

利用近红外光谱技术对葡萄酒原产地进行 Fisher 判别

刘 巍¹ 李德美² 刘国杰³ 黄卫东¹ 战吉成¹

(1.中国农业大学食品科学与营养工程学院,北京 100083;2.北京农学院食品科学学院,北京 102206;

3.中国农业大学农学与生物技术学院,北京 100193)

摘要: 以中国昌黎、烟台和沙城干红葡萄酒样本为试验材料,利用近红外光谱分析技术,建立一种可对葡萄酒产地进行辨别的方法。试验中对原产自上述3地的共计60份干红葡萄酒样本进行近红外光谱采集,通过主成分分析后获得14个主成分,利用逐步回归筛选方法最终确定其中5个主成分,建立 Fisher 判别函数,并进行判别模型验证。结果表明,通过利用原始样本和留一交叉验证法对预测模型进行验证,准确率分别为90%和86.7%。所有预测结果的准确率均达到80%以上,可以满足定性判别的要求。

关键词: 葡萄酒; 原产地; 近红外光谱; 逐步回归分析

中图分类号:O657.33;TS262.6;TS261.4

文献标识码:B

文章编号:1001-9286(2010)07-0065-03

Application of Near Infrared Spectroscopy and Fisher Discriminant Function to Distinguish the Original Production Place of Grape Wine

LIU Wei¹, LI De-mei², LIU Guo-jie³, HUANG Wei-dong¹ and ZHAN Ji-cheng¹

(1.College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083;

2.Department of Food Science, Beijing University of Agriculture, Beijing 102206; 3.College of

Agriculture and Biotechnology, China Agricultural University, Beijing,100193, China)

Abstract: In the experiment, near-infrared spectroscopy (NIRS) was used to develop a method capable of distinguishing the original production place of dry red grape wine. The NIRS preprocessed data were collected from 60 grape wine samples (originally made in Changli, Yantai and Shacheng) and then analyzed by principal components analysis (PCA) and stepwise regression analysis (SRA). Fourteen principal components were obtained in PCA and finally five of them were selected by stepwise regression selection method in order to establish Fisher discriminant function. And Fisher discriminant function was verified. The results showed that the accuracy rates were 90% and 86.7% respectively by using the original sample and the remaining one cross-validation. All predicting accuracy rates were above 80%, which could meet the requirements of qualitative discrimination.

Key words: near-infrared spectroscopy; wine; geographical origin; stepwise regression

葡萄酒有“七分原料,三分工艺”之称,其质量和风格主要取决于酿酒葡萄的质量和与之相适应的加工工艺。而在某一特定的区域内,经过长期的实践,往往会形成一些特定的生产工艺,从而生产出具有特定风格和品质的葡萄酒。自然与人文因素的完美结合,使葡萄酒深深地打上了地域的烙印。我国于1999年由国家质检总局首先颁布《原产地域产品保护规定》,2005年更改为《地理标志产品保护规定》,截至目前,已有河北昌黎、河北沙城、山东烟台、宁夏贺兰山东麓和吉林通化等葡萄酒产区先后为各自产区申请了地理标志产品保护^[1-2]。

作为典型的地域性产品,如何在全球市场竞争中突

出本地葡萄酒的地域特点、提高品牌价值,从而占有市场、为生产者带来更大利润,成为各葡萄酒生产商、生产区域甚至生产国的共同诉求。因此,建立有效的葡萄酒原产地鉴别方法是规范葡萄酒地理标志认证的关键技术。

Marisa C等^[3]研究发现Sr的同位素组成受季节和气候的影响不大,建立的数据库比较稳定,而且葡萄酒中的⁸⁷Sr/⁸⁶Sr比值与土壤中的差异不大。因此,他们一致认为Sr是判断葡萄酒地域来源理想的同位素指标。Rodriguez-Delgado等^[4]发现多酚可以作为葡萄酒的特征成分,利用高效液相色谱技术和主成分分析统计分析方法鉴别了Canary Islands地区不同产地的葡萄酒。Guadar-

基金项目 国家质检公益行业科研专项 地理标志产品检测系统与技术研究示范(编号:10-85)。

收稿日期:2010-05-24

作者简介:刘巍(1984-),男,内蒙古赤峰人,在读硕士,主要从事葡萄酒检测研究。

通讯作者:战吉成(1972-),男,北京人,博士,副教授,研究方向为葡萄与葡萄酒工程,E-mail:jczhan@263.net。

rama 等^[5]用电子鼻实现了对 2 种西班牙红葡萄酒和 1 种白葡萄酒进行检测和区分。秦丽娜等^[6]采用固相微萃取-气相色谱联用方法对天津、宁夏 2 个产区的赤霞珠、梅鹿辄干红葡萄酒的香气成分分析结果也揭示了不同品种和不同产地的葡萄酒具有各自的特性,可以通过少数主要香气组分对葡萄酒的品种和产地进行识别分析。

但是,以往的分析方法成本高,操作复杂,因此,迫切的需要寻找一种快速低成本的葡萄酒原产地鉴别技术。近红外光谱(near-infrared spectroscopy, NIRS)是指波长介于可见区与中红外区之间的电磁波,其波数范围为 12500~4000 cm^{-1} 。近红外谱区主要包含 O-H、C-H、N-H 等含氢等基团原子间振动的倍频与合频信息^[7]。近红外光谱技术作为一种快速无损绿色分析检测技术,已应用于中药材品种产地^[8]、稻谷年份^[9]、羊肉产地^[10]和茶叶产地鉴定^[11]等定性判别领域,但还没有将近红外技术应用于中国干红葡萄酒产地鉴别的报道。本研究利用傅里叶变换近红外光谱分析技术,研究 3 个中国葡萄酒地理标志产区的葡萄酒近红外光谱特征,结合 Fisher 判别函数,探索建立一套根据葡萄酒的近红外光谱特征快速识别其产地的方法的可行性。

1 材料与方 法

1.1 仪器设备

德国布鲁克光学仪器公司 (Bruker Optics Inc.)MPA 傅立叶变换近红外光谱仪,最大分辨率为 2 cm^{-1} ,谱区范围为 12800~4000 cm^{-1} (780~2500 nm),半导体冷却的高灵敏度 InGaAs 检测器。

1.2 样品来源

中国葡萄酒样品采集于按照标准《GB 19049—2003 原产地域产品昌黎葡萄酒》、《GB 19265—2003 原产地域产品沙城葡萄酒》和《GBT 18966—2008 原产地域产品烟台葡萄酒》所示的地区种植的葡萄为原料,按照其规定酿造方法所酿的干红葡萄酒,通过酒厂取样方式获取样品。分别获取原产于昌黎、烟台和沙城的干红葡萄酒样品各 20 份,总计 60 份。

1.3 光谱采集与预处理

利用光程为 2 mm 的石英液体光纤探头,以空气为参比,分辨率为 8 cm^{-1} ,扫描次数为 16 次的参数下采集光谱。所得光谱的波数范围为 12500 cm^{-1} 至 4000 cm^{-1} 。利用光谱仪自带的 OPUS 6.5 软件进行光谱采集和光谱平滑去噪以及微分等预处理。

1.4 数据分析方法

利用软件 Unscrambler 9.7(CAMO,挪威)进行光谱主成分分析(principal components analysis, PCA),SPSS16.0

(SPSS,美国)进行识别函数的建立和识别效果的分析。对所有样本进行 PCA 分析,提取主成分,采用逐步回归判别(stepwise regression analysis, SRA)方法构建 Fisher 判别函数,每次选入 Wilks- Λ 统计量最小的变量进入函数进行逐步回归运算,根据变量进入函数的 F 检验值最小值为 3.84,剔除函数的最大值为 2.71 的标准。即当被加入的变量的值 $F \geq 3.84$ 时,该变量进入函数,当要被移出的变量的值 $F \leq 2.71$ 时,该变量被移出函数。

2 结果与分析

2.1 光谱波段选择和主成分分析

图 1 为原产地为昌黎、烟台和沙城的葡萄酒样品近红外透射图谱。由图 1 可以看出,在波数 12000~4000 cm^{-1} 的近红外谱区内,3 个产区的葡萄酒的响应模式是相同的,但在 7500~4000 cm^{-1} 区域内存在明显差异。这一范围是近红外光谱的一级倍频和合频谱区,由于葡萄酒中含有大量糖和多酚类物质,所以,此部分光谱信息含量大于其他谱区。其中,有 3 个明显的吸收峰,波数 4210 cm^{-1} 处吸收峰信息主要表现为碳水化合物中的 O-H 的倍频吸收,这与葡萄酒中糖类物质有关。波数 5200 cm^{-1} 处吸收峰主要表现为水中 O-H 的合频吸收,此处水的吸收非常强烈,在这种情况下,其他成分的吸收有可能被其所掩盖。波数 6870 cm^{-1} 处吸收峰信息主要表现为水的 O-H 倍频和 N-H 倍频,是在此处水的吸收强度大大低于波数 5200 cm^{-1} 处。分析上述吸收峰,水分吸收所导致的高吸收值占有很大比例,这与葡萄酒的高水分含量相符合。

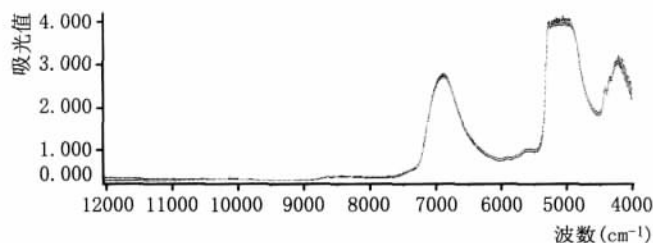


图 1 原产地为昌黎、烟台和沙城的葡萄酒样品近红外透射图谱

图 2 是经过矢量归一化和一阶导数变换的 3 个地区葡萄酒光谱图。由图 2 可以看出,经过一阶导数处理后,光谱的重叠峰可以更加明显的分开,更细致地反映出样品的光谱特征。结果表明,在 4000~5200 cm^{-1} , 6500~7200 cm^{-1} 波段区域,3 个产地葡萄酒的光谱差异较大。此区域以外的光谱对于表征样品的作用较小。

通过对光谱的预处理可以发现光谱差异比较明显的区域,选择此区域建模可以减少运算量并提高建模效率。但是,此区域所对应的光谱点将近 500 个,在 Fisher

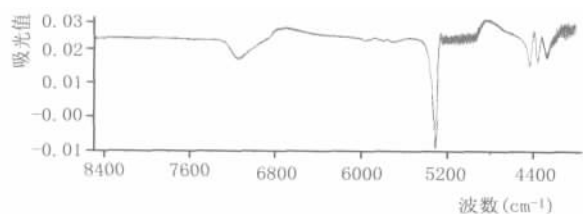


图2 经过矢量归一化和一阶求导处理后的3个产地葡萄酒样品的近红外透射图谱

判别过程中亦需要大量运算才能实现, 为了进一步提高建模效率, 本试验采用主成分分析降维的方法对原始光谱信息进行降维处理。图3为葡萄酒样本光谱前16个主成分的累计贡献率。

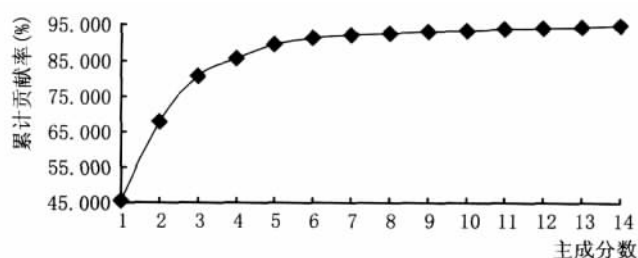


图3 葡萄酒样本光谱前16个主成分的累计贡献率

从图3中可以看出, 通过主成分分析对葡萄酒样本光谱信息提取的主成分数达到14时, 累计贡献率为95.01%, 已经超过95%, 基本可以认为涵盖了原光谱所携带的信息。因此, 本试验中, 采用这14个主成分作为自变量进行逐步回归建模并建立Fisher判别函数。

2.2 Fisher 自变量的确定和判别函数的建立

根据 Wilks- Λ 统计量的大小进行自变量选择顺序的筛选, 每次选择最小值进行逐步回归运算。当被加入的变量的值 $F \geq 3.84$ 时, 该变量进入函数, 当要被移出的变量的值 $F \leq 2.71$ 时, 该变量被移出函数。最终共进行5部筛选, 有5个变量入选, 它们分别是: 第1主成分、第3主成分、第6主成分、第2主成分、第4主成分(分别记为: PC1、PC3、PC6、PC2、PC4)。

表1 入选变量及其 Wilks- Λ 统计量和 F 检验值

进入判别方程顺序	变量名称	Wilks- Λ 统计量	F 检验值
1	PC1	0.477	31.195
2	PC3	0.317	21.756
3	PC6	0.233	19.668
4	PC2	0.171	19.176
5	PC4	0.148	16.955

根据入选变量建立 Fisher 判别函数 F1、F2 和 F3。

$$F1 = 2.043PC1 + 0.22PC2 - 1.466PC3 - 0.623PC4 - 1.302PC6 - 2.709$$

$$F2 = 0.77PC1 - 1.281PC2 + 2.137PC3 + 0.86PC4 + 1.35PC6 - 2.748$$

$$F3 = -2.121PC1 + 1.061PC2 - 0.671PC3 - 0.237PC4 - 0.048PC6 - 2.424$$

将样本相应的主成分分别代入3个方程中求得3个函数值, 并作比较。若 F1 最大, 则将样本产地判定为来自昌黎; 若 F2 最大, 则将样本判定为来自烟台; 若 F3 最大, 则将样本产地判定为来自沙城。

2.3 判别函数对训练集样本的识别和交叉验证

利用建立的判别函数对昌黎、烟台和沙城3地的葡萄酒样本进行识别和留一法交叉验证。

表2 Fisher 判别函数判别昌黎、沙城和烟台葡萄酒样品的结果

验证集合选择	类别	判别正确个数	判别错误个数	正确率 (%)	样本总数
原始样本	昌黎	19	1	95.0	20
	烟台	18	2	90.0	20
	沙城	17	3	85.0	20
	平均	54	6	90.0	60
交叉验证	昌黎	19	1	95.0	20
	烟台	16	4	80.0	20
	沙城	17	3	85.0	20
	平均	52	8	86.7	60

从表2中可以看出, 在原始样本识别中, 昌黎葡萄酒样本的识别率最高, 达到95.0%; 沙城葡萄酒样本的识别率最低, 为85.0%; 平均识别正确率为90.0%。在利用留一交叉验证法对 Fisher 判别模型进行检验中, 昌黎葡萄酒样品的识别率依然最高并与原始样本验证结果一致。烟台样品识别率仅为80.0%; 平均识别正确率为86.7%, 略低于原始样本识别结果。

3 结论与讨论

3.1 本试验初步探讨了近红外光谱技术在中国原产地葡萄酒产地鉴别中的应用并建立昌黎、烟台和沙城的干红葡萄酒样品 Fisher 判别模型, 通过原始样本和留一交叉验证对预测模型进行验证, 准确率分别为90.0%和86.7%。所有预测结果的准确率均达到80.0%以上, 可以满足定性判别的要求。

3.2 试验中采用平滑去噪法和一阶求导对光谱进行预处理, 通过分析与处理后的光谱图可以发现在4000~5200 cm^{-1} 、6500~7200 cm^{-1} 波段区域3个产地葡萄酒的光谱差异较大。采用此范围进行建模可以提高建模效率并减少无用信息对模型的干扰。

3.3 通过主成分分析方法有效对光谱数据进行降维, 所提取的前14个主成分的累计贡献率超过95%, 利用这14个主成分作为逐步回归变量。

3.4 最终确定第1、第2、第3、第4、第6这5个主成分

(下转第69页)

新酒不良气味被酒精吸收稀释, 催陈物促使酒中的各种物质快速缔合转化, 转换不良气味为酒香味, 加速老熟。从而有效地减少了新酒味、酒精味、香精香料的浮香味及其他有害物质, 使配制出的新型白酒更具有固态发酵酒的口感。比传统工艺生产的酒口感绵甜(不用甜味剂), 整体感好, 饮用下咽顺畅, 经动物实验对比, 具有降低酒对机体肝、胃损伤的功效。有减轻和降低白酒产品过量饮用后的不良反应的功能^[2]。

2.2 处理后酒体不浑浊沉淀 过滤方法简单

生产的配制酒任意降度后冷冻至-20℃均无浑浊沉淀。各种卫生指标符合国家标准。如固态法酒质太差或固态酒与食用酒精比例太高, 任意降度检测出现浑浊时, 采用白酒除浊过滤设备滤除即可(但主体香味成分要损失4%~5%), 一般正常情况下可省略这道工序, 采用常规过滤方法即可。

2.3 生产效率高 生产成本低

不需新增设备, 在白酒厂现有设备条件下即可满足生产需求, 而且还减少了活性炭处理酒精过程和新酒老熟过程中过多的储罐占用、资金占用、酒精损耗。提高了生产效率, 降低了生产成本, 降低了劳动强度和环境污染。

2.4 安全可靠

其机理是采用活性物质有效转化酒内的有害物质为有益物质, 而非传统的遮盖原理, 所以生产出的酒质量稳

定, 长期存放不返生。产品所用原料全部来源于食品, 生产和出厂环节均采用仪器检测和动物实验相结合的严格检测方法, 以确保产品安全。产品多次经多家法定机构检测, 均判定为无毒级和合格品。新食品安全法实施后, 作为酿酒辅料使用, 又重新通过有关专家评审和法定机构的登记备案。产品也是首届全国白酒科技大会唯一指定的酒用催陈物。

2.5 技术水平领先 推广应用潜力大

用无毒物转化酒中的有害物为有益物的技术, 在全国处于领先水平, 属国家支持和鼓励发展的项目, 从而为公司奠定了良好的研究发展环境, 使得公司的研发推广工作不断取得新的进展。

3 结束语

经过多年的研究和实践应用证明, 所述新工艺流程短, 物耗低, 操作方法简单, 生产过程中各工艺节点可控, 安全可靠, 能有效提高酒质、降低成本, 又具有节能减排, 降低环境污染的效果, 可广泛推广使用。

以上关于永生宜康在新型白酒生产中的应用介绍, 及研究应用中的一些体会, 愿与广大酒业同仁探讨交流, 共同努力, 为新型白酒的更好发展尽绵薄之力。

参考文献:

- [1] 翟志忠. 浅谈对新型白酒的认识[J]. 酿酒, 2003, (7): 77.
- [2] 傅兰, 等. 白酒加入宜康后减轻对人体损伤的实验[J]. 酿酒, 2004, (1): 85-86.

(上接第 67 页)

为判别变量, 并建立 3 个判别函数。

3.5 后续的研究应该继续对国内外不同产地葡萄酒进行光谱采集和分析, 扩充葡萄酒产地的光谱图谱库, 提高葡萄酒原产地指纹图谱的多样性和代表性, 提高葡萄酒产地鉴别的适用性。

参考文献:

- [1] 姚黎, 孔令梅. 葡萄酒地理标志保护制度探析[J]. 国际贸易问题, 2004, (11): 87-91.
- [2] 管纪尧. 农产品地理标志国际法律保护[D]. 兰州: 兰州大学, 2009.
- [3] Guimet F, Ferre J, Boque R, et al. Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix[J]. Analytica Chimica Acta, 2004, 515(1): 75-85.
- [4] Rodriguez-Delgado MA., Gonzalez-Hernandez G, Conde-Gonzalez J E, et al. Principal component analysis of the polyphenol content in young red wines [J]. Food Chemistry, 2002, 78(4):

523-532.

- [5] Guadarrama A, Fernandez J A, Iiguez M, et al. Array of conducting polymer sensors for the characterisation of wines[J]. Analytica Chimica Acta, 2000, 411(1-2): 193-200.
- [6] 秦丽娜, 倪元颖, 梁方华, 等. 利用香气成分识别葡萄酒品种和产地的初步研究[J]. 酿酒科技, 2008, 164(2): 40-44.
- [7] 陆婉珍, 袁洪福, 徐广通, 等. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2000.
- [8] 刘沐华, 张学工, 周群, 等. 近红外漫反射光谱法和模式识别技术鉴别中药材产地[J]. 光谱学与光谱分析, 2006, 26(4): 629-632.
- [9] 周子立, 张瑜, 何勇, 等. 基于近红外光谱技术的大米品种快速鉴别方法[J]. 农业工程学报, 2009, 25(8): 131-135.
- [10] 张宁, 张德权, 李淑荣, 等. 近红外光谱结合 SIMCA 法溯源羊肉产地的初步研究[J]. 农业工程学报, 2008, 24(12): 309-312.
- [11] 占茉莉, 李勇, 魏益民, 等. 应用 FT-IR 光谱指纹分析和模式识别技术溯源茶叶产地的研究[J]. 核农学报, 2008, 22(6): 829-833.