

基于 SSA – RFR 算法的采棉机测产传感器研究

伟利国^{1,2} 马若飞^{1,2} 周利明^{1,2} 隗立昂^{1,2} 刘阳春^{1,2} 赵 博^{1,2}

(1. 中国农业机械化科学研究院集团有限公司, 北京 100083; 2. 农业装备技术全国重点实验室, 北京 100083)

摘要: 随着棉花种植和收获的机械化程度提高,获取准确的产量图,分析田间产量数据,变得尤为重要,而采棉机作业时 在输棉管道处监测产量是一种有效、可行的方法。现有光电对射式棉花测产传感器在作业中会有粘液遮挡检测通道、环境光影响等问题,面对复杂的田间作业环境,传感器标定普遍采用线性或多项式模型,精度和抗干扰性表现不够理想。针对上述现状,本文首先在传感器的结构和电路设计上做了抗干扰改进。然后在传感器标定过程中,尝试使用随机森林回归模型(Random forest regression, RFR),对实验样本进行训练、测试。在分析模型的表现后,提出了麻雀算法(Sparrow search algorithm, SSA)改进的随机森林回归模型,以均方误差作为适应度,对模型进行优化。经过验证,在相同验证集下,优化后的模型有更好的检测精确度。通过研究寻优上下界范围,平衡运行时间和检测精度,得到最优检测模型。该模型在验证集上表现良好,决定系数 R^2 为 0.99,平均绝对百分比误差(MAPE)为 6.34%。台架实验结果表明,不同风速下最大误差为 9.21%,平均误差为 8.33%,改进后的传感器及检测模型性能良好,能够较准确检测采棉机作业时棉花产量。

关键词: 棉花测产; 质量流量; 麻雀算法; 机器学习

中图分类号: S562 文献标识码: A 文章编号: 1000-1298(2023)09-0154-10 **OSID:** 

Yield Sensor of Cotton Picker Based on SSA – RFR Algorithm

WEI Ligu^{1,2} MA Ruofei^{1,2} ZHOU Liming^{1,2} WEI Li'ang^{1,2} LIU Yangchun^{1,2} ZHAO Bo^{1,2}

(1. Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd., Beijing 100083, China

2. National Key Laboratory of Agricultural Equipment Technology, Beijing 100083, China)

Abstract: With the increase of mechanization of cotton planting and harvesting, it is particularly important to obtain accurate yield map and analyze field yield data, and it is an effective and feasible method to monitor the yield at the cotton conveying pipeline during the operation of cotton picker. The existing photoelectric beam cotton yield measurement sensor has problems such as mucus blocking detection channel and ambient light influence in operation. Facing the complex field working environment, linear or polynomial model is generally used for sensor calibration, and the accuracy and anti-interference performance are not ideal. In view of the above situation, the anti-interference in the structure and circuit design of the sensor was firstly improved. Then, in the process of sensor calibration, random forest regression (RFR) was used to train and test the experimental samples. After analyzing the performance of the model, a stochastic forest regression model based on sparrow search algorithm (SSA) was proposed. The mean square error was used as fitness value to optimize the model. After verification, the optimized model had better detection accuracy under the same verification set. The optimal detection model was obtained by optimizing the range of upper and lower bounds, balancing the running time and detection accuracy. The model performed well on the validation set with a coefficient of determination (R^2) of 0.99 and a mean absolute percentage error (MAPE) of 6.34%. The bench test results showed that the maximum error was 9.21% and the average error was 8.33% at different wind speeds. The improved sensor and detection model had good performance and can accurately detect the cotton quality during the operation of the cotton picker.

Key words: cotton yield measurement; mass flow; sparrow search algorithm; machine learning

收稿日期: 2023-04-26 修回日期: 2023-06-12

基金项目: 中国机械工业集团有限公司重大科技专项(ZDZX2020-2)

作者简介: 伟利国(1978—),男,研究员,博士生导师,主要从事农业智能装备与信息化技术研究,E-mail: weilig78@126.com

0 引言

棉花作为我国主要大田经济作物之一,其产业链长且涉及多个行业,在国民经济发展中占据重要地位。2020 年我国仅新疆维吾尔自治区皮棉总产量 5.161×10^6 t, 占全球总量的 20% 以上。面对日益增长的棉花需求量,实施精准农业,是我国棉花增产增收发展的方向。作物产量监测是精准农业技术的重要组成部分,获取棉花产量分布图,可作为精准农业系统信息反馈,也可检验精准农业技术的实际实施效果,同时为下一季棉花生产精准控制提供原始数据信息。因此,开展棉花产量动态监测系统的研究,具有实际意义^[1-3]。

美国从 20 世纪 90 年代末开始研究棉花的产量监测系统。最先由 WILKERSON 等^[4]提出基于光线密度变化的产量检测方法,用线性关系标定质量流量和传感器输出,决定系数 0.93,在改进后由 Ag Leader 公司投向市场。SUI 等^[5]设计了反射式光电传感器,降低了安装难度,在密西西比州进行了 3 年田间实验,检测误差率 5%。MAILANDER 等^[6]对输棉管道内压差和不同含水率的棉花质量流量进行研究,并建立关系模型,实验显示测量误差不大于 13%。SANH^[7]设计的质量流量计,可以实时检测加工过程中棉花的质量,其由 2 个 S 形应变片组成,经过实验验证,其精度可以达到 $\pm 5\%$ 。国外经过多年发展,现在已有较为成熟的棉花质量在线检测系统,如凯斯 (CASE) 的精准农业系统 (AFS), Ag Leader 的 PF3000 以及 Insight 系统等,都有较好的表现^[8]。

国内学者在棉花测产系统上也有一定的研究成果。何晓峰等^[9]利用近红外方法,提出拟合优度更好的双指数数学拟合,决定系数 R^2 达到 0.99。安光辉^[10]以五通道的光电传感器为基础,通过回归分析,确定每个通道在回归中的权重,台架实验绝对误差平均值 5.67%。周利明等^[11]设计了一种差分结构的电容质量流量传感器,建立含水率、质量流量和电容变化率的二元回归模型,实验验证误差 5.16%。王欢^[12]设计的光电点阵式测产传感器,经过处理后建立流经传感器的棉花团三维模型,换算得到产量,实验结果相对误差 4.31%。目前检测棉花质量流量的方法较多,但大多数方法在田间实验时表现不佳,目前广泛采用的检测方法仍为光电对射法。随着机器学习热潮的到来,其在棉花测产上也有了一定的应用。任玲等^[13]论证了应用神经网络进行棉花质量建模的分析方法的可行性,并分析了神经网络建模的过程。岑红蕾等^[14]在光电测产

传感器上,引入 BP 神经网络标定质量流量和辐射量,实验验证相对误差 7.84%。林昌建等^[15]基于现有的光电测产传感器,引入广义神经元网络进行产量标定,并使用改进遗传算法对网络进行改进,实验验证平均误差最小 7.3%。检测精度较传统方法有所上升,说明该方法具有一定的研究潜力。

在训练机器学习模型时,超参数的选择对模型预测精度影响较大^[16],若在没有先验经验的情况下,能通过寻优算法,高效、准确得到模型超参数,将进一步提高模型检测精确度以及减少模型运算成本。

本文在现有光电对射式测产传感器的基础上,对现有传感器进行优化和改进,并在标定产量关系时,运用将麻雀优化算法和随机森林回归模型融合后的模型来检测棉花产量。最终利用台架实验加以验证,并对比未经优化的随机森林回归模型,分析建立最优模型。

1 传感器优化设计

1.1 传感器检测原理

如图 1 所示,光电对射式测产传感器检测原理是利用棉花通过采棉机输棉管道时,遮挡安装在输棉管道上的传感器发射器发射的红外辐射,接收器通过检测辐射强度的变化,并在完成辐射强度和棉花质量之间的关系标定后,即可在线检测采棉机作业时的棉花产量。

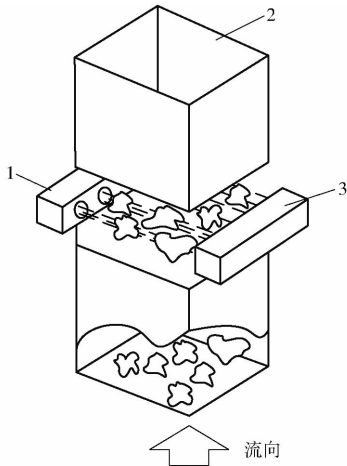


图 1 检测原理示意图

Fig. 1 Schematic of detection principle

1. 发射器 2. 输棉管道 3. 接收器

本传感器通过检测每个设定的单位时间内棉花产量,依次得到作业轨迹上各个小田块内棉花产量,最后整合呈现整块地的产量,检测原理公式为

$$m = g(H_1, H_2, \dots, H_n) \tag{1}$$

式中 m ——单位时间内检测的棉花质量, g

H_i ——单位时间内辐射强度变化向量

1.2 结构优化

本文以中国农业机械化科学研究院集团有限公司机电技术应用研究所团队研发的采棉机测产传感器为基础,针对其在实际作业中存在的问题进行优化改进。

现有的传感器在获取发射器准直光路时,均通过普通凸透镜实现,这种方法需要将光源较为准确地放置在透镜的焦点处,这对安装和加工提出了较高的要求。通过查阅资料,全内反射(TIR)透镜较普通凸透镜有更好的聚光效果,安装时不需人为手动寻找透镜焦点,且形成的光斑也更加自然^[17]。因此,本文用TIR透镜替换凸透镜,并在透镜的安装孔处进行了相应的修改。

周利明^[18]调查发现,因脱叶剂喷洒等原因,造成传感器发射器易被灰尘、叶片以及脱叶剂所形成的黏浆所遮挡,使传感器不能正常工作。因此,本文在传感器上,采用透镜安装孔内嵌的结构,使其不阻碍棉花通过传感器。而特弗伦材料对防止黏浆粘黏有较好的表现^[19]。故在透镜外加装一层特弗伦材质的防护层,保证传感器检测光路不被遮挡。

LED红外发射管会随着工作时间的增加而发热,使发射管发射辐射强度发生变化^[20],导致检测结果出现偏差。因此,本文增加一路校验通道,通过安装在连接板上的隔板使该通道与其他检测通道相隔,只检测空载值。传感器结构示意图如图2所示。

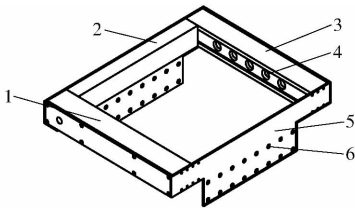


图2 传感器结构示意图

Fig.2 Sensor structure diagram

1. 发射器 2. 校验通道隔板 3. 接收器 4. TIR透镜 5. 连接板 6. 连接孔

因每个发射管均为同一型号且用相同的供电电源,故其因工作时间导致的温度变化基本相同。所以只需将同一时刻的检测通道的数据减去校验通道的数据,即可修正因温度改变辐射源强度而导致的检测偏差,修正原理公式为

$$L^k = W^k - W^0 \quad (k = 1, 2, \cdots, 5) \quad (2)$$

式中 L^k ——某时刻通道 k 补偿后的辐射强度

W^k ——某时刻通道 k 输出的辐射强度

W^0 ——某时刻校验通道输出的辐射强度

1.3 硬件优化

经过验证得到使用 940 nm 红外辐射源时,棉花

品种、含水率对产量检测结果影响较小,而证明年份、色泽对棉花的红外光学特性没有显著影响^[21]。因此,本文的红外发射器选用台湾亿光公司生产的IR333-A型940 nm红外发射管,其结构简单,工作稳定。

环境光传感芯片选用AMS公司的TSL2540环境光芯片,该芯片有可见光和红外光2个感应通道,如图3所示。可单独打开红外感应通道,可在一定程度上减小环境光对检测精度的影响。该芯片通过IIC通信协议和下位机进行通信,向外输出频率为2 Hz,输出的数据结构为高、低8位二进制数。

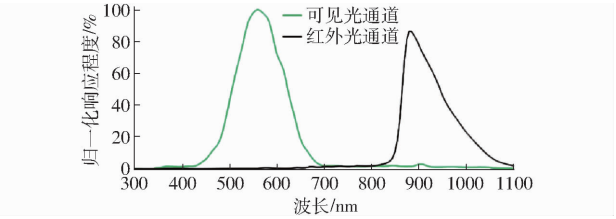


图3 光感芯片特性

Fig.3 Characteristics of light sensing chip

在安装时会因连接板上发射器和接收器之间距离有级调节导致初始辐射值发生变化。因此,本文在发射器中加入电流控制模块,通过接收上位机发送的指令,改变供给给发射管的电流,而调节红外发射器初始辐射强度,调节信号通过CAN通信协议发送至下位机。

使用STM32F103RCT6作为下位机,用于收集各检测通道检测结果和接收上位机指令信号,使用CAN通信协议。在向上位机发送数据时,因CAN数据帧格式要求,需将每个采样周期内的数据分两次向上位机发送,传感器工作原理如图4所示。

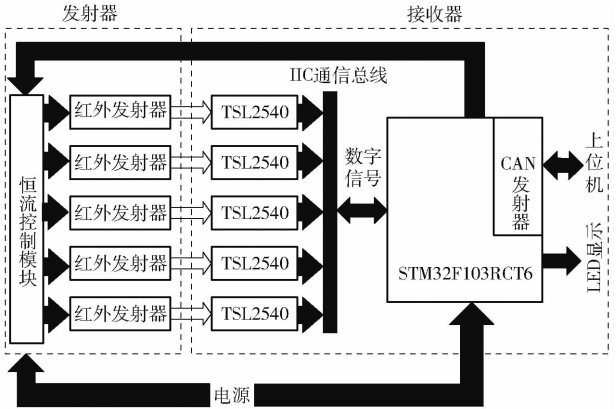


图4 传感器工作原理框图

Fig.4 Sensor working principle block diagram

2 数据集建立

2.1 样本平台

实验棉花样本来源于2021年石河子市采收的新陆早80号,马克隆值4.7,含水率约为7.3%。实

验台架由风机、喂棉传送带、输棉管道和集棉网组成。传送带选用 2 台 DGBL60 型电动滚筒作为动力源,可实现无级调速;风机选用 CF - 2A 型多翼式离心风机,转速 2 800 r/min,风量 1 248 ~ 2 258 m³/h,全压 890 ~ 1 150 Pa。风机可以通过变频器调节风速。棉花经过喂棉传送带进入喂棉口,通过风机的作用在输棉管道内飘动,经过测产传感器,最终被集棉网收集。实验台架示意图如图 5 所示。

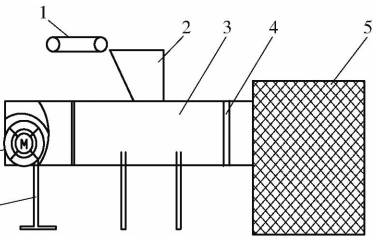


图 5 实验台架示意图

Fig. 5 Schematic of experimental bench

1. 喂棉传送带 2. 喂棉口 3. 输棉管道 4. 测产传感器 5. 集棉网 6. 支架 7. 风机

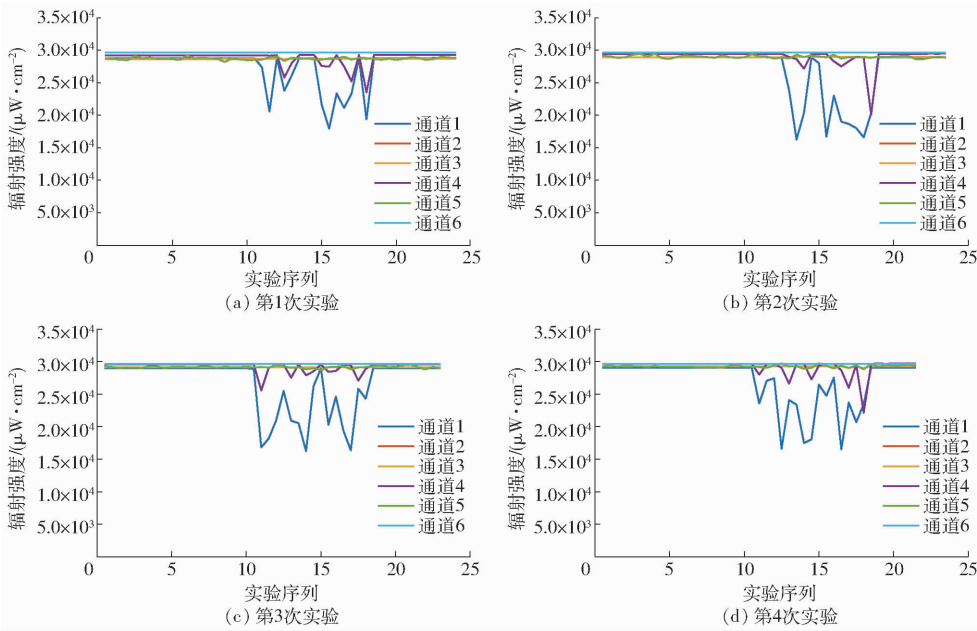


图 6 质量 125 g 时的 4 次实验各通道的辐射强度变化曲线

Fig. 6 Radiation intensity changes of each channel in four tests at mass of 125 g

由图 6、7 可以明显看出,随着喂棉质量的增加,每次实验中各通道辐射强度的最大、最小值之差有明显增加,且辐射强度变化曲线和空载值所围成的面积也呈增加趋势,因此,可以将实验中每个通道辐射变化有效值的极差、方差、积分等参数作为代表实验棉花质量的特征。

2.3 数据预处理

首先对所采集到的数据进行整合、进制转换、空载值修正以及有效值提取等操作,得到有效数据。然后对有效数据提取特征,根据测产传感器检测原理和数据特征,综合上文的分析,最终选择每个通道

2.2 数据获取

每次实验前,用电子秤称量一定量棉花,将其均匀平铺在喂棉传送带上,平铺宽度为 250 mm,长度为 1 000 mm。设置喂棉传送带传送速度为 0.2 m/s,设置风机工作频率为 30 Hz。启动风机,等待上位机稳定显示检测数据后,启动喂棉传送带,将其上的棉花经模拟输棉管道上的喂棉口喂入,在风机作用下,流过光电对射式棉花产量检测传感器,最后被铁丝阻拦网收集。进行多次实验,每次实验的风机变频器频率、喂棉传送带传送速度、喂棉传送带棉花平铺长度均不变,只定量增加输棉传送带上的棉花质量。

总共进行 200 次实验,实验的棉花质量从 125 g 开始,每次实验增加 125 g 棉花,每个质量进行 5 次实验。图 6、7 分别为质量 125、1 000 g 时的 4 次实验各通道的辐射强度变化图,其中通道 6 为校验通道。

的极差(R_1, R_2, R_3, R_4, R_5)、方差(V_1, V_2, V_3, V_4, V_5)和 5 个通道辐射强度的积分(C)作为数据集特征,以每次实验棉花的喂入质量作为该条数据的标签。

为增加数据集的抗干扰性,对现有数据集增加伪数据。将得到的 200 条数据,按照正态分布的方式,以每个数据本身为均值,对每条数据中的每个特征使用相同的种子生成 5 个随机数据。扩展后,数据集总共有 1 200 条数据,并按照 5:1 的比例划分训练集和验证集,其中训练集 1 000 条,验证集 200 条。

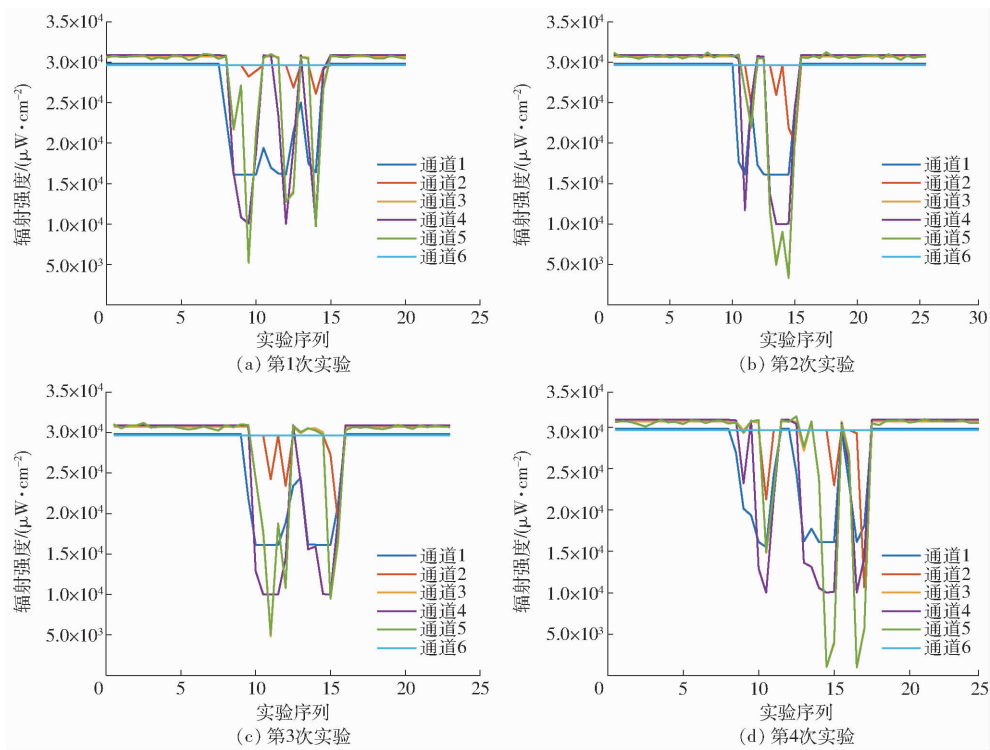


图 7 质量 1 000 g 时的 4 次实验各通道的辐射强度变化曲线

Fig. 7 Radiation intensity changes of each channel in four tests at mass of 1 000 g

3 模型构建和优化

3.1 模型选择和训练

前文所述原始传感器和目前国内的采棉机测产系统普遍采用线性模型进行产量标定,但其抗干扰性和鲁棒性较差,在面对田间复杂的工况有时会出现较大的误差。

而随机森林回归模型有着较好降低泛化误差和抗过拟合的性能^[22],且在其它农作物测产中有实际应用^[23]。因此本文尝试使用该种模型进行产量标定,将前文所建立的训练集数据分别进行线性拟合和随机森林回归(RFR)模型训练,并用验证集检测模型效果。表 1 为 2 种模型在验证集上前 25 个样本的检测结果和残差。图 8 为两种模型的对角误差图。

通过图 8 和表 1 可以明显看出,在验证集上,RFR 模型检测的残差值普遍小于线性模型的检测结果,且在线性模型检测结果在 3 500 g 附近时有较大误差;RFR 模型的预测值比线性模型得到的预测值更加靠近 0 误差对角线,且随着棉花质量的增大,线性模型的检测值相距 0 误差线的距离更远,离散程度更高。综合分析可能是棉花质量的增大,实验中的干扰信息增加,线性模型不能很好地区分干扰和有效信息,而使检测值误差增大。两种模型综合评价指标如表 2 所示。

综上,本文选择随机森林回归模型作为本传感

表 1 模型检测值和残差

Tab. 1 Model prediction result						g
序号	真值	线性拟合 检测值	随机森林 回归检测值	线性模型 残差	随机森林 回归残差	
1	3 407	4 140. 21	3 487. 54	732. 80	80. 14	
2	1 540	1 848. 38	1 482. 32	308. 22	- 57. 84	
3	2 444	2 457. 94	2 584. 53	13. 48	140. 08	
4	3 400	4 271. 40	3 478. 48	871. 40	78. 48	
5	3 626	4 335. 14	3 429. 79	708. 71	- 196. 63	
6	429	868. 90	472. 46	439. 70	43. 26	
7	3 456	4 680. 95	3 473. 37	1 224. 86	17. 28	
8	3 514	4 488. 37	3 466. 00	974. 64	- 47. 73	
9	905	1 294. 49	931. 14	389. 92	26. 57	
10	4 060	3 729. 92	3 958. 28	- 330. 30	- 101. 95	
11	2 020	1 963. 03	2 098. 18	- 56. 83	78. 32	
12	3 235	4 360. 08	3 455. 31	1 125. 02	220. 25	
13	1 625	1 793. 69	1 541. 05	168. 69	- 83. 95	
14	3 631	4 366. 66	3 516. 13	736. 14	- 114. 38	
15	3 205	3 042. 73	2 969. 03	- 162. 03	- 235. 73	
16	2 996	3 077. 42	3 022. 86	80. 99	26. 43	
17	3 570	4 178. 74	3 515. 71	609. 09	- 53. 93	
18	3 550	4 709. 06	3 465. 74	1 159. 54	- 83. 77	
19	4 816	5 160. 85	4 991. 33	344. 56	175. 04	
20	1 125	1 493. 75	978. 19	368. 75	- 146. 81	
21	976	1 296. 09	1 021. 25	320. 39	45. 55	
22	2 926	2 712. 03	3 069. 40	- 213. 48	143. 89	
23	3 493	4 157. 18	3 459. 69	663. 86	- 33. 62	
24	2 506	2 922. 77	2 515. 56	416. 44	9. 23	
25	1 554	1 769. 94	1 562. 78	215. 79	8. 62	

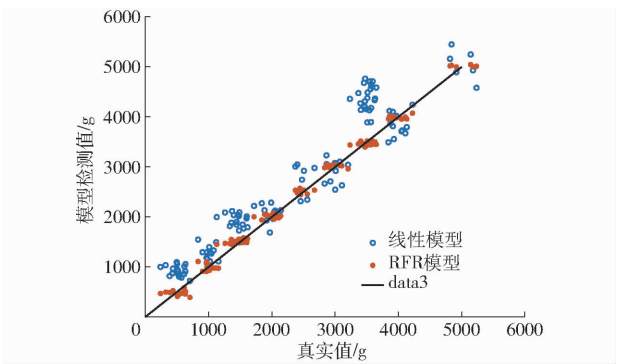


图 8 两种模型对角误差

Fig. 8 Diagonal error diagram of two models

表 2 模型检测结果

Tab.2 Model prediction result

模型	MAE/g	RMSE/g	MAPE/%
线性拟合	47.46	80.38	15.15
RFR	27.98	33.48	7.84

器的标定模型,并进行后续的研究。

3.2 麻雀优化算法

文献[24]指出随机森林回归模型中,决策树数量和树木深度对最终的预测精度有重要影响。但决策树的个数和深度一般为默认参数设置或人为指定,有很强的主观性,不能在应对不同的应用场景时合理改变。

麻雀优化算法(SSA)^[25]是一种基于麻雀觅食和反觅食行为理论启发式算法,使用该算法优化时,具有调整参数少,从开始就能快速收敛的特性。因此,本文采用 SSA 优化算法对 RFR 模型的树木棵数和深度进行优化。

在一个麻雀种群中,分为发现者和追随者两组,发现者拥有较好的适应度,为整个种群寻找更好的觅食区域和方向,追随者利用发现者寻找食物。且当负责警示的麻雀发现危险,即大于所设定的安全阈值,发现者将整个种群引向安全地带。

在优化中,麻雀位置表示为

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & & \vdots \\ x_{q,1} & x_{q,2} & \cdots & x_{q,d} \end{bmatrix} \tag{3}$$

式中 q ——种群内麻雀个数

d ——需要优化的维数

发现者位置更新式为

$$\mathbf{X}_{i,j}^{t+1} = \begin{cases} \mathbf{X}_{i,j}^t e^{\frac{i}{at}} & (G < V) \\ \mathbf{X}_{i,j}^t + Q\mathbf{B} & (G \geq V) \end{cases} \tag{4}$$

式中 $\mathbf{X}_{i,j}^t$ ——第 t 次迭代时,第 i 只麻雀在第 j 维的位置信息

$\mathbf{X}_{i,j}^{t+1}$ ——第 $t+1$ 次迭代时,第 i 只麻雀在第 j 维的位置信息

t ——当前迭代次数

l ——最大迭代次数

a ——在 $(0,1)$ 区间的随机数

Q ——正态分布随机数

$\mathbf{B}$ —— $1 \times d$ 矩阵,其中每个元素都是 1

G ——预警值 V ——安全阈值

当 $G < V$ 时,表示附近没有捕食者,此时发现者继续寻找食物;当 $G \geq V$ 时,表示有部分麻雀发现了捕食者,此时所有麻雀立即飞到其他安全区域寻找食物。

追随者位置更新式为

$$\mathbf{X}_{i,j}^{t+1} = \begin{cases} Q \exp\left(\frac{\mathbf{X}_{\text{worst}}^t - \mathbf{X}_{i,j}^t}{i^2}\right) & (i > q/2) \\ \mathbf{X}_p^{t+1} + \frac{1}{d} \sum_{j=1}^d (\text{rand}(-1,1) \mid \mathbf{X}_{i,j}^t - \mathbf{X}_p^{t+1} \mid) & (i \leq q/2) \end{cases} \tag{5}$$

式中 $\mathbf{X}_p^{t+1}$ ——第 $t+1$ 次迭代时的全局最佳位置

$\mathbf{X}_{\text{worst}}^t$ ——第 t 次迭代时的全局最差位置

当 $i > q/2$ 时,表示第 i 个进入者无法获得食物,需要改变位置来寻找食物。当 $i \leq q/2$ 时,表示第 i 个元素在 $\mathbf{X}_p^t$ 附近,并且在随机觅食。

警示者一般占整个种群的 10% ~ 20%,其位置更新式为

$$\mathbf{X}_{i+j}^{t+1} = \begin{cases} \mathbf{X}_{\text{best}}^t + \beta \mid \mathbf{X}_{i,j}^t - \mathbf{X}_p^{t+1} \mid & (f_i > f_b) \\ \mathbf{X}_{i,j}^t + K_1 \left(\frac{\mid \mathbf{X}_{i,j}^t - \mathbf{X}_{\text{worst}}^{t+1} \mid}{f_i - f_w + \varepsilon} \right) & (f_i = f_b) \end{cases} \tag{6}$$

式中 $\mathbf{X}_{\text{best}}^t$ ——第 t 次迭代时的全局最佳位置

β ——服从 $N(0,1)$ 分布的随机值

K_1 ——麻雀种群的移动方向,在 $[-1,1]$ 内

ε ——小值常数,防止分母为 0

f_i ——第 i 只麻雀的适应度

f_b ——当前麻雀种群的全局最佳适应度

f_w ——当前麻雀种群的全局最差适应度

通过每一次迭代所计算得到的适应度函数,不断改变麻雀种群中发现者和追随者的位置,最终寻得最优参数。

本文通过应用麻雀优化算法对随机森林回归模型的树木棵数和树木深度进行优化,并选择均方误差为适应度函数

$$MSE = \frac{1}{s} \sum_{i=1}^s (\hat{m}_s - m_s)^2 \tag{7}$$

式中 s ——验证集内数据数量
 $\hat{m}_s$ ——模型输出的棉花质量,g
 m_s ——实验喂入的棉花质量,g

根据前文对麻雀优化算法的叙述,可以设计得到 SSA - RFR 工作的流程图,如图 9 所示。

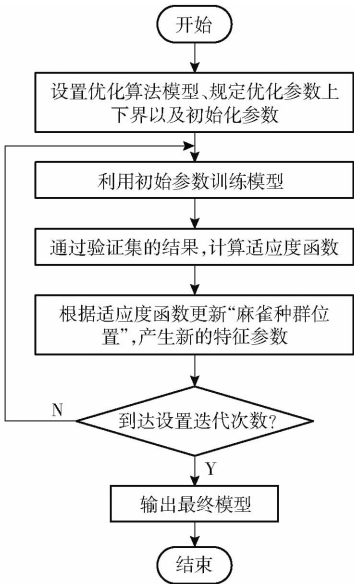


图 9 SSA - RFR 算法流程图

Fig.9 Flowchart of SSA - RFR algorithm

3.3 优化模型训练

在设置优化算法参数时,因麻雀优化算法有收敛速度快的特性,迭代次数取 10;优化参数分别为随机森林回归的树木棵数和树木深度,其范围(5, 50)棵和(2,10)层;预警者比例按照常规取 0.2,安全值取为 0.6。

SSA 进化收敛曲线如图 10 所示。可以看到,通过 10 次寻优迭代,麻雀种群适应度有所下降,且在前 5 次已经达到最优。经过优化的参数选择为 50 棵决策树,深度为 7 层。

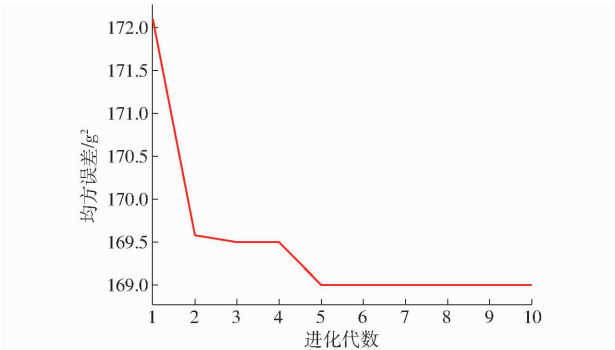


图 10 SSA 进化收敛曲线

Fig.10 Evolutionary convergence curve of SSA

3.4 参数调试

上述实验可以看到,通过麻雀算法找到的最优超参数是在提前规定好的上下界范围内的,有一定局限性。在保持数据集不变的情况下,对 2 个寻优

参数控制变量,不断改变其上界,使其在保证准确性的前提下,限制运算复杂程度,节省运算成本,选择最优的参数。

在树木深度不变的情况下,树木棵数上限每次增加 50 棵,进行 6 次实验,每次实验的搜索时间成本和验证集的检测准确率如图 11 所示。

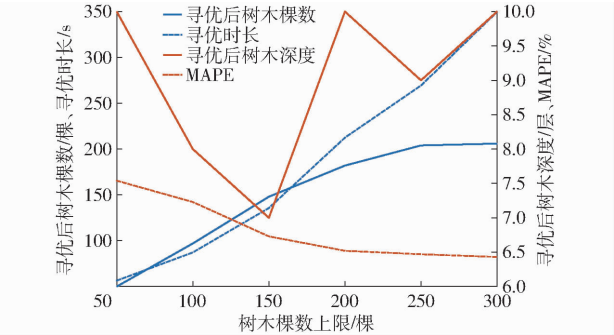


图 11 树木棵数范围对寻优影响

Fig.11 Effect of number of trees on optimization

在树木棵数不变的情况下,树木深度上限每次增加 5,进行 6 次实验,每次实验的搜索时间成本和验证集的检测准确率如图 12 所示。

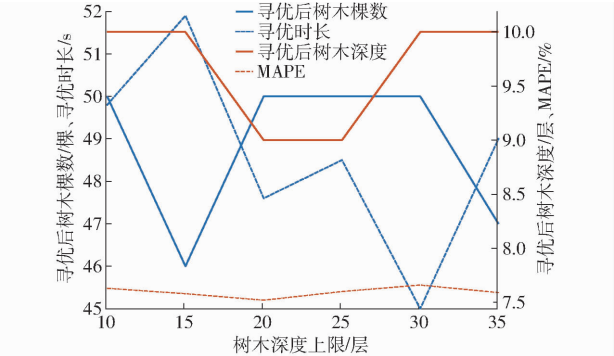


图 12 树木深度对寻优影响

Fig.12 Effect of tree depth on optimization

在保持树木建模深度不变时,随着树木棵数上限的增加,最优树木棵数和寻优时长都在增加,但真实值和检测值的平均绝对百分比误差在减小。在上限到达 200 棵以后,明显看出运算成本继续增加,但平均绝对百分比误差几乎不再减小,此时再增大搜索上限收益较低。当保持树木棵数不变时,增加树木深度上限,各评价参数只在一定范围内波动,并无明显变化趋势,说明树木深度的上限选为 10 层时即满足最优。

经过上述分析得到,最终选择树棵数上限为 250 棵,深度上限为 10 层。最终优化参数为 206 棵决策树,深度为 10 层,以此参数训练后的随机森林回归模型,作为最终传感器的标定模型。优化后对模型验证,结果的决定系数 R^2 为 0.99,平均绝对误差(MAE)为 24.69 g,平均绝对百分比误差(MAPE)为 6.34%,准确率较随机森林回归原始模型有提

高,且运算成本相差不大。寻优参数范围优化前后的模型在相同验证集上的表现如图 13 所示。可以看出,参数优化后的模型的检测值更加集中于 $y=x$ 的 0 误差线上,说明参数优化在一定程度上提升检测准确性。

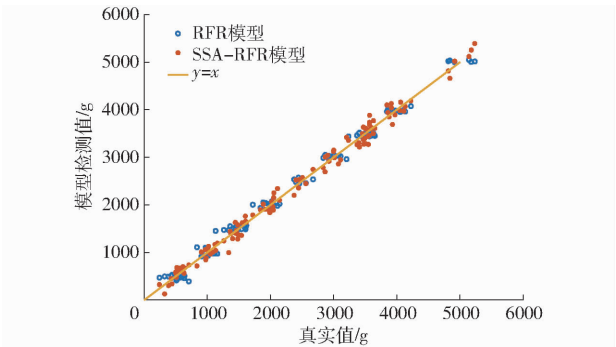


图 13 模型优化前后检测值对比

Fig. 13 Comparison of test values before and after model optimization

4 实验验证

4.1 台架实验

为验证本文所改进的传感器及其模型检测效果,进行采棉机台架实验验证。实验台架仍采用前文所述的台架,如图 14 所示。

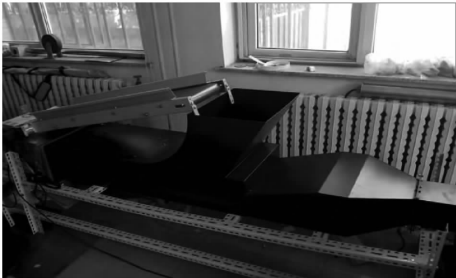


图 14 实验台架

Fig. 14 Experimental bench

台架可以通过变频器输出改变输棉管道内风机风速,通过控制电机改变喂棉速度。为进一步验证模型检测精确度,并模拟田间复杂的作业环境,本文通过重复验证实验和不同风机风速实验来对传感器检测效果进行测试。

每种工况进行一组实验,每组 120 次。各组实验内的每次实验棉花质量对应相同,棉花质量在实验前 0 ~ 5 500 g 的范围内随机生成,在每次实验前使用电子秤称量所需质量的棉花。完成实验和数据采集后,使用和 2.3 节相同的方法对数据进行预处理并提取数据特征,并标注每次实验的真值。

4.2 重复验证实验

为验证改进后的传感器和检测模型的性能是否稳定,先在同训练集相同的实验条件下进行重复验证实验。

共进行 2 组实验,设置喂棉速度为 0.2 m/s,风机输入频率为 30 Hz,按照提前确定的质量进行实验。两组实验数据预处理后,将实验数据输入前文训练好的模型后,输出检测结果。

计算 2 组实验检测值的相对误差,并将结果按照每次实验对应的喂棉质量从低到高进行排列后,得到相对误差直方图和相对误差绝对值趋势曲线,如图 15 所示。在前 25 次实验,喂入棉花质量小于 1 250 g 时,误差大多数为负值,即检测结果偏小,对于此现象,可能是由于传感器的检测通道之间存在间隔,较小的棉花喂入量使棉花可能从间隔处流经,没有使辐射减弱,导致丢失产量信息。还可以看到随着实验棉花质量的增加,检测值的相对误差呈现减小趋势,且最后稳定在 5% 附近,符合实际生产的应用。

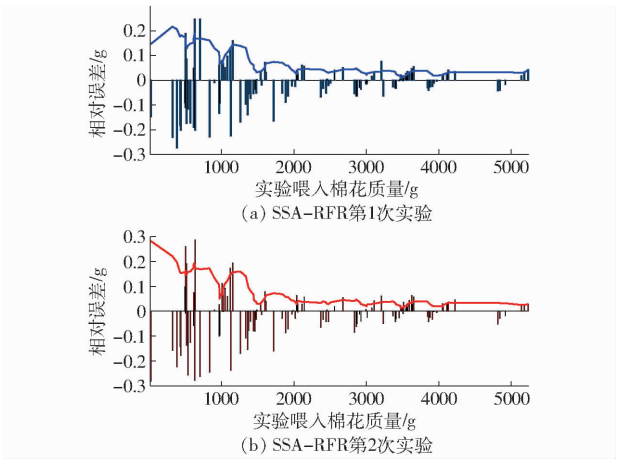


图 15 检测值相对误差

Fig. 15 Relative error of detection value

对 2 组实验检测结果进行评价,2 组实验的平均绝对误差(MAE)分别为 24.08、22.24 g,平均绝对百分比误差(MAPE)为 6.52% 和 6.02%。实验结果与验证集数据结果基本吻合,说明该传感器的性能稳定,可复现检测结果。

4.3 不同风机风速下检测实验

为检验采棉机在田间工作时风机风速改变对检测结果的影响,通过改变变频器的输出频率而调节风机风速,3 组实验分别设置 3 种不同的变频器输出频率为 30、35、40 Hz,每次实验的喂棉时间 5 s 保持不变。

不同风速下的检测值残差箱线图如图 16 所示。前 2 种风速下的检测结果残差中心线基本在 0 附近波动。但可以观察出,随着风速增加,检测结果的残差有向负值变化的趋势,尤其是风机输入频率为 40 Hz 时,该现象更为明显。对于该现象,可能是由于传感器采样频率不高,但随着风速增加,棉花经过测产传感器时间变短,导致一些棉花信息漏采,而出

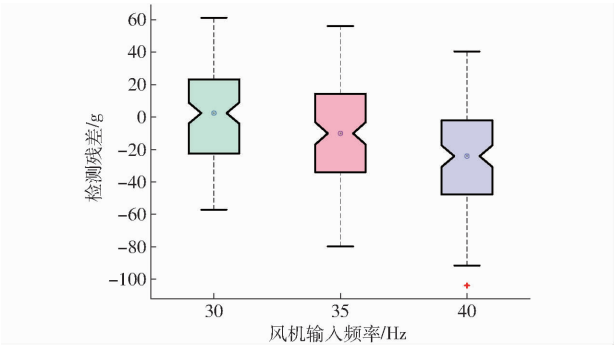


图 16 不同风速下检测值残差箱线图

Fig. 16 Residual boxplot of detected values at different wind speeds

现检测值较真实值偏小的现象。

还可以看出,随着风速增大,残差的分布也越来越分散,且风速越大,残差极大、极小值也越突出,但从误差绝对值看,最大绝对误差为 96.58 g,基本符合检测要求。

对 3 种风速下的检测结果进行指标评价,结果如表 3 所示。在喂棉速度不变的情况下,风机输入频率为 40 Hz 时平均检测误差最大,为 9.21%;3 种风速下平均检测误差 8.33%,检测结果基本符合预期。

表 3 不同风机风速下检测结果

Tab.3 Model prediction result

风机输入 频率/Hz	喂棉速度/ ($\text{m}\cdot\text{s}^{-1}$)	MAE/g	RMSE/g	MAPE/%
30	0.2	26.41	31.22	6.97
35	0.2	27.43	34.48	8.81
40	0.2	30.89	35.74	9.21

5 结论

(1)对现有采棉机测产传感器存在的问题进行优化,分别对其结构和封装材料进行改进,并重新选型了光传感器,增加可变电流控制模块等。

(2)针对在检测模型时不能很好地应对田间复杂环境的问题,本文尝试随机森林回归模型,在分析其检测表现后,提出麻雀算法优化的随机森林回归模型。最后通过验证表明,该模型决定系数 R^2 为 0.99,在测试集的平均绝对百分比误差(MAPE)为 6.34%。

(3)台架实验表明,改进后的传感器和检测算法稳定性较好;且在不同风速下,最大误差 9.21%,平均误差 8.33%,检测效果良好,说明改进后的传感器和检测方法能满足采棉机田间作业时的监测要求。

参 考 文 献

[1] 牛国梁,李斌,刘洋,等.我国采棉机发展历程与研究现状[J].中国农机化学报,2020,41(2):212-218.
NIU Guoliang, LI Bin, LIU Yang, et al. Development and research status of cotton picker in China[J]. Journal of Chinese Agricultural Mechanization, 2020, 41(2): 212-218. (in Chinese)

[2] 吴传云,冯健,陈传强,等.我国棉花产业现状与机械化发展情况分析[J].中国农机化学报,2021,42(5):215-221.
WU Chuanyun, FENG Jian, CHEN Chuanqiang, et al. Analysis of the status quo and mechanization development of cotton producing industry in China[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(5): 215-221. (in Chinese)

[3] 辛明华,王占彪,韩迎春,等.新疆机采棉发展回顾、现状分析及措施建议[J].中国农业科技导报,2021,23(7):11-20.
XIN Minghua, WANG Zhanbiao, HAN Yingchun, et al. Review, status and measures of Xinjiang machine-picked cotton[J]. Journal of Agricultural Science and Technology, 2021, 23(7): 11-20. (in Chinese)

[4] WILKERSON J B, KIRBY J S, HART W E, et al. Real-time cotton flow sensor[C]//American Society of Agricultural Engineers. Meeting (USA), 1994.

[5] SUI R, THOMASSON J A, MEHRLE R, et al. Mississippi cotton yield monitor: beta test for commercialization[J]. Computers and Electronics in Agriculture, 2004, 42(3): 149-160.

[6] MAILANDER M, MORIASI D. Test of pressure transducer for measuring cotton mass flow[J]. The Journal of Cotton Science, 2011, 15: 137-143.

[7] SANH L E. Development of an electronic hopper weight scale[C]//American Society of Agricultural and Biological Engineers Annual International Meeting, 2006:146-157.

[8] American Society of Agriculture and Biological Engineers. Yield monitor performance test standard[R]. ASABE, 2007.

[9] 何晓峰,吴国新,徐守东.棉纤维根数的近红外测量[J].农业工程学报,2009,25(3):119-123.
HE Xiaofeng, WU Guoxin, XU Shoudong. Measurement of cotton fiber numbers by near infrared spectroscopy[J]. Transactions of the CSAE, 2009, 25(3): 119-123. (in Chinese)

[10] 安光辉.基于光电检测原理的棉花质量流量在线测量系统设计与研究[D].石河子:石河子大学,2012.
AN Guanghui. Design and research of online measurement system of cotton mass flow based on optoelectronic detection[D]. Shihezi: Shihezi University, 2012. (in Chinese)

[11] 周利明,李树君,张小超,等.基于电容法的棉管籽棉质量流量检测[J].农业机械学报,2014,45(6):47-52.
ZHOU Liming, LI Shujun, ZHANG Xiaochao, et al. Detection of seed cotton mass flow based on capacitance approach[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(6): 47-52. (in Chinese)

[12] 王欢.基于WSN的采棉机产量在线监测系统设计与研究[D].石河子:石河子大学,2020.

- WANG Huan. Design and research of on-line monitoring system for cotton output based on WSN [D]. Shihezi: Shihezi University, 2020. (in Chinese)
- [13] 任玲, 岑红蕾, 马蓉. 基于神经网络的棉花质量检测系统的可行性分析[J]. 安徽农学通报, 2013(19): 134-135.
REN Ling, CEN Honglei, MA Rong. The feasibility analysis of the cotton quality inspection system based on neural network [J]. Anhui Agri. Sci. Bull., 2013(19): 134-135. (in Chinese)
- [14] 岑红蕾, 任玲, 马蓉. 神经网络在棉花质量流量预测中的建模研究[J]. 农业网络信息, 2013(12): 16-19.
CEN Honglei, REN Ling, MA Rong. Research on modeling of neural networks in the prediction of cotton quality and flow rate [J]. Agriculture Network Information, 2013(12): 16-19. (in Chinese)
- [15] 林昌建, 李彦明, 苗中华, 等. 基于遗传算法和 GRNN 的棉花流量检测模型研究[J]. 机械与电子, 2013(4): 11-14, 18.
LIN Changjian, LI Yanming, MIAO Zhonghua, et al. Research on model of cotton flow measuring based on genetic algorithm and GRNN [J]. Machinery & Electronics, 2013(4): 11-14, 18. (in Chinese)
- [16] 曹桃云. 基于随机森林的变量重要性研究[J]. 统计与决策, 2022, 38(4): 60-63.
CAO Taoyun. Study on the importance of variables based on random forest [J]. Statistics and Decision, 2022, 38(4): 60-63. (in Chinese)
- [17] 王海洋, 金志樑, 王沛沛, 等. 用于 LED 准直的 TIR 透镜与反光杯效果研究[J]. 激光与光电子学进展, 2021, 58(19): 328-333.
WANG Haiyang, JIN Zhiliang, WANG Peipei, et al. Study on effect of TIR lens and reflector for LED collimation [J]. Laser & Optoelectronics Progress, 2021, 58(19): 328-333. (in Chinese)
- [18] 周利明. 基于电容法的棉花产量和播种量检测技术研究[D]. 北京: 中国农业大学, 2014.
ZHOU Liming. Research on detection of yield and seeding rate of cotton based on capacitive method [D]. Beijing: China Agricultural University, 2014. (in Chinese)
- [19] 王运明, 钟耀. 汽车整车厂涂装车间格栅板防粘涂层的应用[J]. 中国设备工程, 2016(7): 40-41.
- [20] XIAO Yuanbin, ZHAO Yanfang, YANG Yanqi, et al. Effects of multi-factors on the junction temperature of LED automotive lamp chips [J]. Microelectronics and Reliability, 2022, 136: 114658.
- [21] 沈雷. 输棉管籽棉流量光电检测模型建立与优化[D]. 石河子: 石河子大学, 2014.
SHEN Lei. The establishment and optimization of photoelectric detection model for seed cotton mass flow in pipe [D]. Shihezi: Shihezi University, 2014. (in Chinese)
- [22] SHAH K, PATEL H, SANGHVI D, et al. A comparative analysis of logistic regression, random forest and KNN models for the text classification [J]. Augmented Human Research, 2020, 5: 1-16.
- [23] DANG C, LIU Y, YUE H, et al. Autumn crop yield prediction using data-driven approaches: support vector machines, random forest, and deep neural network methods [J]. Canadian Journal of Remote Sensing, 2021, 47(2): 162-181.
- [24] PROBST P, WRIGHT M N, BOULESTEIX A L. Hyperparameters and tuning strategies for random forest [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9(3): e1301.
- [25] 薛建凯. 一种新型的群智能优化技术的研究与应用[D]. 上海: 东华大学, 2020.
XUE Jiankai. Research and application of a novel swarm intelligence optimization technique: sparrow search algorithm [D]. Shanghai: Donghua University, 2020. (in Chinese)