

# 基于扩散与时频注意力的时间序列插补方法<sup>①</sup>

王 槃, 曾倩欣, 杨 欢

(华南师范大学 软件学院, 佛山 528225)

通信作者: 杨 欢, E-mail: [huanyang2019@163.com](mailto:huanyang2019@163.com)



**摘 要:** 时间序列插补旨在根据现有数据填补缺失值以恢复数据的完整性。目前基于 RNN 的插补方法存在较大的误差, 并且增加网络层数容易出现梯度爆炸和消失问题, 而基于 GAN 和 VAE 的插补方法经常面临训练困难和模式崩溃的挑战。为解决上述问题, 本文提出了一种基于扩散与时频注意力的时间序列插补模型 DTFA (diffusion model and time-frequency attention), 通过反向扩散实现从高斯噪声中重建缺失数据。具体而言, 本研究利用多尺度卷积模块与二维注意力机制捕获时域数据中的时间依赖性, 并利用 MLP 与二维注意力机制学习频域数据的实部与虚部信息。此外, 本研究通过线性插补模块对现有的观测数据进行初步的数据增强, 从而更好地指导模型的插补过程。最后, 本研究通过最小化真实噪声与估计噪声的欧氏距离来训练噪声估计网络, 并利用反向扩散实现对时序数据的缺失插补。本研究的实验结果表明, DTFA 在 ETTm1、WindPower 和 Electricity 这 3 个公开数据集上的插补效果均优于近年主流的基线模型。

**关键词:** 时间序列插补; 扩散模型; 注意力机制; 线性插补; 多尺度卷积

引用格式: 王槃, 曾倩欣, 杨欢. 基于扩散与时频注意力的时间序列插补方法. 计算机系统应用, 2024, 33(11): 90-100. <http://www.c-s-a.org.cn/1003-3254/9669.html>

## Time Series Imputation Method Based on Diffusion and Temporal-frequency Attention

WANG Pan, ZENG Qian-Xin, YANG Huan

(School of Software, South China Normal University, Foshan 528225, China)

**Abstract:** Time series imputation aims to restore data integrity by filling in missing values based on existing data. Currently, RNN-based imputation methods suffer from large errors, and increasing the number of network layers often leads to exploding and vanishing gradients. Additionally, GAN-based and VAE-based imputation methods frequently encounter challenges such as training difficulties and pattern collapse. To address these challenges, this study proposes a time series imputation model named diffusion model and time-frequency attention (DTFA), which reconstructs missing data from Gaussian noise through reverse diffusion. Specifically, this study utilizes multi-scale convolutional modules and two-dimensional attention mechanisms to capture temporal dependencies in time-domain data and employs MLPs and two-dimensional attention mechanisms to learn real and imaginary parts of frequency-domain data. This study also implements a linear imputation module to augment the existing observed data, thereby providing better guidance for model imputation. Finally, this study trains a noise estimation network by minimizing the Euclidean distance between real noise and estimated noise and then utilizes reverse diffusion to fill in the missing values in time series data. The experimental results demonstrate that DTFA outperforms mainstream baseline models in terms of imputation effectiveness on three public datasets: ETTm1, WindPower, and Electricity.

**Key words:** time series imputation; diffusion model; attention mechanism; linear imputation; multi-scale convolution

① 基金项目: 2020 年批次佛山高等教育高层次人才项目; 广东省基础与应用基础研究基金 (区域联合基金-青年基金) (2020A1515110783)

收稿时间: 2024-04-20; 修改时间: 2024-05-14; 采用时间: 2024-05-23; csa 在线出版时间: 2024-09-24

CNKI 网络首发时间: 2024-09-25

## 1 引言

时间序列数据是按照时间顺序排列的数据序列,具有较强的时间相关性和连续性,是工程、医疗、气象等众多领域中常见的数据类型,因此时序数据对于人们理解系统的演变过程具有重要的研究意义<sup>[1]</sup>.然而在实际应用场景中,时序数据常会因为传感器故障、数据采集错误或通信中断等原因而出现部分数据缺失的情况,这将会对时序数据的分析、预测和分类等下游任务产生不利的影响.

时序数据插补是解决数据缺失问题的重要手段之一.其核心目标是通过捕获现有数据中潜在的信息以推断出缺失位置的数据,实现对残缺时序样本的修复,从而为后续的分析 and 预测任务提供可靠的数据基础.因此研究有效的时序数据插补方法对于众多领域的发展具有广泛的意义,并且在理论方面和实际应用具有重要的价值<sup>[2]</sup>.多元时序数据插补的示例如图1所示,“x”标记的位置表示发生数据缺失,而“.”标记的位置表示模型根据现有数据估计的插补数据,两者的距离越近表明模型的插补准确率越高.

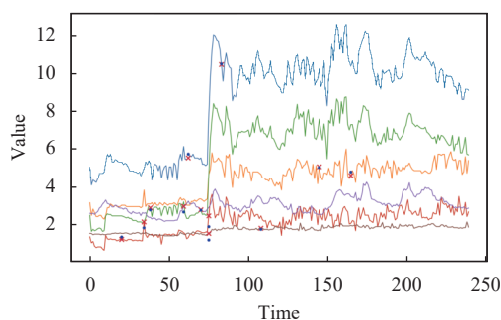


图1 多元时间序列插补示例

早期研究中的传统统计方法如均值插补、中值插补和多项式插值等,对数据分布的假设较为严格,难以适应不同类型的时序数据.而K近邻、回归模型和随机森林等机器学习方法虽然简单高效,但随着大数据时代的到来,这些方法难以捕捉高维数据中复杂的非线性关系<sup>[3]</sup>.此外,在大部分应用场景下略长插补的时间几乎不会对下游任务造成负面影响,因此人们对于时序插补的实时性需求较少,而时序插补的准确率才是首要的考量因素.

近年来,基于深度学习的时序数据插补方法受到了广泛关注,循环神经网络(recurrent neural network, RNN)展示了深度学习模型处理高维时序数据的强大

能力. Che 等人<sup>[4]</sup>通过改进门控循环单元(gated recurrent unit, GRU)设计了一种插补模型 GRU-D,利用时间间隔矩阵记录最后一次观测缺失数据的时间,并根据权重衰减机制反应时序数据的长短期依赖关系. Yoon 等人<sup>[5]</sup>将 RNN 改进为一种双向的时序插补模型 M-RNN (multi-directional recurrent neural network),通过结合前向和后向的信息,实现对多元时间序列动态变化的全面捕获. Cao 等人<sup>[6]</sup>提出的模型 BRITS (bidirectional recurrent imputation for time series) 无需对数据的生成过程强加假设,即可通过 Bi-RNN 实现有效的缺失插补.然而,基于 RNN 的插补方法仍然存在较大的误差,难以满足大部分场景的插补需求,并且随着网络层数的增加容易引发梯度爆炸和消失等问题.

Luo 等人<sup>[7]</sup>提出了一种基于生成对抗网络(generative adversarial network, GAN)的时序插补模型,并设计了 GRUI 以及 E<sup>2</sup>GAN<sup>[8]</sup>.然而,基于 GAN 的模型训练方式通常比较复杂,生成器与鉴别器之间的博弈过程容易引发训练不稳定等问题. Fortuin 等人<sup>[9]</sup>提出一种基于变分自编码器(variational autoencoder, VAE)的时序插补模型 GP-VAE.然而基于 VAE 的插补模型输出的样本质量容易受到潜在空间结构的限制.因此,相比于之前的研究工作,基于 GAN 和 VAE 架构的模型虽然成功地将时序插补的准确率提升到一个全新的高度,但是两者在训练阶段中仍可能遭遇模式崩溃等问题<sup>[10]</sup>. Vaswani 等人<sup>[11]</sup>于 2017 年提出的 Transformer 在各种自然语言处理任务中证明了其出色的性能,并被广泛应用于其他研究领域<sup>[12]</sup>,包括时间序列插补.例如, Shan 等人<sup>[13]</sup>提出的 NRTSI 模型直接利用 Transformer 编码器部分对不规则采样的时间序列进行建模.另一方面,随着人们对概率建模的研究逐渐深入,扩散模型在生成任务中取得了一系列卓越的成就.例如 Tashiro 等人<sup>[14]</sup>提出的一种基于扩散的时序插补模型 CSDI 可以实现更精准的时序插补,并且与其他确定性插补方法不同, CSDI 还支持概率性插补.然而,上述的时序插补方法仅支持学习时域特征中的潜在信息,这将导致模型学习的能力无法得到充分的发挥.

为了进一步提升时序插补的准确率,同时尽可能避免梯度爆炸和模式崩溃等问题,本文提出了一种基于扩散与时频注意力的时序插补模型 DTFA.受扩散模型的启发,本研究在训练过程中通过随机掩码策略移除部分观测值,并使用前向扩散得到的高斯噪声填

补缺失部分.接着 DTFA 中的噪声估计网络将学习根据现有数据来估计缺失部分的高斯噪声,在此过程中实现对时序数据的建模.在真实插补场景中,DTFA 将使用高斯噪声  $\epsilon$  填补时序数据的缺失片段,预训练的噪声估计网络则根据现有数据推测出一个估计噪声  $\epsilon_\theta$ ,最后通过逆向扩散对高斯噪声  $\epsilon$  进行去噪,逐渐将其转换为插补数据,实现从噪声  $\epsilon$  中重建数据.

为使 DTFA 兼顾学习时域和频域信息,本研究在噪声估计网络中分别引入了时域注意力模块和频域注意力模块,以便 DTFA 更全面地学习多维特征在时域和频域中的时间依赖性与特征相关性,以及更好地识别和理解复杂的数据模式.除此之外,为了给噪声估计网络提供更好的全局背景,DTFA 还采用了一种增强控制条件的方法.该方法通过线性插补模块对残缺的时序样本计算出一个粗略但有效的线性插补结果,并将其作为噪声估计网络的控制条件之一.

本文的贡献主要有如下几点.

(1) 本文提出了一种基于扩散与时频注意力时序插补模型 DTFA.该模型通过时域注意力模块和频域注意力模块捕获时序数据在时域和频域中的时间依赖性与特征相关性,并进行了有效的特征融合.

(2) 本文设计了一种增强控制条件的方法,通过对样本中缺失的部分进行初步的线性插补,以提高数据的完整性,并为噪声估计网络提供了更强的全局信息.

(3) 本文在 EETm1<sup>[15]</sup>、Electricity<sup>[16]</sup>和 Wind Power<sup>[17]</sup>这3个公开的数据集上对 DTFA 进行验证,并使用9种不同缺失程度的 EETm1 数据集进行更细粒度的对比,误差指标均低于一系列的基线模型,证明了 DTFA 的通用性和可靠性.

## 2 相关工作

### 2.1 时间序列插补

多元时间序列是一种具有时间依赖性的数据,被广泛应用于现实世界中的电力系统监测、交通状态监测和生态系统监测<sup>[18-20]</sup>等场景中.然而,传感器故障引起时序数据丢失是一个常见问题,这将导致下游任务难以开展和推进.时序插补的主要目标是捕获现有数据的时空依赖性并推理缺失位置的数据.在此,本文总结了几类先前的时序插补相关工作.

起初,由于 RNN 的结构特性比较适用于序列数据,因此被研究人员广泛应用于时序插补任务中. Che

等人<sup>[4]</sup>提出一种基于 RNN 的插补方法 GRU-D,它属于 GRU 的一种变体.同时,作者还提出了时间衰减机制的概念,即利用最后一次观测值与整体平均值作为插补的依据. Yoon 等人<sup>[5]</sup>提出了 M-RNN,尝试通过从双向 RNN 的隐藏状态中捕获缺失特征之间的相关性来插补缺失值.但是由于 M-RNN 将缺失值视为常量,意味着它忽略了缺失变量之间的联系.与 M-RNN 类似, Cao 等人<sup>[6]</sup>提出的 BRITS 也同样利用了双向 RNN 来对时间序列进行建模,然而 BRITS 将缺失值视为 RNN 图的变量,并且还特别关注了特征之间的相关性.另外, M-RNN 和 BRITS 同样采用了 GRU-D 的时间衰减机制以提高插补性能.

随着数据和计算需求的指数级增长,神经网络也在深度和广度上稳步增长.以 RNN 为主要架构的模型因其顺序处理特性或内存限制,在处理较长序列数据时可能面临可扩展性问题,甚至引起梯度消失或爆炸.因此,研究者们尝试将其他研究领域的优秀研究方法迁移到时序插补领域之中,如基于 GAN 的方法和基于 VAE 的方法.

Yoon 等人<sup>[21]</sup>提出一种基于 GAN 的插补方法 GAIN (generative adversarial imputation net),利用提示向量进行数据生成,但是在时序数据上的适用性欠佳. Luo 等人<sup>[7]</sup>提出了一种 RNN 单元,名为 GRUI (gated recurrent unit for data imputation),并利用 GRUI 层和全连接层构建生成器与鉴别器,以共同组成 GRUI-GAN 模型整体结构.然而该方法需要分为两个阶段来训练,并且第2阶段需要投入大量时间寻找最佳匹配的输入向量.因此, Luo 等人<sup>[8]</sup>又提出了一种端到端的模型 E<sup>2</sup>GAN (end-to-end GAN-based imputation model),其生成器由基于 GRUI 的自动编码器构成,以避免 GRUI-GAN 中的“噪声”优化阶段.另外,作者采用了压缩和重构策略,以减少训练的时间并降低训练的难度. Fortuin 等人<sup>[9]</sup>受 GPPVAE 和 HI-VAE 的启发,提出了第1个基于 VAE 的时序插补方法 GP-VAE.它使用高斯过程先验来捕捉时间动态,并且可以实现概率性插补.然而,基于 GAN 的模型可能会遇到难以收敛和模式崩溃的问题<sup>[22]</sup>,基于 VAE 的模型也可能因隐变量与数据分布不匹配而导致难以学习复杂数据的情况,从而使模型的插补误差较大.

### 2.2 基于注意力的时间序列插补

以注意力结构为基础的 Transformer 结构在许多

任务中都表现出了卓越的性能。目前,越来越多研究者致力于利用自注意力机制来设计时序插补模型,如 GLIMA、CDSA、DeepMVI 和 MTSIT 等<sup>[23]</sup>。最近,Shan 等人<sup>[13]</sup>提出了一种更加新颖的时序插补模型 NRTSI (non-recurrent time series imputation),将时间序列表示为一个集合,直接利用 Transformer 编码器进行建模,并通过一种分层插补的策略来进行缺失插补。Du 等人<sup>[24]</sup>提出了一种基于联合优化训练的方法 SAITS (self-attention-based imputation for time series),通过两个对角掩码自注意力块来进行插补。然而 SAITS 对于非线性关系的建模能力较弱,且对超参数的选择较为敏感。

### 2.3 基于扩散的时间序列插补

扩散概率模型作为新兴的深度生成模型之一,在图像生成和音频生成等多个领域均证明了其优秀的性能。目前扩散模型在时间序列领域的预测任务中已有一些应用先例<sup>[17,25]</sup>,然而在时序插补任务中的应用仍然较少。Kong 等人<sup>[26]</sup>提出一种用于音频生成的模型 DiffWave,由于音频属于一种特殊的时间序列,因此将其进行适当的修改,即可应用于时序插补领域。例如, Tashiro 等人<sup>[14]</sup>提出了一个多元时间序列的概率插补模型 CSDI (conditional score-based diffusion models for imputation),初步展示了扩散模型在时序插补领域中的性能优势。该模型将现有数据作为观测条件来插补缺失数据,并结合注意力模块捕获数据之间的关联性。Liu 等人<sup>[27]</sup>提出了 PriSTI (spatiotemporal imputation with enhanced prior modeling) 模型,通过条件特征提取模块将提取的时空依赖性作为全局上下文先验。

然而,所有上述包含 RNN 的插补模型都容易受到误差累积的影响<sup>[28]</sup>,即当前的插补结果可能是由一系列错误的历史插补结果依次推理的,而基于 GAN 与 VAE 的模型也容易引发模式崩溃和难以收敛的问题。除此之外,上述所有模型仅考虑从时域中获取信息,并没有挖掘频域中的隐含信息,这将不利于模型学习深层的数据关系。受上述相关工作的启发,本论文提出一种基于扩散与时频注意力的时间序列插补模型 DTFA。该模型在已捕获时域特征的基础之上,充分挖掘了频域特征,通过从更丰富的角度学习时序数据的时间依赖性和特征相关性,从而提高插补的准确率。此外,本文还提出了一种增强控制条件的方法,通过线性插补模块对现有数据进行初步插补,从而为噪声估计网络提供更完整的控制条件以及更有效的全局信息。

## 3 方法

本文提出的基于扩散与时频注意力的时间序列插补模型 DTFA 整体结构如图 2 所示。本节将首先介绍时序插补任务中的一些关键定义,然后简要介绍 DTFA 中使用扩散思想插补时序数据的原理和设计思路,最后将介绍 DTFA 中核心组成部分的结构,如线性插补模块、噪声估计网络及其内部的时域注意力模块和频域注意力模块。

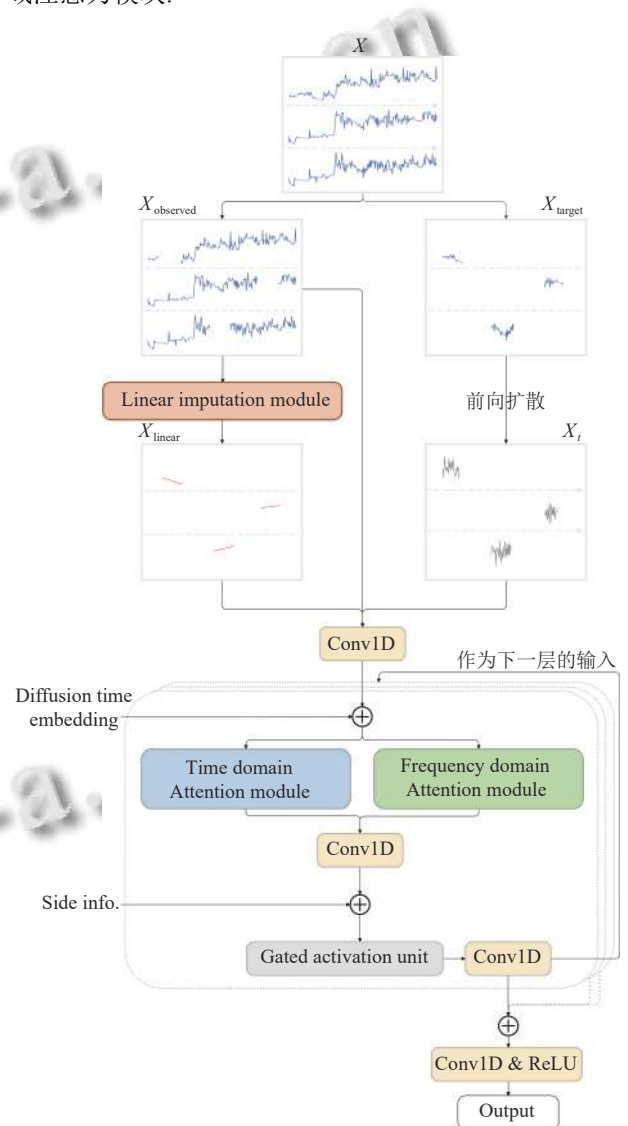


图 2 DTFA 整体结构图

### 3.1 任务定义

本文将每个多元时序样本表示为  $X \in \mathbb{R}^{L \times D}$ , 其中  $L$  表示样本的采样点数,  $D$  表示样本的特征维数, 则  $X^{l,d}$  表示在第  $l$  时刻所观测到第  $d$  维特征的原始数据。

MLM (masked language modeling) 在自然语言处

理领域中是一种常见的自监督学习方法<sup>[29]</sup>,在本研究的时序插补任务中依然可以发挥作用.其基本思想是在时间序列样本中随机选择一部分数据进行掩码标记,并利用模型预测掩码位置的数据.本研究将通过该自监督训练的方式,迫使 DTFA 学习时间序列中的模式以及上下文信息.具体而言,本研究对每个样本 $X$ 都利用与之对应的掩码矩阵 $M \in \mathbb{R}^{L \times D}$ 进行随机掩码标记,其中 $M$ 的每个元素定义如下:

$$M^{l,d} = \begin{cases} 1, & \text{若 } X^{l,d} \text{ 可以观测} \\ 0, & \text{若 } X^{l,d} \text{ 视为缺失} \end{cases} \quad (1)$$

因此, $X$ 可由 $M$ 划分为两个部分:

$$X = X_{\text{observed}} + X_{\text{target}} \quad (2)$$

$$X_{\text{observed}} = M \odot X \quad (3)$$

$$X_{\text{target}} = (1 - M) \odot X \quad (4)$$

其中, $\odot$ 表示矩阵点乘. $X_{\text{observed}}$ 将作为模型的观测条件,而 $X_{\text{target}}$ 在训练阶段将通过 $t$ 步前向扩散计算带噪数据 $X_t$ ,在测试阶段则用于评估模型插补结果 $X_{\text{impute}} = \text{DTFA}(X_{\text{observed}})$ 的误差.

最后,可以将插补后的样本表示为:

$$\hat{X} = M \odot X_{\text{observed}} + (1 - M) \odot X_{\text{impute}} \quad (5)$$

### 3.2 线性插补模块

线性插补模块将对残缺的样本进行初步的插补,以保证观测条件的完整性和连续性.具体而言,在给定的 $X_{\text{observed}}$ 中,对于任意缺失片段左右两个已知数据点 $(l_1, X_{\text{observed}}^{l_1, d_n})$ 和 $(l_2, X_{\text{observed}}^{l_2, d_n})$ ,线性插补模块将通过两者之间的直线以粗略地估计中间缺失区域每个点的数值 $(l, X_{\text{linear}}^{l, d_n})$ ,其中 $l_1 < l < l_2$ ,  $1 \leq n \leq D$ ,公式如下:

$$X_{\text{linear}}^{l, d_n} = X_{\text{observed}}^{l_1, d_n} + \frac{(l - l_1)(X_{\text{observed}}^{l_2, d_n} - X_{\text{observed}}^{l_1, d_n})}{l_2 - l_1} \quad (6)$$

这种基于线性插补的观测条件增强方法具有计算高效、保留数据趋势和时空一致性的优点,特别是在缺失片段中数据变化平缓或者缺失值较少的情况下,它能够简单高效地解决了观测信息完整性的问题,并为下游的噪声估计网络提供可靠的全局先验信息.另外, $X_{\text{linear}}$ 与 $X_{\text{observed}}$ 在 DTFA 训练和测试的全过程中都是可见的.

### 3.3 DTFA 模型

基于物理热力学扩散思想的扩散模型,类似一个更深层次的变分自编码器,包括了前向扩散和反向扩

散两个过程<sup>[30]</sup>.

在 DTFA 的训练阶段,本研究首先对数据集执行滑动窗口取样,将原始样本记为 $X$ ,并根据随机掩码将 $X$ 分解为互补的 $X_{\text{observed}}$ 和 $X_{\text{target}}$ 两部分.接着使用 $\varepsilon \sim N(0, I)$ 对 $X_{\text{target}}$ 执行前向扩散过程得到带噪数据 $X_t$ .由于扩散模型是基于马尔可夫过程的数学假设的,因此根据马尔可夫链的性质可知,只要向样本中加入足够多次的噪声,最终将得到纯随机噪声分布的数据 $X_t$ ,概率分布 $q(X_t | X_{\text{target}})$ 也将趋于稳定的标准高斯分布.本文将 $q_{\text{data}}(X_0)$ 定义为原始数据的分布, $X_t$ 则表示原始数据 $X_0$ 经过了 $t$ 次前向扩散后的带噪数据,其中 $t \in [1, T]$ , $T$ 是人为设定的超参数.则前向扩散过程可以表示为:

$$q(X_{1:T} | X_0) = \prod_{t=1}^T q(X_t | X_{t-1}) \quad (7)$$

$$q(X_t | X_{t-1}) = N(X_t; \sqrt{1 - \beta_t} X_{t-1}, \beta_t I) \quad (8)$$

其中, $\beta_t$ 表示所添加噪声的方差,是可以自行调整的超参数.为了方便表示,本文将 $1 - \beta_t$ 记为 $\hat{\alpha}_t$ ,将 $\prod_{i=1}^t \hat{\alpha}_i$ 记为 $\alpha_t$ ,标准高斯噪声记为 $\varepsilon \sim N(0, I)$ .则采样 $X_t$ 具有以下闭合解:

$$q(X_t | X_0) = N(X_t; \sqrt{\alpha_t} X_0, (1 - \alpha_t) I) \quad (9)$$

$$X_t = \sqrt{\alpha_t} X_0 + (1 - \alpha_t) \varepsilon \quad (10)$$

接着,线性插补模块将根据残缺的 $X_{\text{observed}}$ 计算出线性插补条件 $X_{\text{linear}}$ .最后将采用一维卷积模块对 $X_{\text{linear}}$ 、 $X_{\text{observed}}$ 与 $X_t$ 进行特征融合后输入到噪声估计网络之中,以估计 $X_t$ 在各个时间步中的噪声 $\varepsilon$ .

训练扩散模型的本质是训练其中的噪声估计网络 $\varepsilon_\theta$ .本研究设计的噪声估计网络由3个残差层组成,并在每层都引入了时域注意力模块与频域注意力模块,便于更细粒度地提取数据中的深层时频特征,以实现更准确的噪声估计.本文将在第3.4节与第3.5节对时域注意力模块与频域注意力模块的内部结构进行详细介绍. DTFA 的训练目标 $\min_{\theta} \mathcal{L}(\theta)$ 可表示为:

$$\min_{\theta} E_{X_0 \sim q_{\text{data}}(X_0), \varepsilon \sim N(0, I)} \|\varepsilon - \varepsilon_\theta(X_t, X_{\text{observed}}, X_{\text{linear}}, t)\|^2 \quad (11)$$

DTFA 的测试阶段与训练阶段不同之处在于,带噪数据 $X_t$ 不再通过对 $X_{\text{target}}$ 执行前向扩散获取,取而代之的是随机采样的高斯噪声 $\varepsilon \sim N(0, I)$ ,并且 $X_{\text{target}}$ 将作

为 ground truth 以评估 DTFA 的插补误差。另外,噪声估计网络根据输入的带噪数据  $X_t$ 、 $X_{\text{observed}}$  和  $X_{\text{linear}}$  得到估计噪声  $\varepsilon_\theta$ , 将用于对  $X_t$  执行反向扩散过程进行去噪, 以实现掩码数据的重建。

DTFA 的反向扩散过程是一种对高斯噪声样本  $\varepsilon \sim N(0, I)$  逐步去噪的生成过程。与前向扩散的不同之处在于, 反向扩散过程是由一系列神经网络参数化的高斯分布所组成的马尔可夫链实现的, 即本研究的噪声估计网络  $\varepsilon_\theta$ 。

预训练的噪声估计网络  $\varepsilon_\theta$  在  $X_{\text{observed}}$  和  $X_{\text{linear}}$  引导下, 将反向扩散过程直接应用于标准高斯噪声  $\varepsilon = X_t \sim N(0, I)$ , 使  $X_t$  转变为略微去噪后的带噪数据  $X_{t-1}$ , 该反向扩散过程可以表示为:

$$p_\theta(X_{0:T} | X_{\text{observed}}, X_{\text{linear}}) = p(X_T) \prod_{t=1}^T p_\theta(X_{t-1} | X_t, X_{\text{observed}}, X_{\text{linear}}) \quad (12)$$

$$p_\theta(X_{t-1} | X_t, X_{\text{observed}}, X_{\text{linear}}) = N(X_{t-1}; \mu_\theta(X_t, X_{\text{observed}}, X_{\text{linear}}, t), \sigma_t^2 I) \quad (13)$$

根据 Ho 等人<sup>[31]</sup>提出的一种参数化  $\mu_\theta$  和  $\sigma_t^2$  的方法, 本研究将其表示为:

$$\mu_\theta(X_t, X_{\text{observed}}, X_{\text{linear}}, t) = \frac{1}{\sqrt{\alpha_t}} \left( X_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \varepsilon_\theta(X_t, X_{\text{observed}}, X_{\text{linear}}, t) \right) \quad (14)$$

$$\sigma_t^2 = \frac{1-\alpha_{t-1}}{1-\alpha_t} \beta_t \quad (15)$$

同理, DTFA 将  $X_{t-1}$  再按照上述的反向扩散过程迭代去噪  $t-1$  次, 最终得到一个高度符合数据集分布的时序样本  $X_0$ , 将其记为  $X_{\text{impute}}$  以作为模型的插补结果。本文将在第 4.2 节中介绍根据插补结果  $X_{\text{impute}}$  与 ground truth  $X_{\text{target}}$  评估模型插补性能的方法。

### 3.4 时域注意力模块

本文提出了一种采用多尺度卷积核结构的时域注意力模块, 其主要目标是捕捉时序数据在时域中的重要特征与时空关系, 具体结构如图 3 所示。

本研究设置了 8 个不同大小的卷积核, 分别为 1、2、4、8、16、32、64 和 128。得益于每个卷积核具有不同大小的感受野, DTFA 可以通过多个卷积核的组合识别时序数据中多种尺度的特征, 从而可以更准确地捕获不同时间跨度的依赖性。其中, 较小的卷积核专

注于捕获局部模式与短期依赖关系, 而较大的卷积核则负责捕获全局模式与长期依赖关系。此外, 该设计在特定情况下有助于模型更好地适应不同尺度的数据变化, 可减少 DTFA 对单一尺度特征的过度依赖, 从而提高模型的鲁棒性并降低过拟合的风险。因此, 该特性对于处理具有不同周期性与时序特性的时序数据尤为重要。

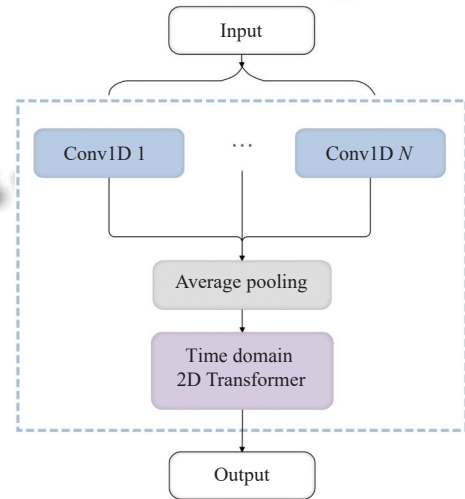


图 3 时域注意力模块

提取到的多尺度特征将通过平均池化层合并为一个统一的表示, 接着输入到时间域二维 Transformer 层中。时间域二维 Transformer 层由时间 Transformer 层和特征 Transformer 层组成, 每层都使用 Transformer 编码器学习整个序列的上下文信息。时间 Transformer 层将沿着时间轴的方向学习各个特征的时间依赖性, 而特征 Transformer 层则沿着特征轴的方向学习各个时刻的特征相关性。

### 3.5 频域注意力模块

由于时域特征仅能表示数据中局部的趋势和模式, 而频域特征通常还包含关于数据周期性和波动性等更丰富的信息, 所以仅依靠时域特征进行学习的模型无法充分发挥其学习能力。因此, 学习时序数据的频域表示将更有利于提取深层特征并提高插补模型的准确率, 特别是对于具有复杂频率特征的时序数据效果更加显著。

本文提出了一种频域注意力模块, 主要包括快速傅里叶变换层 (FFT)、多层感知机 (MLP)、Transformer 层和快速傅里叶逆变换层 (IFFT), 如图 4 所示。

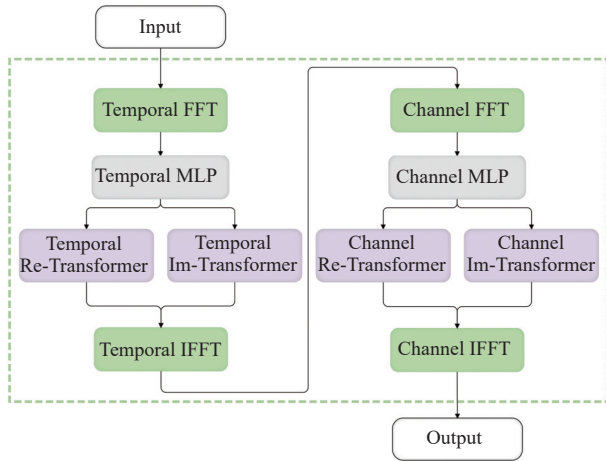


图4 频域注意力模块

在该模块中, Temporal FFT层与 Channel FFT层将时序数据分别在时间维度和通道维度上进行时频域之间的转换. 转换后的频域数据  $Y$  将分别输入 Temporal MLP 与 Channel MLP 进行学习, 实现数据中周期性和频率成分特征的提取, 以及捕捉不同通道之间的潜在关系和共变模式. 该 MLP 的输出可以统一表示为:

$$Y_{MLP} = \sigma(YW + B) \quad (16)$$

其中,  $\sigma$  表示激活函数,  $Y_{MLP} \in \mathbb{C}^{n \times m}$ ,  $Y \in \mathbb{C}^{n \times m}$ ,  $W \in \mathbb{C}^{m \times m}$ ,  $B \in \mathbb{C}^m$ . 由于  $Y$  与  $W$  均为复数矩阵, 根据复数乘法法则, 可以将式 (16) 表示为:

$$\begin{aligned} Y_{MLP} &= \sigma(YW + B) \\ &= \sigma((Y_r + jY_i)(W_r + jW_i) + (B_r + jB_i)) \\ &= \sigma(Y_r W_r + jY_r W_i + jY_i W_r - Y_i W_i + B_r + jB_i) \\ &= \sigma(Y_r W_r - Y_i W_i + B_r) + j\sigma(Y_r W_i + Y_i W_r + B_i) \end{aligned} \quad (17)$$

因此, 本研究将单独计算  $Y_{MLP}$  的实部与虚部部分. 对于计算得到的实数矩阵与虚数矩阵通过 Re-Transformer 层与 Im-Transformer 层学习关键的频率分量特征与模式, 以及抑制噪声频率的影响. 最后将整合后的复数矩阵通过 Temporal IFFT 层或 Channel IFFT 层转换为时域表示.

## 4 实验

### 4.1 数据集

本研究使用 3 个公开的多元时间序列数据集 EETm1、Electricity 和 Wind Power 对 DTFA 进行插补准确率的评估.

EETm1<sup>[15]</sup>: 电力变压器温度数据集中的每个采样点都包含了关于油温以及 6 种不同类型的外部电源负载特征的信息, 时间戳范围为 2016 年 7 月 1 日-2018 年 6 月 26 日. EETm1 数据集为 ETT 数据集预处理之后的一种版本, 数据以 15 min 为间隔均匀采样, 共计 69 680 个采样点. 本研究将 EETm1 的前 4 个月数据作为测试集, 接下来的 4 个月数据作为验证集, 剩余的 16 个月数据作为训练集. 并且本研究通过滑动窗口来进行多元时序样本的取样, 其中窗口大小设置为 24, 滑动步长设置为 12.

Wind Power<sup>[16]</sup>: 该风力发电数据集的时间戳范围为 2020 年 1 月 1 日-2021 年 5 月 22 日, 共计 7 个特征, 数据以 15 min 为间隔均匀采样, 共计 48 673 个样本点. 数据集的划分比例和取样方法与 EETm1 数据集的相同, 即窗口大小设置为 24, 滑动步长设置为 12.

Electricity<sup>[17]</sup>: 源自 UCI 存储库中的 Electricity 电力数据集包含从 370 个客户端收集的用电量数据, 单位为千瓦时. 该数据集以 15 min 为间隔均匀采样, 共计 140 256 个采样点, 每个采样点具有 370 个特征. 在本研究对该数据集通过大小为 100 和步长为 100 的窗口进行取样.

### 4.2 评价指标与实验细节

本研究使用 3 个时序插补领域中常用的评价指标来衡量 DTFA 的性能: 均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 和连续分级概率评分 (continuous ranked probability score, CRPS), 三者的数值越小表明模型的插补性能越好.

RMSE 与 MAE 量化了真实值与预测值之间的差异, 主要用于评估模型在回归任务中的性能. RMSE 偏向关注异常离群值, 而 MAE 可以更好地体现误差的实际情况. 本研究中的 RMSE 和 MAE 计算方式可以表示为:

$$RMSE = \sqrt{\frac{\sum_{l=1}^L \sum_{d=1}^D [(X_{target} - (1-M) \odot X_{impute})^2]^{l,d}}{\sum_{l=1}^L \sum_{d=1}^D (1-M)^{l,d}}} \quad (18)$$

$$MAE = \frac{\sum_{l=1}^L \sum_{d=1}^D |X_{target} - (1-M) \odot X_{impute}|^{l,d}}{\sum_{l=1}^L \sum_{d=1}^D (1-M)^{l,d}} \quad (19)$$

CRPS 是一种评估数据与概率分布差异的指标, 因

此特别适用于评估 DTFA 的插补准确性. 本文参考了 Tashiro 等人<sup>[14]</sup>的计算方法, 将其定义为分位数损失  $\Lambda_\alpha(D^{-1}(\alpha), X) = (\alpha - X_{<D^{-1}(\alpha)}) (X - D^{-1}(\alpha))$  的积分, 其中分位数水平  $\alpha \in [0, 1]$ ,  $D^{-1}(\alpha)$  是分布  $D$  的分位数,  $\mathbb{I}$  为指示函数.

由于本研究设置生成 100 个样本以近似缺失值的分布, 因此将计算 0.05 刻度的离散分位数水平的分位数损失. 最后, 本研究对所有插补结果的 CRPS 计算平均值作为评估指标. 具体计算方式如下:

$$CRPS(D^{-1}, x) = \int_0^1 2\Lambda_\alpha(D^{-1}(\alpha), x) d\alpha$$

$$\approx \sum_{i=1}^{19} 2\Lambda_{i \times 0.05}(D^{-1}(i \times 0.05), x) / 19 \quad (20)$$

$$CRPS(D, \tilde{X}) = \frac{\sum_{\tilde{x} \in \tilde{X}} CRPS(D^{-1}, \tilde{x})}{|\tilde{X}|} \quad (21)$$

本研究中的 DTFA 是基于 PyTorch 框架搭建的, 并在 16 GB 显存的 NVIDIA GeForce RTX 4070 Ti SUPER GPU 计算资源上进行训练、验证以及测试. 实验的参数设置如表 1 所示.

表 1 实验参数设置

参数	设置值	说明
Batchsize	32	批处理大小
Optimizer	Adam	优化器
Learning rate	0.001	学习率
Attention head	8	注意力头数
Layers	4	残差层数
Num steps	50	扩散步数

### 4.3 基线方法

为了验证 DTFA 的性能优越性, 本研究选取了一系列基于深度学习的基线方法与之对比.

M-RNN<sup>[5]</sup>: 该模型使用双向 RNN 络进行插值, 并采用了 GRU-D 的时间衰减函数来捕获信息缺失.

BRITS<sup>[6]</sup>: 该模型同样使用了双向 RNN 以及 GRU-D 的时间衰减函数来对时间序列进行建模, 不同之处在于它将缺失值作为 RNN 图的变量, 接着使用 RNN 的隐藏状态进行插补.

Transformer<sup>[11]</sup>: 在本研究中仅使用了 Transformer 的编码器部分, 去除了 Transformer 的解码器部分.

SAITS<sup>[24]</sup>: 该模型采用了一种掩码插补和观测值重建的联合优化训练方法, 并使用两个对角屏蔽自注意力块和一个加权组合块以对时间序列建模.

GP-VAE<sup>[9]</sup>: 该模型在潜在空间中利用高斯先验过程捕获时间动态, 其 ELBO 仍依赖现有数据的特征.

CSDI<sup>[14]</sup>: 该模型是一种基于分数的条件扩散模型, 通过使用二维注意力机制以进一步捕获时间和特征的相关性.

### 4.4 对比实验

本研究在 ETTm1、Wind Power 和 Electricity 数据集上进行了插补效果的对比实验, 如表 2 所示. 其中测试集的缺失率设置为 10%, 最优的指标通过加粗来表示. 该对比实验表明, DTFA 在 ETTm1、Wind Power 和 Electricity 数据集上插补误差均都低于其他基线模型, 充分证明了 DTFA 通用性和泛化能力. 具体而言, DTFA 在 ETTm1 数据集上的 RMSE 与 MAE 分别为 0.103 和 0.065, 比最优基线模型的误差分别降低了 7.21% 和 4.41%; 在 Wind Power 数据集上的 RMSE 与 MAE 分别为 0.127 和 0.022, 比最优基线模型的误差分别降低了 3.05% 和 15.38%; 在 Electricity 数据集上的 RMSE 与 MAE 分别为 0.122 和 0.053, 比最优基线模型的误差分别降低了 6.87% 和 5.36%.

表 2 DTFA 与基线模型在 ETTm1、Wind Power 和 Electricity 数据集上插补的 RMSE 与 MAE

模型	ETTm1	Wind Power	Electricity
M-RNN	1.395/0.989	0.840/0.668	1.869/1.269
BRITS	0.265/0.142	0.196/0.081	1.346/0.883
Transformer	0.184/0.116	0.199/0.091	1.294/0.821
SAITS	0.183/0.123	0.178/0.060	1.148/0.737
GP-VAE	0.465/0.334	0.408/0.299	1.565/1.094
CSDI	0.111/0.068	0.131/0.026	0.131/0.056
DTFA (Ours)	<b>0.103/0.065</b>	<b>0.127/0.022</b>	<b>0.122/0.053</b>
降低百分比 (%)	7.21/4.41	3.05/15.38	6.87/5.36

M-RNN 因其无法考虑缺失变量之间的关系, 导致插补误差较大, 尽管 BRITS 通过将缺失值视为 RNN 图的变量以捕获特征的相关性, 两者的插补性能也无法摆脱 RNN 结构特性的限制, 从而难以大幅降低插补误差. GP-VAE 可能因隐变量与数据分布匹配困难而导致误差较大. Transformer 与 SAITS 受益于注意力机制, 在 3 个数据集上的插补性能上实现新的突破. CSDI 与 DTFA 是基于扩散架构的插补模型, 两者相较于其他基线模型在 RMSE 和 MAE 指标上具有明显优势, 在特征维度较高的 Electricity 数据集上效果更为显著. 综上所述, DTFA 实现了扩散架构与时频注意力的有效结合, 并对于多种数据场景具有较好的泛化性.



为了进一步比较 DTFA 与基线模型在不同缺失率数据集上的插补性能,本研究向 ETTm1 测试集分别引入 10%–90% 数据缺失以进行验证,表 3–表 5 分别详细展示了 *RMSE*、*MAE* 和 *CRPS* 的评估结果. 该实验结果表明, 尽管面对不同程度缺失的数据集, DTFA 的

插补效果依然优于其他基线方法, 且完整度较高的数据集可以更充分地发挥 DTFA 的学习能力. 具体而言, DTFA 的 *RMSE* 相较于最优的基线模型降低了 0.58%–8.53%, *MAE* 相较于最优的基线模型降低了 0.98%–7.79%, *CRPS* 相较于最优的基线模型降低了 1.08%–7.35%.

表 3 缺失率不同时, DTFA 与基线模型在 ETTm1 数据集上的 *RMSE* 对比实验结果

模型	10%	20%	30%	40%	50%	60%	70%	80%	90%
M-RNN	1.395	1.415	1.420	1.420	1.419	1.415	1.406	1.417	1.405
BRITS	0.265	0.290	0.328	0.357	0.378	0.418	0.448	0.503	0.542
Transformer	0.184	0.225	0.223	0.219	0.221	0.260	0.277	0.313	0.343
SAITS	0.183	0.196	0.208	0.221	0.236	0.265	0.284	0.316	0.350
GP-VAE	0.465	0.534	0.599	0.612	0.670	0.734	0.802	0.912	1.060
CSDI	0.111	0.114	0.121	0.129	0.139	0.152	0.173	0.223	0.362
DTFA (Ours)	<b>0.103</b>	<b>0.107</b>	<b>0.112</b>	<b>0.118</b>	<b>0.130</b>	<b>0.146</b>	<b>0.167</b>	<b>0.217</b>	<b>0.341</b>
降低百分比 (%)	7.21	6.14	7.44	8.53	6.47	3.95	3.47	2.70	0.58

表 4 缺失率不同时, DTFA 与基线模型在 ETTm1 数据集上的 *MAE* 对比实验结果

模型	10%	20%	30%	40%	50%	60%	70%	80%	90%
M-RNN	0.989	1.012	1.019	1.027	1.028	1.030	1.024	1.034	1.029
BRITS	0.142	0.160	0.182	0.205	0.222	0.250	0.270	0.303	0.331
Transformer	0.116	0.139	0.136	0.137	0.143	0.162	0.174	0.193	0.216
SAITS	0.123	0.129	0.139	0.147	0.160	0.180	0.193	0.205	0.231
GP-VAE	0.334	0.384	0.432	0.448	0.491	0.548	0.594	0.683	0.795
CSDI	0.068	0.071	0.077	0.081	0.087	0.095	0.108	0.132	0.204
DTFA (Ours)	<b>0.065</b>	<b>0.068</b>	<b>0.071</b>	<b>0.075</b>	<b>0.081</b>	<b>0.090</b>	<b>0.103</b>	<b>0.128</b>	<b>0.202</b>
降低百分比 (%)	4.41	4.23	7.79	7.41	6.90	5.26	4.63	3.00	0.98

表 5 缺失率不同时, DTFA 与基线模型在 ETTm1 数据集上的 *CRPS* 对比实验结果

模型	10%	20%	30%	40%	50%	60%	70%	80%	90%
GP-VAE	0.386	0.446	0.501	0.519	0.571	0.638	0.692	0.796	0.928
CSDI	0.060	0.062	0.068	0.071	0.077	0.084	0.096	0.118	0.185
DTFA (Ours)	<b>0.057</b>	<b>0.060</b>	<b>0.063</b>	<b>0.066</b>	<b>0.072</b>	<b>0.080</b>	<b>0.092</b>	<b>0.114</b>	<b>0.183</b>
降低百分比 (%)	5.00	3.23	7.35	7.04	6.49	4.76	4.17	3.40	1.08

#### 4.5 消融实验

为了深入研究 DTFA 中不同组件对整体性能的提升, 本研究设计了 3 组消融实验, 并在缺失率为 10% 的 ETTm1 数据集上进行验证, 实验结果如表 6 所示.

表 6 DTFA 在 ETTm1 数据集上的消融实验结果

消融项	<i>RMSE</i>	<i>MAE</i>	<i>CRPS</i>
w/o 时域注意力模块	0.126	0.072	0.081
w/o 频域注意力模块	0.134	0.076	0.114
w/o 线性插补模块	0.107	0.069	0.073
DTFA (Ours)	<b>0.103</b>	<b>0.065</b>	<b>0.057</b>

该消融实验表明, 当去除线性插补模块后, DTFA 仅以现有数据作为辅助条件进行训练, 导致模型的插补误差增大. 该现象可以印证线性插补模块在一定程度上为 DTFA 提供有效的全局信息和缺失数据的潜在变化趋势. 而当去除时域注意力或频域注意力模块后,

模型的准确率大幅下降, 该现象可以印证结合时频特征进行学习有助于更好地挖掘数据的时间依赖性和特征相关性. 综上所述, 线性插补模块、时域注意力模块和频域注意力模块均对 DTFA 的插补性能有着不同程度的性能增益.

#### 4.6 残差层数的对比

为了进一步探索 DTFA 的最优性能与残差层数之间的关系, 本研究将残差层从 1 层逐渐增加至 6 层分别进行验证, 在缺失率为 10% 的 ETTm1 数据集上的实验结果如表 7 所示.

为了突出 3 种评价指标的变化趋势, 本研究对表 7 数据可视化之前对数据进行了最小-最大归一化处理, 如图 5 所示, 其中蓝色线表示 *RMSE*, 橙色线表示 *MAE*, 绿色线表示 *CRPS*. 根据图 5 中各曲线整体趋

势可知,当残差层数仅为1时,DTFA无法充分捕获时间序列中的动态规律,从而无法充分发挥其插补性能。随着残差层数的增加,3种误差指标在整体上具有下降的趋势。当残差层数为4时,模型表现出最佳的性能,而当残差层数为5或更多时,模型的误差呈上升趋势,出现了过拟合的现象,不仅浪费计算资源,而且对模型的性能产生了负面的影响。因此,选择合适的残差层数,可在一定程度上降低模型的插补误差。

表7 缺失率为10%时,不同残差层数的DTFA在ETTm1数据集上的RMSE、MAE与CRPS的对比结果

残差层数	RMSE	MAE	CRPS
1	0.1075	0.0677	0.0600
2	0.1042	0.0649	0.0574
3	0.1030	0.0652	0.0575
4	<b>0.1028</b>	<b>0.0645</b>	<b>0.0568</b>
5	0.1035	<b>0.0645</b>	0.0569
6	0.1033	0.0649	0.0571

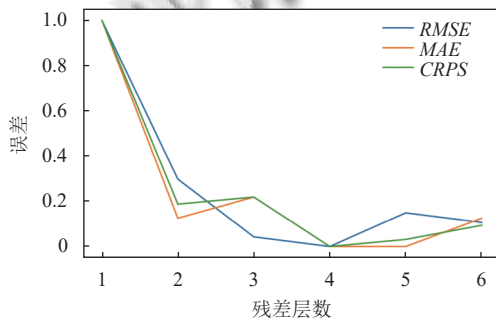


图5 不同残差层数对3种评价指标误差的影响

## 5 结论与展望

本文提出一种基于扩散与时频注意力的时间序列插补模型DTFA,通过线性插补模块进行初步的缺失插补以保证数据的完整性和连续性。为捕捉时序数据在时域和频域中的重要特征和时空关系,本研究将分别通过时域注意力模块与频域注意力模块提取特征信息以学习更高质量的表示。其中DTFA的时域注意力模块利用多尺度卷积核的结构识别时间序列中多种尺度的特征,并利用Transformer编码器捕获不同时间跨度的时间相关性。而DTFA的频域注意力模块则使用FFT获取时序数据的频域表示,并利用MLP与Transformer编码器学习频域特征中潜在的相关性以及依赖性。最后,本研究通过对比实验证明DTFA在ETTm1、WindPower和Electricity数据集上的插补性能均优于其他基线模型,同时通过消融实验证明了DTFA核心

模块的有效性。

在数字经济蓬勃发展与物联网设备不断进步的推动下,人们迫切需要更高准确率的插补方法。由于目前主流的基线模型暂时无法通过简单地加深网络层数实现更准确的插补,因此DTFA成功地解决了时序插补领域中最具有挑战性的难题。在今后的研究中将考虑提高DTFA针对更大规模数据集的扩展性,以及提高DTFA的训练和推理的速度。

## 参考文献

- Wen QS, Sun L, Yang F, *et al.* Time series data augmentation for deep learning: A survey. Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal: Artificial Intelligence Organization, 2021. 4653–4660. [doi: [10.24963/IJCAI.2021/631](https://doi.org/10.24963/IJCAI.2021/631)]
- 葛秋洲, 许翔, 杨锁荣, 等. 序列数据的数据增强方法综述. 计算机科学与探索, 2021, 15(7): 1207–1219. [doi: [10.3778/j.issn.1673-9418.2012062](https://doi.org/10.3778/j.issn.1673-9418.2012062)]
- Yang JS, Shao YH, Li CN, *et al.* A multistage deep imputation framework for missing values large segment imputation with statistical metrics. Applied Soft Computing, 2023, 146: 110654. [doi: [10.1016/j.asoc.2023.110654](https://doi.org/10.1016/j.asoc.2023.110654)]
- Che ZP, Purushotham S, Cho K, *et al.* Recurrent neural networks for multivariate time series with missing values. Scientific Reports, 2018, 8(1): 6085. [doi: [10.1038/s41598-018-24271-9](https://doi.org/10.1038/s41598-018-24271-9)]
- Yoon J, Zame WR, Van Der Schaar M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. IEEE Transactions on Biomedical Engineering, 2019, 66(5): 1477–1490. [doi: [10.1109/TBME.2018.2874712](https://doi.org/10.1109/TBME.2018.2874712)]
- Cao W, Wang D, Li J, *et al.* BRITS: Bidirectional recurrent imputation for time series. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 6776–6786. [doi: [10.48550/arXiv.1805.10572](https://doi.org/10.48550/arXiv.1805.10572)]
- Luo YH, Cai XR, Zhang Y, *et al.* Multivariate time series imputation with generative adversarial networks. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 1603–1614.
- Luo YH, Zhang Y, Cai XR, *et al.* E<sup>2</sup>GAN: End-to-end generative adversarial network for multivariate time series imputation. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Palo Alto: Artificial Intelligence Organization, 2019. 3094–3100.
- Fortuin V, Baranchuk D, Räetsch G, *et al.* GP-VAE: Deep probabilistic time series imputation. Proceedings of the 23rd International Conference on Artificial Intelligence and

- Statistics. PMLR, 2020. 1651–1661.
- 10 Brophy E, Wang ZW, She Q, *et al.* Generative adversarial networks in time series: A systematic literature review. *ACM Computing Surveys*, 2023, 55(10): 1–31. [doi: [10.1145/3559540](https://doi.org/10.1145/3559540)]
  - 11 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010. [doi: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762)]
  - 12 Wen QS, Zhou T, Zhang CL, *et al.* Transformers in time series: A survey. *Proceedings of the 32nd International Joint Conference on Artificial Intelligence*. Macao: Artificial Intelligence Organization, 2023. 6778–6786. [doi: [10.24963/ijcai.2023/759](https://doi.org/10.24963/ijcai.2023/759)]
  - 13 Shan SY, Li Y, Oliva JB. NRTSI: Non-recurrent time series imputation. *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing*. Rhodes Island: IEEE, 2023. 1–5. [doi: [10.1109/ICASSP49357.2023.10095054](https://doi.org/10.1109/ICASSP49357.2023.10095054)]
  - 14 Tashiro Y, Song JM, Song Y, *et al.* CSDI: Conditional score-based diffusion models for probabilistic time series imputation. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc., 2021. 24804–24816. [doi: [10.48550/arXiv.2107.03502](https://doi.org/10.48550/arXiv.2107.03502)]
  - 15 Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient Transformer for long sequence time-series forecasting. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI, 2021. 11106–11115. [doi: [10.1609/aaai.v35i12.17325](https://doi.org/10.1609/aaai.v35i12.17325)]
  - 16 Nie YQ, Nguyen NH, Sinthong P, *et al.* A time series is worth 64 words: Long-term forecasting with Transformers. *Proceedings of the 11th International Conference on Learning Representations*. OpenReview.net, 2023. [doi: [10.48550/arXiv.2211.14730](https://doi.org/10.48550/arXiv.2211.14730)]
  - 17 Li Y, Lu XJ, Wang YQ, *et al.* Generative time series forecasting with diffusion, denoise, and disentanglement. *Proceedings of the 36th Conference on Neural Information Processing Systems*. New Orleans: OpenReview.net, 2022. [doi: [10.48550/arXiv.2301.03028](https://doi.org/10.48550/arXiv.2301.03028)]
  - 18 郑欣彤, 边婷婷, 张德强, 等. 基于深度学习的温度观测数据长时间缺失值插补方法. *计算机系统应用*, 2022, 31(4): 221–228. [doi: [10.15888/j.cnki.csa.008493](https://doi.org/10.15888/j.cnki.csa.008493)]
  - 19 Iwana BK, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. *PLoS One*, 2021, 16(7): e0254841. [doi: [10.1371/journal.pone.0254841](https://doi.org/10.1371/journal.pone.0254841)]
  - 20 Iglesias G, Talavera E, González-Prieto Á, *et al.* Data augmentation techniques in time series domain: A survey and taxonomy. *Neural Computing and Applications*, 2023, 35(14): 10123–10145. [doi: [10.1007/s00521-023-08459-3](https://doi.org/10.1007/s00521-023-08459-3)]
  - 21 Yoon J, Jordon J, van der Schaar M. GAIN: Missing data imputation using generative adversarial nets. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018. 5689–5698. [doi: [10.48550/arXiv.1806.02920](https://doi.org/10.48550/arXiv.1806.02920)]
  - 22 Miao XY, Wu YY, Wang J, *et al.* Generative semi-supervised learning for multivariate time series imputation. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI, 2021. 8983–8991. [doi: [10.1609/aaai.v35i10.17086](https://doi.org/10.1609/aaai.v35i10.17086)]
  - 23 Yıldız AY, Koç E, Koç A. Multivariate time series imputation with Transformers. *IEEE Signal Processing Letters*, 2022, 29: 2517–2521. [doi: [10.1109/LSP.2022.3224880](https://doi.org/10.1109/LSP.2022.3224880)]
  - 24 Du WJ, Côté D, Liu Y. SAITS: Self-attention-based imputation for time series. *Expert Systems with Applications*, 2023, 219: 119619. [doi: [10.1016/j.eswa.2023.119619](https://doi.org/10.1016/j.eswa.2023.119619)]
  - 25 Rasul K, Seward C, Schuster I, *et al.* Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021. 8857–8868. [doi: [10.48550/arXiv.2101.12072](https://doi.org/10.48550/arXiv.2101.12072)]
  - 26 Kong ZF, Ping W, Huang JJ, *et al.* DiffWave: A versatile diffusion model for audio synthesis. *Proceedings of the 2021 International Conference on Learning Representations*. Vienna: OpenReview.net, 2021. [doi: [10.48550/arXiv.2009.09761](https://doi.org/10.48550/arXiv.2009.09761)]
  - 27 Liu MZ, Huang H, Feng H, *et al.* PriSTI: A conditional diffusion framework for spatiotemporal imputation. *Proceedings of the 39th IEEE International Conference on Data Engineering*. Anaheim: IEEE, 2023. 1927–1939. [doi: [10.1109/ICDE55515.2023.00150](https://doi.org/10.1109/ICDE55515.2023.00150)]
  - 28 Liu YK, Yu R, Zheng S, *et al.* NAOMI: Non-autoregressive multiresolution sequence imputation. *Proceedings of the 33rd Conference on Neural Information Processing Systems*. Vancouver: ACM, 2019. 11238–11248. [doi: [10.48550/arXiv.1901.10946](https://doi.org/10.48550/arXiv.1901.10946)]
  - 29 Dong JX, Wu HX, Zhang HR, *et al.* SimMTM: A simple pre-training framework for masked time-series modeling. *Proceedings of the 37th Conference on Neural Information Processing Systems*. New Orleans: OpenReview.net, 2024. [doi: [10.48550/arXiv.2302.00861](https://doi.org/10.48550/arXiv.2302.00861)]
  - 30 Lin LQ, Li ZK, Li RK, *et al.* Diffusion models for time-series applications: A survey. *Frontiers of Information Technology & Electronic Engineering*, 2024, 25(1): 19–41. [doi: [10.1631/FITEE.2300310](https://doi.org/10.1631/FITEE.2300310)]
  - 31 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 6840–6851. [doi: [10.48550/arXiv.2006.11239](https://doi.org/10.48550/arXiv.2006.11239)]

(校对责编: 张重毅)