

doi:10.6041/j.issn.1000-1298.2022.02.024

# 基于 SPA - SSA - BP 的小麦秸秆含水率检测模型

孟志军<sup>1,2</sup> 刘淮玉<sup>1,2</sup> 安晓飞<sup>2,3</sup> 尹彦鑫<sup>2,3</sup> 金诚谦<sup>4</sup> 张安琪<sup>2,3</sup>

(1. 黑龙江八一农垦大学工程学院, 大庆 163319; 2. 国家农业智能装备工程技术研究中心, 北京 100097;

3. 北京市农林科学院智能装备技术研究中心, 北京 100097; 4. 农业农村部南京农业机械化研究所, 南京 210014)

**摘要:** 为提高基于电容法的小麦秸秆含水率检测模型的检测精度, 扩大含水率检测范围, 提高模型适应性, 本文以小麦秸秆为研究对象, 使用 LCR 数字电桥, 测量含水率为 10.43% ~ 25.89% 的秸秆在频率 0.05 ~ 100 kHz、容积密度 90.03 ~ 179.42 kg/m<sup>3</sup> 和温度 25 ~ 40℃ 内的电容, 利用连续投影法 (Successive projections algorithm, SPA) 和主成分分析法 (Principal component analysis, PCA) 对原始数据进行预处理, 提取特征频率, 选用反向传播神经网络 (Back propagation neural network, BPNN) 在全频率及 2 个特征频率下分别建立秸秆含水率、容积密度、温度的定量分析模型, 引入麻雀搜索算法 (Sparrow search algorithm, SSA) 优化反向传播神经网络模型。试验结果表明, 基于全频率构建的模型较基于 SPA 算法构建的模型预测效果略好, 综合考虑模型复杂度和预测性能, 本研究选用基于 SPA 算法结合 SSA 算法优化后的 BP 神经网络模型 (SPA - SSA - BP) 作为小麦秸秆含水率的检测模型, 其预测集  $R_p^2$ 、RMSEP 和 RPDP 分别为 0.983 2、0.005 50 和 7.715。利用该模型对 13 个含水率为 10.62% ~ 25.59% 的秸秆样本进行预测, 含水率预测结果的相对误差为 -5.27% ~ 5.52%, 其中 96.8% 的预测误差在 ±5% 以内。由此说明, 模型具有较高的准确性和较好的鲁棒性。

**关键词:** 小麦; 秸秆; 含水率; 检测模型; 电容; 麻雀搜索算法

中图分类号: S233

文献标识码: A

文章编号: 1000-1298(2022)02-0231-08

OSID:



## Prediction Model of Wheat Straw Moisture Content Based on SPA - SSA - BP

MENG Zhijun<sup>1,2</sup> LIU Huaiyu<sup>1,2</sup> AN Xiaofei<sup>2,3</sup> YIN Yanxin<sup>2,3</sup> JIN Chengqian<sup>4</sup> ZHANG Anqi<sup>2,3</sup>

(1. College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China

2. National Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China

3. Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

4. Nanjing Research Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China)

**Abstract:** In order to improve the detection accuracy of the wheat straw moisture content prediction model based on capacitance method, expand the detection range of moisture content and improve the adaptability of the model, taking wheat straw as the research object and using LCR digital bridge, the capacitance data of straw with 10.43% ~ 25.89% moisture content were measured in the frequency range of 0.05 ~ 100 kHz, volume density range of 90.03 ~ 179.42 kg/m<sup>3</sup> and temperature range of 25 ~ 40℃. The original data were preprocessed by using the successive projections algorithm (SPA) and principal component analysis (PCA) to extract characteristic frequencies, BP neural network was used to establish quantitative analysis models of straw moisture content, volume density, temperature and capacitance at full frequency and two characteristic frequencies respectively, and sparrow search algorithm (SSA) was introduced to optimize the BP neural network model. The experimental results showed that the prediction effect of the model based on full frequency was slightly better than that of the model based on SPA algorithm. Considering the model complexity and prediction performance, the BP neural network model (SPA - SSA - BP) optimized based on SPA algorithm and SSA algorithm was selected as the prediction model of wheat straw moisture content. The  $R_p^2$ , RMSEP and RPDP of prediction sets were 0.983 2,

收稿日期: 2021-11-05 修回日期: 2021-11-25

基金项目: 国家重点研发计划项目(2019YFB1312304)、北京市农林科学院创新能力建设专项(KJXC20200416)和江苏省农业科技自主创新专项(CX(20)1007)

作者简介: 孟志军(1975—),男,研究员,博士,主要从事农机智能装备研究,E-mail: mengzj@nercita.org.cn

通信作者: 张安琪(1989—),男,高级工程师,博士,主要从事农机智能装备研究,E-mail: zhangaq@nercita.org.cn

0.005 50 and 7.715, respectively. The model was used to predict 13 straw samples with water content ranging from 10.62% to 25.59%, and the relative error of water content prediction results was within  $-5.27\%$  to  $5.52\%$ , 96.8% of which was within  $\pm 5\%$ . This showed that the model had high accuracy and good robustness, and the method can provide an idea and theoretical reference for other crop straw water content prediction.

**Key words:** wheat; straw; moisture content; prediction model; capacitance; sparrow search algorithm

## 0 引言

秸秆是农业生产过程中重要的生物质资源,其富含氮、磷、钾、镁、钙等重要元素以及粗纤维和有机质,是一种具有多用途的可再生生物资源,具有很大的利用价值。对秸秆资源的开发利用,不仅可有效改善人居环境,还能在一定程度上缓解全球能源紧缺的问题<sup>[1-2]</sup>。含水率是衡量秸秆品质的重要指标,含水率在一定程度上决定了秸秆的利用价值<sup>[3-5]</sup>。由于过去对其不够重视及缺乏相关检测技术,常造成不必要的经济损失和资源浪费。因此,研究一种快速、准确的小麦秸秆含水率检测方法,对于秸秆资源的高效利用具有重要意义。

目前,应用于秸秆含水率检测的方法主要有电容法、电阻法、近红外光谱法、微波法等。皇才进等<sup>[6]</sup>采用近红外光谱技术结合 LOCAL 算法建立秸秆含水率(5.13%~20.44%)和热值的近红外光谱模型;FALBO<sup>[7]</sup>基于电阻法设计了平板板式含水率在线检测系统;万舟等<sup>[8]</sup>基于微波法通过对微波的衰减量和相移量的测量可测得秸秆的水分含量;郭文川等<sup>[9]</sup>基于电容法探究秸秆含水率(10.6%~19.6%)、温度(5~35℃)和容积密度(77.2~103.6 kg/m<sup>3</sup>)对输出电容的影响;FONSECA 等<sup>[10]</sup>采用近红外光谱仪对秸秆含水率进行检测,研究表明旋转扫描方法检测精度更加稳定。其中,电容法具有适应性强、可靠性高、动态响应性好和结构简单易维护等优点<sup>[11]</sup>,是目前农业物料含水率检测的主要技术手段。

基于电容法的秸秆含水率检测研究多考虑温度、含水率对电容的影响,较少考虑容积密度因素,且已有模型多为线性模型<sup>[9]</sup>,存在容积密度适用上限低、含水率检测范围窄等问题。基于此,本文采用电容法原理,通过设计电容采集装置,使用 LCR 数字电桥采集不同含水率小麦秸秆在不同频率、不同容积密度与不同温度下的电容数据,采用 BP 神经网络结合特征频率筛选算法构建秸秆含水率定量分析模型,并引入麻雀搜索算法(SSA)对模型进行优化,分析模型的预测效果,确定小麦秸秆含水率的最佳建模方法。以实现小麦秸秆含水率的快速、准确定量分析。

## 1 材料与方法

### 1.1 试验材料

以小麦秸秆为试验对象,秸秆样本于2021年6月在国家精准农业研究示范基地采集。由于试验盒内部尺寸为180 mm×125 mm×70 mm,为保证每份装入盒中的样本密度均匀,不因过于疏松或致密导致试验盒中不同样本的质量差距过大,因此将采集的秸秆粉碎成长度(30±5) mm的枝干状,并随机分为56个样本,每个样本80 g,装于自封袋中保存在25℃的室温环境下。

### 1.2 试验方案

#### 1.2.1 电容采集装置设计

电容采集装置结构如图1所示,主要包括亚克力试验盒,TH2830型LCR数字电桥及配套数据采集软件(常州同惠电子股份有限公司),DHG-9240A型鼓风干燥箱(上海一恒科学仪器有限公司),WDW-01S型微机控制电子万能拉力试验机及配套软件(常州三丰仪器科技有限公司),MTB2000D型精准电子天平(深圳市美孚电子有限公司),西玛-AT1150型红外线测温仪(东莞万创电子制品有限公司)。

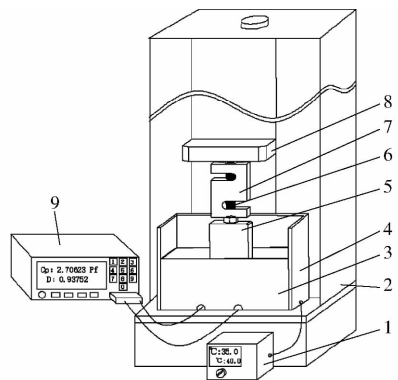


图1 电容采集装置结构示意图

Fig.1 Schematic of capacitance acquisition device

1. 温控箱 2. 拉压力试验机 3. 亚克力试验盒 4. 硅橡胶加热板 5. 支柱 6. 螺栓 7. 拉压力传感器 8. 升降台 9. 数字电桥

亚克力试验盒内部尺寸为180 mm×125 mm×70 mm,将待测秸秆在亚克力试验盒内填满,为使得秸秆样本各处受力均匀,将压板覆盖于秸秆样本上;设计2块尺寸为120 mm×50 mm×2 mm的铜板对称布置在试验盒底部,二者相距3 mm,用于采集秸

秆电容数据;设计硅橡胶加热板包裹在试验盒外壁,通过调节温控箱按钮,预设试验环境温度,达到改变试验盒内温度的目的;用螺栓将试验机压力台、拉压力传感器和支柱连接到一起,通过上位机软件控制试验机压力台升降,实现对试验盒内秸秆压力的调控,以达到改变秸秆容积密度的目的。

### 1.2.2 样品制备

从秸秆样本中随机取出1份并置于105℃的电热鼓风干燥箱中4h至质量恒定,测得小麦秸秆的初始湿基含水率为10.43%<sup>[12]</sup>。为配制不同含水率的样本,取56个(每个80g)秸秆样本,通过添加不同质量的去离子水,得到56个不同含水率的样本并编号。为保证样品吸收水分均匀,将配好的样品装入PE自封袋中,置于25℃的室温环境下1~2d。期间每天取出样本2~3次,充分搅拌后倒回袋中,以使水分分布均匀。

### 1.2.3 数据获取

电容数据采集前,先将室内温度调至25℃恒定,将LCR数字电桥测量电极与2块铜板外接线路相连,再开机预热30min并校准清零。采集时,由于试验盒容积所限,从样本中取出40g左右秸秆填满试验盒即可,剩余样本装回袋中继续密封;通过温控箱将试验环境温度依次设置为25、30、35、40℃,使用手持式红外线测温仪对电容采集装置内部温度进行定期测量;通过万能拉压力试验机改变对秸秆压力的方式来调节秸秆的容积密度,将试验机对秸秆的压力设置在500~3600N之间,故秸秆的容积密度在90.08~179.42kg/m<sup>3</sup>范围内,每个秸秆样本在同一温度下从上述容积密度范围内随机选取3个容积密度,确保所有样本组合选取的容积密度在90.03~179.42kg/m<sup>3</sup>之间均匀分布;在0.05~100kHz共取100个呈对数正态分布的频率点,作为LCR数字电桥的检测频率,测量待测秸秆样本在上述不同容积密度、温度、频率点组合下的电容数据。

秸秆含水率测量装置主要由电子天平、电热鼓风干燥箱组成。测量前,先将电子天平开机预热15min后并校准清零,测量时,将试验盒中样本与自封袋中剩余样本充分混合后分为2份,并分别置于105℃的电热鼓风干燥箱内干燥4h,根据干燥前样本的鲜质量与干燥后样本的干质量可计算出各样本的实际含水率。

## 1.4 数据处理及建模方法

### 1.4.1 样本集划分

试验共配制了56份样本,样本的湿基含水率在10.43%~25.89%之间,每个样本又在4个温度及3个容积密度条件下分别采集100频率点下的电

容,共得到672组电容数据。通过对数据进行观察,采用马氏距离法剔除一些因仪器设备和试验操作原因产生的异常数据,剩余656组数据。采用留出法,随机选取样本数据中3/4的数据作为建模的校正样本,另外1/4的数据作为测试样本。划分结果见表1。

表1 校正集和预测集统计结果

Tab.1 Statistics results of moisture content in calibration and prediction sets

样本集	样本数	平均值/ %	最大值/ %	最小值/ %	标准偏差/%
校正集	492	18.76	25.89	10.43	4.243
预测集	164	18.73	25.89	10.43	4.246

### 1.4.2 数据处理方法

因研究中所选的多个频率间可能会存在较强的线性相关,它们会包含较多的冗余信息,增加模型的复杂度,通过特征提取方法能够从原始数据中提取主要信息,达到在不损失过多原频率信息的基础上,降低模型的复杂程度<sup>[13-14]</sup>。本研究采用连续投影法(SPA)及主成分分析法(PCA)来选取特征频率。SPA是一种能够有效解决变量间共线性问题的变量筛选方法,利用向量投影可优选出冗余度低、共线性小又能反映样本频谱重要信息的特征频率。PCA是一种被广泛使用的数据降维算法。其主要思想是通过某种线性投影,将高维空间的数据映射到低维空间中,并通过计算特征频率的方差优选出信息量大的特征频率,以达到使用较少的数据维度就可保留住较多信息的目的<sup>[15]</sup>。PCA是对频谱数据进行主成分分析,在累计贡献率大的前几个主成分相应的权值系数曲线中,权值系数的绝对值与其相对应频率的贡献程度成正比,故选择曲线中波峰和波谷处对应的频率为特征频率<sup>[16]</sup>。

### 1.4.3 建模方法

常用的建模方法主要分为线性方法和非线性方法。线性建模方法主要有偏最小二乘回归(PLSR)、多元线性回归(MLR)和主成分回归等;非线性建模方法主要有支持向量回归(SVR)<sup>[17]</sup>、BP神经网络等。考虑到试验数据量大,线性建模方法不能很好地拟合非线性数据,支持向量回归(SVR)不善于处理大量数据。因此,本文选用适合处理大量数据,并且非线性拟合能力强的BP神经网络建立含水率检测模型。

BP神经网络是一种多层前馈神经网络。该算法处理信息的神经元可分为3层:输入层、隐含层、输出层,每层的神经元状态只影响下一层神经元状态。其主要特点是信号向前传递,误差反向传播,通

过反向传播来不断调整网络的权值和阈值,使神经网络的误差平方和趋于最小<sup>[18]</sup>,该算法具有很强的非线性映射能力,可映射任意复杂的非线性关系,并具有很强的鲁棒性和自适应能力。

影响小麦秸秆电容的主要因素有秸秆含水率、容积密度和环境温度,因此,设定小麦秸秆容积密度、环境温度和选取的特征频率为神经网络的输入参数,小麦秸秆含水率为输出参数,如图2所示,采用单隐含层的3层网络结构,隐含层节点数计算公式为

$$l < \sqrt{a(m+n)} + a \quad (1)$$

式中  $n$ ——输入层节点数

$l$ ——隐含层节点数

$m$ ——输出层节点数

$a$ ——1~10之间的常数

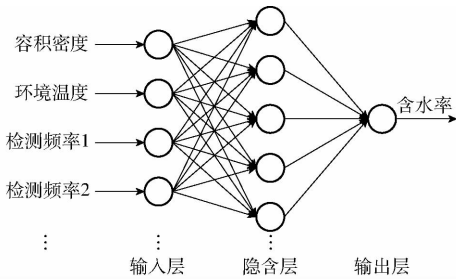


图2 BP神经网络结构

Fig.2 BP neural network structure

通过反复试验确定输入层到隐含层的传递函数为S型正切函数  $\text{tansig}$ ,隐含层到输出层的传递函数为线性函数  $\text{purelin}$ 。常用的训练函数包括  $\text{trainlm}$ 、 $\text{trainrp}$ 、 $\text{trainsecg}$  等,其中  $\text{trainlm}$  具有收敛速度快、误差小、训练效果优的特点,因此本文采用  $\text{trainlm}$  作为训练函数<sup>[19]</sup>。

BP神经网络训练前,为减弱各主控因素不同量纲的数据对网络模型训练与预测值的影响,各主控数据做归一化处理。计算式为

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min} \quad (2)$$

式中  $x$ ——原始数据  $y$ ——归一化值

$x_{\max}$ ——同一影响因素响应值的最大值

$x_{\min}$ ——同一影响因素响应值的最小值

$y_{\max}$ ——归一化最大值,取1

$y_{\min}$ ——归一化最小值,取0

在此基础上,采用BP神经网络结合特征频率筛选算法构建基于容积密度、温度和电容的秸秆含水率定量分析模型,以预测集决定系数  $R_p^2$ 、预测集均方根误差(Root mean square error of prediction, RMSEP)和预测集相对分析误差(Residual predictive deviation of prediction, RPDP)作为评价模型性能的

指标。 $R_p^2$ 用于评估样本含水率预测值与实际值之间的相关程度, $R_p^2$ 越接近于1,预测值与实际值之间的相关程度越好;RMSEP用于评价模型对测试集的预测能力,RMSEP越小,表明模型的泛化能力越强;RPDP用于评价所建模型的稳定性,RPDP越大,表明模型稳定性越好,通常RPDP大于2.5时,模型精度较优<sup>[20]</sup>。

#### 1.4.4 麻雀搜索算法优化BP神经网络模型

麻雀搜索算法(SSA)是一种新型的群智能优化算法,该算法主要是受麻雀觅食行为和逃避捕食者行为启发而设计<sup>[21-24]</sup>。麻雀在觅食过程中会分为3种类型:发现者、加入者和侦察者,利用这三者间的关系及麻雀遇到捕食者时的行为可达到优化搜索的目的。同近年来新兴的群智能优化算法相比,麻雀搜索算法具有较好的全局搜索和局部开发的能力,在寻优过程中,能促进麻雀种群向全局最优值移动,有效避免易早熟收敛、收敛速度慢等缺点,具有良好的鲁棒性和收敛速度。整体过程如图3所示。

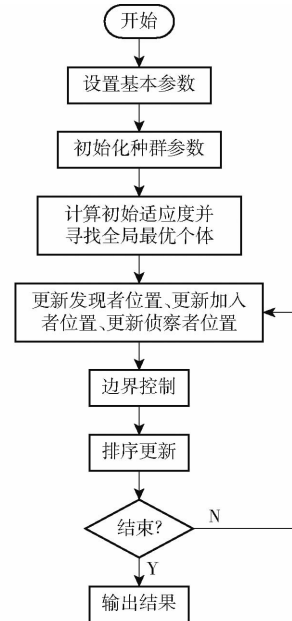


图3 麻雀优化算法流程图

Fig.3 Process chart of sparrow optimization algorithm

SSA-BP算法的实施过程如下:

(1)读取数据。确定BP神经网络模型的校正集和测试集样本,对数据进行归一化处理,将本研究中不同量纲数据归一化到0~1之间,并根据式(1)确定最佳隐含层节点数。

(2)网络参数配制。将训练次数设置为1000次,学习速率设置为0.01,训练目标最小误差设置为0.0001。

(3)初始化参数。设置SSA算法的初始种群规模  $N$  和最大迭代次数  $T$ ,本文将初始种群规模  $N$  设为30次,最大迭代次数  $T$  设为50次;设定种群中不

同类型的麻雀比重和安全值, 本文将安全值  $S_T$  设为 0.6, 发现者数量  $N_{pd}$  设为 0.7, 侦察者数量设为 0.2。

(4) 计算初始适应度。根据适应度计算出全局中最优个体。

(5) 根据适应度将麻雀种群分类, 并对不同类型的麻雀个体位置进行更新。

(6) 若迭代次数达到最大迭代次数, 学习过程结束, 输出最优参数和适应度, 否则返回步骤(5)重复上述过程。

## 2 结果与分析

### 2.1 各因素对秸秆电容的影响

图 4 为不同含水率的秸秆在相同环境温度和容积密度条件下电容随频率的变化曲线。在频率 0.05 ~ 100 kHz 范围内, 秸秆的电容随频率的增大而减小。这是因为随着频率的不断增大导致偶极子的振动速度滞后于电场的变化, 所以样本的电容会随着介电常数的减小而不断减小<sup>[25]</sup>。同一频率下, 不同样本的电容受湿基含水率的影响不同, 主要表现为: 秸秆样本湿基含水率越高, 电容越高。

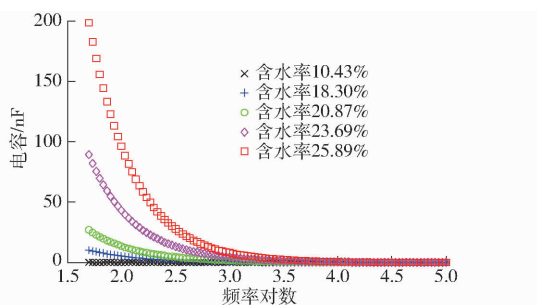


图 4 不同含水率秸秆在不同频率下电容变化曲线  
Fig.4 Capacitance curves of straw with different moisture contents at different frequencies

图 5 为容积密度和温度对秸秆电容的影响曲面。由图 5a 可以看出, 当频率和温度一定时, 随着容积密度的升高, 秸秆电容呈单调递增的趋势, 这是因为秸秆受到挤压后密度增大, 单位体积内秸秆量随之增加, 并可以储存更多的电场能, 所以测量仪器会测得更大的电容<sup>[26]</sup>。由图 5b 可以看出, 当频率和容积密度一定时, 秸秆的电容随温度的升高而增加, 这是因为温度升高有利于加速秸秆内极性分子的取向运动和自由水的布朗运动, 致使秸秆的相对介电常数增加, 故秸秆电容也随之增加<sup>[27]</sup>。

### 2.2 特征频率选取

#### 2.2.1 连续投影法特征频率选取

为保证模型性能可靠, 设置选取的频率数量为 2 ~ 30, 以不同频率数量所对应的 RMSE 作为最佳的特征频率数的指标。当选取的频率数量为 5 时, RMSE 最低(0.025 002), 如图 6 所示。考虑到过多

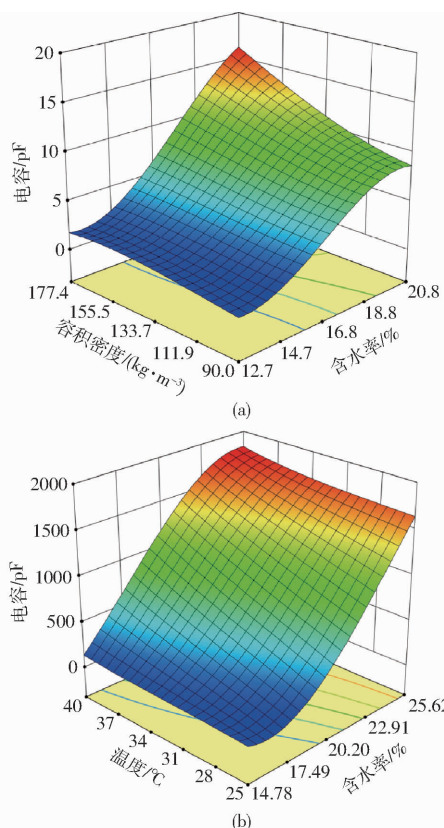


图 5 秸秆容积密度和温度对电容的影响曲面  
Fig.5 Influence of bulk density and temperature of straw on capacitance

的特征频率会导致模型复杂度上升, 故选取电容的 5 个频率作为 SPA 选取的特征频率。SPA 算法选取的特征频率如图 7 所示。

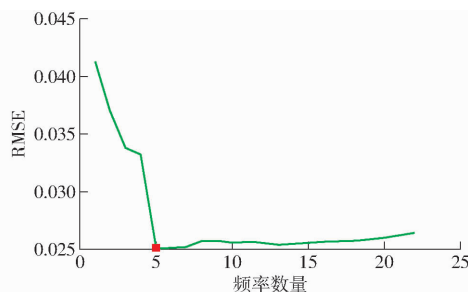


图 6 RMSE 随 SPA 选取频率数量的变化曲线  
Fig.6 Change of RMSE with selected characteristic frequency by SPA

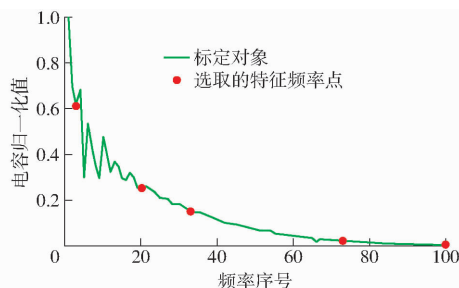


图 7 采用 SPA 算法选取的特征频率点  
Fig.7 Characteristic frequency points selected by SPA algorithm

### 2.2.2 主成分分析法特征频率选取

在将 PCA 算法用于原始频率数据的提取过程中,主成分分析得到的前 3 个主成分累计方差贡献率接近 100%,说明前 3 个主成分能够较为全面地反映绝大部分原始信息,从前 3 个主成分的权值系数曲线中提取 4 个特征频率。采用 PCA 法选取的特征频率如图 8 所示。全变量、SPA 和 PCA 所选出的特征频率如表 2 所示。

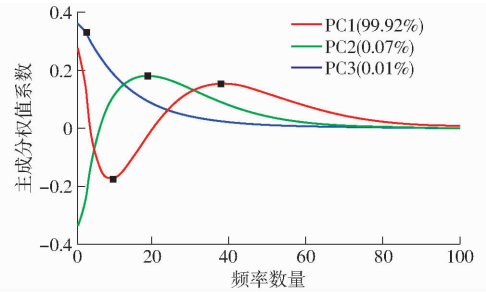


图 8 采用 PCA 算法选取的特征频率

Fig. 8 Characteristic frequency selected by PCA algorithm

表 2 SPA 与 PCA 选取的特征频率

Tab. 2 Characteristic frequency selected by successive projections algorithm and principal component analysis

频率选取方法	频率数量	频率/kHz
全频率	100	0.050、0.054、0.058、0.063、0.068、0.073、0.079、0.086、0.092、0.100、0.108、0.116、0.126、0.136、0.146、0.158、0.171、0.184、0.199、0.215、0.232、0.251、0.271、0.292、0.316、0.341、0.368、0.397、0.429、0.463、0.500、0.540、0.583、0.630、0.680、0.735、0.793、0.856、0.925、0.999、1.078、1.164、1.257、1.358、1.466、1.583、1.709、1.846、1.993、2.152、2.324、2.509、2.709、2.925、3.159、3.411、3.683、3.977、4.294、4.637、5.007、5.407、5.838、6.304、6.807、7.350、7.937、8.570、9.254、9.993、10.790、11.651、12.581、13.585、14.669、15.840、17.104、18.469、19.943、21.534、23.253、25.108、27.112、29.275、31.611、34.134、36.858、39.799、42.975、46.405、50.108、54.107、58.424、63.087、68.121、73.557、79.427、85.766、92.610、100.000
SPA	5	0.058、0.215、0.583、12.581、100.000
PCA	4	0.058、0.100、0.199、0.856

### 2.3 建模并比较分析

为了选取最佳检测模型,以全频率、SPA 和 PCA 分别选取的特征频率与容积密度、环境温度组合作为建模分析的自变量,秸秆样本含水率为因变量。选用 BP 神经网络算法分别与上述 3 种变量组合构建模型,建模结果如表 3 所示。

表 3 基于不同特征频率选取方法的 BP 建模结果

Tab. 3 BP modeling results based on different characteristic frequency selection methods

频率选取方法	频率数量	校正集			预测集		
		$R_c^2$	RMSEC	RPDC	$R_p^2$	RMSEP	RPDP
全频率	100	0.983 2	0.005 44	7.715	0.978 6	0.006 14	6.836
SPA	5	0.983 2	0.005 46	7.715	0.980 5	0.005 90	7.161
PCA	4	0.944 1	0.009 68	4.230	0.948 8	0.009 34	4.419

从表 3 不同模型预测结果来看,采用 SPA 选取特征频率与容积密度、环境温度组合建立的检测模型最佳,模型性能评价指标  $R_p^2$ 、RMSEP 和 RPDP 分别为 0.980 5、0.005 90 和 7.161,较其他特征频率所建模型有更高的含水率预测精度。

从特征频率选取角度来看,基于 SPA 构建的模型较基于 PCA 构建的模型更好,其中校正集  $R_c^2$  提高了 0.039 1, RMSEC 降低了 0.004 22, RPDC 提高了 3.485;预测集  $R_p^2$  提高了 0.031 7, RMSEP 降低了 0.003 44, RPDP 提高了 2.742。

从模型复杂程度来看,基于全频率构建的 BP 模型中存在大量冗余信息,增加了模型复杂程度。基于 SPA 和 PCA 算法能有效提取全频率中的重要信息,会大大简化模型复杂程度,减少模型的运算量并提高程序运行速度。

建模结果表明,提取特征频率有效地减少大量冗余信息和损害模型的信息,在大幅降低模型复杂度的基础上,依旧保持较高的预测精度。因此,基于电容法采用 SPA 和 PCA 提取特征频率并分别与容积密度、环境温度组合建立的 BP 神经网络模型均具有较高的含水率预测精度和可靠性。

### 2.4 SSA 算法优化模型

为进一步探究 SSA 算法对 BP 模型预测精度的影响,引入 SSA 算法对基于不同特征频率选取方法构建的 BP 模型进行优化,并继续比较分析。

由表 3 和表 4 可以看出,SSA-BP 建模方法中模型  $R^2$  较 BP 建模方法中模型的  $R^2$  更高, RMSE 更低, RPD 则更高,这表明经 SSA 算法优化后的模型

表 4 SSA-BP 模型结果

Tab. 4 Results of SSA-BP model

频率选取方法	频率数量	校正集			预测集		
		$R_c^2$	RMSEC	RPDC	$R_p^2$	RMSEP	RPDP
全频率	100	0.990 4	0.004 16	10.206	0.983 7	0.005 43	7.833
SPA	5	0.985 9	0.005 03	8.422	0.983 2	0.005 50	7.715
PCA	4	0.977 6	0.006 32	6.682	0.976 8	0.006 48	6.565

具有更高的预测精度和可靠性。

通过对比可知,采用全频率构建的模型预测效果最好,预测集  $R_p^2$ 、RMSEP 和 RPDP 分别为 0.983 7、0.005 43 和 7.833。基于 SPA 构建的模型预测精度与基于全频率构建的定量分析模型相差不大,并且较基于 PCA 构建的模型预测效果略好,预测集的  $R_p^2$  高出 0.006 4, RMSEP 降低了 0.000 98, RPDP 提高了 1.150。考虑到基于全频率构建的模型存在大量的冗余信息,模型过于复杂,本研究选用基于 SPA 算法提取特征频率,并结合 SSA 算法优化后的 BP 神经网络构建秸秆含水率检测模型。

## 2.5 模型验证

为探究最佳模型对不同秸秆样本含水率的预测效果,随机配制了含水率为 10.62% ~ 25.59% 的 13 个秸秆样本,将环境温度分别设置为 25、30、35、40℃,在容积密度 90.03 ~ 179.42 kg/m<sup>3</sup> 范围内随机选择 4 个容积密度和频率 0.058、0.215、0.583、12.581、100.000 kHz 下采集上述秸秆样本的电容,共形成 156 组数据,对模型进行验证。图 9 为 SPA - SSA - BP 的预测结果,可以看出样本集中于回归线 ( $y = x$ ) 附近,预测效果较佳。因此,最终选择 SPA - SSA - BP 作为小麦秸秆含水率的检测模型。图 10 统计了样本含水率预测值与烘干法测得的实际值的相对误差,相对误差为 -5.27% ~ 5.52%,其中 96.8% 的预测误差集中在  $\pm 5\%$  范围内,说明该模型具有较高的准确性和较好的鲁棒性。

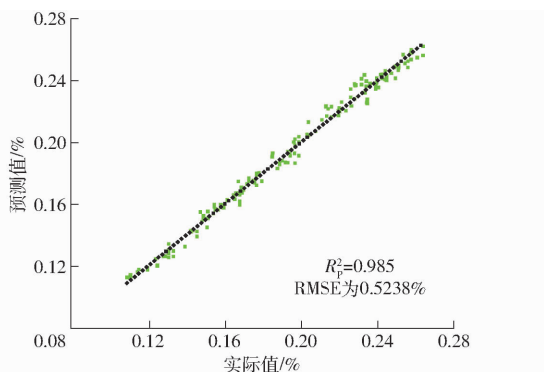


图 9 SPA - SSA - BP 模型的含水率预测结果

Fig. 9 Moisture content predicted results of SPA - SSA - BP model

## 3 结论

(1) 探究了不同含水率、容积密度、环境温度、频率下秸秆电容的变化规律,在含水率 10.43% ~

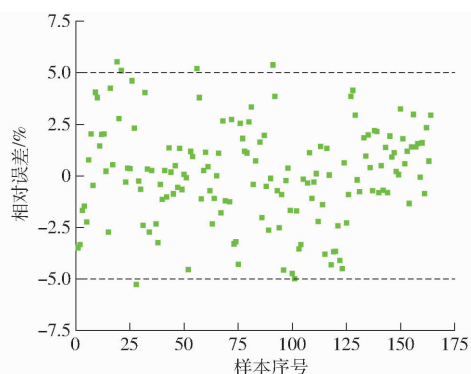


图 10 模型对不同含水率秸秆的预测相对误差

Fig. 10 Model prediction relative error of different moisture contents of straw

25.89% 范围内,电容随样本含水率的增大而增大;在频率 0.05 ~ 100 kHz 范围内,电容随频率的增大而减小,当频率大于 1 kHz 时,减小趋势平缓;在容积密度 90.03 ~ 179.42 kg/m<sup>3</sup> 范围内,电容随容积密度的增大而增大;在温度 25 ~ 40℃ 范围内,秸秆电容随环境温度的升高而增大。结果表明含水率、容积密度和环境温度对电容的影响极显著。

(2) 利用连续投影法 (SPA) 和主成分分析法 (PCA) 对原始数据进行预处理,提取特征频率,采用 BP 神经网络结合特征频率筛选算法构建基于容积密度、温度和电容的秸秆含水率定量分析模型,与全频率模型相比,特征频率下的模型不仅复杂程度大幅降低,模型的预测性能也有所提高。其中,基于 SPA 选取 5 个特征频率后构建的检测模型最佳,其预测集  $R_p^2$  为 0.980 5, RMSEP 为 0.590%, RPDP 为 7.161。

(3) 引入麻雀搜索算法优化 BP 神经网络模型,该模型采用 3 层网络结构。试验结果表明,选用基于 SPA 算法结合 SSA 算法优化后的 BP 神经网络模型 (SPA - SSA - BP) 作为小麦秸秆含水率的检测模型,其预测性能最好,预测集  $R_p^2$ 、RMSEP 和 RPDP 分别为 0.983 2、0.005 50 和 7.715。

(4) 对 13 个含水率为 10.62% ~ 25.59% 的秸秆样本进行预测,结果表明,模型预测结果相对误差为 -5.27% ~ 5.52%,其中 96.8% 的预测误差集中在  $\pm 5\%$  范围内,模型具有较高的准确性和较好的鲁棒性,本文提出的方法可进一步提高小麦秸秆含水率检测模型的检测精度,并为其他作物秸秆含水率预测提供了思路和理论参考。

## 参 考 文 献

[1] 许山晶,尹晓青. 我国农村秸秆资源利用的综合效应评价[J]. 重庆社会科学, 2021(2): 19 - 32.

XU Shanjing, YIN Xiaqing. The effect evaluation of comprehensive utilization of straw resources in China[J]. Chongqing Social

- Sciences, 2021(2):19–32. (in Chinese)
- [2] 毕于运,高春雨,王红彦,等. 我国农作物秸秆离田多元化利用现状与策略[J]. 中国农业资源与区划, 2019,40(9):1–11. BI Yuyun, GAO Chunyu, WANG Hongyan, et al. Research on the present situation and corresponding strategies of off-field straw collection and comprehensive utilization in China[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2019,40(9):1–11. (in Chinese)
- [3] 廖娜,韩鲁佳,黄光群,等. 含水率和压缩频率对秸秆开式压缩能耗的影响[J]. 农业工程学报, 2011,27(增刊):318–322. LIAO Na, HAN Lujia, HUANG Guangqun, et al. Effects of moisture content and compression frequency on straw open compression energy consumption[J]. Transactions of the CSAE, 2011,27(Supp.):318–322. (in Chinese)
- [4] 陈树人,蒋成宠,姚勇,等. 水稻秸秆压块热值模型构建及其影响因子相关性分析[J]. 农业工程学报, 2014,30(24):200–208. CHEN Shuren, JIANG Chengchong, YAO Yong, et al. Establishment of predicted model of calorific value for rice straw briquetting and analysis of correlation of its influencing factors[J]. Transactions of the CSAE, 2014,30(24):200–208. (in Chinese)
- [5] GILL N, DOGRA R, DOGRA B. Influence of moisture content, particle size, and binder ratio on quality and economics of rice straw briquettes[J]. Bioenergy Research, 2018,11(1):54–68.
- [6] 皇才进,韩鲁佳,刘贤,等. 基于 LOCAL 算法的秸秆水分和热值的近红外光谱快速分析[J]. 红外与毫米波学报, 2009,28(3):184–187. HUANG Caijin, HAN Lujia, LIU Xian, et al. Rapid measurement of moisture and calorific value of straw by near infrared spectroscopy with local algorithm[J]. Journal of Infrared and Millimeter Waves, 2009,28(3):184–187. (in Chinese)
- [7] FALBO J. System and method for determining moisture content in a bale of hay: 112562652[P]. 2010–03–24.
- [8] 万舟,黄益锦,吴建德,等. 一种基于微波的秸秆水分检测装置: 201020670716.X[P]. 2011–08–31.
- [9] 郭文川,刘驰,杨军. 小麦秸秆含水率测量仪的设计与试验[J]. 农业工程学报, 2013,29(1):33–40. GUO Wenchuan, LIU Chi, YANG Jun. Design and experiment on wheat straw moisture content meter[J]. Transactions of the CSAE, 2013,29(1):33–40. (in Chinese)
- [10] FONSECA F G, FUNKE A, SAECHUA W, et al. Precision test for the spectral characteristic of FT–NIR for the measurement of water content of wheat straw[J]. IOP Conference Series: Earth and Environmental Science, 2019,301(1):1–6.
- [11] 王金武,汤天勇,唐汉,等. 联合收获机电容式稻谷含水率在线检测装置设计与试验[J]. 农业机械学报, 2021,52(3):143–152. WANG Jinwu, TANG Tianyong, TANG Han, et al. Design and experiment of on-line detection device for capacitive paddy rice moisture content of combine harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(3):143–152. (in Chinese)
- [12] 农业部规划设计研究院. 农作物秸秆资源调查与评价技术规范: NY/T 1701—2009[S]. 北京:中国标准出版社, 2009.
- [13] NAZARLOO A S, SHARABIANI V R, GILANDEH Y A, et al. Evaluation of different models for non-destructive detection of tomato pesticide residues based on near-infrared spectroscopy[J]. Sensors, 2021,21(9):3032.
- [14] GUO W, SHANG L, ZHU X, et al. Nondestructive detection of soluble solids content of apples from dielectric spectra with ANN and chemometric methods[J]. Food and Bioprocess Technology, 2015,8(5):1126–1138.
- [15] LI C, LI L, WU Y, et al. Apple variety identification using near-infrared spectroscopy[J]. Journal of Spectroscopy, 2018(8):1–7.
- [16] 樊阳阳,裴正军,陈俭,等. 基于近红外高光谱成像技术的干制红枣品种鉴别[J]. 光谱学与光谱分析, 2017,37(3):836–840. FAN Yangyang, QIU Zhengjun, CHEN Jian, et al. Identification of varieties of dried red jujubes with near-infrared hyperspectral imaging[J]. Spectroscopy and Spectral Analysis, 2017,37(3):836–840. (in Chinese)
- [17] ONIFAFE O F W, AKINYEMI D J. A review on the suitability of machine learning approaches to facial age estimation[J]. Modern Education and Computer Science, 2015,12:17–28.
- [18] 王小川,史峰,郝磊,等. MATLAB 神经网络 43 个案例分析[M]. 北京:北京航空航天大学出版社, 2013.
- [19] 刘雪,李亚妹,刘娇,等. 基于 BP 神经网络的鲜鸡蛋货架期预测模型[J]. 农业机械学报, 2015,46(10):328–334. LIU Xue, LI Yamei, LIU Jiao, et al. BP neural network based prediction model for fresh egg's shelf life[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015,46(10):328–334. (in Chinese)
- [20] 刘秀英,余俊茹,王世华. 光谱特征变量和 BP 神经网络构建油用牡丹种子含水率估算模型[J]. 农业工程学报, 2020,36(22):308–315. LIU Xiuying, YU Junru, WANG Shihua. Estimation of moisture content in peony seed oil using spectral characteristic variables and BP neural network[J]. Transactions of the CSAE, 2020,36(22):308–315. (in Chinese)
- [21] XUE J, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020,8(1):22–34.
- [22] 吕鑫,慕晓冬,张钧. 基于改进麻雀搜索算法的多阈值图像分割[J]. 系统工程与电子技术, 2021,43(2):318–327. LÜ Xin, MU Xiaodong, ZHANG Jun. Multi-threshold image segmentation based on improved sparrow search algorithm[J]. System Engineering and Electronics, 2021,43(2):318–327. (in Chinese)