

文章编号:1005-0523(2009)03-0033-05

南丰蜜桔可溶性固形物非线性模型研究

刘燕德,孙旭东,章海亮,熊 坚

(华东交通大学 机电工程学院,江西 南昌 330013)

摘要:探索应用近红外透射光谱无损检测技术,结合径向基神经网络和反向传播神经网络两种人工智能算法,建立南丰蜜桔可溶性固形物快速无损检测的非线性稳定预测模型。通过43个未参与建模的南丰蜜桔样品作为外部验证数据,验证径向基函数神经网络模型的性能,预测模型的相关系数 $R=0.92$ 。研究结果表明:径向基函数神经网络比反向神经网络获得的预测结果更好,校正模型的预测均方根误差从 0.72°Brix 降低到 0.65°Brix 。径向基函数神经网络方法是解决近红外光谱检测南丰蜜桔可溶性固形物中非线性建模问题的有效方法。

关键词:非线性模型;径向基神经网络;近红外透射光谱;南丰蜜桔;可溶性固形物

中图分类号:TP242.6

文献标识码:A

南丰蜜桔品质优良,果实皮薄、无核或少核、色泽金黄、风味独特、质量上胜而深受广大消费者欢迎^[1]。可溶性固形物、酸度和维生素C等成分是评价南丰蜜桔品质的重要指标,但其传统破坏式测量方法耗时、费工、很难做到快速测定。近红外光谱技术是一种快速测定物质化学组分含量的现代光谱技术,具有快速、无损和廉价等优点,已被广泛用于石化、农业及食品行业中^[2,3]。

在南丰蜜桔的可溶性固形物、酸度和维生素C的近红外光谱建模过程中,常采用逐步回归分析法、主成分回归法和偏最小二乘法等算法来建立其化学组分含量与光谱之间的线性模型^[4~6]。但光谱与样品化学组分含量测定值之间常具有一定的非线性。而径向基函数神经网络具有最佳逼近和全局最优的性能,是一种有效地解决非线性问题的前馈式神经网络^[7]。在应用非线性建模算法解决南丰蜜桔近红外光谱检测方面的非线性问题还未见有相关报道。

本文应用近红外透射光谱技术在600~980 nm的光谱范围内,研究了径向基函数神经网络(RBFNN)和反向传播神经网络(BPNN)在建立南丰蜜桔可溶性固形物含量的非线性数学模型方面的可行性和有效性,并对比分析了两种模型的预测性能。

1 材料与方法

1.1 实验装置

本实验采用的近红外透射光谱采集系统是由CCD光谱仪(USB4000,美国海洋公司)、达美风行光源(220 V/50 W、佛山照明)、光纤(SMA905、美国海洋公司)、USB数据线、样品支座和计算机组成,系统组成如图1所示。

实验采用聚四氟乙烯白板作为参比材料,采集参比光谱时,光谱仪参数设置为:积分时间10 ms、平滑次数10次、扫描次数10次、采样间隔0.2 nm。实验时样品果柄朝上采集光谱,采集样品光谱时,光谱仪参数设置为:积分时间25 ms、平

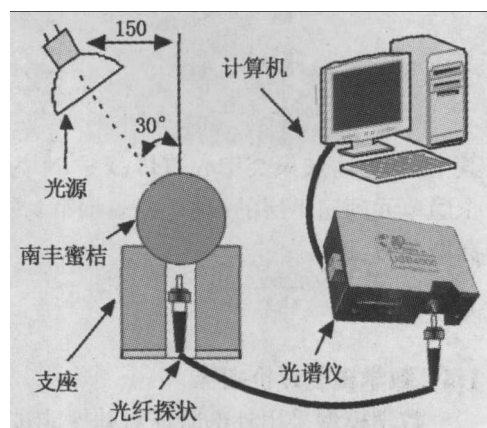


图1 近红外透射光谱采集系统

收稿日期:2009-02-16

基金项目:国家自然科学基金项目(60844007);国家科技支撑计划项目(2008BAD96B04、2006BAD11A12-07);江西省青年基金(2008GQN0029、2007GZN0266);教育部新世纪优秀人才资助项目(NCET-06-0575)

作者简介:刘燕德(1967-),女,江西泰和人,博士,教授,研究方向为光机电技术及应用。

滑次数 10 次、扫描次数 10 次、采样间隔 0.2 nm。光谱采集后,应用采集光谱软件 OOIbase32 系统(美国海洋公司),转化为吸光度光谱,再结合 Matlab7.0 软件中的神经网络工具箱,建立南丰蜜桔可溶性固形物含量非线性定量分析数学模型。

1.2 实验材料

本次实验共用 150 个南丰蜜桔样品,样品来自江西省南丰县果园场,样品运回实验室后在常温 25 ℃ 下左右放置。采集光谱前须将样品表面清理干净,剔除损伤或表面有缺陷样品后再开始采集样品的漫透射光谱。

实验样品的理化指标分析参照食品卫生检验方法的理化部分总则(GB/T5009.1-2003),可溶性固形物的测定采用折射式数字糖度计测量(PR-101a,日本)测量,具体统计数据见表 1。

表 1 南丰蜜桔可溶性固形物指标统计数据

测量指标	校正样本			预测样本		
	数量	平均值	范围	数量	平均值	范围
可溶性固形物/°Brix	107	13.9	9.6~18.0	43	14.0	10.5~17.9

1.3 径向基函数神经网络(RBFNN)

目前,在近红外光谱定量分析中常用的线性校正方法有逐步回归分析、主成分回归、偏最小二乘方法等。这些方法都是以朗伯-比尔定律为基础,假设样品的吸光度与其成分的浓度具有线性相关性,通过对光谱数据和样品的浓度建立线性校正模型,利用此模型根据光谱数据即可预测样品成分的浓度。而在实际应用中,由于生物样品一般是由多组分组成的混合物体系,其中各组分相互作用、仪器的噪声和基线漂移等使朗伯-比尔定律成立的条件改变,致使吸光度和样品的浓度之间呈现出非线性关系^[8]。常用的非线性校正方法是人工神经网络方法。在化学计量中常采用的一种人工神经网络结构是反向误差传播网络,在实际应用中反向误差传播网络存在训练速度慢、容易陷入局部极小等问题^[9]。本文采用在逼近能力和学习速度等方面优于反向误差传播网络的径向基函数网络(RBFNN)来建立南丰蜜桔中可溶性固形物含量的非线性模型。

RBFNN 的典型结构是一个 3 层前向网络,其中隐含层单元取径向基函数的局部感受单元。RBFNN 可描述为

$$Z = \frac{\omega_0 + \sum_{i=1}^N \omega_i \phi_i(\|x - c_i\|/\sigma_i)}{\sum_{i=1}^N \phi_i(\|x - c_i\|/\sigma_i)} \quad (1)$$

其中: N 为隐层聚类中心数目; $x \in R^n$ 为 n 维输入向量; $c_i \in R^n$ 为第 i 个隐单元中心的权向量; ω_i 为第 i 个隐单元到输出层的权值; σ_i 表示第 i 个隐单元的宽度, $\|\cdot\|$ 为欧几里德泛数; ϕ_i 为径向基函数,通常取高斯函数,即

$$\phi_i(\|x - c_i\|/\sigma_i) = e^{-\frac{\|x - c_i\|^2}{\sigma_i^2}} \quad (2)$$

1.4 数学模型评价指标

数学模型采用外部验证对其性能进行评价,由相关系数(R)和模型预测均方根误差(RMSEP)进行评价,相关系数越高且预测均方根误差越小,模型的稳健性和预测能力越高。

2 结果与讨论

2.1 近红外透射光谱主成分分析

近红外透射光谱往往有上千个数据点,若直接作为输入变量,则网络规模非常庞大。另外,近红外透射光谱数据点可能存在的共线性,不仅使网络运算速度慢,更易陷入局部最小,因此,有必要对近红外透射光谱数据进行降维处理。主成分分析具有良好的数据简化和降维效果,因此采用主成分分析压缩光谱数

据作为神经网络的输入变量。

从表 2 可知,随着主成分数的增加,方差逐渐减小,累计贡献率逐渐增大,当主成分数达到 6 个以后,从第 7 个开始,其方差很小,贡献率不大,因此本文选择前 6 个主成分作为神经网络的输入层变量。

表 2 前 7 个主成分的方差及累计贡献率

主成分数	01 - PC	02 - PC	03 - PC	04 - PC	05 - PC	06 - PC	07 - PC
方 差	79.521	7.637	12.545	0.182	0.041	0.049	0.003
累积贡献率/%	79.521	87.158	99.703	99.885	99.926	99.975	99.978

2.2 基于 RBFNN 和 BPNN 网络的非线性模型建立

根据主成分分析的结果,分别应用 107 个南丰蜜桔光谱的前 6 个主成分,作为 RBFNN 和 BPNN 网络的输入变量,建立南丰蜜桔可溶性固形物含量快速无损检测定量数学模型。RBFNN 网络的参数设置为:输入层、隐含层和输出层节点数分别为 6,14,1,传播速度 2.2,目标误差 0.0,训练迭代次数为 1 000 次;BPNN 网络的参数设置为:各层传递函数采用 Sigmoid 函数,输入层、隐含层和输出层节点数分别为 6,14,1,目标误差设为 0.001,网络学习速率 0.2,训练迭代次数为 1 000 次。校正模型的性能及预测结果如表 3 所示。

表 3 BPNN 和 RBFNN 训练和预测结果统计表

建模方法	输入层	隐含层	输出层	训练样本		验证样本	
				相关系数	建模均方根误差	相关系数	建模均方根误差
反向传播神经网络	6	14	1	0.95	0.41	0.90	0.72
径向基函数神经网络	6	14	1	0.97	0.37	0.92	0.65

为检验模型的预测能力,采用 43 个未参与建模的南丰蜜桔样品作为外部验证数据,其中 RBFNN 网络的预测模型的相关系数 $R = 0.92$,预测均方根误差 RMSEP 为 0.65°Brix (见图 2);BPNN 网络的相关系数 $R = 0.90$,预测均方根误差 RMSEP 为 0.72°Brix (见图 3)。可见,依据较高的相关系数 R 和较低的预测均方根误差 RMSEP 评价准则,在南丰蜜桔可溶性固形物近红外光谱分析中,RBFNN 和 BPNN 方法都有较强的预测能力,但 RBFNN 方法的预测能力更强。

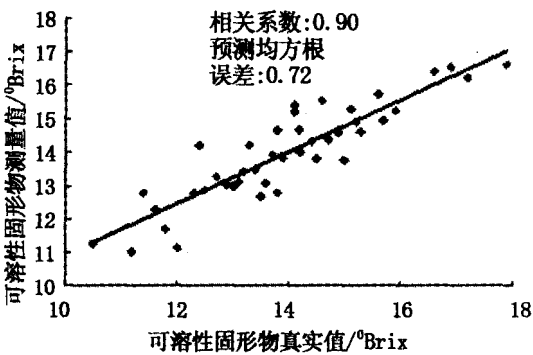
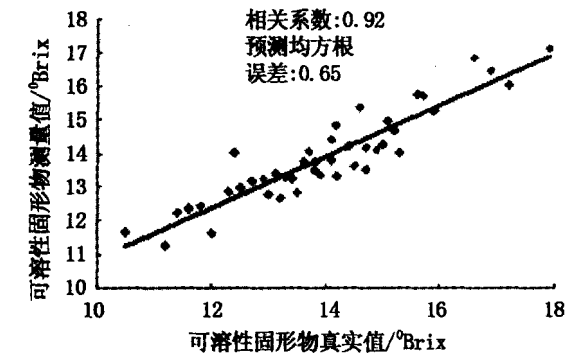


图 2 43 个未参与建模样品径向基函数神经网络预测模型 图 3 43 个未参与建模样品反向传播神经网络预测模型

3 结论

通过本次实验研究结果表明:通过 RBFNN 和 BPNN 两种神经网络方法建立的南丰蜜桔可溶性固形物近红外光谱快速无损检测校正模型具有较好的稳健性,模型具有较高的相关系数 R 和较低的预测均方根误差 RMSEP。另外,通过 43 个未参与建模的样品对模型进行验证,检验 RBFNN 和 BPNN 两种神经网络建模方法的预测能力,相关系数 R 和预测均方根误差 RMSEP 也比较理想,但 RBFNN 方法更优于 BPNN 方法。RBFNN 神经网络与 BPNN 神经网络相比,相关系数 R 从 0.90 提高到 0.92,预测均方根误差 RMSEP 从

0.72 °Brix 降低到 0.65 °Brix。今后本研究考虑到环境温度和光谱仪器本身电子电路噪声的影响,要建立更稳健、预测能力更强的非线性模型,还需进一步对南丰蜜桔的近红外光谱参数与其可溶性固形物化学测定值之间的非线性关系进行探讨。

参考文献:

- [1] 袁展汽,肖运平.南丰蜜桔主产区产地环境分析及果品质量安全生产建议[J].江西农业学报,2006,18(6):222-224.
- [2] 陆婉珍.现代近红外光谱分析技术[M].北京:中国石化出版社,2007.19-26.
- [3] 严衍禄,赵龙莲,韩东海,等.近红外光谱分析基础与应用[M].北京:中国轻工业出版社,2005.286.
- [4] 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.
- [5] 刘燕德,欧阳爱国,罗吉,等.近红外漫反射光谱检测赣南脐橙可溶性固形物的研究[J].光谱学与光谱分析,2007,27(11):2190-2192.
- [6] 夏俊芳,李小昱,李培武,等.基于小波变换的柑橘维生素C含量的近红外光谱无损检测方法[J].农业工程学报,2007,23(6):170-174.
- [7] 丁涛,周惠成.基于径向基函数神经网络的预测方法研究[J].哈尔滨工业大学学报,2005,37(2):272-275.
- [8] 陈华才,吕进,陈星旦,等.基于径向基函数网络的茶多酚总儿茶素近红外光谱检测模型研究[J].光学精密工程,2006,14(1):58-62.
- [9] 张乃尧,阎平凡.神经网络与模糊控制[M].北京:清华大学出版社,1998.171-180.

A Research on Nonlinear Model of Soluble Solids Content Based on Radial Basis Function Neural Network in Nanfeng Mandarin Fruits

LIU Yan-de, SUN Xu-dong, ZHANG Hai-liang, XIONG Jian

(School of Mechanical and Electrical Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: The research aims to develop nonlinear mathematical model for rapid determining soluble solids content (SSC) of Nanfeng mandarin fruit nondestructively using near infrared (NIR) transmittance spectroscopy technology, combining with artificial intelligent algorithms of basis function neural network (RBFNN) and back propagation neural network (BPNN). The 43 unknown Nanfeng mandarin fruits are used for external verifying data, the performance of RBFNN model is verified with Correlation coefficient ($R = 0.92$). The experimental results show that the predicting result of RBFNN is better than that of BPNN. Root mean square error of prediction (RMSEP) decreases from 0.72 °Brix to 0.65 °Brix. So RBFNN method is an effective method to solve the nonlinear problem in NIR spectra measurement SSC of Nanfeng mandarin.

Key words: nonlinear model; RBFNN; near infrared transmittance spectroscopy; Nanfeng mandarin fruit; soluble solids content

(责任编辑:刘棉玲)