

基于分治策略和频域调制的轻量级多元长时序预测方法

张亚浩^{1,2,3},周晓锋^{1,2},张宜弛^{1,2},李 帅^{1,2},刘舒锐^{1,2,3}

¹(中国科学院 网络化控制系统重点实验室,沈阳 110016)

²(中国科学院 沈阳自动化研究所,沈阳 110016)

³(中国科学院大学,北京 100049)

E-mail:zhouxf@sia.cn

摘要:多元长时序预测的主流框架是编码器-解码器结构,然而关于解码器结构的研究略显不足.采用简单的全连接层作为解码器不仅导致模型参数数量过大,且忽略了多元时间序列的通道异质性和时序依赖性.针对上述问题,本文提出了一种基于分治策略和频域调制的轻量级多元长时序预测方法.首先,频域分段编码器将历史时间序列划分为时间片段并映射至频域,利用频域混合器提取时间片段内部特征和片段间相关性;然后,基于分治策略将长时序预测任务分解为各频率分量的子预测任务,并通过模型参数共享机制实现了解码器的轻量化设计;最后,使用通道调制头和时序调制头对初步预测结果进行频域特异性调制,弥补了解码过程在建模通道异质性和时序依赖性方面的不足.五个时序数据集上的实验结果表明该方法的预测精度优于现有的八个最新基线方法,同时展现出优异的轻量化特性.

关键词:多元长时序预测;解码器;分治策略;通道异质性;时序依赖性

中图分类号: TP183

文献标识码: A

文章编号: 1000-1220(2026)02-0265-09

Lightweight Multivariate Time Series Forecasting Method Based on Divide-and-conquer Strategy and Frequency Domain Modulation

ZHANG Yahao^{1,2,3}, ZHOU Xiaofeng^{1,2}, ZHANG Yichi^{1,2}, LI Shuai^{1,2}, LIU Shurui^{1,2,3}

¹(Key Laboratory of Networked Control Systems, Chinese Academy of Sciences, Shenyang 110016, China)

²(Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China)

³(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The encoder-decoder architecture serves as the predominant framework for multivariate long-term time series forecasting. However, research on decoder structures remains relatively underexplored. The use of a simple fully connected layer as the decoder not only leads to a large number of model parameters but also neglects the channel heterogeneity and temporal dependencies inherent in multivariate time series. To address these issues, this paper proposes a lightweight multivariate long-term time series forecasting method based on divide-and-conquer strategy and frequency-domain modulation. First, the frequency-domain segmentation encoder divides the historical time series into segments and maps them to the frequency domain, and employs a frequency-domain mixer to extract intra-segment features and inter-segment correlations. Then, the divide-and-conquer strategy decomposes the long-term forecasting task into sub-forecasting tasks for each frequency component, achieving a lightweight decoder design through parameter sharing. Finally, a channel modulation head and a temporal modulation head are introduced to refine the preliminary forecasting results in the frequency domain, compensating for the decoder's shortcomings in modeling channel heterogeneity and temporal dependency. Experiments on five time series datasets demonstrate that the proposed method achieves superior forecasting accuracy compared to existing eight baseline models while exhibiting outstanding lightweight characteristics.

Keywords: multivariate long-term time series forecasting; decoder; divide-and-conquer; channel heterogeneity; temporal dependency

0 引言

随着传感器技术和数据采集系统的不断发展,越来越多的多元时间序列数据被生成并存储,这些数据涵盖了多个变量随时间变化的复杂动态特征.多元长时间序列预测旨在基于这些历史数据揭示未来较长时间段内各变量的变化趋势.

该技术广泛应用于气象预测^[1]、金融市场分析^[2]、边缘负载预测^[3]、交通流量预测^[4]等领域.准确的时间序列预测结果不仅有助于提升决策质量,还能为资源优化、风险预警等方面提供有力支持.因此,开发一种高效、精确的多元长时序预测技术,对于各行业在不确定的环境中做出科学且及时的决策具有重要意义.

收稿日期:2025-01-13 收修改稿日期:2025-02-27 基金项目:辽宁省应川基础研究计划项目(2022JH2/101300255)资助;中国科学院沈阳自动化研究所基础研究计划项目(2022000346)资助. 作者简介:张亚浩,男,1998年生,硕士研究生,研究方向为长时间序列预测、多元时间序列预测;周晓锋(通信作者),女,1978年生,博士,研究员,研究方向为机器学习与大数据分析;张宜弛,男,1986年生,硕士,副研究员,研究方向为工业大数据分析;李 帅,男,1988年生,博士,副研究员,研究方向为大数据智能分析与挖掘;刘舒锐,男,1993年生,博士研究生,研究方向为数据挖掘与长时间序列预测.

时间序列数据通常具有复杂的非线性关系^[5]、周期性波动^[6]以及高维输入特征^[7],研究者们通常使用复杂的模型提取深度时序特征,如RNN^[8,9]、CNN^[10,11]、Transformer^[12,13]和线性模型^[14-16].这些多元长时序预测模型通常基于图1所示的编码器-解码器框架^[12,17-21]:输入数据首先通过编码器提取深层时序特征,随后由解码器生成预测结果.尽管这些研究在一定程度上提升了多元长时序预测模型的性能,但大多数工作主要集中在编码器的改进,而对解码器的优化关注较少.目前主流方法通常直接使用如图2所示的简单全连接层作为解码器^[8,12,15,19,22].然而,这种解码方式存在3个显著问题:1)在历史窗口和预测窗口长度较大时,解码器的参数量会显著增加,在资源受限的场景(如边缘设备或嵌入式系统)下难以实际部署;2)解码器采用统一的线性层处理所有通道,未充分考虑不同通道之间的异质性.时间序列数据中不同通道往往具有不同的统计特性和动态模式,这种统一处理方式无法有效捕捉这些差异,可能导致关键信息的丢失;3)非自回归类解码器忽略了时间步之间的模式关联,这可能导致生成的预测序列缺乏一致性,难以准确捕捉时间序列的顺序性和动态变化.

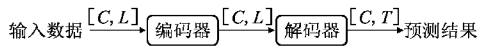


图1 多元长时序预测模型基本组成框架

Fig. 1 Basic structure of multivariate long-term time series forecasting models

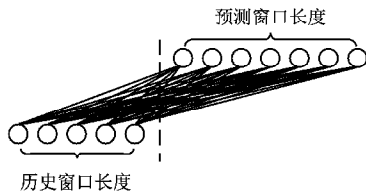


图2 常用多元长时序预测解码器

Fig. 2 Commonly used decoder for long-term multivariate time series forecasting

针对上述问题,本文提出了一种基于分治策略和频域调制的多元长时序预测网络(Divide-and-Conquer and Frequency Modulation Network, DCFMNet). DCFMNet 首先通过片段划分模块将时间序列分割为固定长度片段,并利用快速傅里叶变换将其映射至频域提取深层频域信息;然后基于分治策略将长时序预测任务分解为各频率分量的子预测任务,并通过对每个频率分量上共享相同的模型参数,显著减少了解码器的参数量.此外,为弥补解码阶段对通道特异性信息和时序依赖关系建模的不足,本文引入频域通道调制头和频域时序调制头对通道和时间片进行特异性频域调制,进一步提升了模型的预测性能.

本文的主要贡献总结如下:

1) 提出了一个轻量级多元长时间序列预测模型,该模型通过将时间序列转换到频域进行分析,并设计了3个核心模块——频域分段编码器、频域分治解码器和频域调制头.在极少计算资源的条件下,有效提升了多元长时间序列预测的性能.

2) 提出了一种基于分治策略的时序预测解码器,将时序预测任务分解成对时间片段中不同频率分量的子预测任务,并通过共享模型参数的方式,大幅减少了解码器的参数量.

3) 提出了两个对预测结果进行信息修正的频域调制头.通过对编码器-解码器架构生成的初步预测结果在频域内进行乘性和加性信号调制,增强了模型在捕获多元时间序列通道异质性和时序依赖性的能力.

4) 在多个真实的时间序列数据集上的实验结果表明,本文提出的方法在使用极小参数量的同时能够有效提升多元长时间序列预测的准确性,优于现有基线方法.

1 相关工作

由于时间序列预测任务的重要意义,国内外学者展开了丰富的研究.统计模型是时间序列预测研究的早期重要工具,例如,ARIMA模型^[23]通过差分将非平稳序列平稳化,结合自回归与移动平均建模当前值与历史值及误差的关系;Holt-Winters模型^[24]通过分解时间序列中的水平、趋势和季节性因素捕捉时间序列的全局趋势和周期模式,动态更新每个时间步的参数并生成未来的预测值.然而,这些统计类模型基于固定的先验规律假设,在处理具有复杂模式的长时间序列预测任务时效果不尽人意,且不适用于多变量序列.

随着计算能力的增强和数据量的增加,深度学习模型成为时间序列预测的热点.自回归的模型如RNN^[25]等直观上是建模时间依赖关系的主要方法,然而,RNN类模型存在梯度消失和爆炸的问题,且在长时间序列预测中存在误差累积效应.近年来,Transformer^[26]凭借自注意力机制捕获长程依赖的能力,受到了广泛关注.然而,其解码器推理过程的迭代生成机制同样会导致误差累积现象.为解决这一问题,Zhou^[27]等人通过引入引导向量和占位符,采用单次前向过程生成序列输出,缓解了误差累积现象;随后,Wu^[20]等人和Zhou^[21]等人在一步生成策略的基础上,进一步优化了自注意力机制,分别通过自相关机制和频域增强技术提升了预测精度,但其解码阶段参数量仍然庞大,推理速度较慢.为降低长时序预测的时间复杂度,Nic^[12]等人将时间序列分割成多个小片段,改善了单个时间点语义信息不足的问题,同时以全连接层代替自注意力机制作为解码器,在降低了计算复杂度的同时提升了性能.随后,大多数研究均采用全连接层作为解码器^[8,12,15,19,22],然而全连接层解码器在多元长时序预测中仍面临参数量较大、通道区分性不足以及时序依赖性差等问题.

针对使用全连接层作为多元长时序预测解码器存在的缺点,研究者们提出了多种改进方法.例如,针对解码器参数量大的问题,Xu^[28]等人利用频域信息紧凑的特点,通过低通滤波器在频域中预测时间序列的低频成分,大幅降低了参数量,然而该方法的预测效果仍然存在改进空间.为解决通道区分性不足的问题,Zeng^[14]等人提出了一种方法,为每个通道单独设计全连接层以生成预测结果.尽管该方法在一定程度上提高了通道间的区分能力,但在变量数量较多时,参数量显著增加,同时未能充分利用变量间的共享特性,从而对预测精度造成一定影响.针对时序依赖性建模问题,一些模型^[27,29]通过位置编码嵌入方式引导解码器生成具有时序季节信息的预

测结果,容易导致原有特征向量语义的损失;Xiong^[30]等人通过改进损失函数,将预测结果的差分序列与实际值的差分序列对比,以捕获时序信息损失,但并未从根本上解决预测结果生成过程中时序依赖性不足的问题。

综上所述,现有多元长时序预测解码器仍存在参数量大、通道区分性不足、时序依赖性差的缺点,在实际应用时存在一定的局限性。针对这些挑战,本文提出了一种基于分治策略和频域调制的轻量级多元长时序预测方法。该方法首先将时间序列片段转换到频域,使用多个复数全连接层作为编码器提取深层频率特征;随后利用分治策略,将长时序预测问题分解为时间序列片段的各频率分量的子预测问题,实现轻量化预测。最终,利用频域通道调制头和频域时序调制头对结果进行针对性调制,避免了解码阶段通道信息和时序信息的缺失。

2 DCFMNet 模型设计

本文提出的 DCFMNet 的整体框架如图 3 所示。DCFNet 由频域分段编码器、频域分治解码器和频域调制头 3 个模块组成。具体来说,输入数据首先通过如图 4 所示的编码器-解码器结构生成初步预测结果。该结构采用通道独立策略,有助于提高模型的鲁棒性,帮助模型捕获各变量之间的共



图 3 DCFMNet 模型整体结构

Fig. 3 Overall structure of DCFMNet

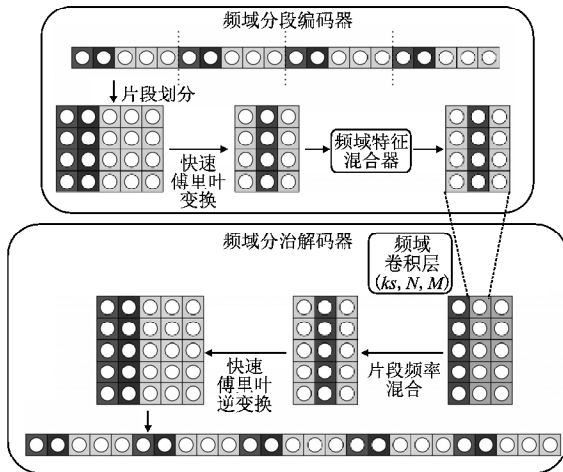


图 4 DCFMNet 的编码器与解码器结构

Fig. 4 Structure of the encoder and decoder of DCFMNet

有趋势特征。在此基础上,频域调制头进一步对初步预测结果进行细化处理,通过在通道和时序上进行特异性调制,使预测结果更具针对性和准确性。接下来各节将对问题定义及各个模块详细介绍。

2.1 问题定义

多元时间序列预测旨在根据过去多个变量的观测数据预测未来时刻的多个变量的值。具体来说,给定一个时间序列 X

$= [x_1, x_2, \dots, x_C] \in \mathbb{R}^{C \times L}$, 其中 $x_i \in \mathbb{R}^L$ 表示第 i 个变量对应的历史时间序列, L 表示历史时间序列的长度, C 表示变量的个数。多元时间序列预测的目标就是寻找一个模型 f , 将输入序列映射到未来 T 个时间步的序列 $Y = [y_1, y_2, \dots, y_C] \in \mathbb{R}^{C \times T}$, 其中 $y_i \in \mathbb{R}^T$ 表示第 i 个变量对应的预测序列。长时序预测往往涉及到更大的预测步数, 即 $T \gg 1$ 。

2.2 频域分段编码器

如图 4 所示,频域分段编码器通过片段划分模块将历史时间序列划分为长度一致的片段,随后利用傅里叶变换将时域信号转换到频域,并通过频域特征混合器提取片段内部及片段间的深层特征。

2.2.1 片段划分模块

在时间序列中,单个数据点仅表示某一时刻的观测值,其语义信息往往有限,难以体现时间序列的上下文特征。而通过片段划分,将单个数据点与其邻近的点划分到同一片段中,模型能够更好地捕捉短时间范围内的趋势和波动模式,从而丰富单个数据点的语义表达。此外,使用基于周期性的片段划分方法可以帮助模型学习时间序列的周期模式和全局趋势信息。

给定一个片段长度 W , 片段划分模块将具有 C 个变量、长度为 L 的时间序列 $X \in \mathbb{R}^{C \times L}$ 划分成 N 个长度为 W 的片段, 表示为 $X_w \in \mathbb{R}^{C \times N \times W}$, 其中 $N = \frac{L}{W}$:

$$X_w = \text{Patching}(X)_{\text{patch_len}=W} \mathbb{R}^{C \times L} \rightarrow \mathbb{R}^{C \times N \times W} \quad (1)$$

由于频域可以直观地揭示信号中蕴含的周期性和趋势信息,更适合用于时间序列分析。因此本文利用快速傅里叶变换将每个时间片段从时域映射到频域。具体地,对于每个时间片段 $X_w^{i,j}$:

$$X_F^{i,j} = \mathcal{F}(X_w^{i,j}) \mathbb{R}^W \rightarrow \mathbb{C}^{\frac{W}{2}+1} \quad (2)$$

其中, $X_w^{i,j}$ 和 $X_F^{i,j}$ 分别表示第 i 个变量的第 j 个时间序列片段的时域和频域表示向量, $\mathcal{F}(\cdot)$ 表示快速傅里叶变换操作。

2.2.2 频域特征混合器

如图 5 所示,为了深入挖掘历史序列的频域信息,本文在频域特征混合器通过 3 个步骤捕获频域的深层特征:特征增强层、片段间信息提取层和特征降维层。

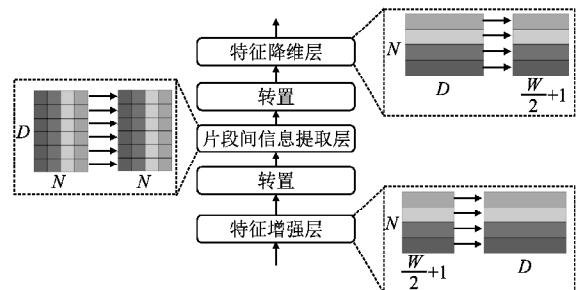


图 5 频域特征混合器结构

Fig. 5 Structure of the frequency domain feature mixer

首先,特征增强层使用复数全连接层将数据从原始的维度 $\mathbb{C}^{C \times N \times (\frac{W}{2}+1)}$ 投影到更高的维度 $\mathbb{C}^{C \times N \times D}$, 以获得丰富的频域特征。接着,考虑到时间序列通常具有显著的周期特征,不同片段之间存在相似性和相互影响,本文将数据矩阵转置并

通过权重矩阵维度为 $\mathbb{C}^{N \times N}$ 的复数全连接层在每个频率分量上整合所有片段的信息以捕获全局特性. 这里对每个频率分量使用相同的权重, 帮助模型学习频谱中各频率分量的普遍模式. 随后再次对数据矩阵进行转置, 最后特征降维层再次利用复数全连接层将特征维度从 $\mathbb{C}^{C \times N \times D}$ 映射为 $\mathbb{C}^{C \times N \times (\frac{W}{2}+1)}$. 具体过程如下:

$$X_1 = FLinear(X_F) \mathbb{C}^{C \times N \times (\frac{W}{2}+1)} \rightarrow \mathbb{C}^{C \times N \times D} \quad (3)$$

$$X_2 = Transpose(FLinear(Transpose(X_1))) \mathbb{C}^{C \times N \times D} \rightarrow \mathbb{C}^{C \times N \times D} \quad (4)$$

$$X_{enc_out} = FLinear(X_2) \mathbb{C}^{C \times N \times D} \rightarrow \mathbb{C}^{C \times N \times (\frac{W}{2}+1)} \quad (5)$$

其中 $FLinear(\cdot)$ 表示复数全连接层, $Transpose(\cdot)$ 表示张量转置操作, $X_{enc_out} \in \mathbb{C}^{C \times N \times (\frac{W}{2}+1)}$ 为编码器的输出结果.

2.3 频域分治解码器

由于时间序列固有的周期性特征, 时间片段通常表现出相似的性质, 且时间序列的趋势变化也与周期密切相关. 因此, 与传统解码器通过全连接层将整个历史时间序列直接映射到未来时间窗口的方式不同, 本方法提出了一种基于时间片段的预测策略, 即通过历史时间片段组预测未来时间片段组. 同时, 由于时间片段中的各频率分量往往具有相似的变化趋势, 本文将整个预测任务划分为各频率分量的子预测任务, 这些子任务共享同一模型及参数, 从而有效降低模型的参数量.

图4展示了频域分治解码器的工作流程. 该解码器对每个频率分量使用相同的预测模型; 完成对各频率分量的子预测任务后, 将未来所有时间片段的各频率分量预测结果整合, 恢复完整的预测输出. 具体来说, 对于预测长度为 T 的任务, 本文将未来序列划分为与编码器片段划分长度 W 相同的时间片段, 因此未来的时间片段的个数为 $M = \frac{T}{W}$. 对于第 k 个频率分量, 本文根据过去 N 个时间窗口预测未来 M 个时间窗口的该频率分量. 此外, 由于从模拟信号采样生成离散信号的过程中部分频谱信息可能会丢失, 该解码器在预测第 k 个频率分量时同时考虑第 $k-p$ 到 $k+p$ (p 默认为1)个频率分量以弥补离散傅里叶变换的频谱稀疏问题. 为保证维度的正确性, 本文首先在频率维度对边界值进行复制填充, 然后使用一维复数卷积层映射到未来时间片段. 该方式可以保证每个频率分量预测过程使用相同的结构和参数, 进而减少模型的参数量:

$$X_{conv} = FConv1d(Padding(X_{enc_out}, dim=3, num=p), in_channels=N, out_channels=M, dim=2, kernel_size=2p+1) \mathbb{C}^{C \times N \times (\frac{W}{2}+1)} \rightarrow \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \quad (6)$$

其中 $Padding(\cdot)$ 表示边界值填充操作, $FConv1d(\cdot)$ 表示复数一维卷积层, 历史数据的片段个数为 N , 预测数据的片段个数为 M . 随后, 在片段内部使用全连接层, 将各个频率分量的预测结果进一步整合:

$$X_{out} = FLinear(X_{conv}) \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \rightarrow \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \quad (7)$$

最后, 本文对输出结果进行逆傅里叶变换并重塑形状, 得到初步预测结果, 如式(8)和式(9)所示, 其中 $\mathcal{F}^{-1}(\cdot)$ 表示快速傅里叶逆变换操作, $Reshape(\cdot)$ 表示张量重塑操作:

$$X_{dec_out} = \mathcal{F}^{-1}(X_{out}) \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \rightarrow \mathbb{R}^{C \times M \times W} \quad (8)$$

$$Y_{first} = Reshape(X_{dec_out}) \mathbb{R}^{C \times M \times W} \rightarrow \mathbb{R}^{C \times T} \quad (9)$$

通过这一分治策略, 即使用共享参数的网络对历史时间序列和未来时间序列分片段分频率预测, 随后将结果汇总生成最终预测输出, 该频域分治解码器相比于全连接层解码器将参数的空间复杂度从 $O(L \cdot T)$ 降至 $O(\frac{(L+T) \cdot T}{W^2})$.

2.4 频域调制头

传统解码器在预测过程中通常通过全连接层将特征信息从历史窗口直接映射到未来窗口. 这种方式既未对每个变量进行差异化处理, 也忽略了时序信息. 为此, 本文再次引入分治策略: 在编码器和解码器中, 本文采用通道独立的策略, 并未注入任何时序编码信息; 而是在生成初步预测结果后, 再对各变量和时间片段的频域特征进行针对性的信号调制. 为实现这一目标, 本文设计了两种频域调制头.

2.4.1 频域通道调制头

如图6所示, 在使用编码器-解码器得到初步预测结果之后, 首先将预测结果进行频域转换:

$$Y_{first_F} = \mathcal{F}(Y_{first}) \mathbb{R}^{C \times T} \rightarrow \mathbb{C}^{C \times (\frac{T}{2}+1)} \quad (10)$$

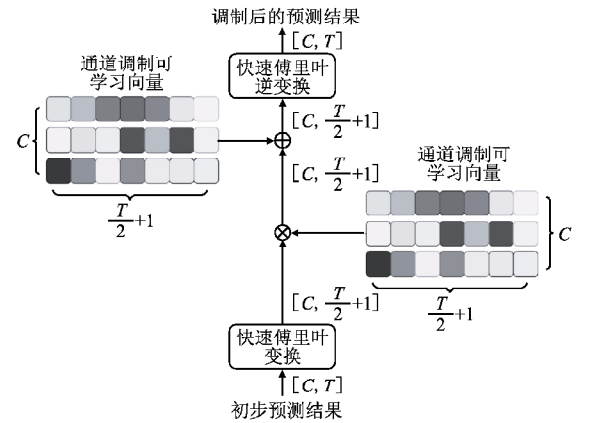


图6 频域通道调制头

Fig. 6 Frequency domain channel modulation head

接下来, 本文设计了一种频域调制机制, 分别从乘性和加性两个角度对预测结果进行细粒度调整. 首先, 引入一组频域调制向量 $CM = [cm_1, cm_2, \dots, cm_C] \in \mathbb{C}^{C \times (\frac{T}{2}+1)}$, 其中 $cm_i \in \mathbb{C}^{\frac{T}{2}+1}$ 表示与第 i 个变量对应的频域调制向量, 通过哈达马积 \odot 将该向量与第 i 个变量的初步预测结果逐点相乘, 实现对每个变量的频域乘性调制. 在完成乘性调制后, 本文进一步引入另一组频域向量 $CA = [ca_1, ca_2, \dots, ca_C] \in \mathbb{C}^{C \times (\frac{T}{2}+1)}$, 用于对各个通道进行加性调制. 整个频域调制过程如下:

$$Y_{F1} = Y_{first_F} \odot CM \quad (11)$$

$$Y_{F2} = Y_{F1} + CA \quad (12)$$

最后, 通过快速傅里叶逆变换将调制后的预测结果转换回时域, 得到最终的预测结果:

$$Y = \mathcal{F}^{-1}(Y_{F2}) \mathbb{C}^{C \times (\frac{T}{2}+1)} \rightarrow \mathbb{R}^{C \times T} \quad (13)$$

2.4.2 频域时序调制头

对于某一个时刻 t , 预测未来 T 个时间点的值时, 不同时间点距离当前时刻 t 的远近会导致其预测需求存在差异, 同

时每个时间步还应包含其前后点的时序依赖特征,为此本文引入频域时序调制头。如图 7 所示,本文采用分片策略,将初步预测结果分割为时间片形式 $Y_p \in \mathbb{R}^{C \times M \times W}$,以便对不同时间片进行分区调整。其中 $M = \frac{T}{W}$ 表示未来时间片段的个数, W 表示时间片的长度。接着,通过傅里叶变换转换到频域以方便进行调制:

$$Y_p = \text{Patching}(Y_{\text{first}})_{\text{patch_len}=W} \quad (14)$$

$$Y_{P,F} = \mathcal{F}(Y_p) \mathbb{R}^{C \times M \times W} \rightarrow \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \quad (15)$$

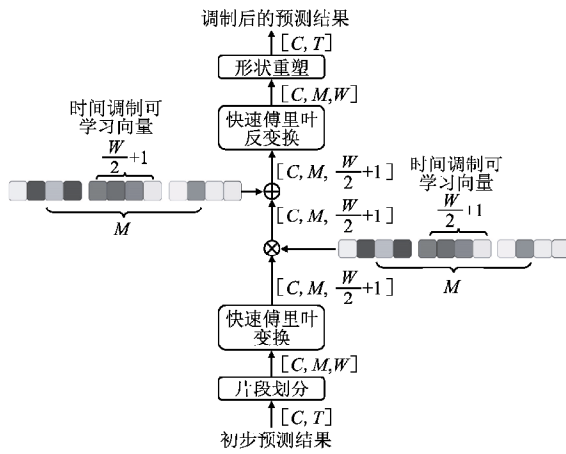


图 7 频域时序调制头

Fig. 7 Frequency domain temporal modulation head

接着,本文引入了一种时间维度上的调制机制,以实现针对不同时间片的细粒度调整。具体而言,通过定义一组时间调制向量 $TM = [tm_1, tm_2, \dots, tm_M] \in \mathbb{C}^{M \times (\frac{W}{2}+1)}$,其中 $tm_i \in \mathbb{C}^{\frac{W}{2}+1}$ 表示与第 i 个时间片段对应的频域调制向量,该向量与初步预测结果进行逐点相乘;除了乘性调制,本文进一步引入了一组时间加性调制向量 $TA = [ta_1, ta_2, \dots, ta_M] \in \mathbb{C}^{M \times (\frac{W}{2}+1)}$,用于对每个时间片的预测结果进行频域偏移调整:

$$Y_{F1} = Y_{P,F} \odot TM \quad (16)$$

$$Y_{F2} = Y_{F1} + TA \quad (17)$$

最后,通过傅里叶逆变换将调制后的时间片结果还原到时域,并进行片段拼接,得到最终预测结果:

$$Y_{F2,T} = \mathcal{F}^{-1}(Y_{F2}) \mathbb{C}^{C \times M \times (\frac{W}{2}+1)} \rightarrow \mathbb{R}^{C \times M \times W} \quad (18)$$

$$Y = \text{Reshape}(Y_{F2,T}) \mathbb{R}^{C \times M \times W} \rightarrow \mathbb{R}^{C \times T} \quad (19)$$

2.5 可逆归一化

时间序列通常具有非平稳性,其均值和方差等统计特性会随时间的推移而变化。为减少不同输入序列之间分布差异对模型性能的影响, Kim 等人^[31]提出了可逆归一化(RevIN)用于去除非平稳信息。具体而言,对于每个输入序列,先将该序列标准化为均值为 0、标准差为 1 的时间序列,并在网络的输出结果上将原始序列的均值和标准差添加回预测序列中。本文中,除了主干网络的处理流程外,还在输入和输出序列上引入了 RevIN 的平稳化和反平稳化操作,以提升模型的泛化性。

2.6 算法流程

综合上述讨论,本文提出了一种基于分治策略和频域调

制的多元长时序预测方法(DCFMNet),该方法包括频域分段编码器、频域分治解码器以及频域通道与时序调制头等关键模块,其算法关键步骤如下所示。

算法. DCFMNet 模型概览

输入: 历史时间序列 $X \in \mathbb{R}^{C \times L}$

输出: 预测时间序列 $Y \in \mathbb{R}^{C \times T}$

初始化: 输入序列长度 L , 预测长度 T , 变量个数 C , 时间片段长度 W , 隐含层维度 D , 卷积核半径 p

1. 历史时间序列平稳化: $X = \text{RevIN}(X, \text{norm})$ $X \in \mathbb{R}^{C \times L}$

2. 频域分段编码器:

3. 将时间序列划分为片段并转换到频域:

$$X_F = \mathcal{F}(\text{Patching}(X)) \quad X_F \in \mathbb{C}^{C \times N \times (\frac{W}{2}+1)}$$

4. 使用频域特征混合器捕获片段内和片段间频域特征:

$$X_{\text{enc_out}} = \text{Mixer}(X_F) \quad X_{\text{enc_out}} \in \mathbb{C}^{C \times N \times (\frac{W}{2}+1)}$$

5. 频域分治解码器:

6. 利用轻量化解码器生成预测序列片段的频域表示:

$$X_{\text{out}} = \text{FLinear}(\text{FConv1d}(X_{\text{enc_out}})) \quad X_{\text{out}} \in \mathbb{C}^{C \times M \times (\frac{W}{2}+1)}$$

7. 将预测序列片段转换到时域并重塑形状生成预测结果:

$$Y_{\text{first}} = \text{Reshape}(\mathcal{F}^{-1}(X_{\text{out}})) \quad Y_{\text{first}} \in \mathbb{C}^{C \times T}$$

8. 频域调制头:

9. 使用频域通道调制头注入通道特异性信息:

$$Y = \text{ChannelModulation}(Y_{\text{first}}) \quad Y \in \mathbb{C}^{C \times T}$$

10. 使用频域时序调制头注入时序特异性信息:

$$Y_{\text{first}} = Y \quad Y_{\text{first}} \in \mathbb{C}^{C \times T}$$

$$Y = \text{TemporalModulation}(Y_{\text{first}}) \quad Y \in \mathbb{C}^{C \times T}$$

11. 预测序列反平稳化: $Y = \text{RevIN}(Y, \text{denorm})$ $Y \in \mathbb{C}^{C \times T}$

12. Return Y

3 实验

3.1 数据集介绍

为了验证 DCFMNet 在真实数据集上的有效性,本文在实验中选用了 5 个常用的时间序列预测数据集: ETTh1、ETTh2、ETTm1、ETTm2 和 Weather。以下是对各数据集的具体描述: ETT(Electricity Transformer Temperature)数据集记录了电力变压器系统的多变量运行数据,涵盖了负荷、温度、湿度等 7 个变量。ETT 数据集分为两个时间分辨率版本:其中 ETTh1 和 ETTh2 的采样频率为 1 小时,均包含 17420 个时间

表 1 时间序列数据集属性概览

Table 1 Overview of the properties of time series datasets

数据集	ETTh1	ETTh2	ETTm1	ETTm2	Weather
变量数	7	7	7	7	21
时间点数	17420	17420	69680	69680	52696
采样频率	1 小时	1 小时	15 分钟	15 分钟	1 小时

步的数据,两个数据集来源于不同的变压器系统;ETTm1 和 ETTm2 分别对应 ETTh1 和 ETTh2,但采样频率提高到 15 分钟,数据规模扩展至 69680 个时间步。此外, Weather 数据集来源于德国的天气观测站,包含气温、湿度、风速、降雨量等 21 个气象变量,数据以每小时为采样频率记录,共包含 52696 个时间步。表 1 展示了各数据集的变量数、时间步数及采样频率的详细信息。

3.2 评价指标

为了对模型进行有效的验证,在实验中,本文采用时间序列预测任务中常用的误差衡量标准:均方误差(Mean Squared Error, MSE)和平均绝对误差(Mean Absolute Error, MAE),作为评价模型预测性能的主要指标. MSE和MAE的定义如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (21)$$

其中, y_i 和 \hat{y}_i 分别表示第 i 个点实际值和预测值, n 表示时间序列的长度.

3.3 实验设置

为了全面评估 DCFMNet 的性能,本文选择了多个具有代表性的多元长时间序列预测模型进行对比,包括: PatchTST^[12]、FITS^[28]、DLinear^[14]、iTransformer^[19]、MICN^[11]、TimesNet^[5]、FEDformer^[21]、Autoformer^[20]. 实验中,历史时间序列长度设置为 720,待预测序列长度包括 96、192、336、720 共 4 个场景. 时间序列片段划分长度 W 设置为 48,中间层的维度 D 为 128,卷积半径范围 p 设置为 1,即卷积核的大小为 3. 训练过程中,采用 Adam 优化器进行参数更新,并使用平均绝对误差(MAE)作为损失函数. 批次大小设置为 32,模型使

用 Pytorch 框架实现并在 NVIDIA GeForce RTX 3090 24GB GPU 上训练.

3.4 实验结果

表 2 展示了提出的方法与其它基线模型的性能对比,最优和次优结果分别用粗体和下划线标注. 从整体表现来看, DCFMNet 在 5 个数据集中的不同预测步长上均展现出稳定且优异的预测精度. 在大多数实验中, DCFMNet 在 MSE 和 MAE 两个指标上实现了最低误差. 具体而言,相比于其它基线模型, DCFMNet 在 MSE 和 MAE 上平均降低了 16.6% 和 12.3%, 在 40 个评估指标中获得了 37 个最优结果,充分体现了其出色的预测性能. 相比之下,其他模型在解码器部分的处理存在明显不足. 例如, PatchTST、FITS、DLinear、iTransformer 和 MICN 均直接使用简单的时域或频域线性层进行预测,未充分考虑时间序列的时间依赖性和局部周期性. 此外,这些模型的解码器采用共享参数的方式,未能有效区分多元变量之间的分布差异. TimesNet 虽然在值嵌入后引入了二维卷积层以提取周期性特征,但其值嵌入过程将时间序列长度映射为 $L+T$,这一过程数据丢失了原始的周期信息,并进一步增大了模型的参数量. FEDformer 和 Autoformer 在解码器中使用历史序列的实际值和位置编码作为初始输入,但由于未能提供足够的上下文信息,交叉注意力机制在建模历史序列与未

表 2 不同方法的预测结果性能对比

Table 2 Forecasting performance of the proposed model and baseline models

方法	DCFNet		PatchTST		FITS		DLinear		iTransformer		MICN		TimesNet		FEDformer		Autoformer		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	0.346	0.378	0.370	0.400	0.374	0.395	0.375	0.399	0.387	0.405	0.421	0.431	0.384	0.402	0.376	0.419	0.449	0.459
	192	0.387	0.404	0.413	0.429	0.407	<u>0.414</u>	<u>0.405</u>	0.416	0.441	0.436	0.474	0.487	0.436	0.429	0.420	0.448	0.500	0.482
	336	0.403	0.419	<u>0.422</u>	0.440	0.429	<u>0.428</u>	0.439	0.443	0.488	0.459	0.569	0.551	0.491	0.469	0.459	0.465	0.521	0.496
	720	0.413	0.439	0.447	0.468	0.425	<u>0.446</u>	0.472	0.490	0.501	0.488	0.770	0.672	0.521	0.500	0.506	0.507	0.514	0.512
ETTh2	96	0.270	0.328	0.274	<u>0.337</u>	<u>0.271</u>	<u>0.337</u>	0.289	0.353	0.300	0.350	0.299	0.364	0.340	0.374	0.358	0.397	0.346	0.388
	192	0.329	0.369	0.341	0.382	<u>0.332</u>	<u>0.375</u>	0.383	0.418	0.380	0.399	0.441	0.454	0.402	0.414	0.429	0.439	0.456	0.452
	336	0.367	0.400	0.329	0.384	<u>0.354</u>	<u>0.396</u>	0.448	0.465	0.424	0.432	0.654	0.567	0.452	0.452	0.496	0.487	0.482	0.486
	720	0.373	0.417	0.379	<u>0.422</u>	<u>0.378</u>	0.423	0.605	0.551	0.429	0.447	0.956	0.716	0.462	0.468	0.463	0.474	0.515	0.511
ETTm1	96	0.287	0.328	<u>0.293</u>	0.346	0.303	0.345	0.299	<u>0.343</u>	0.341	0.376	0.316	0.362	0.338	0.375	0.379	0.419	0.505	0.475
	192	0.330	0.358	<u>0.333</u>	0.370	0.337	<u>0.365</u>	0.335	<u>0.365</u>	0.380	0.394	0.363	0.390	0.374	0.387	0.426	0.441	0.553	0.496
	336	0.359	0.377	0.369	0.392	<u>0.366</u>	<u>0.385</u>	0.369	0.386	0.418	0.418	0.408	0.426	0.410	0.411	0.445	0.459	0.621	0.537
	720	0.414	0.408	0.416	0.420	0.416	<u>0.412</u>	0.425	0.421	0.487	0.457	0.481	0.476	0.478	0.450	0.543	0.490	0.671	0.561
ETTm2	96	0.159	0.245	0.166	0.256	<u>0.163</u>	<u>0.253</u>	0.167	0.260	0.185	0.271	0.179	0.275	0.187	0.267	0.203	0.287	0.255	0.339
	192	0.215	0.284	0.223	0.296	<u>0.217</u>	<u>0.291</u>	0.224	0.303	0.251	0.312	0.307	0.376	0.249	0.309	0.269	0.328	0.281	0.340
	336	0.268	0.322	<u>0.274</u>	0.329	0.268	<u>0.326</u>	0.281	0.342	0.316	0.351	0.325	0.388	0.321	0.351	0.325	0.366	0.339	0.372
	720	0.352	0.375	0.362	0.385	0.349	<u>0.378</u>	0.397	0.421	0.409	0.405	0.502	0.490	0.408	0.403	0.421	0.415	0.433	0.432
Weather	96	0.147	0.193	<u>0.149</u>	<u>0.198</u>	0.170	0.224	0.176	0.237	0.174	0.214	0.161	0.229	0.172	0.220	0.217	0.296	0.266	0.336
	192	0.188	0.233	<u>0.194</u>	<u>0.241</u>	0.213	0.260	0.220	0.282	0.227	0.259	0.220	0.281	0.219	0.261	0.276	0.336	0.307	0.367
	336	0.239	0.271	0.245	0.282	0.258	0.295	0.265	0.319	0.280	0.297	0.278	0.331	0.280	0.306	0.339	0.380	0.359	0.395
	720	0.308	0.321	<u>0.314</u>	<u>0.334</u>	0.320	0.339	0.323	0.362	0.359	0.350	0.311	0.356	0.365	0.359	0.403	0.428	0.419	0.428
最佳个数	37		2		2		0		0		0		0		0		0		

来序列关系时缺乏明确的指导. 相比之下, DCFMNet 在编码器中充分捕捉了时间序列周期片段间的关系和周期内的局部变化. 在解码器中, DCFMNet 通过分片段预测的方式, 充分利用时间序列的周期特性, 同时还利用频域调制头兼顾通道异质性和时间依赖关系, 取得了显著优于其他模型的效果.

图 8 展示 DCFMNet 与性能较好的基线模型 (PatchTST、FITS 和 DLinear) 在 ETTh1 数据集上的预测曲线对比, 预测步长设为 96. 为了便于观察, 这里仅展示了一个通道的预测结果. 可以看出, 所有模型在总体趋势上均能较好地跟踪真实值的周期性变化, 但在关键的峰值和谷值区域存在明显差异. 其

中,DCFMNet 在多个峰值片段的预测中表现出与真实值更为

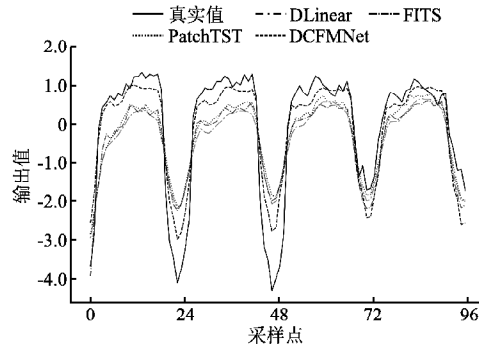


图 8 不同模型在 ETTh1 数据集上单通道预测曲线对比

Fig. 8 Comparison of signal-channel forecasting curves of different models on the ETTh1 dataset

接近的趋势,而其他基线模型在这些区域出现了较大的偏离;此外,对于真实曲线谷值附近的显著波动,DCFMNet 能够更准确地捕捉其幅度变化,而其他模型则有所偏差.总体而言,DCFMNet 展现了其在长时序数据建模中的优越性,能够提供更高的预测精度和更小的波动误差.

3.5 消融实验

为了验证 DCFMNet 提出的各模块的有效性,本文设计了多组消融实验,分别评估频域分治解码器、频域通道调制头和频域时序调制头对模型性能的影响.

3.5.1 频域分治解码器

为了验证频域分治解码器的有效性,本文将频域分治解码器替换成全连接层,该变体使用 DCFMNet \dagger 表示.具体的,通过编码器提取时间序列的高层特征后,将各片段变换回时域并拼接成完整的序列,随后利用全连接层直接从历史时间窗口映射为未来序列,实现解码过程.具体结果如表 3 所示.

表 3 解码器消融实验结果

Table 3 Ablation study results of decoders

方法	DCFMNet		DCFMNet \dagger		
	MSE	MAE	MSE	MAE	
ETTh1	96	0.346	0.378	0.349	0.379
	192	0.387	0.404	0.387	0.405
	336	0.403	0.419	0.404	0.427
	720	0.413	0.439	0.416	0.442
ETTh2	96	0.270	0.328	0.269	0.328
	192	0.329	0.369	0.329	0.369
	336	0.368	0.400	0.370	0.401
	720	0.373	0.417	0.377	0.418

结果表明,频域分治解码器在预测效果上优于全连接层,这是因为频域分治解码器通过分片段的预测策略,有效挖掘不同时间片段间的规律,更充分地利用了时间序列的周期.

3.5.2 频域调制头

为评估两种频域调制头的有效性,本小节设计了 3 个模型变体进行:1) DCFMNet-CT:移除所有调制头;2) DCFMNet-C:仅保留时序调制头;3) DCFMNet-T:仅保留通道调制头.

表 4 展示了频域调制头消融实验的结果.结果表明,两种调制头对模型预测性能均有提升.其中,对于 Weather 数据

集,通道调制头的作用尤为显著,这是因为该数据集上通道异质性强,需要对每个变量区分性处理,而编码器-解码器的通道独立策略没有对通道进行有效区分.此外,当两种调制头配合使用时,模型能够达到最佳效果,充分验证了它们在建模通道异质性和时间依赖关系上的有效性.

表 4 频域调制头消融实验

Table 4 Ablation study results of frequency domain modulation heads

方法	DCFMNet		DCFMNet-CT		DCFMNet-C		DCFMNet-T		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	0.346	0.378	0.356	0.380	0.355	0.380	0.348	0.380
	192	0.387	0.404	0.397	0.406	0.397	0.406	0.388	0.405
	336	0.403	0.419	0.410	0.421	0.411	0.423	0.407	0.424
	720	0.413	0.439	0.418	0.439	0.418	0.439	0.416	0.441
ETTh2	96	0.270	0.328	0.279	0.331	0.279	0.331	0.270	0.329
	192	0.329	0.369	0.356	0.379	0.356	0.380	0.330	0.370
	336	0.367	0.400	0.387	0.408	0.391	0.410	0.370	0.402
	720	0.373	0.417	0.396	0.429	0.390	0.427	0.379	0.418
Weather	96	0.147	0.193	0.170	0.216	0.170	0.212	0.147	0.195
	192	0.188	0.233	0.212	0.254	0.212	0.250	0.188	0.233
	336	0.239	0.271	0.259	0.291	0.258	0.287	0.239	0.271
	720	0.308	0.321	0.322	0.337	0.321	0.335	0.307	0.323

为进一步验证所提出的频域调制头的有效性,本文将频域调制头集成到 DLinear 和 iTransformer 模型中,生成改进模

表 5 频域调制头与其它基线结合性能结果

Table 5 Performance results of the frequency domain modulation head combined with other baselines

方法	DLinear*		DLinear		iTransformer*		iTransformer		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	0.295	0.341	0.299	0.343	0.334	0.371	0.341	0.376
	192	0.331	0.363	0.335	0.365	0.375	0.389	0.380	0.394
	336	0.368	0.384	0.369	0.386	0.413	0.414	0.418	0.418
	720	0.424	0.417	0.425	0.421	0.482	0.452	0.487	0.457
ETTh2	96	0.162	0.231	0.176	0.237	0.172	0.212	0.174	0.214
	192	0.206	0.274	0.220	0.282	0.221	0.256	0.227	0.259
	336	0.256	0.313	0.265	0.319	0.278	0.296	0.280	0.297
	720	0.322	0.363	0.323	0.362	0.353	0.347	0.359	0.350

型,记为 DLinear* 和 iTransformer*. 实验结果如表 5 所示,结果表明,加入频域调制头后,两个模型预测精度均得到提升,进一步验证了所提模块的普适性和有效性.

3.6 模型参数量对比

除了展现出卓越的性能,DCFMNet 还具备极其轻量化的特性.表 6 展示了在 ETTh2 数据集上预测窗口长度为 720 时各模型在参数量与 MACs (Multiply-Accumulate Operations, 乘法加法运算次数) 两方面的对比结果,其中 DCFMNet \dagger 代表将 DCFMNet 的频域分治解码器替换成全连接层的变体.具体来说,DCFMNet 的参数量仅为 13.94k,远低于其他模型;其 MACs 也仅为 1.09M,在轻量化方面表现突出.相比之下,DCFMNet \dagger 的参数量和 MACs 分别增至 531.72k 和 4.54M,可见频域分治解码器采用的分治策略和参数共享机制可以有效降低解码器参数开销,使得模型能够更高效地适配计算资

源受限的设备环境.

表6 模型参数数量和 MACs 对比实验结果

Table 6 Experimental results of parameters and MACs for models

模型	参数量	MACs
MICN	12.34M	2.67G
FEDformer	10.54M	3.5G
Autoformer	1.47M	504.88M
TimesNet	666k	2.4G
PatchTST	4.28M	155.11G
DLinear	485.28k	3.4M
iTransformer	304.72k	3.34M
FITS	77.22k	540.57k
DCFNet†	531.72k	4.54M
DCFNet	13.94k	1.09M

DCFNet 的参数量仅为以轻量化著称的 FITS 的 18.1%。尽管 FITS 在 MACs 上表现更小,这是因为其为了减少模型参数量直接舍弃了编码器,但这一设计导致其预测性能较差.相比之下,DCFNet 不仅在参数量上进一步缩减,还展现出更为优异的表现.

3.7 权重可视化分析

为了更清晰地展示频域分治解码器如何利用频域特征进行预测,本文进行了针对解码器预测过程的权重可视化实验,实验采用 ETTh2 数据集训练模型,历史时间长度和预测长度均设置为 336,片段长度为 48,卷积核大小为 3.

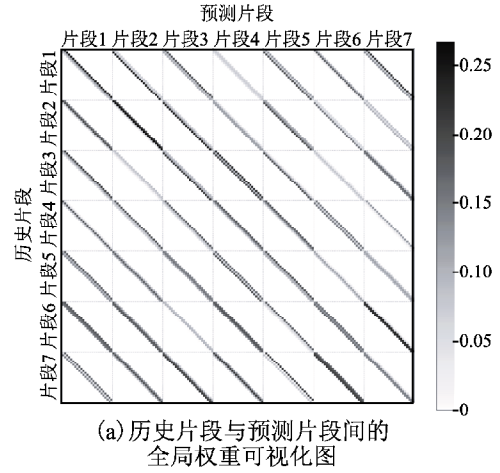
实验结果如图 9 所示.图 9 (a) 展示了预测过程中所有历史时间片段和预测时间片段之间的权重分布,图 9 (b) 进一步展示了其中一对历史片段和预测片段之间的权重分布.图中呈现明显的对角化稀疏分布模式,每个预测片段的频率成分仅依赖于该频率及其相邻频率的历史变化,这种稀疏化的权重分布模式减少了模型对不相关频率成分的过多关注,并有效降低了解码器的参数量;此外,通过应用一维卷积层,本文确保了解码器在预测任意频率时,对其相邻频率的权重分配与预测其他频率时相同,实现了一致的预测模式.这种共享参数的设计有助于模型捕获所有频率分量的共同变化趋势,同时也进一步减少了解码器的参数量.

4 结束语

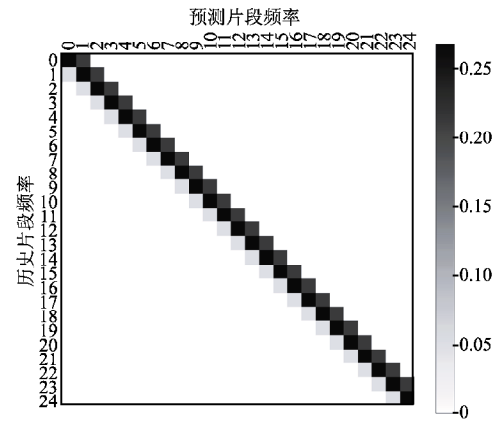
为提升多元长时序预测模型的准确性与轻量化特性,本文提出了一种基于分治策略和频域信号调制的方法 DCFNet.该方法针对多元长时序预测任务,对解码器进行了优化改进.通过将时间序列预测任务分解为多个时间片段的各频率分量的了预测任务并共享模型参数,DCFNet 显著降低了解码器的参数量;同时,通过引入频域调制头对初步预测结果进行特异性调制,为预测结果注入了各通道和时间片段的特有信息.实验结果表明,DCFNet 在多个数据集上表现出优异的预测精度,且在参数量上远低于其它基线模型.此外,将频域调制头集成到其他模型中也展现了良好的效果,进一步验证了该模块的普适性.

尽管该研究在多元长时序预测方面取得了一定的进展,模型仍存在一些待改进之处.未来的工作将考虑在编码器

中引入通道间动态关系的建模机制,以更清晰地捕捉变量间



(a) 历史片段与预测片段间的全局权重可视化图



(b) 特定历史片段与预测片段间的权重可视化图

图9 解码器权重可视化

Fig.9 Visualization of the decoder's weights

的依赖关系;同时还可以利用频域分析实现片段划分长度的自动选择.

References:

- [1] Zhang W, Huang J, Wang R, et al. Integration of mamba and transformer-MAT for long-short range time series forecasting with application to weather dynamics [EB/OL]. <http://arxiv.org/abs/2409.08530>, 2024.
- [2] Preeti, Bala R, Singh R P. A dual-stage advanced deep learning algorithm for long-term and long-sequence prediction for multivariate financial time series [J]. Applied Soft Computing, 2022, 126: 109317, doi:10.1016/j.asoc.2022.109317.
- [3] JIANG Q N, XU H R, CHEN Z Y, et al. Edge load prediction with multi-variable spatio-temporal inverted transformer [J/OL]. Journal of Chinese Computer Systems, 1-11, <http://kns.cnki.net/kcms/detail/21.1106.tp.20240729.1943.006.html>, 2024.
- [4] ZHANG A Q, QIN T. Traffic flow prediction model based on multi-channel spatial-temporal encoder [J]. Application Research of Computers, 2024, 41 (1) : 83-87 + 93.
- [5] Wu H, Hu T, Liu Y, et al. TimesNet: temporal 2D-variation modeling for general time series analysis [C] // Proceedings of the 11th

- International Conference on Learning Representations, 2023; 1-23.
- [6] Dai T, Wu B, Liu P, et al. Periodicity decoupling framework for long-term series forecasting [C] // Proceedings of the 12th International Conference on Learning Representations, 2024; 1-12.
- [7] Zhang T, Zhang Y, Cao W, et al. Less is more: fast multivariate time series forecasting with light sampling-oriented mlp structures [EB/OL]. <http://arxiv.org/abs/2207.01186>, 2022.
- [8] Lv C, Zhou X, Hu G, et al. Multiscale temporal-variable patches network for long-term prediction of industrial key parameters [J]. IEEE Sensors Journal, 2024, 24(22): 37237-37248.
- [9] Hou H, Yu F R. RWKV-TS: beyond traditional recurrent neural network for time series tasks [EB/OL]. <https://arxiv.org/abs/2401.09093v1>, 2024.
- [10] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. <http://arxiv.org/abs/1803.01271>, 2018.
- [11] Wang H, Peng J, Huang F, et al. MICN: multi-scale local and global context modeling for long-term series forecasting [C] // Proceedings of the 11th International Conference on Learning Representations, 2023; 1-22.
- [12] Nie Y, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: long-term forecasting with transformers [C] // Proceedings of the 11th International Conference on Learning Representations, 2023; 1-24.
- [13] Shen L, Wei Y, Wang Y. GBT: two-stage transformer framework for non-stationary time series forecasting [J]. Neural Networks, 2023, 165: 953-970, doi: 10.1016/j.neunet.2023.06.044.
- [14] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting? [C] // AAAI Conference on Artificial Intelligence, 2023; 11121-11128.
- [15] Chen S A, Li C L, Yoder N, et al. TSMixer: an all-mlp architecture for time series forecasting [EB/OL]. <http://arxiv.org/abs/2303.06053>, 2023.
- [16] Zhao L G, Che C, Zhao T M. Long-term time series forecasting with channel fusion and sequence stabilization [J/OL]. Journal of Chinese Computer Systems, 2024; 1-8, <http://kns.cnki.net/kcms/detail/21.1106.TP.20240301.1421.010.html>, 2024.
- [17] Das A, Kong W, Leach A, et al. Long-term forecasting with TiDE: time-series dense encoder [EB/OL]. <http://arxiv.org/abs/2304.08424>, 2023.
- [18] Shao Z, Wang F, Zhang Z, et al. HUTFormer: hierarchical u-net transformer for long-term traffic forecasting [EB/OL]. <http://arxiv.org/abs/2307.14596>, 2023.
- [19] Liu Y, Hu T, Zhang H, et al. iTransformer: inverted transformers are effective for time series forecasting [C] // Proceedings of the 12th International Conference on Learning Representations, 2024; 1-25.
- [20] Wu H, Xu J, Wang J, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting [C] // Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021; 22419-22430.
- [21] Zhou T, Ma Z, Wen Q, et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting [C] // Proceedings of the 39th International Conference on Machine Learning, 2022; 27268-27286.
- [22] Wang S, Wu H, Shi X, et al. TimeMixer: decomposable multiscale mixing for time series forecasting [C] // Proceedings of the 12th International Conference on Learning Representations, 2024; 1-27.
- [23] Nkongolo M. Using ARIMA to predict the growth in the subscriber data usage [J]. Eng, 2023, 4(1): 92-120.
- [24] Gardner Jr E S. Exponential smoothing: the state of the art [J]. Journal of Forecasting, 1985, 4(1): 1-28.
- [25] Lipton Z C, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning [EB/OL]. <http://arxiv.org/abs/1506.00019>, 2015.
- [26] 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.
- [27] 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.
- [28] Xu Z, Zeng A, Xu Q. FITS: modeling time series with 10k parameters [C] // Proceedings of the 12th International Conference on Learning Representations, 2024; 1-24.
- [29] Zeng P, Hu G, Zhou X, et al. Seformer: a long sequence time-series forecasting model based on binary position encoding and information transfer regularization [J]. Applied Intelligence, 2023, 53(12): 15747-15771.
- [30] Xiong Q, Tang K, Ma M, et al. TDT loss takes it all: integrating temporal dependencies among targets into non-autoregressive time series forecasting [EB/OL]. <http://arxiv.org/abs/2406.04777>, 2024.
- [31] Kim T, Kim J, Tae Y, et al. Reversible instance normalization for accurate time-series forecasting against distribution shift [C] // Proceedings of the 10th International Conference on Learning Representations, 2022; 1-25.

附中文参考文献:

- [3] 江庆南, 许浩然, 陈哲毅, 等. 基于多变量时空反转 Transformer 的边缘负载预测方法 [J/OL]. 小型微型计算机系统, 2024; 1-11, <http://kns.cnki.net/kcms/detail/21.1106.tp.20240729.1943.006.html>, 2024.
- [4] 张安勤, 秦 添. 基于多通道时空编码器的交通流量预测模型 [J]. 计算机应用研究, 2024, 41(1): 83-87 + 93.
- [16] 赵龙港, 车 超, 赵天明. 使用通道融合和序列平稳化策略的长期时间序列预测方法 [J/OL]. 小型微型计算机系统, 2024; 1-8, <http://kns.cnki.net/kcms/detail/21.1106.TP.20240301.1421.010.html>, 2024.