



研究与开发

## 基于元启发式RIME算法的深度时间序列预测模型优化方法

诸葛斌<sup>1</sup>, 蔡晓丹<sup>1</sup>, 潘婷婷<sup>1</sup>, 许云汉<sup>1</sup>, 王正贤<sup>1</sup>, 张子天<sup>1</sup>, 董黎刚<sup>1</sup>, 蒋献<sup>1</sup>, 于晓<sup>2</sup>  
(1.浙江工商大学信息与电子工程学院(萨塞克斯人工智能学院), 浙江 杭州 310018;  
2.浙江力积存储科技股份有限公司, 浙江 杭州 310018)

**摘要:** 时间序列预测在金融、电力、网络等关键领域具有重要应用价值。深度学习模型在该任务中展现出强大的拟合能力,但其性能高度依赖结构设计与超参数选择。传统的调参方法(如网格搜索和人工经验)存在效率低、易陷入局部最优等问题。为此,构建了基于双向时间卷积网络和双向门控循环单元的注意力机制模型(bidirectional temporal convolutional network-bidirectional gated recurrent unit-attention, BiTCN-BiGRU-Attention)的深度时间序列预测模型,并引入新型元启发式优化算法——霜冰优化算法(RIME)对其进行优化。RIME算法模拟霜冰自然生长机制,结合软霜搜索策略、硬霜刺破机制与正贪婪选择策略,实现了全局探索与局部开发的有效平衡。实验在标准基准函数和多个真实数据集上对算法性能进行了全面评估。结果表明,RIME算法优化后的预测模型在精度、收敛速度与稳定性方面均优于未优化模型,为深度时序预测模型的高效自动化优化提供了新的思路与实践路径。

**关键词:** 霜冰优化算法;深度学习;超参数优化;时间序列预测

**中图分类号:** TP393

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.DXKX250559

## Optimization method of deep time series prediction model based on meta-heuristic RIME algorithm

Zhugue Bin<sup>1</sup>, Cai Xiaodan<sup>1</sup>, Pan Tingting<sup>1</sup>, Xu Yunhan<sup>1</sup>, Wang Zhengxian<sup>1</sup>, Zhang Zitian<sup>1</sup>,  
Dong Ligang<sup>1</sup>, Jiang Xian<sup>1</sup>, Yu Xiao<sup>2</sup>

(1.School of Information and Electronic Engineering (Sussex Artificial Intelligence Institute), Zhejiang Gongshang University, Hangzhou 310018, China

2.Hangzhou Liji Storage Technology Co., Ltd., Hangzhou 310018, China)

收稿日期: 2025-09-18; 修回日期: 2025-12-29

通信作者: 蒋献, jiangxian@zjgsu.edu.cn

基金项目: 国家自然科学基金资助项目(No.W2421086, No.61871468); 桐乡通用人工智能研究院项目(No.TAGI2-B-2024-0014); 浙江省新型网络标准及应用技术重点实验室项目(No.2013E10012); 浙江省科技创新重点项目(No.2023R5211); 浙江省自然科学基金资助项目(No.LZ23F010003); 浙江省重点研发计划项目(No.2025C02038)

**Foundation Items:** The National Natural Science Foundation of China(No.W2421086, No.61871468), Tongxiang General Artificial Intelligence Research Institute Project(No.TAGI2-B-2024-0014), Zhejiang Provincial Key Laboratory of New Network Standards and Application Technology (No.2013E10012), Key Science and Technology Innovation Project of Zhejiang Province(No.2023R5211), The Natural Science Foundation of Zhejiang Province (No. LZ23F010003), Key Research & Development Program of Zhejiang Province(No.2025C02038)

**Abstract:** Time series prediction was recognized as having significant application value in critical fields such as finance, power, and networks. Deep learning models were demonstrated to possess strong fitting capabilities for this task, but their performance was found to be heavily dependent on structural design and hyperparameter selection. Traditional parameter-tuning methods, such as grid search and manual experience, were criticized for their low efficiency and tendency to fall into local optima. To address these issues, a deep time series prediction model based on the BiTCN-BiGRU-Attention architecture was constructed, and a novel metaheuristic optimization algorithm, RIME, was introduced for optimization. The RIME was designed to simulate the natural growth mechanism of rime ice, combining soft rime search strategies, hard rime piercing mechanisms, and positive greedy selection strategies to achieve an effective balance between global exploration and local exploitation. In the experimental section, the algorithm's performance was comprehensively evaluated on standard benchmark functions and multiple real-world datasets. The results show that the RIME-optimized prediction model was superior to the unoptimized model in terms of accuracy, convergence speed, and stability. New insights and practical pathways were provided for the efficient and automated optimization of deep time series prediction models.

**Key words:** RIME, deep learning, hyperparameter optimization, time series prediction

## 0 引言

在当今数据驱动的决策环境中，时间序列预测应用范围涵盖金融投资分析、能源消耗预测、气候变化监测以及供应链资源调配等多个关键领域。然而，传统时序分析方法如指数平滑（exponential smoothing, ES）法、自回归积分移动平均（autoregressive integrated moving average, ARIMA）模型<sup>[1]</sup>等在面对非线性关系、长期依赖性和多尺度特征等复杂数据特性时，往往表现出明显的局限性。

近年来，深度学习在时序建模领域取得突破性进展，双向时序卷积网络（bidirectional temporal convolutional network, BiTCN）通过扩张卷积捕获序列中的长程依赖，双向门控循环单元（bidirectional gated recurrent unit, BiGRU）则利用其门控机制与双向信息流，递归地增强序列上下文理解。自注意力机制的引入使模型能通过计算序列内部元素间的关联性，动态地为不同时间步分配权重，从而自适应聚焦关键时间点，显著提升预测精度与可解释性。

然而，深度模型性能对超参数高度敏感，网络层数、神经元数量、学习率与正则化等配置会

直接影响收敛稳定性与泛化能力。传统调参方法（如网格搜索）在高维空间代价昂贵，随机搜索效率不稳定，基于梯度的策略易受离散、非光滑参数空间限制，导致调参成本高且难以获得稳健最优解。因此，开发高效、自适应的超参数优化方法已成为深度时序预测领域的迫切需求。

针对上述挑战，本文提出一种基于霜冰优化算法（RIME）的深度时间序列预测框架，构建BiTCN-BiGRU-Attention混合模型并以RIME算法实现关键超参数的自动搜索与自适应更新。本研究的主要贡献如下。

（1）构建了BiTCN-BiGRU-Attention混合深度模型，通过有机融合3种先进神经网络技术，形成互补增强的时序特征提取体系，以提升多尺度特征提取与关键依赖建模能力。

（2）首次将RIME算法应用于深度时序预测模型的超参数优化，通过“全局探索+局部开发”的平衡机制，在有限评估次数下获得更优参数组合。

（3）实验在基准函数与多源真实数据集上验证了方法的有效性。结果表明，RIME算法优化后的模型在平均绝对误差（mean absolute error, MAE）、均方根误差（root mean square error, RMSE）、平均绝对百分比误差（mean absolute



percentage error, MAPE) 等指标上整体优于未优化模型; 在网络流量数据上均方误差 (mean square error, MSE) 相对提升约 33.2%, 并在股票数据上取得接近 1 的决定系数 ( $R^2$ ), 显示出良好的精度与泛化能力。

## 1 研究现状

### 1.1 时间序列预测模型

时间序列预测是根据历史数据预测未来趋势和模式的关键技术。随着数据复杂性的增加, 时间序列预测方法经历了从经典统计模型向深度学习模型的演进。传统的统计时间序列模型, 如 ARIMA 和向量自回归模型 (vector auto regression, VAR)<sup>[2]</sup>, 在处理平稳线性数据时表现良好, 但在面对具有非线性、长程依赖及高噪声特征的复杂数据时, 其拟合能力往往受限。

近年来, 深度学习模型在时间序列预测领域取得了显著进展, 能够有效处理传统统计模型难以应对的复杂数据特性。为解决网络入侵检测领域多分类准确率不高的问题, 文献[3]结合注意力机制与 BiTCN, 提升了网络入侵检测性能。文献[4]基于门控循环单元 (bidirectional gated recurrent unit, BiGRU) 神经网络提出一种 BiGRU 神经网络和人工神经网络堆叠的网络模型, 提升了流量峰值的预测准确度。文献[5]通过基于时序卷积网络 (temporal convolutional network, TCN) -BiGRU 的异常检测模型提取数据的空间和时序特征, 有效识别网络流量的攻击类型。

### 1.2 超参数优化方法

深度学习模型的性能高度依赖其超参数的设置, 超参数优化是保证算法性能最佳的必要步骤<sup>[6]</sup>。

传统的超参数优化方法 (如网格搜索和随机搜索) 实现简单, 但在高维搜索空间下成本高、效率有限。贝叶斯优化算法<sup>[7]</sup>通过代理模型在有限评估次数下提高了搜索效率, 已在多类深度模型调参任务中得到应用。如文献[8]基于贝叶斯优

化对 CNN-LSTM-Attention 推荐模型进行调参以提升推荐效果。文献[9]则结合异方差非平稳高斯过程改进贝叶斯优化以适应不确定性建模需求。

除贝叶斯优化外, 元启发式算法 (如粒子群优化算法、遗传算法等) 凭借全局探索与局部开发能力, 被广泛用于超参数、层数、神经元数与学习率等关键变量的寻优<sup>[10]</sup>。在与 RIME 算法相关研究中, 文献[11]将 RIME 算法用于 CNN-LSTM-AM 模型超参数寻优以提升训练效率并降低成本。文献[12]采用 RIME—变分模态分解—麻雀搜索算法—长短期记忆网络 (RIME-variational mode decomposition-sparrow search algorithm-long short term memory, RIME-VMD-SSA-LSTM) 组合算法进行短期来水流量预测并提升精度。文献[13]提出混合估计霜冰优化 (hybrid estimation RIME, HERIME) 算法, 采用基于概率模型的估计分布算法采样方法, 以提升霜冰种群的质量并增强其全局探索能力。

## 2 模型构建

为评估 RIME 算法在深度模型超参数寻优中的有效性, 本文选择构建一个结构较复杂且对超参数较敏感的预测模型作为优化对象。综合多尺度依赖建模与序列上下文表达需求, 本文采用 BiTCN-BiGRU-Attention 混合模型: 首先由 BiTCN 提取多尺度时序特征, 其次再由 BiGRU 融合双向上下文, 最后通过注意力机制对关键时间步进行加权, 以提升预测性能与稳定性。

### 2.1 模型设计

#### 2.1.1 BiTCN

TCN 是一种专门为序列建模设计的卷积神经网络架构。TCN 通过因果卷积保证预测仅依赖历史信息, 并利用扩张卷积在不显著增加参数的情况下扩大感受野以捕获长期依赖。

本文采用的 BiTCN 结构进一步增强了模型对时间序列信息的利用。BiTCN 包含两个并行的

TCN分支：一个正向TCN分支处理原始顺序的时间序列，另一个反向TCN分支处理时间反向的序列。两个分支的输出在后续层进行融合，从而为模型提供更全面的时间依赖信息。

TCN中的因果卷积操作可以表示如下：

$$y_t = \sum_{i=0}^{k-1} w_i x_{t-d-i} \quad (1)$$

其中， $y_t$ 是在时间步 $t$ 的输出， $x$ 是输入序列， $w_i$ 是卷积核的权重， $k$ 是卷积核的大小， $d$ 是扩张因子，当 $d=1$ 时为标准因果卷积。

### 2.1.2 网络结构

GRU是一种基于循环神经网络（recurrent neural network, RNN）和LSTM的改进方法，旨在解决传统RNN在处理长序列时遇到的梯度消失和梯度爆炸问题。GRU通过引入更新门和重置门来控制信息的流动和保留。更新门决定前一时间步隐藏状态的保留程度，而重置门决定该状态的忽略程度。GRU的计算如下。

重置门：

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2)$$

更新门：

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (3)$$

候选隐藏状态：

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t \odot h_{t-1}, x_t] + b_{\tilde{h}}) \quad (4)$$

当前隐藏状态：

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (5)$$

其中， $r_t$ 是重置门的输出， $z_t$ 是更新门的输出， $\tilde{h}_t$ 是输入 $x_t$ 和上一隐藏层状态的汇总， $h_t$ 是当前时刻的隐藏状态， $x_t$ 是当前时间步的输入， $h_{t-1}$ 是前一时间步的隐藏状态， $W_r$ 、 $W_z$ 、 $W_{\tilde{h}}$ 是权重矩阵， $b_r$ 、 $b_z$ 、 $b_{\tilde{h}}$ 为对应的偏置向量， $\sigma$ 是Sigmoid激活函数， $\tanh$ 是双曲正切激活函数， $\odot$ 表示元素级乘法。

BiGRU由两个独立的GRU层组成：一个按

时间正向处理输入序列，另一个按时间反向处理输入序列。这种结构使得模型能够同时利用过去和未来的上下文信息，从而更全面地理解序列中的依赖关系。

### 2.1.3 注意力机制

注意力机制最初在自然语言处理领域取得了巨大成功，随后被广泛应用于各种序列建模任务，包括时间序列预测。本文采用的自注意力机制是注意力机制的一种特殊形式，它计算序列内部各元素之间的相关性，并根据这些相关性为每个元素生成加权的表示。

自注意力机制的计算如下：

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

其中， $Q$ 、 $K$ 、 $V$ 是输入序列的3个不同线性变换， $\sqrt{d_k}$ 是缩放因子，用于防止点积过大导致Softmax函数梯度过小。Softmax函数将注意力权重归一化，使得所有权重之和为1。通过这种方式，模型可以学习到输入序列中不同部分的重要性。

在本文的BiTCN-BiGRU-Attention模型中，注意力机制被应用于BiGRU层的输出之上。这意味着模型可以根据BiGRU提取的上下文特征，自适应地判断哪些时间步的特征对于最终的预测结果更为关键。

## 2.2 模型结构

BiTCN-BiGRU-Attention模型结构如图1所示。数据处理流程如下：首先，通过输入层接收归一化时间序列，其中核心特征提取部分采用BiTCN结构，利用双向扩张因果卷积与残差连接机制，在扩大感受野的同时有效捕获双向时序依赖；其次使特征经由BiGRU层进一步处理，以融合全局时序上下文；再次，为提升预测精度，模型引入单头自注意力机制，对BiGRU输出的特征进行加权处理，突出关键时间步信息；最后，经过全连接层映射，由回归层输出最终预测结果。

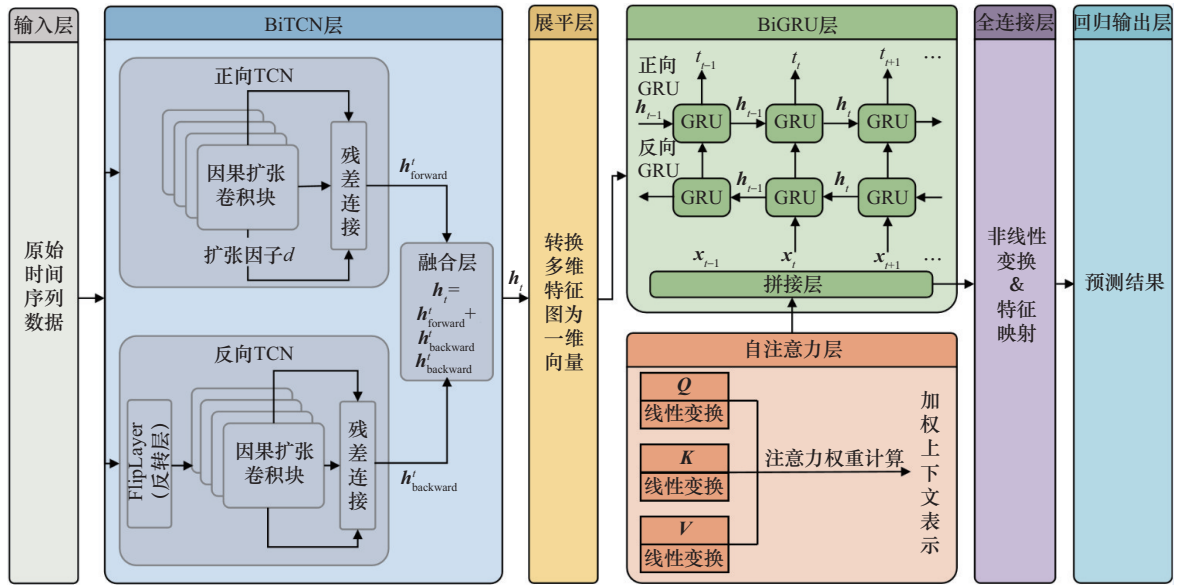


图1 BiTCN-BiGRU-Attention模型结构

### 3 基于RIME算法的深度时间序列预测模型优化

#### 3.1 RIME算法

RIME算法<sup>[14]</sup>是一种基于霜冰自然形成现象的优化策略，通过模拟霜晶在寒冷物体表面的附着和穿透性生长行为，在全局探索与局部开发之间取得平衡。算法动态结合软霜的聚集倾向和硬霜的精细挖掘行为，引入刺穿机制随机扰动粒子位置以逃离局部最优，逐步逼近全局最优解。在本文中，RIME算法被创新性地应用于优化BiTCN-BiGRU-Attention深度时间序列预测模型的超参数，包括学习率、BiTCN卷积核大小、BiGRU的神经元个数、注意力机制的键值以及正则化参数等关键参数。

##### 3.1.1 种群初始化

受自然界霜冰生长行为的启发，该方法将每个霜冰智能体视为算法的搜索智能体，将所有智能体形成的霜冰种群作为算法的种群。首先初始化整个霜冰种群 $R$ ，其中霜冰种群由 $n$ 个霜冰智能体 $S_i$ 组成，每个霜冰智能体由 $d$ 个霜冰粒子 $x_{ij}$ 组成，表示如下：

$$R = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_i \end{bmatrix}, S_i = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{ij}] \quad (7)$$

因此，霜冰种群 $R$ 可以直接表示为霜冰粒子 $x_{ij}$ ，如下：

$$R = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix} \quad (8)$$

其中， $i$ 为霜冰智能体的序号， $j$ 为霜冰粒子的序号。此外， $S_i$ 表示每个霜冰智能体的生长状态，即元启发式算法中智能体的适应度值。

##### 3.1.2 软霜搜索策略

受软霜增长的启发，本文提出一种利用霜冰粒子的强随机性和覆盖性的软霜搜索策略，使算法在迭代初期能够快速覆盖整个搜索空间，不易陷入局部最优。

由于环境因素的影响，软霜不会无限生长，最终会达到稳态。霜冰粒子的位置表示如下：

$$R_{ij}^{new} = R_{best,j} + r_1 \cdot \cos \theta \cdot \beta \cdot (h \cdot (UB_{ij} - LB_{ij}) + LB_{ij}), r_2 < E \quad (9)$$

其中， $R_{ij}^{new}$ 为更新后粒子的新位置， $ij$ 表示第 $i$ 个

霜冰智能体的第  $j$  个粒子； $\mathbf{R}_{\text{best},j}$  为霜冰种群  $\mathbf{R}$  中最佳软霜的第  $j$  个粒子； $r_1$  为  $(-1, 1)$  范围内的随机数，控制粒子的运动方向  $\cos \theta$  随迭代次数发生变化； $\beta$  为环境因子，它跟随迭代次数模拟外界环境的影响，用来保证算法的收敛性； $h$  为附着度，是一个在  $(0, 1)$  范围内的随机数，用来控制两个粒子中心之间的距离； $\mathbf{UB}_{ij}$  和  $\mathbf{LB}_{ij}$  分别为逃逸空间的上界和下界，它们限制了粒子运动的有效区域； $E$  为附着系数，它影响软霜的凝结概率，随着迭代次数的增加而增大； $r_2$  为在  $(0, 1)$  范围内的随机数，它与  $E$  一起控制粒子是否凝聚，即粒子位置是否更新。

$\theta$ 、 $\beta$ 、 $E$  的表达式如下：

$$\theta = \pi \cdot \frac{t}{10T} \quad (10)$$

$$\beta = 1 - \left[ \frac{\omega \cdot t}{T} \right] / \omega \quad (11)$$

$$E = \sqrt{t/T} \quad (12)$$

其中， $t$  为当前迭代次数， $T$  为算法的最大迭代次数， $\omega$  默认值为 5，用于控制阶跃函数的分段数。

### 3.1.3 硬霜刺破机制

在强风条件下，硬霜的生长比软霜的生长更简单，更有规律。该现象对应优化过程中的“强收敛行为”，即粒子以全局最优解为中心，经少量扰动迅速靠拢优良区域。硬霜刺破机制作用于搜索中后期，可加速粒子聚集。其更新通过引入归一化适应度与全局最优粒子，在保证多样性的同时实现方向性收敛，增强了 RIME 算法在超参数空间中的开发能力：

$$\mathbf{R}_{ij}^{\text{new}'} = \mathbf{R}'_{\text{best},j}, r_3 < F^{\text{normr}}(S_i) \quad (13)$$

其中， $\mathbf{R}_{ij}^{\text{new}'}$  为更新后的粒子的新位置， $\mathbf{R}'_{\text{best},j}$  为霜冰种群  $\mathbf{R}$  中最佳硬霜的第  $j$  个粒子， $F^{\text{normr}}(S_i)$  为当前智能体适应度值的归一化值，表示第  $i$  个霜冰智能体被选中的概率， $r_3$  为  $(-1, 1)$  范围内

的随机数。

### 3.1.4 正贪婪选择机制

本文提出了一种积极的贪婪选择机制来参与种群更新，以提高全局探索效率。具体是将更新后的智能体适应度值与更新前的进行比较，若更优则进行替换。该操作可用于确保种群在每次迭代时向更优方向进化，实现如下：

$$\begin{cases} \mathbf{R}_i = \mathbf{R}_i^{\text{new}}, F(\mathbf{R}_i^{\text{new}}) < F(\mathbf{R}_i) \\ \mathbf{R}_{\text{best}} = \mathbf{R}_i^{\text{new}}, F(\mathbf{R}_i^{\text{new}}) < F(\mathbf{R}_{\text{best}}) \end{cases} \quad (14)$$

其中， $\mathbf{R}_i$  为种群中第  $i$  个霜冰粒子的位置， $\mathbf{R}_{\text{best}}$  为霜冰粒子的最佳位置， $\mathbf{R}_i^{\text{new}}$  为第  $i$  个霜冰粒子更新后的位置， $F(\cdot)$  为智能体适应度值。

## 3.2 模型配置

在 RIME 算法优化之前，模型采用以下初始配置进行训练。

### (1) 超参数初值

初始学习率设置为 0.01，同时采用 Adam 优化器来自动调整学习率，适应模型的训练需求。批量大小设定为 128，最大训练次数设置为 50，梯度阈值设定为 1，L2 正则化参数设定为 0.000 1，这是基于实验性分析确定的，用以优化计算资源的使用和训练的稳定性。这些参数是模型训练的基础配置。

此外，BiTCN-BiGRU-Attention 模型的内部结构参数也进行了预设：BiTCN 层的卷积核数量为 64，卷积核大小为 5，空间丢弃率为 0.1，残差块数量为 1。BiGRU 层的神经元个数设置为 16，自注意力机制设置为单头注意力。

### (2) RIME 算法优化

利用 RIME 算法对 BiTCN-BiGRU-Attention 模型的关键超参数进行寻优。RIME 算法的优化目标是 minimized 模型的预测误差（适应度函数）。优化的超参数及其搜索范围见表 1，适应度函数以这些超参数作为输入，并返回模型在测试集上



的预测误差。

表1 优化的超参数及其搜索范围

超参数	搜索范围
学习率	[0.000 1, 0.01]
BiTCN卷积核大小	[1, 10]
BiGRU的神经元个数	[10, 50]
注意力机制的键值	[2, 50]
L2正则化参数	[0.000 01, 0.001]

### 3.3 RIME 算法模型优化设计

本文将 RIME 算法应用于 BiTCN-BiGRU-Attention 深度时间序列预测模型的超参数优化。RIME 算法通过模拟霜冰的自然形成过程，结合软霜搜索策略、硬霜刺破机制和正贪婪选择机制，以最小化模型的预测 MAPE 为目标，对模型的关键超参数进行寻优。基于 RIME-BiTCN-BiGRU-Attention 的深度时间序列预测模型的超参数优化流程如图2所示。

基于RIME算法优化BiTCN-BiGRU-Attention模型的伪代码如下。

**算法1** 基于RIME算法的深度时间序列预测模型优化

**输入：** 训练集  $vp\_train$ 、 $vt\_train$ ，测试集  $vp\_test$ 、 $T\_test$ ，输出数据归一化参数  $ps\_output$ ，种群规模  $popsiz$ e、最大迭代次数  $maxgen$ ，超参数搜索范围  $lb$ （下限）、 $ub$ （上限）

**输出：** 最佳超参数组合  $Best\_pos$ ，最佳适应度值  $Best\_score$ ，优化后的模型预测结果  $optimize\_T\_sim$

(1) 初始化 RIME 种群  $popsiz$ e 个霜冰粒子，每个粒子代表一组超参数配置，在  $lb$  和  $ub$  范围内随机生成；

(2) 计算初始适应度值：对于每个霜冰粒子，构建并训练 BiTCN-BiGRU-Attention 模型，

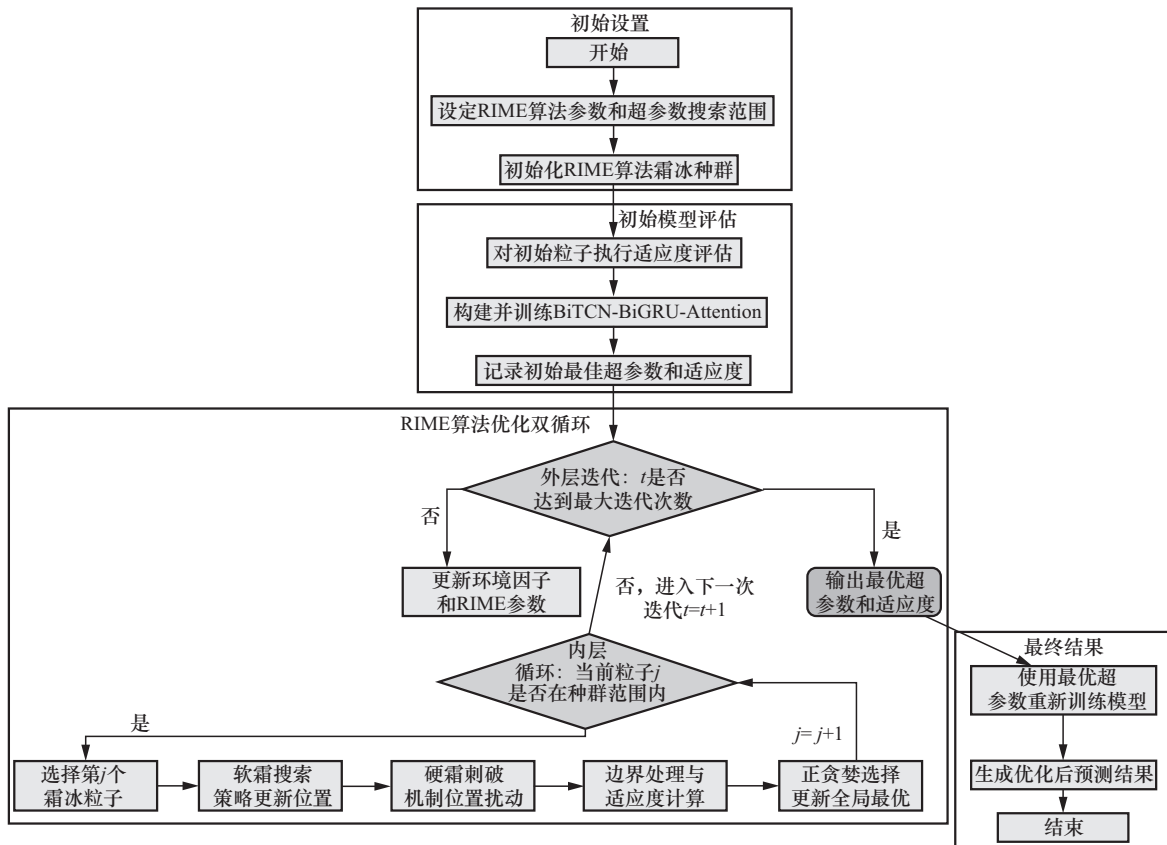


图2 超参数优化流程

计算其在测试集上的MAPE作为适应度值；

(3) 确定初始最佳粒子：记录当前种群中的最佳适应度值 Best\_score 和对应的最佳超参数组合 Best\_pos；

(4) or  $i=1$  to maxgen do

(5) 更新环境因子  $E$  和霜冰因子 RimeFactor；

(6) 归一化当前种群适应度值；

(7) for  $j=1$  to popsize do

(8) 生成随机数  $r_1$  和  $r_2$ ；

(9) if  $r_1 < E$  then

(10) 根据软霜搜索策略更新  $j$  的位置；

(11) end if

(12) if  $r_2 < \text{归一化适应度值}[j]$  then

(13) 根据硬霜刺破策略更新  $j$  的位置；

(14) end if

(15) 对更新后的粒子位置进行边界处理，计算更新后粒子  $j$  的新适应度值；

(16) if 新的适应度值  $<$  旧的适应度值 then

(17) 根据正贪婪选择机制更新粒子  $j$  的位置和适应度值；

(18) if 粒子  $j$  的适应度值  $<$  Best\_score then

(19) 根据正贪婪选择机制更新 Best\_score 和 Best\_pos；

(20) end if

(21) end if

(22) end for

(23) end for

(24) return Best\_pos、Best\_score；

(25) 使用最佳超参数组合 Best\_pos 重新训练模型，并获取优化后的模型预测结果 optimize\_T\_sim；

(26) end

最终，模型的性能通过多个指标进行评估，包括MSE、RMSE和MAE。这些指标可以全面反映模型在时间序列预测任务上的准确性和泛化能力。此外，为了评估模型的实用性，本文还将考

察模型在不同时间尺度和多变量预测任务上的表现。

## 4 实验仿真与分析

### 4.1 数据收集和处理

本文使用的数据集来源于Kaggle平台，涵盖了交通流量、股票市场和计算机网络流量3个领域的数据集，具体如下。

(1) 交通流量预测数据集：有48 000 h的观测数据，涵盖4个不同交叉路口的车辆流量信息，包括日期时间、路口编号、唯一ID和车辆数目。

(2) 股票市场数据集：涵盖纽约证券交易所、纳斯达克和纽约证券交易所的所有美国股票和交易所交易基金(exchange traded funds, ETF)的日常价格和交易量数据，记录自最后更新日起，包含开盘价、最高价、最低价、收盘价、成交量及未平仓合约数，且已对价格进行股息和拆分调整。

(3) 计算机网络流量数据集：记录了2万条在3个月内涉及10个本地工作站IP地址的网络流量数据，包括日期、本地IP地址、远程自治系统编号及流量信息，其中部分IP曾遭受攻击并一度成为僵尸网络的一部分。

在模型训练前，本文对上述时间序列数据进行了预处理，包括数据清洗(缺失值填补与异常值修正)和特征工程(构建滚动窗口统计量与差分特征)，以提升输入特征的有效性与训练稳定性。在数据分割方面，采用90%的数据作为训练集，10%作为测试集<sup>[15]</sup>，确保模型在未见数据的性能可以得到有效评估。

### 4.2 基准函数寻优

为了验证RIME算法的寻优性能，本文将其与黑翅鸢算法(black-winged kite algorithm, BKA)<sup>[16]</sup>、算数优化算法(arithmetic optimization algorithm, AOA)<sup>[17]</sup>、灰狼优化(grey wolf



optimizer, GWO) 算法<sup>[18]</sup>、改进麻雀搜索算法(improved sparrow search algorithm, ISSA)<sup>[19]</sup>进行了对比分析。本文选择了CEC2017基准函数集中的11个代表性函数进行测试,包括单峰函数( $F_1$ 、 $F_3$ )、简单多峰函数( $F_5$ 、 $F_{10}$ )以及组合函数( $F_{23}$ 、 $F_{24}$ )。为保证仿真实验控制变量的原则一致,设置每种算法种群数设为30,最大迭代数为500,测试的维度按照给定的基准函数进行设定,通过最优值、平均值和标准差这3个指标进行对比,算法测试结果对比见表2,其中每个函数对应的最优值用粗体标出。

从表2可知,RIME算法在大多数基准函数上都表现出了卓越的寻优能力。对于单峰函数 $F_1$ 、 $F_3$ ,RIME算法在最优值、平均值和标准差方面均显著优于对比算法,表明其在收敛速度和寻优精度上具有明显优势;对于多峰函数 $F_5$ 、 $F_{10}$ ,RIME算法同样展现出强大的全局搜索能力,能够有效避免陷入局部最优;对于组合函数 $F_{23}$ 、 $F_{24}$ ,RIME算法依然保持竞争力,进一步验证了其在复杂优化环境下的适应性和有效性。

综上所述,RIME算法在基准函数测试中表现出优异的寻优性能,尤其在求解精度、收敛速度和跳出局部最优能力方面均优于BKA、AOA、GWO和ISSA等对比算法,这为RIME算法在深度时间序列预测模型超参数优化中的应用奠定了

坚实的基础。

#### 4.3 实验仿真设计

时间序列预测模型的实验环境配置见表3。

表3 实验环境配置

名称	配置信息
操作系统	Ubuntu 18.04.5 LTS
开发语言	Python 3.6.13
框架	Pytorch 1.8.1 + cu11.6
CPU	Intel(R) Xeon(R) CPU E5-2620 v4 @2.10 GHz
GPU0	NVIDIA TITAN Xp 12 GB
GPU1	NVIDIA TITAN Xp 12 GB
运行内存	32 GB
硬盘内存	2 048 GB
开发软件	PyCharm 2023.1.2 (专业版本)

#### 4.4 性能指标

为了全面评估所提出的基于RIME-BiTCN-BiGRU-Attention的深度时间序列预测模型的性能,本文采用5类常用的回归与预测类评价指标进行分析,包括MAE、MAPE、MSE、RMSE、 $R^2$ 。这些指标从不同角度刻画预测误差的大小、方向与分布特性,能够比较全面地反映模型的拟合能力与预测精度。各指标计算如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

表2 算法测试结果对比

函数	$F_1$			$F_3$			$F_5$			$F_{10}$			$F_{23}$			$F_{24}$		
	最优值	平均值	标准差	最优值	平均值	标准差	最优值	平均值	标准差	最优值	平均值	标准差	最优值	平均值	标准差	最优值	平均值	标准差
BKA	5.58×10 <sup>7</sup>	6.74×10 <sup>9</sup>	1.09×10 <sup>10</sup>	1.26×10 <sup>4</sup>	2.60×10 <sup>4</sup>	1.96×10 <sup>4</sup>	6.54×10 <sup>2</sup>	7.36×10 <sup>2</sup>	3.76×10 <sup>1</sup>	4.08×10 <sup>3</sup>	5.28×10 <sup>3</sup>	8.19×10 <sup>2</sup>	2.97×10 <sup>3</sup>	3.14×10 <sup>3</sup>	1.14×10 <sup>2</sup>	3.09×10 <sup>3</sup>	3.27×10 <sup>3</sup>	9.36×10 <sup>1</sup>
AOA	4.01×10 <sup>10</sup>	5.35×10 <sup>10</sup>	9.94×10 <sup>9</sup>	5.30×10 <sup>4</sup>	7.70×10 <sup>4</sup>	1.29×10 <sup>4</sup>	7.95×10 <sup>2</sup>	8.90×10 <sup>2</sup>	3.01×10 <sup>1</sup>	6.41×10 <sup>3</sup>	7.48×10 <sup>3</sup>	<b>4.65</b> ×10 <sup>2</sup>	3.26×10 <sup>3</sup>	3.55×10 <sup>3</sup>	1.74×10 <sup>2</sup>	3.46×10 <sup>3</sup>	3.89×10 <sup>3</sup>	1.84×10 <sup>2</sup>
GWO	1.67×10 <sup>8</sup>	2.89×10 <sup>9</sup>	2.32×10 <sup>9</sup>	2.94×10 <sup>4</sup>	5.28×10 <sup>4</sup>	1.19×10 <sup>4</sup>	5.65×10 <sup>7</sup>	6.22×10 <sup>2</sup>	3.96×10 <sup>1</sup>	3.54×10 <sup>3</sup>	4.77×10 <sup>3</sup>	1.04×10 <sup>3</sup>	2.72×10 <sup>3</sup>	2.80×10 <sup>3</sup>	5.59×10 <sup>1</sup>	<b>2.87</b> ×10 <sup>3</sup>	2.95×10 <sup>3</sup>	6.09×10 <sup>1</sup>
ISSA	2.76×10 <sup>10</sup>	5.02×10 <sup>10</sup>	8.89×10 <sup>9</sup>	4.83×10 <sup>4</sup>	6.01×10 <sup>4</sup>	1.03×10 <sup>4</sup>	7.68×10 <sup>2</sup>	8.94×10 <sup>2</sup>	3.73×10 <sup>1</sup>	5.56×10 <sup>3</sup>	7.20×10 <sup>3</sup>	9.34×10 <sup>2</sup>	3.22×10 <sup>3</sup>	3.54×10 <sup>3</sup>	1.77×10 <sup>2</sup>	3.33×10 <sup>3</sup>	3.68×10 <sup>3</sup>	1.71×10 <sup>2</sup>
RIME	<b>9.81</b> ×10 <sup>2</sup>	<b>9.44</b> ×10 <sup>3</sup>	<b>7.22</b> ×10 <sup>3</sup>	<b>4.80</b> ×10 <sup>3</sup>	<b>1.46</b> ×10 <sup>4</sup>	<b>6.30</b> ×10 <sup>3</sup>	<b>5.53</b> ×10 <sup>2</sup>	<b>5.93</b> ×10 <sup>2</sup>	<b>2.02</b> ×10 <sup>1</sup>	<b>3.09</b> ×10 <sup>3</sup>	<b>4.45</b> ×10 <sup>3</sup>	6.90×10 <sup>2</sup>	<b>2.70</b> ×10 <sup>3</sup>	<b>2.77</b> ×10 <sup>3</sup>	<b>3.66</b> ×10 <sup>1</sup>	2.88×10 <sup>3</sup>	<b>2.93</b> ×10 <sup>3</sup>	<b>2.99</b> ×10 <sup>1</sup>

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (16)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

其中, MAE、MAPE、MSE、RMSE、 $R^2$ 分别为平均绝对误差、平均绝对百分比误差、均方误差、均方根误差和决定系数,  $y_i$ 为真实值,  $\hat{y}_i$ 为预测值,  $\bar{y}$ 为平均值。

### 4.5 实验结果分析

#### 4.5.1 网络流量数据预测

网络流量数据的真实值与预测值对比如图3所示。由图3可以看出, 优化后的模型预测曲线与真实值曲线的拟合度显著提高, 尤其是在峰值和谷值处, 优化模型能够更紧密地跟随真实值的变化, 减少了滞后和偏差。

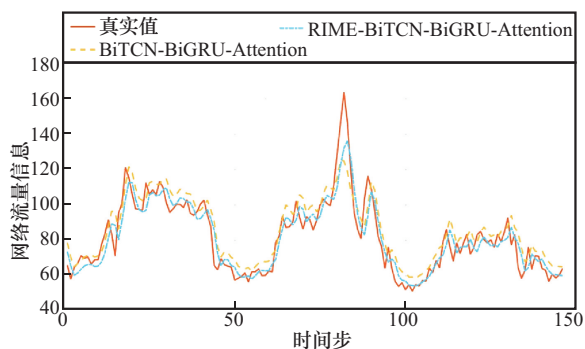


图3 网络流量数据的真实值与预测值对比

模型优化前、后的性能比较(网络数据集)如图4所示, 模型性能评估雷达图(网络数据集)如图5所示。由图4、图5可知, 经过RIME算法优化后的BiTCN-BiGRU-Attention模型在各项指标上均取得了显著提升。具体而言, MAE从7.743降低到5.977, 实现了22.8%的性能提升;

MAPE从9.77%降低到7.18%, 提升幅度达2.59个百分点; MSE从95.63降低到63.87, 改善了33.2%; RMSE从9.78降低到7.99, 提升了18.3%。同时,  $R^2$ 从77.65%提升至85.07%, 提升了7.42个百分点。

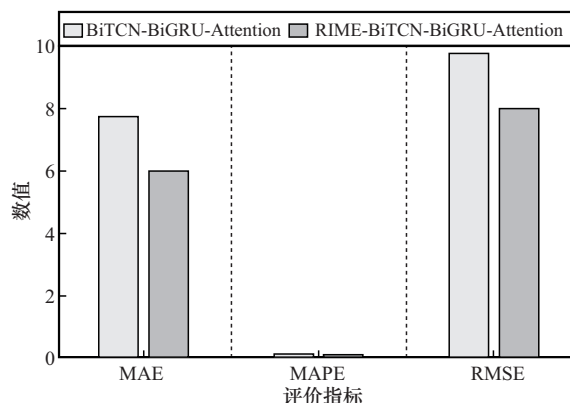


图4 模型优化前、后的性能比较(网络数据集)

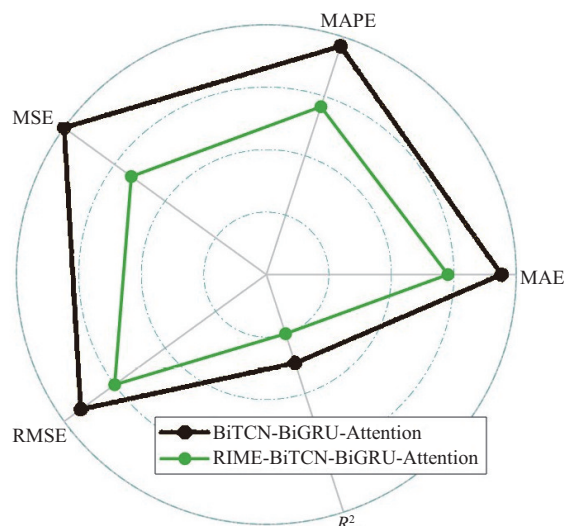


图5 模型性能评估雷达图(网络数据集)

本文采用的网络流量数据具有典型通信网络的业务特征, 反映了互联网主机间真实数据交互的动态变化。该类流量预测任务广泛存在于通信网络的拥塞预警、链路负载均衡、软件定义网络/网络功能虚拟化调度策略生成等典型应用场景中。RIME-BiTCN-BiGRU-Attention模型在该数据集上的显著性能提升表明, 该模型能够为通信网络中的智能管理、资源调度与安全防



护提供有效的预测支撑，具备明确的行业应用价值。

### 4.5.2 股票市场数据预测

股票市场数据真实值与预测值的对比如图6所示，预测曲线不仅能准确捕捉长期趋势，在局部快速上涨或下跌时，也能更迅速、更精确地调整预测，与真实值曲线的贴合度更高。

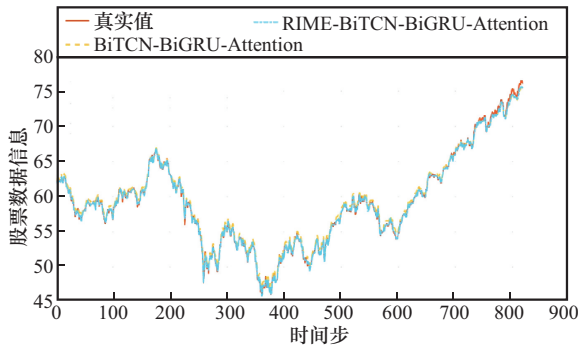


图6 股票市场数据真实值与预测值的对比

模型优化前、后的性能比较（股票市场数据集）如图7所示，模型性能评估雷达图（股票市场数据集）如图8所示。RIME算法优化后的模型在股票市场预测中展现出卓越的性能提升。MAE从0.474大幅降低至0.241，实现了49.2%的显著改善；MAPE从0.827%降低至0.408%；MSE从0.311降低至0.096，改善了69.1%；RMSE从0.557降低至0.31，提升了44.3%。 $R^2$ 值从99.294%提升至99.781%。

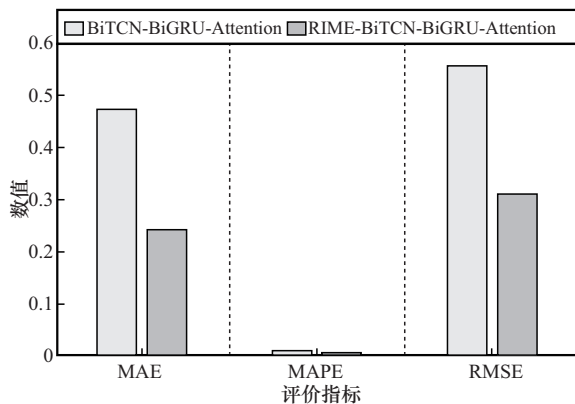


图7 模型优化前、后的性能比较(股票市场数据集)

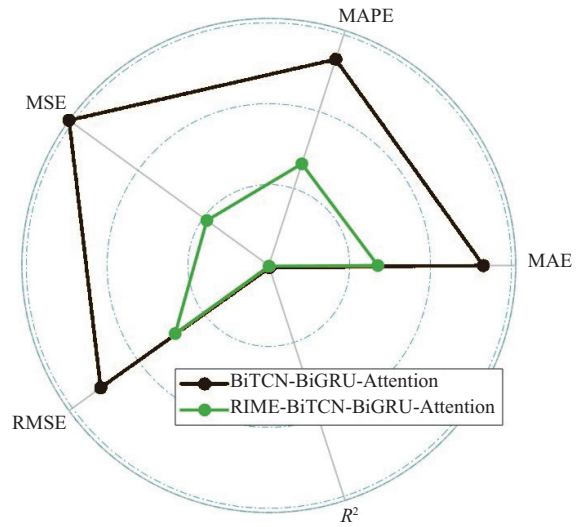


图8 模型性能评估雷达图(股票市场数据集)

股票市场数据以其高噪声、非线性和强随机性著称，对预测模型的鲁棒性要求极高。RIME算法的引入使得模型能够更有效地在复杂且不确定的金融市场数据中找到最优的超参数配置，显著提高了预测精度和模型的解释能力。尤其是在 $R^2$ 指标上接近100%的表现，进一步验证了优化模型在捕捉股票价格波动趋势上的强大能力。

### 4.5.3 交通流量数据预测

交通流量数据真实值与预测值的对比如图9所示。由图9可知，交通流量数据集中，模型在保持对交通流量周期性捕捉能力的同时，显著提升了对局部细节和异常波动的预测能力。预测曲线在峰值和谷值处与真实值更加吻合，能够更好地反映交通流量的瞬时变化。

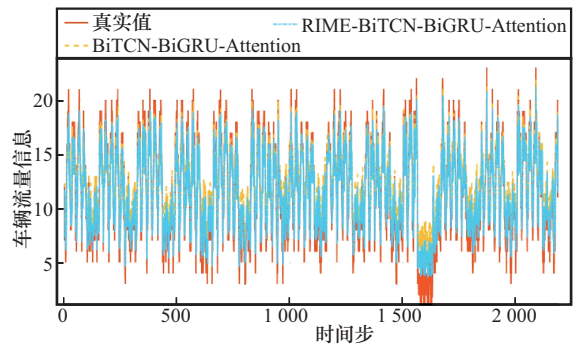


图9 交通流量数据真实值与预测值的对比

模型优化前、后的性能比较（交通流量数据集）如图10所示，模型性能评估雷达图（交通流量数据集）如图11所示。在交通流量预测任务中，RIME算法优化后的模型同样表现出显著的性能改善。MAE从2.335降低至2.083，改善了10.8%；MAPE从30.335%降低至23.950%，改善了6.385个百分点；MSE从8.339降低至6.851，改善了17.8%；RMSE从2.888降低至2.617，改善了9.4%。 $R^2$ 值从56.593%提升至64.342%，提升了7.749个百分点。

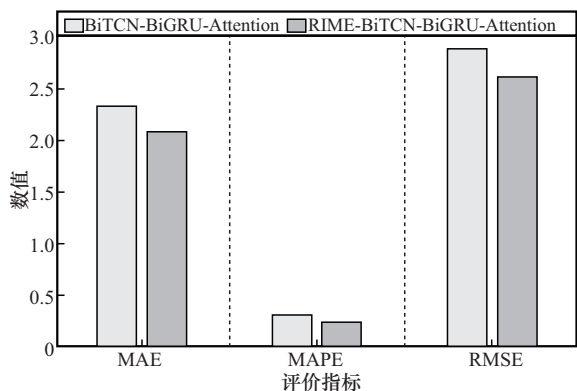


图10 模型优化前、后的性能比较(交通流量数据集)

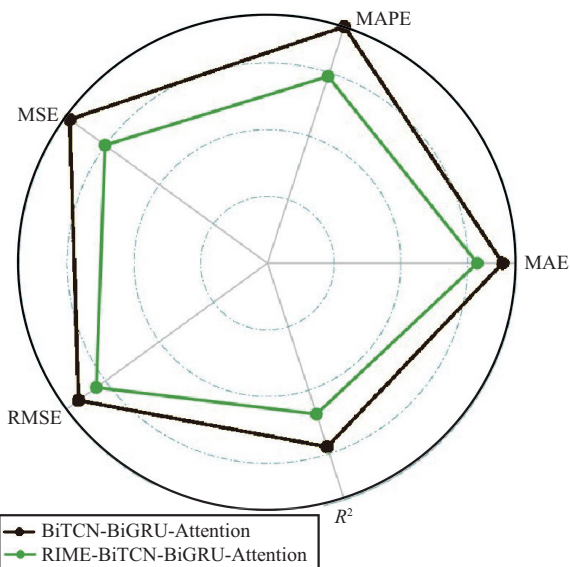


图11 模型性能评估雷达图(交通流量数据集)

交通流量数据通常具有明显的周期性、季节性和随机性，且易受外部事件影响。RIME算法

的优化使模型能够更好地适应这些复杂的时序模式，尤其是在处理交通流量的非线性动态方面表现出更强的能力。

综合来看，RIME算法在不同类型的时间序列数据集上均展现出强大的优化能力，这充分证明了RIME算法在超参数优化方面的有效性和泛化能力，为深度时间序列预测模型在实际应用中的性能提升提供了坚实的基础。

### 5 结束语

本文针对深度时间序列预测模型超参数优化的挑战，构建了BiTCN-BiGRU-Attention混合深度模型架构，并结合RIME算法进行超参数优化，实现了时序预测性能的显著提升。本文的主要贡献在于创新性地设计了BiTCN-BiGRU-Attention混合深度学习模型，有效融合了三者优势来捕获多尺度时序特征和长期依赖；将RIME算法应用于深度时序预测模型的超参数优化，通过其独特的搜索策略高效探索参数空间；在多个真实数据集上验证了优化后模型在RMSE、MAE和MAPE等关键指标上的显著性能提升，充分证明了所提方法的有效性，为时间序列预测技术发展提供了新思路和实践价值。

### 参考文献：

- [1] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.
- [2] 何勇, 李艳婷. 基于向量自回归模型的移动通信基站流量预测[J]. 工业工程与管理, 2017, 22(4): 79-84.  
He Y, Li Y T. Forecasting the traffic flow of base station based on vector auto-regression[J]. Industrial Engineering and Management, 2017, 22(4): 79-84.
- [3] 孙红哲, 王坚, 王鹏, 等. 基于 Attention-BiTCN 的网络入侵检测方法[J]. 信息安全, 2024, 24(2): 309-318.  
Sun H Z, Wang J, Wang P, et al. Network intrusion detection method based on attention-BiTCN[J]. Netinfo Security, 2024, 24(2): 309-318.



- [4] 徐海兵, 郭久明. 基于双向GRU模型的网络流量预测的研究[J]. 电子技术应用, 2022, 48(2): 19-22, 27.  
Xu H B, Guo J M. Research on network traffic prediction based on Bi-GRU model[J]. Application of Electronic Technique, 2022, 48(2): 19-22, 27.
- [5] 尹春勇, 曹儒商, 王琪凯. 基于TCN-BiGRU的网络异常检测研究[J]. 微电子学与计算机, 2025, 42(8): 120-131.  
Yin C Y, Cao R S, Wang Q K. Research on network anomaly detection based on TCN-BiGRU[J]. Microelectronics & Computer, 2025, 42(8): 120-131.
- [6] Morales-Hernández A, Van Nieuwenhuysse I, Rojas Gonzalez S. A survey on multi-objective hyperparameter optimization algorithms for machine learning[J]. Artificial Intelligence Review, 2023, 56(8): 8043-8093.
- [7] Snoek J, Larochelle H, Adams R P. Practical Bayesian optimization of machine learning algorithms[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2 (NIPS'12). 2012:2951 - 2959.
- [8] 路梦雨, 毛经坤, 赵洪阳. 基于贝叶斯优化的CNN-LSTM-Attention的个性化商品推荐算法[J]. 天津理工大学学报, 2024-04-30.  
Lu M Y, Mao J K, Zhao H Y. Personalized product recommendation algorithm based on Bayesian optimization of CNN-LSTM-Attention[J]. Journal of Tianjin University of Technology, 2024-04-30.
- [9] Fakhrmoosavi F, Kamjoo E, Kaviani-pour M, et al. A stochastic framework using Bayesian optimization algorithm to assess the network-level societal impacts of connected and autonomous vehicles[J]. Transportation Research Part C: Emerging Technologies, 2022, 139: 103663.
- [10] Kaveh M, Mesgari M S. Application of meta-heuristic algorithms for training neural networks and deep learning architectures: a comprehensive review[J]. Neural Processing Letters, 2023, 55(4): 4519-4622.
- [11] 杜伟, 王圣, 李健, 等. 基于CNN-LSTM-AM模型的储能锂离子电池荷电状态预测[J]. 电工技术学报, 2025, 40(9): 2982-2995.  
Du W, Wang S, Li J, et al. Prediction of state of charge for energy storage lithium-ion batteries based on CNN-LSTM-AM model[J]. Transactions of China Electrotechnical Society, 2025, 40(9): 2982-2995.
- [12] 段宇, 黄君, 杨关友, 等. 基于RIME-VMD-SSA-LSTM法研究非生态因素影响的来水预报模型[J]. 云南水力发电, 2024, 40(5): 44-50.  
Duan Y, Huang J, Yang G Y, et al. Research on inflow forecasting model of non-ecological factors based on RIME-VMD-SSA-LSTM method[J]. Yunnan Water Power, 2024, 40(5): 44-50.
- [13] Li W K, Yang X, Yin Y C, et al. A novel hybrid improved RIME algorithm for global optimization problems[J]. Biometrics, 2025, 10(1): 14.
- [14] Su H, Zhao D, Heidari A A, et al. RIME: a physics-based optimization[J]. Neurocomputing, 2023, 532: 183-214.
- [15] 蒋卓宇, 关维维, 孔祥力, 等. 基于IHOA-BiTCN-BiGRU-Attention的城市负荷信息聚合预测[J]. 电源学报, 2025-04-15.  
Jiang Z Y, Guan W W, Kong X L, et al. Urban load information aggregation forecasting based on attention mechanisms[J]. Journal of Power Supply, 2025-04-15.
- [16] Wang J, Wang W C, Hu X X, et al. Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems[J]. Artificial Intelligence Review, 2024, 57(4): 98.
- [17] Abualigah L, Diabat A, Mirjalili S, et al. The arithmetic optimization algorithm[J]. Computer Methods in Applied Mechanics and Engineering, 2021, 376: 113609.
- [18] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61.
- [19] Ren Q Y, Zhuge B, Zhang Z T, et al. Improved sparrow algorithm based virtual machine placement[J]. Cluster Computing, 2024, 27(5): 6511-6525.

## [作者简介]



葛斌 (1976-), 男, 博士, 浙江工商大学信息与电子工程学院教授, 主要研究方向为网络和通信技术、互联网技术和网络安全。



蔡晓丹 (2001-), 女, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为智慧网络和网络资源调度。



潘婷婷 (2000-), 女, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为智慧网络和网络资源调度。



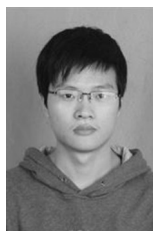
许云汉（2000-），男，浙江工商大学信息与电子工程学院硕士生，主要研究方向为云计算任务分配与资源管理、深度学习。



董黎刚（1972-），男，博士，浙江工商大学信息与电子工程学院教授，主要研究方向为智能网络、在线教育。



王正贤（2000-），男，浙江工商大学信息与电子工程学院硕士生，主要研究方向为计算机网络、深度学习和机器学习。



蒋献（1988-），男，浙江工商大学信息与电子工程学院讲师、实验员，主要研究方向为在线教育。



张子天（1988-），男，博士，浙江工商大学信息与电子工程学院副研究员，主要研究方向为基于机器学习的网络流量预测与资源管理。



于晓（1981-），男，浙江力积存储科技股份有限公司总经理，主要研究方向为存储芯片、高带宽内存技术。