

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

赵龙港¹, 车超^{1,2}, 赵天明¹

¹(大连大学 先进设计与智能计算省部共建教育部重点实验室, 辽宁 大连 116622)

²(大连大学 软件工程学院, 辽宁 大连 116622)

E-mail:zhao_longgang@qq.com

摘要: 长期时间序列预测在现实场景中扮演重要角色。先前的研究表明,基于 Transformers 的模型采用的逐点自注意力会增加计算复杂度,而基于线性结构和通道独立的模型可以获得更高的效率和准确性。然而,长期时间模式在不同通道之间也存在难以抽取的依赖关系。为了解决计算复杂度高和复杂时间模式难以捕捉的问题,该文提出了通道融合和序列平稳化模型,模型结合了通道独立与通道依赖的训练策略,基于线性结构发掘序列单个通道的相关性,并使用由傅里叶运算启发的卷积结构来自适应地融合不同的通道。同时,通过堆叠序列通道融合-分解模块,进一步提高模型的预测性能。此外,该文在子序列级别引入了平稳化与反平稳化模块,从而提高了模型的泛化能力。在长期预测方面,所提模型在3个通用时序数据集上的准确度超越了其他基准模型。

关键词: 时间序列预测;线性模型;周期分解;通道融合卷积;平稳化

中图分类号: TP183

文献标识码: A

文章编号: 1000-1220(2025)05-1120-07

Long-term Time Series Forecasting with Channel Fusion and Sequence Stabilization

ZHAO Longgang¹, CHE Chao^{1,2}, ZHAO Tianming¹

¹(Key Laboratory of Advanced Design and Intelligent Computing, Ministry of Education, Dalian University, Dalian 116622, China)

²(College of Software Engineering, Dalian University, Dalian 116622, China)

Abstract: Long-term time series forecasting plays a significant role in real-world scenarios. Previous studies showed that Transformer-based models used point-by-point self-attention, which increased computational complexity. In contrast, models based on linear structures and channel independence could achieve higher efficiency and accuracy. However, the long-term temporal patterns also had dependencies between different channels that are difficult to extract. To address the issues of high computational complexity and the difficulty of capturing complex temporal patterns, this paper proposed Channel Fusion and Sequence Stabilization Network, which combines channel independent and channel dependent training strategies. It explored correlations of individual sequence channels based on linear structures and adaptively fuses different channels using a convolutional structure inspired by Fourier operations. Meanwhile, by stacking sequential channel fusion-decomposition modules, the predictive performance of the model is further enhanced. In addition, this paper introduced stabilization and anti-stabilization modules, operating at the subsequence level to improve the generalization ability of the model. In terms of long-term forecasting, the proposed model outperforms other baseline models in accuracy on three common time series datasets.

Keywords: time series forecasting; linear model; periodic decomposition; channel fusion convolution; stabilization

0 引言

现实世界每天都在产生大量的时间序列数据,对这些数据的分析和挖掘能够揭示蕴含在数据中的潜在规律,帮助人们更清晰地认识事物。长期时间序列预测在投资分析、电力调度、气象预报、交通流量预测等领域有着广泛的应用,从而引起了人们极大的研究兴趣^[1]。通常,序列中单个时间点提供的信息有限,因此许多工作关注一段时间内的复杂变化,如连续性、周期性、趋势性等。这些变化相互混合和重叠,更能反映

时序数据的内在特性,使得对时间变化的建模极具挑战性。

目前,在长期时间序列预测领域主要采用的主流深度学习模型包括循环神经网络(RNN, Recurrent Neural Network)^[2,3]、时域卷积网络(TCN, Temporal Convolutional Network)^[4-6]、Transformer^[7-15]以及线性模型。基于RNN的时间序列预测方法将过去的信息压缩为内部记忆状态,并在每个时间步递归地更新。但是,梯度消失、爆炸问题和低效的训练过程极大地限制了基于RNN模型的应用;基于TCN的方法则利用卷积神经网络来在时间维度上提取信息。由于时间序

收稿日期:2024-02-08 收修改稿日期:2024-03-25 基金项目:国家自然科学基金面上项目(62076045)资助;辽宁省教育厅高校基本科研项目揭榜挂帅服务地方项目(LJKFZZ20220290)资助;高等学校学科创新引智计划项目(D23006)资助;大连大学学科交叉项目(DLUXK-2023-YB-003, DLUXK-2023-YB-009, DLUXK-2022-ZD-003)资助。作者简介:赵龙港,男,1998年生,硕士研究生,研究方向为时间序列预测;车超,男,1981年生,博士,教授,CCF会员,研究方向为医学信息学、自然语言处理和数据挖掘;赵天明(通信作者),男,1990年生,博士,讲师,研究方向为机器学习、深度学习、图像处理。

列数据的局部相关性体现在一个时隙内的连续变化,因此卷积滤波器可以有效地捕捉这种局部特征.但卷积核的局部性,导致其难以捕获长期依赖信息;最近,基于注意力机制的Transformer结构及其变体在多个领域表现出色,它可以捕获时间点之间的时间依赖性.但在处理长序列时,具有二次复杂度的注意力机制难以从冗长的序列中找出可靠的关系;最近有研究表明经过精巧设计地简单线性模型可以胜过所有基于Transformer的方法^[16].然而,该线性模型采用了与以往的研究不同的通道无关的训练策略,没有将所有通道视为一个整体,而是训练了一个在所有通道之间共享的单变量预测模型.而在实际的时间序列预测问题中,不同通道之间存在一定的相互关联性.

为了克服线性模型难以应对复杂的时间变化的问题,本文提出了通道融合与时序平稳化网络(Channel Fusion and sequence Stabilization Network, CFSNet). CFSNet通过堆叠一系列的CFSNBlock,对时间序列进行周期和趋势的分解,并通过卷积模块实现了不同通道的自适应融合.为了更加高效地捕获序列间的依赖关系,本文采用线性结构分别对周期项和趋势项进行处理.此外,还引入了平稳化和反平稳化模块,以提高模型的泛化能力.通过对未来的预测,该模型可以预估未来事物发展的趋势,为决策提供支撑.

本文的主要贡献总结如下:

- 1) 提出一种基于傅里叶启发卷积与线性模型的多变量时间序列预测模型 CFSNet,它将通道融合-分解架构整合为CFSNBlock并堆叠起来.使模型获得了快速和精准的预测能力.
- 2) 使用了序列平稳化和反平稳化策略,两者相辅相成,为CFSNBlock提供了更加平稳的输入,提高了模型的泛化能力.
- 3) 在3个真实的时间序列数据集上的实验结果表明,CFSNet的预测性能超越了其他基准模型,并具有良好的泛化性.

1 相关工作

由于时间序列预测在各领域具有重要作用,各种模型得到了很好的发展.时间序列预测最初起步于经典算法.

早期的传统方法假设时间变化遵循预定义的模式.例如,ARIMA^[17]模型通过寻找历史数据之间的前后关联,将非平稳过程转换为平稳过程来预测未来;Holt-Winter^[18]使用3次指数平滑法,将历史时序数据输入3个递推序列,再由3个序列的递推值来推算出预测数据的值. Prophet^[19]采用可分解的时间序列模型,以减少需要构建的特征.然而,现实世界时间序列的变化通常过于复杂,无法被预定义的模式覆盖,这限制了这些传统方法的适用性.

随着深度学习在不同领域的应用愈发广泛,深度学习模型在建模特征间的非线性关系上具有更大的优势.因此,基于深度学习的时间序列预测方法引起了广泛的关注.目前广泛使用的深度学习时间序列预测方法包括基于递归神经网络(Recurrent Neural Network, RNN)、卷积神经网络(Convolutional Neural Network, CNN)、Transformer和多层感知机(Multilayer Perceptron, MLP)的方法.

基于RNN的方法在马尔可夫假设^[2,20,21]的基础上对连续时间点进行建模.例如,Salinas^[3]等人提出的多步概率预测模型DeepAR可以预测未来时序的分布和处理多条序列之间变动幅度对模型的影响,从而提高准确性. Rangapuram^[2]等人提出的深度状态空间模型(Deep State Space Model, DSSM)只建模连续两个隐藏状态的关系,解决了训练和预测不一致的问题.但是,RNN的方法单步预测的特点限制了它的性能,还有可能出现梯度爆炸的现象.

基于CNN的方法通过使用时间卷积网络^[4,22]等技术提取沿时间维度的变化信息和特征. S Bai^[4]等人采用空洞卷积结合因果卷积的网络结构,在复杂度不变的同时引入更长的历史信息. Liu^[5]等人提出的SCI-Net将数据下采样为两个子序列,使用不同的卷积滤波器提取每个子序列特征. MICN^[23]提出了一种串行结构来对时间序列进行长期依赖建模,尝试使用卷积操作将局部信息与全局信息聚合,但是增加了参数量.总之,卷积结构只能捕获局部信息,限制了它在长期序列预测上的表现.

基于Transformer的模型得益于自注意力机制,在许多领域取得了最好的表现.例如,LogTrans^[12]使用LogSparrse结构降低空间复杂度,利用因果卷积在自注意力层中产生键值和查询. Informer^[11]使用蒸馏技术,在堆叠的注意力模块中逐步提取最重要的信息来减少参数量. Autoformer^[11]使用了传统时间序列分解和自相关的思想,将原始序列分解为趋势循环部分和周期循环部分,在自相关模块对周期项执行特征挖掘,有效地增强了预测效果. FEDformer^[13]在频域应用注意力机制,增强对噪声的鲁棒性. Pyraformer^[15]应用金字塔注意力模块,可以很好地捕获不同尺度的时间依赖性,并降低了时间复杂度.与直接从输入中建模局部信息不同,注意力机制不需要堆叠许多层来提取全局信息,使得在时间序列预测中取得了显著成果.尽管更能学习长期依赖关系,但是注意力机制的复杂度更高,学习难度也更大.

基于MLP的方法采用线性结构沿时间维度进行处理,将时间依赖性转换为固定参数的线性层,显著提高了模型计算的速度.例如,N-BEATS^[24]使用多个块串联,每个块只学习序列的一部分信息. DLinear使用了一层线性层,并采用通道独立的训练策略,在时间复杂度和预测效果上超过了基于Transformer的模型.本文在DLinear使用的线性结构的基础上对通道独立策略进行改进,并结合序列平稳化与反平稳化方法,实现更具泛化性和精确性的预测.

2 通道融合与序列平稳化网络

为了解决长期序列预测中复杂的时间模式难以利用与计算效率低下的困难,在DLinear架构的启发下,本文提出了CFSNet时间序列预测模型,整体模型架构如图1所示.模型将分解作为内置操作引入序列处理的同时,还引入了通道融合编码CFSNBlock模块和序列平稳化与反平稳化模块.接下来将对时间序列问题的定义作出说明,并对模型每个模块进行介绍.

2.1 问题定义

时间序列 $\mathbf{x}^i = \{x_1^i, x_2^i, \dots, x_T^i\}$ 记录变量 i 在时间戳 T 处的

观测值,是一组按照时间先后顺序排列的数据点序列.在多元时间序列中,存在多个与时间相关的变量,每个时间步对应多个变量值.多元时间序列表示为 $\mathbf{X} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{d_x} \mid \mathbf{x}^{d_x} \in \mathbb{R}^{T_x}\}$, d_x 描述了变量的维度,即通道的个数,一般情况下,两个连续观测值之间的时间间隔是固定的.

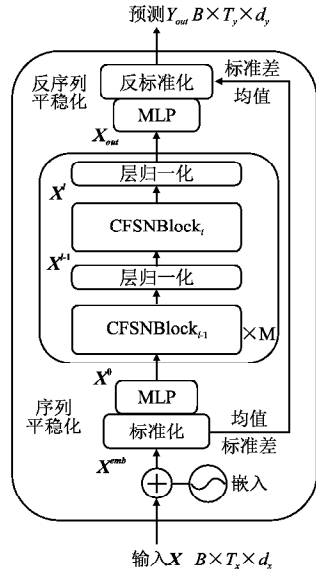


图1 CFSNet 模型框架

Fig. 1 CFSNet model framework

给定固定输入的时间窗口 T^x , 有输入 \mathbf{X} , 输出是相应的序列 $\mathbf{Y} = \{y^1, y^2, \dots, y^{d_y} \mid y^{d_y} \in \mathbb{R}^{T_y}\}$. 为了实现滚动预测, 本文的目标是基于 \mathbf{X} 预测 \mathbf{Y}_{T+h} . $h \in \mathbb{R}^+$ 是晚于当前时间戳的时间范围, T_y 是输出窗口, 其长度可能与 T^x 不同. 因此, 时间序列预测的定义是: 对于输入为 $\mathbf{X} = \{x_1, x_2, \dots, x_{T_x} \mid x_{T_x} \in \mathbb{R}^{d_x}\}$ 的矩阵, 预测目标为 $\mathbf{Y}_{T+h} \in \mathbb{R}^{d_y}$. 长期序列预测的目的是预测长期的未来, 即更大的 h .

2.2 通道融合编码

通道融合编码模块分为3部分: 序列分解、通道融合与线性聚合. 序列分解可以将原序列根据周期特点分解为两个分序列, 使特征更加精细从而易于下一步操作; 通道融合可以对分序列中的通道特征进行自适应挖掘; 线性聚合在时间复杂度低的同时对结果进行预测.

2.2.1 序列分解

每个时间点同时涉及与其相邻区域、不同时期之间相位的两种时间变化, 即周期内变化和周期间变化. 为了学习长期

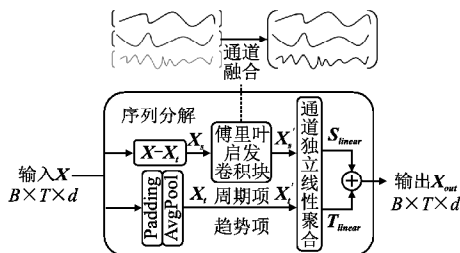


图2 CFSNBlock 结构

Fig. 2 Structure of CFSNBlock

预测任务下复杂的时间变化, 本文采用分解的思想^[17,25], 可

以将序列分解为趋势项部分和周期项部分.

由于未来的数据是未知的, 因此分解未来的序列是无法实现的. 为了达到近似的效果, 模型首先将序列进行分解, 作为 CFSNet 的预先处理 (见图2), 通过调整移动平均窗口逐步从时间序列中提取周期性波动以及长期趋势. 对于长度为 T^x 的输入序列 $\mathbf{X} = \{x_1, x_2, \dots, x_{T_x} \mid x_{T_x} \in \mathbb{R}^{d_x}\}$, 过程为:

$$\begin{aligned} X_t &= \text{AvgPool}(\text{Padding}(\mathbf{X})) \\ X_s &= \mathbf{X} - X_t \end{aligned} \quad (1)$$

其中 $X_t, X_s \in \mathbb{R}^{T_x \times d_x}$ 分别表示趋势部分和周期部分, 本文采用和 Autoformer 相同的 AvgPool(·) 进行移动平均, 窗口大小为 25, 并使用 Padding(·) 将时间序列沿时间维度用零进行填充操作以保持序列长度不变.

2.2.2 通道融合

不同通道之间的周期变化可能线性相关, 也可能不存在明显的关系. 因此, 如图3所示, 通道融合模块作为模型的主

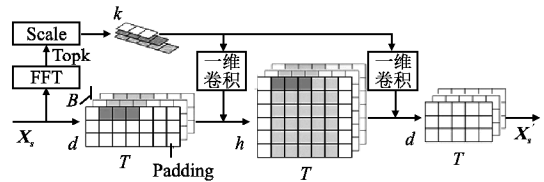


图3 傅里叶启发的通道融合卷积块结构

Fig. 3 Fourier-inspired channel fusion convolutional block structure

要操作, 使用不同的卷积核大小来捕获周期项中不同通道的特征. 根据通道之间的相互影响程度, 来自适应地训练出所需的卷积核, 过程为:

$$\begin{aligned} X'_s &= \text{Cove}(X_s) \\ X'_s &= X_s \end{aligned} \quad (2)$$

其中, X'_s, X'_t 表示对周期项和趋势项处理后的结果. 由于趋势项相对更加稳定, 在这里只对周期项进行通道融合编码. Cove(·) 操作包含两次步长为 1 的一维卷积, 为了保持卷积前后序列长度不变, 使用了 Padding 策略. 两次卷积之间的激活函数使用 GELU. 第 1 次卷积先将输入数据的通道从 d 维映射到隐藏层, 维度为 h , 接着第 2 次卷积又将通道从 h 维映射回 d 维, 以实现不同通道的信息融合, 在这里 n 的大小取决于数据集的维度. 在每一次卷积中, 利用快速傅里叶运算得到序列中普遍的周期信息, 卷积核大小等于序列中最明显的 k 个周期. 具体来说, 首先通过快速傅里叶变换得到频率及其对应的振幅, 接着选取振幅最大的 k 个频率, 最后根据频率计算出周期大小. 同时, 为了防止卷积核尺寸过大导致信息丢失, 本文通过放缩操作将卷积核尺寸限制在一定的范围内, 并且为了方便 Padding 操作, 卷积核大小需要为奇数. 过程如下:

$$\begin{aligned} A &= \text{Amp}(\text{FFT}(X'_s)) \\ \{f_1, \dots, f_k\} &= \text{argTopk}(A) \\ ks_i &= \lceil \frac{L}{f_i} \rceil, i \in \{1, 2, \dots, k\} \\ ks'_i &= \text{Scale}(ks_i) \end{aligned} \quad (3)$$

其中, FFT(·) 表示快速傅里叶变换, Amp(·) 代表计算幅度值, f_i 代表排行前 k 的幅度值代表的序列频率, L 代表序列长

度, ks' 为放缩得到的卷积核大小. 这些卷积核经过挑选, 可以更加准确地对有关依赖关系的通道特征进行融合.

2.2.3 线性聚合

最后, 本文使用两个线性层分别作用于 X'_s, X'_t , 并将其结果相加以获得最终预测 X_{out} . 过程为:

$$\begin{aligned} S_{linear}, T_{linear} &= \text{Linear}(X'_s), \text{Linear}(X'_t) \\ X_{out} &= S_{linear} + T_{linear} \end{aligned} \quad (4)$$

2.3 CFSNBlock 与序列平稳化、反平稳化模块

如图 1 所示, 本文将上述通道融合编码模块整合进 CFSNBlock, 并对输入序列分别进行平稳化、嵌入、多个 CFSNBlock 处理、反平稳化操作.

对于长度为 T_x 的时间序列 $X \in \mathbb{R}^{T_x \times d_x}$, 通过嵌入层得到 X^{emb} , 其中嵌入操作包括时间嵌入和位置嵌入:

$$X^{emb} = \text{Embed}(X) \quad (5)$$

本文使用 CFSNBlock 代表一个通道融合编码模块, 将 M 个 CFSNBlock 块堆叠起来, 在更加平稳地捕获周期项和趋势项的同时, 提取深层次的特征. 每经过一次堆叠就进行层归一化, 对于 CFSNet 第 l 层, 该过程可以形式化为:

$$X^l = \text{CFSNBlock}_l(X^{l-1}) \quad (6)$$

另外, 受 Stationary 模型中标准化与反标准化化模块的启发, 本文使用了改进的平稳化与反平稳化模块, 在原来的基础上, 增加了线性层. 其中, 平稳化模块可以使输入序列更加平稳, 通过一个滑动窗口, 实现在时间维度上的平稳化. 对于输入序列 $X = \{x_1, x_2, \dots, x_{T_x} | x_{T_x} \in \mathbb{R}^{d_x}\}$, 平稳化模块操作如下:

$$\begin{aligned} \mu_x &= \frac{1}{T_x} \sum_{i=1}^{T_x} x_i, \sigma_x^2 = \frac{1}{T_x} \sum_{i=1}^{T_x} (x_i - \mu_x)^2, x'_i = \frac{1}{\sigma_x} \odot (x_i - \mu_x) \\ X'' &= \text{MLP}(X') \end{aligned} \quad (7)$$

其中, $\mu_x, \sigma_x \in \mathbb{R}^{d_x \times 1}$ 分别代表标准差与均值, $\frac{1}{\sigma_x}$ 代表逐元素除法以及 \odot 代表逐元素乘积, X'' 代表最终平稳化的结果. 通过平稳化模块, 每个输入的时间序列之间的分布差异减小, 因此模型的输入更加稳定. 本文使用 $\text{Station}(\cdot)$ 表示对 X 的每个元素进行平稳化, 平稳化模块作用于模型最开始的输入:

$$X^0 = \text{Station}(X^{emb}) \quad (8)$$

最后, 对序列输出用反平稳化模块进行处理, 作为最终的预测结果, 可以使输出序列符合原始序列的特征. 反平稳化模块过程如下:

$$\begin{aligned} X_{out} &= \text{CFSNBlock}_M(X^{M-1}) \\ Y &= \text{MLP}(X_{out}) \\ y_i^{out} &= \sigma_x \odot (y_i + \mu_x) \end{aligned} \quad (9)$$

此处, X_{out} 是最后一个 CFSNBlock 的输出. 反平稳化模块由线性层和反标准化操作组成. 在这里, Y 和 X 都拥有相同的通道数 C , 即 $d_x = d_y$.

通过平稳化与反平稳化两个阶段, 模型可以接收平稳化的输入, 这种平稳化会增强模型的泛化能力, 从而有利于现实世界的序列预测.

3 实验

3.1 实验设置

本文在 Electricity、Weather、ETTm2 3 个广泛使用的时间

序列数据集上进行了实验, 它们都是多元时间序列. 其中 Electricity 数据集包含 2012 年 ~ 2014 年 321 个客户的每小时用电量; Weather 数据集是 2020 年全年每 10 分钟记录一次气象数据, 包括气温、湿度等指标. ETTm2 数据集包含从电力变压器中每 15 分钟记录一次的 2016 年 7 月 ~ 2018 年 7 月期间的负载和油温. 表 1 给出了 3 个数据集的概况.

表 1 3 个通用数据集内容一览

Table 1 Overview of three general datasets

数据集	通道数量	采样频率	序列长度
Electricity	321	1 小时	26304
Weather	21	10 分钟	52695
ETTm2	7	15 分钟	69681

本文使用均方误差 (Mean Squared Error, MSE) 和平均绝对值误差 (Mean Absolute Error, MAE) 作为比较性能的核心指标, 其计算公式如下:

$$\begin{aligned} MSE &= \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \\ MAE &= \frac{1}{m} \sum_{i=1}^m |\hat{y}_i - y_i| \end{aligned} \quad (10)$$

其中, \hat{y}_i 为预测值, y_i 为真实值, m 表示测试集的样本总数. MSE 与 MAE 的值越小, 代表预测值和真实值越接近, 模型的预测准确度越高.

表 2 实验环境

Table 2 Experimental environment

实验环境	环境配置
操作系统	Linux 3.10.0
处理器	Intel(R) Xeon(R) Gold 5218R
显卡	P100-SXM2-16GB
编程语言	Python3.8
深度学习框架	Pytorch1.7.1

本文使用深度学习框架 Pytorch 进行实验, 采用 GPU 加速, 具体环境配置如表 2 所示. 模型使用批量梯度下降法进行训练, 数据分批次送入模型, 批量样本大小 B 设置为 32, 选用 Adam 优化器和 MSE 损失函数, 初始学习率设置为 0.1. 在傅里叶启发的通道融合卷积模块中, 模型的隐藏层维度 h 默认设置为 32. 训练轮次设置为 20, 并使用早停策略, 当模型连续 3 次在验证集上的损失值未获得改善, 则提前停止训练.

3.2 对比实验与结果分析

本文使用了 5 种方法作为对比, 其中包括一种新的基于线性模型的方法、一种新的基于 CNN 的方法、4 种新的基于 Transformer 的方法与一种传统方法, 各种方法介绍如下:

DLinear^[16]: 该模型使用简单的经过设计的线性模型可以获得更快的计算速度和更准确的结果;

MICN^[23]: 使用串行的卷积结构捕获数据中的全局-局部信息;

FEDformer^[13]: 采用专家混合设计分解周期项与趋势项, 并在频域进行注意力计算.

Autoformer^[11]: 在 Transformer 模型基础上, 结合序列分解的思想来更加准确地提取时序中的周期变化;

Informer^[14]: 通过使用蒸馏的方法降低 Transfoermer 的时

间和空间复杂度;

Pyraformer^[15]引入了低复杂度的金字塔注意机制,用于长时间序列数据的建模与预测.

最后,本文使用 LSTM 作为一个简单的基线.

为了公平比较不同模型的性能,本文使用相同的输入长度 96,并选取了 4 种预测长度:96,192,336,720.

多变量预测结果如表 3 所示.在上述设置中,针对单步预测设计的 LSTM 的效果最差,是由于其太过简单,在长期序列依赖捕获方面的能力不足,不适用于多步预测.本文的模型

超过了基线中所有基于 Transformer 的模型,在基于 Transformer 的模型中,Autoformer 由于其周期提取能力和自相关机制可以更好地处理具有周期性的数据,取得了最好的效果.

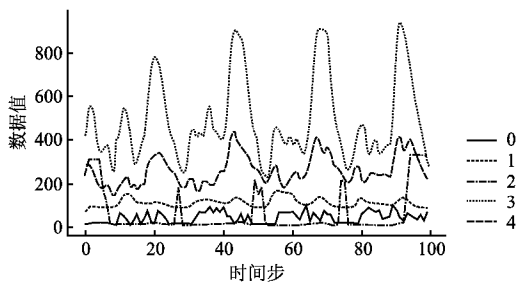
本文模型与 DLinear 相比,在 Electricity 与 Weather 数据集上的结果十分接近,这是因为这两个数据集单个通道的曲线变化特征易于捕捉.在 Electricity 数据集中,每个通道代表一名用户的用电数据,图 4(a)选取了其中 5 名用户的数据,可以看出,尽管用户之间的周期波动具有相似性,但是单个用户的用电数据也是十分规律的,因此只要精心设计,通道独立

表 3 多变量不同预测长度结果

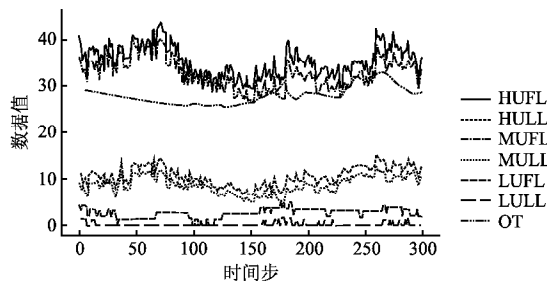
Table 3 Results for different prediction lengths in multivariate setting

模型	指标	Electricity				Weather				ETTm2			
		96	192	336	720	96	192	336	720	96	192	336	720
CFSNet	MSE	0.183	0.196	0.211	0.241	0.165	0.238	0.286	0.359	0.185	0.257	0.342	0.426
	MAE	0.288	0.298	0.311	0.333	0.214	0.276	0.308	0.355	0.271	0.316	0.369	0.413
DLinear	MSE	0.195	0.197	0.209	0.245	0.196	0.237	0.283	0.345	0.193	0.284	0.369	0.554
	MAE	0.282	0.285	0.301	0.334	0.255	0.296	0.335	0.381	0.292	0.362	0.427	0.522
MICN	MSE	0.207	0.206	0.219	0.252	0.195	0.304	0.315	0.695	0.194	0.401	0.320	0.502
	MAE	0.298	0.299	0.314	0.343	0.255	0.363	0.372	0.569	0.291	0.423	0.376	0.486
FEDformer	MSE	0.193	0.201	0.214	0.246	0.217	0.276	0.339	0.403	0.203	0.269	0.325	0.427
	MAE	0.308	0.315	0.329	0.355	0.296	0.336	0.380	0.428	0.287	0.328	0.371	0.415
Autoformer *	MSE	0.201	0.222	0.231	0.254	0.266	0.307	0.359	0.419	0.255	0.281	0.339	0.433
	MAE	0.317	0.334	0.338	0.361	0.336	0.367	0.395	0.428	0.339	0.340	0.372	0.432
Informer *	MSE	0.274	0.296	0.300	0.373	0.300	0.598	0.578	1.059	0.365	0.533	1.363	3.379
	MAE	0.368	0.386	0.394	0.439	0.384	0.544	0.523	0.741	0.453	0.563	0.887	1.338
Pyraformer	MSE	0.386	0.378	0.376	0.376	0.622	0.739	1.004	1.420	0.435	0.730	1.201	3.625
	MAE	0.449	0.443	0.443	0.445	0.556	0.624	0.753	0.934	0.507	0.673	0.845	1.451
LSTM *	MSE	0.375	0.442	0.439	0.980	0.369	0.416	0.455	0.535	2.041	2.249	2.568	2.720
	MAE	0.437	0.473	0.473	0.814	0.406	0.435	0.454	0.520	1.073	1.112	1.238	1.287

注: * 表示数据来源于 Autoformer^[11]的原始论文



(a) Electricity数据集5名用户用电曲线



(b) ETTm2数据集7个通道数据曲线

图 4 Electricity 与 ETTm2 数据集中不同通道数据曲线

Fig. 4 Different channel data curves in Electricity and ETTm2 datasets

策略也可以准确预测用户未来用电数据,达到与通道融合策

略近似的效果.模型在 ETTm2 数据集上表现最好,从图 4(b)可以看出,该数据集中一些通道的数据曲线周期十分相似,但是单个通道的变化规律不够明显,因此通过通道融合-独立策略,可以使模型综合相似通道的变化特征,进行更准确的预测.

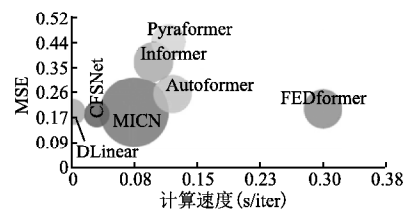


图 5 模型效率比较

Fig. 5 Model efficiency comparison

本文还从预测性能、训练速度和参数量 3 个方面综合评估模型的效率.图 5 显示了在具有代表性的 ETTm2 数据集上,设置输入为 96,预测长度为 96 的情况下的各模型效率.其中圆的半径越大,训练所需参数量越大,圆心越靠近原点,代表模型训练越快、预测越准确.与预测准确率较高的 MICN、Autoformer、FEDformer 相比,本文模型在 ETTm2 数据集中分别节省了 57.7%、73.4%、89.2% 的训练时间,而仅占用其 0.4%、0.9%、0.9% 的参数量,使预测性能提升了 4%、37.9%、24.3%.与 DLinear 模型相比,CFSNet 的计算速度略

有降低,参数量有所增加,但是 MSE 改善了 4%.

总体而言,CFNet 在平均性能上超过其他基线模型,它不仅相对于基于 Transformer 的模型在计算速度上有优势,其通道融合-独立的策略也使它比基于线性结构的通道独立模型更加准确.同时,模型还能适应数据集的特点,在不同领域的数据集集中表现出良好的泛化性.

3.3 消融实验

为了验证通道融合结构和平稳化与反平稳化结构的有效性,本文在保持输入序列长度不变的情况下,在初始通道独立的线性模型基础上,分别增加传统的标准化与反标准化模块、改进的平稳化与反平稳化模块、傅里叶启发的卷积模块进行消融实验,同时研究了卷积核选取策略对预测效果的影响,以表明这些模块的重要性.

通道融合结构的有效性:在初始模型的基础上,当增加了卷积结构后,模型可以进行通道融合,使预测性能得到提升,具体结果见表 4.这是因为在初始模型的通道独立策略下,模型只是执行多个单序列预测,通道之间无法进行信息交流.而引入傅里叶启发的通道融合卷积模块后,通道之间能够相互传递信息,发掘出通道之间的相关性.这表明模型中的 CFNet-Block 可以自适应地提取各通道的信息,有效地关联数据中具有相关性的通道特征.

表 4 消融实验结果

Table 4 Model ablation experiment results

策略	指标	ETTM2		Weather	
		96	192	96	192
初始	MSE	1.695	1.732	0.274	0.301
	MAE	1.149	1.161	0.365	0.380
+ 标准化	MSE	0.199	0.261	0.198	0.241
	MAE	0.281	0.318	0.246	0.280
+ 平稳化	MSE	0.201	0.260	0.189	0.240
	MAE	0.284	0.315	0.236	0.280
+ 卷积	MSE	1.491	1.910	0.210	0.277
	MAE	1.029	1.090	0.295	0.339
CFNet	MSE	0.185	0.257	0.165	0.238
	MAE	0.271	0.316	0.214	0.276

注: + 代表在初始模型上增加相应的模块

为了证实根据快速傅里叶变换选取 k 个卷积核的有效性,本文研究了卷积核大小的选取策略对预测效果的影响.表 5 展示了模型在数据集 ETTm2 中的 3 种选取策略下的预测

表 5 卷积核的 3 种选取策略及其性能

Table 5 Three strategies for convolutional kernel selection and their performance

策略	指标	96	192	336	720
FFT + Scale	MSE	0.185	0.257	0.342	0.425
	MAE	0.271	0.316	0.369	0.413
FFT	MSE	0.193	0.262	0.329	0.435
	MAE	0.277	0.321	0.357	0.420
传统方式	MSE	0.191	0.259	0.326	0.426
	MAE	0.273	0.318	0.360	0.415

效果.其中,FFT 代表使用快速傅里叶变换计算得到 k 个周期作为卷积核尺寸,FFT + Scale 代表将卷积核尺寸放缩到小于 30 步长的大小.在传统方式下,卷积核被依次指定为 1, 3, 5... 从表中可以得知,仅使用 FFT 策略得到的卷积核不如传统方式,这是因为有可能选取到较大的周期,使卷积核也变

大,导致 Padding 操作时引入大量的空白,同时感受野过大也会造成信息损失.当进行放缩后,卷积核被限制在适当的尺寸,相比传统方式,更容易捕获时间序列中的周期信息.

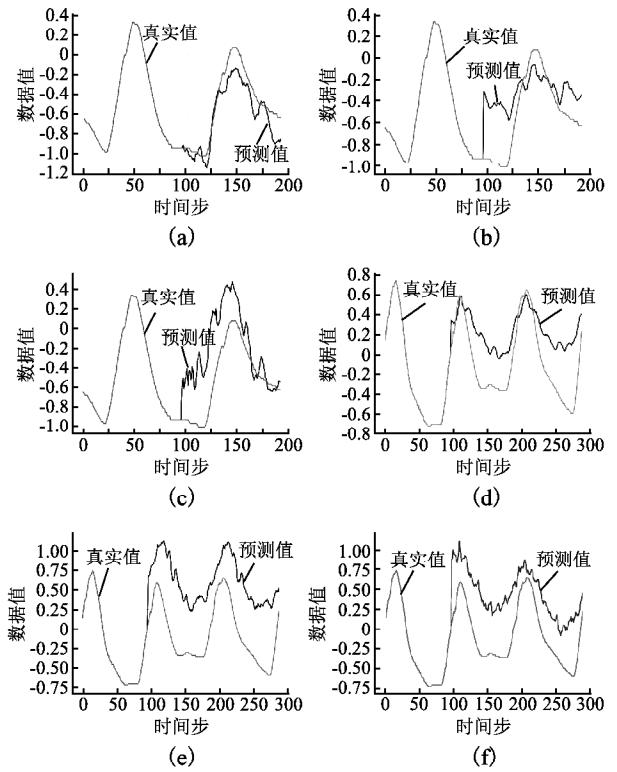


图 6 数据集 ETTm2 中平稳化与反平稳化模块有效性的对比

Fig. 6 Comparison of the effectiveness of stabilization and destabilization modules in ETTm2 dataset

平稳化与反平稳化结构的有效性:在图 6 中,截取了 ETTm2 测试集中的两段数据作为输入序列.图 6(a)、图 6(b)和图 6(c)代表了预测长度为 96 的情况,图 6(d)、图 6(e)和图 6(f)代表了预测长度为 192 的情况.图 6(a)和图 6(d)表示原始模型的预测结果,图 6(b)和图 6(e)代表移除了反平稳化结构的预测结果,图 6(c)和图 6(f)是移除了平稳化和反平稳化结构后的预测结果.从图中可以观察到,在原始模型中,预测曲线与输入序列有相似的形状,然而去除反平稳化结构后,预测曲线的准确度显著下降,虽然在周期上有一定相似性,但是在趋势上与真实情况有着很大的偏差,这是因为移除了反平稳化结构后,预测曲线没有得到修正,不能有效遵循输入序列的分布.而接着去除了平稳化结构后,模型的预测曲线的波动更加剧烈,这是因为输入序列平稳性被打破,模型学习相关特征将会更加费力,泛化能力降低.根据表 4 可知,初始模型增加了传统的标准化与反标准化模块后,模型预测能力得到提高,在 ETTm2 数据集中表现得尤为明显.使用了改进的平稳化与反平稳化模块后,平均性能比传统方法提升了 1%.以上实验表明平稳化结构可以使输入序列平稳,使模型更加高效地发现规律,捕获所需信息;反平稳结构可以使预测曲线遵循输入序列的模式.

3.4 超参数敏感度

本文在公式(3)中引入了超参数 k 来选择卷积核的个

数,在公式(9)中引入了超参数 M 来确定 CSFNBlock 堆叠的次数.图7是对这两个超参数进行敏感性分析的结果,选用的数据集是ETTm2.从图中可以发现,模型在不同 k 和 M 的选

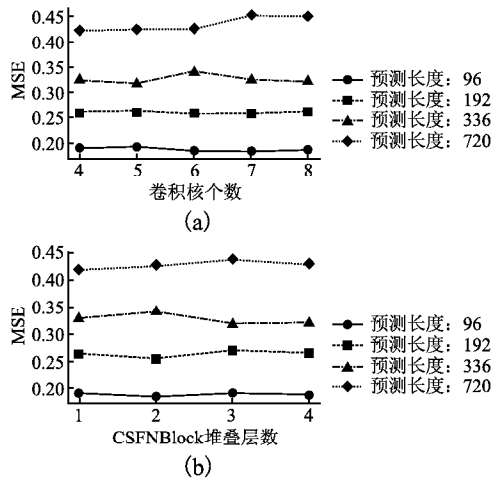


图7 卷积核个数与模型堆叠层数对预测性能的影响
Fig.7 Impact of the number of convolutional kernels and the stacking layers on prediction performance

择下都可以呈现较为稳定的性能.随着 k 的增加,模型对周期依赖的捕获性能逐渐增强,但这种增强越来越不明显.出于对性能和效率的考量,最终模型选择 $k=6, M=2$.

4 总结与展望

本文针对已有工作中难以捕获复杂多变的时间模式以及基于 Transformer 的模型时间复杂度高的困难,提出了一种新的长期时间序列预测模型 CFSNet. CFSNet 可以通过线性架构提取复杂的时间变化,同时能够利用卷积融合不同通道的周期信息.另外,平稳化与反平稳化策略使来自不同数据集的输入更加平稳,进而提高模型的泛化能力.在 Electricity, Weather, ETTm2 数据集上的实验结果表明,本文提出的模型相比基准模型能够提供更加精准的预测结果,并具有良好的泛化性能.

尽管该研究在利用深度学习技术进行时间序列预测的研究上取得了一定进展,未来还可以在可解释性方面进行更深入的研究,比较不同可解释方法的有效性.同时,可以进一步研究多通道融合的新方法,以获得更加稳定和准确的预测.

References:

[1] Lim B, Zohren S. Time-series forecasting with deep learning: a survey[J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209, doi:10.1098/rsta.2020.0209.

[2] Rangapuram S S, Seeger M, Gasthaus J, et al. Deep state space models for time series forecasting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), 2018: 7796-7805.

[3] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181-1191.

[4] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:180301271, 2018, doi:10.48550/arXiv.1803.01271.

[5] Liu M, Zeng A, Chen M, et al. SCINet: time series modeling and forecasting with sample convolution and interaction[J]. Advances in Neural Information Processing Systems, 2022, 35: 5816-5828, doi:10.48550/arXiv.2106.09305.

[6] He Y, Zhao J. Temporal convolutional networks for anomaly detection in time series[C]//Proceedings of the Journal of Physics: Conference

Series, 2019: 042050, doi:10.1088/1742-6596/1213/4/042050.

[7] KANG R X, NIU B N, LI X, et al. Predicting stock pieces using LSTM with the self-attention mechanism and multi-source data[J]. Journal of Chinese Computer Systems, 2023, 44(2): 326-333.

[8] LI Z X, LIU H Y. Combing global and sequential patterns for multivariate times series forecasting[J]. Chinese Journal of Computers, 2023, 46(1): 70-84.

[9] LI W, CHEN J W, LIU R X, et al. T-Transformer model for predicting tensor time series[J]. Computer Engineering and Applications, 2023, 59(11): 57-62.

[10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30, doi:10.1109/ICASSP39728.2021.9413901.

[11] 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 (NIPS), 2021: 22419-22430.

[12] Li S, Jin X, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS), 2019: 5243-5253.

[13] Zhou T, Ma Z, Wen Q, et al. Fedformer: frequency enhanced decomposed transformer for long-term series forecasting[C]//Proceedings of the International Conference on Machine Learning, 2022: 27268-27286.

[14] 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.

[15] Liu S, Yu H, Liao C, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2021, doi:10.34726/2945.

[16] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting? [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 11121-11128.

[17] Anderson O D. Time-series[J]. Journal of the Royal Statistical Society Series D: The Statistician, 2018, 25(4): 308-310.

[18] Hyndman R J, Athanasopoulos G. Forecasting: principles and practice[M]. Melbourne: OTexts, 2018.

[19] Taylor S J, Letham B. Forecasting at scale[J]. The American Statistician, 2018, 72(1): 37-45.

[20] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[21] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches[C]//Proceedings of the SSST-8, 8th Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014: 103-111.

[22] Franceschi J Y, Dieuleveut A, Jaggi M. Unsupervised scalable representation learning for multivariate time series[C]//Proceedings of the 33rd Conference on Neural Information Processing Systems (NIPS), 2019: 4650-4661.

[23] 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 Eleventh International Conference on Learning Representations (ICLR), 2023.

[24] Oreshkin B N, Carpov D, Chapados N, et al. N-BEATS: neural basis expansion analysis for interpretable time series forecasting[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2019, doi:10.48550/arXiv.1905.10437.

[25] Cleveland R B, Cleveland W S, Mcrae J E, et al. STL: a seasonal-trend decomposition[J]. Journal of Official Statistics, 1990, 6(1): 3-73.

附中文参考文献:

[7] 康瑞雪, 牛保宁, 李显, 等. 融合多源数据输入具有自注意力机制的 LSTM 股票价格预测[J]. 小型微型计算机系统, 2023, 44(2): 326-333.

[8] 李兆玺, 刘红岩. 融合全局和序列特征的多变量时间序列预测方法[J]. 计算机学报, 2023, 46(1): 70-84.

[9] 李文, 陈佳伟, 刘瑞雪, 等. 张量时间序列预测 T-Transformer 模型[J]. 计算机工程与应用, 2023, 59(11): 57-62.