

苹果酒发酵过程中糖度近红外光谱检测模型的建立

彭帮柱, 岳田利*, 袁亚宏, 高振鹏

西北农林科技大学食品科学与工程学院, 陕西 杨凌 712100

摘要 在苹果酒发酵过程中, 由于糖度变化幅度很大且酒体基质始终处于动态变化之中, 文章采用分阶段处理结合人工神经网络法对不同发酵阶段的糖度检测、监测近红外光谱模型的建立进行了探讨。不同建模方法的比较结果表明采用减去一条直线法预处理光谱, 阶段 建模光谱范围选择 $7\ 502 \sim 6\ 472.1\ \text{cm}^{-1}$, 阶段 建模光谱范围选择 $6\ 102 \sim 5\ 446.2\ \text{cm}^{-1}$ 时, 阶段 模型的 R^2 为 98.93%, RMSECV 为 $4.42\ \text{g} \cdot \text{L}^{-1}$; 阶段 模型的 R^2 为 99.34%, RMSECV 为 $1.21\ \text{g} \cdot \text{L}^{-1}$; 进一步对模型进行验证和评价, 结果表明阶段模型验证集的 RMSEP 为 $4.07\ \text{g} \cdot \text{L}^{-1}$; 阶段 模型验证集的 RMSEP 为 $1.13\ \text{g} \cdot \text{L}^{-1}$ 。本研究结果表明利用近红外光谱法建立的模型具有良好的预测效果, 能满足苹果酒工业生产中检测、监测精度要求。

关键词 近红外光谱; 苹果酒; 糖度; 人工神经网络; 偏最小二乘法

中图分类号: S37; TP216 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2009)03-0652-04

引言

苹果酒发酵过程中, 糖度是一个需要不断检测、监测的指标, 许多工艺步骤需要根据糖度的变化而进行调整; 糖度也是影响成品酒质量的重要理化指标之一, 其常规分析方法一般有如下两种: 一是菲林试剂直接滴定法, 另一个是高效液相色谱法, 这两种方法都操作繁琐, 费时费力。随着酿造技术的发展和质控要求的提高, 如何对其进行快速、准确的测定对降低生产成本、提高产品在线监测效率等都具有重要意义; 目前苹果酒酿造行业迫切需要引入更为先进的在线快速分析工具, 以便及时得到可靠的反馈数据, 用于指导生产过程。近红外光谱分析技术, 适合在线检测, 是一种能满足生产过程检测和监测的现代分析技术^[1], 在成品酒常规理化指标快速检测中已有应用报道^[2-6], 但是用于发酵过程的监测、检测的应用研究较少。在果酒发酵过程中, 一般糖度变化幅度较大, 从发酵起始的 $200\ \text{g} \cdot \text{L}^{-1}$ 左右下降到发酵结束时的 $4\ \text{g} \cdot \text{L}^{-1}$ 左右(干型酒要求残糖 $< 4\ \text{g} \cdot \text{L}^{-1}$ ^[7]), 并且由于发酵过程的不断进行, 酒体基质也始终处于动态变化之中, 从而导致利用常规方法构建的糖度近红外光谱检测模型的精度不高。鉴于此, 本研究通过将苹果酒发酵过程分段处理的思想来解析近红外光谱图, 选取了不同发酵阶段的建模

特征谱区, 比较了不同光谱数据处理方法对建模效果的影响, 旨在建立苹果酒发酵过程中糖度的近红外光谱法快速检测、监测模型, 期望为糖度的在线检测、监测提供一套高效、准确、快速的分析方法。

1 实验条件

1.1 样品制备

在苹果酒发酵过程中, 前 72 h(定义为阶段) 每 12 h 采集样品 1 个, 以后至发酵结束(定义为阶段) 每 24 h 采集样品 1 个, 每个发酵过程采集酒样 11 个。利用不同原料的苹果汁, 接种不同的酵母发酵 15 个批次, 共采集发酵过程中不同的酒样 165 个。然后分段归类不同酒样, 阶段 的 90 个样品归为一类, 用于建立发酵前 72 h 的糖度近红外检测模型(), 阶段 的 75 个酒样用于建立发酵 72 h 后的糖度近红外检测模型()。

1.2 糖度化学值测定

糖度化学值测定参照国家标准 GB/T 15038—2005, 采用菲林试剂直接滴定法^[7]。

1.3 光谱采集

利用德国布鲁克光学仪器公司 MPA TM 傅里叶变换近红外光谱仪采集光谱。InGaAs 检测器。选用光程为 2 mm 液

收稿日期: 2007-09-10, 修订日期: 2007-12-26

基金项目: 国家食品安全关键技术重大项目(2006BA K02A24), 国家自然科学基金项目(20806062)和西北农林科技大学青年学术骨干支持计划项目资助

作者简介: 彭帮柱, 1978 年生, 西北农林科技大学食品科学与工程学院讲师 e-mail: pengbangzhu@tom.com

通讯联系人 e-mail: ytl6503@tom.com

体光纤探头,以空气为参比,扫描波数为 12 000 ~ 4 000 cm^{-1} ,分辨率 8 cm^{-1} ,重复扫描次数 20 次,取平均光谱。

1.4 光谱波长范围选择及数据处理

由于不同官能团的基频、倍频及合频的特征吸收波长不同,建立近红外光谱定标模型时,需要筛选出与官能团定量关系最佳的特征谱段来建立数学模型^[8]。本试验中采用 OPUS-QUANT2 定量分析软件(德国布鲁克公司)的自动优化功能选择最佳特征波长范围和光谱预处理方法。光谱数据利用多元线性回归法(MLR)、主成分分析法(PCA)、偏最小二乘法(PLS)和人工神经网络法(ANN)进行计量学分析。其中,人工神经网络法应用误差反向传播神经网络(BP)算法进行计算,采用 Levenberg-Marquardt 法优化^[9],网络结构为 3 层,隐含层节点数为 2,首先输入建模谱段的近红外光谱数据,然后将光谱数据进行主成分变换,取前 7 个主成分作为神经网络的输入,建立模型。其中多元线性回归法(MLR)和主成分分析法(PCA)用数据处理软件 SAS8.0(美国 SAS 软件研究所)处理,偏最小二乘法(PLS)用数据处理软件 OPUS-QUANT2 处理,人工神经网络法(ANN)用 MATLAB 7.0(美国 MathWorks 公司)软件工具箱编程处理。

1.5 模型质量判断

定标模型质量的衡量指标为模型的决定系数(R^2)和交叉验证均方根差(RMSECV), R^2 越大,RMSECV 越小,模型质量越好^[10]。模型预测性能用预测集均方根差(RMSEP)来表示,预测标准偏差越小,模型预测能力越好。其中

$$R^2 = \left[1 - \frac{\sum_{i=1}^n (y_{Ti} - y_{Pi})^2}{\sum_{i=1}^n (y_{Ti} - y_a)^2} \right] \times 100\%$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_{Ti} - y_{Pi})^2}{n-1}}$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^m (y_{Ti} - y_{Pi})^2}{m}}$$

式中: n 为定标集样本个数; m 为验证集样本个数; y_{Ti} 为样

本的化学实测值; y_{Pi} 为样本的预测值; y_a 为样本化学实测值的平均值。

2 结果与分析

2.1 糖度化学实测值的分析

酒样糖度化学实测值分析结果见表 1,由校正集和验证集中糖度含量范围、平均值和变异系数可知,选取样品中的各组分含量范围较宽,代表性强,符合近红外光谱建模要求。

2.2 光谱预处理及波长选择

光谱预处理就是为了减小数据的共线性、维数、背景影响、噪声水平和光谱的基线漂移偏移而采取的光谱转换方法^[11]。转换的目的是消除背景干扰,使光谱数据更利于分析和提取有用的信息^[12]。建模时要选择最佳的特征波长范围建立模型,以消除系统误差,提高模型的预测精度。本试验中,定标集光谱利用 OPUS-QUANT2 定量分析软件(德国布鲁克公司)进行建模条件自动优化选择,结果表明最优条件为光谱经中心均值化后利用减去直线法(SLS)进行预处理,阶段 特征波长范围选择为 7 502 ~ 6 472.1 cm^{-1} ,阶段 特征波长范围选择为 6 102 ~ 5 446.2 cm^{-1} 。

2.3 定标模型的建立

模型建立的过程就是将适当预处理后的近红外光谱特征谱段与样品中相关组分含量进行关联,找出相关关系的一种方法^[10]。在建模过程中,模型的决定系数(R^2)偏高为好,内部交叉验证均方根差(RMSECV)偏小为好。本试验中不同数据处理方法建模质量的比较结果见表 2,由表 2 可知,利用人工神经网络法建立的糖度模型的 R^2 最大,RMSECV 最小,所以采用人工神经网络法建立定标模型。结果表明采用减去直线法预处理光谱,主成分为 7,光谱范围选 7 502 ~ 6 472.1 cm^{-1} (阶段)和 6 102 ~ 5 446.2 cm^{-1} (阶段)时,阶段 模型和阶段 模型的 R^2 最大,RMSECV 最小。其中,阶段 模型的 R^2 为 98.93%,RMSECV 为 4.42 $\text{g} \cdot \text{L}^{-1}$;阶段 模型的 R^2 为 99.34%,RMSECV 为 1.21 $\text{g} \cdot \text{L}^{-1}$ 。模型预测值和实测值相关图如图 1 和图 2 所示。

Table 1 Analysis for true values of sugar content in samples

	糖度(阶段)		糖度(阶段)	
	校正集	验证集	校正集	验证集
样品数/个	70	20	60	15
含量范围/($\text{g} \cdot \text{L}^{-1}$)	40.7 ~ 208	50.70 ~ 182.3	1.9 ~ 58.7	5.7 ~ 46.8
平均值/($\text{g} \cdot \text{L}^{-1}$)	92.8	78.9	31.2	27.3
变异系数/%	38.6	34.2	26.7	24.8

Table 2 Comparison of different data pretreatment methods for establishing calibration model

数据预处理方法	糖度(阶段)		糖度(阶段)	
	R^2 /%	RMSECV/($\text{g} \cdot \text{L}^{-1}$)	R^2 /%	RMSECV/($\text{g} \cdot \text{L}^{-1}$)
多元线性回归法	90.57	8.84	92.18	3.26
主成分分析法	92.13	7.10	93.35	2.19
偏最小二乘法	95.65	5.06	96.73	1.65
人工神经网络法	98.93	4.42	99.34	1.21

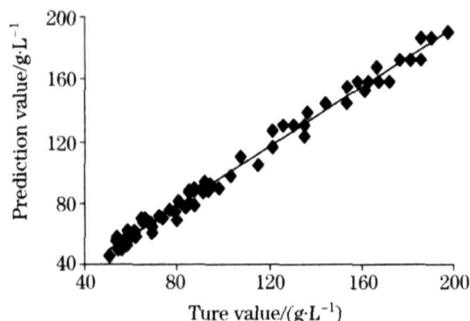


Fig. 1 Diagram between prediction and true values of sugar content in calibration set of stage I

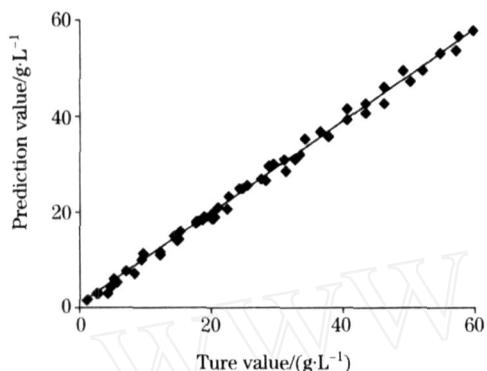


Fig. 2 Diagram between prediction and true values of sugar content in calibration set of stage II

2.4 模型的验证与评价

利用建立的定标模型预测验证集中不同阶段的样品糖度,验证集预测值和实测值的相关图(图 5 和图 6)可以看出,近红外光谱预测值与实测值基本一致。其中,阶段 I 模型验证集 RMSEP 为 $4.07 \text{ g} \cdot \text{L}^{-1}$,阶段 II 模型验证集的 RMSEP

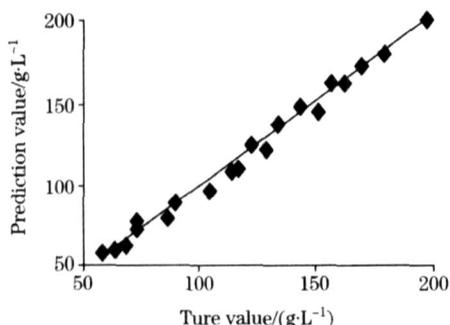


Fig. 3 Diagram between prediction and true values of sugar content in test set of stage I

为 $1.13 \text{ g} \cdot \text{L}^{-1}$,说明建立的定标模型的预测效果很好,能满足苹果酒生产中糖度的检测、监测要求。将预测值和实测值进一步通过配对 t 检验确认,在 0.05 显著性水平下,两种方法不存在显著性差异,表明近红外光谱法与化学方法检测的糖度值差异不显著,建立的定标模型具有良好的预测能力,可以达到常规分析方法的精度要求。

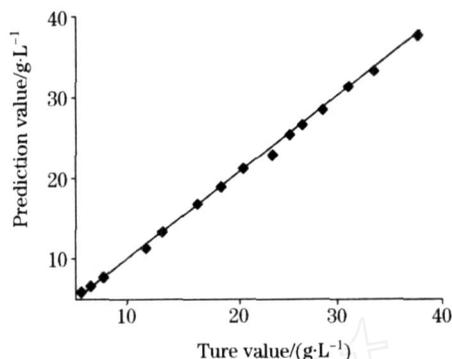


Fig. 4 Diagram between prediction and true values of sugar content in test set of stage II

3 结论与讨论

利用近红外光谱技术结合人工神经网络法建立了苹果酒发酵过程中不同阶段的糖度检测、监测的定量分析模型。其中,阶段 I 模型的 R^2 为 98.93%,RMSECV 为 $4.42 \text{ g} \cdot \text{L}^{-1}$;阶段 II 模型的 R^2 为 99.34%,RMSECV 为 $1.21 \text{ g} \cdot \text{L}^{-1}$;进一步对模型进行验证和评价,阶段 I 模型验证集的 RMSEP 为 $4.07 \text{ g} \cdot \text{L}^{-1}$,阶段 II 模型验证集的 RMSEP 为 $1.13 \text{ g} \cdot \text{L}^{-1}$,表明模型的预测效果很好,能满足苹果酒工业生产中检测、监测精度要求。

人工神经网络是近年来颇受关注的人工智能方法,它是解决非线性问题的有力工具^[13],它具有高度的并行性、良好的容错能力、自组织和自学习能力,处理复杂信息数据的能力是传统方法难以比拟的。在近红外光谱建立定标模型时,建模条件的选取一定要适当,即要保证预测具有较高的精度(RMSEP 要小),又要避免出现过拟合现象^[8]。选择最佳的特征波长范围,可以避免近红外光谱数据中存在的波峰重叠和共线现象等引起的误差,可以提高预测的全面性和预测的精度。近红外光谱技术在酒类工业中的应用研究已进入了快速发展期,除了建立常规指标的定量分析模型外,也可以用于酒类工业中原材料相关成分的检测和发酵过程的在线监控上,还可以为生产管理 with 酒质评价体系的建立等方面提供了重要的技术支持。

参 考 文 献

- [1] Blanco M, Villarroya I. Trends in Analytical Chemistry, 2002, 21(4) : 240.
- [2] Jos ÉL ús Moreira, L úcia Santos. Analytica Chimica Acta, 2004, 513: 263.
- [3] Urbano-Cuadrado M, Luque de Castro M D, Perez-Juan P M, et al. Analytica Chimica Acta, 2004, 527: 81.
- [4] Cozzolino D, Kwiatkowski M J, Parker M, et al. Analytica Chimica Acta, 2004, 513: 73.
- [5] PENG Bang-zhu, LONG Ming-hua, YUE Tian-li, et al(彭帮柱, 龙明华, 岳田利, 等). Transactions of Chinese Society Agricultural Engineering (CSAE) (农业工程学报), 2006, 22(12) : 216.
- [6] CHEN Bin, WANG Hao, LIN Song, et al(陈斌, 王豪, 林松, 等). Transactions of Chinese Society Agricultural Engineering (CSAE) (农业工程学报), 2005, 21(7) : 99.
- [7] GUO Xir-guang, MA Pei-xuan, REN Yi-ping(郭新光, 马佩选, 任一平). GB/ T 15038—2005. Analytical Methods of Wine and Fruit Wine (GB/ T 15038—2005 葡萄酒、果酒适用分析方法). Beijing: Chinese Standards Press(北京: 中国标准出版社), 2005.
- [8] LI Dai-xi, WU Zhi-yong, XU Duan-jun, et al(李代禧, 吴智勇, 徐端钧, 等). Chinese Journal of Analytical Chemistry (分析化学), 2004, 32(8) : 1070.
- [9] Fredric M Ham, Ivica Kostanic. Principles of Neurocomputing for Science and Engineering. Beijing: China Machine Press, 2003.
- [10] LU Wan-zhen, YUAN Hong-fu, XU Guang-tong, et al(陆婉珍, 袁洪福, 徐广通, 等). Modern Near Infrared Spectroscopy Analysis Technique(现代近红外光谱分析技术). Beijing: Chinese Press of Petroleum Chemical Industry(北京: 中国石化出版社), 2000.
- [11] Larrechi M, Callao M. Trends Anal. Chem., 2003, 22: 634.
- [12] XU Gang-tong, LU Wan-zhen, YUAN Hong-fu(徐广通, 陆婉珍, 袁洪福). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(4) : 459.
- [13] YUAN Cen-ren(袁曾任). Artificial Neural Networks and Applications(人工神经网络及应用). Beijing: Tsinghua University Press(北京: 清华大学出版社), 2003.

Modeling of Sugar Content Based on NIRS During Cider-Making Fermentation

PENG Bang-zhu, YUE Tian-li*, YUAN Ya-hong, GAO Zhen-peng

College of Food Science and Engineering, Northwest A & F University, Yangling 712100, China

Abstract The sugar content and the matrix always are being changed during cider-making fermentation. In order to measure and monitor sugar content accurately and rapidly, it is necessary for the spectra to be sorted. Calibration models were established at different fermentation stages based on near infrared spectroscopy with artificial neural network. NIR spectral data were collected in the spectral region of 12 000-4 000 cm^{-1} for the next analysis. After the different conditions for modeling sugar content were analyzed and discussed, the results indicated that the calibration models developed by the spectral data pretreatment of straight line subtraction (SLS) in the characteristic absorption spectra ranges of 7 502-6 472. 1 cm^{-1} at stage and 6 102-5 446. 2 cm^{-1} at stage were the best for sugar content. The result of comparison of different data pretreatment methods for establishing calibration model showed that the correlation coefficients of the models (R^2) for stage and were 98. 93 % and 99. 34 % respectively, and the root mean square errors of cross validation (RMSECV) for stage and were 4. 42 and 1. 21 $\text{g} \cdot \text{L}^{-1}$ respectively. Then the models were tested and the results showed that the root mean square error of prediction (RMSEP) was 4. 07 $\text{g} \cdot \text{L}^{-1}$ and 1. 13 $\text{g} \cdot \text{L}^{-1}$ respectively. These demonstrated that the models the authors established are very well and can be applied to quick determination and monitoring of sugar content during cider-making fermentation.

Keywords Near infrared spectroscopy; Cider; Sugar content; Artificial neural network; Partial least square

(Received Sep. 10, 2007; accepted Dec. 26, 2007)

* Corresponding author