

利用麻雀算法优化随机森林提高持液率预测精度

王亚宁^{1*}, 许梅¹, 王珍², 张永才¹, 王鑫¹, 于庆印¹, 吝少康¹, 陈丹琳¹

¹重庆科技大学石油与天然气工程学院, 重庆

²中国石油化工股份有限公司江汉油田分公司, 湖北 潜江

收稿日期: 2025年7月7日; 录用日期: 2025年8月16日; 发布日期: 2025年9月11日

摘要

为提高气液两相流持液率的预测精度, 本文提出一种基于麻雀搜索算法(SSA)优化随机森林(RF)模型的持液率预测方法。通过整合1219组来自不同学者多源实验数据, 构建了包含管径、气相折算速度、液相折算速度、压力、温度及管道倾角等多维特征的高质量持液率数据库。采用 3σ 法则进行异常值处理, 确保模型训练数据的稳健性。利用SSA算法自动优化RF模型的关键超参数, 避免传统调参的主观性和局限性, 提升模型拟合能力与泛化性能。实验结果表明, SSA-RF模型在训练集与测试集上分别实现了 R^2 为0.935和0.913, RPD值分别为3.81和3.18。研究成果可为复杂气液两相流工况下持液率的高精度预测及工程优化设计提供重要技术支持。

关键词

机器学习, 气液两相流, 管道, 优化算法, 持液率

Using Sparrow Algorithm to Optimize Random Forest to Improve the Accuracy of Liquid Holdup Prediction and Identification

Yaning Wang^{1*}, Mei Xu¹, Zhen Wang², Yongcai Zhang¹, Xin Wang¹, Qingyin Yu¹, Shaokang Lin¹, Danlin Chen¹

¹School of Petroleum Engineering, Chongqing University of Science & Technology, Chongqing

²Sinopec Jiangnan Oilfield, Qianjiang Hubei

Received: Jul. 7th, 2025; accepted: Aug. 16th, 2025; published: Sep. 11th, 2025

*第一作者。

文章引用: 王亚宁, 许梅, 王珍, 张永才, 王鑫, 于庆印, 吝少康, 陈丹琳. 利用麻雀算法优化随机森林提高持液率预测精度[J]. 矿山工程, 2025, 13(5): 981-991. DOI: 10.12677/me.2025.135112

Abstract

To improve the prediction accuracy of the liquid holding rate of gas-liquid two-phase flow, this paper proposes a liquid holding rate prediction method based on the Sparrow Search algorithm (SSA) to optimize the random forest (RF) model. By integrating 1219 sets of multi-source experimental data from different scholars, a high-quality liquid holdup database was constructed, which includes multi-dimensional features such as pipe diameter, gas phase conversion rate, liquid phase conversion rate, pressure, temperature and pipe inclination Angle. The 3σ rule is adopted for outlier handling to ensure the robustness of the model training data. The SSA algorithm is utilized to automatically optimize the key hyperparameters of the RF model, avoiding the subjectivity and limitations of traditional parameter adjustment, and enhancing the model's fitting ability and generalization performance. The experimental results show that the SSA-RF model achieves R^2 of 0.935 and 0.913 respectively on the training set and the test set, and the RPD values are 3.81 and 3.18 respectively. The research results can provide important technical support for the high-precision prediction of liquid holdup and engineering optimization design under complex gas-liquid two-phase flow conditions.

Keywords

Machine Learning, Gas-Liquid Two-Phase Flow, Pipeline, Optimization Algorithm, Liquid Holdup

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在石油和化工领域,液体在气井或油气集输管道中的滞留显著阻碍了气体的流动,导致井底压力升高和产量下降,严重时甚至造成停喷,还会加剧管道腐蚀和结垢。增加运行能耗与维护成本。而持液率是描述液体在管道内所占体积分布的重要参数,直接影响流动特性、压降、传热传质效率及工程设备设计等方面。持液率定义为单位长度管道中被液体所占的体积分数,反映了液体在气液两相共存流动中的空间分布情况。在油气井生产、油气集输等工程中,准确掌握和预测持液率对合理调控流动状态、优化生产工艺具有重要意义。由于气液界面复杂,流型多样,持液率通常依赖实验测量、经验模型估算,而实验测量较为繁琐,经验模型计算存在误差。

目前,已有大量学者进行了油气井或管道中持液率与气液两相流动测量相关实验,Kokal [1]通过在直径 50.8 mm、长度 21.3 m 的有机玻璃试验段中注入空气-水和空气-甘油水混合物等介质,在倾角范围为 -5° 至 $+5^\circ$ 条件下测量了不同气液速比下的持液率与压降变化。Mandhane [2]开展了空气-水两相在水平管道中流动实验,研究不同气液表观速度组合下的流型变化规律。识别并归类了泡状流、弹状流、波状流、分层流等典型流型,并根据实验结果绘制了详细的流型图谱。Hughmark [3]在直径 38.1 mm、长 5.49 m 的垂直管道内进行了系统的气液两相流实验,系统测量了不同气体表观速度、液体表观速度范围内的持液率与压降。基于实验数据建立了经验关联式并修正了 Zuber-Findlay 等经典模型。Minami [4]基于直径 101.6 mm 的水平和微倾斜管道,考虑三种实验介质(煤油/空气、水/空气、水-表面活性剂/空气混合物),在气速 10~40 m/s、极低液率(液气比 < 0.01)的工况下,采用 Gamma 射线与电容探测测量了瞬时与时间平均持液率分布。Kokal [5]在一段直径 38.1 mm、长度 18 米的可调角度实验管道中,系统研究了空

气-水两相流在微倾斜(-2° 至 $+5^{\circ}$)条件下的流动特性。通过控制气速与液速范围,测量了各工况下的持液率与压降,并识别出不同流型的演化路径。Beggs [6]在直径 38.1 mm、长度 18.3 m 的透明管段中,观察了空气-水两相在微倾斜角($\pm 5^{\circ}$)下的流动特性。通过调节气速与液速,测得了各流型下的液体持液率与压降数据。最终建立了以气液速率与倾角为变量的经验模型。Mukherjee [7]直径 50.8 mm、长度 9.1 m 的倾斜管道中开展空气-水两相流实验,测量了倾角从 -10° 到 $+10^{\circ}$ 条件下的流型演化、液体持液率和沿程压降特性。结果表明,在上倾时更容易形成液体堆积。Abdul-Majeed [8]基于空气-煤油两相混合物在水平管道中的流动,采用直径 50.8 mm、长 36 m 的水平塑料管道,测量了压力、温度和流速等参数,并通过快速关闭阀门法获取了 89 组持液率数据(范围 0.009~0.61),并修正了 Taitel-Dukler 模型在分层平滑流中持液率计算较低的缺点。Ajani [9]基于空气-水-表面活性剂混合体系,考虑表面活性剂类型(阴离子、两性、磺酸盐、阳离子)、浓度(200~3000 ppm)、管径(50.8 mm 和 101.6 mm)、倾角(90° 垂直管)等因素,通过小规模实验(表面张力、稳定性和液体卸载测试)和大规模实验(垂直管中的持液率、压降、泡沫持率及气相截留分数测量),基于泡沫多相流实验数据建立了持泡率模型。Kong W [10]考虑了流体流动特性、光纤探针阵列的有效测量区域、重复干扰区域以及流体扰动等因素。通过气液两相流实验平台,验证了该传感器在分层流和波状流条件下测量持液率的准确性和可靠性。Alsarkhi [11]综合分析了 1340 个实验数据点,涵盖不同管径(12.7~156 mm)、流体性质(密度 759~1000 kg/m³、黏度 1~962 cP)、倾角(-9° 至 90°)和表面张力(0.033~0.074 N/m)。揭示了现有 HLS 关联式的存在缺陷,并提出以 V_{SL}/V_{SG} 为关键参数的新模型。Al-Safran E [12]基于自主搭建的 50.8 mm 内径水平管道实验系统对高粘度液体(180~587 mPa·s)与气体在水平管道中两相流动的实验进行研究,采用 Bent-Hyperbola 非线性回归,拟合得到经验关联式。

除了传统的持液率测量实验与模型建立,许多学者还研究了利用机器学习(ML)和深度学习(DL)方法来预测持液率。这些研究通常基于多传感器数据融合(如电导、电容、超声波、光纤探针等)或流型图像分析,结合监督学习(如支持向量机 SVM、随机森林 RF)和深度学习(如卷积神经网络 CNN、长短期记忆网络 LSTM)进行建模。Zhang X [13]收集 1390 组公开实验数据,采用 PGNN 模型来预测气液两相流中的持液率和压降,PGNN 的持液率预测 MAPE (16.99%)显著低于其他模型。He H [14]提出一种结合 CNN、LSTM 和注意力机制的深度学习模型(CNN-LSTM-Attention),采用 1800 组实验数据,覆盖管道直径(11~152 mm)、倾角(-90° 至 90°)、油/水表观速度(0.01~7.58 m/s)及油相黏度(1.88~228 mPa·s)等参数,构建了包含 10 类流动模式(如分层流 ST、分散流 D o/w)和持油率预测的训练集。通过特征相关性分析(油相速度影响最大)和模型优化,该模型在流动模式分类中准确率达 98.3%,持油率预测误差 < 15%。Azizi [15]从文献获取 468 组实验数据,涵盖不同倾角、油水表观速度及流动模式,采用人工神经网络(ANN)模型,预测精度 R^2 为 99.8,但仅考虑了油水表观速度和倾角。Xu L [16]采用一种双圆形电导探针阵列获取实验数据作者在工业级实验装置(16 m 长、125 mm 内径水平管道)中采集了 220 种流量条件下的探针响应数据,建立了非线性支持向量回归(SVR)模型和线性最小二乘回归(LSR)模型。非线性 SVR 模型表现最优,其拟合优度(R^2 均值 0.9909)显著优于线性 LSR 模型(R^2 均值 0.9686)。El-Sebakhy E A [17]通过模糊规则生成和最小二乘参数估计,实现了流型识别(分类准确率 98.85%)和持液率预测(平均绝对误差 7.346%)。对比实验表明,该 ANFIS-TSK 模型在 RMSE (3.183)和相关系数(0.9929)上均优于传统神经网络和 12 种经验关联式。

目前持液率研究主要依赖实验测量、经验模型估算以及机器学习预测,但存在以下局限性:

- (1) 传统实验方法如快速关闭阀门法、Gamma 射线测量等虽然精度较高,但操作繁琐、成本高昂,且难以实现实时监测。
- (2) 基于实验数据建立的 Beggs-Brill、Eaton、Mukherjee 等经验关联式在特定工况下表现良好,但普遍存在外推性差、适用范围窄的问题。

(3) 近年来兴起的机器学习方法(如 ANN、SVM、CNN-LSTM 等)虽显示出强大潜力,但仍存在现有研究多关注单一流型或有限工况,缺乏通用性强的预测框架。

针对上述问题,本文提出了一种基于 SSA-RF (麻雀算法优化随机森林)的持液率预测模型采用麻雀优化算法(SSA)自动优化 RF 的关键参数,避免人工调参的主观性,构建多尺度特征选择机制,综合考虑流体物性、管道几何等多维度参数,引入识别模块,实现不同流型下的持液率预测。

2. 持液率预测模型建立

2.1. 数据库预处理

2.1.1. 数据库建立

在持液率数据库的建立过程中,为了支撑具有广泛适用性和高精度的持液率预测模型开发,本文整合了 1219 组来自国内外不同学者的气水、气油两相流持液率实验数据。数据集系统涵盖了管径、气相折算速度、液相折算速度、压力、温度、管道倾角、持液率等关键影响参数,构成了一个多维度、跨工况条件的数据集。其中数据来源包括 Mukherjee、Minami、Beerens、Abdul-Majeed 等多项经典两相流实验文献,实验条件覆盖了多种管径(38.1~77.93 mm)、气液速度范围(气相 0.004~275.1 m/s,液相 0.0024~50.29 m/s)、倾角区间(0°~90°)、压力范围(142.72~919.1 kPa)及温度区间(4.44~74.16°C) (表 1)。

Table 1. Experimental range of gas-liquid two-phase flow holdup by different scholars

表 1. 不同学者气液两相流持液率实验范围

文献来源	介质	液相速度 /(m·s ⁻¹)	气相速度 /(m·s ⁻¹)	倾角/(°)	管径/(mm)	压力/(kPa)	温度/(°C)	实验 组数
Mukherjee [7]	空气 - 煤油	0.0024~50.29	0.015~41.45	0°、30°、45°、50°、 60°、70°、80°、90°	38.1	142.72~646	4.44~74.16	591
Minami [4]	空气 - 煤油 空气 - 水	0.005~0.95	0.47~16.59	0°	77.93	301.3~641.9	24.44~47.22	111
Beggs [6]	空气 - 水	0.052~5.54	4.64~275.1	0°、5°、15°、20°、 35°、50°、55°、 75°、85°、90°	38.1	444.64~682.37	18.05~37.2	461
Abdul-Majeed [8]	空气 - 水	0.315~45.87	0.004~1.53	0°	50.8	216.5~919.1	27.3~48.3	56
参数范围	空气 - 煤油 空气 - 水	0.0024~50.29	0.004~275.1	0°~90°	38.1~77.9	142.72~919.1	4.44~74.16	1219

2.1.2. 数据预处理

在机器学习建模过程中,为提升模型训练的准确性与稳健性,本文对持液率数据库中的原始数据进行了系统的数据清洗处理。主要采用 3σ 法则(3-Sigma)进行异常值识别[18],即基于各特征数据的均值和标准差,判别落在 $\mu \pm 3\sigma$ 之外的数据为异常值。对于检测出的异常值,本文采取替换为该特征均值的方式进行修正,以减少极端值对模型学习过程的干扰,保证数据整体分布的合理性和模型训练的稳定性。通过该方法,确保数据集在清洗后具备良好的统计特性,为后续持液率预测模型的训练提供了高质量、无严重异常值污染的数据基础。

1) 计算数据的均值和标准差

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (2)$$

2) 计算数据的均值和标准差

上限: $\mu + 3\sigma$

下限: $\mu - 3\sigma$

数据点落在 $[\mu - 3\sigma, \mu + 3\sigma]$ 区间之外的即为异常值。

2.1.3. 相关性分析

如图 1, 相关性分析结果显示, 持液率与各输入特征之间呈现出不同程度的相关性特征。其中, 液相速度与持液率呈现负相关(相关系数-0.41), 表明随着液相速度增加, 管道内液体滞留明显, 整体持液率增加; 气相折算速度与持液率呈现低至中等正相关(0.31), 说明不同气体流速下, 持液率变化较大。管径与持液率存在弱负相关(-0.18), 不同管径下持液率有略微变化。其他参数如温度(-0.1)、压力(0.0075)、倾角(0.084)与持液率相关性均较弱, 影响相对有限。整体来看, 气相速度、液相速度是影响持液率的主要特征变量。

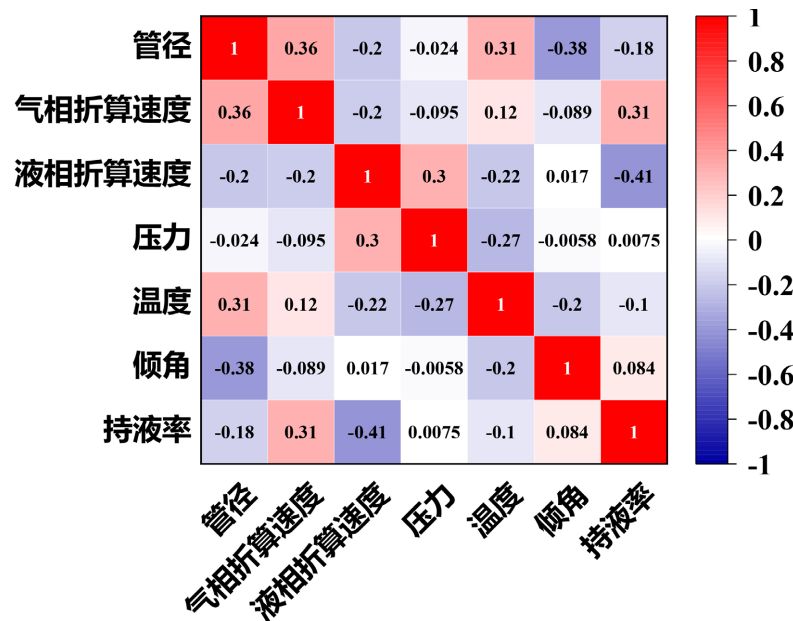


Figure 1. Correlation analysis of different features

图 1. 不同特征相关性分析

2.2. 模型评估指标

在评估机器学习模型训练精度和性能时, 这些评价指标各自反映了模型不同的误差或拟合能力。均方根误差 RMSE 与均方误差 MSE 用来确定模型预测性能, 平均绝对误差 MAE 来确定模型可靠性, 决定系数 R^2 评估模型的拟合效果与泛化能力。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_j| \quad (3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_j - y_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_j)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_j)^2}{\sum_{i=1}^n (y_i - y_m)^2} \tag{6}$$

$$RPD = \frac{y_i - y_j}{|S|} \tag{7}$$

式中： n 是样本数量； y_i 是第 i 个样本的真实值； y_j 第 i 个样本的预测值； S 为残差的标准差； y_m 是样本平均值。

2.3. 模型原理

麻雀搜索算法(SSA)是一种新兴的群体智能优化算法，由薛建凯等人于 2020 年提出[19]，灵感来源于麻雀的觅食和警戒行为。SSA 具有简单易实现、全局搜索能力强、收敛速度快等优点，广泛用于解决各类全局优化问题[20] (图 2)。



Figure 2. Sparrow algorithm flow chart
图 2. 麻雀算法流程图

麻雀种群分为两类角色：
发现者(Discoverer)：主导搜索过程，负责寻找食物，具备更高的适应度；
加入者(Joiner)：跟随发现者，在局部区域内进行搜索；
警戒机制：当发现环境中存在捕食者(高风险区域)时，麻雀群体会调整搜索策略以避险，提高算法跳出局部最优的能力。

(1) 初始化种群

随机生成一组候选解 $X = \{x_1, x_2, \dots, x_n\}$ ，每个个体代表一组优化变量，并计算初始种群适应度值；

(2) 位置更新

发现者负责全局搜索其位置更新如下所示：

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot e^{-\frac{i}{\alpha \cdot T_{\max}}}, & R < ST \\ x_{i,j}^t + Q \cdot L, & R \geq ST \end{cases} \quad (8)$$

式中: $x_{i,j}^{t+1}$ 是第 $t+1$ 代第 i 个麻雀在第 j 维上的位置; α 是常数, 通常设为 1; $T_{\max}$ 是最大迭代次数; $R \sim U(0, 1)$ 是随机数; ST 属于 $[0.5, 1)$ 之间, 为环境安全值。 Q 是正态分布随机数; L 为 1 向量;

加入者主要执行局部搜索, 更新公式:

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot e^{\frac{x_{\text{worst},j} - x_{i,j}^t}{i^2}}, & f(x_i) > f(x_g) \\ x_{i,j}^t + |x_{i,j}^t - x_g| \cdot K, & f(x_i) \leq f(x_g) \end{cases} \quad (9)$$

式中: $x_{\text{worst},j}$ 是当前种群中第 j 维最差个体的位置; x_g 是当前全局最优位置; K 服从 $U(-1, 1)$ 的随机数; $f(x_i)$ 是第 i 个麻雀对应的适应度值; $f(x_g)$ 是当前全局最优个体适应度值。

警戒者调整机制, 为增强算法跳出局部最优能力, 采用如下更新策略:

$$x_{i,j}^{t+1} = \begin{cases} x_g + \beta \cdot |x_{i,j}^t - x_g| & f(x_i) \neq f(x_g) \\ x_g + K \cdot \frac{x_{i,j}^t - x_{\text{worst},j}}{|f(x_i) - f_{\text{worst}}| - \varepsilon} & f(x_i) = f(x_g) \end{cases} \quad (10)$$

式中: $\beta \sim N(0, 1)$ 符合正太分布; K 为 $[-1, 1]$ 之间的随机数; ε 为一个较小的数字, 防止分母为 0。

重复上述步骤, 直到达到最大迭代次数 $T_{\max}$, 输出全局最优解。

随机森林(RF)是一种基于 Bagging 策略的集成学习方法, 由 Leo Breiman 和 Adele Cutler 在 2001 年提出[21], 由多棵回归树组成, 通过对不同子样本集训练多棵树, 再将各树预测结果进行平均, 提升模型的泛化能力和预测精度[22]-[24]。

(1) 对原始训练集进行放回采样, 生成 T 个子训练集 D_t 。

(2) 对每个子训练集 D_t , 训练一颗决策树, 在每个树节点划分时, 随机选取部分特征进行最优划分。

对于回归树, 最常用划分指标为均方误差如公式(4)。

(3) 训练完成后, 形成一个由 T 棵树组成的森林 $\{h_1(x), h_2(x), h_3(x), \dots, h_T(x)\}$

(4) 通过投票分类和回归任务确定最终结果。

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_T(x)) \quad (11)$$

$$y_n = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (12)$$

式中: $h_t(x)$ 是第 t 颗决策书对输入 x 预测值; T 是森林中决策树的数量; y_n 是最终预测值。

(5) 特征重要性评估

常用指标是平均不纯度减少(Mean Decrease Impurity, MDI), 例如基于 Gini 指数或 MSE 的减少量。

$$\text{Importance}(j) = \frac{1}{T} \sum_{t=1}^T \sum_{n \in \text{nodes}(t,j)} p(n) \cdot \Delta I(n) \quad (13)$$

式中: $\Delta I(n)$ 是节点 n 分裂后对应指标的减少值; $p(n)$ 是节点样本占比; $\text{node}(t, j)$ 是第 t 颗树种使用了特征 j 的所有节点集合。

本研究采用的数据集系统涵盖影响持液率预测的多项关键参数, 包括管径、气相折算速度、液相折算速度、压力、温度及管道倾角, 构建基于麻雀搜索算法-随机森林(SSA-RF)的组合预测模型。首先,

按照 80%和 20%的比例将数据集划分为训练集与测试集,训练集用于模型训练及参数优化,测试集用于模型性能评估。随后,采用麻雀搜索算法(SSA)对随机森林(RF)模型的超参数进行全局优化,利用训练集数据在每次迭代中通过适应度函数(以回归误差指标为评价标准)指导群体搜索,迭代更新最优参数组合。最终,基于优化后的 RF 模型对测试集数据进行持液率预测,验证模型的预测性能与泛化能力(图 3)。

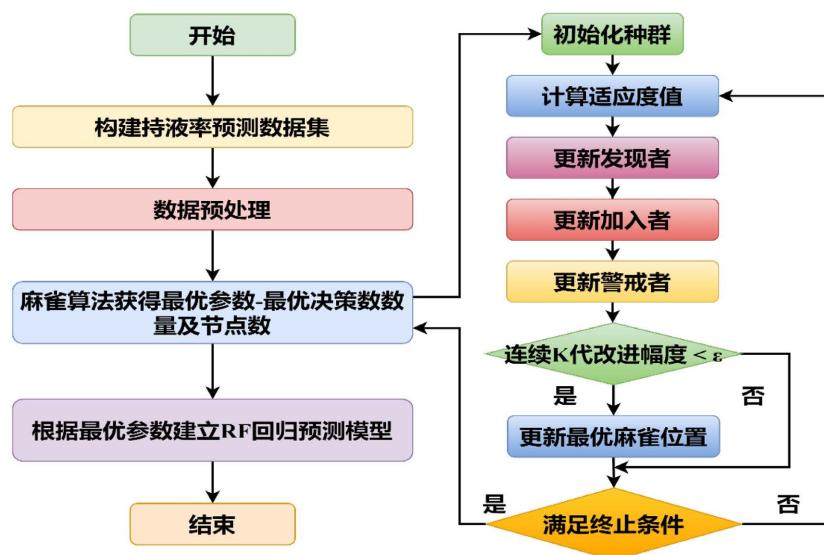


Figure 3. Training process
图 3. 训练流程

3. 结果与讨论

模型预测

将数据清洗后的 1219 条持液率样本数据作为本模型的输入集,以管径、气相折算速度,液相折算速度,压力,温度,倾角作为本模型的训练特征,持液率作为本模型的输出特征。

首先按照训练集与测试集比例为 7:3 进行划分,其中 852 组训练集,366 组测试集。将训练集与测试集分别归一化处理,确保所有特征范围在 0~1 之间,其次采用 SSA 算法优化 RF 随机森林的最优决策树数量以及最优决策树节点数两个参数。然后将优化后的参数输出至 RF 模型中。设定初始种群个数为 15,最大迭代次数 50 次进行模型训练,最终优化参数最优决策树数量为 37、最优决策树节点数为 1。当满足适应度值误差小于设定值时或达到最大迭代次数终止计算。

图 4 所示为 SSA-RF 模型在训练集与测试集上的拟合结果。训练集拟合 R^2 达到 0.93513,模型拟合趋势与实际数据高度一致,残差较小,未出现明显过拟合现象。测试集拟合 R^2 同样达到 0.913721,残差水平与训练集保持一致,表明 SSA-RF 模型具有优良的泛化能力与稳健性。在不同气液两相流动工况下,模型均能保持较高预测精度,验证了麻雀优化算法在 RF 参数调优中的有效性。

图 5 展示了 SSA-RF 模型优化过程中适应度值随迭代轮次的变化趋势。可以看出,模型在优化初期,适应度值继续缓慢下降,SSA 开始在局部优区进行精细搜索。最终在第 5 次迭代后适应度值趋于稳定,表明算法成功收敛至最优解且优化过程无明显震荡,收敛性良好、稳定性高。

表 2 给出了 SSA-RF 模型在训练集与测试集上的主要评价指标。从表中可以看出,训练集和测试集上的 MAE 分别为 0.0466 和 0.0572, RMSE 分别为 0.0642 和 0.0716,表明模型整体预测误差较低,具备良好的拟合能力和泛化性能。RPD 值分别为 3.81 和 3.18,均大于 3,说明模型在训练集和测试集上均具

有较高的预测精度，属于优秀预测模型范畴。决定系数 R^2 在训练集和测试集上分别达到 0.935 和 0.913，拟合优度较高，进一步证明模型能够有效捕捉复杂输入特征与持液率之间的关系。整体来看，SSA-RF 模型在持液率预测任务中表现出良好的准确性、稳健性与泛化能力，具有较好的工程应用前景。

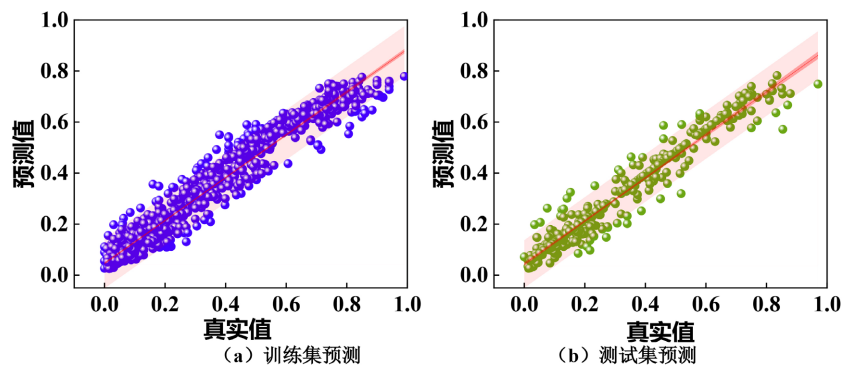


Figure 4. Prediction fitting diagram
图 4. 预测拟合图

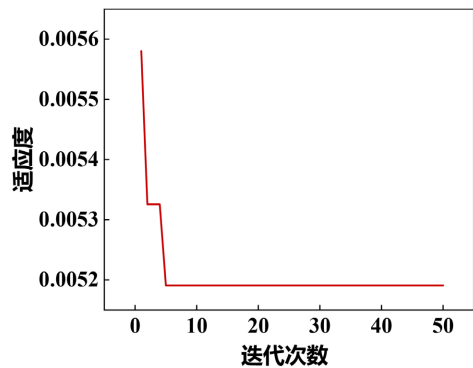


Figure 5. Optimal fitness
图 5. 最佳适应度

Table 2. Evaluation indicators of the model
表 2. 模型的评价指标

评价指标	MAE	MSE	RMSE	RPD	R^2
训练集	0.46557	0.00432	0.064187	3.8066	0.93513
测试集	0.057167	0.00618	0.071585	3.1752	0.91321

4. 结论

- (1) 本文针对复杂气液两相流持液率预测问题，提出了一种基于麻雀搜索算法优化随机森林(SSA-RF)的持液率预测模型。研究构建包含 1219 组多维持液率数据的高质量数据库，并通过 3σ 法则优化数据质量，保障模型训练的稳定性。
- (2) 实验结果表明，SSA-RF 模型在训练集和测试集上分别取得了 $R^2 = 0.935$ 、 $R^2 = 0.913$ ，预测误差低，稳定性强，优于传统机器学习和深度学习模型。

基金项目

重庆科技大学研究生创新计划项目“泡沫排水采气中泡沫携液机理及模型建立研究”(编号：

YKJCX2420144)。

参考文献

- [1] Kokal, S.L. and Stanislav, J.F. (1989) An Experimental Study of Two-Phase Flow in Slightly Inclined Pipes—II. Liquid Holdup and Pressure Drop. *Chemical Engineering Science*, **44**, 681-693. [https://doi.org/10.1016/0009-2509\(89\)85043-2](https://doi.org/10.1016/0009-2509(89)85043-2)
- [2] Mandhane, J.M., Gregory, G.A. and Aziz, K. (1974) A Flow Pattern Map for Gas-Liquid Flow in Horizontal Pipes. *International Journal of Multiphase Flow*, **1**, 537-553. [https://doi.org/10.1016/0301-9322\(74\)90006-8](https://doi.org/10.1016/0301-9322(74)90006-8)
- [3] Hughmark, G.A. and Pressburg, B.S. (1961) Holdup and Pressure Drop with Gas-liquid Flow in a Vertical Pipe. *AIChE Journal*, **7**, 677-682. <https://doi.org/10.1002/aic.690070429>
- [4] Minami, K. and Brill, J.P. (1987) Liquid Holdup in Wet-Gas Pipelines. *SPE Production Engineering*, **2**, 36-44. <https://doi.org/10.2118/14535-pa>
- [5] Kokal, S.L. (1987) An Experimental Study of Two-Phase Flow in Inclined Pipes.
- [6] Beggs, H.D. (1972) An Experimental Study of Two-Phase Flow in Inclined Pipes. The University of Tulsa.
- [7] Mukherjee, H. (1979) An Experimental Study of Inclined Two-Phase Flow. The University of Tulsa.
- [8] Abdul-Majeed, G.H. (1996) Liquid Holdup in Horizontal Two-Phase Gas-Liquid Flow. *Journal of Petroleum Science and Engineering*, **15**, 271-280. [https://doi.org/10.1016/0920-4105\(95\)00069-0](https://doi.org/10.1016/0920-4105(95)00069-0)
- [9] Ajani, A., Kelkar, M., Sarica, C. and Pereyra, E. (2016) Foam Flow in Vertical Gas Wells under Liquid Loading: Critical Velocity and Pressure Drop Prediction. *International Journal of Multiphase Flow*, **87**, 124-135. <https://doi.org/10.1016/j.ijmultiphaseflow.2016.09.001>
- [10] Kong, W., Li, S., Hao, H., Yan, P., Zhuo, C. and Li, H. (2024) Measurement Method of Gas Holdup in Horizontal Gas-Liquid Two-Phase Flow Based on Fiber-Optic Probe Array. *Flow Measurement and Instrumentation*, **97**, Article ID: 102588. <https://doi.org/10.1016/j.flowmeasinst.2024.102588>
- [11] Alsarkhi, A. and Fdleseed, A. (2024) Unmasking the Mystery: The Path to Accurate Liquid Slug Holdup Prediction. *Geoenergy Science and Engineering*, **243**, Article ID: 213331. <https://doi.org/10.1016/j.geoen.2024.213331>
- [12] Al-Safran, E., Kora, C. and Sarica, C. (2015) Prediction of Slug Liquid Holdup in High Viscosity Liquid and Gas Two-Phase Flow in Horizontal Pipes. *Journal of Petroleum Science and Engineering*, **133**, 566-575. <https://doi.org/10.1016/j.petrol.2015.06.032>
- [13] Zhang, X., Hou, L., Zhu, Z., Yang, M., Hu, Z. and Liu, J. (2024) Prediction of Liquid Holdup and Pressure Drop in Gas-Liquid Two-Phase Flow Based on Integrating Mechanism Analysis and Data Mining. *Geoenergy Science and Engineering*, **239**, Article ID: 212971. <https://doi.org/10.1016/j.geoen.2024.212971>
- [14] He, H., Zhou, Z., Sun, B., Li, X. and Sun, X. (2025) Study on the Prediction Method of Oil-Water Two-Phase Flow Pattern and Oil Holdup. *Geoenergy Science and Engineering*, **246**, Article ID: 213627. <https://doi.org/10.1016/j.geoen.2024.213627>
- [15] Azizi, S., Awad, M.M. and Ahmadloo, E. (2016) Prediction of Water Holdup in Vertical and Inclined Oil-Water Two-Phase Flow Using Artificial Neural Network. *International Journal of Multiphase Flow*, **80**, 181-187. <https://doi.org/10.1016/j.ijmultiphaseflow.2015.12.010>
- [16] Xu, L., Zhang, W., Zhao, J., Cao, Z., Xie, R., Liu, X., et al. (2017) Support-Vector-Regression-Based Prediction of Water Holdup in Horizontal Oil-Water Flow by Using a Bicircular Conductance Probe Array. *Flow Measurement and Instrumentation*, **57**, 64-72. <https://doi.org/10.1016/j.flowmeasinst.2017.08.003>
- [17] El-Sebakhy, E.A. (2010) Flow Regimes Identification and Liquid-Holdup Prediction in Horizontal Multiphase Flow Based on Neuro-Fuzzy Inference Systems. *Mathematics and Computers in Simulation*, **80**, 1854-1866. <https://doi.org/10.1016/j.matcom.2010.01.002>
- [18] Wei, X., Huang, W., Liu, L., Wang, J., Cui, Z. and Xue, L. (2024) Low-Rank Coalbed Methane Production Capacity Prediction Method Based on Time-Series Deep Learning. *Energy*, **311**, Article ID: 133247. <https://doi.org/10.1016/j.energy.2024.133247>
- [19] 薛建凯. 一种新型的群智能优化技术的研究与应用——麻雀搜索算法[D]: [硕士学位论文]. 上海: 东华大学, 2020.
- [20] 贾立山, 许海军. 基于改进麻雀算法的机器人路径规划[J]. 信息与控制, 2025(3): 536-544.
- [21] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [22] 江铭, 邹清腾, 肖壮, 等. 浅层页岩气井间压窜影响因素与防治优化[J]. 石油实验地质, 2025, 47(3): 693-704.
- [23] 胡秋嘉, 刘春春, 张建国, 等. 基于机器学习的煤层气井产能预测与压裂参数优化[J]. 油气藏评价与开发, 2025,

15(2): 266-273+299.

- [24] 费世祥, 崔越华, 李小锋, 等. 鄂尔多斯盆地中、东部深层煤岩气水平井高效开发主控因素[J]. 石油与天然气地质, 2025, 46(1): 273-287.