

DOI: 10.19783/j.cnki.pspc.250608

基于时空卷积-注意力网络的分布式光伏缺失数据插补方法

余思杪¹, 许玉格¹, 樊淼嘉², 刘俊峰¹, 曾君²

(1. 华南理工大学自动化科学与工程学院, 广东 广州 510641; 2. 华南理工大学电力学院, 广东 广州 510641)

摘要: 光伏系统运行数据受设备故障、通信中断等客观因素影响, 普遍存在数据缺失问题, 导致数据质量下降。有效的数据修复是实现高精度光伏功率预测和电站运行优化的重要基础。传统数据修复方法难以有效处理分布式光伏数据的复杂时空相关性, 且对气象数据缺失等特殊场景的适应性不足, 导致修复精度有限。为此, 提出了一种时空卷积-注意力插补网络(spatial-attention temporal-convolution network, SATCN), 使用一维卷积神经网络学习数据的时间依赖, 借鉴 Transformer 的自注意力机制构建空间注意力网络作为空间聚合器, 利用少量可观测数据, 充分挖掘数据的时空依赖性, 实现对多分布式光伏集群数据的高质量修复。同时采用联合优化训练方法, 考虑观测值重建任务和人为缺失插补任务, 避免模型只关注可观测数据而忽略了缺失插补。实验结果表明, 所提方法在不同缺失率下均具有良好的插补性能, 能够在无监督的情况下实现对缺失数据的高精度修复。

关键词: 缺失数据插补; 时空依赖; 注意力机制; 联合优化训练; 分布式光伏

Missing data imputation method for distributed photovoltaic systems based on spatial-attention temporal-convolution network

YU Simiao¹, XU Yuge¹, FAN Miaojia², LIU Junfeng¹, ZENG Jun²

(1. School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China;

2. School of Electric Power Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: Operational data of photovoltaic (PV) systems are often affected by equipment failure, communication interruption, and other objective factors, leading to widespread data missing issues and degraded data quality. Effective data restoration is an important foundation for achieving high-precision PV power forecasting and optimal power plant operation. Traditional data imputation methods struggle to effectively capture the complex spatiotemporal correlations in distributed PV data and exhibit limited adaptability to special scenarios such as missing meteorological data, resulting in unsatisfactory imputation accuracy. Therefore, a spatial-attention temporal-convolution network (SATCN) is proposed. A one-dimensional convolutional neural network is used to learn temporal dependencies, while a spatial attention network inspired by the self-attention mechanism of the Transformer is constructed as a spatial aggregator. By leveraging limited observable data, the proposed model fully exploits spatiotemporal dependencies to achieve high-quality restoration of data from multiple distributed PV clusters. In addition, a joint optimization training strategy is adopted, incorporating both observed value reconstruction and artificially masked data imputation tasks, thereby preventing the model from focusing solely on observed data while neglecting missing data imputation. Experimental results show that the proposed method has good imputation performance under various missing rates and achieves high-precision restoration of missing data without supervision.

This work is supported by the National Natural Science Foundation of China (No. 62173148 and No. 52377186).

Key words: missing data imputation; spatiotemporal dependence; attention mechanism; joint optimization training; distributed photovoltaic system

0 引言

分布式光伏是光伏的重要发展形式, 在“双碳”

目标的推动下, 近年来发展迅速。根据国家能源局最新数据统计, 截至 2024 年 12 月, 全国光伏发电装机容量达到 8.86 亿 kW, 同比增长 45%, 其中集中式光伏 5.11 亿 kW, 分布式光伏 3.75 亿 kW^[1]。随着分布式光伏规模的不断扩大, 光伏电站的功率预测与性能评估成为光伏有序健康发展的重要支撑^[2-3]。

基金项目: 国家自然科学基金项目资助(62173148, 52377186);
广东省自然科学基金项目资助(2024A1515012428)

其中, 光伏历史运行数据的获取、分析与挖掘是重要基础。然而在实际应用中, 光伏数据的采集会因仪表、设备或组件故障等原因中断^[4], 导致数据出现缺失。缺失数据的存在会降低时间序列模型的准确性, 进而影响光伏功率预测精度和性能评估准确性^[5]。直接剔除存在缺失的数据^[6]不仅会造成样本量减少, 还可能破坏真实数据分布。因此, 对光伏数据进行高质量修复具有重要意义。

缺失数据插补是数据修复的核心部分。传统的缺失数据插补方法大致可以分为插值方法和统计学方法两大类。插值方法包括线性插值法^[7]、样条插值法^[8]和 Hermite 插值法^[9]等。这些方法主要利用数据的分布特性和变量间的相关性等统计属性进行插补, 具有局限性, 对数据的线性性质以及平滑程度有严格要求。统计学方法则包括自回归方法^[10]、K 最近邻(K nearest neighbor, KNN)^[11]、随机森林^[12]、链式方程多重插补^[13](multivariate imputation by chained equations, MICE)等。然而, 上述方法均存在数据利用率低和插补精度有限的问题。

随着人工智能和大数据技术的发展, 深度学习已被广泛应用于电力数据插补任务中, 其中生成对抗网络(generative adversarial networks, GAN)^[14-15]及其改进模型应用较为广泛^[16-18]。通过引入语义感知^[19]、时序信号编码^[20]、提示机制^[21]、Wasserstein 散度^[22]、门控循环单元^[23]、长短期记忆网络^[24-25]、卷积神经网络^[26-27]等机制, GAN 系列模型在配电网量测^[19-21, 26]、光伏功率^[22]、电动汽车充电负荷^[23]、风电场监测数据^[24]、相量测量单元^[25]和城市负荷^[27]等多种电力数据缺失场景中有效提升了插补精度。

文献[28]采用注意力机制增强的双向长短期记忆网络(bidirectional long and short term memory, BiLSTM)实现风电机组缺失数据的修复。基于 BiLSTM, 文献[29]构建了双向循环插补网络^[30], 学习分布式光伏场站间的相关性, 并引入注意力机制提高数据修复精度。文献[31]进一步引入了图注意力网络, 更好地挖掘数据的空间依赖关系。

大量分布式光伏采集数据分析表明, 分布式光伏缺失数据在实际应用中存在以下关键特征。

1) 多种缺失模式: 分布式光伏功率数据通常同时存在单点缺失、连续片段缺失、全天缺失等模式, 这些模式难以用单一模型捕获和估算。在连续片段缺失、全天缺失模式下, 时间序列的连续性被破坏, 模型难以有效利用时间依赖性, 导致插补精度受限。

2) 缺乏站点气象监测数据: 分布式光伏场站规模小、布局分散且投资成本受限, 通常不具备站点级气象监测设备, 因而普遍缺乏气象数据支持, 致

使单个场站难以获得充足的多维数据。这种数据维度的缺失使得传统方法在插补时面临显著局限。

3) 具有复杂的空间依赖性: 相邻场站间存在功率波动的传递效应, 并且与节点位置分布具有强关联。现有的插补方法往往仅考虑局部空间相关性, 忽略场站网络的全局拓扑结构。

因此, 尽管前述研究为分布式光伏数据修复提供了重要基础, 但未能充分兼顾上述分布式光伏缺失数据的特点。为此, 本文提出一种基于时空卷积-注意力网络(spatial-attention temporal-convolution network, SATCN)的分布式光伏缺失数据插补方法, 通过深度时空特征级联, 自适应捕捉场站间的时空交互规律, 有效应对混合缺失模式, 在无气象数据的条件下准确建模复杂时空依赖关系, 实现高精度缺失数据插补。本文的主要贡献如下。

1) 借鉴 Transformer 构建空间注意力网络, 通过动态图编码融合物理先验, 从而实现对场站间复杂的全局空间依赖关系进行精准建模。

2) 采用一维卷积神经网络(1D convolutional neural network, 1D-CNN)作为时间聚合器, 捕获时域依赖。注意力网络(spatial attention networks, SAN)作为空间聚合器, 通过不同场站的空间聚合和跨时刻的信息传递进行联合学习, 从而捕获时空依赖关系。

3) 采用联合优化训练方法, 引入人为缺失值, 通过对人为缺失的插值与观测值重构的协同训练, 实现无需完整数据集、不依赖特定数据分布假设的无监督分布式光伏缺失数据插补。

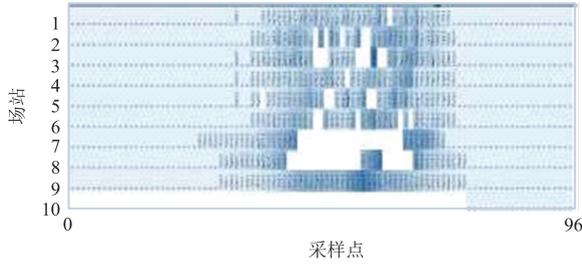
实验结果显示, 所提出的基于 SATCN 的分布式光伏缺失数据插补方法在不同缺失率下均能实现缺失数据的高质量修复。

1 问题建模

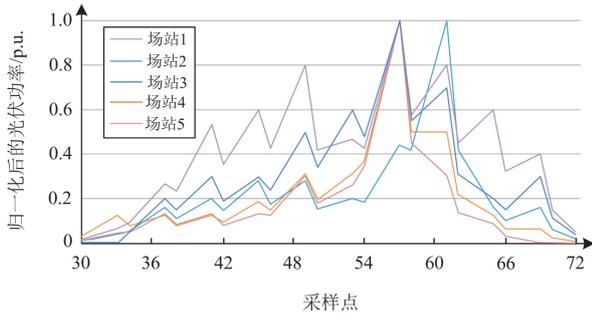
以我国广东省某光伏站群(含 10 个分布式场站)的实测数据为例, 其缺失模式和相关性分析如图 1 所示。由图 1(a)可知, 数据中同时存在单点缺失和连续片段缺失等多种缺失模式。图 1(b)展示了 5 个场站同一天的出力曲线。这些曲线在形态和变化趋势上高度一致, 表明场站间存在显著的时空相关性。此外, 云层移动会依次在不同场站上产生阴影, 从而导致出力波动在部分场站间呈现明显的传递性。

因此, 深入挖掘场站间的时空依赖关系是提高分布式光伏缺失数据插补精度的关键。本文将 T 个采样时刻的各光伏场站数据表示为全连接加权图, 其拓扑结构由 \mathcal{A} 表示, $\mathcal{A} \in \mathbb{R}^{N \times N}$, N 为节点数量。 \mathcal{A} 中元素 a_{ij} 由场站 i 、 j 之间的地理距离计算得到, 表示它们的空间邻近程度。由于地理位置固定, 故

在各时刻保持一致。



(a) 某天10个光伏场站数据缺失情况



(b) 5个光伏场站某天功率示意图

图1 分布式光伏缺失模式和相关性示意图

Fig. 1 Schematic diagram of distributed PV missing mode and correlation

T 个时刻 N 个节点的多维时间序列定义为

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_T) \in \mathbb{R}^{N \times T} \quad (1)$$

式中: \mathbf{x}_t 表示 t 时刻的 N 个节点的量测值。由于多维时间序列 \mathbf{X} 中存在缺失值, 构建掩码向量 \mathbf{M} 描述 \mathbf{X} 中缺失值的位置。其中, \mathbf{M} 与 \mathbf{X} 一一对应, 二者维度相同, $m_{n,t}$ 与 $x_{n,t}$ 分别表示 \mathbf{M} 和 \mathbf{X} 中第 n 个场站在 t 时刻的元素。 $m_{n,t} = 0$ 时表示 $x_{n,t}$ 缺失, 而 $m_{n,t} = 1$ 时表示 $x_{n,t}$ 未出现缺失。

插补后的数据记为 $\hat{\mathbf{X}}$, 分布式光伏缺失数据插补过程可以表示为

$$[\mathbf{X}, \mathbf{M}; \mathbf{A}] \xrightarrow{f} \hat{\mathbf{X}} \quad (2)$$

式中: $\hat{\mathbf{X}}$ 为插补后的数据; f 表示插补模型, 是解决缺失数据插补问题的关键。

2 时空卷积-注意力插补网络

针对分布式光伏数据中复杂的时空耦合特性, 提出了一种时空卷积-注意力插补框架, 其结构如图2所示。模型采用晴空模型 \mathbf{P} 作为输入之一。该输入定义为基于光伏场站的经纬度和时间戳, 通过 Ineichen-Perez 模型^[32]模拟理想晴空条件下的太阳位置与大气传输过程, 所得到的理想发电归一化值。

该框架通过场站间的空间聚合和时间片之间的时态传递来捕获时空依赖关系, 充分获取数据的时空特征。数据重建模块基于学习到的深层特征, 恢复出完整且符合真实分布的光伏功率数据。

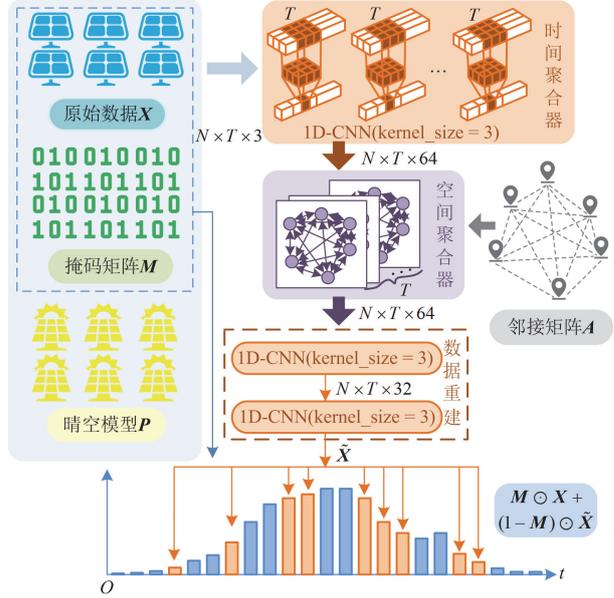


图2 SATCN 整体架构

Fig. 2 Overall architecture of SATCN

2.1 时间聚合器

为捕捉分布式光伏数据的时序特征, 本文采用 1D-CNN 构建时间聚合器。与循环神经网络结构相比, 1D-CNN 对局部依赖关系的捕捉能力更能应对光伏出力的强波动性。

具体地, 本文使用卷积核宽度为 3 的一维卷积层, 通过聚合相邻时刻信息, 有效捕捉局部时间趋势。如图3所示, 节点的信号通过合并相邻时刻上的信息进行更新^[33], 卷积计算表达式为

$$\tilde{\mathbf{H}} = \text{ReLU}(\mathbf{W} * \mathbf{H} + \mathbf{b}) \in \mathbb{R}^{N \times D \times T'} \quad (3)$$

式中: $\tilde{\mathbf{H}}$ 为 1D-CNN 计算输出; \mathbf{H} 为模型输入, $\mathbf{H} = \mathbf{X} \parallel \mathbf{M} \parallel \mathbf{P} \in \mathbb{R}^{N \times 3 \times T}$; “*” 表示卷积运算; \mathbf{W} 、 \mathbf{b} 分别为可学习的权重矩阵和偏置项; $\text{ReLU}(\cdot)$ 为激活函数; D 表示特征通道维度; T' 表示卷积后的输出时间维度。本文采用对称补充策略在序列首尾补 0, 确保输出时间维度 T' 与输入时间维度 T 保持一致, 避免信息损失。

2.2 空间聚合器

为充分获取全局空间依赖关系, 本文构建空间注意力网络作为空间聚合器, 如图4所示。鉴于 Transformer 的自注意力机制已被广泛验证具有强大的建模能力, 本文基于其基础架构, 在图学习任务中引入自注意力机制, 并将场站间的距离信息作

为先验知识注入该模块。该机制生成的注意力得分直接量化节点间的影响强度，从而动态地建模空间关系。

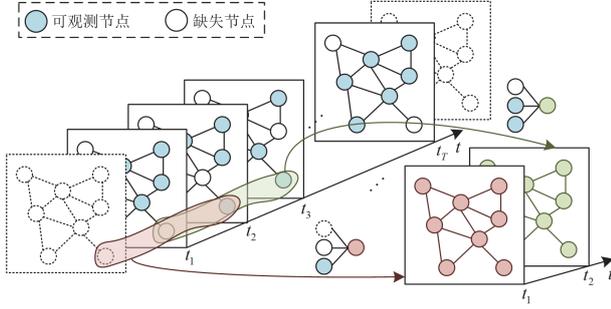


图 3 时间特征聚合器示意图

Fig. 3 Schematic diagram of time feature aggregator

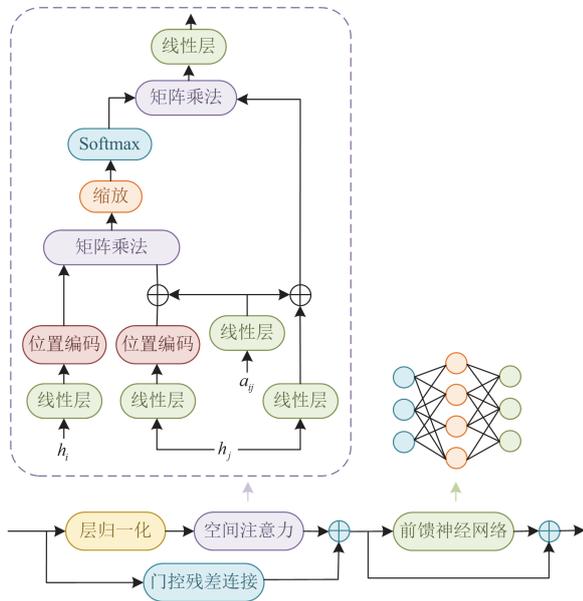


图 4 空间注意力模块示意图

Fig. 4 Schematic diagram of spatial attention network module

具体而言，记 t 时刻的节点特征为 \tilde{H}_t ， $\tilde{H}_t = [h_1, h_2, \dots, h_N]^T \in \mathbb{R}^{N \times D}$ 。将场站 i 、 j 的特征 h_i 、 h_j 经独立线性层和旋转位置编码分别生成查询向量 q_i 和键向量 k_j ，以区分场站在注意力机制中的角色，并保留彼此间的相对位置关系。然后计算节点 i 到其他节点的注意力权重，即场站间的空间相关性，如式(4)所示。

$$\beta_{ij} = \frac{\exp(q_i^T (k_j + e_{ij}) / \sqrt{d})}{\sum_{u \in \mathcal{N}(i)} \exp(q_i^T (k_u + e_{iu}) / \sqrt{d})} \quad (4)$$

式中： e_{ij} 为先验空间关系，是邻接矩阵 A 的线性投影，通过残差连接形式注入键向量，以显式融合物

理拓扑信息； k_u 、 e_{iu} 分别为节点 u 的键向量和其与节点 i 的先验空间关系； $\mathcal{N}(i)$ 为节点 i 的邻居节点集合； d 为 q_i 和 k_j 的维度。

获得图注意力后，对节点进行信息聚合，使各场站学习到邻近场站的信息，如式(5)所示。

$$h'_i = \sum_{j \in \mathcal{N}(i)} \beta_{ij} (v_j + e_{ij}) \quad (5)$$

式中： h'_i 为注意力输出； v_j 为值向量， $v_j \in \mathbb{R}^{N \times d}$ ，由可训练参数 W_v 、 b_v 对 h_j 编码得到。

此外，本文引入门控残差连接机制^[34]。该机制通过可学习的门控单元，以残差连接的形式将时间聚合器输出的时序特征与空间注意力模块输出的空间特征相融合。这种设计能自适应调节两类特征的贡献比例，同时防止特征过度平滑。具体操作为

$$\gamma_i = \text{sigmoid}(W_g (h'_i \| h_i \| h'_i - h_i)) \quad (6)$$

$$s_i = \text{ReLU}(\text{LayerNorm}((1 - \gamma_i)h'_i + \gamma_i h_i)) \quad (7)$$

式中： γ_i 为动态门控系数，用于融合原始特征与注意力输出； W_g 为门控权重矩阵； $\text{LayerNorm}(\cdot)$ 表示层归一化操作； s_i 为门控融合后的节点表示。接着采用前馈神经网络 $F(\cdot)$ 进一步实现精细化处理，如式(8)所示。

$$F(s_i) = W_2 \text{GeLU}(W_1 s_i + b_1) + b_2 \quad (8)$$

式中： W_1 、 W_2 、 b_1 、 b_2 为可训练参数； $\text{GeLU}(\cdot)$ 为高斯误差线性单元激活函数。

2.3 联合优化训练方法

为训练面向分布式光伏数据的时空注意力插补模型，本文设计了基于插补与重建的联合优化训练方法。该方法包括两个学习任务：人为缺失插补任务(artificial missing imputation task, AMIT)和观测值重建任务(observed reconstruction task, ORT)。

ORT 与基于循环神经网络的模型兼容性良好。然而，卷积和注意力机制本质上是非自回归的，所有输入数据并行。因此，若只对 ORT 进行训练，其将只关注最小化观测值的重建误差，忽略缺失值填充效果。为此，本文引入 AMIT，与 ORT 组成联合优化训练架构，如图 5 所示。

AMIT 与真实缺失插补不同，该方法通过对可观测值人为地添加缺失，构建一个有监督学习的插补问题。经人为掩码后，实际输入时间序列记为 \hat{X} ，其对应的缺失掩码向量为 \hat{M} ， $\hat{m}_{n,t}$ 与 $\hat{x}_{n,t}$ 分别表示 M 和 X 中第 n 个场站在 t 时刻的元素。 $\hat{m}_{n,t} = 0$ 表示 $\hat{x}_{n,t}$ 缺失，否则未缺失。为区分真实缺失和人为缺失，引入指示掩码向量 I ，其中元素 $i_{n,t} = 1$ 表示 $\hat{x}_{n,t}$ 为人为缺失，否则为非人为缺失。

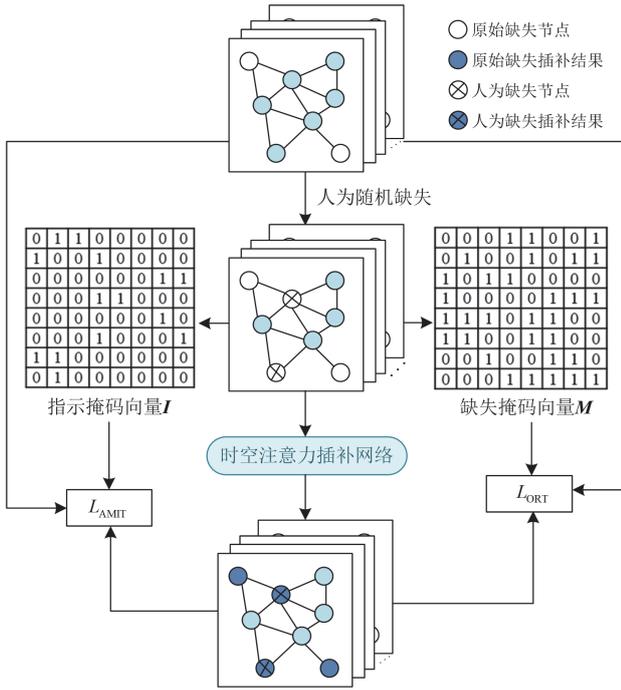


图5 联合优化训练架构示意图

Fig. 5 Diagram of joint optimization training architecture

模型在训练过程中最小化以下两个损失函数。

1) 人为缺失插补损失: 仅针对训练阶段人为构造的缺失位置评估插补准确性。

$$L_{AMIT} = \frac{\sum_{n=1}^N \sum_{t=1}^T |(\hat{x}_{n,t} - x_{n,t}) \cdot i_{n,t}|}{\sum_{n=1}^N \sum_{t=1}^T i_{n,t}} \quad (9)$$

式中: $\hat{x}_{n,t}$ 为第 n 个场站 t 时刻的插补结果。

2) 观测重建损失: 衡量模型在所有未缺失位置上的重建精度。

$$L_{ORT} = \frac{\sum_{n=1}^N \sum_{t=1}^T |(\hat{x}_{n,t} - x_{n,t}) \cdot \hat{m}_{n,t}|}{\sum_{n=1}^N \sum_{t=1}^T \hat{m}_{n,t}} \quad (10)$$

为应用联合优化训练, 本文模型损失设计为

$$L = L_{ORT} + \lambda L_{AMIT} \quad (11)$$

式中: L 为 SATCN 模型的总损失, 通过最小化该总损失实现模型训练; λ 为可调参数, 设置为 1。需要注意的是, 人为缺失仅应用于模型的训练阶段。

3 实验

实验在搭载 AMD Ryzen 5 5600H CPU 与 NVIDIA GeForce RTX 3050 GPU 的硬件平台上进行, 并采用 PyCharm 作为编译环境, 以 Python 3.8、PyTorch 1.13 与 CUDA 11.6 作为开发框架。

3.1 数据准备

本文实验部分使用美国国家可再生能源实验室 (national renewable energy laboratory, NREL) 发布的

光伏数据集^[35]。该数据集包含美国 38 个州 2006 年的光伏数据^[36]。本研究使用纽约州 8 个光伏场站(装机容量为 33~35 MW)2006 年共 366 天出力数据, 原数据集采集时间步长为 5 min, 本研究将其统一设置为 15 min, 即每天 96 个采样点。

缺失数据修复是一种无监督学习过程, 对有缺失的数据集无法验证插补效果。为定量评估算法性能, 本文在原本完整的无缺失数据集上生成缺失数据, 并将这些缺失视为需要修复的真实缺失。具体地, 本文基于实际光伏出力数据缺失的常见情况设计了单点缺失和连续片段缺失两种光伏出力数据缺失场景。为进一步模拟现实中连续缺失的随机性, 采用随机长度和随机位置生成连续片段缺失, 并将其与单点缺失按 1:1 的比例混合。需要指出的是, 用于训练的数据无需划分训练集与测试集。

3.2 评价指标

本文采用均方根误差(root mean square error, RMSE) e_{RMSE} 、平均绝对误差(mean absolute error, MAE) e_{MAE} 和平均相对误差(mean relative error, MRE) e_{MRE} 来评估插补性能。

$$e_{RMSE} = \sqrt{\frac{\sum_{n=1}^N \sum_{t=1}^T ((\hat{x}_{n,t} - x_{n,t}) \cdot (1 - m_{n,t}))^2}{\sum_{n=1}^N \sum_{t=1}^T (1 - m_{n,t})}} \quad (12)$$

$$e_{MAE} = \frac{\sum_{n=1}^N \sum_{t=1}^T |(\hat{x}_{n,t} - x_{n,t}) \cdot (1 - m_{n,t})|}{\sum_{n=1}^N \sum_{t=1}^T (1 - m_{n,t})} \quad (13)$$

$$e_{MRE} = \frac{\sum_{n=1}^N \sum_{t=1}^T |(\hat{x}_{n,t} - x_{n,t}) \cdot (1 - m_{n,t})|}{\sum_{t=1}^T |x_{n,t} \cdot (1 - m_{n,t})|} \quad (14)$$

3.3 实验设置

3.3.1 基准插补方法

本文选取了以下 5 种基准方法。1) KNN: 依据最近邻样本进行插值; 2) MICE: 通过链式方程进行多重插补; 3) 生成对抗插补网络 (generative adversarial imputation network, GAIN)^[37]: 基于生成对抗网络的插补方法, 参数设置参照文献[21]; 4) 双向递归时间序列插补 (bidirectional recurrent imputation for time series, BRITS)^[30]: 基于 BiLSTM 的多阶段循环插补方法, 挖掘时间依赖性和多维数据间的潜在耦合关系; 5) 基于自注意力机制的时间序列插补 (self-attention-based imputation for time series, SAITS)^[38]: 从对角掩码自注意力块的加权组合中学习缺失值, 实现多维时间序列插补。

3.3.2 实验设计

本文设计了对比实验、联合优化训练实验、消融实验和稳健性实验来测试 SATCN 的性能。训练

阶段在可观测点按 1:1 生成随机单点缺失和连续片段缺失, 后者基于随机缺失数据长度和位置生成, 引入人为缺失。在无特殊说明时, 所有实验均在人工缺失率(artificial missing ratio, AMR)为 0.2 的设置下进行训练。

模型训练中, 超参数的设置至关重要。本文采用网格搜索确定模型维度为 64。使用 Adam 优化器, 学习率为 0.001, 权重衰减为 0.0001, beta 参数为(0.9, 0.99), 批量大小为 16。为保证实验结果的可复现性, 所有实验均将随机种子固定为 26。

3.4 实验结果

3.4.1 对比实验

为验证所提方法的有效性, 针对不同的真实缺失率(real missing rate, RMR)进行实验, 缺失率设定为 0.1, 0.2, ..., 0.8。表 1 展示了各插补方法在不同缺失率下的插补误差, 并将其可视化如图 6。可见, 随着缺失率增大, 所有方法插补性能均下降, 本文所提方法在所有缺失率下均显著优于其他基准方法。

图 7 展示了场站 3 和场站 7 在 0.2 缺失率下, 不同方法对某 3 天(记为 DAY1、DAY2、DAY3)光伏功率的插补结果与真实值的对比。图 8 为 0.7 缺

表 1 不同缺失率下的插补误差(RMSE/MAE/(MRE%))

Table 1 Imputation errors under different missing rate (RMSE/MAE/(MRE%))

方法	缺失率							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
KNN	1.57/0.66/44.3	1.75/0.77/54.6	1.84/0.83/65.0	2.01/0.94/77.5	2.08/0.98/87.7	2.31/1.09/94.6	2.75/1.30/104.2%	3.59/1.85/118.7
MICE	1.36/0.59/45.7	1.41/0.61/52.4	1.43/0.61/50.5	1.74/0.76/56.8	1.90/0.88/63.3	2.32/1.21/74.5	2.90/1.62/98.0	6.34/4.91/238.8
GAIN	1.55/0.74/61.4	1.52/0.80/70.5	1.57/0.76/74.2	1.70/0.90/66.0	1.76/0.90/74.7	1.91/0.98/76.5	2.19/1.17/81.6	2.53/1.36/107.7
BRITS	1.17/0.51/41.2	1.15/0.56/46.8	1.20/0.53/44.2	1.21/0.60/41.2	1.35/0.67/45.5	1.38/0.69/51.5	1.59/0.88/56.8	1.80/1.00/60.7
SAITS	1.46/0.70/50.1	1.41/0.74/50.7	1.44/0.70/50.6	1.48/0.84/50.9	1.46/0.78/49.9	1.45/0.77/51.6	1.51/0.81/51.8	1.66/0.93/58.3
SATCN	1.11/0.46/34.8	1.09/0.44/38.1	1.14/0.46/38.3	1.15/0.48/37.3	1.16/0.47/37.2	1.19/0.50/42.5	1.26/0.54/43.4	1.48/0.64/45.5

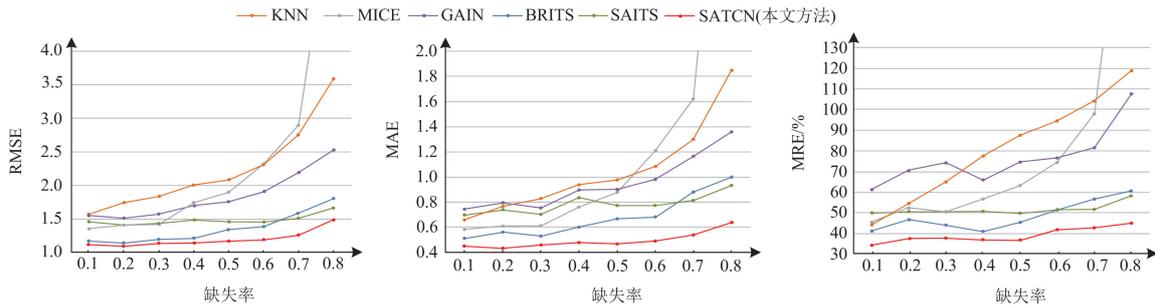
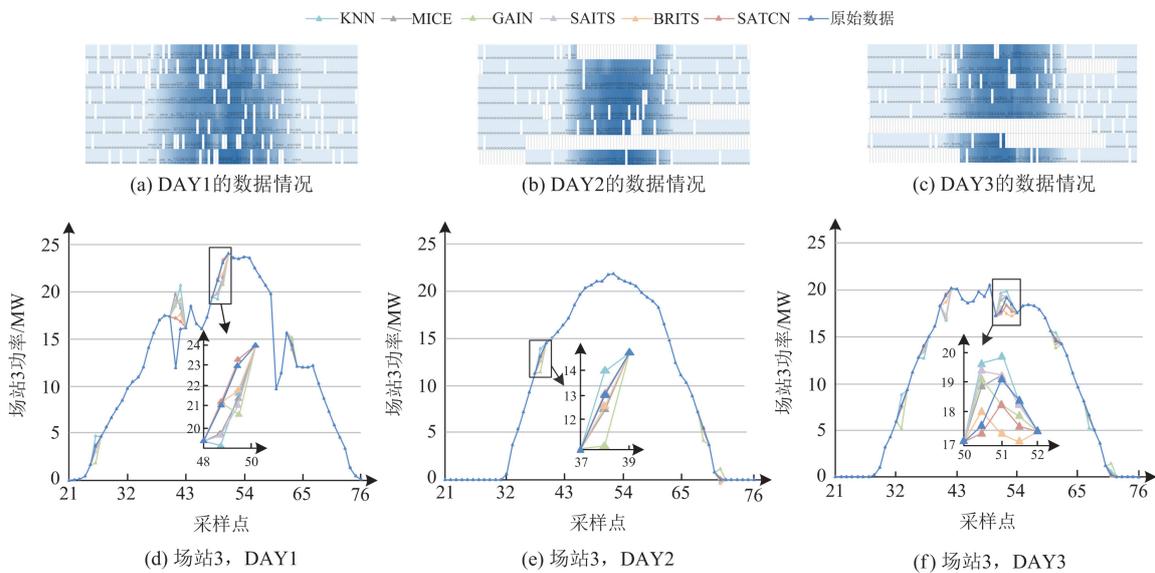


图 6 对比实验误差情况

Fig. 6 Error situation of comparative experiment



(d) 场站3, DAY1

(e) 场站3, DAY2

(f) 场站3, DAY3

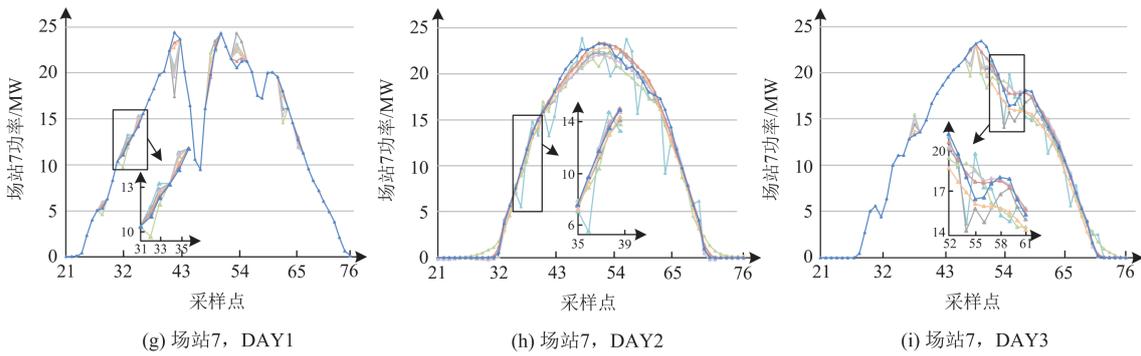


图 7 0.2 缺失率下数据插补效果对比图

Fig. 7 Comparison chart of data interpolation effects at a 0.2 missing rate

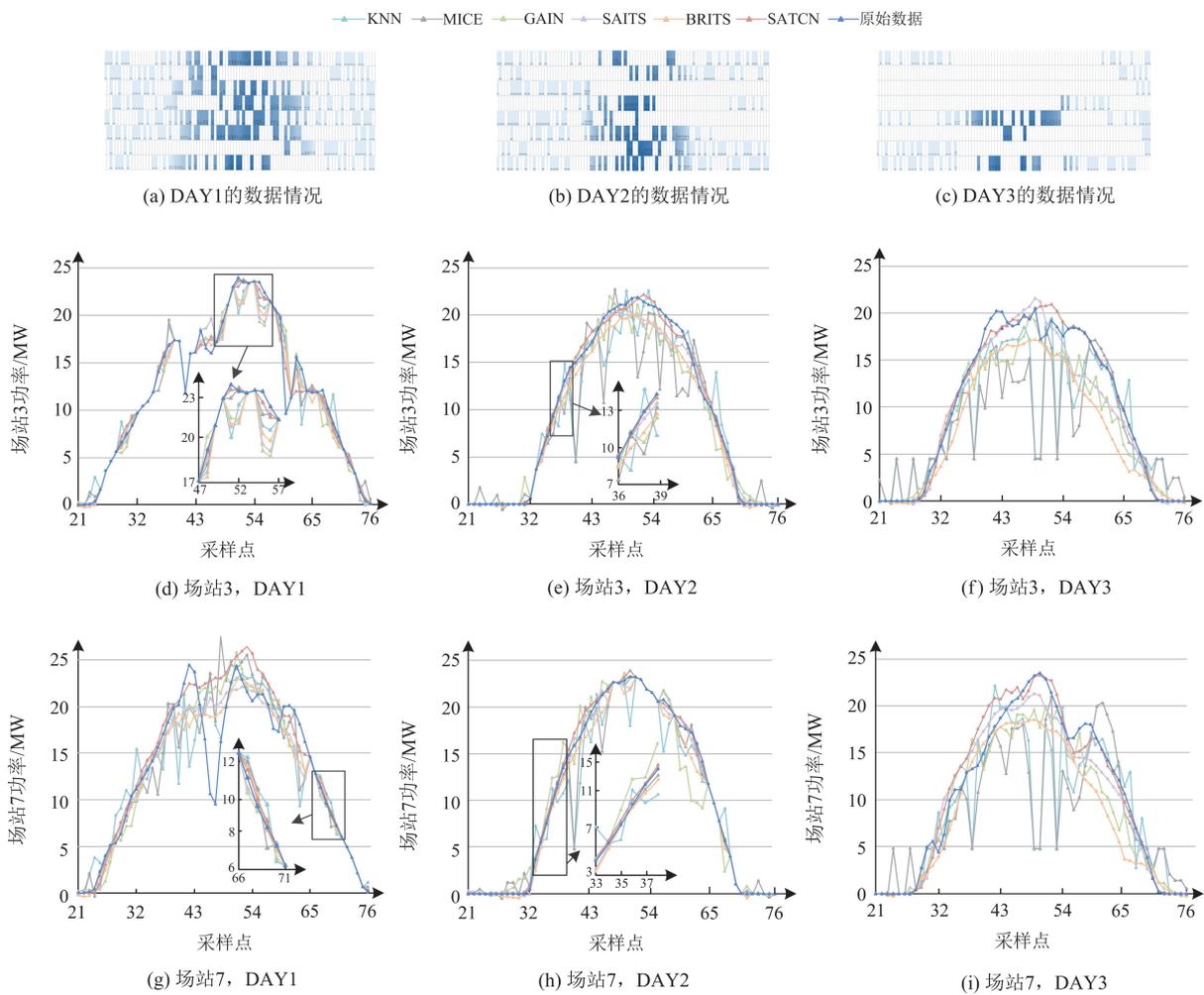


图 8 0.7 缺失率下数据插补效果对比图

Fig. 8 Comparison chart of data interpolation effects at a 0.7 missing rate

失率下的对比结果。图 8(i)中各曲线稀疏, 无需进行局部放大。在两种缺失率下, 传统方法均存在较大波动, 特别在连续缺失时无法复原数据原本的趋势。BRITS 在较低缺失率下表现更佳, 而 SAITS 在高缺失率下更具优势。相比之下, 本文提出的

SATCN 能很好地应对各种缺失率下的复杂缺失情况。即使在数据连续长时间缺失的情况下, 本文方法也能更准确地还原光伏功率的波动趋势。

从图 9 可以看出, 由于多个光伏场站在地理位置、装机容量和气象条件等方面存在显著差异, 各

插补方法在不同场站的误差分布存在显著差异。其中本文所提方法的平均绝对误差雷达图轮廓保持近似正八边形，显示出较强的鲁棒性，且在各场站的插补效果均为最优。

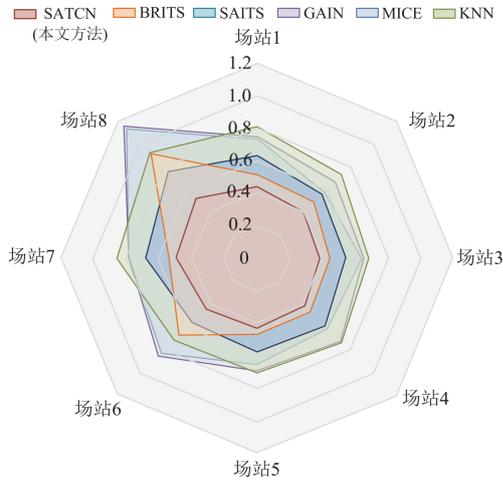
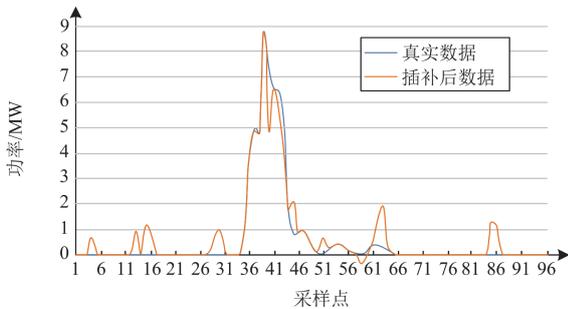


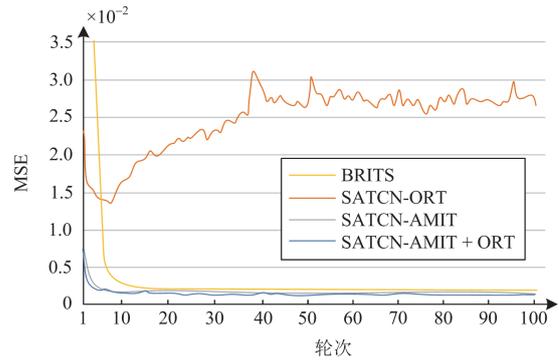
图 9 0.2 真实缺失率下各场站不同插补方法的 MAE
Fig. 9 MAE of different imputation methods at each station under 0.2 RMR

3.4.2 联合优化训练分析

图 10(a)展示了某光伏场站的插补结果，可见仅考虑 ORT 的 SATCN 对观测值的重建效果良好，但忽略了缺失值。对 BRITS 和结合不同学习任务的 SATCN 分别记录 100 轮插补结果的误差。该误差定义为归一化后的插补结果与真实值之间的均方误差 (mean square error, MSE)，结果如图 10(b)所示，仅考虑 ORT 时，SATCN 的插补误差在前几个训练轮次中先下降后上升，最终在高误差区间内波动。而引入 AMIT 后，SATCN 在第一次迭代就表现出较好的插补效果，并随着迭代次数的增加迅速收敛至更优的插补效果。相比之下，BRITS 虽插补性能也较好，但其收敛速度较慢。从表 2 可以看出，ORT 与 AMIT 结合即联合优化训练的插补效果最佳。上述实验验证了联合优化训练的有效性和必要性。



(a) 仅考虑观测值重建任务的插补结果



(b) 联合优化训练过程误差示意图

图 10 插补结果与训练误差示意图

Fig. 10 Diagram of imputation results and training errors

表 2 联合优化训练插补误差(RMSE/MAE/(MRE%))

Table 2 Imputation errors of joint optimization training (RMSE/MAE/(MRE%))

RMR	ORT	AMIT	ORT+AMIT
0.2	5.20/2.49/87.4	1.09/0.44/38.8	1.09/0.44/38.1
0.7	6.181/3.220/80.2	1.296/0.578/46.8	1.258/0.540/43.4

为了确定最佳的 AMR，表 3 展示了 RMR 为 0.2 和 0.7，AMR 为 10%、20%和 30%时 SATCN 的插补性能。可见，RMR 较低时，AMR 变化对精度影响较小。而 RMR 较高时，本文所提方法在 AMR 为 20%时性能最佳，故使用该 AMR 进行训练。

表 3 不同 RMR 与 AMR 下的插补误差

Table 3 Imputation errors under different RMR and AMR

RMR	AMR(RMSE/MAE/(MRE%))		
	10%	20%	30%
0.2	1.11/0.45/37.4	1.093/0.436/38.1	1.09/0.45/37.4
0.7	1.30/0.57/43.8	1.26/0.54/43.4	1.26/0.54/48.5

3.4.3 消融实验

本文提出的 SATCN 主要由时间聚合器和空间聚合器构成。为评估各模块的贡献，分别在原始 SATCN 上移除各个子模块，构建以下消融实验。1) Original: 完整 SATCN; 2) -SAN: 移除空间聚合器; 3) -1D-CNN1: 移除时间聚合器; 4) -1D-CNN2&3: 移除数据重建模块; 5) -ClearSky: 移除晴空模型输入。消融实验结果如表 4 所示，SAN 模块对性能的影响最大，表明在混合缺失场景下，连续片段缺失的插补效果高度依赖于对空间依赖关系的建模。同时，移除晴空模型输入后插补性能下降，这证实了物理先验的有效性。

3.4.4 稳健性实验

在现实世界的数据集中，噪声是不可避免的。因此，本文实验向分布式光伏功率数据添加了服从

高斯分布 $N \in (0, \delta^2)$ 的随机噪声来测试 SATCN 的稳健性, δ 为噪声标准差。如图 11 所示, 添加噪声对 SATCN 的性能影响较小, 表明本文方法在应对数据噪声时具有较好的鲁棒性。

表 4 消融实验结果

Table 4 Result of ablation experiment

消融实验	RMR (RMSE/MAE/(MRE%))	
	0.2	0.7
Original	1.095/0.441/38.1	1.258/0.540/43.4
-SAN	4.029/1.613/119.4	4.358/2.167/119.6
-1D-CNN1	1.133/0.509/38.4	1.366/0.606/57.0
-1D-CNN2&3	1.210/0.529/40.6	1.454/0.659/48.7
-ClearSky	1.104/0.445/38.3	1.301/0.581/44.4

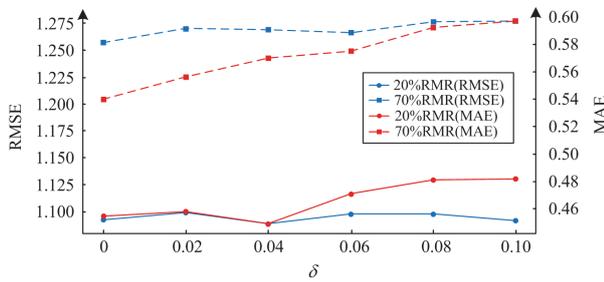


图 11 高斯分布噪声下的插补误差

Fig. 11 Imputation error under Gaussian distribution noise

4 结论

分布式光伏缺失数据插补是保障光伏功率预测准确性与系统性能评估准确性的关键环节, 其精度直接影响电力系统运行分析与调度决策的有效性。本文提出了一种融合 1D-CNN 和空间注意力网络的新型插补模型 SATCN, 修复分布式光伏场站功率缺失数据。通过引入联合优化训练架构, 有效克服了传统单一观测重建任务的训练局限性。实验结果表明, 在混合缺失模式且无气象监测数据的严苛场景下, 所提方法能够显著提高光伏数据复杂时空关联特征的捕捉能力; 在任意缺失率下, 其插补精度较基准方法均有所提高。

参考文献

[1] 国家能源局. 2024 年可再生能源并网运行情况[EB/OL]. [2025-01-27]. <https://www.nea.gov.cn/20250221/e10f363cabe3458aaf78ba4558970054/c.html>

[2] 杨康, 李蓝青, 李艺丰, 等. 一种新型分布式光伏出力区间预测方法[J]. 发电技术, 2024, 45(4): 684-695.
YANG Kang, LI Lanqing, LI Yifeng, et al. A novel distributed photovoltaic output interval prediction method[J]. Power Generation Technology, 2024, 45(4): 684-695.

[3] 宗炫君, 袁越, 蒋科, 等. 基于晴空指数与波动特性的光

伏电站可靠性分析[J]. 电力工程技术, 2019, 38(1): 36-41.
ZONG Xuanjun, YUAN Yue, JIANG Ke, et al. Reliability analysis for the PV power plant based on the clearness index and variability characteristic[J]. Electric Power Engineering Technology, 2019, 38(1): 36-41.

[4] GENES C, ESNALA I, PERLAZA S M, et al. Robust recovery of missing data in electricity distribution systems[J]. IEEE Transactions on Smart Grid, 2019, 10(4): 4057-4067.

[5] ROMERO-FIANCES I, LIVERA A, THERISTIS M, et al. Impact of duration and missing data on the long-term photovoltaic degradation rate estimation[J]. Renewable Energy, 2022, 181: 738-748.

[6] 黄泽, 毕贵红, 谢旭, 等. 基于 MBI-PBI-ResNet 的超短期光伏功率预测[J]. 电力系统保护与控制, 2024, 52(2): 165-176.
HUANG Ze, BI Guihong, XIE Xu, et al. Ultra-short-term PV power prediction based on MBI-PBI-ResNet[J]. Power System Protection and Control, 2024, 52(2): 165-176.

[7] 余凯峰, 吐松江·卡日, 张紫薇, 等. 基于级联 MCNN-MMLP 双残差网络的短期负荷预测[J]. 电力系统保护与控制, 2025, 53(2): 151-162.
YU Kaifeng, TUSONGJIANG·Kari, ZHANG Ziwei, et al. Short-term load forecasting based on a cascade MCNN-MMLP double residual network[J]. Power System Protection and Control, 2025, 53(2): 151-162.

[8] 张帅, 杨晶显, 刘继春, 等. 基于多尺度时序建模与估计的电力负荷数据恢复[J]. 电工技术学报, 2020, 35(13): 2736-2746.
ZHANG Shuai, YANG Jingxian, LIU Jichun, et al. Power load recovery based on multi-scale time-series modeling and estimation[J]. Transactions of China Electrotechnical Society, 2020, 35(13): 2736-2746.

[9] 韩晓, 王涛, 韦晓广, 等. 考虑阵列间时空相关性的超短期光伏出力预测[J]. 电力系统保护与控制, 2024, 52(14): 82-94.
HAN Xiao, WANG Tao, WEI Xiaoguang, et al. Ultrashort-term photovoltaic output forecasting considering spatiotemporal correlation between arrays[J]. Power System Protection and Control, 2024, 52(14): 82-94.

[10] 王子伟, 杨国林, 刘涛. 一种用于多维时间序列数据修复的贝叶斯张量分解算法[J]. 兰州交通大学学报, 2023, 42(6): 117-126.
WANG Ziwei, YANG Guolin, LIU Tao. A Bayesian tensor decomposition algorithm for repairing multi-dimensional time-series data[J]. Journal of Lanzhou Jiaotong University, 2023, 42(6): 117-126.

[11] 李永飞, 张耀, 林帆, 等. 基于气候特征分析及改进

- XGBoost 算法的中长期光伏电站发电量预测方法[J]. 电力系统保护与控制, 2024, 52(11): 84-92.
- LI Yongfei, ZHANG Yao, LIN Fan, et al. Medium- and long-term power generation forecast based on climate characterisation and an improved XGBoost algorithm for photovoltaic power plants[J]. Power System Protection and Control, 2024, 52(11): 84-92.
- [12] 梅玉杰, 李勇, 周王峰, 等. 基于机器学习的配电网异常缺失数据动态清洗方法[J]. 电力系统保护与控制, 2023, 51(7): 158-169.
- MEI Yujie, LI Yong, ZHOU Wangfeng, et al. Dynamic data cleaning method of abnormal and missing data in a distribution network based on machine learning[J]. Power System Protection and Control, 2023, 51(7): 158-169.
- [13] 杜婧涵, 胡明华, 尹嘉男. 基于链式方程多重插补的机场场面数据补全[J]. 航空计算技术, 2022, 52(5): 1-4.
- DU Jinghan, HU Minghua, YIN Jianan. Airport surface data completion based on multiple imputation by chained equations[J]. Aeronautical Computing Technique, 2022, 52(5): 1-4.
- [14] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[C] // Proceedings of the 27th International Conference on Neural Information Processing Systems, Cambridge: MIT Press, 2014: 2672-2680.
- [15] 周秋阳, 高辉, 李炜卓, 等. 基于 GAN 数据增强与改进 Bi-LSTM 的充电桩故障预测方法[J]. 电力工程技术, 2025, 44(6): 49-61.
- ZHOU Qiuyang, GAO Hui, LI Weizhuo, et al. A fault prediction method for charging pile based on GAN data enhancement and improved Bi-LSTM[J]. Electric Power Engineering Technology, 2025, 44(6): 49-61.
- [16] 王守相, 陈海文, 潘志新, 等. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法[J]. 中国电机工程学报, 2019, 39(1): 56-64.
- WANG Shouxiang, CHEN Haiwen, PAN Zhixin, et al. A reconstruction method for missing data in power system measurement using an improved generative adversarial network[J]. Proceedings of the CSEE, 2019, 39(1): 56-64.
- [17] 吕朋蓬, 卜强生, 郭野, 等. 基于梯度惩罚生成对抗网络的配电网缺失数据修复方法[J]. 太阳能学报, 2025, 46(5): 185-192.
- LÜ Pengpeng, BU Qiangsheng, GUO Ye, et al. Method for missing data imputation in distribution network based on gradient penalty generative adversarial network[J]. Acta Energetica Solaris Sinica, 2025, 46(5): 185-192.
- [18] 臧海祥, 郭镜玮, 黄蔓云, 等. 基于改进 Wasserstein 生成式对抗网络的电力系统不良数据辨识[J]. 电力自动化设备, 2022, 42(9): 50-56, 110.
- ZANG Haixiang, GUO Jingwei, HUANG Manyun, et al. Bad data identification of power system based on WGAN-GP[J]. Electric Power Automation Equipment, 2022, 42(9): 50-56, 110.
- [19] 杨玉莲, 齐林海, 王红, 等. 基于生成对抗和双重语义感知的配电网量测数据缺失重构[J]. 电力系统自动化, 2020, 44(18): 46-54.
- YANG Yulian, QI Linhai, WANG Hong, et al. Reconstruction of missing measurement data in distribution network based on generative adversarial network and double semantic perception[J]. Automation of Electric Power Systems, 2020, 44(18): 46-54.
- [20] 刘科研, 周方泽, 周晖. 基于时序信号图像编码和生成对抗网络的配电网台区数据修复[J]. 电力系统保护与控制, 2022, 50(24): 129-136.
- LIU Keyan, ZHOU Fangze, ZHOU Hui. Missing data imputation in a transformer district based on time series imaging encoding and a generative adversarial network[J]. Power System Protection and Control, 2022, 50(24): 129-136.
- [21] 刘科研, 周方泽, 周晖, 等. 基于改进生成对抗网络的台区采集数据修复[J]. 电网技术, 2022, 46(8): 3231-3239.
- LIU Keyan, ZHOU Fangze, ZHOU Hui, et al. Missing data imputation in transformer district based on improved generative adversarial network[J]. Power System Technology, 2022, 46(8): 3231-3239.
- [22] 殷豪, 丁伟锋, 陈顺, 等. 基于生成对抗网络和纵横交叉粒子群算法的光伏数据缺失重构方法[J]. 电网技术, 2022, 46(4): 1372-1381.
- YIN Hao, DING Weifeng, CHEN Shun, et al. Reconstruction method for missing data in photovoltaic based on generative adversarial network and crisscross particle swarm optimization algorithm[J]. Power System Technology, 2022, 46(4): 1372-1381.
- [23] 赵厚翔, 沈晓东, 吕林, 等. 基于 GAN 的负荷数据修复及其在 EV 短期负荷预测中的应用[J]. 电力系统自动化, 2021, 45(16): 143-151.
- ZHAO Houxiang, SHEN Xiaodong, LÜ Lin, et al. Load data restoration based on generative adversarial network and its application in short-term load forecasting of electric vehicle[J]. Automation of Electric Power Systems, 2021, 45(16): 143-151.
- [24] 管迎春, 何君, 牟令, 等. 风电场电能质量监测数据的时频多尺度特征修复方法[J]. 供用电, 2025, 42(8): 20-29, 60.
- GUAN Yingchun, HE Jun, MOU Ling, et al. Time-frequency multi-scale feature restoration method for wind

- farm power quality monitoring data[J]. *Distribution & Utilization*, 2025, 42(8): 20-29, 60.
- [25] 郭小龙, 李子康, 刘灏, 等. 基于增强生成对抗网络的PMU丢失数据恢复方法[J]. *电网技术*, 2022, 46(6): 2114-2121.
GUO Xiaolong, LI Zikang, LIU Hao, et al. PMU missing data recovery algorithm based on enhanced generative adversarial network[J]. *Power System Technology*, 2022, 46(6): 2114-2121.
- [26] 李富盛, 林丹, 余涛, 等. 基于改进生成式对抗网络的电气数据升频重建方法[J]. *电力系统自动化*, 2022, 46(3): 105-112.
LI Fusheng, LIN Dan, YU Tao, et al. Frequency-increased reconstruction method for electrical data based on improved generative adversarial network[J]. *Automation of Electric Power Systems*, 2022, 46(3): 105-112.
- [27] 刘志坚, 陶韵旭, 刘航, 等. 基于双通道生成对抗网络的城用电负荷缺失数据补全方法[J]. *电力系统自动化*, 2024, 48(17): 161-170.
LIU Zhijian, TAO Yunxu, LIU Hang, et al. Completion method for missing urban power load data based on double-channel generative adversarial networks[J]. *Automation of Electric Power Systems*, 2024, 48(17): 161-170.
- [28] 余晓霞, 汤宝平, 王伟影, 等. 复杂工况条件下多头注意力双向长短时记忆网络的风电机组缺失数据修复方法研究[J]. *机械工程学报*, 2023, 59(14): 1-9.
YU Xiaoxia, TANG Baoping, WANG Weiyang, et al. Repairing deteriorated data of wind turbines by multi-head attention bi-directional long short time memory networks under complex working conditions[J]. *Journal of Mechanical Engineering*, 2023, 59(14): 1-9.
- [29] 廖若愚, 刘友波, 沈晓东, 等. 基于双向循环插补网络的分布式光伏集群时序数据耦合增强方法[J]. *电网技术*, 2024, 48(7): 2784-2794.
LIAO Ruoyu, LIU Youbo, SHEN Xiaodong, et al. Time series data coupling enhancement method of distributed photovoltaic cluster based on bidirectional recurrent imputation network[J]. *Power System Technology*, 2024, 48(7): 2784-2794.
- [30] CAO W, WANG D, LI J, et al. BRITS: bidirectional recurrent imputation for time series[C] // *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, December 4-9, 2017, Long Beach, USA: 6775-6785.
- [31] SHEN G, ZHOU W, ZHANG W, et al. Bidirectional spatial-temporal traffic data imputation via graph attention recurrent neural network[J]. *Neurocomputing*, 2023, 531: 151-162.
- [32] INEICHEN P, PEREZ R. A new air mass independent formulation for the Linke turbidity coefficient[J]. *Solar Energy*, 2002, 73(3): 151-157.
- [33] HUANG L, HUANG J, LI H, et al. Robust spatial temporal imputation based on spatio-temporal generative adversarial nets[J]. *Knowledge-based Systems*, 2023, 279.
- [34] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks[C/OL] // *Proceedings of the 37th International Conference on Machine Learning*, July 13-18, 2020: 1725-1735.
- [35] NREL. The national renewable energy laboratory [EB/OL]. [2023-01-12]. <https://www.nrel.gov/grid/solar-power-data.html>
- [36] ZHANG P, WEI Z, HUANG Z, et al. A review of public datasets for photovoltaic power generation forecasting[J]. *Electric Power Information and Communication Technology*, 2023, 21(8): 16-21.
- [37] YOON J, JORDON J, SCHAAR M. GAIN: missing data imputation using generative adversarial nets[C] // *Proceedings of the 35th International Conference on Machine Learning*, July 10-15, 2018, Stockholm, Sweden: 5689-5698.
- [38] DU W, CÔTÉ D, LIU Y. SAITS: self-attention-based imputation for time series[J]. *Expert Systems with Applications*, 2023, 219.
-
- 收稿日期: 2025-06-05; 修回日期: 2025-11-25
作者简介:
余思杪(2001—), 女, 硕士研究生, 研究方向为深度学习、电力大数据处理和电力预测; E-mail: 2972260469@qq.com
许玉格(1978—), 女, 博士, 副教授, 研究方向为智能计算、深度学习和故障诊断。E-mail: xuyuge@scut.edu.cn
(编辑 张颖)