

doi:10.3788/gzxb20134201.0069

# 基于支持向量机的假酒近红外光谱识别分类研究

谭琨<sup>1</sup>, 叶元元<sup>1</sup>, 杜培军<sup>2</sup>

(1 江苏省资源环境信息工程重点实验室(中国矿业大学), 江苏 徐州 221116)

(2 南京大学 地理信息科学系, 南京 210093)

**摘 要:**一般的检测假酒中甲醇的化学方法虽然结果较准确,但操作复杂、费用昂贵且对实验的环境条件要求严格.为此,提出了一种基于支持向量机对掺甲醇的假酒光谱进行识别与分类的方法.采用 ASD FieldSpec 3 光谱仪测量了样品溶液的反射光谱;通过对反射光谱进行平滑、导数等预处理并进行相关性分析和单变量回归分析,得出假酒中甲醇光谱不被乙醇光谱掩盖的特征峰作为特征谱带;最后用特征谱带训练分类模型并得到分类结果.结果表明:以甲醇含量小于等于 3% 为真酒的总体分类准确率为 85%,以甲醇含量小于等于 5% 为真酒的总体分类准确率为 97.5%;证明该方法是可行的且具有较高的分类准确度.

**关键词:**支持向量机;相关性分析;单变量回归分析;假酒

中图分类号:O433.4

文献标识码:A

文章编号:1004-4213(2013)01-0069-5

## Identification and Classification of Near-infrared Spectrum of Adulterated Wine Based on Support Vector Machine

TAN Kun<sup>1</sup>, YE Yuan-yuan<sup>1</sup>, DU Pei-jun<sup>2</sup>

(1 Jiangsu Key Laboratory of Resources and Environment Information Engineering  
(China University of Mining and Technology), Xuzhou, Jiangsu 221116, China)

(2 Department of Geographical Information Science, Nanjing University, Nanjing 210093, China)

**Abstract:** Although usual chemical methods have more accurate results in detecting the methanol in adulterated wine, they are complex, expensive and requiring rigorous environment condition. A novel identified and classified spectrum of adulterated wine was proposed based on the support vector machine. The spectra of samples were measured by the ASD FieldSpec 3 spectrometer; reflection spectra were pretreated and correlation analysis and univariate regression analysis were carried out, so the peaks of methanol spectra as the characteristic bands which is not over shadowed by the ethanol were obtained; the characteristic bands were used to train classification model, the result was obtained. The result shows that, the classification accuracy is 85% while the content of methanol is less than or equal to 3% as the true wine, and the classification accuracy is 97.5% while the content of methanol is less than or equal to 5% as the true wine. So, this method is available and has higher classification accuracy.

**Key words:** Support vector machine; Correlation analysis; Univariate regression analysis; Adulterated wine

## 0 引言

在检测假酒中甲醇含量的方法里,一般的气相色谱法及品红-亚硫酸比色法检测较精确,可用于食

品卫生检测.但是实验的费用昂贵、过程复杂且对环境条件要求严格<sup>[1]</sup>,故不能得到广泛的应用和实现快速识别假酒的功能.

近些年发展起来的光谱分析技术具有操作简单

基金项目:国家自然科学基金(No. 41101423)、中央高校基本科研业务费专项资金(No. 2010QNA18)、中国博士后科学基金(Nos. 2011M500128, 2012T50499)和江苏高校优势学科建设工程项目资助

第一作者:谭琨(1981-),男,博士,主要研究方向为高光谱遥感、环境遥感、模式识别. Email: tankun@cumt.edu.cn

收稿日期:2012-08-24;录用日期:2012-10-16

方便,准确度高,快速稳定等优点,并在化工、农业和医学等领域得到了广泛应用<sup>[2-4]</sup>.就食品类的安全检测领域,国内外先后有学者将光谱分析技术与反向传播(Back Propagation, BP)神经网络、偏最小二乘法(Partial Least Squares, PLS)等算法结合起来应用于食用油、牛奶、蜂蜜、苹果汁、烘焙咖啡、酒精饮料等的掺杂掺假检测与分类的研究<sup>[5-10]</sup>.但是将支持向量机(Support Vector Machine, SVM)分类方法与光谱分析技术结合起来应用于酒类掺假快速检测分类还没有分析研究.

传统的图像分类法存在计算量过大、分类准确率低、推广能力差等缺点<sup>[11-12]</sup>,而 SVM 是 20 世纪 90 年代形成的一种新的模式识别方法,这一基于统计学习理论结构风险最小化准则的信息处理方法<sup>[13-14]</sup>,已表现出许多优于其它模式识别方法的性能<sup>[15-17]</sup>.鉴于此,本文提出一种新的基于 SVM 的假酒光谱识别分类的研究方法.

## 1 材料与方法

### 1.1 实验仪器

本实验采用美国 ASD 公司生产的 FieldSpec 3 光谱仪在暗室中采集样品光谱,仪器的光谱覆盖范围为 350~2 500 nm. 350~1 000 nm 区间的采样间隔为 1.4 nm,光谱分辨率为 3 nm; 1000~2 500 nm 区间的采样间隔为 2 nm,光谱分辨率为 10 nm.

### 1.2 实验样品

根据实验需要,首先配置乙醇含量为 38% 的水溶液作为实验酒,然后往实验酒中分别掺比例为 1%、3%、5%、10%、15%、20%、25% 的甲醇,每种比例配置五个样品,外加 5 个真酒样品,共 40 个样品溶液.

### 1.3 光谱采集

在暗室内,架好 ASD 光谱仪,试验中使用 ASD 公司配置的专用卤素灯,将装有样品溶液的石英玻

璃皿放置在黑色天鹅绒中心,调整好光源与样品溶液表面之间的距离以及几何方位角,并在实验过程中保持不变.每个样品的光谱反射率采集十次.每个样品的十条光谱平均后作为该样品溶液的最终反射光谱.

## 2 结果与讨论

### 2.1 光谱预处理

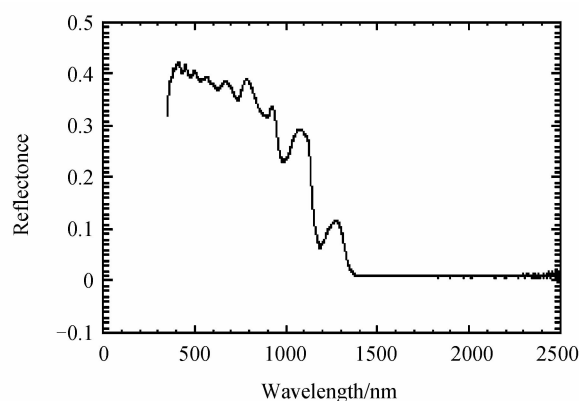
FieldSpec 3 光谱仪所采集的光谱信息,除了包含样品的自身信息外,还包含了其它无关信息和噪音.因此,在建立模型及分类前消除光谱数据中的无关信息和噪音的预处理过程变得十分必要和关键.光谱分析技术中常用的预处理方法有断点修正、平滑、导数方法、连续统去除、标准正态变量变换、多元散射校正等<sup>[18]</sup>.

为了保证后续模型建立及分类的准确度要求,本实验对原始光谱进行了平滑、断点修正、导数方法、连续统去除等预处理.部分预处理效果如图 1.

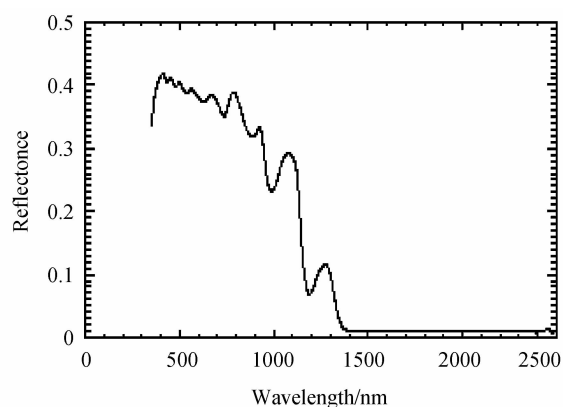
### 2.2 相关性分析

1)对 350~2 500 nm 波长区间内的平滑和断点修正后光谱反射率与甲醇含量进行相关性分析,可获得它们的相关关系随波长变化图(图 2(a)).通过观察原始光谱,首尾的光谱特征不明显,故本实验选取 900~1 500 nm 之间的光谱作为分析区间.甲醇含量与样本反射率在此区间呈正相关.表明随着甲醇浓度的增加时,在 900~1 500 nm 之间光谱反射率强度亦增加.波长 935 nm 处相关性为 0.82,达到最大相关.

2)对 350~2 500 nm 波长区间内一阶导数光谱反射率与甲醇含量进行同样的相关性分析,可获得它们的相关关系随波长变化图(图 2(b)).在 900~1 500 nm 波长区间内,波长 1 270 nm 处相关性为 0.97,达到最大相关.



(a) Original spectrum



(b) Smooth and spline correction

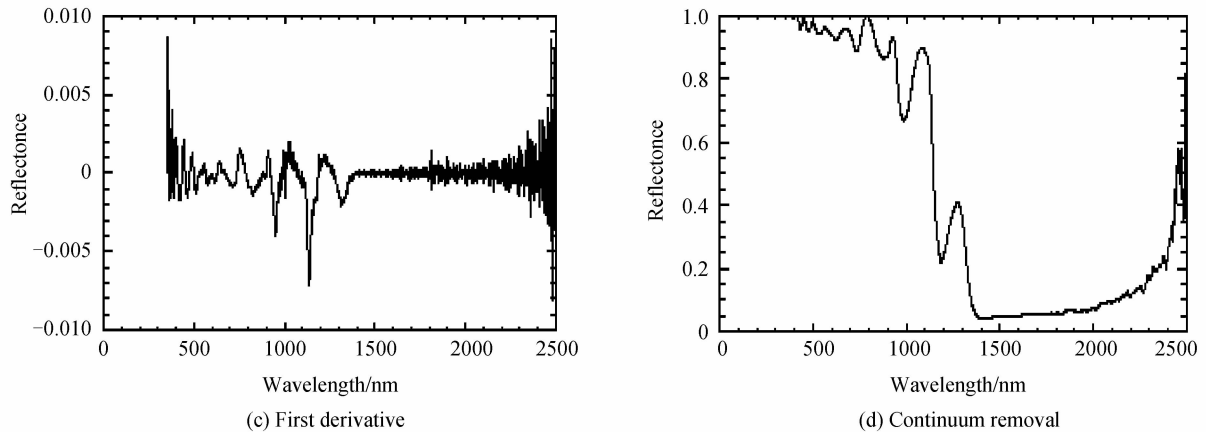


图 1 原始光谱预处理效果图  
Fig. 1 The preprocessing of original spectrum

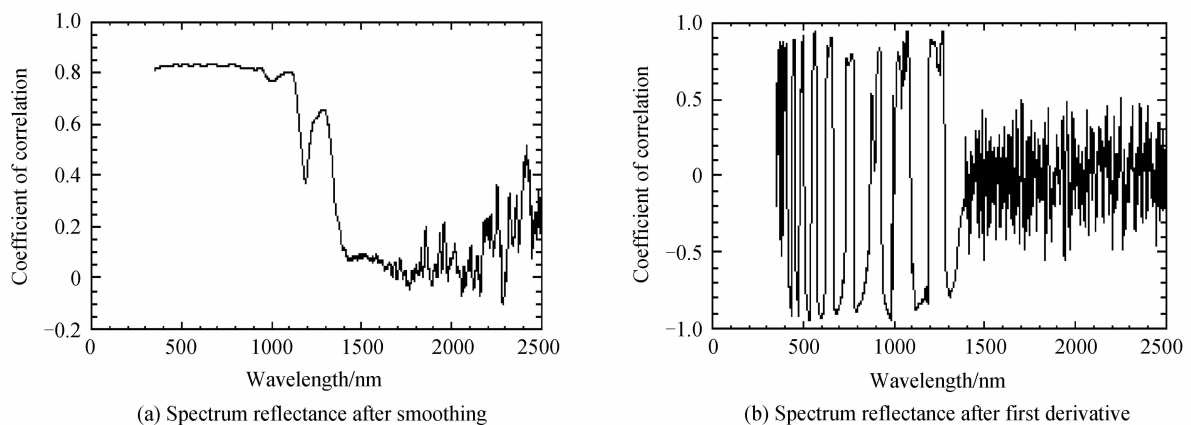


图 2 甲醇含量与平滑、一阶导数后光谱反射率相关系数随波长变化图  
Fig. 2 The correlation coefficient of methanol content and the spectrum reflectance after smoothing and first derivative with the wavelength varies

### 2.3 单变量回归分析

以甲醇含量与光谱反射率最大相关波长处的反射率为回归方程的自变量  $X$ 、甲醇含量为因变量  $Y$  进行简单回归拟合分析. 分别对平滑光谱和一阶导

数光谱进行回归分析,其回归方程如表 1 所示,此处由于甲醇含量与连续统去除光谱相关性较差而没有进行回归分析.

表 1 光谱与甲醇含量的单变量回归方程

Spectrum	Wavelength/nm	Correlation coefficient	Regression equation	$R^2$
Smooth	935	0.82	$Y = 1.0897X + 0.296$	0.68
First derivative	1270	0.97	$Y = 0.0021X - 0.000007$	0.90

图 3(a)、(c) 分别表示平滑光谱和一阶导数光谱的最大相关波长处光谱反射率与甲醇含量的回归模型. 图 3(b)、(d) 分别表示平滑光谱和一阶导数光谱的相应回归模型预测的残差分布图. 实验结果表明一阶导数光谱最大相关波长处的反射率与甲醇含量的回归模型准确度更高.

### 2.4 SVM 分类

根据分析结果,得到特征峰在 1 270 nm. 因而以其为中心向两边扩展,选取 1 245~1 294 nm 作为特征波段进行分析,能很好地表示特征. 建模时候选取的 1 270 nm 波长附近 (1 245~1 294 nm) 的一

阶导数光谱值作为光谱特征训练分类模型,在均匀选取 11 个训练样品的基础上对测试样本进行 SVM 分类,样本选择时候考虑不同的甲醇含量(例如,在 0%、1%、3%、5%、10%、15%、20%、25% 各类别中各选一个做训练样本,由于 0%、1%、3%,差别较小,故在这三个类别中各取两个训练样本,共 11 个). 分类中选择径向基核函数,对于核函数的参量选择应用 Grid 搜索方法. 分类结果如表 2 所示,总体分类准确度及 Kappa 系数见表 3. 由于 SVM 分类器的良好的小样本学习能力,所以本研究得到了较好的识别结果,能够进行假酒的初步预警.

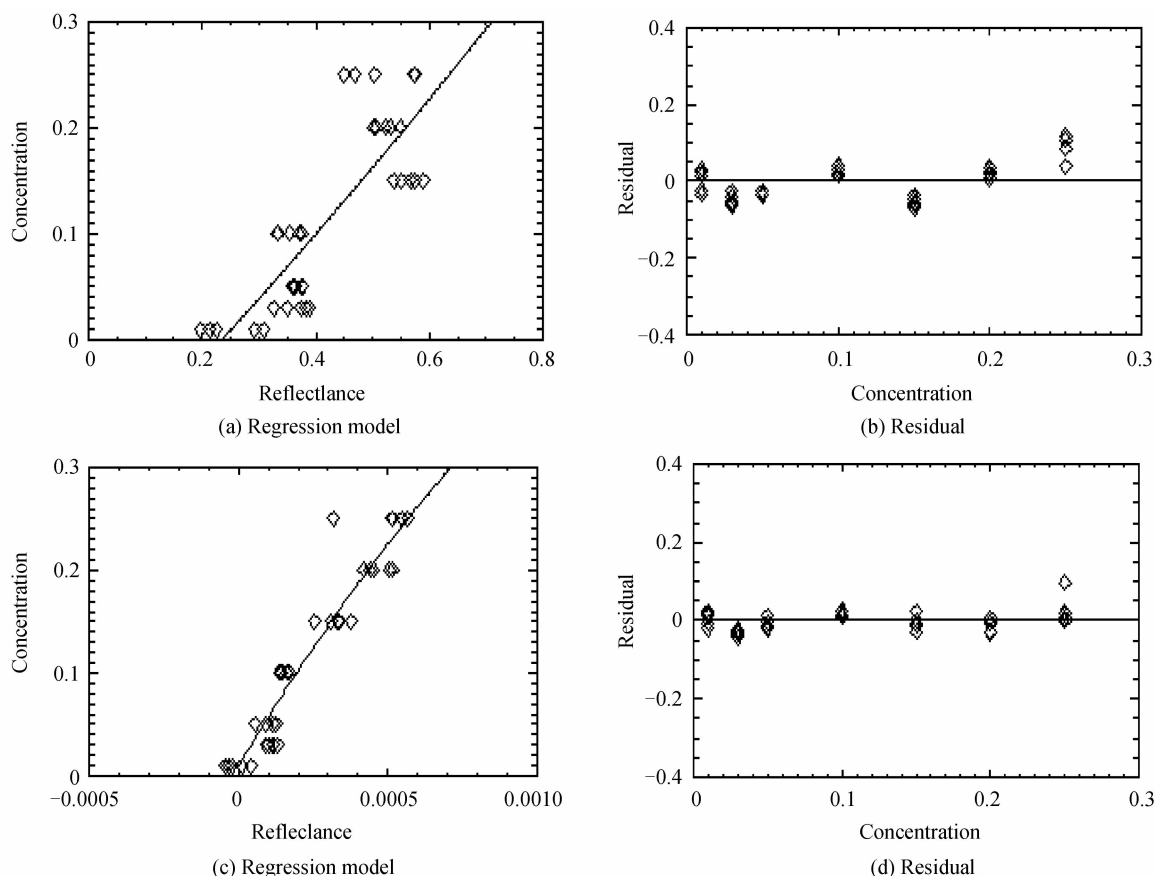


图3 光谱与甲醇含量的单变量回归分析比较

Fig. 3 The univariate regression analysis of spectrum and the methanol content

表2 两个不同基准的分类结果

Table 2 The classification results of two criterion

Criterion of classification	Classification results
Content of methanol $\leq 3\%$ is the true wine	4 of true be divided into false, 2 of false be divided into true; the rest of the classification results are true
Content of methanol $\leq 5\%$ is the true wine	1 of false be divided into true; the rest of the classification results are true

表3 两个不同基准的分类准确度

Table 3 The classification accuracy of two criterion

Criterion of classification	Classification accuracy	Kappa coefficient
Content of methanol $\leq 3\%$ is the true wine	85%	0.67
Content of methanol $\leq 5\%$ is the true wine	97.5%	0.95

### 3 结论

本文分析了基于 SVM 分类方法在可见-近红外波段定性检测假酒中甲醇含量并对其进行分类的可行性. 通过对 FieldSpec 3 光谱仪采集的光谱数据进行预处理、相关性和单变量回归分析, 建立了基于 SVM 判别酒的真伪的分类方法并获得了分类准确度. 实验结果表明, 以甲醇含量小于等于 3% 为真酒的总体分类准确度为 85%, kappa 系数为 0.67; 以甲醇含量小于等于 5% 为真酒的总体分类准确度为 97.5%, kappa 系数为 0.95. 所以基于 SVM 分类方法在可见-近红外波段定性检测假酒中甲醇含量并

对其进行分类是可行的, 且具有简单、快速、费用低的优势, 这些优势可以为食品的质量控制检测工作带来很大的便利.

#### 参考文献

- [1] YAO Jie, YANG Qian, SUN Cai-yun, *et al.* Qualitative analysis of methanol in the forged wine using infrared spectrophotometer [J]. *Chinese Journal of Spectroscopy Laboratory*, 2000, **17**(1): 35-37.  
姚杰, 杨倩, 孙彩云, 等. 红外光谱法定性分析假酒中的甲醇 [J]. *光谱实验室*, 2000, **17**(1): 35-37.
- [2] XU Zheng-hong, ZHANG Zhen-xia, WANG Jing, *et al.* Spectral characteristics of voltage sensitive dye di-4-ANEPPS in perfusates [J]. *Acta Photonica Sinica*, 2008, **37**(3): 552-554.  
徐正红, 张镇西, 王晶, 等. 电压敏感染料 di-4-ANEPPS 在灌注液中的光谱特性 [J]. *光子学报*, 2008, **37**(3): 552-554.

- [3] WANG Zhi-ping, YANG Jian-feng, XUE Bin, *et al.* Using hyperspectral data for land cover identification and classification[J]. *Acta Photonica Sinica*, 2008, **37**(3): 561-565.  
王志平, 杨建峰, 薛彬, 等. 利用高光谱数据进行地物识别分类研究[J]. *光子学报*, 2008, **37**(3): 561-565.
- [4] WANG Le-xin, ZHAO Zhi-min, YAO Hong-bing, *et al.* Infrared absorption spectrum analysis and its application to blood[J]. *Spectroscopy and Spectral Analysis*, 2002, **22**(6): 980-982.  
王乐新, 赵志敏, 姚红兵, 等. 血液的红外吸收光谱分析及应用研究[J]. *光谱学与光谱分析*, 2002, **22**(6): 980-982.
- [5] WENG Xin-xin, LU Feng, WANG Chuan-xian, *et al.* Discriminating and quantifying potential adulteration in virgin olive oil by near-infrared spectroscopy with BP-ANN and PLS [J]. *Spectroscopy and Spectral Analysis*, 2009, **29**(12): 3283-3287.  
翁欣欣, 陆锋, 王传现, 等. 近红外光谱-BP神经网络-PLS法用于橄榄油掺杂分析[J]. *光谱学与光谱分析*, 2009, **29**(12): 3283-3287.
- [6] HAN Dong-hai, LU Chao, LIU Yi, *et al.* Identification of reconstructed milk in raw milk using near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(3): 465-468.  
韩东海, 鲁超, 刘毅, 等. 生鲜乳中还原乳的近红外光谱法鉴别[J]. *光谱学与光谱分析*, 2007, **27**(3): 465-468.
- [7] CHEN Lan-zhen, ZHAO Jing, YE Zhi-hua, *et al.* Determination of adulteration in honey using near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(11): 2565-2568.  
陈兰珍, 赵静, 叶志华, 等. 蜂蜜真伪的近红外光谱鉴别研究[J]. *光谱学与光谱分析*, 2008, **28**(11): 2565-2568.
- [8] LEON L, KELLY J D, DOWNEY G. Detection of apple juice adulteration using near-infrared transmittance spectroscopy [J]. *Applied Spectroscopy*, 2005, **59**(5): 593-599.
- [9] PIZARRO C I, ESTEBAN-DIEZ I, GONZALEZ-SAIZ J M. Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near-infrared spectroscopy[J]. *Analytical Chemical Acta*, 2007, **585**(2): 266-276.
- [10] PONTES M J C, SANTOS R B, ARAJO M C U, *et al.* Classification of distilled alcoholic beverages and verification of adulteration by near-infrared spectrometry [J]. *Food Research International*, 2006, **39**(2): 182-189.
- [11] LUO Jian-cheng, ZHOU Cheng-hu, YANG Yan. ANN remote sensing classification model and its integration approach with Geo-knowledge [J]. *Journal of Remote Sensing*, 2001, **5**(2): 122-129.  
骆剑承, 周成虎, 杨艳. 神经网络遥感影像分类模型及其与知识集成方法研究[J]. *遥感学报*, 2001, **5**(2): 122-129.
- [12] MA Huai-zhi, WU Qing-xiao, HAO Ying-ming. Steelmaking materials recognition based on lifting wavelet transform and SVM classification[J]. *Computer Engineering Design*, 2010, **31**(18): 4093-4096.  
马怀志, 吴清潇, 郝颖明. 基于提升小波和SVM分类的炼钢物料识别[J]. *计算机工程与设计*, 2010, **31**(18): 4093-4096.
- [13] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer, 1995.
- [14] BAI Peng, LI Yan, ZHANG Bin, *et al.* Key technologies research of mixture gas infrared spectrum analysis based on SVM[J]. *Acta Photonica Sinica*, 2008, **37**(3): 566-572.  
白鹏, 李彦, 张斌, 等. 基于SVM的混合气体红外光谱分析关键技术研究[J]. *光子学报*, 2008, **37**(3): 566-572.
- [15] TAN Kun, DU Pei-jun. Wavelet support vector machines based on reproducing kernel hilbert space for hyperspectral remote sensing image classification [J]. *Acta Geodaetica et Cartographica Sinica*, 2011, **40**(2): 142-147.  
谭琨, 杜培军. 基于再生核 Hilbert 空间小波核函数支持向量机的高光谱遥感影像分类[J]. *测绘学报*, 2011, **40**(2): 142-147.
- [16] 杜培军, 谭琨, 夏俊士. 高光谱遥感影像分类与支持向量机应用研究[M]. 北京: 科学出版社, 2012.
- [17] CHEN Quan-sheng, ZHAO Jie-wen, ZHANG Hai-dong, *et al.* Identification of authenticity of tea with near-infrared spectroscopy based on support vector machine [J]. *Acta Optica Sinica*, 2006, **26**(6): 933-937.  
陈全胜, 赵杰文, 张海东, 等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. *光学学报*, 2006, **26**(6): 933-937.
- [18] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen, *et al.* Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, **16**(4): 528-542.  
褚小立, 袁洪福, 陆婉珍, 等. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, **16**(4): 528-542.