

# 基于超参数优化的重质碳酸钙粉体粒度预测研究<sup>\*</sup>

黄 何<sup>1,2</sup> 邹 帅<sup>1,2</sup> 杨 靖<sup>1,2</sup> 黄福川<sup>1,2</sup>

(1. 广西大学机械工程学院 南宁 530004; 2. 广西石化资源加工及过程强化技术重点实验室 南宁 530004)

**摘要:** 重质碳酸钙粉磨系统中,粒度是衡量产品质量的关键指标之一,准确预测粒度有助于控制产品质量并指导立磨(VRM)进行参数调节。因此,该研究提出了一种基于常春藤算法(IVYA)的双向时间卷积网络(BiTCN)和双向门控循环单元(BiGRU)相结合的重钙粉体粒度预测模型。首先对特征和标签数据进行预处理,利用时变滤波经验模态分解联合小波阈值去除分级机电流中的高频噪声;然后通过BiTCN从前向两个方向挖掘时间序列中多维特征间的关联性,在BiGRU输出端融入注意力模块赋予每一个位置不同的权重,从而有效关注序列中的关键数据。其次,在整个模型上引入IVYA寻找神经网络中关键超参数的最优解。最后,以某碳酸钙粉磨工厂实测数据为例进行模型实验。实验表明,IVYA优化后的模型相比较于其他单一模型和组合模型具有更高的预测性能,其均方根误差、平均绝对误差、平均相对百分误差和决定系数分别为:0.824 4、0.423 0、1.295 4%、98.95%。

**关键词:** 双向时间卷积网络;双向门控循环单元;超参数优化;粒度预测

中图分类号: TP183; TN06 文献标识码: A 国家标准学科分类代码: 520.2060

## Research on the prediction of particle size of heavy calcium carbonate powder based on hyperparameter optimization

Huang He<sup>1,2</sup> Zou Shuai<sup>1,2</sup> Yang Jing<sup>1,2</sup> Huang Fuchuan<sup>1,2</sup>

(1. School of Mechanical Engineering, Guangxi University, Nanning 530004, China;

2. Guangxi Key Laboratory of Petrochemical Resource Processing and Process Intensification Technology, Nanning 530004, China)

**Abstract:** In the grinding system of heavy calcium carbonate, particle size was esteemed as a crucial metric for assessing product quality. The precise prediction of particle size was deemed instrumental in controlling product quality and informing the adjustments of vertical roller mill (VRM) parameters. Therefore, the study proposed a model for predicting the particle size of heavy calcium powder, which combined the bidirectional time convolutional network (BiTCN) based on the ivy algorithm (IVYA) with the bidirectional gated recurrent unit (BiGRU). Initially, feature and label data were subjected to preprocessing, employing time-varying filtering empirical mode decomposition in conjunction with wavelet thresholding to eliminate high-frequency noise from the primary current of the vertical mill. Then, the correlations between multidimensional features in the time series were explored from both forward and backward directions through BiTCN. At the output end of BiGRU, an attention module was incorporated to assign different weights to each position, thereby effectively focusing on the key data in the sequence. Finally, the IVYA was integrated into the overall model to ascertain the optimal solutions for critical hyperparameters within the neural network. Actual measured data from a carbonate powder mill factory were subsequently employed to simulate particle size prediction. The experimentation indicates that the optimized model using the IVYA exhibits superior predictive performance compared to other single and combined models. Specifically, it achieves the following metrics: root mean square error of 0.824 4, mean absolute error of 0.423 0, mean relative percentage error of 1.295 4%, and coefficient of determination of 98.95%.

**Keywords:** bidirectional temporal convolutional network; bidirectional gated recurrent unit; hyperparameter optimization; particle size prediction

## 0 引言

重质碳酸钙(简称重钙)产品的粒度是衡量粉磨过程的

关键指标。有效的粒度预测可以更精确地调整和控制粉磨过程中的运行参数,确保产品达到所需的质量标准。随着技术的进步,市场对重质碳酸钙产品粒度的要求逐渐趋向

于更加细致和精确<sup>[1]</sup>。因此,在重钙生产阶段需要根据市场需求不断的调整生产设备参数。广泛用于重质碳酸钙粉体加工的先进设备主要是立式辊磨机,该设备集破碎、粉磨、烘干、分级选粉、气力输送于一体,因而具备较高的生产效率<sup>[2]</sup>。目前,大多数重钙生产企业仍然采用人工离线化验的方法对重钙粒度进行检测,但这种方法存在重钙粒度检测滞后性大的问题,导致在没有实时检测的情况下难以进行闭环优化控制<sup>[3]</sup>。加之立磨粉磨过程是一个复杂的动态系统,涉及多个物理和化学交互过程,具有非线性、变时延和强耦合的特点。因此,采用合适的方法实现重钙粉体粒度的精确预测尤为必要。

重钙立磨系统运行参数的变化是一个典型的时间序列问题。传统的参数模型和浅层神经网络算法在处理时间序列数据时已难以达到高效准确的水平,因此使用深度学习算法来挖掘时间序列中的有效信息成为许多学者关注的重点<sup>[4]</sup>。深度学习是基于数据驱动的预测方法,具有很好的非线性拟合性能和自主学习特征表示的能力,在流程工业领域已经有广泛的应用。Liu 等<sup>[5]</sup>提出了一种二维卷积神经网络(2D-convolutional neural network, 2D-CNN),用于预测立磨研磨过程中的生料细度。Bao 等<sup>[6]</sup>提出一种基于DPCA 相似性的时间序列分割方法,并通过支持向量机进行生料细度预测。

循环神经网络(recurrent neural network, RNN)常用于处理时间序列问题,但其同 CNN 一样,在网络架构存在很大的局限性。传统的 CNN 和 RNN 在反向更新权重过程中,会存在梯度消失和梯度爆炸的情况。故在 CNN 和 RNN 的基础上分别衍生出了时间卷积网络(temporal convolutional network, TCN)和门控循环单元(gated recurrent unit, GRU)。TCN 模型<sup>[7]</sup>凭借其独特的残差块和因果膨胀卷积结构,可以有效的解决上述局限性。郝椿淋等<sup>[8]</sup>利用 TCN 模型与其他模型组合对交通流进行预测,实验证明 TCN 模型比传统 CNN 和 RNN 模型具有更好的预测性能。GRU 模型由 Cho 等<sup>[9]</sup>于 2014 年提出,目的是简化长短期记忆网络(long short-term memory, LSTM)的网络结构,但仍保持其核心机制。GRU 模型中每个循环单元能够自适应地捕捉不同时间尺度的依赖关系,这使它广泛应用于处理时间序列数据。Cahuantzi 等<sup>[10]</sup>论证了 GRU 与其他递归单元在序列建模上的优越性。Tan 等<sup>[11]</sup>提出了 GRU 结合 TCN 模型对轴承剩余使用寿命进行预测,在对特征充分提取之后再进行时间序列的预测。

深度学习模型趋向于更深更宽层的复杂方向发展,而具有复杂结构的模型意味着需要调整更多的超参数。对于广泛使用的简单模型,研究人员可以基于经验和反复实验手动调整其中的超参数,这种方法依赖于对模型行为的直觉理解和实验结果的反馈,并且经验通常提供的是可行的超参数集,而不是最佳的超参数集。对于超参数间相互关联的复杂模型,涉及的超参数组合众多,而且超参数调节需

要大量的试错过程,非常消耗时间和计算资源,手动优化将会不切实际<sup>[12]</sup>。Hanifi 等<sup>[13]</sup>采用 3 种先进优化算法对 CNN 和 LSTM 模型的超参数进行优化,证明优化算法调参具有更高的准确性。韩素敏等<sup>[14]</sup>通过引入改进的蜣螂搜索算法对双向 LSTM 模型超参数进行优化,有效地提高逆变器中 IGBT 模块的老化过程的预测精度。

然而,上述的深度神经网络预测模型在网络结构优化和超参数调整方面仍有改进空间,普遍存在输入特征提取不充分、长时间依赖特性捕获不足以及超参数设置不合理等问题。故本文将改进的 BiTCN 模型与 BiGRU 结合搭建训练网络。将从 BiTCN 层中提取出的特征信息输入到 BiGRU 层进行双向训练。训练模型就是不断的更新模型中的权重和偏置,将注意力机制加入到 BiGRU 网络层中可以有效地调整时间序列中各时间节点的权重。最后,采用 IVYA 对其进行关键超参数寻优,提高模型的预测精度和收敛速度。

## 1 立磨重钙粉磨工艺

本文结合立式辊磨机(简称立磨)介绍其生产工艺流程。立磨粉磨系统主要由进料、破碎、研磨、分级、输送、集尘和控制系统等组成,其中研磨系统是其核心组件,包括磨盘、磨辊、驱动装置、磨辊的加压装置等部件,其工艺流程如图 1 所示。

重钙原料由皮带输送机和螺旋机定量喂入立磨主机,落至磨盘中央。立磨主电机通过减速机带动磨盘转动,磨辊借助与磨盘的摩擦力自转。在离心力的作用下,物料从磨盘中央向边缘均匀移动,经过磨盘上的辊道区域时,大块物料由磨辊碾压粉碎,而细粒物料则在挤压后形成料层,进行粒间粉磨。粉磨后的物料继续向磨盘边缘移动,大颗粒物料经外循环输送斗提再次与原料回磨,较细物料则随热气流上升至磨盘上方的分级机,边烘干边分级。在分级机转子叶片的作用下,较粗的颗粒被叶片打落,重新进入磨盘内进行二次粉磨,而合格的细粉则随负压气流进入除尘设备,经过集粉器收集成为成品。其余大颗粒物料和杂物通过磨机下腔排渣口排出,以确保成品质量。

## 2 IVY 优化算法

常春藤算法(IVY algorithm, IVYA)是由 Ghasemi 等<sup>[15]</sup>于 2024 年发表的一种新型仿生元启发式算法。该算法以常春藤为研究对象,通过模拟它在大自然中的生长过程来建立优化算法。

### 2.1 种群初始化

在算法运行时,IVYA 种群在搜索空间中的初始位置使用式(1)随机确定。

$$I_i = I_{\min} + \text{rand}(1, D) \odot (I_{\max} - I_{\min}), i = 1, \dots, N_{pop} \quad (1)$$

式中:  $N_{pop}$  和  $D$  分别为总体种群数量和问题的决策参数

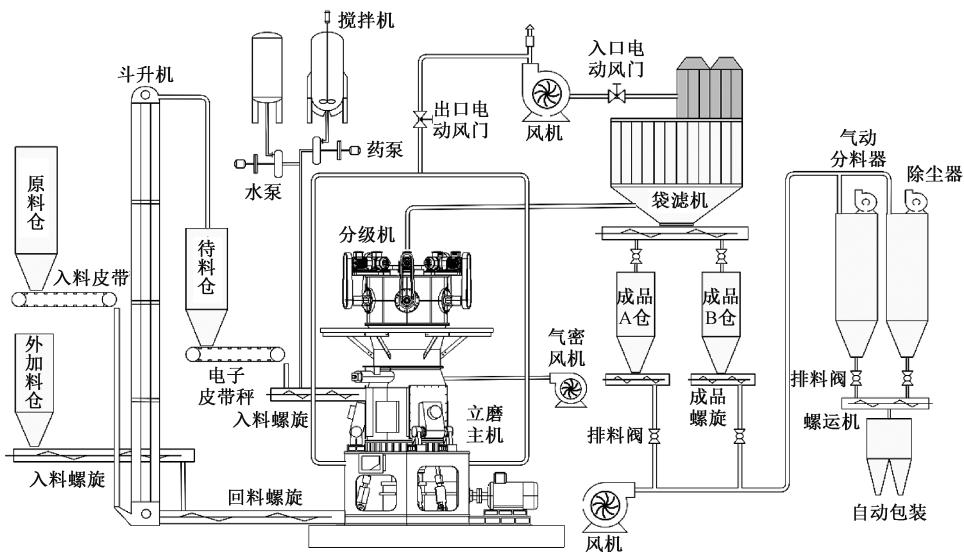


图 1 重钙粉磨工艺流程图

Fig. 1 Flowchart of the heavy calcium powder grinding process

的数量。 $rand(1, D)$  表示一个维度为  $D$  的随机向量, 其中每个分量都是从区间  $[0, 1]$  均匀分布的随机数。 $I_{\max}$  和  $I_{\min}$  分别为搜索空间的上界和下界, 运算符“ $\odot$ ”表示两个向量的 Hadamard 乘积。

## 2.2 协调有序的种群生长

假设常春藤植物的生长速率  $Gv$  是时间的函数, 则其微分方程为:

$$\frac{dGv(t)}{dt} = \psi \cdot Gv(t) \cdot \varphi(Gv(t)) \quad (2)$$

式中:  $\psi$  和  $\varphi$  分别表示生长速度、指示生长偏差的修正因子。在 IVY 中, 方程(2)是基于数据密集型的实验和仿真过程, 并通过下面的差分方程来描述种群成员  $I_i$  的生长速度  $Gv_i(t)$ :

$$\Delta Gv_i(t+1) = rand^2 \odot (N(1, D) \odot \Delta Gv_i(t)) \quad (3)$$

式中:  $\Delta Gv_i(t)$  和  $\Delta Gv_i(t+1)$  分别表示在离散时间系统中  $t$  和  $t+1$  时的生长速率,  $rand$  是分布在区间  $[0, 1]$  的均匀随机实数, 而  $rand^2$  是来自随机变量的随机数, 其概率密度函数为  $1/(2\sqrt{x})$ 。 $N(1, D)$  表示维度为  $D$  的随机向量, 其中每个分量都来自标准高斯分布。

## 2.3 通过生长来获取阳光资源

为了在自然环境中生存下去, 常春藤必须要找到附着表面(如树木、墙壁或岩石), 以便向阳光攀爬。在森林中, 幼年的常春藤会选择朝着最近的树木生长, 而这颗树木往往也生长着老常春藤, 如图 2 所示。通过这种方式, 常春藤可以逐渐覆盖大片的森林区域。常春藤的这种攀爬行为使得只有最强壮的常春藤能够存活, 且几乎不受年龄影响。

图 2 常春藤攀爬示意图<sup>[15]</sup>Fig. 2 Diagram of IVY climbing<sup>[15]</sup>

下面的公式描述了种群成员  $I_i$  如何使用成员  $I_{ii}$  沿着光源的方向有逻辑地攀爬和移动:

$$\begin{aligned} I_i^{\text{new}} &= I_i + |N(1, D)| \odot (I_{ii} - I_i) + N(1, D) \odot \Delta Gv_i, \\ (i &= 1, \dots, N_{\text{pop}}) \end{aligned} \quad (4)$$

其中,

$$\Delta Gv_i = \begin{cases} I_i \oslash (I_{\max} - I_{\min}), & \text{Iter} = 1 \\ rand^2 \odot (N(1, D) \odot \Delta Gv_i), & \text{Iter} > 1 \end{cases} \quad (5)$$

式中:  $|N(1, D)|$  是一个向量, 其分量是向量  $N(1, D)$  分量的绝对值, 运算符“ $\oslash$ ”表示两个向量的 Hadamard 除法。

## 2.4 常春藤种群的传播与进化

种群成员  $I_i$  在搜索空间中遍历全局以寻找最接近、最具生命力的邻居  $I_{ii}$  之后, 种群成员  $I_i$  尝试跟随整个种群中最好的成员  $I_{\text{Best}}$ , 这种行为相当于依附于  $I_{\text{Best}}$  以便在其周围找到最优解。该阶段用数学公式表示如下:

$$I_i^{\text{new}} = I_{\text{Best}} \odot (rand(1, D) + N(1, D) \odot \Delta Gv_i) \quad (6)$$

式(6)中的种群成员  $I_i^{\text{new}}$  的生长速率新值  $\Delta Gv_i^{\text{new}}$  由式(7)计算。

$$\Delta Gv_i^{\text{new}} = I_i^{\text{new}} \oslash (I_{\max} - I_{\min}) \quad (7)$$

## 2.5 幸存者选择

常春藤植物具有两个交替生长阶段,即幼叶阶段和成年叶阶段。常春藤在幼叶阶段会快速向上生长,而在成年叶阶段会扩展新枝条的宽度以及开花。在 IVYA 中将会采用以下方法来建立常春藤在这两个交替生长阶段的模型,即当种群成员  $I_i$  的目标函数值  $f(I_i)$  小于  $f(I_{\text{Best}})$  的  $\beta = (2 + \text{rand})/2$  时,常春藤将会扩展枝条和叶子的宽度,其数学表达式如式(4)所示。反之,则其数学表达式如式(6)所示。

在每次迭代结束后,IVYA 将前一轮迭代中计算得到的种群  $\vec{I}$  与当前迭代中新生成的种群  $\vec{I}_{\text{new}}$  合并,形成一个新的种群集合:

$$\vec{I}^{\text{Merged}} = \langle \vec{I}, \vec{I}_{\text{new}} \rangle \quad (8)$$

合并后的种群  $\vec{I}^{\text{Merged}}$  包含了两个来源的所有成员,通过计算适应度值对这些合并后的种群成员进行排序,以确定它们的相对优劣。选择前  $N_{\text{pop}}$  个成员作为下一轮迭代的种群,其中  $N_{\text{pop}}$  是算法中设定的种群大小。

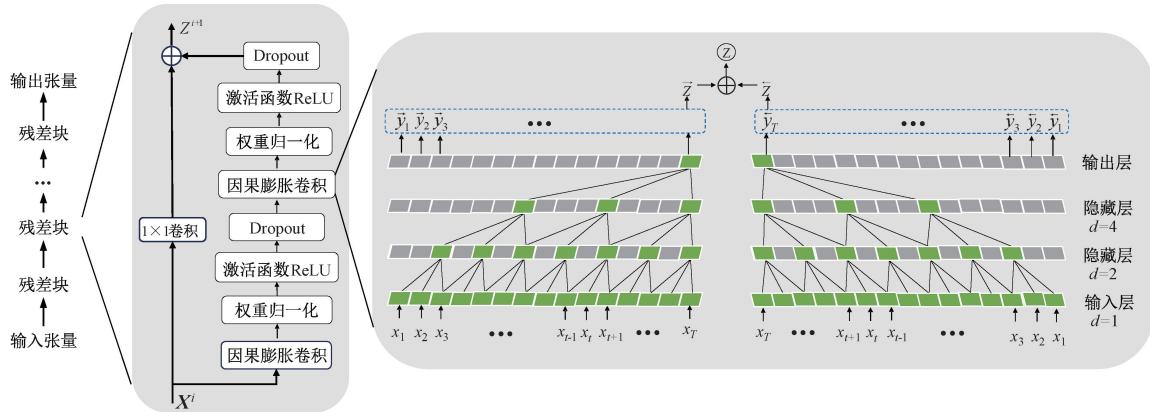


图 3 BiTCN 模型网络结构

Fig. 3 The BiTCN model network structure

## 3.2 双向门控循环单元架构

双向门控循环单元是 RNN 的一种特殊形式,主要是通过整合前向 GRU 层和后向 GRU 层的数据输入,使得模型能够同时从两个方向处理时间序列数据,从而捕捉到更加全面和丰富的时序特征。与 LSTM 相比,GRU 在设计上更简洁,其网络架构仅设计了两个门:重置门和更新门,其结构如图 4 所示。图 4 中  $R_t$  表示重置门,重置门的值通过一个 sigmoid 激活函数生成,其输出范围在 0~1,值越接近 0 表示模型将几乎不考虑之前的信息,值越接近 1 则表示保留较多的过去信息。 $Z_t$  表示更新门,更新门的值也通过一个 sigmoid 激活函数生成,值越接近 0 则表示更多地引入新信息,值越接近 1 表示模型保留更多的旧信息<sup>[17]</sup>。

## 3 模型结构原理

### 3.1 双向时间卷积网络构架

TCN 模型借鉴 CNN 的卷积操作,通过引入因果卷积和膨胀卷积使其适应时间序列数据。因果卷积模块通过限制卷积核的移动方向,确保时间序列预测的因果性。膨胀卷积通过引入膨胀因子,在卷积核的权重矩阵上产生间隔,使得卷积操作能够在不增加参数量的前提下覆盖更大的感受野。此外,在 TCN 中引入了残差块,通过残差路径的添加,有效缓解梯度消失和爆炸问题,提高了网络的训练效率和稳定性。通过增加双向信息处理机制,对逆向序列中的特征进行学习,可以显著提高模型对时间序列数据的理解和预测能力<sup>[16]</sup>。

给定输入序列  $X_t = [x_1, x_2, \dots, x_t, \dots, x_T] \in \mathbb{R}^{T \times V}$ , 并将其记作前向输入向量  $\vec{X}$ , 将  $X_t$  翻转得到逆向序列  $\vec{X}$ , 即  $\vec{X} = [x_T, \dots, x_t, \dots, x_2, x_1] \in \mathbb{R}^{T \times V}$ , 如图 3 所示。将这两个序列输入到残差块中进行特征提取,经过一系列操作后得到的特征向量分别表示为  $\vec{Z}$  和  $\vec{Z}$ , 最后将前向和后向特征在加法层中按元素相加,融合成一个新的特征向量,融合后的特征向量  $Z$  的维度与单个特征维度相同。图中变量的上标箭头只表示方向。

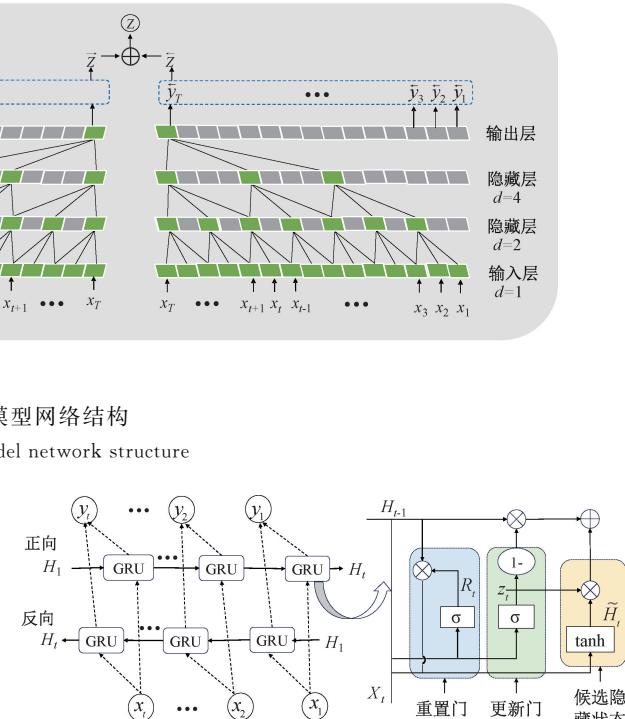


图 4 BiGRU 和单个 GRU 模型网络结构

Fig. 4 BiGRU and single GRU model network structure

基于 GRU 神经网络的架构,其前向传播的数学公式如下:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad (9)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad (10)$$

$$\tilde{H}_t = \tanh[X_t W_{xh} + (R_t \times H_{t-1}) W_{hr} + b_h] \quad (11)$$

$$H_t = (1 - Z_t) \times H_{t-1} + Z_t \times \tilde{H}_t \quad (12)$$

式中:  $X_t$  为输入,  $H_{t-1}$  为前一个时间步的隐藏状态信息,  $W_{\text{下标}}$  表示要学习的参数权重,  $b_{\text{下标}}$  表示偏置,  $\tilde{H}_t$  表示候选隐藏状态, “ $\times$ ”表示向量的逐元素乘法,  $\sigma(\cdot)$  和  $\tanh(\cdot)$  为激活函数。

### 3.3 自注意力机制

自注意力机制是深度学习领域中的一种重要技术, 作为 Transformer 架构的核心组件, 极大地提升了模型在处

理序列数据和捕捉长程依赖关系方面的能力。它通过对输入特征赋予不同的权重, 选择性地强调重要特征, 并将权重分布应用于神经网络训练过程中的中间结果。本文主要是将自注意力机制融入到 BiGRU 的隐藏状态中, 允许自注意力机制对 BiGRU 的隐藏状态进行加权<sup>[18]</sup>。

### 3.4 IVYA 优化 BiTCN-BiGRU-Attention 模型预测流程

本文采用 IVYA 对 BiTCN-BiGRU-Attention(BBA) 模型中的关键超参数进行优化。通过不断更新种群个体的位置, 评估适应度以及选择优秀个体, 逐步逼近优化问题的最优解。优化流程如图 5 所示。

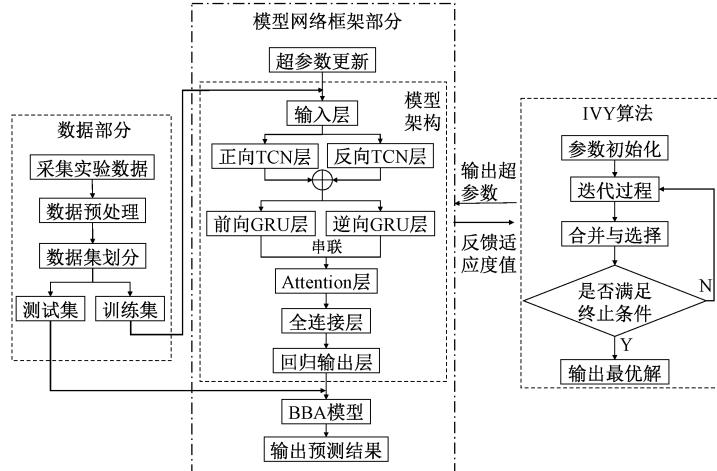


图 5 IVYA 优化 BBA 模型流程

Fig. 5 IVYA optimization BBA model process

具体步骤如下:

1) 数据预处理, 对采集到的原始数据进行异常数据的剔除以及进行归一化处理, 并将数据集划分为训练集和测试集。

2) 选择网络模型中对预测结果有重要影响的超参数, 并确定其搜索空间。

3) 初始化 IVYA 相关参数。包括种群规模、最大迭代次数、搜索空间维度。生成初始种群, 计算其在目标函数上的值, 用于评估适应度。

4) 在迭代过程中, 对种群规模中的每个个体执行以下操作: (1) 基于适应度策略选择优秀的个体。(2) 更新个体的位置。在更新个体位置之前, 需要计算两个关键参数  $\beta$  和  $\Delta Gv_i$ , 用于计算阈值和个体位置。当选中的个体适应度满足相对应的阈值时, 则按相对应的公式更新个体位置。(3) 计算更新后的个体适应度值和  $\Delta Gv_i$ , 结束当前个体的迭代, 直至所有个体都经历上述操作。

5) 在所有个体完成更新后, 合并当前种群和新生种群, 并排序选择最优个体。

6) 更新迭代次数, 判断是否满足终止条件。若满足, 则输出最优解。若不满足, 返回到步骤 4)。

## 4 重钙粉磨实例分析

### 4.1 特征选择

本文的实验数据来源于某重质碳酸钙粉磨工厂生产数据, 部分实验样本数据如表 1 所示。该数据集采样周期为 10 min, 共筛选出 4 989 组样本数据, 并按照 80% 和 20% 的比例分别将其划分为训练集和测试集。

为避免特征数据冗余, 选出与重钙粒度联系最紧密的输入特征。本文使用 PEARSON 相关系数法对输入特征和标签进行相关性分析。其过程特征包括磨辊的油压、油温, 分级机的电流、频率和转速, 立磨的出口温度、主机电流、入口温度、入口压力、压差压力、振幅, 风机的电流、频率、出口风门开度、风速、入口风门开度、转速, 皮带输送机的电流、频率, 加水泵的电流、频率, 药泵的电流、频率、温度, 进料螺旋机电流等 25 个特征维度, 在热力图中依次记作 1~25; 标签为重钙粒度的 2  $\mu\text{m}$  下累积值, 在热力图中记作序号 26。通过计算输入特征与标签之间的相关性系数, 剔除了磨辊油压和油温、立磨的主机电流、振幅, 加水泵的电流、频率, 进料螺旋机电流, 药泵的频率、温度等 9 个系数较低的值, 其相关性分析如图 6 所示。

表 1 实验样本数据集

Table 1 Experimental sample dataset

采样 编号	磨辊油压/ Kgf·cm <sup>-2</sup>	磨辊 油温/℃	分级机 电流/A	分级机 频率/Hz	分级机 转速/rpm	立磨出口 温度/℃	立磨主机 电流/A	立磨入口 温度/℃	…	下累积结果 (2 μm)
1	45.9	39.9	12.24	19.0	1 014.6	51.5	341.5	53.2	…	26.83
2	46.0	40.0	9.48	18.5	1 011.2	51.4	338.7	54.0	…	26.36
3	44.0	40.1	12.22	18.5	986.0	49.0	296.1	54.6	…	26.07
4	44.1	40.2	12.48	18.5	983.4	47.1	270	53.5	…	25.93
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4 989	44.4	38.5	16.32	37.5	2 001.0	55.2	396.7	57.2	…	47.09

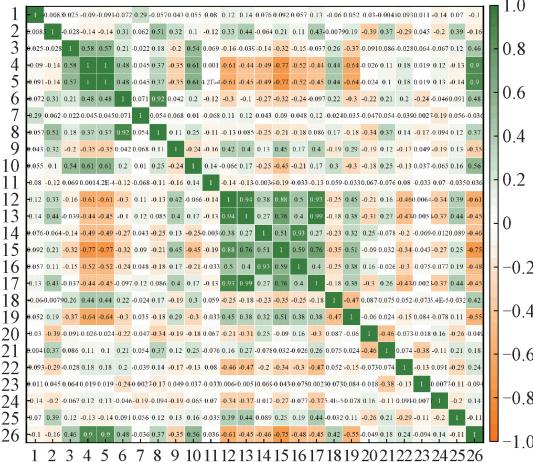


图 6 相关性分析热力图

Fig. 6 Correlation analysis heatmap

#### 4.2 数据预处理

根据工厂对重钙粉体粒度的检测记录,剔除生产过程中停机、跳机、不良等工况的前后数据点。对于数据集中的缺失值,采用三次样条插值法进行填补。在数据集中存在不同的量纲特征,需要进行归一化处理,将不同尺度的数据转化到一个统一的尺度,避免特征间不平等的影响。本文使用 max-min 归一化将数据转化到[0,1]之间。

在生产过程中,对分级机的电流进行数据采集时,电流会受到所处电磁环境中电磁噪声的干扰,导致采集的电流信号在产生突变点时可能会被噪声覆盖<sup>[19]</sup>。因此,在模型训练之前,需要采用滤波算法处理电流信号。本文采用时变滤波经验模态分解法与小波阈值混合滤除噪声信号,可以有效解决传统 EMD 中的模态混叠和模态过多等问题。其处理前后对比如图 7 所示。

#### 4.3 预测模型评价指标

为了定量地描述预测模型的性能,本文采用了 4 个评价指标,即均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)、平均绝对值百分比误差(mean absolute percentage error, MAPE)、拟合系数(R-square, R<sup>2</sup>)。具体计算公式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (13)$$

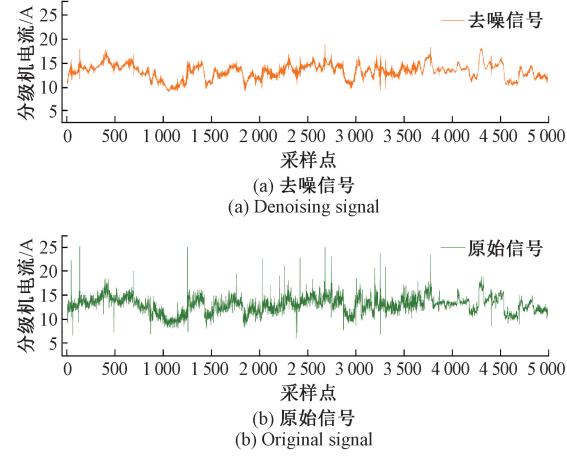


图 7 时变滤波经验模态与小波阈值混合去噪

Fig. 7 The TVF-EMD and wavelet thresholding mixed denoising

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (15)$$

$$R^2 = 1 - \left( \sum_{i=1}^n (\hat{y}_i - y_i)^2 / \sum_{i=1}^n (\bar{y}_i - y_i)^2 \right) \quad (16)$$

式中:  $\hat{y}_i$ 、 $y_i$  和  $\bar{y}_i$  分别为  $i$  时刻的预测值、真实值和均值,  $n$  为样本数量。

#### 4.4 实验平台与模型参数配置

本文所使用的训练平台选择 GPU, 型号为 NVIDIA GeForce GTX 4060, 处理器为 Intel i7-8300H 2.30 GHz, 操作系统为 Window10, 实验环境为 MATLAB 2023b 版。

在进行模型训练之前,先设置各模块的初始参数值。IVYA 和 BiTCN-BiGRU-Attention 模型的相关参数如表 2 所示。

#### 4.5 预测结果分析

本文仿真实验以 TCN、GRU、CNN-GRU、TCN-GRU、BiTCN-BiGRU、BiTCN-BiGRU-Attention 等基准模型的预测数据作为参考,记作实验对照组,与所采用的模型预测结果进行对比。对于上述模型,由于没有加入优化算法,则模型中的超参数需要人为设定,这里分别对各模型进行多次仿真实验,取模型的最优预测结果。各模型在测

表 2 模型和算法相关参数设置

Table 2 Parameter settings for models and algorithms

类别	参数	数值
IVYA	种群数量	10
	进化代数	8
	变量个数	4
	学习率	[0.0001, 0.01]
	GRU 神经元个数	[10, 100]
	卷积核个数	[10, 120]
	正则化参数	[0.00001, 0.005]
BiTCN-BiGRU-Attention 模型	卷积核大小	5
	DropoutFactor	0.1
	numBlocks	2
	膨胀因子	[1, 2]
	最大训练次数	500
	注意力头数	1
	注意力键值	50
优化器	Adam	

试集上的性能对比如表 3 所示。

从上述实验结果可知, 本文的初始模型是基于 TCN 和 GRU 模型, 其在所有的评价指标上数据表现最差。与单一模型相比, TCN-GRU 模型的 RMSE、MAE、MAPE 分别减少 19.16%、35.48%、47.84%, R<sup>2</sup> 提高 3.83%。说明 TCN 模型的长时间依赖捕捉能力与 GRU 模型的动态记忆和门控机制结合, 能够提供更好的预测性能和模型稳定性。另外, 将 CNN 和 TCN 分别与 GRU 结合, 实验结果表明 TCN-GRU 模型优于 CNN-GRU 模型。比起 CNN 通过多层卷积来扩展感受野, TCN 模型利用膨胀卷积覆盖更长的序列长度对处理长期复杂的时序数据更契合。在 TCN-GRU 模型的基础上, 引入双向特性, 增强模型对前后数据的理解。实验结果表明 BiTCN-BiGRU 模型的 RMSE、MAE、MAPE 分别减少 20.09%、31.26%、26.46%, R<sup>2</sup> 提高 1.87%。在 BiGRU 隐藏状态中融入 Attention 机制, 其模型与 BiTCN-BiGRU 模型相比, RMSE、MAE、MAPE 分别减少 7.71%、7.32%、3.0%, R<sup>2</sup> 提高 1.2%, 表明 Attention 机制对 BiGRU 隐藏状态进行加权在一定程度上增强了模型对关键数据的关注能力。

表 3 不同模型的性能对比

Table 3 Performance comparison of different models

预测模型	RMSE	MAE	MAPE/%	R <sup>2</sup> /%
TCN	2.811 7	2.441 1	8.315 6	90.06
GRU	2.596 9	2.029 2	6.310 6	91.52
CNN-GRU <sup>[20-21]</sup>	2.282 5	1.686 5	5.250 3	92.79
TCN-GRU <sup>[22]</sup>	2.273 0	1.575 1	4.336 3	93.51
BiTCN-BiGRU <sup>[23]</sup>	1.816 4	1.082 7	3.189 1	95.26
BiTCN-BiLSTM <sup>[24-25]</sup>	1.755 9	1.255 5	3.847 3	95.13
BiTCN-BiGRU-Attention	1.676 3	1.003 4	3.093 5	96.40
IVYA-BiTCN-BiGRU-Attention	0.824 4	0.423 0	1.295 4	98.95

经过超参数优化处理后, 其 RMSE、MAE、MAPE 分别减少 50.8%、57.8%、58.1%, R<sup>2</sup> 提高 2.6%。这充分说明采用 IVYA 能够为模型引入最佳的参数设置, 进而大幅提升其预测性能。图 8 及其 A、B 两区域局部放大图展示了上述模型在测试集上的拟合效果, 其基准模型预测结果与真实值变化趋势相近, 表明这些模型能够捕捉到粒度的变化趋势。然而, 各模型均存在较大的误差。相比之下, 本文采用的模型在拟合效果上表现最佳。

同时, 为了验证 IVYA 的优化性能, 本文选取了北方苍鹰(NGO)、鹅(GOOSE)、冠豪猪(CPO)、黑翅鸢(BKA)等最新发表的优化算法对 BiTCN-BiGRU-Attention 模型进行了对比。具体评价指标和可视化结果如表 4 和图 8 所

示。实验表明, 采用 IVYA 优化 BiTCN-BiGRU-Attention 模型在所有的评价指标上均优于其他优化算法, 即 RMSE、MAE、MAPE 分别减少 50.82%、57.84%、58.13%, R<sup>2</sup> 提高 2.6%。由图 9 及其 C、D 两区域局部放大图可知, 4 种优化算法都可以预测粒度的变化趋势, 但 IVYA 的拟合系数最高, 预测曲线趋势和原始曲线最相近。在图 9 中 C、D 两处区域存在一定误差, 这是因为数据本身在这两处呈断崖式下降, 在保证全局最优解的情况下, 局部最优解会存在偏差, 这也是后期需要研究优化的方向。

表 5 展示了 5 种优化算法对模型中 4 个超参数优化结果以及各算法的适应度值。本文以最小化损失函数为优化目标, 则适应度值越小, 算法的优化效果越好。

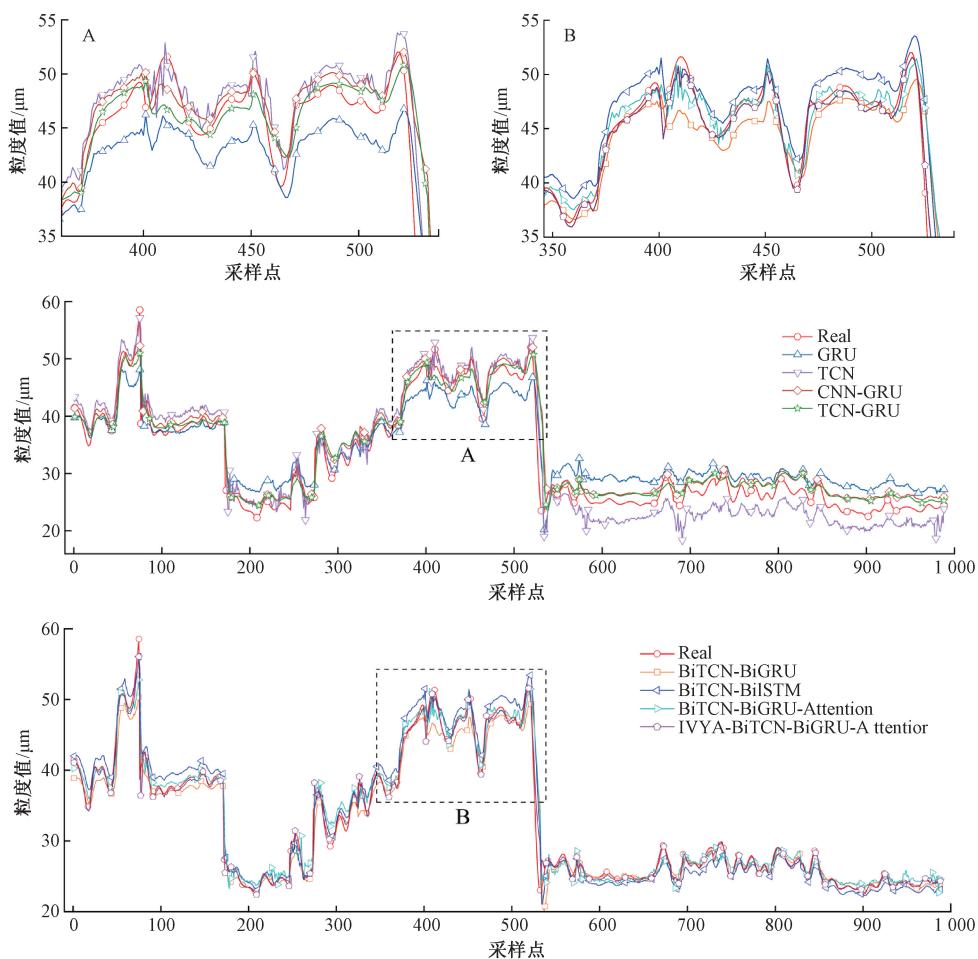


图 8 不同模型预测结果对比

Fig. 8 Comparison of prediction results from different models

表 4 不同优化算法的性能对比

Table 4 Performance comparison of different optimization algorithms

预测模型	RMSE	MAE	MAPE/%	R <sup>2</sup> /%
CPO-BiTCN-BiGRU-Attention	0.892 1	0.464 5	1.460 0	98.76
GOOSE-BiTCN-BiGRU-Attention	1.023 0	0.571 1	1.794 3	98.44
BKA-BiTCN-BiGRU-Attention	1.018 6	0.571 4	1.820 0	98.50
NGO-BiTCN-BiGRU-Attention	0.971 6	0.515 4	1.538 8	98.61
IVYA-BiTCN-BiGRU-Attention	0.824 4	0.423 0	1.295 4	98.95

表 5 不同优化算法超参数优化结果

Table 5 Optimization results of different algorithms

优化算法	适应度值	超参数			
		学习率	神经元个数	卷积核个数	正则化参数
NGO	0.971 6	$5.1 \times 10^{-4}$	57	94	$1.2 \times 10^{-5}$
GOOSE	1.023 0	$5.3 \times 10^{-3}$	86	47	$2.66 \times 10^{-4}$
CPO	0.892 1	$6.1 \times 10^{-4}$	49	20	$9.8 \times 10^{-5}$
BKA	1.018 6	$2.9 \times 10^{-3}$	29	52	$3.0 \times 10^{-5}$
IVYA	0.824 4	$1.2 \times 10^{-3}$	45	20	$1.5 \times 10^{-5}$

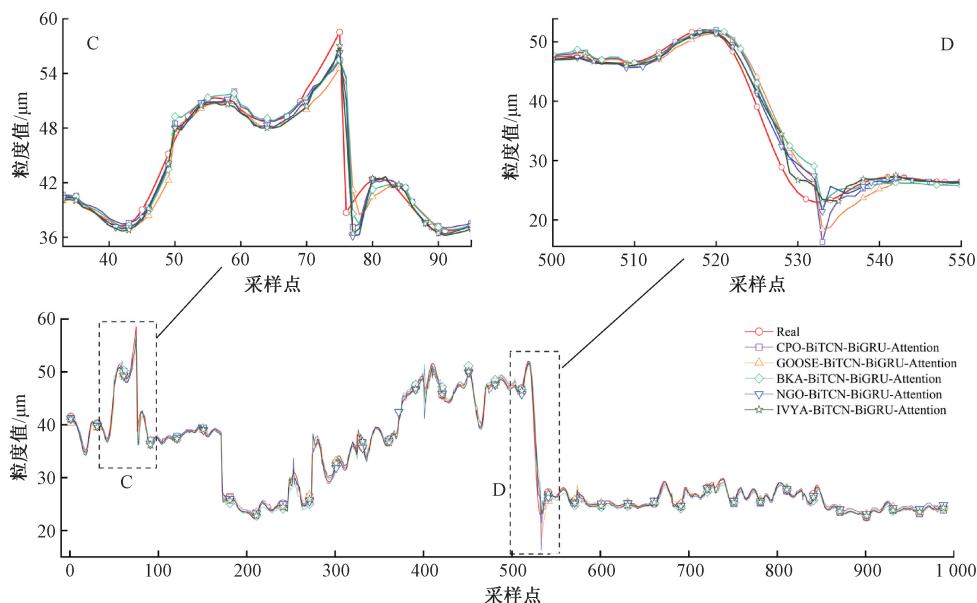


图9 不同优化算法预测结果对比

Fig. 9 Comparison of prediction results from different optimization algorithms

## 5 结论

针对重质碳酸钙粉磨系统中粉体粒度的预测问题,本文提出了一种基于超参数优化和Attention机制的BiTCN-BiGRU预测模型,通过设计多种模型和优化算法进行横向和纵向实验对比,从而验证预测模型的有效性和先进性。主要研究结论如下:通过比较六种模型的RMSE、MAE、MAPE、 $R^2$ 等评价指标,确定BiTCN-BiGRU-Attention模型的预测精度最高,对整体特征有出色的拟合性能,并且模型的稳定性较好,重复实验的决定系数误差小于1%。针对立磨运行参数间的非线性、变时延和耦合性问题,BiTCN与BiGRU结合可以解决变量间的耦合特征以及处理时序数据中的变时延问题,而模型中的激活函数可以构造复杂的非线性关系。

实验结果表明,相较于NGO、GOOSE、CPO、BKA算法,IVYA可以寻找模型中关键超参数的全局最优解,提高了预测精度,避免了局部最优解。有利于准确地预测工艺参数的最佳设置,减少手动调节的误差。根据重钙立磨粉体粒度分布和立磨运行参数之间的相关数据,所建立的预测模型可以有效的解决工艺参数调节困难的问题。本研究结论可为生产重钙产品制定出运行参数范围提供参考价值。

## 参考文献

[1] 曾令威,卢乐民,梁逸昊,等.重质碳酸钙应用于建筑涂料中的研究进展[J].当代化工,2023,52(11):2761-2766.

ZENG L W, LU L M, LIANG Y H, et al. Research progress of application of heavy calcium carbonate in

architectural coatings [J]. Contemporary Chemical Industry, 2023, 52(11):2761-2766.

[2] 胡泽武,胡其军.重钙粉立磨工艺系统运行参数研究[J].中国非金属矿工业导刊,2023(2):73-77.

HU Z W, HU Q J. Research on operation parameters of vertical roller mill process system of fine ground calcium carbonate powder [J]. China Non-metallic Mining Industry, 2023(2):73-77.

[3] 黎芬芳,崔桂梅.基于PSO优化T-S模糊神经网络的磨矿粒度预测[J].烧结球团,2022,47(4):64-70.

LI F F, CUI G M. Grinding size prediction of optimization of T-S fuzzy neural network based on PSO [J]. Sintering and Pelletizing, 2022, 47 (4): 64-70.

[4] 梁宏涛,刘硕,杜军威,等.深度学习应用于时序预测研究综述[J].计算机科学与探索,2023,17(6):1285-1300.

LIANG H T, LIU SH, DU J W, et al. Review of deep learning applied to time series prediction [J]. Journal of Frontiers of Computer Science & Technology, 2023, 17(6):1285-1300.

[5] LIU G, OUYANG ZH Y, HAO X CH, et al. Prediction of raw meal fineness in the grinding process of cement raw material: A two-dimensional convolutional neural network prediction method [J]. Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2021, 235(6): 823-838.

[6] BAO Y Y, ZHU Y M, DU W L, et al. A distributed

PCA-TSS based soft sensor for raw meal fineness in VRM system[J]. Control Engineering Practice, 2019, 90: 38-49.

[7] BAI SH J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. ArXiv preprint arXiv:1803.01271, 2018.

[8] 郝椿淋,张剑.基于自注意力机制TCN-BiGRU的交通流预测[J].电子测量技术,2024,47(8):61-68.  
HAO CH L, ZHANG J. Traffic flow prediction based on self-attention mechanism TCN-BiGRU [J]. Electronic Measurement Technology, 2024, 47 (8): 61-68.

[9] CHO K, MERRIENBOER B V, BAHdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches[J]. ArXiv preprint arXiv: 1409.1259, 2014.

[10] CAHUANTZI R, CHEN X, GÜTTEL S. A comparison of LSTM and GRU networks for learning symbolic sequences [C]. Science and Information Conference, Cham: Springer Nature Switzerland, 2023: 771-785.

[11] TAN H, SHANG Y J, LUO H J, et al. A combined temporal convolutional network and gated recurrent unit for the remaining useful life prediction of rolling element bearings[C]. International Conference on the Efficiency and Performance Engineering Network, Cham: Springer Nature Switzerland, 2023: 853-862.

[12] YANG L, SHAMI A. On hyperparameter optimization of machine learning algorithms: Theory and practice [J]. Neurocomputing, 2020, 415: 295-316.

[13] HANIFI S, CAMMARONO A, ZARE-BEHTASH H. Advanced hyperparameter optimization of deep learning models for wind power prediction [J]. Renewable Energy, 2024, 221: 119700.

[14] 韩素敏,赵国帅,尚志豪,等.基于改进DBO优化BiLSTM的IGBT老化预测模型[J].电子测量技术,2024,47(1):46-54.  
HAN S M, ZHAO G SH, SHANG ZH H, et al. IGBT aging prediction model based on improved DBO optimization BiLSTM [J]. Electronic Measurement Technology, 2024,47(1):46-54.

[15] GHASEMI M, ZARE M, TROJOVSKY P, et al. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm[J]. Knowledge-Based Systems, 2024, 295: 111850.

[16] 梁浩鹏,曹洁,赵小强.基于并行双向时间卷积网络和双向长短期记忆网络的轴承剩余使用寿命预测方法[J].控制与决策,2024,39(4):1288-1296.  
LIANG H P, CAO J, ZHAO X Q. Remaining useful life prediction method for bearing based on parallel bidirectional temporal convolutional network and bidirectional long and short-term memory network[J]. Control and Decision, 2024,39(4):1288-1296.

[17] 逯静,张燕茹,王瑞.基于BWO-VMD和TCN-BiGRU的短期风功率预测[J/OL].工程科学与技术,1-14 [2024-12-18]. <http://kns.cnki.net/kcms/detail/51.1773.TB.20240604.1704.013.html>.  
LU J, ZHANG Y R, WANG R. Basing on the integration of BWO-VMD and TCN-BiGRU for short-term wind power forecasting [J/OL]. Engineering Science and Technology, 1-14 [2024-12-18]. <http://kns.cnki.net/kcms/detail/51.1773.TB.20240604.1704.013.html>.

[18] 祝玉华,张钰涵,李智慧,等.基于TCN-BiGRU结合自注意力机制的储粮温度预测研究[J].中国农机化学报,2024,45(12):133-139.  
ZHU Y H, ZHANG Y H, LI ZH H, et al. Research on grain storage temperature prediction based on TCN-BiGRU combined with self-attention mechanism [J]. Journal of Chinese Agricultural Mechanization, 2024, 45(12):133-139.

[19] 孙克仲,鲁迎春,杨笑笑,等.改进小波阈值函数在高功率电源信号去噪中的应用[J/OL].电源学报,1-12 [2024-12-18]. <http://kns.cnki.net/kcms/detail/12.1420.TM.20221206.1420.001.html>.  
SUN K ZH, LU Y CH, YANG X X, et al. Application of modified wavelet threshold function in high power supply signal[J/OL]. Journal of Power Supply, 1-12 [2024-12-18]. <http://kns.cnki.net/kcms/detail/12.1420.TM.20221206.1420.001.html>.

[20] 王楠,胡永涛,王康杰,等.融合TVF-EMD和CNN-GRU锅炉受热面能效预测[J].电子测量与仪器学报,2024,38(9):67-75.  
WANG N, HU Y T, WANG K J, et al. Fusion of TVF-EMD and CNN-GRU boiler heat surface energy efficiency prediction [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (9): 67-75.

[21] 孙兴伟,杨铜铜,杨赫然,等.基于CNN-GRU组合神经网络的数控机床进给系统热误差研究[J].仪器仪表学报,2023,44(10):219-226.  
SUN X W, YANG T T, YANG H R, et al. Research on thermal error of CNC machine tool feed system based on CNN-GRU combined neural network [J].

Chinese Journal of Scientific Instrument, 2023, 44(10):219-226.

[22] 项新建,许宏辉,谢建立,等.基于 VMD-TCN-GRU 模型的水质预测研究[J].人民黄河,2024,46(3):92-97.  
XIANG X J, XU H H, XIE J L, et al. Water quality prediction based on VMD-TCN-GRU model [J]. Yellow River, 2024,46(3):92-97.

[23] HUA ZH G, YANG Q, CHEN J W, et al. Degradation prediction of PEMFC based on BiTCN-BiGRU-ELM fusion prognostic method [J]. International Journal of Hydrogen Energy, 2024, 87: 361-372.

[24] 方祖春,汪繁荣.基于非侵入式负荷分解的有色金属冶炼工序识别[J].国外电子测量技术,2023,42(1): 170-177.  
FANG Z CH, WANG F R. Process identification of non-ferrous metal smelting based on non-invasive load decomposition [J]. Foreign Electronic Measurement Technology, 2023,42(1):170-177.

[25] 高萌,鲁玉军.基于 Bi-TCN-LSTM 的滚动轴承剩余使用寿命预测方法[J].轻工机械,2024(3):66-73,79.  
GAO M, LU Y J. Prediction method for remaining useful life of rolling bearings based on bidirectional temporal convolutional network and long short-term memory network [J]. Light Industry Machinery,2024(3):66-73,79.

### 作者简介

黄何,硕士研究生,主要研究方向为过程控制与优化。

E-mail:2211391017@st.gxu.edu.cn

黄福川(通信作者),博士,教授,主要研究方向为非金属材料加工及其相关设备的研发、油气储运等。

E-mail:huangfuchuan@gxu.edu.cn