

# 基于 BIRCH 聚类的 L-Transformer 分布式光伏短期发电功率预测

董俊<sup>1,2</sup>, 刘瑞<sup>1,2</sup>, 束洪春<sup>1,2</sup>, 罗琨<sup>1,2</sup>, 刘壮<sup>1,2</sup>

(1. 昆明理工大学电力工程学院, 昆明 650504;

2. 云南省绿色能源与数字电力量测及控保重点实验室, 昆明 650504)

**摘要:** 精准的分布式光伏短期发电功率预测有助于电力系统运行与功率就地平衡。该文提出一种基于 BIRCH(balanced iterative reducing and clustering using hierarchies)相似日聚类的 L-Transformer(LSTM-Transformer)模型进行短期光伏功率预测。首先使用 BIRCH 无监督聚类算法对历史数据聚类得到 3 种典型天气, 根据聚类结果划分测试集对模型进行训练。为提高不同天气类型下的预测精度, 采用双层架构的 L-Transformer 模型, 首层通过长短期记忆网络(long short term memory, LSTM)的门控单元机制捕捉时间序列中的长期依赖关系; 次层结合 Transformer 模型的自注意力机制聚焦于当前任务更关键的特征量, 通过多注意力头与光伏数据特征量相结合生成向量, 注意力头并行计算, 从而高效、精确地预测短期光伏功率。实测数据验证结果表明 L-Transformer 模型对于不同天气类型功率预测泛化性优异、精确度高, 气象数据波动大时鲁棒性强。

**关键词:** 深度学习; 自注意力机制; 多头注意力; BIRCH 聚类; 短期光伏功率预测; 特征融合

## Short-term Distributed Photovoltaic Power Generation Prediction Based on BIRCH Clustering and L-Transformer

DONG Jun<sup>1,2</sup>, LIU Rui<sup>1,2</sup>, SHU Hongchun<sup>1,2</sup>, LUO Kun<sup>1,2</sup>, LIU Zhuang<sup>1,2</sup>

(1. School of Electric Power Engineering, Kunming University of Science and Technology, Kunming 650504, China;

2. Yunnan Key Laboratory of Green Energy and Digital Power Measurement Control and Protection, Kunming 650504, China)

**Abstract:** Accurate short-term power prediction of distributed photovoltaic systems is helpful to the operation of power systems and local power balancing. This paper proposes an LSTM-Transformer (L-Transformer) model based on BIRCH (balanced iterative reducing and clustering using hierarchies) for short-term photovoltaic power forecasting. First, the BIRCH unsupervised clustering algorithm is used to cluster historical data into three typical weather patterns, and the test set is divided according to the clustering results to train the model. To improve the prediction accuracy under different weather types, a two-layer L-Transformer model is adopted. The first layer captures long-term dependencies in the time series by using the gating unit mechanism of long short term memory (LSTM). The second layer combines the self-attention mechanism of the transformer model, focusing on the more critical features of the current task. By integrating multiple attention heads with photovoltaic data feature quantities to generate vectors, the parallel computation of attention heads will efficiently and accurately predict short-term photovoltaic power. The results, combined with actual measurement data, show that the L-Transformer model has excellent generalization and high accuracy for power prediction under different weather types, and the robustness is strong when meteorological data fluctuations are remarkable.

**Key words:** deep learning; self attention mechanism; multi-head attention; BIRCH clustering; short-term photovoltaic power prediction; feature fusion

## 0 引言

随着“光伏整县推进”, 电力系统用户侧大量

光伏接入配电网, 不同于传统能源发电技术, 光伏发电出力受实发辐照度、温度、风速和湿度等气象因素影响, 间歇性、随机性和波动性大<sup>[1]</sup>, 在并网过程中对电网电能质量、供电可靠性产生影响。通过准确地预测光伏功率, 电力系统管理人员可以调整光伏电站发电计划, 或提前制定光伏就地消纳计

基金资助项目: 云南省重大科技专项计划(202002AF080001)。

Project supported by the Major Special Science and Technology Project of Yunnan Province(202002AF080001).

划, 从而实现电网的平衡调度。

目前常见的光伏功率预测主要有: 基于时间序列预测、人工神经网络预测和机器学习预测。其中基于时间序列预测比较常见的模型有自回归移动平均模型(autoregressive integrated moving average, ARIMA)、季节性分解模型和指数平滑模型等<sup>[2-4]</sup>; 常见的人工神经网络预测有循环神经网络(recurrent neural network, RNN)和多层感知机(multilayer perceptron, MLP)等<sup>[5-7]</sup>; 常见的机器学习预测有决策树、随机森林和支持向量机等<sup>[8-9]</sup>。不同的预测方法适用于不同的场景和数据集。文献[10]提出一种结合模糊均值聚类和长短期记忆网络(long short-term memory, LSTM)的光伏功率超短期预测方法, 定义了3种特征指标提取波动规律, 采用模糊C均值(fuzzy C-means, FCM)方法进行时序片段聚类, 在聚类后重构数据集, 结合LSTM模型进行光伏功率预测。文献[11]提出了一种日前组合预测模型, 通过聚类法将该光伏系统所在地的天气类型划分为不同的天气模型, 根据预测日期的天气状况选择适合的天气模型进行预测, 并引入了狼群优化算法预测光伏有功出力的波动状态, 以减少气象数据变化较大时带来的误差。文献[12]详细阐述了影响光伏功率预测的难点, 对光伏功率超短期和短期预测的主要技术要点进行了详细评述。基于神经网络和遗传算法等人工智能技术预测光伏发电功率的方法, 可处理复杂的光伏相关数据, 具有更高的精准性, 探索了光伏发电影响因素内部耦合关系。其中最常用的模型为LSTM模型, 在光伏功率预测中LSTM可以良好地处理数据中非线性关系, 并捕捉数据之间长期依赖性。

基于LSTM模型, 一些文献融入其他神经网络形成混合预测模型, 通过不同神经网络之间的优势互补, 实现更精确的多维度数据预测<sup>[13-14]</sup>。文献[15]提出了一种结合卷积神经网络(convolutional neural network, CNN)与LSTM网络的混合模型, 通过CNN提取输入数据中的特征向量, 与时间序列相结合后作为LSTM网络的输入, 从而实现更精确的预测。文献[16]提出了一种组合预测模型, 将LSTM、门控循环单元、RNN 3个单一模型分别预测后, 将3者结果进行加权组合, 再用Q学习算法对组合权重进行优化。文献[17]提出基于相似日理论和纵横交叉算法与改进的高斯过程回归算法相结合的短期光伏发电功率预测方法。在短期光伏发电功率预测模

型中, 机器学习与神经网络相结合, 提高了模型处理数据能力和预测性能<sup>[18-21]</sup>。

目前, 短期光伏功率预测技术已取得一定成果, 但仍在以下问题亟待解决: (1)传统模型对噪声和异常数据较为敏感, 在长时间序列中易对时序噪声过拟合, 鲁棒性需要提高; (2)传统模型对高维度的大量历史光伏数据泛化性不足, 需要进行大量的特征工程和参数调整才能实现稳定的泛化能力。

传统的单一模型很难有效解决以上问题, 针对多变量时序预测场景的光伏功率预测技术, 本文提出一种基于BIRCH聚类的L-Transformer模型, 通过BIRCH算法, 在无监督聚类下将天气场景通过气象因素自发聚类为3种类型, 针对不同的天气场景, L-Transformer模型自动选择对应的气象数据作为训练集, 学习不同场景下特有的时序特征, L-Transformer模型第1层通过LSTM的门控机制感知序列中与光伏功率有关的耦合特征, 得到全面的依赖关系后传入到Transformer层, 通过自注意力机制聚焦于更关键的任务信息, 在大量数据上进行自监督训练, 抵抗噪声和异常数据的影响, 突出多变量的长时间序列中重要因素, 多头注意力层在面对气象数据波动剧烈时仍保持良好的时序特征捕捉能力和鲁棒性, 从而提高预测精度与模型泛用性, 适用于多种天气类型下光伏功率预测, 精度更高, 泛用性强。

## 1 光伏发电功率多变量相关性分析

光伏发电系统的功率输出受多种因素影响, 其中一些因素可能相互关联, 影响相互作用复杂。为了更好地理解这些影响因素之间的相关性, 本文采用Kendall<sup>[21]</sup>、Spearman和Pearson<sup>[22]</sup>分别进行相关性分析, 通过3种相关系数对比得出影响光伏发电功率的多变量相关性, 计算式为:

$$P_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

$$S_{xy} = 1 - \frac{6 \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \quad (2)$$

$$K_r = \frac{2(P-B)}{T(T-1)} \quad (3)$$

式中:  $P_{xy}$  为 Pearson 相关系数;  $\bar{x}$ 、 $\bar{y}$  分别为信号

中的均值;  $S_{xy}$  为 Spearman 相关系数;  $R(x_i)$ 、 $R(y_i)$  分别为  $x_i$ 、 $y_i$  信号值在序列中的顺序;  $n$  为信号  $x$  和  $y$  的维数;  $i$  为信号的序号;  $K_r$  为 Kendall 秩相关函数;  $P$  为 Kendall 中一致对外数;  $B$  为 Kendall 中非一致对外数;  $T$  为函数中观察值数。

考虑的天气相关因素包括辐照度、风速、风向、温度、压强以及湿度。采用热力图绘制 3 种相关系数下各个影响因素对光伏发电功率的相关性系数。3 种相关系数的热力图在附录 A 中进行了详细阐述, 光伏发电功率与各环境变量的 3 种相关系数如表 1 所示。

通过对比发现, 光伏发电功率与辐照度、温度和风向呈正相关, 与压强、风速和湿度呈负相关。分析相关系数绝对值可知, 光伏发电功率与辐照度相关性最大, 因此在后文预测光伏功率中, 剔除风向与压强数据, 从而缩减输入量的维度与数据规模, 提高 L-Transformer 模型运算速度。

## 2 基于 BIRCH 算法的相似日聚类

在天气相似日聚类中, 天气数据受多变量气象因素影响, 数据量大, 数据维度高。传统的聚类方式具有以下缺点: (1)需要人为划分聚类簇数, 从而引入了主观聚类误差; (2)对大数据聚类收敛慢。因此采用 BIRCH 算法(balanced iterative reducing and clustering using hierarchies, BIRCH)对天气进行分类。

BIRCH 聚类基于无监督层次聚类, 在不需要预先指定簇数的情况下对大规模数据集进行聚类, 并高效处理高维数据。通过附录 B 发现, 相比其他聚类方式, BIRCH 聚类在没有已知标签或分类信息的情况下对大数据集进行分组, 无监督聚类揭示数据中存在的潜在模式和结构, 更好地了解数据的分布特性, 有助于 L-Transformer 模型在训练过程中, 从自发的分类结果中捕捉数据内部耦合关系。同时, BIRCH 聚类通过构造聚类特征树, 可实现快速、并行处理大规模数据集, 大大提高运算速度和收敛速度。

### 2.1 模型介绍

天气相似日聚类需要处理大量高维度数据, 本文选用云南省某地区环境监测仪获得的气象数据与装机容量 10 kW 的分布式光伏测点数据进行聚类, 数据样本 39 381 条, 涉及辐照度、风速、风向、温度、压强和湿度等 6 个维度。因此, 本文采用 BIRCH

表 1 发电功率与各环境变量相关系数

Table 1 Correlation coefficient between generating power and environmental variables

气象因素	Pearson 相关系数	Spearman 相关系数	Kendall 相关系数
辐照度	0.90	0.90	0.76
温度	0.23	0.24	0.16
风向	0.02	0.07	0.05
压强	-0.03	-0.06	-0.04
风速	-0.17	-0.23	-0.15
湿度	-0.42	-0.39	-0.27

算法对天气的全特征进行无监督聚类, 以确定最佳聚类数。BIRCH 无监督层次聚类算法通过聚类特征描述数据中关键信息, 通过聚类特征树存储数据, 对树结构遍历聚类数据, 不需要对整个数据集进行扫描, 减少计算复杂度。

### 2.2 天气相似日场景聚类

BIRCH 算法采用树模型确定好最佳聚类数据值  $k$ , 采用均值模型做最后聚类。首先通过轮廓系数  $S_i$  与方差比准则指标(Calinski-Harabasz index, CH)综合判别, 选出最佳分层数, 二者指标数值越大, 聚类的数目就越优, 研判详细过程见附录 B。可以看出, 当  $k=3$  时, 轮廓系数和 CH 值综合水平达到最佳, 聚类层数为 3, 均值层将天气场景聚类为晴天、阴天和雨天, 聚类详细过程见附录 B。全年数据 39 381 条, 经过聚类后, 晴天聚类数 21 767 条, 阴天聚类数 12 615 条, 雨天聚类数 5000 条。聚类结果如图 1 所示。

通过图 1 可以看出, 在晴天场景下, 光伏出力图像较为平滑, 幅值在 10 kW 上下波动。此场景下由于数据周期性、趋势性和季节性强, 预测精度相比其他场景最高; 阴天场景下, 由于一天中阴晴变化较多, 光伏出力波动幅度大, 最高值仍在 10 kW 上下波动, 趋势性明显减弱, 数据特征不明显。此场景下传统的预测模型难以捕捉气象与光伏处理之间的耦合关系, 预测难度大; 雨天场景下, 由于辐照度与温度下降, 光伏出力变化大, 最高值在 7 kW 左右, 趋势性最为不明显, 且雨天的气象数据本身规律性不强, 难以寻找与光伏功率之间长期依赖关系, 预测难度达到最大。

3 种聚类结果中, 随着季节变更, 光照时段与光照时长均有所变化。在本文提出的 L-Transformer

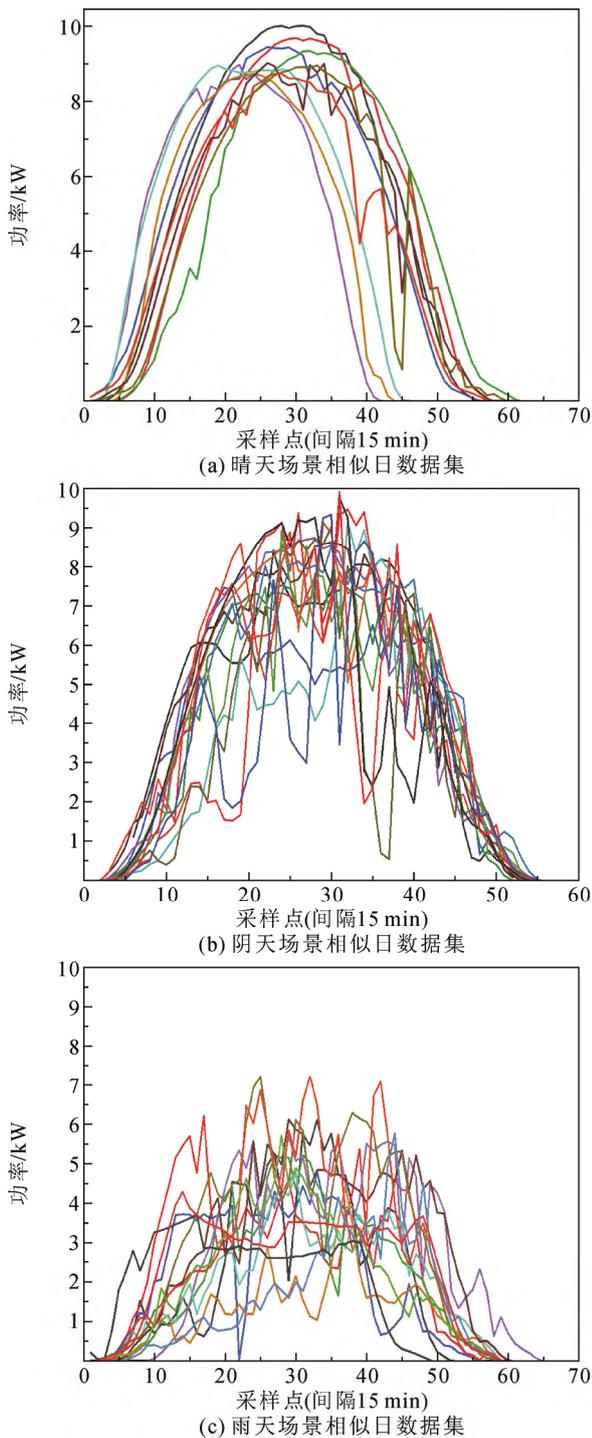


图1 天气聚类数据图

Fig.1 Clustering data graph of weather

模型中,通过自注意力机制自动捕捉随光照时长变化引起的光伏功率对应变化。注意力头感知历史时间序列中辐照度与光伏功率之间的长期依赖关系,学习历史数据中光照与光伏功率之间的耦合模式后根据将要预测的辐照度进行光伏功率预测,从而高效地抵抗季节变化和天气变化带来的光照时长不同的干扰。

### 3 基于 L-Transformer 光伏功率预测模型

不同的人工神经网络进行时间序列预测时,模型内部网络结构层、训练算法和超参数设置存在差异,这些不同的结构会影响网络模型的拟合能力、泛化能力、计算效率、误差大小,甚至对于相同的数据会得到不同的预测结果。在选择预测模型时需要考虑不同场景、数据集特点和具体需求。

#### 3.1 数据预处理

历史光伏数据和气象数据存在潜在特征,若直接将数据代入模型,则难以挖掘出其中的特征关系,导致模型难以收敛,进而减弱预测效果。因此需要将数据进行预处理,本文采用 Datapre(data pre-processing)对原始历史光伏和气象数据进行预处理,以使数据更适合用于训练和测试模型。

Datapre 作为数据预处理的库,对输入多维长时序数据提供数据清洗、特征提取、数据归一化等功能,以便于模型的训练和推理。光伏电站在实际运行过程中,由于设备故障、操作误差和环境影响等因素,收集到的数据可能存在异常值、缺失值、重复值等问题,对模型计算产生影响,导致计算结果不收敛,因此通过 Datapre 对数据进行清洗,以天为单位删除缺失值、填充异常值和剔除重复值。

进行数据清洗后,为消除不同特征之间的量纲差异,避免特征值之间因量纲不同而产生的误差,使用 Datapre 对特征进行最小、最大归一化处理。将数据缩放到指定的区间范围内,可以保留原始数据的分布信息,同时使数据更易于处理和理解。数据转换到相同的尺度上,使得不同尺度的特征量在模型训练中能够被平等对待。

通过 Datapre 对多维长时序输入数据进行预处理,从而提高 L-Transformer 模型在训练集、测试集和验证集运行时的效率和准确性。

#### 3.2 LSTM 神经网络

长短期预测模型(LSTM)在循环神经网络(RNN)的基础上,引入了细胞状态的3种门(输入门、遗忘门、输出门)来保持和控制信息的流动,并学习数据之间的长期依赖性,解决了 RNN 在训练数据时出现梯度消失和梯度爆炸问题。RNN 对数据进行深层训练计算过程中,当神经网络的后层的梯度值逐渐变小直至接近于 0 时,产生梯度消失现象,网络训练变得非常缓慢,导致模型无法学习到时间序列中的深层特征,不能充分利用历史数据中的信息,使

得预测误差变大。预测模型的后层梯度值逐渐变大, 会产生梯度爆炸现象, 导致网络训练不稳定, 模型过拟合, 计算结果难以收敛, 对历史数据过于敏感, 导致预测结果波动较大。LSTM 根据输入序列中的信息, 自适应地更新细胞状态中的内容, 将每个时间步骤的输入与先前的记忆状态组合, 生成一个新的记忆状态, 保证了模型的记忆能力和推理能力。

LSTM 的循环神经网络结构如图 2 所示, LSTM 循环单元包括 2 个状态, 即单元状态  $C_t$ (cell state) 和隐状态  $h_t$ (hidden state);  $\sigma$  表示 S 激活函数 (sigmoid)<sup>[23]</sup>, 作用是将门控的输出值映射到 0~1 之间, 从而使得门控可以控制信息在神经网络中的流动。tanh(hyperbolic tangent function)表示 LSTM 中的双曲正切函数的变化;  $x_r$  表示第  $r$  个序列的输入信号; LSTM 详细运算原理见附录 C。

### 3.3 Transformer 神经网络

Transformer 是一种基于注意力机制的序列到序列(seq2seq)模型, 由编码器(Encoder)和解码器(Decoder)组成, Encoder 将输入的历史光伏数据、气象数据编码为一个高维的特征向量, 以便 Decoder 在输出生成序列时能够根据序列的上下依赖关系做出更准确的预测。Encoder 由多个相同的层组成, 每层包含多头自注意力机制(multi-head self-attention)和前馈神经网络(feedforward neural network) 2 个子层, 同时还包含残差连接(residual connection)和层归一化(layer normalization)等技术<sup>[24]</sup>。

Transformer 模型中的自注意力机制对输入历史光伏数据序列中的每个位置进行加权聚合, 对影响光伏功率的因素赋予更高的权重, 使得模型能够关注输入光伏出力序列中的任意位置, 而不是只关注输入时间序列的固定部分。如图 3 所示, Transformer 模型中的多头注意力机制让模型同时关注输入序列矩阵  $X$  中的不同部分, 拥有对时间序列的全局感知能力, 从而提高模型的表达能力, 更好地处理光伏功率数据序列中的局部和全局信息。

注意力机制由查询矩阵  $Q$ 、键矩阵  $K$  和价值矩阵  $V$  组成, 用于计算输入数据的加权和。图 3 中,  $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$  分别表示第  $i$  个注意力头中  $Q$ 、 $K$ 、 $V$  所对应的权重矩阵;  $Q_i$ 、 $K_i$ 、 $V_i$  分别表示第  $i$  个注意力头中的  $Q$ 、 $K$ 、 $V$ ;  $Z$  表示自注意力机制的输出矩阵;  $N$  表示注意力头的序号;  $Z_N$  表示第  $N$  个注意力头的输出矩阵。Transformer 详细运算原理见附录 D。

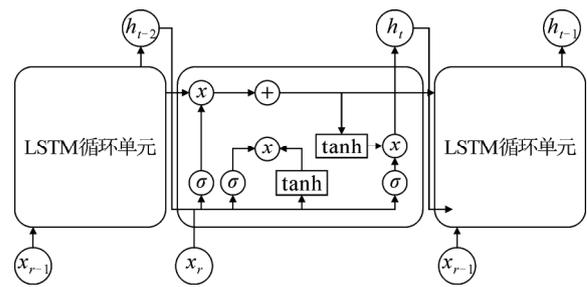


图 2 LSTM 循环神经网络示意图

Fig.2 Schematic diagram of LSTM cyclic neural network

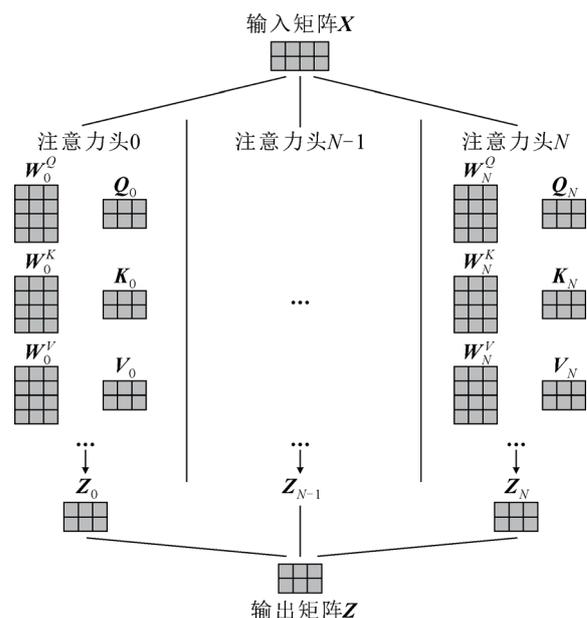


图 3 多头注意力结构图

Fig.3 Structure chart of multi-head attention

### 3.4 L-Transformer 混合神经网络

光伏功率预测是一个多变量时序预测, 光伏的输出功率与温度、直接辐射度、风速、湿度等多个天气因素有关。LSTM 可以可靠地捕捉多变量时序特征, LSTM 在处理数据时对多门进行计算, 这使得计算复杂度增高。在长时间序列集上训练时, LSTM 训练时间变长、容易过拟合、难以并行化, 部分时长的信息会被遗忘或者淹没在时序噪声中, 导致 LSTM 不能良好地处理长时序的信息。而 Transformer 的自注意力机制和前馈网络层可以在一次处理中同时考虑输入序列的所有位置, 从而在计算时不受特征之间距离干扰, Transformer 的并行计算机制可以大幅提高模型训练速度。

结合以上问题, 本文提出一种 L-Transformer 模型, 结合 Transformer 的自注意力机制和 LSTM 的序列建模能力, 使用 LSTM 编码器对输入序列进

行编码,使用 Transformer 解码器生成输出序列。在这种结构下,LSTM 解码器可以通过逐步生成输出序列来保留时序特征,而 Transformer 编码器可以有效地捕捉多变量时间序列中不同位置之间的关系。

Transformer 结构图如图 4 所示,在 Transformer 的多头自适应(multi-head attention, MHA)前端插入 LSTM 模块,其中  $X_i^r$  为 LSTM 模块处理后的时间序列组;  $h_i^r$  为记忆模块输出信号;  $X^r$  为经过数据清洗和模态分解后的各环境因素量与历史光伏功率等多变量时序数据,输入的多维度历史时间序列首先被单层 LSTM 处理,通过 LSTM 的记忆单元和遗忘门等机制,使得 L-Transformer 更好地捕捉时间序列数据中的长期依赖关系和非线性模式。LSTM 层通过门机制,控制细胞状态中多维气象因素与光伏功率的舍弃量、细胞状态添加量和细胞状态输出量,从而在大量的多变量数据中提取到有价值的时间特征量作为隐状态向量。此时输出隐状态向量与原输入经过融合层结合,将 LSTM 的多个层和方向中学习到不同特征进行融合,得到更全面、更准确的表示,从而作为输入传入多头注意力层之中,计算 MHA 时将上个模块的输入与本次的输入进行共同计算。通过 Transformer 的自注意力机制,聚焦于当前预测任务更关键的特征量,降低对其他环境因素特征的关注度,学习先前时间步的状态,并结合当前观察到的天气等因素来预测未来的光伏发电量。

计算过程为:

$$X_i^r, h_i^r = L(X^r, g(h_i^{r-1})) \tag{4}$$

$$\tilde{X}_i^r = F(X_i^r, X^r) \tag{5}$$

$$G^r = T(X_i^r) \tag{6}$$

式中: $L$ 表示每次从 LSTM 模块中输出的  $X_i^r$  和  $h_i^r$  不做梯度更新后输入 MHA 层,不断将前 LSTM 模块中隐藏状态引入当前的输入中; $g$ 表示隐藏状态的解码函数; $\tilde{X}_i^r$ 表示融合层输出量; $F$ 表示将融合层中可选的数值  $X_i^r$  和  $X^r$  进行组合操作; $G^r$ 表示 L-Transformer 模型输出的光伏功率预测值; $T$ 表示在 MHA 中添加的向量已经经历了相同的计算过程<sup>[25]</sup>。在训练期间分段递归应用于 MHA 模块,每一次外部输入从最后一个分段的 LSTM 层输出中提取,但不做梯度更新,通过 LSTM 模块中的隐藏状态使得 L-Transformer 模型在训练预测过程中鲁棒性更强。在融合层中,LSTM 层的每个输出向量与相应位置的输入向量级联后,融合的向量传送到激活函数层,使模型可以在不同的输入特征之间分配不同的注意力权重。以高斯误差线性单元作为激活函数,历史光伏功率数据通常具有较大的波动性和非线性特征,因此需要一个具有灵活性和可重构性的激活函数来适应其不同的分布情况,高斯误差线性单元可以动态调整其权重和偏置来适应不同的输入分布和任务需求,使 L-Transformer 模型能够学习和表示

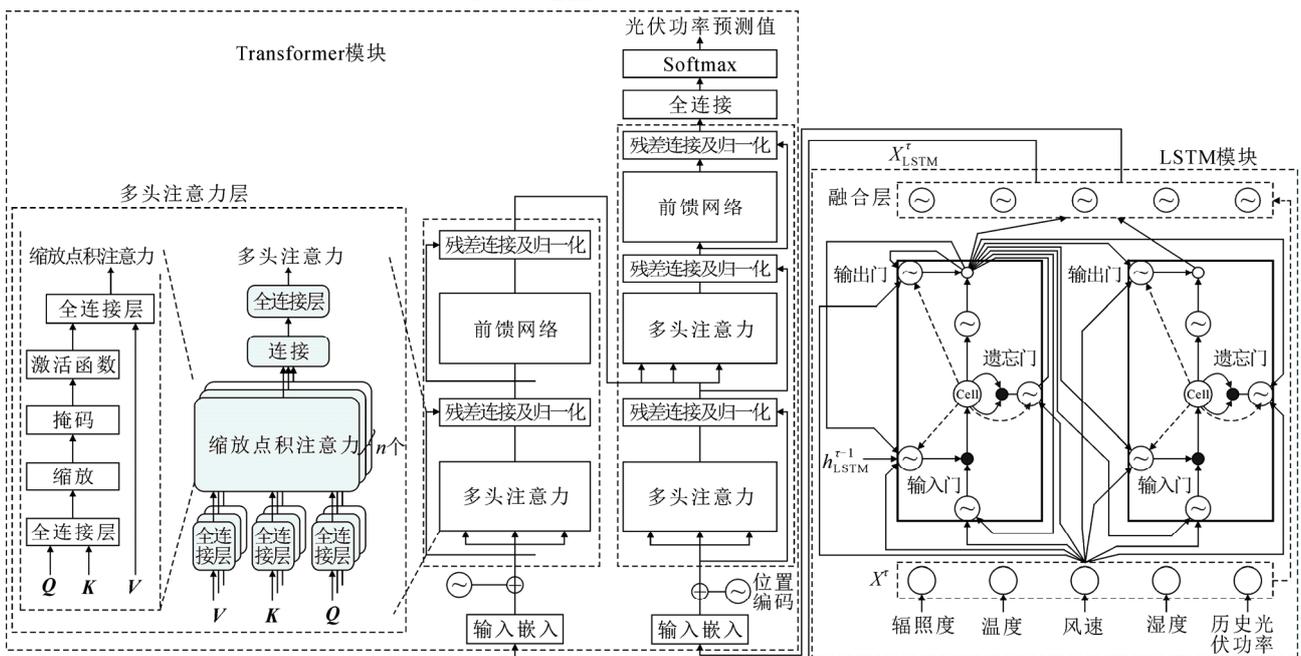


图 4 L-Transformer 结构图

Fig.4 Structure diagram of L-Transformer

光伏功率预测过程中更复杂的函数和特征, 从而提高光伏功率预测的准确性和鲁棒性。由于每个子层都使用残差连接和规范化, 因此子层输出为:

$$O_u = N(j + b(j)) \quad (7)$$

式中:  $O_u$  表示下一子层的输出与上一层规范化输出与子层之间的融合关系;  $j$  表示输出量;  $N$  表示本层输出;  $b(j)$  表示上一层输出。传统 Transformer 模型运行机制通过  $Q$  与  $K$  内积后传入激活函数层进行归一化处理。在进行光伏功率预测时,  $Q$  中大部分标记值不具备预测的鲜明特征, 因此 L-Transformer 模型通过计算均匀分布差异选取活跃度高的  $Q$ , 再与  $K$  进行内积, 使用 Softmax 激活函数进行概率化近似。相比于传统 Transformer 模型, L-Transformer 模型在允许降低计算复杂度的同时, 仍能够捕获输入序列中的上下文信息。

在 L-Transformer 模型中, 第 1 层 LSTM 起到模型编码器的作用, 将时间序列数据编码为向量表示, 而第 2 层 Transformer 则起到解码器的作用。使用 LSTM 来提取序列数据的局部特征, 然后使用 Transformer 来捕捉序列数据的全局关系, 从而更准确地进行光伏预测。

## 4 仿真分析

### 4.1 评价指标

为了评估模型预测的准确性, 本文采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)和决定系数(R-squared,  $R^2$ )等指标评价模型的光伏功率预测性能。具体计算式为:

$$\lambda_{MAE} = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \quad (8)$$

$$\lambda_{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (9)$$

$$\lambda_R = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - d_i)^2} \times 100\% \quad (10)$$

式中:  $\lambda_{MAE}$ 、 $\lambda_{RMSE}$ 、 $\lambda_R$  分别表示 MAE、RMSE、 $R^2$  指数值;  $m$  表示样本数量;  $i$  表示样本序列号;  $y_i$  表示第  $i$  个样本的真实值;  $\hat{y}_i$  表示第  $i$  个样本的预测值;  $d_i$  表示真实值的平均值。其中 MAE 与 RMSE 用以评估预测值与真实值之间的误差, 其值越小则表示预测结果越准确。 $\lambda_R$  是衡量预测值与真实值之间相关性, 取值范围为 0~1, 越接近 1 表示预测结

果与真实值越相关。

### 4.2 光伏功率预测

为探索 L-Transformer 模型对光伏功率的预测能力, 使用该模型对云南省某光伏阵区 2019 年全年历史光伏实测数据以及该地区环境监测仪获得的气象数据进行预测, 气象数据经过 Datapre 进行预处理, 采样间隔为 15 min。本文采用 PyTorch 框架搭建了 L-Transformer 模型、LSTM 模型和 Transformer 模型, 处理器为 Intel(R) Core i7-8700K CPU @ 3.70 GHz, 内存为 8 GB。预测流程图如图 5 所示。

为验证 L-Transformer 模型预测光伏功率的准确性, 文章选用机器学习中的经典模型 SVM 和深度学习中的经典模型 BiLSTM 与本文提出的 L-Transformer 模型进行对比研究。3 种模型参数设置如表 2 所示。

通过 BiLSTM 模型、SVM 模型分别对光伏功率进行预测, 最后将 3 种模型预测结果进行对比。考虑到影响光伏功率的天气因素, 将光伏发电场景聚类为晴天、阴天和雨天, 分别使用 L-Transformer、BiLSTM 模型与 SVM 模型对不同场景进行预测。

模型所用数据采样步长为 15 min, 聚类后晴天场景下, 采用第 1~201 天为训练样本; 阴天场景下第 1~116 天为训练样本; 雨天场景下第 1~45 天为训练样本。3 种场景下测试样本均为最后一天。3 种场景下的预测结果与 3 种不同算法的预测误差指标图如图 6 所示。

不同场景下 L-Transformer 模型、BiLSTM 模型与 SVM 模型的预测误差值对比如表 3 所示。

通过分析表 3 数据可以看出:

1) 晴天场景下数据平滑、幅值固定、波动性小且相似日之间差异性最小, 数据稳定, 周期性、趋势性比雨天和阴天场景更为明显, 3 种模型的预

表 2 3 种模型主要参数

Table 2 Main parameters of the three models	
预测模型	主要参数
BiLSTM	拆分比率为 0.7, 迭代周期为 50, 批大小为 25, 权重衰减为 $1 \times 10^{-3}$ , 学习率为 $1 \times 10^{-2}$
SVM	拆分比率为 0.7, 迭代周期为 50, 正则化参数为 4.0, 径向基函数为 0.8
L-Transformer	拆分比率为 0.7, 迭代周期为 50, 头数为 8, 输入维度为 8, 隐藏层数为 64, 前馈神经网络层数为 5

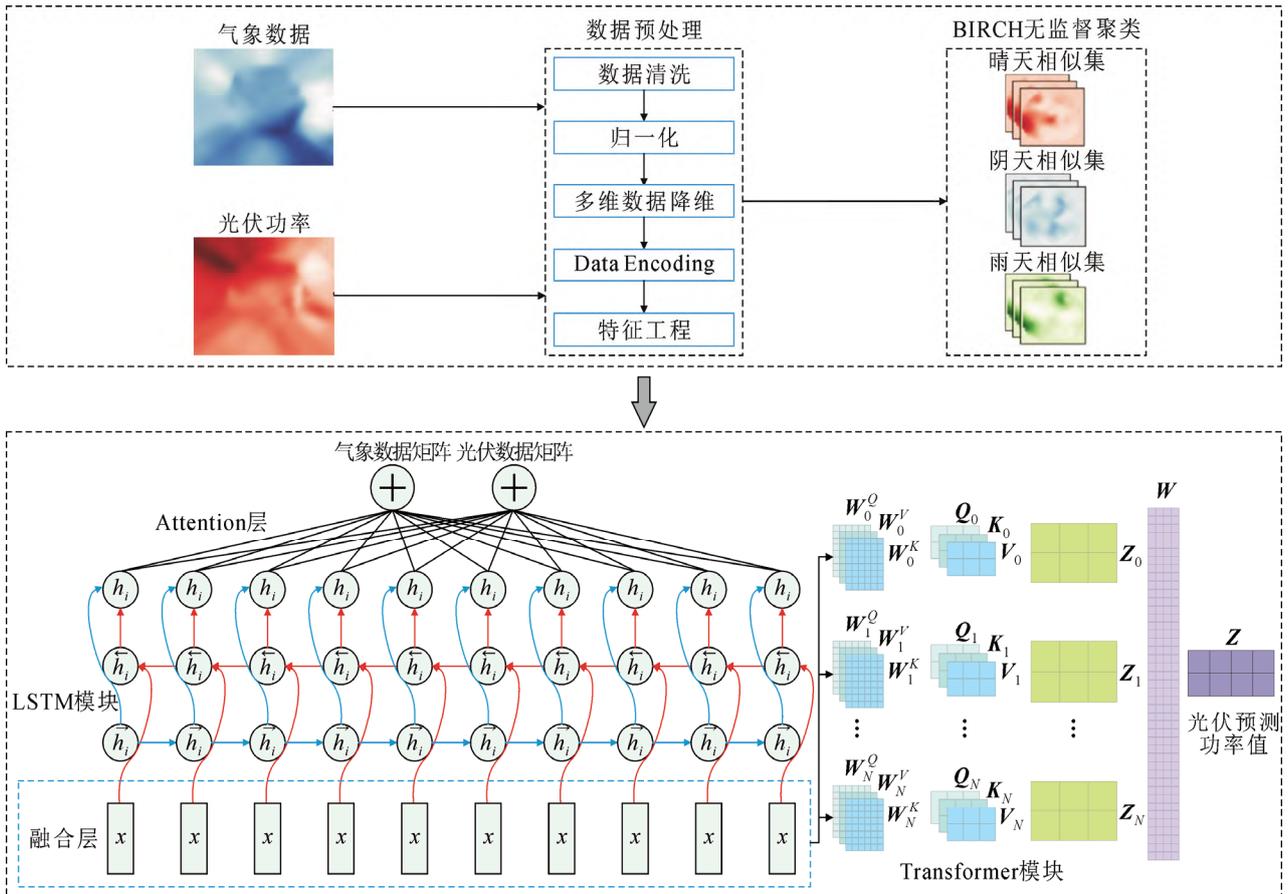


图5 L-Transformer 预测流程图

Fig.5 Forecast flowchart of L-Transformer

测值可以较好地跟踪数据波动变化，从总体指标而言，L-Transformer 预测效果最接近真实值，误差指标均低于传统 SVM 模型和 BiLSTM 模型。

2) 在阴天场景下，样本数据剧烈波动，幅值略低于晴天场景，天气特征不稳定，周期性与趋势性减弱，由于阴天多云，降低了辐照度与光伏功率之间的耦合性，减弱了长时间序列之间的长期依赖关系，导致预测精度有所下降。L-Transformer 模型的误差指数仍然最低，但相比于晴天场景，L-Transformer 的 RMSE 值上升 28.6%，MAE 上升 25.9%，R<sup>2</sup> 降低 2.5%；BiLSTM 模型的 RMSE 值上升 30.8%，MAE 上升 65.9%，R<sup>2</sup> 降低 3.8%；SVM 模型的 RMSE 值上升 23.1%，MAE 上升 21.9%，R<sup>2</sup> 下降 3.5%。从阴天场景与晴天场景下 3 种模型在训练集的变化幅度增大后，误差均有所提高，但 L-Transformer 模型的误差指标参数仍低于 SVM 模型与 BiLSTM 模型，在阴天场景下 L-Transformer 模型相比另外 2 个传统模型表现出较好的鲁棒性和泛用性。

3) 雨天场景与晴天场景相比，数据波动更为剧烈，且幅值降低到仅有 5 kW 左右，各种气象数据与晴天场景相比均发生较大变化，各种气象因素与光伏功率之间的依赖关系减弱，变化特征难以捕捉，通过预测曲线图可以看出，传统的 BiLSTM 模型与 SVM 模型预测虽能捕捉整体变化趋势，但在幅值上相差较大，远远偏离实际值。L-Transformer 的 RMSE 值上升 26.5%，MAE 上升 24.1%，R<sup>2</sup> 下降 1.0%；BiLSTM 模型的 RMSE 值上升 54.6%，MAE 上升 109.5%，R<sup>2</sup> 下降 11.2%；SVM 模型的 RMSE 值上升 98.3%，MAE 上升 103.1%，R<sup>2</sup> 下降 26.2%。相比之下 L-Transformer 的误差增量小于传统预测模型，并保持预测精度最高。在雨天场景下 L-Transformer 模型鲁棒性和泛用性优于传统 BiLSTM 模型与 SVM 模型。

阴天和雨天场景下，气象数据波动大，周期性和趋势性减弱，传统的预测模型难以捕捉此类场景下的时序特征。此时提高 L-Transformer 模型对数据全局观察，其多头注意力结构能更好地捕捉阴天和

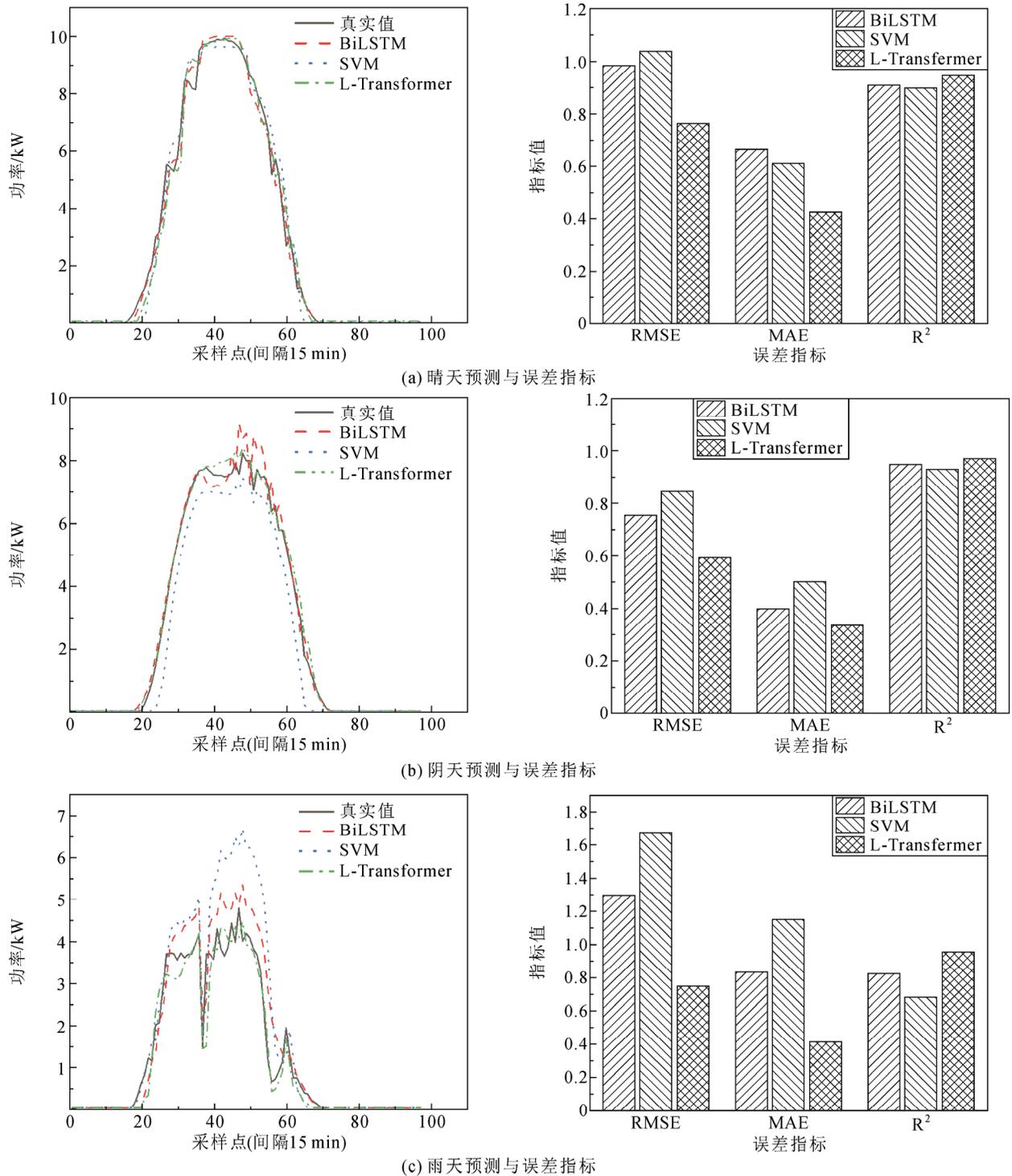


图6 不同天气下 L-Transformer 模型预测情况

Fig.6 Forecast situation of L-Transformer model under different weather conditions

雨天场景下, 数据波动大的时间序列中气象因素与光伏功率之间的耦合关系, 通过自注意力机制, 重点关注数据变化剧烈条件下光伏与气象数据之间的依赖关系。阴雨场景下气象数据与光伏功率之间的耦合关系减弱, L-Transformer 模型的多层自注意力结构捕获多维度长时间序列中不同部分之间微弱的

特征关系, 学习到更丰富的特征表示, 相比于传统模型在阴雨场景下具有更高的预测精度。

BiLSTM 模型受限于递归结构, SVM 模型需人工设计特征, 特征抽取的质量取决于特征工程。二者的泛用性、鲁棒性和预测精度均低于 L-Transformer 模型。

表3 模型预测误差对比

Table 3 Comparison of model prediction errors

场景	指标	BiLSTM	SVM	L-Transformer
晴天场景	RMSE	0.7517	0.8435	0.5929
	MAE	0.3990	0.5010	0.3352
	R <sup>2</sup>	0.9435	0.9288	0.9686
阴天场景	RMSE	0.9832	1.0378	0.7617
	MAE	0.6621	0.6108	0.4221
	R <sup>2</sup>	0.9078	0.8973	0.9447
雨天场景	RMSE	1.2970	1.6727	0.7503
	MAE	0.8350	1.1516	0.4162
	R <sup>2</sup>	0.8254	0.6854	0.9588

## 5 结论

本文针对大规模光伏并网引起的电压波动问题,提出了一种基于 L-Transformer 模型的短期光伏功率预测模型,通过预测短期光伏功率,结合云南省某光伏阵区 2019 年全年历史光伏实测数据以及该地区环境监测仪获得的气象数据进行验证,得到以下结论:

1) 通过 BIRCH 无监督聚类算法将历史数据划分为 3 种相似天气类型,探索天气因素与光伏发电功率之间的特征值,模型基于相似日数据集对光伏发电功率进行预测,在不同天气下仍保持较高预测精度,验证了 L-Transformer 模型的泛用性。

2) 在多维度输入、长时间序列的光伏功率预测上,与传统 BiLSTM 模型、SVM 模型等单层模型相比, L-Transformer 模型能更好地捕捉气候因素与光伏功率之间的耦合关系,在气象数据变化剧烈的情况下,其预测精确度更高,验证了模型的鲁棒性。

3) L-Transformer 模型结合自注意力机制与门控制,专注于长时序、多维度数据中更关键的特征量,对时序噪声的抗干扰能力更强,验证了模型的准确性。

附录见本刊网络版(<http://hve.epri.sgcc.com.cn>)。

## 参考文献 References

- [1] 黎静华, 骆怡辰, 杨舒惠, 等. 可再生能源电力不确定性预测方法综述[J]. 高电压技术, 2021, 47(4): 1144-1155.  
LI Jinghua, LUO Yichen, YANG Shuhui, et al. Review of uncertainty forecasting methods for renewable energy power[J]. High Voltage Engineering, 2021, 47(4): 1144-1155.
- [2] 叶林, 程文丁, 李卓, 等. 光伏集群有功功率分层预测控制策略[J]. 电力系统自动化, 2023, 47(2): 42-52.

- YE Lin, CHENG Wending, LI Zhuo, et al. Hierarchical prediction control strategy of active power for photovoltaic cluster[J]. Automation of Electric Power Systems, 2023, 47(2): 42-52.
- [3] 龚莺飞, 鲁宗相, 乔颖, 等. 光伏功率预测技术[J]. 电力系统自动化, 2016, 40(4): 140-151.  
GONG Yingfei, LU Zongxiang, QIAO Ying, et al. An overview of photovoltaic energy system output forecasting technology[J]. Automation of Electric Power Systems, 2016, 40(4): 140-151.
- [4] 刘大贵, 王维庆, 张慧娥, 等. 基于隐马尔科夫修正的光伏中长期电量预测及调度计划应用[J]. 高电压技术, 2023, 49(2): 840-848.  
LIU Dagui, WANG Weiqing, ZHANG Huie, et al. Mid-long term available quantity of electricity forecasting with error calibration by Hidden Markov Model in photovoltaic and application of dispatching plan[J]. High Voltage Engineering, 2023, 49(2): 840-848.
- [5] BIN SHAMS M, HAJI S, SALMAN A, et al. Time series analysis of Bahrain's first hybrid renewable energy system[J]. Energy, 2016, 103: 1-15.
- [6] 孟安波, 陈嘉铭, 黎湛联, 等. 基于相似日理论和 CSO-WGPR 的短期光伏发电功率预测[J]. 高电压技术, 2021, 47(4): 1176-1184.  
MENG Anbo, CHEN Jiaming, LI Zhanlian, et al. Short-term photovoltaic power generation prediction based on similar day theory and CSO-WGPR[J]. High Voltage Engineering, 2021, 47(4): 1176-1184.
- [7] HUANG C, WANG L, YEUNG R S C, et al. A prediction model-guided Jaya algorithm for the PV system maximum power point tracking[J]. IEEE Transactions on Sustainable Energy, 2018, 9(1): 45-55.
- [8] DE JESÚS D A R, MANDAL P, CHAKRABORTY S, et al. Solar PV power prediction using a new approach based on hybrid deep neural network[C]//2019 IEEE Power & Energy Society General Meeting (PESGM). Atlanta, USA: IEEE, 2019: 1-5.
- [9] ZHANG M, ZHEN Z, LIU N, et al. Optimal graph structure based short-term solar PV power forecasting method considering surrounding spatio-temporal correlations[J]. IEEE Transactions on Industry Applications, 2023, 59(1): 345-357.
- [10] YANG M, HUANG X. Ultra-short-term prediction of photovoltaic power based on periodic extraction of PV energy and LSH algorithm[J]. IEEE Access, 2018, 6: 51200-51205.
- [11] SUCCETTI F, ROSATO A, ARANEO R, et al. Deep neural networks for multivariate prediction of photovoltaic power time series[J]. IEEE Access, 2020, 8: 211490-211505.
- [12] 程启明, 张强, 程尹曼, 等. 基于密度峰值层次聚类的短期光伏功率预测模型[J]. 高电压技术, 2017, 43(4): 1214-1222.  
CHENG Qiming, ZHANG Qiang, CHENG Yinman, et al. Short-term photovoltaic power prediction model based on hierarchical clustering of density peaks algorithm[J]. High Voltage Engineering, 2017, 43(4): 1214-1222.
- [13] 魏斌, 韩肖清, 李雯, 等. 融合多场景分析的交直流混合微电网多时间尺度随机优化调度策略[J]. 高电压技术, 2020, 46(7): 2359-2369.  
WEI Bin, HAN Xiaoqing, LI Wen, et al. Multi-time scale stochastic optimal dispatch for AC/DC hybrid microgrid incorporating multi-scenario analysis[J]. High Voltage Engineering, 2020, 46(7): 2359-2369.
- [14] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(8): 131-137.  
LU Jixiang, ZHANG Qipei, YANG Zhihong, et al. Short-term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Automation of Electric Power Systems, 2019, 43(8): 131-137.

- [15] 孟安波, 许炫淙, 陈嘉铭, 等. 基于强化学习和组合式深度学习模型的超短期光伏功率预测[J]. 电网技术, 2021, 45(12): 4721-4728. MENG Anbo, XU Xuancong, CHEN Jiaming, et al. Ultra short term photovoltaic power prediction based on reinforcement learning and combined deep learning model[J]. Power System Technology, 2021, 45(12): 4721-4728.
- [16] 闫钊汛, 王丽婕, 郭洪武, 等. 基于多特征分析和提取的短期光伏功率预测[J]. 高电压技术, 2022, 48(9): 3734-3743. YAN Yixun, WANG Lijie, GUO Hongwu, et al. Short-term photovoltaic power prediction based on multi-feature analysis and extraction[J]. High Voltage Engineering, 2022, 48(9): 3734-3743.
- [17] 朱琼锋, 李家腾, 乔骥, 等. 人工智能技术在新能源功率预测的应用及展望[J]. 中国电机工程学报, 2023, 43(8): 3027-3047. ZHU Qiongfeng, LI Jiateng, QIAO Ji, et al. Application and prospect of artificial intelligence technology in renewable energy forecasting[J]. Proceedings of the CSEE, 2023, 43(8): 3027-3047.
- [18] 卢俊杰, 蔡涛, 郎建勋, 等. 基于集群划分的光伏电站集群发电功率短期预测方法[J]. 高电压技术, 2022, 48(5): 1943-1951. LU Junjie, CAI Tao, LANG Jianxun, et al. Short-term power output forecasting of clustered photovoltaic solar plants based on cluster partition[J]. High Voltage Engineering, 2022, 48(5): 1943-1951.
- [19] 殷豪, 张铮, 丁伟锋, 等. 基于生成对抗网络和 LSTM-CSO 的少样本光伏功率短期预测[J]. 高电压技术, 2022, 48(11): 4342-4351. YIN Hao, ZHANG Zheng, DING Weifeng, et al. Short-term prediction of small-sample photovoltaic power based on generative adversarial network and LSTM-CSO[J]. High Voltage Engineering, 2022, 48(11): 4342-4351.
- [20] 王育飞, 杨启星, 薛花, 等. 考虑混沌特征的增强型大脑情绪神经网络光伏发电功率超短期预测模型[J]. 高电压技术, 2021, 47(4): 1165-1175. WANG Yufei, YANG Qixing, XUE Hua, et al. Ultra-short-term prediction model of enhanced brain emotional neural network considering chaotic characteristics for photovoltaic power generation[J]. High Voltage Engineering, 2021, 47(4): 1165-1175.
- [21] 王开艳, 杜浩东, 贾嵘, 等. 基于相似日聚类 and QR-CNN-BiLSTM 模型的光伏功率短期区间概率预测[J]. 高电压技术, 2022, 48(11): 4372-4384. WANG Kaiyan, DU Haodong, JIA Rong, et al. Short-term interval probability prediction of photovoltaic power based on similar daily clustering and QR-CNN-BiLSTM model[J]. High Voltage Engineering, 2022, 48(11): 4372-4384.
- [22] 毕贵红, 赵鑫, 陈臣鹏, 等. 基于多通道输入和 PCNN-BiLSTM 的光伏发电功率超短期预测[J]. 电网技术, 2022, 46(9): 3463-3476. BI Guihong, ZHAO Xin, CHEN Chenpeng, et al. Ultra-short-term prediction of photovoltaic power generation based on multi-channel input and PCNN-BiLSTM[J]. Power System Technology, 2022, 46(9): 3463-3476.
- [23] 朱乔木, 李弘毅, 王子琪, 等. 基于长短期记忆网络的风电场发电功率超短期预测[J]. 电网技术, 2017, 41(12): 3797-3802. ZHU Qiaomu, LI Hongyi, WANG Ziqi, et al. Short-term wind power forecasting based on LSTM[J]. Power System Technology, 2017, 41(12): 3797-3802.
- [24] 黄旭锐, 于丰源, 杨波, 等. 基于 Transformer 网络和多任务学习的园区综合能源系统电-热短期负荷预测方法[J]. 南方电网技术, 2023, 17(1): 152-160. HUANG Xurui, YU Fengyuan, YANG Bo, et al. Short-term electric-thermal load forecasting method for park-level integrated energy system based on transformer network and multi-task learning[J]. Southern Power System Technology, 2023, 17(1): 152-160.
- [25] SUN G, ZHANG C, WOODLAND P C. Transformer language models with LSTM-based cross-utterance information representation[C]/Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, Canada: IEEE, 2021: 7363-7367.



DONG Jun  
Ph.D. candidate  
Associate professor



SHU Hongchun  
Ph.D., Professor  
Corresponding author

#### 董俊

1977—, 男, 博士生, 副教授  
研究方向为配电网故障检测、保护与自愈  
E-mail: dongjun\_kust@sina.com

#### 束洪春(通信作者)

1961—, 男, 博士, 教授, 博导  
主要从事电力系统新型继电保护与故障测距、故障录波、数字信号处理及 DSP 应用等方面的研究  
E-mail: kmshc@sina.com.cn

收稿日期 2023-05-10 修回日期 2023-12-11 编辑 程子丰