

融合多尺度特征和注意力机制的时间序列预测模型

潘建, 汪绪豪

引用本文

潘建, 汪绪豪. 融合多尺度特征和注意力机制的时间序列预测模型[J]. 计算机科学, 2026, 53(2): 180-186.

PAN Jian, WANG Xuhao. [Time Series Forecasting Model Integrating Multi-scale Features and Attention Mechanism](#) [J]. Computer Science, 2026, 53(2): 180-186.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[聚焦边界和多尺度特征融合的脑卒中病灶分割](#)

Boundary-focused Multi-scale Feature Fusion Network for Stroke Lesion Segmentation

计算机科学, 2026, 53(2): 264-272. <https://doi.org/10.11896/jsjcx.250300137>

[D-LINet:融合双线性层与双向归一化的时间序列预测框架](#)

D-LINet:Time Series Forecasting Framework Integrating Dual-linear Layersand Dual Normalization

计算机科学, 2026, 53(2): 170-179. <https://doi.org/10.11896/jsjcx.250100137>

[GTKT:融合联通主义学习和多层时序图Transformer的知识追踪模型](#)

GTKT:Knowledge Tracing Model Integrating Connectivism Learning and Multi-layer TemporalGraph Transformer

计算机科学, 2026, 53(2): 78-88. <https://doi.org/10.11896/jsjcx.250700188>

[基于KAN的无监督多元时间序列异常检测网络](#)

KAN-based Unsupervised Multivariate Time Series Anomaly Detection Network

计算机科学, 2026, 53(1): 89-96. <https://doi.org/10.11896/jsjcx.241200190>

[层次时间序列预测方法与应用综述](#)

Comprehensive Review of Hierarchical Time Series Forecasting Methods and Applications

计算机科学, 2025, 52(11A): 241000139-7. <https://doi.org/10.11896/jsjcx.241000139>

融合多尺度特征和注意力机制的时间序列预测模型

潘建^{1,2} 汪绪豪²

1 浙江工业大学之江学院 浙江 绍兴 312030

2 浙江工业大学计算机科学与技术学院 杭州 310023

摘要 目前,在时间序列预测任务的研究中,基于 Transformer 的模型主要关注从时序数据中提取全局性和局部性特征,并通过改进注意力机制以降低模型的复杂度。然而,现有方法往往忽略了时间序列在多个尺度上展现出的不同粒度特征。针对这一问题,提出了一种融合多尺度特征和注意力机制的时间序列预测模型——MTSformer。首先,通过对原始序列进行下采样,得到多个尺度的子序列,使模型能够融合多个尺度的特征信息,从而增强泛化能力;其次,使用多预测头代替传统的解码器,在提升预测速度的同时降低模型的复杂度;最后,在 5 个基准数据集上进行了实验,结果显示,与现有的方法相比,MTSformer 模型在时间序列预测上的 MSE 平均降低了 24.51%,MAE 平均降低了 17.84%。

关键词: 时间序列预测;多尺度特征;Transformer;多预测头;下采样

中图分类号 TP391

Time Series Forecasting Model Integrating Multi-scale Features and Attention Mechanism

PAN Jian^{1,2} and WANG Xuhao²

1 Zhijiang College of Zhejiang University of Technology, Shaoxing, Zhejiang 312030, China

2 College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract Currently, in the research of time series forecasting tasks, Transformer-based models primarily focus on extracting global and local features from time series data and improving attention mechanisms to reduce model complexity. However, existing methods often overlook the different granularity features exhibited by time series at multiple scales. To address this issue, this paper proposes a time series forecasting model that integrates multi-scale features and the attention mechanism, called MTSformer. Firstly, by down-sampling the original sequence, multiple scale subsequences are obtained, enabling the model to integrate multi-scale feature information and enhance generalization ability. Then, a multi-prediction head structure is used to replace the traditional decoder, which improves prediction speed while reducing model complexity. Finally, experiments are conducted on five benchmark datasets, and the results show that compared with existing methods, the MTSformer model achieves average reductions of 24.51% in MSE and 17.84% in MAE for time series forecasting.

Keywords Time series forecasting, Multi-scale features, Transformer, Multi-head prediction, Downsampling

1 引言

时间序列(Time Series)是按照时间维度以固定的采样频率生成的一组或者多组变量^[1]。时间序列预测(Time Series Forecasting)是一种利用历史数据来预测未来时间点上的数值或趋势的方法,它在许多领域中都有广泛的应用,例如医学、金融、气象、农业、电力等^[2-4]。由于时间序列数据往往具有复杂的动态性和高度的噪声,传统的基于统计学特征和数学模型的时间序列预测方法在处理这些数据时面临着诸多挑战,如难以捕捉数据中的潜在模式和趋势,且容易受到异常数据的干扰等。

在面对大规模、高维度、异构性的时间序列数据时,基于深度学习的方法比传统方法具有更好的性能。近年来,得益于 Transformer^[5]模型优秀的捕获长距离依赖的能力以及强大的拟合能力,基于 Transformer 的模型已被广泛应用于时间序列预测。但由于时间序列数据往往具有复杂的动态性和高度的噪声,普通的点状注意力机制无法充分捕获时间序列的局部以及全局特征,更无法兼顾序列在多个时间尺度上的特征,并且 Transformer 模型复杂的编码器-解码器结构也会带来较大的内存消耗和较高的时间复杂度。

为了解决上述问题,本文提出了一种融合多尺度特征和

到稿日期:2025-01-17 返修日期:2025-05-06

基金项目:浙江省自然科学基金探索项目(LGF20F020015)

This work was supported by the Natural Science Foundation of Zhejiang Province, China(LGF20F020015).

通信作者:潘建(pj@zjut.edu.cn)

注意力机制的时间序列预测模型 MTSformer (Transformer with Multi-scale Features)。该模型对输入序列进行多尺度时序编码,然后通过编码器的多头自注意力机制和残差连接提取多尺度时序特征,输入到模型解码器中,使用多预测头预测多个尺度的输出,最后将多个尺度的输出进行融合得到最终的预测结果。

本文的主要贡献如下:

1)提出了一种融合多尺度特征和注意力机制的时间序列预测模型 MTSformer,有效增强了模型的泛化能力;

2)在多个时间序列数据集上取得了较好的结果,在多个预测长度上的预测性能优于 PatchTST, Timesnet, Dlinear 等基准模型。

2 相关工作

时间序列预测是统计学和数据科学中的一个关键领域,旨在通过分析过去的模式来预测未来的值。传统的基于统计学特征的时间序列预测方法包括移动平均法(MA)、指数平滑法、自回归模型(AR)和自回归积分移动平均模型(ARIMA)等。近年来,得益于深度学习模型强大的非线性建模能力,研究人员提出了许多基于深度学习的时间序列预测模型。

基于马尔可夫假设的循环神经网络^[6] (Recurrent Neural Network, RNN)及其变体门控循环单元^[7] (Gate Recurrent Unit, GRU)、长短时记忆网络^[8] (Long Short-Term Memory, LSTM)等,通过对连续时间点进行建模来预测时间序列,但存在梯度消失、梯度爆炸等问题,在处理长时间序列数据时效果不理想。

基于卷积神经网络的模型,如 Lea 等^[9]提出的 TCN,使用因果卷积和膨胀卷积来建模序列的长时间依赖关系。Wu 等^[10]提出的 Timesnet,使用 KL 散度计算序列的周期性,将 1 维序列转换为 2 维,使用 2 维卷积处理时间序列数据。但基于卷积神经网络的方法都无法充分捕捉长期依赖信息,同时,深层的卷积结构会带来较大的计算开销和模型复杂性。

基于 Transformer 的模型,如 Zhou 等^[11]提出的 Informer,使用稀疏注意力机制、注意力蒸馏以及生成式解码器提高模型性能,降低模型复杂度。Wu 等^[12]提出的 Autoformer,使用序列分解块和自相关机制将 Transformer 更新为一个深度分解结构,提高信息利用率以提高模型性能和效率。Kitaev 等^[13]提出的 Reformer,引入局部敏感哈希注意力、分块处理和可逆网络,在保证模型性能的同时降低模型的空间和时间复杂度。Zhou 等^[14]提出的 FEDformer,将时间序列分解并和 Transformer 相结合,并在频域上进行分析,取得了良好的效果。Nie 等^[15]提出的 PatchTST,引入计算机视觉领域中常用的补丁机制来处理时间序列数据,将输入分割为若干个补丁并将补丁编码输入 Transformer 中,取得了良好的效果。Chen 等^[16]于 2024 年提出的 Pathformer 也是基于补丁机制来处理时间

序列数据的,将不同的补丁划分为不同的分辨率,并基于每个尺度的划分对补丁进行双重关注来捕获全局和局部特征。Zeng 等^[17]提出的 Dlinear 模型证明了线性模型的预测能力与 Transformer 类模型相当,同时模型结构更加简单,模型收敛速度更快。

上述基于 Transformer 的模型主要关注降低模型的复杂度或捕获时间序列本身的周期性特征,都未充分考虑到序列整体的多个时间尺度,没有从序列整体的时间跨度进行不同粒度的观察。

3 问题定义

给定一个时间序列 $\mathbf{X} = \{x_t\}_{t=1}^T$, 其中 x_t 表示在时间 t 的观测值,长时间序列预测问题旨在使用过去 N 个时间步的数据来预测未来 M 个时间步的值。具体来说,输入序列 $\mathbf{X} = \{x_{T-N+1}, x_{T-N+2}, \dots, x_T\}$ 是长度为 N 的历史观测值,输出序列 $\hat{\mathbf{X}} = \{\hat{x}_{T+1}, \hat{x}_{T+2}, \dots, \hat{x}_{T+M}\}$ 是长度为 M 的预测值。预测模型将历史观测值作为输入,并生成未来的预测值。

4 MTSformer 模型

MTSformer 模型对输入序列进行多尺度时序编码,使得注意力机制能够融合多尺度时序特征,并使用多预测头代替传统的解码器,降低模型的复杂度。模型的总体流程如下:

步骤 1 对长度为 N 的历史观测值序列 \mathbf{X} 进行 k 次下采样得到 k 个尺度的子序列;

步骤 2 将原始序列 \mathbf{X} 和 k 个不同尺度子序列经多尺度时序编码后输入编码器中,通过多头,注意力机制提取到 $k+1$ 个多尺度的时序特征;

步骤 3 将这 $k+1$ 个多尺度的时序特征经解码器的多预测头得到不同尺度的预测输出;

步骤 4 将 $k+1$ 个尺度的预测输出进行融合,得到模型的最终预测值 $\hat{\mathbf{X}}$ 。

4.1 总体结构

MTSformer 模型的总体结构如图 1 所示。嵌入层对原始输入序列进行多尺度时序编码,编码器使用多头注意力机制来提取多尺度时序特征,解码器使用多个线性预测头来代替原生 Transformer 解码器的注意力模块。

MTSformer 模型通过下采样生成层次化子序列,在本质上形成了对原始时间序列的多粒度逼近。粗尺度子序列捕获低频趋势(如周期性规律),细尺度子序列保留高频细节(如突发性波动),从而满足信号重构的完备性条件,即通过线性组合各尺度特征能以预设误差精度还原原始序列。多头注意力机制则通过获取粗尺度和细尺度的信息增益得到更强的注意力连接,在跨尺度的关键信息交互时过滤无关的噪声扰动。传统自回归解码器因逐步生成预测序列,其计算复杂度随预测步长呈二次增长,导致长时预测效率低下。本文采用的并行化线性预测头可将复杂度降至线性级别。

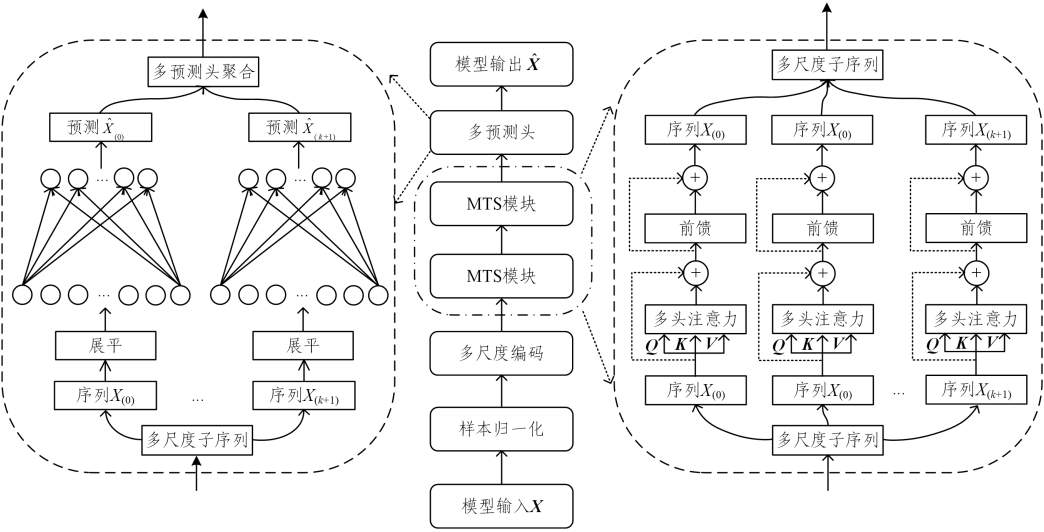


图1 模型结构

Fig.1 Structure of model

4.2 多尺度时序编码

首先,对原始序列进行下采样得到多个尺度的子序列并进行时序编码,过程如图2所示。

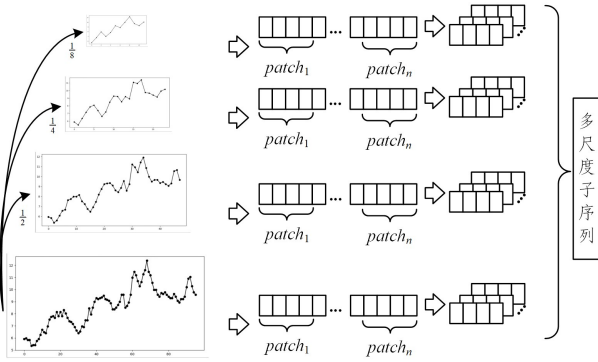


图2 多尺度时序编码

Fig.2 Multi-scale temporal encoding

从时间索引1开始、步长为 N 的历史观测值序列 $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ 进行 k 次下采样操作,得到 k 个不同尺度的时间序列,第 i 次下采样后的子序列 $\mathbf{X}_{(i)}$ 表示为:

$$\mathbf{X}_{(i)} = \{x_1, x_2, \dots, x_{\frac{N}{2^i}}\}, i \in \{1, 2, \dots, k\} \quad (1)$$

下采样的过程使用步长为2的1维均值池化,对于第 i 个时间窗口 $\mathbf{W}_i = \{x_{2i-1}, x_{2i}\}$,将其均值作为采样后的子序列的第 i 个时刻的值 \bar{x}_i :

$$\bar{x}_i = \frac{1}{2} (x_{2i-1} + x_{2i}), i \in \left\{1, 2, \dots, \frac{N}{2}\right\} \quad (2)$$

原始序列加上下采样后的子序列,共有 $k+1$ 个序列,分别对这 $k+1$ 个序列采用长度为 P 、步长为 S 的滑动窗口,将输入的时间序列划分为 L 个补丁:

$$L = \left\lfloor \frac{N-P}{S} \right\rfloor + 2 \quad (3)$$

展开和重塑补丁,得到第 i 个尺度的时序编码 $\hat{\mathbf{X}}_{(i)}$:

$$\hat{\mathbf{X}}_{(i)} = \text{Unfold}(\text{Pad}(\mathbf{X}_{(i)}, P, S), P, S), i \in \{1, 2, \dots, k\} \quad (4)$$

4.3 MTS 模块

MTSformer 模型使用多头注意力机制提取多尺度时序

编码中的时序特征。

首先,使用线性的投影 \mathbf{W}_p 将多尺度编码映射为 D 维的张量,并附加位置编码 \mathbf{W}_{pos} 来帮助模型捕获序列的时间顺序,得到编码器的输入 $\hat{\mathbf{X}}'_{(i)}$:

$$\hat{\mathbf{X}}'_{(i)} = \mathbf{W}_p(\hat{\mathbf{X}}_{(i)}) + \mathbf{W}_{\text{pos}}(\hat{\mathbf{X}}_{(i)}) \quad (5)$$

然后,多头注意力的每个头对其进行线性投影,得到3个参数矩阵 \mathbf{Q}, \mathbf{K} 和 \mathbf{V} ,并计算多头注意力 $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right) \times \mathbf{V} \quad (6)$$

模型在经过多尺度时序编码和多头注意力机制的处理之后,得到多尺度的时序特征 \mathbf{X} :

$$\mathbf{X} = \{\mathbf{X}_{\text{enc}}^{(0)}, \mathbf{X}_{\text{enc}}^{(1)}, \dots, \mathbf{X}_{\text{enc}}^{(k)}\} \quad (7)$$

其中, $\mathbf{X}_{\text{enc}}^{(i)}$ 是每个尺度的时序数据经过编码器后的输出特征, $\mathbf{X}_{\text{enc}}^{(0)}$ 则是原始序列经过编码器后的输出特征。

原始序列包含不同物理尺度的混合模式,对单一尺度的序列进行建模会导致特征丢失。全局注意力机制在长序列中会产生大量无关的查询键值对交互,例如短期波动特征与长期趋势特征的有效关联。多尺度时序编码模块生成尺度递减的子序列,再由 MTS 模块对每个尺度的序列进行查询键值对交互,能够在避免多头注意力机制产生无关联交互的同时保留多头注意力机制强大的捕获特征的能力,从而使模型能够动态捕获各个关键尺度的信息流。

4.4 多预测头

与 Transformer 模型采用注意力机制的解码器不同, MTSformer 模型使用多个线性的预测头对编码器的输出进行预测,极大地降低了模型的复杂度,提高了模型的预测速度。

不同尺度的时序特征可以反映原始时间序列中粗粒度和细粒度的特征,为了充分融合这些多尺度特征,采用线性的预测头将这些不同尺度的序列特征映射为原始输入序列的长度,并进行特征融合,得到解码器的输出,即最终的预测结果 \hat{X}_{dec} :

$$\hat{X}_{\text{dec}}^{(i)} = \text{Preject}(\mathbf{X}_{\text{enc}}^{(i)}), i \in \{1, 2, \dots, k\} \quad (8)$$

$$X_{\text{dec}} = \sum_{i=0}^k \hat{X}_{\text{enc}}^{(i)} \quad (9)$$

其中, $\hat{X}_{\text{enc}}^{(i)}$ 表示将第 i 个尺度映射到原始尺度长度的预测值。

采用多预测头取代传统解码器后,模型的时间复杂度分析如下。

首先设原始序列的序列长度为 L , 编码器层数 $N_e = 3$, 解码器层数 $N_d = 3$, 特征维度为 d , 多头注意力头数 $h = 8$, 预测步长 $T = 24$, 多尺度编码的序列数为 3。

传统基于编码器解码器结构的模型时间复杂度主要来自于编码器和解码器的多头注意力机制, 因此编码器和解码器的时间复杂度分别为 $O(N_e L^2 d)$ 和 $O(N_d L T d)$, 总的时间复杂度为 $O(N_e L^2 d + N_d L T d) \approx 3 L^2 d + 72 L d$ 。

本文提出的 MTS 模块的时间复杂度为 $O\left(\sum_{i=1}^3 \left(\frac{L}{s_i}\right)^2 d\right)$, 多预测头的时间复杂度为 $O(3dT)$, 因此总的时间复杂度为 $O\left(\sum_{i=1}^3 \left(\frac{L}{s_i}\right)^2 d + 3dT\right) \approx 1.31 L^2 d + 72d$, 其中 s_i 表示第 i 个尺度 ($s_1 = 1, s_2 = 2, s_3 = 4$)。

当序列长度 $L > 100$ 时, 本文模型的时间复杂度相较于传统的基于编码器解码器结构的模型降低约 64.6%, 可见融合多尺度时序编码和多头注意力机制并使用多预测头代替传统解码器, 可显著提高模型预测速度, 增强模型性能。

5 实验及结果分析

5.1 实验设置

5.1.1 数据集

为了评估 MTSformer 模型的性能, 本文在变压器油温 (ETT)、电力^[18] (Electricity)、天气^[19] (Weather) 和交通^[20] (Traffic) 4 个公共数据集上进行了对比实验。

变压器油温 (ETT) 数据集包含多个负载列和一个油温列。ETT 数据集按照采样频率 (1 小时, 15 分钟) 的不同分为 ETTh 和 ETTm, 且每种都包含两个子集, 因此共有 4 个子数据集, 分别是 ETTh1, ETTh2, ETTm1 和 ETTm2。

电力 (Electricity) 数据集包含多个用户的电力消耗, 每一列都对应一个用户。

天气 (Weather) 数据集包含德国一年内的 21 个气象指标。

交通 (Traffic) 数据集描述了道路占用率, 包含 2015 年至

2016 年旧金山高速公路传感器记录的每小时数据。表 1 统计了 7 个公开数据集的特征。

表 1 数据集特征

Table 1 Dataset characteristics

数据集	时间步	特征数	采样频率
ETTh1	17 420	7	1 h
ETTh2	17 420	7	1 h
ETTm1	69 680	7	15 min
ETTm2	69 680	7	15 min
Electricity	26 304	321	1 h
Weather	52 696	21	10 min
Traffic	17 544	862	1 h

5.1.2 评价指标

本文使用均方误差 (Mean Squared Error, MSE) 损失函数和平均绝对误差 (Mean Absolute Error, MAE) 损失函数来衡量模型的预测性能。MSE 和 MAE 的计算式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

其中, n 是样本数, y_i 是第 i 个样本的真实值, \hat{y}_i 是第 i 个样本的预测值。

5.1.3 超参数设置

对比实验均使用 Adam 优化器, 学习率设置为 0.0001, 批次大小为 32, 训练轮次为 10。若训练损失在验证集上 3 个轮次未下降, 则提前停止训练。

5.1.4 基准模型

对比的基准模型包括 Pathformer, PatchTST, FEDformer, Crossformer, Autoformer 和 Informer 共 6 个基于 Transformer 的模型, 以及基于卷积神经网络的模型 Timesnet 和基于多层感知器的模型 Dlinear。

5.2 实验结果分析

5.2.1 对比实验

实验设置历史回望窗口的长度为 96, 对每个数据集分别预测 96, 192, 336 和 720 的输出长度。

实验结果如表 2 所列, 本文提出的 MTSformer 模型在 ETTh1, ETTm2, Weather, Electricity 和 Traffic 这 5 个数据集上的均方误差 (MSE) 和平均绝对误差 (MAE) 相较于 8 个基准模型均有所降低, 其中 MSE 平均降低 24.51%, MAE 平均降低 17.84%, 预测性能有较大的提升。

表 2 5 个数据集上的多变量长期预测结果

Table 2 Multivariate long-term series forecasting results on five datasets

方法 指标	MTSformer		Pathformer		PatchTST		TimesNet		Dlinear		Crossformer		FEDformer		Autoformer		Informer		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
Traffic	96	0.477	0.308	0.479	0.310	0.526	0.347	0.593	0.321	0.650	0.587	0.644	0.429	0.587	0.366	0.613	0.388	0.719	0.391
	192	0.475	0.307	0.484	0.295	0.522	0.332	0.617	0.336	0.598	0.604	0.665	0.431	0.604	0.373	0.616	0.382	0.696	0.379
	336	0.491	0.312	0.503	0.310	0.517	0.334	0.629	0.336	0.605	0.621	0.674	0.420	0.621	0.383	0.622	0.337	0.777	0.420
	720	0.528	0.333	0.537	0.334	0.552	0.352	0.640	0.350	0.645	0.626	0.683	0.424	0.626	0.382	0.660	0.408	0.864	0.472
	AVG	0.493	0.315	0.500	0.312	0.529	0.341	0.620	0.336	0.625	0.610	0.667	0.426	0.610	0.376	0.628	0.379	0.764	0.416
Weather	96	0.176	0.215	0.167	0.199	0.182	0.223	0.172	0.220	0.195	0.217	0.195	0.271	0.217	0.296	0.266	0.336	0.300	0.384
	192	0.224	0.258	0.216	0.246	0.227	0.260	0.219	0.261	0.237	0.276	0.209	0.277	0.276	0.336	0.307	0.367	0.598	0.544
	336	0.278	0.297	0.268	0.302	0.280	0.306	0.246	0.337	0.282	0.339	0.273	0.332	0.339	0.380	0.359	0.395	0.578	0.523
	720	0.349	0.346	0.346	0.343	0.356	0.349	0.365	0.359	0.345	0.403	0.379	0.401	0.403	0.428	0.419	0.428	1.059	0.741
	AVG	0.257	0.280	0.249	0.272	0.261	0.285	0.251	0.294	0.265	0.309	0.264	0.320	0.309	0.360	0.338	0.382	0.634	0.548

(续表)

方法 指标	MTSformer		Pathformer		PatchTST		TimesNet		Dlinear		Crossformer		FEDformer		Autoformer		Informer		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
Electricity	96	0.169	0.259	0.165	0.247	0.190	0.269	0.168	0.272	0.210	0.193	0.219	0.314	0.193	0.308	0.201	0.317	0.327	0.411
	192	0.177	0.267	0.177	0.269	0.199	0.304	0.184	0.322	0.210	0.201	0.231	0.322	0.201	0.315	0.222	0.334	0.296	0.386
	336	0.205	0.294	0.199	0.287	0.217	0.319	0.198	0.300	0.223	0.214	0.246	0.337	0.214	0.329	0.231	0.443	0.300	0.394
	720	0.245	0.326	0.246	0.329	0.258	0.352	0.220	0.320	0.258	0.222	0.280	0.363	0.222	0.355	0.254	0.361	0.373	0.439
	AVG	0.199	0.287	0.197	0.283	0.216	0.311	0.193	0.304	0.225	0.208	0.244	0.334	0.208	0.327	0.227	0.364	0.311	0.397
ETTh1	96	0.376	0.397	0.387	0.403	0.378	0.399	0.384	0.402	0.397	0.395	0.423	0.448	0.395	0.424	0.449	0.459	0.865	0.713
	192	0.422	0.428	0.443	0.431	0.425	0.427	0.436	0.429	0.446	0.469	0.471	0.474	0.469	0.470	0.500	0.482	1.008	0.792
	336	0.465	0.452	0.468	0.449	0.468	0.456	0.638	0.469	0.489	0.530	0.570	0.546	0.530	0.499	0.521	0.496	1.107	0.809
	720	0.505	0.488	0.487	0.474	0.514	0.494	0.521	0.500	0.513	0.598	0.653	0.621	0.598	0.544	0.514	0.512	1.181	0.865
	AVG	0.422	0.441	0.446	0.439	0.446	0.444	0.495	0.450	0.461	0.498	0.529	0.522	0.498	0.484	0.496	0.487	1.040	0.795
ETTM2	96	0.177	0.259	0.177	0.261	0.183	0.270	0.187	0.267	0.193	0.203	0.287	0.366	0.203	0.287	0.255	0.339	0.365	0.453
	192	0.241	0.302	0.239	0.305	0.255	0.314	0.249	0.309	0.284	0.268	0.414	0.492	0.268	0.328	0.281	0.340	0.533	0.563
	336	0.244	0.303	0.293	0.331	0.309	0.347	0.321	0.351	0.382	0.325	0.597	0.542	0.325	0.366	0.339	0.372	1.363	0.887
	720	0.407	0.417	0.397	0.389	0.412	0.404	0.408	0.403	0.558	0.422	1.730	1.042	0.422	0.415	0.422	0.419	3.379	1.338
	AVG	0.267	0.320	0.277	0.322	0.290	0.334	0.291	0.333	0.354	0.305	0.757	0.611	0.305	0.349	0.324	0.368	1.410	0.810

5.2.2 消融实验

1) 设置多尺度时序编码阶段的尺度数量 k 为 1, 2, 3, 回望窗口大小为 96, 预测 4 个不同长度的输出, 并在 ETTh1 和 ETTm2 数据集上进行实验, 结果如表 3 和表 4 所列。

表 3 ETTh1 上的实验结果

Table 3 Experimental results on ETTh1

尺度数	$k=1$		$k=2$		$k=3$	
	MSE	MAE	MSE	MAE	MSE	MAE
96	0.384	0.402	0.382	0.401	0.379	0.397
192	0.439	0.438	0.432	0.436	0.422	0.428
336	0.470	0.454	0.482	0.458	0.465	0.452
720	0.518	0.496	0.499	0.488	0.474	0.467

表 4 ETTm2 上的实验结果

Table 4 Experimental results on ETTm2

尺度数	$k=1$		$k=2$		$k=3$	
	MSE	MAE	MSE	MAE	MSE	MAE
96	0.178	0.262	0.180	0.261	0.177	0.259
192	0.250	0.310	0.242	0.305	0.241	0.302
336	0.308	0.344	0.309	0.348	0.317	0.353
720	0.410	0.406	0.409	0.406	0.407	0.403

随着编码尺度数的递增, 模型的预测精度呈现出显著的上升趋势。这种在不同时间尺度数据集上的改善, 有力地验证了本文的核心理论假设: 多尺度时间序列编码能够有效捕获时间序列在不同时间粒度上的特征模式, 从而提供更为全面和深入的时序信息表示。

2) 为了研究不同长度的回望窗口对模型预测性能的影响, 对 Electricity 和 Weather 两个数据集使用不同的历史回望窗口来进行预测。设置回望窗口大小 $L=24, 48, 96, 192$, 预测长度 $T=96, 720$ 。在 Electricity 数据集上的实验结果如图 3 所示。

在 Electricity 数据集上的实验结果表明, 当逐步增加历史回望窗口的大小时, 大多数基准模型的预测精度的提升并不显著, 呈现出明显的性能瓶颈。相比之下, 本文提出的 MTSformer 模型却展现出更为优越的性能提升趋势, 预测精度随着历史回望窗口的扩大而持续改善。这种显著的性能差异主要源于本文提出的多尺度时序编码模块的独特优势: 该模块能够在更长的历史回望窗口下有效捕获并整合不同时间粒度的序列特征, 从而充分利用扩展的历史信息。

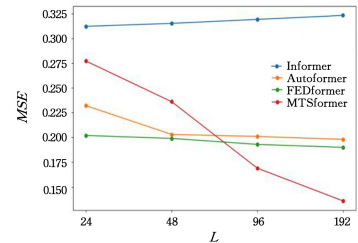
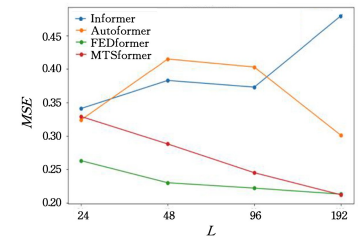
(a) $T=96$ (b) $T=720$

图 3 Electricity 数据集上的实验结果

Fig. 3 Experimental results on Electricity dataset

在 Weather 数据集上的实验结果如图 4 所示。

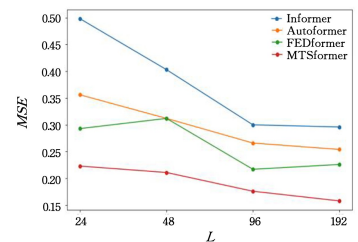
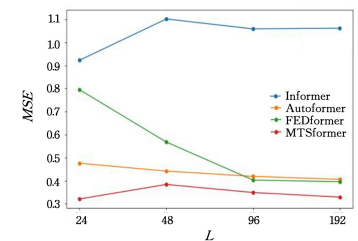
(a) $T=96$ (b) $T=720$

图 4 Weather 数据集上的实验结果

Fig. 4 Experimental results on Weather dataset

在 Weather 数据集上进行的长序列预测(预测窗口长度为 96)实验结果表明,输入序列长度(回望窗口)与模型性能呈现显著的正相关关系。具体而言,随着回望窗口从 24 逐步增加至 192,所有对比模型的预测精度均呈现上升趋势。值得注意的是,MTSformer 模型在不同回望窗口大小的设置下均优于基准模型,展现出了较强的预测能力和鲁棒性。

在超长序列预测任务(预测窗口长度为 720)的实验中,MTSformer 在所有预测长度配置下均保持了性能优势,相较于基准模型都有较大的性能提升。这主要源于本文提出的多尺度时序编码模块的独特优势:多尺度子序列保留了更加完整和全面的时序特征,使得模型能更好地捕获粗粒度和细粒度特征,在更长的历史回望窗口下有效捕获并整合不同时间粒度的序列特征,从而进一步提升模型的预测性能。

3)将多尺度时序编码模块添加到 4 个 Transformer 类模型中,包括 Transformer,Informer,Autoformer 和 Crossformer。

本文实验保留了 Transformer 类模型整体的编码器-解码器结构,并在小数据集 ETTh2 和大数据集 Weather 中分别进行实验,实验设定历史回望窗口大小为 96,分别预测 4 个输出长度(96,192,336 和 720)。

Transformer 实验结果如表 5 和表 6 所列,可以看到,将多尺度时序编码模块集成到原始 Transformer 中,模型在 MSE 和 MAE 上均实现了较大的性能提升,表明该模块能有效增强 Transformer 模型捕获原始序列中跨尺度时序特征的能力。

表 5 Transformer-ETTh2 实验结果

Table 5 Experimental results of Transformer-ETTh2

输出长度	Transformer		Transformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	2.466	1.267	1.755	1.095
192	5.042	1.823	3.345	1.534
336	5.585	1.954	3.007	1.374
720	2.861	1.349	2.803	1.457

表 6 Transformer-Weather 实验结果

Table 6 Experimental results of Transformer-Weather

输出长度	Transformer		Transformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.382	0.426	0.375	0.410
192	0.530	0.503	0.512	0.488
336	0.704	0.570	0.646	0.544
720	0.922	0.741	0.852	0.697

Informer 实验结果如表 7 和表 8 所列,可以看出,该模块同样显著提升了 Informer 的预测性能,验证了该模块与 Informer 模型的稀疏注意力机制的良好兼容性,可为其他变体注意力机制提供更好的支持。

表 7 Informer-ETTh2 实验结果

Table 7 Experimental results of Informer-ETTh2

输出长度	Informer		Informer ⁺⁺	
	MSE	MAE	MSE	MAE
96	1.549	0.952	1.185	0.845
192	3.792	1.542	3.211	1.181
336	4.215	1.642	3.048	1.311
720	3.656	1.619	3.389	1.157

表 8 Informer-Weather 实验结果

Table 8 Experimental results of Informer-Weather

输出长度	Informer		Informer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.300	0.384	0.278	0.396
192	0.598	0.544	0.563	0.512
336	0.578	0.523	0.548	0.497
720	1.059	0.741	0.962	0.683

Autoformer 实验结果如表 9 和表 10 所列,结果同样证实了该模块在 Autoformer 模型中的适用性,表明该模块能够与时间序列分解架构有效协同。

表 9 Autoformer-ETTh2 实验结果

Table 9 Experimental results of Autoformer-ETTh2

输出长度	Autoformer		Autoformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.332	0.368	0.317	0.348
192	0.426	0.434	0.411	0.428
336	0.477	0.479	0.452	0.447
720	0.453	0.490	0.433	0.497

表 10 Autoformer-Weather 实验结果

Table 10 Experimental results of Autoformer-Weather

输出长度	Transformer		Transformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.266	0.336	0.253	0.320
192	0.307	0.367	0.296	0.347
336	0.359	0.395	0.364	0.377
720	0.419	0.428	0.391	0.399

Crossformer 实验结果如表 11 和表 12 所列,进一步验证了该模块在 Crossformer 模型的跨变量注意力机制中的适用性,实验表明,该模块能够与跨变量注意力机制有效协同。

表 11 Crossformer-ETTh2 实验结果

Table 11 Experimental results of Crossformer-ETTh2

输出长度	Crossformer		Crossformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.745	0.584	0.656	0.520
192	0.877	0.656	0.796	0.647
336	1.043	0.731	0.864	0.677
720	1.104	0.763	0.991	0.699

表 12 Crossformer-Weather 实验结果

Table 12 Experimental results of Crossformer-Weather

输出长度	Crossformer		Crossformer ⁺⁺	
	MSE	MAE	MSE	MAE
96	0.195	0.271	0.189	0.268
192	0.209	0.277	0.219	0.265
336	0.273	0.332	0.275	0.320
720	0.379	0.401	0.368	0.390

上述实验结果表明,将多尺度时序编码模块添加到 4 个 Transformer 类的模型中,能够增强模型捕获输入序列的多尺度特征,从而进一步提升模型的预测性能。

结束语 本文提出了一种融合多尺度特征和注意力机制的时间序列预测模型 MTSformer,采用多尺度时序编码提高多头注意力机制捕获时间序列多尺度特征的能力,同时采用多预测头替代传统的复杂解码器提高模型的预测速度。该模型在 5 个基准数据集上进行了实验,并与 Pathformer, PatchTST, FEDformer, Crossformer, Autoformer, Timesnet

和 Dlinear 等主流模型进行了比较, MTSformer 模型在时间序列预测上的 MSE 平均降低了 24.51%, MAE 平均降低了 17.84%, 具有较佳的性能。

未来的工作主要有两个方面: 一是在多尺度编码中引入可学习的下采样因子, 动态调整尺度的粒度, 以适配不同场景下数据分布的时变特性; 二是面向边缘计算场景的轻量化部署需求, 研究基于神经架构搜索的模型压缩方法, 在保持多尺度特征交互能力的同时进一步降低计算能耗。

参 考 文 献

- [1] MAO Y H, SUN C C, XU L Y, et al. A Review of Time Series Forecasting Methods Based on Deep Learning [J]. *Microelectronics & Computer*, 2023, 40(4): 8-17.
- [2] CHENG W H, CHE W G. Research on Financial Time Series Forecasting Algorithm Based on Secondary Decomposition and LSTM [J]. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2022, 34(4): 638-645.
- [3] KAN G Y, YANG J. Research on Meteorological Time Series Forecasting Based on Wavelet Transform and LSTM Hybrid Model [J]. *Computer Science and Application*, 2022, 12(3): 682-689.
- [4] XU S, LIU D D. Power Load Forecasting Based on Time Series Combination Model [J]. *Electronic Design Engineering*, 2023, 31(23): 1-6.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017: 6000-6010.
- [6] ZAREMBA W, ILYA S, ORIOL V. Recurrent neural network regularization [J]. *arXiv*: 1409. 2329, 2014.
- [7] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. *arXiv*: 1412. 3555, 2014.
- [8] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [9] LEA C, FLYNN M D, VIDAL R, et al. Temporal convolutional networks for action segmentation and detection [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 156-165.
- [10] WU H, HU T, LIU Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis [C] // *The Eleventh International Conference on Learning Representations*. 2022.
- [11] 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: 11106-11115.
- [12] 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: 22419-22430.
- [13] KITAEV N, KAISER Ł, LEVSKAYA A. Reformer: The efficient transformer [J]. *arXiv*: 2001. 04451, 2020.
- [14] ZHOU T, MA Z, WEN Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting [C] // *International Conference on Machine Learning*. PMLR, 2022: 27268-27286.
- [15] NIE Y, NGUYEN N H, SINTHONG P, et al. A time series is worth 64 words: Long-term forecasting with transformers [J]. *arXiv*: 2211. 14730, 2022.
- [16] CHEN P, ZHANG Y, CHENG Y, et al. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting [J]. *arXiv*: 2402. 05956, 2024.
- [17] ZENG A L, CHEN M X, ZHANG L, et al. Are transformers effective for time series forecasting? [C] // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023.
- [18] ASUNCION A, NEWMAN D. UCI machine learning repository [EB/OL]. <https://github.com/uci-ml-repo>.
- [19] CHOLLET F. *Deep learning with Python* [M] // *Manning Publications*. 2021: 45-52.
- [20] ZHAO L, GKOUNTOUNA O, PFOSER D. Spatial auto-regressive dependency interpretable learning based on spatial topological constraints [J]. *ACM Transactions on Spatial Algorithms and Systems*, 2019, 5(3): 1-28.



PAN Jian, born in 1976, Ph.D, associate professor, master's supervisor, is a member of CCF (No. 26947M). His main research interests include natural language processing, intelligent information processing and Internet of Things.

(责任编辑: 何杨)