

李伟伟,罗华平,孔维楠. 高光谱成像技术结合遗传算法和 BP 神经网络的南疆骏枣总糖含量建模分析[J]. 江苏农业科学,2018,46(1):173-176. doi:10.15889/j.issn.1002-1302.2018.01.046

# 高光谱成像技术结合遗传算法和 BP 神经网络的南疆骏枣总糖含量建模分析

李伟伟<sup>1</sup>, 罗华平<sup>1,2,3</sup>, 孔维楠<sup>1</sup>

(1. 塔里木大学机械电气化工程学院, 新疆阿拉尔 843300; 2. 新疆维吾尔自治区普通高等学校现代农业工程重点实验室, 新疆阿拉尔 843300; 3. 新疆农业农机化研究中心, 新疆阿拉尔 843300)

**摘要:**为考察南疆骏枣的总糖含量与高光谱数据的对应关系,采用高光谱成像技术结合遗传算法和反向传播(back propagation,简称 BP)神经网络的方法对南疆骏枣进行总糖含量建模分析,采用不同预处理方法处理原始光谱数据,并用遗传算法进行波长变量的选择,回归分析分别采用偏最小二乘法(PLS)和 BP 神经网络(BP-ANN),从线性和非线性回归方式上建立校正模型分析。结果表明,用多元散射校正(MSC)进行预处理好于其他预处理方式。在回归分析上,BP-ANN 预测南疆骏枣总糖含量的模型好于 PLS,预测相关系数( $R_p$ )和预测标准偏差(RMSEP)分别为 0.918 4 和 0.931 0。由结果可知,利用近红外高光谱成像技术对南疆骏枣总糖含量进行预测是可行的。

**关键词:**高光谱成像技术;遗传算法;偏最小二乘;BP 神经网络;多元散射校正

**中图分类号:**S665.101;S126 **文献标志码:**A **文章编号:**1002-1302(2018)01-0173-04

南疆骏枣营养价值高,其品质深受人们关注。由于南疆特殊的地理环境和气候条件,特别是昼夜温差大、光照时间长、有效积温高等条件,孕育了南疆骏枣优良的品质,如粒大饱满、营养价值高,深受人们的喜爱。总糖含量是南疆骏枣品质信息的常规性指标,它对红枣进一步分级和销售有重要的影响。传统的检测方法不仅耗时费力,还对样品产生侵入式

破坏,且分析速度慢、损害环境,严重影响红枣的分级和销售质量。因此迫切需要一种简单、快速、无损的品质检测方法,保证南疆骏枣的发展。

近年来,高光谱成像技术应用在农产品品质检测方面已成为一个研究热点<sup>[1]</sup>。传统的光谱技术只能对农产品的光谱信息进行处理,不能获取样品的图像信息。新一代的无损检测技术——高光谱成像技术的诞生,弥补了近红外(near infrared, NIR)光谱技术的弊端,同时获得样品的图像信息和光谱信息,可以综合反映农产品的品质信息<sup>[2]</sup>。目前,在农产品检测方面,高光谱图像技术应用于鲜枣的损伤、可溶性固形物含量检测<sup>[3]</sup>、小黄瓜的水分无损检测<sup>[4]</sup>、苹果的硬度与糖度的检测<sup>[5]</sup>、马铃薯的淀粉含量检测<sup>[6]</sup>、柑橘的缺陷检测<sup>[7]</sup>、长枣表面农药残留检测<sup>[8]</sup>等内外部品质的检测。研究结果表明,高光谱图像技术在农产品品质检测方面拥有广阔的前景。

收稿日期:2016-08-10

基金项目:国家自然科学基金(编号:11164023,11464039);塔里木大学校长基金(编号:TDZKTG201501,TDGRI201503);塔里木大学研究生科研创新项目(编号:TDGRI201619)。

作者简介:李伟伟(1985—),男,山东临沂人,硕士研究生,主要从事农产品无损检测方面的研究。E-mail:liweimei8503@163.com。

通信作者:罗华平,教授,主要从事农产品品质无损检测、光谱技术应用及快速无损检测装备开发研究。E-mail:luohuaping739@163.com。

[2]王冬. 基于物联网的智能农业监测系统的设计与实现[D]. 大连:大连理工大学,2013.

[3]武洪峰. 物联网技术在农机作业监控中的应用[J]. 现代化农业,2015(1):64-66.

[4]魏卓,郑琪,张俐,等. 基于 GPS-OEM 的车载农田面积测量系统的设计[J]. 东北农业大学学报,2010,41(5):145-149.

[5]王陈陈. 基于 GPS 的农田面积测量系统的研究[D]. 淄博:山东理工大学,2013.

[6]陈立飞. 综合型汽车智能诊断仪的研究与开发[D]. 天津:河北工业大学,2014.

[7]廖奎旭. GPS/北斗双模接收机前端的设计与实现[D]. 成都:电子科技大学,2013.

[8]辛德奎. 基于北斗/GPS 双模的田间作业机车工况监测系统[D]. 大庆:黑龙江八一农垦大学,2015.

[9]辛德奎,黄操军,甘龙辉. 基于北斗和 GPS 的双模田间作业机车定

位系统[J]. 黑龙江八一农垦大学学报,2014,26(1):85-87,112.

[10]刘荣. 基于北斗卫星导航系统的采棉机监控终端的设计与研究[D]. 新疆:石河子大学,2014.

[11]刘兴科,张之孔. “北斗一号”车辆监控管理系统的设计[J]. 测绘与空间地理信息,2012,35(6):15-18.

[12]李瑞欣. 基于网络和状态监测的设备管理理论与方法研究[D]. 天津:天津大学,2004,62-79.

[13]王川. 农业机械装备信息管理系统的设计和研究[D]. 合肥:安徽农业大学,2009,25-57.

[14]张玉宝. 基于网络的农业机械化信息系统的研究与开发[D]. 杨凌:西北农林科技大学,2003,17-27.

[15]许志宏,闫福鑫,余志军,等. 基于状态机的 CAN 2.0 总线协议的实现[J]. 计算机技术与发展,2012,22(5):141-144.

[16]孙巍. 基于 ZIGBEE 协议的无线土壤温、湿度检测系统的研究[D]. 长春:吉林农业大学,2012.

## 1 材料与方法

### 1.1 仪器设备及软件

Hyperspectral Sorting System 高光谱分选系统(900 ~ 1 700 nm, 光谱分辨率 5 nm, 256 个波段)。如图 1 所示, 该系统由高光谱成像光谱仪、电荷耦合器件(CCD)相机、探测器(类型 InGaAs)、成像镜头、电控位移平台、计算机组成。高光谱图像数据分析软件为 ENVI 4.6 和 Matlab 2013a。

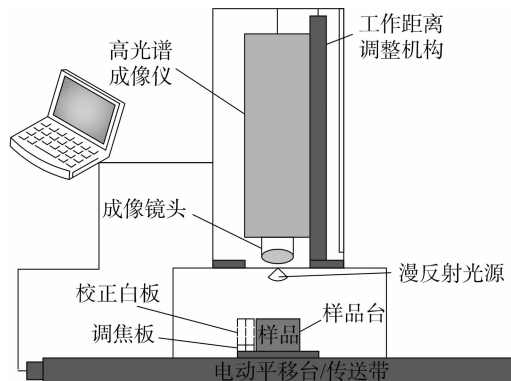


图1 高光谱分选系统

### 1.2 试验材料与总糖测定

试验材料为 2015 年 9 月采集于新疆生产建设兵团阿拉尔市 10 团的白熟期骏枣, 选出大小均匀、无破损的红枣 120 颗, 进行编号及除尘等处理。

红枣总糖含量测定采用国标法, 即直接滴定法, 按照 GB/T 12456—2008《食品中总糖的测定》执行。

### 1.3 高光谱图像的采集及反射光谱的获取

1.3.1 高光谱图像的采集 采集南疆骏枣样品的高光谱图像系统设置: 光谱扫描范围 900 ~ 1 700 nm, 光谱分辨率 5 nm, 扫描速度 100 张/s, 相机像素 320 × 256, CCD 相机的曝光时间 20 ms, 电控位移平台速度 17.3 mm/s, 相机高度 36 cm, 行程 25 cm。

由于光源的强度在各个波段下分布不均匀、样品形状不规则及暗电流的存在, 需对所获高光谱图像进行黑白校正, 消除噪声影响。样品采集相同系统条件下, 扫描白色校正板得到的全白图像  $W$ , 关闭相机镜头进行图像采集得到的全黑图像  $B$ , 然后再采集样品光谱图像<sup>[9]</sup>。黑白校正公式如下:

$$R = \frac{I - B}{W - B}$$

式中:  $R$  为校正后高光谱图像;  $I$  为原始高光谱图像。

1.3.2 反射光谱的获取 校正后的南疆骏枣高光谱图像在 ENVI 4.6 中打开, 选取表面 1 个感兴趣区域(简称 ROI)(60 ~ 100 个像素点), 计算其平均反射光谱, 得到原始反射光谱<sup>[10]</sup>, 如图 2 所示。

### 1.4 算法及原理

遗传算法(GA)借鉴生物界的自然选择和遗传机制, 利用选择、交换和突变等算子操作, 随着遗传迭代的不断进行, 使得目标函数值较优的变量被保留, 较差的变量被淘汰, 最终得到最优的结果。遗传算法的实现包括 5 个基本要素: 参数编码、群体的初始化、适应度函数的设计、遗传操作设计、收敛判断和变量的选取<sup>[11]</sup>。具体的实现流程如图 3 所示。

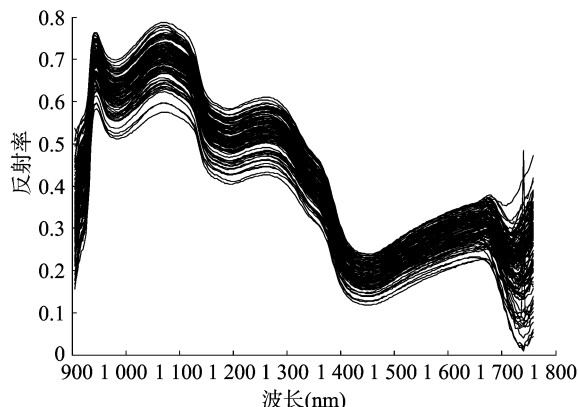


图2 原始反射光谱曲线

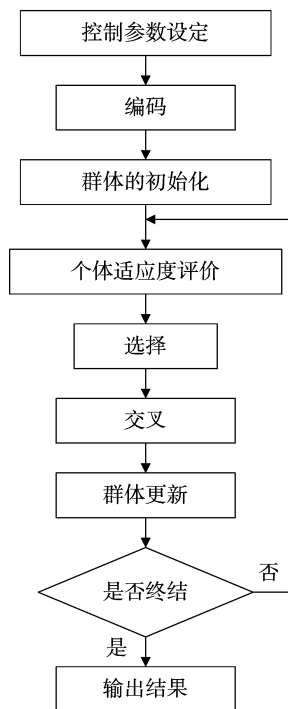


图3 遗传算法实现流程

反向传播(back propagation, 简称 BP)神经网络(neural network)是由非线性变换神经单元组成的前馈型多层神经网络, 主要由 3 部分组成: 输入层、隐含层和输出层。输入信息按从输入层经隐含层到输出层计算的方向进行, 阈值和权值则按照输出到输入的方向进行修正, 从而实现了信息的正向传递和误差的反向传播<sup>[12]</sup>。

## 2 结果与分析

### 2.1 反射光谱的预处理

分别采用均值中心化(mean centering, MC)、矢量归一化(vector normalization)、标准正态变量变换(standard normal variate transformation, SNV)、多元散射校正(multiplicative scatter correction, MSC)、导数法对光谱进行预处理。其中多元散射校正处理后的光谱见图 4。

### 2.2 波长变量选择方法

由于高光谱图像波长覆盖范围宽、波段多, 每次采集大量

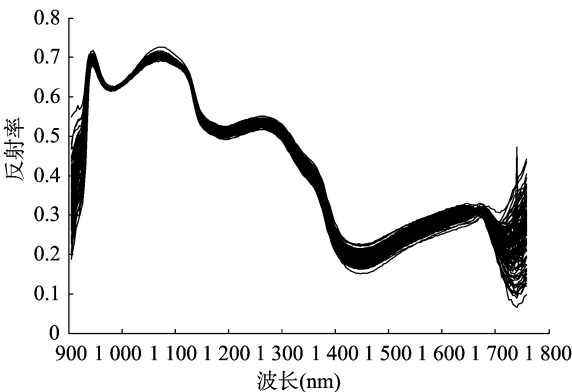


图4 多元散射校正处理的光谱

光谱信息,而不同的原始光谱数据对待测样品的品质信息贡献不同,通过特定的波长变量的筛选方法对波长变量进行优选,可以减少建模时间,简化建模过程,最重要的是可以剔除无关的信息变量,从而建立鲁棒性强、预测能力好的定量校正模型<sup>[13]</sup>。

遗传算法参数设定初始种群为 100,交叉概率  $P_c = 0.8$ ,变异概率  $P_m = 0.01$ ,遗传迭代次为 100 次。图 5 为遗传算法各变量被选用的频次,图 6 为交叉验证选择出的变量数分布。

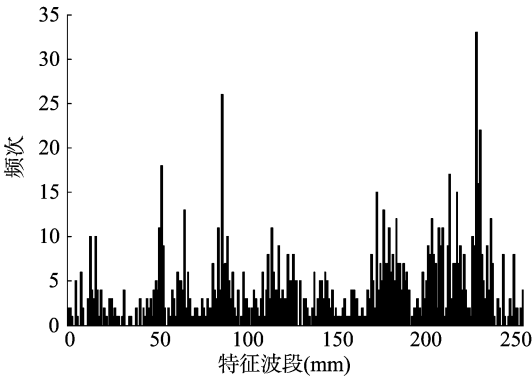


图5 各变量被选用的频次

2.3 建模方法的选择

目前主要采用的线性校正方法是偏最小二乘法 (PLS)、非线性校正方法是 BP 人工神经网络 (BP - ANN)。BP -

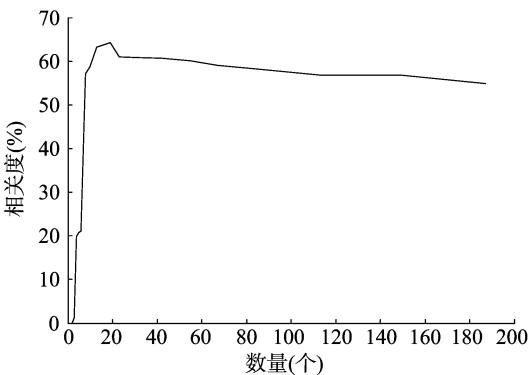


图6 交叉验证选择出的变量数分布

ANN 是一个前向多层网络,它利用误差反向传播算法对网络进行训练。由于其结构简单、可塑性强,包含了神经网络理论中最精华的部分,因此 BP - ANN 在非线性校正模型中得到了广泛应用<sup>[14]</sup>。

2.4 界外样本的剔除及样本的总糖信息

120 颗待测红枣经过马氏距离 (阈值  $e = 2.1$ ) 和浓度残差 (阈值  $h = 2.8$ ) 对骏枣光谱和总糖含量进行剔除。

剔除后样品数为 112 个,选择校正集与预测集的样本比例为 3 : 1 (校正集 84 个,预测集 28 个)。南疆骏枣总糖信息如表 1 所示。

表 1 南疆骏枣样本总糖信息

样品集	样品数 (个)	总糖含量 (%)		标准偏差 (%)	变异系数 (%)
		范围	均值		
校正集	84	11.59 ~ 23.87	16.83	3.06	18.16
预测集	28	12.23 ~ 23.33	17.88	2.40	13.40

2.5 预处理方法的选择结果

采集的红枣样品信息除了有待测样品的原始化学信息外,还有其他外在的干扰信息和噪声,这些信息导致化学信息与真实信息存在一定差异。为了提高信噪比,尽量减少误差,选择合适的预处理方法至关重要<sup>[15]</sup>。

经过建模分析,不同预处理方法对南疆骏枣样品的建模精度和预测能力的影响如表 2 所示。

表 2 不同预处理方法南疆骏枣样品的预测能力

建模方法	预处理方法	GA 选择的变量数 (个)	主成分数 (个)	$R_p$	RMSEP
PLS	MC	10	8	0.844 4	1.260 5
	Vector Normalization	15	7	0.898 7	1.031 9
	SNV	11	6	0.864 0	1.184 7
	MSC	12	8	0.900 2	1.024 5
	First Derivative	17	7	0.686 6	1.710 6
BP - ANN	MC	10		0.877 3	1.129 2
	Vector Normalization	15		0.913 6	0.956 7
	SNV	11		0.899 0	1.030 3
	MSC	12		0.918 4	0.931 0
	First Derivative	17		0.795 9	1.424 5

2.6 建模分析

运用多元散射校正进行预处理,遗传算法选择特征波长变量。分别用偏最小二乘法和 BP 神经网络建模,真实值与

预测值对比结果如图 7、图 8 所示。

2.7 模型评价

2.7.1 模型的验证参数评价 模型验证评价参数如表 3

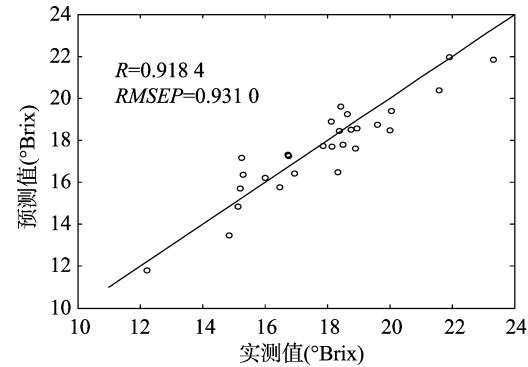


图7 新疆骏枣糖度 PLS 建模预测结果

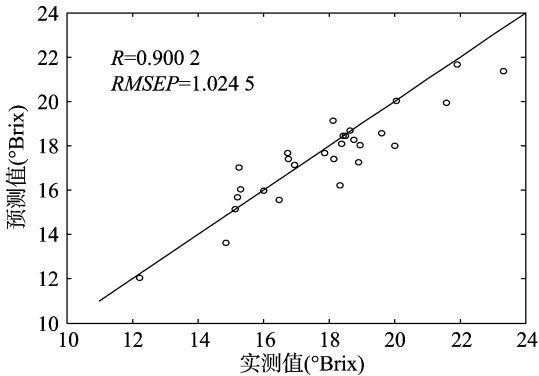


图8 新疆骏枣糖度 BP-ANN 建模预测结果

表 3 不同建模方法的验证评价参数

建模方法	$R_p$	RMSEP	RPD
PLS	0.900 2	1.024 5	2.296
BP-ANN	0.918 4	0.931 0	2.527

所示。

2.7.2 显著性检验 选择成对  $t$  检验,其实质是判断 2 种方法之间的偏差的均值是否与期望值(0)存在显著差异<sup>[14]</sup>。用 BP-ANN 模型预测近红外高光谱方法测定总糖含量,给定显著水平  $\alpha=0.05$ ,如果  $|t| < t_{\alpha}$ ,则说明 2 种方法无显著差异。 $t$  检验结果显示,国标法直接滴定法与近红外高光谱法测定新疆骏枣总糖含量的  $n=28, \bar{d}=-0.246\ 3, s=0.914\ 3, |t|=1.425\ 5, t_{(27,0.05)}=2.052$ 。

3 结论

(1) 运用多元散射校正(MSC)预处理方法好于其他预处理方法。

(2) 遗传算法(GA)筛选的特征波长变量作为输入变量的定量校正模型鲁棒性强、预测能力更好。

(3) 应用 BP 神经网络(BP-ANN)建立的校正模型预测能力好于偏最小二乘法(PLS)建立的校正模型。

本试验选取 112 个样本作为研究对象,试验结果比较理想,但后续工作需要进一步增加样本数量及样本的品种类型,完善新疆红枣的建模预测模型。红枣的外部品质信息也是下一步工作的研究内容,从而实现红枣内外部品质的同时检测。

参考文献:

[1] 徐爽,何建国,易东,等. 基于高光谱图像技术的长枣糖度无损检测[J]. 食品与机械,2012,28(6):168-170.

[2] 吴龙国,何建国,刘贵珊,等. 基于近红外高光谱成像技术的长枣含水量无损检测[J]. 光电子·激光,2014,25(1):135-140.

[3] 薛建新,张淑娟,张晶晶. 高光谱成像技术对鲜枣内外部品质检测的研究[J]. 光谱学与光谱分析,2015,35(8):2297-2302.

[4] 李丹,何建国,刘贵珊,等. 基于高光谱成像技术的小黄瓜水分无损检测[J]. 红外与激光工程,2014,43(7):2393-2397.

[5] 王爽. 基于高光谱散射图像的苹果内部品质预测建模[D]. 无锡:江南大学,2012.

[6] 吴晨,何建国,贺晓光,等. 基于近红外高光谱成像技术的马铃薯淀粉含量无损检测[J]. 河南工业大学学报(自然科学版),2014,35(5):11-16.

[7] 章海亮,高俊峰,何勇. 基于高光谱成像技术的柑橘缺陷无损检测[J]. 农业机械学报,2013,44(9):177-181.

[8] 刘民法,张令标,何建国,等. 基于高光谱成像技术的长枣表面农药残留无损检测[J]. 食品与机械,2014,30(5):87-92.

[9] 黄培贤,姚志湘,栗晖,等. 高光谱图像技术在食品无损检测中的研究进展[J]. 食品工业科技,2012,33(15):412-417.

[10] 徐爽. 基于高光谱图像技术的红枣品质无损检测研究[D]. 银川:宁夏大学,2013.

[11] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京:化学工业出版社,2011.

[12] 杨杰,占君,张继传,等. MATLAB 神经网络 30 例[M]. 北京:电子工业出版社,2014.

[13] 彭云发. 近红外光谱技术在新疆红枣品质快速无损检测中的应用[D]. 阿拉尔:塔里木大学,2015.

[14] 葛哲学,孙志强. 神经网络理论与 MATLAB R2007 实现[M]. 北京:电子工业出版社,2007.

[15] 胡晓男. 新疆红枣品质数字化快速无损检测技术推广应用[D]. 阿拉尔:塔里木大学,2015.