不同品牌白

酒的差异有一定的难度,需要引入化学计量学的方法来对白酒的品牌进行自动、准确的鉴别。

3.2 平行因子分解

激发发射矩阵按照样本叠加成三维矩阵(81×586×120)。通过核一致诊断法结合拟合残差分析(表1),可以确定当组分数 $F=3$ 时,可释方差为 94.265%,残差(SSR)为 0.855,核一致达到 92.567%,模型为三线性模型,平行因子分解准确性较高。

表1 平行因子组分估计

Table 1 PARAFAC component number estimation

	Number of component F			
	1	2	3	4
Explained variance /%	47.42	88.55	94.2654	96.47
Core consistency /%	100.0	100.0	92.567	36.011
SSR /10 ¹³	16.5	3.6	0.855	0.599

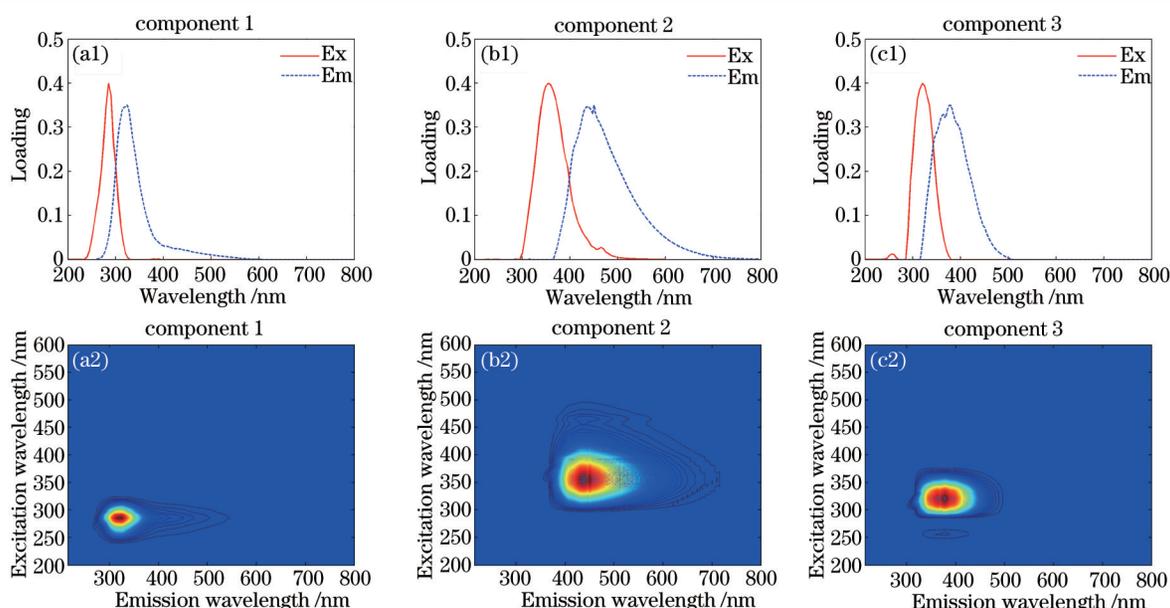


图2 PARAFAC模型鉴定出的白酒的三个荧光组分

Fig.2 Three fluorescent components of Chinese liquors by PARAFAC model

利用PARAFAC方法对三维矩阵进行分析,可以得到三种组分(component 1, component 2, component 3)的激发和发射光谱以及对应的浓度得分。图2是三种组分的激发和发射轮廓图及荧光光谱等高图。从图中可以看出 component 1 激发/发射波长为 285/324; component 2 激发/发射波长为 355/438; component 3 激发/发射波长为 320/380。根据文献[17-19]可以初步判定, component 1 的荧光峰归属为微量成分与乙醇-水形成的团簇分子, component 2 荧光峰归属为酯类物质, component 3 荧光峰归属为酸类物质。具体为哪几种单体成分的综合体现,还有待进一步研究。

借助浓度得分矩阵,可以获得每个样本的浓度得分,如此便可以将不同品牌白酒训练样本的聚类结果可视化,如图3(a)所示。从图中可以看出,不同香型白酒之间比较疏远,同种香型之间较为靠拢。特别是浓香型白酒由于浓度得分较为相似,导致聚合得较为密集。因此需要单独对聚合较密集的浓香型白酒和豉香型白酒进行聚类分析,如图3(b)所示。从图3(b)可以看到豉香A与浓香型白酒得分较开,三种浓香型白酒基本上也能够分开。通过上述分析可以说明浓度得分可以看成样本的特征量,将其作为支持向量机的输入具有一定的可行性。

3.3 基于遗传算法获取测试样本浓度得分

通过平行因子分解得到载荷矩阵 A 和 B , 结合遗传算法获得了40个测试样本对应的三个组分浓度得分(Score 1, Score 2, Score 3)。遗传算法的参数设置:种群规模为100,交叉概率为0.8,变异概率为0.35,设

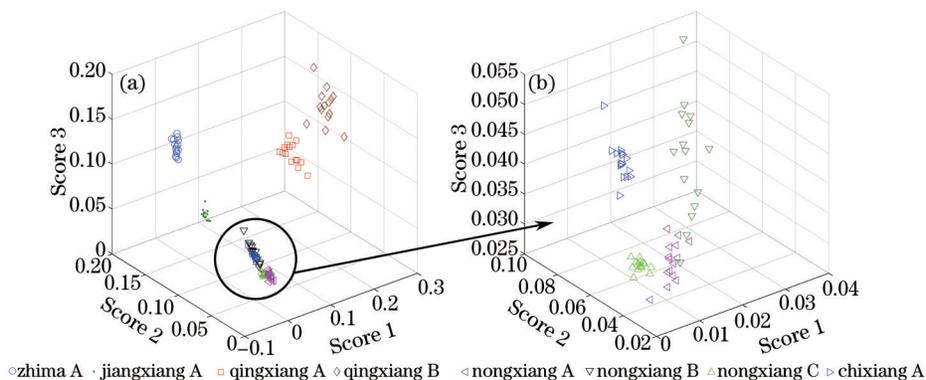


图3 白酒样本的得分分布图

Fig.3 Score plots of Chinese liquors

定最大迭代次数为1600。为了验证通过遗传算法获得的验证样本浓度值的准确性,将其与平行因子分解得到的结果做比较,具体结果如图4所示。横坐标为通过平行因子分解获得值,纵坐标为通过遗传算法获得值。从图中可以看到,两种方法获得的得分值非常接近,相关系数 $R=0.9998$ 。说明通过遗传算法获取浓度得分方法的可行性。

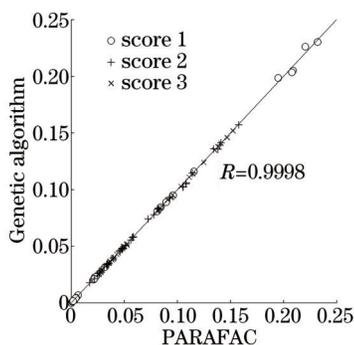


图4 两种方法的比较

Fig.4 Comparison of two methods

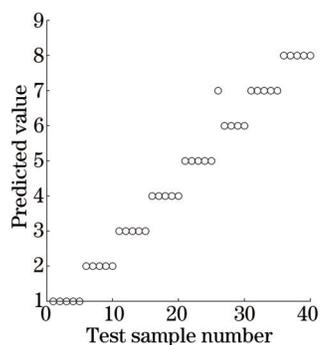


图5 不同品牌白酒的预测结果

Fig.5 Output results of Chinese liquors

3.4 支持向量机鉴别分析

将样本浓度得分作为SVM的输入,建立鉴别模型。采取交叉验证和网格搜索算法寻找最优惩罚因子 c 和核参数 σ 参变量。将白酒的品牌作为模型的输出值,设定虚拟变量1~8代表不同的品牌,其中1为芝麻A、2为酱香A、3为清香A、4为清香B、5为浓香A、6为浓香B、7为浓香C、8为豉香A。由以上参数构建的SVM模型对测试集样本的输出结果如图5所示。从图中可以看出测试集里除了第26个样本误判为浓香C外,其余都判断正确,鉴别准确率达97.5%。作为比较,将训练样本和测试样本放在一起进行PARAFAC分解,结合SVM进行品牌预测,准确率也为97.5%,但是整个分解测试过程所需时间是58.8 s。而利用PARAFAC-GA-SVM方法需时仅为29.4 s。对大批量样本需要进行品牌预测的情况下,PARAFAC-GA-SVM方法具有一定优势。

为了检验PARAFAC-GA-SVM建模方法对同一香型不同品牌白酒的预测能力,对3种浓香型白酒采用留一法进行交叉验证,即每次留一个样本对由剩余样本建立的模型进行验证,此过程重复60次,模型预测能力为60个预测准确率的平均值。该研究构建的判别模型交叉验证的预测准确率达96.7%,能够正确预测白酒样品的品牌。造成模型鉴别错误的主要原因是用于构建模型的训练集酒样少。因此,若要提高模型的预测准确性,增加样本数是必不可少的过程。

4 结 论

探索建立了一种中国白酒品牌快速鉴别的方法。利用三维荧光光谱结合平行因子和遗传算法分别获得训练样本和测试样本的浓度得分,并将其作为支持向量机的输入,建立中国白酒品牌鉴别的数学模型,预

测准确率达到97.5%。PARAFAC分解方法能有效提取白酒的特征光谱信息,同时也降低了SVM输入变量的维数。遗传算法的引入,增强了鉴别模型的实用性。这一方法也为多组分混合溶液的快速定量分析提供了新的思路。

参考文献

- 1 Shen Yifang. Chinese liquor sensory quality and evaluation of the history and development of technology[J]. *Liquor Making*, 2006, 33(4): 3-4.
沈怡方. 中国白酒感官品质及品评技术历史与发展[J]. *酿酒*, 2006, 33(4): 3-4.
- 2 Liu M, Han X, Tu K, *et al.*. Application of electronic nose in Chinese spirits quality control and flavour assessment[J]. *Food Control*, 2012, 26(2): 564-570.
- 3 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.
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*, 2014, 39(4): 516-520.
- 4 Cheng P, Fan W, Xu Y. Determination of Chinese liquors from different geographic origins by combination of mass spectrometry and chemometric technique[J]. *Food Control*, 2014, 35(1): 153-158.
- 5 Cheng P, Fan W, Xu Y. Quality grade discrimination of Chinese strong aroma type liquors using mass spectrometry and multivariate analysis[J]. *Food Research International*, 2013, 54(2): 1753-1760.
- 6 Zhao L, Wang P, Huang C, *et al.*. Application of Vis/NIR spectroscopy for Chinese liquor discrimination[J]. *Food Analytical Methods*, 2014, 7(6): 1337-1344.
- 7 Wang Yutian, Xu Jing, Zhou Yanling. Study on mineral oil identification based on a dimension reduction method of three dimensional fluorescence spectra[J]. *Chinese J Lasers*, 2014, 41(12): 1215002.
王玉田, 徐靖, 周艳岭. 基于三维荧光光谱降维处理的矿物油识别研究[J]. *中国激光*, 2014, 41(12): 1215002.
- 8 Liu Jing, Liu Wenqing, Zhao Nanjing, *et al.*. Phytoplankton chlorophyll fluorescence characteristics excited by various light qualities and intensities[J]. *Acta Optica Sinica*, 2013, 33(9): 0930001.
刘晶, 刘文清, 赵南京, 等. 浮游植物在不同光质和光强激发下的叶绿素荧光特性[J]. *光学学报*, 2013, 33(9): 0930001.
- 9 Xiao Haibin, Zhao Jinhui, Liu Muhua, *et al.*. Application study of three-dimensional fluorescence spectroscopy combined with PLSR for carbaryl residues in duck meat[J]. *Laser & Optoelectronics Progress*, 2014, 51(4): 043001.
肖海斌, 赵进辉, 刘木华, 等. 三维荧光光谱技术结合PLSR在鸭肉中西维因残留检测的应用[J]. *激光与光电子学进展*, 2014, 51(4): 043001.
- 10 Bro R, Kiers H A L. A new efficient method for determining the number of components in PARAFAC models[J]. *Journal of Chemometrics*, 2003, 17(5): 274-286.
- 11 Liu Yanfang, Su Rongguo, Zhou Qianqian, *et al.*. Rapid modeling offshore eutrophication technique using optical parameters of CDOM[J]. *Chinese J Lasers*, 2014, 41(12): 1215001.
刘艳芳, 苏荣国, 周倩倩, 等. 基于CDOM光学参数的近海富营养化快速评价技术[J]. *中国激光*, 2014, 41(12): 1215001.
- 12 Yang Lili, Wang Yutian, Lu Xinqiong. Identification and measurement of petroleum pollutant by three-dimensional matrix fluorescence with second-order calibration methods[J]. *Chinese J Lasers*, 2013, 40(6): 0615002.
杨丽丽, 王玉田, 鲁信琼. 三维荧光光谱结合二阶校正法用于石油类污染物的识别和检测[J]. *中国激光*, 2013, 40(6): 0615002.
- 13 Bro R. PARAFAC: Tutorial and applications[J]. *Chemometrics and Intelligent Laboratory Systems*, 1997, 38(2): 149-171.
- 14 Belousov A I, Verzakov S A, Von F J. A flexible classification approach with optimal generalization performance: Support vector machines[J]. *Chemometrics and Intelligent Laboratory Systems*, 2002, 64(1): 15-25.
- 15 Wu Sanduo. Research on correlations and trace components of five flavor types of liquor[J]. *Liquor-Making Science and Technology*, 2001, (4): 82-85.
吴三多. 五大香型白酒的相互关系与微量成分浅析[J]. *酿酒科技*, 2001, (4): 82-85.
- 16 Li Weiqing. Flavor and flavor styles of liquor[J]. *Liquor Making*, 2007, 34(2): 5-7.
李维青. 白酒的香气与香型[J]. *酿酒*, 2007, 34(2): 5-7.
- 17 Qiao H, Zhang W, Wang W. Fluorescence spectroscopic and viscosity studies of hydrogen bonding in Chinese Fenjiu[J]. *Journal of Bioscience and Bioengineering*, 2013, 115(4): 405-411.
- 18 Yang Jianlei, Zhu Tuo, Xu Yan, *et al.*. Study on ultraviolet fluorescence spectra of monomers of distilled spirits[J]. *Spectroscopy and Spectral Analysis*, 2009, 29(12): 3339-3343.
杨建磊, 朱拓, 徐岩, 等. 白酒单体物质紫外荧光光谱研究[J]. *光谱学与光谱析*, 2009, 29(12): 3339-3343.
- 19 Zheng Lei. Study on Fluorescence Spectrum about Several Kinds of Wine and Liquor Monomer Material[D]. Wuxi: Jiangnan University, 2011: 19-20.
郑磊. 若干种白酒及白酒单体物质的荧光指纹图谱研究[D]. 无锡: 江南大学, 2011: 19-20.