

DOI:10.19651/j.cnki.emt.2518423

基于双重时间依赖学习的多元长时间序列预测*

衡红军 周晨晓

(中国民航大学计算机科学与技术学院 天津 300300)

摘要: 考虑到时间序列数据因其复杂的长短期模式和多周期特征,为预测带来了独特的挑战。传统的固定尺度分片方法难以有效捕捉多周期信息,同时周期性和趋势性变化进一步增加了建模难度,影响了预测精度和可解释性。基于上述问题,本文提出了基于双重时间依赖学习的多周期模型 MDTDNet。该模型首先通过傅里叶变换自适应获取多周期信息;然后对于每个周期,结合季节趋势增强模块,通过周期分片设计、频域季节性增强和时域趋势性增强,提升了子序列的语义表达;引入双重时间依赖模块,通过长期变化提取器和局部波动提取器,分别捕捉分片间和分片内的不同时间依赖模式,实现特征提取和融合;实验结果表明,在六个数据集上的模型实验效果均超越了当前的最优模型 PatchTST,在 ETTh1 数据集上均方误差(MSE)平均下降了9%,最高下降了10.14%。

关键词: 时间序列预测;多周期;季节趋势增强;双重时间依赖

中图分类号: TP391;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Multivariate long-term series forecasting based on dual time-dependent learning

Heng Hongjun Zhou Chenxiao

(College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Time series data pose unique challenges for forecasting due to their complex long- and short-term patterns and multi-period characteristics. Traditional fixed-scale patching methods are difficult to effectively capture multi-period information, while periodicity and trend changes further increase the modeling difficulty, affecting forecast accuracy and interpretability. Based on the above problems, this paper proposes a multi-periodic model MDTDNet based on dual time-dependent learning. The model firstly acquires multi-period information adaptively by Fourier transform; then for each period, combined with the seasonal trend enhancement module, it improves the semantic expression of the subsequence through period patch design, frequency domain seasonal enhancement and time domain trend enhancement. A dual time-dependence module is introduced to realize feature extraction and fusion by capturing different time-dependent patterns inter- and intra-patches by means of a long-term change extractor and a local fluctuation extractor, respectively. The experimental results show that the experimental results of the models in all six datasets outperform the current optimal model, PatchTST, with an average decrease of 9% in the mean square error (MSE) on ETTh1 dataset, with a maximum decrease of 10.14%.

Keywords: time series forecasting; multi-period; seasonal trend enhancement; dual time-dependence

0 引言

多元时间序列的长期预测在各个现实行业中发挥着重要作用,例如交通状态估计^[1]、天气预报^[2]、能源消耗预测^[3]等。然而随着数据量的增加和复杂性的提升,传统的时间序列分析方法^[4-5]的局限性逐渐显现,深度学习的出现为处理复杂的非线性时间序列数据提供了新的解决途

径^[6],从而提出了一系列性能卓越的预测模型^[7-8]。然而,多元时间序列数据的低语义特征以及内部错综复杂的周期关系^[9]为长时间序列预测分析带来了很大挑战,如何有效建模它们成为了研究的重点内容。

首先,多元时间序列数据本质上具有低语义特征,因此提取局部语义信息对于分析它们之间的关联至关重要。许多工作^[10]采用序列分片方法,通过将时间序列以固定窗口

收稿日期:2025-03-24

* 基金项目:国家自然科学基金民航联合研究基金(U2333204)项目资助

大小切割为多个连续子序列片段,从而降低输入数据维度并减少冗余特征。这种方法通过关注多个时间步所展示的时间变化,模拟子序列间的时间依赖关系,进而更好地捕捉时间序列的内在规律,有效增强局部特征并减少模型输入长度。PatchTST^[11]模型采用分片思想,并使用注意力机制建模时间片之间的关系,提高了时间序列预测的精度。李岚皓等^[12]提出的 ScaleNN 模型对分片后的序列使用移动平均方法进行压缩,使其可以处理更长的历史数据并获取更具解释性的趋势信息,随后进行全局和局部的特征提取。Pathformer^[13]采用多尺度划分策略,使用不同大小的块将时间序列划分为不同的时间分辨率。基于每个尺度的划分,对这些补丁执行双重关注,以捕获全局和局部时间依赖关系。然而,上述方法使用根据先验知识所确定的固定长度的分片大小,这样很容易丢失时间的边界信息,例如完整的峰值和周期信息。所以在选择分片长度时,确保分片长度与时间序列的周期特征相匹配是很重要的。

其次,这种固定窗口的方法忽略了现实世界的时间序列数据常常呈现多周期的变化。例如,天气观测数据同时以日和周为不同周期变化,因此需要从多周期的角度进行时间依赖学习。MTPNet^[14]使用基于分片的多尺度金字塔网络建模序列的时间依赖关系,但在尺度选择上存在限制,可能无法捕捉所有时间依赖性。Scaleformer^[15]通过用不同的采样率将时间序列从低到高进行采样,迭代地在越来越精细的尺度上优化预测的时间序列。MICN^[16]采用多尺度分支结构来建模不同的潜在模式,并通过下采样卷积和等距卷积分别提取局部特征和全局相关性。而这些模型的尺度选择通常为不同的固定窗口大小,因此这些模型所使用的多尺度划分方式忽略了不同数据集的多周期时间特性,缺乏泛化和自适应能力。

最后,周期性和趋势性是时间序列的重要特征^[17],它们捕获了潜在的时间模式,并作为驱动数据随时间变化的关键因素。如何有效地模拟长期和短期时间依赖性并同时捕获周期性仍然是一个挑战^[18]。现有的基于季节趋势分解的模型,像 Autoformer^[19]和 DLinear^[20]使用移动平均方法将时间序列分解为趋势项和季节项,并通过对其分别建模获取到了更加全面的时间依赖关系。徐仲武^[21]提出的 MPDF 模型使用频谱图的主导频率确定潜在的多周期信息,并以此作为滑动窗口大小将时间序列分解为季节和趋势项,然后分别建模提取时间特征。曾渝等^[22]提出的多尺度门控膨胀卷积网络使用不同大小的均值滤波器将序列分解,并分别预测周期项和趋势项。FEDformer^[23]使用混合专家分解块实现季节趋势分解,它使用一组不同大小的平均滤波器从输入信号中提取多个趋势分量并加权组合为最终趋势。然而,这些方法忽视了序列在分解后的时间信息会被削弱的问题。因其固有的技术局限性,可能会在分解过程中引入不准确性的时间特征,导致分解后的趋势和季节信息不够充分,甚至可能会减弱这些成分的有效性。而

在此基础上对季节趋势直接进行建模可能会面临难以精确捕捉季节性和趋势性特征的问题,从而影响预测结果的准确性和可靠性。

为了解决上述问题,本文提出了基于双重时间依赖学习的多周期模型 MDTDNet(Multi-Periodic Dual Temporal Dependency Network,MDTDNet),对具有复杂时间模式的多元时间序列进行长期预测。本文的主要工作如下:

1)针对时间序列低语义特征和多周期问题,模型基于傅里叶变换自适应获取多周期信息,提出了季节趋势增强模块。包含特有的周期分片设计、频域上的季节性增强机制以及时域上的趋势性增强机制,用以对时间序列进行子序列划分、以及增强子序列语义信息。

2)针对时间序列长短期依赖协同学习问题,提出了双重时间依赖模块,包括长期时间变化提取器和局部波动提取器,用以对分片后的时间序列产生出的两种时间依赖模式进行特征提取和特征融合。

3)在 6 个时间序列数据集的实验结果表明,本文模型的预测精度全面超越了目前主流的最优模型。

1 模型结构

本文提出的时序预测模型 MDTDNet 结构如图 1 所示。模型首先对时间序列进行归一化处理,然后通过周期提取模块对归一化后的输入序列进行多周期获取,同时将归一化后的序列进行全局上的季节趋势分解,对得到的季节项和趋势项进行多周期建模。在每一个周期中,首先在季节趋势增强模块中对时间序列进行周期分片和子序列语义增强操作,然后在双重时间依赖模块中提取分片的时间依赖关系,得到该周期下的特征表示。最后融合多周期的结果得到最终的预测结果。

1.1 问题定义

多元时间序列预测任务是指对于输入序列 $\mathbf{X} = [x_1^1, x_1^2, \dots, x_1^d]^T, \mathbf{X} \in \mathbb{R}^{d \times l}$, 预测出长度为 L 的未来序列,其中 l 表示历史序列的长度, d 为多元时间序列的变量个数。

本文使用通道独立的方式进行多元时间序列的预测,使用单变量时间序列进行描述,通过独立考虑单变量,很容易推广到多变量。将输入序列 \mathbf{X} 分成 d 个独立的通道序列 $\mathbf{X}' = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^d\}$, 对于给定时间序列 $\mathbf{X}^i = [x_1, x_2, \dots, x_l]$, 即 $\mathbf{X}^i \in \mathbb{R}^l$, 预测长度为 L 的未来值 $\mathbf{X}_{pre}^i = [x_{l+1}, x_{l+2}, \dots, x_{l+L}]$ 。

1.2 序列分解

为了更准确地捕捉时间序列中的复杂时间模式,在对时间序列进行归一化处理之后,采用移动平均法对其进行分解,从而提取出趋势项和季节项。如式(1)和(2)所示。

$$\mathbf{X}_t = \text{AvgPool}(\text{Padding}(\mathbf{X})) \quad (1)$$

$$\mathbf{X}_s = \mathbf{X} - \mathbf{X}_t \quad (2)$$

其中, $\mathbf{X}_t, \mathbf{X}_s \in \mathbb{R}^{d \times l}$ 分别表示序列分解后得到的季节部分和趋势部分。采用 $\text{AvgPool}(\cdot)$ 进行移动平均,并使用

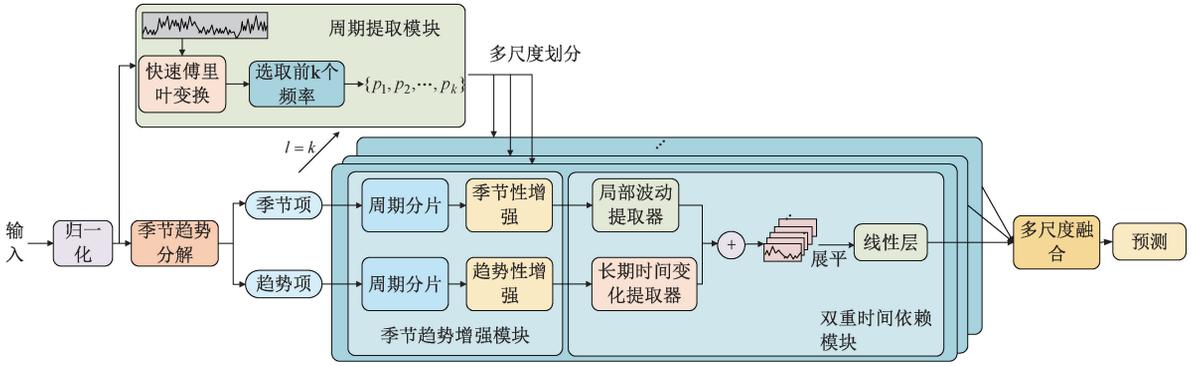


图 1 MDTDNet 模型结构

Fig. 1 MDTDNet model structure

填充操作 $\text{Padding}(\cdot)$ 来保持序列长度不变。

1.3 周期提取模块

为了获取时间序列中的多周期信息,模型通过傅里叶变换检测出关键周期,并将其作为划分尺度的依据。对于给定了 d 个变量的长度为 l 的时间序列 $\mathbf{X} \in \mathbb{R}^{d \times l}$,通过快速傅里叶变换在频域中分析时间序列,如式(3)所示。

$$\begin{cases} \mathbf{A} = \text{Avg}(\text{Amp}(\text{FFT}(\mathbf{X}))) \\ \{f_1, f_2, \dots, f_k\} = \text{argTopk}(\mathbf{A}) \\ p_i = \lfloor \frac{l}{f_i} \rfloor, i \in \{1, \dots, k\} \end{cases} \quad (3)$$

这里 $\text{FFT}(\cdot)$ (fast Fourier transform)和 $\text{Amp}(\cdot)$ 表示快速傅里叶变换和幅值的计算, \mathbf{A} 表示每个频率的幅度,它是从 d 维上通过 $\text{Avg}(\cdot)$ 求平均得到的。考虑到频域的稀疏性,本文只选择前 k 个最显著的幅度值(k 的个数由 2.4 节中实验确定)并获得对应的 k 个频率 $\{f_1, f_2, \dots, f_k\}$,这些选定的频率对应了相应的周期长度 $\{p_1, p_2, \dots, p_k\}$ 。

模型采用 k 的数量作为多周期分层数量划分的依据。使用选定的 k 个频率所对应的周期长度作为后续每一周期分层的周期分片大小。

模型将获得的周期数量 k 作为尺度数量划分也就是层数划分的依据,即 $l = k$ 。分别建模相对应周期的信息,每一层都主要关注该尺度下的时间依赖。 $\mathbf{X}_s^{i,l}, \mathbf{X}_t^{i,l}$ 记为由序列分解后得到的第 i 个通道的第 l 个尺度的季节项和趋势项。

1.4 季节趋势增强模块

季节和趋势性是时间序列的重要组成部分,经 1.2 节序列分解之后可以获得一个全局的季节项和趋势项。为了更好地捕捉时间序列中的复杂季节性和趋势性特征,本文设计了季节趋势增强模块,如图 2 所示。该模块首先对时间序列进行周期分片,然后通过对分片后的时间序列进行季节性和趋势性的增强处理,进一步提高数据特征表示,提升模型的预测能力。

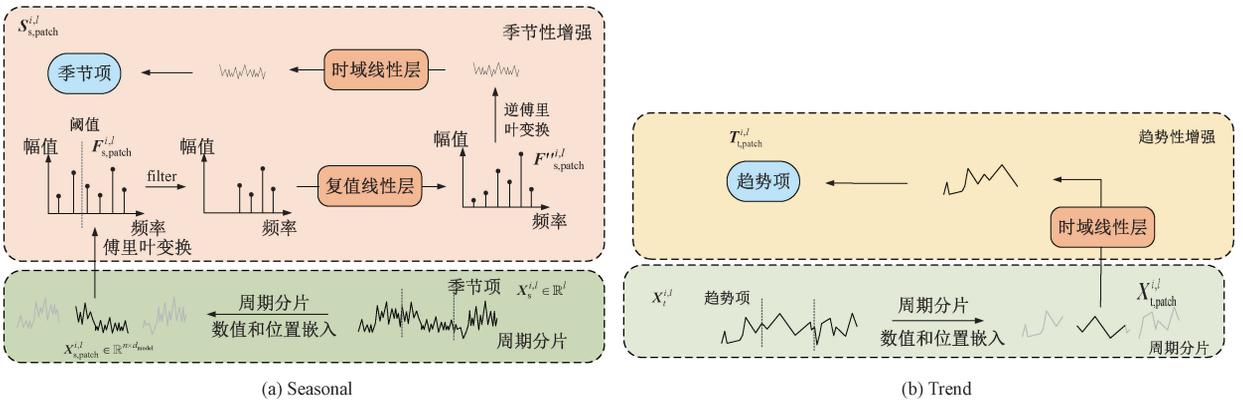


图 2 季节趋势增强模块

Fig. 2 Seasonal trend enhancement module

1) 周期分片

先前使用分片技术的工作将分片大小作为一个给定的参数,但是这样会降低模型的可解释性,并且会产生信息损失。本文模型使用周期提取模块得到的周期 p_i 作为

分片大小,建模不同周期的特有模式。首先,模型对季节项和趋势项进行周期分片,即根据时间序列的周期 p_i 将其分割成多个固定长度的分片。这一步骤能够使模型在每个分片内专注于局部周期性和趋势性特征,从而更精确地

捕捉每个周期中的重要模式。给定独立通道的历史序列 $\mathbf{X}_s^{i,l}, \mathbf{X}_s^{i,l} \in \mathbb{R}^l$, 切割为重叠或不重叠的分片 $\mathbf{X}_{s,\text{patch}}^{i,l}, \mathbf{X}_{t,\text{patch}}^{i,l} \in \mathbb{R}^{n \times p_i}$, 其中 n (分片个数) 由分片长度 p_i 和步长 s (两个连续分片之间的非重叠区域) 确定, 即 $n = \lfloor (l - p_i) / s \rfloor + 2$ 。

随后每个分片通过参数共享的可学习线性层映射到 d_{model} 维向量进行数值嵌入。同时为了减少位置信息的丢失, 每个分片会嵌入一个位置向量。构建位置编码的具体过程如式(4)和(5)所示, 其中 pos 表示当前位置的索引。

$$PE_{(pos, 2j)} = \sin\left(\frac{pos}{10000^{2j/d_{\text{model}}}}\right) \quad (4)$$

$$PE_{(pos, 2j+1)} = \cos\left(\frac{pos}{10000^{2j/d_{\text{model}}}}\right) \quad (5)$$

采用正弦 $\sin(\cdot)$ 和余弦 $\cos(\cdot)$ 函数生成一系列固定大小的向量, 这些向量与数值嵌入相加后, 共同构成序列中每个位置的输入表示 $\mathbf{X}_{s,\text{patch}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 和 $\mathbf{X}_{t,\text{patch}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 。

然而, 分片操作可能会丢失一些长期依赖关系, 因此, 后续的增强处理尤为重要。

2) 季节性增强

季节项通常具有较为复杂的周期性特征, 而时间序列在频域上往往能够以更加清晰的形式展示出其周期性规律。基于这一特点, 模型设计将分片后的季节项转换到频域, 以充分挖掘其潜在的周期信息。具体来说, 首先通过傅里叶变换将季节项转化为频域表示, 提取出其对应的频率分布, 如式(6)所示。

$$\mathbf{F}_{s,\text{patch}}^{i,l} = \text{FFT}(\mathbf{X}_{s,\text{patch}}^{i,l}) \quad (6)$$

其中, $\mathbf{F}_{s,\text{patch}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}/2+1}$ 是第 l 个尺度中的季节项分片经由傅里叶变换后得到的频域季节项分片。随后通过滤波器, 在频域中筛选出中高频成分, 从而强化重要的周期性模式, 如式(7)和(8)所示。

$$H(\theta) = \begin{cases} 1, & \theta_1 \leq \theta \leq \theta_2 \\ 0, & \text{其他} \end{cases} \quad (7)$$

$$\mathbf{F}'_{s,\text{patch}}^{i,l} = H(\theta) \cdot \mathbf{F}_{s,\text{patch}}^{i,l} \quad (8)$$

由于高频噪声的原因, 所以选择的 $H(\theta)$ 为带通滤波器。 θ_1, θ_2 为超参数, 分别表示划分的下截止频率和上截止频率, 即只保留在 $[\theta_1, \theta_2]$ 范围内的频率。之后, 为了进一步增强频域季节项的特征表达能力, 模型引入了复值线性层, 对频域季节项进行线性变换。具体来说, 经过滤波器处理的频域季节项分片 $\mathbf{F}'_{s,\text{patch}}^{i,l}$ 被输入到复值线性层, 进行特征的加权和放大操作, 其公式为:

$$\mathbf{F}''_{s,\text{patch}}^{i,l} = \text{fre_linear}(\mathbf{F}'_{s,\text{patch}}^{i,l}) = \mathbf{W} \cdot \mathbf{F}'_{s,\text{patch}}^{i,l} + \mathbf{b} \quad (9)$$

其中, $\text{fre_linear}(\cdot)$ 为复值线性层, $\mathbf{F}''_{s,\text{patch}}^{i,l}$ 为经过线性变换后的频域季节项, \mathbf{W} 初始化为复数均匀分布, 其实部和虚部独立采样。该层未使用复数激活函数, 主要为了保持频域特征的线性变换特性。 \mathbf{b} 为偏置项。在反向传播过程中, 复值梯度的计算通过将复值权重的梯度分解为实部和虚部的偏导数, 分别对其进行更新。随后, 再通过逆

傅里叶变换将其转换回时域。在逆变换之后, 由于数据可能仍然存在一些波动或不平稳, 为此, 模型进一步在时域中引入线性层进行优化处理。时域线性层通过对数据的线性变换, 能够有效平滑结果, 最终生成更稳定的季节项表示。公式为:

$$\mathbf{S}_{s,\text{patch}}^{i,l} = \text{linear}(\mathbf{F}^{-1}(\mathbf{F}''_{s,\text{patch}}^{i,l})) \quad (10)$$

其中, $\text{linear}(\cdot)$ 为时域线性层, $\mathbf{F}^{-1}(\cdot)$ 为逆傅里叶变换。 $\mathbf{S}_{s,\text{patch}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 为最终稳定的季节项, 作为后面双重时间依赖模块的输入。

3) 趋势性增强

对于趋势项, 由于其变化通常是平滑且线性的, 所以直接在时域上使用线性层对趋势项进行增强。时域线性层通过加权求和等简单的线性变换, 扩大趋势项的变化幅度, 强化长期趋势信息。这种线性操作使得趋势项的变化更容易被模型捕捉, 同时, 线性层的简单性也使得其输出结果易于解释和理解。过程如式(11)所示。

$$\mathbf{T}_{t,\text{patch}}^{i,l} = \text{linear}(\mathbf{X}_{t,\text{patch}}^{i,l}) \quad (11)$$

其中, $\text{linear}(\cdot)$ 为时域线性层。 $\mathbf{T}_{t,\text{patch}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 作为双重时间依赖模块的输入。

1.5 双重时间依赖模块

双重时间依赖模块由两部分组成, 分别为长期时间变化提取器和局部波动提取器, 用于实现时间特征的提取, 如图 3 所示。

本文通过采用分支结构并行建模长期和短期特征, Transformer 在捕捉时间序列的全局依赖性方面表现优异, 尤其适合提取分片间存在的长距离时间步相关性, 因此非常适合作为长期时间变化提取器。相比之下, 一维卷积网络更擅长捕捉分片内的局部波动, 能够有效提取短期变化特征, 因此更适合作为局部波动提取器。双重时间依赖模块的设计可以概括为式(12)~(14)。

$$\mathbf{X}_{\text{long-term}}^{i,l} = \text{Trans}(\mathbf{T}_{t,\text{patch}}^{i,l}) \quad (12)$$

$$\mathbf{X}_{\text{short-term}}^{i,l} = \text{Conv}(\mathbf{S}_{s,\text{patch}}^{i,l}) \quad (13)$$

$$\mathbf{X}^{i,l} = \text{linear}(\text{Flatten}(\mathbf{X}_{\text{short-term}}^{i,l} + \mathbf{X}_{\text{long-term}}^{i,l})) \quad (14)$$

其中, $\text{Trans}(\cdot)$ 指的是 Transformer 编码器(具体如式(15)~(19)), 用于建模分片间的长期时间依赖, 得到长期时间依赖特征 $\mathbf{X}_{\text{long-term}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 。 $\text{Conv}(\cdot)$ 由 Conv1d 和非线性激活层组成, 用于提取分片内的局部波动性, 得到短期时间依赖特征 $\mathbf{X}_{\text{short-term}}^{i,l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 。最后, 获取到的长期依赖 $\mathbf{X}_{\text{long-term}}^{i,l}$ 和短期波动 $\mathbf{X}_{\text{short-term}}^{i,l}$ 会经融合操作得到整体信息, 通过 $\text{Flatten}(\cdot)$ 沿时间维度展平时间序列, 最后将序列通过线性层进行预测, 得到双重时间依赖模块的最终输出 $\mathbf{X}^{i,l}$ 。

1) 长期时间变化提取器

长期时间变化提取器由 Transformer 编码器组成, 通过计算分片间的注意力得到时间片之间的时间依赖关系, 捕获序列的全局相关性和长期时间模式。如图 4(a)所示。

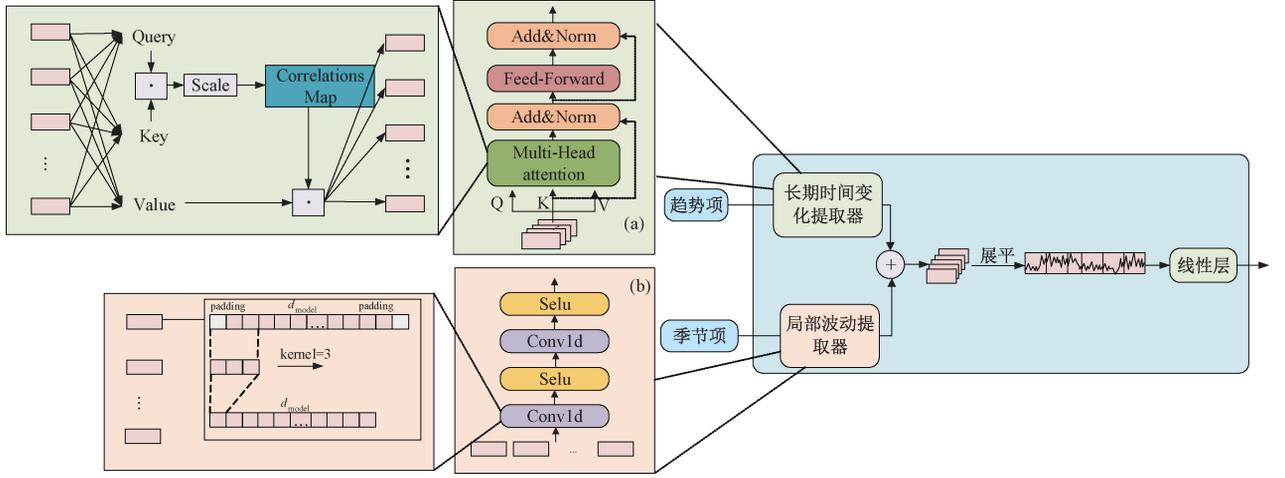


图 3 双重时间依赖模块

Fig. 3 Dual time-dependence module

多头注意力层的每个头 ($h = 1, 2, \dots, H$) 将 $\mathbf{T}_{t, \text{patch}}^{i, l}$ 这些分片与可训练的线性投影 \mathbf{W} 相乘, 转换成查询矩阵 $\mathbf{Q}_h^{i, l} = \mathbf{T}_{t, \text{patch}}^{i, l} \mathbf{W}_h^Q$ 、键矩阵 $\mathbf{K}_h^{i, l} = \mathbf{T}_{t, \text{patch}}^{i, l} \mathbf{W}_h^K$ 、值矩阵 $\mathbf{V}_h^{i, l} = \mathbf{T}_{t, \text{patch}}^{i, l} \mathbf{W}_h^V$, 之后进行注意力的计算。计算过程如式(15)和(16)所示。

$$\mathbf{X}_h^{i, l} = \text{Att}(\mathbf{Q}_h^{i, l}, \mathbf{K}_h^{i, l}, \mathbf{V}_h^{i, l}) = \text{softmax}\left(\frac{\mathbf{Q}_h^{i, l} \times (\mathbf{K}_h^{i, l})^T}{\sqrt{d_{\text{model}}}}\right) \mathbf{V}_h^{i, l} \quad (15)$$

$$\mathbf{X}_{\text{mha}}^{i, l} = \text{Concat}(\mathbf{X}_1^{i, l}, \mathbf{X}_2^{i, l}, \dots, \mathbf{X}_H^{i, l}) \quad (16)$$

其中, $\mathbf{X}_h^{i, l} \in \mathbb{R}^{n \times d_{\text{model}}/h}$ 表示单变量 i 时间序列的第 h 个头的输出, $\mathbf{Q}_h^{i, l} \times (\mathbf{K}_h^{i, l})^T$ 用于查询和键的相似度计算, $\sqrt{d_{\text{model}}}$ 为缩放因子用于避免点积值过大导致梯度消失问题。softmax(\cdot) 函数对每一行的权重进行归一化, 得到注意力权重矩阵。根据注意力权重矩阵和值矩阵, 计算最终的注意力输出 $\mathbf{X}_h^{i, l}$ 。 $\mathbf{X}_{\text{mha}}^{i, l} \in \mathbb{R}^{n \times d_{\text{model}}}$ 是每个头的输出拼接在一起形成最终的输出。

通过子层跳跃连接和前馈层计算特征值, 将特征值作为编码器最终的输出, 计算过程如式(17)~(19)所示。

$$\hat{\mathbf{X}}_{\text{mha}}^{i, l} = \text{LayerNorm}(\mathbf{X}_{\text{mha}}^{i, l} + \mathbf{T}_{t, \text{patch}}^{i, l}) \quad (17)$$

$$\mathbf{X}_{\text{feed}}^{i, l} = \text{FeedForward}(\hat{\mathbf{X}}_{\text{mha}}^{i, l}) \quad (18)$$

$$\mathbf{X}_{\text{long-term}}^{i, l} = \text{LayerNorm}(\mathbf{X}_{\text{feed}}^{i, l} + \hat{\mathbf{X}}_{\text{mha}}^{i, l}) \quad (19)$$

其中, LayerNorm(\cdot) 是层归一化方法, 它对子层的输出进行标准化, 使得模型对输入数据的变化更加稳定和鲁棒。FeedForward(\cdot) 是前馈层, 用于对每个位置的特征进行非线性变换和映射, 它由两个线性变换和一个 Relu 激活函数组成, 可以增强模型的表达能力和学习能力。

2) 局部波动提取器

局部波动提取器由 Conv1d 和非线性激活层组成, 如图 4(b) 所示。现有的基于分片的模型通常只建模分片间的关系, 分片内的关系通常是隐式学习的, 并没有特地去

关注分片内的局部波动。这样可能会导致模型无法准确捕捉到时间序列的细节和局部模式, 从而影响模型的性能。为了获取时间序列的局部模式, 本文采用 Conv1d 来对局部信息进行建模, 同时添加激活函数 SELU(\cdot) 引进非线性关系。计算过程如式(20)和(21)所示。

$$\mathbf{S}_{s, \text{patch}}^{i, l} = \text{SELU}(\text{Conv1d}(\mathbf{S}_{s, \text{patch}}^{i, l})) \quad (20)$$

$$\mathbf{X}_{\text{short-term}}^{i, l} = \text{SELU}(\text{Conv1d}(\mathbf{S}_{s, \text{patch}}^{i, l})) \quad (21)$$

模型采用两层卷积层, $\mathbf{S}_{s, \text{patch}}^{i, l}$ 为第一层卷积的输出, 同时也作为第二层卷积的输入。本文设置了较小的卷积核来建模分片内的局部波动, 第一层的卷积核大小为 1, 第二层的卷积核大小为 3。这是由于小卷积核可以更好地关注分片内的局部信息, 并捕捉到细微的变化, 在一定程度上能够增加模型对分片内局部波动的敏感性。

1.6 多周期融合

多层的预测结果会进行特征融合, 如式(22)所示。

$$\mathbf{X}_{\text{pre}}^i = \sum_{l=1}^k \mathbf{X}^{i, l} \quad (22)$$

$\mathbf{X}_{\text{pre}}^i \in \mathbb{R}^L$ 是将每一层的预测结果加和后得到的最终结果。

2 实验与结果分析

2.1 数据集

为评估本模型的效果, 本文在包含能源、天气、汇率等领域的 6 个公开数据集上进行了全面实验。使用的数据集包括油温数据集 ETT(以不同采样频率记录的 4 个数据集)、天气数据集 Weather 和汇率数据集 Exchange。表 1 为数据集的详细信息。

本文实验按照时间顺序将所有数据集分为训练集、验证集和测试集, 其中 ETT 数据集的比例为 6:2:2, 其他数据集的比例为 7:1:2。

2.2 实验设置

本文选择最优模型和有代表性的长时间序列预测

表 1 数据集详情
Table 1 Dataset details

数据集	长度	维度	采集频率
ETTh1	17 420	8	1 h
ETTh2	17 420	8	1 h
ETTm1	69 680	8	15 min
ETTm2	69 680	8	15 min
Weather	52 556	22	10 min
Exchange	7 588	9	1 d

模型作为对比实验的基准模型:

1) PatchTST^[11]: 最优模型。将时间序列分片后, 使用通道独立方法在 Transformer 编码器上进行学习。

2) FEDformer^[23]: 基于 Transformer 结构, 提出频域注意力机制, 具有线性复杂度。

3) Crossformer^[24]: 提出了新颖的变量间和变量内的两阶段注意力机制, 同时实现了从细粒度到粗粒度的分层建模。

4) TimesNet^[25]: 将时间序列从 1D 转为 2D 并使用卷积神经网络获取周期和周期内的时间依赖。

5) MICN^[16]: 提出结合局部特征和全局相关性来捕获时间序列(例如波动和趋势)的整体视图, 两者都实现了多尺度建模。

6) DLinear^[20]: 仅使用线性层就取得了很好的效果。

本文实验采用 L2 损失函数进行训练, 并使用 ADAM 优化器来优化模型参数, 初始学习率为 10^{-4} 。使用固定步长衰减策略, 步长按训练轮次比例划分, 批次大小设置为 32。训练过程在 10 个轮次内提前结束, 若在一定轮次的测试集上未获得更低的验证集误差即提前停止训练。

实验硬件环境为至强 E5-2678 v3 2.5 GHz 48 核 96 线程 CPU, RTX 2080 Ti GPU; 软件环境为 Ubuntu 18.04 系统, PyTorch 1.12.1 深度学习框架, Python 3.10.13。

所有实验均使用均方误差(mean squared error, MSE)和平均绝对误差(mean absolute error, MAE)作为评价指标, 计算公式如式(23)和(24)所示。

$$\text{MSE} = \frac{1}{L} \sum_{i=1}^L (y_i - \hat{y}_i)^2 \quad (23)$$

$$\text{MAE} = \frac{1}{L} \sum_{i=1}^L |y_i - \hat{y}_i| \quad (24)$$

其中, L 是预测长度, y_i 和 \hat{y}_i 分别对应第 i 时刻的真实值和预测值。MSE 和 MAE 都是用来衡量模型预测值与实际观测值之间的差异的指标。较小的 MSE 和 MAE 表示模型的预测值与实际观测值之间的差异较小, 即模型的预测精度较好。

2.3 不同模型比较

为了更好地和基准模型作比较, 本文的实验设置与 Autoformer 一致, 输入长度固定为 96, 对每个数据集预测

的输出长度各取[96, 192, 336, 720]。PatchTST 原论文实验的输入长度为 336, 本文为了实验的公平性, 将其输入长度调整为 96 后再进行实验, 并未拉低最优模型的实验结果。实验结果如表 2 所示, 最优结果加粗显示。

从结果可以看出, 本文提出的模型在 6 个公开数据集中全面超越了现在的最优模型 PatchTST, 这体现了本文模型在长期时间序列预测任务上的优越性。特别地, 在 ETTh1 上本文模型与最优模型 PatchTST 相比, MSE 最大下降了 10.14%, MAE 最大下降了 6.68%。在 Exchange 数据集中, 本文模型的 MSE 最大下降了 7.95%, MAE 最大下降了 5.37%。在其他数据集上结果虽然有差异, 但整体呈现精度提升的趋势。虽然 ETTh1 数据集并不能显著地观测到周期模式, 本文模型却依旧可以在这个数据集上取得很好的效果。模型在 Weather 数据集上的部分结果并没有超过 Crossformer, 可能是因为 Weather 数据集变量间存在复杂的相关性, 而模型并没有显式建模变量间的关系。未来的工作将会重点探索变量间的依赖关系。

2.4 对比试验

1) 分层对比实验

本文模型通过周期的数量确定最终层数, 从多周期的角度对时间序列进行分析与建模。为了评估不同周期数量对模型效果的影响, 本文在 ETTh1、ETTh2 以及 Exchange 数据集上进行了对比实验, 结果如图 4 所示, 横坐标为模型层数, 纵坐标分别为精度 MSE 和 MAE。实验结果表明, 周期层数确定为三层时模型效果最好。随着层数的增加, 模型的精度逐渐提高, 但是在三层之后, 模型的精度逐渐下降, 并且退化得较为明显。这表明, 在选择模型的层数时, 过度增加层数并不会带来积极的效果。相反, 盲目增加层数可能会导致模型的过拟合问题, 进而降低预测精度并增加模型复杂度。

2) 多周期分片实验

为了进一步证明本文所提出的多周期分层的重要性, 本文做了以下对比实验。如图 5 所示, 图 5 左图是直接对整个时间序列使用 Transformer 进行注意力权重的计算; 图 5 中间图和右图表示通过周期提取模块, 在 Exchange 数据集上获取到前两个主要周期, 分别为 48 和 32, 经过周期分片模块后, 将时间序列使用长期变化提取器提取长期时间依赖, 利用 Transformer 模型提取分片间特征, 获得注意力权重矩阵。

从图 5 中可以看到, 直接对原始序列使用 Transformer 模型获取时间依赖, 注意力较为均匀, 缺乏明显的集中区域。而注意力均匀分布可能导致模型未能有效识别和关注时间序列中的关键模式或特征。对比之下, 对时间序列进行周期长度为 48 和 32 的分片操作, 并使用长期变化提取器提取注意力的做法能够使注意力的分布集中于特定区域, 显示出明显的模式, 同时不同周期的设置使注意力模式呈现较大差异, 表明模型在不同周期设置下能够更好地

表2 6个数据集上的预测结果

Table 2 Forecasting results on six datasets

模型		MDTDNet		PatchTST		FEDformer		Crossformer		TimesNet		MICN		DLinear	
数据集	指标	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.371	0.391	0.414	0.419	0.376	0.419	0.423	0.448	0.384	0.402	0.398	0.427	0.386	0.400
	192	0.421	0.422	0.460	0.445	0.420	0.448	0.471	0.474	0.436	0.429	0.430	0.453	0.437	0.432
	336	0.454	0.440	0.501	0.466	0.459	0.465	0.570	0.546	0.491	0.469	0.440	0.460	0.481	0.459
	720	0.459	0.461	0.500	0.488	0.506	0.507	0.653	0.621	0.521	0.500	0.491	0.509	0.519	0.516
ETTh2	96	0.288	0.339	0.302	0.348	0.358	0.397	0.745	0.584	0.340	0.374	0.332	0.377	0.333	0.387
	192	0.371	0.392	0.388	0.400	0.429	0.439	0.877	0.656	0.402	0.414	0.422	0.441	0.477	0.476
	336	0.407	0.427	0.426	0.433	0.496	0.487	1.043	0.731	0.452	0.452	0.447	0.474	0.594	0.541
	720	0.423	0.441	0.431	0.446	0.463	0.474	1.104	0.765	0.462	0.468	0.442	0.467	0.831	0.657
ETTm1	96	0.318	0.357	0.329	0.367	0.379	0.419	0.404	0.426	0.338	0.375	0.360	0.399	0.345	0.372
	192	0.366	0.381	0.367	0.385	0.426	0.441	0.450	0.451	0.374	0.387	0.402	0.426	0.380	0.389
	336	0.392	0.402	0.399	0.410	0.445	0.459	0.532	0.515	0.410	0.411	0.403	0.437	0.413	0.413
	720	0.451	0.437	0.454	0.439	0.543	0.490	0.666	0.589	0.478	0.450	0.459	0.464	0.474	0.453
ETTm2	96	0.175	0.257	0.175	0.259	0.203	0.287	0.287	0.366	0.187	0.267	0.203	0.287	0.193	0.292
	192	0.239	0.300	0.241	0.302	0.269	0.328	0.414	0.492	0.249	0.309	0.262	0.326	0.284	0.362
	336	0.299	0.339	0.305	0.343	0.325	0.366	0.597	0.542	0.321	0.351	0.305	0.353	0.369	0.427
	720	0.398	0.395	0.402	0.400	0.421	0.415	1.730	1.042	0.408	0.403	0.389	0.407	0.554	0.522
Weather	96	0.174	0.214	0.177	0.218	0.217	0.296	0.158	0.230	0.172	0.220	0.183	0.250	0.196	0.255
	192	0.223	0.257	0.225	0.259	0.276	0.336	0.206	0.277	0.219	0.261	0.246	0.317	0.237	0.296
	336	0.277	0.296	0.278	0.297	0.339	0.380	0.272	0.335	0.280	0.306	0.293	0.335	0.283	0.335
	720	0.350	0.343	0.354	0.348	0.403	0.428	0.398	0.418	0.365	0.359	0.373	0.399	0.345	0.381
Ex-change	96	0.081	0.194	0.088	0.205	0.148	0.278	0.256	0.367	0.107	0.234	0.173	0.297	0.088	0.218
	192	0.170	0.292	0.176	0.299	0.271	0.315	0.470	0.509	0.226	0.344	0.324	0.408	0.176	0.315
	336	0.293	0.392	0.301	0.397	0.460	0.427	1.268	0.883	0.367	0.448	0.639	0.598	0.313	0.427
	720	0.848	0.693	0.901	0.714	1.195	0.695	1.767	1.068	0.964	0.746	1.218	0.862	0.839	0.695

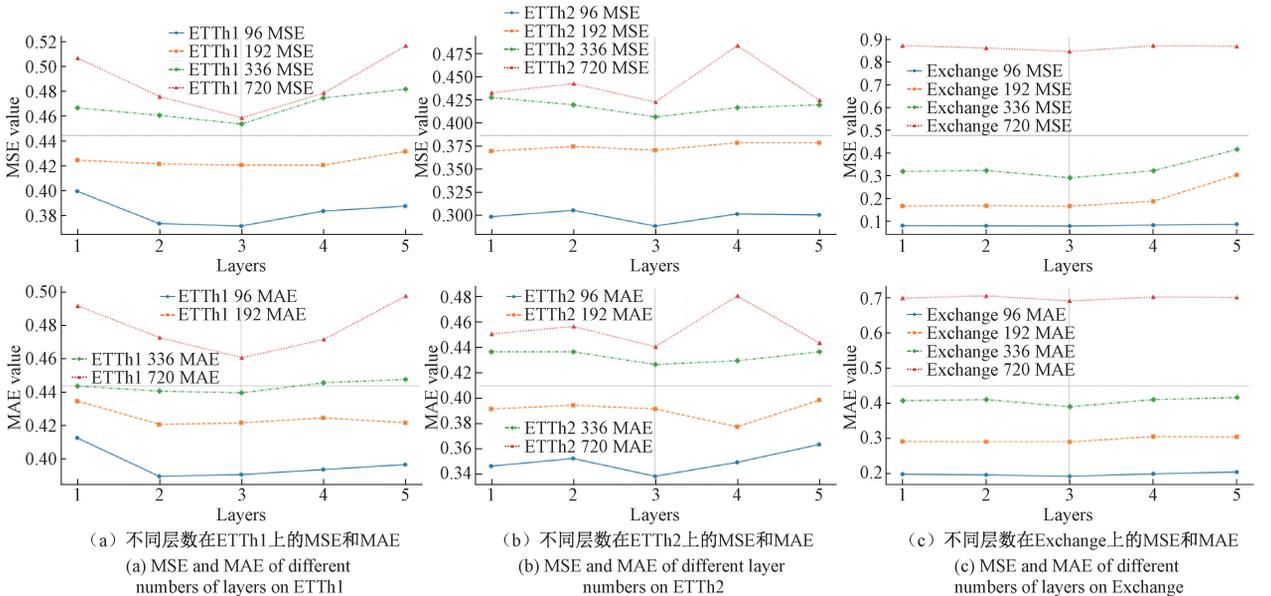


图4 分层对比实验

Fig. 4 Layered comparison experiment

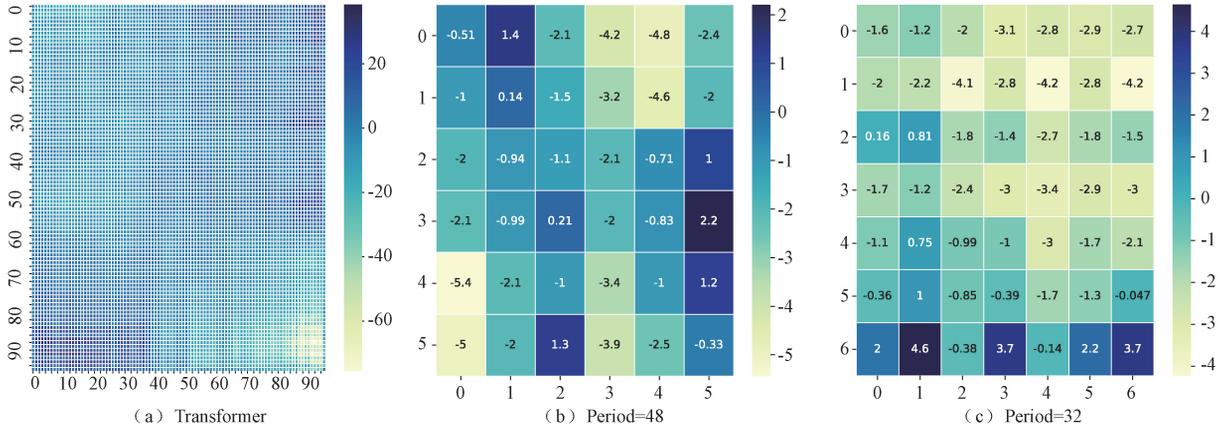


图 5 多周期分片实验

Fig. 5 Multi-periodic patching experiment

地识别时间序列中的不同周期性特征。

实验结果证明,多周期分层通过多个不同周期的特征提取可以有效提高模型对时间序列中周期模式的识别能力。不同的周期设置让模型能够捕获多样的周期性特征,提高了模型的表现和解释性。

2.5 消融试验

为了进一步证明模型提出的季节趋势增强模块和双重时间依赖模块的有效性,本文在 ETTh1、ETTh2 以及 Exchange 数据集上进行了消融实验。

1) 季节趋势增强模块的消融实验

季节趋势增强模块由周期分片、趋势性增强、季节性增强 3 个子模块组成。模型变体如下:

(1) PatchTST (period): 开展周期分片,将 PatchTST 模型分片长度改为周期长度,选取由傅里叶变换后得到的最重要的周期。

(2) MDTDNet (non-period): 去掉周期分片,将模型 MDTDNet 的分片长度的更改为基于分片思想的模型常用分片大小[2、8、12]。

(3) MDTDNet (de-TE): 其他条件不变,将 MDTDNet 模型的趋势性增强部分去除。

(4) MDTDNet (de-SE): 其他条件不变,将 MDTDNet 模型的季节性增强部分去除。

(5) MDTDNet (de-STE): 其他条件不变,将 MDTDNet 模型的数据增强部分去除,仅保留周期分片部分。

如表 3 所示,采用周期长度分片的 PatchTST (period) 模型在 MAE 和 MSE 上均显著优于原始(不开展周期分片) PatchTST 模型,其中在 ETTh1 数据集上 MSE 最高降低了 8.8%,平均降低了 5.3%,这说明周期分片模块不仅有效还具有普适性。而 MDTDNet (non-period) 的 MSE 和 MAE 均没有超过 MDTDNet 的模型结果。

表 3 季节趋势增强模块的消融实验

Table 3 Ablation experiment on seasonal trend enhancement module

模型	MDTDNet (non-period)	PatchTST (period)		MDTDNet (de-TE)		MDTDNet (de-SE)		MDTDNet (de-STE)		MDTDNet		PatchTST			
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE		
ETTh1	96	0.379	0.396	0.378	0.396	0.378	0.390	0.377	0.390	0.384	0.395	0.372	0.391	0.414	0.419
	192	0.426	0.427	0.422	0.424	0.429	0.420	0.428	0.429	0.435	0.424	0.421	0.422	0.460	0.445
	336	0.469	0.449	0.457	0.444	0.471	0.441	0.470	0.440	0.462	0.451	0.454	0.440	0.501	0.466
	720	0.533	0.491	0.522	0.498	0.497	0.478	0.504	0.485	0.5212	0.509	0.459	0.461	0.500	0.488
ETTh2	96	0.305	0.302	0.295	0.342	0.294	0.345	0.297	0.347	0.292	0.343	0.289	0.339	0.302	0.348
	192	0.388	0.402	0.380	0.405	0.377	0.398	0.378	0.398	0.379	0.397	0.371	0.392	0.388	0.400
	336	0.425	0.435	0.429	0.439	0.412	0.429	0.413	0.430	0.415	0.427	0.407	0.427	0.426	0.433
Ex-change	720	0.462	0.465	0.446	0.455	0.427	0.447	0.426	0.446	0.425	0.449	0.423	0.441	0.431	0.446
	96	0.082	0.178	0.101	0.223	0.083	0.199	0.089	0.199	0.085	0.201	0.081	0.194	0.088	0.205
	192	0.183	0.303	0.180	0.300	0.188	0.305	0.188	0.305	0.175	0.295	0.170	0.292	0.176	0.299
	336	0.300	0.396	0.343	0.426	0.342	0.423	0.343	0.424	0.342	0.423	0.293	0.392	0.301	0.397
	720	0.844	0.691	0.892	0.714	0.886	0.711	0.887	0.711	0.851	0.694	0.848	0.693	0.901	0.714

此外,与完整的 MDTNet 模型相比,去掉季节趋势增强模块中的任一部分,模型的性能都会受到显著的负面影响。这表明,该模块在提升模型捕捉时间特征的能力方面起到了至关重要的作用,是 MDTNet 能够取得优异预测性能的关键组成部分。

周期分片通过按照合适的周期长度切分时间序列,有助于缓解时间序列低语义特征的问题,同时结合周期性和趋势性增强模块,进一步提升模型对长期和短期依赖关系的建模能力。具体来说,周期分片方法帮助模型更好地捕捉序列的周期特征,而增强模块则在片内有效提取更多的时序信息。消融实验表明,周期分片设计在 PatchTST 上也适用,因此可以推广应用至其他基于分片思想的时间序列预测模型中,以进一步提升模型性能。对比结果展示了利用周期性信息和增强机制在提升时间序列预测模型性能方面的潜力。

2) 双重时间依赖模块的消融实验

为了验证本文提出的双重时间依赖模块的有效性,实

验中设置模型变体如下:

(1)MDTNet(de-long):保持原模型不变,去除长时间变化提取器,只建模分片内的局部时间模式。

(2)MDTNet(de-short):保持原模型不变,去除局部波动提取器,只建模分片间的长期时间模式。

(3)MDTNet(de-dual):保持原模型不变,去除长时间变化提取器和局部波动提取器,将时间序列进行分片后不进行长短期特征提取,直接沿时间维度进行展平,然后通过一个线性层输出预测结果。

表 4 是双重时间依赖模块的消融实验结果。局部波动提取器和长时间变化提取器是该模块的两个重要组成部分。从结果可以看到,原模型(MDTNet)在不同预测窗口下都能保持优势,并在消融实验中展现了最优结果,这说明本文模型通过双重时间依赖模块可以有效地捕捉分片间长期依赖关系和分片内短期波动来提高预测准确性,在处理时间依赖关系方面具有较好的性能。

表 4 双重时间依赖模块的消融实验

Table 4 Ablation experiments on dual temporal dependency module

模型		MDTNet(de-dual)		MDTNet(de-short)		MDTNet(de-long)		MDTNet	
数据集	指标	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.382	0.392	0.376	0.391	0.380	0.393	0.372	0.391
	192	0.433	0.421	0.422	0.428	0.431	0.422	0.421	0.422
	336	0.476	0.444	0.465	0.466	0.474	0.446	0.454	0.440
	720	0.478	0.470	0.493	0.478	0.475	0.470	0.459	0.461
ETTh2	96	0.291	0.340	0.292	0.342	0.289	0.338	0.289	0.339
	192	0.384	0.399	0.369	0.394	0.375	0.391	0.371	0.392
	336	0.420	0.430	0.415	0.427	0.425	0.435	0.407	0.427
Exchange	720	0.497	0.485	0.438	0.453	0.425	0.443	0.425	0.442
	96	0.087	0.203	0.089	0.206	0.083	0.199	0.081	0.194
	192	0.173	0.294	0.177	0.298	0.174	0.294	0.170	0.292
	336	0.364	0.435	0.374	0.442	0.344	0.423	0.293	0.392
	720	0.858	0.698	0.859	0.696	0.851	0.695	0.848	0.693

消融实验去除局部波动提取器(MDTNet(de-short))和去除长时间变化提取器(MDTNet(de-long))的实验结果显示,这两个子模块的去除都导致了模型性能的一定程度的下降。这表明建模局部波动和长时间变化对于模型预测精度的提升有一定的贡献。值得注意的是,MDTNet(de-long)的精度下降明显,这说明此模块可以识别出较长时间范围内的变化和趋势,并且对模型在各个窗口的预测性能保持稳定起着关键作用。通过保持这种稳定性,模型能够更好地适应不同时间尺度的数据变化,并提高其在各种预测情境下的准确性和可靠性。

综上所述,双重时间依赖模块对于模型的精度提升起到了关键作用,这个模块全面地考虑了将时间序列分片后产生的两种时间依赖模式,能够有效地捕捉局部波动和长

期时间依赖关系,从而提高预测精度。

3) 消融模块的显著性检验

为了进一步验证所提模块的有效性,使用在 ETTh1 数据集上的五次重复实验的结果 MSE,进行配对的 t 检验,比较消融模块后的模型和 MDTNet 模型的显著性差异,并通过 p 值来判断是否有显著性差异。如果 p 值小于 0.05,则认为该比较结果是显著的。结果如图 6 所示。

这些显著性差异表明,文章所提模块会显著影响模型的表现。例如,删除某些模型主要模块(如季节趋势增强模块、双重时间依赖模块)会导致模型性能的显著下降。这些结果对于进一步的模型优化和改进具有重要意义。

2.6 效率分析

本文将 MDTNet 模型在 ETTh1 数据集上训练一次

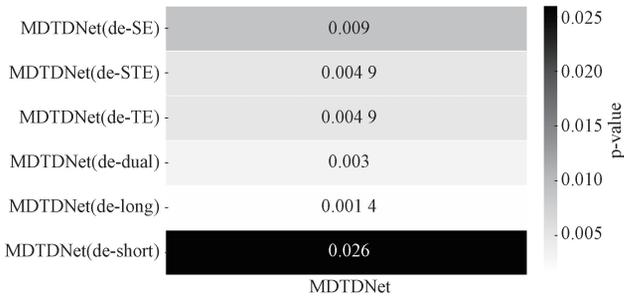


图 6 显著性检验
Fig. 6 Significance test

的运行时间与 PatchTST、FEDformer、DLinear 等模型进行比较,结果如图 7 所示,横坐标为预测窗口大小,纵坐标为时间。可以看到,MDTDNet 的训练时间在不同预测长度下相对较短,显示出较高的训练效率。同时,大多数模型的训练时间会随着序列长度的增加而增加。FEDformer 和 Crossformer 的训练时间增长显著,尤其在较长的序列长度下,表明这些更复杂的模型,对长序列的计算需求较高。DLinear 由于仅通过线性层来进行预测,所以耗时最短。虽然 MDTDNet 训练时间比 DLinear 更长,但是精度却远远高于 DLinear,这表明模型同时兼顾了效率和精度,展现了其在时间序列预测任务中的优越性。

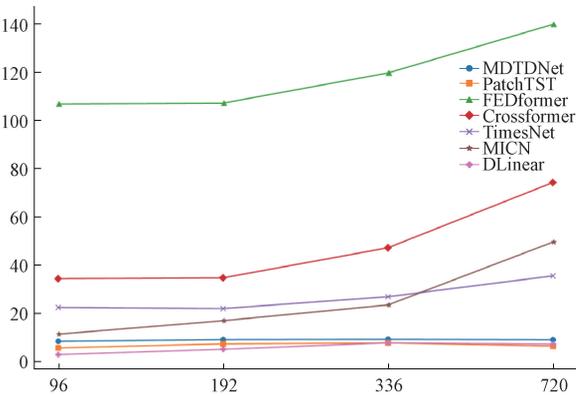


图 7 ETTh1 数据集训练时长 (s/epoch)

Fig. 7 The training time of ETTh1 dataset (s/epoch)

3 结 论

本文提出了基于双重时间依赖学习的多周期分层模型 MDTDNet。首先,模型通过周期提取模块自适应地识别重要周期,并划分为多个不同的分层。随后,在每个分层内部,模型对时间序列进行周期分片并结合季节性增强和趋势性增强模块,有效提取时间序列的内在特征。最后,通过双重时间依赖模块,分别捕捉序列的长期和局部时间模式,从而全面建模时间序列的动态特性。在六个数据集上的对比实验和消融实验表明,MDTDNet 不仅具有较高的预测精度,还具备良好的普适性和鲁棒性。但是,MDTDNet 采用通道独立策略,未显式建模变量间的复杂

依赖关系。未来的研究将重点探索多元时间序列中变量间依赖的建模方法,以进一步提升模型的预测性能和可解释性。

参 考 文 献

- [1] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting [C]. International Conference on Learning Representations, 2018.
- [2] ANGRYK R A, MARTENS P C, AYDIN B, et al. Multivariate time series dataset for space weather data analytics[J]. Scientific Data, 2020, 7(1):227.
- [3] DEB C, ZHANG F, YANG J J, et al. A review on time series forecasting techniques for building energy consumption[J]. Renewable and Sustainable Energy Reviews, 2017, 74: 902-924.
- [4] ZHANG G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50: 159-175.
- [5] CHATFIELD C. The holt-winters forecasting procedure[J]. Journal of the Royal Statistical Society: Series C(Applied Statistics), 1978, 27(3): 264-279.
- [6] 肖霞,马强,杨震.结合注意力和多尺度特征的电动汽车负荷预测[J].电子测量技术,2025,48(5):57-64.
- [7] XIAO X, MA Q, YANG ZH. Electric vehicle load prediction combining attention and multiscale features[J]. Electronic Measurement Technology, 2025, 48(5): 57-64.
- [8] KITAEV N, KAISER L, LEVSKAYA A. Reformer: The efficient transformer[C]. International Conference on Learning Representations, 2020.
- [9] 林泓,刘桂雄,戈燕红,等.单变量时间序列的 MHAGRU-MCCE 分类方法[J].电子测量技术,2024, 47(16):83-91.
- [10] LIN H, LIU G X, GE Y H, et al. MHAGRU-MCCE classification method for univariate time series [J]. Electronic Measurement Techniques, 2024, 47(16): 83-91.
- [11] 龚小龙,孔玲爽,袁川来,等.非线性时间序列粒子群优化 B 样条网络预测模型[J].电子测量与仪器学报, 2017,31(12):1890-1895.
- [12] GONG X L, KONG L SH, YUAN CH L, et al. Nonlinear time series particle swarm optimization B-spline network prediction model [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(12): 1890-1895.
- [13] LIU Y, HU T G, ZHANG H R, et al. Itransformer: Inverted transformers are effective for time series forecasting[C]. The Twelfth International Conference

- on Learning Representations, 2024.
- [11] NIE Y Q, NGUYEN N H, SINHTHONG P, et al. A time series is worth 64 words: Long-term forecasting with transformers [C]. The Eleventh International Conference on Learning Representations, 2023.
- [12] 李岚皓, 严皓钧, 周号益, 等. 基于神经网络的多尺度信息融合时间序列长期预测模型[J/OL]. 计算机应用, 1-12[2025-05-15]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20241202.1032.002.html>.
- LI L H, YAN H J, ZHOU H Y, et al. Long term prediction model of multi-scale information fusion time series based on neural network [J]. Computer Applications, 1-12[2020-02-07]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20241202.1032.002.html>.
- [13] CHEN P, ZHANG Y Y, CHENG Y Y, et al. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting[C]. The Twelfth International Conference on Learning Representations, 2024.
- [14] ZHANG Y F, WU R, DASCALU S M, et al. Multi-scale transformer pyramid networks for multivariate time series forecasting[C]. IEEE Access, 2024.
- [15] SHABANI A, ABDI A, MENG L, et al. Scaleformer: Iterative multi-scale refining transformers for time series forecasting[C]. The Eleventh International Conference on Learning Representations, 2023.
- [16] WANG H Q, PENG J, HUANG F H, et al. Micn: Multi-scale local and global context modeling for long-term series forecasting[C]. The Eleventh International Conference on Learning Representations, 2023.
- [17] WEN Q S, ZHOU T, ZHANG CH L, et al. Transformers in time series: A survey[J]. ArXiv preprint arXiv:2202.07125, 2022.
- [18] WEN Q S, YANG L X, ZHOU T, et al. Robust time series analysis and applications: An industrial perspective [C]. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022: 4836-4837.
- [19] WU H X, XU J H, WANG J M, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. Advances in Neural Information Processing Systems, 2021, 34: 22419-22430.
- [20] 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, 37(9): 11121-11128.
- [21] 徐仲武. 基于周期分解和局部依赖的多变量长时间序列预测[D]. 济南: 山东大学, 2024.
- XU ZH W. Multivariate long time series prediction based on periodic decomposition and local dependence[D]. Jinan: Shandong University, 2024.
- [22] 曾渝, 张洋, 曾尚, 等. 基于多尺度门控膨胀卷积网络的时间序列预测算法[J]. 计算机应用, 2024, 44(11): 3427-3434.
- ZENG Y, ZHANG Y, ZENG SH, et al. Time series prediction algorithm based on multi-scale gated dilated convolutional network [J]. Computer Applications, 2024, 44(11): 3427-3434.
- [23] ZHOU T, MA Z Q, WEN Q S, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]. International Conference on Machine Learning. PMLR, 2022: 27268-27286.
- [24] ZHANG Y H, YAN J CH. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]. The Eleventh International Conference on Learning Representations, 2023.
- [25] WU H X, HU T G, LIU Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis[C]. The Eleventh International Conference on Learning Representations, 2023.

作者简介

衡红军, 博士, 副教授, 主要研究方向为多元时间序列预测、时间序列异常检测。

E-mail: henghjcauc@163.com

周晨晓(通信作者), 硕士研究生, 主要研究方向为多元时间序列预测。

E-mail: zchenxiao1999@163.com