



DOI:10.12404/j.issn.1671-1815.2507461

引用格式:史加荣,王姣姣,栗雪娟.融合 BiMamba2 与非局部注意力的光伏电站群功率预测时空图网络模型[J].科学技术与工程,2026,26(26):11318-11325.

Shi Jiarong, Wang Jiaojiao, Li Xuejuan, et al. Temporal-spatial graph network model with BiMamba2 and non-local attention for photovoltaic cluster power prediction[J]. Science Technology and Engineering, 2026, 26(26): 11318-11325.

电工技术

融合 BiMamba2 与非局部注意力的光伏电站 群功率预测时空图网络模型

史加荣^{1,2}, 王姣姣², 栗雪娟²

(1. 绿色建筑全国重点实验室/西安建筑科技大学, 西安 710055; 2. 西安建筑科技大学理学院, 西安 710055)

摘 要 随着光伏发电占比提升,其功率的精准预测对电网稳定与消纳至关重要。然而,传统时间序列预测方法对光伏功率复杂动态的刻画能力有限,而现有基于图神经网络的时空预测模型在长时序依赖与全局空间相关性的捕捉上仍存在一定挑战,尤其在多步预测中易产生误差累积。为此,提出了一种改进的时空图神经网络模型:MTGNN-BiMamba2-NL。该模型在多元时间序列预测框架(multivariate time series forecasting framework based on graph neural networks, MTGNN)中,以双向状态空间模型(BiMamba2)替代时间卷积,以强化长时序特征提取。同时,在图卷积层后融入非局部(Non-local, NL)注意力机制,以增强站点间的全局空间关联建模。在两个真实光伏数据集上的实验表明,所提模型在多步预测任务中能有效缓解误差累积,其预测精度与稳定性均优于现有基线方法。

关键词 光伏功率预测;多元时间序列;图神经网络;非局部注意力;BiMamba2

中图法分类号 TM615;

文献标志码 A

Temporal-spatial Graph Network Model with BiMamba2 and Non-local Attention for Photovoltaic Cluster Power Prediction

SHI Jia-rong^{1,2}, WANG Jiao-jiao², LI Xue-juan²

(1. State Key Laboratory of Green Building, Xi'an University of Architecture and Technology, Xi'an 710055, China;

2. School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China)

[Abstract] With the increasing penetration of PV (photovoltaic) power generation, accurate prediction of PV output is essential for grid stability and power accommodation. However, traditional time series forecasting methods exhibit limited capability in characterizing the complex dynamics of PV power, while existing graph neural network based spatiotemporal forecasting models still face certain challenges in capturing long-term temporal dependencies and global spatial correlations, particularly in multi-step forecasting where error accumulation easily occurs. To address these limitations, an improved spatiotemporal graph neural network model, termed MTGNN-BiMamba2-NL, was proposed. Within MTGNN (multivariate time series forecasting framework based on graph neural networks), the temporal convolution module was replaced by a bidirectional state space model (BiMamba2) to enhance long-term temporal feature extraction. Meanwhile, a NL (Non-local) attention was incorporated after the graph convolution layers to strengthen the modeling of global spatial correlations among sites. Experiments were conducted on two real-world PV datasets. The results demonstrate that the proposed model effectively alleviates error accumulation in multi-step forecasting and achieves superior prediction accuracy and stability compared with existing baseline methods.

[Keywords] photovoltaic power forecasting; multivariate time series; graph neural network; non-local attention; BiMamba2

在全球能源结构绿色转型深入推进的背景下,光伏发电作为清洁能源体系的核心支柱,其装机容量和渗透率的持续提升对电力系统的安全稳定运行带来

了严峻挑战^[1]。光伏输出功率具有显著的间歇性和波动性特征,这使得高精度功率预测,特别是多站点协同预测,成为保障电网消纳能力、优化调度策略与

收稿日期:2025-10-27; 修订日期:2026-02-06

基金项目:绿色建筑全国重点实验室自主研究课题(LSZZ-Y202414);国家自然科学基金(12471410)

第一作者:史加荣(1979—),男,汉族,山东聊城人,博士,教授。研究方向:可再生能源预测。E-mail:shijiarong@xauat.edu.cn。

投稿网址:www.stae.com.cn

稳定电力市场交易的关键技术^[2]。光伏电站群功率预测的核心挑战在于其复杂的时空耦合关系:在时间维度上,功率序列具有强烈的非平稳与非线性特征;空间维度上,各站点虽地理分散,却因共同的气象条件而形成复杂的动态关联网络。

为应对光伏功率预测中的时序特征挑战,研究者提出了多种深度学习方法。例如,梁亚峰等^[3]结合完全经验模态分解与长短时记忆网络,将光伏功率预测中的多模态气象特征进行分解与建模,实现对不平稳时间序列的精确预测。柳天虹等^[4]基于分位数回归对门控循环单元(gated recurrent unit, GRU)进行结构改进,通过自适应权重调节增强了模型对新能源功率序列不确定性和突变特征的刻画能力。刘俊宏等^[5]将优化的时间卷积网络(temporal convolutional network, TCN)与注意力机制相结合,实现了光伏功率预测精度的提升,但其时间特征提取仍受限于卷积结构。Wang 等^[6]采用改进的 Transformer 模型捕捉长序列全局依赖,但其计算复杂度,难以适应大规模数据场景。近年来,状态空间模型(state space model, SSM),尤其是 Mamba 系列^[7],凭借线性计算复杂度与选择性机制,为长序列建模提供了新路径。基于此, Mamba2 进一步提升了计算效率与序列处理能力^[8]。

在空间特征提取方面,图神经网络(graph neural network, GNN)^[9]为多站点系统中的空间依赖关系刻画提供了有效的解决方案。该类方法将站点映射为图节点,并通过边表示站点间的关联关系。早期研究多使用预定义的邻接矩阵构建图结构。文献[10]基于 GraphSAGE 构建交通速度预测模型,借助邻居特征聚合机制,实现对恢复后缺失数据的高效利用,改善了稀疏路网下的短期交通速度预测性能。文献[11]利用图注意力网络(graph attention network, GAT)动态学习节点间关联权重,增强了空间特征表征的灵活性。

随着时空预测问题复杂度的提升,研究者开始将 GNN 与时序模型相结合,形成统一的时空预测范式。时空图卷积网络(spatio-temporal graph convolutional network, STGCN)将图卷积网络与一维时序卷积交替堆叠,实现了时空特征的联合提取^[12]。基于注意力机制的时空图卷积网络(attention-based STGCN, ASTGCN)通过时空注意力机制动态捕捉时空相关性,显著提升了模型表征能力^[13]。

然而,现有方法仍存在较大改进空间:在处理极长序列时 TCN 感受野有限,且无法充分利用双向上下文信息;传统的图卷积操作往往局限于局部邻域信息聚合,难以捕捉全局空间依赖。为此,现对

基于图神经网络的多元时间序列预测(multivariate time series forecasting with GNN, MTGNN)^[14]架构的卷积模块和注意力机制进行改进,并将其应用于多站点光伏功率预测中,以期能有效缓解多步预测误差累积,增强复杂场景下的预测鲁棒性,为高比例光伏并网调度提供可靠支撑。

1 模型原理

1.1 问题定义

在多站点光伏功率预测任务中,假设共有 N 个光伏电站节点,每个节点在时间步 t 的特征向量记为 $\mathbf{x}_t^n \in \mathbf{R}^c$, 其中, c 为输入特征维度, $n = 1, 2, \dots, N$ 。将所有节点在连续 T 个时间步的观测组成输入张量为

$$\mathbf{X}_{t-T:t} = [\mathbf{x}_{t-T+1}^n, \mathbf{x}_{t-T+2}^n, \dots, \mathbf{x}_t^n]_{n=1}^N \in \mathbf{R}^{T \times N \times c} \quad (1)$$

基于历史输入 $\mathbf{X}_{t-T:t}$, 学习一个非线性映射函数 $F(\cdot)$, 用于预测未来 τ 个时间步的功率输出。

$$\hat{\mathbf{Y}}_{t+1:t+\tau} = F(\mathbf{X}_{t-T:t}) \in \mathbf{R}^{\tau \times N} \quad (2)$$

1.2 图学习层

光伏电站输出功率受云层运动和局部气象变化等时空因素动态影响,传统的基于固定地理距离或简单相关系数的图结构难以准确刻画这种时变关联。而图学习层通过可训练的节点嵌入自适应地捕捉动态关联模式。同时考虑到光伏功率传播具有明显方向性(如上游站点对下游站点存在延迟影响而反向作用较弱),图学习层能够提取反映物理机制的单向关系,从而为后续图卷积操作提供合理的邻接结构。

基于站点经纬度构建初始邻接矩阵 $\mathbf{A} = (a_{ij})_{N \times N}$, 以提供物理意义上的空间先验, 其中: $a_{ij} = \exp[-d_{ij}^2/(2\sigma^2)]$, d_{ij} 为节点 i 与 j 的球面距离, σ 为距离标准差。再引入可训练的节点嵌入以自适应更新邻接矩阵, 计算公式为

$$\begin{cases} \mathbf{M}_1 = \tanh(\alpha \mathbf{E}_1 \boldsymbol{\Theta}_1) \\ \mathbf{M}_2 = \tanh(\alpha \mathbf{E}_2 \boldsymbol{\Theta}_2) \end{cases} \quad (3)$$

$$\mathbf{A}^* = \text{ReLU} \{ \tanh[\alpha (\mathbf{M}_1 \mathbf{M}_2^T - \mathbf{M}_2 \mathbf{M}_1^T)] \} \quad (4)$$

式中: $\mathbf{E}_1, \mathbf{E}_2 \in \mathbf{R}^{N \times d}$ 为随机初始化的节点嵌入; $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ 为可学习参数矩阵; α 为控制激活强度; $\mathbf{A}^*$ 为邻接矩阵; $\tanh(\cdot)$ 和 $\text{ReLU}(\cdot)$ 分别为双曲正切、修正线性单元激活函数。

式(4)通过反对称矩阵运算保证邻接矩阵 $\mathbf{A}^* = (a_{ij}^*)_{N \times N}$ 的非对称性,从而刻画光伏功率在时空网络传播的方向特征。

为提升计算效率,采用基于相似度的稀疏化策略,确保模型关注物理意义最显著的节点。稀疏化邻接矩阵 $\mathbf{A}_{\text{sp}} = (a_{ij}^{\text{sp}})_{N \times N}$ 的定义为

$$a_{ij}^{\text{sp}} = \begin{cases} a_{ij}^*, & j \in \text{Topk}(A_i^*, K) \\ 0, & j \notin \text{Topk}(A_i^*, K) \end{cases} \quad (5)$$

式(5)中: $\text{Topk}(A_i^*, K)$ 为第 i 个节点的邻接权重向量 A_i^* 中数值最大的 K 个索引集合。

1.3 图卷积模块

图卷积模块将每个节点的信息与其邻居信息结合,以捕捉空间依赖关系。该模块采用两个混合跳跃(Mix-hop)传播架构:正向和反向。以正向传播为例,每个 Mix-hop 传播层包含信息传播和信息选择两个关键计算阶段,结构如图 1 所示,其中 MLP 表示多层感知机。信息传播阶段通过多轮迭代传递邻居节点信息,逐步融合来自不同跳的邻居特征。该过程保留节点自身特征,以缓解过平滑现象,第 m 步传播后节点特征为

$$H^m = \beta H_{\text{in}} + (1 - \beta) \tilde{A} H^{m-1} \quad (6)$$

式(6)中: m 为信息传播深度,即信息从节点出发通过图的邻接关系传播的步数; β 为超参数,控制原始节点信息的留存比例; H_{in} 为输入特征。

取 $H^0 = H_{\text{in}}$, $\tilde{A} = \tilde{D}^{-1}(A_{\text{sp}} + I)$ 为归一化的邻接矩阵,其中, $\tilde{D}$ 和 I 分别为度矩阵和单位矩阵。

信息选择步骤采用可学习的线性变换,对不同传播步的特征进行自适应加权融合,定义为

$$H_{\text{out}} = \sum_{m=0}^M H^m W^m \quad (7)$$

式(7)中: W^m 为第 m 步可学习参数矩阵; M 为最大传播深度。

该设计可动态调节各跳邻居贡献度,从而更精准地捕捉局部与全局的空间依赖模式。

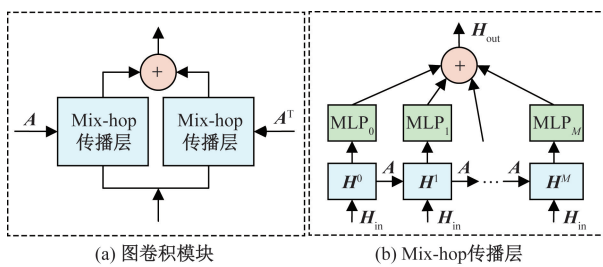


图 1 图卷积模块结构

Fig. 1 Graph convolution module structure

1.4 时间卷积模块

时间卷积模块从时间维度上捕捉动态依赖关系。其核心设计融合了扩张卷积与 Inception 结构:在同一层中设置多个不同尺度的卷积核,能够同时提取短期与长期时序特征,并在特征维度上进行拼接,以增强表示能力。此外,该模块还引入了门控机制, $\sigma(\cdot)$ 作为门控,二者逐点相乘以控制信息流动。时间卷积模块结构如图 2 所示。

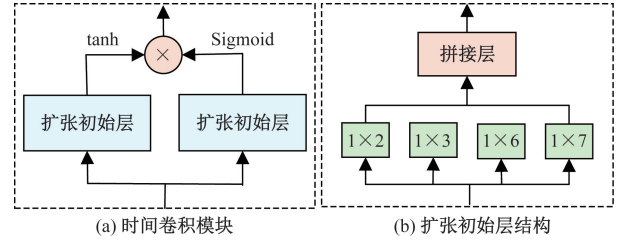


图 2 时间卷积模块结构

Fig. 2 Temporal convolution module structure

2 光伏电站群功率预测的 MTGNN-BiMamba2-NL 模型

2.1 MTGNN

基于图神经网络的多元时间序列预测 (multivariate time series forecasting with graph neural networks, MTGNN) 是一种端到端的深度学习模型。它集成了自适应图学习、图卷积与时间卷积,同时捕捉变量间的空间依赖关系和时间动态特征,且不依赖预定义的图结构,其框架如图 3 所示。在多站点光伏功率预测任务中,MTGNN 自动识别站点间由气象与地理因素驱动的潜在关联,并利用时序特征提取功率数据中的周期与波动模式,从而显著提升多站点协同预测的准确性与鲁棒性。

MTGNN 由 1 个图学习层、多层堆叠的图卷积与时间卷积模块以及 1 个输出模块构成。其中,图学习层自适应地推断节点间的隐含关系,并生成邻接矩阵,以供后续图卷积使用。图卷积与时间卷积交替排列,分别用于提取空间与时间维度上的特征。MTGNN 采用了残差连接,将时间卷积的输入直接与随后的图卷积输出相加,以缓解梯度消失问题。同时,在每个时间卷积后引入跳跃连接,以增强信息流动。最终,输出模块将提取到的隐特征映射至目标预测维度。

2.2 BiMamba2 模块

不同于基于扩张卷积的时间处理方式,从时序范式层面在 MTGNN 中引入双向状态空间模型 BiMamba2^[15],构建面向长时序依赖的时间表征机制。该方法不受有限感受野约束,通过状态空间递推建模序列动态,稳定高效地捕捉光伏功率的长程时间相关性。

Mamba2 基于 SSM,在保持线性复杂度的同时,利用选择性机制和参数动态化设计来提升对长时依赖的表征能力。其结构如图 4 所示。对于输入特征 x_t^n ,首先通过线性映射生成中间表示,即

$$\begin{cases} u_t^n = W_u x_t^n + b_u \\ v_t^n = W_v x_t^n + b_v \end{cases} \quad (8)$$

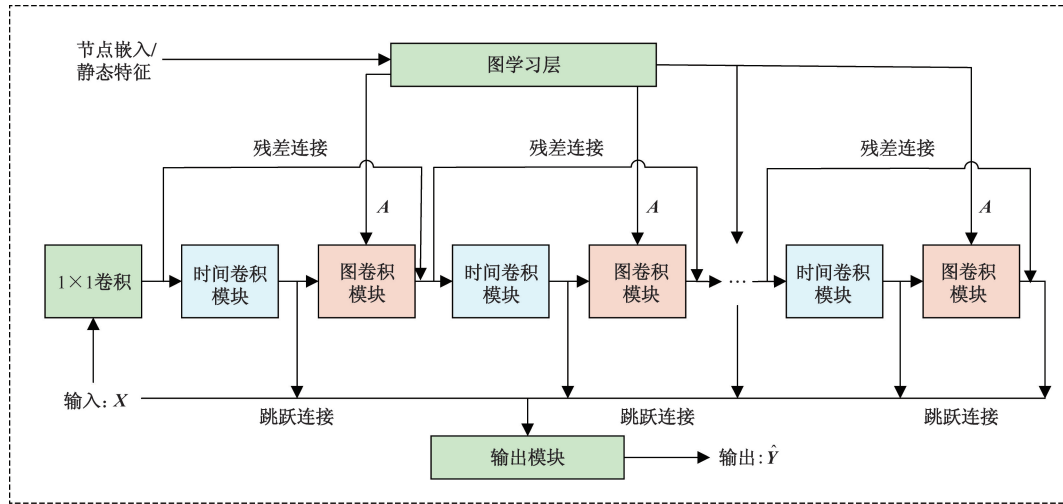


图3 MTGNN 模型框架

Fig.3 MTGNN model framework

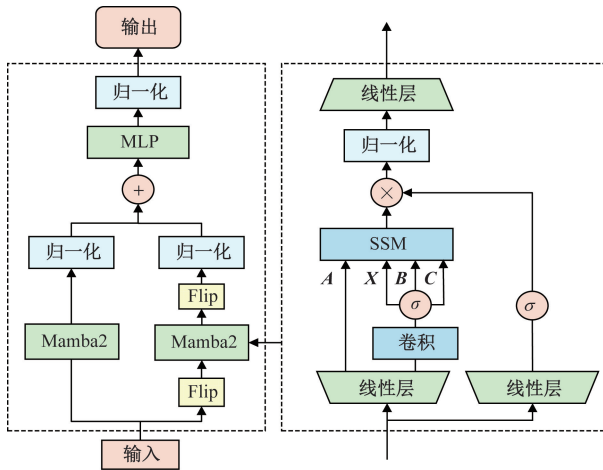


图4 BiMamba2 网络结构

Fig.4 BiMamba2 network structure

式(8)中: u_t^n 用于驱动状态空间模型的动态建模过程; v_t^n 用于生成输入相关的门控信号; W_u 、 W_v 为权重矩阵; b_u 、 b_v 为偏置向量。

接着, Mamba2 内部基于 SSM 对序列动态进行递推建模, 其离散化可表示为

$$s_t^n = A(u_t^n) s_{t-1}^n + B(u_t^n) u_t^n \quad (9)$$

$$y_t^n = C(u_t^n) s_t^n \quad (10)$$

式中: s_t^n 为当前时刻的隐状态; $A(\cdot)$ 、 $B(\cdot)$ 、 $C(\cdot)$ 为输入驱动生成的状态转移矩阵、输入投影矩阵和输出投影矩阵; y_t^n 为 SSM 输出的中间时序表示。

为增强模型对关键信息的选择性表达能力, Mamba2 进一步引入输入驱动的门控机制, 对状态空间模型的输出进行调制, 即

$$g_t^n = \sigma(W_g v_t^n + b_g) \quad (11)$$

式(11)中: g_t^n 为选择性门控向量; W_g 和 b_g 分别为门控层的可学习权重矩阵与偏置向量。

随后, 将 SSM 的输出与门控信号进行逐元素调制, 并通过输出投影、残差连接及层归一化操作, 得到 Mamba2 的最终输出特征为

$$o_t^n = \text{LN}[x_t^n + W_o(y_t^n \odot g_t^n)] \quad (12)$$

其中: W_o 为输出投影矩阵; $\odot$ 表示逐元素乘法; $\text{LN}(\cdot)$ 为层归一化操作。

为进一步捕捉时间序列的正向与反向依赖特征, 在 Mamba2 基础上构建了一种双向架构: BiMamba2, 其由两个并行 Mamba2 通道组成, 即

$$\vec{o}_t^n = \text{Mamba2}(x_t^n) \quad (13)$$

$$\overleftarrow{o}_t^n = \text{Mamba2}[\text{Flip}(x_t^n)] \quad (14)$$

$$o_t^n = \vec{o}_t^n + \text{Flip}(\overleftarrow{o}_t^n) \quad (15)$$

式中: $\vec{o}_t^n$ 、 $\overleftarrow{o}_t^n$ 分别为前向、反向依赖; $\text{Flip}(\cdot)$ 为时间维度上的反向操作。

对最终的双向表示 o_t^n 进行归一化处理, 经 MLP 得到最终预测结果。

2.3 非局部注意力机制

非局部(non-local, NL)注意力机制^[16]通过显式刻画任意节点对之间的相关性, 实现全局特征的加权聚合, 从而在单次运算中捕捉跨节点的长程空间依赖。而传统图卷积主要依赖预定义邻接关系来聚合局部邻域信息, 对远距离节点间潜在空间关联的表达能力有限。基于此, 在图卷积特征提取基础上引入 NL 注意力机制, 以增强对跨节点全局空间关系的表征能力。

在某一时间步, 图卷积模块的输出特征设为 $H_{\text{out}} = [h_1, h_2, \dots, h_N]$, 其中 h_n 为站点 n 的特征向量。聚合后的特征为

$$y_n = \frac{1}{C(H_{\text{out}})} \sum_{j=1}^N f(h_n, h_j) g(h_j) \quad (16)$$

式(16)中: $f(\cdot, \cdot)$ 为相似度函数; $g(\mathbf{h}_j) = \mathbf{W}_g \mathbf{h}_j$ 为特征线性映射; $\mathbf{W}_g$ 为可学习的权重矩阵; $C(\mathbf{H}_{out})$ 为归一化因子。点积相似度为

$$f(\mathbf{h}_n, \mathbf{h}_j) = \theta(\mathbf{h}_n)^T \varphi(\mathbf{h}_j) \quad (17)$$

$$\theta(\mathbf{h}) = \mathbf{W}_\theta \mathbf{h}, \varphi(\mathbf{h}) = \mathbf{W}_\varphi \mathbf{h} \quad (18)$$

式中: $\mathbf{W}_\theta$ 、 $\mathbf{W}_\varphi$ 均为可学习的权重矩阵。

最终输出的残差形式为

$$\mathbf{z}_n = \mathbf{W}_z \mathbf{y}_n + \mathbf{h}_n \quad (19)$$

式(19)中: $\mathbf{W}_z$ 为可学习的权重矩阵。

2.4 跳跃连接与输出模块

为保持特征传递过程中维度匹配,并促进不同层级间信息的有效整合,在模型中设计了跳跃连接层。采用 $1 \times L_i$ 的一维标准卷积核,其中 L_i 表示第 i 层跳跃连接输入的序列长度,使其在时间维度上保持一致,从而实现多层特征的累加并作为输出模块的输入。输出模块由两个 1×1 标准卷积层组成,负责将通道维度转换至预期的输出维度。

2.5 MTGNN-BiMamba2-NL 模型框架

为增强模型对光伏电站群功率序列中长期时空依赖特征的表征能力,在 MTGNN 框架下设计了多站点光伏功率预测模型 MTGNN-BiMamba2-NL,通过融合 BiMamba2 与 NL 注意力机制,实现对时间与空间特征的统一建模。具体而言,BiMamba2 模块以状态空间递推方式捕捉序列动态,显著提升对长时间序列依赖的捕捉能力;NL 注意力模块在图卷积特征基础上对跨站点空间关系进行全局强化,从而有效提升模型对光伏功率时空特征的综合表征能力。所提模型结构如图 5 所示。

3 算例分析

3.1 数据描述与参数设置

为验证模型在不同规模光伏站点群的适用性,

选取两个在站点数量上有显著差异的数据集:含 86 个光伏电站的亚利桑那州数据集和涵盖 16 个站点的特拉华州分布式光伏数据集,它们均源自美国国家可再生能源实验室(NREL, <https://www2.nrel.gov/grid/solar-power-data>)。这两个数据集覆盖 2006 年全年,时间分辨率为 10 min,其站点分布如图 6 所示。依据日出日落时间,剔除每日 18:00 至次日 04:00 的无发电时段,仅保留有效发电区间。预测任务为用过去 12 个时间步的数据预测未来 8 步的光伏出力,训练集、验证集和测试集的比例为 7:1:2。亚利桑那州单个光伏电站的装机容量介于 10 ~ 100 MW,特拉华州单个光伏电站容量则介于 7 ~ 70 MW。

在 MTGNN-BiMamba2-NL 模型中,设置图卷积模块和 BiMamba2 模块的数量均为 3,图卷积信息传播深度 $M = 3$,BiMamba2 模块的隐藏状态和状态空间的维数分别设为 32、16。在训练过程中,采用均方误差作为损失函数,使用 Adam 优化器,设学习率为 0.001、批量大小为 32。

3.2 评价指标

为评估模型的预测性能,采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)和决定系数(R -square, R^2)作为评价指标。

$$\text{MAE} = \frac{1}{N\tau} \sum_{n=1}^N \|\mathbf{y}^n - \hat{\mathbf{y}}^n\|_1 \quad (20)$$

$$\text{RMSE} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{\tau} \|\mathbf{y}^n - \hat{\mathbf{y}}^n\|_2^2} \quad (21)$$

$$R^2 = 1 - \frac{\sum_{n=1}^N \|\mathbf{y}^n - \hat{\mathbf{y}}^n\|_2^2}{\sum_{n=1}^N \|\mathbf{y}^n - \frac{1}{N\tau} \sum_{n=1}^N \mathbf{y}^n\|_2^2} \quad (22)$$

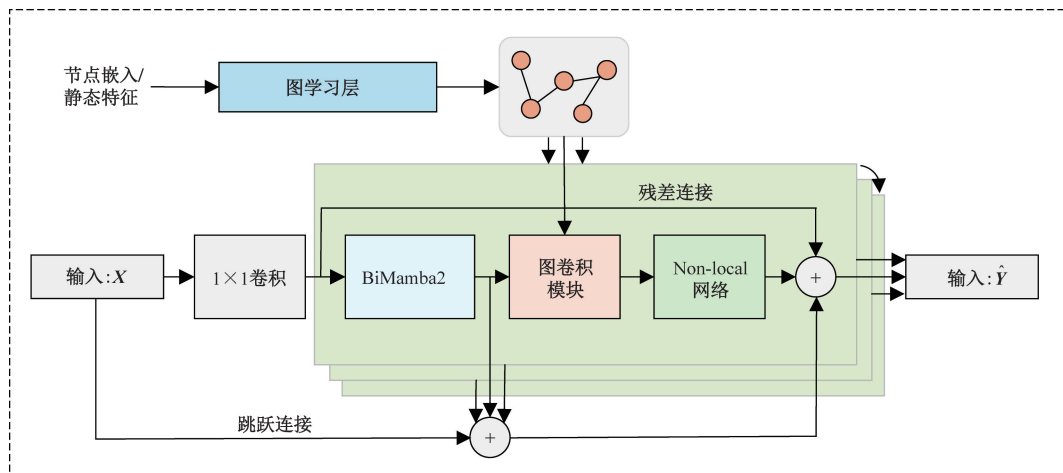


图 5 MTGNN-BiMamba2-NL 模型结构

Fig. 5 MTGNN-BiMamba2-NL model structure

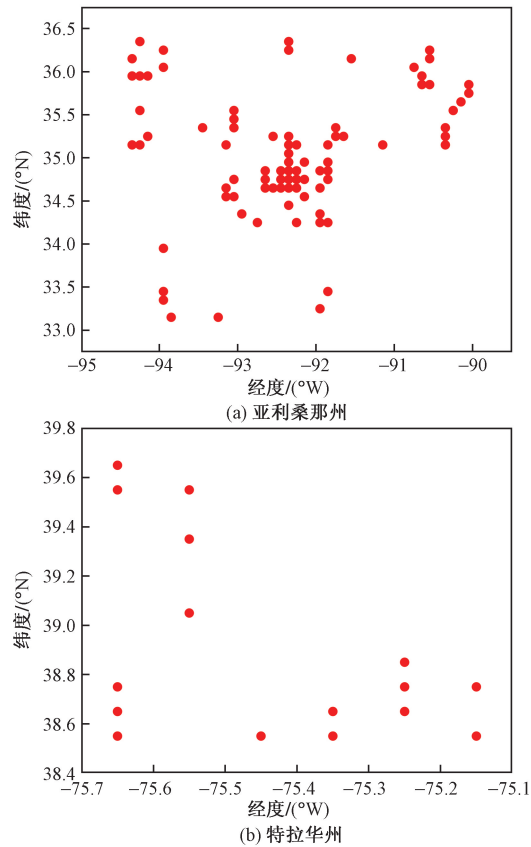


图6 两个数据集的站点分布

Fig. 6 Site distribution of the two datasets

式中: y^n 和 $\hat{y}^n$ 分别为第 n 个站点的真实值和预测值。

3.3 预测结果及分析

为全面评估所提模型在多站点光伏预测中的性能,选取以下3类基线模型进行对比。

①时序模型(GRU和TCN):专注挖掘单站点时序依赖,未考虑站点间空间关联,通过对各电站单独建模并累加得到预测结果;②图神经网络(GCN、GAT和GraphSAGE):能显式建模站点空间结构,但对动态时序变化建模不足;③时空模型(STGCN和

MTGNN):作为主流时空序列预测方法,可同时捕捉时空依赖性。

所有对比模型参数如表1所示,其优化器、批大小、学习率均与MTGNN-BiMamba2-NL一致,训练100轮。实验采用相同数据划分与输入条件,并设置预测步长为1、4和8,以检验模型在不同时间尺度下的预测能力。

两个数据集上不同步长下的预测性能如表2和表3所示,可以看出:所有模型的预测精度均随预测步长的增加而下降;本文方法在所有预测步长上均表现最优,且在多步预测任务中优势更为显著。

在对比实验中,单一时序模型在单步预测中表现尚可,但在4步、8步预测中性能显著下降,明显劣于引入空间关系的图神经网络。图神经网络虽在短期精度上不及时空模型,但其多步预测指标退化幅度更平缓,稳定性更强。相比之下,时空图模型性能显著优于前两类,验证了时空建模架构的有效性。

以亚利桑那州数据集为例,与MTGNN相比,MTGNN-BiMamba2在4步、8步预测的MAE分别降低9.3%和14.3%,验证了BiMamba2在捕获长时序依赖上的优势。进一步引入NL模块后,MTGNN-BiMamba2-NL在单步预测中MAE再降低9.1%,且多步预测各指标均最优,表明NL模块增强全局空间特征提取的能力。

表1 对比模型参数设置

Table1 Parameter settings of the compared models	
模型	参数值
GRU	隐藏层维度为50
TCN	层数为2,卷积核个数及大小分别为32、3
GCN	层数为2,隐藏层维度为64
GAT	层数为2,注意力头数为8,隐藏层维度为64
GraphSAGE	层数为2,每一层采样的邻居数分别为25、10,聚合器为平均聚合
STGCN	层数为2,时序卷积核及图卷积核大小均为3,卷积层数为2

表2 亚利桑那州数据集预测结果比较

Table 2 Comparison of prediction results for the Arizona dataset									
模型	1步			4步			8步		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
GRU	0.736 6	1.819 8	0.974 6	2.009 2	3.590 7	0.900 9	3.378 6	5.385 8	0.777 2
TCN	0.823 2	1.879 5	0.972 9	1.958 5	3.558 9	0.902 7	3.312 8	5.298 0	0.784 4
GCN	2.230 1	3.866 0	0.885 3	2.798 3	4.365 8	0.853 7	3.300 4	4.900 2	0.815 8
GAT	1.959 7	3.947 4	0.880 4	2.647 9	4.341 2	0.855 3	3.543 2	5.237 8	0.789 4
GraphSAGE	1.074 6	1.982 1	0.969 8	1.943 8	3.195 4	0.921 6	2.738 4	4.217 4	0.863 5
STGCN	1.527 2	2.513 5	0.964 5	2.409 3	3.746 1	0.907 9	2.913 2	4.893 5	0.808 9
MTGNN	0.773 6	1.719 3	0.977 3	1.534 2	3.037 9	0.929 1	2.324 1	4.190 8	0.865 1
MTGNN-BiMamba2	0.792 3	1.742 9	0.977 5	1.392 1	2.822 8	0.934 6	1.991 7	3.893 2	0.883 6
本文方法	0.720 2	1.709 9	0.978 7	1.378 6	2.918 9	0.938 8	1.945 7	3.769 6	0.890 9

表3 特拉华州数据集预测结果比较

Table 3 Comparison of prediction results for the Delaware dataset

模型	1 步			4 步			8 步		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
GRU	0.771 1	1.707 9	0.956 0	1.911 4	3.125 5	0.852 8	3.126 8	4.479 5	0.697 6
TCN	0.836 4	1.755 6	0.953 6	1.914 1	3.133 7	0.852 0	3.169 4	4.519 8	0.692 2
GCN	1.932 6	3.046 5	0.860 4	2.364 9	3.470 4	0.818 6	2.849 2	4.026 1	0.755 9
GAT	1.690 1	2.712 0	0.889 4	2.120 1	3.117 3	0.853 7	2.679 9	3.816 4	0.781 0
GraphSAGE	0.928 6	1.737 9	0.954 6	1.758 8	2.716 3	0.889 0	2.697 9	3.749 4	0.788 6
STGCN	0.889 9	1.666 0	0.953 5	1.729 1	2.806 9	0.868 1	2.874 0	4.084 0	0.720 6
MTGNN	0.791 0	1.639 3	0.959 5	1.557 4	2.683 9	0.891 5	2.215 5	3.504 1	0.815 0
MTGNN-BiMamba2	0.770 7	1.621 1	0.960 4	1.518 8	2.635 7	0.895 3	2.166 4	3.453 3	0.820 3
本文方法	0.748 1	1.604 9	0.961 2	1.388 7	2.540 6	0.902 7	2.105 6	3.345 4	0.831 4

亚利桑那州各站点单步预测的评价指标分布如图7所示。可以看出:MAE与RMSE整体处于较低水平,且大部分站点误差集中于较小范围,表明预测方法稳定性强。同时 R^2 普遍高于0.95,中位数接近0.98,说明模型在各站点均能有效拟合光伏功率变化趋势,具有良好泛化能力。

特拉华州16个站点单步预测的预测性能如图8所示。整体上,多数站点MAE与RMSE较低,表明本文方法在各站点均具有良好的误差控制能

力。个别站点(如第7号)误差略高,反映局部预测存在波动,但未影响整体稳定性。此外,大多数站点 R^2 高于0.95,部分接近0.99,说明模型能准确捕捉光伏功率变化趋势。综上,本文方法在特拉华州各站点均表现出较高的预测精度与良好的鲁棒性。

图9展示了两个数据集下所有站点总功率单步预测的观测值与预测值对比结果,直观呈现所提模型在测试集上的预测性能。结果显示,亚利桑那州的数据点集中分布于参考线 $y=x$ 附近,拟合直线斜率接近1,说明所提方法在总功率预测方面具有优异的精度和稳定性。相对而言,特拉华州的数据点偏离参考线程度较大,拟合斜率明显偏低,即预测不确定性较大。可能原因是该地区位于美国东北部沿海,其地形平坦、易受局地气象扰动、极端天气及微气候效应影响,导致光伏功率波动加剧,预测难度增大。

4 结论

针对多站点光伏功率预测问题,提出一种改进的时空图神经网络模型:MTGNN-BiMamba2-NL。此模型基于MTGNN框架,通过BiMamba2模块增强长时序依赖捕捉能力。同时,在图卷积模块后引入非局部注意力机制以增强空间特征的全局聚合能力,从而更全面地捕捉光伏功率的时空动态特征。实验结果表明:较MTGNN,所提模型在亚利桑那州数据集的8步预测中,其MAE降低约16.3%;在特拉华州数据集的4步预测中,RMSE降低约5.3%。模型不仅在多步预测精度上显著优于其他方法,还在多站点场景下展现了较强的稳定性与适用性。研究结果为大规模光伏电站群功率的短期预测提供了一种有效思路,也为后续电力系统调度与可再生能源消纳提供了参考。未来的研究可进一步结合气象预测信息,引入不确定性建模方法,以提升模型在极端气候条件下的鲁棒性与应用价值。

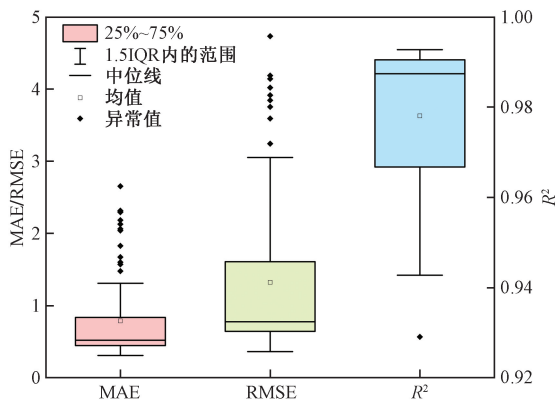


图7 亚利桑那州评价指标箱线图

Fig. 7 Box plot of evaluation metrics in Arizona

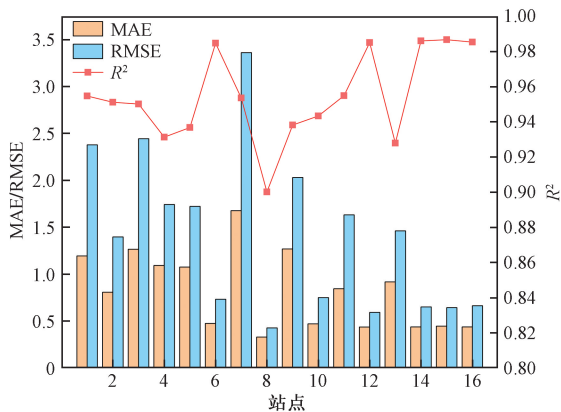


图8 特拉华州各站点评价指标

Fig. 8 Evaluation metrics for stations in Delaware

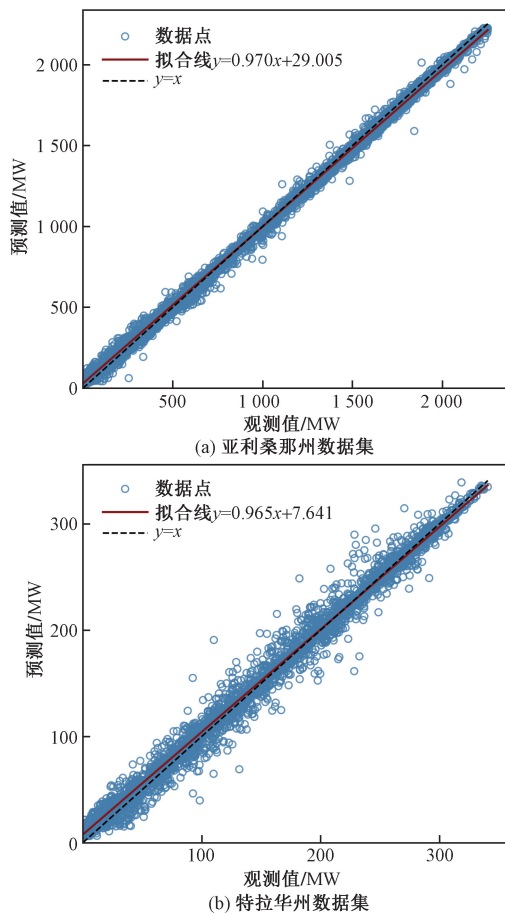


图9 总功率观测值与预测值

Fig.9 Observed and predicted total power

参 考 文 献

- [1] Wang J Z, Yu Y, Zeng B, et al. Hybrid ultra-short-term PV power forecasting system for deterministic forecasting and uncertainty analysis[J]. Energy, 2024, 288: 129898.
- [2] Yang J, Niu M B. WaveGRU: a framework with frequency-domain spatial attention for accurate solar PV and wind power forecasting [J]. Sustainable Energy Technologies and Assessments, 2025, 83: 104572.
- [3] 梁亚峰, 马立红, 邱剑洪, 等. 基于 CEEMD-LSTM 光伏短期功率预测[J]. 科学技术与工程, 2024, 24(13): 5396-5405.
Liang Yafeng, Ma Lihong, Qiu Jianhong, et al. Short-term photovoltaic power forecasting based on CEEMD-LSTM [J]. Science Technology and Engineering, 2024, 24(13): 5396-5405.
- [4] 柳天虹, 齐胜利, 裔扬, 等. 基于分位数回归的改进权重 GRU 风电功率区间预测[J]. 太阳能学报, 2024, 45(12): 291-298.
- [5] Liu Tianhong, Qi Shengli, Yi Yang, et al. Interval forecasting of wind power based on an improved weighted GRU with quantile regression [J]. Acta Energiæ Solaris Sinica, 2024, 45 (12): 291-298.
- [6] 刘俊宏, 富斯源, 王亚君. 基于优化 TCN 组合模型的短期光伏功率预测[J]. 科学技术与工程, 2025, 25(15): 6378-6388.
Liu Junhong, Fu Siyuan, Wang Yajun. Short-term photovoltaic power forecasting based on an optimized TCN ensemble model[J]. Science Technology and Engineering, 2025, 25(15): 6378-6388.
- [7] Wang S, Sun Y H, Zhang W J, et al. Very short-term wind power forecasting considering static data: an improved Transformer model [J]. Energy, 2024, 312: 133577.
- [8] Zhou N, Sun L F. A dual-layer state space network for multivariate time series forecasting with dimensional interdependency [J]. Applied Soft Computing, 2025, 183: 113452.
- [9] Alkilane K, He Y H, Lee D H. Expert denoising for enhanced diffusion-based time series forecasting[J]. Expert Systems with Applications, 2026, 299: 130269.
- [10] 曾程, 肖茂然, 夏元铁, 等. 基于分位数时空图神经网络的分布式光伏聚合不确定性表征方法[J]. 科学技术与工程, 2024, 24(34): 14726-14733.
Zeng Zeng, Xiao Maoran, Xia Yuanyi, et al. A distributed photovoltaic aggregation uncertainty characterization method based on quantile spatiotemporal graph neural network[J]. Science Technology and Engineering, 2024, 24(34): 14726-14733.
- [11] Liu J L, Ong G P, Chen X Q, et al. GraphSAGE-based traffic speed forecasting for segment network with sparse data[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(3): 1755-1766.
- [12] Liu J C, Li T. Multi-step power forecasting for regional photovoltaic plants based on ITDE-GAT model [J]. Energy, 2024, 293: 130468.
- [13] Cao J, Liu C Q, Chen C L, et al. A short-term load forecasting method for integrated community energy system based on STGCN [J]. Electric Power Systems Research, 2024, 232: 110265.
- [14] Guo S N, Lin Y F, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 922-929.
- [15] Liu R, Huang C Z. Data-driven modeling approach of heavy-duty gas turbine with physical constraint by MTGNN and Transformer [J]. Control Engineering Practice, 2024, 151: 106014.
- [16] Jiang T L, Liu W Y, Lin J Q, et al. Research on an hourly heat load forecasting model for district heating systems based on heterogeneous model fusion[J]. Case Studies in Thermal Engineering, 2025, 73: 106500.
- [17] Go J, Ryu J B. Spatial bias for attention-free non-local neural networks[J]. Expert Systems with Applications, 2024, 238: 122053.