WTConv-Informer 结合频率增强通道注意力的温度预测方法

王 曦， 张轩睿， 黄 海
（哈尔滨理工大学 计算机科学与技术学院，哈尔滨 150080）

摘 要：深度学习在温度预测中展现了建模非线性关系和复杂模式的优势，但在多维度气象数据处理及长期依赖性捕捉方面存在局限性。针对此问题，提出了基于 WTConv-Informer 结合频率增强通道注意力的温度预测方法。采用小波卷积与 Informer 模型结合，并引入频率增强通道注意力机制，实现全局特征的深层提取并精准提取时间序列中的低频主导特征和高频变化。实验采用马克思·普朗克生物地球化学研究所气象站数据集，结果显示：模型在预测精度上显著优于传统方法，其平均绝对误差为 0.206 4，均方误差为 0.085 4。研究表明：该模型在温度预测中精度高、泛化能力强，在多变量气象数据处理上具有显著优势。

关键词：温度预测；Informer；小波卷积；频率增强通道注意力；时间序列预测
DOI：10.15938/j.jhust.2026.04.008
中图分类号：TP301.6 **文献标志码**：A **文章编号**：1007-2683(2026)04-0082-12

Temperature Prediction Method Based on WTConv-Informer with Frequency Enhanced Channel Attention

WANG Xi， ZHANG Xuanrui， HUANG Hai
(School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: Deep learning has shown strong capabilities in modeling nonlinear relationships and complex patterns in temperature prediction. However, it faces limitations in processing multidimensional meteorological data and capturing long-term dependencies. To tackle this issue, this study proposes a WTConv-Informer based temperature prediction method enhanced with frequency-aware channel attention. By integrating wavelet convolution with the Informer model and incorporating a frequency-enhanced channel attention mechanism, the method enables deep extraction of global features and precise identification of dominant low-frequency trends and fine-grained high-frequency variations in time series. Experiments conducted on the Max Planck Institute for Biogeochemistry weather station dataset show that the proposed model outperforms traditional approaches in prediction accuracy, achieving a mean absolute error (MAE) of 0.206 4 and a mean squared error (MSE) of 0.085 4. The findings indicate that the proposed model delivers high accuracy and robust generalization in temperature prediction and demonstrates notable advantages in handling multivariable meteorological data.

Keywords: temperature forecasting; Informer; wavelet convolution; frequency-enhanced channel attention; time series forecasting

收稿日期：2024-12-26
基金项目：国家重点研发计划重点专项(2023YFB4403500).
作者简介：张轩睿(2000—)，男，硕士研究生；
黄 海(1982—)，男，教授，博士研究生导师.
通信作者：王 曦(1984—)，女，讲师，硕士研究生导师，E-mail：19990714510@163.com.

0　引　言

随着全球气候变化的加剧,极端天气事件频发,精确预测温度变化已成为应对气候变化的基础性任务[1]。然而,温度预测一直是气象领域中的一大挑战,传统的预测方法常受到非线性关系和长期依赖性捕捉能力不足的限制[2]。

随着机器学习技术的迅速发展,支持向量回归(support vector regression, SVR)[3]、K 近邻算法(K-nearest neighbor, KNN)[4]、随机森林(RF)[5]、支持向量机(support vector machine, SVM)[6]等算法开始被应用于时间序列预测中。这些方法虽能自学习数据规律,但在长序列建模时计算效率低,对长期依赖关系捕捉不足,难以处理大规模数据,影响实际应用的预测精度和鲁棒性[7]。

近年来,深度学习技术在温度预测中取得了显著进展。尤其是长短期记忆网络(long short-term memory, LSTM)[8-9]、卷积神经网络(convalutional neural network, CNN)[10]、变换器(Transformer)[11]及其变体等深度学习模型,由于其强大的非线性建模能力和自适应特征提取能力,已在时间序列预测中表现出较高的准确性。然而,这些模型在处理高维、多变量数据时,仍面临计算复杂度高、误差累积及局部极端变化捕捉不足等问题[12]。例如 Transformer 的变体 Informer[13]采用概率稀疏自注意力机制和生成式解码器进行长序列预测,提升了推理速度并降低了复杂度,但面对更加复杂的预测需求时,其对局部短期变化和高频波动的捕捉能力仍显不足。为解决这些问题,融合模型逐渐兴起,旨在结合不同模型的优势以提升预测性能。

陶晔等[14]将随机森林(RF)和 LSTM 结合,提高精度的同时也具备较强的泛化能力。杨迎新等[2]提出基于主成分分析(PCA)和改进粒子群算法优化门控循环单元(GRU)的递归神经网络时间序列预测模型用于气温预测,与传统神经网络预测模型对比,具有更高的预测精度和稳定性。JUN 等[15]提出基于 Informer 的气温预测模型,该模型融合实测观测数据与数值天气预报数据展开温度预测。构造时间周期特征并将其输入网络,使模型学习时序周期规律,有效提升了温度预测精度。王文豪等[7]提出了高斯过程驱动的 Autoformer(GPAformer)模型用于温度预测研究,结合高斯过程算子和自相关机制,增强了对温度时间序列变化的建模能力。

针对传统方法预测精度差、计算复杂的问题,本研究提出了结合小波卷积(wavelet convolutions, WTConv)[16]和 Informer 的温度预测模型(FECAM-WTConv-Informer),通过引入频率增强通道注意力机制(frequency enhanced channel attention mechanism, FECAM)[17],增强模型对温度变化频率成分的感知,从而更好地捕捉序列中的细节信息。WTConv 通过小波变换实现多尺度特征提取,结合小核卷积有效捕捉长序列中的局部波动和非平稳性,同时具备去噪能力。FECAM 强调全局频率信息,提升特征提取效率与全局建模能力,降低计算复杂度。相比传统深度学习方法,该模型在多变量气象数据处理、频率噪声处理和依赖捕捉方面表现出更强的泛化能力与预测稳定性,为温度预测提供了新的思路和技术支持。

1　模型原理

1.1　改进的小波卷积

小波卷积(WTConv)使用级联的小波变换(wavelet transform, WT)[18]对输入数据进行分解并应用一组小核卷积提取不同频带的特征,其原理是在逐步扩大的感受野上聚焦输入数据的不同频率分量[15],最初被应用于图像识别任务。为了使其适用于一维时间序列预测任务,需要对小波卷积的结构和操作方式进行相应调整。

对于 WTConv,给定的图像采用 Haar 小波变换[18],在 1 个空间维度(宽度或高度)上进行一级小波变换,通过组合 2 个维度的操作实现二维小波变换。通过文[13]给出的公式可知,使用以下 4 个过滤器,产生 1 个步幅为 2 的深度卷积:

$$\left.\begin{aligned} \boldsymbol{f}_{\mathrm{LL}}=\frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix},\boldsymbol{f}_{\mathrm{LH}}=\frac{1}{2}\begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix}\\ \boldsymbol{f}_{\mathrm{HL}}=\frac{1}{2}\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix},\boldsymbol{f}_{\mathrm{HH}}=\frac{1}{2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\end{aligned}\right\} \tag{1}$$

WT 通过上述卷积核对输入进行卷积和下采样,逆小波变换(IWT)通过转置卷积恢复原始输入,具体来说,单层的 WTConv 公式如下:

$$\boldsymbol{Y}=\mathrm{IWT}(\mathrm{Conv}(\boldsymbol{W},\mathrm{WT}(\boldsymbol{X}))) \tag{2}$$

式中:$\boldsymbol{X}$ 为输入张量;$\boldsymbol{W}$ 为 $k\times k$ 深度核的权张量;$\boldsymbol{Y}$ 为经过逆小波变换后的输出张量。再引入多层级

递归分解和频率组合完成小波卷积过程。

在二维空间中,小波变换使用二维卷积核来处理图像的高频和低频分量(即 $\boldsymbol{f}_{\mathrm{LL}}$ $\boldsymbol{f}_{\mathrm{LH}}$ $\boldsymbol{f}_{\mathrm{HL}}$ $\boldsymbol{f}_{\mathrm{HH}}$)。为应用于时间序列预测任务,将小波变换的卷积核转化为一维滤波器,使之能处理一维时序信号。

维度调整策略上将二维卷积转化为一维卷积,使其只作用于时间方向,并将小波分解后的高低频通道中所有卷积核的维度也调整对齐。二维中每个分解层生成4个分量(LL、LH、HL、HH)。对于一维时序信号,只需捕获低频和高频2个成分。因此,将二维滤波器进行维度压缩,经过归一化,生成2个分量($\boldsymbol{L}$ 和 $\boldsymbol{H}$)。定义低通滤波器 $\boldsymbol{L}$ 与高通滤波器 $\boldsymbol{H}$ 如下:

$$\boldsymbol{L}=\frac{1}{\sqrt{2}}[1,1],\boldsymbol{H}=\frac{1}{\sqrt{2}}[1,-1] \tag{3}$$

在小波分解上做出了调整,一维小波变换只需作用于1个维度,将分解后的结果存储为两部分:$\boldsymbol{X}_{\mathrm{L}}$(低频分量)和 $\boldsymbol{X}_{\mathrm{H}}$(高频分量)。同样的,逆小波变换过程及输入输出形状在实现方式上做出对应调整。

通过以上操作,在支持一维信号处理的需求上,做到了高效的频率特征提取,保留了适应非平稳信号的能力,并降低了计算量,节省了计算资源。

1.2 频率增强通道注意力

FECAM 通过离散余弦变换(discrete cosine transform, DCT)自适应建模通道间频率相互依赖关系,有效避免傅里叶变换中因周期问题导致的高频噪声[17](吉布斯现象)。该现象表现为在跳变点超调,无法通过增加正弦项消除从而引入噪声[19]。FECAM 通过 DCT 和通道注意力机制来解决这个问题。与传统注意力机制(如 SENet)相比,后者通过全局平均池化获取通道间的统计信息,并通过2个全连接层学习通道权重,通过对通道之间的相互依赖关系建模,从而增强重要特征通道并抑制冗余信息[20]。FECAM 不仅融合了空域信息,还引入了频域特征,提取输入特征在频率维度上的响应,捕捉时间序列中潜在的周期性与频率模式。这种融合方式使得 FECAM 可更全面地挖掘特征的时序结构。因此,FECAM 在对通道重要性进行评估时,可结合时间频率特征增强通道选择的准确性,显著提升模型性能。

通道注意力机制是通过挤压激发(squeeze-and-excitation, SE)块[21]利用全局信息对特征映射通道之间的相互依赖关系进行建模,并对特征映射进行重新校准。对于时间序列信号 $\boldsymbol{X}\in\boldsymbol{R}^{C\times L}$($C$ 为通道数,L 为时间序列长度),挤压步骤是在时间维度上生成通道描述符,而激发步骤就是使用2个具有瓶颈架构和非线性的完全连接层来建模通道依赖关系。

一维离散余弦变换基函数可表示为

$$B_{li}=\cos\left(\frac{\pi l}{L_s}\left(i+\frac{1}{2}\right)\right) \tag{4}$$

通过基函数计算离散余弦变换后的频率分量,则一维离散余弦变换可表示为

$$\boldsymbol{f}_{1\mathrm{d},l}=\sum_{i=0}^{L_s-1}\boldsymbol{x}_{i,1\mathrm{d}}B_{li} \tag{5}$$

式中:B_{li}为一维离散余弦变换的基函数值,$l\in\{0,1,\cdots,L_s-1\}$;$\boldsymbol{f}_{1\mathrm{d},l}\in\boldsymbol{R}^{L}$ 为一维离散余弦变换的频谱;$\boldsymbol{x}_{1\mathrm{d}}\in\boldsymbol{R}_{L}$ 为输入信号;l 为 $\boldsymbol{x}_{1\mathrm{d}}$的长度;$L_s$ 为时间维度。

FECAM 的处理过程如下:首先,FECAM 输入的特征图谱沿通道维度拆分为 n 个子组,即$[\boldsymbol{v}_0,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-1}]$,其中 $\boldsymbol{v}_i=\boldsymbol{R}^{1\times L}$,$i\in\{0,1,\cdots,n-1\}$,$n=N_v$。随后,每个子组将被处理为对应的离散余弦变换($f_{\mathrm{DCT}}$)频率分量,频率范围从低到高,每个通道将使用相同的频率分量进行处理,得到:

$$\boldsymbol{F}_i=f_{\mathrm{DCT},j}(\boldsymbol{v}_i)=\sum_{j=0}^{j=L_s-1}(\boldsymbol{V}_{i:,l})\boldsymbol{B}_{lj} \tag{6}$$

式中:$\boldsymbol{F}_i\in\boldsymbol{R}^{L}$ 为离散余弦变换后的 L 维向量,$i\in\{0,1,\cdots,N_v-1\}$;$j\in\{0,1,\cdots,L_s-1\}$ 为对应于 $\boldsymbol{v}_i$ 的频率分量一维指标。整个信道矢量可以通过叠加运算得到:

$$\boldsymbol{F}=f_{\mathrm{DCT}}(\boldsymbol{v})=\mathrm{stack}([\boldsymbol{F}_0,\boldsymbol{F}_1,\cdots,\boldsymbol{F}_{n-1}]) \tag{7}$$

stack 的作用是将向量沿通道维度堆叠起来。一旦得到了 $\boldsymbol{F}$,注意权值就可以通过神经结构作为 SE 块来学习。整个频率增强通道注意机制框架可以写成:

$$\boldsymbol{F}_c-\boldsymbol{F}_{\mathrm{att}}=\sigma(W_2\delta(W_1f_{\mathrm{DCT}}(\boldsymbol{v}))) \tag{8}$$

式中:$\boldsymbol{F}_{\mathrm{att}}\in\boldsymbol{R}^{\mathrm{C}}$ 为激励步骤学习到的关注向量;σ 和 δ 分别为 ReLU 和 sigmoid 激活函数。每个信道特征与每个频率分量相互作用,从频域全面获取重要的时间信息,从而增强提取特征的多样性[17]。

1.3 Informer 预测模型

Informer 模型是基于 Transformer 模型的高效长序列时间序列预测模型,在长时间序列建模中表现出显著优势[22]。其主要特点包括:采用概率稀疏自注意力机制,提升序列依赖关系捕捉性能;对级联层输入进行降维,突出主要注意力区域,有效处理超长输入序列;引入生成式解码器,1次向前计算完成长序列预测,显著提升推理效率[13]。在多变量非平稳

的温度时间序列预测中,Informer 可有效避免误差累计。在编码层中,概率稀疏自注意力机制替代传统的自注意力机制,通过自注意力蒸馏降低特征维度从而提升模型的运行效率。在解码层中,引入带掩码的多头注意力机制,减少解码时间并有效避免整个网络的误差累积[23]。模型结构图如图 1 所示,包含编码器和解码器 2 个主要模块。

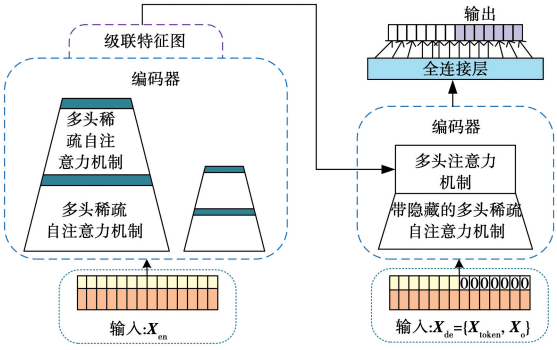

图 1　Informer 模型网络结构

Fig. 1　Informer model network architecture

编码器由多头概率稀疏自注意力层和蒸馏层组成,自注意力机制的计算基于查询、键和值的乘积,通常通过缩放点积来实现。概率稀疏自注意力计算需要利用输入序列经过线性变换得到的 3 个矩阵:查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$。其中,$\boldsymbol{Q}\in\boldsymbol{R}_{L_Q\times d}$、$\boldsymbol{K}\in\boldsymbol{R}^{L_K\times d}$、$\boldsymbol{V}\in\boldsymbol{R}^{L_V\times d}$,$L_Q$、$L_K$、$L_V$ 和 d 分别为查询矩阵、键矩阵、值矩阵和输入序列的维度[24]。而概率稀疏自注意力机制通过限制每个键只关注最重要的查询来实现自注意力计算的稀疏化。因此,Informer只关注前 u 个重要性强的 $\boldsymbol{Q}$,如式(9)所示。

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}\left(\frac{\bar{\boldsymbol{Q}}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{9}$$

式中:$\bar{\boldsymbol{Q}}$ 为对 $\boldsymbol{Q}$ 概率稀疏化后得到的值,即前 u 个查询矩阵的稀疏矩阵;Attention($\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$)为概率稀疏自注意力机制;$\boldsymbol{K}^{\mathrm{T}}$ 为键矩阵的转置;d 为输入维度。蒸馏机制的原理如下:

$$\boldsymbol{X}_{j+1,t}=\text{MaxPool}(\text{ELU}(\text{Conv1d}([\boldsymbol{X}_{j,t}]_{AB}))) \tag{10}$$

式中:$[\,.\,]_{AB}$为注意块;$[\boldsymbol{X}_{j,t}]_{AB}$包含多头稀疏自注意力机制;MaxPool 为最大池化操作;ELU 为激活函数;Conv1d 为序列上的一维卷积操作。

最后生成式解码器通过 1 个前向过程生成输出序列,如下:

$$\boldsymbol{X}_{\text{de},t}=\text{Concat}(\boldsymbol{X}_{\text{token},t},\boldsymbol{X}_{0,t})\in\boldsymbol{R}^{(L_{\text{token}}+L_y)\times d_{\text{model}}} \tag{11}$$

式中:$\boldsymbol{X}_{\text{de},t}$ 为 Informer 解码器的输入时间序列;$\boldsymbol{X}_{\text{token},t}$为开始标记,即有 L_{token}个时间步,每个时间步的维度为 d_{model};$\boldsymbol{X}_{0,t}$为目标序列的占位符;Concat 为拼接操作;L_y 为目标序列长度。

解码器通过掩码自注意力机制确保预测仅依赖已知数据,结合编码器输出并经全连接层生成长序列预测结果。针对温度预测的时间与季节性因素,Informer 将时间特征与位置向量作为输入,利用多头注意力机制减少过拟合。解码器一次性生成完整预测序列,显著缩短解码时间并提升预测效率。

2　WTConv-Informer 结合 FECAM 的温度预测模型

2.1　数据描述及预处理

为了对提出的模型进行实验验证,历史天气数据集取自马克斯·普朗克生物地球化学实验室[25]的 Beutenberg 气象站及 Saaleaue 气象站 2023 年 1 月 1 日 00:10 至 2024 年 1 月 1 日 00:00 的观测数据,数据每 10 min 更新 1 次。其中 Beutenberg 气象站数据集包括时间日期及 21 个气象相关属性,Saaleaue 气象站数据集包括时间日期以及 30 个气象相关属性。

由于温度要素计量单位和取值范围不同,数值范围较大的特征可能导致模型权重分配不均,影响预测性能。通过归一化将特征值缩放至[0,1]区间,可减少梯度下降中的振荡,加速训练并避免数值不稳定性。归一化公式如下:

$$T^*=\frac{T-T_{\min}}{T_{\max}-T_{\min}} \tag{12}$$

式中:T^* 为归一化后的数据;T 为原始数据;$T_{\min}$为各个温度要素值的最小值;$T_{\max}$为各个温度要素值的最大值。

2.2　预测模型构建

温度时间序列数据具有长期趋势、非平稳性、季节性周期和不规则变化等特征。为提升预测精度和稳定性,提出 FECAM-WTConv-Informer 模型。

FECAM 通过 DCT 提取频率分量,增强通道注意力,强调全局频率信息并解决吉布斯问题;WTConv 实现多尺度特征提取,分离低频趋势与高频细节,有效抑制噪声,捕获复杂动态变化;结合 Informer 在长序列建模和周期性特征预测上的优势,实现高效准确的温度预测。构建的温度预测模型流程图如图 2 所示。

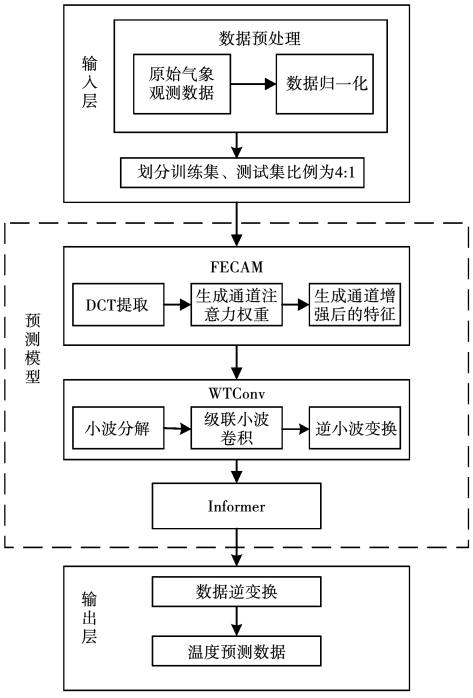

图 2 FECAM-WTConv-Informer 预测模型流程图

Fig. 2 FECAM-WTConv-Informer prediction model flowchart

模型首先对原始气象数据进行归一化处理并划分训练集与测试集,训练集与测试集比例为 4∶1。随后,FECAM 提取频率特征并增强通道注意力,WTConv 与 Informer 模型结合完成多尺度特征提取和长序列建模。最终,将预测结果反归一化后进行误差评估并输出。

3 实验分析

3.1 实验环境及参数

本文所有模拟方法均使用 Python 3.8 版本和 PyTorch 深度学习框架实现,并运行在 14 代英特尔 Intel(R) Core(TM) Ultra 9 185H 2.30 GHz 处理器和 NVIDIA GeForce RTX 4060 GPU(8 GB)的计算环境上。模型内部参数:使用批次大小为 64,丢弃率为 0.1,学习率为0.000 1,模型维度为 32,时间频率为 t,注意力因子为 5,使用 GeLU 激活函数,并使用了蒸馏。

3.2 评价指标

本文以决定系数(coefficient of determination, R^2)、均方误差(mean square error, MSE)、平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE)作为评价预测结果的指标。

决定系数 R^2 是评估模型解释能力的指标,表示预测结果能够解释实际数据总体方差的比例,计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_l)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_l)^2} \tag{13}$$

式中:y_i 为第 i 个样本的真实值;$\bar{y}_i$ 为样本真实值的均值;$\hat{y}_l$ 为第 l 个样本的模型预测值。

均方误差 MSE 预测误差平方的平均值,反映预测值与真实值之间的偏差程度。MSE 越低,说明模型的预测准确性越高,计算公式如下:

$$\alpha_{\mathrm{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_l)^2 \tag{14}$$

平均绝对误差 MAE 预测误差绝对值的平均值,直接表示预测值与真实值的平均差距。相比 MSE 和 RMSE,MAE 对异常值不敏感,不会因少数大误差的平方放大结果,计算公式如下:

$$\alpha_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - \hat{y}_l \right| \tag{15}$$

均方根误差 RMSE 是预测误差平方的平均值开方后的结果,表示均方根差异,用于衡量预测值与实际值之间的差距大小。与 MSE 相比,RMSE 保持了与原数据相同的量纲,更便于直观理解误差大小,计算公式如下:

$$\alpha_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{16}$$

式(13)~(16)中:y_i 为真实数据中的第 i 个样本值(即真实值);$\hat{y}_i$ 为使用模型得到的相应预测值。

3.3 实验结果分析

3.3.1 消融实验对比分析

为验证模型局部模块对预测效果的影响,消融实验将改进的小波卷积和 FECAM 分别与基础 Informer 模型结合,并与主模型 FECAM-WTConv-Informer对比,验证主模型的预测效果。

在 Beutenberg 气象站数据集上 FECAM-WTConv-Informer 主模型与 WTConv-Informer 模型的局部预测对比如图 3 所示,与 FECAM-Informer 模型局部预测对比如图 4 所示。

由图 3 可知,WTConv 与 Informer 结合的预测效果优于单独的 Informer 模型,但整体不及主模型。

由图 4 可知,FECAM 与 Informer 结合的预测效果不如主模型,基本优于基础 Informer 模型,进一步验证了 FECAM 在主模型中的作用。实验评价指标对比如表 1 所示。

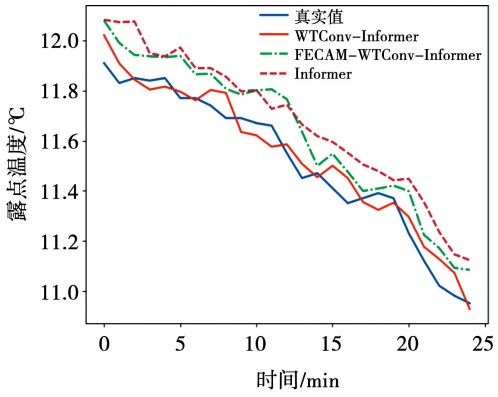

图 3　主模型与 WTConv-Informer 模型对比图

Fig. 3　Comparison chart between the main model and the WTConv-Informer model

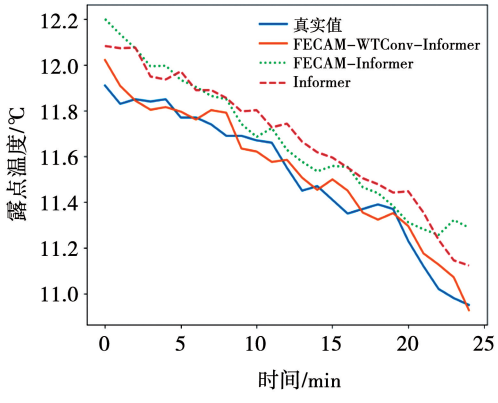

图 4　主模型与 FECAM-Informer 模型对比图

Fig. 4　Comparison chart between the main model and the FECAM-Informer model

表 1　Beutenberg 数据集消融实验结果对比表

Tab. 1　Ablation experiment results comparison on Beutenberg dataset

模型	R^2	α_{MAE}	α_{MSE}	α_{RMSE}
主模型	0.997 4	0.137 4	0.042 2	0.205 5
FECAM-Informer	0.994 2	0.214 5	0.094 6	0.307 6
WTConv-Informer	0.997 3	0.149 7	0.043 5	0.208 6
Informer	0.996 8	0.153 5	0.051 6	0.227 1

由表 1 可知,主模型各项指标对比 Informer 均有较大提升,单独的 WTConv 对模型提升较小,而只结合 FECAM 反而会造成预测结果下降和预测波动大的问题。

在拥有更复杂特征值的 Saaleaue 气象站数据集上,主模型与 WTConv-Informer 模型的局部预测对比结果如图 5 所示,与 FECAM-Informer 模型的局部预测对比结果如图 6 所示。

由图 5 可见,随着数据集特征复杂度的提升,主模型依然表现出优势,而 WTConv-Informer 模型局部表现较基础 Informer 模型提升不大。

由图 6 可见,仅结合 FECAM 无法大幅提升基础 Informer 模型的预测能力,并且在局部会引发较大的高频偏差,而主模型则可以保持较好的稳定性。更详细的实验评价指标如表 2 所示。

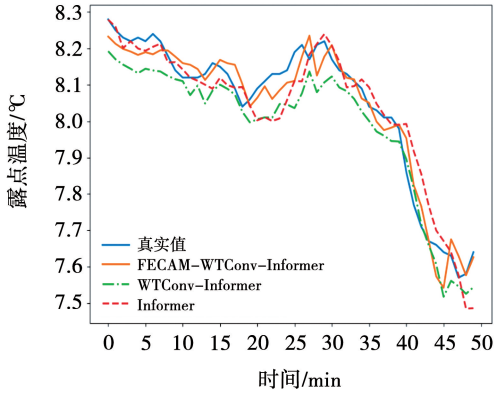

图 5　主模型与 WTConv-Informer 模型对比图(Saaleaue 数据集)

Fig. 5　Comparison chart between the main model and the WTConv-Informer model (Saaleaue dataset)

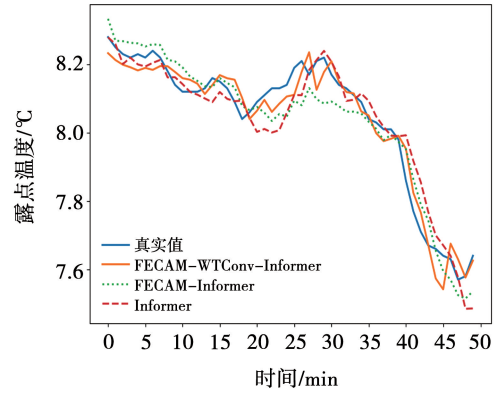

图 6　主模型与 FECAM-Informer 模型对比图(Saaleaue 数据集)

Fig. 6　Comparison chart between the main model and the FECAM-Informer model (Saaleaue dataset)

由表 2 可知,FECAM-WTConv-Informer 模型在各项指标上均优于 Informer,按照降幅计算差值与原值的比可知,R^2 提升了 0.34%,α_{MAE} 降低了 19.28%,α_{MSE} 降低了 39.39%,α_{RMSE} 降低了 22.16%。

而单独结合 FECAM 虽然在预测结果局部上有所提升,但是预测性能数据有所下降,低于 Informer 模型。

表 2 Saaleaue 数据集消融实验结果对比表

Tab. 2 Ablation experiment results comparison on Saaleaue dataset

模型	R^2	α_{MAE}	α_{MSE}	α_{RMSE}
主模型	0.994 8	0.206 4	0.085 4	0.292 2
FECAM-Informer	0.989 7	0.283 3	0.167 2	0.408 9
WTConv-Informer	0.993 0	0.234 9	0.114 3	0.338 1
Informer	0.991 4	0.255 7	0.140 9	0.375 4

从 2 个数据集的消融实验结果可知,主模型 FECAM-WTConv-Informer 在预测指标上优于基础 Informer 模型及其他 2 个消融实验模型,表明模型中各个模块的联合使用可有效提升预测性能。

当单独将 FECAM 与 Informer 结合时,模型的表现在局部上较 Informer 模型虽有提升,但整体数据指标略差。造成这一现象的根本原因在于 FECAM 通过 DCT 提取频率分量,并利用通道注意力机制对频率信息进行增强,特征过于复杂时缺乏对时序信息的有效捕捉,在没有结合多尺度特征提取方法的情况下,FECAM 容易过度关注高频信息的提取与增强。由于高频部分通常包含较多的噪声或局部波动,这种做法会导致模型在捕捉长期趋势时易受到严重干扰,进而引发预测结果的波动。并且,FECAM 的通道注意力机制在强化频率特征时容易忽略对低频成分的有效提取和建模,使得模型在处理复杂的时间序列数据时,难以保持频率信息与时序信息之间的平衡。

因此,在数据特征过于复杂的情况下单独使用 FECAM 会破坏时间序列特征中的频率 – 时域平衡,使得模型无法准确捕捉全局的低频趋势和局部的高频变化。而将 FECAM 与 WTConv 结合,可在强调全局频率信息的同时,保证时序信息的稳定提取与建模,实现二者的优势互补,从而协同提升模型的预测性能。

3.3.2 不同预测模型对比实验分析

在 2 个数据集上对比主模型与 CNN、LSTM、CNN-LSTM、GRU、TCN 等基础模型的表现,通过不同窗口长度可以更全面了解模型在预测任务中的差距。Beutenberg 气象站数据集上不同窗口局部结果对比如图 7 ~9 所示。

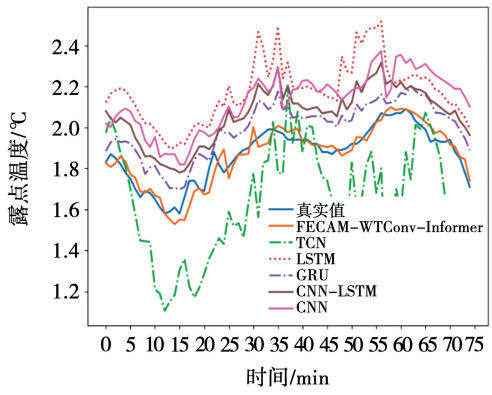

图 7 窗口大小为 16 的模型对比图

Fig. 7 Comparison chart of models with window size of 16

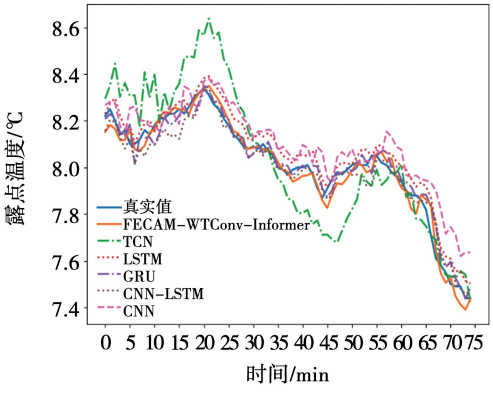

图 8 窗口大小为 24 的模型对比图

Fig. 8 Comparison chart of models with window size of 24

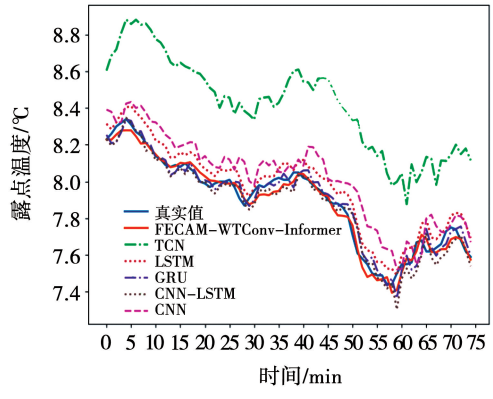

图 9 窗口大小为 32 的模型对比图

Fig. 9 Comparison chart of models with window size of 32

由图 7 ~9 可见,FECAM-WTConv-Informer 模型的预测精度高于其他预测模型,表现出明显预测优势。Beutenberg 气象站数据集具体对比实验评价指标如表 3 所示。

表 3　Beutenberg 数据集模型对比实验结果

Tab. 3　Comparative experimental results of models on Beutenberg dataset

窗口大小	评价指标	主模型	TCN	LSTM	GRU	CNN-LSTM	CNN
16	R^2	0.998 7	0.985 3	0.994 6	0.998 1	0.996 6	0.993 4
	α_{MAE}	0.099 5	0.382 7	0.205 5	0.125 8	0.173 4	0.243 9
	α_{MSE}	0.02 2	0.243 6	0.089 8	0.030 9	0.052 7	0.101 7
	α_{RMSE}	0.148 4	0.493 6	0.299 7	0.175 9	0.229 6	0.318 8
24	R^2	0.998 4	0.982 5	0.993 7	0.997 9	0.996 3	0.992 8
	α_{MAE}	0.108 8	0.431 9	0.235 0	0.133 9	0.181 9	0.254 5
	α_{MSE}	0.027 2	0.288 4	0.104 7	0.033 7	0.056 5	0.109 8
	α_{RMSE}	0.164 9	0.537 0	0.323 6	0.183 5	0.237 7	0.331 4
32	R^2	0.998 1	0.977 2	0.992 9	0.997 9	0.996 7	0.994 1
	α_{MAE}	0.122 9	0.498 6	0.261 5	0.131 1	0.168 4	0.228 2
	α_{MSE}	0.030 6	0.375 8	0.117 0	0.032 8	0.050 8	0.090 9
	α_{RMSE}	0.175 0	0.613 0	0.342 0	0.181 2	0.225 3	0.301 3

由表 3 可见,在不同预测窗口长度下,主模型均展现出良好的性能。随着预测窗口长度的增加,其预测精度保持稳定,表现出较强的鲁棒性。

在窗口长度为 32 的条件下,与 TCN、LSTM、GRU、CNN-LSTM、CNN 等模型相比,主模型在平均绝对误差 α_{MAE} 上分别降低了 75.35%、53%、6.25%、27.02%、46.14%,表明该模型在实际值与预测值的误差控制上有显著优势。

此外,在反映误差平方平均值的均方根误差 α_{RMSE} 上,主模型相较上述模型分别降低了 71.45%、48.83%、3.42%、22.33%、41.92%。证明了该模型在复杂动态时间序列预测任务中的优越性,尤其在长序列预测和多变量条件下表现突出。

Saaleaue 数据集上不同窗口局部结果对比如图 10 ~ 12 所示。

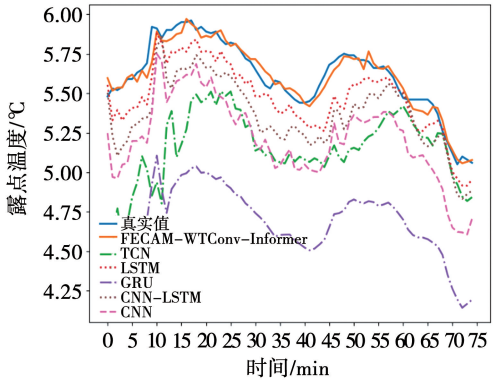

图 10　窗口大小为 16 的模型对比图(Saaleaue 数据集)

Fig. 10　Comparison chart of models with window size of 16 (Saaleaue dataset)

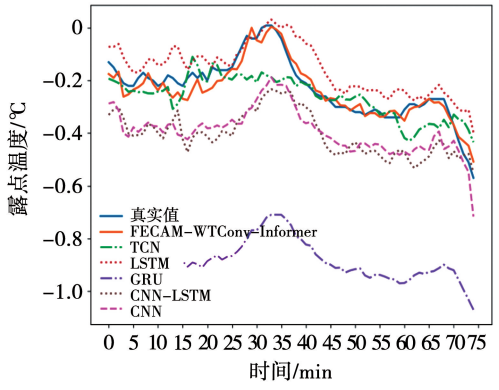

图 11　窗口大小为 24 的模型对比图(Saaleaue 数据集)

Fig. 11　Comparison chart of models with window size of 24 (Saaleaue dataset)

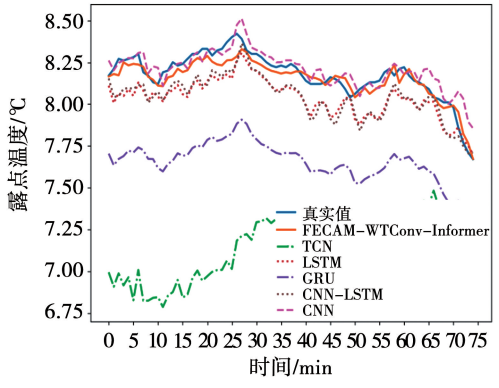

图 12　窗口大小为 32 的模型对比图(Saaleaue 数据集)

Fig. 12　Comparison chart of models with window size of 32 (Saaleaue dataset)

由图 10 ~ 12 可知,FECAM-WTConv-Informer 模型的预测精度高于其他预测模型,在预测结果及稳

定性上体现出明显预测优势,并且在对比实验中可以看出,随着预测窗口长度的增加,模型依然保持着较大的优势。Saaleaue 气象站数据集对比实验评价指标如表 4 所示。

表 4 Saaleaue 数据集模型对比实验结果

Tab. 4 Comparative experimental results of models on Saaleaue dataset

窗口大小	评价指标	主模型	TCN	LSTM	GRU	CNN-LSTM	CNN
16	R^2	0.997 0	0.883 9	0.993 9	0.979 2	0.985 4	0.985 4
	α_{MAE}	0.151 7	1.109 8	0.224 3	0.483 5	0.389 4	0.401 5
	α_{MSE}	0.048 5	1.909 0	0.099 8	0.364 4	0.254 1	0.250 2
	α_{RMSE}	0.220 2	1.381 6	0.315 9	0.603 6	0.504 1	0.500 2
24	R^2	0.996 3	0.840 5	0.993 8	0.976 4	0.986 7	0.989 0
	α_{MAE}	0.168 3	1.313 9	0.225 5	0.508 4	0.346 2	0.332 7
	α_{MSE}	0.060 0	2.617 6	0.100 7	0.417 5	0.220 8	0.183 9
	α_{RMSE}	0.245 0	1.617 9	0.317 3	0.646 1	0.469 9	0.428 8
32	R^2	0.994 8	0.874 4	0.993 7	0.978 2	0.986 2	0.987 0
	α_{MAE}	0.206 4	1.118 0	0.227 7	0.475 5	0.343 3	0.360 6
	α_{MSE}	0.085 4	2.058 4	0.102 5	0.382 0	0.223 7	0.216 0
	α_{RMSE}	0.292 2	1.434 7	0.320 2	0.618 1	0.473 0	0.464 8

由表 4 可知,与其他模型相比,FECAM-WT-Conv-Informer 模型在多个预测窗口长度下表现出明显的优势,尤其是在窗口长度为 32 时,α_{MAE} 和 α_{MSE} 分别较 TCN、LSTM、GRU、CNN-LSTM 和 CNN 模型降低了 81.54%、9.35%、56.58%、39.88%、42.76% 及 95.85%、16.68%、77.64%、61.82%、60.46%。主模型在收敛速度和精度上表现出更强的能力,可实现更理想的预测效果。

除以上主流基础模型对比外,在与相同预测任务的复合模型对比中,主模型依然能在长序列复杂时间预测任务中表现出一定的优势,与 Autoformer 相比,在 Beutenberg 数据集上的对比试验局部结果对比如图 13 ~ 15 所示。

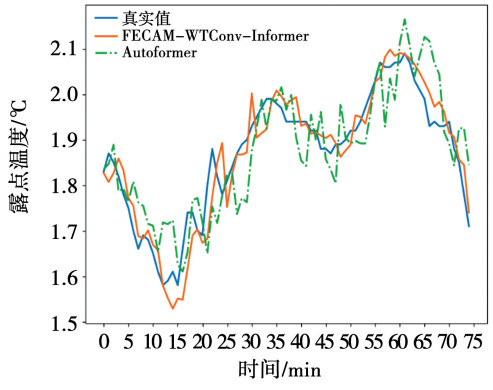

图 13 窗口大小为 16 的 Autoformer 对比图

Fig. 13 Comparison diagram of Autoformer with window size of 16

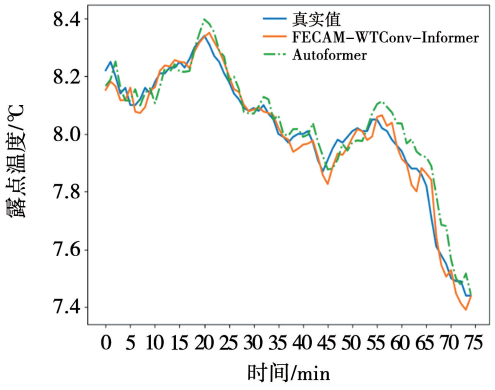

图 14 窗口大小为 24 的 Autoformer 对比图

Fig. 14 Comparison diagram of Autoformer with window size of 24

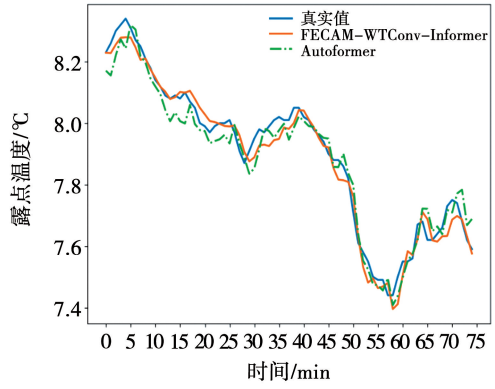

图 15 窗口大小为 32 的 Autoformer 对比图

Fig. 15 Comparison diagram of Autoformer with window size of 32

由图 13 ~ 15 可知,主模型比 Autoformer 表现出了更高的稳定性和更精确的预测结果,Beutenberg 气象站数据集 Autoformer 对比实验评价指标如表 5 所示。

表 5　Beutenberg 数据集 Autoformer 对比实验表

Tab. 5　Comparison table of Autoformer experiments on the Beutenberg dataset

模型	窗口大小	R^2	α_{MAE}	α_{MSE}	α_{RMSE}
主模型	16	0.998 7	0.099 5	0.022 0	0.148 4
	24	0.998 4	0.108 8	0.027 2	0.164 9
	32	0.998 1	0.206 5	0.085 4	0.292 2
Autoformer	16	0.997 6	0.122 9	0.030 6	0.175 0
	24	0.997 0	0.157 3	0.049 1	0.221 6
	32	0.996 3	0.175 1	0.061 3	0.247 6

由表 5 可知,主模型在各指标上的表现均优于 Autoformer。在数据特征更加复杂,数据波动更加明显的 Saaleaue 数据集上进行对比试验,不同窗口局部结果对比如图 16 ~ 18 所示。

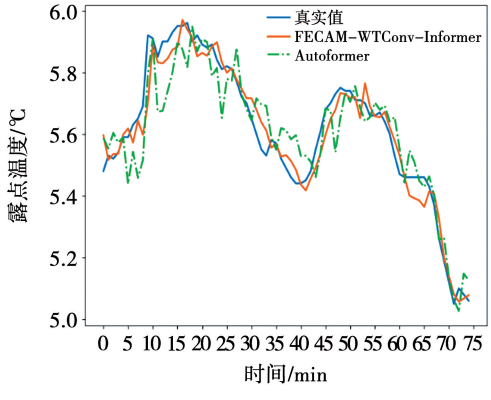

图 16　窗口大小为 16 的 Autoformer 对比图(Saaleaue 数据集)

Fig. 16　Comparison diagram of Autoformer with window size of 16 on Saaleaue dataset

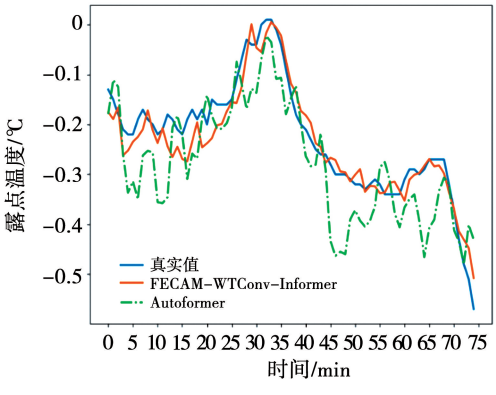

图 17　Saaleaue 窗口大小为 24 的 Autoformer 对比图(Saaleaue 数据集)

Fig. 17　Comparison diagram of Autoformer with window size of 24 on Saaleaue dataset

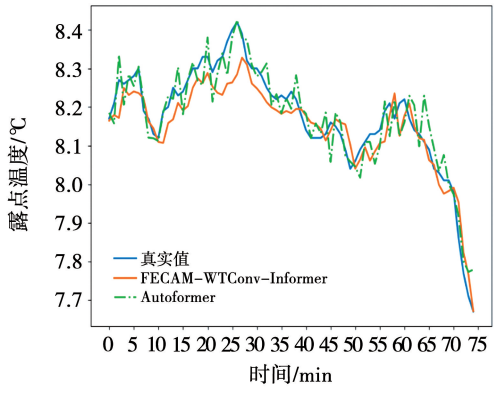

图 18　窗口大小为 32 的 Autoformer 对比图(Saaleaue 数据集)

Fig. 18　Comparison diagram of Autoformer with window size of 32 on Saaleaue dataset

由图 16 ~ 18 可知,主模型的预测效果优于 Autoformer,并且在窗口增长下表现出了良好的稳定性。Autoformer 在同预测目标下、同等参数条件下的波动要高于 FECAM-WTConv-Informer 模型。Saaleaue 气象站数据集 Autoformer 对比实验评价指标如表 6 所示。

表 6　Saaleaue 数据集 Autoformer 对比实验表

Tab. 6　Comparison table of Autoformer experiments on the Saaleaue dataset

模型	窗口大小	R^2	α_{MAE}	α_{MSE}	α_{RMSE}
主模型	16	0.997 0	0.151 7	0.048 5	0.220 3
	24	0.996 3	0.168 3	0.060 0	0.245 1
	32	0.994 8	0.206 5	0.085 4	0.292 2
Autoformer	16	0.996 1	0.175 9	0.064 0	0.253 1
	24	0.995 4	0.194 4	0.076 0	0.275 7
	32	0.994 2	0.218 4	0.094 7	0.307 7

由表 6 可知,FECAM-WTConv-Informer 模型在各项指标上较 Autoformer 均有一定程度的提升,在窗口大小为 16 时,R^2 虽然提升不大,但是 α_{MAE} 降低了 13.76%,α_{MSE} 降低了 24.22%,α_{RMSE} 降低了 12.96%。并且从实验结果可知,虽然随着窗口的增长提升效果有所下降,但是在稳定性上,主模型保持着更低的波动。因此,与同类复合模型相比,FECAM-WTConv-Informer 模型有其独特的优势。

4　结　语

针对复杂温度预测任务,本文提出了 FECAM-WTConv-Informer 模型,将改进 WTConv 的多尺度特征提取、Informer 的高效长序列建模能力及 FECAM 的全局频率感知和通道依赖建模优势相结合,显著

提升了温度预测的精度与稳定性。实验结果表明:该模型在2个不同气象站数据集上的性能优于Informer模型和其他传统预测模型,特别是在高频动态变化的预测中表现出强大的优势。综上所述,FECAM-WTConv-Informer模型不仅可高效捕捉气象时间序列中的长期依赖和局部波动特性,还具有良好的泛化能力和鲁棒性,为高维多变量气象数据预测提供了高效可靠的解决方案。未来研究可以进一步优化模型复杂度与运行效率,以适应更大规模和实时性的应用场景。

参 考 文 献:

[1] 王岩, 王昊, 崔鹏, 等. 气候变化的灾害效应与科学挑战[J]. 科学通报, 2024, 69(2): 286.
WANG Yan, WANG Hao, CUI Peng, et al. Disaster Effects of Climate Change and the Associated Scientific Challenges[J]. Chinese Science Bulletin, 2024, 69(2): 286.

[2] 杨迎新, 杜景林, 武艳. 基于PCA和改进PSO-GRU神经网络的气温预测[J]. 现代电子技术, 2022, 45(1): 89.
YANG Yingxin, DU Jinglin, WU Yan. Air Temperature Prediction Based on PCA and Improved PSO-GRU Neural Network[J]. Modern Electronics Technique, 2022, 45(1): 89.

[3] 曾子健, 肖慧, 徐哈宁, 等. 基于CRITIC赋权法与PSO-SVR模型的滑坡地表位移预测[J]. 科学技术与工程, 2024, 24(35): 14940.
ZENG Zijian, XIAO Hui, XU Haning, et al. Landslide Surface Displacement Prediction Based on CRITIC Weight Method and PSO-SVR Model[J]. Science Technology and Engineering, 2024, 24(35): 14940.

[4] CAI P, WANG Y, LU G, et al. A Spatiotemporal Correlative K-nearest Neighbor Model for Short-term Traffic Multistep Forecasting[J]. Transportation Research Part C: Emerging Technologies, 2016, 62: 21.

[5] 张欣, 李林. 基于随机森林的季节性电动汽车充电负荷预测[J]. 软件工程, 2024, 27(11): 11.
ZHANG Xin, LI Lin. Seasonal Electric Vehicle Charging Load Prediction Based on Random Forest[J]. Software Engineer, 2024, 27(11): 11.

[6] DU K L, JIANG B, LU J, et al. Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions[J]. Mathematics, 2024, 12(24): 3935.

[7] 王文豪, 李秀芹. GPAformer模型在气温预测中的应用研究[J]. 计算机技术与发展, 2025, 35(3): 187.
WANG Wenhao, LI Xiuqin. Research on Application of a GPAformer Model for Temperature Prediction[J]. Computer Technology and Development, 2025, 35(3): 187.

[8] ZHANG M, ZHANG Z, WANG X, et al. The Use of Attention-enhanced CNN-LSTM Models for Multi-indicator and Time-series Predictions of Surface Water Quality[J]. Water Resources Management, 2024, 38(15): 6103.

[9] 吴劲松, 张少峰, 徐向民, 等. 计及PUE的数据中心短期电力负荷预测方法[J]. 哈尔滨理工大学学报, 2021, 26(6): 1.
WU Jinsong, ZHANG Shaofeng, XU Xiangmin, et al. Short-term Power Load Forecasting Method of Data Center Considering PUE[J]. Journal of Harbin University of Science and Technology, 2021, 26(6): 1.

[10] 王旭, 托娅, 李鑫, 等. 基于GRU的数据中心温度预测[J]. 哈尔滨理工大学学报, 2024, 29(6): 101.
WANG Xu, TUO Ya, LI Xin, et al. Temperature Prediction Model for Power Data Centers Based on GRU[J]. Journal of Harbin University of Science and Technology, 2024, 29(6): 101.

[11] HAN K, XIAO A, WU E, et al. Transformer in Transformer[J]. Advances in Neural Information Processing Systems, 2021, 34: 15908.

[12] 张涛, 张晓利, 任彦. Transformer与CNN融合的单目图像深度估计[J]. 哈尔滨理工大学学报, 2022, 27(6): 88.
ZHANG Tao, ZHANG Xiaoli, REN Yan. Monocular Image Depth Estimation Based on the Fusion of Transformer and CNN[J]. Journal of Harbin University of Science and Technology, 2022, 27(6): 88.

[13] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-series Forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106.

[14] 陶晔, 杜景林. 基于随机森林的长短期记忆网络气温预测[J]. 计算机工程与设计, 2019, 40(3): 737.
TAO Ye, DU Jinglin. Temperature Prediction Using Long Short Term Memory Network Based on Random Forest[J]. Computer Engineering and Design, 2019, 40(3): 737.

[15] JUN J, KIM H K. Informer-based Temperature Prediction Using Observed and Numerical Weather Prediction Data[J]. Sensors, 2023, 23(16): 7047.

[16] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet Convolutions for Large Receptive Fields[C]// European Conference on Computer Vision, 2024: 363.

[17] JIANG M, ZENG P, WANG K, et al. FECAM: Frequency Enhanced Channel Attention Mechanism for Time Series Forecasting[J]. Advanced Engineering Informatics, 2023, 58: 102158.

[18] SALEH E H, BADAWY W, FOUAD M, et al. Textile Inspection Based on à Trous Wavelet Transform[J]. Wireless Personal Communications, 2024, 138(3): 1405.

[19] ASHKENAZI M, FINDER S E, TREISTER E, et al. Wavelet Feature Maps Compression for Image-to-image CNNS[C]//Advances in Neural Information Processing Systems, 2022: 20592.

[20] 杜睿山, 周长坤, 解红涛, 等. SE-CMT 语义信息补充的特征提取模型[J]. 哈尔滨理工大学学报, 2024, 29(6): 74.
DU Ruishan, ZHOU Changkun, XIE Hongtao, et al. Feature Extraction Model of SE-CMT Semantic Information Supplement[J]. Journal of Harbin University of Science and Technology, 2024, 29(6): 74.

[21] HU J, SHEN L, SUN G. Squeeze-and-excitation Networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132.

[22] 滕陈源, 丁逸超, 张有兵, 等. 基于 VMD-Informer-BiLSTM 模型的超短期光伏功率预测[J]. 高电压技术, 2023, 49(7): 2961.
TENG Chenyuan, DING Yichao, ZHANG Youbing, et al. Ultra-short-term Photovoltaic Power Prediction Based on VMD-informer-BiLSTM Model[J]. High Voltage Engineering, 2023, 49(7): 2961.

[23] 王颉, 刘兴杰, 梁英, 等. 一种基于 MGWO-Informer 的超短期风电功率预测方法[J]. 太阳能学报, 2024, 45(11): 477.
WANG Jie, LIU Xingjie, LIANG Ying, et al. An Ultra-short-term Wind Power Prediction Method Based on mgwo-informer[J]. Acta Energiae Solaris Sinica, 2024, 45(11): 477.

[24] 马愈昭, 张宇航, 王凌飞. TimeGAN-Informer 长时机场能见度预测[J]. 安全与环境学报, 2024, 24(7): 2517.
MA Yuzhao, ZHANG Yuhang, WANG Lingfei. Long-term Prediction of Visibility Using TimeGAN-informer[J]. Journal of Safety and Environment, 2024, 24(7): 2517.

[25] WU H, XU J, WANG J, et al. Autoformer: Decomposition Transformers with Auto-correlation for Long-term Series Forecasting[J]. Advances in Neural Information Processing Systems, 2021, 34: 22419.

(编辑:温泽宇)