

文章编号: 0253-2239(2010)02-0574-05

应用可见-近红外光谱快速识别沙棘汁品牌

张海红 张淑娟 王凤花 介邓飞 赵华民

(山西农业大学工学院, 山西 太谷 030801)

摘要 为了实现沙棘汁品牌的快速无损鉴别,提出了采用可见-近红外光谱分析技术(NIR)鉴别沙棘汁品牌的方法。采用 FieldSpec3 光谱仪对三种沙棘汁进行光谱分析,各获取 40 个样本数据。采用平均平滑法和多元散射校正(MSC)方法对样本数据进行预处理,再用主成分法(PCA)对光谱数据进行聚类分析并获得各主成分数据。将 120 个沙棘汁样本随机分成 90 个建模样本和 30 个预测样本,把基于累计可信度选择的建模样本的 8 个主成分(PCs)数据作为 BP 网络的输入变量,沙棘汁品牌作为输出变量,建立三层反向传播(BP)神经网络鉴别模型,并对 30 个预测样本进行预测。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%。所以应用近红外光谱技术结合主成分分析和 BP 神经网络算法识别沙棘汁品牌是一种有效的方法。

关键词 可见-近红外光谱;主成分分析;人工神经网络;品牌;沙棘汁

中图分类号 S123 文献标识码 A doi: 10.3788/AOS20103002.0574

Study on Fast Discrimination of Seabuckthorn Juice Varieties Using Visible-Nir Spectroscopy

Zhang Haihong Zhang Shujuan Wang Fenghua Jie Dengfei Zhao Huamin

(College of Engineering, Shanxi Agricultural University, Taigu, Shanxi 030801 China)

Abstract In order to achieve non-destructive variety identification of seabuckthorn juice, a fast discrimination method based on visible-near infrared reflectance (NIR) spectroscopy was put forward. A Field Spec 3 spectroradiometer was used for collecting 40 sample spectral data of three varieties of seabuckthorn juice separately. Average smoothing method and multiplicative scattering correction (MSC) method were used to complete the pretreatment of sample data. Then principal component analysis (PCA) was used to process the spectral data after pretreatment. A total of 120 seabuckthorn juice samples were divided into calibration set and validation set randomly, the calibration set had 90 samples and validation set had 30 samples. Eight principal components (PCs) were selected based on accumulative reliabilities which would be taken as the inputs of the three-layer back-propagation neural network, and seabuckthorn juice varieties were selected as the outputs of back propagation (BP) neural network. Then this model was used to predict 30 samples in the validation set. The result showed that a 100% recognition ratio was achieved with the threshold predictive error ± 0.1 . It could be concluded that PCA combined with BP neural network was an available method for varieties recognition of seabuckthorn juice based on NIR spectroscopy.

Key words visible-NIR spectroscopy; principal component analysis (PCA); BP neural network; variety; seabuckthorn juice

1 引言

近红外漫反射光谱分析技术,可充分利用全波段下的光谱数据进行定性和定量分析。由于光谱技

术具有无需预处理、速度快、无污染、无破坏、多成分同时分析、结果重现性好和适于在线分析等优点,被广泛应用于农产品品质及品种的鉴别研究^[1~14]。

收稿日期: 2009-02-23; 收到修改稿日期: 2009-04-25

基金项目: 山西省科技攻关项目(2007031109-2)资助课题。

作者简介: 张海红(1983—),女,硕士研究生,主要从事数字农业技术与装备等方面的研究。E-mail: fyzh888@163.com

导师简介: 张淑娟(1963—),女,教授,博士生导师,主要从事数字农业技术与装备等方面的研究。

E-mail: zsujuan@263.net(通信联系人)

目前,国内已有学者利用近红外光谱技术对酱油品牌^[1]、杨梅汁品种^[2]和茶叶品种^[3]等进行研究,但未见有人对沙棘汁品牌进行鉴别研究。沙棘汁作为一种绿色饮料,其口感、糖分含量、酸度和维生素含量等内部品质,都得到了人们广泛的重视,从目前市场看,品牌繁多,其口感和品质差异也较显著。因此,针对沙棘汁的品质检测技术的发展和无疑是非常重要的,尤其是研究一种简单、快速、无损的沙棘汁品牌鉴别技术更是非常必要的。反向传播(BP)神经网络是一种基于误差反向传播算法的强有力的学习系统,具有高度非线性映射能力,可很好解决复杂的非线性模式识别问题,但在应用中关于隐含层节点数确定的问题还没有很好地解决^[1,2]。主成分分析法(PCA)可在不丢失主要信息的前提下,选择较少的新变量来代替原来较多的变量,解决由于谱带的重叠而无法分析的困难,是目前应用较广泛的光谱分析数学方法。引入BP神经网络中隐含层节点数定量快速确定的方法,采用主成分分析和BP神经网络相结合的方法建立不同品牌沙棘汁的可见-近红外光谱预测模型,以实现沙棘汁品牌的鉴别。

2 实验部分

2.1 可见-近红外光谱数据的获取

实验装置由计算机、光谱仪、卤光灯和校正白板等组成。使用美国 Analytical Spectral Device (ASD)公司的 FieldSpec 3 光谱仪,光谱采样间隔为 1 nm,采样范围为 350~2500 nm,扫描次数 30 次,探头视场角为 25°,采用漫反射方式进行样品光谱采样;光源采用与光谱仪配套的 14.5 V 卤素灯。光谱数据以 ASCII 码形式导出进行处理,分析软件为 ASD View Spec Pro V5.0,Unscramble V9.7 和数据处理系统(DPS)。

对从市场购买的同批次同一日期生产的三个品牌的沙棘汁:夏普赛尔沙棘汁、维仕杰沙棘汁和汇源沙棘汁各 40 个样本,合计 120 个样本,随机分为建模集 90 个样本和预测集 30 个样本。将沙棘汁装入高 1.4 cm,直径 6.5 cm 的培养皿内,液面高度 10 mm。再将光谱仪探头置于沙棘汁样本的上方,距沙棘汁表面 50 mm,与液面成 45°固定。光源距盛样容器中心 350 mm,与液面成 45°照射。三种沙棘汁的典型可见-近红外漫反射光谱曲线见图 1。

2.2 光谱数据预处理

为了去除来自高频随机噪音、基线漂移、样本不均匀和光散射等影响,采用平滑法进行光谱预处理,

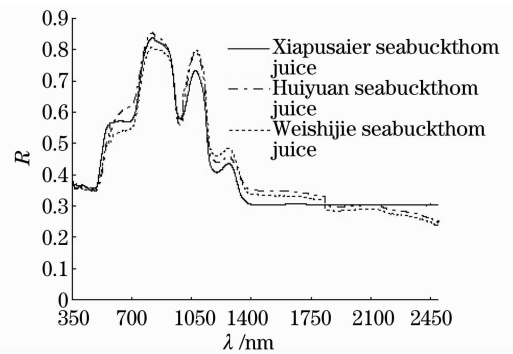


图 1 三种不同品牌沙棘汁的可见-近红外漫反射光谱曲线

Fig. 1 Visible-near infrared diffuse reflectance spectra of three different varieties of Seabuckthorn juice

选用平滑点数为 9,可很好滤除各种因素产生的高频噪音,再进行多元散射校正(MSC)处理^[3]。

3 结果与分析

3.1 光谱曲线分析

三个品牌沙棘汁的典型近红外漫反射光谱曲线如图 1 所示。图 1 中横坐标为波长,范围是 350~2500 nm,纵坐标为光谱反射率。从图 1 中可以看出,不同品牌沙棘汁的光谱图不同,具有一定的特征性和指纹性。应用 ASD ViewSpec Pro V5.0 软件将所测样本的光谱反射率数据取平均值并转换成 ASCII 码导出,再通过 Unscramble V9.7 软件对其进行主成分分析。

3.2 沙棘汁的主成分定性聚类分析

主成分分析的目的是将数据降维,以消除众多信息共存中相互重叠的信息部分,通过对原始大量光谱变量进行转换,使数目较少的新变量成为原变量的线性组合,而且新变量能最大限度的表征原变量的数据结构特征^[1,2]。利用 Unscramble V9.7 软件,对三种沙棘汁的 90 个样本进行主成分分析聚类结果见图 2。(A₁: 汇源沙棘汁, A₂: 夏普赛尔沙棘汁, A₃: 维仕杰沙棘汁)

图 2 表示 90 个建模样本的主成分 1,2 得分图,图中横坐标表示每个样本的第一主成分得分值,纵坐标表示每个样本的第二主成分得分值。图 2 中“夏普赛尔沙棘汁”、“维仕杰沙棘汁”和“汇源沙棘汁”明显分成 3 类,说明主成分 1,2 对 3 种沙棘汁有较好的聚类作用。从图 2 可以看出,汇源沙棘汁的 30 个样本聚合度较好,分布于第二象限并聚合在 X 轴附近。维仕杰沙棘汁的 30 个样本聚合度较好,大多分布于第三象限并聚合在 Y 轴附近。夏普赛尔沙棘汁其样本聚合度不如前两种,但也分布在第一、

四象限。这三种沙棘汁基本没有重叠,能够区分开。分析表明前 2 个主成分对 3 种沙棘汁有一定的聚类作用,能定性区别不同品牌的沙棘汁,但不能给出定量的鉴别模型,因此,考虑建立沙棘汁品牌鉴别的基于多个主成分的 BP 神经网络模型。

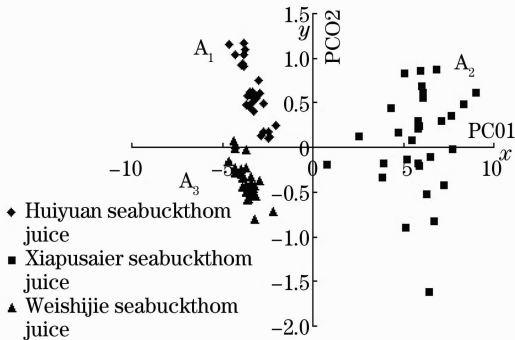


图 2 90 个沙棘汁的主成分 1 和主成分 2 得分图

Fig. 2 PCA scores plots (PC1×PC2) for Seabuckthorn juice sample across the entire spectral region

表 1 前 8 个主成分的累计可信度

Table 1 Accumulative reliabilities of the first 8 PCs

No. of PC	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Accumulative reliability /%	97.997	99.315	99.720	99.877	99.965	99.981	99.986	99.992

3.3.2 基于 BP 神经网络的沙棘汁品牌鉴别

把由 PCA 提取的 8 个主成分(分别选取了 2~8 个主成分的得分建模分析比较,取 8 个主成分的得分预测精度最高)的得分作为 BP 神经网络的输入变量建立鉴别模型,因而确定网络输入层节点数为 8,再通过调整隐含层的节点数来优化网络结构。依据隐含层节点数计算公式^[10]

$$n_1 = \sqrt{n+m} + a, \quad (1)$$

式中 n_1 为隐含层节点数, n 为输入层节点数, m 为输出层节点数, a 为 1~10 之间的常数。隐含层节点数计算如下:

$$\sqrt{8+1} + 1 \leq n_1 \leq \sqrt{8+1} + 10 \text{ 即: } 4 \leq n_1 \leq 13.$$

根据 n_1 的计算值,由小到大改变节点数训练并检验其精度,当随着节点数的增加误差不再进一步减小时,其临界值即为应采用的数值。经依次实验后(结果见表 2),确定最佳隐含层节点数为 6,输出

3.3 沙棘汁品牌鉴别的主成分分析和 BP 神经网络建模

3.3.1 基于主成分分析的主成分数提取

样本光谱波段从 350~2500 nm 共有 2150 个点,采用全谱段计算时,计算量大,而且由于有些区域样品的光谱信息很弱,与样品的组成或性质间缺乏相关关系。所以通过 PCA 分析,选取对于不同沙棘汁敏感的新变量作为输入建立神经网络品牌鉴别模型。

8 个主成分的累计可信度见表 1。累计可信度表示主成分对原始变量的解释程度。前 8 个主成分的累计可信度达到 99.99%,表示这 8 个主成分能够解释原始波长变量的 99.99%。

表 1 表明主成分分析是一种非常有效的数据挖掘方法,它把原来的 2000 多个波长变量压缩成了彼此正交的 8 个新变量,这 8 个新变量彼此间互不影响,而且能代表绝大部分原变量包含的信息。

层节点数为 1(品牌值)。网络设定训练迭代次数为 1000 次,目标误差设定为 0.0001。随机选取 90 个样品为建模集,30 个样品为预测集。对 90 个建模样本的拟合残差为 2.95×10^{-4} 。对未知的 30 个样本进行预测,预测结果见表 3。在阈值设定为 ± 0.1 的情况下,未知样品的品牌识别正确率达到了 100%。BP 神经网络分析由 DPS 软件完成。

表 2 BP 神经网络模型中隐含层节点数和拟合残差

Table 2 Hidden layer nodes and fitting residual of BP neural network model

Hidden layer node	Fitting residual / 10^{-4}	Hidden layer node	Fitting residual / 10^{-4}
4	6.10	9	3.82
5	6.48	10	7.42
6	2.95	11	8.33
7	3.39	12	6.65
8	3.44	13	6.01

表 3 BP 神经网络模型对未知样品的预测结果

Table 3 Prediction results for unknown samples by BP neural network model

Sample number	Real value	Predicted value	Bias	Sample number	Real value	Predicted value	Bias
1	1	1.0041	0.0041	16	2	1.9801	-0.0199
2	1	1.0009	0.0009	17	2	2.0038	0.0038
3	1	1.0007	0.0007	18	2	1.9545	-0.0455
4	1	1.0017	0.0017	19	2	2.0836	0.0836

(续表)

Sample number	Real value	Predicted value	Bias	Sample number	Real value	Predicted value	Bias
5	1	1.0155	0.0155	20	2	2.0065	0.0065
6	1	1.0155	0.0155	21	3	2.9968	-0.0032
7	1	1.0871	0.0871	22	3	2.9950	-0.0050
8	1	1.0623	0.0623	23	3	2.9962	-0.0038
9	1	1.0572	0.0572	24	3	2.9900	-0.0100
10	1	1.0175	0.0175	25	3	2.9951	-0.0049
11	2	1.9192	-0.0808	26	3	2.9901	-0.0099
12	2	1.9574	-0.0426	27	3	2.9926	-0.0074
13	2	1.9452	-0.0548	28	3	2.9795	-0.0205
14	2	1.9591	-0.0409	29	3	2.9925	-0.0075
15	2	1.9752	-0.0248	30	3	2.9927	-0.0073

Note: 1-Huiyuan Seabuckthom juice; 2-Xiapusaier Seabuckthom juice; 3-Weishijie Seabuckthom juice

4 结 论

在对三种沙棘汁的漫反射光谱数据采集和分析的基础上,采用主成分分析法和 BP 神经网络相结合,建立了沙棘汁品牌鉴别模型,该模型的预测效果很好,对未知沙棘汁样品的预测偏差在±0.1 以下,识别率达到 100%,说明了运用可见-近红外光谱可以快速、准确的对沙棘汁品牌进行鉴别。将从 PCA 分析中得到的主成分作为 BP 神经网络的输入,不但减少了神经网络的计算量,加快了训练速率,而且因为去除了光谱干扰信息,也提高了 BP 神经网络预测的正确率。在 BP 神经网络建模过程中引入隐含层节点数定量快速确定的方法,也大大节约了建模时间。

参 考 文 献

- 1 Tong Xiaoling, Bao Yidan, He Yong. Study on fast discrimination of soy sauce using near infrared spectra [J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(3): 597~601
- 童晓星, 鲍一丹, 何勇. 应用近红外光谱技术快速检测酱油品牌的研究[J]. *光谱学与光谱分析*, 2008, **28**(3): 597~601
- 2 Cen Haiyan, Bao Yidan, He Yong. Fast discrimination of varieties of bayberry juice based on spectroscopy technology [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(3): 503~506
- 岑海燕, 鲍一丹, 何勇. 基于光谱技术的杨梅汁品种快速鉴别方法的研究[J]. *光谱学与光谱分析*, 2007, **27**(3): 503~506
- 3 Li Xiaoli, He Yong, Qiu Zangjun. A new method to fast discrimination of tea varieties using visible/ near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(2): 279~282
- 李晓丽, 何勇, 裴正军. 一种可见-近红外光谱快速鉴别茶叶品种的新方法[J]. *光谱学与光谱分析*, 2007, **27**(2): 279~282
- 4 Liu Yande, Sun Yandong, Chen Xiangmiao. Research on the soluble solids content of pear internal quality index by near infrared diffuse reflectance spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(4): 797~800
- 刘燕德, 孙旭东, 陈兴苗. 近红外漫反射光谱检测梨内部指标可溶

- 性固形物的研究 [J]. *光谱学与光谱分析*, 2008, **28**(4): 797~800
- 5 Liu Yande, Luo Ji, Chen Xingmiao. Analysis of soluble solid content in nanfeng tangerine with visible near infrared spectroscopy [J]. *J. Infrared Millim. Waves*, 2008, **27**(2): 119~122
- 刘燕德, 罗吉, 陈兴苗. 可见/近红外光谱的南丰蜜桔可溶性固形物含量定量分析 [J]. *红外与毫米波学报*, 2008, **27**(2): 119~122
- 6 Zhao Jiewen, Zhang Haidong, Liu Muhua. Non-destructive determination of sugar contents of apples using near infrared diffuse reflectance [J]. *Transactions of the CSAE*, 2005, **21**(3): 162~165
- 赵杰文, 张海东, 刘木华. 利用近红外漫反射光谱技术进行苹果糖度无损检测的研究 [J]. *农业工程学报*, 2005, **21**(3): 162~165
- 7 Li Guifeng, Zhao Guojian, Wang Xiangdong *et al.*. Nondestructive measurement and fingerprint analysis of apple texture quality based on NIR spectra [J]. *Transactions of the CSAE*, 2008, **24**(6): 169~173
- 李桂峰, 赵建国, 王向东等. 苹果质地品质近红外无损检测和指纹分析 [J]. *农业工程学报*, 2008, **24**(6): 169~173
- 8 Zou Xiaobo, Zhao Jiewen, Xia Rong. Near infrared determination of sugar content in apples based on multiresolution decomposition and interval Partial Least Square (iPLS) method [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2006, **37**(6): 79~82
- 邹小波, 赵杰文, 夏蓉等. 苹果糖度近红外光谱小波去噪和 iPLS 建模 [J]. *农业机械学报*, 2006, **37**(6): 79~82
- 9 Wang Gang, Zhu Shiping, Kan Jianquan *et al.*. Nondestructive detection of volatile oil content in zanthoxylum bungeagum maxim by near infrared spectroscopy [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2008, **39**(3): 79~85
- 王刚, 祝诗平, 阚健全等. 花椒挥发油含量的近红外光谱无损检测 [J]. *农业机械学报*, 2008, **39**(3): 79~85
- 10 Xiao Gangluo, Chen Liyao. Discussion about BP neural network structure and choice of training parameter [J]. *J. Wuhan University of Technology*, 1997, **19**(2): 108~110
- 肖钢路, 陈立耀. 神经网络结构与训练参数选取 [J]. *武汉工业大学学报*, 1997, **19**(2): 108~110
- 11 Liu Yande, Chen Xingmiao, Ouyang Aiguo. Non-destructive measurement of soluble solid content in gannan navel oranges by visible/near-infrared spectroscopy [J]. *Acta Optica Sinica*, 2008, **28**(3): 478~481
- 刘燕德, 陈兴苗, 欧阳爱国. 可见/近红外光谱法无损检测赣南脐橙可溶性固形物 [J]. *光学学报*, 2008, **28**(3): 478~481
- 12 Han Donghai, Wang Jiahua. Review of nondestructive measurement of fruit quality by means of near infrared spectroscopy [J]. *Chinese J. Lasers*, 2008, **35**(8): 1123~1131

- 韩东海,王加华. 水果内部品质近红外光谱无损检测研究进展[J]. 中国激光, 2008, **35**(8): 1123~1131
- 13 Liu Fei, He Yong, Wang Li. Methods for the prediction of sugar content of rice wine using visible-near infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(11): 2054~2058
- 刘 飞,何 勇,王 莉. 黄酒糖度预测的可见近红外光谱方法研究[J]. 光学学报, 2007, **27**(11): 2054~2058
- 14 Zhao Jiewen, Guo Zhiming, Chen Quansheng *et al.*. Feasibility study on use of near-infrared spectroscopy in quantitative analysis of catechins in green tea[J]. *Acta Optica Sinica*, 2008, **28**(12): 2302~2306
- 赵杰文,郭志明,陈全胜 等. 近红外光谱法快速检测绿茶中儿茶素的含量[J]. 光学学报, 2008, **28**(12): 2302~2306