

# 控制与决策

Control and Decision

一种新的轻量化生成对抗网络及其在风电数据插补中的应用

武甲, 齐咏生, 马然, 高胜利, 刘慧文

引用本文:

武甲, 齐咏生, 马然, 等. 一种新的轻量化生成对抗网络及其在风电数据插补中的应用[J]. *控制与决策*, 2024, 39(12): 4141-4150.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1544>

---

您可能感兴趣的其他文章

Articles you may be interested in

[基于聚类簇结构特性的自适应综合采样法在入侵检测中的应用](#)

Toward intrusion detection via cluster structure-based adaptive synthetic sampling approach  
*控制与决策*. 2021, 36(8): 1920-1928 <https://doi.org/10.13195/j.kzyjc.2019.1672>

[基于自注意力生成对抗网络的图像超分辨率重建](#)

Image super-resolution reconstruction based on self-attention GAN  
*控制与决策*. 2021, 36(6): 1324-1332 <https://doi.org/10.13195/j.kzyjc.2019.1290>

[基于数据驱动的非线性网络系统自适应迭代学习控制](#)

Data driven adaptive learning control of nonlinear network system  
*控制与决策*. 2021, 36(6): 1523-1528 <https://doi.org/10.13195/j.kzyjc.2019.1182>

[面向工业软测量应用的定制化生成对抗数据填补模型](#)

Customized generative adversarial data imputation model for industrial soft sensing  
*控制与决策*. 2021, 36(12): 2929-2936 <https://doi.org/10.13195/j.kzyjc.2020.0974>

[基于生成对抗网络的大规模路网交通流预测算法](#)

Traffic flow forecasting algorithm for large-scale road network based on GAN  
*控制与决策*. 2021, 36(12): 2937-2945 <https://doi.org/10.13195/j.kzyjc.2020.0333>

# 一种新的轻量化生成对抗网络及其在 风电数据插补中的应用

武甲<sup>1,2,3</sup>, 齐咏生<sup>1,2,3†</sup>, 马然<sup>1,2,3</sup>, 高胜利<sup>4</sup>, 刘慧文<sup>1,2,3</sup>

1. 内蒙古工业大学 电力学院, 呼和浩特 010051;
2. 大规模储能技术教育部工程研究中心, 呼和浩特 010080;
3. 内蒙古自治区高等学校智慧能源技术与装备工程研究中心, 呼和浩特 010080;
4. 内蒙古北方龙源风力发电有限责任公司, 内蒙古 乌兰察布 013550)

**摘要:** 针对风电场监控和数据采集系统(supervisory control and data acquisition, SCADA)数据在采集传输过程中常遇到的数据丢失问题,提出一种新的自适应轻量化生成对抗网络插补策略(adaptive transformer slim GAIN, AT-SGAIN),旨在增强数据完整性。AT-SGAIN通过简化GAIN模型结构,显著提高了计算效率;采用双判别器结构,分别用于真实数据和生成数据的鉴别,保障了速度提升过程中插补精度的维护。算法集成了Transformer(变压器模型)编码器,增强了对风电数据时间序列特征的捕捉能力,并通过自适应双分支注意力机制,精准调整通道和空间注意力权重,提升了网络对局部信息的敏感度。实验结果证明,所提算法在多项对比测试中均显著优于现有经典方法。

**关键词:** 风电数据插补; 监控和数据采集系统数据; 轻量化; 变压器模型; 自适应注意力机制; 自适应轻量化生成对抗插补网络

中图分类号: TM614 文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1544

引用格式: 武甲, 齐咏生, 马然, 等. 一种新的轻量化生成对抗网络及其在风电数据插补中的应用[J]. 控制与决策, 2024, 39(12): 4141-4150.

## A new lightweight generative adversarial network and its application in wind power data interpolation

WU Jia<sup>1,2,3</sup>, QI Yong-sheng<sup>1,2,3†</sup>, MA Ran<sup>1,2,3</sup>, GAO Sheng-li<sup>4</sup>, LIU Hui-wen<sup>1,2,3</sup>

1. School of Electric Power, Inner Mongolia University of Technology, Hohhot 010051, China;
2. Engineering Research Center of the Ministry of Education for Large Scale Energy Storage Technology, Hohhot 010080, China;
3. Intelligent Energy Technology and Equipment Engineering Research Center of Higher Education Institutions in Inner Mongolia Autonomous Region, Hohhot 010080, China;
4. Inner Mongolia North Longyuan Wind Power Co., Ltd., Ulanqab 013550, China)

**Abstract:** A lightweight generative adversarial network interpolation strategy based on adaptive transformer slim GAIN(AT-SGAIN) is proposed to address the common problem of data loss in the collection and transmission of supervisory control and data acquisition(SCADA) data in wind farms, aiming to enhance data integrity. AT-SGAIN simplifies the GAIN model structure, significantly improves computational efficiency, and adopts a dual discriminator structure for distinguishing between real data and generated data, ensuring the maintenance of interpolation accuracy during the speed improvement process. This model integrates a Transformer encoder, enhancing the ability to capture time series features of wind power data. Through an adaptive dual branch attention mechanism, it accurately adjusts channel and spatial attention weights, improving the network's sensitivity to local information. The experimental results show that, in multiple comparative tests, the algorithm proposed is significantly better than existing classical methods.

**Keywords:** wind power data interpolation; SCADA data; lightweight; Transformer; adaptive attention mechanism; AT-SGAIN

收稿日期: 2023-11-06; 录用日期: 2024-04-07.

基金项目: 国家自然科学基金项目(62363029, 62241309); 内蒙古自治区高等学校科学研究项目(NJZY22365); 内蒙古自然科学基金项目(2023LHMS06005, 2022MS06018).

†通讯作者. E-mail: qys@imut.edu.cn.

## 0 引言

风电场监控和数据采集系统(supervisory control and data acquisition, SCADA)数据对于风电机组的运行与维护至关重要<sup>[1]</sup>,尤其在状态监测<sup>[2]</sup>和故障诊断<sup>[3]</sup>方面发挥了关键作用.然而,传感器故障和通信错误引发的数据缺失问题,直接影响了数据的有效应用.现有主流插补方法包括数据删除法<sup>[4]</sup>和数据插补法<sup>[5]</sup>,前者可能丢失关键信息,后者旨在最小化信息损失.传统插补方法,如插值法<sup>[6]</sup>、相似性法<sup>[7]</sup>及参数估计法<sup>[8]</sup>,虽适用于特定场景,但面对大规模或多属性数据时准确性有限.近年来,机器学习及深度学习方法,例如循环神经网络(RNNs)<sup>[9]</sup>、随机森林<sup>[10]</sup>、反向传播(BP)<sup>[11]</sup>和长短期记忆网络(LSTM)<sup>[12]</sup>,因其较高的准确度而受到重视,但计算复杂度较高.

近年,生成对抗网络(GAN)的深度学习策略在数据插补领域取得显著成就,尤其是在模拟数据潜在分布以估计缺失值方面<sup>[13]</sup>.尽管面临着需要大量完整数据训练的挑战以及在长时序数据处理上有局限性<sup>[14]</sup>,研究界仍提出多种改进模型以增强性能.GAMIN<sup>[15]</sup>旨在通过深度模型(如CNN)的反向攻击提高数据连贯性;GAGIN<sup>[16]</sup>通过Global-Impute-Net等结构提升模型稳定性,但在处理复杂数据时仍有挑战;Two-stage GAN<sup>[17]</sup>通过双阶段训练提升插补准确性,但训练过程较为缓慢;E2GAN<sup>[18]</sup>和IMGAN<sup>[19]</sup>分别通过端到端生成模型和逆映射学习网络提升插补效率,但可能受到训练数据量的限制;WGAIN<sup>[20]</sup>可以有效维持数据分布,但在RMSE评价指标上可能逊色,由于其过度专注于分布模拟而忽略了误差最小化;GRAPE<sup>[21]</sup>虽展现出较强的泛化性,但性能易受数据缺失率的影响;Smoothed LSTM-AE<sup>[22]</sup>结合了LSTM和AE优化时序数据缺失值处理,展现了稳定性和准确性,提供了复杂数据插补的新视角.

基于此,本文提出一种新的自适应轻量化生成对抗插补网络(adaptive transformer slim GAIN, AT-SGAIN),AT-SGAIN通过对GAIN网络结构进行轻量化优化(即SGAIN),并引入自适应注意力机制的Transformer编码器(即AT),极大提升了对长时间序列风电数据的分析效率,并增强了局部信息处理能力.实验验证于内蒙古某风电场的真实SCADA数据集,表明了其在实际应用中的有效性.

## 1 算法基础

### 1.1 GAIN模型

GAIN是一种基于生成对抗网络GAN的数据插补方法<sup>[23]</sup>,将数据插补问题转变为数据生成问题.

生成器的主旨在于推测数据的缺失部分,通过结合原始数据、随机噪声,产生估计的插补值.判别器则评估数据的真实性,估算数据为真实的概率.这两者的互动和竞争促进了数据的高效插补.

生成器的核心公式可以表示为

$$\bar{X} = G(\tilde{X}, M, (1 - M) \odot Z), \quad (1)$$

$$\hat{X} = M \odot \tilde{X} + (1 - M) \odot \bar{X}. \quad (2)$$

其中: $\bar{X}$ 是有缺失的原始数据, $M$ 是掩码向量, $Z$ 是随机噪声, $\tilde{X}$ 是生成器输出的估计值, $\hat{X}$ 是最终的插补数据.

判别器模型的角色是区分生成数据与真实数据.为了增强判别能力,引入提示向量 $H$ ,将 $H$ 和 $\hat{X}$ 一起输入给判别器.

### 1.2 注意力机制

风电数据插补中,引入注意力机制,模仿了人类对关键信息的集中关注<sup>[24]</sup>.主要的机制包括通道注意力、空间注意力和混合注意力.

通道注意力,受SENet网络启发,通过权重捕获关键的时间依赖性.空间注意力,基于空间变换网络STN,突出关键空间位置.混合注意力如CBAM,综合利用通道和空间信息,筛选出关键特征,从而有效地强化了有价值的信息,并抑制了不重要的信息.

## 2 基于自适应轻量化的GAIN插补网络

### 2.1 总体框架

本文提出AT-SGAIN,旨在提高风电SCADA数据插补的准确性和效率,总体框架如图1所示.

AT-SGAIN网络通过融合轻量化的SGAIN模型与自适应注意力机制的Transformer,将Transformer编码器融入到生成器和判别器中,有效捕获风电数据的长期序列特性及其动态性.同时,引入并行的通道与空间注意力机制,提升对局部变异的敏感度.此外,AT-SGAIN对编解码器结构进行优化,结合多层感知机(MLP)和层归一化(LN)技术,增强处理长期依赖和复杂数据结构的能力,保证了插补的精度.

经处理后的风电数据集被拆分为缺失数据矩阵、随机噪声矩阵及掩码矩阵,其中,随机噪声矩阵进一步按照10个特征分解为10个子矩阵,依次选取某一矩阵为主矩阵,并取其余9个子矩阵构成9组特征矩阵,辅助插补.生成器首先插补第1个特征,通过对第1维特征的9组特征矩阵进行特征提取,并利用通道注意力机制计算9组通道的权重参数,将复合评价指标(composite ratio, CR)与9组通道权重参数进行相乘,得到最终的通道权重.依次再插补第2维特征,直

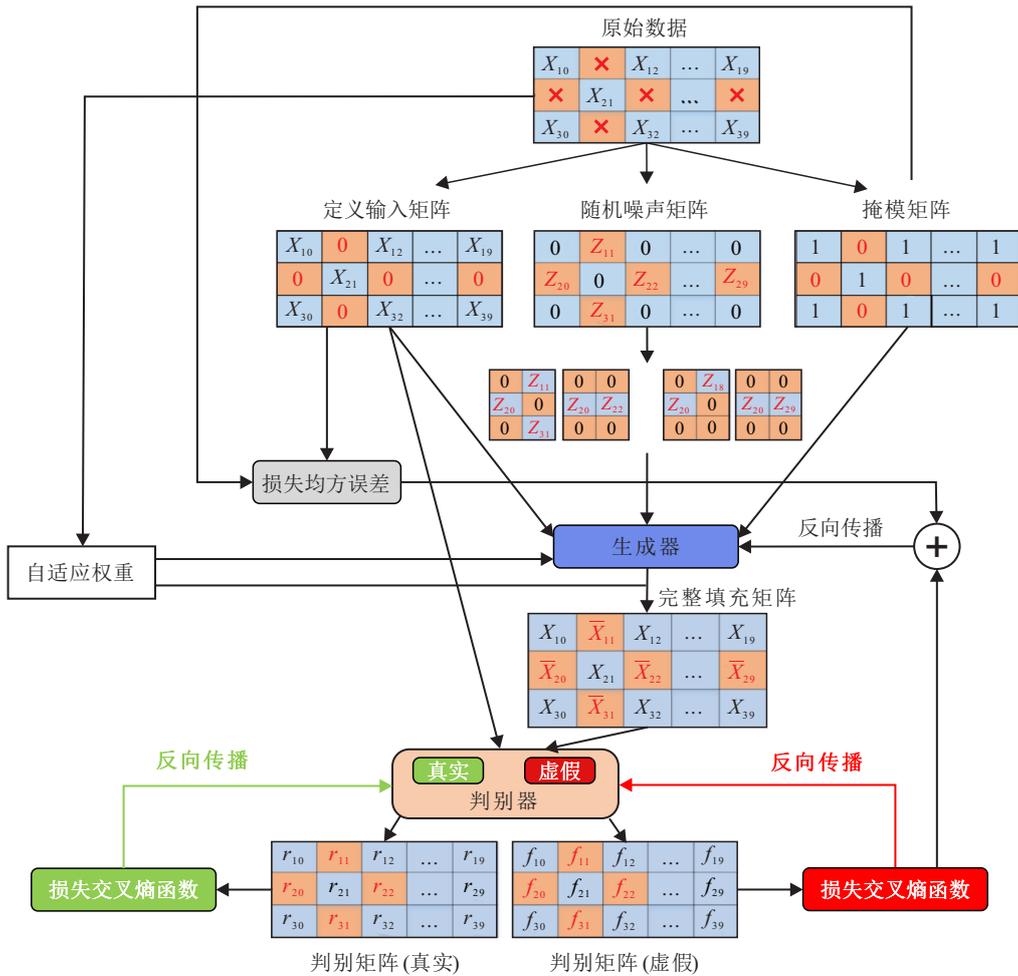


图1 AT-SGAIN总体框架结构

到最后一维结束. 然后通过空间注意力机制结合缺失率(miss rate, MR)优化空间权重,增强插补质量. 此过程生成的插补矩阵由双判别器验证,促进 Loss 函数收敛.

### 2.2 SGAIN 自适应训练与结构设计

SGAIN 模型采用了基于Transformer编码器的创新训练策略,优化了生成器和判别器的结构,实现了高效的数据插补. 模型结构如图2所示.

模型以原始数据、掩模矩阵和随机噪声为输入,生成器通过多层感知机预处理SCADA数据,随后Transformer编码器提取特征. 经过两次上采样和编码,与其生成器和判别器的层数从原始的3层简化至两层的策略保持一致. 判别器用于执行数据拆分并应用Transformer编码器进行特征分析,最终输出一个代表生成数据与真实数据相似度的概率指标. 为提高计算效率和插补准确性,对网络进行了优化,移除了提示矩阵生成器,并采纳双判别器结构进行评估.

选择双曲正切(tanh)作为激活函数,其导数在整

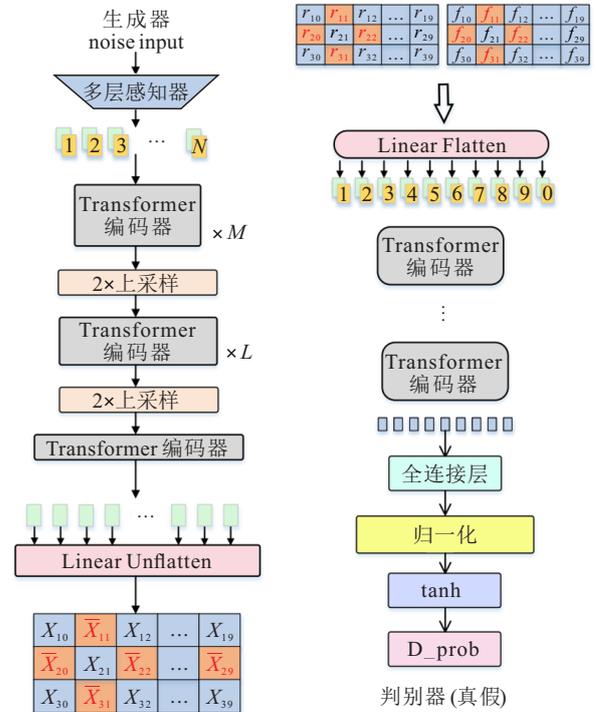


图2 SGAIN模型结构

个输入范围内均超过sigmoid,加速了梯度下降的收敛过程,显著提升了训练效率,对比如图3所示.

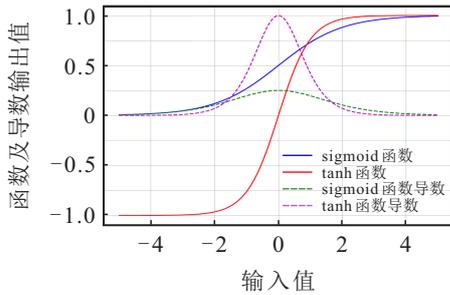


图3 sigmoid和tanh函数对比曲线

SGAIN通过双判别器和移除提示矩阵生成器的策略,及双分支自适应注意力机制,提高了特征关联捕捉与局部特征细化能力,保证了插补任务的高效执行.自适应注意力机制在Transformer网络中的应用将在后文详述.

### 2.2.1 评价指标设计

在建立双分支注意力机制的自适应权重系数之前时,引入两个评价指标,分别为MR和CR. CR是通过加权几何平均数的方法将变异系数(coefficient of variation, CV)和皮尔逊相关系数(pearson correlation coefficient,  $r$ )合并而得的,直接展示了数据的完整性状况,而CR能够整体反映风电数据的稳定性和一致性.为同量级衡量这两个指标,采用模糊隶属度函数进行归一化处理.

CV是标准差和均值的比值,用于描述数据的相对波动程度,其计算公式为

$$CV = \sigma(X)/\mu(X). \quad (3)$$

其中: $\sigma(X)$ 是样本 $X$ 的标准差, $\mu(X)$ 是样本 $X$ 的均值.

皮尔逊相关系数 $r$ 是度量两个变量 $X$ 与 $Y$ 之间线性关系强度和方向的统计量,公式为

$$r = \frac{\sum [(X_i - \mu(X))(Y_i - \mu(Y))]}{[n\sigma(X)\sigma(Y)]}. \quad (4)$$

其中: $X_i$ 和 $Y_i$ 是变量 $X$ 和 $Y$ 的观察值, $\mu(X)$ 和 $\mu(Y)$ 是变量 $X$ 和 $Y$ 的均值, $\sigma(X)$ 和 $\sigma(Y)$ 是变量 $X_i$ 和 $Y_i$ 的标准差, $n$ 是观察数.相关系数范围为 $[-1, 1]$ ,反映参数间的正负相关性;绝对值越大,相关性越强.

变异系数与皮尔逊相关系数经加权几何平均后形成复合评价指标,公式如下:

$$CR = (CV^{w_1} \times r^{w_2})^{\frac{1}{w_1+w_2}}. \quad (5)$$

对于缺失率评价指标(MR),有

$$F_{MR}(MR) = \begin{cases} 1, & MR < a_{MR}; \\ \frac{b_{MR} - MR}{b_{MR} - a_{MR}}, & a_{MR} \leq MR \leq b_{MR}; \\ 0, & MR > b_{MR}. \end{cases} \quad (6)$$

其中: $w_1$ 和 $w_2$ 是权重,反映了变异系数和皮尔逊相关系数在复合评价指标中的重要性; $M$ 代表缺失数据点的数量, $R$ 代表总数据点的数量,MR代表缺失率,是 $M$ 和 $R$ 的比值.模糊隶属度函数定义如下:

$$F_{MR}(x) = \begin{cases} 1, & x < a_{MR}; \\ \frac{b_{MR} - x}{b_{MR} - a_{MR}}, & a_{MR} \leq x \leq b_{MR}; \\ 0, & x > b_{MR}; \end{cases} \quad (7)$$

$$F_{CR}(x) = \begin{cases} 1, & x < a_{CR}; \\ \frac{b_{CR} - x}{b_{CR} - a_{CR}}, & a_{CR} \leq x \leq b_{CR}; \\ 0, & x > b_{CR}. \end{cases} \quad (8)$$

其中: $x$ 代表具体的复合评价指标(CR)的值,是模糊隶属度函数的参数,用于确定函数的形状和取值范围; $a_{MR}$ 是缺失率评价指标的最佳值, $b_{MR}$ 是缺失率评价指标的最差值; $a_{CR}$ 是复合评价指标的最佳值, $b_{CR}$ 是复合评价指标的最差值; $F_{MR}$ 和 $F_{CR}$ 代表复合评价指标的归一化值.

为了确保这两个指标可以在相同的量度尺度上进行评价,采用标准化的衡量方法来实现指标间的一致性.用模糊隶属度归一化到同一区间 $[0, 1]$ .对于缺失率,选用偏小型模糊隶属度函数,对于复合评价指标,选用偏大型模糊隶属度函数.

### 2.2.2 融合自适应注意力机制的Transformer模型

鉴于Transformer在局部信息感知和风电机组时间序列数据捕捉方面的局限性,本研究引入双分支自适应注意力机制,结合空间注意力与通道注意力,旨在提升对局部特征和时间序列模式的识别及捕捉能力.

自适应Transformer(AT)通过卷积块注意模块(CBAM),融合空间位置信息,构建通道与空间双分支并行的注意力机制.进一步通过动态加权法,为通道注意力与空间注意力分支分配权重系数 $F_{MR}$ 和 $F_{CR}$ ,由缺失率、变异系数及数据相关性综合评价指标确定,加强两分支的作用强度.AT模型处理含非缺失值和随机噪声的数据矩阵及缺失位置矩阵,通过深度学习赋予相应注意力权重,生成新矩阵 $T(Z, M)$ .模型如图4所示.

在SGAIN模型中,将生成器 $G$ 和判别器 $D$ 从原始的全连接神经网络替换为改进的Transformer模型,改进的公式如下:

$$\bar{X} = G(Z, M), \quad (9)$$

$$P = D(\bar{X}, M). \quad (10)$$

其中: $\bar{X}$ 表示由生成器 $G$ 生成的数据, $P$ 表示输入数据为真实数据的概率.

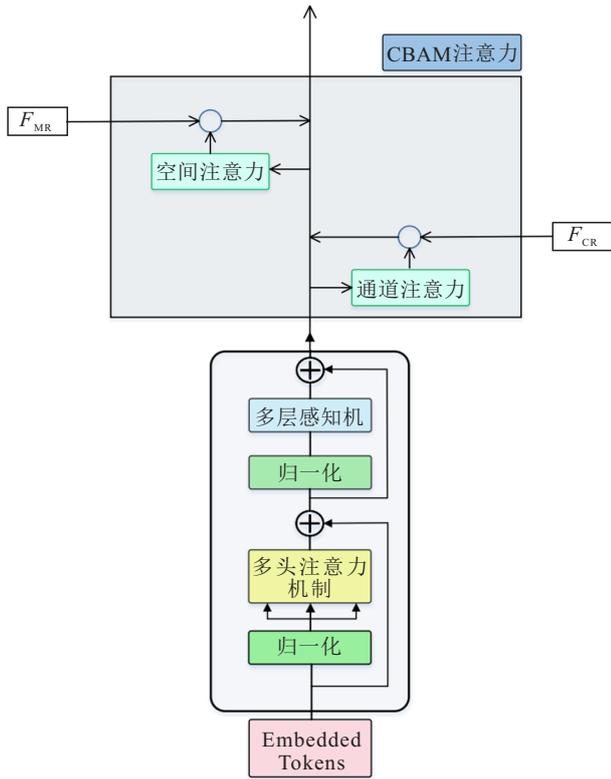


图4 AT模型

融合自适应注意力机制的Transformer模型针对传统Transformer在局部信息处理方面的不足进行改进,有效提升了对风电时间序列数据的分析性能.该架构自适应分配不同特征关系的注意力,结合自适应权重层,实现动态权重调整,增强了学习能力的深度与广度.实验结果表明,此模型在风电数据插补中表现出色,证实了其高精度与鲁棒性.

### 3 算例分析

#### 3.1 数据准备

本研究中,采用内蒙古西部某风场1.5 MW风电机组的实际运行数据.数据集包含风速、风轮转速、发电机转速、齿轮箱油温、机舱内温度、齿轮箱输入轴温度、齿轮箱输出轴温度、齿轮箱侧主轴温度、齿轮箱入口油温、润滑油过滤器入口压力10个重要运行参数.原始数据集为17 229组,每组为10维数据,原始数据通过DBSCAN剔除异常点之后,由于各种因素,还会包括一些传感器故障和通信错误.数据中共有1 301组错误数据,并通过随机缺失,缺失了1 026组数据,总缺失率约为20%.

#### 3.2 AT-SGAIN参数设置

在本研究中,AT-SGAIN模型的超参数精选以提升对风电数据处理的效率和准确度.选择两个自适应Transformer层,利用其卓越的序列数据处理和长距离依赖捕捉能力.为防止过拟合,基于模型性能测

试和调优结果,设置0.2的Dropout率.生成器和判别器的全连接层分别配置10个和1个神经元,匹配输入特征维度和输出真实性概率.激活函数采用Tanh,以应对梯度消失问题.经20 000次训练迭代的Adam优化器选择,基于性能评估确保模型稳定性和收敛.这些超参数设置综合考虑模型结构和数据需求,确保AT-SGAIN的高效、精确插补性能.

#### 3.3 比较方法和指标

本研究采用平均绝对误差(MAE)和均方根误差(RMSE)作为评价指标,以全面评价AT-SGAIN模型的性能.MAE衡量预测与实际值的平均偏差,反映模型精度;RMSE量化预测误差的标准差,反映模型预测的准确性.数学表达式定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i), \quad (11)$$

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

其中: $n$ 代表样本数量, $y_i$ 代表真实值, $\hat{y}_i$ 代表预测值.

#### 3.4 算法测试

##### 3.4.1 结果与分析

在消融实验前,通过DBSCAN聚类算法对数据进行预处理,设置算法参数MinPts = 100,  $\epsilon = 0.2$ .这一步骤旨在区分数据中的错误值、有效值和异常值.在此过程中,DBSCAN算法有效地将原始数据划分成不同类别,识别并排除了离群点.随后,利用PCA降维,图5中深紫色表示正常数据,其他颜色用于区分异常数据.

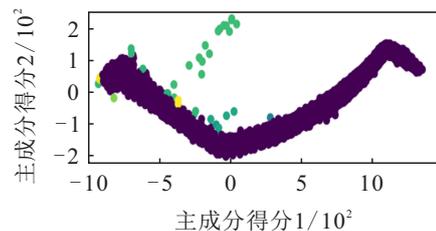


图5 DBSCAN聚类效果

使用DBSCAN剔除异常值并进行数据随机缺失后,采用AT-SGAIN模型对数据进行插补.图6(a)展示了经过PCA筛选后的异常点与缺失数据点的分布情况,图6(b)显示了通过AT-SGAIN模型进行插补后的数据分布,图6(c)则将插补前后的数据进行了可视化对比.图7通过聚焦序列末端的局部放大视图,进一步突出模型在恢复数据精度和细节方面的能力,展示了模型处理累积误差和边界效应的稳定性和鲁棒性.

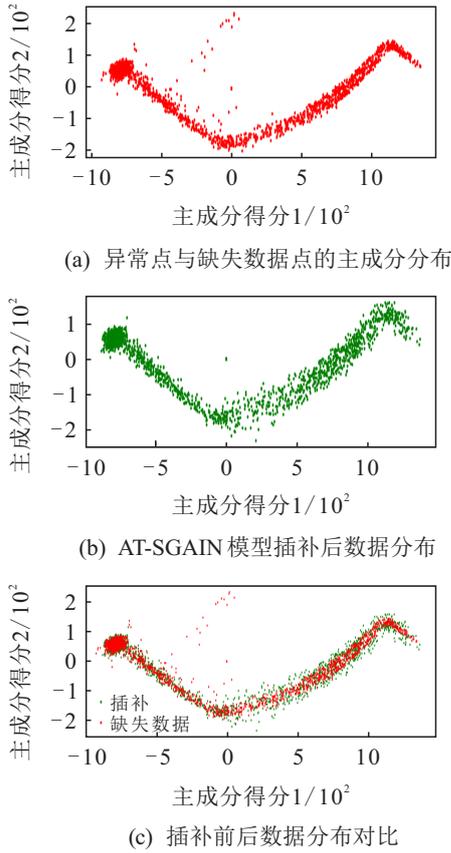


图6 AT-SGAIN插值数据可视化效果

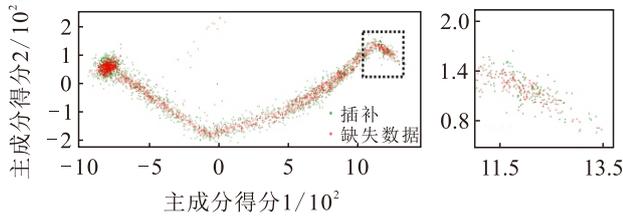


图7 AT-SGAIN模型数据插补精细化分析

### 3.4.2 消融实验

本研究首先通过网络复杂度的消融实验,深入分析AT-SGAIN模型的参数优化及计算效率。之后,针对模型的整体性能进行进一步评估,特别在特定数据集上实施数据插补性能的消融实验,以准确探究模型各组成部分在数据插补效果上的作用及影响。

模型的参数量(Parameters)是衡量模型复杂性和存储需求的重要指标,计算公式如下:

$$\text{Parameters} = \frac{K_H \times K_W \times C_{in} \times C_{out}}{G} \quad (13)$$

浮点运算量(FLOPs)公式反映了模型中卷积核的参数数量,其中分组卷积的情况下每个分组内有对应的卷积核,对应公式如下:

$$\text{FLOPs} = \text{FLOPs}_{\text{conv}} + \text{FLOPs}_{\text{Add}}, \quad (14)$$

$$\text{FLOPs}_{\text{Conv}} =$$

$$\frac{H_{out} \times W_{out} \times C_{out} \times C_{in} \times K_H \times K_W}{G}, \quad (15)$$

$$\text{FLOPs}_{\text{Add}} =$$

$$\frac{H_{out} \times W_{out} \times C_{out} \times (C_{in} \times K_H \times K_W - 1)}{G} \quad (16)$$

其中:  $K_H$  和  $K_W$  表示卷积核的高度和宽度,  $C_{in}$  表示输入通道数,  $C_{out}$  表示输出通道数,  $G$  表示分组卷积的组数。

在GAIN模型的基础上,进行了一系列改进,分别为SGAIN-I至SGAIN-V,每一步改进都建立在前一步改进的基础之上。SGAIN-I通过减少生成器和判别器的层数简化了模型复杂度,并保持性能;SGAIN-II换用Tanh激活函数加速优化过程收敛;SGAIN-III进一步简化模型,去除了提示矩阵生成器;SGAIN-IV引入双判别器架构,分别处理真实和虚假数据;SGAIN-V(AT-SGAIN)结合自适应注意力机制和Transformer编码器,提高风电数据插补精度。模型参数量、计算量和插补精度对比在表1中详述。

表1 模型参数量和计算量对比

模型	参数量	计算量/G	RMSE
GAIN	2 223 604	19.32	0.113
SGAIN-I	1 213 802	9.67	0.152
SGAIN-II	1 213 802	9.08	0.161
SGAIN-III	893 486	8.76	0.201
SGAIN-IV	1 039 436	10.91	0.127
SGAIN-V	1 339 684	11.53	0.047

SGAIN网络作为GAIN网络的改进版本,显著降低了计算成本和模型复杂度,同时简化了训练流程。尽管这些优化可能带来精度的轻微下降,但引入的AT-SGAIN模型有效弥补了这一缺陷,实现了在保持快速计算的同时提高精度的目标。AT-SGAIN在提高运算速度的基础上,进一步优化了模型性能。经过对模型的参数优化和计算复杂性分析,本研究还对其数据插补效果进行了消融实验,全面评估模型的有效性和实用性。

图8~图13展示了在逐步移除AT-SGAIN模型的关键组成部分,包括自适应注意力机制(A)、Transformer(T)网络以及轻量化生成对抗插补网络(SGAIN)后的测试结果图。将这些结果与图6和图7中的完整AT-SGAIN插补效果进行详细对比,可以看出,A和SGAIN都对模型的插补性能有着重要的影响。当移除A时,插补模型对于噪声数据的适应性较差,插补性能降低,模型的上下文捕捉能力显著下降,这表明了A在捕捉数据的时序特性和处理缺失值上的重要性。当移除SGAIN时,虽然模型仍然能够进行插补,但由于网络结构简化,参数量降低,其性能也有所下降,这表明了SGAIN在优化插补结果上的重要作用。详细的实验结果如表2所示。

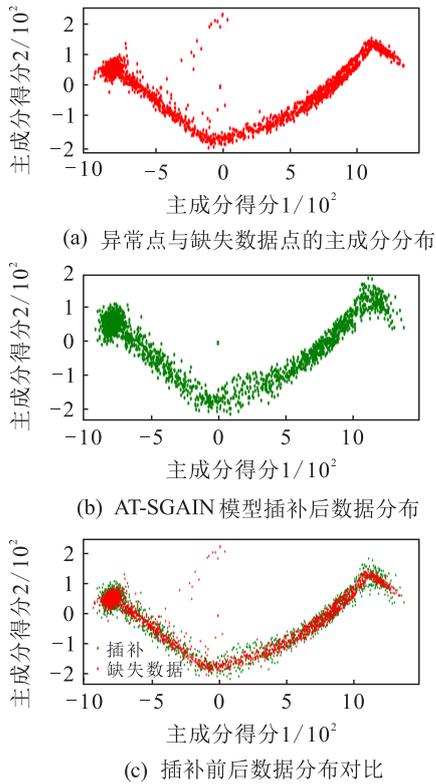


图8 T-SGAIN 插值数据可视化效果

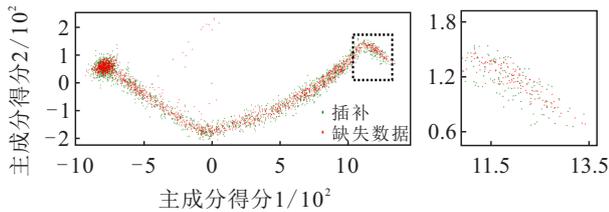


图9 T-SGAIN 模型数据插补精细化分析

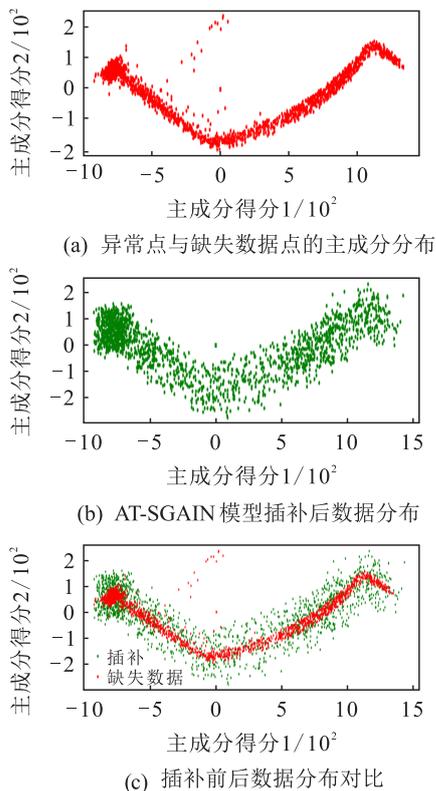


图10 SGAIN 插值数据可视化效果

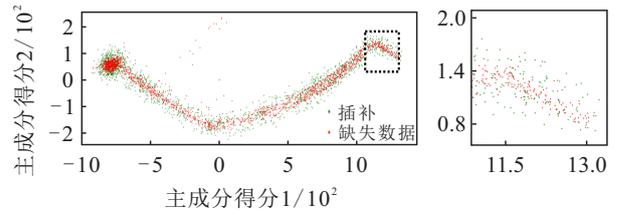


图11 SGAIN 模型数据插补精细化分析

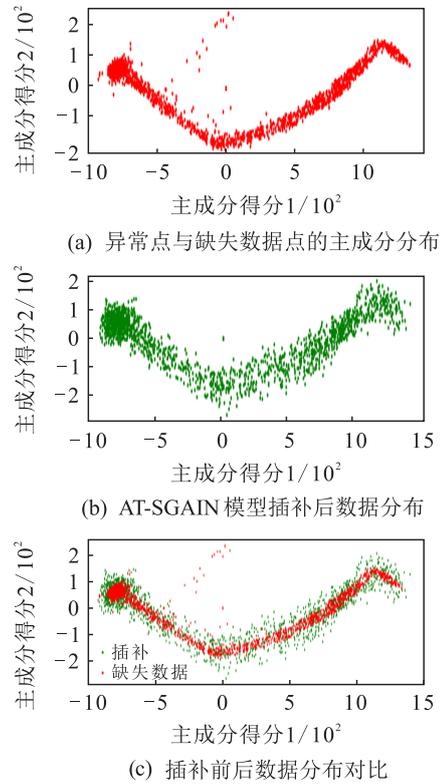


图12 GAIN 插值数据可视化效果

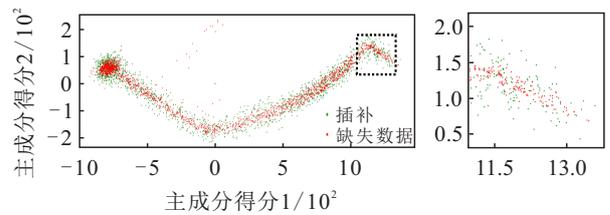


图13 GAIN 模型数据插补精细化分析

表2 AT-SGAIN网络消融实验

模型	RMSE	MAE	t/ms
AT-SGAIN	0.047	0.043	2776
T-SGAIN	0.098	0.103	2589
SGAIN	0.133	0.12	2258
GAIN	0.113	0.111	3442

如表2所示,各模型在数据插补精度上均有所表现,但AT-SGAIN模型精确度尤为突出.对比各模型的误差指标,T-SGAIN、SGAIN和GAIN模型的RMSE指标分别为0.098、0.133和0.113,而在MAE指标上分别为0.103、0.120和0.111.尽管这些模型已经展现出相对可观的性能,但AT-SGAIN模型的RMSE和MAE分别达到了0.047和0.043,精度明显优于其

他模型.

进一步考虑计算速度,AT-SGAIN模型的插补所需时间为2776 ms.相较于GAIN模型的3442 ms,其插补速度提高了19.34%;SGAIN缺少A和Transformer结构,实现了最快的插补速度,仅为2258 ms.这表明,SGAIN在速度上确实具有优势,但是AT-SGAIN在SGAIN的基础上引入自适应注意力机制和Transformer编码器后,不仅提高了插补精度,同时也保证了较为理想的计算速度.

#### 4 AT-SGAIN模型有效性验证

使用AT-SGAIN模型生成数据来插补原始SCADA数据集,本文采用多种方法验证AT-SGAIN模型的有效性和可靠性.

##### 4.1 直接验证对比

###### 4.1.1 插补模型效果对照分析

使用相同的数据集和评价指标,将AT-SGAIN与一些其他主流插补方法进行对比,包括MisGAN<sup>[25]</sup>、SAE-CD<sup>[26]</sup>、PC-GAIN<sup>[27]</sup>等深度学习模型.

缺失率对比测试图结果显示(见图14),在缺失率为22.3%~58.1%区间时,AT-SGAIN模型的插补效果优于其余插补模型,这表明AT-SGAIN模型在风电机组数据插补任务中具有优越的性能.

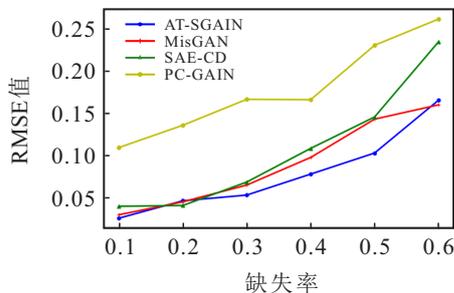


图14 缺失率对比效果

###### 4.1.2 AT-SGAIN模型多工况性能评估

在2023年4月1日~12月31日间,本研究对AT-SGAIN数据插补模型在各种环境条件下的表现进行了全面评估.4月1日~4月15日为沙尘暴天气,5月1日~5月15日为大风环境,7月10日~7月25日和12月1日~12月15日为雨雪天气,这些特殊环境工况可能对数据收集和风电机组的运行稳定性产生不利影响.这些测试展示了AT-SGAIN模型在应对不同环境工况时的适应性和效能,为其在实际应用中提供了重要参考.

图15~图18中,红色点代表缺失数据,绿色点代表模型插补后的数据,右侧为局部放大视图.由图可知,AT-SGAIN模型在雨雪、低温与沙尘暴工况下表现出较好的插补效果,数据点密集,一致性较高,相

较之下,大风工况插补效果相对较差.从图16可以看出,大风工况下数据点的波动最为明显,这可能是由于风速的极端不稳定性对数据收集造成的影响,这指出模型在大风极端条件下处理数据的能力还有待提高.

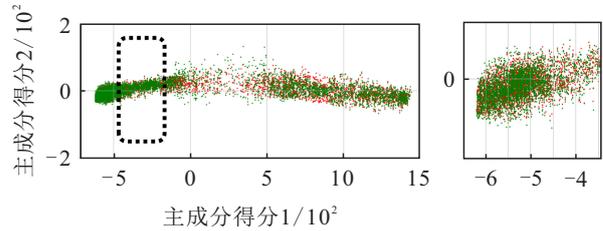


图15 沙尘暴工况数据插补可视化效果

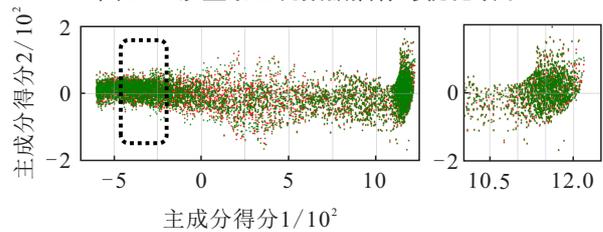


图16 大风工况数据插补可视化效果

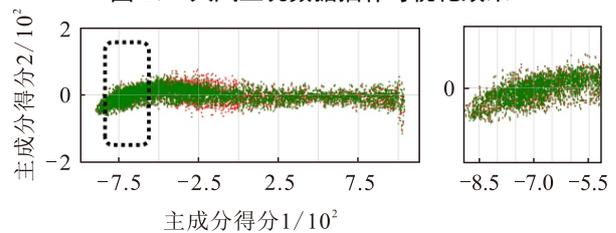


图17 雨雪工况数据插补可视化效果

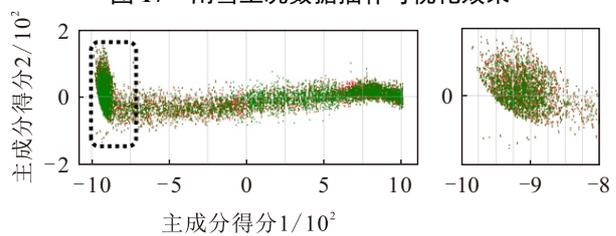


图18 低温工况数据插补可视化效果

##### 4.2 间接验证对比

基于AT-SGAIN插补模型与其他各主流方法插补后的数据集,使用时间序列预测的回归模型——长短期记忆模型,进行后续风功率分析和预测对比.预测模型采用三层LSTM结构,单元数依次为32、64和96,通过全连接层输出预测结果.任务基于回归,利用RMSE和MAE评估误差,使用Adam优化器整合AdaGrad和RMSProp技术,确保训练的稳定与收敛.

图19和图20的结果显示,使用AT-SGAIN模型插补后的数据进行LSTM风功率预测,功率预测效果显著优于使用其他模型(如MisGAN、SAE-CD和PC-GAIN)插补后的数据.在AT-SGAIN插补数据的支持下,风电功率的预测结果与实际功率高度一致,精度明显优于其他插补方法.具体表现为RMSE和MAE指标分别低至0.113和0.077,充分表明了AT-SGAIN

在风电数据插补方面的较好性能和应用潜力。

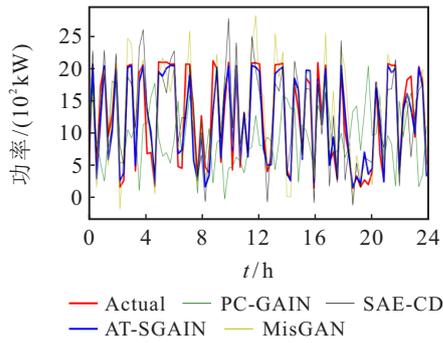


图19 风功率预测对比

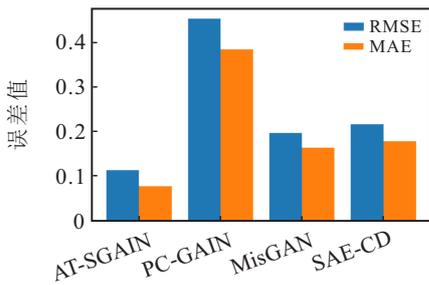


图20 误差精度对比

4.3 跨数据集的直接验证对比

为验证AT-SGAIN模型的泛化能力,引入内蒙古东部某风场2MW风电机组实际数据进行测试. 总计23183组10维数据,错误数据约1200组,由于设备老化和通信障碍,随机缺失1500组. 利用DBSCAN剔除异常后,缺失率约22%. 图21和图22展示了AT-SGAIN模型跨数据集插补效果.

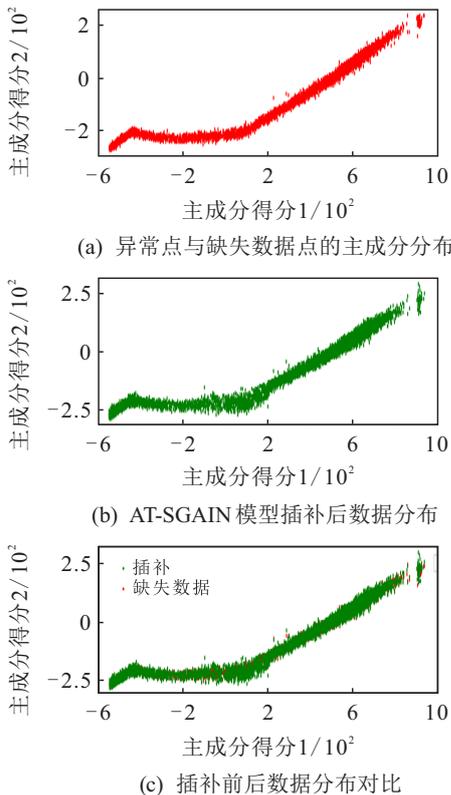


图21 AT-SGAIN跨数据集插补可视化效果

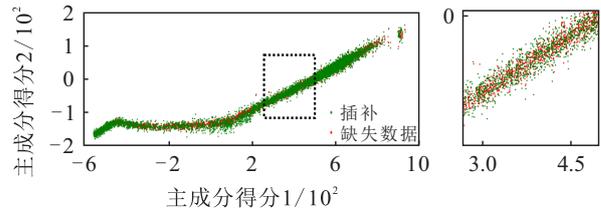


图22 跨数据集插补精细化分析

在实验方法上,继续采用与之前相同的模型配置和评估指标,但对AT-SGAIN的参数进行了针对性的调整以适应新数据集的特点. 同时,本次测试也将AT-SGAIN模型的表现与MisGAN<sup>[25]</sup>、SAE-CD<sup>[26]</sup>和PC-GAIN<sup>[27]</sup>等其他主流插补方法进行了对比. 图23为基于缺失率对比测试图.

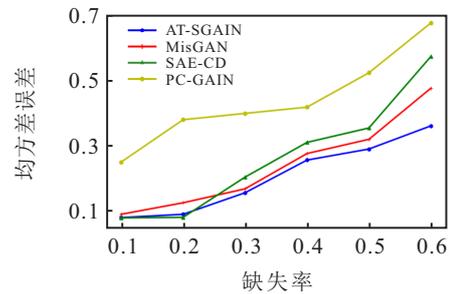


图23 跨数据集缺失率对比效果

从图23可以看出,AT-SGAIN插补模型与其他主流插补方法相比,在处理不同缺失率数据时表现出色. 缺失率对比测试图显示,在缺失率逐渐增高的情况下,AT-SGAIN模型保持了较低的RMSE,特别是在高缺失率环境下,其优势更为明显. 这一结果表明AT-SGAIN模型在处理大规模数据缺失时,具有较强的鲁棒性和优越的插补精度,验证了其在风电数据插补应用中的潜力和有效性.

5 结论

本文提出基于自适应轻量化生成对抗插补网络(AT-SGAIN),其SGAIN中生成器和判别器均源于Transformer的编码器设计,结合自适应注意力机制,该网络能够深度捕获时间序列特性及其局部信息,从而确保高效、精准的数据插补效果. 为验证其性能,首先通过消融实验评估其插补精度,并与主流算法进行插补对比;随后基于各插补方法的插补结果,利用LSTM进行风电功率预测对比实验,结果显示AT-SGAIN明显优于当前主流的插补方法. 本文还扩展了实验范围,包括在多种风电机组工况及不同风电场数据集上的对比测试. 综合考虑,AT-SGAIN在风电数据插补方面表现出色,有较高的应用价值.

参考文献(References)

[1] Habibullah, Khan M A, Atif M, et al. Numerical analysis of airfoils used in an omni-directional-guide vane structure of vertical axis wind turbine for high-rise

- buildings[J]. *Modern Physics Letters B*, 2024, 38(16): 1185-1192.
- [2] Xiang L, Yang X, Hu A J, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks[J]. *Applied Energy*, 2022, 305: 117925.
- [3] Li X, Li X, Li S, et al. Fault diagnosis based on meta learning and neural architecture search for new modes with small samples[J]. *Control and Decision*, 2023, 38(11): 3175-3183.
- [4] Choate L A, Koleilat A, Harris K, et al. Confirmation of insertion, deletion, and deletion-insertion variants detected by next-generation sequencing[J]. *Clinical Chemistry*, 2023, 69(10): 1155-1162.
- [5] Plaksienko A, Di Lena P, Nardini C, et al. methylImp2: Faster missing value estimation for DNA methylation data[J]. *Bioinformatics*, 2024, 40(1): btae001.
- [6] Chen J Y, Zou J F, Pang L, et al. Anti-interpolation based stealthy poisoning attack method on deep neural networks[J]. *Control and Decision*, 2023, 38(12): 3381-3389.
- [7] Becker R, Thrän D. Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors[J]. *Applied Energy*, 2017, 208: 252-262.
- [8] Eklund N H W. Using genetic algorithms to estimate confidence intervals for missing spatial data[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2006, 36(4): 519-523.
- [9] Yoon J, Zame W R, van der Schaar M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks[J]. *IEEE Transactions on Biomedical Engineering*, 2019, 66(5): 1477-1490.
- [10] Deng W, Guo Y X, Liu J, et al. A missing power data filling method based on improved random forest algorithm[J]. *Chinese Journal of Electrical Engineering*, 2019, 5(4): 33-39.
- [11] Chen F B, Kang W, Shu Z R, et al. Predicting roof-surface wind pressure induced by conical vortex using a BP neural network combined with POD[J]. *Building Simulation*, 2022, 15(8): 1475-1490.
- [12] Liang H P, Cao J, Zhao X Q. Remaining useful life prediction method for bearing based on parallel bidirectional temporal convolutional network and bidirectional long and short-term memory network[J]. *Control and Decision*, 2024, 39(4): 1288-1296.
- [13] Ryu S, Kim M, Kim H. Denoising autoencoder-based missing value imputation for smart meters[J]. *IEEE Access*, 2020, 8: 40656-40666.
- [14] Qu F M, Liu J H, Ma Y J, et al. A novel wind turbine data imputation method with multiple optimizations based on GANs[J]. *Mechanical Systems and Signal Processing*, 2020, 139: 106610.
- [15] Yoon S, Sull S. GAMIN: Generative adversarial multiple imputation network for highly missing data[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 8453-8461.
- [16] Wang W, Chai Y M, Li Y. GAGIN: Generative adversarial guider imputation network for missing data[J]. *Neural Computing and Applications*, 2022, 34(10): 7597-7610.
- [17] Andreini P, Ciano G, Bonechi S, et al. A two-stage GAN for high-resolution retinal image generation and segmentation[J]. *Electronics*, 2021, 11(1): 60.
- [18] Luo Y H, Zhang Y, Cai X R, et al. E<sup>2</sup>GAN: End-to-end generative adversarial network for multivariate time series imputation[C]. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Edinburgh, 2019: 3094-3100.
- [19] Wu Z J, Ma C, Shi X C, et al. Imputing missing indoor air quality data with inverse mapping generative adversarial network[J]. *Building and Environment*, 2022, 215: 108896.
- [20] Friedjungová M, Vařata D, Balatsko M, et al. Missing features reconstruction using a Wasserstein generative adversarial imputation network[C]. International Conference on Computational Science. Cham: Springer, 2020: 225-239.
- [21] You J X, Ma X B, Ding D Y, et al. Handling missing data with graph representation learning[J/OL]. 2020, arXiv: 2010.16418.
- [22] Li D, Li L H, Li X L, et al. Smoothed LSTM-AE: A spatio-temporal deep model for multiple time-series missing imputation[J]. *Neurocomputing*, 2020, 411: 351-363.
- [23] Jia X D, Jin C, Buzza M, et al. Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves[J]. *Renewable Energy*, 2016, 99: 1191-1201.
- [24] Jiang G Q, Li W Y, Bai J R, et al. SCADA data-driven blade icing detection for wind turbines: An enhanced spatio-temporal feature learning approach[J]. *Measurement Science and Technology*, 2023, 34(5): 054004.
- [25] Li S C X, Jiang B, Marlin B. MisGAN: Learning from incomplete data with generative adversarial networks[J/OL]. 2019, arXiv: 1902.09599.
- [26] Liu X, Zhang Z J. A two-stage deep autoencoder-based missing data imputation method for wind farm SCADA data[J]. *IEEE Sensors Journal*, 2021, 21(9): 10933-10945.
- [27] Wang Y F, Li D, Li X, et al. PC-GAIN: Pseudo-label conditional generative adversarial imputation networks for incomplete data[J]. *Neural Networks*, 2021, 141: 395-403.

## 作者简介

武甲(1996—),男,硕士,主要研究方向为风电机组性能预测与健康状态评估, E-mail: wujia\_1130@163.com;

齐咏生(1975—),男,教授,博士生导师,主要研究方向为复杂工业过程的故障监测与诊断, E-mail: qys@imut.edu.cn;

马然(1982—),女,博士,主要研究方向为风力发电机组状态评估, E-mail: mranmarlena@163.com;

高胜利(1973—),男,教授,博士,主要研究方向为风力发电机组状态监测系统, E-mail: gaoslemail@163.com;

刘慧文(1981—),男,博士,主要研究方向为新能源电力系统规划设计与运行优化, E-mail: liuhuiwen@imut.edu.cn.