

引文格式:周亚罗,姚飞宇,刘文广,等. 基于 UNGO-CNN-BiLSTM-Self Attention 的烧结混合料含水量预测模型研究[J]. 中国有色冶金, 2026, 55(3): 163-178.

ZHOU Yaluo, YAO feiyu, LIU Wenguang, et al. Research on a sinter mix moisture content prediction model based on UNGO-CNN-BiLSTM-Self Attention[J]. China Nonferrous Metallurgy, 2026, 55(3): 163-178.

基于 UNGO-CNN-BiLSTM-Self Attention 的烧结混合料含水量预测模型研究

周亚罗¹, 姚飞宇¹, 刘文广², 张瑞成¹

(1. 华北理工大学 电气工程学院, 河北 唐山 063210; 2. 首钢京唐钢铁联合有限责任公司, 河北 唐山 063200)

[摘 要] 针对目前烧结机混合料含水量预测精度不足与收敛速度慢等问题, 本文提出一种基于改进北方苍鹰优化算法(UNGO)的 CNN-BiLSTM-Self Attention 混合预测模型。模型首先利用卷积神经网络(CNN)提取输入特征的空间关联信息, 再通过双向长短期记忆网络(BiLSTM)捕捉混合料含水量的时序依赖特征, 并引入自注意力机制(Self Attention)以强化关键特征权重分配和长距离依赖建模, 从而提升预测的稳定性与精度。为降低模型复杂度并提高运行效率, 采用 Spearman 相关系数对输入特征进行筛选, 剔除相关性较低的变量。针对传统优化算法易陷入局部最优的问题, 在北方苍鹰算法基础上引入混沌初始化与非线性正余弦衰减策略, 显著增强了全局寻优与局部搜索能力, 实现了学习率、神经元个数、正则化系数及注意力键值等超参数的自适应优化。实验以河北某大型钢厂烧结生产数据为样本, 结果表明, 在相同数据条件下, CNN-BiLSTM-Self Attention 模型的误差指标 MSE、RMSE、MAE 分别为 0.057 5、0.239 8 和 0.192 3, R^2 达到 0.790 8, 均优于 CNN、LSTM、CNN-LSTM 及 CNN-BiLSTM 等对比模型, 表明 CNN-BiLSTM-Self Attention 模型具有更强的特征表达与时序建模能力。经 UNGO 算法优化后, 相较原模型, MSE、RMSE、MAE 分别降低至 0.022 4、0.149 8 和 0.118 5, R^2 达到 0.918 4, 整体性能优于 NGO、PSO 和 WOA 等算法优化结果, 表现出更快的收敛速度和更高的预测精度。进一步与 UNGO-Transformer、UNGO-BP 及 UNGO-XGBoost 等主流预测模型对比, 所提模型在预测精度、泛化能力和鲁棒性方面均具有更优表现。综上, UNGO-CNN-BiLSTM-Self Attention 模型在结构设计与参数优化层面均展现出良好的综合性能, 可有效实现烧结混合料含水量的高精度预测, 为烧结过程的智能控制与能效优化提供了可靠的技术支撑。

[关键词] 烧结混合料; 含水量预测; CNN-BiLSTM-Self Attention; 改进北方苍鹰算法(UNGO); 自注意力机制(Self Attention); Tent 混沌初始化策略; Spearman 相关系数

[中图分类号] TF046; TP391.9 [文献标识码] A [文章编号] 1672-6103(2026)03-0163-16
DOI: 10.19612/j.cnki.cn11-5066/tf.2026.03.017

0 引言

钢铁工业烧结过程^[1]中, 混合料的水分含量对烧结矿质量和产量具有重要影响。水分过高或过

低均会导致料层透气性变差、成品率下降等问题, 因此适宜的水分控制区间通常为 6.75%~7.5%^[2]。实现水分的准确预测对优化烧结工艺尤为关键。

早期研究多采用统计模型进行水分预测。周林等^[3]建立的一维线性回归模型预测误差为 0.1%; 丁园等^[4]提出的二维区间自回归模型在 MAE 和 RMSE 上优于传统方法, 但对非线性关系处理能力有限。随着机器学习发展, 孙学波等^[5]利用 RBF 神经网络预测烧结终点, 误差控制在 ± 0.15 以内; 杨宁

[收稿日期] 2025-12-30 [修订日期] 2026-02-12

[录用日期] 2026-02-15 [出版日期] 2026-06-01

[第一作者] 周亚罗(1982—), 女, 河北衡水人, 副教授, 主要从事复杂系统的建模与控制方面研究。

[通信作者] 姚飞宇(1998—), 男, 河南许昌人, 研究生, 主要研究方向为控制技术及应用方面研究。

等^[6]采用NARX神经网络预测水分, MSE 为0.052, R^2 达0.96,精度显著提升。为进一步提升预测性能,学者引入智能优化算法以增强模型能力。尹学等^[7]结合GA和PSO优化BP神经网络,使 $RMSE$ 、 MAE 和 MSE 大幅降低;王瑞林等^[8]利用鲸鱼优化算法(WOA)优化BP网络,但水分预测精度仍待提高;王自赞等^[9]采用改进遗传算法(IGA)优化LSTM超参数, MAE 和 MSE 降低超过50%。为克服单向LSTM无法捕捉未来信息的问题,吕世轩等^[10]融合CNN和BiLSTM,显著优化了 $RMSE$ 、 MAE 和 R^2 ;李小根等^[11]在CNN-LSTM中引入注意力机制, R^2 达到0.972,但超参数优化仍依赖传统方法,效率有限。近年来,深度学习方法因其端到端的自动特征提取能力,被广泛应用于烧结过程建模与状态预测中。

针对烧结混合料含水量具有强非线性、多因素耦合及动态变化特性的特点,传统建模方法在特征表达和预测精度方面存在局限,难以有效刻画混合料含水量随时间变化的动态特征。为此,深度学习方法凭借端到端的自动特征提取能力,逐渐成为烧结过程建模的重要手段。卷积神经网络^[12]在烧结过程数据识别与特征提取中表现出较强的局部特征学习能力,能够有效挖掘原料参数、配料比例及环境条件等因素之间的空间相关性。长短时记忆网络及其改进形式双向LSTM(BiLSTM)具有优秀的时序建模能力,能够捕捉混合料含水量随时间变化的动态特征,被广泛用于烧结过程的状态预测与趋势分析^[13]。注意力机制(Attention)通过自适应分配特征权重,可突出关键时序信息与重要变量,在处理多变量耦合和非线性特征时表现出较强的选择性与解释性^[14]。然而,深度学习模型在实际应用中仍对超参数设置较为敏感,训练过程依赖经验调整,容易出现过拟合及预测稳定性不足的问题。

因此本文提出一种改进北方苍鹰算法的CNN-BiLSTM-SelfAttention预测模型,该模型结合CNN的局部特征提取、BiLSTM的双向时序建模以及自注意力机制的关键信息聚焦能力,并采用改进的北方苍鹰优化算法对超参数进行全局寻优,以提升预测精度与响应速度。

1 模型介绍

1.1 CNN模型

卷积神经网络(Convolutional Neural Network,

CNN)是一种能够自动提取局部特征的深度学习模型,其通过多层结构对输入数据进行逐层特征提取,能有效捕捉空间相关信息,从而增强数据的表达能力,提高模型的预测精度和稳定性。

为了进一步增强特征提取能力,本文在卷积模块中引入了迭代扩张卷积(Iterative Dilated Convolution, ID-CNN)结构。扩张卷积通过在卷积核元素之间加入间隔,实现感受野的扩大,在不增加参数数量的情况下提取更大范围的特征,其计算方法见式(1)。

$$y[i] = \sum_{k=1}^K x[i+r \cdot k] \cdot \omega[k] \quad (1)$$

式中: r 为扩张率; K 为卷积核大小; $\omega[k]$ 为第 k 个卷积核权重。

当 $r=1$ 时退化为普通卷积, r 增大时可捕捉更长距离的依赖信息。迭代扩张卷积通过多层逐级递增扩张率的堆叠与残差连接,实现多尺度特征融合,提升了CNN在复杂非线性时序数据中的特征表达能力。

1.2 BiLSTM网络

1.2.1 LSTM网络

长短期记忆网络(Long Short-Term Memory, LSTM)是一种特殊的循环神经网络,专为处理和预测时间序列数据而设计,能够有效捕捉长期依赖关系,解决传统RNN在长序列学习中易出现的梯度消失问题,因此在回归预测和时间序列分析等任务中表现出色。其核心在于由遗忘门(f_t)、输入门(i_t)和输出门(o_t)调控的记忆单元(C_t)。记忆单元作为信息传送带,在序列中传递状态。三个门控机制通过 sigmoid 和 Tanh 函数,基于前一时刻隐藏状态与当前输入进行计算。其中,遗忘门表示决定从记忆单元中丢弃多少过往信息;输入门表示控制将多少新信息写入记忆单元;输出门表示决定将多少记忆单元的信息输出到当前隐藏状态。LSTM神经网络的模型结构如图1所示。

图1中, X_{t-1} 、 X_t 、 X_{t+1} 分别为 $t-1$ 时刻、 t 时刻、 $t+1$ 时刻的输入向量; h_t 、 h_{t-1} 、 h_{t+1} 分别为 t 时刻、 $t-1$ 时刻、 $t+1$ 时刻的输出向量; C_t 、 C_{t-1} 分别为 t 时刻、 $t-1$ 时刻的长时记忆信息。 $\otimes$ 为遗忘门、输入门、输出门所选择的激活函数, σ 表示 sigmoid 函数,输出范围 $[0,1]$ 。最终输出中,采用 tanh 激活函数,将已生成且待记忆的输出值限定在 $[-1,1]$ 范围。计

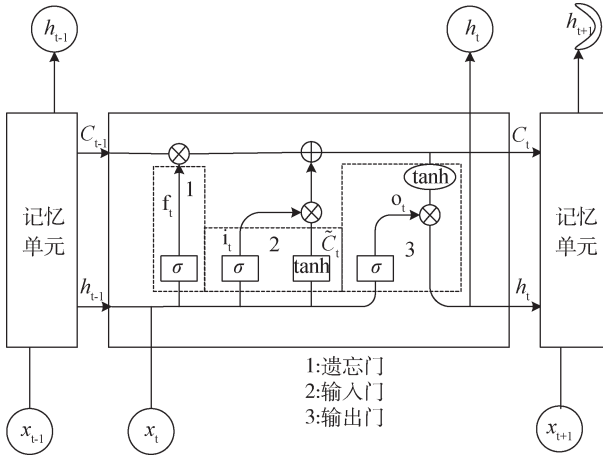


图1 LSTM模型结构

Fig. 1 LSTM model structure

算公式见式(2)-(7)。

遗忘门 f_t :

$$f_t = \sigma(W_f(h_{t-1}, X_t) + b_f) \quad (2)$$

输入门 i_t :

$$i_t = \sigma(W_i(h_{t-1}, X_t) + b_i) \quad (3)$$

候选值 $\tilde{C}_t$:

$$\tilde{C}_t = \tanh(W_c(h_{t-1}, X_t) + b_c) \quad (4)$$

记忆单元 C_t :

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (5)$$

输出门 O_t :

$$O_t = \sigma(W_o(h_{t-1}, X_t) + b_o) \quad (6)$$

最终输出 h_t :

$$h_t = O_t \odot \tanh(C_t) \quad (7)$$

式中: W_f, W_i, W_c, W_o 分别表示遗忘门、输入门、输出门和计算单元状态的权重矩阵; b_f, b_i, b_c, b_o 表示对应门的偏置项; $\odot$ 表示逐元素累积; σ 为sigmoid激活函数。

1.2.2 BiLSTM双向长短期记忆网络

双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)在传统LSTM基础上引入双向信息处理机制,从正向和反向两个方向同时处理序列数据。BiLSTM由前向LSTM和后向LSTM两个独立的LSTM层组成,分别按照时间顺序和时间逆序处理输入序列,每个时间步的最终输出是前向和后向LSTM的隐藏状态的拼接或加权组合。BiLSTM神经网络示意图如图2所示。

图2中, y_{t-1}, y_t, y_{t+1} 分别为BiLSTM神经网络在 t 时刻、 t 前一时刻、 t 后一时刻的输出向量。在

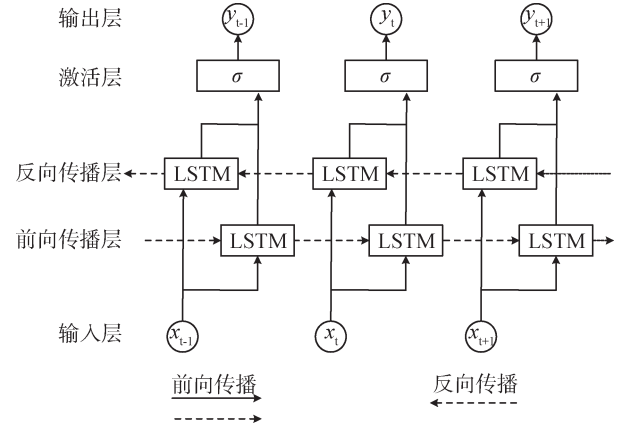


图2 BiLSTM神经网络示意图

Fig. 2 Schematic diagram of BiLSTM neural network

BiLSTM中, t 时刻的输出向量 y_t 通常等同于 t 时刻的隐藏状态 H_t ,它是前向和后向两个LSTM的隐藏状态拼接或加权后的结果,计算公式见式(8)-(10)。

前向(1~ t)传播层:

$$\vec{h}_t = \overrightarrow{\text{LSTM}}(h_{t-1}, X_t, C_{t-1}) \quad (8)$$

反向($t \sim 1$)传播层:

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}(h_{t+1}, X_t, C_{t+1}) \quad (9)$$

t 时刻的隐藏状态:

$$H_t = [\vec{h}_t + \overleftarrow{h}_t] \quad (10)$$

式中: $\vec{h}_t$ 和 $\overleftarrow{h}_t$ 分别为向前和向后隐藏状态; H_t 为 t 时刻的隐藏状态。

BiLSTM采用逐步递归计算,计算复杂度较高,且信息传递时可能会丢失一些细节。注意力机制可以聚焦更为关键的信息,降低对其他信息的关注,过滤掉无关信息,可以提高任务处理的效率和准确性。

1.3 自注意力机制Self Attention

传统注意力机制解决了序列建模中的信息压缩与长距离依赖问题。该机制在目标序列与源序列之间建立关联,通过计算当前解码状态与输入序列各隐状态的相似度动态分配权重,并对输入序列进行加权求和,生成一个整合后的信息表示用于当前输出。其权重计算见式(11)。该方式使模型在每一时间步都能聚焦于最相关的输入信息,从而克服固定长度编码的局限性,显著提升模型处理长距离依赖的能力。

$$a_i = \text{softmax}(\tanh(W_Q Q + W_K K))V \quad (11)$$

式中: Q 是当前目标的查询; K 是源序列的键; V 是源

序列的值; W_Q 和 W_K 是可学习的权重矩阵; softmax 用于将相似度转换为权重分布。

自注意力机制^[15]不依赖外部目标序列,而是在输入序列内部直接建立位置间的关联,以捕捉长距离依赖关系。该机制将输入序列的每个位置映射为查询 Q 、键 K 和值 V 向量,通过点积计算查询与所有键之间的相似度,并利用 Softmax 函数将其归一化为权重,见式(12)。最终通过对值向量进行加权求和,得到每个位置的输出表示。

$$a_i = \text{softmax} \left(\frac{QK}{\sqrt{dk}} \right) V \quad (12)$$

式中: $\sqrt{dk}$ 为缩放因子。

缩放因子 $\sqrt{dk}$ 的引入,一方面缩小点积值的范围,防止 softmax 输出过于尖锐,缓解梯度消失的问

题;另一方面使点积结果更接近期望为0、方差为1的分布,起到类似归一化的效果,提升模型训练的稳定性。

1.4 CNN-BiLSTM-Self Attention 模型

烧结混合料含水量预测模型采用双通道结构,通道1通过多层卷积神经网络提取局部特征,学习输入数据的高级特征表示。通道2则保留原始输入的基本信息,防止在深度特征提取过程中丢失重要的基础信息,并用于输出辅助结果。通道1提取的特征与通道2保留的信息共同传递至BiLSTM。两个通道相辅相成:通道1侧重于提取局部特征,提升模型表达能力;通道2则起到信息补充与校正的作用,增强模型对输入的全面理解。双通道模型流程如图3所示。

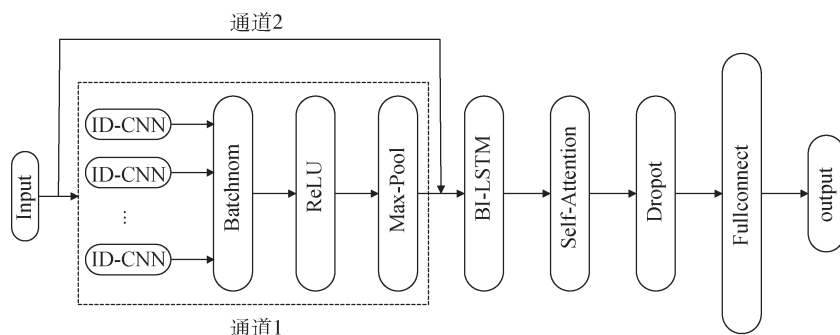


图3 双通道 CNN-BiLSTM-Self Attention 模型流程图

Fig. 3 Flowchart of the dual-channel CNN-BiLSTM-Self Attention model

图3中,模型整合了输入数据与通道1的特征,通过堆叠ID-CNN自注意力机制和Dropout层,进一步强化特征表达能力。BiLSTM模块用于捕捉序列的长期依赖关系,其输出作为自注意力机制的输入结果进行拼接,最后通过全连接层生成最终的预测结果。

双通道卷积神经网络-双向长短期记忆网络-自注意力机制模型(CNN-BiLSTM-Self Attention)结构图如图4所示。

图4中,输入数据首先进入卷积层(Conv1d层),通过多个卷积核提取局部特征,生成特征图。随后,这些特征图被传递至MaxPooling层,进行下采样操作,以降低维度并保留显著特征。经过池化后的特征序列被标记为 X_t (其中 t 表示时间步长),并被输入至双向长短期记忆网络。BiLSTM同时处理正向和反向序列信息,输出每个时间步的隐藏状态 H_t 。

Self Attention层对每个时间步的隐藏状态 h_t 计算注意力权重 α_t ,最终,BiLSTM隐藏层的输出经过Self Attention层加权后,其输出特征计算公式见式(13)。

$$y = \sum_{t=1}^n \alpha_t \times h_t \quad (13)$$

1.5 改进北方苍鹰算法优化混合料含水量预测模型

烧结混合料含水量预测模型中,BiLSTM的学习率、神经元个数、Self-Attention中的键值等关键超参数的设置,对烧结矿混合料含水量预测模型的预测速度和精度有重要影响。采用改进北方苍鹰算法UNGO对模型参数进行寻优,以期实现更精准的烧结混合料含水量预测。

原始NGO算法^[16]在参数优化过程中存在两个主要问题:一是采用随机初始化导致种群早期搜索多样性不足,易陷入局部最优;二是搜索策略中步

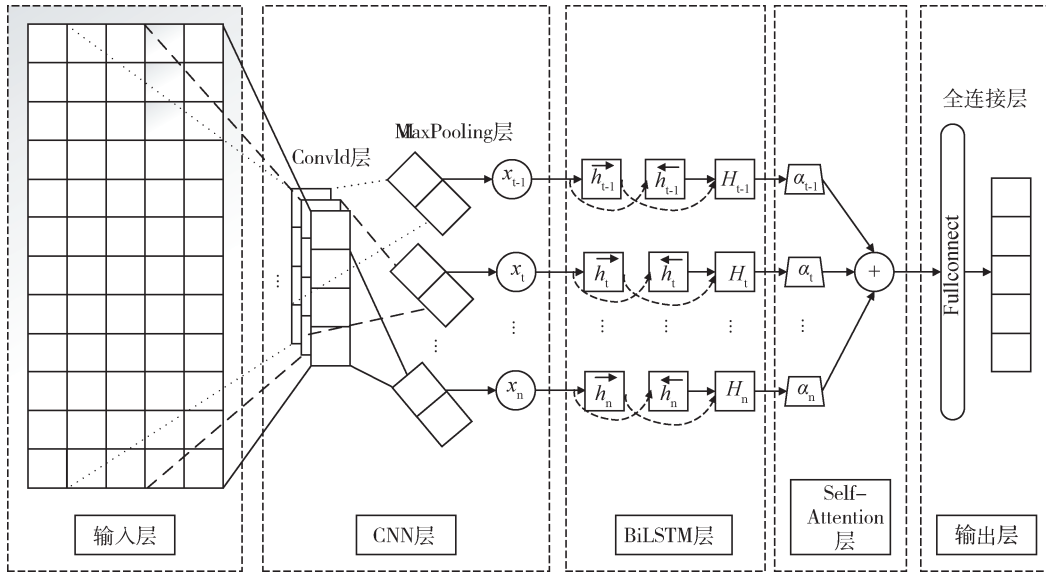


图4 双通道CNN-BiLSTM-Self Attention结构图

Fig. 4 Dual-channel CNN-BiLSTM-Self Attention structure diagram

长调整机制不灵活,导致后期精细搜索效率低。为此,本文引入 *Tent* 混沌初始化策略以增强种群分布的多样性,并结合非线性正余弦衰减机制以动态调整搜索强度,从而提高全局探索与局部精度平衡能力。

1.5.1 *Tent* 混沌映射

Tent 混沌映射 (*Tent* Chaotic Map) 是一种常见的离散一维混沌映射,用于生成在区间 $[0, 1]$ 上分布均匀、具有随机性的序列。它广泛应用于智能优化算法中的种群初始化或参数扰动等场景,以增强算法的多样性和跳出局部最优的能力。

Tent 分布使初始种群分布更加均匀分散,利用 *Tent* 混沌映射生成的值来填充种群位置,使种群的初始阶段覆盖更大的搜索空间,避免早期收敛局部最优解。*Tent* 计算公式见式(14)。

$$Tent(i, j) = \begin{cases} Tent(i, j-1), & Tent(i, j-1) < tent \\ \frac{1 - Tent(i, j-1)}{1 - tent}, & Tent(i, j-1) \geq tent \end{cases} \quad (14)$$

式中: $Tent(i, j-1)$ 表示矩阵中第 i 个个体在第 $j-1$ 维度的值,决定了当前个体 i 在第 j 维度的值,即通过迭代 $j-1$ 维度的值来影响 j 维度的映射行为, $tent$ 为混沌系数,设定值为 1.1。

将映射后的值应用到种群初始位置公式,见式(15)。

$$x = lb + Tent(ub - lb) \quad (15)$$

式中: x 为第 i 个苍鹰个体的位置; ub 和 lb 分别为搜

索空间的最大值和最小值。

随机初始化和 *Tent* 混沌映射初始化种群对比如图5所示。

图5中,左图(随机初始化)展示的是种群在使用常规随机方式初始化时的分布情况。由于随机性,个体分布可能不均匀,存在聚集或空洞,可能导致初始种群多样性较差,从而影响优化效果。右图(*Tent* 混沌初始化)展示的是种群在使用 *Tent* 混沌映射初始化时的分布情况,由于 *Tent* 映射具有良好的覆盖性和分布均匀性,个体在解空间中分布更为均匀,有助于提升种群的初始多样性,增强算法的全局搜索能力。

1.5.2 非线性正余弦衰减策略

在搜索阶段,算法通过调整搜索半径递减因子来控制个体在局部区域内的搜索范围,从而提升局部搜索的精度,逐步逼近最优解。搜索阶段引入正余弦策略,利用正弦与余弦函数的震荡特性调整个体在搜索阶段的移动幅度,且在正余弦策略中引入非线性递减因子,动态调整搜索过程中的步长和收敛速度。具体实现过程见式(16)。

$$x'_{i,j} = \begin{cases} x_{i,j} + r_1 \times \sin(r_2) \times |r_3 P_j^i - x_{i,j}|, & r_4 > 0.5 \\ x_{i,j} - r_1 \times \cos(r_2) \times |r_3 P_j^i - x_{i,j}|, & r_4 < 0.5 \end{cases} \quad (16)$$

式中: $x_{i,j}$ 和 $x'_{i,j}$ 分别表示第 i 只北方苍鹰在 t 和 $t+1$ 次迭代时的位置在第 j 维的分量; P_j^i 为当前全局最优

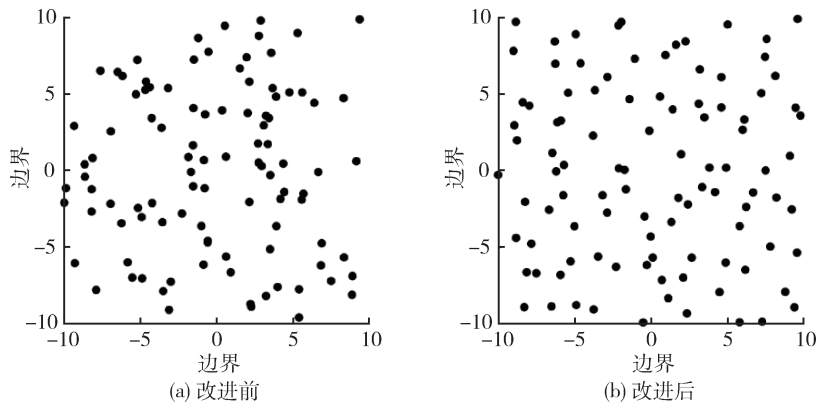


图5 初始化种群对比图

Fig. 5 Comparison of initial populations

解的位置;递减因子 r_1, r_2, r_3 和 r_4 为随机参数,其中 $r_1 = 2(1 - t/T)$, r_2 和 $r_3 \in [0, 2\pi]$, 而 $r_4 \in [0, 4]$ 。在算法的迭代过程中, r_1 搜索递减因子采用线性递减策略可能在迭代后期过早减弱搜索能力。为此,本文引入基于Sigmoid函数的非线性衰减策略该策略使得搜索强度在前期保持平稳,中期快速过渡,后期趋于收敛。并引入权重因子 w 控制全局扰动强度,增强收敛稳定性,具体实现过程,计算公式见式(17)–(19)。

$$r_1' = \frac{1}{(1 + \exp(-\alpha(t - \beta)))} \quad (17)$$

$$w = \frac{1}{1 + \gamma \cdot t} \quad (18)$$

$$x'_{i,j} = \begin{cases} w \times x_{i,j} + r_1' \times \sin(r_2) \times |r_3 P^j - x_{i,j}|, & r_4 > 0.5 \\ w \times x_{i,j} - r_1' \times \cos(r_2) \times |r_3 P^j - x_{i,j}|, & r_4 < 0.5 \end{cases} \quad (19)$$

式中: α, γ 均为为控制衰减速度的参数,取值为1.2和0.1; β 表示函数的中心位置,设为总迭代次数的一半(即 $\beta = T/2$); r_1' 为新的衰减因子。

1.6 基于混合模型的烧结混合料含水量预测流程

为进一步提升烧结混合料含水量预测模型的性能,本文引入改进的北方苍鹰优化算法(UNGO)对CNN-BiLSTM-Self Attention网络的关键超参数进行联合优化。优化参数包括学习率、BiLSTM层神经元数量、正则化系数以及Self-Attention机制中的键值维度。该方法通过算法搜索替代人工经验选参,能够有效减少参数设置的盲目性与随意性,实现模型参数的自适应优化。在优化过程中,将四个待优化参数映射为改进北方苍鹰种群的个体位置向量,个体当前位置对应一组网络参数取值。随着

种群的迭代演化,北方苍鹰算法不断更新个体位置,逐步逼近全局最优解。最终获得的最优位置即为CNN-BiLSTM-Self Attention模型的最优参数组合。该优化策略可有效提高模型的收敛速度与预测精度,并增强其在复杂烧结过程数据下的泛化能力。基于UNGO的CNN-BiLSTM-Self Attention烧结混合料含水量预测模型的实验步骤如下所述。

步骤1):选取UNGO种群数量为 N 个,将其位置作为待优化参数,根据改进型Tent混沌映射初始化种群。

步骤2):将粒子的位置作为模型的学习率参数正则化参数BiLSTM神经元个数和自注意力机制的键值数,通过UNGO计算出烧结混合料含水量的结果,求得与真实值的均方误差(MSE)。

步骤3):通过引入正余弦策略替换搜索阶段的方式表示苍鹰位置更新公式,采用动态搜索范围推动算法走向全局最优解。根据每组MSE更新苍鹰与猎物的位置。

步骤4):重复步骤2)和3),同时记录 N 组粒子中的最优位置,直到满足MSE达到设定的范围或达到最大迭代次数的要求。

步骤5):将最优位置作为最优参数应用到CNN-BiLSTM-Attention模型中,以便进行优化后的训练和预测。

本研究将模型预测值与真实值之间的均方误差(MSE)作为UNGO算法的适应度函数,基于上述优化策略,所构建的UNGO-CNN-BiLSTM-Self Attention模型总体流程如图6所示。

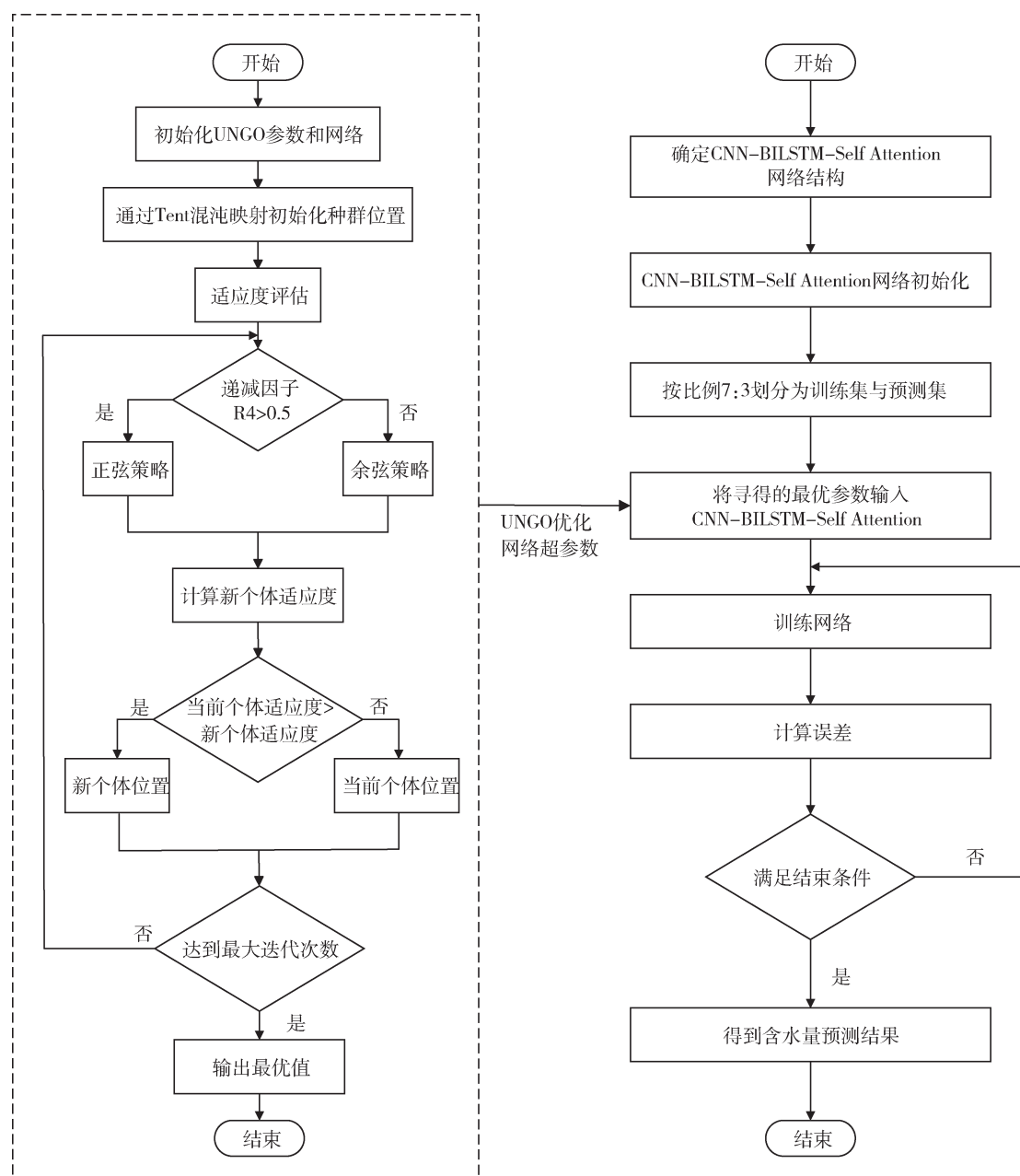


图6 基于UNGO-CNN-BiLSTM-Self Attention混合料含水量预测流程

Fig. 6 Prediction process of mixture moisture content based on UNGO-CNN-BiLSTM-Self Attention

2 实验结果与分析

研究采用河北某大型钢厂2024年4月1日至10月31日的实际生产数据进行仿真实验。基于烧结过程机理分析与专家经验,筛选出影响混合料含水量的21个关键参数,最终确定的变量见表1。从生产数据库中提取500组过程参数样本,采样间隔均为1 min。为提升数据质量,对原始数据进行了预处理:首先识别并剔除明显超出合理范围的异常值;

随后,针对少量缺失数据,采用线性插值方法进行填补,计算式见式(20)。将所提取的500组数据样本,按7:3的比例划分为训练集(350组)与测试集(150组)。模型训练以适应度函数均方误差(MSE)作为优化目标。

$$y = y_a + (y_b - y_a) \cdot \alpha \quad (20)$$

式中: y 为待填补的缺失值估计结果; y_a 和 y_b 分别为缺失值位置前后最近的有效数据; α 为填补位置相

对于 y_a 和 y_b 的比例系数。

表1 配料系统的各项原料参数

Table1 Raw material parameters of batching system

参数种类	参数名称
料仓物料的百分比含量	高炉返矿配比/%
	烧结返矿配比/%
	焦粉配比/%
	白云石配比/%
	自产白灰配比/%
	混合白灰配比/%
	外购白灰配比/%
	高炉灰配比/%
	除尘灰配比/%
	混匀矿配比/%
混合料的成分含量	混料二氧化硅/%
	混料氧化亚铁/%
	混料氧化钙/%
	混料氧化镁/%
	混料氧化铝/%
	混料含水量/%
	碳含量/%
其他	全铁/%
	一混二混加水量/(t·h ⁻¹)

此方法能够在不引入剧烈波动的前提下,有效保证时间序列数据的连续性。

采用Spearman 相关系数进行相关性分析,为了更直观展现 17 个特征变量对烧结混合料含水量的

影响程度,利用 feature importance 排序对 17 个特征变量与烧结混合料含水量之间的相关性进行评估,如图 7 所示。

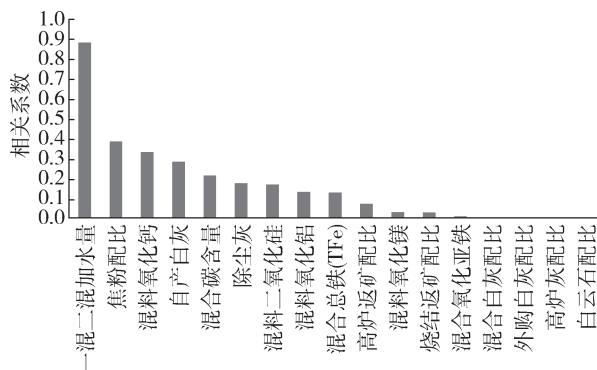


图7 feature importance 排序

Fig. 7 Feature importance

由图 7 可知,与烧结混合料含水量有强相关的特征变量为一混二混加水量,当相关系数在 0.0 时,可以认为两特征之间没有相关性,因此对相关系数为 0.0 的特征变量进行剔除,即对混合白灰配比、外购白灰配比、高炉灰配比以及白云石配比进行剔除。此外,由于混匀矿配比是一个固定值,因此将其用于训练没有太大意义,所以对其也进行剔除。剩余的量作为混合料含水量预测的输入样本。

2.1 改进算法的有效性分析

为验证算法的寻优能力,分别选择 2 个典型的单峰测试函数、2 个典型的多峰测试函数和 2 个典型的固定维多峰测试函数,对算法的寻优能力进行测试。选取的基准测试函数相关参数见表 2。

表2 基准测试函数

Table2 Benchmark functions

基准测试函数	维度	搜索范围	理论最小值
$f_1(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	[-100,100]	0
$f_2(x) = \sum_{i=1}^n (\sum_{j=1}^i x_j)^2$	30	[-10,10]	0
$f_3(x) = \sum_{i=1}^n -x_i \sin(\sqrt{ x_i })$	30	[-500,500]	0
$f_4(x) = \sum_{i=1}^n [x_i^2 - 10\cos(2\pi x_i) + 10]$	30	[-5.1,5.1]	0
$f_5(x) = \sum_{i=1}^n [a_i - \frac{x_1(b_i^2 + b_1 x_2)}{b_i^2 + b_1 x_3 + x_4}]^2$	4	[-5,5]	0.1484
$f_6(x) = -\sum_{j=1}^4 c_j \exp[-\sum_{j=1}^6 a_{ij}(x_j - p_{ij})^2]$	6	[0,1]	-3

为验证 UNGO 算法的优越性,选取 6 个基本测试函数,通过与北方苍鹰优化算法(NGO)、粒子群算法(PSO)、鲸鱼算法(WOA)寻优比对来验证 UNGO 的优越性,种群规模设置为 30,最大迭代次数

为 500。为降低随机误差对结果的影响,对 6 个测试函数分别进行 30 次独立实验,求平均值及标准差,计算结果见表 3。

表 3 UNGO 与其他算法对比

Table3 Comparison of UNGO with other algorithms

函数	指标	PSO	WOA	NGO	UNGO
$f_1(x)$	平均值	1.32E-28	1.18E-74	1.40E-88	1.57E-282
	标准差	4.21E-27	1.57E-87	6.07E-87	0
$f_2(x)$	平均值	1.59E-30	1.68E-51	1.71-45	0
	标准差	9.16E-29	2.29E-51	2.19E-45	0
$f_3(x)$	平均值	4.05E-51	1.00E-112	8.68E-171	0
	标准差	8.49E-51	5.37E-112	0	0
$f_4(x)$	平均值	1.97E-3	4.43E-3	5.96E-4	4.67E-5
	标准差	9.82E-4	6.13E-3	2.76E-4	4.01E-5
$f_5(x)$	平均值	5.36E-4	8.40E-37	2.56E-46	0
	标准差	6.55E-4	4.60E-36	3.03E-46	0
$f_6(x)$	平均值	1.97E-7	4.95E-1	4.41E-9	0
	标准差	6.56E-7	1.13E-1	1.06E-8	0

UNGO 算法在单峰、多峰、复合峰函数中均能达到理论最优值,表明 UNGO 具有更强的全局寻优和局部搜寻能力,且在最值、平均值和标准差的计算上同样优于其他算法。

为验证算法的收敛速度、寻优精度,分别用 4 种智能算法对 6 个基准测试函数进行迭代寻优,各函数的迭代曲线如图 8 所示。

NGO 算法、WOA 算法和 PSO 算法收敛曲线变化相对平稳,而 UNGO 算法收敛曲线先大幅度下降后趋于稳定,收敛精度和速度明显优于其他算法。这是由于对正余弦策略的步长搜索因子进行改进,对其最优位置不断进行调整,增强了其全局寻优和局部搜寻的能力,且用正余弦策略替换原始苍鹰算法的勘察阶段的位置,可加强北方苍鹰位置的灵动性,进一步提升算法后期的局部优化能力,提高了收敛精度。由上述分析可知,UNGO 算法的寻优能力、收敛速度及寻优精度等优于 NGO、WOA、PSO 算法。

2.2 算法对比实验

本实验以烧结混合料含水量为预测目标,模型的输入由 12 个特征组成,分别为一混加水量、二混加水量、焦粉配比、混料氧化钙配比、自产白灰配

比、混合碳含量、除尘灰配比、混料二氧化硅含量、混料氧化铝含量、混料总铁含量、高炉返矿配比、混料氧化镁含量和烧结返矿配比。为提升模型预测性能,分别将 CNN-BiLSTM-Self Attention 模型与 NGO、WOA、PSO 三种优化算法结合,优化的四个关键超参数依次为学习率、神经元个数、正则化系数和自注意力键值,其参数上下界分别设为 $[0.001, 1, 1, 0.0001]$ 与 $[0.01, 50, 100, 0.01]$ 。优化过程中设定初始种群规模为 10,最大迭代次数为 20 代,每种算法独立运行 50 次以获得稳定结果,运行结果见表 4。

图 9 为经过 UNGO、PSO、NGO、WOA 优化后的预测模型的收敛曲线,结果表明,UNGO 达到收敛稳定需要 4 代,适应度曲线趋近于 0.022。PSO 达到收敛稳定需要 6 代,适应度曲线近似于 0.048。NGO 达到收敛稳定需要 15 代,适应度曲线趋近于 0.037。WOA 达到收敛稳定需要 9 代,适应度曲线趋近于 0.04。通过比较,经过 UNGO 优化后的预测模型性能优于 NGO、PSO、WOA,训练时间更短,收敛速度更快。

为对比北方苍鹰优化后的 CNN-BiLSTM-Self Attention 不同算法优化效果,将 PSO、UNGO、NGO、

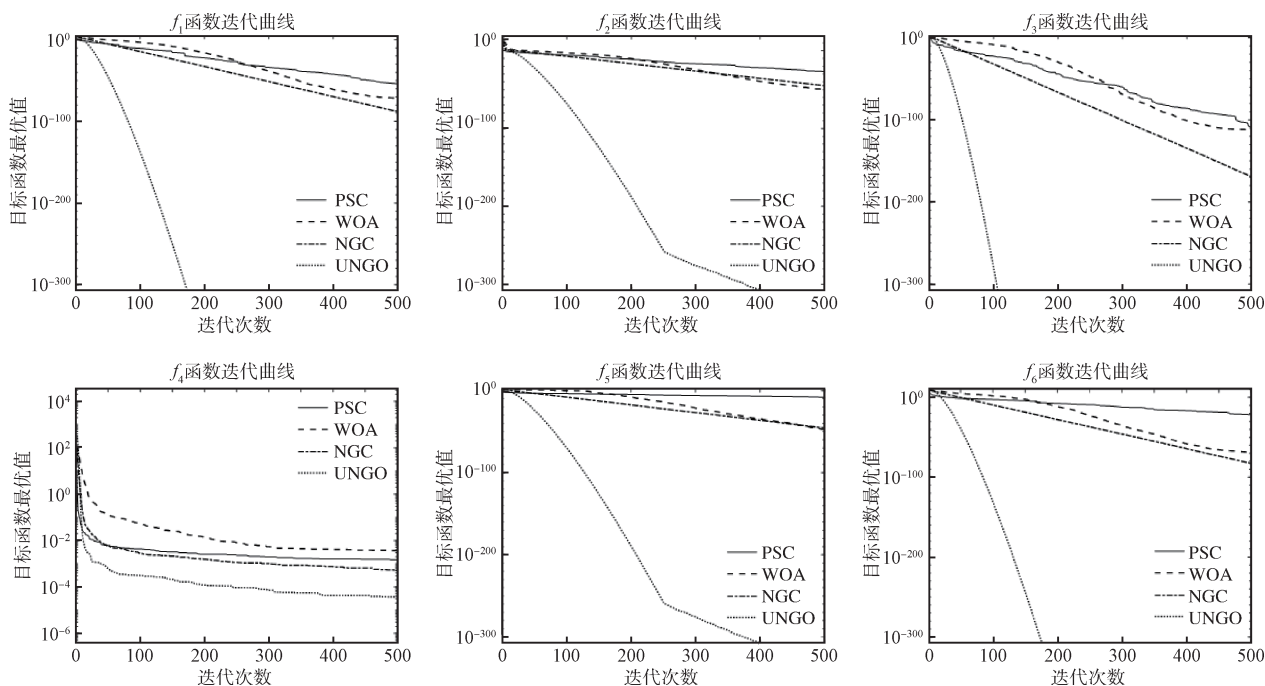


图8 性能测试函数收敛对比图

Fig. 8 Convergence comparison of performance test functions

表4 参数设置列表

Table.4 Parameter setting list

参数名称	优化前设定值	UNGO优化后数值
学习率	0.01	0.053
神经元个数	25	19.096 4
正则化系数	0.001	0.004 2
自注意力键值	2	48.422 5

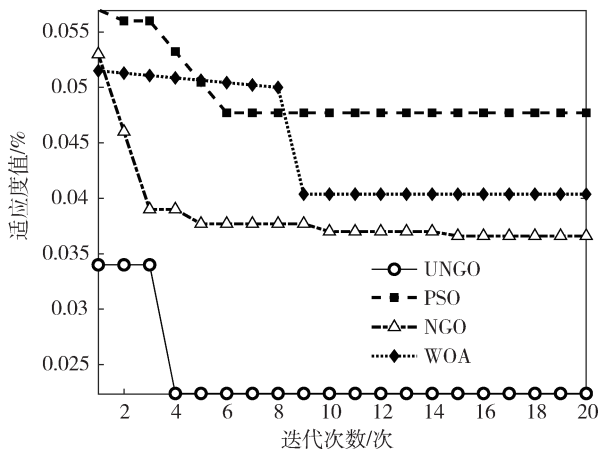


图9 收敛曲线对比图

Fig. 9 Convergence curve comparison

WOA优化后的预测模型用于混合料含水量预测中,对以上模型进行30次实验,预测结果对比如图10

所示。图10表明,UNGO经过优化的预测模型预测精度最高,预测结果和真实值最接近。

对以上模型进行30次实验,并采用MSE(均方误差, Mean Square Error)、RMSE(均方根误差, Root Mean Square Error)、MAE(平均绝对误差, Mean Absolute Error)、 R^2 (决定系数, Coefficient of Determination)4种性能指标进行评价,计算公式见式(21)~(24)。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{real} - y_{pro})^2 \quad (21)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{real} - y_{pro})^2} \quad (22)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{real} - y_{pro}| \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{real} - y_{pro})^2}{\sum_{i=1}^n (y_{real} - \bar{y})^2} \quad (24)$$

式中: y_{real} 为真实值; y_{pro} 为预测值; $\bar{y}$ 为目标变量的均值。

最终取各指标平均值作为评价指标。不同算法对比见表5。

由表5可知,UNGO优化模型与原始NGO、PSO、WOA优化模型相比,拥有更高的 R^2 ,同时

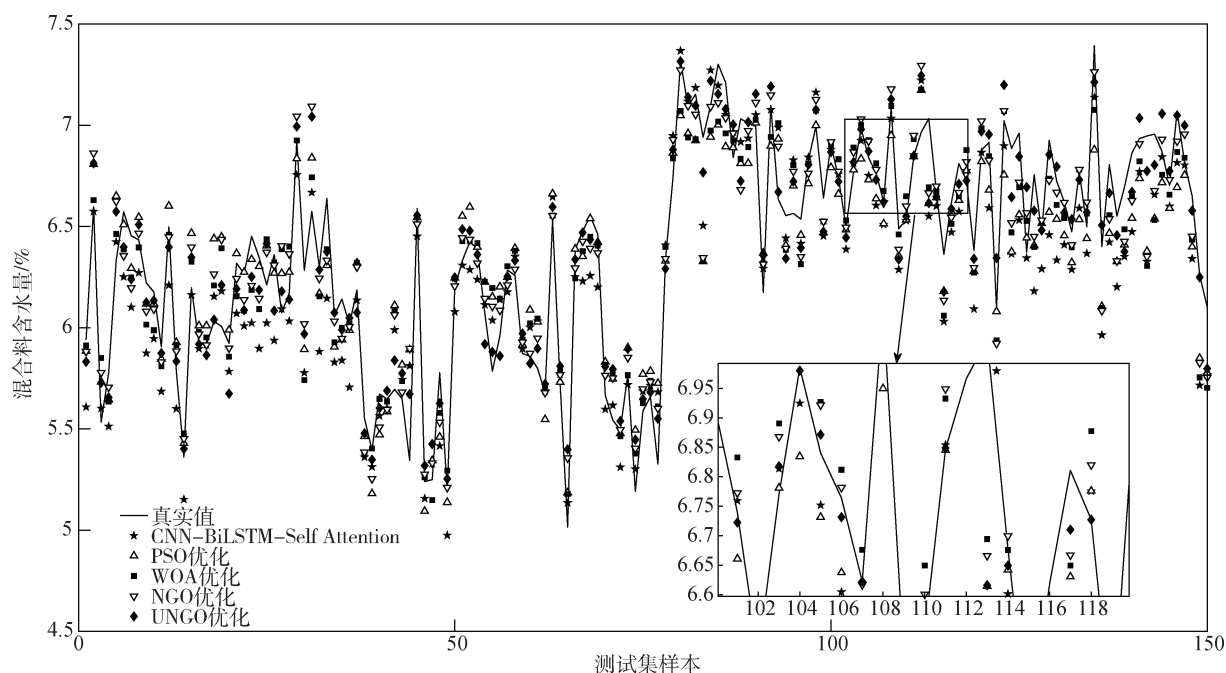


图10 预测结果对比

Fig. 10 Comparison of prediction results

表5 不同算法评估指标对比

Table.5 Comparison of mainstream models

算法	R^2	MSE	$RMSE$	MAE
CNN-BiLSTM-ATT	0.790 8	0.057 5	0.239 8	0.192 3
UNGO 优化	0.918 4	0.022 4	0.149 8	0.118 5
PSO 优化	0.826 4	0.047 7	0.218 5	0.174 7
NGO 优化	0.867 0	0.036 6	0.191 2	0.145 2
WOA 优化	0.853 1	0.040 37	0.200 1	0.154 9

MSE 、 $RMSE$ 、 MAE 误差指标更低。数据结果表明, UNGO 优化模型的精度最好, 经过 UNGO 评价指标 MSE 、 $RMSE$ 、 MAE 相比于 NGO 模型分别降低了 38.80%、21.66%、18.39%, UNGO 相比于 PSO 模型分别降低了 53.04%、31.45%、32.17%, UNGO 相比于 WOA 模型分别降低了 44.53%、25.13%、23.50%, 均优于其他算法寻优结果, 预测效果最佳。

2.3 模型对比实验

2.3.1 实验一

本文仿真实验均基于 Windows 11(64bit), 处理器 Intel(R) Core(TM) i7-13650HX (2.60 GHz), 内存 16.0GB, 仿真平台 MatlabR2024b。本实验旨在剥离优化算法的影响, 从结构层面, 对本文提出的 CNN-BiLSTM-Self Attention 混合模型进行系统性对比验证。为了验证 CNN-BiLSTM-Self Attention 模

型的结合效果, 以及更直观的观察所提模型的优越性能并验证模型的泛化能力, 实验展示了五种模型 (CNN、LSTM、CNN-LSTM、CNN-BiLSTM 和 CNN-BiLSTM-Self Attention) 的误差指标 (MAE 、 $MAPE$ 、 $RMSE$ 和 R^2) 对比。图 11 所示的对比结果表明, 本文提出的 CNN-BiLSTM-Self Attention 模型在各项评估指标中均位居首位。具体来看, 该模型在 MAE 、 $MAPE$ 和 $RMSE$ 三项误差指标上均达到最低, 说明其在控制预测偏差方面表现稳定, 尤其对较大误差具备更好的抑制能力。同时, 该模型的 R^2 值也高于其他所有对比模型, 反映出其对数据整体趋势具有更强的解释力, 拟合效果更为理想。在参与对比的模型中, CNN-BiLSTM 的综合性能仅次于本文模型, CNN-LSTM 表现处于中等水平, 而 LSTM 模型在所有指标上的结果均不够理想, 特别是在 MAE 与 $RMSE$ 上误差明显偏高, 表明其结构在处理此类预测任务时存在一定局限。通过对这些误差指标的比较, 证实 CNN-BiLSTM-Self Attention 模型最优。

2.3.2 实验二

实验二为验证 UNGO-CNN-BiLSTM-Self Attention 模型的优越性和有效性, 本文将其与 UNGO-Transformer、UNGO-BP、UNGO-XGBoost 等主流预

测模型进行对比实验。实验首先在训练集上对各个模型进行训练,并通过UNGO算法对各模型的超参数进行寻优,然后在测试集上进行验证。由于不同模型结构所涉及的超参数种类和数量各不相同,例如Transformer模型包含注意力头数和编码层数等结构参数,BP神经网络涉及隐藏层节点数和层数等配置参数,而XGBoost模型则包括学习率、最大深

度、子样本比率、最小分裂损失等控制树结构与正则化的关键参数。为保证比较的公平性,统一采用UNGO算法分别对各模型的关键超参数进行优化寻优。为减小随机性对实验结果的影响,进行10次重复实验,并对各次实验结果取平均值,以确保实验结果的可靠性。各模型预测结果如图12所示,预测绝对误差分布情况见图13。

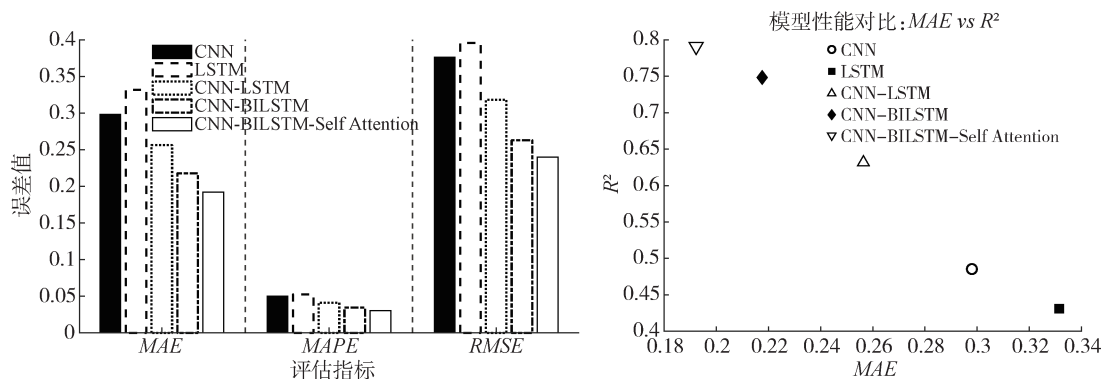


图11 性能多维评估图

Fig. 11 Multi-dimensional Performance Assessment

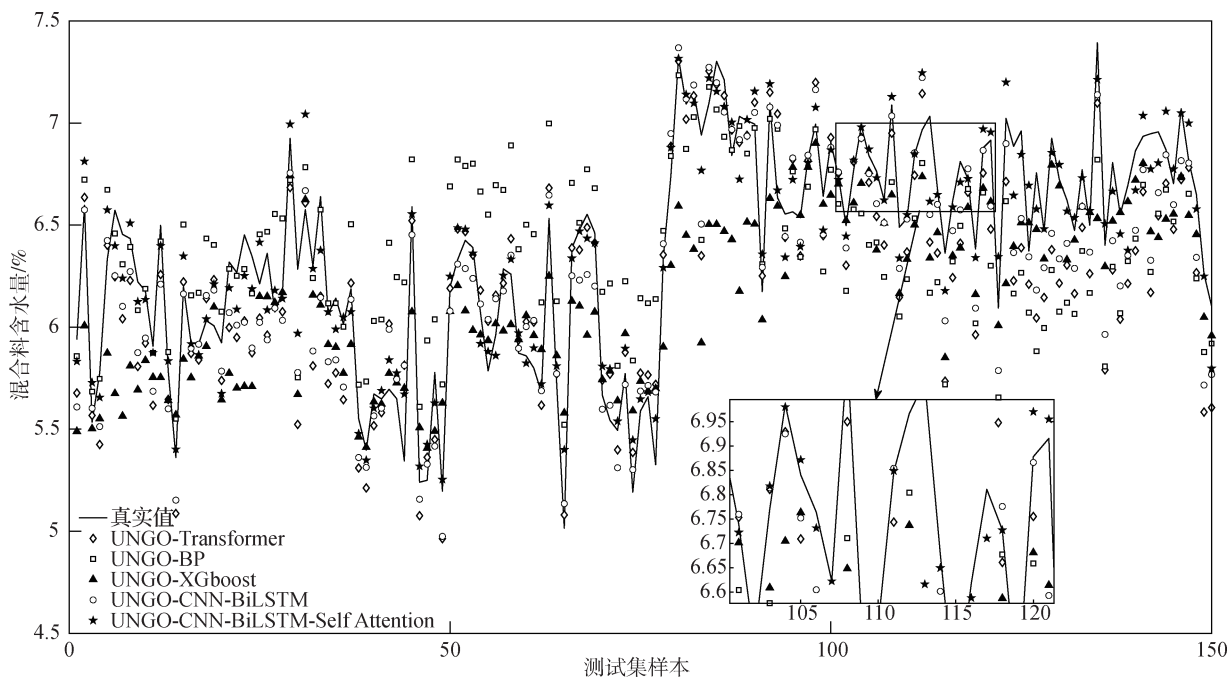


图12 各模型预测结果对比图

Fig. 12 Comparison chart of prediction results of each model

从图中的预测结果可以明显看出,UNGO-CNN-BiLSTM-Self Attention模型的预测曲线最接近真实值,且误差曲线平滑、波动较小,充分验证了自注意力机制在捕捉局部特征方面的有效性,显著提

升了模型的鲁棒性。相比之下,BP和XGBoost模型表现相对较弱,预测误差较大,尤其在混合料含水量剧烈变化的区间,难以准确捕捉数据的变化趋势,导致预测结果与真实值存在明显偏差,反映出

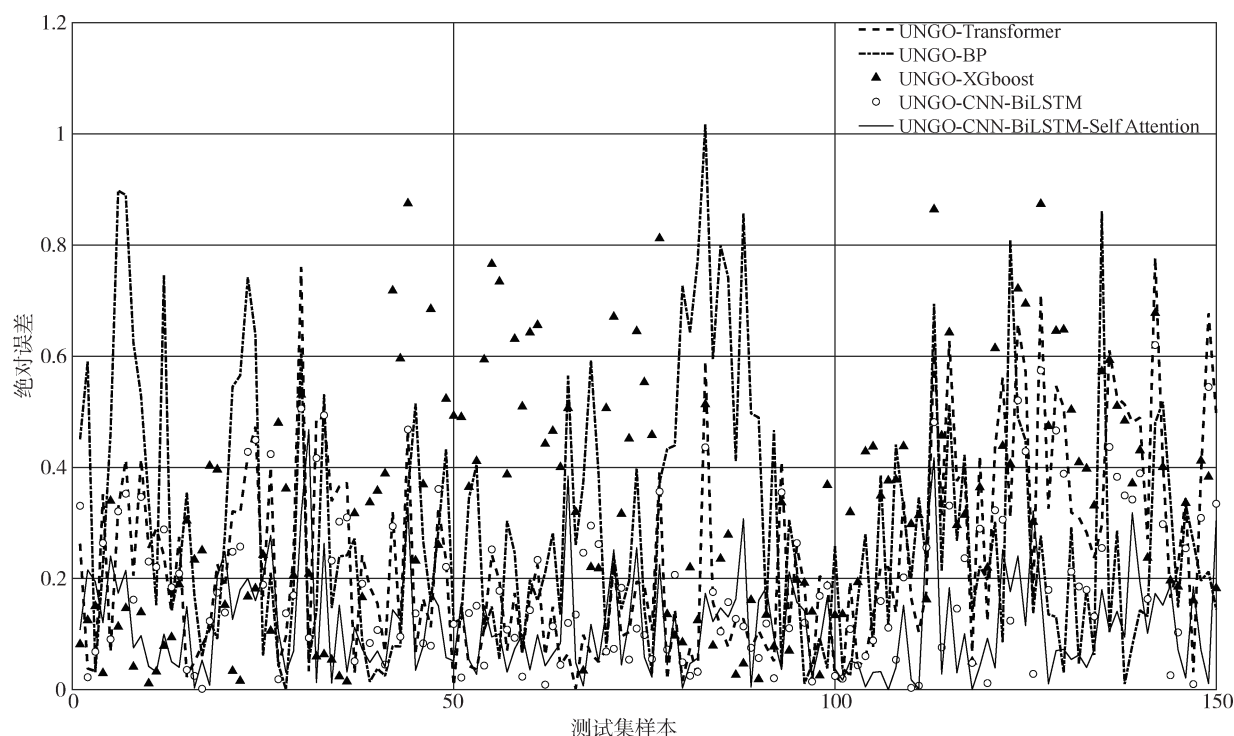


图13 各模型预测结果绝对误差

Fig. 13 Absolute errors of the prediction results of each model

传统神经网络在处理时序特征及高维非线性问题时的局限性。UNGO-Transformer模型虽同样具备全局注意力机制,在建模长距离依赖方面表现良好,但由于结构复杂且参数众多,对小样本数据较为敏感,部分波动区间出现过拟合或欠拟合现象,导致预测曲线出现一定偏移。进一步对比UNGO-CNN-BiLSTM-Self Attention与UNGO-CNN-BiLSTM模型,前者融合自注意力机制后,能够更精准地捕捉输入特征中的关键时序信息和局部依赖关系,使预测结果更加平滑且误差更小,表现出更强的泛化能力与鲁棒性。整体来看,融合卷积、双向循环单元与自注意力机制的UNGO-CNN-BiLSTM-Self Attention模型,在稳定性和拟合精度之间取得了更优的平衡,展现出卓越的泛化性能。

3 结论

针对目前烧结混合料含水量预测精度不足与收敛速度慢等问题,本文提出并构建了一种基于改进北方苍鹰优化算法(UNGO)的CNN-BiLSTM-Self Attention混合模型,并将其与其他算法的寻优能力、收敛速度及寻优精度等进行了对比,采用样本进行

了预测结果的有效性验证,得到以下主要结论。

1)在样本预处理阶段,通过异常值剔除与缺失值插补提升了数据质量;采用Spearman相关系数,将输入量的特征变量与烧结混合料含水量之间的相关性进行评估,剔除相关性低的变量,用来减小模型的复杂度,减小模型运行时间。

2)算法对比实验结果表明,经UNGO算法优化后,模型的 MSE 、 $RMSE$ 、 MAE 分别降至0.022 4、0.149 8和0.118 5, R^2 提升至0.918 4。与NGO、PSO、WOA等算法优化结果相比, MSE 、 $RMSE$ 、 MAE 分别降低了38.80%、21.66%和18.39%,表明UNGO算法在超参数全局优化中具有更高的寻优效率和更优的收敛性能。

3)模型性能通过两组对比实验得以验证。一方面,在剥离算法优化的基础模型测试中,CNN-BiLSTM-Self Attention模型的误差指标分别为: $MSE=0.057 5$ 、 $RMSE=0.239 8$ 、 $MAE=0.192 3$,决定系数 $R^2=0.790 8$,整体性能优于CNN、LSTM、CNN-LSTM及CNN-BiLSTM等模型,说明该模型在特征提取和时序建模方面具有较强优势。另一方面,与UNGO-Transformer等主流模型对比时,UNGO优化后的

CNN-BiLSTM-Self Attention 模型在预测精度、误差波动等方面均表现最佳,体现出其优越的综合性能。

综上所述,本文所提出的算法运行速度较快,满足混合料含水量预测模型的时效性。该组合预测模型也为提高混合料含水量预测精度提供了一定参考价值。接下来,可以进一步进行拓补该模型的结构和参数,探索更多更新的算法和优化技术,以满足实际应用的需求,并促进烧结机领域的发展。

[参考文献]

- [1] 吴岳明, 聂慧远, 吴朝霞. 基于非线性自回归神经网络的烧结混合料水分智能控制[J]. 冶金自动化, 2021, 45(1): 27-33.
WU Yueming, NIE Huiyuan, WU Zhaoxia. Intelligent control of sintering mixture moisture based on NARX neural network [J]. Metallurgical Industry Automation, 2021, 45(1): 27-33.
- [2] JIANG Y, YANG N, YAO Q, et al. Real-time moisture control in sintering process using offline-online NARX neural networks [J]. Neurocomputing, 2020, 396: 209-215.
- [3] 周林. 烧结混合料制粒适宜水分预测模型研究[J]. 山东冶金, 2021, 43(2): 30-32.
ZHOU Lin. Study on prediction model of suitable moisture content of sintering mixture granulation [J]. Shandong Metallurgy, 2021, 43(2): 30-32.
- [4] 丁园, 王斌, 鄢进冲, 等. 基于二维区间自回归模型的烧结终点预测[J]. 烧结球团, 2017, 42(03): 1-6, 15.
DING Yuan, WANG Bin, YAN Jinchong, et al. Prediction of burning through point based on two-dimensional interval autoregressive model [J]. Sintering and Pelletizing, 2017, 42(3): 1-6, 15.
- [5] 孙学波, 屈文涛. 基于RBF神经网络的烧结终点预测模型[J]. 世界科技研究与发展, 2008(1): 21-23.
SUN Xuebo, QU Wentao. Burning through point prediction model based on RBF neural network [J]. World Sci-Tech R & D, 2008(1): 21-23.
- [6] 杨宁. 烧结混合料水分智能测控系统研究[D]. 沈阳: 东北大学, 2020.
YANG Ning. Research on intelligent measurement and control system of moisture content of sintering mixture [D]. Shenyang: Northeastern University, 2020.
- [7] 尹学, 李绍铭, 闫成忍. 基于GA-PSO-BP神经网络的烧结终点预测模型[J]. 烧结球团, 2020, 45(6): 35-39, 46.
YIN Xue, LI Shaoming, YAN Chengren. Prediction model of BTP based on GA-PSO-BP neural network [J]. Sintering and Pelletizing, 2020, 45(6): 35-39, 46.
- [8] 王瑞林. 烧结混合料水分优化控制方法及实现[D]. 武汉工程大学, 2023.
WANG Ruilin. Optimization control method and realization of sintering mixture moisture [D]. Wuhan University of Technology, 2023.
- [9] 王自赞, 刘萍先, 陈天荣, 等. 基于IGA-LSTM的大坝变形预测模型研究[J]. 水电能源科学, 2025, 43(6): 134-137.
WANG Ziyun, LIU Pingxian, CHEN Tianrong, et al. Research on dam deformation prediction model based on IGA-LSTM [J]. Water Resources and Power, 2025, 43(6): 134-137.
- [10] 吕世轩, 郑丽君, 刘鑫, 等. 基于DTW和CNN-BiLSTM的新型电力系统等效惯量评估方法[J/OL]. 南方电网技术, 1-12 [2026-05-25]. <https://link.cnki.net/urlid/44.1643.tk.20250421.1656.006>.
LYU Shixuan, ZHENG Lijun, LIU Xin, et al. Equivalent Inertia Estimation Method for New Power System Based on DTW and CNN-BiLSTM [J/OL]. Southern Power System Technology, 1-12 [2026-05-25]. <https://link.cnki.net/urlid/44.1643.tk.20250421.1656.006>.
- [11] 李小根, 刘泓辰, 付景保, 等. 基于CNN-Attention-LSTM神经网络的地下水位预测——以河南某地为例[J/OL]. 水资源保护, 1-15 [2026-05-25]. <https://link.cnki.net/urlid/32.1356.TV.20250407.1755.026>.
LI Xiaogen, LIU Hongchen, FU Jingbao, et al. Groundwater level prediction utilizing a hybrid CNN-Attention-LSTM neural network framework: a case study in Henan [J/OL]. Water Resources Protection, 1-15 [2026-05-25]. <https://link.cnki.net/urlid/32.1356.TV.20250407.1755.026>.
- [12] 靳亚涛, 李宏扬, 刘小杰, 等. 基于大数据和机器学习的烧结终点预测[J]. 冶金自动化, 2024, 48(S1): 101-105.
JIN Yatao, LI Hongyang, LIU Xiaojie, et al. Sintering End Point Prediction Based on Big Data and Machine Learning [J]. Metallurgical Automation, 2024, 48(S1): 101-105.
- [13] 刘梦园, 吴朝霞, 王金杨, 等. 基于TST-LSTM模型的烧结料层透气性预测[J]. 东北大学学报(自然科学版), 2024, 45(10): 1379-1385.
LIU Mengyuan, WU Zhaoxia, WANG Jinyang, et al. Prediction of sinter bed permeability based on TST-LSTM Model [J]. Journal of Northeastern University (Natural Science), 2024, 45(10): 1379-1385.
- [14] 边会媛, 姬嘉骏, 段朝伟, 等. 基于贝叶斯优化的CNN-BiLSTM-Attention的煤体结构识别方法[J]. 煤田地质与勘探, 1-10 [2025-10-24].
BIAN Huiyuan, JI Jiajun, DUAN Chaowei, et al. A Bayesian optimization-based CNN-BiLSTM-Attention Method for coal structure identification [J]. Coal Geology & Exploration, 1-10 [2025-10-24].
- [15] 唐俊飞, 邢海龙, 李溯, 等. 结合改进CNN与自注意力机制的煤矿轮式机器人目标检测技术[J]. 煤矿安全, 2025, 56(3): 224-232.
TANG Junfei, XING Hailong, LI Su, et al. Target detection

technology of coal mine wheeled robot combining improved CNN and self attention mechanism[J]. Safety in Coal Mines, 2025, 56(3): 224-232.

[16] 吴昌寅, 张军. 混合策略改进的北方苍鹰优化算法及其应用

[J]. 智能计算机与应用, 1-11.

WU Changyin, ZHANG Jun. Improved northern goshawk optimization algorithm with hybrid strategy and its applications

[J]. Intelligent Computers and Applications, 1-11.

Research on a sinter mix moisture content prediction model based on UNGO-CNN-BiLSTM-Self Attention

ZHOU Yaluo¹, YAO feiyu¹, LIU Wenguang², ZHANG Ruicheng¹

(1. School of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China;

2. Shougang Jingtang Limited Iron & Steel Co., Ltd., Tangshan 063200, China)

Abstract: To address the issues of insufficient prediction accuracy and slow convergence in sinter mixture moisture content forecasting, this study proposes a hybrid prediction model based on an Improved Northern Goshawk Optimization algorithm (UNGO) integrated with a CNN-BiLSTM-Self Attention architecture. The model first employs a Convolutional Neural Network (CNN) to extract spatial correlation features from the input data, then utilizes a Bidirectional Long Short-Term Memory network (BiLSTM) to capture the temporal dependencies of the sinter mixture moisture content. Meanwhile, a Self-Attention mechanism is incorporated to strengthen key feature weighting and long-range dependency modeling, thereby improving prediction stability and accuracy. To reduce model complexity and enhance computational efficiency, the Spearman correlation coefficient is applied to screen input features and remove variables with low relevance. To overcome the tendency of traditional optimization algorithms to fall into local optima, the proposed UNGO introduces chaotic initialization and a nonlinear sine-cosine decay strategy, significantly enhancing both global exploration and local exploitation capabilities. This enables adaptive optimization of key hyperparameters, including the learning rate, number of BiLSTM neurons, regularization coefficient, and attention key dimension. Experimental results based on production data from a large steel plant in Hebei Province demonstrate that, under identical data conditions, the CNN-BiLSTM-Self Attention model achieves $MSE = 0.0575$, $RMSE = 0.2398$, $MAE = 0.1923$, and $R^2 = 0.7908$, outperforming CNN, LSTM, CNN-LSTM, and CNN-BiLSTM models. This confirms the model's superior capability in feature representation and temporal dependency learning. After optimization with the UNGO algorithm, the performance metrics are further improved to $MSE = 0.0224$, $RMSE = 0.1498$, $MAE = 0.1185$, and $R^2 = 0.9184$, showing faster convergence and higher prediction accuracy than models optimized by NGO, PSO, and WOA. Furthermore, compared with other mainstream prediction models such as UNGO-Transformer, UNGO-BP, and UNGO-XGBoost, the proposed model exhibits superior prediction accuracy, generalization, and robustness. In summary, the UNGO-CNN-BiLSTM-Self Attention model demonstrates excellent overall performance in both structural design and parameter optimization, enabling high-precision prediction of sinter mixture moisture content. It provides a reliable technical foundation for intelligent control and energy efficiency optimization in the sintering process.

Key words: sintering mixture; moisture content prediction; CNN-BiLSTM-Self Attention; improved Northern Goshawk Optimization (UNGO); Self-Attention Mechanism; Tent chaotic initialization strategy; Spearman correlation coefficient