

# 基于拉曼光谱和支持向量机的黄酒品质快速分析

王维琴<sup>1</sup>, 汪丽<sup>2</sup>, 于海燕<sup>2</sup>

(1. 上海宏海食品有限公司, 上海 201611) (2. 上海应用技术学院香料香精技术与工程学院, 上海 201418)

**摘要:** 本研究使用拉曼光谱分析技术采集不同产地和不同酒龄的黄酒样品指纹信息, 对比判别分析 (DA) 和最小二乘支持向量机 (LS-SVM) 所建黄酒品质快速模型性能, 确定最优模型以实现快速准确地评价黄酒品质。本研究在全波段范围利用主成分分析对拉曼光谱数据降维, 计算降维谱图间马氏距离, 基于ward's算法建立判别分析模型; 将全波段范围作为最小二乘支持向量机的输入量, 选择能较好处理非线性问题的RBF为核函数, 同时采用交互验证方式优化RBF核函数参数, 基于优化RBF核函数, 建立最小二乘支持向量机鉴别模型。研究表明: 拉曼光谱结合最小二乘支持向量机鉴别模型对黄酒产地和酒龄的鉴别正确率均为100%; 拉曼光谱结合判别分析鉴别模型对嘉善、绍兴和上海黄酒的鉴别正确率分别为100%、80%和80%, 对黄酒酒龄的鉴别正确率均为100%; 最小二乘支持向量机模型性能优于判别分析模型。拉曼光谱结合化学计量学方法可快速、准确评价黄酒品质。

**关键词:** 拉曼光谱; 支持向量机; 判别分析; 黄酒; 产地; 酒龄

文章编号: 1673-9078(2015)3-255-259

DOI: 10.13982/j.mfst.1673-9078.2015.3.042

## Rapid Quality Analysis of Chinese Rice Wine by Raman Spectroscopy and Support Vector Machines

WANG Wei-qin<sup>1</sup>, WANG Li<sup>2</sup>, YU Hai-yan<sup>2</sup>

(1. Shanghai Honghai Foodstuffs Co., Ltd, Shanghai 201611, China)

(2. School of Perfume and Aroma Technology, Shanghai Institute of Technology, Shanghai 201418, China)

**Abstract:** In this study, Raman spectra analysis technology was used to rapidly and accurately evaluate various Chinese rice wines. Fingerprint information was collected from Chinese rice wine samples from different sources and of different ages, followed by a comparison of the prediction performances of quality evaluation models established by discriminant analysis (DA) and least-squares support vector machine (LS-SVM) in order to select the optimal model. Principal component analysis was used in the entire wavelength range to reduce the dimension of Raman spectroscopy data. The Mahalanobis distance between reduced-dimension spectra was calculated and discrimination analysis models were established based on ward's calculation. The entire wavelength range was used as the input of LS-SVM, to select a kernel radial basis function (RBF) that could alleviate non-linear issues. The kernel RBF was optimized at the same time via interactive authentication. The LS-SVM discrimination model was established based on the optimized kernel RBF function. The result showed that Raman spectroscopy coupled with LS-SVM could 100% accurately identify both production origin and wine age. The correction rates of Raman spectroscopy coupled with DA to discriminate Chinese rice wines from among Jiashan, Shaoxing, and Shanghai origins were 100%, 80%, and 80%, respectively, while that to determine wine age was 100% for all samples. Thus, the LX-SVM model was superior to the DA model. It was concluded that Raman spectroscopy in combination with chemometrics could be used for the rapid and accurate evaluation of Chinese rice wines.

**Key words:** Raman spectroscopy; support vector machine; discriminant analysis; Chinese rice wine; production origin; wine age

黄酒, 是世界上最古老的酒类之一, 它源于中国

收稿日期: 2014-07-22

基金项目: 国家自然科学基金资助项目 (21105065); 全国优秀博士学位论文作者专项资金资助项目 (201059)

作者简介: 王维琴 (1979-), 女, 硕士, 工程师, 研究方向: 食品品质快速检测

通讯作者: 于海燕 (1979-), 女, 博士, 副教授, 研究方向: 食品风味分析和食品品质快速检测

绍兴, 是我国的民族特产。黄酒具有显著的商业价值和社会重要性, 也是最有希望走向世界并占有一席之地的酒品。黄酒是以大米、黍米为原料, 通过酒曲及酒药等共同作用而酿成的低度发酵原酒<sup>[1]</sup>。黄酒的品质主要取决于原料, 之相适应的加工工艺和陈酿时间。由于原料、生产工艺等条件的不同, 所酿造出来的黄酒也就具有不同的风味, 因而, 黄酒是一种明显带有地理标志特征的产品。目前, 绍兴黄酒、嘉善黄酒获

准使用地理标志。通过贮存使黄酒老熟的过程通常称为陈酿。经过一段时间的陈酿,酒体丰满协调,酒质芳香醇和。为了有效保护黄酒,也为了保护消费者权益和厂家的信誉,快速、准确、无损地评价黄酒的品质显得尤为重要。

传统上是通过部分化学成分的测定来分析黄酒品质,如今已发展到运用高效、灵敏、样品用量少的现代仪器来定性、定量全面分析黄酒的品质阶段<sup>[2]</sup>。近几年,近红外光谱作为一种快速分析技术应用于黄酒产地和酒龄鉴别的研究正在开展<sup>[3]</sup>。然而,水在红外光谱区吸收强度大,为有效信息提取带来干扰<sup>[4]</sup>。拉曼光谱分析技术是分析分子的振动或转动信息的光谱技术<sup>[5]</sup>。水在拉曼光谱区没有伸缩振动频率带,因而拉曼光谱分析技术适用于水溶液体系的分析<sup>[4]</sup>。拉曼光谱分析技术已应用于乳制品<sup>[6]</sup>和饮料酒<sup>[7]</sup>等的分析,并取得较好结果。迄今未检索到利用拉曼光谱鉴别黄酒产地和酒龄的研究报道。

本文采用拉曼光谱分析技术采集不同产地和不同酒龄黄酒样品指纹信息,采用判别分析(DA)和最小二乘支持向量机(LS-SVM)对拉曼光谱指纹信息进行分类建模,对比两种方法所建黄酒产地和酒龄鉴别模型性能,以实现黄酒产地和酒龄的快速鉴别。该方法快速、准确、无需样品前处理,可为黄酒快速甄别提供新的分析技术。

## 1 材料与方法

### 1.1 材料

不同产地黄酒分别源自浙江绍兴、浙江嘉善和上海某一黄酒厂的5年陈加饭酒,其中绍兴黄酒5瓶,嘉善黄酒3瓶,上海黄酒5瓶;不同酒龄黄酒源自浙江绍兴某一酒厂的1年陈、3年陈、5年陈和8年陈加饭酒,每个年份各5瓶样品,取自不同批次基酒。

### 1.2 仪器

采用LabRAM HR Evolution拉曼光谱仪(HORIBA Jobin Yvon, 法国),使用倒置显微镜。技术参数分别为:光谱范围:100~3050  $\text{cm}^{-1}$ ;焦距:800 mm;激光功率控制:多级激光功率衰减片;探测器:研究级大芯片尺寸空冷CCD。

### 1.3 光谱采集

室温条件下,在5 mm光程比色皿中采集拉曼光谱。试验参数为:光栅600  $\text{g/mm}$ 、采集时间5 s、累加次数9、使用50x LWD、镜头采集、波长范围

100~3050  $\text{cm}^{-1}$ 。

## 1.4 数据处理方法

判别分析是在有先验知识的前提下,即知道各样品所属类别的情况下,对原始数据向量进行线性变换,使得各类样品能够更好的区分。本研究在全波段范围利用主成分分析对拉曼光谱数据降维,计算降维谱图间马氏距离,基于ward's算法建立判别分析模型;最小二乘支持向量机(Least-squares support vector machine, LS-SVM)是由Suykens等<sup>[8]</sup>提出的机器学习方法,主要用于定性分类和定量分析。它通过求解一个线性方程组来代替标准支持向量机(support vector machine, SVM)<sup>[9]</sup>中求解较复杂的二次规划问题,大大减少了计算量,降低了复杂度,同时弥补了支持向量机在分类正确率上的不足。本研究选用拉曼光谱全波段范围作为最小二乘支持向量机的输入量,采用较好处理非线性问题的径向基(RBF)为核函数。利用Matlab7.0软件(美国MathWorks公司,马萨诸塞州,美国)进行数据处理。

## 2 结果与讨论

### 2.1 浙江绍兴、浙江嘉善和上海黄酒的拉曼光谱特性

拉曼光谱是利用散射光强度和拉曼位移作图而产生的,并获得基于键的延伸和弯曲的振动模式的信息。拉曼光谱的分析方法是基于光谱的线性叠加原理,混合物的拉曼光谱等效于各组分拉曼光谱的线性加权,加权系数由样本各成分的浓度决定<sup>[10]</sup>。与近红外光谱相比,拉曼光谱谱峰尖锐清晰,且信息量非常丰富,具有一定的物理意义,一般是各个待测组分在拉曼谱图中的特征频段;通过分析拉曼谱图中的特征谱峰,就可以分析出样品中的组分。图1为浙江绍兴、浙江嘉善以及上海三产地黄酒样品的平均拉曼光谱图。从图中可以看出三产地黄酒样品十分丰富的信息。1000~1100  $\text{cm}^{-1}$ 附近的吸收峰产生于C-C-O面外伸缩;884  $\text{cm}^{-1}$ 附近的吸收峰产生于C-C-O面内伸缩<sup>[11]</sup>。885  $\text{cm}^{-1}$ 、1050  $\text{cm}^{-1}$ 、1097  $\text{cm}^{-1}$ 、1277  $\text{cm}^{-1}$ 和1457  $\text{cm}^{-1}$ 的吸收峰则与乙醇的吸收有关<sup>[12]</sup>。

由图1还可以看出:来自浙江绍兴、浙江嘉善和上海的黄酒样品,虽然产地不同,但它们的拉曼光谱图趋势基本一致;浙江绍兴和浙江嘉善黄酒的拉曼光谱相似度较高。浙江黄酒以传统黄酒为主,其风味醇厚、香气浓郁;上海黄酒则以清爽型黄酒为主,其风

味柔和、鲜爽。

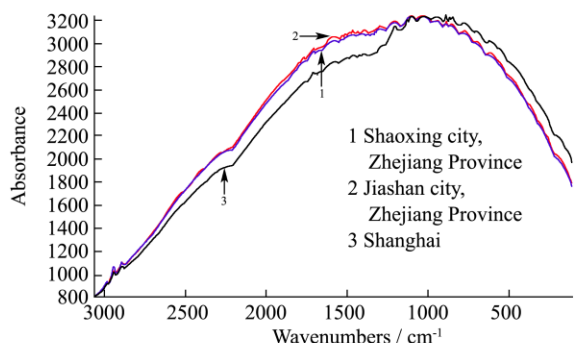


图1 浙江绍兴、浙江嘉善以及上海三产地黄酒样品的平均拉曼光谱图

Fig.1 Average Raman spectra of Chinese rice wines from Shaoxing in Zhejiang Province, Jiashan in Zhejiang Province, and Shanghai

## 2.2 1年陈、3年陈、5年陈和8年陈黄酒的拉曼光谱特性

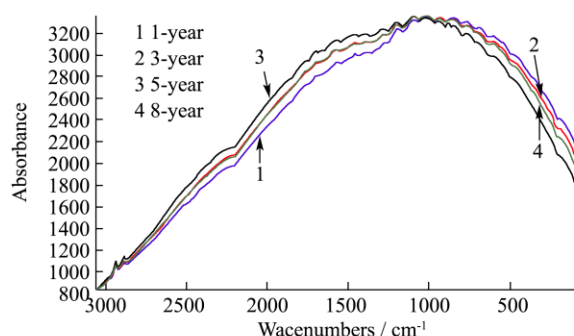


图2 1年陈、3年陈、5年陈和8年陈黄酒的平均拉曼光谱图  
Fig.2 Average Raman spectra of Chinese rice wines aged by 1, 3, 5, and 8 years

图2为1年陈、3年陈、5年陈和8年陈黄酒的平均拉曼光谱图。由图2可以看出：不同酒龄黄酒样品的主要特征峰一致，小于 $1000\text{ cm}^{-1}$ 范围内各个年份吸收强度不同。黄酒在陈酿过程中，一些醇、醛类物质会被氧化成有机酸，酸的增加促进了酯化反应；醇类、氨基酸态氮类物质含量呈减少趋势，总酸、酯类、固形物含量呈增加趋势<sup>[13]</sup>。

## 2.3 黄酒产地鉴别模型建立与优化

利用主成分分析对来自浙江绍兴、浙江嘉善和上海的13个样品进行总体分析，所得主成分分析图如图3所示。主成分分析可在低维空间内描述高维空间最大方差信息，提取的是线性信息。由图3可以看出，同一产地生产的黄酒其样品点基本上都聚集在一个相

近区域，形成各自产地的样品区域块（上海产地黄酒样品点中的SH-1离其它所有样品点都很远，这可能是由于实验误差所造成的，由于另外四个上海黄酒的样品点都聚集在相近区域，故将SH-1样品点剔除在上海区域块之外）。由图3：浙江嘉善与浙江绍兴和上海产地的黄酒能够明显地区分开来，浙江嘉善黄酒的样品点在一个单独的区域，与另外两个产地的黄酒样品点区域没有重叠；而浙江绍兴黄酒样品点区域（SX-2和SX-3样品点重合）与上海黄酒样品点区域部分重叠，不能完全区分开来，说明浙江绍兴、浙江嘉善和上海三产地黄酒样品点的主成分分析可以部分地将它们区分开来。

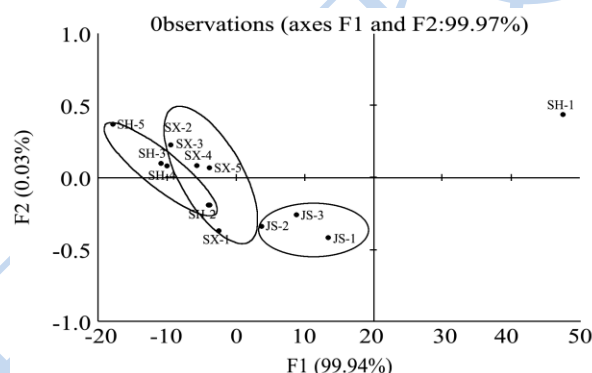


图3 不同产地黄酒的主成分分析二维分类图

Fig.3 Two-dimensional classification plot of the principal components of Chinese rice wines produced at different locations

支持向量机性能的好坏主要取决于合适的核函数和优化的核参数。常用的核函数有4种，包括线性核、多项式核、RBF核和Sigmoid核，其中应用最广泛的是RBF核<sup>[14]</sup>。采用交互验证方式优化RBF核函数参数：首先以较大步长搜索，然后减小步长搜索直至均方根误差小于阈值。图4所示为黄酒产地鉴别中RBF核函数参数优化迭代过程。最优参数回归误差( $\gamma$ )和核参数( $\sigma^2$ )分别为2.531和0.022。

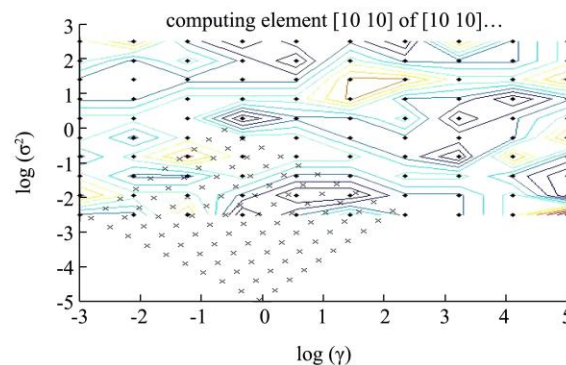


图4 黄酒产地鉴别中RBF核函数参数优化迭代过程图

Fig.4 Iterative process of kernel RBF parameter optimization to differentiate Chinese rice wines of different origins



基于主成分降维信息,建立了判别分析黄酒产地鉴别模型,鉴别模型性能如表1所示;基于优化RBF核函数,建立了最小二乘支持向量机黄酒产地鉴别模型,性能如表1所示。由表1判别分析模型和最小二乘支持向量机模型性能对比分析可得,不同产地黄酒拉曼光谱呈现一定的非线性分布规律,最小二乘支持向量机能完全提取有效信息。最小二乘支持向量机在解决小样本、非线性及高维模式识别中表现出特有的优势。

表1 判别分析法和最小二乘支持向量机所建黄酒产地鉴别模型的性能比较

Table 1 Comparison of models to differentiate Chinese rice wine origins based on DA and LS-SVM

产地	判别分析	最小二乘支持向量机
浙江绍兴	80 %	100 %
浙江嘉善	100 %	100 %
上海	80%	100 %

### 2.4 黄酒酒龄鉴别模型建立与优化

利用主成分分析对来自浙江绍兴的四种不同酒龄的20个样品(每种酒龄5个样品)进行总体分析,所得主成分分析图如图5所示。由图5可知:同一酒龄的样品点聚集在相近的区域,四种酒龄的样品点基本上可以分别形成各自的样品点区域;其中1年陈样品点区域与另外三种酒龄的样品点区域相隔一段距离,能够完全与它们区分开来;3年陈、5年陈和8年陈黄酒样品点区域相隔很近,其中3年陈和8年陈、5年陈和8年陈样品点区域还有部分重叠,彼此不能完全区分开,说明四种酒龄黄酒样品点的主成分分析只能部分地将它们区分开来。

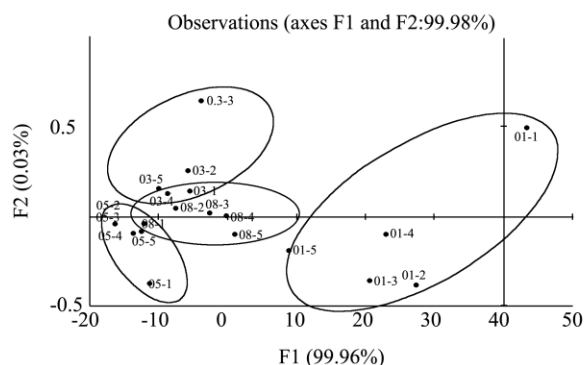


图5 1年陈、3年陈、5年陈和8年陈黄酒样品的主成分分析图  
Fig5 Principal component analysis of 1-, 3-, 5-, and 8-year-old Chinese rice wine samples

基于主成分降维信息(前4个主成分),建立了判别分析黄酒酒龄鉴别模型;基于优化RBF核函数( $\gamma=0.132, \sigma^2=0.015$ ),建立了最小二乘支持向量机黄

酒酒龄鉴别模型。判别分析模型和最小二乘支持向量机模型中1年陈、3年陈、5年陈和8年陈的鉴别正确率均为100%。图6所示为判别分析分类图,不同年份样品集可以清晰鉴别。

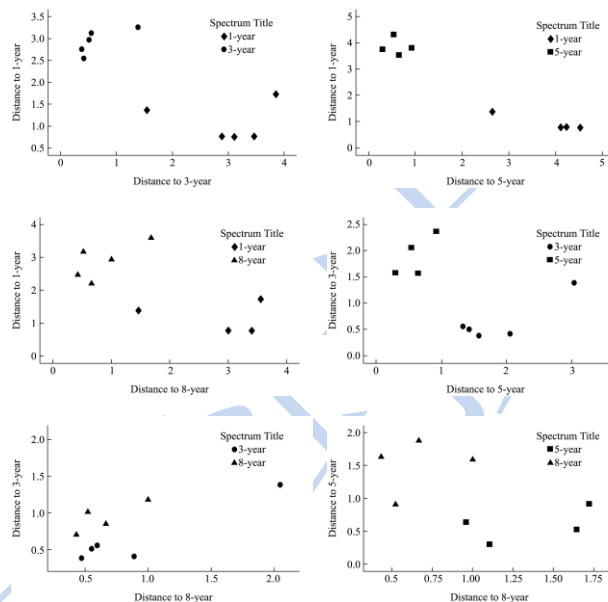


图6 1年陈、3年陈、5年陈和8年陈黄酒样品两两之间的判别分析分类图

Fig6 DA pairwise classification of 1-, 3-, 5-, and 8-year-old Chinese rice wine samples

### 3 结论

利用拉曼光谱结合化学计量学方法分析了不同产地和不同酒龄黄酒。研究表明:拉曼光谱结合最小二乘支持向量机鉴别上海、浙江绍兴和浙江嘉善三个产地黄酒的正确率分别为100%,拉曼光谱结合判别分析或最小二乘支持向量机鉴别1年、3年和5年陈黄酒的正确率分别为100%。拉曼光谱分析技术将为黄酒产地和酒龄甄别提供一种准确、快速、简便的方法,将为保障消费者权益和厂家声誉提了可靠的分析技术。

### 参考文献

[1] 于海燕.黄酒品质和酒龄的近红外光谱分析方法的研究[D].浙江:浙江大学,2007  
YU Hai-yan. Methods for quality and wine age determination of Chinese rice wine based on near infrared spectroscopy [D]. Zhejiang: Zhejiang University, 2007

[2] 苏海荣,安冬梅,王学民,等.现代仪器分析技术在黄酒分析研究中的应用[J].酿酒科技,2012,3:67-71  
SU Hai-rong, AN Dong-mei, WANG Xue-min, et al. Application of modern instrument analysis techniques in the

- analysis of yellow rice wine [J]. *Liquor-making Science & Technology*, 2012, 3: 67-71
- [3] Shen F, Ying YB, Xu HR, et al. Identification of aging status of chinese rice wine using Fourier transform Near-Infrared Spectroscopy [J]. *Transactions of the ASABE*, 2011, 54(5): 1857-1862
- [4] 李琼. 微型拉曼光谱仪的拉曼光谱数据处理方法研究[D]. 重庆:重庆大学,2008
- LI Qiong. Study on data processing of raman spectrum based on mini-spectroscopy [D]. Chongqing: Chongqing University, 2008
- [5] 祁龙凯,林励,陈地灵,等.荔枝蜜拉曼光谱指纹图谱的研究[J].现代食品科技,2014,30(3):201-205
- QI Long-kai, LIN Li, CHEN Di-ling, et al. Fingerprint of the Litchi Honey by Raman Spectroscopic[J]. *Modern Food Science and Technology*, 2014, 30(3): 201-205
- [6] Ma P, Liang F, Sun Y, et al. Rapid determination of melamine in milk and milk powder by surface-enhanced Raman spectroscopy and using cyclodextrin-decorated silver nanoparticles [J]. *Microchimica Acta*, 2013, 180: 1173-1180
- [7] He L, Deen B D, Pagel A H, et al. Concentration, detection and discrimination of Bacillus anthracis spores in orange juice using aptamer based surface enhanced Raman spectroscopy [J]. *Analyst*, 2013, 138(6): 1657-1659
- [8] Suykens J A K, Van G T, De B J, et al. Least-Squares support vector machine [M]. Singapore: World Scientific, 2002
- [9] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995
- [10] 周秀军.基于拉曼光谱的食用植物油定性鉴别与定量分析[D].浙江:浙江大学,2013
- ZHOU Xiu-jun. Qualitative identification and quantitative analysis of edible vegetable oil based on raman spectroscopy [M]. Zhejiang: Zhejiang University, 2013
- [11] 刘文涵,杨未,吴小琼,等.激光拉曼光谱内标法直接测定乙醇浓度[J].分析化学研究简报,2007,35(3):416-418
- LIU Wen-han, YANG Wei, WU Xiao-qiong, et al. Direct determination of ethanol by laser raman spectra with internal standard method [J]. *Chinese Journal of Analytical Chemistry*, 2007, 35(3): 416-418
- [12] 张玉,方立铭.激光喇曼光谱在检测假酒中的应用[J].常州工学院学报,2012,25(5):37-40
- ZHANG Yu, FANG Li-ming. Application of laser raman spectrum in detection of adulterated liquor [J]. *Journal of Changzhou Institute of Technology*, 2012, 25(5): 37-40
- [13] 许荣年,鲍忠定,潘兴祥,等.黄酒的陈化[J].酿酒,2003,30(2):50-52
- XU Rong-nian, BAO Zhong-ding, PAN Xing-xiang, et al. Mautization of Chinese rice wine [J]. *Liquor Making*, 2003, 30(2): 50-52
- [14] 林升梁,刘志.基于RBF核函数的支持向量机参数选择[J].浙江工业大学学报,2007,35(2):163-167
- LIN Sheng-liang, LIU Zhi. Parameter selection in SVM with RBF kernel function [J]. *Journal of Zhejiang University of technology*, 2007, 35(2): 163-167