

DOI:10.19651/j.cnki.emt.2519134

基于 GW-Attention-DFNN 的烧结矿质量预测研究^{*}

杨兴旺 杨爱民 薛涛

(华北理工大学理学院 唐山 063210)

摘要: 烧结矿的性能指标可以充分反映烧结矿的质量情况,而烧结矿的质量情况又能提高高炉生产效率,降低能耗和燃料比,促进绿色冶炼与环保,在进行烧结矿质量预测的过程中传统深度神经网络会面临可解释性差,而具有较强解释性的模糊神经网络又容易面临规则膨胀、参数调优困难等问题。本文构建了基于模糊神经网络和深度神经网络相结合的预测模型,首先通过改进 CBAM 通道注意力模块,对输入特征进行通道和空间两种注意力进行计算,进行有效特征的融合;提高了模型对复杂非线性关系的有效建模以及对特征重要性的动态分配能力,并通过改进灰狼优化算法对模型进行优化,提高了模型的预测准确度。最后在烧结矿转鼓指数、烧结矿碱度、 $RDI_{+3.15}$ 预测上进行实验研究,取得了较高的准确度,验证了所提出模型及算法的可行性。通过对 GW-FNN、GW-DFNN、Attention-DFNN 和 GW-Attention-DFNN 4 种模型进行了比较,结果表明 GW-Attention-DFNN 模型所预测的转鼓指数 R^2 为 0.968 2、烧结矿碱度(R)的 R^2 为 0.975 0、 $RDI_{+3.15}$ 的 R^2 为 0.964 2,结果表明该模型在预测烧结矿质量性能方面表现较好。

关键词: 烧结矿性能指标;模糊神经网络;深度神经网络;灰狼优化算法

中图分类号: TP183;TN911 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Research on sinter quality prediction based on GW-Attention-DFNN

Yang Xingwang Yang Aimin Xue Tao

(College of Science, North China University of Science and Technology, Tangshan 063210, China)

Abstract: The performance indicators of sintered ore fully reflect its quality, and the quality of sintered ore, in turn, enhances blast furnace production efficiency, reduces energy consumption and fuel ratios, and promotes green smelting and environmental protection. In the process of predicting sintered ore quality, traditional deep neural networks suffer from poor interpretability, while fuzzy neural networks, which offer strong interpretability, are prone to issues such as rule explosion and difficulties in parameter tuning. This paper constructs a predictive model that combines fuzzy neural networks with deep neural networks. First, by improving the CBAM channel attention module, the model calculates both channel and spatial attention for input features to fuse effective features; this enhances the model's ability to effectively model complex nonlinear relationships and dynamically allocate feature importance. Furthermore, by optimizing the model using an improved Grey Wolf optimization algorithm, the model's predictive accuracy is improved. Finally, experimental studies were conducted on the prediction of sintered ore drum index, sintered ore alkalinity, and $RDI_{+3.15}$, achieving high accuracy and validating the feasibility of the proposed model and algorithm. A comparison of the four models—GW-FNN, GW-DFNN, Attention-DFNN, and GW-Attention-DFNN—revealed that the GW-Attention-DFNN model achieved an R^2 of 0.968 2 for the drum index, 0.975 0 for sintered ore alkalinity (R), and 0.964 2 for $RDI_{+3.15}$. These results indicate that this model performs well in predicting the quality performance of sintered ore.

Keywords: properties and indicators of sintered ore; fuzzy neural network; deep neural network; grey wolf optimization algorithm

0 引言

文献[1]提出钢铁制造作为现代社会当中的支柱型产

业之一,为当代建筑业、工业和各种制造业提供了不可或缺的原材料在烧结矿的日常生产和钢铁冶炼的过程中,烧结矿作为炼铁的主要原料之一,其质量的预测是提高炼铁质

收稿日期:2025-06-18

* 基金项目:国家自然科学基金(52074126)、河北省自然科学基金(E2022209110)项目资助

量的重要环节,但在烧结矿的日常生产和钢铁冶炼的过程中,烧结矿的质量预测仍然面临若干工业痛点,现有预测方法在复杂多变的工况下往往精度不足,从而导致烧结矿的转鼓指数、碱度(R)、还原粉化指数(RDI_{+3.15})等关键指标波动,在增加了能耗的同时也提高了生产的成本;烧结过程涉及多因素耦合与强非线性关系,诸如原料配比复杂、燃料粒度与配比波动、风箱负压与风量调控、料层厚度及点火制度设定、水分控制等因素均会显著影响反应速率与温度场分布,使得控制参数优化难度大,生产调控高度依赖操作经验,难以实现稳定可控;部分传统模型缺乏良好的泛化能力,在原料性质或工况条件发生波动时预测性能显著下降。因此,构建兼具高精度、强稳定性与泛化能力强的烧结矿质量预测模型,对于保障高炉顺行和提升整体生产效率具有重要的现实意义。

在目前,对于烧结矿质量预测的相关模型主要分为 3 类,分别为基于智能优化算法的预测模型、基于数学机理的预测模型、基于规则的专家系统预测模型^[2],传统的烧结矿质量预测系统通常依托于专家的经验知识和理论知识来构建模型^[3],但烧结矿的生产过程中影响因素较多,生产环境非常复杂,这种方式易受到人为操作方式,知识获取困难等因素影响,最终模型的预测精度难以保证。另一方面,依托于数学机理的预测模型构建通常基于物理、化学知识等原理来预测烧结矿的质量^[4]。Pahlevaninezhad 等^[5]提出化学反应中的气化和固化反应综合考虑起来,对烧结过程进行了数值研究。通过广泛的参数,包括焦炭和石灰石粒度,入口气流速度和焦炭含量进行了全面的研究。还有许多的学者将其他的优化算法例如粒子群算法(particle swarm optimization, PSO)、蚁群算法(ant colony optimization, ACO)、差分进化算法(differential evolution, DE)等应用到烧结矿质量预测当中,近年来,随着人工智能领域的发展,大量研究人员在预测烧结矿的化学成分和冶金性的方面进行了深度的研究,并获得了显著的研究成果^[6-10]。例如,刘加达等^[11]通过优化焦炭配比,结合全铁品位和碱度等指标的预测,实现对烧结矿质量的精准控制并在此基础上建立了基于 BP 神经网络(BP neural network, BPNN)和循环神经网络(recurrent neural network, RNN)的烧结矿质量预测模型。Gao 等^[12]利用主成分分析法(principal component analysis, PCA)和遗传算法(genetic algorithm, GA)对用于预测转鼓强度的神经网络进行了改进,最终达到了 95.1% 的准确率。张乐文等^[13]采用加权的支持向量机(support vector machines, SVM)的方法根据烧结过程参数和断面的图像信息预测了转鼓指数,其实际中标率达到了 91%。Liu 等^[14]针对烧结数据具有时间序列的相关性。提出深度神经网络(deep neural networks, DNN)和长短期记忆网络(long short-term memory, LSTM)相结合的预测模型,成功的解决了烧结矿预测其化学成分的问题,并且取得了较好的效果。Li 等^[15]利用 Optuna 框架对烧结矿的质

量预测模型进行了优化,并通过实验证明了优化算法的良好性能。王紫荆等^[16]将改进鲸鱼优化算法(whale optimization algorithm, WOA)应用于工业 CT 图像增强并取得了良好效果。

针对过去传统的烧结矿性能研究模型存在通用性与迁移性差、缺乏对烧结性能评价的因素考虑、在大数据背景下模型的泛化能力不强等问题。本文提出基于通道注意力模块(convolutional block attention module, CBAM)的深度模糊神经网络(deep fuzzy neural network, DFNN)^[17-18]并使用改进灰狼优化算法对模型进行参数寻优^[19],即基于改进灰狼算法与注意力增强的深度模糊神经网络(grey wolf optimizer attention deep fuzzy neural network, GWO-Attention-DFNN)。研究具体内容为:

1)对实验数据进行数据预处理,选择转鼓指数、烧结矿碱度(R)和还原粉化指数(RDI_{+3.15})作为输出指标,并使用随机森对每个输出指标进行特性相关分析,从而实现样本集的构建。

2)将模糊神经网络算法、深度神经网络算法、注意力机制和灰狼优化算法相结合,构建 GW-Attention-DFNN 模型,对 GW-Attention-DFNN 进行消融实验,并选择 MSE、MAE、MAPE、R² 作为模型的评价指标,将模型的预测效果进行展示并分析。本文旨在准确的预测烧结矿质量,进而为生产过程提供更为精准的控制和优化方案,确保烧结矿石冶炼质量,从而提高生产效率。

1 模型架构

本文提出了一个改进灰狼优化算法与改进 CBAM 模块结合 GW-Attention-DFNN。GW-Attention-DFNN 的设计灵感来自对模糊神经网络和深度神经网络的综合理解, GW-Attention-DFNN 融合了模糊神经网络(feedforward neural network, FNN)、深度神经网络(deep neural networks, DNN)^[20]和灰狼优化算法(grey wolf optimizer, GWO)、以及改进 CBAM 通道注意力机制。首先通过并行计算通道注意力以及空间注意力,引入自适应融合策略,充分利用两种注意力机制的优势,避免传统串联方式中的信息衰减和误差累计,同时结合了模糊神经网络中隶属函数层和规则生成层,增强对数据中的不确定性和关系复杂性的建模能力。提升模型的特征表示能力和整体性能机制。在优化的过程中,采用了灰狼优化算法作为超参数优化工具,通过搜索超参数的最佳组合,自动优化模型性能,使模型在训练过程更好的拟合数据特征。改进后的深度神经模糊神经网络提高了模型的预测精度,增强了模型的鲁棒性与可解释性,保证了模型可以更加适应现实生产环境。

1.1 改进灰狼优化算法

原始灰狼优化算法存在初始化具有随机性的问题,容易导致初始种群分布不均匀及权重因子固定造成算法易陷入局部最优的问题。因此,本研究提出了一种改进的灰狼

优化算法(dynamic condition logistic grey wolf optimizer, DL-GWO),通过引入混沌初始化种群与动态权重因子,有效地提升算法的全局搜索能力与收敛性能。

为增强种群的初期探索能力,本研究采用 Logistic 混沌映射初始化种群个体位置。混沌映射能够产生遍历搜索空间且敏感于初值的非线性序列,种群初始化公式为:

$$x_{n+1} = \mu \cdot x_n(1 - x_n) \quad (1)$$

$$x_i = LB + x_n \cdot (UB - LB) \quad (2)$$

其中, $\mu = 4.0$ 为混沌参数, LU 与 UB 分别表示变量的上下界。通过这种混沌映射方法,初始个体位置能够均匀遍历搜索空间,显著改善算法对全局最优解的探索能力。并且针对灰狼算法中 α, β, δ 这 3 只头狼位置更新权重固定的问题,本研究还提出了一种动态权重因子策略,根据头狼适应度实时调整各头狼对种群的引导贡献。权重计算公式为:

$$W_\alpha = \frac{\frac{1}{f_\alpha}}{\frac{1}{f_\alpha} + \frac{1}{f_\beta} + \frac{1}{f_\delta}} \quad (3)$$

$$W_\beta = \frac{\frac{1}{f_\beta}}{\frac{1}{f_\alpha} + \frac{1}{f_\beta} + \frac{1}{f_\delta}} \quad (4)$$

$$W_\delta = \frac{\frac{1}{f_\delta}}{\frac{1}{f_\alpha} + \frac{1}{f_\beta} + \frac{1}{f_\delta}} \quad (5)$$

将上述动态权重因子用于位置更新,改进后的位置更新公式表示为:

$$x_i = W_\alpha X_\alpha + W_\beta X_\beta + W_\delta X_\delta \quad (6)$$

相比传统等权更新方式,本方法能够更好地协调各头狼位置的优势信息,优化搜索方向并提高整体算法性能。

1.2 CM-CBAM 模块

为进一步提升模型对关键特征的感知能力,并充分挖掘不同尺度下的信息表示,本文在传统注意力机制 CBAM 的基础上,设计了一种多尺度通道-特征注意力融合模块(channel multiscale convolutional block attention module, CM-CBAM),用于一维特征向量的重要性加权,如图 1 所示。首先,在通道注意力分中,给定输入特征矩阵 $\mathbf{X} \in \mathbf{R}^{N \times C}$,其中 N 为批处理大小, C 为输入特征的维度,通过对特征维度执行全局平均池和全局最大池操作,分别获取两个全局特征向量:

$$\mathbf{X}_{\text{avg}} = \text{Avg pool}(\mathbf{x}) \in \mathbf{R}^{N \times C \times 1} \quad (7)$$

$$\mathbf{X}_{\text{max}} = \text{Max pool}(\mathbf{x}) \in \mathbf{R}^{N \times C \times 1} \quad (8)$$

随后,利用共享的多层感知分别处理这两个向量,并对输出进行逐元素求和后,再经 Sigmoid 函数获得通道注意力权重 $\mathbf{M}_c(\mathbf{X})$:

$$\mathbf{M}_c(\mathbf{X}) = \sigma(\text{MLP}(\mathbf{X}_{\text{avg}}) + \text{MLP}(\mathbf{X}_{\text{max}})) \quad (9)$$

其中, Sigmoid 函数定义为 $\sigma(x) = \frac{1}{1 + e^{-x}}$ 。最后,原

始输入特征通过该注意力权重进行加权,以强调对预测任务有显著贡献的特征通道:

$$\mathbf{X}_{\text{channel}} = \mathbf{X} \otimes \mathbf{M}_c(\mathbf{X}) \quad (10)$$

其次,考虑到传统 CBAM 模块中空间注意力机制无法直接用于无空间结构的数据,因此本研究提出了特征维度多尺度注意力机制。在烧结工艺监测过程中,各类工艺参数的变化频率存在显著差异。例如,原料组分在配料阶段呈现中等频率扰动,料层温度则表现为较长周期的缓变趋势,而震动、风速等干扰信号则常伴随高频波动。针对上述多尺度动态特性,本文在特征注意力机制中引入核大小为 ($k=3, 5, 7$) 的多尺度一维卷积结构,以适应不同尺度下的特征响应。该设计有助于模型从高频扰动、过程中频变化到低频趋势演化等多个层面挖掘关键特征,提升对复杂冶金状态的建模能力。首先对加权后的通道特征 $\mathbf{x}_{\text{channel}} \in \mathbf{R}^{N \times C}$ 进行特征级统计:

$$\mathbf{X}_{\text{global_avg}} = \text{Mean}(\mathbf{X}_{\text{channel}}, \text{dim} = 1) \in \mathbf{R}^{N \times 1} \quad (11)$$

$$\mathbf{X}_{\text{global_max}} = \text{Max}(\mathbf{X}_{\text{channel}}, \text{dim} = 1) \in \mathbf{R}^{N \times 1} \quad (12)$$

接着,将上述两个统计特征拼接为二维张量,引入多个一维卷积核 ($k=3, 5, 7$),分别对上述拼接特征进行并行卷积操作,得到多个尺度的特征注意力响应:

$$\mathbf{T}_k = \text{Conv1D}_k(\mathbf{f}_T), k \in \{3, 5, 7\} \quad (13)$$

将每个尺度卷积输出堆叠后,通过可学习的融合权重 $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ 进行 Softmax 归一化,并加权融合得到最终的特征注意力向量:

$$\mathbf{W}_T = \sigma\left(\sum_k \hat{\alpha} \times \mathbf{T}_k\right), \hat{\alpha}_k = \frac{e^{\alpha_k}}{\sum_j e^{\alpha_j}} \quad (14)$$

将该注意力向量作用于通道加权特征,最终输出经多尺度注意力增强的特征向量为:

$$\mathbf{X}_{\text{final}} = \mathbf{X}_{\text{channel}} \odot \mathbf{W}_T \quad (15)$$

1.3 GW-Attention-DFNN 模型构建

本文提出的预测模型由数据预处理层、模糊逻辑分支、深度学习分支(含 CM-CBAM 注意力机制)和融合预测层 4 个主要部分组成,每一部分的具体功能如下:

输入与数据预处理的主要功能是从原始数据集中导入数据,并进行归一化处理,以确保各个输入特征的数值处于相同的数量级,从而提高模型训练稳定性和收敛速度。同时,将数据随机分割为训练集和测试集,训练集用于模型训练,测试集用于模型性能验证。模糊推理分支主要利用模糊逻辑理论捕捉输入特征的模糊性和不确定性信息,该层包含高斯隶属度函数模块,通过将输入数据映射到模糊空间,并通过高斯函数定义隶属度公式为:

$$\mu(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \quad (16)$$

其中, $\mu(x)$ 表示模糊隶属度, x 为输入特征, c, σ 为模糊高斯函数的中心与宽度参数,通过神经网络训练自适应进行调整。模糊规则生成与解模糊化接收上一层的模糊隶属度,生成模糊规则的输出,并进行解模糊化处理,首先利

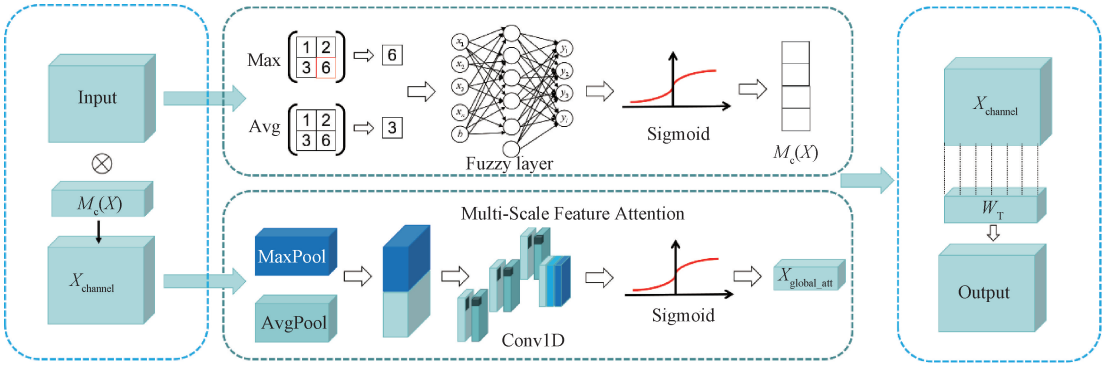


图 1 CM-CBAM 模块

Fig. 1 Improvement of the cbam module

用隶属度输出生成模糊规则:

$$\text{RuleOutputs} = \text{MembershipOutputs} \times \text{Weights} \quad (17)$$

在模糊规则生成过程中,为防止规则数量过多导致计算冗余,本文采用如下设定策略:规则数量设定:每个输入变量分配 m 个高斯隶属度函数(本文设 $m=3$)若输入特征维度为 d ,则所有可能的一阶组合数为:规则数 $\approx m \times d$ 若考虑两两交互组合(二阶组合),总数为:规则数 $\approx m^2 \times d \times (d-1)/2$ 实际使用中根据任务复杂度与资源限制,通过经验设定或网格搜索,控制总规则数不超过几百条。规则权重初始化方式:所有模糊规则的权重参数 w 在模型训练前采用 Xavier 均匀分布进行初始化,其计算方式为:初始化范围 $\in [-\sqrt{\frac{6}{n_{in}+n_{out}}}, \sqrt{\frac{6}{n_{in}+n_{out}}}]$ 其中: n_{in} 表示上一层输入的维度, n_{out} 表示当前层输出的维度、权重参数(Weights)为可学习参数;规则输出表示每个输入数据在各模糊规则下的激活程度,然后将规则输出通过全连接层(线性映射)进行“解模糊化”,获得明确的特征表示:

$$\mathbf{Y}_{defuzz} = \mathbf{W} \cdot \text{Flatten}(\text{RuleOutput}) + \mathbf{b} \quad (18)$$

为了提高模型效率并减少无效规则的干扰,本文设计了如下剪枝机制:在训练过程中,统计每条规则的平均激活程度和平均权重绝对值,分别记作 avg_activation 和 avg_weight 当某条规则同时满足 $\text{avg_activation} < 0.01$ 和 $\text{avg_weight} < 0.02$ 两个条件时,判定为无效规则,可从模型中删除;剪枝过程每隔 100 轮训练执行一次。

CM-CBAM 注意力机是原始 CBAM(卷积注意力模块)的改进版本,适用于处理表格数据(非时序),包含两个并行分支:通道注意力分通过全局平均池化和全局最大池化,分别捕获输入特征的平均与极值信息;再通过两层多层感知器,结合 Sigmoid 函数学习出各个特征的重要性权重。公式可表示为:

$$\mathbf{M}_c(\mathbf{x}) = \sigma(\text{MLP}(\text{Avgpool}(\mathbf{x})) + \text{MLP}(\text{Maxpool}(\mathbf{x}))) \quad (19)$$

多尺度特征注意力分为进一步建模特征维度间的全局交互关系,引入多尺度特征注意力机制。在通道加权后

的特征基础上,首先计算特征在样本维度上的全局平均值与最大值,并拼接形成统计向量:

$$\mathbf{f}_{feat} = \text{Concat}[\text{Mean}(\mathbf{X}_{channel}, \text{dim} = 1), \text{Max}(\mathbf{X}_{channel}, \text{dim} = 1)] \quad (20)$$

将该特征统计张量输入多个并行的一维卷积通道,卷积核尺寸分别设为 $k=3, 5, 7$ 以提取不同尺度下的局部与全局响应特征。每个尺度的卷积输出通过融合权重 $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ 进行归一化加权整合,生成最终的特征注意力权重 $\mathbf{W}_T \in \mathbf{R}^{N \times 1}$, 公式为:

$$\mathbf{W}_T = \sigma\left(\sum_k \hat{\alpha} \cdot \text{Conv1D}_k(\mathbf{f}_{feat})\right) \quad (21)$$

通道注意力与特征注意力所生成的权重分别作用于输入特征的两个维度,形成双重加权机制:

$$\mathbf{X}_{final} = \mathbf{X}_{channel} \odot \mathbf{W}_T \quad (22)$$

在多尺度注意力增强模块之后,本文采用深度全连接网络作为主干提取器,对表征后的特征进行进一步建模。该分支包含多个线性变换,每一层后接 ReLU 激活函数,对于第 l 层的输入特征,通过如下流程处理:

$$\mathbf{H}^{(l+1)} = \text{ReLU}(\text{Linear}(\text{MultiScale-CBAM}(\mathbf{H}^{(l)}))) \quad (23)$$

模型的输出端包含模糊神经网络分支与上述深度注意力网络分支。在特征提取完成后,二者的输出向量通过拼接操作形成融合表征:

$$\mathbf{F}_{fusion} = \text{Concat}[\mathbf{F}_{fuzzy}, \mathbf{F}_{deep}] \quad (24)$$

融合特征输入至最终全连接层进行回归或分类预测:

$$\hat{y} = \text{Linear}(\mathbf{F}_{fusion}) \quad (25)$$

融合层充分发挥模糊逻辑的推理能力与深度神经网络的特征表示能力,实现更准确、稳健的预测结果。最后使用 DL-GWO 算法自动优化深度分支的隐藏层节点数以及学习率等关键超参数;初始化随机候选解种群,进行适应度计算;迭代更新种群,直至搜索到使得模型预测性能最优的一组超参数组合;有效避免人工调参带来的局限性和随意性,提高了模型的鲁棒性与泛化能力。图 2 所示为 GW-Attention-DFNN 模型整体流程图,该模型伪代码如算法 1 所示。

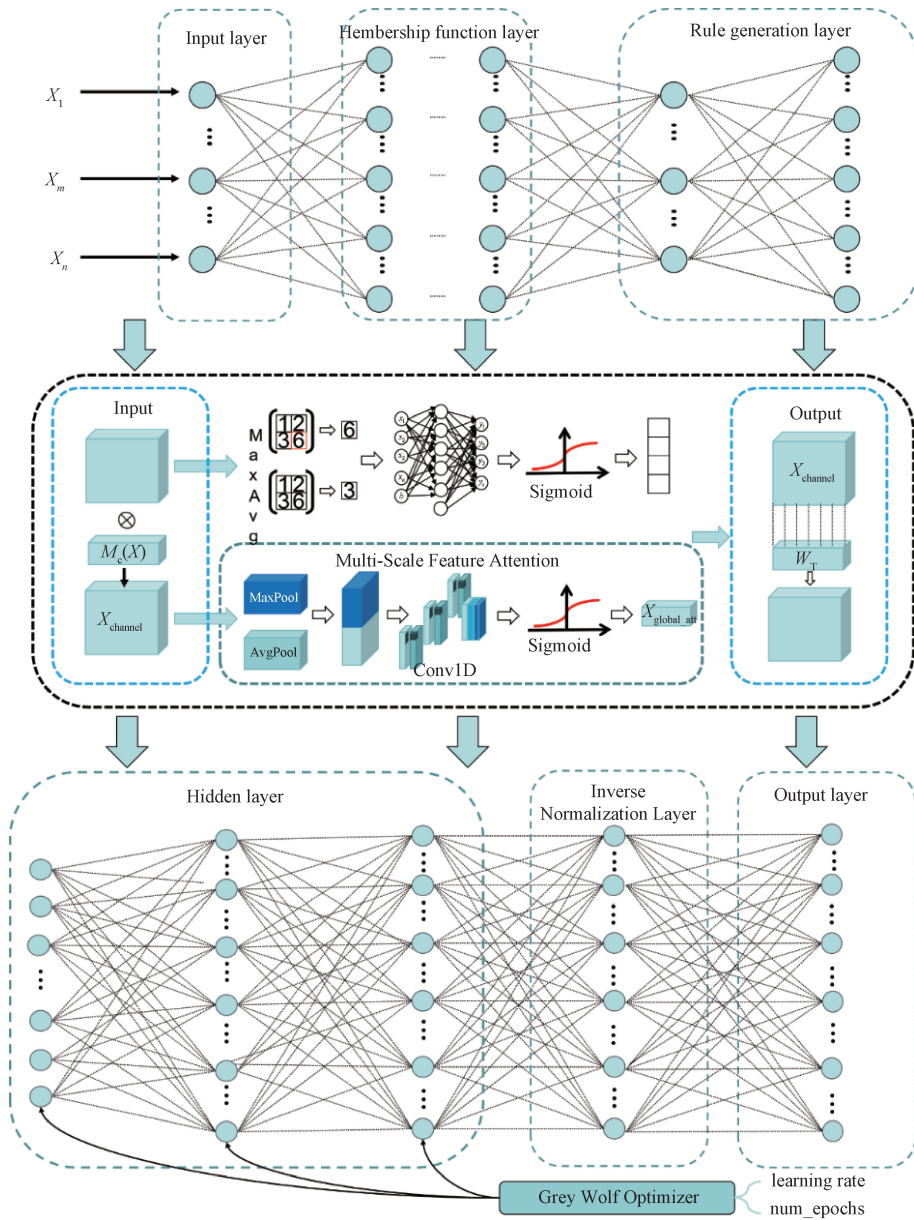


图 2 GW-Attention-DFNN 模型流程图

Fig. 2 Grey wolf optimizer-attention-deep feedforward neural network

2 实验及结果

2.1 数据预处理及样本集构建

1) 数据预处理

首先对数据中的缺失值进行处理,由于本文中的数据均来自实验数据,因此缺失值较少,所以采用直接删除的方式处理数据缺失值。然后对异常的数据进行处理。将所有数据点的核函数叠加,并对叠加结果进行归一化,形成一个平滑的密度估计。公式上,如果有数据点 x_1, x_2, \dots, x_n , 估计的密度函数可表示为:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (26)$$

其中, k 是核函数, h 是带宽,决定了每个核函数的宽度和光滑程度。

然后对数据进行标准化,并且缩放到某些相似的范围中,避免某些特征主导模型预测的结果,根据数据分布的特征,采用 Z-Score 的方法将数据进行标准化,公式为:

$$Z = \frac{(X_i - \mu)}{\sigma} \quad (27)$$

其中, μ 为每列特征的均值, σ 为每列特征的标准差, X_i 为原始的特征数据点的值, Z 则为归一化之后的结果。

2) 样本集构建

根据日常烧结生产数据的特点,为了提高预测模型的准确性和可解释性,可以选择适当的特征参数,在进行烧

算法 1: GW-Attention-DFNN 训练流程

输入: $X = \{X_1, X_2, \dots, X_n\}$ 结构化数据特征矩阵,

输出: GA-Attention-DFNN 最优参数 (最优 $hidden_sizes, learning_rate, num_epochs$)

1. 初始化变量: CM-CBAM 卷积核、规则数、隶属度函数、 $hidden_size, learning_rate, num_epochs$;
2. 初始化模糊规则分支 $FuzzyBranch$;
3. 计算高斯隶属度 $\mu(x)$;
4. 激活模糊规则: $R = \mu \times W_r$;
5. 展平后送入线性层, 输出 F_fuzzy ;
6. 初始化深度分支 $DeepFuzzyNN_CBAM_Branch$;
7. **for** $i=1$ to 3:
8. 全连接映射 $Z_i = ReLu$, 通道注意力计算 W_C ;
9. 特征维度注意力(多尺度 $Conv1D$):
10. 分别用卷积核 $k=3, 5, 7$ 处理 $Softmax$ 加权融合得到 W_T ;
11. 最终输出深度特征 F_deep ;
12. **End**
13. 拼接 F_fuzzy 与 F_deep , 形成融合特征 F_fusion ;
14. 输出层预测: $\hat{y} = Linear(F_fusion)$;
15. 使用灰狼优化算法(GWO)搜索最优超参数;
16. 设定搜索空间 $[H_1, H_2, H_3, \epsilon]$, 设适应度函数为验证集 MSE ;
17. 迭代更新群体位置, 输出最优结构参数组;
18. **End**
19. 根据模型评估指标评估模型预测性能

结矿质量预测时, 由于研究数据涵盖了较多维度, 其中包括混合料粒级及成分、烧结过程控制参数等, 为了提升模型的性能, 需要挑选关联性最强的特征进行预测, 本文采用随机森林的方法对特征数据进行相关度分析来确定预测模型最终的特征变量集合^[21-23]。为了更加便于绘图, 特征变量用 $C1 \sim C17$ 表示, 各预测指标的输入参数特征重要性如图 3(a)所示为烧结矿转鼓指数输入参数特征重要性排序; 图 3(b)为 $RDI_{+3.15}$ 输入参数特征重要性排序; 图 3(c)所示为烧结矿碱度输入参数特征重要性排序。

随机森林作为一种集成学习方法, 通过构建大量的决策树模型, 能够在面对复杂非线性关系和特征交互效应时, 有效地评估各输入变量对目标变量预测性能的贡献。通过对数据进行多层次分割, 能够自动捕捉输入特征之间的非线性关联和潜在的相互作用, 并且通过每个特征在分裂节点上降低不同纯度或误差的累计贡献, 为特征相关性提供了一种更加灵敏的和全面的度量手段, 从而能够更加有效的识别呈现非线性关系但对模型预测贡献显著的变量。在所有相关特征变量当中, 筛选后的各预测指标及其输入参数如表 1 所示。

为进一步提升模型建模效率并增强输入变量的可解释性, 本文基于随机森林算法对特征参数进行了重要性分析。通过统计每个输入特征在所有树模型中用于划分节点所带来的“信息增益”或“误差下降量”的累计贡献, 衡量其在整个预测过程中的相对重要性。为了筛选出对目标变量具有显著影响的关键特征, 本文设置特征重要性筛选阈值为 0.05, 即仅保留在随机森林模型中平均重要性权重大于 0.05 的特征变量作为后续建模输入。该阈值结合前

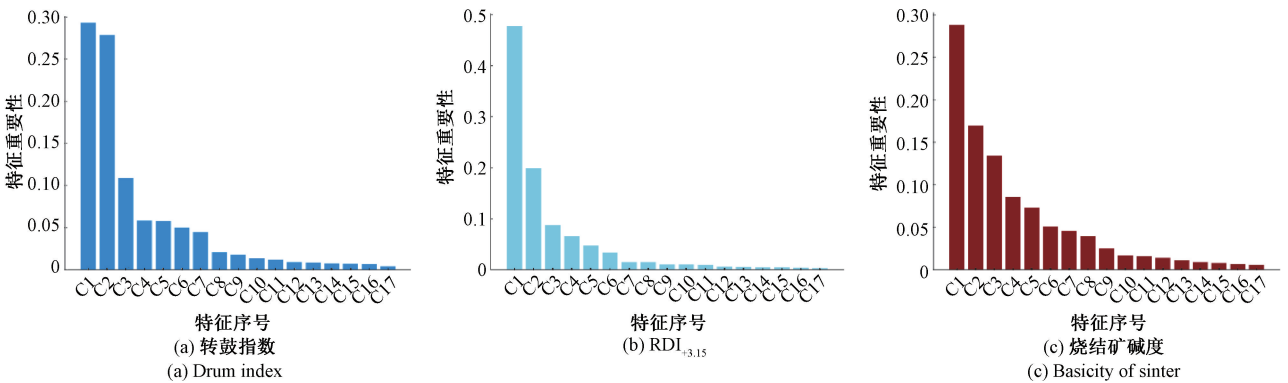


图 3 各指标相关性分析
Fig. 3 Correlation analysis of various indicators

期多轮训练经验设置, 可有效剔除噪声变量和冗余信息, 同时保留具备非线性建模价值的关键输入因素。

以转鼓指数预测任务为例, 经重要性评估后, 最终保留的主要输入变量包括加水率(%)、混匀料粒度 > 8 mm 占比(%)、点火负压、烧结负压、烧结时间、湿度、 SiO_2 平均粒径, 均对预测性能有显著正向贡献, 验证了所设阈值的有效性 with 合理性。

在数据划分上, 本文首先采用 8:2 的比例将数据集划分为训练集与测试集, 用于模型的初步训练与性能评估。为降低单次划分带来的偶然性偏差, 本文还引入 K 折交叉验证对模型进行综合评估, 将整体数据集随机划分为 K 个大小相近的子集, 其中 $K-1$ 个子集作为训练集, 其余 1 个子集作为验证集, 循环进行 K 次实验, 直至每个子集均作为一次验证集。最终以各折实验结果的平均值和方差作

表 1 预测指标及输入参数

Table 1 Prediction indicators and input parameters

预测指标	特征参数名称
转鼓指数	加水率%、混匀料粒度 > 8 mm 占比 (%)、点火负压、烧结负压、烧结时间、湿度、SiO ₂ 平均粒径
	TFe、水分、加水率%、MgO、Al ₂ O ₃ 平均粒径、烧结时间 min
RDI _{+3.15}	Cao、SiO ₂ 、Al ₂ O ₃ 点火负压 Kpa、烧结时间 min、点火温度℃

为模型的综合性能指标,从而更全面地反映模型在未知数据上的稳定性与泛化能力。

3) 评价指标

为了评价模型的预测能力,选取了 MAE、MSE、MAPE 和 R² 4 个预测模型评价指标来评估模型,预测模型评价指标的计算方式如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \quad (28)$$

$$MSE = \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N} \quad (29)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100\% \quad (30)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (31)$$

其中, N 为测试集中样本的数量, Y_i 为实际的输出, \hat{Y}_i 为该模型的预测输出值, \bar{Y} 表示将实际输出取平均值, MAE 表示为预测误差的平均绝对值, MSE 表示均方误差, MAPE 表示预测误差的平均百分比, MAE、MSE、MAPE 的值越小,则表示模型在总体上对于数据的拟合效果更为优秀,模型的预测误差越小,则表明模型泛化效果越好, R² 则是衡量回归模型拟合优度的指标,它反应了自变量对因变量的解释程度,取值范围在 [0, 1] 之间。越接近 1 则表示模型与数据拟合的越好。综合多个指标来评价模型的预测性能,可以更好的反应模型的泛化能力和预测效果。

2.2 实验结果与比较

本节综合比较了基于改进灰狼与注意力增强的深度神经网络(grey wolf optimizer fuzzy neural network, GW-FNN)、基于改进灰狼与注意力增强的深度模糊神经网络(grey wolf optimizer deep fuzzy neural network, GW-DFNN)、注意力增强的深度模糊神经网络(attention deep fuzzy neural network, Attention-DFNN)、GW-Attention-DFNN 4 种模型所得到转鼓指数、烧结矿碱度 (R)、RDI_{+3.15} 的预测情况。模型对比实验:首先选取 GW-FNN、GW-DFNN、Attention-DFNN、GW-Attention-DFNN 这 4 个模型,针对烧结矿质量中的转鼓指数进行了预测,并分别采用 MAE、MSE、MAPE 和 R²; 4 个模型评价指标对模型的预测性能进行评价,其中 GW-FNN 为初始模型, GW-DFNN、Attention-DFNN、GW-Attention-DFNN 为本文所提出预测模型,评价结果如表 2 所示。

表 2 各预测模型评价指标

Table 2 Evaluation indicators of each prediction model

预测指标	模型	MSE	MAE	MAPE	R ²
转鼓指数	GW-FNN	0.175 5	0.106 5	0.161 2	0.852 6
	GW-DFNN	0.141 1	0.092 3	0.139 6	0.926 0
	Attention-DFNN	0.073 4	0.091 6	0.138 5	0.946 6
	GW-Attention-DFNN	0.011 8	0.085 7	0.129 8	0.968 2
烧结矿碱度 (R)	GW-FNN	0.119 4	0.869 0	0.488 1	0.845 4
	GW-DFNN	0.107 3	0.677 8	0.380 6	0.922 3
	Attention-DFNN	0.372 0	0.453 1	0.249 5	0.956 1
	GW-Attention-DFNN	0.127 5	0.278 7	0.154 2	0.975 0
RDI _{+3.15}	GW-FNN	0.454 0	0.167 0	0.222 7	0.839 7
	GW-DFNN	0.426 1	0.164 2	0.218 8	0.900 9
	Attention-DFNN	0.416 2	0.156 9	0.209 0	0.951 4
	GW-Attention-DFNN	0.353 2	0.150 2	0.200 0	0.964 2

从表 2 可以看出, GW-Attention-DFNN 模型在预测烧结矿转鼓指数、烧结矿碱度 (R) 和低温还原粉化指数

RDI_{+3.15} 方面均表现出了其良好的性能。具体来说, 本文所设计模型即 GW-Attention-DFNN 模型在转鼓指数预测

方面 MSE 为 0.011 8、MAE 为 0.085 7、MAPE 为 0.129 8、 R^2 的值为 0.968 2；在烧结矿碱度(R)预测方面 MSE 为 0.127 5、MAE 为 0.278 7、MAPE 为 0.154 2、 R^2 的值为 0.975 0 在 $RDI_{+3.15}$ 预测方面 MSE 为 0.353 2、MAE 为 0.150 2、MAPE 为 0.200 0、 R^2 的值为 0.964 2。

为避免单次 80/20 随机划分导致的结果波动,并更全面地验证模型在未知数据上的稳定性,本文进一步引入五折交叉验证,并在表 3 中给出了各模型在不同指标下的平

均性能表现。结果显示,所有模型在五折中的表现差异较小,方差水平有限,证明了模型在不同训练/测试集划分下的鲁棒性。其中,GW-Attention-DFNN 在转鼓指数、烧结矿碱度和 $RDI_{+3.15}$ 3 个目标上均表现出较低的 MSE 与 MAE 以及较高的 R^2 ,进一步验证了该方法在泛化能力上的优势。这说明本文提出的融合模型不仅在单次划分下取得了优异精度,而且在多折验证下依然保持稳定性能,具备较强的工业应用可靠性。

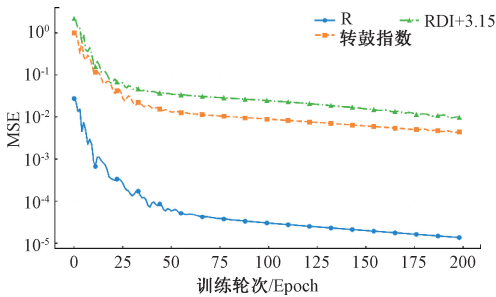
表 3 五折交叉验证下各预测模型评价指标

Table 3 Evaluation indicators of each prediction model under five-fold cross-validation

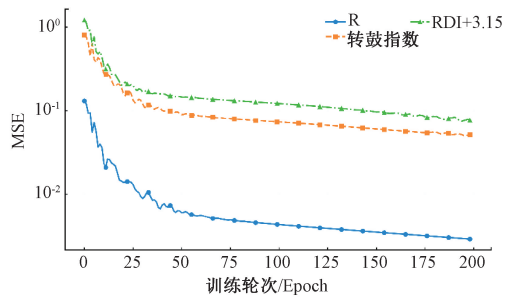
预测指标	模型	MSE	MAE	MAPE	R^2
转鼓指数	GW-FNN	0.184 4	0.107 5	0.162 7	0.928 8
	GW-DFNN	0.141 2	0.093 3	0.141 3	0.985 7
	Attention-DFNN	0.043 0	0.092 8	0.138 8	0.985 8
	GW-Attention-DFNN	0.012 9	0.091 3	0.132 2	0.984 9
烧结矿碱度(R)	GW-FNN	0.117 7	0.855 8	0.481 8	0.895 6
	GW-DFNN	0.144 1	0.503 9	0.282 8	0.908 4
	Attention-DFNN	0.125 3	0.266 4	0.147 5	0.981 4
	GW-Attention-DFNN	0.110 1	0.244 0	0.136 7	0.966 1
$RDI_{+3.15}$	GW-FNN	0.481 2	0.172 2	0.229 4	0.978 1
	GW-DFNN	0.410 5	0.159 2	0.212 0	0.988 3
	Attention-DFNN	0.399 1	0.157 3	0.209 4	0.982 4
	GW-Attention-DFNN	0.394 1	0.150 1	0.200 9	0.981 8

如图 4(a)所示表示了训练过程当中 MSE 的收敛曲线,图 4(b)展现了训练过程当中 MAE 的收敛曲线,图 4(c)

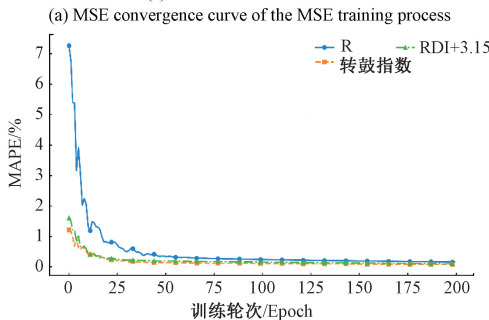
和(d)分别展现了模型训练后 MAPE 和 R^2 的训练过程收敛曲线,各种结果明显表明了该模型在数据预测方面



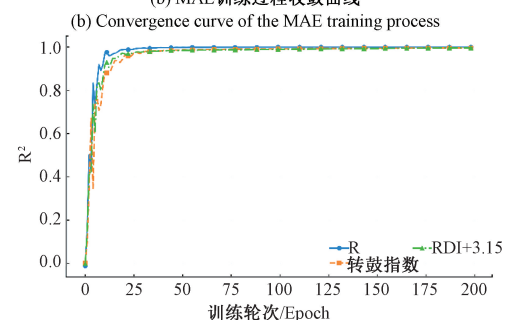
(a) MSE训练过程收敛曲线



(b) MAE训练过程收敛曲线



(c) MAPE训练过程收敛曲线



(d) R^2 训练过程收敛曲线

(c) Convergence curve of the MAPE training process

(d) Convergence curve of the R^2 training process

图 4 模型训练过程各评价指标

Fig. 4 Evaluation indicators of each stage in model training process

具有良好的性能,综合说明了该模型在针对烧结矿质量预测能够具有广泛适用性和具有良好的性能,在转鼓指数、烧结矿碱度(R)、 $RDI_{+3.15}$ 的预测出中均展现出了良好的性能和显著的优势。

表 4 中各预测模型的隐藏层神经元数量由 DL-GWO 自动搜索获得。为保证搜索空间既具备足够表达能力,又避免搜索维度过大造成计算负担,本文依据以下原则设置各层神经元数量的搜索范围:第 1 隐藏层设置区间为[32, 512],考虑到输入特征维度较大;烧结数据包含多种化学组分、工艺参数,初始特征映射层需具备较强的表征能力,

但同时受限于计算资源:显存容量与训练耗时,故上限设为 512;第 2 隐藏层区间为[16, 256],用于压缩与抽象第 1 层输出的特征,考虑到中层特征需保留信息又不能过宽,依据经验设定该范围;第 3 隐藏层区间为[8, 128],作为最终的低维特征提取层,维度设定需兼顾模型表达能力与计算效率,避免过深网络导致梯度消散;输出层神经元数量视具体预测任务设定,通过 DL-GWO 联合优化。此外,搜索范围的设定也充分考虑了实际工业建模中对模型训练效率、推理速度的限制,确保深度神经网络在标准 GPU 平台上训练时间控制在合理范围内。

表 4 模型训练过程参数

Table 4 Parameters of model training process

参数指标	转鼓指数	烧结矿碱(R)	$RDI_{+3.15}$
Hidden_sizes	[480, 118, 70]	[509, 224, 20]	[512, 256, 112]
Num_epochs	2 329	884	1 832
Learning rate	0.005 290 433	0.001 609 288	0.002 499 480

综上所述,各层神经元搜索范围结合了数据维度特性、经验建模规律与计算资源约束,具有合理性与适应性。

如图 5(a)和(b)所示,分别展示了该模型在预测 $RDI_{+3.15}$ 训练集和测试集时的表现,从图中可以看出, $RDI_{+3.15}$ 的模型预测结果和真实值之间取得了较好的结果。模型的预测结果与真实值之间展现了较好的拟合效果。通过对比预测值与实际值,模型在训练过程中能够有

效捕捉到目标变量 R 的变化趋势,并对输入特征进行了有效的建模。评估指标如均方误差(MSE)为 0.127 5、 R^2 为 0.975 0、绝对误差(MAE)为 0.278 7 各种结果均表明,模型在准确性和稳定性方面都取得了令人满意的表现。特别是在预测过程中,模型的输出与实际观察值之间的差异较小,且预测值分布紧密贴合真实值,进一步表明模型能够很好地解释目标变量的方差,预测结果与实际情况高度一致。

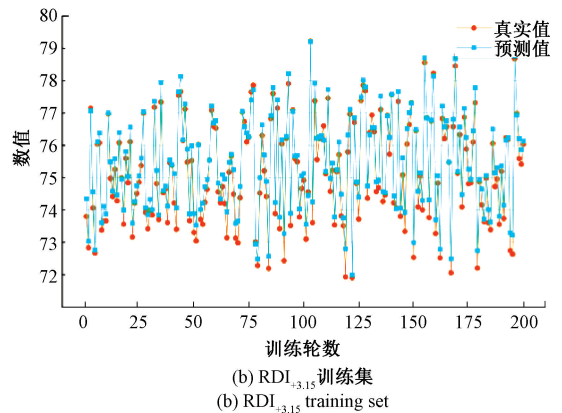
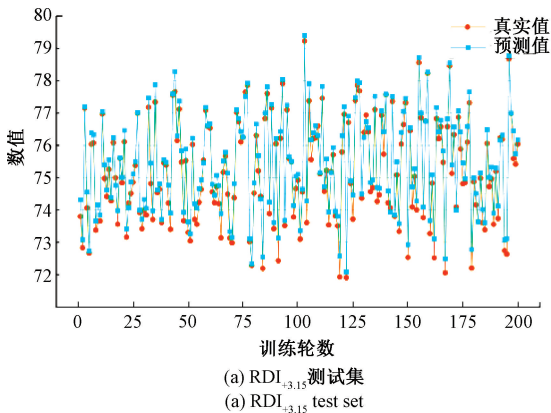


图 5 真实值与预测值指数比较

Fig. 5 Comparison of the index between the actual value and the predicted value

如图 6(a)和(b)所示可以看到烧结矿碱度(R)的真实值与预测值之间的对比,图像清晰地展示了模型在实际应用中的出色表现。图像中真实值与预测值的分布几乎重合,表明模型能够精确地预测烧结矿的碱度值,并且与实际生产数据高度一致。这种高一致性反映了模型在捕捉烧结矿质量特征方面的强大能力,能够在多变的工业环境中准确预测关键质量指标。

间的对比图像有效地展示了模型在实际应用中的优异性能。从图像中可以看到,真实值与预测值之间呈现出较强的线性关系,表明模型在对转鼓指数的预测方面取得了很好的效果。从图中可以体现出该模型具有较强的应用价值,能够为烧结矿生产过程中参数调控提供科学依据,进而提高产品质量和生产效率。

2.3 运行效率与部署可行性分析

为评估所提方法在工业场景下的可部署性,本文进一

如图 7(a)和(b)所示,转鼓指数的真实值与预测值之

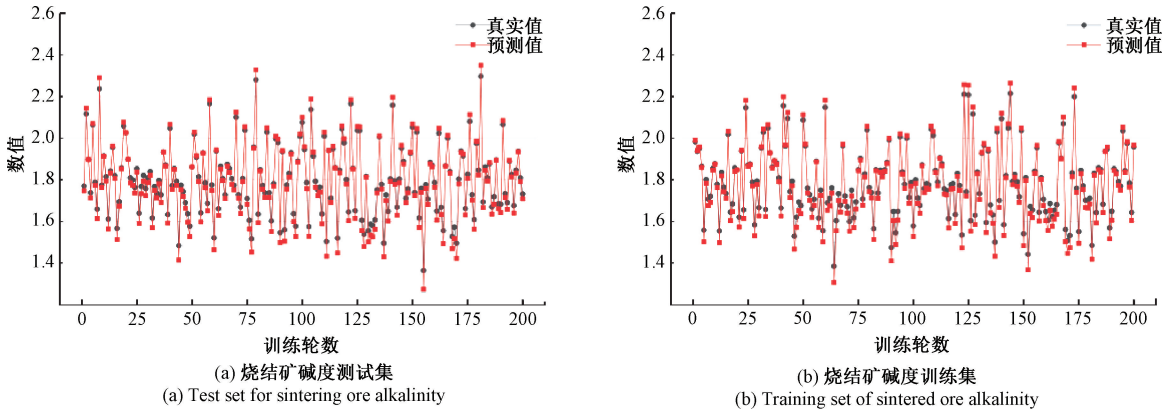


图 6 烧结矿碱度真实值与预测值对比

Fig. 6 Comparison chart of actual and predicted values of sinter alkalinity

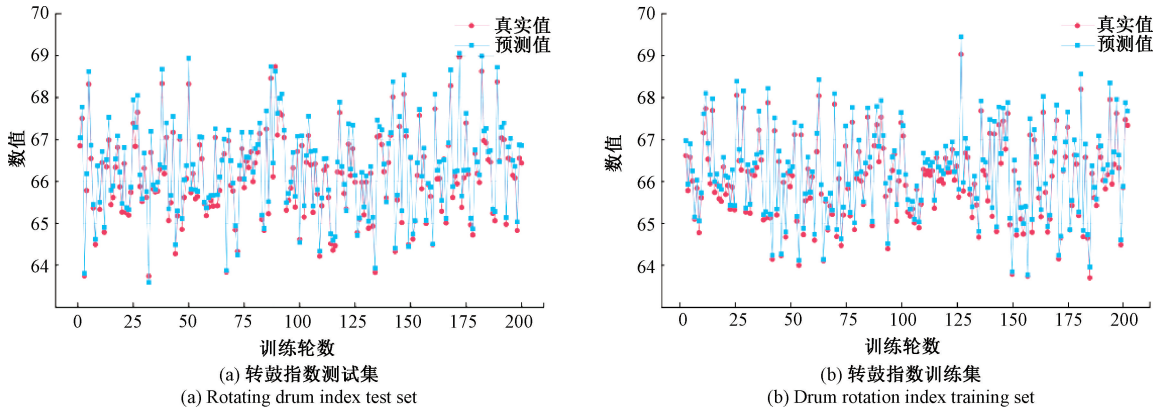


图 7 烧结矿转鼓指数真实值与预测值对比

Fig. 7 Comparison chart of actual and predicted values of sinter rotation drum index

步统计了模型在标准 GPU 平台 (NVIDIA GeForce RTX 4060 Laptop GPU, CUDA 12.1, PyTorch 2.3.0) 上的运行效率指标。实验结果表明:在 5 折交叉验证下,模型单折训练总耗时约 3.09 ± 0.23 s, 平均每个 epoch 的开销约 15.4 ms, 峰值显存占用仅 0.054 ± 0.006 GB。在推理阶段,单样本延迟约 2.37 ± 0.02 ms, 小批量 64 条样本的平均延迟降至 0.054 ± 0.027 ms, 对应吞吐率可达 2.17×10^4 样本/s, 推理时显存需求约 0.019 GB。此外,灰狼优化 (GWO) 在超参数搜索过程中表现出较高的收敛效率, 平均 20 次迭代即可稳定收敛, 单次迭代耗时 7.48 ± 0.33 s, 总搜索时间约 74.8 ± 3.3 s。

综合来看,该方法在保持高预测精度的同时,具备较低的训练与推理时延以及显存占用,且 DL-GWO 搜索过程收敛迅速,显示出良好的工程可行性与工业部署潜力。

3 结 论

本文通过建立模糊神经网络算法、深度神经网络、注意力机制和灰狼优化算法相结合,建立了基于烧结大数据的烧结矿质量预测模型 GW-Attention-DFNN 预测模型,

该模型在数据预处理和特征提取过程中引入了通道注意力机制,能够在每一隐藏层中自适应的调整各特征通道的重要性,从而更加精准的捕捉关键数据特征,同时模型利用深度网络对高维数据进行非线性建模,通过反向传播自动更新各层连接权值,显著的提升了模型在不同场景下的泛化能力,此外模型采用了可学习的隶属度函数,其质心和宽度参数在训练过程中在训练的过程中通过梯度下降的方式自适应的进行更新,有效的刻画了输入数据的模糊特性,实验结果表明,该模型在转鼓指数、烧结矿碱度 (R) 以及 $RDI_{+3.15}$ 的预测任务当中,分别达到了 R^2 为 0.968 2、0.975 0、0.964 2; MSE 为 0.011 8、0.127 5、0.353 2; MAPE 为 0.129 8、0.154 2、0.200 0 转鼓指数以及 $RDI_{+3.15}$ 的相对误差均控制在 1.05% 以内,其中转鼓指数的相对误差仅为 0.13%, 各种指标均表明了该模型的预测精度高,泛化能力强,当面对大量的复杂数据是可以表现出其优秀的预测能力,凭借着高预测精度和出色的泛化能力,该模型能够有效的应对烧结生产过程当中海量负责的数据所带来的挑战,通过实现对烧结质量的提前预测,为生产高品质的烧结矿提供技术支撑,同时降低环境污染并为企业创造经济效益。

参考文献

- [1] ZHOU H, ZHANG H F, YANG C J. Hybrid-model-based intelligent optimization of ironmaking process[J]. IEEE Transactions on Industrial Electronics, 2019, 67(3): 2469-2479.
- [2] CHEN X X, CHEN X, SHE J H, et al. A hybrid just-in-time soft sensor for carbon efficiency of iron ore sintering process based on feature extraction of cross-sectional frames at discharge end [J]. Journal of Process Control, 2017, 54: 14-24.
- [3] SHAO H, YI Z, CHEN Z, et al. Application of artificial neural networks for prediction of sinter quality based on process parameters control [J]. Transactions of the Institute of Measurement and Control, 2020, 42(3): 422-429.
- [4] LI Z P, FAN X H, CHEN G, et al. Optimization of iron ore sintering process based on ELM model and multi-criteria evaluation [J]. Neural Computing and Applications, 2017, 28(8): 2247-2253.
- [5] PAHLEVANINEZHAD M, EMAMI M D, PANJEPOUR M. The effects of kinetic parameters on combustion characteristics in a sintering bed [J]. Energy, 2014, 73: 160-176.
- [6] WANG S H, LI H F, ZHANG Y J, et al. A hybrid ensemble model based on ELM and improved AdaBoost. RT algorithm for predicting the iron ore sintering characters [J]. Computational Intelligence and Neuroscience, 2019, 2019(1): 4164296.
- [7] KUMAR V, PATIL D, MRUNMAYA K P, et al. Impact of sinter basicity and alumina on softening melting behavior in blast furnace [J]. Transactions of the Indian Institute of Metals, 2020, 73(2): 365-375.
- [8] LI Y F, ZHANG Q W, ZHU Y, et al. A model study on raw material chemical composition to predict sinter quality based on GA-RNN [J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 3343427.
- [9] 李泽政, 刘卫星, 李飞, 等. 基于数据增强的烧结矿转鼓强度预测研究 [J]. 烧结球团, 2023, 48(6): 62-68.
- [10] REN X Y, YANG B, LUO N, et al. The prediction of sinter drums strength using hybrid machine learning algorithms [J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 4790736.
- [11] 刘加达, 于帆, 温冶, 等. 基于 BPNN 和 RNN 模型的烧结矿质量预测方法对比及分析 [J]. 冶金自动化, 2020, 44(5): 20-26.
- [12] LIU J D, YU F, WEN ZH, et al. Comparison and analysis of sinter quality prediction methods based on BPNN and RNN models [J]. Metallurgical Automation, 2020, 44(5): 20-26.
- [13] GAO Q J, WANG H, PAN X Y, et al. A forecast model of the sinter tumble strength in iron ore fines sintering process [J]. Powder Technology, 2021, 390: 256-267.
- [14] 张乐文, 陈新兵, 侯东旭, 等. 基于加权支持向量机对烧结过程中 FeO 含量和转鼓指数的预测研究 [J]. 冶金动力, 2017(8): 1-3.
- [15] ZHANG L W, CHEN X B, HOU D X et al. Prediction of FeO content and drum index in sintering process based on weighted support vector machine [J]. Metallurgical Power, 2017(8): 1-3.
- [16] LIU S, LIU X, L Y Q, et al. Comprehensive system based on a DNN and LSTM for predicting sinter composition [J]. Applied Soft Computing, 2020, 95: 106574.
- [17] LI Y F, CAO Y P, YANG J T, et al. Optuna-DFNN: An optuna framework driven deep fuzzy neural network for predicting sintering performance in big data [J]. Alexandria Engineering Journal, 2024, 97: 100-113.
- [18] 王紫荆, 刘卫星, 杨爱民. 基于改进鲸鱼优化算法的工业 CT 图像增强方法 [J]. 电子测量技术, 2024, 47(11): 160-168.
- [19] WANG Z J, LIU W X, YANG AI M. Industrial CT image enhancement method based on improved whale optimization algorithm [J]. Electronic Measurement Technology, 2024, 47(11): 160-168.
- [20] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [21] DENG Y, REN Z Q, KONG Y Y et al. A hierarchical fused fuzzy deep neural network for data classification [J]. IEEE Transactions on Fuzzy Systems, 2016, 25(4): 1006-1012.
- [22] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer [J]. Advances in Engineering Software, 2014, 69: 46-61.
- [23] SZE V, CHEN Y H, YANG T J, et al. Efficient processing of deep neural networks: A tutorial and survey [J]. Proceedings of the IEEE, 2017, 105(12): 2295-2329.
- [24] SENTHILNATHAN S. Usefulness of correlation analysis [J]. Social Science Research Network, 2019, DOI:10.2139/ssrn.3416918.
- [25] FRANZESE M, IULIANO A. Correlation analysis [M]. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. Elsevier, 2018, 1:706-721.
- [26] GOGTAY N J, THATTE U M. Principles of correlation analysis [J]. Journal of the Association of Physicians of India, 2017, 65(3): 78-81.

作者简介

杨兴旺, 硕士研究生, 主要研究方向为数据挖掘、人工智能和烧结矿质量预测。

E-mail: yangxingwang@stu.ncst.edu.cn

杨爱民(通信作者), 博士, 教授, 博士生导师, 主要研究方向为计算机视觉、钢铁大数据分析和智能冶炼等。

E-mail: aimin@ncst.edu.cn

薛涛, 副教授, 硕士, 主要研究方向为机器学习、深度学习和烧结流程智能化。

E-mail: xuetao@ncst.edu.cn