

基于Bootstrap-GA-ELM算法的清香型白酒发酵过程 酒醅淀粉和水分含量区间预测模型

张贵宇, 向星睿*, 张磊, 王怡博, 严俊, 张云龙

(四川轻化工大学自动化与信息工程学院, 人工智能四川省重点实验室, 四川宜宾 644000)

摘要: 针对清香型白酒发酵过程酒醅淀粉和水分含量信息难以获取的困境, 且传统点预测模型仅提供单一预测值而难以量化结果的可靠度的问题, 提出一种发酵过程酒醅淀粉和水分含量区间预测模型。首先, 从黄水理化参数和酒醅温度等易测潜在解释变量中, 通过相关系数、最小角回归 (LARS) 综合分析确定酒醅淀粉和水分预测模型的输入参数。其次, 通过遗传算法 (GA) 优化极限学习机 (ELM) 的隐层权值和阈值, 建立精度较高的酒醅淀粉和水分点预测模型。最后, 通过 4 种不同自助法 (Bootstrap) 对点预测结果、系统误差以及随机误差进行估计, 以此构建不同置信度下的酒醅淀粉和水分含量预测区间。结果表明, 在 99% 置信度下, 基于 Residual Bootstrap 方法建立的区间预测模型效果最好, 在酒醅淀粉和水分测试集上, 点预测的判定系数 (R^2) 和均方根误差 (RMSE) 分别为 0.998 9、0.109 1, 0.920 3、0.802 1, 预测区间的区间覆盖率 (PICP) 和平均预测区间宽度 (MPIW) 分别为 100%、0.728 0%, 100%、4.339 9%。该研究可对清香型白酒发酵过程酒醅淀粉和水分含量进行可靠预测, 为白酒生产指导和发酵过程透明化提供参考。

关键词: 酒醅; 淀粉; 水分; 黄水; 相关性; 点预测; 区间预测; Bootstrap 方法

文章编号: 1673-9078(2026)3-211-222

DOI: 10.13982/j.mfst.1673-9078.2026.3.0016

Interval Prediction Model for Starch and Moisture Content in Fermenting Grains during Light-flavor Baijiu Fermentation Based on the Bootstrap-GA-ELM Algorithm

ZHANG Guiyu, XIANG Xingrui*, ZHANG Lei, WANG Yibo, YAN Jun, ZHANG Yunlong

(School of Automation & Information Engineering, Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644000, China)

Abstract: To address the challenges of obtaining starch and moisture content information from fermented grains during the fermentation of light-flavor Baijiu and the limitations of traditional point prediction models, an interval prediction model was proposed for measuring starch and moisture content in fermented grains during fermentation. Firstly, input parameters for the starch

引文格式:

张贵宇, 向星睿, 张磊, 等. 基于Bootstrap-GA-ELM算法的清香型白酒发酵过程酒醅淀粉和水分含量区间预测模型[J]. 现代食品科技, 2026, 42(3): 211-222.

ZHANG Guiyu, XIANG Xingrui, ZHANG Lei, et al. Interval prediction model for starch and moisture content in fermenting grains during light-flavor Baijiu fermentation based on the Bootstrap-GA-ELM algorithm [J]. Modern Food Science and Technology, 2026, 42(3): 211-222.

收稿日期: 2025-01-04; 修回日期: 2025-03-14; 接受日期: 2025-03-17

基金项目: 四川省科技计划项目 (2022YFS0554); 酿酒生物技术及应用四川省重点实验室开放课题 (NJ2022-06); 泸州老窖研究生创新基金 (LZCX2023-7); 四川轻化工大学科技成果转化专项项目 (HXJY01); 五粮液产学研合作项目 (CXY2022ZR007)

作者简介: 张贵宇 (1987-), 男, 博士, 副教授, 研究方向: 白酒自动化、人工智能, E-mail: gyz_118@163.com

通讯作者: 向星睿 (1999-), 男, 硕士研究生, 研究方向: 智能酿造, E-mail: 1460863207@qq.com

and moisture prediction models were determined through a comprehensive analysis of easily measurable potential explanatory variables, such as physicochemical parameters of Huangshui and fermented grain temperature, using correlation coefficients and least-angle regression. Secondly, a high-accuracy point prediction model for starch and moisture content in fermented grains was established by optimizing hidden layer weights and thresholds of the extreme learning machine using a genetic algorithm. Finally, four different Bootstrap methods were employed to estimate point prediction results, systematic errors, and random errors, thereby constructing prediction intervals for starch and moisture content in fermented grains at different confidence levels. Results indicated that, at a 99% confidence level, the interval prediction model based on the Residual Bootstrap method exhibited superior performance. On the test set for starch and moisture content in fermented grains, the coefficients of determination for points predictions were 0.998 9 and 0.109 1, respectively, and the root-mean-square error for point predictions were 0.920 3 and 0.802 1, respectively. The prediction interval coverage probabilities were 100% and 0.728 0%, and the mean prediction interval widths were 100% and 4.339 9%, respectively. This study provides reliable predictions for starch and moisture content in fermented grains during the production of light-flavor Baijiu, offering valuable references for production guidance and improving transparency in the fermentation process.

Key words: fermented grains; starch; moisture; Huangshui; relevance; point prediction; interval prediction; bootstrap method

白酒作为中国国酒,拥有悠久的历史传承^[1]。其生产过程包括蒸粮、打量水、摊晾撒曲、入窖、发酵、出窖、蒸酒等步骤,其中发酵是决定白酒口感和品质的关键环节^[2]。酒醅作为发酵过程中能量流动、信息传递和物质循环的载体,其淀粉和水分含量直接影响微生物的生长、繁殖和代谢,进而决定酒精和风味物质的生成^[3,4]。因此,酒企常通过监测酒醅的理化参数来控制发酵质量^[5]。但由于白酒固态发酵的密封工艺限制,无法在发酵过程中取出酒醅进行检测,导致酒醅理化参数信息脱节,酒企主要进行入窖和出窖阶段的酒醅检测,目前可监测的过程参数主要集中在温度上^[6],难以及时掌握发酵过程酒醅淀粉和水分含量信息,这给白酒生产指导带来了困难。

黄水是酒醅发酵过程中产生的棕黄色液体副产物,作为窖内发酵信息与物质的载体,与酒醅发酵质量密切相关^[7]。现有研究表明,黄水与酒醅在理化参数、风味物质和微生物上具有较强相关性,可通过检测黄水指标间接反映窖内酒醅的发酵情况^[8,9],且黄水取样方便,不会破坏发酵环境。此外,酒醅温度作为发酵过程另一易测参数,是影响发酵状况的主要因素之一,故可以结合黄水理化参数和酒醅温度,建立酒醅淀粉和水分的间接预测模型,为发酵过程酒醅监测提供新方法。

传统点预测方法只能提供确定性预测值,无法评估预测结果的可靠度,容易导致决策错误^[10]。相比之下,区间预测通过构造预测区间量化预测可靠度,能够有效弥补点预测的不足。目前,区间预测方法包括 Delta 法^[11]、Bayesian 法^[12]、均值方差估计法^[13]、上下限估计法^[14]和 Bootstrap 法^[15]等。与前四种相比,Bootstrap 法无需对数据分布进行假设,通过反复抽样将小样本

问题转化为大样本问题,具有计算简便、预测区间精度高的优势^[16,17],已在滑坡位移预测^[18]、风力发电预测^[19]、价格预测^[20]等领域广泛应用,但在白酒领域的应用研究仍较少。

综上所述,本文选用发酵周期较短(12 d)的清香型白酒发酵过程为研究对象,从发酵过程黄水酒精度、pH 值、还原糖、淀粉等常规理化参数和酒醅温度入手。初步把它们作为发酵过程酒醅淀粉和水分的潜在解释变量,通过相关系数、最小角回归(Least Angle Regression, LARS),分析验证其与酒醅淀粉和水分的关联性,确定酒醅淀粉和水分预测模型的输入参数。采用遗传算法(Genetic Algorithm, GA)优化极限学习机(Extreme Learning Machine, ELM)建立发酵过程酒醅淀粉和水分点预测模型。并引入区间预测思想,通过 Bootstrap 方法构建区间预测模型来量化点预测的可靠度,为发酵过程提供兼具精度与可靠度的酒醅淀粉和水分含量信息。

1 材料与方法

1.1 材料与试剂

酒醅和黄水,依托四川某清香型白酒企业的生产糟车进行,该糟车中部内置可插入式温度传感模块,底部设计有配备滤网结构的黄水收集室。该企业采用小曲清香型固态发酵工艺,采样时的酒醅入窖参数约为:温度 22~25 °C、水分含量 53%~60%、淀粉含量 32%~38%、酸度 0.5~0.8。采样周期设定为黄水产生后的 10 d 内,两天为一个发酵阶段进行取样,即选用第 4、6、8、10、12 天的酒醅和黄水作为研究对象。采用如图 1 所示的五点取样法,在糟车中心和四周的采样点范围内采集中层酒醅样本,以干净玻璃棒搅拌混合备

用,同时通过阀门接取黄水样本,每次采集酒醅 100 g 和黄水 150 mL。本次实验共采集了 50 个糟车的发酵过程样本,累计获得 250 个酒醅和黄水样品。

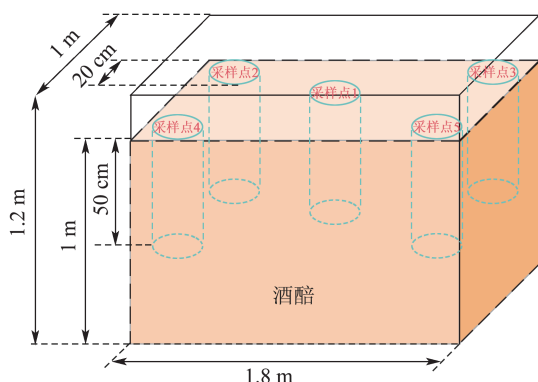


图 1 酒醅五点取样示意图

Fig.1 Schematic diagram of five point sampling of fermented grains

NaOH、葡萄糖、五水硫酸铜、次甲基蓝、酒石酸钾钠、亚铁氰化钾、HCl 溶液 (均为分析纯), 酚酞溶液 ($10 \text{ g} \cdot \text{L}^{-1}$), NaOH 标准滴定溶液 ($0.1 \text{ mol} \cdot \text{L}^{-1}$), 成都市科隆化学有限公司。

DHG-9245A 电热鼓风干燥箱, 上海一恒科学仪器有限公司; 雷磁 PHS-3C 型 pH 计, 上海仪电科学仪器有限公司; BSA223S-CW 电子天平, 赛多利斯科学仪器有限公司。

1.2 方法

1.2.1 酒醅和黄水理化数据采集

依据发酵周期对黄水、酒醅样品进行采集, 并记录每天的酒醅发酵温度。采用安徽省地方标准 DB34/T 1728-2012《白酒固态发酵黄水中常规指标的分析方法》中的蒸馏法和葡萄糖反滴定法对黄水酒精度、淀粉、还原糖等理化参数进行测定, 以及《酿酒分析与检测》中 pH 计法测定黄水 pH; 采用安徽省地方标准 DB34/T 2264-2014《固态发酵酒醅分析方法》中葡萄糖反滴定法和恒温干燥法对酒醅淀粉、水分含量进行测定。

1.2.2 酒醅淀粉和水分与潜在解释变量相关性分析方法

针对发酵过程酒醅淀粉和水分的解释变量选择, 考虑到单一方法的片面性, 本研究采用两种方法综合验证酒醅淀粉和水分与其潜在解释变量相关性, 为确定酒醅淀粉和水分预测模型的输入参数提供参考。先从发酵过程变量间整体相关性角度出发, 通过相关系数衡量酒醅淀粉和水分与其各潜在解释变量的相关性, 然后从自变量对目标变量模型预测能力贡献度的角度, 通过 LARS 迭代计算各潜在解释变量的回归

系数, 验证其与酒醅淀粉和水分相关性的同时剔除不重要变量。

(1) 相关系数: 最常用的相关系数是 Pearson 相关系数, 用来分析两个连续变量的线性相关程度, 当数据不能同时满足线性关系、连续、正态分布时, Pearson 相关系数就不再适合, 可采用 Spearman 相关系数代替, 其适用条件更广, 只要数据满足单调关系就能够使用。

(2) 最小角回归: LARS 方法通过构造一阶惩罚函数确定某些变量的回归系数为零, 并当作无效变量剔除, 在变量筛选上具有较强可解释性^[21]。其思想为: 首先把所有自变量系数设为 0, 在条件约束下, 沿着最小化角即自变量与残差之间的夹角方向, 逐步添加自变量到模型中, 每次添加与当前残差相关性最大的自变量, 通过调节回归系数使得真实值与回归值的总平方误差最小^[22]。

1.2.3 发酵过程酒醅淀粉和水分区间预测概况

在发酵过程酒醅淀粉和水分含量预测上, 预测结果的准确性和可靠度易受多源不确定性因素影响, 这些不确定性主要来源于两个方面。其一源于数据采集环节: 实验人员取样时的抽样偏差、理化指标检测阶段的操作误差 (如称量、滴定误差等), 以及实验设备的精度限制或噪声干扰引发的系统误差, 这些因素均会导致输入数据的不确定性; 其二源于模型构建层面: 预测模型选择的主观性及参数设置的局限性所产生的模型不确定性。这些不确定性因素很难完全避免, 只通过点预测模型实现发酵过程酒醅淀粉和水分含量预测, 其结果必然会受影响, 且无法量化预测结果的可靠度, 因此有必要引入区间预测方法来解决这一问题。区间预测可采用数学模型近似发酵过程酒醅淀粉、水分与解释变量之间的相互作用关系, 进一步估算其在一定置信度下的潜在取值范围, 以区间的形式展示预测结果, 为预测结果可靠度提供参考。

设数据样本集为 $D=[x_i, y_i]$, $i=1, 2, \dots, N$, 其中, x_i 表示第 i 个酒醅淀粉和水分的解释变量向量集, 作为为预测模型的输入参数, y_i 表示第 i 个酒醅淀粉、水分的实测值, 作为模型的输出。则输入参数 x_i 与输出参数 y_i 之间的数学模型为:

$$y(x_i) = f(x_i) + \varepsilon(x_i) \quad (1)$$

式中:

$f(x_i)$ ——理论上的回归值, 表示输入与输出的映射关系;

$\varepsilon(x_i)$ ——随机误差, 由输入数据不确定性导致, 一般服从均值为 0, 方差为 $\sigma_\varepsilon^2(x_i)$ 的标准正态分布。

实际情况下, 用人工智能算法建立的回归模型

的预测值 $t(x_i)$ 与理论回归值 $f(x_i)$ 会存在模型误差 $f(x_i) - t(x_i)$, 又叫系统误差, 主要由于模型选择和参数设置的不确定性造成。预测误差可表示为:

$$y(x_i) - t(x_i) = f(x_i) - t(x_i) + \varepsilon(x_i) \quad (2)$$

式中:

$y(x_i) - t(x_i)$ ——预测总误差, 区间预测是对预测总误差进行估计, 假设系统误差和随机误差是相互独立的, 则实测值 $y(x_i)$ 与预测值 $t(x_i)$ 之间预测总误差的方差为:

$$\sigma_y^2(x_i) = \sigma_t^2(x_i) + \sigma_\varepsilon^2(x_i) \quad (3)$$

式中:

$\sigma_t^2(x_i)$ ——系统误差方差;

$\sigma_\varepsilon^2(x_i)$ ——随机误差方差。

将 $\sigma_y^2(x_i)$ 通过概率预测的方法得到预测目标的概率分布, 便可得到对应置信水平下的预测区间, 在显著性为 α 下的预测区间 $[L_i^\alpha, U_i^\alpha]$ 可表示为:

$$L_i^\alpha = t(x_i) - Z_{1-\alpha/2} \sqrt{\sigma_y^2(x_i)} \quad (4)$$

$$U_i^\alpha = t(x_i) + Z_{1-\alpha/2} \sqrt{\sigma_y^2(x_i)} \quad (5)$$

式中:

L_i^α ——预测区间下界;

U_i^α ——预测区间上界;

$Z_{1-\alpha/2}$ ——标准正态分布下的 $1-\alpha/2$ 的百分位数。

采取适合的方法对 $\sigma_y^2(x_i)$ 进行估计, 对于构造的预测区间质量至关重要, 针对发酵过程酒醅和黄水理化样本数据具有样本量小且分布特性不明确的特点, 本研究选择 Bootstrap 方法来构造发酵过程酒醅淀粉和水分含量的预测区间。常用的 4 种 Bootstrap 方法为, Pairs Bootstrap 法、Block Bootstrap 法、Residual Bootstrap 法以及 Wild Bootstrap 法, 这些方法通过对原始数据进行不同抽样操作生成 Bootstrap 伪样本。

1.2.4 基于GA-ELM的酒醅淀粉和水分点预测模型

本研究采用 ELM 作为发酵过程酒醅淀粉和水分点预测模型, ELM 为 2004 年提出的一种单隐层前馈神经网络^[23], 相较于其他神经网络, 采用随机给定的隐层权重与阈值, 在训练过程中该网络的权重和阈值无需更新, 只需通过最小二乘准则和摩尔-彭洛斯广义逆矩阵理论计算输出层权重。该方法凭借快速的学习能力、泛化性好、训练误差小, 在定量预测上得到广泛应用。

考虑到 ELM 隐层神经元个数和权重与阈值的随机性对预测精度的影响, 先通过交叉验证找到最优神经元个数, 然后使用具有较强全局搜索能力和适应性的 GA, 对隐层权重和阈值进行参数寻优。GA 优化 ELM 建模过程如图 2 所示, 可归纳为以下步骤:

(1) 将样本集 D 按照 4:1 划分为训练集和测试集, 并将数据进行归一化处理消除量纲影响;

(2) 通过 10 折交叉验证, 在 $[10, L]$ 范围内寻找最优隐层神经元个数;

(3) 设置 ELM 隐层权重和阈值的搜索范围、种群规模、迭代次数、交叉与变异概率等初始值, 并通过实数编码产生初始种群;

(4) 将 ELM 预测值与实测值之间的均方根误差作为适应度函数, 计算种群中个体适应度, 当达到训练误差或迭代次数阈值时, 跳转至第 (6) 步, 否则进入第 (5) 步;

(5) 未达到终止条件时, 通过上代群体的适应度值, 分别进行选择、交叉、变异 3 种操作, 产生下一代群体, 把新种群转置第 (4) 步重新计算适应度;

(6) 迭代停止, 将得到的最优种群个体的参数赋予 ELM 的隐层权重与阈值, 利用该最优模型完成测试集样本的预测。

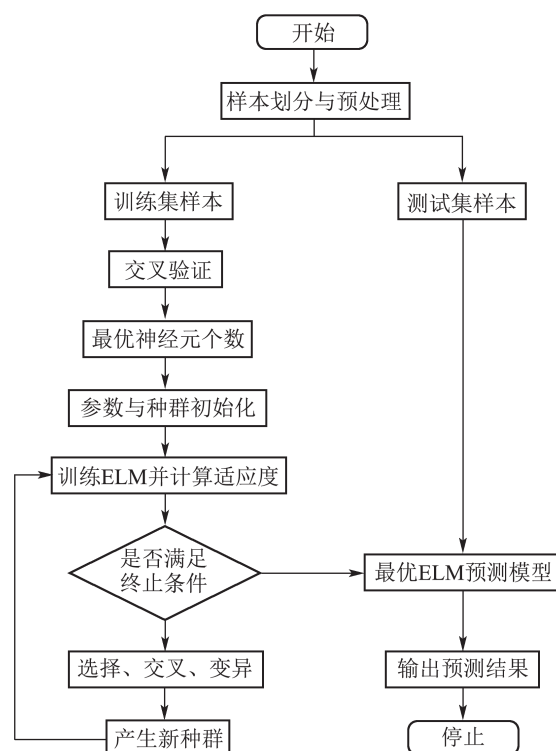


图 2 GA-ELM 点预测模型的流程图

Fig.2 GA-ELM point prediction model flowchart

1.2.5 基于Bootstrap-GA-ELM的酒醅淀粉和水分区间预测模型

为实现发酵过程酒醅淀粉和水分含量的区间预测, 本研究以 GA-ELM 点预测模型为基础, 同时采用 Bootstrap 方法构建预测区间, 量化预测结果的可靠度, 整个 Bootstrap-GA-ELM 区间预测模型框架如图 3 所示^[24], 其核心思路如下:

对于 1.2.3 节中所述发酵过程原始数据 $D=[x_i, y_i]$,

$i=1,2,\dots,N$, 采用 X-Y 共生距离法 (Sample Set Partitioning Based on Joint X-Y Distance, SPXY) 方法将其按照 4:1 划分为训练集样本 D_{train} 和测试集样本 D_{test} , 然后通过前文提到的 4 种不同 Bootstrap 方法对 D_{train} 进行有放回的随机重抽样, 重复进行 B 次, 获得 B 组伪样本 $D_{train}^*=\{D_1^*, D_2^*, \dots, D_B^*\}$ 。

按照 1.2.4 节中介绍的建模流程, 将新得到的 B 组伪样本来训练 B 个 GA-ELM 点预测模型, 并对原始数据 D 进行预测, 根据 Bootstrap 原理, 把 B 组预测值通过公式 (6)、(7) 得到最终的点预测结果 $\hat{t}(x_i)$ 与其对应的系统误差方差估计值 $\sigma_t^2(x_i)$ 。

$$\hat{t}(x_i) = \frac{1}{B} \sum_{b=1}^B t_b(x_i) \quad (6)$$

$$\sigma_t^2(x_i) = \frac{1}{B-1} \sum_{b=1}^B [t_b(x_i) - \hat{t}(x_i)]^2 \quad (7)$$

接着还需要对由输入数据不确定性产生的随机误差方差 $\sigma_e^2(x_i)$ 进行估计, 其估算公式如下:

$$\sigma_e^2(x_i) \cong E\{[y(x_i) - \hat{t}(x_i)]^2 - \sigma_t^2(x_i)} \quad (8)$$

为进一步估计随机误差方差 $\sigma_e^2(x_i)$, 需建立新的数据集 $D_{r2}=[x_i, r_i^2(x_i)]$, $i=1,2,\dots,N$, x_i 为原始数据 D 的自变量 $r_i^2(x_i)$ 为估计 $\sigma_e^2(x_i)$ 的平方残差值, 因其值不可为负数, 通过下式得到:

$$r^2(x_i) = \max\{[y(x_i) - \hat{t}(x_i)]^2 - \sigma_t^2(x_i), 0\} \quad (9)$$

然后利用数据集 D_{r2} 训练第 $B+1$ 个 GA-ELM 模型来预测 $\sigma_e^2(x_i)$, 为提高 $\sigma_e^2(x_i)$ 在样本中出现的概率, 可通过极大似然估计方法建立如下目标函数, 将其作为遗传算法里的适应度函数。

$$f_N = \frac{1}{2} \sum_{i=1}^N \left[\frac{r^2(x_i)}{\sigma_e^2(x_i)} + \ln(\sigma_e^2(x_i)) \right] \quad (10)$$

完成对 $\sigma_t^2(x_i)$ 与 $\sigma_e^2(x_i)$ 的估计后, 根据式 (3)~(5) 构建发酵过程酒醅淀粉和水分含量预测区间, $\hat{t}(x_i)$ 作为其点预测结果。

1.2.6 预测结果评价指标

点预测模型的准确度是很重要的, 本研究采用判定系数 (Coefficient of Determination, R^2) 与均方根误差 (Root Mean Square Error, RMSE) 作为发酵过程酒醅淀粉和水分点预测结果评价指标; 另外为更好衡量区间预测模型的预测区间质量, 采用区间覆盖率 (Prediction Interval Coverage Probability, PICP)、平均预测区间宽度 (Mean Prediction Interval Width, MPIW)、覆盖宽度指数 (Coverage Width Criterion, CWC) 作为发酵过程酒醅淀粉和水分区间预测结果评价指标, 其详细定义参考文献^[25]。

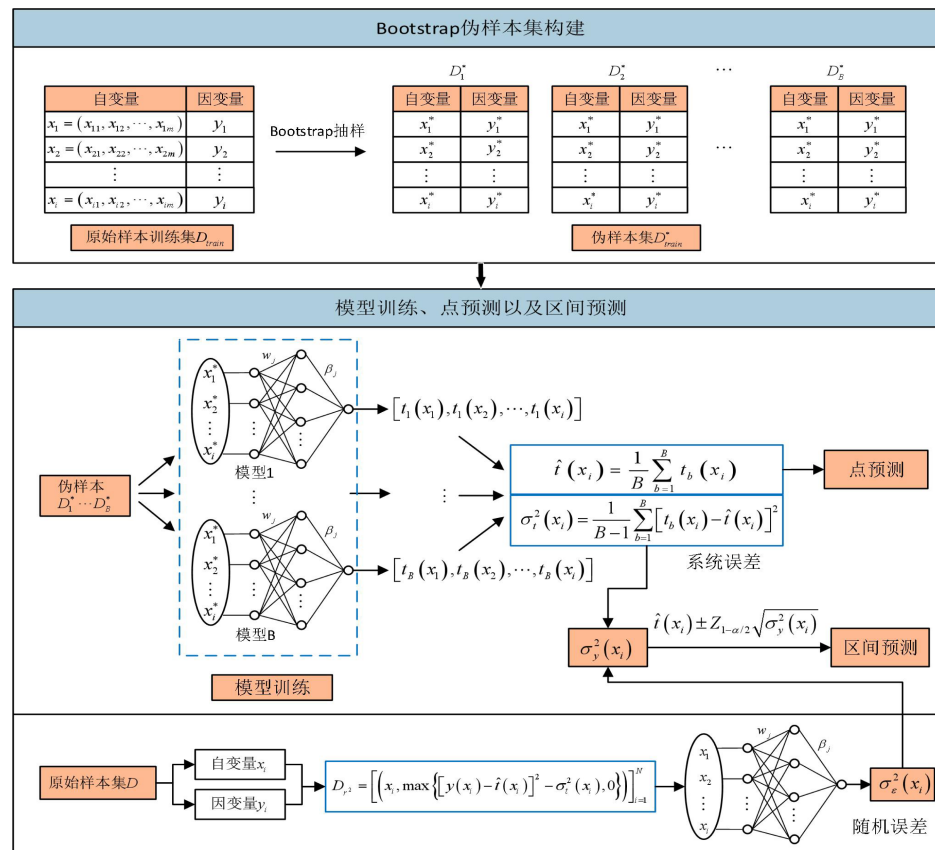


图 3 Bootstrap-GA-ELM 区间预测模型框架

Fig.3 Bootstrap-GA-ELM interval prediction model framework

1.2.7 数据分析

表格采用 Excel 2019 绘制，用 MATLAB 2020b 软件对数据进行相关性分析、点预测和区间预测模型建立，以及完成图形绘制，另外结构框图和流程图采用 Visio 2013 绘制。本文区间预测模型训练的硬件配置为：Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz 处理器，6 核 12 线程，以及 8 GB 内存，单次完整模型训练耗时约 46 min。

2 结果与讨论

2.1 发酵过程酒醪和黄水理化参数变化分析

温度是影响酒醪糖化发酵的重要指标，从图 4a 可以看出，发酵周期内酒醪温度变化基本呈“前缓、中挺、后缓落”的趋势，符合发酵过程酒醪温度变化规律^[26]。在酒曲中的酵母菌作用下，在发酵第 1~3 天内，酒醪温度从 23.13 ℃ 上升到最高 31.25 ℃，第 2 天温度上升幅度较第 1 天更快，第 3~5 天内温度维持在最高点附近，之后酒醪温度开始缓慢下降，从 30.56 ℃ 降低到 25.58 ℃。

在发酵前中期，淀粉会通过 α -淀粉酶、糖化酶等多种水解酶系的协同作用被糖化分解，生成葡萄糖等还原糖，为窖内微生物提供能源，同时酵母等微生物会消耗这些能源物质产生代谢水，从图 4b 可以看出在第 4~8 天内，酒醪水分含量迅速上升，从 63.52% 上升到 68.67%，淀粉含量迅速下降，从 33.41% 降到 26.24%。到发酵后期，受 CO₂、酒精、酸等物质堆积以及温度下降等因素的影响，微生物代谢活动逐渐放慢，淀粉分解和水分生产速度变缓，而整个发酵过程水分都会受重力作用落入窖池底部^[27]，使得在之后的第 8~12 天内酒醪水分含量上升变缓甚至可能降低，第 12 天达到最大值 69.13%，淀粉含量下降速度也趋于平缓，第 12 天降到最小值 24.83%。Kang 等^[28]的研究证明了此变化趋势。

黄水中淀粉和还原糖大部分来源于酒醪，由图 4c 可看出，与酒醪淀粉一样，黄水淀粉和还原糖在整个周期内呈现出被消耗的趋势。黄水淀粉和还原糖含量在第 4 天处于最高点，分别为 7.48%、6.67%，这是由于发酵的前几天酒醪中的淀粉被分解从而不断地产生还原糖，在酒醪中的水分达到一定量后会将部分淀粉和还原糖带入到窖池底部^[29]。在第 4~8 天内，因为酒醪和黄水中淀粉和还原糖被微生物代谢持续消耗，以及大量水分产生，使得黄水的淀粉和还原糖含量占比有较大幅度的下降，第 8 天分别降到 6.03%、5.73%。由于后期微生物代谢活动变慢，黄水中淀粉和还原糖

含量在第 8~12 天内下降速度变慢。该黄水淀粉、还原糖变化趋势与母雯竹等^[3]和罗琪等^[29]的研究一致。

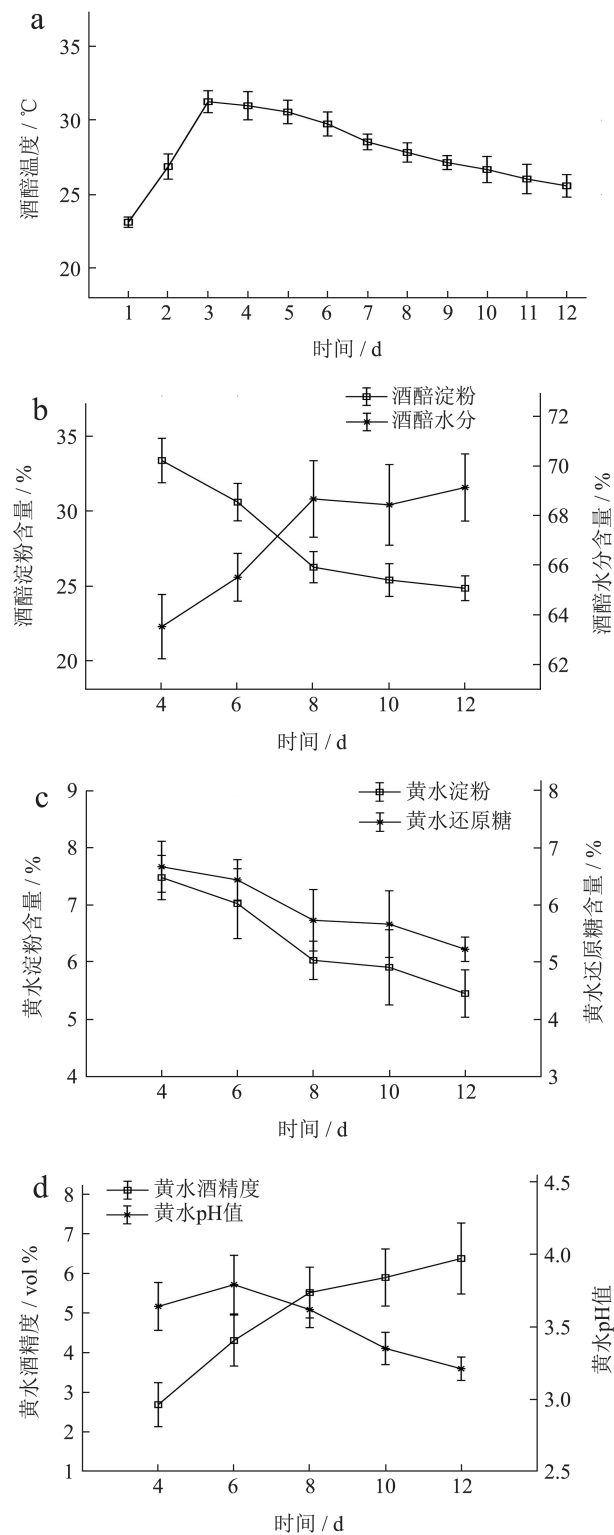


图 4 酒醪和黄水理化参数变化趋势

Fig.4 Change trends of physicochemical parameters for fermenting grains and Huangshui

黄水中酒精和酸主要来自酒醪产生的酒精和酸的沉降，以及黄水中微生物代谢产生。在发酵周期内，

微生物利用还原糖等营养物质代谢不断产生酒精和酸类物质^[30]。从图 4d 反映的变化趋势可知,在第 4~12 天内,黄水中的酒精度不断升高,从 2.69vol% 升到 6.38vol%,但增长速度在变缓,这与发酵过程中给酵母菌的活性变化有关。在第 4~6 天内,主要产生酒精,产酸较弱,且黄水量增大,导致黄水 pH 有些许上升,第 6 天之后伴随着酸类物质的增多,黄水酸度逐渐升高,使得黄水 pH 在发酵过程中整体上呈现下降趋势^[31],由 3.64 降到 3.21。

2.2 酒醅淀粉和水分预测模型输入参数确定

从前面发酵过程酒醅和黄水理化参数变化分析中可以看出,这些理化参数是有紧密联系的,本研究从黄水酒精度 (HJ)、pH (HS)、还原糖 (HH)、淀粉 (HD) 以及酒醅温度 (JW) 这些易测参数中,选择发酵过程酒醅淀粉 (JD) 和水分 (JS) 的解释变量作为预测模型的输入参数。

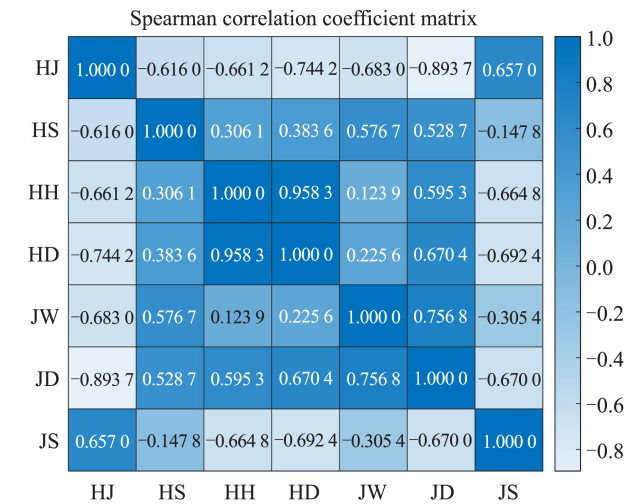


图 5 酒醅和黄水理化参数之间的 Spearman 相关系数

Fig.5 Spearman correlation coefficient between physicochemical parameters of fermented grains and Huangshui

对于相关系数方法的选择,在分别对 HJ、HS、HH、HD、JW、JD、JS 进行 Jarque-Bera 正态分布检验后,结果表明在 0.05 显著水平下并非所有变量都满足正态分布,因此选择 Spearman 相关系数计算这些理化参数之间的相关系数,其结果如图 5 所示。可以看出对于 JD 来说,其与潜在解释变量之间均有较显著的相关性 ($P<0.01$),相关系数绝对值均在 0.5 之上,其中,与 HJ 之间的相关系数高达 $-0.893\ 7$ ($P<0.01$),呈极强的单调负相关性,与 JW 之间的相关系数为 $0.756\ 8$ ($P<0.01$),呈较强的单调正相关性。对于 JS 来说,其与 HJ、HH、HD 之间均有较显著的相关性 ($P<0.01$),相关系数绝对值均在 0.6 之上,但与 JW

和 HS 之间的单调相关性较低,相关系数分别为 $-0.305\ 4$ ($P<0.01$)、 $-0.147\ 8$ ($P<0.05$)。

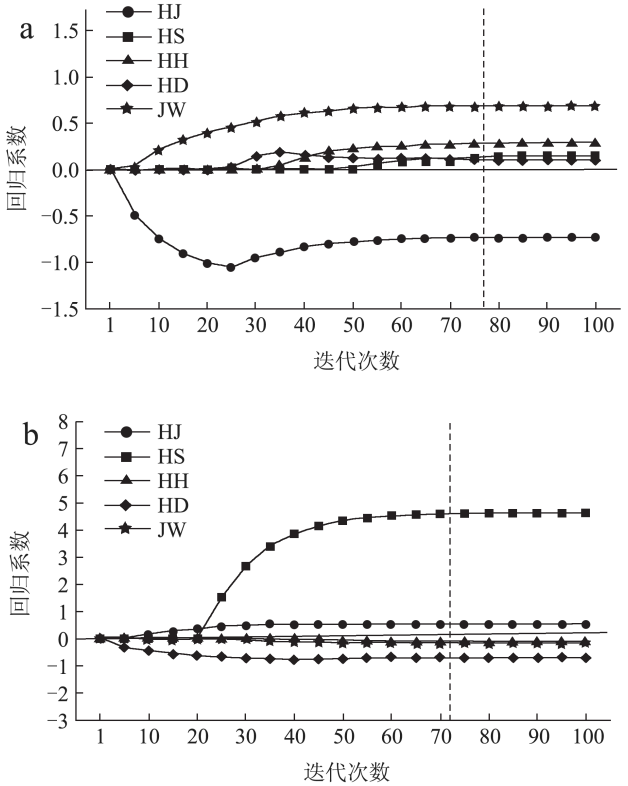


图 6 潜在解释变量的 LARS 回归系数变化

Fig.6 The change of LARS regression coefficient of potential explanatory variables

前面通过 Spearman 相关系数,验证了 JD 和 JS 与大部分潜在解释变量均具有较高的相关性,进一步使用 LARS 方法,验证潜在解释变量在模型预测 JD 和 JS 能力上的贡献度。将 JD 和 JS 作为预测输出,潜在解释变量作为输入变量集,计算各输入变量系数,其结果如图 6 所示。图 6a 为 JD 回归模型系数迭代图,可以看出在迭代到第 77 次时,预测平方误差最小,潜在解释变量被全部选中,此时系数分别为 $\beta_1=-0.734\ 4$ 、 $\beta_2=0.129\ 2$ 、 $\beta_3=0.278\ 3$ 、 $\beta_4=0.110\ 3$ 、 $\beta_5=0.678\ 8$; 图 6b 为 JS 回归模型系数迭代图,可以看出在迭代到第 72 次时,预测平方误差最小,潜在解释变量也被全部选中,此时系数分别为 $\beta_1=0.538\ 4$ 、 $\beta_2=4.612\ 8$ 、 $\beta_3=-0.112\ 6$ 、 $\beta_4=-0.697\ 1$ 、 $\beta_5=-0.143\ 4$ 。该结果表明所有潜在解释变量在模型预测 JD 和 JS 能力上均贡献其作用,在迭代过程中,各系数均未有接近 0 的趋势,表明这些潜在解释变量是 JD 和 JS 的重要变量,验证了其相关性。另外,与 JS 相关系数较低的 HS 和 JW,不会因其引入导致预测效果变差,证明了其对于模型的解释作用,并非无效变量,因此综合来看,可以把 HJ、HS、HH、HD、JW 作为发酵

过程 JD 和 JS 的解释变量，即 JD 和 JS 预测模型的输入参数。

2.3 样本集划分

利用 SPXY 法，将样本集 D 按照 4:1 的比例划分训练集和测试集，训练集样本有 200 个，测试集样本 50 个。从表 1 中可以看出，淀粉和水分训练集的分布范围覆盖了测试集的分布范围，且训练集和测试集的平均值和标准差相差较小，说明训练集和测试集的数据分布较一致，划分结果合理，有助于提升模型的泛化能力。

表 1 酒醅淀粉和水分样本集划分结果 (%)

Table 1 The division results of the starch and moisture sample sets for fermented grains (%)					
理化名称	数据集	最小值	最大值	平均值	标准差
淀粉	训练集	23.69	34.12	27.94	3.34
	测试集	22.98	33.41	28.01	3.31
水分	训练集	60.15	72.32	67.16	2.62
	测试集	60.30	71.43	67.04	2.87

2.4 GA-ELM点预测模型参数设置及结果分析

点预测模型的质量是区间预测好坏的重要保障，模型算法选择和参数设置很关键，经过多次调参验证，对 1.2.4 节中所述 GA-ELM 点预测模型结构中参数设置如下：JD 和 JS 点预测模型中隐层神经元个数的交叉验证寻优范围分别为[10,50]、[10,45]，隐层神经元激活函数使用 Sigmoid 函数，隐层权值和阈值的 GA 寻优范围为[-1,1]，种群规模为 150，最大迭代次数分别为 80、70，交叉概率为 0.8，变异概率为 0.2。

为验证 ELM 模型的预测精度，采用偏最小二乘回归 (Partial Least Squares Regression, PLSR)、支持向量机 (Support Vector Machine, SVM)、随机森林 (Random Forest, RF) 等算法与其对比，重复进行 10 次预测后取均值，结果如表 2 所示。从 4 种建模算法的结果来看，PLSR 在 JD 和 JS 预测精度上均为最差，特别是 JS 预测上表现出极低的精度， R^2 为 0.875 6、0.588 2，RMSE 为 1.175 0、1.740 8，说明 JD 和 JS 与解释变量之间并非简单的线性关系，PLSR 这种线性模型难以解释它们之间的非线性关系。ELM 凭借强大的非线性逼近能力，使得其在 JD 和 JS 预测上表现优于其他 3 种方法， R^2 为 0.983 7、0.904 7，RMSE 为 0.418 8、0.877 3，而且相对于 SVM、RF 等机器学习算法，ELM 模型结构简单，所需参数设置更少，计

算速度更快，对于后续构建预测区间需要建立大量子模型来讲，ELM 在建模效率和精度上更占优势。

接着通过麻雀搜索算法 (Sparrow Search Algorithm, SSA)、鲸鱼优化算法 (Whale Optimization Algorithm, WOA)、霜冰优化算法 (Rime Optimization Algorithm, RIME)、冠豪猪优化算法 (Crested Porcupine Optimizer, CPO) 等近些年提出的新算法与 GA 这种传统经典算法进行对比，验证 GA 在优化 JD 和 JS 点预测模型 (ELM) 隐层权值和阈值上的适配性。4 种算法的种群规模、最大迭代次数以及权阈值寻优范围与 GA 保持一致，SSA 中安全阈值、响应阈值、探索者比例等参数分别设置为 0.8、0.2、0.2，WOA 中螺旋系数、包围概率分别设置为 1、0.5，RIME 中结霜概率、温度衰减率、邻域搜索步长分别设置为 0.5、0.95、0.1，CPO 中防御系数、觅食概率、搜索强度分别设置为 0.3、0.2、0.5。图 7a 和 7b 分别为 5 种算法在 JD 和 JS 预测上的最优适应度迭代图，可以看到在迭代初期 (前 20 次)，与其他 4 个算法相比，GA 的适应度 (均方根误差) 呈现快速下降趋势，说明其可以快速逼近最优解，有效避免在早期迭代寻找最优解时存在搜索效率低的问题。并且 GA 最终适应度明显低于其他算法，证明其通过选择、交叉、变异操作有效探索解空间，可避免早熟收敛，不易陷入局部最优。另外从表 2 易发现，通过 GA 寻优权阈值的 ELM，在 JD 和 JS 测试集上点预测结果均为最优，表明其更适合 JD 和 JS 点预测模型的高维权阈值优化问题，相较于 ELM 模型预测精度进一步提升， R^2 为 0.998 8、0.912 9，RMSE 为 0.111 1、0.838 6。

表 2 不同算法在测试集上的点预测精度

Table 2 The point prediction accuracy on the test set for different algorithms				
模型名称	JD		JS	
	R^2	RMSE	R^2	RMSE
PLSR	0.875 6	1.175 0	0.588 2	1.740 8
SVM	0.956 0	0.688 1	0.884 5	0.948 3
RF	0.971 8	0.551 1	0.888 4	0.949 5
ELM	0.983 7	0.418 8	0.904 7	0.877 3
GA-ELM	0.998 8	0.111 1	0.912 9	0.838 6
SSA-ELM	0.998 1	0.143 3	0.910 8	0.844 4
WOA-ELM	0.997 7	0.157 1	0.910 5	0.849 4
RIME-ELM	0.998 1	0.143 4	0.911 9	0.842 9
CPO-ELM	0.997 2	0.173 2	0.906 1	0.869 7

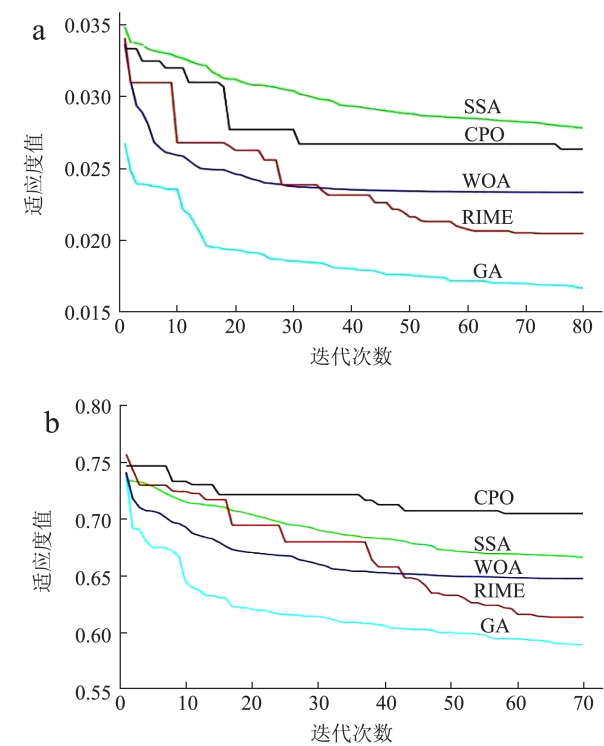


图 7 不同优化算法在 JD 和 JS 预测上的适应度迭代图
Fig.7 The fitness iteration chart of different optimization algorithms in JD and JS prediction

2.5 区间预测结果分析

在完成点预测模型参数设置后，按照 1.2.5 节所提方法对发酵过程 JD 和 JS 进行区间预测，其中，Bootstrap 重复抽样次数 B 设置为 50，Block Bootstrap 法中数据块长度设置为 5，Wild Bootstrap 法中随机权

重根据 Rademacher 分布生成，本研究所提出的区间预测方法在 90%、95%、99% 三种置信度下得到的区间预测结果如表 3 和表 4 所示。原则上构建的预测区间的 PICP 应大于其对应的置信度，但 Wild Bootstrap 法对 JS 测试集构建的预测区间的 PICP，在 90% 和 99% 置信度上均未达到这一要求，分别为 84%、98%。从整体结果可以看出，随着置信度的提升，MPIW、CWC 都会升高，MPIW 的增长一般会提高 PICP，而 Wild Bootstrap 法的 PICP 增长跟不上 MPIW 增长，导致其 CWC 明显增大，且大于 MPIW 值，可认为该方法构建的 JS 预测区间可靠性上不足，不予考虑。

其余 3 种 Bootstrap 方法的 PICP 均大于对应置信度，即所构预测区间均能满足期望覆盖，除了考虑预测区间的可靠度外，还需结合区间的清晰度分析，也就是在 PICP 最大情况下 MPIW 尽可能越小。在 99% 置信度下，对于 JD 和 JS 的训练集、测试集，这 3 种方法的 PICP 均取得最大值 100%，其中 Residual Bootstrap 法的 MPIW 最小，JD 分别为 0.646 2%、0.728 0%，JS 分别为 4.241 5%、4.339 9%。在更低置信度下，MPIW 会取得更小的结果，可以看到在 90% 置信度下，Block Bootstrap 法的 PICP 在 JD 训练集和测试集上都取得了最大 100%，MPIW 也比 Residual Bootstrap 法在 99% 置信度下的结果更小，但其 PICP 在 JS 训练集和测试集上并未达到 100%。综合考虑后，选择 Residual Bootstrap 法作为 Bootstrap 重采样的方法，并将其在 99% 置信度下取得的结果作为发酵过程 JD 和 JS 区间预测的最终结果。

表 3 不同置信度下 JD 的区间预测结果

Table 3 The interval prediction results of JD under different confidence levels										
数据集	Bootstrap 方法	90% 置信度			95% 置信度			99% 置信度		
		PICP	MPIW	CWC	PICP	MPIW	CWC	PICP	MPIW	CWC
训练集	Residual Bootstrap	97.5	0.412 6	0.412 6	100	0.491 6	0.491 6	100	0.646 2	0.646 2
	Pairs Bootstrap	99.5	0.501 6	0.501 6	99.5	0.597 3	0.597 3	100	0.785 6	0.785 6
	Block Bootstrap	100	0.566 2	0.566 2	100	0.674 7	0.674 7	100	0.886 7	0.886 7
	Wild Bootstrap	98	0.429 3	0.429 3	99.5	0.511 6	0.511 6	100	0.672 3	0.672 3
测试集	Residual Bootstrap	96	0.464 9	0.464 9	98	0.553 9	0.553 9	100	0.728 0	0.728 0
	Pairs Bootstrap	98	0.557 4	0.557 4	100	0.655 8	0.655 8	100	0.872 9	0.872 9
	Block Bootstrap	100	0.658 7	0.658 7	100	0.784 8	0.784 8	100	1.031 5	1.031 5
	Wild Bootstrap	98	0.472 3	0.472 3	100	0.562 8	0.562 8	100	0.739 7	0.739 7

表 4 不同置信度下JS的区间预测结果

Table 4 The interval prediction results of JS under different confidence levels										
数据集	Bootstrap 方法	90% 置信度			95% 置信度			99% 置信度		
		PICP	MPIW	CWC	PICP	MPIW	CWC	PICP	MPIW	CWC
训练集	Residual Bootstrap	98	2.765 6	2.765 6	99.5	3.295 2	3.295 2	100	4.241 5	4.241 5
	Pairs Bootstrap	100	3.884 1	3.884 1	100	4.627 9	4.627 9	100	6.082 4	6.082 4
	Block Bootstrap	99.5	3.511 6	3.511 6	100	4.184 0	4.184 0	100	5.499 0	5.499 0
	Wild Bootstrap	98	2.782 0	2.782 0	100	3.314 7	3.314 7	100	4.356 5	4.356 5
测试集	Residual Bootstrap	96	2.924 3	2.924 3	100	3.484 2	3.484 2	100	4.339 9	4.339 9
	Pairs Bootstrap	98	4.298 6	4.298 6	100	5.121 7	5.121 7	100	6.731 4	6.731 4
	Block Bootstrap	94	3.728 2	3.728 2	98	4.442 1	4.442 1	100	5.838 1	5.838 1
	Wild Bootstrap	84	2.846 1	60.012 3	98	3.391 1	3.391 1	98	4.456 9	11.805 2

表 5 Bootstrap-GA-ELM模型在JD和JS测试集上的点预测结果

Table 5 The point prediction results of the Bootstrap-GA-ELM model on the JD and JS test sets

模型名称	JD		JS	
	R ²	RMSE	R ²	RMSE
GA-ELM	0.998 8	0.111 1	0.912 9	0.838 6
Bootstrap-GA-ELM	0.998 9	0.109 1	0.920 3	0.802 1

最终确定的 Bootstrap-GA-ELM 模型对 JD 和 JS 点预测和区间预测结果如表 5、图 8 所示，显而易见，无论是对于 JD 还是 JS，采用 Bootstrap 估计的点预测结果精度均高于 GA-ELM 模型， R^2 为 0.998 9、0.920 3，RMSE 为 0.109 1、0.802 1，这说明本文提出的方法在重采样过程中并未出现信息丢失的情况，且能更准确地预测发酵过程 JD 和 JS 含量。从图 8 可以看出，Bootstrap-GA-ELM 模型构建的预测区间能够将 JD 和 JS 实测值曲线完全包含在内，区间覆盖率达到 100%，区间宽度也较窄，特别是 JD 的预测区间平均宽度只有 0.728 0%。因此，本文提出的 Bootstrap-GA-ELM 区间预测模型，既可为发酵过程 JD 和 JS 提供一个较为准确的含量预测值，又能提供一个可靠性和清晰度高的预测区间，可以有效评估 JD 和 JS 可能会达到的范围，量化预测结果的可靠度，区间宽度越大，可靠度越低。特别地，预测区间宽度的变化也能提供一定信息，如果过程中出现样本点的区间宽度急剧增大，表明预测结果遭受较大影响导致不确定程度突变，结果可信度低，这时应该采取重新获取输入参数样本数据、调整模型参数、使用其他模型对比分析等措施，控制不确定性因素对区间预测的干扰。

总的来说，本文构建 Bootstrap-GA-ELM 区间预测模型，对酒醅和黄水理化样本的数量和数据分布无特殊要求，对这种小样本数据（200 个样本左右）也适合，只需满足样本数据完整性，以及保证训练集和

测试集数据分布一致性，便可实现对清香型白酒发酵过程酒醅淀粉和水分含量的可靠预测。

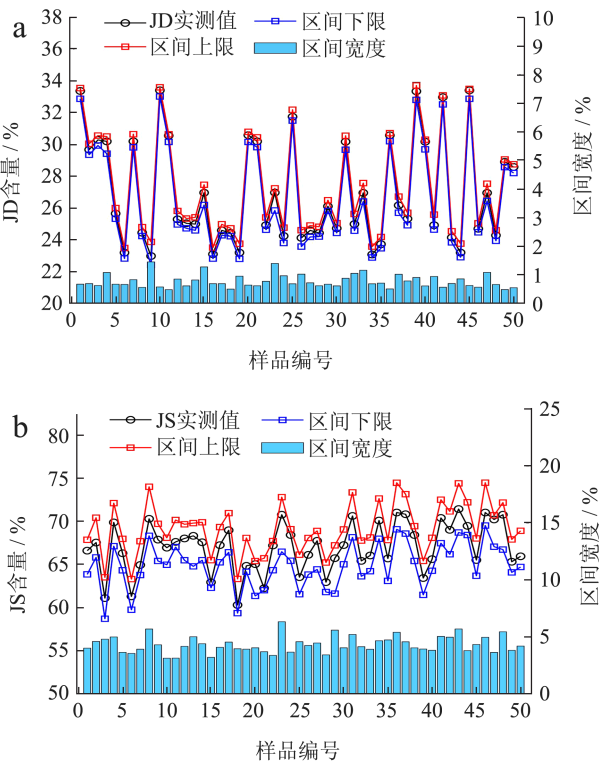


图 8 Bootstrap-GA-ELM 模型在 JD 和 JS 测试集上的区间预测结果

Fig.8 Interval prediction results of Bootstrap GA ELM model on JD and JS test sets

目前，对于清香型白酒发酵过程酒醅淀粉和水分含量的间接预测模型研究相对较少。王昆等^[32]选择发酵过程产生的 O₂、CO₂、湿度和温度作为自变量，基于偏最小二乘回归建立清香型酒醅淀粉和水分含量的软测量模型，模型预测淀粉和水分的 RMSE 分别为 1.93、1.89。在各自淀粉和水分含量量级范围相差不大的情况下，本文提出的区间预测模型具有更低的

RMSE, 淀粉和水分分别为 0.109 1、0.802 1, 且这种传统确定性预测模型, 无法衡量其预测结果的可靠度, 难以获取发酵过程酒醅淀粉和水分含量的不确定信息。

3 结论

针对清香型白酒发酵过程酒醅淀粉和水分含量信息难获取、传统点预测可靠度难量化的双重挑战, 本文以 Bootstrap、GA、ELM 等方法为基础, 构建了 Bootstrap-GA-ELM 区间预测模型, 对发酵过程酒醅淀粉和水分含量进行点预测和区间预测: 通过 Spearman 相关系数和 LARS 综合分析发现, 酒醅淀粉和水分与黄水酒精度、pH、还原糖、淀粉以及酒醅温度这些易测潜在解释变量几乎都具有较高的相关性, 且这些潜在解释变量均在酒醅淀粉和水分模型预测能力上发挥作用, 故可将其作为解释变量输入模型, 为发酵过程酒醅淀粉和水分含量的间接预测提供可行性。并采用 GA-ELM 建立酒醅淀粉和水分含量点预测模型, 对比发现在模型预测精度上, ELM 相较于 PLSR、SVM、RF 等线性或机器学习算法具有更好的表现, 以及 GA 在优化 ELM 隐层权值和阈值上, 相较于近些年提出的优化算法效果更好。然后利用 4 种 Bootstrap 方法构建区间预测模型量化点预测结果的可靠度, 从 PICP、MPIW、CWC 三种评价指标综合分析对比各 Bootstrap 方法的区间预测效果, 结果表明在 99% 置信度下, 基于 Residual Bootstrap 方法构建的区间预测模型, 不但具有较好的点预测效果, 其预测区间也表现出不俗的结果, 能够将酒醅淀粉和水分实测值完全包络在内。其中, 在酒醅淀粉和水分测试集上的点预测精度分别为 $R^2=0.998\ 9$ 、 $RMSE=0.109\ 1$, $R^2=0.920\ 3$ 、 $RMSE=0.802\ 1$, 预测区间的 PICP 和 MPIW 分别为 100%、0.728 0%, 100%、4.339 9%。本文提出的 Bootstrap-GA-ELM 区间预测模型, 将难以直接获取的酒醅理化信息转化为可量化可靠度的预测结果, 可为清香型白酒发酵过程酒醅淀粉和水分含量信息监测提供参考手段, 为发酵过程透明化提供新思路。

参考文献

- [1] 邓杰,张曼,周杰,等.2种浓香型白酒大曲中微生物群落结构及功能研究[J].食品科学技术学报,2024,42(4):75-85.
- [2] 刘瑞华,刘勇,杨甲平,等.白酒酒醅水分检测技术研究和应用综述[J].酿酒,2024,51(2):40-46.
- [3] 母雯竹,张贵宇,张维,等.基于CARS-SPA特征提取的黄水淀粉近红外光谱定量模型优化[J].食品科学,2024,45(19):8-14.
- [4] 韩韶.清香型酒醅发酵过程理化性质变化规律研究[J].中国食品工业,2024,10:104-106.
- [5] HU X L, TIAN R J, WANG K L, et al. The prokaryotic community, physicochemical properties and flavors dynamics and their correlations in fermented grains for Chinese strong-flavor Baijiu production [J]. Food Research International, 2021, 148: 110626.
- [6] 李慧娟,金元浩,李群庆,等.白酒发酵过程代谢气体与原酒品质的相关性分析[J].食品科技,2022,47(4):105-110.
- [7] GUO Q Y, ZHAO J J, PENG J B, et al. Revealing the flavor compositions, microbial diversity, and biological functions of Huangshui from different production workshops [J]. LWT, 2024, 207: 116683.
- [8] 张榆俊,张宿义,张楷正,等.浓香型白酒发酵过程中黄水与酒醅的相关性[J].中国酿造,2023,42(5):35-40.
- [9] 车路萍.浓香型白酒发酵过程中酒醅与黄水相关性研究[D].自贡:四川轻化工大学,2020.
- [10] 李凯,任炳昱,关涛,等.帷幕灌浆量区间预测的Bootstrap-IGWO-SVM模型研究[J].水力发电学报,2022,41(10):18-29.
- [11] ABBAS K, SAEID N, DOUG C, et al. Comprehensive review of neural network-based prediction intervals and new advances [J]. IEEE Transactions on Neural Networks, 2011, 22(9): 1341-1356.
- [12] YE T R, GUI W H. Estimation and Bayesian prediction of the generalized Pareto distribution in the context of a progressive Type-II censoring scheme [J]. Applied Sciences, 2024, 14(18): 8433.
- [13] ZHANG Y, ZHANG W W, LI Y L, et al. AF-OS-ELM-MVE: A new online sequential extreme learning machine of dam safety monitoring model for structure deformation estimation [J]. Advanced Engineering Informatics, 2024, 60: 102345.
- [14] WANG J J, HE M L, JIANG W Y. Causal carbon price interval prediction using lower upper bound estimation combined with asymmetric multi-objective evolutionary algorithm and long short-term memory [J]. Expert Systems with Applications, 2024, 236: 121286.
- [15] REN Q B, LI M C, KONG R, et al. A hybrid approach for interval prediction of concrete dam displacements under uncertain conditions [J]. Engineering with Computers, 2021, 39(2): 1-19.
- [16] SONG L Y, MINKU L L, YAO X. Software effort interval prediction via Bayesian inference and synthetic Bootstrap resampling [J]. ACM Transactions on Software Engineering and Methodology (TOSEM), 2019, 28(1): 1-46.
- [17] ZENG W J, LI X Y, GUO C, et al. Development and uncertainty analysis of the dynamic simulation for HFETR with BMUS framework [J]. Reliability Engineering and System Safety, 2025, 256: 110715.
- [18] MA J W, TANG H M, LIU X, et al. Probabilistic forecasting of landslide displacement accounting for epistemic uncertainty: a case study in the three gorges reservoir area, China [J]. Landslides, 2018, 15(6): 1145-1153.
- [19] AFSHARI-IGDER M, NIKNAM T, KHOOBAN M. Probabilistic wind power forecasting using a novel hybrid intelligent method [J]. Neural Computing and Applications, 2018, 30(2): 473-485.
- [20] LOIZIDIS S, KYPRIANOU A, GEORGHIOU E G. Electricity market price forecasting using ELM and Bootstrap analysis: A case study of the German and Finnish Day-Ahead markets [J]. Applied Energy, 2024, 363: 123058.

[21] LI D, REN X F. Carbon price prediction based on LsOALEO feature selection and time-delay least angle regression [J]. Journal of Cleaner Production, 2023, 416: 137853.

[22] 董泽,贾昊,姜炜.基于特征选择的数据驱动软测量方法[J].热力发电,2019,48(9):83-89.

[23] HUANG G B, ZHOU H M, DING X J, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society, 2012, 42(2): 513-529.

[24] 林平.基于Bootstrap和遗传算法优化LSSVM的滑坡位移区间预测研究[D].重庆:重庆交通大学,2023.

[25] 闫长斌,汪鹤健,周建军,等.基于Bootstrap-SVR-ANN算法的TBM施工速度预测[J].岩土工程学报,2021,43(6):1078-1087.

[26] 贾丽艳,郭晋田,刘帅,等.清香型白酒发酵过程中微生物及理化指标的变化规律[J].中国食品学报,2020,20(8):162-167.

[27] 杨帆,周兴鹏,张敬铨,等.混蒸混烧生产清香型白酒工艺[J].食品研究与开发,2018,39(9):76-81.

[28] KANG J M, HU Y N, JIA L L, et al. Response of microbial community assembly and succession pattern to abiotic factors during the second round of light-flavor Baijiu fermentation [J]. Food Research International, 2022, 162(PA): 111915.

[29] 罗琪,庾先国,张贵宇,等.基于化学计量学方法的黄水还原糖预测模型研究[J].食品安全质量检测学报,2022,13(18):6026-6031.

[30] CHAI L J, QIAN W, ZHONG X Z, et al. Mining the evolutionary patterns of pit mud microbiome and driving factors under the impact of long-term strong-flavor baijiu production [J]. Applied and Environmental Microbiology, 2021, 87(17): AEM0088521.

[31] 彭杰,张宿义,敖宗华,等.浓香型白酒发酵过程中黄水理化指标与微生物群落的相关性研究[J].酿酒科技,2024,6:65-74.

[32] 王昆,庾先国,张贵宇,等.清香型酒醅水分、酸度和淀粉含量的在线预测[J].食品与发酵工业,2022,48(7):85-90,96.