

融合通道注意力与SDFFT的同质多变量时序预测模型

王凯欣, 杨磊, 赵品杰, 江宇涵, 李康顺, 刘启健

(华南农业大学数学与信息学院, 广东 广州 510642)

摘要: 在同质多变量时间序列长期预测任务中, 现有模型大多存在着特征提取时未充分利用变量间内在关联信息、对数据的时效性考虑不足、季节性和随机波动信息提取不到位等问题。为此, 提出了一种融合多阶段跨时间依赖和拐点信息的同质多变量时间序列预测模型(HMTS-TI), 设计了RGCA模块, 在变量间通道注意力中引入分组排序的策略, 将学习到的关联信息转化为注意权重, 增强对多变量时间序列变量间相关信息的提取; 构建了SDFFT模块, 通过子序列分解与快速傅里叶变换, 挖掘长期趋势、短期季节模式及随机波动信息, 加强跨时间维度依赖特征提取; 将拐点数量统计值作为预测调整因子, 提取不同时间序列变量中的波动信息, 对模型拟合产生正向调节作用。在Traffic、Solar、PeMS04 3个公共同质多变量时间序列数据集上的实验结果表明, 相较于对比实验中的最优模型, HMTS-TI在96、192、336和720共4个预测长度上整体均方误差、平均绝对误差分别下降了12.1%和6.8%, 决定系数提升了4.1%, 具有较高的预测准确率。此外, 在具体应用案例中, 该模型对广东省内92个观测点的臭氧数据预测长度为336个时间单位任务上的均方误差相比最优对比模型下降了19.0%。

关键词: 同质多变量时间序列; 长期预测; 通道注意力; 子序列划分; 数据分解; 快速傅里叶变换

DOI: 10.11907/rjdk.241803

扫描二维码阅读全文:



中图分类号: TP18

文献标识码: A

文章编号: 1672-7800(2026)001-0083-12

Homogeneous Multivariate Time Series Prediction Model Integrating Channel Attention and SDFFT

WANG Kaixin, YANG Lei, ZHAO Pinjie, JIANG Yuhuan, LI Kangshun, LIU Qijian

(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: For long-term prediction of homogeneous multivariate time series, existing models often fail to fully utilize the inherent correlations between variables during feature extraction, adequately consider the timeliness of the data, and effectively extract seasonal and random fluctuation information. In order to improve upon these deficiencies, we propose the Homogeneous Multivariate Time Series Prediction Model Incorporating Multi-stage Temporal Dependencies and Inflection Point Information (HMTS-TI). An RGCA module implements a ranked grouping strategy within inter-variable attention mechanisms, transforming insights on correlations into weight assignments, thereby enhancing the extraction of inter-variable correlation information within multivariate time series. SDFFT module extracts trends, seasonal patterns, and fluctuations via subsequence decomposition and fast Fourier transform, reinforcing cross-temporal dependency feature extraction. Inflection point statistics serve as a predictive adjustment factor, enhancing model fitting by incorporating information on variable-specific fluctuations. The experimental outcomes on three public homogeneous multivariate time series datasets—Traffic, Solar, and PeMS04—demonstrate that, in comparison to the best model in the comparative experiments, HMTS-TI achieves an overall reduction in mean squared error and mean absolute error.

收稿日期: 2024-11-03

基金项目: 广东省自然科学基金项目(2020A1515010691); 广东省科技创新专项资金项目(SDZX2023032)

作者简介: 王凯欣(1999-), 女, CCF会员, 华南农业大学数学与信息学院硕士研究生, 研究方向为数据挖掘、时间序列预测; 杨磊(1978-), 男, 华南农业大学数学与信息学院讲师、硕士生导师, 研究方向为数据挖掘、智能计算; 赵品杰(2000-), 男, 华南农业大学数学与信息学院硕士研究生, 研究方向为数据挖掘、时间序列预测; 江宇涵(2000-), 男, 华南农业大学数学与信息学院硕士研究生, 研究方向为数据挖掘、时间序列预测; 李康顺(1962-), 男, 华南农业大学数学与信息学院教授、博士生导师, 研究方向为数据挖掘、人工智能; 刘启健(2003-), 男, 华南农业大学数学与信息学院学生, 研究方向为数据挖掘、时间序列预测。本文通讯作者: 杨磊。

ror by 12.1% and 6.8%, respectively, across four prediction horizons of 96, 192, 336, and 720. Concurrently, the coefficient of determination is augmented by 4.1%, validating the model's high predictive accuracy. In addition, in the specific application case, the model achieved a mean squared error reduction of 19.0% compared to the best-performing benchmark model in the task of predicting ozone data from 92 observation points in Guangdong Province over a length of 336 time units.

Key Words: homogeneous multivariate time series; long-term prediction; channel attention; subsequence segmentation; data decomposition; fast Fourier transform

0 引言

多变量时间序列(Multivariate Time Series, MTS)由一系列单变量序列构成,每个单变量时间序列都是在相同的时间点上记录的一系列观察值,承载着特定现象随时间演进的轨迹,蕴含极为丰富的信息。这样的数据集中,每个时间点都涉及多个变量的观察值,通常包括变量、时间点、观测数值等要素。这些序列数据来源十分广泛,覆盖多个关键领域。例如:能源管理领域,电网中多个节点的实时功率输出及电压、电流等数据^[1];农业管理领域,农田中部署的多个传感器收集的土壤湿度、温度、光照强度等数据;交通运输领域,高速公路上不同传感器测量的道路占用率、车流量数据^[2];环境监测领域,空气中氮氧化物、二氧化硫、臭氧、一氧化碳等污染物监测数据等。

上述应用场景中的多变量时间序列数据具有共同特性,即来源于相同的数据生成过程(Data Generating Processes, DGP),Hewamalage 等^[3]将这种特性称为同质性(Homogeneity),并认为数据集内的各项数据随着时间的变化表现出一致的行为模式,并且它们之间的比较是合理且有意义的,各变量存在着协同变化或相互制约等联系。

从变量要素角度看,传统的时间序列预测方法大多是将每个时间序列视为独立的数据集,并单独进行预测,如指数平滑方法、自回归移动平均模型。这些方法未能充分利用跨变量学习模式的巨大潜力,即使如近年来较为热门的TCN^[4]、Transformer^[5]、TimesNet^[6]等深度学习预测模型对所有序列使用相同的参数集,但仍未考虑对变量间的关系进行建模。

时间序列的跨变量依赖性是指在一个多变量时间序列数据集中,不同变量之间存在的相互依赖关系。具体如:一个变量在某个时间点上的值可能依赖于另一个变量在过去某个时间点上的值;多个变量可能遵循相似的趋势或周期性模式。这种依赖性表明,一个变量在某个时间点上的值可能受到另一个或多个变量在同一时间点或先前时间点上值的影响。对于各变量存在着协同变化或相互制约等高度相关性的同质多变量时间序列而言,降低其在变量维度的冗余性,有效地提取不同时间序列变量间的依赖性具有积极意义。

利用时间序列的跨变量依赖性进行跨变量学习模式具有很大潜力。首先,通过充分利用信息共享,可以分析多个序列之间的相关性,将一个序列中的信息用于改善另

一个序列的预测。其次,跨变量学习能够增强模式识别,识别出多个序列之间的共同模式或趋势,从而有助于更好地理解系统的整体行为。此外,利用序列间的相互作用能够加强预测能力,提高预测稳健性和准确性,尤其是在处理非平稳或复杂动态的情况下。同时,这种方法有助于减少过拟合风险,通过整合来自多个序列的信息,可以在一定程度上缓解因单一序列数据不足而导致的过拟合问题。最后,跨序列学习能够提升模型可解释性,在某些情况下,它有助于揭示变量间的真实关系,从而提供更有意义的预测结果解释。

从时间要素角度看,根据指数平滑法理论,不同时间段的数据具有不同的参考价值^[7]。在LSTM中同样通过设置遗忘机制,根据当前输入和先前状态决定哪些旧的信息不再重要,可以被遗忘,从而有效地忽略远期数据中的不相关信息,专注于近期更有意义的信息^[8]。基于这些理论,模型在预测未来短期波动时,近期数据往往比远期数据更具参考意义,而在预测长期趋势时,则需全面考虑整个序列的变化特征。但目前较为流行的基于MLP或者Transformer的多变量时间序列预测模型大多忽视了这个因素。

此外,现有方法的鲁棒性和泛化能力通常存在一些问题。比如:当时间序列数据包含复杂的季节性模式和长期趋势时,单一模型难以正确地识别和处理这些成分,既可能错误预测出季节性的峰值或谷底,导致预测误差显著增加,也可能导致异常点或突变被其他成分掩盖;序列中的随机波动增加了数据的噪声水平,可能导致模型在处理含有大量随机波动的数据时,为了减少噪声影响而变得过于保守,无法充分拟合数据中的真实模式。

基于上述分析,本文从同质多变量时间序列的长期预测研究入手,提出了一种融合多阶段跨时间依赖和拐点信息的同质多变量时间序列预测模型(Homogeneous Multivariate Time Series Prediction Model Incorporating Multi-stage Temporal Dependencies and Inflection Point Information, HMTS-TI)。在时间维度上捕捉数据的时序特性,而在多变量层面挖掘序列间的内在联系,并将时间序列拐点数量作为预测结果的调整因子,通过对序列转折点的统计分析,提炼序列波动与变化趋势的关键信息,进而优化预测结果,有效提升模型预测精度。

本文主要贡献如下:①针对同质多变量时间序列数据提出HMTS-TI模型用于长期预测研究,设计基于排序分组的通道注意力(Channel Attention Based on Ranked Group-

ing, RGCA), 引入排序分组策略, 基于变量间的通道注意力学习跨变量维度的依赖, 以提取更加全面的跨变量信息并用于预测; ②构造基于子序列分解和快速傅里叶变换的特征提取方法 (Subsequence Decomposition and Fast Fourier Transform Feature Extraction Module, SDFFT), 子序列划分帮助模型从不同输入长度上捕获数据的动态变化, 同时融合数据时效信息, 通过数据分解和傅里叶变换深入地挖掘子序列中的季节与随机波动成分的隐藏信息, 实现跨时间依赖信息的学习; ③引入时间序列统计域中的拐点数量参数, 通过神经网络提取不同时间序列变量中的波动信息, 作为调整预测结果的一个附加值, 使之对模型拟合产生正向调节作用, 并最终提高预测精度。

1 相关工作

在多变量时间序列预测领域, 学者们提出了一些较具代表性的模型。例如: Vaswani 等^[5]提出 Transformer, 基于注意力机制和 Seq2Seq 架构实现时间序列的长期预测; Wu 等^[9]提出 Autoformer, 设计了基于序列周期性的自相关机制, 在子序列层面进行相关性发现和表示聚合; Wang 等^[10]提出 MICN, 结合局部特征和全局相关性以捕获多变量时间序列隐藏的时序信息。尽管这些模型在多变量长时间序列预测中展现出了良好性能, 但构建模型的焦点集中在跨时间维度依赖关系上, 忽视了极为重要的跨变量依赖关系, 仅通过在同一时间点将所有维度的数据投影到一个特征向量中加以粗略处理, 未能深入挖掘并有效利用各变量间的内在关联信息。很多研究者注意到了这些问题, 并以此为改进方向。对于如何处理好跨变量依赖关系, 其主要关注点在于对变量中同质化特征的捕捉、分析及利用等相关问题。Hewamalage 等^[3]引入了同质多变量的概念, 其具体应用领域中较具代表性的研究有: Qi 等^[11]提出 GC-LSTM, 通过预测不同站点以预测 PM2.5 浓度变化; Wu 等^[12]利用图嵌入技术和马氏距离度量进行加利福尼亚近海风力农场多站点的风速预测; Li 等^[13]提出了一个信息融合图和并行 TCN 以进行交通流量预测等。在对同质化特征建模策略的研究方面, 具有代表性的研究有: Crossformer 设计了两阶段注意力, 利用变量间的相关性构建模型^[14]; MTS-Mixer 采用因子分解策略以降低变量间冗余, 从而捕捉变量间的相关信息^[15]。分析已有文献可知, 同质多变量研究成果主要体现在垂直化方向的应用领域。目前存在的主要问题体现在两方面: 一是更多体现在方法和技巧层面, 还未能形成系统性的总结、归纳; 二是应用领域虽然较为广泛, 但可泛化应用且性能较好的预测模型样本仍不多见。

此外, 现有模型对于沿时间维度信息的捕捉, 大多情况下未能充分考虑到数据的时效性, 倾向于对整个输入序列进行统一的信息提取。例如: Zhou 等^[16]提出 Informer,

采用 ProbSparse 自注意力机制和生成式解码器的结构, 全局一致地处理超长输入的多变量时间序列。Zeng 等^[17]提出 DLinear, 采用两个单层线性网络对输入到模型中的完整多变量时间序列进行分解并预测其趋势和残差成分, 这种做法未能平衡数据的即时性与长期趋势, 导致在预测过程中可能忽略掉重要的短期波动细节。而在长期预测任务中, 同时捕捉短期波动与长期演变规律, 整合数据局部与全局信息, 对于提高预测精度具有极其重要的意义。

同时, 时间序列的标准差、信息熵等一些统计域特征难以直接通过模型训练提取, 需要人为进行设置引导。Liu 等^[18]设计了 Non-stationary Transformer, 通过学习序列的标准差提取时间序列中的非平稳特征; Wang 等^[19]提出了一种基于多元离散熵的方法, 以探究股票序列的结构特征。受其启发, 本文在对同质多变量时间序列数据集建模研究时, 引入时间序列统计域中的拐点数量参数, 学习序列波动程度, 以提高模型预测准确性。

2 本文模型

2.1 整体模型

模型 HMTS-TI 的整体框架如图 1 所示。HMTS-TI 由 3 个核心模块组成: 跨时间信息提取模块 SDFFT、跨变量信息提取模块 RGCA, 以及序列拐点信息提取模块。SDFFT 将输入序列切分为不同长度的子序列, 利用滑动窗口技术分解为趋势性分量和季节性及残差分量。趋势性分量通过多层感知机提取整体特征; 季节性及残差分量则通过傅里叶变换转至频域, 提取周期和波动信息后再逆变换回时域。最终, 将包含周期和波动信息的子序列与趋势信息整合。RGCA 结合排序分组策略与通道注意力, 自适应学习多个序列的相关信息, 提升变量间特征表征能力。序列拐点信息提取模块利用数据的拐点数量提取波动信息, 通过网络学习生成优化预测结果的附加值, 干预并修正预测结果, 从而提升准确性。

模型的数据处理流程如下: 首先, 将预处理后的同质多变量时间序列输入模型; 其次, 经过正则化层、SDFFT 和 RGCA 后, 输出数据通过残差连接并重复 $N-1$ 次; 再次, 与序列拐点信息提取模块的附加值相加, 相加结果通过全连接神经网络映射到输出长度, 得到时间序列预测结果。

2.2 跨时间信息提取模块 SDFFT

在时间序列预测领域, 数据的时效性, 即输入序列的不同阶段, 对预测精度有着至关重要的影响^[20-21], 该思想对各种类型时间序列数据的研究都具有适用性。在预测距离输入点较近的时间点信息时, 近期数据相较于远期数据对预测结果的贡献更为显著。预测未来较为久远时间点的信息时, 除近期数据外, 整个序列的历史趋势也至关重要, 需综合考量以捕捉潜在的演变规律。因此, 在进行

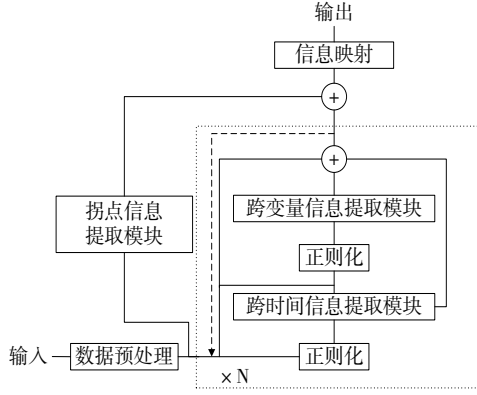


Fig. 1 The overall framework of HMTS-TI

图1 HMTS-TI的整体框架

时间序列的长期预测时,需要平衡近期数据的即时性与历史数据的全局视野,通过融合短期波动与长期趋势,捕捉系统演化的多层次特征。跨时间信息的提取过程具体如下:将输入长度为 L 的序列依据其采样率设定子序列长度 l ,从靠近预测点的一侧逆向划分子序列,划分长度为 $l, 2l, 3l, 4l, \dots$,的共 n 个子序列,其中 $n = \frac{L}{l}$,对每一个子序列在时间维度上对季节性或不规则扰动信息进行提取,映射到一个新的同样长度的序列,再将所有新序列按照位置相加获得一个原始长度的序列。通过如上设置,最靠近预测点的一段长为 l 的序列将提取包含 n 个子序列提取的信息,而最远离预测点的一段长为 l 的序列提取包含1个子序列提取的信息。

在深入挖掘时间序列潜在模式的过程中,进一步采用时间序列分解技术,对于每个子序列,应用滑动平均法,动态调整窗口尺寸以匹配序列长度,将复杂序列拆解为长期趋势与季节性加随机扰动两个关键组成部分。这一策略有助于在最终整合的新序列中保留多尺度的季节信息,增强预测模型的综合表现。

Autoformer^[9]、FEDformer^[22]、FCVAE^[23]等证明了将序

列从时域转化为频域,对具有周期性的数据信息提取的有效性。时间序列中分离的季节性加随机扰动成分,可以表征序列中的周期性和不规则波动,通过快速傅里叶变换(Fast Fourier Transform, FFT)将时域信号转换至频域,将原始数据表示为不同频率正弦波与余弦波的叠加,有利于提取出时间序列中的周期性特征。对给定序列 $X_t \in \mathbb{R}^{N \times L}$,其中 N 表示变量数量, L 表示输入的时间长度,当 L 为偶数时,对于每一个变量的时间序列 $x \in \mathbb{R}^L$ 进行快速傅里叶变化,将数据从时域转化为频域,如式(1)所示。

$$x' = FFT(x) = \sum_{n=0}^{L-1} x_n \cdot e^{-j \frac{2\pi kn}{L}}, k = 0, 1, \dots, \frac{L}{2} \quad (1)$$

其中, j 为虚数单位, n 为时域索引, k 为频率索引, $k = 0$ 时表示没有波动信息, $k = 1$ 时表示序列中最缓慢的波动信息,其周期恰好等于序列长度 L , $k = \frac{L}{2}$ 反映最高解析频率下的特性。对于第 k 个频率分量,其实部和虚部可分别表示为如式(2)、式(3)所示。

$$x'_r = \sum_{n=0}^{L-1} x_n \cdot \cos\left(\frac{2\pi kn}{L}\right), k = 0, 1, \dots, \frac{L}{2} \quad (2)$$

$$x'_i = -\sum_{n=0}^{L-1} x_n \cdot \sin\left(\frac{2\pi kn}{L}\right), k = 0, 1, \dots, \frac{L}{2} \quad (3)$$

经过FFT变换后获得复数值中的实部和虚部,它们相互独立,但又组合表示频率数据振幅和相位角信息。分别对实部与虚部应用多层感知机模块进行映射,提取原始数据在多个频率下的振幅相位信息,再通过逆傅里叶变换将处理后的复数序列还原至时域,能够更加全面地提取季节和波动信息,随后与完整序列分离的趋势性成分的映射结果相加,以重建序列。该模块的具体实现过程如图2所示。

2.3 跨变量信息提取模块RGCA

SENet是一种基于注意力机制的通道注意力模型,最初被提出是用以解决在计算机视觉领域卷积神经网络(Convolutional Neural Networks, CNNs)倾向于平等地对待

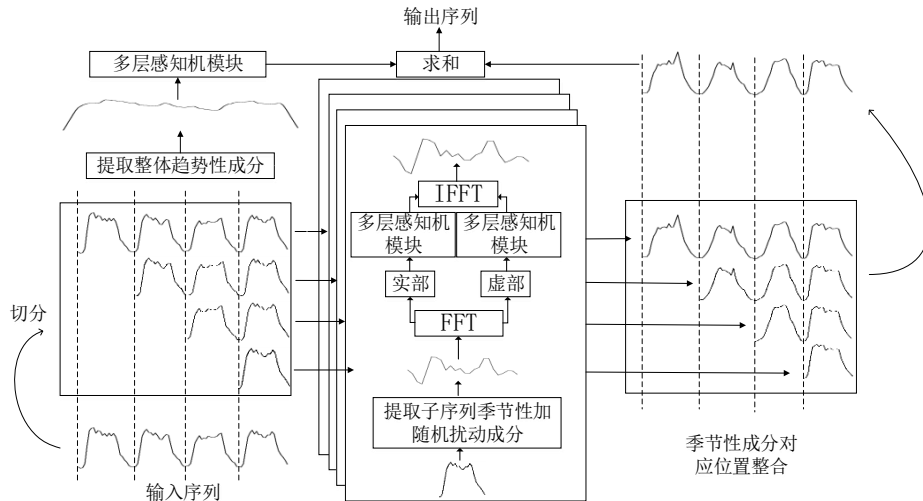


Fig. 2 SDFFT module implementation detail

图2 SDFFT模块实现细节

图像中的所有特征通道,而忽视了不同通道之间的相关性这一问题^[24]。在多变量时间序列预测中,类似问题也存在,过去的模型往往平等对待所有变量序列,忽视了它们之间的相关性,或仅通过降维减少冗余。因此,在变量间相关性较强的同质多变量时间序列预测任务中,引入通道注意力机制能够有效提升模型对关键特征的捕捉能力,强化对相关信息的利用与整合。

SENet 的一个重要观点是通过全连接层学习通道间的关联,将这些关联转化为注意权重,以强化重要特征并增强对通道间相关信息的利用。与时间维度常用的自注意力机制相比,这种通道注意力不需要为每个通道单独定义 Q 、 K 、 V (Query-Key-Value) 矩阵,只需关注特征张量在通道维度上的权重分配,从而大幅减少参数数量并降低计算成本^[5]。本文研究基于这一观点,在通道注意力中引入多头的概念,设计了同质多变量时间序列数据的跨变量信息提取模块,使不同注意力头能关注不同的隐藏信息,组合后提供更全面丰富的表示。

对于从上一层网络中输出的数据 $X_c \in R^{N \times L}$,首先通过自适应池化层 A 对序列中的数据进行压缩,去除冗余信息,提取共性特征得到 $Y \in R^{F \times L}$,其公式如式(4)所示, F 表示压缩后的特征向量维度数目。

$$Y = A(X_c) \quad (4)$$

然后进行分组处理,将 Y 分为 $\frac{F}{G}$ 组,用 $Y_g =$

$[Y_1, Y_2, \dots, Y_{\frac{F}{G}}]$ 表示。其中, G 表示每组特征的数量,对每个分组 Y_i 利用对应的分组卷积层 C_i ,以及 GELU 激活函数进行信息提取,并将获取的信息作为注意权重 $Y_i' \in R^{N \times L}$,其公式如式(5)所示。

$$Y_i' = GELU(C_i(Y_i)), i = 1, 2, \dots, \frac{F}{G} \quad (5)$$

将排序后的原始序列 X_{sorted} 与第 i 个注意权重 Y_i' 进行点乘,计算原始序列对于每个注意力头提取到的关联信息的关注程度,并将所有关注度分数进行整合,获得输出序列 $S \in R^{N \times L}$,相关公式如式(6)、式(7)所示。

$$S_i = Y_i' \cdot X_{sorted} \quad (6)$$

$$S = \sum_{i=1}^{\frac{F}{G}} S_i \quad (7)$$

尽管同质多变量时间序列具有高度相关性和动态一致性,但由于缺乏固有顺序,不同序列间的波动幅度和数值范围存在差异,这可能影响模型学习到的特征一致性和有效性。为提高池化操作所提取的共性信息的代表性,并确保每个注意力头的信息更加平均,本文模型在读取数据后,首先对不同变量维度的时间序列进行平均高度排序,以确保相邻时间序列在平均高度上更为相似。与图像处理中难以量化和排序的 RGB 通道不同,时序预测领域的序列具有可量化特性,因此排序是必要的。排序后的时间序列被送入池化层,并按固定间隔划分为多个分组,从而提高模型性能和特征提取的准确性。该模块整体流程如图 3 所示。

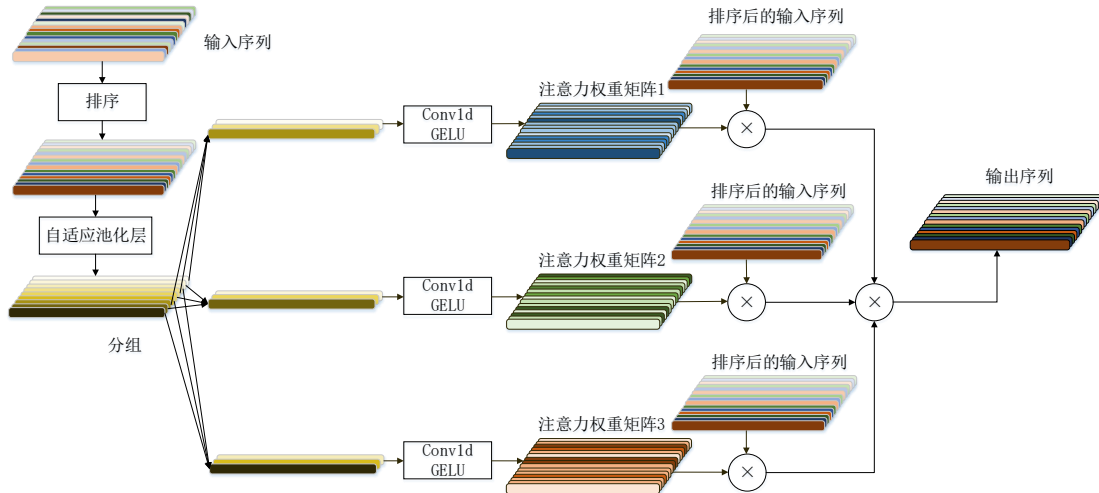


Fig. 3 RGCA module implementation schematic

图3 RGCA 模块实现示意图

2.4 拐点信息提取模块

拐点指时间序列中数据走势发生显著变化的点,标志着数据轨迹由增转减或由减转增的转折时刻,其存在与否直接反映了序列走势的动态特性。高密度的拐点分布表示数据波动剧烈且缺乏稳定趋势,显示出高变异性 and 随机扰动;而稀疏的拐点布局则预示着数据走势平缓,波动幅度较低且规律性强。不同序列的稳定性和拐点数量存在

差异,探究各类时间序列的拐点信息有助于量化和挖掘季节性与随机波动。因此,本文研究将序列的拐点数量作为一个序列的统计域特征参数,通过参数的介入以优化模型预测准确性。

对于输入的多变量时间序列 $X_{sorted} \in R^{N \times L}$,计算序列间的差异,即每个时间点与其前一时间点之间的差值,得到差分张量 $\Delta T \in R^{N \times (L-1)}$,其公式如式(8)所示。

$$\Delta T = X_{sorted}[:, 1:] - X_{sorted}[:, :-1] \quad (8)$$

继而,找到拐点位置,即差分张量中从正转负或从负转正的点。此过程可以通过检查连续两个时间点的差分乘积是否小于零加以实现。定义拐点掩码张量 $M \in R^{N \times (L-2)}$,其公式如式(9)所示。

$$M = (\Delta T[:, :-1] \cdot \Delta T[:, 1:]) < 0 \quad (9)$$

然后,统计每个序列的拐点数量,即求掩码张量 M 中为真的元素总数。定义拐点计数张量 $T_{turn} \in R^n$,其公式如式(10)所示。

$$T_{turn} = \sum_{k=0}^{L-3} M[:, k] \quad (10)$$

最后,将拐点计数张量 T_{turn} 输入到网络,与预先定义的初始权重为1的权重矩阵 $W \in R^L$ 进行爱因斯坦求和,获得拐点信息附加值 $AV \in R^{L \times N}$,其公式如式(11)所示。

$$AV = einsum(T_{turn}, W, "N, L \rightarrow LN") \quad (11)$$

3 实验与结果分析

3.1 数据集

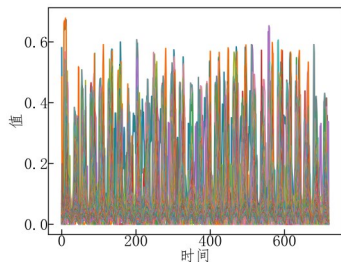
为了验证本文 HMTS-TI 的性能,选取了3个公开的多变量数据集 Traffic^[9]、Solar^[21]和 PeMS04^[25]进行测试,其简要描述如表1所示。

Table 1 Homogeneous multivariate datasets used in experiments

表1 实验中使用的同质多变量数据集

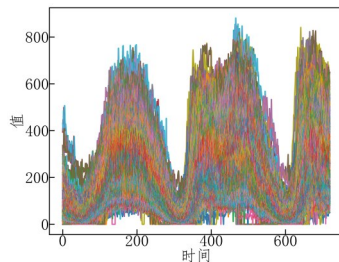
数据集	维度	时间步	采样时间	描述
Traffic	862	17 544	1h	旧金山海湾高速路占用率
Solar	137	52 560	10min	阿拉巴马州光伏发电厂太阳能发电记录
PeMS04	307	16 992	5min	加州高速交通流量数据

以上3个数据集中不同维度的变量序列均来自于同样的数据生成过程。其中,Traffic和PeMS04数据集包含某段高速公路中多个相同的传感器对某一时间内的道路占用率和交通流量信息的采集,Solar数据集包含在同一地区多个发电厂收集的太阳能输出数据。这些数据在结构上协调一致,且在动态特性上展现出了高度的同步性和可比性。对序列数据进行截取片段可视化操作后,如图4所示。



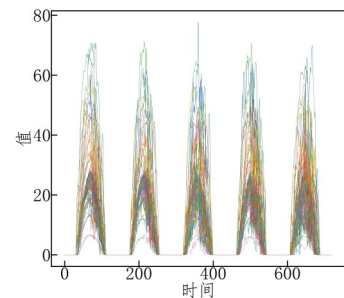
(a) Traffic dataset visualization

(a) Traffic数据集可视化



(b) Solar dataset visualization

(b) Solar数据集可视化



(c) PeMS04 dataset visualization

(c) PeMS04数据集可视化

Fig. 4 Visualization of original datasets

图4 原始数据可视化

为了进一步说明以上同质数据集变量间的相关性,本文对标准化操作后的数据进行了两两变量维度间的皮尔逊相关系数计算,计算数值越接近正负1,表示两者正负相关性越强,计算结果如图5所示。为了更直观地加以对比说明,图5(d)可视化了一个非同质变量数据集(深圳遥感站气象数据)的相关性热力图。可以看出,同质数据集变量间大多有较强的正相关性,且不同维度间的相关性差别较小,而非同质多变量数据集的变量相关性可能存在正相关,也可能存在负相关,且不同变量间的相关性差距较大。

3.2 性能评价指标与实验设置

在评估预测模型准确性时,本文选择了3种被广泛接受且具有互补性的度量指标——均方误差(Mean Squared Error, MSE)、平均绝对误差(Mean Absolute Error, MAE)和决定系数(Coefficient of Determination, R^2),以量化模型预测与实际观测值之间的差异,其公式分别如式(12)一式(14)所示。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

本文模型采用ADAM优化算法,批量大小为16,数据按7:1:2比例划分为训练集、验证集和测试集,并使用StandardScaler进行标准化处理。输入窗口大小为96个时间单位,输出预测长度为96、192、336和720。滑动窗口大小设置为序列整体长度的 $\frac{1}{4}$ 。训练在验证集性能提升停滞3个连续周期时自动中止,以避免过拟合。优化目标基于Log-Cosh损失函数,在3个随机种子下进行3轮独立实验。其他超参数设置如表2所示。

其中,d_ff为输入变量维度的池化压缩维度,group为RGCA中每组的维度数量,均通过网格搜索确定;SDFFT模块中的时间序列切分参数cut参考数据的采样率和周期设定,并通过d_cut个隐藏层神经元提取特征。

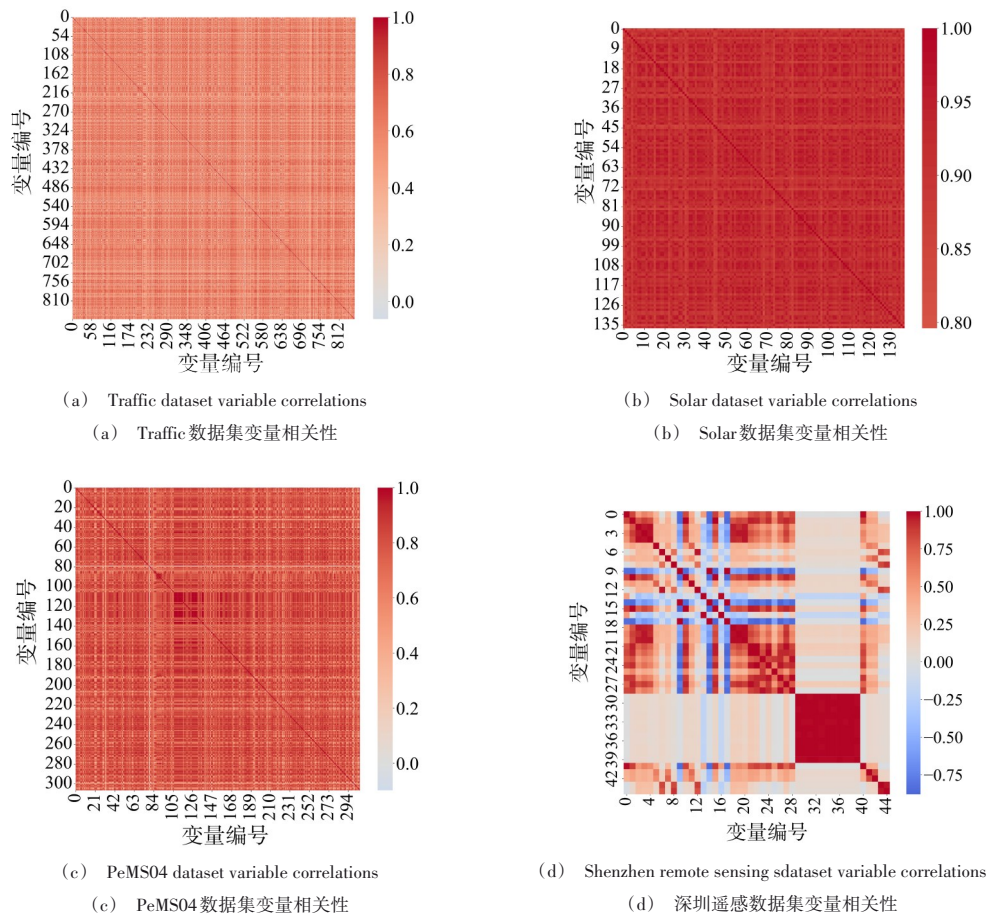


Fig. 5 Heatmap of pearson correlation coefficients among variables

图5 变量间皮尔逊相关系数热力图

Table 2 Hyperparameters used in the experiments

表2 实验中使用的超参数

数据集	learning_rate	d_ff	group	d_cut	cut
Traffic	0.020	64	16	128	16
Solar	0.001	2	2	2 048	96
PeMS04	0.010	32	4	128	12

3.3 对比模型与实验结果

为验证 HMTS-TI 在时间序列长期预测任务上的实验效果,采用当前常见且性能较优的时间序列长期预测模型与 HMTS-TI 进行了对比实验,包括基于 MLP 的 MTS-Mixer、Dlinear、LightTS^[26]、基于 CNN 的 MICN 和基于 Transformer 的 FEDformer。实验对比结果如表 3 所示,其中加粗表示最优实验结果,本文模型在 36 个预测指标中有 33 个均取得了相对更优效果。分析可知,与对比模型中预测准确性最高的模型 MTS-Mixer 相比,本文模型在 3 个数据集的 4 个预测长度上 MSE、MAE 分别下降了 12.1% 和 6.8%,R²提升了 4.1%。其中,在 Traffic 数据集上提升最为明显,在 4 个预测长度上 MSE、MAE 分别下降了 15.1% 和 13.0%,R²提升了 9.2%。

结合上述分析可知,MTS-Mixer 模型通过变量间的信息映射在 3 个数据集上表现相对优越,但其仅采用多层感知机结构进行时序信息分析,提取能力不足,尤其在季节性明显的 Traffic 数据集中,缺乏分解策略限制了潜在时间依赖关系的探究。尽管 MICN、Dlinear 和 FEDformer 引入了分解策略,但未充分考虑变量间的依赖关系,导致在 Solar 和 PeMS04 数据集上的预测误差较大,且未考虑输入序列不同位置对预测序列的影响差异。LightTS 通过连续和间隔采样捕获时间维度的局部特征与长期依赖,但同样未进行分解以探究季节及趋势规律,且未学习变量间的相互依赖关系。

对于 RGCA 中不同分组学习的学习情况,将 Traffic 数据集上 4 个分组的卷积网络权重矩阵可视化,如图 6 所示。其中每一行代表一个降采样后的序列数据,每一列表示经过不同的卷积核操作。可以看出,每一个分组会特别关注该分组中的某些序列(亮色表示较大的权重),且每个分组关注的降采样序列也不同,如第 1 个分组关注序列 4 的信息,第 2 个分组关注序列 9 的信息,第 3 个分组关注序列 7 的信息,第 4 个分组关注序列 12 的信息,通过赋予这些序列不同的关注程度,进而对模型预测结果产生影响,与 RGCA 的设计思路相对应。

Table 3 Prediction results of different models on three datasets across four forecast lengths
表 3 3 个数据集在 4 个预测长度上不同模型的预测结果

模型	数据集	指标	Traffic				Solar				PeMS04			
			96	192	336	720	96	192	336	720	96	192	336	720
HMTS-TI	MSE		0.448	0.462	0.484	0.512	0.156	0.184	0.199	0.203	0.120	0.133	0.138	0.165
	MAE		0.299	0.309	0.315	0.335	0.234	0.258	0.266	0.266	0.234	0.249	0.253	0.281
	R ²		0.693	0.685	0.673	0.657	0.789	0.752	0.732	0.728	0.883	0.871	0.866	0.840
MTS-Mixer	MSE		0.550	0.531	0.555	0.608	0.187	0.214	0.223	0.208	0.122	0.141	0.142	0.165
	MAE		0.358	0.346	0.359	0.383	0.253	0.275	0.276	0.267	0.236	0.252	0.256	0.277
	R ²		0.623	0.638	0.625	0.594	0.747	0.712	0.700	0.720	0.881	0.863	0.862	0.840
MICN	MSE		0.518	0.546	0.538	0.594	0.208	0.232	0.299	0.252	0.214	0.246	0.221	0.246
	MAE		0.306	0.316	0.313	0.324	0.269	0.283	0.322	0.295	0.316	0.325	0.311	0.333
	R ²		0.646	0.629	0.637	0.603	0.719	0.688	0.599	0.661	0.792	0.761	0.785	0.762
Dlinear	MSE		0.650	0.599	0.606	0.646	0.285	0.316	0.349	0.354	0.422	0.441	0.377	0.439
	MAE		0.397	0.371	0.375	0.396	0.372	0.393	0.412	0.410	0.484	0.497	0.445	0.486
	R ²		0.555	0.592	0.590	0.568	0.615	0.575	0.532	0.524	0.589	0.572	0.633	0.574
LightTS	MSE		0.597	0.604	0.625	0.668	0.205	0.232	0.250	0.249	0.193	0.228	0.237	0.278
	MAE		0.383	0.393	0.406	0.423	0.280	0.301	0.312	0.309	0.316	0.345	0.348	0.379
	R ²		0.591	0.589	0.578	0.548	0.723	0.687	0.664	0.666	0.812	0.778	0.770	0.732
FEDformer	MSE		0.575	0.602	0.617	0.631	0.304	0.347	0.408	0.408	0.149	0.171	0.197	0.279
	MAE		0.359	0.372	0.381	0.385	0.386	0.412	0.459	0.458	0.271	0.290	0.315	0.379
	R ²		0.606	0.590	0.583	0.579	0.589	0.532	0.452	0.451	0.855	0.834	0.808	0.729

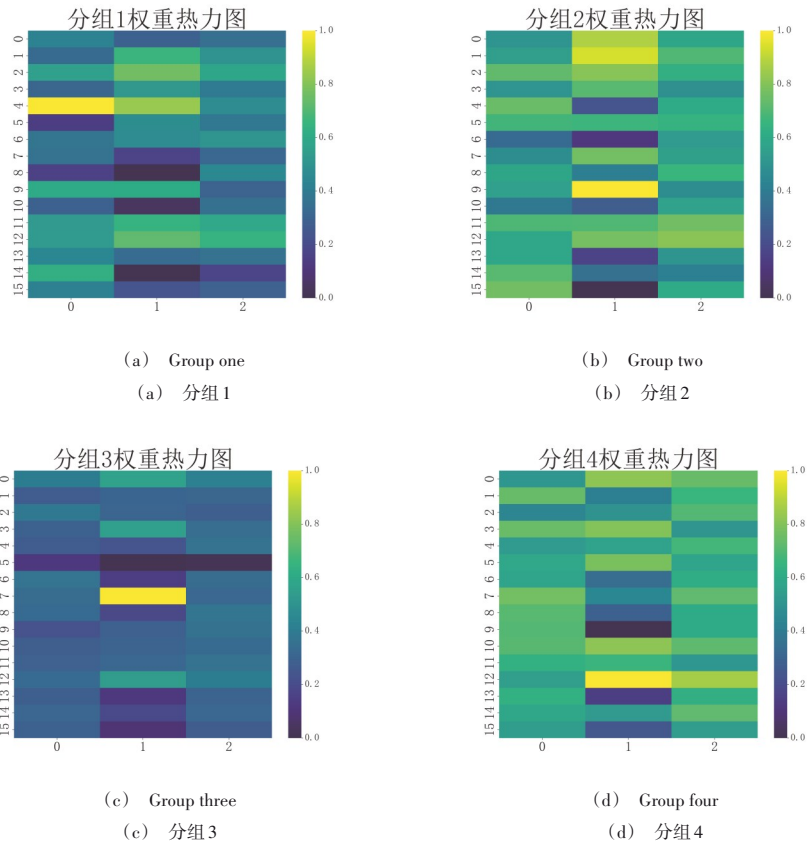


Fig. 6 Visualization of network weight matrices

图 6 网络权重矩阵可视化

为了更加直观地展示模型预测效果,将 3 个数据集最后一个变量维度的序列在 720 预测长度上的预测结果进行可视化,如图 7 所示,其中蓝色线表示真实值,橙色线表示预测值,展示了从输入的 96 个数据到一次性输出的 720

个数据的完整曲线。可以看出,无论是单周期的 Solar 数据集、多周期的 Traffic 数据集还是整体走势变化较为缓慢的 PeMS04 数据集,模型均能较好地拟合出序列变化,并较为精确地识别和预测出时间序列整体趋势。

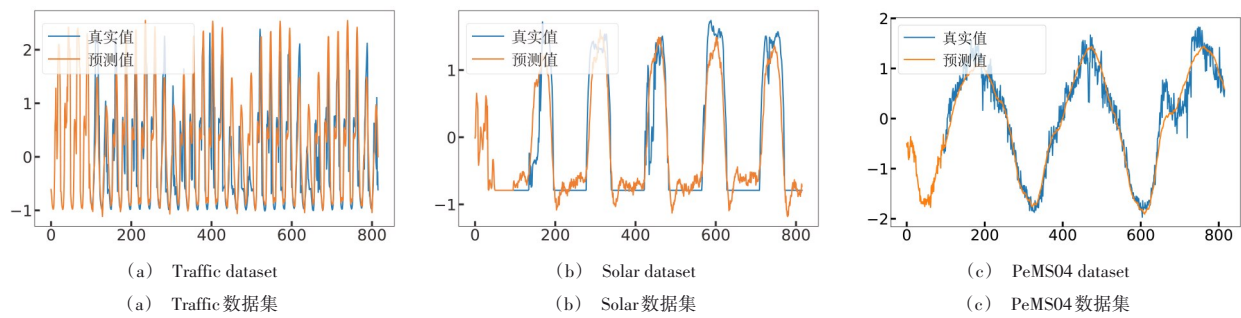


Fig. 7 Visualization of HMTS-TI prediction results
图 7 HMTS-TI 预测结果可视化

3.4 消融实验

为分析 HMTS-TI 的有效性, 本文开展消融实验, 探究 HMTS-TI 各模块对时间序列预测准确率的影响。实验结果如表 4 所示, 其中 HMTS-TI-I 表示在原模型的基础上去除拐点信息提取模块, 以探究将拐点数量作为补充信息对

预测准确性的影响; HMTS-TI-T 表示在原模型的基础上去除 SDFFT, 以探究跨时间依赖信息对预测准确性的影响; HMTS-TI-H 表示在原模型的基础上去除 RGCA, 以探究跨变量依赖信息对模型预测准确性的影响。实验结果中的最优结果用粗体表示。

Table 4 Comparison of ablation results metrics for three datasets across four forecast lengths
表 4 3 个数据集在四个预测长度上消融结果指标对比

模型	数据集 指标	Traffic				Solar				PeMS04			
		96	192	336	720	96	192	336	720	96	192	336	720
HMTS-TI	MSE	0.448	0.462	0.484	0.512	0.156	0.184	0.199	0.203	0.120	0.133	0.138	0.165
	MAE	0.299	0.309	0.315	0.335	0.234	0.258	0.266	0.266	0.234	0.249	0.253	0.281
	R ²	0.693	0.685	0.673	0.657	0.789	0.752	0.732	0.728	0.883	0.871	0.866	0.840
HMTS-TI-I	MSE	0.457	0.470	0.492	0.521	0.157	0.188	0.200	0.204	0.125	0.136	0.143	0.175
	MAE	0.302	0.306	0.317	0.340	0.236	0.259	0.267	0.266	0.240	0.253	0.258	0.292
	R ²	0.687	0.680	0.668	0.651	0.788	0.747	0.731	0.725	0.879	0.868	0.861	0.830
HMTS-TI-T	MSE	0.599	0.579	0.588	0.640	0.223	0.245	0.268	0.263	0.159	0.172	0.183	0.245
	MAE	0.379	0.356	0.362	0.390	0.301	0.315	0.322	0.311	0.278	0.293	0.303	0.355
	R ²	0.589	0.606	0.603	0.572	0.699	0.670	0.640	0.646	0.845	0.833	0.822	0.762
HMTS-TI-H	MSE	0.547	0.547	0.564	0.604	0.175	0.197	0.206	0.204	0.138	0.167	0.197	0.217
	MAE	0.371	0.370	0.376	0.394	0.242	0.258	0.267	0.258	0.253	0.274	0.298	0.316
	R ²	0.625	0.628	0.619	0.596	0.763	0.735	0.724	0.726	0.865	0.838	0.809	0.805

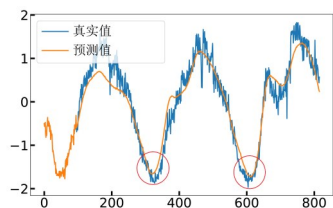
对数据进行分析可知, 在去除拐点信息提取模块、SDFFT 和 RGCA 后, 模型预测的 MSE 在 3 个数据集上整体分别增加了 2.0%、30.0% 和 17.4%。由此可以得出, 在相关性较高的同质多变量数据集中跨时间依赖信息以及跨变量依赖信息对模型结果具有较为显著的影响, 而拐点数量的附加统计信息对模型预测结果有正向调节作用。

在 PeMS04 数据集预测长度为 720 个时间单位的情况下对未引入拐点信息及引入拐点信息的预测结果进行可视化对比, 如图 8(a)、(b) 所示。对于具有高密度的拐点分布真实数据, HMTS-TV 只能大致拟合出其较为明显的季节性波动, 预测曲线整体十分平滑, HMTS-TI 会产生一些微小波动以尝试拟合数据中的不规则扰动, 如圆圈中所指部分。对图 8(a)、(b) 中左下角圆圈进行放大, 如图 8(c)、(d) 所示, 可以清晰看出两者区别。

为了验证模型中各结构的有效性, 本文在 PeMS04 数据集中增加了一组对比实验。对于 RGCA 模块, 为了研究引入通道注意力以及搭配分组排序的有效性, 设计了变体 1 和变体 2。其中, 变体 1 使用两层全连接神经网络代替下

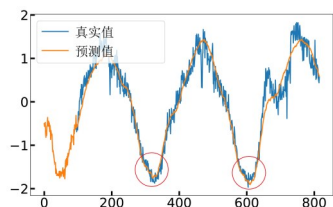
采样池化和变量间信息提取过程, 并保持隐藏神经元数量不变, 以此确保比较的公平性; 变体 2 在模型 HMTS-TI 的基础上, 删除分组排序操作, 仅通过池化和通道注意力提取变量间特征。对于 SDFFT 模块设计了变体 3、变体 4 和变体 5, 分别反映子序列划分提取时效信息、傅里叶变换将序列从时域转化为频域、数据分解提取隐藏信息对预测结果的正向影响。其中, 变体 3 不使用划分子序列的操作, 对序列进行整体信息提取; 变体 4 对季节性波动信息的提取不经过傅里叶变化, 直接通过两层全连接神经网络进行映射, 并保持隐藏神经元数量与原模型一致; 变体 5 在模型 HMTS-TI 的基础上不使用分解, 即对每个子序列利用傅里叶变换提取时序信息, 用以显示在模型中分离隐藏成分的重要性。记录实验的 MSE、MAE 和 R² 结果如表 5 所示, 其中黑体加粗表示该组实验的最佳数值。

由结果可以看出, 本文模型 HMTS-TI 在整体上具有最小的 MSE、MAE 和最高的 R²。从求和数据看, 变体 2、变体 3 和变体 5 的预测准确性下降最为明显。对 5 种变体的预测结果进行分析, 其中变体 1 说明了仅仅使用全连接神经



(a) HMTS-TI-I output results

(a) HMTS-TI-I预测结果



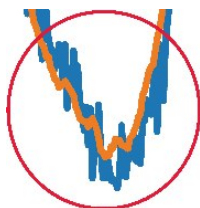
(b) HMTS-TI output results

(b) HMTS-TI预测结果



(c) HMTS-TI-I partial outputs

(c) HMTS-TI-I结果局部



(d) HMTS-TI partial outputs

(d) HMTS-TI结果局部

Fig. 8 Impact of inflection point information on HMTS-TI predictions

图8 拐点信息对HMTS-TI模型预测结果影响

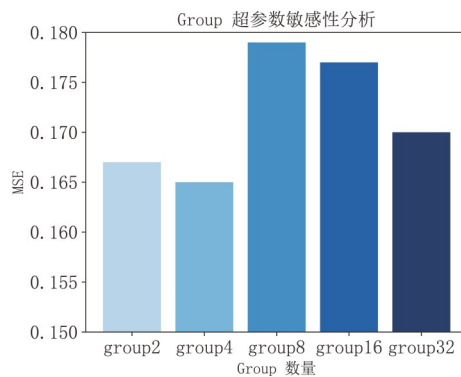
网络不如使用池化及分组通道间注意力的效果。在变体2的基础上加入了池化和通道注意力明显提升了变量间信息提取模块的效果。变体3验证了逆向子序列划分思路的有效性,说明了模型对不同预测阶段的适应能力,即对于预测输出的早期阶段和晚期阶段,模型能够精准识别并有效利用输入序列中对应阶段的特征,从而在整体上提升了预测精度。变体4和变体5的实验也分别验证了数据分解对隐藏其中的趋势和季节及残差部分进行分别建模及傅里叶变化,将序列从时域转化为频域的操作对于预测准确率提升具有一定效果。

为了进一步分析 group 和 cut 超参数对模型结果的影响,同样在 PeMS04 预测长度为 720 个时间单位的设置下进行超参数敏感实验,结果如图 9 所示。group 的参数选择范围设定为 2、4、8、16、32, cut 的参数选择范围设定为 96、

Table 5 Comparison and analysis of HMTS-TI model variants on the PeMS04 dataset

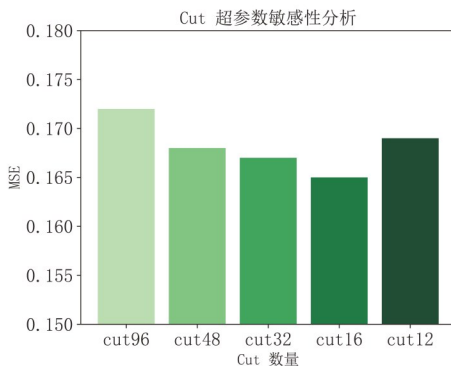
表5 PeMS04数据集上HMTS-TI模型结构变体对比分析

指标	模型变体	96	192	336	720	求和
MSE	HMTS-TI	0.120	0.133	0.138	0.165	0.556
	变体1	0.123	0.134	0.146	0.165	0.568
	变体2	0.126	0.141	0.139	0.175	0.581
	变体3	0.129	0.140	0.143	0.170	0.582
	变体4	0.125	0.136	0.145	0.170	0.576
	变体5	0.120	0.135	0.144	0.182	0.581
MAE	HMTS-TI	0.234	0.249	0.253	0.281	1.017
	变体1	0.246	0.250	0.258	0.278	1.032
	变体2	0.241	0.256	0.252	0.288	1.037
	变体3	0.245	0.258	0.259	0.287	1.049
	变体4	0.242	0.253	0.263	0.286	1.044
	变体5	0.234	0.251	0.258	0.292	1.035
R ²	HMTS-TI	0.883	0.871	0.866	0.840	3.460
	变体1	0.881	0.870	0.858	0.840	3.449
	变体2	0.877	0.863	0.865	0.830	3.435
	变体3	0.875	0.864	0.861	0.834	3.434
	变体4	0.878	0.868	0.859	0.835	3.440
	变体5	0.883	0.869	0.860	0.823	3.435



(a) Group hyperparameter analysis

(a) Group超参数分析



(b) Cut hyperparameter analysis

(b) Cut超参数分析

Fig. 9 Hyperparameter sensitivity analysis

图9 超参数敏感性分析

48、32、16、12。可以看出, group 的数量相比 cut 对模型影响更大, 当 group 参数取 4 时具有最低误差; 而 cut 超参数取

16 时,即以 16 的倍数为间隔划分子序列具有最低误差,并且在该数据集上,随着子序列长度的增加,误差也在逐步增大。

3.5 模型性能分析

为了分析模型性能,分别计算 HMTS-TI 与第 3.3 小节中提及的对比模型在数据量最大的 Traffic 数据集上预测 720 个时间单位长度的结果所需的存储资源和所耗时间。数据包含 862 个变量维度 17 544 个时间点的信息,结果如表 6 所示。

Dlinear、LightTS 和 MTS-Mixer 均属于时间序列预测领域的轻量级模型,拥有较少的参数量,需要较短的训练时间。相比之下,FEDformer 和 MICN 拥有更多的参数量,需要更长的训练时间。这是因为 Dlinear、LightTS 和 MTS-Mixer 在时间序列特征的提取上主要采用多层感知神经网络结构。而 FEDformer 使用了基于 Q 、 K 、 V 的传统注意力机制的运算框架,这种机制虽然有助于捕捉长距离依赖关系,但同时也导致了大量的计算和存储需求;MICN 通过多尺度混合分解和局部全局建模的方式,利用多次卷积以捕捉数据中的复杂特征,这也导致了模型参数量的增加。

从表 6 数据可以看出,HMTS-TI 的参数量数量级远小于 FEDformer 和 MICN,同时在训练时间上也表现良好,与轻量级模型相当。这意味着 HMTS-TI 能够在保持较低资源消耗的同时,处理时间序列数据时表现出较高的效率。

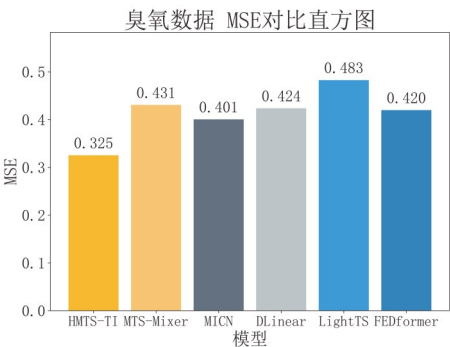
Table 6 Comparison and analysis of model computational performance

表 6 模型计算性能对比分析						
模型名称	HMTS-TI	MTS-Mixer	MICN	Dlinear	LightTS	FEDformer
模型参数量	492 824	285 186	50 868 270	139 680	965 940	20 681 582
单个 epoch	66	54	241	51	79	592
训练时长/s						

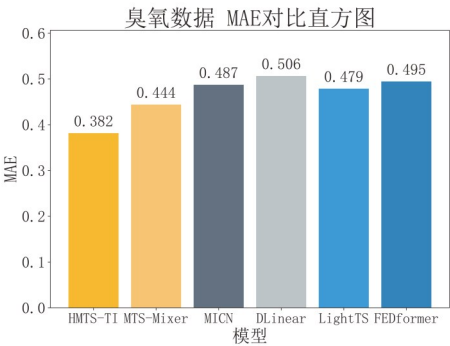
3.6 现实应用——广东省臭氧浓度预测

广东省位于中国南部沿海,是中国经济最为发达的省份之一,也是人口密度较高的区域。珠三角地区作为广东乃至全国的重要经济区,其工业化和城市化进程迅速,伴随而来的是环境污染问题,尤其是空气污染问题。臭氧(O_3)是一种常见的空气污染物,在地面附近,高浓度的臭氧会对人体健康造成伤害,引起呼吸系统疾病,加剧哮喘病人的症状,并损害肺功能。同时,臭氧也会对农作物和其他植物造成危害,减少农作物产量,影响生态系统健康。本文从国家环境气象服务平台网站获取 2023 年 12 月 31 日—2024 年 7 月 31 日广东省全省 92 个监测点的臭氧逐小时数据。数据间的皮尔逊相关系数达到 0.7,具有较高的相关性,且来源于相同的采集方式,符合同质多变量数据特征。实验通过 K 最近邻法进行空缺值填补,将数据按照 7:1:2 划分为训练集、验证集和测试集,利用提出的模型基于过去 96 个时间点(2 天)数据预测未来 336 个时间点(14 天)的信息,以探究模型在实际生活场景中的应用价值。

预测结果在 5 个模型上的 MSE 和 MAE 如图 10 所示。可以看出,HMTS-TI 的预测误差明显小于其他对比模型,其中 MSE 相比于对比模型中最优的 MICN 下降了 19.0%,MAE 相比于对比模型中最优的 MTS-Mixer 下降了 14.0%。将预测结果进行可视化,如图 11 所示,可见模型能够清晰识别出数据的周期,并判断波峰和波谷位置。由此可以得出,本文通用模型在同质多变量的长期预测任务中具有一定的泛化能力和实际应用价值。



(a) O₃ data MSE results comparison
(a) O₃数据MSE结果对比



(b) O₃ data MAE results dataset
(b) O₃数据MAE结果对比

Fig. 10 Comparison histogram of O₃ data prediction results
图 10 臭氧数据预测结果对比直方图

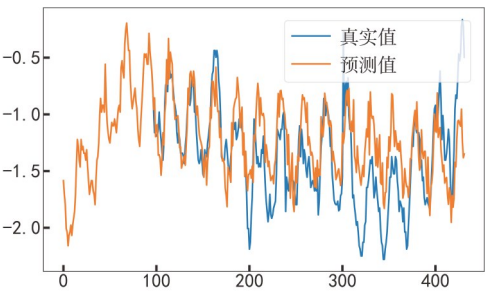


Fig. 11 Visualization of O₃ data prediction result
图 11 臭氧数据预测结果可视化

4 结语

本文提出了融合多阶段跨时间依赖和拐点信息的同

质多变量时间序列预测模型(HMTS-TI),以改善现有模型在同质多变量时间序列(HMTS)长期预测任务中的局限性,增强跨时间信息的提取能力,利用变量间关联信息的学习能力,提升对序列中趋势性、季节性和随机波动等复杂成分的处理能力。通过在模型中构建 RGCA 模块、SDFFT 模块并引入拐点数量统计值,实现如下目标:在变量间通道注意力中引入分组排序策略,有效捕捉并利用了序列间的跨变量维度依赖性,增强对多变量时间序列间相关信息的提取与应用能力;采取逆向子序列划分、多尺度分解以及结合傅里叶变换的信息提取操作深入挖掘时间序列的长期趋势、短期季节模式和随机波动特性,增强了对跨时间维度依赖特征的提取;引入拐点数量统计值作为预测调整因子,对模型拟合产生了正向调节作用。

本文在 3 个公共同质多变量时间序列数据集上进行了不同预测步长的实验,并通过设置多个变体模型证明了各模块的有效性。实验结果表明,本文模型在这些数据集表现出了更强的特征提取能力和更好的预测精度。此外,将模型应用于广东地区 92 个空气检测站点的臭氧浓度预测任务,证实了模型在解决现实问题上的良好效果。

基于现阶段同质多变量时间序列的长期预测研究现状,本文一是在建模策略层面进行了一些探索和创新,二是试图建立一个具有一定泛化能力的预测模型,这两方面均取得了一定成效。但由于种种因素的限制,模型未能进一步在更大规模数据集上进行验证,无法适配更多更广泛的应用场景,这有待在后续的拓展研究中去尝试验证。

参考文献:

- [1] CHEN R H, DAI H, ZHENG G N, et al. Urban electricity load forecasting method based on discrepancy compensation and short-term sampling contrasting loss[J]. Computer Science, 2024, 51(4): 158-164.
陈润桓,戴华,郑桂能,等.基于差异补偿和短期采样对比损失的城市电力负荷预测方法[J].计算机学报,2024,51(4):158-164.
- [2] WANG C, TIAN R, HU J, et al. A trend graph attention network for traffic prediction[J]. Information Sciences, 2023, 623: 275-292.
- [3] HEWAMALAGE H, BERGMEIR C, BANDARA K. Global models for time series forecasting: a simulation study[J]. Pattern Recognition, 2022, 124: 108441.
- [4] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[DB/OL]. <https://arxiv.org/abs/1803.01271>, 2018.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
- [6] WU H, HU T, LIU Y, et al. Timesnet: temporal 2d-variation modeling for general time series analysis[DB/OL]. <https://arxiv.org/abs/2210.02186>, 2023.
- [7] HYNDMAN R, KOEHLER A B, ORD J K, et al. Forecasting with exponential smoothing: the state space approach[M]. Berlin: Springer Science & Business Media, 2008.
- [8] HOCHREITER S. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1880.
- [9] 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.
- [10] WANG H, PENG J, HUANG F, et al. MICN: multi-scale local and global context modeling for long-term series forecasting[C]//Kigali: International Conference on Learning Representations, 2023.
- [11] QI Y, LI Q, KARIMIAN H, LIU D. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory [J]. Science of the Total Environment, 2019, 664: 1-10.
- [12] WU H, CHEN H. Multi-site wind speed prediction based on graph embedding and cyclic graph isomorphism network (GIN-GRU) [J]. Energies, 2024, 17(14): 3516.
- [13] LI M, ZHU Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 4189-4196.
- [14] ZHANG Y H, YAN J C. Crossformer: transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//Kigali: International Conference on Learning Representations, 2023.
- [15] LI Z, RAO Z, PAN L, XU Z. Mts-mixers: multivariate time series forecasting via factorized temporal and channel mixing[DB/OL]. <https://arxiv.org/abs/2302.04501>, 2023.
- [16] 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.
- [17] 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.
- [18] LIU Y, WU H, WANG J, et al. Non-stationary transformers: exploring the stationarity in time series forecasting[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 9881-9893.
- [19] WANG Z, SHANG P J. Generalized entropy plane based on multiscale weighted multivariate dispersion entropy for financial time series[J]. Chaos, Solitons & Fractals, 2021, 142: 110473.
- [20] HEWAMALAGE H, BERGMEIR C, BANDARA K. Recurrent neural networks for time series forecasting: current status and future directions[J]. International Journal of Forecasting, 2021, 37(1): 388-427.
- [21] LAI G, CHANG W, YANG Y, et al. Modeling long- and short-term temporal patterns with deep neural networks [C]//International ACM SIGIR Conference on Research and Development in Information Retrieval, 2018: 95-104.
- [22] ZHOU T, MA Z, WEN Q, et al. Fedformer: frequency enhanced decomposed transformer for long-term series forecasting[C]//International Conference on Machine Learning, 2022: 27268-27286.
- [23] WANG Z, PEI C, MA M, et al. Revisiting VAE for unsupervised time series anomaly detection: a frequency perspective [C]//WWW '24: Proceedings of the ACM Web Conference, 2024: 3096-3105.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [25] ZHU J, WANG Q, TAO C, et al. AST-GCN: attribute-augmented spatiotemporal graph convolutional network for traffic forecasting[J]. IEEE Access, 2021, 9: 35973-35983.
- [26] CAMPOS D, ZHANG M, YANG B, et al. LightTS: lightweight time series classification with adaptive ensemble distillation [C]//Proceedings of the ACM on Management of Data, 2023: 1-27.

(责任编辑:孙娟)