

近红外光谱结合变量优选和 GA-ELM 模型的 干制哈密大枣水分含量研究

王文霞^{1,2}, 马本学^{1*}, 罗秀芝^{1,2}, 李小霞^{1,2}, 雷声渊^{1,2}, 李玉洁^{1,2}, 孙静涛³

1. 石河子大学机械电气工程学院, 新疆 石河子 832003
2. 农业农村部西北农业装备重点实验室, 新疆 石河子 832003
3. 石河子大学食品学院, 新疆 石河子 832003

摘要 水分含量是哈密大枣干制过程中的重要指标, 对其外观、口感、贮藏和运输具有重要的影响。因此, 为实现哈密大枣水分含量的准确预测, 采用近红外光谱结合变量优选方法, 建立干制哈密大枣水分含量的 GA-ELM 预测模型。为提高模型的稳定性和预测精度, 开展并讨论了核函数和神经元个数对 GA-ELM 预测模型的影响。采用多种预处理方法对全波段光谱进行处理, 对比分析发现标准正态变换方法(SNV)效果最佳。对标准正态变换处理后的光谱利用连续投影算法(SPA)、联合区间偏最小二乘(si-PLS)和遗传算法(GA)及其组合算法分别从全波段 927.77~2 501.14 nm 范围内筛选特征波长, 并建立对应 GA-ELM 预测模型, 同时与全波段的 GA-ELM 模型效果相比较, 采用 SNV+SPA 筛选的 14 个特征波长建立的 GA-ELM 模型效果最佳, 预测结果 R_c 和 R_p 分别为 0.984 2 和 0.967 5, RMSEC 和 RMSEP 分别为 0.006 1 和 0.007 9, RPD 为 3.678 8。研究结果表明: SNV+SPA+GA-ELM 方法可实现干制哈密大枣水分含量的准确预测, 为近红外光谱技术应用于干制哈密大枣在线检测提供了参考。

关键词 近红外光谱; 哈密大枣; 水分含量; 特征波长; 极限学习机(ELM)

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)02-0543-07

引言

哈密大枣是新疆特色优势果品, 个大肉厚, 外观紫红有光泽, 食之有药香, 是上等的滋补食品和药用食品^[1]。哈密大枣的鲜食期较短, 大量的鲜枣收获后进行干制, 将鲜枣的水分含量从 70% 降低到 25% 左右, 使其可溶性固形物的浓度达到微生物难以生存和利用的程度^[2], 以便贮藏和运输。水分含量(MC)作为干制哈密大枣重要品质参数之一, 干制过程中若含水量过少, 会使硬度增加, 口感变差; 若水分含量过高, 会使细菌容易繁殖, 在贮藏和运输过程中易腐烂变质。因此, 如何实现干制哈密大枣水分含量的快速、无损检测就显得尤为重要。

近红外光谱在农产品品质定量分析领域中得到广泛应用^[3-6]。由于光谱变量中含有大量冗余信息, 从全波长中提取出对模型有效的特征波长变量, 可以进一步提高模型的稳定性和准确性^[7]。彭云发等^[8]利用近红外光谱结合遗传算法

(genetic algorithm, GA), 对白熟期灰枣的总糖含量进行预测, 建立的偏最小二乘(partial least squares, PLS)预测模型 R_p 达到 0.9583; 胡晓男等^[9]使用近红外光谱结合无信息变量消除法(uninformative variable elimination, UVE), 对晚熟期骏枣总酸进行预测, 建立的 PLS 模型 R_p 为 0.8778; 彭海根等^[10]利用近红外光谱结合联合区间偏最小二乘法(synergy interval partial leset squares, si-PLS), 对南疆灰枣进行糖度预测, 建立的 PLS 预测模型, 相关系数达到 0.942; 以上研究主要以鲜枣为研究对象, 而对干制哈密大枣的研究却鲜有报道。另外, 有国内外学者通过近红外高光谱对农产品水分含量进行研究。Anisur Rahman 等^[11]使用高光谱成像对番茄水分含量进行预测, PLS 模型相关系数为 0.81; 杨传得等^[12]使用近红外光谱技术对鲜食花生水分含量进行预测, PLS 模型决定系数达到 0.936 2。他们的工作对干制哈密大枣水分检测具有一定的启示和借鉴作用。

采用多种预处理方法结合变量筛选方法, 将筛选出的与

收稿日期: 2018-12-17, 修订日期: 2019-04-09

基金项目: 国家自然科学基金项目(61763043)资助

作者简介: 王文霞, 1992 年生, 石河子大学机械电气工程学院硕士研究生 e-mail: wwx_0817@163.com

* 通讯联系人 e-mail: mbx_shz@163.com

水分含量相关性较高波长变量,作为遗传算法优化极限学习机(GA-ELM)的输入量,建立干制哈密大枣水分含量的预测模型,并评价相应的模型预测效果,以此验证了GA-ELM模型应用于干制哈密大枣水分含量预测的可行性,实现快速、科学和准确的干制哈密大枣水分含量检测。

1 实验部分

1.1 材料

选取 220 个新疆哈密五堡果园采摘的哈密大枣为研究样本,果实成熟后采用自然晾晒干,颜色深红,果实饱满有弹性,大小均匀,无损伤,平均果重 9 g 左右。将挑选好的干制哈密大枣表面擦拭干净,并进行编号,在室温条件下(28

~31 °C, 27~30% RH)放置 48 h,使样本温度和环境温度达到一致。在样本赤道上选取 3 个点(间隔约 120°)作为光谱采集点,每个点采集 1 次,对 3 条光谱取平均,作为该样本的代表光谱。使用 SPXY 方法将 220 个干制哈密大枣样本分为 170 个校正集和 50 个预测集样本,用于建模分析。

1.2 仪器

近红外光谱系统主要由光谱仪(NIRQuest 256-2.5, Ocean Inc., USA),光源(卤钨灯, 30 W, 12 V, QP400-1-VIS-NIR, VIVO, Ocean Optics Inc., USA),漫反射光纤,暗箱和计算机组成,如图 1 所示。光谱范围为 869.19~2 501.14 nm,共 256 个波长,分辨率为 9.5 nm。设置积分时间为 18 ms,扫描次数和平滑度分别为 32 和 3,最后将样本放置于 VIVO 光源载物台上进行近红外漫反射光谱采集。

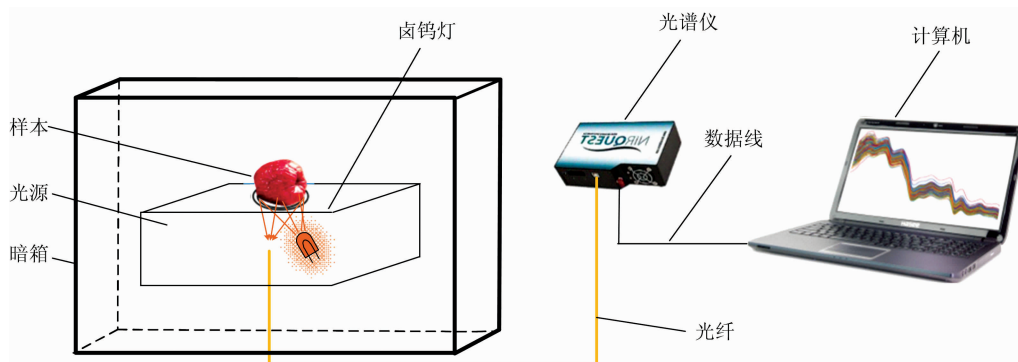


图 1 近红外光谱采集系统的结构示意图

Fig. 1 Schematic diagram of near infrared reflectance spectrum acquisition system

1.3 干制哈密大枣水分含量的测量

在近红外光谱数据采集后将完整的干制哈密大枣使用电子秤(JY/YP30002, 上海乐平科学仪器有限公司,精度为 0.000 1 g)测量干制哈密大枣的湿重。然后放入烘干箱(DHG-9070 A, 220 V, 50 Hz, 1 550 W, 温控范围: 10~250 °C)进行脱水,干燥温度设定为 105 °C, 4 h 后取出称重,然后继续放入烘干箱干燥 0.5 h,重复操作直至前后两次重量误差不超过 2 mg,记录此时重量作为样本的干重值。将湿重与干重的差值,除以湿重,计算出每个样本的水分含量。表 1 为测得 220 个干制哈密大枣样本划分为校正集和预测集的水分含量统计结果。

$$\text{水分含量}(\%) = \frac{\text{湿重} - \text{干重}}{\text{湿重}} \times 100\%$$

表 1 干制哈密大枣水分含量统计分析

Table 1 Statistical analysis of MC of Hami big jujubes samples

理化指标	样本集	样品数	最大值 /%	最小值 /%	平均值 /%	标准 偏差/%
水分含量	校正集	170	27.97	10.70	18.28	3.51
	预测集	50	23.38	13.01	17.48	2.80

1.4 数据处理方法

1.4.1 预处理方法

由于受到外部环境和仪器暗电流的影响,在光谱数据采集过程中存在谱线重叠、噪音信号和基线漂移等问题,影响

后续建模精度和稳定性^[13]。通过分析,在 869.19~927.77 nm 光谱范围内存在较大的噪音,因此采用 927.77~2 501.14 nm 范围内的光谱(共 247 个波段)用于下一步的数据分析。为进一步消除干扰因素影响,应用标准正态变换(standard normal variate, SNV)、多元散射(multiplicative scatter correction, MSC)和矢量归一化(normalization, Norm)等方法处理。

1.4.2 变量优选方法

在建立预测模型时,若模型中存在较多与干制哈密大枣水分含量无关的光谱信息,会影响模型准确性和计算速度。光谱特征波段的筛选对后续建模和预测效果具有较大影响。因此,采用联合区间偏最小二乘法(si-PLS)、遗传算法(GA)、连续投影算法(successive projections algorithm, SPA)及其算法组合等方法来提取特征波长以简化模型。

1.4.3 GA-ELM 建模方法

极限学习机(extreme learning machine, ELM)是一种求解单隐层前馈型神经网络(SLFN)的学习算法,其学习速度快,泛化性能好,具有高效处理非线性数据回归拟合问题的能力^[14]。ELM 与传统的函数逼近理论不同,其预测精度与隐层节点的个数密切相关。输入层与隐层间的连接权重 w_i 和隐层神经元阈值 b_i 是随机生成的,在给定参数时可能存在部分数值为 0 的情况,导致隐层输出矩阵不满秩,使得部分隐层节点失效,进而降低模型的预测精度。

为增强 ELM 网络稳定性和预测精度，用遗传算法(GA)对 ELM 神经网络的 w_j 和 b_j 进行优化选择，以确定最优的 ELM 模型。

训练步骤如下：

(1)根据 ELM 神经网络的基本拓扑结构，将输入权值和阈值级联起来，产生 k 个初始种群。

$$Q^{\lambda} = [\omega_{\lambda 1}^1, \dots, \omega_{\lambda l}^1, \omega_{\lambda 1}^2, \dots, \omega_{\lambda l}^2, \dots, \omega_{\lambda n_1}^l, \omega_{\lambda n_2}^l, \dots, \omega_{\lambda n_l}^l, b_{\lambda 1}^l, \dots, b_{\lambda l}^l] \quad (1)$$

其中， Q^{λ} 为种群中的第 λ ($1 \leq \lambda \leq k$) 个个体， ω_{ij} 为第 i 个神经元与隐含层第 j 个神经元的连接权值， b_l 为第 l 个隐含层节点的阈值，且均为区间 $[-1, 1]$ 中的随机数。适应度函数用于评价种群优劣程度，将模型对验证集的预测均方根误差作为适应度函数，适应度函数越小，则模型精度越高。

$$\text{Fitness} = \frac{\sum_{i,j} |y_{ij} - x_{ij}|}{k} \quad (2)$$

(2)利用 ELM 算法隐含层激活函数(“sig”函数)计算出输出权值矩阵。将归一化后的样本数据代入模型，逐个计算每个个体的适应度(Fitness_best)，根据适应度函数值来确定适应度较优的个体。

(3)对适应度较优的个体利用交叉、变异对种群进行进化，得到新种群。并检查进化代数 g 值，当 g 小于 P 时，返回到第(2)步，直至 g 等于 P 则结束运算，选出最优个体，并将其拆分为初始权值和隐含层阈值，即可确定最优的 ELM 模型。

1.5 模型评价

模型校正和预测性能采用校正相关系数(R_c)、预测相关系数(R_p)、校正均方根误差(RMSEC)、预测均方根误差(RMSEP)表示，以及剩余预测偏差(RPD)作为评价模型预测能力的指标。Nicolar^[15]指出当 RPD 小于 1.5 时，说明模型预测性能较差；当 RPD 值在 1.5~2 之间说明模型对检测指标有一定的预测能力；当在 2~2.5 之间说明可以粗略对检测指标进行定量分析；当在 2.5~3 之间，说明具有较好的预测精度。

2 结果与讨论

2.1 GA-ELM 预测模型参数优化

在 GA-ELM 算法中，参数的设置直接影响算法的寻优速度和质量。其中，核函数的类别和隐含层神经元个数对模型预测能力和稳定性的影响最直接，因此针对 GA-ELM 模型中核函数选择和隐含层神经元个数进行研究。

(1)核函数对网络的影响：核函数的种类一般有“sig”，“sin”和“hardlim”等 3 种。将 220 个干制哈密大枣样本的原始光谱作为 GA-ELM 模型输入量，使用不同核函数建立的 GA-ELM 模型预测结果如表 2 所示。从表中可以看出，使用“sig”核函数建立的 GA-ELM 预测模型相关系数达到 0.9 以上，具有较强的预测能力，因此选择“sig”作为最优核函数。

(2)隐含层神经元个数对模型的影响：隐含层神经元数目太少会导致 ELM 网络“欠拟合”，而数目太多会导致 ELM

网络“过拟合”。分别选取 10, 20, ..., 120 个隐含层节点，得到 12 种不同情况下 GA-ELM 网络的预测结果，再次进行 3 次重复测试，求平均值以比较其拟合情况，结果如图 2 所示。当隐含层神经元个数小于 50 时， R_p 大于 R_c ，模型存在过拟合现象；当隐含层神经元个数在 50~100 时， R_p 和 R_c 呈平稳上升趋势，且 R_c 上升趋势较 R_p 要快，模型稳定性显著提高；当隐含层神经元个数大于 100 时， R_p 急剧下降，RMSEP 急剧上升，模型精度大幅度下降。因此，本文选择隐含层神经元个数为 100 时，作为最优隐含层神经元个数。

表 2 不同核函数建立 GA-ELM 模型的预测结果

Table 2 Prediction results of GA-ELM model established by different kernel functions

核函数	R_c	RMSEC	R_p	RMSEP	RPD
sig	0.937 0	0.012 3	0.937 4	0.009 8	2.870 8
sin	0.892 2	0.015 9	0.867 5	0.013 9	2.010 2
hardlim	0.790 1	0.021 5	0.543 7	0.023 5	1.191 5

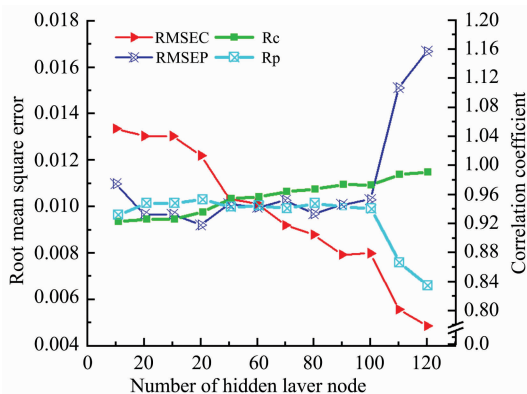


图 2 隐含层神经元个数对网络的影响

Fig. 2 The influence of the number of neurons in the hidden layer on the network

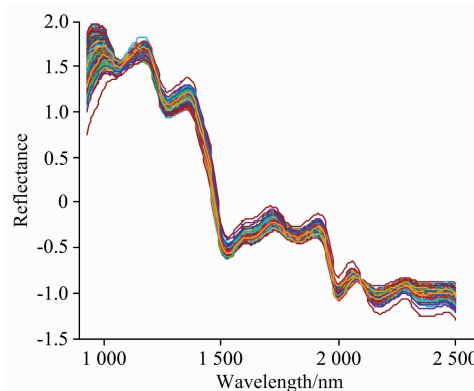


图 3 经过 SNV 后得到的反射光谱

Fig. 3 Reflectance spectra obtained by SNV pretreatment

2.2 光谱预处理

干制哈密大枣样本在光谱数据采集过程中易受到多种因素的干扰，需要对其做必要的预处理^[16]。分别使用标准正态

变换(SNV)、多元散射(MSC)和矢量归一化(Norm)等方法对原始光谱进行预处理来减弱或消除干扰因素,主要处理结果如表 3 所示。

表 3 不同光谱预处理方法的 SVM, ELM 和 GA-ELM 水分含量预测结果

Table 3 Prediction results of SVM, ELM and GA-ELM models by different spectral preprocessing methods

预测模型	预处理方法	校正集		预测集		RPD
		R_c	RMEC	R_p	RMSEP	
SVM	None	0.944 9	0.011 3	0.926 5	0.011 4	2.654 0
	SNV	0.976 5	0.007 5	0.953 0	0.002 0	3.318 4
ELM	None	0.897 3	0.015 2	0.883 7	0.014 1	2.136 1
	Norm	0.922 0	0.013 2	0.920 3	0.012 3	2.555 8
GA-ELM	None	0.971 9	0.008 1	0.928 3	0.011 2	2.689 6
	SNV	0.980 8	0.006 8	0.962 7	0.008 2	3.697 2

通过不同预处理方法处理后建立的 SVM, ELM 和 GA-ELM 建模比较分析,利用 SNV 处理后的 GA-ELM 建模效果最佳,相关系数 R_p 和 R_c 分别达到 0.980 8 和 0.962 7, RMSEC 和 RMSEP 分别达到 0.006 8 和 0.008 2, RPD 达到 3.697 2。因此选 SNV 处理后的光谱进一步进行特征波长筛选。

2.3 特征波长筛选

2.3.1 联合区间偏最小二乘法(si-PLS)

使用 si-PLS 方法将光谱数据划分为 20 个子区间,尝试联合 4 个区间筛选干制哈密大枣水分含量的特征波长。运行结果选择的最优子区间为 [3 11 13 15], 交互验证均方根误差(RMSECV)为 0.010 5, 同时该区间组合所对应的特征波长范围为 1 096.99 ~ 1 175.04, 1 769.98 ~ 1 846.85, 1 936.23 ~ 2 012.57 和 2 101.29 ~ 2 177.00 nm 共 52 个波长

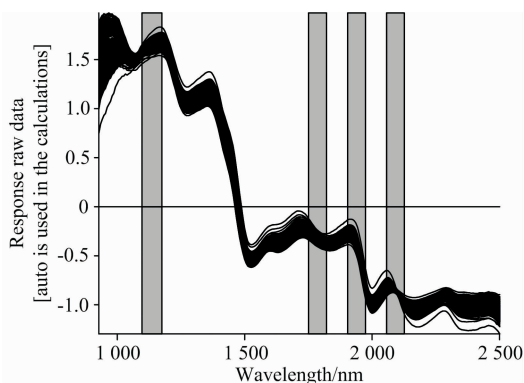


图 4 si-PLS 波长提取

Fig. 4 Wavelengths selected by si-PLS method

变量,占全波段的 21%,如图 4 所示。将选出的特征波长作为输入量,建立 GA-ELM 预测模型,结果如表 4 所示。

2.3.2 遗传算法(GA)

遗传算法(genetic algorithm, GA)是一种通过模拟自然进化过程搜索最优解的方法。本研究中,遗传算法的参数设定为:初始种群 50,变异概率 0.005,遗传迭代次数为 100 和收敛率为 0.5。图 5 为 GA 迭代所选光谱变量的频率图,

图 6 表示被选波长变量数所对应的 RMSECV 值的变化,图中绿色圆点表示当选取 115 个波长变量时,其 PLS 模型的 RMSECV 最低(0.01),所选取的波长变量占全波段的 46.5%,建立 GA-ELM 预测模型,结果如表 4 所示。

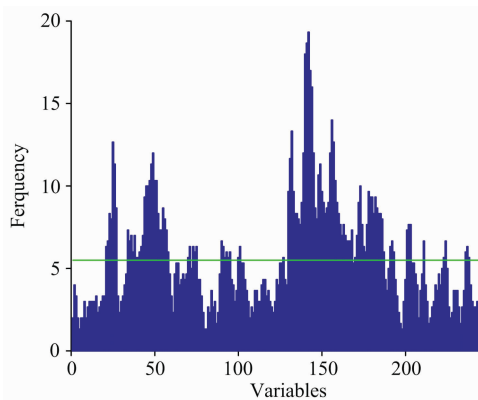


图 5 GA 筛选波段的频率分布

Fig. 5 Frequency distribution of bands selected by GA

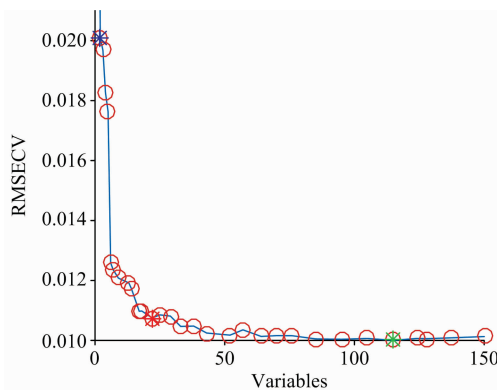


图 6 GA 波长提取

Fig. 6 Selection of wavelengths by GA

2.3.3 连续投影算法(SPA)

连续投影算法(successive projections algorithm, SPA)是一种新兴的特征波长筛选方法,能够有效消除波长变量之间共线性的影响,进而有效提取出特征波长变量^[14]。在干制哈密大枣水分含量特征波长筛选中,指定波长数 N 范围设定为 1~50。图 7 为 RMSECV 值随 SPA 选择变量数增加变化图,当 $m_{max}=14$ 时, RMSECV 值最低,模型拟合效果最佳,筛选出 14 个特征波长(927.77, 966.83, 1 090.48, 1 129.51, 1 298.49, 1 467.04, 1 680.03, 1 789.22, 1 840.45, 1 917.11, 1 974.44, 2 227.3, 2 314.97 和 2 464.1 nm)占全波长的 5.6%。将筛选的特征波长作为输入量建立 GA-ELM 预测模型, R_c 和 R_p 分别达到 0.984 2 和 0.967 5, RMSEC 和 RMSEP 分别达到 0.006 1 和 0.007 9, RPD 为 3.952 8。其中 R_p 较全波长预测模型提高 0.004 8, RMSEP 降低 0.000 3, RPD 提高了 0.255 6,说明 SPA 能够从全波长中删除与水分含量不相关的光谱信息,提取有效光谱信息,提高模型预测精度。图 8 表示所筛选的干制哈密大枣水分含量特征波段在全波段中的分布状况。

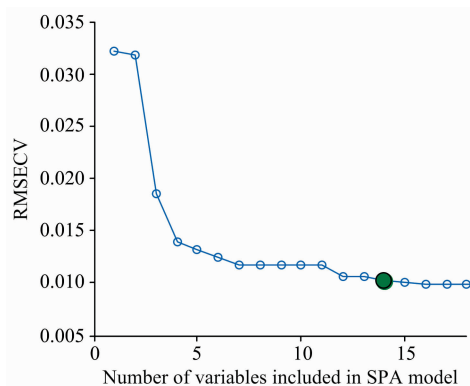


图 7 RMSECV 值随 SPA 选择变量数增加变化图
Fig. 7 RMSECV vs variable numbers selected by SPA

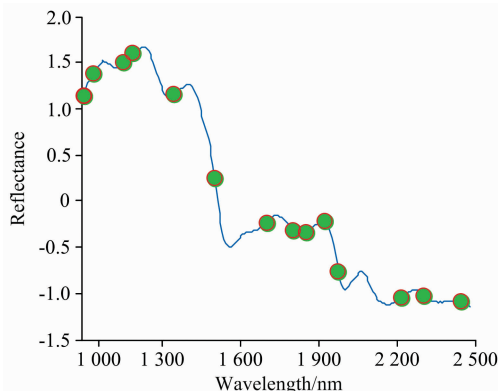


图 8 SPA 选择的 14 个特征波长分布图
Fig. 8 Plot of 14 Wavelengths selected by SPA

2.4 建立干制哈密大枣水分含量预测模型

为比较不同特征波长筛选方法对 GA-ELM 模型性能的影响,对 SNV 处理后的全波段光谱、GA、si-PLS 和 SPA 及组合算法提取的特征波长光谱数据分别建立对应的预测模型,结果如表 4 所示。

表 4 不同波长提取方法的水分含量 GA-ELM 建模结果
Table 4 Results of GA-ELM modeling for MC with different wavelength selection algorithms

方法	特征波长数	校正集		预测集		RPD
		R_c	RMSEC	R_p	RMSEP	
SNV	247	0.980 8	0.006 8	0.962 7	0.008 2	3.697 2
si-PLS	52	0.977 9	0.007 1	0.957 1	0.007 1	3.449 8
si-PLS+GA	42	0.974 8	0.007 6	0.957 4	0.009 3	3.464 3
si-PLS+SPA	11	0.977 7	0.007 2	0.935 8	0.011 5	2.836 7
GA	115	0.977 6	0.007 3	0.952 7	0.007 3	3.291 8
GA+SPA	13	0.982 8	0.006 3	0.962 3	0.006 3	3.678 8
SPA	14	0.984 2	0.006 1	0.967 5	0.007 9	3.952 8

结果显示,采用 SPA 和 GA-SAP 方法筛选出的 14 个和 13 个特征波长建立的 GA-ELM 模型均得到较高的预测结果,且 SPA+GA-ELM 预测结果中 R_c 和 R_p 的 0.984 2 和 0.967 5 均高于全波长的 0.980 8 和 0.962 7,进一步提高了

模型的精度和稳定性,图 9 和图 10 分别为 GA 优化进化过程和 ELM 预测结果。

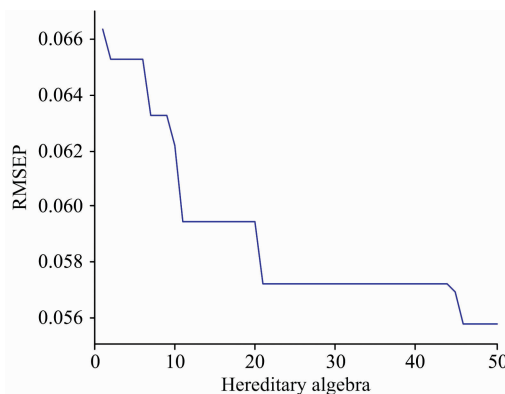


图 9 进化过程

Fig. 9 Evolutionary process

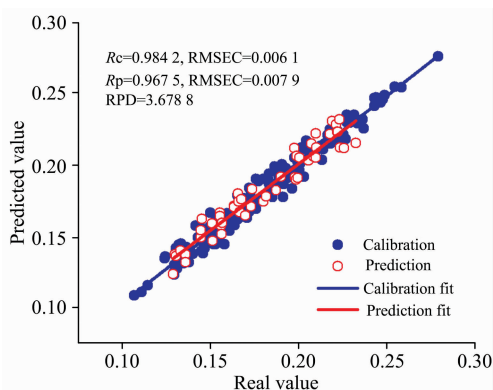


图 10 校正集和预测集干制哈密大枣水分含量预测值与真实值的关系

Fig. 10 Correlation of predicted value and actual value of dried Hami big jujubes MC for calibration set and validation set

3 结 论

采用近红外光谱结合 GA-ELM 神经网络建立干制哈密大枣水分含量预测模型,通过多种预处理方法结合波段筛选方法,能够有效提高模型的预测精度,实现干制哈密大枣水分含量的准确预测。以 220 个干制哈密大枣样本为研究对象,按照 170 : 50 划分校正集和验证集,使用多种预处理方法结合特征波长筛选方法 (si-PLS, GA, SPA 及它们的组合),分别建立 SVM, ELM 和 GA-ELM 定量预测模型。采用 SNV+SPA 组合提取的 14 个特征波长建立的 GA-ELM 预测模型效果最佳, R_c 和 R_p 分别为 0.984 2 和 0.967 5, RMSEC 和 RMSEP 分别为 0.006 1 和 0.007 9, RPD 为 3.678 8,得到了较为理想的预测结果。结果表明:GA-ELM 预测模型具有较好的鲁棒性和泛化能力,通过遗传算法对网络连接权值和阈值进行优化,改善了普通 ELM 连接权值和阈值的随意性所带来预测精度不稳定性。同时由于 ELM 神

经元个数的大幅度缩减,使得网络训练时间大大缩短,且预测精度的有效提高,对解决近红外光谱技术在农产品无损检测中数据量大、数据波动大的问题具有重要的意义。

References

- [1] SUN Jing-tao, MA Ben-xue, DONG Juan, et al(孙静涛, 马本学, 董娟, 等). Modern Food Science and Technology(现代食品科技), 2016, 32(9): 174.
- [2] ZHAO Huan-xia, ZHANG Hai-sheng, LI Qin, et al(赵换霞, 张海生, 李琴, 等). Science and Technology of Food Industry(食品工业科技), 2014, 35(3): 379.
- [3] Xie L, Wang A, Xu H, et al. Transactions of the ASABE, 2016, 59(2): 399.
- [4] YAN Yan-lu(严衍禄). Near Infrared Spectroscopy Principles, Technologies and Applications(近红外光谱分析的原理、技术与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2013.
- [5] Fernández-Espinosa A J. Talanta, 2016, 148: 216.
- [6] Alfatni M S M, Shariff A R M, Abdullah M Z, et al. Journal of Food Engineering, 2013, 116(3): 703.
- [7] HUANG Shuang-ping, HONG Tian-sheng, YUE Xue-jun, et al(黄双萍, 洪添胜, 岳学军, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(5): 132.
- [8] He Jianguo, Luo Yang, Liu Guishan, et al. Advanced Materials Research, 2013, 201.
- [9] HU Xiao-nan, PENG Yun-fa, LUO Hua-ping, et al(胡晓男, 彭云发, 罗华平, 等). Food Engineering(食品工业), 2015, 36(5): 232.
- [10] PENG Hai-gen, PENG Yun-fa, ZHAN Ying, et al(彭海根, 彭云发, 詹映, 等). Food Science and Technology(食品科技), 2014, 39(6): 276.
- [11] Anisur Rahman, Lalit Mohan Kandpal, Santosh Lohumi et al. Applied Sciences, 2017, 7(1): 109.
- [12] YANG Chuan-de, YU Hong-tao, GUAN Shu-yan, et al(杨传得, 于洪涛, 关淑艳, 等). Journal of Peanut Science(花生学报), 2012, 41(1): 6.
- [13] LI Jiang-bo, ZHAO Chun-jian, CHEN Li-ping, et al(李江波, 赵春江, 陈立平, 等). Journal of Agricultural Machinery(农业机械学报), 2013, 44(3): 153, 179.
- [14] ZHU Zhe-yan, LIU Fei, ZHANG Chu, et al(朱哲燕, 刘飞, 张初, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2014, 34(7): 1844.
- [15] Nicolai B N, Beullens K, Bobelyn E, et al. Postharvest Biology and Technology, 2007, 46(2): 99.
- [16] DAI Chun-xia, LIU Fang, GE Xiao-feng(戴春霞, 刘芳, 葛晓峰). Journal of Tea Science(茶叶科学), 2018, 38(3): 281.

Study on the Moisture Content of Dried Hami Big Jujubes by Near-Infrared Spectroscopy Combined with Variable Preferred and GA-ELM Model

WANG Wen-xia^{1,2}, MA Ben-xue^{1*}, LUO Xiu-zhi^{1,2}, LI Xiao-xia^{1,2}, LEI Sheng-yuan^{1,2}, LI Yu-jie^{1,2}, SUN Jing-tao³

1. College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, China

2. Key Laboratory of Northwest Agricultural Equipment, Ministry of Agriculture and Rural Affairs, Shihezi 832003 China

3. College of Food Science, Shihezi University, Shihezi 832003, China

Abstract Moisture content is an important index in the drying process of Hami big jujubes which has an important influence on its appearance, taste, storage and transportation. Therefore, in order to realize the accurate prediction of the moisture content of Hami big jujubes, GA-ELM prediction model of the moisture content of dried Hami big jujubes was studied by using Near-Infrared spectroscopy combined with variable preferred method. In order to improve the stability and prediction accuracy of the model, the effects of kernel function and the number of neurons on the GA-ELM prediction model were discussed. Various pretreatment methods were used to deal with the spectrum of the whole band. The comparison analysis denoted that the standard normal variation (SNV) method was the best. The characteristic wavelengths were screened from the range of 927.77~2 501.14 nm by combining with successive projection algorithm (SPA), the synergy interval partial least squares (si-PLS, genetic algorithm (GA) and their combination algorithms after processing of SNV. Respectively, the corresponding GA-ELM prediction model was established. The GA-ELM model with 14 characteristic wavelengths screened by SNV and SPA had the best effect while compared with the full-band GA-ELM model. Furthermore, the predicted results could be given as follows: R_c and R_p are 0.984

2 and 0.967 5, RMSEC and RMSEP are 0.006 1 and 0.007 9 while RPD is 3.678 8. The results denoted that the SNA+SPA+GA-ELM method can realize the accurate prediction of moisture content of dried Hami big jujubes and provide a reference for the application of near-infrared spectroscopy in the on-line detection of dried Hami big jujubes.

Keywords Near infrared spectroscopy; Hami big jujubes; Water content; Characteristic wavelengths; Extreme learning machine

(Received Dec. 17, 2018; accepted Apr. 9, 2019)

* Corresponding author

《光谱学与光谱分析》对来稿英文摘要的要求

来稿英文摘要不符合下列要求者,本刊要求作者重写,这可能要推迟论文发表的时间。

1. 请用符合语法的英文,要求言简意明、确切地论述文章的主要内容,突出创新之处。
2. 应拥有与论文同等量的主要信息,包括四个要素,即研究目的、方法、结果、结论。其中后两个要素最重要。有时一个句子即可包含前两个要素,例如“用某种改进的 ICP-AES 测量了鱼池水样的痕量铅”。但有些情况下,英文摘要可包括研究工作的主要对象和范围,以及具有情报价值的其他重要信息。在结果部分最好有定量数据,如检测限、相对标准偏差等;结论部分最好指出方法或结果的优点和意义。
3. 句型力求简单,尽量采用被动式,建议经专业英语翻译机构润色,与中文摘要相对应。用 A4 复印纸单面打印。
4. 摘要不应有引言中出现的内容,换言之,摘要中必须写进的内容应尽量避免在引言中出现。摘要也不要对论文内容作解释和评论,不得简单重复题名中已有的信息;不用非公知公用的符号和术语;不用引文,除非该论文证实或否定了他人已发表的论文。缩略语、略称、代号,除相邻专业的读者也能清楚地理解外,在首次出现时必须加以说明,例如用括号写出全称。