

# DTFNet: 基于时频协同分解网络的时间序列预测<sup>①</sup>



魏祥麒, 顾晶晶

(南京航空航天大学 计算机科学与技术学院, 南京 211106)  
通信作者: 顾晶晶, E-mail: [gujingjing@nuaa.edu.cn](mailto:gujingjing@nuaa.edu.cn)

**摘要:** 时间序列预测在气象预报、电力负荷预测与金融管理等领域具有广泛应用, 近年来深度学习方法在该任务中取得显著进展. 然而, 现有模型在应对非平稳性和异质模式建模方面仍存在局限, 主要表现为对趋势与季节分量的同质化建模与分解过程中的模态混叠问题. 本文提出一种名为 DTFNet 的时频协同分解网络, 通过设计时域-频域并行的异构结构, 在时域利用抗噪性强的 MLP 网络建模趋势项的长期演化特性, 在频域采用快速傅里叶变换提取周期性季节成分, 并通过多尺度卷积操作捕捉时频特征间的空间关联. 同时, 本文引入基于离散小波变换 (DWT) 的分解方法, 替代传统移动平均分解, 有效缓解边界效应与模态混叠问题. 在 6 个公开数据集上的实验结果表明, DTFNet 在准确性和鲁棒性方面均优于现有主流模型. 消融实验结果表明, 本文提出的离散小波变换分解模块以及时频协同建模结构在提升时间序列预测精度方面具有显著效果. DTFNet 具备良好的通用性, 能够应用于多种时间序列预测任务, 为电力负荷预测、天气预报等实际应用场景提供有力的支持.

**关键词:** 时间序列分解; 小波变换; 非平稳性; 多尺度分析

引用格式: 魏祥麒, 顾晶晶. DTFNet: 基于时频协同分解网络的时间序列预测. 计算机系统应用, 2026, 35(1): 52-63. <http://www.c-s-a.org.cn/1003-3254/10043.html>

## DTFNet: Time Series Forecasting Based on Time-frequency Cooperative Decomposition Network

WEI Xiang-Qi, GU Jing-Jing

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** Time series forecasting finds widespread applications in such fields as weather forecasting, power load forecasting, and financial management. In recent years, deep learning has made remarkable progress in these tasks. However, existing models still have limitations in struggling with non-stationarity and heterogeneous pattern modeling, which is mainly represented by homogenized modeling of trends and seasonal components, and modal aliasing during decomposition. To this end, this study proposes a time-frequency cooperative decomposition network, termed as DTFNet, which designs a heterogeneous architecture with parallel time and frequency domains. In the time domain, an MLP network with strong noise resistance is employed to model the long-term evolution characteristics of trends, while in the frequency domain, fast Fourier transform is adopted to extract periodic seasonal components, with multi-scale convolution operations employed to capture spatial correlations between time-frequency characteristics. Meanwhile, this study introduces a decomposition method based on discrete wavelet transformation (DWT) to replace conventional moving average decomposition, effectively mitigating boundary effects and modal aliasing. Experiments on six public datasets demonstrate that DTFNet outperforms the current mainstream models in both accuracy and robustness. Ablation experiments show the notable effectiveness of the proposed DWT-based decomposition module and dual-domain time-frequency modeling architecture. Featuring sound generalization ability, DTFNet is applicable to multiple time series

① 收稿时间: 2025-06-17; 修改时间: 2025-07-07; 采用时间: 2025-08-01; csa 在线出版时间: 2025-11-17  
CNKI 网络首发时间: 2025-11-18

forecasting tasks, offering powerful support for real-world applications such as power load forecasting and weather forecasting.

**Key words:** time series decomposition; wavelet transformation; non-stationary; multi-scale analysis

时间序列预测作为时序数据分析的核心任务,广泛应用于气象预报<sup>[1]</sup>、智能电网负荷预测<sup>[2]</sup>、金融风险<sup>[3]</sup>等现实场景中.随着物联网技术的普及,海量时序数据的可获得性推动了深度学习方法的突破性进展.与传统统计模型<sup>[4,5]</sup>相比,基于深度神经网络的预测模型通过端到端学习机制,在复杂模式捕获方面展现出显著优势,从而引发了时序建模方法论的范式革新.

深度时序建模技术经历了从局部到全局、从复杂到简约的架构演进.早期研究以循环神经网络(RNN)<sup>[6]</sup>和时序卷积网络(TCN)<sup>[7]</sup>为主流,分别通过循环结构和膨胀卷积捕获时序动态与长程依赖.随着Transformer<sup>[8]</sup>的引入,自注意力机制凭借全局依赖建模突破局部感受野限制,但计算复杂度与局部特征提取瓶颈催生了注意力维度反转<sup>[9]</sup>、局部划分<sup>[10]</sup>等优化方法.与此同时,多层感知机(MLP)的复兴<sup>[11,12]</sup>揭示了全连接网络在显式时序编码中的潜力,推动模型设计向简约化回归.然而,非平稳性<sup>[13]</sup>引发的分布偏移<sup>[14]</sup>及异质模式建模<sup>[15]</sup>难题,促使技术路线转向序列分解<sup>[16]</sup>.

但现有方法在理论和实践层面均存在显著局限性:从建模维度看,主流方法<sup>[16-18]</sup>对趋势项和季节项采用同质化处理,忽略了二者的物理特性差异<sup>[19]</sup>.具体而言,趋势分量表征序列的长期演化规律,具有低频缓变特性,其时域微分特征(如梯度变化率)包含关键信息;而季节分量反映周期性波动,其频域能量集中特性更适宜通过谱分析提取.现有方法的单一域建模范式导致时频特征解耦不充分,造成模态混叠(modal aliasing)<sup>[20]</sup>.从分解技术看,传统简单移动平均(simple moving average, SMA)存在边界效应和频率分辨率受限的固有缺陷,滑动窗口机制需对序列首尾进行人工填充,易引入边界效应导致分解偏差,作为线性滤波器对高频噪声敏感,在处理非平稳信号时易引发模态混叠,导致分量纯净度下降<sup>[21]</sup>.

针对上述挑战,本文提出一种名为DTFNet的时频协同分解网络(time-frequency cooperative decomposition network),设计时域频域并行处理架构,协同长期依赖和周期性特征,并且对序列分解方法进行改进,

提出通过小波变换提取低频分量作为趋势项实现趋势-季节项的精准分离,其主要贡献有以下方面.

- 双域异构建模架构,在时域采用抗噪性强的MLP网络建模趋势分量的长期演化规律,在频域通过快速傅里叶变换(FFT)提取季节分量的周期特征,并引入卷积建模时频特征的空间相关性.

- 基于离散小波变换的序列分解方法,相较于移动平均分解,小波分解避免了人工填充操作,还具备更优的频率局部化特性,使其在非平稳信号处理中表现更为稳健,可有效抑制边界效应和模态混叠.

- 在6个数据集上的对比实验结果表明,DTFNet具有比基线模型更高的预测准确度.

## 1 相关工作

### 1.1 时间序列预测

时间序列预测是时间序列分析领域中一项关键的任务,旨在利用历史数据中的时间依赖性和模式,预测未来特定时间点或时间段内的数值或趋势.

传统时间序列预测方法主要基于统计模型,如自回归移动平均(ARMA/ARIMA)、季节性ARIMA(SARIMA)和向量自回归(VAR)等.ARIMA<sup>[4]</sup>模型通过自回归、差分和移动平均这3部分建模,适用于具有趋势的非平稳序列;SARIMA<sup>[22]</sup>在此基础上引入季节性差分,以处理周期性变化;VAR<sup>[23]</sup>模型则扩展到多变量序列,建模变量间的相互影响.这些方法结构简单、可解释性强,但在处理非线性、结构变化和长期依赖时表现有限,且递归预测易积累误差.

近年来,深度学习方法凭借大规模数据驱动的优势,已成为时间序列预测任务的主要方法.循环神经网络(RNN)和时序卷积网络(TCN)是两类经典的深度学习方法,广泛应用于时间序列建模任务中.RNN<sup>[6]</sup>通过隐藏状态递归建模,能捕捉短期及部分长期依赖,但循环结构难以捕捉远距离依赖.TCN<sup>[7]</sup>通过因果卷积(causal convolution)和膨胀卷积(dilated convolution)结构,实现了高效捕捉长距离依赖,但受限感受野与缺乏动态记忆,对复杂时序特征的刻画仍有不足.

Transformer<sup>[8]</sup>架构因其卓越的建模能力,已成为时间序列预测领域的主流方法.通过引入自注意力机制,Transformer能够捕捉任意时间点之间的依赖关系,有效建模时间序列中的复杂非线性特征.Informer<sup>[24]</sup>作为首个专为长期时间序列预测设计的Transformer模型,采用了ProbSparse自注意力机制和生成式解码器,降低了时间和空间复杂度.iTransformer<sup>[9]</sup>将时间序列数据的维度进行反转,使得每个变量成为一个时间序列,而时间步则成为每个变量的一个维度,使模型更好地捕捉不同变量之间的关联性.PatchTST<sup>[10]</sup>受视觉Transformer(ViT)的启发,将时间序列划分为等长片段作为输入Token,并实现了通道独立性,提升了模型的表达能力.TimeXer<sup>[25]</sup>模型则引入了外生变量的建模机制,增强了Transformer在处理实际应用中部分观测数据的能力,结合Patch级自注意力和变量级交叉注意力,融合内生和外生信息,提升了预测精度.

除了以上Transformer结构的模型外,一些更简单的模型也被证明可以达到良好的时间序列预测效果.TimesNet<sup>[26]</sup>通过引入傅里叶变换和多周期分量的解耦建模,显式捕捉不同周期的多尺度动态变化,显著增强了模型对跨尺度模式的刻画能力.

时间序列预测的方法经历了从基于统计到基于深度学习的范式变迁,显著提升了对复杂时序数据的建模能力.然而,现有方法在应对非平稳性、异质性等实际问题上仍存在不足,难以有效捕捉时间变化的统计特性与多样化生成机制.如何进一步提升模型对非平稳性与异质模式的适应与建模能力,仍是当前研究的重要方向.

## 1.2 基于序列分解的预测方法

为了解决时间序列数据中的非平稳性和异质模式建模问题,多个模型采用序列分解以更好地捕捉复杂的时间模式.

Autoformer<sup>[16]</sup>将序列分解机制嵌入Transformer基础模块,借助Auto-Correlation机制在子序列层面挖掘周期性依赖,并在各层中实现渐进式的趋势-季节分解.FEDformer<sup>[18]</sup>融合序列分解与频域稀疏表示,不仅利用分解方法来提取全局轮廓,还通过傅里叶变换强化关键频率分量的建模能力.DLinear<sup>[11]</sup>将原始时间序列分解为趋势项与季节项,并分别由两个独立的单层线性(linear)网络进行预测并汇总生成最终结果,其优异表现表明线性模型在学习时间序列特有的时间特征方面

具备独特优势.TimeMixer<sup>[12]</sup>则通过多尺度分解,将时间序列划分为多个季节性成分,并基于季节-趋势混合后的多尺度历史信息,精准建模复杂的时序动态.

上述模型中采用的分解方法均基于移动平均分解,具体而言,在对输入序列两端进行填充后,应用移动平均核提取序列的趋势分量,而原始序列与趋势分量的差值则被视为季节分量.这种分解方式的优势在于,能够以线性时间复杂度高效地分离出序列中的不同特征.然而,该方法存在两大固有缺陷.

首先,由于滑动窗口机制需要对序列首尾进行人工填充,容易引入边界效应,进而导致分解偏差.其根本原因在于,滑动窗口在序列两端因无法获取完整数据而必须依赖人工填充(如零填充或反射填充),这些填充值并非真实数据,会扭曲边界处的趋势估计,产生失真<sup>[21]</sup>.

其次,作为一种线性低通滤波器,移动平均对高频噪声较为敏感,难以克服处理非平稳信号时的固有局限.在面对非平稳序列时,该方法容易引发模态混叠,导致分解出的分量纯净度下降.其理论根源在于移动平均滤波器非理想的频率响应特性,其通带与阻带之间存在一个较宽的过渡带,无法实现频率成分的清晰分离.这导致高频的噪声或季节性细节容易“泄露”到低频的趋势项中,反之亦然,从而引发模态混叠<sup>[20]</sup>.

因此,如何改进现有序列分解技术,以更稳健地应对非平稳和复杂时序特征,仍是亟待解决的重要问题.

## 2 模型结构

### 2.1 问题描述

考虑具有 $C$ 个变量的多元时间序列可表示为张量 $X \in \mathbb{R}^{L \times C}$ ,其中 $L$ 为观测窗口长度, $C$ 为特征维度(变量数).设 $t \in \{1, 2, \dots, L\}$ 为时间索引,则时刻 $t$ 的观测向量为 $\mathbf{x}_t = [x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(C)}]^\top \in \mathbb{R}^C$ ,其中 $x_t^{(i)}$ 表示第 $i$ 个变量在 $t$ 时刻的取值.时间序列预测问题可形式化定义为:给定历史观测序列 $X_{1:L} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L]^\top$ ,预测未来 $T$ 步的序列 $X_{L+1:L+T} = [\mathbf{x}_{L+1}, \mathbf{x}_{L+2}, \dots, \mathbf{x}_{L+T}]^\top$ .主流方法通常对时间序列采用同质化处理,对趋势项和季节项的差异性缺乏针对性建模,单一域建模范式导致时频特征解耦不充分,造成模态混叠.此外传统简单移动平均(SMA)存在边界效应和频率分辨率受限的固有缺陷,在处理非平稳信号时容易引发模态混叠,导致分量

纯净度下降. 因此, 对于趋势和季节性成分的精准分离成为更准确预测的先决条件.

## 2.2 模型整体结构

基于时频域分解的时间序列预测模型结构如图1所示, 该模型的核心思想是通过离散小波变换将时间序列分解为趋势项和季节项, 分别在时域和频域中进行建模. 模型主要由以下3个部分组成: 离散小波变换分解模块、时域趋势项分支和频域季节项分支. 其中小波变换分解模块用于提取时间序列中的趋势项和季节项; 时域趋势项分支使用一系列线性变换和归一化层对趋势项进行建模, 捕捉长时间的变化趋势; 频域季节项分支通过快速傅里叶变换(FFT)和频域MLP模块对季节项建模, 以捕捉短期周期性变化. 最后, 两部分预测结果拼接后, 通过自适应权重融合趋势预测和季节预测, 获得预测结果.

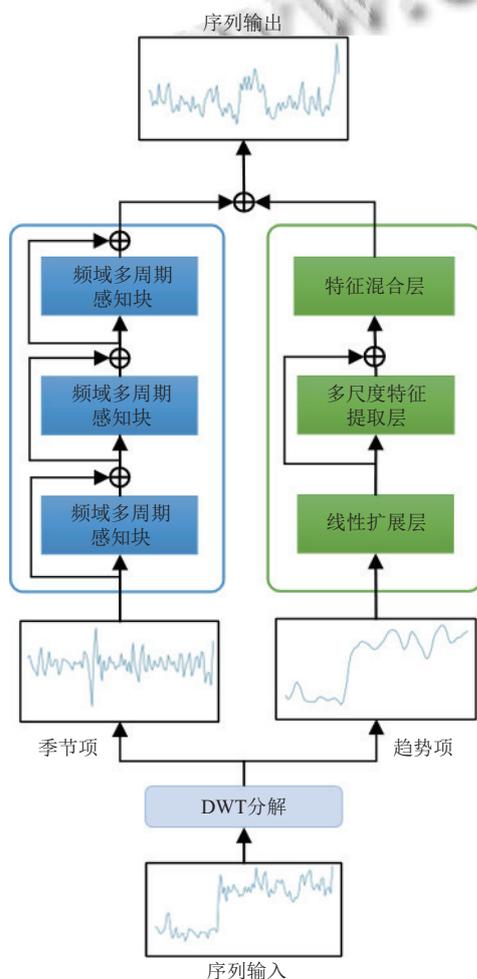


图1 时频双域协同网络模型框架

## 2.3 离散小波变换分解

为了有效捕捉时间序列数据中的趋势和季节性成

分, 本文采用了基于离散小波变换 (discrete wavelet transform, DWT) 的分解方法, 来实现时间序列的多尺度分解. 离散小波变换通过一系列的滤波器组将信号分解为不同频率带宽的小波系数, 其中每一层分解都将原始信号划分为低频近似分量 (approximation coefficient,  $A$ ) 和 高频细节分量 (detail coefficient,  $D$ ). 低频近似分量反映时间序列的长期趋势, 而高频细节分量则包含短期波动、周期性和噪声信息, 模块结构如图2所示.

相比传统的移动平均法, 离散小波变换在处理边界效应和模态混叠问题上具备显著的理论优势. 首先, 在处理边界效应方面, 由于小波基函数具有“紧支撑性 (compact support)”, 即其在时域上是局部化的, 使得 DWT 在处理有限长度信号的端点时无需进行人工填充, 从根本上避免了由填充引入的失真. 其次, 在抑制模态混叠方面, DWT 凭借其“多分辨率分析”的特性, 通过正交滤波器组, 能够将信号投影到不同频率的子空间上, 实现更清晰的频率分离. 这与移动平均法非理想的低通滤波特性形成鲜明对比, 有效防止了不同频率模式间的“泄露”, 从而保证了分解后趋势项与季节项的纯净度.

在本文的模型中, 通过控制小波分解的层数 ( $level$ ), 可以灵活地调整对趋势项和季节项的分离程度. 层数越高, 提取出的趋势越平滑, 对应的季节项则保留了更多的短期变化特征. 具体来说, 对于给定的时间序列  $X$ ,  $level$  层小波分解的数学表达式可表示为:

$$\begin{cases} A_j[k] = \sum_n h[n-2k]x[n] \\ D_j[k] = \sum_n g[n-2k]x[n] \end{cases} \quad (1)$$

其中,  $h[n]$  表示低通滤波器对应尺度函数  $\varphi(t)$  的脉冲响应,  $g[n]$  是高通滤波器对应小波函数  $\psi(t)$  的脉冲响应,  $x[n]$  是输入的时间序列数据,  $j$  代表分解的层级,  $k$  是样本索引,  $A_j[k]$  和  $D_j[k]$  分别表示第  $j$  层低频近似分量和 高频细节分量.

趋势项  $\mathcal{T}$  由最终层低频近似分量  $A_{level}$  反离散小波变换重构获得, 捕捉时间序列的长期变化趋势.  $D_1, D_2, \dots, D_{level}$  分别表示从第1层到第  $level$  层的高频细节分量, 它们共同构成了季节项  $\mathcal{S}$ , 反映了时间序列中的短期波动和周期性特征, 为方便计算由原始序列  $X$  与趋势项  $\mathcal{T}$  通过残差计算得到. 趋势项  $\mathcal{T}$  与季节项  $\mathcal{S}$  数学表达式如下:

$$\mathcal{T} = IDWT(A_{level}) \quad (2) \quad \mathcal{S} = \mathcal{X} - \mathcal{T} \quad (3)$$

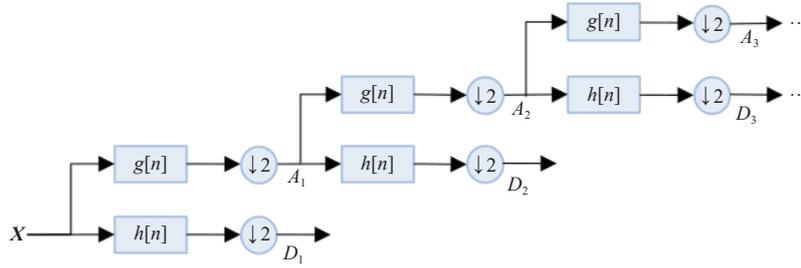


图2 离散小波变换分解示意图

为了进一步提高算法效率并减少重复计算, 本文设计了一个基于哈希值的缓存机制. 每当有新的输入需要进行小波分解时, 首先将其转换为哈希值作为缓存键. 如果该键对应的分解结果已经存在于缓存中, 则直接返回缓存的结果; 否则执行小波分解, 并将结果存储于缓存中以便后续使用.

### 2.4 时域趋势建模网络

趋势项 $\mathcal{T}$ 通常呈现出平稳的长周期变化, 反映了数据的宏观演化规律. 为有效捕捉趋势信息, 本文设计了一种时域趋势建模网络 (temporal trend modeling network), 专门用于处理趋势分量, 提升预测的长期准确性. 网络通过多尺度特征提取架构, 结合线性变换与非线性激活函数, 能够有效捕捉复杂趋势模式并提升预测精度.

时域趋势建模网络主要由以下 3 个部分构成.

(1) 线性扩展层 (temporal projection layer): 为了统一预测时间长度, 首先将趋势项 $\mathcal{T}$ 的时间长度从原始的 $L$ 映射到预测长度 $T$ . 通过对趋势项进行线性变换, 并采用统一缩放方式初始化权重矩阵 $W_{proj}$ , 确保训练初期的稳定性, 同时配合 LayerNorm 正则化处理, 提升网络对数值变化的鲁棒性, 数学表达式如下:

$$\mathcal{T}_T = \mathcal{T}_L \cdot W_{proj} + b_{proj} \quad (4)$$

其中,  $\mathcal{T}_L \in \mathbb{R}^{B \times C \times L}$  表示输入的趋势项,  $B$  是批次大小,  $C$  是特征维度;  $W_{proj} \in \mathbb{R}^{T \times L}$  是可学习的权重矩阵,  $b_{proj}$  是偏置项. 映射后的结果经过归一化和非线性激活函数处理, 增强模型的表达能力.

(2) 多尺度特征提取层 (multi-scale feature extraction layer): 趋势项的建模需要具备对不同时间尺度的表征能力. 为此, 本文引入上采样-下采样路径 (up-down pathway) 与下采样-上采样路径 (down-up pathway) 两

条并行通路, 从不同视角提取特征.

1) 上采样-下采样路径: 先通过两层线性变换将特征维度逐步扩展至高维空间, 再压缩回较低维度, 数学表达式如下:

$$\mathcal{T}_{up-down}^{(1)} = Linear(\mathcal{T}_{prev}, 2 \times T) \quad (5)$$

$$\mathcal{T}_{up-down}^{(2)} = Linear(\mathcal{T}_{up-down}^{(1)}, 4 \times T) \quad (6)$$

2) 下采样-上采样路径: 特征扩展至高维空间后逐步降维, 提供不同的特征视角, 数学表达式如下:

$$\mathcal{T}_{down-up}^{(1)} = Linear(\mathcal{T}_{prev}, 4 \times T) \quad (7)$$

$$\mathcal{T}_{down-up}^{(2)} = Linear(\mathcal{T}_{down-up}^{(1)}, 2 \times T) \quad (8)$$

两条路径通过归一化和激活函数交替作用, 增强了对非线性关系的学习能力, 分别捕捉趋势项在不同尺度上的表示能力, 通过特征维度的升降处理增强网络对多层次变化的敏感性.

(3) 特征混合层 (feature fusion layer): 在完成升降维特征提取后, 将两种路径的输出在特征维度上拼接, 构建一个包含多尺度信息的融合特征张量:

$$\mathcal{T}_{concat} = Concat(\mathcal{T}_{up-down}, \mathcal{T}_{down-up}) \quad (9)$$

接着通过两层线性层 ( $6T \rightarrow 3T \rightarrow T$ ) 逐步压缩冗余特征, 形成最终的趋势表示:

$$\mathcal{T}_{3T} = Linear(\mathcal{T}_{concat}, 3T) \quad (10)$$

$$\mathcal{T}_{final} = Linear(\mathcal{T}_{3T}, T) \quad (11)$$

此外, 网络结构中引入残差初始化机制, 使得网络在融合特征时能够保留原始趋势项的主导信息, 强化趋势建模效果.

### 2.5 频域季节建模网络

季节项 $\mathcal{S}$ 通常包含多周期性振荡模式, 体现了时间序列中的高频与中频变化特征. 传统的时域方法难以精准捕捉其中的隐式周期性规律, 尤其是在处理复

杂、多重叠加的季节性模式时,容易发生模态混叠,影响预测精度.为此,本文提出了一种基于频域多周期感知块 (multi-period perception in frequency domain

block) 的季节建模网络,旨在通过直接在频域对季节性成分进行分析和建模,有效提升对复杂季节成分的建模能力与泛化能力,模块结构如图3所示.

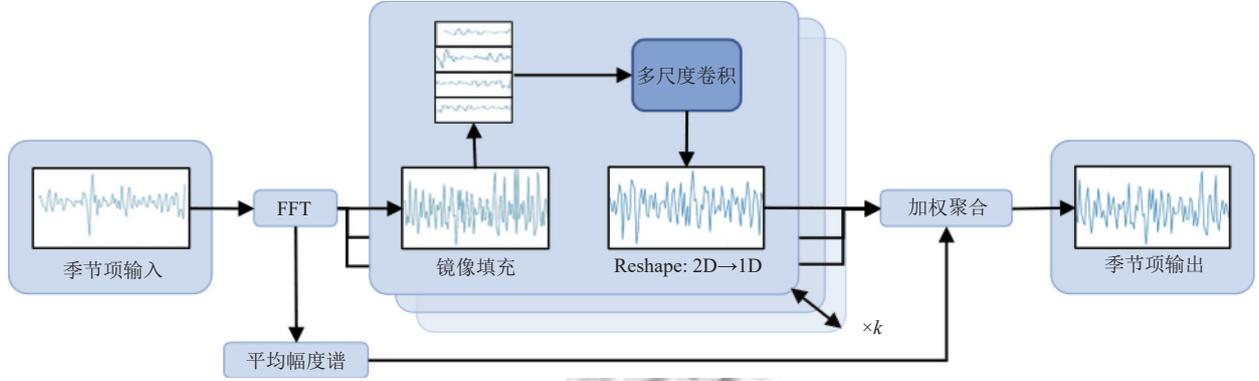


图3 频域多周期感知块示意图

频域多周期感知块通过傅里叶变换在频域中提取主要的周期分量,利用幅度谱平均强度对频率分量进行排序,选取前 $k$ 个最显著周期进行特征重构,构成一个并行周期建模结构.

具体实现中,首先对输入季节项 $S \in \mathbb{R}^{B \times T \times C}$ 沿时间维度对其进行快速傅里叶变换(FFT),将时序数据转换为平均幅度谱 $A(f_i)$ ,通过选取幅值最大的 $k$ 个频率 $f_1, f_2, \dots, f_k$ 对应的周期 $p_1, p_2, \dots, p_k$ 作为主要周期,实现对季节项序列的频域周期识别,计算过程数学表达如下:

$$\mathcal{F}_x = FFT(S) \quad (12)$$

$$A(f_i) = \frac{1}{BC} \sum_{b=1}^B \sum_{c=1}^C |\mathcal{F}_x^{(b,c)}| \quad (13)$$

$$p_i = \frac{T}{f_i}, \quad i = 1, 2, \dots, k \quad (14)$$

随后针对每个周期 $p_i$ ,根据预测长度 $T$ 对输入数据进行镜像填充 $\mathcal{P}$ 以确保后续重塑操作的维度对齐.接着,将填充后的数据 $S_{padding}^{(i)}$ 根据周期 $p_i$ 进行重塑操作 $\mathcal{R}_i$ ,将时间序列沿着其周期维度展开,得到一个二维结构 $S_{reshaped}^{(i)}$ .在此二维结构上,应用一个多尺度二维卷积模块 $Conv$ ,捕捉周期内部和跨周期的复杂模式,处理后的结果 $S_{conv}^{(i)}$ 再通过逆重塑操作 $\mathcal{R}_i^{-1}$ 恢复到序列形状,并截取预测长度的部分,上述过程可表示为:

$$S_{padding}^{(i)} = \mathcal{P}(S, p_i, T) \in \mathbb{R}^{B \times L_i' \times D} \quad (15)$$

$$S_{reshaped}^{(i)} = \mathcal{R}_i(S_{padding}^{(i)}) \in \mathbb{R}^{B \times D \times (L_i'/p_i) \times p_i} \quad (16)$$

$$S_{conv}^{(i)} = Conv(S_{reshaped}^{(i)}) \in \mathbb{R}^{B \times D \times (L_i'/p_i) \times p_i} \quad (17)$$

$$S_{processed}^{(i)} = Truncate(\mathcal{R}_i^{-1}(S_{conv}^{(i)}), T) \in \mathbb{R}^{B \times T \times D} \quad (18)$$

为了整合来自不同周期的处理结果,本文使用这些周期的平均幅度谱 $A(f_i)$ 作为权重,通过Softmax函数归一化后进行加权平均,权重向量 $w \in \mathbb{R}^k$ 和聚合结果 $S_{aggregated}^{(i)}$ 的计算如下:

$$w_i = \frac{\exp(A(f_i))}{\sum_{j=1}^k \exp(A(f_j))}, \quad i = 1, \dots, k \quad (19)$$

$$S_{aggregated}^{(i)} = \sum_{i=1}^k w_i \cdot S_{processed}^{(i)} \in \mathbb{R}^{B \times L \times D} \quad (20)$$

最后,为了保留原始季节项的信息并提升训练稳定性,频域多周期感知块间采用残差连接,得到最终的聚合结果 $S_{out}$ .

$$S_{out} = S_{aggregated} + S \in \mathbb{R}^{B \times L \times D} \quad (21)$$

通过频域多周期感知块对季节项显式周期性的建模,频域季节建模网络能够有效地从时间序列中提取和利用复杂的周期性及频率特征,克服了传统时域方法在处理多重季节性和隐式频率模式上的局限性,为更准确的季节项预测提供了支撑.

### 3 实验与分析

#### 3.1 实验数据集

本文在各种时间序列预测任务中全面评估了所提出的模型,并确认了所提出框架的通用性.实验采用多

个数据集,在 ETT 数据集、Weather 数据集和 Exchange 数据集上展开实验,基本信息如表 1 所示。

表 1 数据集基本信息

数据集	采样间隔	变量维度	样本划分	类型
ETTh1	1 h	7	(8 545, 2 881, 2 881)	电力
ETTh2	1 h	7	(8 545, 2 881, 2 881)	电力
ETTM1	15 min	7	(34 465, 11 521, 11 521)	电力
ETTM2	15 min	7	(34 465, 11 521, 11 521)	电力
Exchange	1 day	8	(4 879, 425, 1 182)	汇率
Weather	10 min	21	(36 792, 5 271, 10 540)	天气

ETT 数据集:包括电力负荷和变压器油温,从我国的两个不同地区采集数据,时间跨度为两年.数据集分为 4 个子集,分别包含时间间隔为 1 h 的数据集 ETTh1、ETTh2 和时间间隔为 15 min 的数据集 ETTM1、ETTM2.

Weather 数据集:包含了 21 个关键天气指标,如空气温度、湿度、风速、降水量等,这些数据在 2020 年期间以每 10 min 的频率记录.

Exchange 数据集:数据集涵盖了 8 个国家 1990–2016 年间每日的货币汇率变化数据,时间跨度为 26 年.

### 3.2 实验设置

本文选取 TimeMixer、PatchTST 等共 6 个模型作为基准模型,本文模型及对比模型均使用 PyTorch 框架实现,使用 AMD Ryzen 7 7500F 处理器, NVIDIA 3060Ti GPU 用于加速模型网络的训练和测试.

详细数据集描述如表 1 所示,其中样本划分中的 3 个数据分别表示训练集、验证集和测试集,输入序列步长为 96,预测序列步长分别为 96、192、336、720.为了保证真实场景中信息不泄露且加速模型收敛,使用 Adam 优化器进行参数优化,批量大小统一设置为 16,训练周期数固定为 10.

### 3.3 评价指标

实验使用均方误差 (mean squared error, MSE) 和平均绝对误差 (mean absolute error, MAE) 作为评价指标来衡量预测值与真实值之间的差异,数值越小代表预测效果越好.评价指标表达式如式 (22)、式 (23):

$$MSE = \frac{1}{T} \sum_{i=1}^T (x_i - \hat{x}_i)^2 \quad (22)$$

$$MAE = \frac{1}{T} \sum_{i=1}^T |x_i - \hat{x}_i| \quad (23)$$

其中,  $T$  为预测序列长度,  $x_i$  和  $\hat{x}_i$  分别对应第  $i$  时刻的真实值和预测值.

### 3.4 实验结果与分析

将本文模型与对比算法的预测结果进行比较,结果如表 2 和图 4 所示.表 2 中加粗字体表示性能最优.次优的结果以下划线表示.图 4 中横坐标表示数据集预测长度,纵坐标表示 MAE 数值.

表 2 各模型在不同数据集下的预测结果

数据集	预测步长	DTFNet		TimeXer		TimeMixer		PatchTST		TimesNet		FEDformer		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	<b>0.372</b>	<b>0.393</b>	0.385	0.402	0.378	0.400	<u>0.377</u>	<u>0.397</u>	0.401	0.421	0.378	0.417	0.439	0.455
	192	<b>0.422</b>	<b>0.423</b>	0.434	0.438	0.428	0.433	0.428	<u>0.432</u>	0.470	0.458	<u>0.425</u>	0.446	0.497	0.490
	336	<b>0.464</b>	<b>0.444</b>	0.481	<u>0.454</u>	0.500	0.455	0.472	0.459	0.512	0.484	<u>0.469</u>	0.468	0.498	0.487
	720	<b>0.473</b>	<b>0.470</b>	0.511	0.487	0.500	<u>0.481</u>	0.522	0.504	0.536	0.508	<u>0.487</u>	0.494	0.508	0.511
	平均	<b>0.433</b>	<b>0.432</b>	0.453	0.445	0.451	<u>0.442</u>	<u>0.450</u>	0.448	0.480	0.468	<u>0.440</u>	0.456	0.486	0.486
ETTh2	96	<b>0.285</b>	<b>0.334</b>	<u>0.288</u>	0.340	0.297	0.344	0.301	0.353	0.323	0.365	0.336	0.383	0.357	0.398
	192	<b>0.369</b>	<b>0.386</b>	<u>0.369</u>	<u>0.393</u>	0.379	0.401	0.381	0.403	0.422	0.422	0.422	0.434	0.454	0.451
	336	<b>0.404</b>	<b>0.418</b>	0.433	0.435	<u>0.417</u>	<u>0.430</u>	0.424	0.436	0.453	0.455	0.471	0.475	0.476	0.478
	720	<b>0.425</b>	<b>0.440</b>	0.440	0.451	<u>0.429</u>	<u>0.445</u>	0.434	0.452	0.450	0.458	0.485	0.492	0.477	0.491
	平均	<b>0.371</b>	<b>0.394</b>	0.383	0.405	<u>0.381</u>	<u>0.405</u>	0.385	0.411	0.412	0.425	0.428	0.446	0.441	0.454
ETTM1	96	<b>0.315</b>	<b>0.353</b>	0.321	0.358	<u>0.318</u>	0.356	0.323	0.364	0.321	0.365	0.380	0.421	0.558	0.500
	192	<u>0.366</u>	<b>0.382</b>	<b>0.365</b>	<u>0.386</u>	0.369	0.387	0.371	0.391	0.374	0.395	0.401	0.431	0.628	0.522
	336	<b>0.391</b>	<b>0.400</b>	<u>0.395</u>	0.409	0.396	<u>0.407</u>	0.397	0.407	0.412	0.417	0.463	0.467	0.665	0.545
	720	<b>0.421</b>	<b>0.428</b>	<u>0.454</u>	<u>0.442</u>	0.463	0.450	0.462	0.448	0.481	0.456	0.513	0.491	0.544	0.512
	平均	<b>0.373</b>	<b>0.391</b>	<u>0.384</u>	<u>0.399</u>	0.386	0.400	0.388	0.402	0.397	0.408	0.439	0.453	0.599	0.520
ETTM2	96	<b>0.168</b>	<b>0.248</b>	<u>0.172</u>	<u>0.258</u>	0.177	0.260	0.185	0.268	0.183	0.265	0.200	0.287	0.384	0.398
	192	<b>0.237</b>	<b>0.293</b>	<u>0.238</u>	<u>0.299</u>	0.240	0.302	0.247	0.307	0.247	0.304	0.263	0.324	0.421	0.421
	336	<b>0.295</b>	<b>0.330</b>	0.299	0.341	<u>0.299</u>	<u>0.340</u>	0.307	0.347	0.318	0.348	0.330	0.367	0.447	0.441
	720	<b>0.392</b>	<b>0.388</b>	0.399	0.398	<u>0.396</u>	<u>0.396</u>	0.411	0.405	0.417	0.403	0.423	0.423	0.463	0.477
	平均	<b>0.273</b>	<b>0.315</b>	<u>0.277</u>	<u>0.324</u>	0.278	0.325	0.287	0.332	0.291	0.330	0.304	0.350	0.429	0.434

表2 各模型在不同数据集下的预测结果(续)

数据集	预测步长	DTFNet		TimeXer		TimeMixer		PatchTST		TimesNet		FEDformer		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange	96	<b>0.085</b>	<u>0.204</u>	0.095	0.214	0.089	0.207	<u>0.085</u>	<b>0.202</b>	0.110	0.238	0.168	0.295	0.147	0.277
	192	<u>0.183</u>	<u>0.304</u>	0.197	0.314	<b>0.174</b>	<b>0.297</b>	0.185	0.305	0.229	0.344	0.280	0.385	0.281	0.386
	336	<u>0.335</u>	<b>0.414</b>	0.340	0.422	0.352	0.428	<b>0.330</b>	<u>0.416</u>	0.378	0.443	0.464	0.500	0.486	0.519
	720	<b>0.842</b>	<b>0.696</b>	<u>0.865</u>	<u>0.702</u>	0.890	0.709	0.895	0.713	0.968	0.749	1.203	0.846	1.084	0.808
	平均	<b>0.361</b>	<b>0.406</b>	0.374	0.413	0.376	0.410	<u>0.374</u>	<u>0.409</u>	0.421	0.444	0.529	0.507	0.499	0.497
Weather	96	<b>0.155</b>	<b>0.198</b>	<u>0.158</u>	<u>0.206</u>	0.162	0.209	0.175	0.216	0.166	0.214	0.197	0.259	0.340	0.375
	192	<b>0.201</b>	<b>0.241</b>	<u>0.204</u>	<u>0.248</u>	0.207	0.251	0.227	0.262	0.238	0.276	0.236	0.294	0.302	0.359
	336	<b>0.261</b>	<b>0.283</b>	<u>0.262</u>	<u>0.291</u>	0.264	0.293	0.281	0.299	0.285	0.306	0.282	0.332	0.360	0.394
	720	<b>0.339</b>	<b>0.337</b>	<u>0.342</u>	<u>0.341</u>	0.344	0.344	0.356	0.348	0.360	0.354	0.347	0.384	0.437	0.442
	平均	<b>0.239</b>	<b>0.265</b>	<u>0.241</u>	<u>0.271</u>	0.244	0.274	0.260	0.281	0.262	0.287	0.266	0.317	0.360	0.393
平均值	<b>0.342</b>	<b>0.367</b>	<b>0.352</b>	<b>0.376</b>	0.353	0.376	0.357	0.381	0.377	0.394	0.401	0.422	0.469	0.464	

如图4所示,本文提出的DTFNet模型在所有数据集和预测步长上的综合性能优于所有基准模型.根据表2,本文模型在所有测试任务上的平均MSE为0.342,平均MAE为0.367,均达到最优.相较于次优的TimeXer模型,本文模型在平均MSE和平均MAE上

分别降低了约2.84%和2.39%.与基于传统分解方法且平均性能相对靠后的Autoformer相比,本文模型的平均MSE和MAE降低了约27.08%和20.91%.这充分证明了DTFNet在处理不同类型和特征的时间序列数据时的有效性和优越性.

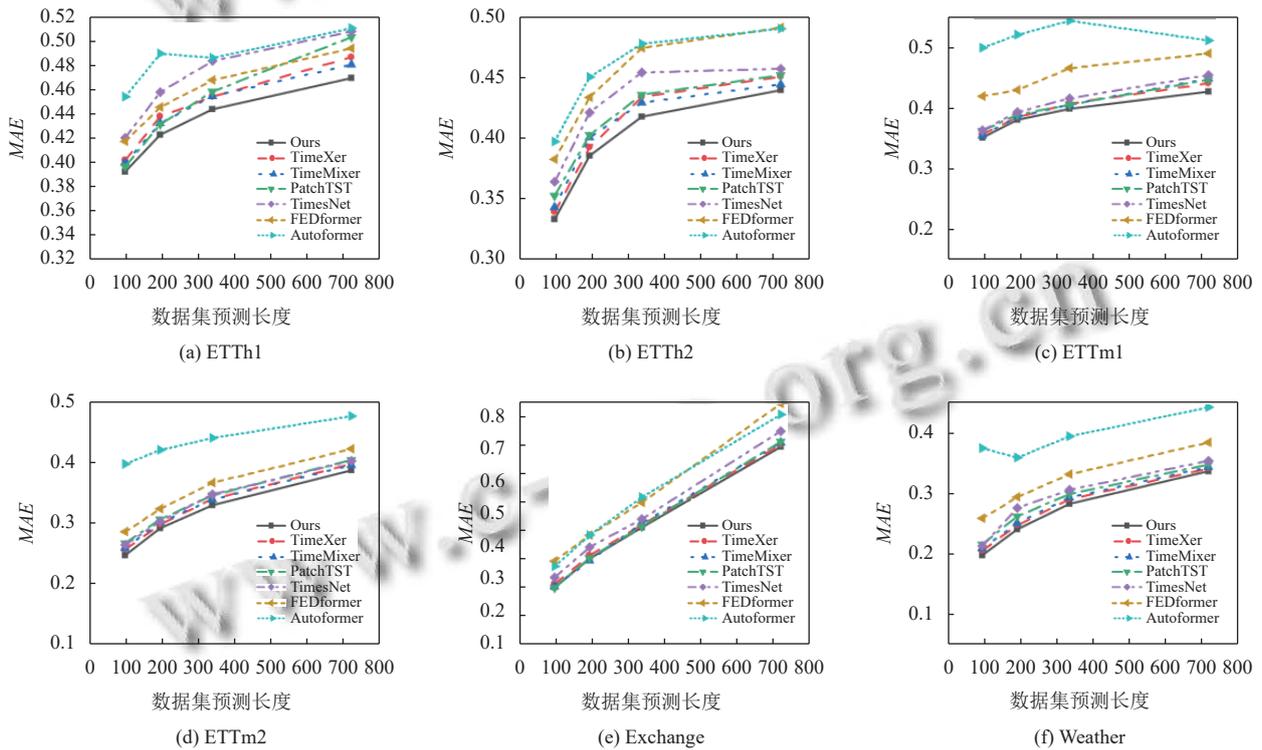


图4 不同数据集下的测试结果图

具体而言,在具有显著非平稳特性的ETTh1数据集上,本文模型与最优基准模型TimeXer相比,本文模型的平均MSE值平均降低了3.99%,MAE值平均降低了2.26%,较传统频域卷积模型TimesNet误差减少了9.79%,与其他模型相比,本文模型的平均MSE值平均降低

了4.41%–10.9%.在高频采样的ETTm1数据集,本文模型的平均MSE较频域专家FEDformer降低15.0%,印证频域卷积模块对快速波动特征的捕获优势.以上实验结果表明本文提出的时频双域协同机制能有效缓解分布偏移问题,特别是在处理电力负荷数据的突变

趋势时展现出更强的模式解耦能力. 此外, 模型在异质数据集上表现出稳定性能. 在强季节特性的 Weather 数据集, 本文模型在该数据集上表现出色, 优于所有基准模型. 相较于次优模型 TimeXer, 本文模型的平均 MAE 降低了约 2.21%. 这表明本文模型能够有效处理不同类型时间序列的复杂模式, 包括强季节性和多变量之间的相互作用. 在相对平缓的 Exchange 数据集中, 本文模型与 PatchTST 性能表现接近, 本文模型的平均 MSE 较 PatchTST 降低了约 3.48%, 平均 MAE 降低了约 0.73%. 相较于其他模型, 优势更为明显, 与 FEDformer 和 Autoformer 相比, 本文模型的平均 MSE 最多降低了约 31.8%. 这验证了本文模型在不同领域数据集上的泛化能力和稳定性.

总体而言, DTFNet 能够在多种数据集和预测长度上取得优异性能. 通过时频信息的有效融合与交互, 模型能够更全面地理解时间序列的复杂结构, 实现更精准的模式解耦, 从而有效应对非平稳性和分布偏移等挑战, 特别是在处理电力负荷数据的突变和天气数据的季节性变化时, 这种协同作用体现得尤为明显.

### 3.5 消融实验

为了验证小波序列分解和时频协同模型设计的有效性, 并探究模型内部各核心组件对预测结果的具体贡献, 本文在 4 个数据集上进行了消融分析. 具体的 3 种模型变体结构如下.

- 模型 1: 保留季节项频域处理不变, 移除时域趋势建模网络, 改用常规线性层替代. 此变体旨在检验专门的趋势建模网络对于捕捉序列长期演化规律的必要性.

- 模型 2: 保留趋势项时域处理不变, 移除频域多周期感知网络, 改用常规卷积层替代. 变体用于评估频域分析在解析和预测周期性波动中的作用.

- 模型 3: 替换小波分解模块为传统移动平均分解模块. 此变体用以证明先进的分解技术是后续趋势和季节精确建模的基础.

表 3 展示了各变体在关键数据集上的性能对比, 结果表明本文所引入的模块能有效地提升时间序列长期预测的准确性, 展现了本文模型在长时间序列预测任务中的有效性和优越性.

表 3 消融实验结果

数据集	预测 步长	DTFNet		模型1		模型2		模型3	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	<b>0.372</b>	<b>0.393</b>	0.391	0.400	0.418	0.427	0.390	0.406
	192	<b>0.422</b>	<b>0.423</b>	0.454	0.439	0.468	0.456	0.436	0.433
	336	<b>0.464</b>	<b>0.444</b>	0.496	0.462	0.491	0.473	0.460	0.449
	720	<b>0.473</b>	<b>0.47</b>	0.515	0.501	0.512	0.513	0.483	0.476
	平均	<b>0.433</b>	<b>0.432</b>	0.464	0.451	0.472	0.467	0.442	0.441
ETTh2	96	<b>0.285</b>	<b>0.334</b>	0.304	0.347	0.320	0.361	0.290	0.340
	192	<b>0.369</b>	<b>0.386</b>	0.391	0.401	0.405	0.411	0.377	0.392
	336	<b>0.404</b>	<b>0.418</b>	0.428	0.435	0.440	0.444	0.426	0.435
	720	<b>0.425</b>	<b>0.44</b>	0.439	0.449	0.446	0.455	0.433	0.446
	平均	<b>0.371</b>	<b>0.394</b>	0.390	0.408	0.403	0.418	0.381	0.403
ETTm1	96	<b>0.315</b>	<b>0.353</b>	0.329	0.358	0.360	0.387	0.323	0.359
	192	<b>0.366</b>	<b>0.382</b>	0.375	0.385	0.391	0.404	0.360	0.378
	336	<b>0.391</b>	<b>0.400</b>	0.428	0.424	0.495	0.459	0.401	0.406
	720	<b>0.421</b>	<b>0.428</b>	0.521	0.473	0.501	0.470	0.465	0.443
	平均	<b>0.373</b>	<b>0.391</b>	0.413	0.410	0.437	0.430	0.387	0.397
ETTm2	96	<b>0.168</b>	<b>0.248</b>	0.186	0.262	0.188	0.263	0.165	0.247
	192	<b>0.237</b>	<b>0.293</b>	0.245	0.298	0.250	0.301	0.240	0.297
	336	<b>0.295</b>	<b>0.33</b>	0.309	0.338	0.317	0.346	0.288	0.328
	720	<b>0.392</b>	<b>0.388</b>	0.414	0.398	0.414	0.399	0.410	0.397
	平均	<b>0.273</b>	<b>0.315</b>	0.288	0.324	0.292	0.327	0.276	0.317
Exchange	96	<b>0.085</b>	<b>0.204</b>	0.103	0.227	0.122	0.250	0.121	0.248
	192	<b>0.183</b>	<b>0.304</b>	0.218	0.334	0.242	0.353	0.239	0.351
	336	<b>0.335</b>	<b>0.414</b>	0.412	0.467	0.425	0.476	0.423	0.475
	720	<b>0.842</b>	<b>0.696</b>	1.061	0.790	1.052	0.787	1.072	0.795
	平均	<b>0.361</b>	<b>0.406</b>	0.448	0.455	0.460	0.467	0.464	0.467

表3 消融实验结果(续)

数据集	预测步长	DTFNet		模型1		模型2		模型3	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	<b>0.155</b>	<b>0.198</b>	0.174	0.223	0.183	0.232	0.180	0.229
	192	<b>0.201</b>	<b>0.241</b>	0.226	0.265	0.233	0.271	0.231	0.268
	336	<b>0.261</b>	<b>0.283</b>	0.279	0.301	0.286	0.307	0.284	0.304
	720	<b>0.339</b>	<b>0.337</b>	0.357	0.349	0.363	0.354	0.360	0.351
	平均	<b>0.239</b>	<b>0.265</b>	0.259	0.284	0.266	0.291	0.264	0.288
平均值		<b>0.342</b>	<b>0.367</b>	0.377	0.389	0.388	0.400	0.369	0.386

具体而言, 移除时域或频域模块均会导致显著的性能衰减. 模型1在ETTh1数据集720个步长预测中MSE相对DTFNet增长8.88%, 表明趋势项时域模块在学习缓变演化轨迹、稳定长期预测方面起关键作用, 趋势项建模对长期趋势预测至关重要; 模型2在ETTh1数据集720个步长预测任务中MSE增加23.8%, 表明季节项频域模块能更好地识别和利用显式或隐式的周期成分, 弥补时域方法对长周期弱敏感的不足, 凸显频域分析对高频波动捕获的关键作用.

将小波分解替换为移动平均分解的模型3同样表现不佳, 尤其在Exchange和Weather数据集上, 性能下降尤为明显. Exchange数据集的平均MSE由0.361上升至0.464, 表明小波分解在处理金融数据等非平稳序列时更具优势, 高质量的小波分解在分离纯净趋势与季节成分、为后续专属网络提供优质输入方面的基础性作用.

综合来看, 完整的模型通过时域和频域的协同学习, 实现了对时间序列特征的全面捕捉. 本文模型实现了两个模块功能互补的有机结合, 时域趋势建模网络善于捕捉局部模式和长期依赖, 为预测提供一个稳定的宏观基线; 而季节频域模块则在此基线之上, 叠加精细的周期性成分和全局波动. 通过将序列的长期趋势与短期细节解耦并分别建模, 本文模型显著提升了模型在长期预测任务中的准确性和稳健性.

### 3.6 小波分解参数实验

为了深入探讨小波分解参数对分量提取效果的影响, 本文从ETTh1数据集中选取长度为96的样本, 分别与移动平均分解(滑动核为25)进行对比, 同时在小波分解中固定母小波为Daubechies-4(db4), 分别设置分解层数为1、2、3级, 实验结果如图5所示.

从图5的对比中可以更直观地看到两种方法的本质差异. 移动平均法(图5(a))生成的趋势线在序列突变点( $t \approx 30$ )处存在明显滞后和过度平滑的问题, 未能

捕捉关键转折. 相比之下, 小波分解的趋势项既能反映整体走势, 又有效保留了此类突变特征, 证明其在处理非平稳信号时的优越性. 在季节项上, 移动平均的残差包含了大量混合噪声, 而小波分解则能清晰地分离出周期性与短期波动信息. 这表明小波变换提供了一种更精细、保真的分解方式, 为后续建模提供了更高质量的特征输入.

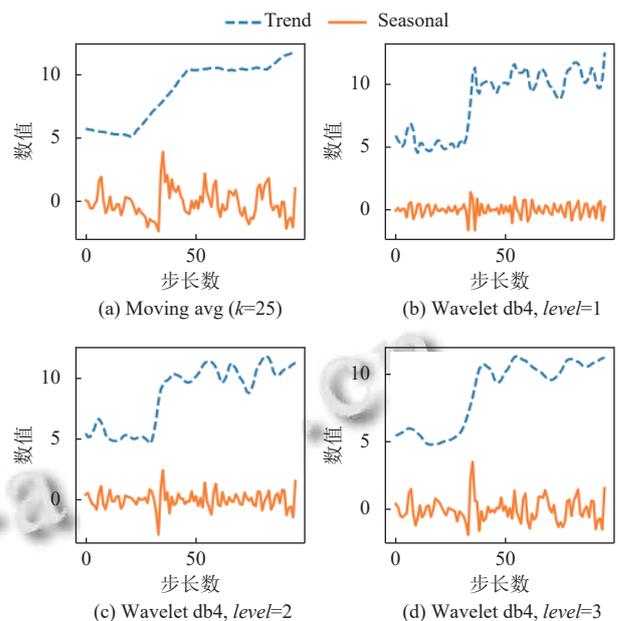


图5 小波分解参数实验图

综上所述, 基于小波变换的分解方法可以更好地控制趋势和季节性成分的行为. 与移动平均法相比, 小波变换分解是一种更灵活的分解方法, 因为它能够自适应地分离不同尺度的时频成分. 这种适应性使小波变换能够更有效地捕捉变化趋势, 因此特别适用于具有动态和演变模式的时间序列.

## 4 总结

本文针对现有时间序列预测方法在非平稳性建模和异质模式解耦方面存在的局限性, 提出了一种时频

双域协同网络 (DTFNet). 该方法通过构建时域与频域并行的异构建模结构, 实现了对趋势与季节分量的差异化建模, 同时引入基于离散小波变换的序列分解模块, 有效缓解了传统移动平均方法在边界效应和模态混叠方面的问题. 在多个公开数据集上的实证研究表明, DTFNet 在长期预测精度和模型鲁棒性方面均优于现有主流模型, 具备良好的通用性和推广能力.

尽管 DTFNet 取得了显著成果, 但仍有值得探索的进阶方向. 首先, 当前模型对趋势和季节分量进行并行处理, 而现实中两者可能存在动态的相互影响. 未来的研究可探索两者间的动态交互机制, 例如引入门控或注意力机制, 使模型能根据趋势的变化自适应地调整对季节模式的建模, 从而更真实地反映时间序列的内在演化规律. 其次, 将模型扩展至跨域预测任务极具挑战性, 核心在于克服不同领域间的数据分布差异和模式鸿沟. 因此, 进一步研究如何增强 DTFNet 的领域自适应能力, 是提升其在更广泛场景下泛化能力的关键.

综上, 未来的研究我们将进一步探索趋势与季节分量之间的动态交互机制, 以及将 DTFNet 扩展至跨域时序预测任务中, 以进一步提升模型在复杂场景下的适应能力与解释性.

### 参考文献

- 1 Espinosa R, Palma J, Jiménez F, *et al.* A time series forecasting based multi-criteria methodology for air quality prediction. *Applied Soft Computing*, 2021, 113: 107850. [doi: [10.1016/j.asoc.2021.107850](https://doi.org/10.1016/j.asoc.2021.107850)]
- 2 Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 2016, 32(3): 914–938. [doi: [10.1016/j.ijforecast.2015.11.011](https://doi.org/10.1016/j.ijforecast.2015.11.011)]
- 3 Idrees SM, Alam MA, Agarwal P. A prediction approach for stock market volatility based on time series data. *IEEE Access*, 2019, 7: 17287–17298. [doi: [10.1109/ACCESS.2019.2895252](https://doi.org/10.1109/ACCESS.2019.2895252)]
- 4 Box GEP, Pierce DA. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 1970, 65(332): 1509–1526. [doi: [10.1080/01621459.1970.10481180](https://doi.org/10.1080/01621459.1970.10481180)]
- 5 Gardner Jr ES. Exponential smoothing: The state of the art. *Journal of Forecasting*, 1985, 4(1): 1–28. [doi: [10.1002/for.3980040103](https://doi.org/10.1002/for.3980040103)]
- 6 Medsker L, Jain LC. *Recurrent neural networks: Design and applications*. Boca Raton: CRC press, 1999.
- 7 Bai SJ, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv:1803.01271*, 2018.
- 8 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 9 Liu Y, Hu TG, Zhang HR, *et al.* iTransformer: Inverted Transformers are effective for time series forecasting. *Proceedings of the 12th International Conference on Learning Representations*. Vienna: OpenReview.net, 2024.
- 10 Nie YQ, Nguyen NH, Sinthong P, *et al.* A time series is worth 64 words: Long-term forecasting with Transformers. *Proceedings of the 11th International Conference on Learning Representations*. Kigali: OpenReview.net, 2023.
- 11 Zeng AL, Chen MX, Zhang L, *et al.* Are Transformers effective for time series forecasting? *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington: AAAI Press, 2023. 11121–11128.
- 12 Wang SY, Wu HX, Shi XM, *et al.* TimeMixer: Decomposable multiscale mixing for time series forecasting. *Proceedings of the 12th International Conference on Learning Representations*. Vienna: OpenReview.net, 2024.
- 13 Brockwell PJ, Davis RA. *Introduction to Time Series and Forecasting*. New York: Springer, 2002.
- 14 Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359. [doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191)]
- 15 Qiu XF, Wu XJ, Lin Y, *et al.* DUET: Dual clustering enhanced multivariate time series forecasting. *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*. Toronto: ACM, 2025. 1185–1196.
- 16 Wu HX, Xu JH, Wang JM, *et al.* Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2021. 1717.
- 17 Cleveland RB, Cleveland WS, McRae JE, *et al.* STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 1990, 6(1): 3–73.
- 18 Zhou T, Ma ZQ, Wen QS, *et al.* FEDformer: Frequency enhanced decomposed Transformer for long-term series forecasting. *Proceedings of the 39th International Conference*

- on Machine Learning. PMLR, 2022. 27268–27286.
- 19 Nava N, Di Matteo T, Aste T. Dynamic correlations at different time-scales with empirical mode decomposition. *Physica A: Statistical Mechanics and its Applications*, 2018, 502: 534–544. [doi: [10.1016/j.physa.2018.02.108](https://doi.org/10.1016/j.physa.2018.02.108)]
- 20 Delage O, Bencherif H, Portafaix T, *et al.* Variability analysis of observational time series: An overview of the decomposition methods for non-stationary and noisy signals. In: Rocha J, Viana CM, Oliveira S, eds. *Time Series Analysis-Recent Advances, New Perspectives and Applications*. IntechOpen, 2023.
- 21 Stitsyuk A, Choi J. xPatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition. *Proceedings of the 39th AAAI Conference on Artificial Intelligence*. Philadelphia: AAAI Press, 2025. 20601–20609.
- 22 Noor TH, Almars AM, Alwateer M, *et al.* SARIMA: A seasonal autoregressive integrated moving average model for crime analysis in Saudi Arabia. *Electronics*, 2022, 11(23): 3986. [doi: [10.3390/electronics11233986](https://doi.org/10.3390/electronics11233986)]
- 23 Sims CA. Macroeconomics and reality. *Econometrica*, 1980, 48(1): 1–48. [doi: [10.2307/1912017](https://doi.org/10.2307/1912017)]
- 24 Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient Transformer for long sequence time-series forecasting. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. AAAI Press, 2021. 11106–11115.
- 25 Wang YX, Wu HX, Dong JX, *et al.* TimeXer: Empowering Transformers for time series forecasting with exogenous variables. *Proceedings of the 38th Annual Conference on Neural Information Processing Systems*. Vancouver: OpenReview.net, 2024.
- 26 Wu HX, Hu TG, Liu Y, *et al.* TimesNet: Temporal 2d-variation modeling for general time series analysis. *Proceedings of the 11th International Conference on Learning Representations*. OpenReview.net, 2023.

(校对责编: 张重毅)