

傅立叶红外光谱技术快速检测黄酒中的氨基酸质量浓度

吴正宗^{1,2}, 王芳^{1,2}, 徐恩波^{1,2}, 徐学明¹, 焦爱权^{*1,2}

(1. 江南大学 食品学院, 江苏 无锡 214122; 2. 食品科学与技术国家重点实验室, 江南大学, 江苏 无锡 214122)

摘要: 氨基酸含量是评价黄酒质量和风味的一项重要指标。为了实现对黄酒中氨基酸质量浓度的快速检测, 作者探索了将傅立叶红外光谱技术(FT-IR)应用于黄酒中的 17 种自由氨基酸质量浓度的快速检测的可行性, 同时采用协同区间偏最小二乘算法(SiPLS)和遗传算法(GA)选出有效波长以期提高模型的预测精度。实验结果显示与基于全波长光谱建立的经典偏最小二乘算法模型相比, 基于 SiPLS 和 GA 选出来的有效光谱变量而建立的回归模型的精度有显著提高, 尤其是对半胱氨酸(Cys)、精氨酸(Arg)和脯氨酸(Pro)。此外, 对于所有的氨基酸, GA-SiPLS 模型效果最好, 交互验证相关系数均大于 0.80, 残余预测偏差均大于 2.00。因此, FT-IR 结合特征谱区筛选方法可以作为常规检测的一种替代手段实现对黄酒中的自由氨基酸的快速检测。

关键词: 波长选择; 傅立叶变换红外光谱; 氨基酸; 遗传算法; 协同区间偏最小二乘算法

中图分类号: TS 251.5 **文献标志码:** A **文章编号:** 1673-1689(2017)01-0034-07

Rapid Determination of Amino Acids in Chinese Rice Wine by Fourier Transform Infrared Spectroscopy

WU Zhengzong^{1,2}, WANG Fang^{1,2}, XU Enbo^{1,2}, XU Xueming¹, JIAO Aiquan^{*1,2}

(1. School of Food Science and Technology, Wuxi 214122, China; 2. State Key Laboratory of Food Science and Technology, Wuxi 214122, China)

Abstract: The content of amino acid in Chinese rice wine (CRW) is one of the most important indexes to evaluate the quality and flavor of Chinese rice wine. In order to rapidly determine the contents of free amino acids in CRW, the possibility of Fourier transform infrared spectroscopy (FT-IR) for the fast detection of 17 different kinds of free amino acids in CRW was discussed. Synergy interval partial least squares (SiPLS) and genetic algorithm (GA) were used to select the most efficient spectral variables to improve the prediction precision of the classical partial least squares (PLS) model based on the full-spectrum. Compared with the PLS model based on the full-spectrum, the prediction accuracy of model based on the spectral variables selected by SiPLS and GA was significantly improved, especially for cysteine, arginine and proline. In addition, GA-SiPLS model showed the most efficient prediction accuracy to all of the free amino acids, with the

收稿日期: 2015-02-05

基金项目: 国家“十二五”科技支撑计划项目(2012BAD37B02; 2012BAD37B06)。

* 通信作者: 焦爱权(1982—), 男, 江苏泰州人, 工学博士, 副教授, 硕士研究生导师, 主要从事食品组分与物性研究。

E-mail: jinlab2008@yahoo.com

引用本文: 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.

correlation coefficient of cross-validation higher than 0.80 and the residual predictive deviation larger than 2.00. The FT-IR combined with efficient variable selection algorithms is confirmed as a useful method to replace the traditional methods for routine analysis of free amino acids in CRW.

Keywords: variable selection, fourier transform infrared spectroscopy, amino acids, genetic algorithm, synergy interval partial least squares

黄酒中的氨基酸也是黄酒的风味物质的前驱体。氨基酸以其具有的鲜、甜、苦、涩、酸等诸多味感赋予黄酒丰富的味觉层次,使其具有鲜美、柔和、浓郁、柔润和协调的特征^[1-3]。此外,黄酒中的氨基酸质量浓度还可被用来实现产地溯源及真假鉴定^[4]。因此,准确的对黄酒中的氨基酸进行定量分析不论对于黄酒的营养价值和风味评估,还是对酒质的控制都具有重要的意义。

目前,黄酒中氨基酸的检测方法主要是高效液相色谱法(HPLC)、毛细管电泳法和氨基酸自动分析仪法。此外,采用气相色谱法检测氨基酸成分的方法也已有报道^[4]。尽管这些分析方法通常精确、可靠,但依然有着诸如需要复杂冗长的样品预处理、耗时等诸多缺点,难以适应现代化的大工业生产。因此,急需一种简便、经济的分析方法来实现氨基酸的快速测定,从而降低成本、提高效率。

红外光谱技术因其快速、无损、绿色的特点引起人们的关注和重视,近年来,其已作为常规化学检测的一种替代手段被广泛应用在组分测定、产地溯源和过程控制等诸多领域^[5-7]。在氨基酸的检测方面,已有许多研究者成功的将近红外光谱技术(FT-NIR)应用到大豆、花生等原料中的氨基酸质量浓度检测中^[8-9],然而,采用傅立叶红外光谱技术对原料中的氨基酸含量进行快速检测的研究则较少。目前,仅 Subramanian 等^[10]采用 FT-IR 对奶酪成熟过程中的有机酸和氨基酸含量进行快速检测。已有学者证明 FT-NIR 更适用于对表征总量的参数的测定,而 FT-IR 则在单个组分质量浓度的预测方面更有优势。此外,之前的研究建立的模型往往是基于全波长的偏最小二乘模型(PLS)。全光谱中包含很多共线变量和无关变量。如果这些变量包含在建模变量中,模型精度将不可避免的受影响。

因此,作者采用 FT-IR 技术来实现对黄酒中氨基酸含量的快速检测,并使用波长选择算法提高模型预测精度,以期对黄酒中氨基酸质量浓度的快速

检测提供帮助。

1 材料与方法

1.1 样品来源

试验共采用 109 份黄酒样品。其中 23 个样品来自“古越龙山”,22 个样品来自“闽族红”,23 个样品来自“女儿红”,20 个样品来自“塔牌”,21 个样品来自“西塘”。为增加回归模型的稳健性,同一品牌的黄酒样品取自不同生产批次(日期)。

1.2 仪器设备

Nicolet iS10 傅立叶红外光谱仪:美国赛默飞公司产品;Ag1100 液相色谱仪:美国安捷伦公司产品。

1.3 实验方法

1.3.1 氨基酸含量的测定 色谱柱采用 ODS HYPERSIL(250 mm×4.6 mm, 5 μ m),柱温保持在 40 $^{\circ}$ C,采用双流动相梯度洗脱。

1.3.2 红外光谱扫描 首先光谱仪开机预热半小时,然后用移液枪取 100 μ L 黄酒样品置于晶片上进行红外光谱扫描。光谱采集条件为:扫描范围 525~4 000 cm^{-1} ,分辨率 4 cm^{-1} ,扫描次数 16,光谱间隔点 1.93 cm^{-1} 。使用去离子水按照与样品相同的条件扫描作为空白对照。每次完成扫描后,使用去离子水清洗晶片,并用纸巾擦拭干净。每个样品采集 3 次光谱,以克服样品的不均匀性。数据采集使用 OMNIC 软件。

1.3.3 多变量回归模型的建立 使用 MATLAB 2010a 软件(美国 MathWorks 公司)对红外光谱数据和氨基酸数据进行定量模型的构建。为了消除基线漂移、颗粒散射及高频随机噪音等影响造成的误差,更有效的提取光谱中的有效信息,建模前,采用平滑(Smooth)、矢量归一化(SNV)、多元散射校正(MSC)、一阶导数(D1)、二阶导数(D2)及基线校正(BC)6 种光谱预处理方法对红外光谱进行处理。选出交互验证均方根误差(RMSECV)最小的方法作为各个氨基酸最佳的预处理方法。同时,采用了主成

分分析(PCA)探索了红外光谱用于不同品牌黄酒区分的可能性。作者建立了3种回归模型。首先建立了基于全波长的经典PLS模型,然后用SiPLS选出最佳的子区间组合,建立SiPLS模型,最后用GA进一步提取有效波长变量,减少冗杂变量,基于这些与氨基酸含量高度相关的几十个变量,建立GA-SiPLS模型。RMSECV、预测均方根误差(RMSEP)、交互验证相关系数(R^2 (cal))、预测相关系数(R^2 (pre))及残余预测偏差(RPD)被用来评价所建立的回归模型的稳健性和预测性能。一般来说,相关系数和残余预测偏差取值越大,均方根误差取值越小,模型的稳健性越好。 R^2 和RPD评价模型的标准如表1所示。各个氨基酸的最佳波长预处理方法如表2所示。

表1 回归模型的评价标准

Table 1 Criteria used for the evaluation of calibration models

R^2	评价	RPD	评价
≥ 0.90	模型极好	>3	精度极好
0.84~0.89	模型很好	2.5~3.0	精度很好
0.75~0.84	模型较好(可准确区分高、中、低值)	2.0~2.5	精度一般
<0.75	模型较差	<2.0	模型不能使用

表2 各个氨基酸的最佳光谱预处理方法

Table 2 Optimal preprocessing methods for spectra detection of different amino acids

氨基酸	最佳预处理	氨基酸	最佳预处理
天冬氨酸	SNV	半胱氨酸	MSC
谷氨酸	SNV	缬氨酸	SNV
丝氨酸	D1	蛋氨酸	Smooth
组氨酸	MSC	苯丙氨酸	SNV
甘氨酸	SNV	异亮氨酸	D1
苏氨酸	D1	亮氨酸	SNV
精氨酸	D2	赖氨酸	MSC
丙氨酸	BC	脯氨酸	SNV
酪氨酸	D2	总量	SNV

2 结果与分析

2.1 氨基酸化学参考值结果分析

通过分析剔除1个异常样本后,剩下的108个黄酒样本采用隔三选一法划分为校正集和预测集。

具体分类方法如下:首先将108个黄酒样品的氨基酸含量按照从高到低的顺序排列,然后将每4个样品作为一个小集合,其中的任意3个划分到校正集中,另一个划分到预测集中,如此循环,最终校正集中含有81个样品,预测集中有27个样品。校正集样本用于黄酒各指标近红外模型的建立,验证集样本用于验证所建模型的准确度与可靠性。两个子集中样品的氨基酸含量的最大值、最小值、平均值和标准偏差如表3所示。各种氨基酸的含量之间差异很大。以校正集中数值为例,甘氨酸(Gly)质量浓度最高,其平均值高达529.00 mg/L,其次是丙氨酸(Ala)和亮氨酸(Leu),它们的平均值都在300.00 mg/L以上,剩下的14种氨基酸质量浓度较少,平均值在16.67~291.75 mg/L之间。除了氨基酸的质量浓度高,黄酒中氨基酸的另一个特点是“质量好”,表2可见,黄酒中富含人体必需的8种氨基酸中的7种(赖氨酸(Lys)、苯丙氨酸(Phe)、甲硫氨酸(Met)、苏氨酸(Thr)、异亮氨酸(Ile)、亮氨酸(Leu)和缬氨酸(Val))。其中赖氨酸尤为重要,因为它对人体的生长有着重要的作用。黄酒各种氨基酸均展现出了很宽的质量浓度范围,这可能是由于各个酒厂采用的酿造技术和生产原料不同。较宽的质量浓度范围十分有利于构建稳健、可靠的模型。此外,校正集中各个氨基酸的含量范围包含了整个预测集的范围,且两个子集的标准偏差基本相同,这说明,样品合理的分配到了两个子集中。

为了更好地理解不同氨基酸之间的关系,作者进行了相关性分析。17种氨基酸之间展现出了很强的相关性。很多氨基酸之间的Pearson相关系数高达0.90以上。

2.2 黄酒的红外光谱谱图解析

由图1可以看出,黄酒的红外光谱中有着丰富的吸收峰。其中1500~1700 cm^{-1} 和3050~3650 cm^{-1} 区域处的两个负的特征吸收峰是由水产生的。1045 cm^{-1} 处的强吸收峰则由乙醇产生,2900 cm^{-1} 和2985 cm^{-1} 附近的吸收峰也是由乙醇中的甲基和亚甲基C-H的伸缩振动所产生的。900~1500 cm^{-1} 处的众多复杂的吸收峰主要来源于糖类、有机酸和蛋白质的中的CH-OH的伸缩振动。不同黄酒样品吸收光谱波形相似,但又不完全重合,既显示了不同样品之间的差异,又显示了大样本群体的基本一致性。

表 3 各个氨基酸的化学参考值

Table 3 Chemical reference values of different amino acids

氨基酸	校正集(N=81)			预测集(N=27)		
	范围	平均值	标准偏差	范围	平均值	标准偏差
天冬氨酸	78.8~349.9	209.7	71.3	102.6~332.9	204.4	72.6
谷氨酸	130.7~492.4	291.8	97.0	130.8~475.5	285.5	96.3
丝氨酸	12.21~66.42	31.7	13.6	12.2~65.3	31.0	13.6
组氨酸	14.5~106.1	49.6	19.8	18.9~97.5	48.1	19.8
甘氨酸	104.8~1 620.0	529.0	464.1	105.1~1 590.0	492.0	436.0
苏氨酸	57.9~177.0	121.4	26.1	59.7~168.9	119.1	26.6
精氨酸	124.0~415.0	293.1	80.0	128.2~411.8	287.8	81.1
丙氨酸	151.3~627.2	392.9	125.4	199.9~611.0	383.5	127.3
酪氨酸	62.5~263.0	166.9	47.1	82.6~247.4	162.3	48.6
半胱氨酸	6.8~75.3	16.7	13.4	6.9~58.6	15.9	11.8
缬氨酸	72.5~297.7	176.0	57.4	85.8~287.7	171.1	58.0
蛋氨酸	8.2~107.4	40.8	26.5	8.4~87.8	39.0	25.3
苯丙氨酸	69.2~255.5	182.3	54.9	89.9~247.8	178.4	57.2
异亮氨酸	47.6~199.6	118.8	41.1	58.7~195.5	115.1	41.1
亮氨酸	119.6~475.3	302.9	87.4	157.1~458.9	295.2	89.2
赖氨酸	51.6~300.1	144.6	52.9	64.7~256.3	140.4	51.7
脯氨酸	56.5~322.9	145.9	51.8	65.4~242.8	140.7	47.9
总量	1 860.0~4 420.0	3 240.0	657.7	1 890.0~4 240.0	3200.0	674.9

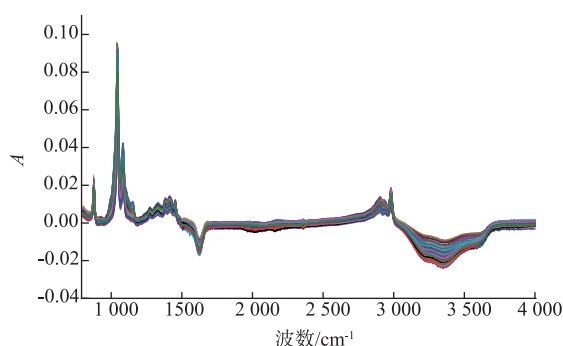


图 1 所有黄酒样品的 FT-IR 光谱图

Fig. 1 Raw FT-IR spectra of all Chinese rice wine samples

2.3 主成分分析

PCA 是一种无监督的模式识别技术,通常用来作为多变量分析的第一步达到降维和提取有效信息的目的。这里,作者采用 PCA 来探索使用 FT-IR 达到对不同品牌的黄酒进行区分的可能性。从图 2 中可以看出,除了“闽族红”,其它 4 个品牌的样品相互聚集、覆盖,没有明显的区分趋势。“闽族红”品牌的黄酒产地是福建省,主要采用红曲作为糖化剂进行发酵,而“古越龙山”、“女儿红”、“塔牌”和“西

塘”4 个黄酒品牌均来自浙江省,它们主要使用麦曲。两者工艺有显著区别,造成化学成分含量和种类的明显不同,因此它们的红外光谱上也会有所不同。

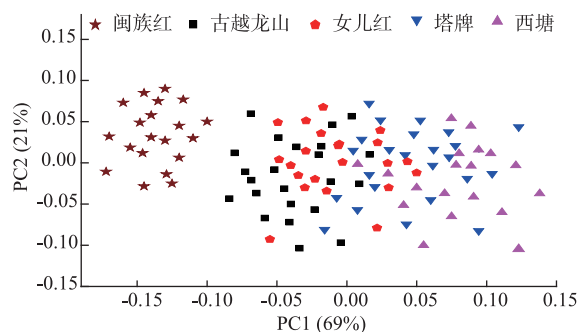


图 2 黄酒样品的前两个主成分的得分图

Fig. 2 Score plot of the first two principal components (PC1 and PC2) of Chinese rice wine samples

2.4 模型构建和评价

2.4.1 协同区间偏最小二乘算法 (SiPLS) SiPLS 是由 Norgaard 等提出的一种特征光谱区间筛选算法,它是在经典偏最小二乘法基础上的拓展和延伸。其具体步骤如下:(1) 将全波长区域分成一定数量的等宽子区间;(2) 对组合区间数目为 2、3、4 的

所有组合建立 PLS 模型;(3) 计算出每个模型的 RMSECV, 选出最小的 RMSECV 对应的子区间组合,用来建立 SiPLS 模型。本实验中,将整个光谱区域分别划分为 11、12、⋯,25 个子区间,以考查不

同子区间划分数目对回归模型稳健性的影响。各个氨基酸的最优区间划分数目、子区间组合及选出的具体波长区间如表 4 所示。图 3 是 SiPLS 为 Val 选出的最佳光谱组合区间。

表 4 17 种氨基酸的最佳 SiPLS 参数
Table 4 Optimal parameters of SiPLS models for 17 amino acids

氨基酸	划分区间	区间组合	变量个数	波长/cm ⁻¹
天冬氨酸	20	[2 4 5 6]	332	960.38~1 118.51; 1 280.50~1 758.76
谷氨酸	22	[2 3 4 6]	304	946.88~1 384.64; 1 533.13~1 677.77
丝氨酸	21	[3 4 7 16]	316	1 106.94~1 409.71; 1 716.34~1 866.76; 3 087.48~3 237.90
组氨酸	20	[2 3 5 6]	332	960.38~1 278.57; 1 440.56~1 758.76
甘氨酸	23	[1 5 6 16]	289	800.31~939.16; 1 363.43~1 639.20; 2 890.77~3 027.69;
苏氨酸	20	[2 3 5 16]	332	960.38~1 278.57; 1 440.56~1 598.70; 3 201.25~3 359.39
精氨酸	23	[3 5 6 7]	289	1 081.87~1 220.72;1 363.43~1 778.05
丙氨酸	23	[2 3 5 6]	290	941.09~1 220.72; 1 363.43~1 639.20
酪氨酸	20	[2 5 6 7]	332	960.38~1 118.51; 1 440.56~1 918.82
半胱氨酸	22	[5 6 16 20]	302	1 386.57~1 677.77; 2 989.12~3 131.83; 3 567.66~3 710.37
缬氨酸	21	[2 3 4 6]	316	954.59~1 409.71; 1 563.99~1 714.41
蛋氨酸	21	[3 4 5 6]	316	1 106.94~1 714.41
苯丙氨酸	22	[3 5 6 7]	264	1 093.44~1 238.08; 1 386.57~1 824.33
异亮氨酸	19	[3 4 5]	264	1 139.72~1 646.91
亮氨酸	18	[2 3 6]	278	979.66~1 336.43;1 695.12~1 870.61
赖氨酸	20	[3 4 5 16]	332	1 120.44~1 598.70; 3 201.25~3 359.39
脯氨酸	19	[1 3 4]	264	800.31~968.09;1 139.72~1 477.21
总量	20	[3 4 5 18]	332	1 120.44~1 598.70; 3 521.38~3 679.51

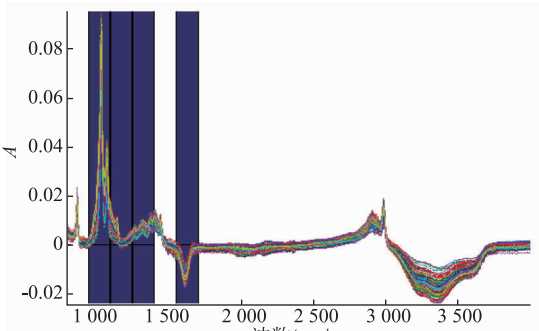


图 3 SiPLS 为 Val 选出的最佳子区间组合
Fig. 3 Optimal combination of subintervals for Val selected by SiPLS

2.4.2 遗传算法(GA) 遗传算法又称进化算法,是一种通过模拟自然进化过程搜索最优解的方法。该算法能自动获取优化空间,自动调整搜索方向,可有效消除无关变量的干扰。然而单纯的采用遗传算法对光谱变量进行处理,往往需要数小时的时间,运行时间过长。用 GA 对 SiPLS 选出的波长变量进

行进一步的提取,并在最终选出的变量的基础上建立 GA-SiPLS 模型。既节省了时间、简化了模型,还可以有效解决 SiPLS 存在的共线性问题。本实验中遗传算法的具体参数设置如下:初始群体大小 30,变异概率 1%,交叉概率 50%,迭代次数 100。由于 GA 选出的波长变量具有随机性的特点,本实验中对于每个氨基酸,GA 均运行 10 次,RMSECV 最小的一组选出的变量用来建立回归模型(GA-SiPLS)。对于 GA 而言,选择哪些变量及选择变量的数目是两个最关键的问题。以 Val 为例,图 4 是基于 SiPLS 选出来的所有变量在 100 次运行中被选择的频率的直方图。在这张图中可以很直观的看出每个变量被选择的频数,直方图越高,代表被选择次数越多,与化学参数(氨基酸)的相关性越大,对模型构建的贡献也就越大。图 5 是变异系数(C.V.)关于包含在模型中的变量个数的函数关系图。随着包含在模型中的变量数目增加,C.V.逐渐增大,最终达到一个最大值保持稳定或略有下降。C.V.达到最大时对应的

变量个数即是最佳变量数目。这时,从图4中按照各变量的频数高低选出相应数目的具体波长变量。

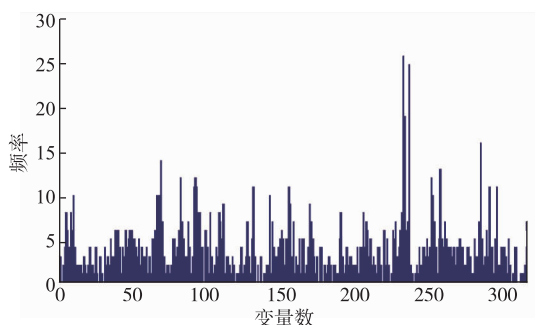


图4 运行100次后所有变量被GA选择的频率的直方图

Fig. 4 Histogram of frequency of selection of each variable after 100 runs by the GA

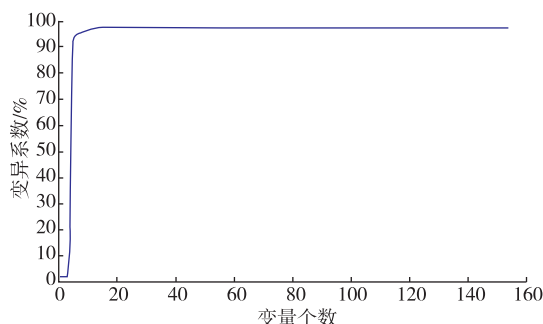


图5 变异系数(C.V.)关于包含在模型中变量个数的函数曲线

Fig. 5 C.V. (%) explained variance as a function of the number of variables included

2.5 模型的建立与评价

在SiPLS和GA筛选结果基础上,分别建立黄酒中各个自由氨基酸的SiPLS模型和GA-SiPLS模型,并与基于全光谱建立的经典PLS模型进行比较。3种回归模型的结果如表5所示。

2.5.1 偏最小二乘模型(PLS) 其中Asp, Glu, His, Gly, Thr, Ala, Val, Phe, Leu的PLS模型的 $R^2(\text{pre})$ 均在0.85~0.86之间,RPD则均在2.50~3.00之间,根据表1中的标准,这些氨基酸的PLS模型得到了很好的预测结果。Ser, Tyr, Met, Ile, Lys的PLS模型的 $R^2(\text{pre})$ 在0.77~0.84之间,RPD则在2.07~2.50之间,这些氨基酸的PLS模型的预测精度一般,只能用于扫描黄酒中的氨基酸质量浓度的高、中、低值。Arg, Cys和Pro的PLS模型结果最差,它们的RPD值均小于2,说明PLS模型不能用于预测这3种氨基酸。Shen等^[15]得到了类似的结果。Pro的结果较差,可能是因为Pro的响应较低,造成参考值测定误差较大。Cys和Arg预测结果也不理想则可能是由

于黄酒中存在的其它复杂化合物干扰了两者的测定。

2.5.2 协同区间偏最小二乘模型(SiPLS) 从表5中可以看出,对于所有的氨基酸,与经典PLS模型相比,基于SiPLS进行波长优化后选出的光谱区间建立的模型的预测精度和模型稳健性均有显著提高。除Ser, Arg, Tyr, Cys和Pro外,其余12种氨基酸的SiPLS模型的RPD均大于3.00,这表明,这12种氨基酸得到了极好的回归模型。由表4可知,对于所有的自由氨基酸,SiPLS选出的波长主要集中在900~1 900 cm^{-1} 的区域里。而这一光谱区间主要与N=O、C=O的伸缩振动及N-H的弯曲振动有关。这些化学键又均与氨基酸密切相关。因此,SiPLS不仅可以去除大部分全光谱中存在的无信息变量和共线变量,而且最大程度的保留了与氨基酸质量浓度相关的光谱变量。因此,相比于经典PLS模型,SiPLS模型的预测性能有明显提高。

2.5.3 协同区间联合遗传算法偏最小二乘模型(GA-SiPLS) 尽管SiPLS淘汰了大部分冗余的光谱变量,但是选择出的相邻光谱区间之间的光谱变量及同一个光谱区间内的变量之间仍存在共线性,少量冗余光谱信息仍然存在。因此,如表5所示,SiPLS的模型精度略低于GA-SiPLS。

经过GA和SiPLS的双重筛选而建立的GA-SiPLS模型与PLS模型相比,模型预测精度有很大提升。对于所有的氨基酸, $R^2(\text{pre})$ 均大于0.80,RPD则均大于2.00,表明经波长筛选后,可所建立的模型可应用于对所有氨基酸的预测。这对于Cys, Pro,和Arg尤其重要,对于Cys,经过波长筛选后,RPD由PLS模型的1.92上升到GA-SiPLS模型的2.50,由不能应用于预测到可很好的用于对Cys质量浓度的预测。对Pro和Arg,基于全波长的PLS模型不能用于对两者质量浓度的预测,而经过波长选择后建立的GA-SiPLS模型则可应用于对Pro和Arg的精确扫描。此外,全波长有1 660个变量,而经过GA和SiPLS双重筛选后,最多含有102个变量(Pro),波长变量减少了93.86%。既简化了回归模型,又节省了模型运行时间。

3 结 语

以成品黄酒为研究目标,综合考虑了多个黄酒品牌,采用傅立叶红外光谱技术对酒中的17种自

表 5 17 种氨基酸的 PLS, SiPLS 和 GA-SiPLS 模型的预测结果

Table 5 Statistic results of FT-IR equations of 17 amino acid contents based on PLS, SiPLS and GA-SiPLS models

氨基酸	偏最小二乘			协同区间偏最小二乘			协同区间联合遗传算法偏最小二乘		
	$R^2(\text{cal})$	$R^2(\text{pre})$	RPD	$R^2(\text{cal})$	$R^2(\text{pre})$	RPD	$R^2(\text{cal})$	$R^2(\text{pre})$	RPD
天冬氨酸	0.885 1	0.850 1	2.58	0.921 0	0.902 7	3.21	0.948 0	0.939 1	4.05
谷氨酸	0.898 0	0.862 8	2.70	0.933 7	0.915 4	3.44	0.964 6	0.955 7	4.75
丝氨酸	0.802 3	0.767 1	2.07	0.853 2	0.834 9	2.46	0.890 9	0.881 9	2.91
组氨酸	0.896 5	0.861 3	2.69	0.942 4	0.923 9	3.63	0.966 4	0.957 6	4.86
甘氨酸	0.897 1	0.861 9	2.69	0.929 4	0.911 2	3.35	0.973 7	0.964 7	5.32
苏氨酸	0.886 8	0.851 6	2.60	0.926 1	0.907 8	3.29	0.965 0	0.955 9	4.76
精氨酸	0.747 9	0.712 6	1.87	0.811 1	0.792 8	2.20	0.838 7	0.829 8	2.42
丙氨酸	0.898 7	0.863 5	2.71	0.919 7	0.901 3	3.18	0.946 8	0.937 8	4.01
酪氨酸	0.838 5	0.803 4	2.25	0.881 2	0.862 9	2.70	0.919 2	0.910 2	3.34
半胱氨酸	0.765 3	0.730 1	1.92	0.822 1	0.803 8	2.26	0.846 1	0.839 7	2.50
缬氨酸	0.902 9	0.867 7	2.75	0.918 9	0.900 6	3.17	0.940 4	0.931 5	3.82
蛋氨酸	0.869 4	0.834 2	2.46	0.918 4	0.900 0	3.16	0.943 4	0.934 2	3.90
苯丙氨酸	0.881 1	0.845 9	2.55	0.919 5	0.901 2	3.18	0.944 0	0.934 8	3.92
异亮氨酸	0.871 6	0.836 3	2.47	0.912 1	0.893 8	3.07	0.937 2	0.928 4	3.74
亮氨酸	0.901 9	0.866 7	2.74	0.927 8	0.909 4	3.32	0.949 6	0.940 4	4.10
赖氨酸	0.871 1	0.836 0	2.47	0.928 7	0.910 2	3.34	0.958 6	0.949 4	4.45
脯氨酸	0.708 1	0.673 0	1.75	0.771 7	0.753 3	2.01	0.808 7	0.800 1	2.23
总量	0.896 2	0.861 2	2.68	0.936 9	0.918 6	3.50	0.969 1	0.960 3	5.02

由氨基酸进行了定量检测,同时利用 SiPLS 和 GA 多波长变量进行优选,以提高模型的稳定性和预测精度。结果表明,经波长筛选,模型精度得到了极大

地提升,同时建模变量数目大大降低,此外,优选出的波长变量与氨基酸特征官能团高度相关,说明波长筛选提高了回归模型的可解释性。

参考文献:

- [1] SHEN F, YING Y Y, LI B B, et al. Multivariate classification of rice wines according to ageing time and brand based on amino acid profiles[J]. **Food Chemistry**, 2011, 129(2): 565-569.
- [2] ZHANG Ying, WANG Jialin, YU Qinfeng. Measurement of total nitrogen content and amino acid content in Miaofu rice wine[J]. **Liquor-Making Science & Technology**, 2011, 208(10): 98-100. (in Chinese)
- [3] LI Bobin, ZENG Jinhong, LIU Xingquan, et al. Study on quantitative relationships between amino acids and sensory taste of yellow rice wine[J]. **Liquor-Making Science & Technology**, 2011, 208(10): 98-100. (in Chinese)
- [4] HASEGAWA H, SHINOHARA Y, MASUDA N, et al. Simultaneous determination of serine enantiomers in plasma using Mosher's reagent and stable isotope dilution gas chromatography-mass spectrometry [J]. **Journal of Mass Spectrometry**, 2011, 46(5): 502-507.
- [5] SHEN F, YING Y Y, LI B B, et al. Prediction of sugars and acids in Chinese rice wine by mid-infrared spectroscopy [J]. **Food Research International**, 2011, 44(5): 1521-1527.
- [6] FENG Yu, GU Xiaohong, TANG Jian, et al. Discrimination of tea varieties by mid-infrared spectroscopy combined with pattern recognition[J]. **Journal of Food Science and Biotechnology**, 2007, 26(2): 7-11. (in Chinese)
- [7] YANG Lijun, LI Zhaojie, WANG Jing, et al. Rapid differentiation and identification of three species of Listeria by FT-IR spectroscopy[J]. **Journal of Food Science and Biotechnology**, 2013, 32(2): 169-173. (in Chinese)
- [8] KOVALENKO I V, RIPPEKE G R, HURBURGH C R, et al. Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy[J]. **Journal of Agricultural and Food Chemistry**, 2006, 54(10): 3485-3491.
- [9] WANG L, WANG Q, LIU H Z, et al. Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy[J]. **Journal of the Science of Food and Agriculture**, 2013, 93(1): 118-124.
- [10] SUBRAMANIAN A, ALVAREZ V B, HARPER W J, et al. Monitoring amino acids, organic acids, and ripening changes in Cheddar cheese using Fourier-transform infrared spectroscopy[J]. **International Dairy Journal**, 2011, 21(6): 434-440.