

RIME-Transformer 模型在复杂时序预测问题中的应用

孙歆怡¹, 郑婷婷^{1,2*}, 孙丽雯¹

(1.安徽大学数学科学学院, 安徽 合肥 230601; 2.安徽大学大学数学教学中心, 安徽 合肥 230601)

摘要:为解决传统 Transformer 在长序列建模和计算效率的不足,本文提改进的 Transformer 模型,该模型在特征提取阶段引入多尺度卷积结构,通过并行卷积核在不同尺度上捕捉短期波动与长期趋势,增强对多层次时序模式的表征能力。随后,模型采用可学习的位置编码代替固定的正弦编码,更好地应对非平稳数据和不规则时间间隔问题。在全局依赖建模过程中,改进的编码器利用多头自注意力机制建立跨时间步的特征交互,动态分配时刻权重以聚焦关键片段,有效降低长序列建模的计算复杂度。Transformer 模型还结合了霜冰优化算法(rime optimization algorithm, RIME)在高维超参数空间中进行高效搜索与优化,提升模型的收敛速度与泛化能力。实验在3个真实复杂数据集上进行,结果表明 RIME-Transformer 模型在多项指标上均优于主流方法研究结果,验证所提模型在复杂时序预测任务中的有效性与优越性。

关键词:Transformer; 霜冰优化算法; 多尺度特征编码; 可学习位置编码; 注意力池化

中图分类号:TP183 **文献标志码:**A

引用格式:孙歆怡,郑婷婷,孙丽雯. RIME-Transformer 模型在复杂时序预测问题中的应用[J]. 山东大学学报(理学版),2026,61(5):79-89.

Application of RIME-Transformer model in complex time series prediction problems

SUN Xinyi¹, ZHENG Tingting^{1,2*}, SUN Liwen¹

(1. School of Mathematical Sciences, Anhui University, Hefei 230601, Anhui, China; 2. Center for University Mathematics Teaching, Anhui University, Hefei 23061, Anhui, China)

Abstract: To address the shortcomings of the traditional Transformer in long-term sequence modeling and computational efficiency, an improved Transformer model is proposed. This model first introduces a multi-scale convolutional structure in the feature extraction stage. Parallel convolution kernels capture both short-term fluctuations and long-term trends at different scales, thereby enhancing the representation of multi-level temporal patterns. Subsequently, the model employs learnable positional encoding instead of fixed sinusoidal encoding to better address the challenges posed by non-stationary data and irregular time intervals. During global dependency modeling, the improved encoder leverages a multi-head self-attention mechanism to establish feature interactions across time steps and dynamically assign moment weights to focus on key segments, effectively reducing the computational complexity of long-term sequence modeling. Furthermore, the model incorporates the rime optimization algorithm RIME for efficient search and optimization in a high-dimensional hyperparameter space, thereby improving the model's convergence speed and generalization ability. Experiments on three real-world complex datasets demonstrate that the RIME-Transformer outperforms mainstream methods across multiple metrics. These results validate the effectiveness and superiority of the proposed model for complex time series prediction tasks.

Key words: Transformer; rime optimization algorithm; multi-scale feature encoding; learnable positional encoding; attention pooling

0 引言

复杂时间序列通常指具有非线性动力学特征、多尺度耦合效应及非平稳演化趋势的时序数据^[1-3],其动

收稿日期:2025-04-10; 网络出版时间:2026-03-12

基金项目:国家自然科学基金资助项目(61806001)

第一作者:孙歆怡(2001—),女,硕士研究生,研究方向为智能计算与时间序列预测. E-mail:741580384@qq.com

*通信作者:郑婷婷(1978—),女,教授,硕士生导师,博士,研究方向为粒计算与知识发现. E-mail:tt-zheng@163.com

态演化过程往往受到多重周期因素(如季节性需求波动)与突发性事件(如地缘政治冲突)的交互影响。这类序列不仅包含传统时间序列的线性趋势和周期性规律,而且突变点密集(如极端价格波动)、长短期依赖交织(如供应链中断的滞后效应)以及时间间隔不规则(如非均匀采样的市场数据)等复杂特性。时间序列预测广泛应用于金融、经济、气象和医疗等领域,通过分析历史数据模式预测未来趋势。传统算法如自回归模型、自回归滑动平均模型和指数平滑法适用于稳定且规律性较强的数据,但难以处理复杂的非线性序列或长时间依赖关系^[4-6]。

传统的统计建模算法如自回归模型(autoregressive model)^[7]、自回归积分滑动平均(autoregressive integrated moving average, ARIMA)模型^[8]等在处理线性和稳定时间序列方面表现良好,但在面对复杂的非平稳、非线性特征时表现不足。为克服传统算法的局限,深度学习因其强大的特征提取和计算能力被广泛使用,在时间序列预测中展现出更高的准确度和适应性。Tokgöz等^[9]使用循环神经网络(recurrent neural networks, RNN)算法在预测土耳其电力负荷任务上取得了较好的效果;Sagheer等^[10]首次将长短期记忆网络(long short-term memory, LSTM)网络应用于石油产量时间序列预测,显著提升预测精度的模型;Kim等^[11]使用双向长短时记忆(bidirectional long short-term memory, BiLSTM)网络算法在多领域多变量的时间序列数据上取得了较好的预测效果。但是,基于RNN和LSTM模型在时序预测中容易面临梯度消失或爆炸问题,尤其是在长序列中,导致训练困难。随着Transformer^[12]模型以其自注意力机制在自然语言处理任务中取得了突破性进展,Transformer模型应用于时间序列预测领域,这个模型全局依赖建模能力使它在复杂时序建模任务中展现出强大的潜力,尤其在多变量建模、长期趋势捕捉方面具有优势,弥补了RNN和LSTM模型存在的不足。Zhou等^[13]通过改进Transformer模型,提出Informer,极大改善长时间预测的能力;Wu等^[14]提出Autoformer模型,提高模型找到长时间序列之间依赖关系的能力;Ahamed等^[15]运用Mamba模型采用选择性状态空间机制进行时序预测,极大减少模型训练和推理所需的计算资源;Lin等^[16]提出PETformer,用于长序列时间序列预测:将历史补丁嵌入与可学习的未来占位字符拼接后输入仅含编码器的Transformer,并以共享的逐字符线性头直接生成未来窗口,同时配合长子序列划分与多通道分离-交互以提升效率与精度。然而,Transformer模型在时间序列应用中仍存在诸多局限:一是其计算复杂度高,难以高效处理长序列数据;二是使用的正弦位置编码机制难以适应时间序列中不规则时间间隔,限制模型捕捉复杂时间依赖关系的能力。

为克服上述问题,本研究提出了一种改进的Transformer模型,引入多尺度特征编码增强模型对不同时间尺度趋势的感知能力,与传统的多尺度特征编码不同,该策略使得不同位置的初始方差相近,避免零初始化导致的表达瓶颈,也优于简单正态初始化在早期训练阶段产生的梯度爆炸风险,通过训练过程自适应调整位置向量,使模型能够捕捉数据驱动的时序位置关系;使用可学习的位置编码替代固定位置编码,以适应不规则时间间隔带来的建模挑战;通过注意力池化机制降低模型计算复杂度,实现对长序列的高效建模。此外,针对改进Transformer模型中的超参数优化问题,本文引入一种新颖的元启发式优化算法——霜冰优化算法(rime optimization algorithm, RIME)^[17]。该算法模拟雾凇冰形成过程中的动态演化机制,结合随机搜索与增量优化策略,能在高维超参数空间中实现高效、稳定的全局搜索,避免陷入局部最优,从而进一步提升模型在不同数据集上的鲁棒性与泛化能力。

为验证所提算法的有效性,本文用不同的时间序列数据为研究对象,将改进的Transformer模型与LSTM、Informer、Autoformer、Mamba和PETformer等典型时间序列预测算法进行对比分析,重点考察改进的Transformer模型在预测精度、计算效率与稳定性方面的表现。

1 预备知识

1.1 Transformer模型

Transformer模型是一种基于自注意力机制的模型,最早应用于自然语言处理^[18]领域,但它在时间序列预测任务中也取得了显著的成功。与传统的RNN或LSTM模型不同,Transformer模型通过自注意力机制(self-attention)处理序列中的所有元素,从而能够捕捉长期依赖关系,并且支持并行计算。Transformer模型主要由编码器和解码器组成。对于时间序列预测任务,通常只使用编码器部分。编码器将时间序列数据转

换为固定长度的表示捕捉输入数据中的模式,然后解码器根据这些表示生成预测值。在自注意力机制中,每个元素(时间步)都通过加权的方式与其他时间步进行交互,权重由输入序列中不同时间步的相关性决定。自注意力为

$$\mathbf{A} = F_{\text{softmax}} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} \right) \mathbf{V}, \quad (1)$$

式中: \mathbf{Q} 是查询矩阵, \mathbf{K} 是键矩阵, \mathbf{V} 矩阵, d 是键向量的维度, F_{softmax} 为softmax函数, T 为时间序列的长度。Transformer模型在时间序列预测中的优势主要体现在其能够捕捉长时间跨度的依赖关系,同时由于其并行计算的特性,也能大幅提高训练速度。

1.2 霜冰优化算法(RIME)

RIME是一种新型的元启发式算法,灵感来源于模拟霜冻冰的形成过程。该算法基于自然现象,同时改进了贪婪选择机制,实现探索和开发行为的优化。算法主要包括软霜搜索策略、硬霜穿刺策略以及正向贪婪机制。

软霜搜索策略模拟软霜粒子在霜冻冰中的运动,提出一种逐步搜索和开发的算法,用于探索算法。这种策略允许算法在大范围探索和小范围开发之间切换,实现高效和高精度的优化。粒子更新后的位置^[17]为

$$R_{ij}^{\text{new}} = R_{\text{best},j} + r_1 \cos \theta \beta \gamma (b_{Uij} - b_{Lij}), \quad (2)$$

式中: i 和 j 表示第 i 个Rime-agent的第 j 个粒子, $R_{\text{best},j}$ 为种群中最佳的第 j 个粒子,参数 r_1 为 $(-1,1)$ 内的随机数,并且 r_1 控制粒子运动的方向, $\cos \theta$ 将随着迭代次数的变化而改变, θ 为粒子的运动角度; β 为环境因子,跟随迭代次数以模拟外部环境的影响,并用于确保算法的收敛, γ 是 $(0,1)$ 内的随机数,表示黏附度,用于控制两个雾气粒子中心之间的距离, b_{Uij} 和 b_{Lij} 分别为逃逸空间的上界和下界,用于限制粒子运动的有效区域。硬霜策略使算法中的粒子能够相互交换,提高算法的收敛速度和跳出局部最优的能力。颗粒间置换的公式^[17]为

$$R_{ij}^{\text{new}} = R_{\text{best},j}, \quad r_2 < F_{\text{norm}}(S_i), \quad (3)$$

式中: F_{norm} 为当前代理适应值的归一化值,指示第 i 个Rime-agents被选择的机会; r_2 为 $(-1,1)$ 内的随机数, S_i 为第 i 个粒子的适应度值。正向贪婪机制用于元启发式优化算法中的种群更新,通过比较个体的更新适应度值与更新前的值,以决定是否替换个体,并同时替换这两个个体的解。这一机制不仅能够持续拥有优秀的个体,提高全局解的质量,还能够确保种群在每次迭代中朝更优方向演化。

RIME嵌入深度学习训练流程,用于自动寻找最优超参数组合。记待调优的超参数维度为 d (包括学习率 η 、批大小 B 、Dropout概率 p 等)。RIME中每个粒子的位置向量在迭代后通过公式(2)和(3)完成更新,随后将其线性映射到实际搜索空间,采用 K -折交叉验证评估目标网络在验证集上的平均损失,定义适应度函数为

$$J(h_i) = \frac{1}{K} \sum_{k=1}^K L_{\text{val}}^k(F(D_{\text{train}}^{(k)}, \mathbf{w}_i)), \quad (4)$$

式中: $F(\sim; \sim)$ 表示以超参数 \mathbf{w}_i 训练网络, \mathbf{w}_i 表示第 i 个粒子的超参数向量, K 为 K -折交叉验证, h_i 表示迭代中第 i 个粒子对应的RIME模型, L_{val}^k 为第 k 折的验证损失, $D_{\text{train}}^{(k)}$ 为第 k 折的训练数据, $k=1,2,\dots,N$ 。RIME通过最小化驱动 $J(\mathbf{w}_i)$ 种群向最优粒子 $h^* = \text{argmin}_i J(\mathbf{w}_i)$ 收敛。

2 基于RIME优化的改进Transformer时间序列预测模型

为了缓解现有算法在复杂时间序列预测任务上精度较低、稳定性较差的问题,本文提出一种基于RIME优化的改进Transformer架构时间序列预测模型,模型通过融合多尺度特征编码、自适应位置嵌入与注意力驱动池化机制,显著提升复杂时序模式的建模能力。如图1所示,模型采用分阶段渐进式设计。首先通过并行卷积提取局部多粒度特征,随后利用Transformer编码器建模全局时序依赖,最终通过内容感知池化动态聚合关键信息完成预测。以下从特征编码、时序建模与预测解码3个阶段详细阐述模型设计,图中 $\mathbf{Z}^{(l)}$ 、 $\mathbf{Z}^{(l+1)}$ 分别为注意力计算过程中的中间变量。

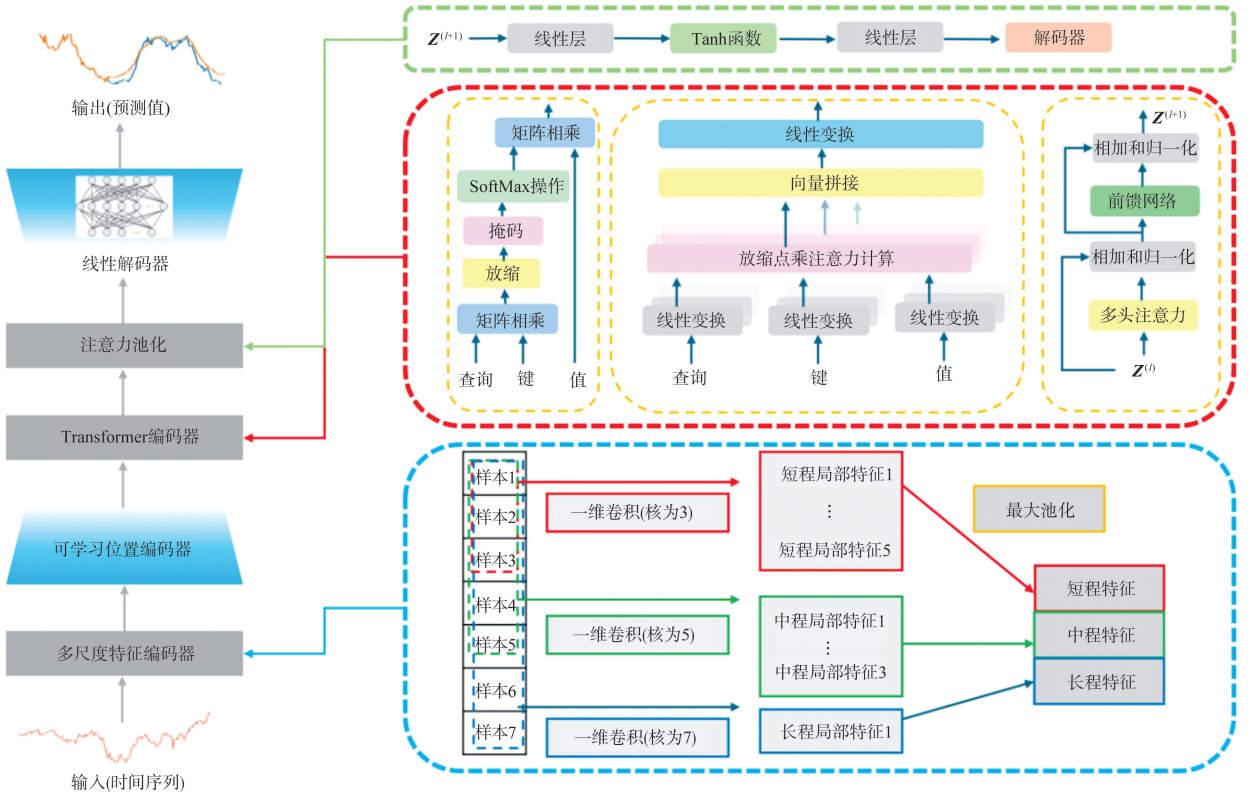


图1 RIME-Transformer 模型结构图

Fig.1 RIME-Transformer model structure

2.1 多尺度特征编码与动态位置嵌入

针对传统 Transformer 模型线性嵌入对局部时序模式不敏感的问题,本文受 Informer 中的 Prob 稀疏自注意力机制以及 Autoformer 中的序列分解机制以及 InceptionNet^[19] 的启发,构建多分支并行卷积结构。与多数图像处理中常见的 InceptionNet 设计不同,不同尺度分支的卷积参数不共享。输入序列 $X \in \mathbf{R}^{B \times L \times d_m}$ (B 为批大小, L 为序列长度, d_m 为每个序列元素的输入特征向量的维度, x 为序列 X 中的单个样本) 分别通过 3 个不同尺度的卷积核(分别为 3、5、7)提取短、中、长程局部特征(如图 1 所示),各分支拥有完全独立的卷积核参数,并且赋予相同的权重,以避免不同尺度特征被同质化。使用卷积核宽度为 c 的一维卷积后得到的特征输出为

$$H_k = F_{\text{RELU}}(\mathbf{W}_c * \mathbf{X} + b_k), \quad c = 3, 5, 7, \quad (5)$$

式中: \mathbf{W}_c 为卷积核参数, $*$ 表示一维卷积运算, b_k 表示对应的偏置项,每分支输出维度为 \mathbf{X} , F_{RELU} 为 RELU 函数。为进一步增强特征多样性,最大池化支路得到的特征输出为

$$\mathbf{H}_p = F_{\text{RELU}}(\mathbf{W} F_{\text{MaxPool}}(\mathbf{X})), \quad (6)$$

式中: $F_{\text{MaxPool}}(\mathbf{X})$ 表示对输入 \mathbf{X} 沿着序列长度维度 L 进行最大池化操作, \mathbf{W} 表示线性变换的参数矩阵,用于池化后特征的映射, F_{RELU} 表示激活函数,增加非线性能力。

通过沿特征维度拼接四分支输出,得到融合多尺度信息的编码结果 $H_{\text{incept}} = H_3; H_5; H_7$ 。为保留时序位置信息,传统 Transformer 模型采用固定频率的正余弦编码,但难以适配非平稳时序的复杂动态。为替代固定正弦位置编码,本文提出可学习位置编码矩阵 $\mathbf{E} = (E_{ij})$, (E_{ij} 表示第 i 个位置、第 j 个维度的编码值) 可学习位置编码矩阵 \mathbf{E} 尺寸等于最大序列长度 L_{max} 与特征维度 d_m 的乘积。为使各位置的初始方差一致并保持训练早期的数值稳定,采用 Glorot-Uniform(Xavier) 策略对该矩阵随机初始化, Glorot-Uniform(Xavier) 策略为

$$E_{ij} \sim U\left(-\sqrt{\frac{6}{L_{\text{max}} + d_m}}, \sqrt{\frac{6}{L_{\text{max}} + d_m}}\right), \quad (7)$$

该策略使得不同位置的初始方差相近,避免了零初始化导致的表达瓶颈,也优于简单正态初始化在早期训练阶段产生的梯度爆炸风险。

通过训练过程自适应调整位置向量,使模型能够捕捉数据驱动的时序位置关系,加上位置编码后的最终

序列表示为

$$\mathbf{Z}_{\text{pos}} = \mathbf{H}_{\text{incept}} + \mathbf{E}_{1:T}, \quad (8)$$

式中: $\mathbf{H}_{\text{incept}}$ 表示多尺度特征提取模块拼接输出的特征张量, $\mathbf{E}_{1:T}$ 表示可学习的位置编码矩阵的前 T 行,作为输入送入后续Transformer模块。该设计使位置编码不再局限于预设频率模式,尤其适用于具有突变点、周期性变化的复杂时序场景。

通过上述对多尺度特征提取模块的描述可以发现,本文提出的多尺度特征编码与Informer通过稀疏化注意力计算减少长序列处理的计算复杂度,以及Autoformer通过分解长期趋势与短期波动的机制提高模型在长序列预测中的效果不同,本文的多尺度特征编码算法采用并行卷积核,从多个尺度捕获局部特征,同时不同尺度的卷积核参数不共享,从而更细粒度地处理序列中多层次的信息。相比于Informer和Autoformer的全局依赖建模,本算法通过多尺度卷积捕捉序列的短期与长期变化,进而更加灵活地应对时序数据中的复杂波动性。

具体而言,Informer的Prob稀疏自注意力机制通过降低计算复杂度处理长序列,但它依赖于对全局依赖关系的筛选,缺乏对序列中不同层次变化的深度建模;Autoformer通过时间序列的分解机制将序列分为长期趋势和短期波动进行建模,具有较强的长时间依赖捕捉能力。然而,这些算法在处理复杂、多尺度的时序数据时可能忽视了局部特征的细节。相比之下,本文的多尺度特征编码模块通过独立卷积操作在多个尺度上捕捉到短期波动和长期趋势的特征,使模型在建模过程中可以同时关注局部细节和全局趋势,从而实现更加全面的时序建模。

2.2 全局依赖建模与自适应特征交互

多尺度特征经位置增强后输入至多层Transformer编码器,通过多头自注意力机制建立跨时间步的全局依赖关系。编码器由3个相同层堆叠而成,每层包含2个子模块:多头自注意力与前馈网络。对于第 l 层的输入,多头自注意力机制首先通过线性投影矩阵生成对应的查询矩阵、键矩阵和值矩阵分别为

$$\mathbf{Q} = \mathbf{Z}^{(l)} \mathbf{W}_Q, \quad (9)$$

$$\mathbf{K} = \mathbf{Z}^{(l)} \mathbf{W}_K, \quad (10)$$

$$\mathbf{V} = \mathbf{Z}^{(l)} \mathbf{W}_V, \quad (11)$$

式中 \mathbf{W}_Q 、 \mathbf{W}_K 、 \mathbf{W}_V 为注意力的投影矩阵。随后,利用缩放点积注意力计算时刻间的相关性,并对特征进行加权聚合,得到注意力输出为

$$\mathbf{A}^{(l)} = F_{\text{softmax}} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} \right) \mathbf{V}. \quad (12)$$

为了稳定训练并防止模型过拟合,在注意力机制后引入层归一化(LayerNorm)与Dropout操作,得到中间表征为

$$\mathbf{Z}_{\text{attn}}^{(l)} = F_{\text{LayerNorm}}(\mathbf{Z}^{(l)} + F_{\text{Dropout}}(\mathbf{A}^{(l)})), \quad (13)$$

该表征随后进入前馈网络,通过两层线性变换及ReLU激活函数进行非线性特征提取为

$$\mathbf{Z}_{\text{ffn}}^{(l)} = F_{\text{ReLU}}(\mathbf{Z}_{\text{attn}}^{(l)} \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2, \quad (14)$$

式中: F_{ReLU} 为ReLU激活函数, \mathbf{W}_1 为前馈网络第一层的权重矩阵, \mathbf{W}_2 为前馈网络第二层的权重矩阵, \mathbf{b}_1 为前馈网络第一层的偏置向量, \mathbf{b}_2 为前馈网络第二层的偏置向量。最终,通过第二层残差连接与归一化处理得到该Transformer层的最终输出为

$$\mathbf{Z}^{(l+1)} = F_{\text{LayerNorm}}(\mathbf{Z}_{\text{attn}}^{(l)} + F_{\text{Dropout}}(\mathbf{Z}_{\text{ffn}}^{(l)})). \quad (15)$$

2.3 内容感知池化与动态权重预测

传统时序模型常使用末时刻状态或简单平均池化进行序列汇总,但此类静态算法难以区分不同时间点的重要性。本文设计注意力驱动池化层,通过可学习参数动态计算各时刻贡献权重。首先,利用可学习参数,计算每个时刻 t 的隐藏得分向量为

$$\mathbf{e}_t = \tanh(\mathbf{W}_a \mathbf{z}_t + \mathbf{b}_a), \quad (16)$$

式中: a 为注意力机制, \mathbf{W}_a 为可学习参数, \mathbf{b}_a 为注意力层的偏置参数, \mathbf{z}_t 为 t 时刻的输入特征向量。通过Softmax函数对时刻 t 的重要性进行归一化处理,得到权重为

$$\alpha_t = \frac{e^{w_e^T \mathbf{e}_t}}{\sum_{t'=1}^T e^{w_e^T \mathbf{e}_{t'}}}, \quad (17)$$

式中: w_e 为可学习参数, T 为时间序列的长度。基于计算得到的权重, 对所有时刻的特征进行加权聚合, 生成汇总后的池化特征为

$$h_{\text{pooled}} = \sum_{t=1}^T \alpha_t h_t, \quad (18)$$

式中 h_t 为 t 时刻的隐藏层特征向量。该机制使模型能够根据输入序列内容自主聚焦关键时段(如趋势转折点或周期性峰值), 抑制噪声干扰。最终, 通过线性解码层将池化后的特征映射至预测空间, 得到最终预测结果为

$$\hat{y} = W_{\text{dec}} h_{\text{pooled}} + b_{\text{dec}}. \quad (19)$$

式中, W_{dec} 为解码层的权重矩阵, b_{dec} 为解码层的偏置向量。

模型训练采用均方误差损失函数, 通过 AdamW 优化器进行端到端优化。实验表明, 多尺度编码器与注意力池化的协同显著提升了模型对非平稳时序的建模能力。

2.4 基于 RIME 的超参数自适应优化

在模型训练过程中, 选择合适的超参数往往对预测性能具有决定性影响。为有效地探索庞大的超参数空间并避免陷入局部最优, 本节在种群中引入 RIME 引入并采用多样性策略并动态调整搜索方向, 实现对候选解的快速收敛, 同时兼顾跳出局部极值点的能力。本文模型能够在较少的迭代次数内识别出最优或接近最优的超参数组合, 提高模型的泛化性能和训练效率。

在模型训练过程中, 超参数选择对预测性能具有决定性影响。为高效探索超参数空间并避免局部最优, 针对多头注意力头数 (h) 及多尺度卷积分支层数 (C) 进行自适应优化。先定义搜索空间: h 的候选值为 2、

4、8、16, 满足 $d_k = \frac{d_m}{h}$ 的维度约束; L 的候选值为 1、2、3, 以平衡特征提取能力与计算复杂度。优化流程中, 随机初始化 50 组候选解 (h, L), 并通过验证集均方误差评估适应度, 目标为最小化均方误差。为维持种群多样性, 若连续 5 代适应度未提升, 则对 20% 个体施加高斯扰动以跳出局部最优; 同时基于历史最优解的梯度信息动态调整搜索步长。若当前代最优解优于全局最优, 则缩小步长以精细搜索, 反之扩大步长以探索新区域。通过锦标赛选择保留高适应度个体, 结合交叉与变异操作迭代更新种群, 直至达到最大迭代次数或收敛阈值。最终选取验证集最优的 (h^*, L^*) 组合重新训练并固定超参数, 部署至测试集进行预测。

实验表明, RIME 在 10 次迭代内即可逼近最优解, 显著提升模型收敛速度与泛化能力, 有效解决传统调参算法依赖经验与资源消耗大的问题, 为复杂时序任务提供了高效可靠的超参数配置方案。

2.5 RIME-Transformer 时间复杂度分析

本文提出的基于 RIME 优化的改进 Transformer 模型的时间复杂度主要由多尺度特征编码、位置嵌入、Transformer 编码器的多头自注意力机制以及注意力池化 4 个部分决定。假设输入序列的长度为 L , 特征维度为 d , Transformer 层数为 F , 注意力头数为 h , 则各部分的计算复杂度如下: 多尺度特征编码的时间复杂度为 $O(Ldc)$, 其中 c 为卷积核的大小; 位置嵌入的计算复杂度为 $O(Ld)$; 多头自注意力机制的计算复杂度为 $O(L^2d)$, 由于每个注意力头的计算可以并行, 因此总复杂度为 $O(L^2d)$; 前馈网络的计算复杂度为 $O(Ld^2)$, 而注意力池化的复杂度为 $O(Td)$ 。综合来看, 模型的总时间复杂度 $O(L^2d + Ld^2)$ 。若不加入多尺度卷积和注意力池化模块, Transformer 模型的总时间复杂度由自注意力机制主导, 总体时间复杂度为 $O(hL^2d)$ 。

在没有采用 RIME 的情况下, 超参数调优通常通过网格搜索或随机搜索进行。例如, 网格搜索的试验次数随着超参数(如注意力头数等)的增加会呈指数增长, 这会大幅提高训练时间。相比之下, 采用 RIME 后, 超参数的调优过程变得更加高效。RIME 结合了随机搜索和增量优化策略, 能够有效缩小超参数搜索空间, 避免陷入局部最优解。因此, RIME 可以通过减少所需的迭代次数, 加速最优配置的搜索过程。例如, RIME 的自适应调整搜索空间能力和有效的全局搜索能力, 使得在较少的迭代次数内就能找到最优或接近最优的超参数配置。

因此, 尽管模型的时间复杂度仍然由多头自注意力机制主导, 但引入 RIME 后, 可以显著减少超参数搜索过程中不必要的计算开销, 尤其在处理长序列时, RIME 能够提升模型的计算效率, 从而更好地应对复杂时序数据的建模挑战。通过减少超参数调优的计算成本, RIME 不仅提升了预测性能, 还增强了计算效率, 使得整个模型训练过程更加高效。

3 实验及相关结果

3.1 数据集与实验环境

为验证模型有效性,实验在秦皇岛港煤炭价格数据集(数据来源于中国煤炭市场网:www.cctd.com.cn)、全球气象变化数据集(数据来源于谷歌云盘:drive.google.com/file/d/1Tc7GeVN7DLEl-RAs-JVwG9yFMf-S8dy/view usp = share_link)及电力变压器温度数据集(数据集来源于 GitHub:github.com/zhouhaoyi/ETDataset)3 类复杂时序场景下进行。其中,秦皇岛港煤炭价格数据集数据集涵盖 2018—2022 年煤炭现货日粒度价格数据,包含价格波动、库存量及宏观经济指标等共 26 维特征,采用 1 000 个时间步;全球气象变化数据集采集全球 50 个城市近 5 年气象数据,涵盖温度、湿度、风速等 15 维特征,采用 3 000 个时间步,具有显著多变量耦合与非平稳特性;电力变压器温度数据集数据集记录电力变压器两年的负载率、油温等特征,步长为 5 000,需建模长期依赖与高频瞬态模式。

对于每个数据集,都按 8:1:1 划分为训练、验证与测试集,并通过 Z-Score 标准化消除量纲差异。

3.2 实验设置与实现细节

实验基于 PyTorch 框架在 NVIDIA RTX 4090 GPU 上完成,模型参数初始化采用 Xavier 正态分布,优化器选用 AdamW 并设置初始学习率为 0.001,配合余弦退火策略动态调节。训练周期为 200 轮,批量大小为 64,早停法用于防止过拟合。对比基线包括 LSTM、Informer、Autoformer、Mamba、PETformer 及标准 Transformer,算法输入窗口长度为 12,预测长度为 6。评估指标采用平均绝对误差、平均绝对百分比误差、均方根误差与平均绝对缩放误差。本文模型通过 RIME 优化超参数,搜索空间涵盖注意力头数(2~8)、卷积核尺寸(3~9)及池化温度系数(0.1~1.0),经 150 轮迭代确定最优配置。

为检验结果的稳健性,本文对所有模型(LSTM、Transformer、Informer、Autoformer、Mamba、PETformer 与 RIME-Transformer)均在 5 个不同随机种子下独立重复训练与测试。具体地,同时设置并固定 Python、NumPy 与 PyTorch 的随机种子;数据划分(训练/验证/测试)在首次实验时固定并在后续复现实验中保持不变;学习率调度与训练轮数等均保持一致。

3.3 多场景预测性能对比

在秦皇岛港煤炭价格数据集、全球气象变化数据集和电力变压器温度数据集 3 类复杂时序数据集上的实验表明,本文提出的 RIME-Transformer 模型相较于主流基线(LSTM、Transformer、Informer、Autoformer、Mamba 以及 PETformer)展现出显著优势,具体测试结果参见表 1—4。

表 1 不同模型在各个数据集上的平均绝对误差

Table 1 Mean absolute errors of different algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
LSTM	33.9	36.1	11.3
Transformer	31.5	32.4	8.0
Informer	21.2	27.7	6.6
Autoformer	26.0	26.5	7.5
Mamba	23.4	25.5	6.9
PETformer	19.2	20.2	6.3
RIME-Transformer	15.6	18.9	6.3

表 2 不同模型在各个数据集上的平均绝对百分比误差

Table 2 Mean absolute percentage errors of different algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
LSTM	4.5	7.8	3.1
Transformer	4.2	7.0	2.3
Informer	2.8	6.0	1.9
Autoformer	3.4	5.7	2.2
Mamba	2.9	5.1	2.0
PETformer	1.6	4.3	1.8
RIME-Transformer	1.3	4.1	1.8

表 3 不同模型在各个数据集上的均方根误差
Table 3 Root mean squared errors of different algorithms on various data sets

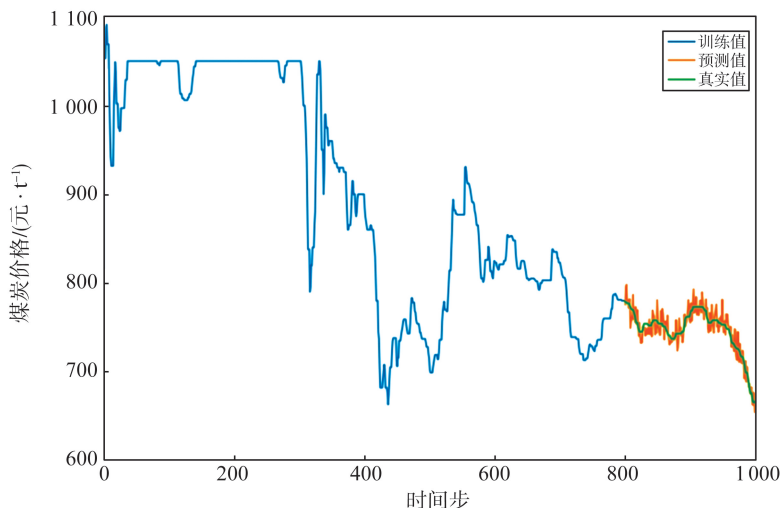
模型	秦皇岛港煤炭价格数据集	全球气候变化数据集	电力变压器温度数据集
LSTM	47.8	50.5	15.0
Transformer	45.7	47.3	14.2
Informer	38.3	42.1	12.6
Autoformer	42.5	44.6	13.3
Mamba	41.0	43.3	13.0
PETformer	39.2	41.0	12.0
RIME-Transformer	35.2	38.6	11.2

表 4 不同模型在各个数据集上的平均绝对缩放误差
Table 4 Mean absolute scaled errors of different algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气候变化数据集	电力变压器温度数据集
LSTM	1.23	1.35	1.29
Transformer	1.17	1.28	1.23
Informer	1.07	1.15	1.15
Autoformer	1.12	1.20	1.18
Mamba	1.10	1.18	1.17
PETformer	1.05	1.12	1.12
RIME-Transformer	0.98	1.05	1.08

在秦皇岛港煤炭价格数据集上,除平均绝对误差与平均绝对百分比误差分别较现有算法显著降低外,RIME-Transformer 的均方根误差达到 35.2、平均绝对缩放误差为 0.98,均为各模型中最佳,表明模型在抑制大误差与跨尺度可比性方面同样具有优势。在全球气候变化数据集上,模型取得均方根误差为 38.6、平均绝对缩放误差为 1.05,相比各基线算法进一步降低整体误差并提升跨场景可比性,配合动态位置编码与注意力池化,在寒潮等突变时段保持更稳定的误差曲线。在电力变压器温度数据集长时预测任务中,模型的均方根误差为 11.2、平均绝对缩放误差为 1.08,延续了在平均绝对误差和平均绝对百分比误差上的最佳值,表明全局注意力与动态权重聚合在长期依赖与高频瞬态并存的场景下同样能有效抑制大偏差。综合 3 类数据集,RIME-Transformer 的均方根误差与平均绝对缩放误差上较小且一致领先与收敛特性强,与平均绝对误差、平均绝对百分比误差的改进相互印证,进一步体现所提算法在复杂时序预测中的稳健性与泛化能力(如表 1—4 所示)。

可视化结果(图 2)显示,模型在煤炭价格拐点与气温突变区域的预测误差较传统模型(如 Transformer 模型)有较大地降低,且避免滞后或过冲现象。



(a) 秦皇岛港煤炭价格数据集

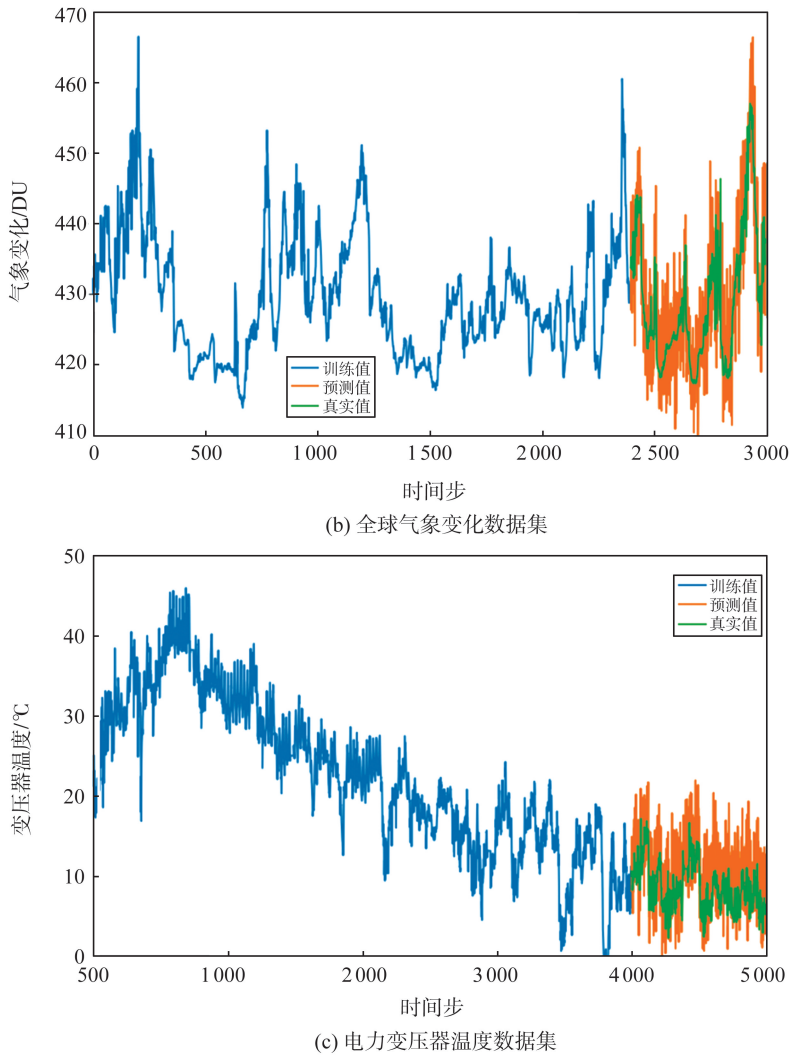


图 2 RIME-Transformer 在 3 种数据集上的预测结果
 Fig.2 RIME-Transformer model prediction results of on three datasets

综上,RIME Transformer 模型通过多尺度特征融合、自适应位置嵌入及 RIME 的超参数配置,在复杂时序场景中实现鲁棒性、精度与泛化能力的全面提升。

3.4 消融实验与结果

为验证模型各模块的有效性及其协同作用,本文设计 4 组消融实验:移除多尺度卷积(M1)、固定位置编码(M2)、替换注意力池化为平均池化(M3)及禁用 RIME 优化(M4)模型。实验表明,如表 5—8 所示,多尺度卷积通过融合短、中、长程特征显著提升模型性能,移除多尺度卷积(M1)在秦皇岛港煤炭价格数据集上使均方根误差、平均绝对缩放误差由 35.2、0.98 分别上升到 36.5、1.04,同时平均绝对误差由 15.6 上升至 18.5,表明模型对抑制大偏差与提升尺度无关性同等关键;固定位置编码(M2)对长期依赖更敏感,在电力变压器温度数据集上均方根误差由 11.2 上升到 12.7,为各消融中最差,同时平均绝对缩放误差由 1.08 上升至 1.13,显示数据驱动的位置表示对长期稳态与瞬态的协同建模的重要性;将注意力池化替换为平均池化(M3)在非平稳、突变更频繁的全球气象变化数据集上影响最大:平均绝对误差、平均绝对百分比误差、平均绝对缩放误差从 15.2、3.4、1.05 分别增大到 22.1、4.9、1.12,尽管均方根误差仅小幅增大(38.6 增大到 39.9),但整体与相对误差同步放大,揭示了内容感知聚合对关键片段的聚焦作用不可替代;禁用 RIME 超参数优化(M4)则在 3 个数据集上带来小幅但一致的退化,如秦皇岛港煤炭价格数据集上的平均绝对误差、平均绝对缩放误差由 15.6、0.98 分别增大到 16.3、1.02,全球气象变化数据集的平均绝对百分比误差、均方根误差也分别由 3.4、38.6 分别增大到 3.9、39.4,体现 RIME 在稳定收敛与抑制大误差方面的表现更好。

表5 不同消融模型在各个数据集上的平均绝对误差
Table 5 Mean absolute errors of different ablation algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
M1	18.5	16.5	6.1
M2	20.3	16.8	5.8
M3	18.1	22.1	5.8
M4	16.3	17.5	5.6
RIME-Transformer	15.6	15.2	5.5

表6 不同消融模型在各个数据集上的平均绝对百分比误差测试结果
Table 6 Mean absolute percentage error test results of different ablation algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
M1	2.5	3.7	1.8
M2	2.7	3.7	1.7
M3	2.4	4.9	1.7
M4	2.1	3.9	1.6
RIME-Transformer	1.3	3.4	1.6

表7 不同消融模型在各个数据集上的均方根误差测试结果
Table 7 Root mean squared error test results of different ablation algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
M1	36.5	40.5	12.5
M2	37.9	41.2	12.7
M3	36.2	39.9	12.1
M4	35.8	39.4	11.8
RIME-Transformer	35.2	38.6	11.2

表8 不同消融模型在各个数据集上的平均绝对缩放误差测试结果
Table 8 Mean absolute scaled error test results of different ablation algorithms on various data sets

模型	秦皇岛港煤炭价格数据集	全球气象变化数据集	电力变压器温度数据集
M1	1.04	1.10	1.11
M2	1.07	1.18	1.13
M3	1.03	1.12	1.11
M4	1.02	1.09	1.08
RIME-Transformer	0.98	1.05	1.08

综合4项指标的整体趋势,RIME-Transformer在3个数据集上相较任何单一组件被削弱的变体,均取得更好的性能。结合表5—8中各消融的变化可以看出:多尺度卷积与可学习位置编码是降低平均绝对误差、平均绝对百分比误差与均方根误差的主力,注意力池化对非平稳场景(如全球气象变化数据集)的误差稳定性尤为关键,而RIME优化提供跨数据集的一致稳健增益。上述结果表明,结构改进与自动化超参搜索在4项指标上形成互补,既能压低平均误差,又能有效抑制大偏差并保持跨尺度可比性,从而实现更全面的时序预测性能提升。

4 总结与展望

本文提出一种结合霜冰优化算法与Transformer模型的改进预测算法,通过引入多尺度特征编码、可学习位置编码以及注意力池化机制,有效增强模型对多尺度信息的表征能力及全局特征的捕捉效率,并利用RIME算法实现超参数的自适应优化。在3个真实数据集上的实验结果表明,改进后的模型平均绝对误差、平均绝对百分比误差、均方根误差和平均绝对缩放误差等各项评价指标均显著优于经典时序模型,验证模型在复杂时序预测任务中的有效性与优越性。尽管已取得良好效果,未来工作仍将致力于提升模型在不同类型数据集上的泛化能力,探索如稀疏注意力或轻量级架构等更高效的计算方案以应对大规模数据需求,并尝

试将该算法扩展至金融、气象及医疗监测等更复杂的多任务时序预测场景中。

参考文献:

- [1] LIM B, ZOHREN S. Time-series forecasting with deep learning: a survey[J]. *Philosophical Transactions of the Royal Society A*, 2021, 379(2194):20200209.
- [2] TORRES J F, HADJOUT D, SEBAA A, et al. Deep learning for time series forecasting: a survey[J]. *Big Data*, 2021, 9(1):3-21.
- [3] MASINI R P, MEDEIROS M C, MENDES E F. Machine learning advances for time series forecasting[J]. *Journal of Economic Surveys*, 2023, 37(1):76-111.
- [4] WANG Zihan, KONG Fanheng, FENG Shi, et al. Is mamba effective for time series forecasting[J]. *Neurocomputing*, 2025, 619:129178.
- [5] BHOGADE V, NITHYA B. Time series forecasting using Transformer neural network[J]. *International Journal of Computers and Applications*, 2024, 46(10):880-888.
- [6] KIM D K, KIM K. A convolutional Transformer model for multivariate time series prediction[J]. *IEEE Access*, 2022, 10:101319-101329.
- [7] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey[J]. *ACM Computing Surveys (CSUR)*, 2022, 54(10s):1-41.
- [8] SUNKI A, SATYAKUMAR C, NARAYANA G S, et al. Time series forecasting of stock market using ARIMA, LSTM and FB prophet[C]//MATEC Web of Conferences.[S.1]: EDP Sciences, 2024, 392:01163.
- [9] TOKGÖZ A, ÜNAL G. A RNN based time series approach for forecasting Turkish electricity load[C]//2018 26th Signal Processing and Communications Applications Conference (SIU). Izmir: IEEE, 2018:1-4.
- [10] SAGHEER A, KOTB M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. *Neurocomputing*, 2019, 323:203-213.
- [11] KIM J, MOON N. BiLSTM model based on multivariate time series data in multiple field for forecasting trading area[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020, 11(2):599-608.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30:5998-6008.
- [13] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. Informer: beyond efficient Transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia: AAAI, 2021, 35(12):11106-11115.
- [14] WU Haixu, XU Jiehui, WANG Jianmin, et al. Autoformer: decomposition Transformers with auto-correlation for long-term series forecasting[J]. *Advances in Neural Information Processing Systems*, 2021, 34:22419-22430.
- [15] AHAMED M A, CHENG Qiang. Timemachine: a time series is worth 4 mambas for long-term forecasting[C]//ECAI 2024: 27th European Conference on Artificial Intelligence. Santiago de Compostela: IOS Press, 2024, 392:1688.
- [16] LIN Shengsheng, LIN Weiwei, WU Wentai, et al. Petformer: long-term time series forecasting via placeholder-enhanced Transformer[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 9(2):1189-1201
- [17] SU Hang, ZHAO Dong, HEIDARI ALI ASGHAR, et al. RIME: a physics-based optimization[J]. *Neurocomputing*, 2023, 532:183-214.
- [18] CAMBRIA E, WHITE B. Jumping NLP curves: a review of natural language processing research[J]. *IEEE Computational Intelligence Magazine*, 2014, 9(2):48-57.
- [19] ZHONG Junliu, PUN Chiman. An end-to-end dense-inceptionnet for image copy-move forgery detection[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 15:2134-2146.

(编辑:陈丽萍)