

文章编号: 1000-7032(2018)09-1310-07

基于 MC-UVE、GA 算法及因子分析 对葡萄酒酒精度近红外定量模型的优化研究

王怡淼^{1,2}, 朱金林³, 张 慧⁴, 赵建新², 顾小红^{1,5*}, 朱华新⁶

(1. 江南大学 食品科学与技术国家重点实验室, 江苏 无锡 214122; 2. 江南大学 食品学院, 江苏 无锡 214122;
3. 浙江大学 控制科学与工程学院, 浙江 杭州 310027; 4. 张家港出入境检验检疫局, 江苏 张家港 215600;
5. 食品安全国际合作联合实验室, 江苏 无锡 214122; 6. 江南大学 理学院, 江苏 无锡 214122)

摘要: 对葡萄酒酒精度偏最小二乘(Partial least squares, PLS)回归模型进行优化研究。使用近红外光谱仪采集葡萄酒样本的光谱数据,用于建立酒精度定量模型,实现在线快速检测。通过蒙特卡罗无信息变量消除(Monte Carlo uninformative variable elimination, MC-UVE)和遗传算法(Genetic algorithm, GA)进行变量选择,基于被选择的变量分别进行 PLS 和因子分析(Factor analysis, FA),建立回归模型。结果表明,MC-UVE-GA-FAR模型预测集相关系数(R^2)为 0.946,预测均方根误差(Root mean square error of prediction, RMSEP)为 0.215,效果优于 MC-UVE-GA-PLS 模型。与基于全范围光谱所建 PLS 回归模型相比,模型效果有所提升,而且模型所选变量个数仅为 6,极大地简化了模型。MC-UVE 和 GA 算法与 FA 分析结合可以实现模型的优化。

关键词: 近红外光谱; 葡萄酒; 遗传算法; 蒙特卡罗无信息变量消除; 因子分析
中图分类号: TN219 **文献标识码:** A **DOI:** 10.3788/fgxb20183909.1310

Optimization of Near Infrared Quantitative Model for Wine Alcohol Content Based on MC-UVE, GA Algorithm and Factor Analysis

WANG Yi-miao^{1,2}, ZHU Jin-lin³, ZHANG Hui⁴, ZHAO Jian-xin², GU Xiao-hong^{1,5*}, ZHU Hua-xin⁶

(1. State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, China;

2. School of Food Science and Technology, Jiangnan University, Wuxi 214122, China;

3. College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China;

4. Zhangjiagang Entry-Exit Inspection and Quarantine Bureau of P. R. C., Zhangjiagang 215600, China;

5. International Joint Laboratory on Food Safety, Jiangnan University, Wuxi 214122, China;

6. School of Science, Jiangnan University, Wuxi 214122, China)

* Corresponding Author, E-mail: guxh@jiangnan.edu.cn

Abstract: The optimization of the PLS regression model of wine alcohol content was studied. The near-infrared spectroscopy was used to collect the spectral data of the wine samples and the data were used to establish the quantitative model of alcohol to achieve rapid on-line detection. PLS regression model and FA model were established based on the selected variables, chosen by MC-UVE and GA. The results show that the MC-UVE-GA-FAR model, which yielded R^2 of 0.946 and RMSEP of 0.215, is superior to the MV-UVE-GA-PLS model. In comparison of the performance of the full-spectra PLS regression model, the model based on the selected wave numbers is much better, and 6 variables in total are selected, which greatly simplifies the model. The study indicates the MC-UVE, GA and FA can optimize the model.

Key words: near-infrared spectroscopy; wine; genetic algorithm; Monte-Carlo uninformative variable elimination; factor analysis

1 引言

葡萄酒在我国起源较早,有着悠久的历史。随着人们生活条件的改善以及我国酿酒技术的提高,葡萄酒越来越被大众所接受,成为餐桌上的必备饮品。葡萄酒的主要成分包括酒精、糖类、醇类、有机酸、维生素、多酚类物质等,每天适量饮用能够增强免疫力,抗氧化^[1],预防癌症、动脉硬化^[2]、糖尿病等。葡萄酒的品质鉴定主要依靠专业的品酒师进行感官评定或是对其进行理化检测,但是这两种方法不仅耗时耗力,而且前期投入大,对样品预处理要求高。因此为了快速智能地对葡萄酒品质进行检测,维护葡萄酒市场的稳定,亟需一种快速无损精确的检测方法。

近年来,国内外已有许多针对葡萄酒快速检测的研究。主要利用红外光谱技术^[3,4]、高效液相色谱法^[5]、气相色谱-质谱联用^[6]等方法结合化学计量学对葡萄酒中的一些成分进行快速检测,其中又以红外光谱技术最为简便快速,能满足工业生产在线无损检测的要求。目前,红外光谱技术已经作为一种常用手段应用于产地溯源、多组分检测、过程控制等领域。吴正宗等^[7]利用傅里叶红外光谱技术实现了黄酒中17种自由氨基酸浓度的快速检测;Basalekou^[8]使用傅里叶红外光谱技术作为检测葡萄酒成熟的工具;Urbano-Cuadrado等^[9]利用近红外光谱技术对葡萄酒样本的15个参数分别建立了PLS模型,大部分模型效果良好;Cozzolino^[10]使用近红外光谱结合化学计量学对葡萄酒中的酚类化合物浓度进行预测。

通常,在建立葡萄酒酒精度定量分析模型时会通过谱图分析选择建模变量,这对光谱分析能力有较高的要求。但如果直接使用全光谱区域建模,往往会包括一些与样品无关的信息,导致模型的稳定性和预测性下降,并且对于大数据量的样本,使用全光谱建模会影响建模及检测效率。因此,选择合适的算法对变量进行筛选是很有必要的,不仅能够简化模型,还能提高建模和检测效率。最常用的变量选择方法有模拟退火(SA)^[11]、遗传算法(GAs)^[12]、人工神经网络(ANN)^[13]、无信息变量消除法(UVE)^[14]和连续

投影算法(SPA)^[15]。本研究主要基于GA和MC-UVE算法选择最有效的特征波长,进一步进行因子分析(FA)建立回归模型,并与PLS回归模型进行比较,探究葡萄酒酒精度定量模型的优化方法。

2 实验

2.1 样品

试验的样品均来自张家港出入境检验检疫局,117个样品分别产自西班牙($n=11$)、智利($n=38$)、澳大利亚($n=16$)、法国($n=42$)、阿根廷($n=4$)、美国($n=3$)、意大利($n=3$),密封保存在4℃的冰箱中。

2.2 近红外光谱采集

使用FT-NIR光谱仪(Thermo Fisher, USA)采集近红外光谱,每次实验前开机预热1h。光谱采集范围为4 000~10 000 cm^{-1} ,扫描次数16次,分辨率8 cm^{-1} 。在透射模式下采集每个样品的光谱信息。将样品在20℃下放置1h,然后将样品置于光程1mm的矩形石英比色皿中,使用RUSULT软件采集光谱数据,光谱数据输出格式为吸光度。每个样品扫描10次,得到平均光谱作为最终谱图。

2.3 参考值的测定

酒精度含量测定采用酒精计法,按照GB/T 15038-2006测定。

2.4 因子分析方法(FA)

因子分析是基于概率回归模型框架下进行的^[16],与传统的回归分析模型相比,概率模型具有以下几个优势^[17]:(1)在概率建模框架下,能够将有效的最大期望(EM)化算法纳入到参数学习中;(2)可以有原则的方式制定概率模型的混合形式;(3)可以通过概率模型解决数据遗漏问题;(4)可以对概率模型进行进一步的贝叶斯(Bayes)处理,用于自动确定潜变量模型的维度。但是,PCA和PCR的概率模型都有一个限制假设:不同的过程变量使用相同的噪声方差。为了解决这个问题,将因子分析引入概率模型中。

传统的FA模型集中在高斯分布的潜变量 t 上,而原始测量变量 x 被视为 t 和噪声 e 的线性组合,其目的是找到最可能的参数集^[18]。传统的

FA 模型只集中在一个无监督的数据集上,而监督的 FA 分析试图在一对数据集 X 和 Y 之间建立模型, $X = [x_1, x_2, \dots, x_n]^T \in R^{n \times m}$, $Y = [y_1, y_2, \dots, y_n]^T \in R^{n \times r}$, 其中 m 代表 X 的测量变量的数量, r 代表 Y 的测量变量的数量, 监督 FA 模型的结构如下述方程式所示:

$$x = A_x t + e_x, \quad (1)$$

$$y = A_y t + e_y, \quad (2)$$

其中 $A_x \in R^{m \times k}$, $A_y \in R^{r \times k}$, 分别是 X 和 Y 的因子载荷矩阵。 $t \in R^{k \times 1}$, 是潜在因子向量, 而 $e_x \in R^{m \times 1}$, $e_y \in R^{r \times 1}$ 分别是 x 和 y 的测量噪声。 在本实验中, x 代表波长, y 代表酒精度。

2.5 蒙特卡罗无信息变量消除法 (MC-UVE)

UVE 可以消除无用的信息变量, 通常使用稳定性来评估每个变量的可靠程度。 在 UVE 算法中, 使用 leave-one-out (LOO) 交叉验证, 而对于 MC-UVE 来说, 是使用 MC 交叉验证来获得稳定性的值, 这两者的区别在于, LOO 交叉验证一次只留一个样本用于验证, 而 MC 交叉验证一次筛选出大部分样品以进行验证。 这增强了验证对建模的影响, 提高了选出最佳模型的可能性^[19]。 通过 MC 算法, 从训练集中随机选择一定量的样本作为用于构建 PLS 子模型的训练子集, 并且该过程重复 M 次。 然后, 计算 PLS 回归系数 $b(M \times p)$ 的矩阵, 通过下式计算每个变量 j 的稳定性:

$$S_j = \frac{\text{mean}(b_j)}{\text{std}(b_j)}, \quad j = 1, \dots, p, \quad (3)$$

其中 $\text{mean}(b_j)$ 和 $\text{std}(b_j)$ 是变量 j 的回归系数的平均值和标准偏差。 稳定性 S_j 的绝对值越大, 相应变量越重要。

2.6 遗传算法 (GA)

遗传算法是基于自然遗传和自然选择过程建立的一种优化方法, 主要包括五个基本步骤: (1) 变量编码; (2) 初始化群体; (3) 计算适应度值; (4) 复制; (5) 变异。 不断重复步骤 (3) ~ (5), 直到达到终止条件^[20]。 应用遗传算法对近红外建模波长进行优化选择是基于该算法能够不断地对模型进行简化以及对有利于提高建模效果的波段或者波长点进行优化组合, 从而达到最佳的预测效果^[21]。

2.7 模型的评估

应用校正相关系数 (R_{cal}^2) 和预测相关系数

(R_{pre}^2), 校正均方根误差 (Root mean square error of calibration, RMSEC)、交互验证均方根误差 (Root mean square error of cross-validation, RMSECV) 和预测均方根误差 (Root mean square error of prediction, RMSEP) 对所开发的模型进行评估。 一般来说, 好的模型应具有较高的 R_{cal}^2 和 R_{pre}^2 值和较低的 RMSEC、RMSECV 和 RMSEP^[22]。

3 结果与讨论

3.1 葡萄酒近红外光谱

图 1 显示了不同产地葡萄酒样品的近红外透射光谱图, 从图中可以看出, 不同产地的葡萄酒具有相似的近红外吸收峰。 其中, 在 $4\,996\text{ cm}^{-1}$ 处的吸收峰, 其吸光度接近 3, 说明此时透过样品的近红外光很微弱, 极易受到噪声等外界环境的影响, 因此在分析时应将其剔除。 同时 $5\,200\text{ cm}^{-1}$ 附近的吸收峰已经饱和, 也应将其剔除。 $6\,862\text{ cm}^{-1}$ 处较强的吸收峰与葡萄酒中碳水化合物和水的一 OH 基团的一级倍频吸收相关, $4\,300 \sim 4\,500\text{ cm}^{-1}$ 波段内的小峰主要与样品中各组分的 C—H 键的合频吸收有关, 而 $5\,250 \sim 6\,000\text{ cm}^{-1}$ 的峰则是与 C—H 键的二倍频吸收有关^[23]。

从以上分析可以看出, 近红外光谱吸收弱, 大多与 O—H、C—H、S—H 和 N—H 等含氢基团的分子振动有关, 并且谱带宽、重叠现象严重, 很难对单一谱带进行精确归属。 因此基于近红外光谱建立定量模型时, 对建模特征变量的选取是有一定难度的, 而光谱区域的选择又会直接影响模型的效果。

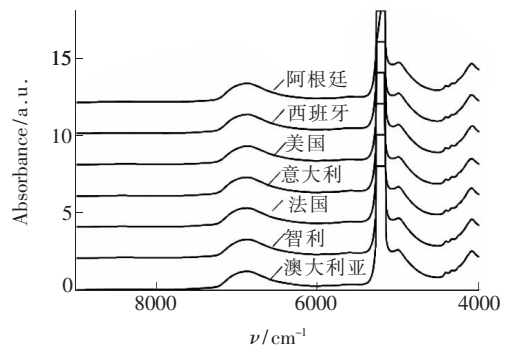


图 1 葡萄酒样本近红外透射光谱图

Fig. 1 NIR transmission spectra of wine samples

3.2 样品的酒精度含量

葡萄酒样品中酒精度含量较为集中, 符合用

于建模的组分含量应尽量均匀分布的原则,因此可直接用于建模。参考国标法测定的葡萄酒样品中酒精度含量,作为建模时的标准数据,产自7个

产地的共117个葡萄酒样品酒精度参考值如表1所示,将117个样品分为78个训练集样品和39个预测集样本。

表1 葡萄酒样品酒精度参考值

Tab.1 Alcohol content of wine samples

Index	Country							
	ALL	ARG	ESP	USA	ITA	FRA	CHI	AUS
Max	14.7	14.7	12.0	14.4	13.0	13.4	14.5	14.5
Min	11.0	13.6	11.3	13.8	11.0	11.4	12.8	12.5
Mean	12.8	14.1	11.5	14.0	12.1	12.1	13.4	13.4
SD	0.90	0.45	0.19	0.26	0.84	0.52	0.44	0.59

3.3 使用全谱结合 PLS 分析方法建立的酒精度回归模型

使用全谱范围进行 PLS 分析,建立回归模型。从表2可以看出,原始光谱建立的模型 R^2 和 RMSECV 分别为 0.891 和 0.301,表明基于近红外光谱建立葡萄酒酒精度 PLS 回归模型具有可行性,但是由于未对建模变量进行筛选,导致红外光谱数据中包含较多的噪声和无关信息,使得模型的预测能力不高。利用预测集样本对该模型进行验证,预测结果见表3。

表2 训练集样本酒精度 PLS 模型参数

Tab.2 Main parameters in PLS models of alcohol content of training samples

Pretreatment	Calibration		Cross-validation	
	RMSEC	R^2	RMSECV	R^2
Full	0.238	0.933	0.301	0.891

表3 不同模型的酒精度预测集

Tab.3 Different models for alcohol content prediction

Model	Number of variables	Prediction	
		R^2_{pre}	RMSEP
Full-PLS	1 557	0.883	0.327
MC-UVE-PLS	29	0.942	0.220
MC-UVE-GA-PLS	6	0.938	0.228
MC-UVE-GA-FAR	6	0.946	0.215

3.4 基于 MC-UVE 方法的变量选择

MC-UVE 算法是用于消除与样本无关的变量。在进行分析时,使用稳定性指标对变量进行评估,以此来决定该变量的重要性。图2所示为通过 MC-UVE 获得的每个变量的稳定性值,虚

线表示阈值,选择绝对值大于阈值的变量用于建模。对于 MC-UVE 算法,选择变量的数量非常关键,如果变量数量较少,有用信息会被忽略;如果将不相关的变量包含进去,将会影响模型的预测性,根据每个变量的稳定性最终选出29个变量。同时,将隐变量作为 PLS 回归模型的输入,建立 MC-UVE-PLS 模型。将 LV 的数目设置为 1~10,根据 R^2_{pre} 确定最佳的 LV 值,当 LV 值为 4 时,获得了最高的 R^2_{pre} ,建立的 MC-UVE-PLS 模型的预测结果如表3所示,预测集 R^2 和 RMSEP 分别为 0.942 和 0.220。

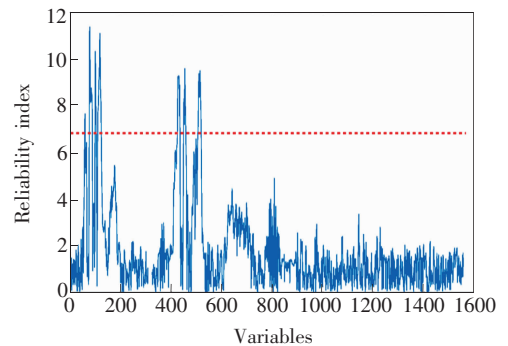


图2 使用 MC-UVE 算法对酒精度预测的变量的稳定性分布

Fig.2 Stability distribution of variables for prediction of the alcohol content by MC-UVE method

3.5 基于 GA 算法的变量选择

在经过 MC-UVE 选择之后,变量数减少到29个,为了对模型进行进一步简化,将 MC-UVE 选择的29个变量作为 GA 算法的输入。在使用 GA 算法时,根据 RMSECV 的值的大小对变量进行选择。从图3可以看出,随着变量数目的增加, RMSECV 呈先减小后增大的趋势,当变量数为6时,

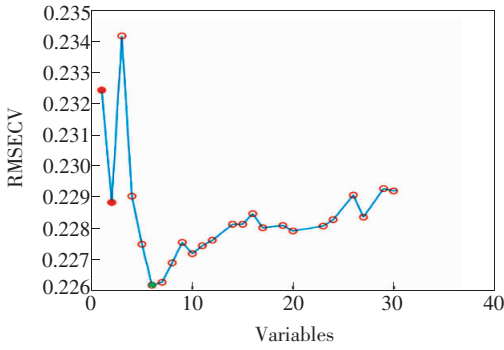


图 3 RMSECV 与所选变量数量的变化关系

Fig. 3 Variation of RMSECV with the number of selected variables

此时 RMSECV 最小。基于所选择的光谱变量建立 PLS 校正模型 (MC-UVE-GA-PLS), 并用未知样品进行预测, R^2_{pre} 及 RMSEP 见表 3。

3.6 因子分析

对经过 MC-UVE-GA 算法选择的变量进行 FA 分析, 建立 MC-UVE-GA-FAR 回归模型, 选定

变量数为 6, 模型的参数 R^2_{pre} 为 0.946, RMSEP 为 0.215, 高于 MC-UVE-GA-PLS 模型的参数。说明 FA 分析方法能够提高模型的性能, 6 个波长被认为是用作 FA 模型输入的最优变量。与基于全光谱的 PLS 模型相比, MC-UVE-GA-FAR 模型的 R^2_{pre} 显著上升, 模型的预测能力有较大提升, 并且 MC-UVE-GA-FAR 模型中使用的变量远低于全谱 PLS 模型, 这将有助于简化预测模型, 提高建模及检测效率, 更能满足实际生产中的需要。

图 4 为两个模型的预测结果的相关性, 从图中可以看出, 采用 MC-UVE-GA-FAR 模型的实测值和预测值的相关性明显高于使用全光谱所建模型, 说明 MC-UVE、GA 算法和 FA 分析能显著提高模型的性能。并且 MC-UVE 和 GA 算法是对变量进行自动选择, 不需要有较强的红外光谱分析能力; 用于建模的变量数仅为 6 个, 有效地提高了建模及检测效率, 更加适合于实际生产应用。

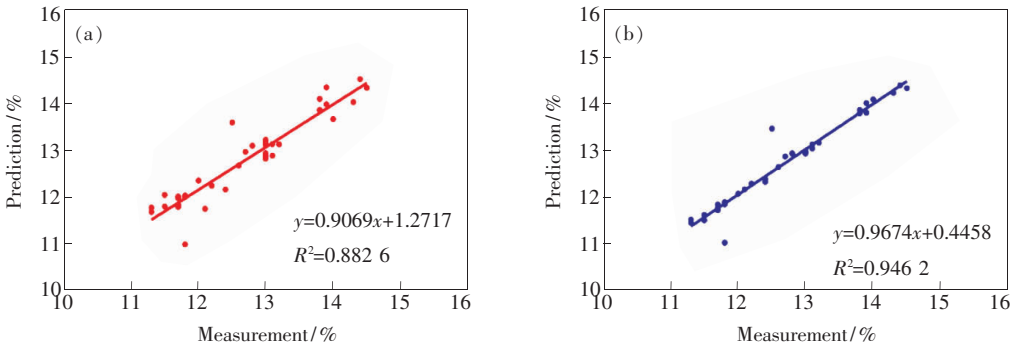


图 4 预测集酒精度实测值与预测值散点图。(a)全光谱 PLS 模型结果;(b)MC-UVE-GA-FAR 模型结果。

Fig. 4 Scatter plot of measured versus predicted alcohol content of the prediction set. (a)Results of full spectra-PLS model. (b) Results of MC-UVE-GA-FAR.

4 结 论

本实验是基于葡萄酒酒精度快速无损检测要求的优化研究。对 117 个葡萄酒样本进行近红外光谱采集, 以采集到的光谱数据的平均值为基础, 利用 MC-UVE 和 GA 算法以选择最佳变量, 同时建立 PLS 回归模型和 FA 回归模型。结果表明, MC-UVE-GA-FAR 模型的效果最佳, 预

测集 R^2 为 0.946, RMSEP 为 0.215, 均优于 MC-UVE-GA-PLS 模型, 说明 FA 分析能够提高模型的预测能力。并且所需的建模变量数仅为 6 个, 简化了模型, 提高了预测效率。本实验证明 MC-UVE 和 GA 算法是变量选择的有效方法, 结合 FA 分析能够实现模型的简化及优化, 满足实际生产应用的要求, 实现葡萄酒酒精度在线快速无损检测。

参 考 文 献:

- [1] FRANKEL E N, GERMAN J B, KINSELLA J E, *et al.*. Inhibition of oxidation of human low-density lipoprotein by phenolic substances in red wine [J]. *The Lancet*, 1993, 341(8843):454-457.

- [2] BIANCA F, ALEXANDRA L, MICHAEL A. Consumption of red wine with meals reduces the susceptibility of human plasma and low-density lipoprotein to lipid peroxidation [J]. *Am. J. Clinic. Nutrit.*, 1995, 61(3):549-554.
- [3] FERRARI E, FOCA G, VIGNALI M, *et al.*. Adulteration of the anthocyanin content of red wines: perspectives for authentication by Fourier transform-near infrared and ¹H NMR spectroscopies [J]. *Anal. Chim. Acta*, 2011, 701(2):139-151.
- [4] URBANO C M, LUQUE D C M, PEREZ J P, *et al.*. Comparison and joint use of near infrared spectroscopy and Fourier transform mid infrared spectroscopy for the determination of wine parameters [J]. *Talanta*, 2005, 66(1):218-224.
- [5] GEANA E, POPESCU R, COSTINEL D, *et al.*. Verifying the red wines adulteration through isotopic and chromatographic investigations coupled with multivariate statistic interpretation of the data [J]. *Food Control*, 2016, 62:1-9.
- [6] 陈士恒, 章晴, 史晓梅, 等. DSPE-GC/MS 快速检测葡萄酒中 52 种农药残留 [J]. *食品科学技术学报*, 2014, 32(3):14-22.
- CHEN S H, ZHANG Q, SHI X M, *et al.*. Fast determination of 52 pesticide residues in grape wine by DSPEGC/MS [J]. *J. Food Sci. Technol.*, 2014, 32(3):14-22. (in Chinese)
- [7] 吴正宗, 王芳, 徐恩波, 等. 傅立叶红外光谱技术快速检测黄酒中的氨基酸质量浓度 [J]. *食品与生物技术学报*, 2017, 36(1):34-40.
- WU Z Z, WANG F, XU E B, *et al.*. Rapid determination of amino acids in chinese rice wine by fourier transform infrared spectroscopy [J]. *J. Food Sci. Biotechnol.*, 2017, 36(1):34-40. (in Chinese)
- [8] BASALEKOU M, PAPPAS C, KOTSERIDIS Y, *et al.*. Red wine age estimation by the alteration of its color parameters: Fourier transform infrared spectroscopy as a tool to monitor wine maturation time [J]. *J. Anal. Methods Chem.*, 2017, 2017:1-9.
- [9] URBANO-CUADRADO M, LUQUE D C M, PEREZ J P, *et al.*. Near infrared reflectance spectroscopy and multivariate analysis in enology [J]. *Anal. Chim. Acta*, 2004, 527(1):81-88.
- [10] COZZOLINO D, KWIATOWSKI M, PARKER M, *et al.*. Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy [J]. *Anal. Chim. Acta*, 2004, 513(1):73-80.
- [11] SWIERENGA H, WULFERT F, NOORD O, *et al.*. Development of robust calibration models in near infrared spectrometric applications [J]. *Anal. Chim. Acta*, 2000, 411(1):121-135.
- [12] LEARDI R, LUPIANEZ G A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them [J]. *Chemometrics Intellig. Lab. Syst.*, 1998, 41(2):195-207.
- [13] BLANCO M, COELL J, ITURRIAGA H, *et al.*. NIR calibration in non-linear systems: different PLS approaches and artificial neural networks [J]. *Chemometrics Intellig. Lab. Syst.*, 2000, 50(1):75-82.
- [14] CENTNER V, MASSART D L, NOORD O, *et al.*. Elimination of uninformative variables for multivariate calibration [J]. *Anal. Chem.*, 1996, 68(21):3851-3858.
- [15] ARAUJO M, SALDANHA T, GALVAO R, *et al.*. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. *Chemometrics Intellig. Lab. Syst.*, 2001, 57(2):65-73.
- [16] ZHENG J, SONG Z, GE Z. Probabilistic learning of partial least squares regression model: theory and industrial applications [J]. *Chemometrics Intellig. Lab. Syst.*, 2016, 158:80-90.
- [17] 陈家益, 赵忠益, 刘飞. 鲁棒 PLS 模型及其在过程监控中的应用 [J]. *化工学报*, 2016, 67(7):2907-2915.
- CHEN J Y, ZHAO Z G, LIU F. Robust PLS model and its applications in process monitoring [J]. *J. Chem. Ind. Eng.*, 2016, 67(7):2907-2915. (in Chinese)
- [18] GE Z. Supervised latent factor analysis for process data regression modeling and soft sensor application [J]. *IEEE Trans. Control Syst. Technol.*, 2016, 24(3):203-209.
- [19] LI C, ZHAO T, LI C, *et al.*. Determination of gossypol content in cottonseeds by near infrared spectroscopy based on Monte Carlo uninformative variable elimination and nonlinear calibration methods [J]. *Food Chem.*, 2017, 221:990-996.
- [20] JIE D, XIE L, FU X, *et al.*. Variable selection for partial least squares analysis of soluble solids content in watermelon using near-infrared diffuse transmission technique [J]. *J. Food Eng.*, 2013, 118(4):387-392.
- [21] 马红辉, 王中江, 袁天军, 等. 烟草中淀粉近红外光谱变量的筛选及校正模型的建立 [J]. *烟草科技*, 2015, 48(8):37-43.

MA H H, WANG Z J, YUAN T J, *et al.*. Screening of NIR spectral variables of starch in tobacco and establishment of calibration model [J]. *Tobacco Sci. Technol.*, 2015, 48(8):37-43. (in Chinese)

[22] COZZOLINO D, CYNKAR W, SHAH N, *et al.*. Multivariate data analysis applied to spectroscopy: potential application to juice and fruit quality [J]. *Food Res. Int.*, 2011, 44(7):1888-1896.

[23] 向伶俐, 李梦华, 李景明, 等. 近、中红外光谱法融合判定葡萄酒产地 [J]. *光谱学与光谱分析*, 2014, 34(10):2662-2666.

XIANG L L, LI M H, LI J M, *et al.*. Determination of wine original regions using information fusion of NIR and MIR spectroscopy [J]. *Spectrosc. Spect. Anal.*, 2014, 34(10):2662-2666. (in Chinese)



王怡森(1992 -),女,江苏无锡人,学士,2015年于扬州大学获得学士学位,主要从事食品快速检测方面的研究。

E-mail: 517604804@qq.com



顾小红(1971 -),女,江苏无锡人,硕士,高级工程师,1999年于无锡轻工大学获得硕士学位,主要从事质谱光谱分析、食品检测方面的研究。

E-mail: guxh@jiangnan.edu.cn