

基于 XGBoost 算法对茶叶揉捻工艺参数的研究

陈侠¹, 柳军^{1,2*}, 杨亦扬³, 孔杰², 皮杰²

(1.南京农业大学人工智能学院, 江苏南京 210031; 2.江苏省农业科学院农业设施与装备研究所, 江苏南京 210014;
3.江苏省农业科学院休闲农业研究所, 江苏南京 210014)

摘要: 茶叶揉捻后的碎茶率、成条率和细胞破坏率是衡量揉捻品质的重要指标。我国茶叶种类繁多,不同种类的茶叶揉捻没有统一标准,受主观能动性影响较大,易造成揉捻品质不一。针对这种情况,文章建立随机森林、XGBoost、支持向量机的茶叶揉捻工艺参数推荐模型,对比三种模型预测结果的评价指标,XGBoost 的 MAE、MAPE、RMSE 指标值更小,其预测效果最好。采用 XGBoost 算法作为推荐系统核心,设计出茶叶揉捻参数推荐系统,并对推荐系统推荐参数进行试验。试验结果表明,按照推荐系统所推荐的揉捻参数进行茶叶揉捻,揉捻后的碎茶率、成条率和细胞破坏率已达到揉捻机作业性能指标。研究为智能揉捻提供一种新方法,对当前茶叶揉捻设备发展具有一定现实意义。

关键词: 机器学习; XGBoost 模型; 揉捻工艺; GridSearch 参数寻优

中图分类号: TS272.3

文献标识码: A

文章编号: 2095-0306(2023)02-0032-07

Research on Tea Rolling Process Parameters Based on XGBoost Algorithm

CHEN Xia¹, LIU Jun^{1,2*}, YANG Yi-yang³, KONG Jie², PI Jie²

(1. College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China;
2. Institute of Facilities and Equipment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China;
3. Institute of Leisure Agriculture, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China)

Abstract: The breaking rate, strips formation rate, and cell damage rate of tea after rolling are important indicators to measure the rolling quality. With various types of tea in this country, there is no unified standard for rolling different types of tea. Human subjective initiative has significant influence on rolling quality. In view of the above shortcomings, this study first establishes the tea rolling process parameter recommendation models of random forest, XGBoost, and support vector machine, and compares the evaluation indicators of the prediction results of the three models. XGBoost has smaller MAE, MAPE, and RMSE indicators, and its prediction effect is the best. Using the XGBoost algorithm as the core of the recommendation system, a tea rolling parameter recommendation system was designed and tested. The experimental results show that the tea breaking rate, sliver formation rate, and cell damage rate after rolling according to the recommended rolling parameters in the recommendation system have reached the performance indicators of the rolling machine. This study provides a new method for intelligent rolling and has practical significance for the current development of tea rolling equipment.

Keywords: Machine learning; XGBoost model; Rolling process; GridSearch parameter optimization

收稿日期: 2023-05-16, 修改日期: 2023-06-16

基金项目: 江苏省重点研发计划项目(BE2019379); 江苏省自主创新资金项目[CX(22)2008]

作者简介: 陈侠(1997-), 女, 安徽宿州人, 硕士研究生, 研究方向为智能农业装备。

* 通讯作者: nkyliu@163.com

茶产业作为我国特色优势产业,承担着支撑茶区经济、满足健康消费、稳定扩大就业、服务乡村振兴的重要任务^[1]。揉捻是茶叶加工的一道关键工序,主要有三个目的^[2]:一是揉捻茶叶成形,为成品茶外形打基础,二是使茶叶细胞破损、茶汁溢出,三是为后续茶叶加工做准备。揉捻机是目前能够完成茶叶揉捻工艺的主要设备,主要由揉捻盘、揉捻桶与加压装置、传动机构等结构组成,揉捻桶在曲柄等机构的带动和控制下,在揉捻盘上做相对偏揉捻轴中心的横向旋转。茶叶在揉桶中随着揉桶的运动不断往复翻转,并在揉盘上进行来回揉搓,逐步卷缩成条。

目前,我国茶叶花色繁多,揉捻工艺要求各有不同,揉捻速度、揉捻压力、揉捻时间设置不能一概而论,难以给出揉捻参数具体确定值^[3]。揉捻茶叶时,制茶者需要有一定的制茶经验,根据实际的鲜叶状况和环境条件对工艺进行选择^[4],揉捻工艺受人主观能动性影响,易造成揉捻品质不一。张问采等^[5]提出利用大数据技术,对茶叶原产地的温度、气候、海拔等地理信息进行系统收集,建立茶叶加工工艺参数库,以确保茶叶揉捻加工工艺的统一性和标准性。新技术与揉捻工艺的结合能够智能选择揉捻工艺参数,对当前茶叶揉捻设备发展具有一定现实意义。

我国茶叶种类较多,不同茶叶的加工生产方式各有不同,但目前揉捻机智能化水平还不高,无法根据茶叶状态智能地选择揉捻参数。文章基于 XGBoost 算法设计茶叶揉捻推荐系统,通过对鲜叶等级、茶叶种类、茶叶产地以及揉捻机型号作为系统输入,能够对揉捻相关参数进行推荐。

1 方法原理

1.1 随机森林算法

随机森林算法(Random Forest)^[6-7]是一种集成学习方法,通过构建多个决策树来进行分类或回归。随机森林是从原始训练样本集 N 中重复抽取 k 个样本生成新的训练样本集合,然后根据自助样本集生成 k 个分类树组成随机森林,新数据的分类结果按分类树投票多少形成的分数而定^[8]。其实质^[9]就是将多个决策树合并在一起,大大提高决策树的运算效率,每棵树都是从一个独立样本中抽取出来的,并且其分布也是一致的,所以分类误

差取决于每棵树的分类能力以及树之间的相关性。CART 决策树是随机森林算法弱分类器的核心部分,优点在于:当数据集的因变量是离散型数值时,此树就是分类树;当数据集的因变量是连续性数值时,此树就是回归树,预测值可以用叶节点观察的均值来表示^[10]。

随机森林算法容易实现,在训练速度方面和训练高度并行化方面也具有明显优势,还能进行模型融合,提高模型的准确性和稳定性。并且由于采用了随机采样,可以训练出方差小、泛化能力强的模型。

1.2 XGBoost 算法

XGBoost (eXtreme Gradient Boosting) 算法是一种强大集成学习方法,同时支持 CART 树和线性分类器为基分类器,基于前向分布算法实现加法模型的集成学习方法。集成模型的基本理念是通过构建一系列弱基础模型来构建一个强大的模型。XGBoost 算法核心思想^[11-12]是通过持续的增加树,不停地进行特征分裂来生长一棵树、添加一个树的过程,实际上就是学习一个新函数的过程,拟合上次预测的残差。构建出 k 棵树,并且每棵树都能够模型化,从而模型化出每个样本的分值,从而达到对未知值的准确估计。通过观察这个样本的特征,会发现它会落在每棵树的一个对应的叶节点上,每个叶节点对应一个分数。最后,只需要将每棵树的相应分数相加,就可以得到样本的预测值^[13]。

XGBoost 算法具有以下优点:(1)简单易用,提供 API 方便用户使用。(2)灵活性高,可应用于多种类型数据集和任务,包括分类、回归、排名和推荐等。(3)准确率高,在分类和回归问题上可以达到其它算法难以匹敌的准确率。(4)可解释强,提供丰富的特征重要性评估方法,可帮助用户理解模型预测过程。

1.3 支持向量机

支持向量机(Support Vector Machine, SVM)^[14]是一种常用的机器学习算法,主要用于分类和回归问题。它的基本原理^[15]是通过找到一个最优的超平面,将数据分为不同的类别。最优超平面是 $n-1$ 维的线性子空间,其中 n 是数据的特征维数。例如,如果数据有两个特征,那么超平面是一维的线性子空间,也就是一个直线;如果数据有三个特

征,那么超平面是个二维的线性子空间,也就是一个平面。在二维空间中,超平面就是一个直线,它将数据分为两个类别。通过 SVM,可以利用一些数学技术,将复杂的高维数据转换为简单的低维数据,从而有效地解决高维数据分析的问题。

支持向量机算法具有适应性广泛、可解释性强、计算复杂度低的优点。支持向量机算法可用于线性和非线性问题,在分类和回归任务中表现出色,预测精度高,还能够清晰表示分类和回归的决策边界和数据分布情况。

2 模型搭建

2.1 模型搭建及训练测试

根据 XGBoost 的算法原理并使用 Python 语言,构建出 XGBoost 算法测试模型框架,如图 1 所示。一共分为四个部分,一是样本数据预处理,二是参数调优,三是模型训练,四是模型预测。

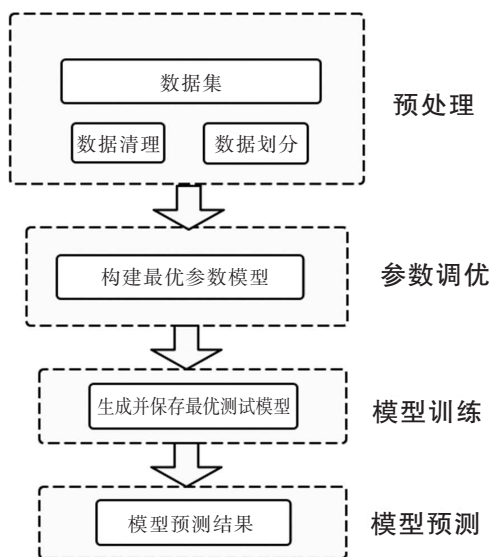


图 1 XGBoost 算法测试模型框架

Fig. 1 XGBoost algorithm testing model framework

2.1.1 数据预处理

文章数据是由广西昭平仙峰茶厂提供,获取时间为 2023 年 3~4 月。在当地 5 位制茶师傅的指导下,采用 6CR-35 型茶叶揉捻机进行数据获取。获取相关数据信息如表 1 所示。试验获取的数据还不能直接使用,因为初始数据中包含揉捻品质指标不符合揉捻机性能指标的揉捻参数数据,会降低模型学习效率和预测精度。在数据使用前需

要对数据进行预处理,此次试验预处理主要为数据清理和数据划分。

表 1 揉捻相关数据信息
Table 1 Rolling related data information

影响揉捻参数相关信息 Information related to the influence of rolling parameters	茶叶揉捻工艺参数 Tea rolling process parameters	揉捻品质指标 Rolling quality indicators
茶叶等级、茶叶种类、揉捻机型号、茶叶产地	揉捻转速、空揉时间、一次轻压大小、一次轻压时间、重压大小、重压时间、二次轻压大小、二次轻压时间	碎茶率、成条率、细胞破坏率

(1) 数据清理。在当地制茶师傅指导下采用 6CR-35 型茶叶揉捻机对三级鲜叶制成红茶春茶来进行揉捻,揉捻过程中受人主观能动性影响。试验获取的 100 组数据中,有 8 组数据的破茶率、成条率、细胞破碎率不符合 35 型茶叶揉捻机标准,这些数据对后续揉捻参数预测没有相关性,模型分析预测数据时需要去除这些数据。

(2) 数据划分。将剩余 92 组数据按照 9:1 的比例分为训练样本和测试样本。训练样本用于算法迭代学习,测试样本用于检测算法预测效果。

2.1.2 参数调优

XGBoost 算法参数^[16]包括通用参数、Booster 参数和学习目标参数。通用参数能够进行宏观函数控制,Booster 参数控制决策树生成和组合过程,学习目标参数能够控制训练目标。

通用参数:(1) booster^[17]:用于选择每次迭代模型的类别。booster 参数有 gbtrees 和 gblinear^[18]两种选择,gbtrees 是基于树结构来构建模型,而 gblinear 是基于线性分类器来构建模型。(2) silent:用于决定运行过程中是否产生输出。

Booster 参数:(1) n_estimators:树的数量,用于控制模型的复杂度和训练时间。(2) learning_rate:学习率,用于控制模型参数的更新速度,并且通过不断降低学习率来提升模型的准确性。(3) gamma:控制节点分裂需要的最小目标函数下降量^[19],用于控制树的生长和防止过拟合。(4) subsample:样本采样比例,用于控制每个树节点上的样本数量^[20]。(5) colsample_bytree:控制每棵树中列的子采样比例。(6) colsample_bylevel:控制每一层中列的子采样比例。(7) max_depth:树的最大深度,用于控制树的复杂度,并避免过拟合。(8) max_delta_step:树

模型权重改变的最大步长,控制每个树模型权重改变的最大步长,防止过拟合。(9)lambda 和 alpha:正则项中 λ 和 γ 的权重,减少模型过拟合。(10)scale_pos_weight:调整正样本的权重,提高模型的分类准确率。

学习目标参数:(1)objective:指训练模型时所采用的损失函数,常见的有 reg:linear、reg:logistic、binary:logistic 等。(2)eval_metric:评估模型性能指标,常见的有 rmse、mae、errors 等。

GridSearch 是一种参数调优方法,通过在指定的参数范围内搜索最优参数组合来优化模型性能。其步骤如下:(1)确定需要调优的参数和参数取值范围。(2)将参数和参数取值范围组合成一个参数网格。(3)通过采用交叉验证的方式,对各种参数组合进行全面的性能评估。(4)选择最优的参数组合,并使用所有训练数据重新训练模型。(5)对数据进行测试,评估模型性能。在实际应用中,由于参数组合数量可能很大,所以使用 GridSearch 时需要注意以下几点:(1)尽量限制参数取值范围,避免搜索空间过大。(2)优先选择影响模型性能最大的参数进行调优。(3)结合模型的特性和实际需求,选择合适的交叉验证方式。(4)可以使用并行计算加速参数搜索。

利用网格寻优调优方法分别对茶叶揉捻工艺中揉捻转速(A)、空揉时间(B)、一次轻压(C)、一次轻压时间(D)、重压(E)、重压时间(F)、二次轻压(G)、二次轻压时间(H)的参数进行网格寻优,其最优参数组合如表 2 所示。

表 2 网格寻优参数组合
Table 2 Combination of grid optimization parameters

参数 Parameter	A	B	C	D	E	F	G	H
colsample_bylevel	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
colsample_bytrees	0.7	0.7	0.5	0.5	0.5	0.7	0.5	0.5
gamma	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
lambda	1	1	1	1	1	1	1	1
learning_rate	0.1	0.05	0.05	0.05	0.1	0.1	0.1	0.1
max_depth	3	3	3	3	3	3	3	3
n_estimators	100	100	200	100	500	100	500	500
subsample	0.5	0.9	0.9	0.5	0.5	0.5	0.5	0.5

2.1.3 模型评价指标

在评价模型的预测能力时,通常采用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Squared Error, RMSE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)作为模型评价指标。平均绝对误差^[21]是预测值与真实值之间误差绝对值的平均值,反映了预测误差的实际情况;均方根误差是预测值与真实值之间误差平方和与样本数量比值的平方根,反映了误差分布的离散程度;平均绝对百分比误差^[22]是比较预测值与真实值之间相对误差绝对值的平均值的大小,反映了预测误差与真实值的相对大小。其表达式如式 1、2、3 所示, RMSE 指标、MAE 指标、MAPE 指标得分越小,预测效果越好。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

式中: n 为样本量; y_i 为真实值; $\hat{y}_i$ 为预测值; $\bar{y}_i$ 为真实值的平均值。

2.2 预测结果分析

三种算法模型对揉桶转速、空揉时间、一次轻压、一次轻压时间、重压、重压时间、二次轻压、二次轻压时间预测评价结果分别如表 3 所示。从 RMSE、MAE、MAPE 三项评价指标结果来看, XGBoost 模型评价指标数值最小、算法优势显著,这表明在最优参数组合的选择下, XGBoost 算法模型的预测能力较好,因此系统将 XGBoost 算法作为茶叶揉捻推荐系统的核心,仅使用 XGBoost 算法模型进行下一步分析预测。

3 推荐系统软件设计

3.1 系统环境

系统硬件环境和软件环境具体介绍如表 4 和 5 所示。

表 3 三种模型对不同揉捻参数预测评价结果
Table 3 Prediction and evaluation results of three models for different rolling parameters

揉捻参数 Rolling parameters	评价指标 Evaluating indicator	模型类别 Model category		
		Random Forest	XGBoost	SVM
揉桶转速	MAE	8.950000	2.749889	5.000000
	RMSE	80.102410	7.561724	80.102450
	MAPE	0.447750	0.109994	0.166667
空揉时间	MAE	0.580000	0.921093	5.000000
	RMSE	0.336400	0.241405	25.000000
	MAPE	1.048333	0.767578	0.500000
一次轻压	MAE	0.052900	0.003461	1.263547
	RMSE	0.230000	0.053473	2.540316
	MAPE	0.025556	0.005940	0.032145
一次轻压时间	MAE	4.320000	0.977710	1.000000
	RMSE	18.662400	0.955918	1.000000
	MAPE	0.288000	0.097771	0.090909
重压	MAE	0.270000	0.242680	0.562301
	RMSE	0.072900	0.058893	0.364120
	MAPE	0.017800	0.016178	0.500000
重压时间	MAE	2.500000	0.553907	4.000000
	RMSE	6.250000	0.306813	16.000000
	MAPE	0.250000	0.039565	0.285714
二次轻压	MAE	0.530000	0.598088	1.346580
	RMSE	0.280860	0.553886	1.105631
	MAPE	0.588888	0.145280	0.198462
二次轻压时间	MAE	0.530000	0.298088	1.726531
	RMSE	0.280860	0.153886	1.026354
	MAPE	0.588888	0.145280	0.172654

表 4 系统硬件环境
Table 4 System hardware environment

硬件 Hardware	型号 Model
中央处理器(CPU)	12th Gen Intel(R) Core(TM) i7-12700H 2.30 GHz
内存(RAM)	16.0 GB
主硬盘	NVMe Micron_3400_MTFDKBA512TFH
显卡	NVIDIA Geforce RTX 3070

表 5 系统软件环境
Table 5 System software environment

软件 Software	版本 Version
操作系统	Windows11(64 位)
Python	3.10.1
PyQt	5
Pycharm	2022.2.1

3.2 系统主要功能实现及展示

3.2.1 信息获取模块

茶叶推荐系统的信息获取界面可读取用户选择或输入的相关参数,完成相关数据输入功能。用户在该界面上选择茶叶种类、茶叶类型等信息,并将上述参数以文本读取方式发送至数据处理层,进行相关揉捻参数预测。

在信息获取界面中需要对茶叶等级、茶叶种类、揉捻机型号、茶叶产地信息进行输入。2016 年,中国工业和信息化部颁布了 JB/T 12835—2016 行业标准确定茶叶等级,并以其特定的长度和比例,对茶叶级别进行了精确的划分。鲜叶长度不大于 30 mm 的占 60%以上的为一级鲜叶、鲜叶长度不大于 40 mm 的占 70%以上的为二级鲜叶、鲜叶长度不大于 50 mm 的占 70%以上的为三级鲜叶、其它鲜叶为四级鲜叶。茶叶种类根据茶叶品种和制茶季节进行划分,主要划分为红茶春茶、红茶秋茶、绿茶春茶、绿茶秋茶四个类别。茶叶揉捻机型号按照揉捻桶直径大小可分为 25 型、35 型、45 型、55 型等。

3.2.2 揉捻参数推荐模块

信息获取层模块将获得的茶叶相关信息按照其所属的茶叶等级、茶叶种类、揉捻机型号以及茶叶产地信息进行分类,找到对应数据库。根据历史制茶经验数据中揉捻品质数据和揉捻工艺数据对 XGBoost 模型训练,此系统内将最优揉捻品质参数作为模型输入来预测茶叶揉捻工艺参数。

此揉捻参数推荐系统还具有数据保存、用户管理、操作日志、帮助文档功能,数据保存指的是可对当前获取信息以及推荐信息按照当前鲜叶等级、茶叶种类、揉捻机型号、茶叶产地的不同进行分类保存。用户管理模块仅涉及个人信息管理,用户可通过设定好的账号密码进行登录。操作日志管理记录系统运行情况,主要是系统运行过程中所产生的信息日志。帮助文档主要阐述了此推荐系统设计目的以及其使用场景。

4 推荐系统测试

4.1 试验目的与试验指标

揉捻推荐系统主要功能是根据不同类型鲜叶状态来推荐其揉捻参数,为了验证推荐揉捻参数的实用性,主要选取了茶叶揉捻后的碎茶率、成条

率、细胞破坏率进行测定。

4.1.1 碎茶率测定

试验碎茶率测定是通过称重完成的。在揉捻试验完后,取出茶叶进行称重,称重后通过网筛进行筛选,再将筛除的碎末碎叶进行称重。碎茶重量再除以茶叶总重得到破碎率,具体公式如式 4 所示。

$$S = \frac{m_s}{m} \quad (4)$$

式中: S 为碎茶率; m_s 为碎茶重量; m 为茶叶总重量。

4.1.2 成条率测定

试验成条率测定也是通过称重完成的。在揉捻试验结束后,取出茶叶进行称重,挑出成条进行称重。成条率可通过成条叶重量除以茶叶总重,具体公式如式 5 所示。

$$C = \frac{m_c}{m} \quad (5)$$

式中: C 为成条率; m_c 为成条茶叶重量; m 为茶叶总重。

4.1.3 细胞破坏率测定

采用计算机图像处理技术来进行茶叶细胞破坏率测定。取揉捻好的茶叶若干,放入 100 ℃水浴处理 15 min;在当前温度和处理时间下,破碎细胞外溢,茶多酚氧化形成茶色素并对受损细胞进行染色。将处理过的茶叶展开在白纸上进行拍照;将图片导入电脑、采用 Grabcut 算法删除背景;根据颜色特征计算染色面积与叶片面积比值,得到细胞破坏率。

4.2 推荐系统参数试验

现对一批鲜叶进行揉捻参数推荐,鲜茶产地来源于广西昭平,鲜叶等级为三级,采用 35 型茶叶揉捻机进行揉捻,所要制成茶叶种类为红茶春茶。揉捻参数推荐系统接收到信息获取层所传送的相关信息,找到鲜叶等级为三级鲜叶、茶叶种类为红茶春茶、揉捻机型号为 35 型、茶叶产地为广西昭平的数据库。由于当前茶叶揉捻机采用的是固定位置揉捻,所以试验中以下降距离为控制参数。对数据库里历史揉捻参数数据进行训练,并推荐出茶叶揉捻工艺参数,揉捻转速 35 r/min、空压揉捻时间为 15 min、一次轻压位置为 8 cm 处、揉捻时间为 11 min、重压压力位置为 13 cm 处、重压

时间 10 min、第二次轻压压力大小为 7 cm 处、揉捻时间为 13 min。将揉捻工艺参数应用到 35 型茶叶揉捻机进行揉捻,揉捻结束后取出部分茶叶,进行筛选、称重,计算破茶率和成条率,并用计算机图像处理技术检测茶叶揉捻后细胞破碎率。

查找茶叶揉捻机作业性能指标表,可以得出 6CR-35 型揉捻机揉捻红茶,揉捻加工后的作业性能指标应满足成条率 $\geq 85\%$ 、碎茶率 $\leq 4.2\%$ 、茶叶细胞破坏率 $\geq 83\%$ 。茶叶揉捻机揉捻结果如表 6 所示,试验 1 是根据 20 次揉捻数据进行推荐,试验 2 是根据 40 次揉捻数据进行推荐,试验 3 是根据 60 次揉捻数据进行推荐。从表中试验 1、试验 2、试验 3 可以看出随着试验次数增加,揉捻叶的成条率逐渐增加、碎茶率逐渐降低、茶叶细胞破碎率逐渐增加。由于现有历史揉捻数据量较少,揉捻参数推荐系统推荐数据还不完善。试验 4 是根据 80 次揉捻数据进行推荐,可以看出揉捻后的碎茶率、成条率以及细胞破坏率已达到揉捻机作业性能指标。因此,随着后续试验数据增加,推荐系统不断进行学习优化,按照此系统推荐参数进行揉捻,茶叶揉捻品质将越来越好。

表 6 实验数据记录表
Table 6 Experimental data record form

试验 Test	取样重 量(g) Sampling weight	碎茶重量 (g) Broken tea weight	成条茶重量 (g) Weight of tea in strips	碎茶率 (%) Broken tea rate	成条率 (%) Strips formation rate	细胞破坏 率(%) Cell damage rate
1	10.02	2.15	7.36			
	10.01	2.27	7.51	21.94	73.88	81.3
	10.10	2.19	7.39			
2	10.09	1.82	7.86			
	10.00	1.75	7.87	17.77	77.80	82.3
	10.01	1.78	7.69			
3	10.00	1.37	7.96			
	10.03	1.29	8.11	13.31	80.30	83.1
	10.02	1.34	8.06			
4	10.00	0.42	8.51			
	10.01	0.39	8.69	4.23	85.81	83.6
	10.01	0.46	8.56			

5 讨论

茶叶揉捻品质易受制茶师傅主观因素影响,通过对鲜叶等级、茶叶种类、茶叶产地以及揉捻机

型号来确定揉捻工艺参数,能够保证茶叶揉捻品质稳定。文章将XGBoost算法用于茶叶揉捻参数推荐,利用真实揉捻工艺相关参数对预测模型进行验证,在参数推荐过程中展现了较强性能。根据此算法设计出茶叶揉捻参数推荐系统,并对推荐系统推荐参数进行试验。从推荐系统参数试验可以看出揉捻后的碎茶率、成条率以及细胞破坏率已达到揉捻机作业性能指标,并且随着试验数据增加,推荐系统不断学习优化,按照推荐参数进行茶叶揉捻,揉捻品质将越来越好。此揉捻推荐系统对当前茶叶揉捻设备发展具有一定现实意义。

参考文献:

- [1] 武雅娇,岳海静,向欢欢.乡村振兴视角下舟山普陀佛茶特色文化产业发展研究[J].对外经贸,2022,332(2):48-51.
- [2] 高帅.北方茶叶生产线揉捻机组的设计与优化[D].泰安:山东农业大学,2019.
- [3] 陈侠,杨亦扬,皮杰,等.茶叶揉捻工艺技术研究现状和发展趋势[J].中国茶叶,2022,44(12):9-15.
- [4] 李荣林,周建涛,彭英.茶叶加工技术的创新发展[J].江西农业学报,2011,23(10):128-130,132.
- [5] 张问采,贾文月,张翔.浅谈茶叶揉捻设备机电一体化的现状与发展[J].南方农机,2017,48(21):25-26,35.
- [6] 张宏利.基于实验的分类挖掘若干问题研究[D].西安:西安理工大学,2020.
- [7] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [8] 曹文哲,应俊,陈广飞,等.基于Logistic回归和随机森林算法的2型糖尿病并发现视网膜病变风险预测及对比研究[J].中

国医疗设备,2016,31(3):33-38,69.

- [9] 仵豪.基于机器学习的糖尿病预测系统研究[D].北京:北京交通大学,2020.
- [10] 邓念东,崔阳阳,郭有金.基于频率比-随机森林模型的滑坡易发性评价[J].科学技术与工程,2020,20(34):13990-13996.
- [11] 李伟娜.基于xgboost算法的学生学习行为分析研究[D].北京:北方工业大学,2022.
- [12] 霍振华.基于XGBoost的电信客户欠费预测模型研究与应用[D].南昌:南昌大学,2022.
- [13] 李翌星.面向急性加重型慢阻肺的计算机辅助诊断模型研究[D].太原:中北大学,2019.
- [14] 伍霜霜.基于过程性可观测信息的信用风险预测[D].成都:西南财经大学,2020.
- [15] 王亚.基于特征提取的多干扰检测与识别方法研究[D].成都:电子科技大学,2022.
- [16] 赵晗,孟晓景,张春勇.基于XGBoost的搜索结果智能排序系统[J].软件导刊,2019,18(12):56-60.
- [17] 赵晨阳.基于集成学习的二手房交易价格预测研究[D].太原:山西大学,2020.
- [18] VINAYAK R K, GILAD-BACHRACH R. DART: Dropouts meet multiple additive regression trees[C]. Artificial Intelligence and Statistics. PMLR, 2015: 489-497.
- [19] 李超.基于空间聚集属性和XGBoost算法的供水管道爆管风险评估研究[D].杭州:浙江大学,2020.
- [20] 许小松.基于集成学习算法的金融机构个人客户违约预测[D].南京:东南大学,2020.
- [21] 闫朋涛.面向盐渍海参品质的图像与力学特征协同检测方法研究[D].大连:大连工业大学,2020.
- [22] 宋先知,姚学喆,李根生,等.基于LSTM-BP神经网络的地层孔隙压力计算方法[J].石油科学通报,2022,7(1):12-23.

(上接第21页)

参考文献:

- [1] 刘阳,李光彩,熊中彩,等.不同茶树品种试制白茶品质初探[J].蚕桑茶叶通讯,2021(3):21-24.
- [2] 李明月.四川白茶加工技术及品质评价研究[D].成都:四川农业大学,2015.
- [3] 王若娴,黄翔翔,李勤,等.不同类型白茶儿茶素、香气成分与感官品质比较[J].食品工业科技,2022,43(5):315-321.
- [4] 倪伯荣.“银猴”新品种简介[J].茶叶,2003(4):224-225.
- [5] 沈华琴,张伟,倪伯荣.茶树新品系“茂绿”选育研究报告[J].中国茶叶,2003(6):26-27.
- [6] 全国茶叶标准化技术委员会.茶叶中茶多酚和儿茶素类含量的检测方法:GB/T 8313—2018[S].北京:中国标准出版社,2018.

- [7] 全国茶叶标准化技术委员会.茶游离氨基酸总量的测定:GB/T 8314—2013[S].北京:中国标准出版社,2013.
- [8] 全国茶叶标准化技术委员会.茶叶感官审评方法:GB/T 23776—2018[S].北京:中国标准出版社,2018.
- [9] 府慧君.茶叶水浸出物含量与品质关系的初步研究[J].茶叶,1957(3):23-25.
- [10] 李凤凤.茶叶的香气化学和加工工艺[D].杭州:浙江大学,2008.
- [11] 赵和涛.红茶绿茶加工中氨基酸变化及与品质关系[J].氨基酸杂志,1990(3):47-49.
- [12] 张泽琴.对茶树早期鉴定品质指标和酚氨比的一点看法[J].茶叶通讯,1991(3):22-25.
- [13] 蔡良绥.白茶审评要点[J].中国茶叶,2005(2):27.
- [14] 杨善岚,吴磊,涂嘉欣,等.自由基致衰老的研究进展[J].中华疾病控制杂志,2022,26(5):589-594.