

doi:10.6041/j.issn.1000-1298.2024.S2.031

基于PSO-Stacking的河蟹投饵量预测模型

李家弟^{1,2} 陈子瑜^{1,2} 高晨^{1,2} 孙龙清^{1,2}

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 农业农村部国家数字渔业创新中心, 北京 100083)

摘要: 河蟹作为我国重要的水产养殖物种之一, 深受消费者喜爱, 在河蟹养殖过程中, 科学的投饵量是保证河蟹健康生长及提高养殖效益的关键因素。本文通过综合分析影响河蟹养殖投饵量的多种因素, 采用集成学习算法建立河蟹养殖投饵量预测模型。搭建数据采集系统, 采集包括河蟹生物量、河蟹数量、性别比例、水体pH值、温度、溶解氧含量以及河蟹摄食量等关键参数数据, 建立投饵量数据集; 运用数据预处理技术对数据集进行平滑处理以及归一化, 减少异常值对预测结果的干扰, 同时消除特征数据不同量纲的影响; 引入粒子群优化算法改进集成学习, 建立了河蟹养殖投饵量预测模型, 实现河蟹养殖投饵量的准确预测。实际应用测试结果表明本文模型平均绝对误差为0.349 71 g, 均方根误差为0.491 14 g, 决定系数达0.903 58。

关键词: 机器学习; 集成学习; 粒子群优化算法; 投饵量预测

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-1298(2024)S2-0303-07

Prediction Model for Feeding Amount of River Crab Based on PSO-Stacking

LI Jiadi^{1,2} CHEN Ziyu^{1,2} GAO Chen^{1,2} SUN Longqing^{1,2}

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

(2. National Digital Fisheries Innovation Center, Ministry of Agriculture and Rural Affairs, Beijing 100083, China)

Abstract: As one of the important aquaculture species in China, river crabs are well-loved by consumers. In the process of river crab aquaculture, scientific baiting is a key factor to ensure the healthy growth of river crabs and improve aquaculture efficiency. By comprehensively analyzing the factors affecting the baiting amount of river crab aquaculture, an ensemble learning algorithm was used to establish a prediction model for the baiting amount of river crab aquaculture. A data collection system was set up to collect key parameters such as river crab biomass, crab population, sex ratio, water pH value, temperature, dissolved oxygen, and crab feeding amounts to establish a baiting data set; data preprocessing techniques were used to smooth and normalize the data set to reduce the interference of outliers on the prediction results, and at the same time to eliminate the influence of different scales of the characteristic data; the particle swarm optimization (PSO) algorithm was introduced to improve the ensemble learning and establish a baiting model for river crab culture. The particle swarm optimization algorithm was introduced to improve the ensemble learning, and the bait quantity prediction model was established to realize the accurate prediction of the bait quantity of river crab aquaculture. The results of practical application tests showed that the average absolute error (MAE) of this model was 0.349 71 g, the root mean square error (RMSE) was 0.491 14 g, and the coefficient of determination (R^2) of key performance reached 0.903 58.

Key words: machine learning; ensemble learning; particle swarm optimization algorithm; prediction of feeding amount

0 引言

在传统河蟹养殖阶段, 养殖人员主要通过经验判断投饵量, 易导致饵料投放量不准确^[1-4]。过量投饵会造

成资源浪费、提高养殖成本, 引发水质污染、增加水体处理成本; 投饵不足则会妨碍河蟹的正常生长发育, 导致

收稿日期: 2024-07-23 修回日期: 2024-09-17

基金项目: 国家重点研发计划项目(2020YFD0900201)

作者简介: 李家弟(1999—), 男, 硕士生, 主要从事水产养殖人工智能技术研究, E-mail: 1903340505@qq.com

通信作者: 孙龙清(1964—), 男, 研究员, 主要从事农业信息化及农业物联网技术研究, E-mail: sunl_q@163.com

营养不良、感染疾病等问题^[5-6]。河蟹摄食量受到多种因素影响,如温度、pH值和溶解氧等,通过经验难以准确判断其摄食量^[7-9]。科学的方法能够提高饵料利用率、节省养殖成本并促进河蟹的健康成长^[10]。不同环境下的养殖需要不同,准确的投饵量不仅能保证河蟹的摄食需求、保护环境,还可以最大程度上提升生产效益,因此,开展河蟹智能投饵量预测对于提升养殖效率和确保河蟹养殖业可持续发展具有重大的实践价值^[11-12]。

机器学习在预测河蟹养殖投饵量方面展现出显著优势,其能够通过分析历史数据,建立影响投饵量多种因素之间的复杂关系,从而提供准确性较高的预测结果。机器学习模型具有强大的适应性和改进潜力,能够持续优化预测算法以适应养殖环境的变化和河蟹成长的需求,算法的抽象能力使得机器学习成为提高河蟹养殖效率、优化投饵策略、减少资源浪费和环境污染的有力工具,为养殖业的可持续发展提供了重要的技术支持^[13-15]。现有研究主要针对河蟹和饵料的识别方法,或通过单一模型对部分特征进行处理预测河蟹摄食量^[16-24]。

本文通过整合影响河蟹投喂量的主要特征参数,开展河蟹投饵量预测模型的研究,提出基于Stacking机器学习框架,引入粒子群优化算法实现集成学习模型超参数寻优,构建基于PSO-Stacking的河蟹投饵量预测模型。

1 材料与方法

1.1 数据获取

1.1.1 数据采集系统搭建

数据采集地点位于江苏省宜兴市河蟹智能立体养殖工厂示范实验基地,研究对象为河蟹。实验过程中,河蟹被安置在专门设计的实验箱中,为了减少河蟹相互攻击的行为,在实验箱中设置蟹舍。实验数据采集系统设计示意图及环境如图1、2所示。

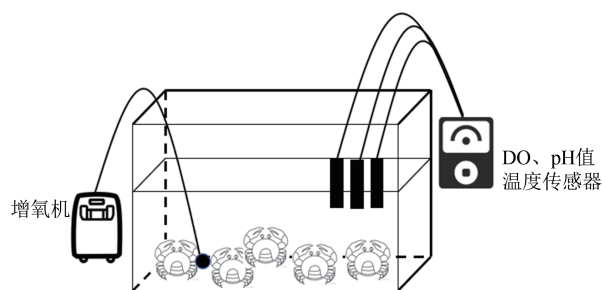


图1 数据采集系统设计示意图

Fig. 1 Design diagram of data acquisition system

为了减少环境突变对河蟹应激的不利影响,干扰实验结果的准确性,采取将河蟹先在实验箱中暂养5 d的措施,使得河蟹能够适应养殖环境,减少由环境改变引起的应激反应。每个实验箱均配备增氧设备,作用是保持水质的新鲜和富氧;并投入自行培养的藻类,模拟了自然环境中的生态系统,确保养殖水中溶解氧含量始终



图2 实验箱环境

Fig. 2 Environment of experiment box

保持充足,为河蟹提供了健康生长的必要条件。此外,实验室配备高级集成水质传感器,用以监测水温、pH值和溶解氧浓度。通过使用加热棒和冷水机可以精确调节养殖水温度。

为保持水质清洁,每日下午进行换水,清除河蟹排泄物和残余饵料,并将实验箱内50%~70%的水抽出,加入新水至原水位,避免亚硝酸盐和氨氮含量超标。在养殖专家的指导下,使用滴管向水中添加硝化细菌净化水质。

使用MR2003型电子比色器(南京特安科贸有限公司)对水中的氨氮、亚硝酸盐、余氯、硫化氢等关键水质指标进行严格检测,确保余氯浓度低于0.2 mg/L,氨氮低于1.2 mg/L(有毒气体 NH_3 浓度低于0.2 mg/L),亚硝酸盐低于0.15 mg/L,硫化氢浓度低于0.05 mg/L,保证河蟹实验环境稳定与健康。

1.1.2 数据集建立

将河蟹饲养在实验箱内,详细记录每个实验箱中的河蟹数量和性别比例。日常饲料投喂安排在每天上午8时和下午18时分别进行。河蟹适应养殖环境后,在每天下午的投喂前,利用传感器对水体pH值、水温以及溶解氧含量这3项关键环境指标进行监测记录,这些因素是影响投饵量的主要参数。鉴于河蟹的生物量不会在短时间内发生显著变化,为了减少人为刺激,采取每5 d对河蟹进行一次称重,记录生物量。实验过程中可能会有河蟹因疾病或其他原因死亡,在每次投喂时记录存活的河蟹数量,有变化及时更新。每次投喂时均记录投喂的饵料量,在清理实验箱中残饵时记录剩余饵料量,投喂饵料量减去投喂后剩余饵料量为河蟹实际摄食量。

通过搭建的数据采集系统成功采集有效的投饵摄食数据1 054条。为了确保模型泛化能力和预测精度,将数据训练集与测试集按比例8:2进行数据划分,即80%的数据用作训练集以构建模型,而剩余的20%则作为测试集以评估模型性能。

1.2 数据预处理

(1) 输入参数选择

在河蟹养殖领域,投饵量影响因素众多。其中,关键的外部因素包括养殖水体温度、溶解氧含量、pH值。在适宜的环境条件下,河蟹对饲料的需求量较大;若水

质参数超出理想范围,河蟹摄食量都会相应减少。河蟹生物量、数量、性别比例也是影响投喂量的主要因素。本文选择河蟹生物量、河蟹数量、性别比例、水体溶解氧含量、温度、pH 值作为模型输入参数,并将投饵量作为模型输出参数。

为了增强模型预测能力,将河蟹质量作为一个特征引入模型,提升模型对数据的拟合效果。

(2) 数据归一化

为提高模型的预测精度,将平滑处理后数据进行归一化处理,确保不同尺度的特征能在统一的标准下进行评估,避免了模型受到特征尺度不一致的负面影响,特别是在使用距离度量或基于梯度的优化算法时尤为重要。归一化后数据同时能够减少数值计算误差,避免模型在某些方向上的过拟合,可以帮助提高模型预测精度。

1.3 预测模型构建

在回归模型问题解决的过程中,单个预测模型往往只能从一个特定的角度实现预测,这会导致对其他因素的忽略,影响预测结果的稳定性和准确性。为了解决这个问题,提出一种基于集成学习的新型河蟹投饵量预测模型^[25]。该模型采用 Stacking 集成策略,将多个不同的预测模型结合起来。本研究构建 3 种不同的基础回归模型,以其预测输出作为输入,传递给上层的元学习器。该元学习器的职责是整合这些初步预测,产出最终结果。

此方法与传统单一预测模型相比较,采用的 Stacking 组合预测策略,通过综合多个模型降低了因模型特定偏见产生的预测偏差,提高预测的准确性。融合不同模型的预测,综合各种因素,增加了预测效果的稳健性。Stacking 多模型叠加预测模型还具有较高的鲁棒性,能够在数据集上取得良好的预测效果。

1.3.1 基学习器

在集成学习领域,基学习器的选择及其组合策略对模型整体性能具有决定性影响。采用极端梯度提升模型 (Extreme gradient boosting, XGBoost)、轻量梯度提升机模型 (Light gradient boosting machine, LightGBM) 和 CatBoost 模型 (Categorical boosting, CatBoost) 作为 Stacking 算法的基学习器。

XGBoost 是一种优化的梯度提升决策树 (Gradient boosting decision tree, GBDT) 算法^[26],通过逐步修正残差的方法,XGBoost 能够逐渐减少模型误差。XGBoost 擅长处理非线性关系,能够捕捉到生物量、溶解氧含量、pH 值、温度等变量与河蟹投饵量之间复杂的相互作用,自动评估各个特征对预测结果的重要性,同时由于内置 L1 和 L2 正则化项,有助于控制模型复杂度,显著提高河蟹摄食量预测可靠性。在 XGBoost 中,模型的目标是最小化目标函数,即

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) + \gamma T \quad (1)$$

$$\text{其中} \quad l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (2)$$

式中 y_i ——实际值 $\hat{y}_i$ ——预测值
 $l(y_i, \hat{y}_i)$ ——损失函数,衡量实际值与预测值之间的差异程度

$\Omega(f_k)$ ——正则化项,用于控制模型的复杂度,防止过拟合

γ ——节点数正则化参数,用于控制树复杂度

T ——叶节点总数

K ——决策树数量

XGBoost 是 Boosting 架构的算法,遵循叠加式训练,其训练公式为

$$\begin{cases} \hat{y}_i^{(0)} = 0 \\ \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(k)} = \hat{y}_i^{(k-1)} + f_k(x_i) \end{cases} \quad (3)$$

式中 $\hat{y}_i^{(k)}$ ——第 k 轮模型预测值

LightGBM 是一种轻量级梯度提升树算法,其主要思想是利用基于直方图的算法来离散化连续的浮点特征,采取一种设定了深度限制的叶级增长 (leaf-wise) 方法,专注于减少当前模型的误差^[27]。算法应用了 EFB (互斥特征捆绑) 算法,它通过捆绑互相排斥的特征来减少总特征数,既简化了模型结构,也加快了训练过程。模型计算首个 CART 回归树的预测误差 (即残差),将其用作构建后续 CART 回归树的训练数据,通过迭代的方法逐步精进模型性能。最终通过对多个 CART 回归树的预测结果进行加权融合,形成综合的预测输出。LightGBM 可以在涉及如生物量、溶解氧等多种不同变量的大规模预测任务中,能够快速处理数据,保持高精度预测,同时可以减少内存占用、提升计算效率,满足河蟹投饵量精准预测的需求。LightGBM 通过确定如何分割特征以最大化模型性能,能够找到最优的分割点,从而有效地构建出决策树。其计算公式为

$$V_{j|O}(d) = \frac{1}{n_0} \left[\frac{\left(\sum_{x_i \in O: x_i \leq d} g_i \right)^2}{n_{lo}^j(d)} + \frac{\left(\sum_{x_i \in O: x_i > d} g_i \right)^2}{n_{ro}^j(d)} \right] \quad (4)$$

其中 $n_0 = \sum I(x_i \in O)$ $n_{lo}^j = \sum I(x_i \in O, x_i \leq d)$

$n_{ro}^j = \sum I(x_i \in O, x_i > d)$

式中 g_i ——梯度

$n_{lo}^j(d)$ ——左子节点数量

$n_{ro}^j(d)$ ——右子节点数量

O ——某固定节点的训练数据集

d ——分割特征 n_0 ——分裂点

CatBoost 是一种基于梯度提升的机器学习算法,同样适用于回归任务^[28]。CatBoost 使用对称决策树作为基学习器,可以处理类别型、数值型特征,采用默认参数就可以获得非常好的结果,可减少在调参上的耗时;在

标准的梯度提升算法中,通常通过建立新树来拟合当前模型的梯度。

CatBoost引入了新颖的过拟合控制方法,采用排序提升技术来对抗训练集中的噪声点,避免梯度估计的偏差,对模型的预测结果进行偏移修正,以提高模型的泛化能力,有效减少过拟合。在对河蟹投饵量预测的实验中,该算法可以充分降低过拟合。

1.3.2 元学习器

元学习器主要处理已由基学习器提取过特征的数据,选择随机森林(Random forest, RF)作为 Stacking 算法的元学习器。随机森林算法是一种基于决策树的集成学习算法,其核心思想是构建多个决策树,通过结合它们的预测结果来提升整体模型性能^[29]。该算法采用 Bagging 集成方法,即通过自助采样方式从原始数据集中抽取多个子集,对每个子集训练一棵决策树。该模型作为经典的机器学习算法,在众多领域中均有所运用^[30-32]。随机森林通过整合多个决策树的预测结果,能显著提高预测的准确性。在 Stacking 框架中,基学习器预测结果被作为特征输入到元学习器中,随机森林能有效地从这些特征中学习,从而提升最终模型性能。在构建每棵决策树时,随机森林会随机选择一部分特征进行分裂,这既增加模型多样性,也提高泛化能力。其算法结构如图3所示。

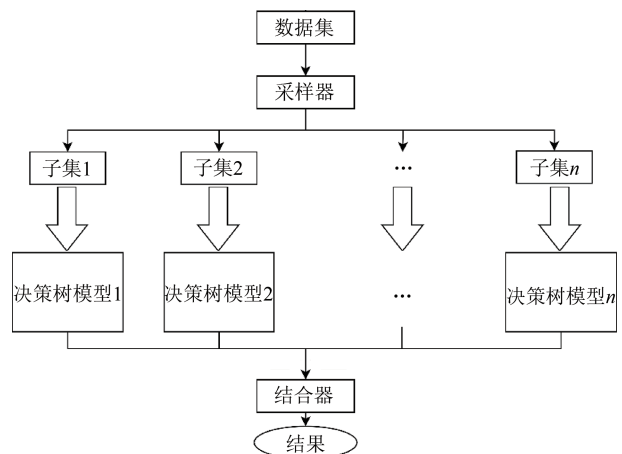


图3 随机森林算法结构图

Fig. 3 Structure diagram of random forest algorithm

在 Stacking 中,多样性是提高集成效果的关键因素之一,随机森林的不同决策树可以提供不同的预测视角,有助于捕捉更全面的数据模式。随机森林还具有内部无偏估计的特性,它可以在没有独立验证集的情况下评估模型的性能。随机森林对于噪声和异常值具有较强的鲁棒性,即使基学习器的输出存在一些偏差,也能减轻这些偏差对最终预测结果的影响。

1.3.3 基于 PSO 优化的 Stacking 多模型组合预测模型

本节将详细阐述基于 Stacking 的多模型组合方法在河蟹投饵量预测中的应用过程。如图4所示,首先需要

算,生成新特征,对数据集执行平滑处理和归一化,增强数据质量,提升预测效果。图4第二个阶段是模型的训练阶段,整个数据处理过程中的关键阶段。在此阶段,经过预处理后的特征数据集被作为输入,输入到 Stacking 多模型融合框架中进行训练,通过组合多个不同的模型来提高整体的预测性能,它在处理复杂问题和提高模型泛化能力方面具有显著优势。

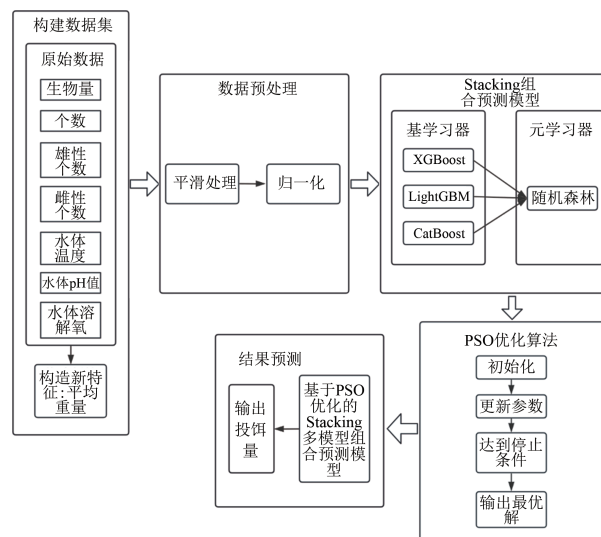


图4 PSO-Stacking 模型预测流程图

Fig. 4 Flowchart of PSO-Stacking model prediction

Stacking 模型设计采用两层结构。图4第1层,模型由3个基础学习器构成,分别是 XGBoost、LightGBM 和 CatBoost。3种模型都是当前机器学习领域中在处理回归问题上广泛应用的算法,这些基础回归模型独立地使用输入的相同特征数据集进行训练,它们各自会从数据中学习并提取出有用的信息,但彼此之间不进行直接的交互。当每个基学习器完成训练后,它们会各自根据输入的特征数据集生成一维向量作为输出。这些一维向量包含模型从数据中学习到的关键信息,它们共同形成新的向量。这个向量被用作 Stacking 模型第2层输入,为下一阶段预测提供基础。

在第2层中,采用随机森林算法作为元学习器。随机森林本身便是一种的集成学习方法,它通过构建多个决策树并综合它们的预测结果来提高整体的预测准确性。在这个框架下,元学习器将第1层3个基学习器产生的输出作为其输入,基于第1层模型提取的特征和信息来进一步学习并对投饵量进行最终预测,通过这种方式,Stacking 模型能够充分利用不同算法优势,提高预测准确性和稳定性。

为了进一步提升模型性能,采用粒子群优化算法(Particle swarm optimization, PSO)对 Stacking 集成学习模型进行超参数寻优。粒子群优化算法是一种模拟自然界群体行为的计算技术,它借鉴了鸟群捕食的协作模式来解决优化问题^[33-34]。在 PSO 算法中,每个潜在的解决方案都被视为搜索空间中的一个“粒子”,它们通过迭代

过程寻求最优解决方案。每个粒子表示一个解决方案，它们在解空间内移动，通过集体指导搜索过程。粒子通过更新位置和速度来逐步接近最优解，这一过程受到惯性权重、个体极值和全局极值的影响。每次迭代过程中，算法会评估粒子适应度，以识别个体历史最佳位置和全体粒子全局最佳位置。粒子群算法中粒子位置和速度更新是算法的核心机制，决定了粒子在解空间中的移动方式。

粒子速度决定了它在下一次迭代中移动方向和距离。速度更新公式包括：前速度影响、个体最优影响以及全局最优影响 3 个部分。具体公式为

$$v_i^{k+1} = v_i^k \omega + c_1 \text{rand}() (pbest_i^k - x_i^k) + c_2 \text{rand}() (gbest_i^k - x_i^k) \quad (5)$$

式中 v_i ——粒子速度

ω ——惯性系数

c_1, c_2 ——学习因子

k ——当前迭代代数

$\text{rand}()$ ——介于 0~1 之间的随机数

$pbest(i)$ ——粒子个体最优位置

$gbest(i)$ ——全局最优位置

$x(i)$ ——粒子当前位置

v_i^{k+1} ——粒子 $k+1$ 轮速度

粒子位置表示问题的一个解，通过加上新的速度进行位置更新，公式为

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (6)$$

式中 x_i^{k+1} ——粒子 $k+1$ 轮位置

x_i^k ——粒子 k 轮位置

通过更新位置，粒子群算法能够有效地在解空间中搜索最优解，不断地通过个体经验和群体经验来调整自己的移动方向和步长。利用 PSO 算法全局搜索能力可避免模型参数陷入局部最优解，帮助其找到全局最优参数，降低误差，使得模型预测效果更加准确。

1.4 模型评价指标

采用决定系数 (R^2)、平均绝对误差 (Mean absolute error, MAE) 和均方根误差 (Root mean squared error, RMSE) 来评价模型性能。

2 结果与分析

2.1 投饵量预测结果

为了评估预测模型性能，将数据集以比例 8:2 划分为训练集和测试集。旨在确保模型在独立的数据集上进行训练和测试，从而有效地评估其泛化能力。

多次实验结果表明，本文预测模型在估计河蟹的投饵量方面，综合多个模型优点，最终实验结果表现出较低的误差水平，MAE 为 0.349 71 g，RMSE 为 0.491 14 g， R^2 达到 0.903 58。这些指标表明，模型具有较好

的预测效果，尤其是 R^2 值接近 1，说明模型能够解释绝大部分的变异。样本预测结果的散点图如图 5 所示，由图 5 可知，大多数数据点紧密地分布在对角线附近，这表明预测值与实际值之间的一致性较高。散点图的这种分布进一步证实了模型在预测河蟹投饵量方面的有效性。其拟合折线对比如图 6 所示。当两条曲线近乎重叠时，这通常意味着模型捕捉到了数据的主要特征和规律，在预测方面表现出了较高的精度。由图 6 可以看出，两条曲线重合程度较高。这表明模型预测数据与真实数据紧密匹配，表明模型具有较好的预测性能。

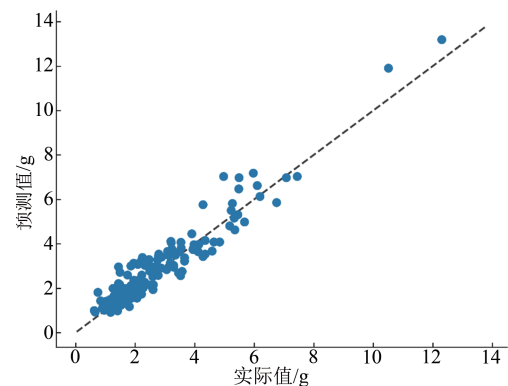


图5 模型预测值与实际值散点图

Fig. 5 Scatterplot of model predicted vs. actual values

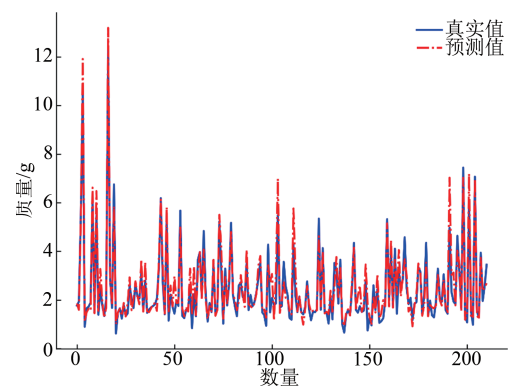


图6 模型预测值与实际值拟合折线对比

Fig. 6 Comparison of model predicted values with fitted line of actual values

2.2 不同模型结果对比

为了验证本文提出的基于粒子群优化的 Stacking 多模型组合投饵量预测模型的有效性，将该组合模型与单一的基学习器和元学习器进行了比较，包括 XGBoost、LightGBM、CatBoost 和随机森林等，旨在评估本文模型相比于单一模型和其他类型的组合模型在投饵量预测方面的优势。同时将提出的粒子群优化组合模型与未添加粒子群优化的组合模型进行比较，目的是探究粒子群优化算法在模型性能提升中的作用。此外还将未添加平均质量的与添加平均质量特征的模型进行对比，探究添加新特征对模型准确率的提升效果。

表1 为所有对比实验结果。对比实验能够全面评估本文提出的 PSO-Stacking 多模型组合投饵量预测模型的有效性,并与现有模型进行性能对比,证明其在投饵量预测方面的优越性。

表 1 不同模型对比实验结果

Tab. 1 Comparative experimental results of different models

模型	算法	R^2	MAE/g	RMSE/g
XGBoost	XGBoost	0.809 01	0.494 82	0.816 98
LightGBM	LightGBM	0.887 63	0.360 71	0.530 20
CatBoost	CatBoost	0.886 40	0.407 00	0.604 95
RandomForest	RandomForest	0.891 62	0.356 35	0.520 70
MLP	MLP	0.834 08	0.430 55	0.643 62
Stacking	XGBoost、LightGBM、CatBoost、RandomForest	0.893 66	0.361 35	0.515 79
PSO-Stacking-7	PSO、XGBoost、LightGBM、CatBoost、RandomForest	0.897 09	0.363 37	0.507 40
PSO-Stacking-8(本文)	PSO、XGBoost、LightGBM、CatBoost、RandomForest	0.903 58	0.349 71	0.491 14

由表 1 可知,PSO-Stacking-8 在各个指标上均优于 XGBoost、LightGBM、CatBoost、随机森林和 MLP 单一模型算法, R^2 分别提高 0.094 57、0.015 95、0.017 18、0.011 96 和 0.069 5。相比于未添加 PSO 寻优算法和未添加平均质量特征的模型,在 R^2 分别提高 0.009 92 和 0.006 49, MAE 和 RMSE 均有所下降。这表明模型参数设置得到了更精细的调整,使模型能够更好地拟合数据,对未知数据的泛化能力得到了提升,添加的新特征也更好地帮助模型理解数据之间的关系,深入学习数据之间的

规律。

为了直观地看出对比模型与本文模型差异,将上述模型各项指标数据以柱状图和折线图的方式绘制展示,图 7 展示了各种模型 3 项指标比较结果。如图 7 所示,在关键性能指标 R^2 上,本文所提的模型表现最好,优于其他对比模型,证明了其预测效能的优越性。在 MAE 以及 RMSE 等评价指标方面,本文模型同样展现出较低的误差水平,进一步验证了其在精确度和稳定性方面相较于其他模型的优势。

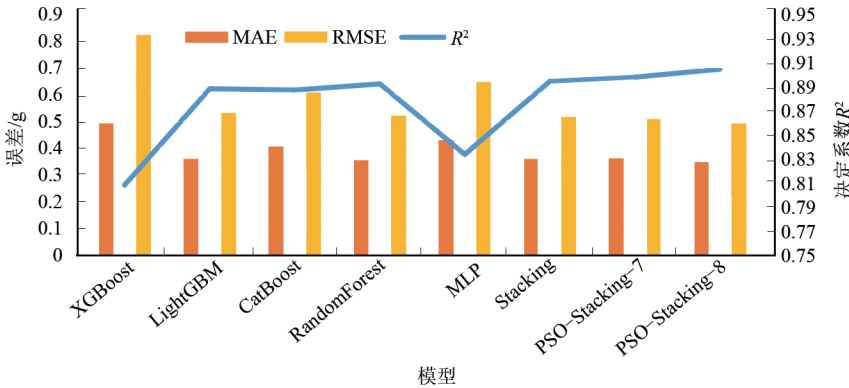


图 7 不同模型各项指标数据对比

Fig. 7 Comparison of various metrics for different models

3 结论

(1) 提出了基于粒子群优化的 Stacking 多模型组合方法,用于预测河蟹投饵量。将河蟹个数、性别比例、生物量、水体温度、水体 pH 值以及水体溶解氧含量作为重要参数输入模型用于投饵量预测。

(2) 采用 XGBoost、LightGBM 和 CatBoost 作为基学习器,随机森林作为元学习器,通过 Stacking 的方法进行组合。这种多模型组合方法旨在克服单一模型在投饵量

预测上的局限性。使用粒子群算法对组合模型进行参数寻优,能够最大化地发挥模型性能,有效提高投饵量预测精度。实验结果表明,本文模型平均绝对误差(MAE)为 0.349 71 g,均方根误差(RMSE)为 0.491 14 g,决定系数(R^2)到达 0.903 58。

(3) 本文模型与 XGBoost、LightGBM、CatBoost、随机森林、MLP 等算法模型进行了对比。对比结果显示,本文模型性能优于这些单一模型,能够更准确地实现投饵量预测。

参 考 文 献

- [1] 陈福平. 蟹塘套养青虾生态养殖技术[J]. 农村新技术, 2023(3):31-33.
- [2] 杨蓓蓓. 水产养殖信息化关键技术研究现状与趋势[J]. 农业工程技术, 2021,41(36):84-85.
- [3] LI H, CHEN Y, LI W, et al. An adaptive method for fish growth prediction with empirical knowledge extraction [J]. Biosystems Engineering, 2021,212:336-346.
- [4] 梁惠, 李晓玉, 关沛. 新型水产养殖模式介绍[J]. 渔业致富指南, 2020(22):36-37.
- [5] ZHANG L, WANG J, LI B, et al. A MobileNetV2-SENet-based method for identifying fish school feeding behavior [J]. Aquacultural Engineering, 2022,99:102288.
- [6] 丰慧. 我国内陆水产健康养殖发展现状与建议研究[J]. 现代农机, 2022(1):105-106.
- [7] 孙月平, 陈祖旭, 赵德安, 等. 池塘河蟹养殖精准投饵系统设计与试验[J]. 农业机械学报, 2022,53(5):291-301.
SUN Yueping, CHEN Zuxu, ZHAO Dean, et al. Design and experiment of precise feeding system for pond crab culture [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 291-301.(in Chinese)
- [8] 孙月平, 袁必康, 詹婷婷, 等. 基于改进 YOLO v5s 的河蟹与饵料检测方法[J]. 农业工程学报, 2023,39(22):178-187.
SUN Yueping, YUAN Bikang, ZHAN Tingting, et al. Detecting river crab and bait using improved YOLO v5s [J]. Transactions of the CSAE, 2023, 39(22): 178-187.(in Chinese)
- [9] MRAMBA R P, KAHINDI E J. Pond water quality and its relation to fish yield and disease occurrence in small-scale aquaculture in arid areas [J]. Heliyon, 2023,9(6):e16753.
- [10] DE VERDAL H, KOMEN H, QUILLET E, et al. Improving feed efficiency in fish using selective breeding: a review [J]. Reviews in Aquaculture, 2018,10(4):833-851.
- [11] CAPETILLO-CONTRERAS O, PEREZ-REYNOSO F D, ZAMORA-ANTUNANO M A, et al. Artificial intelligence-based aquaculture system for optimizing the quality of water: a systematic analysis [J]. Journal of Marine Science and Engineering, 2024,12(1): 161.
- [12] DOYLE J M, MEROVITCH N, WYETH R C, et al. A simple automated system for appetitive conditioning of zebrafish in their home tanks [J]. Behavioural Brain Research, 2017,317:444-452.
- [13] 田世杰, 张一名. 机器学习算法及其应用综述[J]. 软件, 2023,44(7):70-75.
- [14] 何津民, 张丽珍. 基于自注意力机制和 CNN-LSTM 深度学习的对虾投饵量预测模型[J]. 大连海洋大学学报, 2022,37(2):304-311.
HE Jinmin, ZHANG Lizhen. Prediction model of shrimp feeding amount based on self-attention mechanism and CNN-LSTM deep learning [J]. Journal of Dalian Ocean University, 2022, 37(2): 304-311.(in Chinese)
- [15] 周文英, 史文崇. 机器学习在渔业研究中的应用进展与展望[J]. 渔业研究, 2022,44(4):407-414.
- [16] ABANA E, BARICAUA M, CASIBANG R J, et al. Tetra-parameter fish feeding machine [J]. International Journal of Circuits, 2020, 2020(14):923-931.
- [17] 赵思琪, 赵三琴, 国振淇, 等. 池塘水产养殖精准投喂系统研制与试验[J]. 农业工程学报, 2023,39(22):27-34.
ZHAO Siqi, ZHAO Sanqin, GUO Zhenqi, et al. Development of the precise feeding system for pond culture [J]. Transactions of the CSAE, 2023, 39(22): 27-34.(in Chinese)
- [18] 孙龙清, 王新龙, 王泊宁, 等. 基于 ResNet-CA 的鱼群饱腹程度识别方法[J]. 农业机械学报, 2022,53(增刊2):219-225.
SUN Longqing, WANG Xinlong, WANG Boning, et al. Identification method of fish satiation level based on ResNet-CA [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(Supp.2): 219-225.(in Chinese)
- [19] 段青玲, 陈鑫, 许冠华, 等. 基于多维度特征和 LightGBM 的大闸蟹质量估算方法[J]. 农业机械学报, 2022,53(8):353-360.
DUAN Qingling, CHEN Xin, XU Guanhua, et al. Chinese mitten crab weight estimation method based on multi-dimensional features and LightGBM [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 353-360.(in Chinese)
- [20] 詹婷婷. 基于机器视觉的河蟹养殖投饵量精准确定方法研究[D]. 镇江:江苏大学, 2023.
ZHAN Tingting. Research on accurate determination of feeding quantity for river crab farming based on machine vision [D]. Zhenjiang: Jiangsu University, 2023. (in Chinese)
- [21] 陈永彤. 室内青蟹养殖自动投喂系统设计与控制研究[D]. 杭州:浙江科技学院, 2023.
CHEN Yongtong. Research on design and control of automatic feeding system for indoor mud crab breeding [D]. Hangzhou: Zhejiang University of Science and Technology, 2023. (in Chinese)
- [22] 王嘉煜. 基于深度学习的养殖水温与河蟹摄食行为关联模型研究[D]. 北京:中国农业大学, 2023.
WANG Jiayu. The correlation model between aquaculture water temperature and feeding behavior of river crabs based on deep learning [D]. Beijing: China Agricultural University, 2023. (in Chinese)
- [23] 陈纪溪. 基于机器学习的河蟹日摄食量预测模型研究及系统开发[D]. 北京:中国农业大学, 2024.
CHEN Jixi. Research and system development of daily food intake prediction model of river crab based on machine learning [D]. Beijing: China Agricultural University, 2024. (in Chinese)
- [24] 潘晨. 基于计算机视觉的河蟹行为识别与行为评价模型研究及系统设计[D]. 北京:中国农业大学, 2024.
PAN Chen. System design of river crab behavior recognition and behavior evaluation model based on computer vision [D]. Beijing: China Agricultural University, 2024. (in Chinese)
- [25] SAGI O, ROKACH L. Ensemble learning: a survey [J]. Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, 2018,8:e1249.

- ZHOU Zeyu. Metabolomics analysis of *Medicago sativa* L. in response to saline and alkali stress[D]. Harbin: Harbin Normal University, 2022. (in Chinese)
- [28] 李朋亮. 枸杞干制中黄酮类化合物变化规律及其抗氧化活性研究[D]. 银川: 宁夏大学, 2014.
LI Pengliang. Study on varying patterns of flavonoids of *Lycium barbarum* L. during drying and its antioxidant activities [D]. Yinchuan: Ningxia University, 2014. (in Chinese)
- [29] 李宁宁. 热风干制过程中枸杞褐变反应研究[D]. 天津: 天津科技大学, 2016.
LI Ningning. Research on *Lycium barbarum* browning reaction in hot air drying processing[D]. Tianjin: Tianjin University of Science and Technology, 2016. (in Chinese)
- [30] 张雨, 韩延超, 刘瑞玲, 等. FIRD-HPD 方式对蓝莓干燥特性及品质的影响[J]. 中国食品学报, 2024, 24(6): 237 - 247.
ZHANG Yu, HAN Yanchao, LIU Ruiling, et al. Effect of FIRD-HPD method on drying characteristics and quality of blueberry [J]. Journal of Chinese Institute of Food Science and Technology, 2024, 24(6): 237 - 247. (in Chinese)
- [31] DENG L, YANG X, MUJUMDAR S A, et al. Red pepper (*Capsicum annuum* L.) drying: effects of different drying methods on drying kinetics, physicochemical properties, antioxidant capacity, and microstructure[J]. Drying Technology, 2018, 36(8): 893 - 907.
- [32] 仇徐亮, 冯蕾, 聂梅梅, 等. 超声预处理对芥菜微波干燥品质的影响[J]. 农业工程学报, 2024, 40(6): 155 - 167.
- [33] XIAO H W, GAO Z J, HAI L, et al. Air impingement drying characteristics and quality of carrot cubes [J]. Journal of Food Process Engineering, 2010, 33(5): 899 - 918.
- [34] 谭宏渊, 凌玉钊, 黄丽琪, 等. 不同预处理对热风干燥山药片品质特性及微观结构的影响 [J]. 食品工业科技, 2023, 44(20): 43 - 52.
- [35] SILA D N, BUGGENHOUT S V, DUVETTER T, et al. Pectins in processed fruits and vegetables: part II —structure-function relationships[J]. Comprehensive Reviews in Food Science & Food Safety, 2010, 8(2): 86 - 104.
- [36] XIAO M, YI J Y, BI J F, et al. Modification of cell wall polysaccharides during drying process affects texture properties of apple chips [J]. Journal of Food Quality, 2018, 2018: 1 - 11.
- [37] DENG Y, WU Y, LI Y F. Changes in firmness, cell wall composition and cell wall hydrolases of grapes stored in high oxygen atmospheres [J]. Food Research International, 2005, 38(7): 769 - 776.
- [38] ROECK A D, SILA D N, DUVETTER T, et al. Effect of high pressure/high temperature processing on cell wall pectic substances in relation to firmness of carrot tissue[J]. Food Chemistry, 2008, 107(3): 1225 - 1235.
- [39] ZHOU M, BI J F, LYU J, et al. Structural conversion of pectin fractions during heat processing in relation to the ability of inhibiting lipid digestion: a case study of hawthorn pectin[J]. Food hydrocolloids, 2021, 117(6): 106721.
- [40] SILA D N, DOUGLA E, SMOUT C, et al. Pectin fraction interconversions: Insight into understanding texture evolution of thermally processed carrots[J]. Journal of Agricultural and Food Chemistry, 2006, 54(22): 8471 - 8479.
- [41] LIU Y X, ZHANG W D, YANG W T, et al. The effects of drying temperature on the kinetics, color, structure, and pectin composition of *Zizyphus jujuba* Mill. Cv. Junzao[J]. Journal of Food Processing and Preservation, 2021, 45(11): e15942.

~~~~~

( 上接第 309 页 )

- [26] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016:785 - 794.
- [27] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [J]. Advances in Neural Information Processing Systems, 2017,30:3146 - 3154.
- [28] PROKHORENKOVA L, GUSEV G, VOROBAY A, et al. CatBoost: unbiased boosting with categorical features[J]. Advances in Neural Information Processing Systems, 2018,31:6639 - 6649.
- [29] BIAU G. Analysis of a random forests model[J]. Journal of Machine Learning Research, 2012,13:1063 - 1095.
- [30] LUO J, FENG Y, WU X, et al. ForestSubtype: a cancer subtype identifying approach based on high-dimensional genomic data and a parallel random forest[J]. BMC Bioinformatics, 2023,24(1):289.
- [31] WANG Y, CAI D, CHEN L, et al. A downscaling methodology for extracting photovoltaic plants with remote sensing data: from feature optimized random forest to improved HRNet[J]. Remote Sensing, 2023,15(20):4931.
- [32] WANG Z, BI Y, LU G, et al. Monitoring forest diversity under moso bamboo invasion: a random forest approach [J]. Forests, 2024, 15(2):318.
- [33] 韩丁, 马子寅, 王俊林, 等. 基于粒子群算法的超材料吸波体传感器逆设计[J]. 中国激光, 2022,49(17):168 - 175.  
HAN Ding, MA Ziyin, WANG Junlin, et al. Inverse design of metamaterial absorber sensor based on particle swarm optimization [J]. Chinese Journal of Lasers, 2022, 49(17): 168 - 175. (in Chinese)
- [34] COBO A, LLORENTE I, LUNA L, et al. A decision support system for fish farming using particle swarm optimization[J]. Computers and Electronics in Agriculture, 2019,161:121 - 130.