

应用可见-近红外光谱技术进行白醋品牌和 pH 值的快速检测

王莉, 刘飞, 何勇*

浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029

摘要 提出了一种基于可见-近红外透射光谱技术快速判别白醋品牌和测定 pH 值的方法。应用可见-近红外透射光谱获取不同品牌白醋的透射光谱曲线, 并对获得的原始光谱数据进行平滑、变量标准化以及一阶导数等预处理, 然后利用主成分分析对原始光谱数据进行聚类分析, 根据主成分的累计贡献率选取主成分数, 并将所选取的主成分作为三层 BP 神经网络的输入。通过定标集样本对 BP 神经网络进行训练, 得到三层优化神经网络结构: 5 输入层节点, 6 隐含层节点和 2 输出层节点, 各层传递函数均采用 Sigmoid 函数。利用该模型对预测集样本进行预测。实验结果表明在阈值设定为 ± 0.1 的情况下该模型对预测集样本品牌鉴别准确率达到 100%, pH 预测值与实际测量值偏差小于 5%, 得到了理想的结果。所以利用可见-近红外光谱技术结合主成分分析和神经网络算法能够快速准确的判定白醋品牌和 pH 值。

关键词 可见-近红外光谱; 白醋; pH; 主成分分析; BP 神经网络

中图分类号: TS264.2 **文献标识码:** A **文章编号:** 1000-0593(2008)04-0813-04

引言

食醋是人们日常生活中不可或缺的调味佳品, 它不仅具有很高的食用价值, 而且还有广泛的保健治疗、杀菌消毒等功效。白醋适合于西式菜肴和淡色菜肴的调味, 目前市场的白醋可分为酿造白醋和配制白醋, 其中配制白醋采用乙酸稀释, 不具有营养价值, 很少用于食用。目前国内白醋品牌繁多, 不乏假冒伪劣产品, 而常规通过化学分析的方法区别品牌费时费力。此外, pH 作为醋的一项基本指标, 对醋的 pH 值进行快速、简单、准确测定无论对于生产厂家还是消费者来说都具有非常重要的意义^[1]。

现代近红外光谱分析技术, 可充分利用全谱或多波长下的光谱数据进行定性或定量分析。由于近红外光谱技术分析具有速度快、效率高、成本低、测试重现性好、测量方便等特点, 已经被越来越多地应用于食品工业、石油化工、制药工业等领域^[2-6]。在定性分析中, 有学者研究利用近红外光谱技术鉴别茶叶真伪^[7]、杨梅品种^[8]、葡萄酒产地^[9]以及中药材产地^[10]等; 在组分测定方面, 有利用该技术测定食用醋中还原糖含量^[11]、烟草中总糖含量^[12]、葡萄酒化学成分测定^[13]、啤酒酒精度测定^[14]以及白酒酒精度测定^[15]等。

主成分分析是多元统计中的一种数据挖掘技术。在不丢

失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量, 解决了由于谱带的重叠而无法分析的困难。BP 神经网络是一种基于误差反向传播算法的神经网络, 是一种强有力的学习系统, 具有高度非线性映射的能力。由于此种模型可以逼近任何连续的非线性曲线, 现已在各领域得到了广泛的应用, 取得了令人满意的结果^[16], 目前已经有研究将主成份分析和神经网络应用于光谱分析^[17-20]。

我们选用主成分分析 (PCA) 和基于误差反向传播算法 (Back Propagation, BP) 多层前馈神经网络建立不同品牌白醋的近红外光谱预测模型。采用一系列光谱预处理方法消除噪音, 并通过主成分分析得到主成分的累计贡献率和各个主成分的得分, 将选出的主要主成分得分作为多层前馈神经网络的输入, 对未知样本品牌和 pH 值的预测作为输出, 建立 BP 神经网络模型。

1 材料与方法

1.1 仪器设备

实验使用美国 ASD (analytical spectral device) 公司的 Handheld FieldSpec 光谱仪, 其光谱采样间隔 (波段宽) 1.5 nm, 测定范围 325 ~ 1 075 nm, 扫描次数 30 次, 分辨率 3.5 nm, 探头视场角为 20°。使用上海大普仪器有限公司生产的

收稿日期: 2006-09-16, 修订日期: 2006-12-16

基金项目: 国家科技支撑项目 (2006BAD10A04), 国家自然科学基金项目 (30671213) 和高等学校优秀青年教师教学科研奖励计划项目 (02411) 资助

作者简介: 王莉, 1983 年生, 浙江大学生物系统工程与食品科学学院硕士研究生 *通讯联系人 e-mail: yhe@zju.edu.cn

PHS-4CT 型酸度计, pH 测量范围 0~14, 分辨率 0.001, 输出电压范围 -1 999.9~1 999.9 mV。光谱数据以 ASCII 码形式导出进行处理, 分析软件为 ASD View Spec Pro, Unscramble V9.6 和 DPS (data procession system for practical statistics)。

1.2 样品来源及光谱的获取

从超市选得三种典型白醋, 恒顺白醋(江苏镇江)、马大嫂白醋(浙江杭州)、上海鼎丰白醋(上海)共计 240 个样本, 每个品牌 80 个样本。样本随机分成定标集和预测集, 定标集设置为 180 个, 每个品牌 60 个; 预测集设置为 60 个, 每个品牌 20 个。实验在恒温 25℃ 下进行, 采用透射光谱法, 光程 2 mm。光谱仪置于样品池正上方, 对每一个样品扫描 30 次, 从其中取 3 幅图谱, 并将其取平均, 由此得到一个样本光谱。

1.3 光谱数据预处理

为了去除来自高频随机噪音、基线漂移、样本不均匀、光散射等影响, 需要进行光谱预处理来消除噪音。采用 Moving average smoothing 平滑, 选用平滑点数为 3, 此时能很好滤除各种因素产生的高频噪音; 再进行 SNV (standard normal variate) 处理^[21]。

2 实验结果与分析

2.1 不同品牌白醋的可见-近红外光谱图

三种白醋的可见-近红外光谱如图 1 所示(为图形显现清晰, 每个品牌的白醋随意选取 5 条光谱曲线)。图 1 中横坐标为波长范围 325~1 075 nm, 纵坐标为光谱吸光度。从图中可以看出, 不同品牌白醋的光谱曲线交错重叠, 较难直观地看出品牌间的差别, 必须要结合化学计量学方法对所得的光谱数据进行处理, 从而建立起白醋品牌判别及 pH 值检测模型。

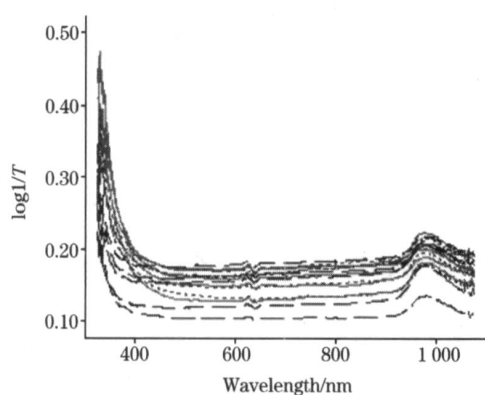


Fig 1 Vis/ NIR absorbance transmittance spectra of three varieties white vinegar

1: Hengshun; 2: Madasao; 3: Shanghai dingfeng

2.2 不同品牌白醋的聚类分析

从图 1 中可以看到光谱曲线的首端和末段都有较大的噪音, 为消除这些影响, 选取 350~1 050 nm 波段的光谱进行研究。实验采用主成分分析法对三个品牌共 240 个样本白醋进行分析。图 2 中 X 轴表示样本的第一主成分得分 (PC1),

Y 轴表示样本第二主成分得分 (PC2), Z 轴表示样本第三主成分得分 (PC3)。图 2 显示了不同品牌白醋聚类效果, 定性的描述了三种白醋的特征差异。说明主成分分析是品牌鉴别的一种有效方法, 但是边缘样品区分不明显。为达到更高的预测精度, 实验中用 BP 神经网络结合主成分分析建立白醋品牌鉴别及其 pH 值检测分析模型。

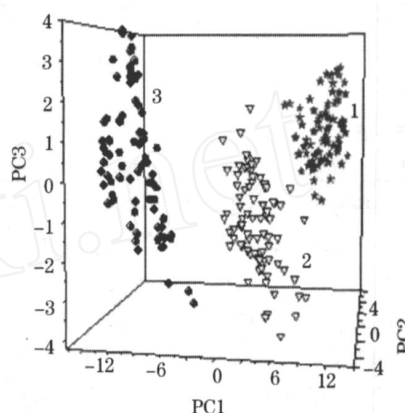


Fig 2 Principal component scores stereonet (PC1 × PC2 × PC3) of three varieties white vinegar
1: Hengshun; 2: Madasao; 3: Shanghai dingfeng

经主成分分析光谱数据后得到前 10 个主成分累计贡献率如表 1 所示, 前 5 个主成分的累计贡献率已经达到了 99.678%, 而且这之后, 随着主成分数的增加, 累计贡献率增加相当缓慢(低于 0.1%), 所以每个样本的光谱数据可以用 5 个主成分代替。

Table 1 Accumulative reliabilities (AR) of principal components (PCs)

主成分	PC1	PC2	PC3	PC4	PC5
累计贡献率	92.348 %	95.667 %	98.783 %	99.133 %	99.678 %
主成分	PC6	PC7	PC8	PC9	PC10
累计贡献率	99.723 %	99.773 %	99.802 %	99.826 %	99.842 %

2.3 基于神经网络的白醋品牌鉴别及 pH 值预测模型

将 5 个主成分的得分输入神经网络, 通过反复训练得到三层最佳网络结构: 5 输入层节点, 6 隐含层节点和 2 输出层节点, 目标误差设定为 0.000 1, 限定最大迭代次数为 1 000。表 2 为预测结果, 拟合误差为 5.05×10^{-3} , 设定品牌预测结果偏差在 ± 0.1 内为预测正确。从表 2 可以看出, 利用该模型对白醋品牌的鉴别率可以达到 100%, 对 pH 值的预测最大偏差为 4.79×10^{-2} 。

3 结 论

应用光谱技术的关键就是如何从大量重叠的光谱数据中提取有用的信息, 建立预测效果优秀的模型。本文中应用主成分分析和 BP 神经网络建立预测模型, 对白醋预测样本品牌鉴别率达到了 100%, pH 值的预测误差也控制在 5% 范围内。说明所以该模型能够对白醋的种类及其 pH 值作出快速

准确的判定。也说明了可见-近红外光谱技术对白醋品牌判
别以及组分预测是可行的,为今后开发相应的仪器设备奠定

Table 2 Prediction result for 60 unknown white vinegar samples by BP-ANN model

样本号	实际值		预测值		pH 偏差($\times 10^{-3}$)	样本号	实际值		预测值		pH 偏差($\times 10^{-3}$)
	品牌	pH	品牌	pH			品牌	pH	品牌	pH	
(1)	1	2.603	1.004	2.601	0.768	(31)	2	3.135	1.994	3.157	7.02
(2)	1	2.617	1.004	2.599	6.88	(32)	2	3.143	2.000	3.135	2.55
(3)	1	2.615	1.004	2.598	6.50	(33)	2	3.135	2.001	3.163	8.93
(4)	1	2.609	1.004	2.595	5.37	(34)	2	3.137	1.999	3.160	7.33
(5)	1	2.600	1.004	2.597	1.15	(35)	2	3.138	2.001	3.162	7.65
(6)	1	2.601	1.004	2.593	3.08	(36)	2	3.138	1.998	3.152	4.46
(7)	1	2.602	1.004	2.596	2.31	(37)	2	3.135	1.997	3.175	1.28
(8)	1	2.584	1.008	2.601	6.58	(38)	2	3.142	2.007	3.139	0.955
(9)	1	2.592	1.005	2.603	4.24	(39)	2	3.142	1.996	3.162	6.37
(10)	1	2.594	1.004	2.593	0.386	(40)	2	3.139	1.996	3.129	3.19
(11)	1	2.594	1.007	2.595	0.386	(41)	3	2.592	2.985	2.575	6.56
(12)	1	2.591	1.019	2.662	27.4	(42)	3	2.595	2.985	2.571	9.25
(13)	1	2.587	1.058	2.711	47.9	(43)	3	2.592	2.985	2.576	6.17
(14)	1	2.591	1.007	2.619	10.8	(44)	3	2.589	2.986	2.577	4.63
(15)	1	2.591	1.009	2.623	12.4	(45)	3	2.593	2.985	2.574	7.33
(16)	1	2.585	1.014	2.638	20.5	(46)	3	2.587	2.985	2.564	8.89
(17)	1	2.580	1.021	2.652	27.9	(47)	3	2.569	2.985	2.573	1.56
(18)	1	2.597	1.038	2.675	30.0	(48)	3	2.590	2.985	2.574	6.18
(19)	1	2.585	1.027	2.664	30.6	(49)	3	2.582	2.986	2.577	1.94
(20)	1	2.583	1.038	2.669	33.3	(50)	3	2.594	2.985	2.574	7.71
(21)	2	3.138	2.028	3.111	8.60	(51)	3	2.590	2.985	2.568	8.49
(22)	2	3.142	1.998	3.164	7.00	(52)	3	2.595	2.985	2.533	23.9
(23)	2	3.139	2.002	3.165	8.28	(53)	3	2.594	2.985	2.574	7.71
(24)	2	3.138	2.002	3.145	2.23	(54)	3	2.585	2.985	2.562	8.90
(25)	2	3.142	1.997	3.167	7.96	(55)	3	2.596	2.986	2.578	6.93
(26)	2	3.141	2.001	3.155	4.46	(56)	3	2.588	2.985	2.572	6.18
(27)	2	3.140	2.000	3.160	6.37	(57)	3	2.590	2.985	2.575	5.79
(28)	2	3.141	2.002	3.151	3.18	(58)	3	2.588	2.985	2.573	5.80
(29)	2	3.136	1.996	3.134	0.638	(59)	3	2.591	2.982	2.568	8.88
(30)	2	3.140	2.004	3.146	1.91	(60)	3	2.593	2.988	2.577	6.17

注意: 1- 恒顺白醋; 2- 马大嫂白醋; 3- 上海鼎丰白醋

参 考 文 献

- [1] SHEN Zhi-yuan(沈志远). Food Science(食品科学), 2005, 26(8): 483.
- [2] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Analytical Instrumentation(分析仪器), 2006, (2): 1.
- [3] WANG Duo-jia, ZHOU Xiang-yang, JIN Tong-ming, et al(王多加, 周向阳, 金同铭, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(4): 447.
- [4] CHEN Hai-yan, HE Yong. Trends in Food Science & Technology, 2007, 18(2): 72.
- [5] SHI Su-jia(史苏佳). Food Research and Development(食品研究与开发), 2007, 28(3): 177.
- [6] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). The Foundation and Application of the Near Infra-red Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [7] CHEN Quan-sheng, ZHAO Jie-wen, ZHANG Hai-dong, et al(陈全胜, 赵杰文, 张海东, 等). Acta Optica Sinica(光学学报), 2006, 26(6): 933.
- [8] HE Yong, LI Xiao-li(何 勇, 李晓丽). Journal of Infrared and Millimeter Waves(红外与毫米学报), 2006, 25(3): 192.
- [9] Liu L, Cozzolino D, Cynkar W U, et al. Journal of Agricultural and Food Chemistry, 2006, 54(18): 6754.
- [10] LIU Shu-hua, ZHANG Xue-gong, ZHOU Qun, et al(刘沐华, 张学工, 周 群, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(4): 629.
- [11] CHEN Bin, HUANG Chuan-xu, LI Guo-quan, et al(陈 斌, 黄传旭, 李国权, 等). Transactions of the Chinese Society of Agricultural

- Machinery(农业机械学报), 2005, 36(5): 65.
- [12] CHEN Da, WANG Fang, SHAO Xue-guang, et al(陈 达, 王 芳, 邵学广, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(6): 672.
- [13] Urbano-Cuadrado M, Luque de Castro M D, Perez-Juan P M, et al. Analytica Chimica Acta, 2004, 527(1): 81.
- [14] CHEN Bin, WANG Hao, LIN Song, et al(陈 斌, 王 毫, 林 松, 等). Transactions of the Chinese Society of Agriculture Engineering(CASE)(农业工程学报), 2005, 21(7): 99.
- [15] WANG Zhen-zuo, LU Jia-hui, GUO Wei-liang, et al(王贞佐, 逯家辉, 郭伟良, 等). Science and Technology of Food Industry(食品工业科技), 2005, 26(11): 163.
- [16] BAO Yi-dan, WU Yan-ping, HE Yong(鲍一丹, 吴燕萍, 何 勇). Journal of Agricultural Mechanization Research(农机化研究), 2004, (3): 162.
- [17] HE Yong, LI Xiao-li, SHAO Yong-ni(何 勇, 李晓丽, 邵咏妮). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(5): 850.
- [18] HUANG Min, HE Yong, HUANG Ling-xia, et al(黄 敏, 何 勇, 黄凌霞, 等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2006, 25(5): 342.
- [19] ZHAO Mei-fang, LUO A-li, WU Fu-chao, et al(赵梅芳, 罗阿理, 吴福朝, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(2): 377.
- [20] HE Yong, FENG Shui-juan, LI Xiao-li, et al(何 勇, 冯水娟, 李晓丽, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2006, 26(11): 2021.
- [21] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.

Fast Detection of White Vinegar Varieties and pH by Vis/ NIR Spectroscopy

WANG Li, LIU Fei, HE Yong *

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

Abstract White vinegar is a condiment indispensable in our life, but our understanding of the white vinegar and evaluation of its quality and function has been gained through routine chemical and physical analysis. It is called for to develop more time- and cost-efficient methodologies for white vinegar detection. Visible and near infrared spectroscopy (Vis/ NIR) is a nondestructive, fast and accurate technique for the measurement of chemical components based on overtone and combination bands of specific functional groups. Vis/ NIR transmittance spectroscopy and chemometrics methods were utilized in classification and pH mensuration of white vinegar in the present study. First, the spectral curves of white vinegar were obtained by handheld Vis/ NIR spectroradiometer, then principal component analysis (PCA) was used to process the spectral data after pretreatment. Five principal components (PCs) were selected based on accumulative reliabilities (AR), and these selected PCs would be taken as the inputs of the three-layer back-propagation artificial neural network (BP-ANN). A total of 240 white vinegar samples were divided into calibration set and validation set randomly, the calibration set had 180 samples with 60 samples of each variety, and the validation set had 60 samples with 20 samples of each variety. The BP-ANN was trained using samples in calibration set, the optimal three-layer BP-ANN model with 5 nodes in input layer, 6 nodes in hidden layer, and 2 nodes in output layer would be obtained, and the transfer function of sigmoid was used in each layer. Then, this model was used to predict the samples in the validation set. The result indicated that a 100% recognition ration was achieved with the threshold predictive error ± 0.1 , the bias between predictive value and standard value was lower than 5%. It could be concluded that PCA combined with BP-ANN was an available method for varieties recognition and pH mensuration of white vinegar based on Vis/ NIR transmittance spectroscopy.

Keywords Vis/ NIR spectroscopy; White vinegar; pH; Principal component analysis; BP neural networks

(Received Sep. 16, 2006; accepted Dec. 16, 2006)

* Corresponding author