

DOI: 10.12158/j.2096-3203.2026.03.010

基于相似日和双层校正 LSTM 的光伏功率短期预测方法

石宇龙¹, 彭乔¹, 刘天琪¹, 陈刚², 曾雪洋², 李燕²

(1. 四川大学电气工程学院, 四川 成都 610065;

2. 国网四川省电力公司电力科学研究院, 四川 成都 610041)

摘要: 光伏功率的准确预测对于电力系统的调度、决策至关重要。为提高光伏功率预测精度, 文中基于数据驱动原理, 提出基于相似日和双层校正长短时记忆(long short-term memory, LSTM)的光伏功率短期预测方法。首先, 对光伏功率和相关气象数据进行归一化处理, 并通过皮尔逊系数确定影响光伏功率的关键因素, 降低训练数据维度。然后, 利用弗雷歇距离算法匹配待预测日的相似日, 提升训练数据质量。最后, 根据数值天气预报, 基于特征学习方法在基准层 LSTM 中得到光伏功率一次预测值, 并基于时间序列方法在校正层 LSTM 中得到光伏功率误差预测值, 对一次预测值进行校正, 得到最终预测值。以实地采集的真实数据为例, 选取晴天、多云、雨天等不同天气下的参考日作为待预测对象进行算例分析。结果表明, 使用文中所提模型与方法在不同条件下均能有效预测未来 24 h 的短期光伏功率, 相比已有方法能大幅提升预测精度。

关键词: 光伏系统; 短期预测; 相似日; 长短时记忆(LSTM)网络; 双层校正模型; 弗雷歇距离算法

中图分类号: TM615

文献标志码: A

文章编号: 2096-3203(2026)03-0085-10

0 引言

近年来, 光伏等新能源在电力系统中的渗透率不断提高, 而光伏发电受辐射度、温度、湿度等因素的影响, 具有很强的波动性与间歇性, 给电网的安全稳定运行带来了挑战^[1]。因此, 研究高精度的光伏功率预测方法对于降低弃光率、提升光伏功率消纳、增强系统运行稳定性具有重要意义^[2-3]。

按照预测机理的不同, 光伏功率预测可以分为 2 类: 基于物理模型和基于历史数据。前者根据光伏电池转换机理创建发电模型, 结合数值天气预报(numerical weather prediction, NWP)的数据集对光伏电站的产出功率进行估算^[4-6]。然而, 构建精确的物理模型在实际应用中非常困难, 因为其依赖于详细的组件参数和场站地理信息, 模型参数难以确定。随着光伏设备老化和环境变化, 模型的预测精度进一步降低, 基于物理模型的方法在面对复杂气象条件时表现不佳^[7]。后者利用人工智能和机器学习技术, 建立光伏功率与气象数据之间的映射关系, 能够较好地处理非线性问题, 具有灵活性高、成本低、适用性强的优势。常用方法有马尔科夫链^[8-9]、支持向量机法^[10]、多元线性回归^[11]、神经网络法^[12]、时间序列法^[13]等。其中, 长短时记忆(long short-term memory, LSTM)网络作为一种处理时间

序列数据的深度学习模型, 凭借其对于长期依赖性数据的处理能力, 成为光伏功率预测的热门方法^[14-16]。然而, 单层 LSTM 在面对多特征数据时存在处理复杂和信息丢失等问题, 难以适应日益复杂的光伏功率预测任务^[17-18]。

针对上述问题, 许多学者开始结合特征提取方法与 LSTM 进行组合模型优化。文献^[19]使用主成分分析(principal component analysis, PCA)法结合经验模态分解法筛选影响光伏功率变化的关键特征, 降低输入维度。在此基础上, 文献^[20]为避免使用 PCA 可能导致的信息损失或其他复杂、不稳定功率信号的扰动, 结合变异模态分解和多维特征建模, 通过双向 LSTM 进行预测, 改善复杂天气条件下的预测精度。此外, 文献^[21-22]将注意力机制引入 LSTM 模型中, 将输入特征重建并映射到预测的输出中, 提升了训练速度和预测准确率。但这些方法增加了模型设计的复杂性和参数设置的困难性, 尤其是数据量较大时, 特征提取的计算代价会显著提升。

为提高训练数据的质量, 许多研究者引入相似日匹配方法, 通过筛选与预测日气象条件相似的历史数据, 提高预测模型的训练效果。文献^[23]基于互信息熵(mutual information entropy, MIE)筛选出与预测日气象条件最接近的历史日, 虽然提高了相似日的筛选精度, 但在面对大数据集时计算冗杂。为此, 文献^[24-25]基于灰色关联度分析(gray relation analysis, GRA)对天气数据进行处理, 简化了计算过

收稿日期: 2025-07-04; 修回日期: 2025-09-19

基金项目: 国家自然科学基金资助项目(52207218); 国家电网有限公司科技项目(5108-202299262A-1-0-ZB)

程,但该方法的关联度阈值设置具有主观性,影响模型的泛化能力和适用性。此外,部分研究采用聚类算法划分天气类型并筛选历史数据。如模糊C均值(fuzzy C-means, FCM)^[26-27]、K-means++^[18]、自适应K-means^[28]、近邻传播(affinity propagation, AP)^[29]等,将历史数据划分为不同气象条件下的天气类型,但这些聚类方法对聚类中心和数量的选择敏感,需要反复调整,容易引入模型不稳定性。

现有光伏功率预测方法在处理复杂气象条件下的天气数据时,往往面临数据量不足、算法复杂度高和计算成本大等问题。为了应对这些挑战,文中提出一种基于相似日和双层校正LSTM的光伏功率短期预测模型。该模型首先利用皮尔逊系数筛选出影响光伏功率的关键气象特征,然后通过弗雷歇距离算法进行相似日匹配。在双层LSTM结构中,基准层LSTM通过特征学习方法生成初步预测结果,校正层LSTM则利用时间序列数据对初步预测结果进行误差校正。实验结果表明,文中所提模型能够有效提高不同天气条件下的光伏功率预测精度。

1 光伏功率和相关气象数据预处理

数据集中,由于各特征的数值范围不同,为避免数值范围差异对模型训练的影响,需要进行归一化处理。文中采用最大最小归一化法对数据进行处理,计算如下:

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中: x' 为归一化后的数据; x_i 为该数据序列中第*i*个原始数据; x_{\min} 为该数据序列的最小值; x_{\max} 为该数据序列的最大值。

光伏功率通常受辐射度、环境温度、湿度、气压和风速等多种外部因素影响,且影响程度各不相同,文中利用皮尔逊系数筛选出影响光伏功率的关键因素。皮尔逊系数作为一种衡量2个随机变量间线性关联程度的指标,在预测研究中被广泛应用^[30]。其优势在于对变量的位置和测量尺度变化不敏感,能保证相关性度量的稳定性。对于数据序列*X*和*Y*,其皮尔逊系数计算如下:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

式中: $\rho_{X,Y}$ 为*X*和*Y*的皮尔逊系数, $\rho_{X,Y} \in [-1,1]$; $\text{cov}(X,Y)$ 为*X*和*Y*的协方差; σ_X 、 σ_Y 分别为*X*、*Y*的标准差。 $|\rho_{X,Y}|$ 越趋近1,*X*和*Y*的相关性越强; $|\rho_{X,Y}|$ 越趋近0,*X*和*Y*的相关性越弱。 $\rho_{X,Y}^2 > 0.5$ 时,认为*X*和*Y*具有强相关性^[31]。

2 基于弗雷歇距离算法的相似日匹配

2.1 离散化弗雷歇距离算法

衡量数据序列相似程度的常见方法是欧氏距离算法,但该算法可能忽略数据形状的相似性和时序特征。弗雷歇距离算法改进了欧氏距离算法,通过考虑数据的位置和点的顺序来计算2条曲线的相似度,因此常用于时间序列和轨迹序列的相似度量^[32]。

假设*X*和*Y*是空间*S*中的2条连续曲线,即: $X: [a_1, a_2] \rightarrow \mathbf{R}^n$, $Y: [b_1, b_2] \rightarrow \mathbf{R}^n$; 又设有2个参数 $t_A \in [a_1, a_2]$ 和 $t_B \in [b_1, b_2]$, 分别表示曲线*X*和*Y*上的点,则弗雷歇距离 $F_{X,Y}$ 为:

$$F_{X,Y} = \inf_{\alpha, \beta} \max_{t \in [0,1]} \|X_{\alpha(t)} - Y_{\beta(t)}\| \quad (3)$$

式中: $\alpha: [0,1] \rightarrow [a_1, a_2]$ 、 $\beta: [0,1] \rightarrow [b_1, b_2]$ 为单位区间的2个重采样函数,用于参数化曲线*X*和*Y*; $X_{\alpha(t)}$ 、 $Y_{\beta(t)}$ 分别为曲线*X*、*Y*在参数 $\alpha(t)$ 、 $\beta(t)$ 处的点。

NWP数据一般是离散的,因此需要将弗雷歇距离算法离散化。首先,参考日NWP数据曲线*P*和历史日*i*的NWP数据曲线*Q_i*可以表示为:

$$\begin{cases} P = \{P_1, P_2, \dots, P_n, \dots, P_N\} \\ Q_i = \{Q_{i,1}, Q_{i,2}, \dots, Q_{i,m}, \dots, Q_{i,M}\} \end{cases} \quad (4)$$

式中: $P_n = (x_n, y_n)$, $Q_{i,m} = (u_m, v_m)$, x_n 、 y_n 分别为*P*上第*n*个采样点的时间数据和气象数据, u_m 、 v_m 分别为*Q_i*上第*m*个采样点的时间数据和气象数据。计算得到距离矩阵*D*为:

$$D = \begin{bmatrix} d_{1,1} & \dots & d_{1,N} \\ \vdots & & \vdots \\ d_{M,1} & \dots & d_{M,N} \end{bmatrix} \quad (5)$$

式中: $d_{m,n}$ 为*Q_{i,m}*到*P_n*的距离。

$$d_{m,n} = \sqrt{(u_m - x_n)^2 - (v_m - y_n)^2} \quad (6)$$

找出距离矩阵中的最大距离 d_{\max} 和最小距离 d_{\min} , 初始化目标距离 $f = d_{\min}$, 设置循环间隔 $r = (d_{\max} - d_{\min})/100$ 。将距离矩阵*D*中的元素和目标距离*f*依次进行比较,更新距离矩阵为:

$$D' = \begin{bmatrix} d'_{1,1} & \dots & d'_{1,N} \\ \vdots & & \vdots \\ d'_{M,1} & \dots & d'_{M,N} \end{bmatrix} \quad (7)$$

式中: D' 为更新后的距离矩阵。更新后的 $d'_{m,n}$ 满足:

$$d'_{m,n} = \begin{cases} 1 & d_{m,n} \leq f \\ 0 & d_{m,n} > f \end{cases} \quad (8)$$

接着,在*D'*中搜索路径*R*,满足以下条件:

$$d'_{1,1} \times \dots \times d'_{m,n} \times d'_{(m+k),(n+k')} \times \dots \times d'_{M,N} = 1 \quad (9)$$

式中: $k \in \{0,1\}$, $k' \in \{0,1\}$ 。如果在该过程中无法找

到符合条件的任何路径, 则修改新目标距离为 $f = f + r$, 直至找到一条合适的路径或已经达到目标距离 $f = d_{\max}$ 。因此, 参考日 NWP 数据曲线 P 和历史日 i 的 NWP 数据曲线 Q_i 之间的弗雷歇距离 $F_{x,y} = f$ 。

2.2 相似日匹配

根据弗雷歇距离算法, 可定义参考日 NWP 数据和历史某日 NWP 数据的相似程度 K 为:

$$K = \frac{1 - F'_{x,y}}{F'_{x,y}} \times 100\% \quad (10)$$

式中: $F'_{x,y}$ 为归一化处理后的弗雷歇距离值。

此外, 历史 NWP 数据曲线集合 Z 可以表示为:

$$Z = \{Q_1, Q_2, \dots, Q_j\} \quad (11)$$

式中: j 为历史日期数。将集合 Z 中的曲线与参考日 NWP 数据曲线 P 进行弗雷歇距离的遍历计算, 可获得历史 NWP 数据集合与参考日 NWP 数据的相似程度。

3 基于双层校正 LSTM 的光伏功率预测

3.1 LSTM 原理

LSTM 神经网络通过增设门控机制控制特征的循环和丢失, 解决了循环神经网络 (recurrent neural network, RNN) 中存在的梯度爆炸和长期序列信息“遗忘”问题^[24], 其模型原理概述如下。

(1) 遗忘阶段决定需要从细胞状态汇总丢弃的信息, 可以表示为:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (12)$$

式中: t 为时间步; f_t 为当前时刻遗忘门激活向量; σ 为 Sigmoid 激活函数, 将输出值压缩到 (0,1) 区间; W_f 、 b_f 分别为遗忘门的权重矩阵和偏置项; h_{t-1} 为上一时刻的隐藏状态; x_t 为当前时刻的输入向量。

(2) 输入阶段决定新信息在细胞状态中的存储程度, 由 Sigmoid 层和 Tanh 层组成。

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (13)$$

$$C'_t = \text{Tanh}(W_c \times [h_{t-1}, x_t] + b_c) \quad (14)$$

式中: i_t 为当前时刻输入门激活向量; C'_t 为当前时刻候选细胞状态; W_i 、 b_i 分别为输入门的权重矩阵和偏置项; W_c 、 b_c 分别为候选细胞状态的权重矩阵和偏置项; Tanh 为激活函数, 将候选细胞状态压缩到 (-1, 1) 区间。

(3) 由遗忘阶段和输入阶段更新细胞状态:

$$C_t = f_t \times C_{t-1} + i_t \times C'_t \quad (15)$$

式中: C_t 为当前时刻细胞状态; C_{t-1} 为上一时刻细胞状态。

(4) 输出阶段输出信息并更新下一隐藏阶段。

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (16)$$

$$h_t = o_t \times \text{Tanh}(C_t) \quad (17)$$

式中: o_t 为当前时刻输出门激活向量; h_t 为当前时刻隐藏状态; W_o 、 b_o 分别为输出门的权重矩阵和偏置项。LSTM 总体结构和运行机制如图 1 所示。

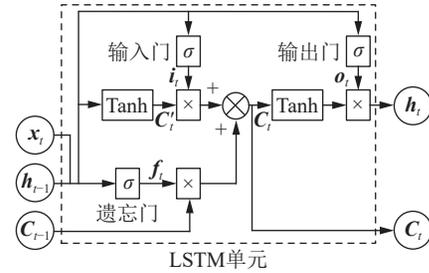


图 1 LSTM 结构和运行机制

Fig.1 Structure and operational mechanism of LSTM

3.2 基于双层校正 LSTM 的光伏预测模型

基于 LSTM 的光伏预测模型可以分为基于特征学习和基于时间序列 2 类, 如图 2 所示。

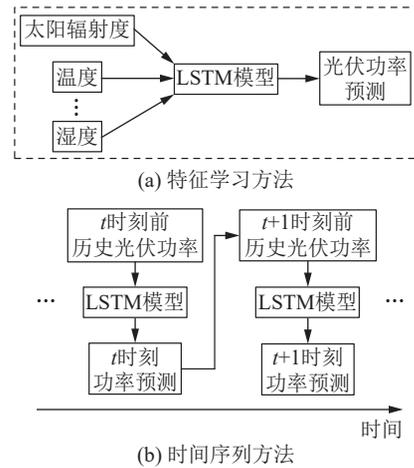


图 2 基于不同方法的 LSTM 光伏功率预测模型

Fig.2 LSTM photovoltaic power forecasting models based on different methods

基于特征学习的光伏预测模型是指在不考虑光伏设备自身老化、损坏等物理变化的影响时, 光伏功率主要受到太阳辐射度、温度、湿度等气象因素的直接影响, 其相关数值可由 NWP 与本地气象监测仪提供, 因此可以利用历史气象数据和光伏功率之间的映射关系进行未来功率预测。

基于时间序列的光伏预测模型主要针对历史光伏功率数据进行趋势学习建模, 其时间序列特征指以当前时刻为时间节点, 将之前不同时刻下的历史光伏出力数据作为特征, 预测下一时刻的光伏功率。与基于特征学习的方法不同, 时间序列模型能够捕捉到光伏设备随时间的物理变化, 这些变化体现在时间序列数据中, 弥补了单纯依赖气象特征的模型的不足。

因此,基于 LSTM 的特征学习模型可以通过气象预报值预测光伏功率,但气象预报值与实时气象值之间难免存在偏差,此时光伏预测的准确度无法得到保证。而基于 LSTM 的时间序列模型突出了时间因素和面板物理因素在预测中的作用,但在遇到极端气象变化时会出现大幅波动,导致预测精度大大下降。文中所提双层校正模型将上述 2 种预测模型融合,采取基于相似日的双层 LSTM 预测模型,预测模型的具体步骤如下。

步骤 1: 数据收集与清洗。采集光伏功率数据及光伏电站的相关气象数据,分析数据质量,使用上下四分位距法^[33]筛选异常值,并采用线性插值法替换少量异常值及补全缺失数据。

步骤 2: 数据归一化。将数据按天划分,对特征量和光伏功率归一化处理,以统一数据尺度。具体而言,太阳辐射度、温度、湿度以及光伏功率被标准化到[0,1]区间,从而去除不同量纲的影响。

步骤 3: 气象特征量降维。通过皮尔逊系数对特征量与光伏功率进行相关性分析,从而确定适合该数据集的气象因子权重系数。将特征量输入维度进行有效减少,厘清影响光伏功率的关键因素。

步骤 4: 参考日确定。根据数据来源地的历史天气记录和光伏功率曲线,确定用于实验预测的参考日数据。

步骤 5: 相似日匹配。利用弗雷歇距离算法将参考日的气象数据作为被比较对象,对参考日之前的日期的气象数据进行遍历循环,根据计算的弗雷歇距离折算得到相似程度 K ,找出相似度最高的 7 个日期作为预测模型的训练数据来源。

步骤 6: 基准层 LSTM 模型的训练与预测。利用相似日气象数据作为输入特征,相似日光伏功率作为输出,对 LSTM 模型进行训练。将最新 NWP 提供的参考日气象特征量作为基准层 LSTM 模型的预测输入,计算得到光伏功率的一次预测值 P_1^{pre} 。同时,将临近参考日日期的历史 NWP 数据作为基准层 LSTM 模型的预测输入,计算得到临近日期的光伏功率预测值 P_2^{pre} 。

步骤 7: 校正层 LSTM 模型的训练与预测。在训练集中,临近日期的误差值 ε 可以表示为:

$$\varepsilon = P_{\text{similar}} - P_2^{pre} \quad (18)$$

式中: P_{similar} 为临近日期的光伏功率。将 ε 作为带有时间序列特征的已知项,构造出历史临近日期误差项的变化趋势。在预测集中,依据光伏面板物理特性变化的连续性,基于时间连续性对参考日误差项进行预测,计算得到光伏功率的误差预测值 ε' 。

步骤 8: 光伏功率预测值修正。将基准层和校正层 LSTM 模型的预测结果叠加,得到参考日的光伏功率预测结果,可以表示为:

$$P^{pre} = P_1^{pre} + \varepsilon' \quad (19)$$

式中: P^{pre} 为参考日的光伏功率预测值。从步骤 6—步骤 8 可知,基于双层校正 LSTM 的光伏功率短期预测模型结构包含基准层和校正层 2 部分,如图 3 所示。

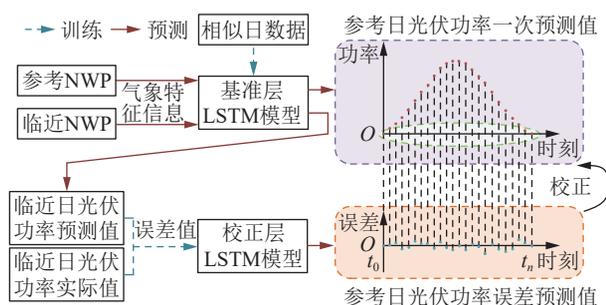


图 3 基于双层校正 LSTM 的短期预测模型

Fig.3 Short-term forecasting model based on bi-layer correction LSTM

步骤 9: 误差分析与比较。对不同天气类型下不同模型得到的光伏功率短期预测数据与真实功率进行误差分析。图 4 为从数据预处理到最终结果分析的所有关键步骤,确保了方法的完整性和系统性。

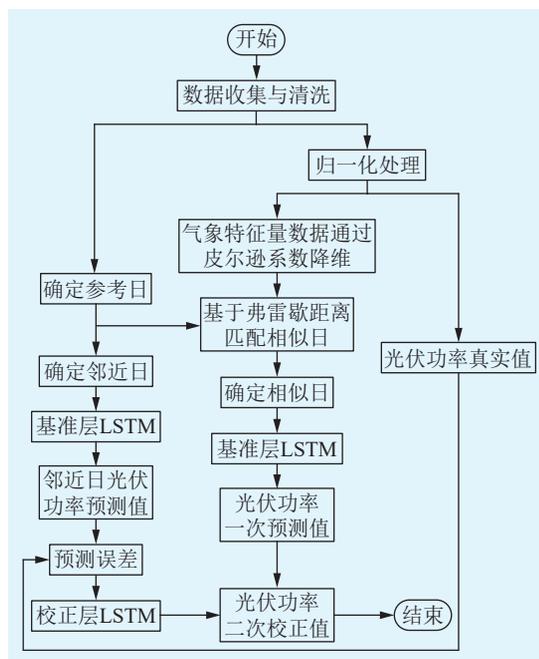


图 4 光伏功率短期预测流程

Fig.4 Flow chart of PV short-term forecasting method

4 算例分析

4.1 数据介绍

文中所用实验数据为澳大利亚沙漠知识太阳

能研究中心(Desert Knowledge Australia Solar Centre, DKASC)和国家太阳辐射数据库(National Solar Radiation Database, NSRDB)2019年1月1日至2020年12月31日的历史光伏功率数据和相关气象数据,时间分辨率为5 min,单日采样点数量为288。并依据澳大利亚历史天气统计数据^[34],分别选取晴天(2020年9月9日)、多云(2020年1月6日)和雨天(2020年12月10日)作为待预测的参考日。此外,以参考日之前的7 d作为临近日。为验证文中所提方法的有效性,将该预测模型和多种预测模型在同一数据集下进行对比实验,且所有模型的非关键参数设置全部相同。

4.2 误差指标

均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和平均绝对百分比误差(mean absolute percentage error, MAPE)能有效评价预测性能^[35-37],因此使用这3个指标评估模型预测的准确度。计算如下:

$$\begin{cases} E_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2} \\ E_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i| \\ E_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - y'_i}{y_i} \right| \times 100\% \end{cases} \quad (20)$$

式中: E_{RMSE} 、 E_{MAE} 、 E_{MAPE} 分别为预测值与真实值的RMSE、MAE和MAPE; N 为测试的光伏功率样本点总个数; y_i 为第*i*个光伏功率真实值; y'_i 为第*i*个光伏功率预测值。

4.3 算例结果分析

基于DKASC和NSRDB的数据集,采用皮尔逊系数衡量气象因素对光伏电站功率的影响程度,具体见表1。

表1 不同影响因素和光伏电站功率的皮尔逊系数
Table 1 Person coefficients for various factors influencing PV output power

影响因素	皮尔逊系数
太阳辐射度	0.96
温度	0.38
湿度	-0.41
风速	-0.15
气压	-0.04

根据判定系数 $\rho_{X,Y}^2 > 0.5$ 的标准,基准层LSTM光伏预测模型选取太阳辐射度作为特征输入量,并对所选参考日的太阳辐射度和历史太阳辐射度进

行遍历,计算弗雷歇距离,获得对应的相似日日期及相似程度,如表2所示。

表2 相似日匹配结果和相似程度
Table 2 Selection results of similar days and similarity degree

参考日	相似日日期	相似程度K/%
2020年 9月9日 (晴天)	2020年4月4日	99.3
	2019年9月2日	99.3
	2020年9月8日	99.2
	2019年9月8日	99.1
	2019年8月30日	99.0
	2019年8月31日	98.9
2020年 1月6日 (多云)	2020年9月6日	98.9
	2019年6月12日	88.8
	2019年9月23日	86.6
	2019年4月30日	86.4
	2019年9月24日	85.2
	2019年6月14日	84.3
2020年 12月10日 (雨天)	2019年6月15日	84.3
	2019年4月17日	82.6
	2020年3月3日	91.0
	2020年5月20日	91.0
	2020年3月2日	89.4
	2020年5月26日	89.3
	2019年6月12日	88.8
	2020年4月9日	88.7
	2020年6月14日	88.1

选取6种模型的预测结果进行对比分析。模型1:单层LSTM,使用数据集中的参考日之前的太阳辐射度数据和带有时间特征的历史光伏功率数据训练模型;设置1层隐藏层,神经元个数为128,激活函数为Sigmoid,学习率为0.001,最大训练迭代次数为1000,dropout率为0.2。模型2:相似日-LSTM,对数据集中参考日之前的太阳辐射度数据基于弗雷歇距离算法匹配相似日后,使用相似日数据训练模型,参数设置与模型1类似,但神经元数为64,dropout率为0.15。模型3:变分模态分解(variational mode decomposition,VMD)-LSTM,基于VMD技术提取光伏功率数据关键特征后,使用LSTM进行预测,VMD部分模态数为3,惩罚参数为2000,频带的中心频率为0.1,LSTM部分设置为1层隐藏层,神经元个数为200,学习率为0.001,最大训练迭代次数为1000。模型4:Attention-LSTM,在单层LSTM中引入注意力机制,增强模型对关键时间步的识别能力,注意力机制中使用8个注意力头,权重通过Softmax函数调整,LSTM设置

为2层结构,每层100个神经元,学习率为0.0005, dropout率为0.1,最大训练迭代次数为1000。模型5:相似日-多层LSTM(multi-layer LSTM, MLT-LSTM),使用3层结构,设置3层隐藏层,神经元个数分别为128、64、32,激活函数使用Tanh,学习率为0.0005, dropout率为0.2,最大训练迭代次数为2000。模型6:文中所提双层校正LSTM模型,基准层使用128个神经元进行初步预测,校正层使用64个神经元修正误差,激活函数均为Sigmoid,学习率为0.0012, dropout率为0.18,最大训练迭代次数为2000。参考日下各模型的预测结果如图5所示,相关预测误差分析见表3—表5。

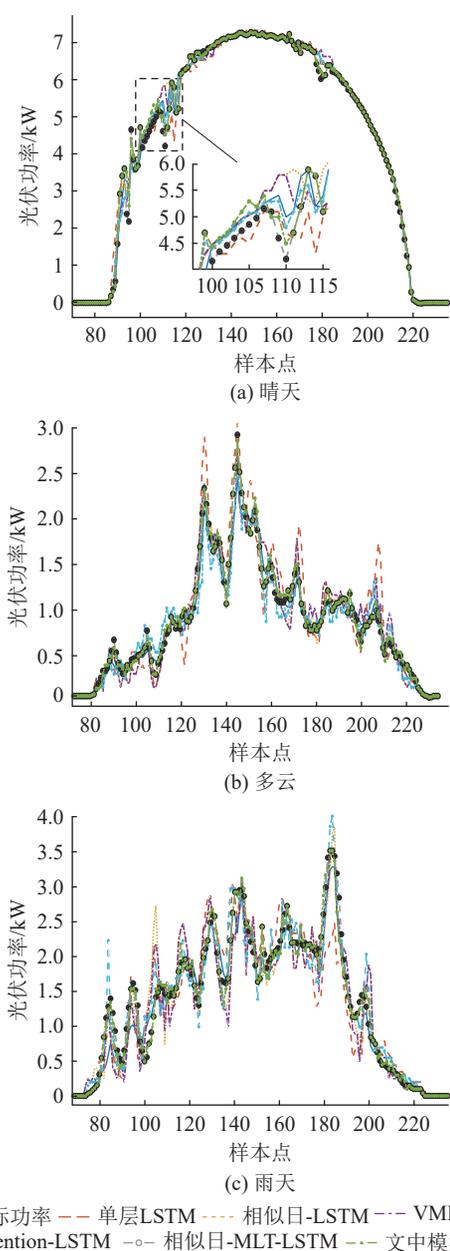


图5 不同模型在不同天气类型下的光伏功率预测曲线
Fig.5 The PV power forecasting curves of different models under various weather types

表3 晴天时不同模型预测误差

Table 3 Errors in forecasting by different models under sunny weather

预测模型	E_{RMSE}/kW	E_{MAE}/kW	$E_{MAPE}/\%$
单层LSTM	0.2718	0.1519	25.2
相似日-LSTM	0.1833	0.0864	17.3
VMD-LSTM	0.1764	0.0934	16.9
Attention-LSTM	0.1253	0.0673	14.3
相似日-MLT-LSTM	0.1324	0.0502	9.0
文中模型	0.1212	0.0487	8.6

表4 多云时模型预测误差

Table 4 Errors in forecasting by different models under cloudy weather

预测模型	E_{RMSE}/kW	E_{MAE}/kW	$E_{MAPE}/\%$
单层LSTM	0.2681	0.1868	23.6
相似日-LSTM	0.1518	0.1154	18.6
VMD-LSTM	0.1532	0.1165	19.2
Attention-LSTM	0.1357	0.1028	15.5
相似日-MLT-LSTM	0.1259	0.1081	16.0
文中模型	0.1230	0.0855	11.0

表5 雨天时模型预测误差

Table 5 Errors in forecasting by different models under rainy weather

预测模型	E_{RMSE}/kW	E_{MAE}/kW	$E_{MAPE}/\%$
单层LSTM	0.3434	0.2704	44.3
相似日-LSTM	0.3165	0.2333	29.9
VMD-LSTM	0.3127	0.2379	28.9
Attention-LSTM	0.2924	0.1853	24.8
相似日-MLT-LSTM	0.2621	0.1644	23.6
文中模型	0.2236	0.1481	17.4

为验证文中模型的简约性,比较各模型在雨天条件下的训练时间和预测时间,如表6所示。

表6 雨天时不同模型的时间成本

Table 6 The time cost of different models under rainy weather

预测模型	训练时间/s	预测时间/s
单层LSTM	65.3	0.8
相似日-LSTM	82.5	0.9
VMD-LSTM	100.2	1.0
Attention-LSTM	115.6	1.0
相似日-MLT-LSTM	125.7	1.1
文中模型	90.4	0.9

可以看出,文中所提双层校正LSTM模型在不同天气条件下均表现出最佳的预测性能。在晴天条件下,由于光伏功率波动较小,日间辐射度变化

趋势较为稳定,功率曲线相对平滑,因此所有模型的预测精度均较为理想。相比于相似日-LSTM 等传统模型,文中模型引入校正层 LSTM,基于临近日光伏功率连续变化的微小误差对基准层 LSTM 预测结果进行动态修正,显著减少了噪声干扰,从而提升了预测精度。而模型 5 使用 3 层 LSTM 结构,增强了对时间序列深层信息的提取能力,因此在晴天这种较为平稳的环境下表现出较高的预测精度,其 E_{MAPE} 为 9.0%。

在多云或雨天条件下,由于云层的频繁移动和降雨的发生和结束,光伏功率的非线性变化显著增加。模型 1 依赖于固定结构来处理时间序列数据,在训练过程中未能充分处理差异性较大的天气数据,导致训练数据中包含过多的噪声和异常值,削弱了模型捕捉光伏功率变化趋势的能力,预测误差显著增加。模型 2 引入相似日匹配技术减少训练数据中的噪声和不相关信息,但因相似日数据的时间序列不连续,模型在捕捉光伏设备的实际物理变化和气象条件的剧烈波动时表现不足。模型 3 利用 VMD 技术在一定程度上降低训练数据中的噪声干扰,将数据特征分解为多个维度进行叠加,提升了整体预测精度,但多模态的合成增加了模型的计算复杂性和资源消耗。模型 4 对非线性特征的深度挖掘,使得其在面对突发天气变化时能够保持较高的预测精度。模型 5 虽然具备较强的深层特征提取能力,但多层次设计增加了梯度消失和爆炸的风险,特别在多云和雨天时表现出不稳定性。相比之下,文中所提模型是在模型 2 的基础上进行优化,针对时间序列不连续性问题,通过引入含有光伏面板状态和环境变化信息的临近日数据,进一步提升了数据的时序连续性。基准层 LSTM 学习光伏功率的主要变化趋势,校正层 LSTM 通过捕捉误差进行动态调整,尤其是在面对雨天等剧烈波动的气象条件时,能够显著减少预测误差。文中模型与模型 2 相比,晴天、多云、雨天下的 E_{RMSE} 分别降低了 0.062 1、0.028 8、0.092 9 kW, E_{MAE} 分别降低了 0.037 7、0.029 9、0.085 2 kW, E_{MAPE} 分别降低了 8.7、7.6、12.5 个百分点。

从表 6 可以看出,各模型的计算成本随着结构复杂度的增加而提升。模型 1 的训练时间为 65.3 s,预测时间为 0.8 s,效率最高。模型 2 的训练时间为 82.5 s,模型 3 由于引入模态分解,训练时间增加至 100.2 s。模型 4 和模型 5 的复杂性更高,训练时间分别为 115.6 s 和 125.7 s。相比之下,文中所提双层校正 LSTM 模型在保证高预测精度的同时,训练时

间为 90.4 s,预测时间为 0.9 s,较好地实现了计算效率与精度平衡。

4.4 不同相似日匹配方法分析

为了进一步评价基于弗雷歇距离算法的相似日匹配方法的简约性和有效性,针对出力变化较为剧烈的雨天,使用不同的相似日匹配方法与文中所提双层校正 LSTM 模型进行组合,预测结果如表 7 所示。

表 7 雨天时模型的预测误差
Table 7 Prediction errors of models under rainy weather

相似日匹配方法	E_{RMSE}/kW	$E_{MAPE}/\%$	计算时间/s
MIE	0.231 5	19.6	0.037 7
GRA	0.218 8	18.1	0.035 2
K-means	0.299 2	22.4	0.025 5
FCM	0.301 6	22.8	0.028 9
AP	0.293 1	22.1	0.038 9
弗雷歇距离	0.223 6	17.4	0.036 6

由表 7 可知, MIE 算法通过量化气象变量与光伏功率的非线性相关性选择相似日,但因需要计算多维概率分布,复杂度高,计算时间较长。GRA 算法通过线性特征评估相似性,虽然处理不确定性较好,但无法充分捕捉非线性关系,导致其相对预测精度略低于弗雷歇距离算法。K-means 聚类固定聚类中心,难以应对非线性天气波动,且聚类中心选择随机,误差较大。FCM 算法使用模糊隶属度处理数据,但在复杂天气下,模糊性无法准确反映波动,预测精度不佳。AP 算法虽然能够处理非线性和噪声,但其工作原理是通过数据点之间的相似性传播选择代表点,随着数据复杂度增加(如雨天时气象数据的剧烈波动),相似性传播过程需要处理更多的距离计算和迭代更新,导致计算时间较长(0.038 9 s)。此外, AP 算法在面对高度波动的气象条件时,由于依赖相似性传播,对于突发性的气象变化捕捉不足,导致其预测精度不高。而文中提出的弗雷歇距离算法通过衡量历史日和预测日的形状相似性,兼顾时间序列的局部变化和整体趋势,能够更好地应对复杂天气条件,计算简单且高效,在保证精度的同时,有效缩短了相似日匹配的计算时间。

5 结论

针对现有光伏功率预测方法算法过于复杂或准确度不足的问题,文中提出基于弗雷歇距离算法的相似日匹配和双层校正 LSTM 光伏功率预测方法,主要结论如下:

- (1) 基于弗雷歇距离算法的相似日匹配方法能

简单、有效地提高训练数据的相关性和可用性,改善预测精度。

(2) 与单层 LSTM 模型相比, 双层校正 LSTM 模型中的校正层 LSTM 能够有效针对时间序列中隐含的气象、光伏面板变化趋势信息对预测误差进行推演, 从而校正基准层 LSTM 预测值, 在保证模型简约性的同时有效提高预测精度。基于实地采集的真实数据进行算例分析, 结果表明文中方法在晴天、多云、雨天天气下的预测结果的 MAPE 分别为 8.6%、11.0%、17.4%, 相比已有方法预测精度更高。

参考文献:

- [1] 卓振宇, 张宁, 谢小荣, 等. 高比例可再生能源电力系统关键技术及发展挑战[J]. *电力系统自动化*, 2021, 45(9): 171-191.
ZHUO Zhenyu, ZHANG Ning, XIE Xiaorong, et al. Key technologies and development challenges of high proportion renewable energy power system[J]. *Automation of Electric Power Systems*, 2021, 45(9): 171-191.
- [2] ZHEN H, NIU D X, WANG K K, et al. Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information[J]. *Energy*, 2021, 231: 120908.
- [3] 杨锡运, 杨岩, 孟令卓越, 等. 计及 NWP 信息缺失的数据共享与 GRA 权重优化的分布式光伏电站功率预测[J]. *电测与仪表*, 2025, 62(4): 172-179.
YANG Xiyun, YANG Yan, MENG Lingzhuochao, et al. Data sharing and GRA weight optimization for power prediction of distributed photovoltaic power plant considering missing NWP information[J]. *Electrical Measurement & Instrumentation*, 2025, 62(4): 172-179.
- [4] MAYER M J. Influence of design data availability on the accuracy of physical photovoltaic power forecasts[J]. *Solar Energy*, 2021, 227: 532-540.
- [5] 董存, 王铮, 白捷予, 等. 光伏发电功率超短期预测方法综述[J]. *高电压技术*, 2023, 49(7): 2938-2951.
DONG Cun, WANG Zheng, BAI Jieyu, et al. Summary of ultra-short-term forecasting methods for photovoltaic power generation[J]. *High Voltage Engineering*, 2023, 49(7): 2938-2951.
- [6] 战文华, 车建峰, 王勃, 等. 基于网格化数值天气预报的区域光伏发电多输出功率预测方法[J]. *中国电力*, 2024, 57(3): 144-151.
ZHAN Wenhua, CHE Jianfeng, WANG Bo, et al. A grid-based numerical weather prediction method for multi-output prediction of regional photovoltaic power[J]. *Electric Power*, 2024, 57(3): 144-151.
- [7] 杨建, 常学军, 姚帅, 等. 基于 WT-CNN-BiLSTM 模型的日前光伏功率预测[J]. *南方电网技术*, 2024, 18(8): 61-69, 79.
YANG Jian, CHANG Xuejun, YAO Shuai, et al. Prediction of photovoltaic power in recent days based on WT-CNN-BiLSTM model[J]. *Southern Power System Technology*, 2024, 18(8): 61-69, 79.
- [8] BAI X Y, LIANG L, ZHU X Q. Improved Markov-chain-based ultra-short-term PV forecasting method for enhancing power system resilience[J]. *The Journal of Engineering*, 2021(2): 114-124.
- [9] 刘迪, 吴林林, 巩宇, 等. 基于滚动采样马尔可夫链模型的光伏时序功率模拟研究[J]. *电力科学与技术学报*, 2024, 39(3): 207-216.
LIU Di, WU Linlin, GONG Yu, et al. Study on time series power simulation of photovoltaic output based on rolling sampling Markov chain model[J]. *Journal of Electric Power Science and Technology*, 2024, 39(3): 207-216.
- [10] 陈元峰, 马溪原, 程凯, 等. 基于气象特征量选取与 SVM 模型参数优化的新能源超短期功率预测[J]. *太阳能学报*, 2023, 44(12): 568-576.
CHEN Yuanfeng, MA Xiyuan, CHENG Kai, et al. Ultra-short-term power prediction of new energy based on meteorological feature selection and SVM model parameter optimization[J]. *Acta Energetica Solaris Sinica*, 2023, 44(12): 568-576.
- [11] 杨茂, 贾梦琦, 张薇, 等. 基于功率重构和时序特性约束的长预见期光伏集群功率预测[J]. *电力系统自动化*, 2024, 48(15): 102-111.
YANG Mao, JIA Mengqi, ZHANG Wei, et al. Power prediction of long foresight photovoltaic clusters based on power reconstruction and time series constraints[J]. *Automation of Electric Power Systems*, 2024, 48(15): 102-111.
- [12] YANG M, ZHAO M, HUANG D W, et al. A composite framework for photovoltaic day-ahead power prediction based on dual clustering of dynamic time warping distance and deep autoencoder[J]. *Renewable Energy*, 2022, 194: 659-673.
- [13] 杨开通, 刘甲云, 蒋瑞祥, 等. 基于深度学习的电网安全态势感知方法[J]. *电力大数据*, 2023(1): 69-75.
YANG Kaitong, LIU Jiayun, JIANG Ruixiang, et al. Method of power grid security situation awareness based on deep learning[J]. *Power Systems and Big Data*, 2023(1): 69-75.
- [14] KUMAR DUBEY A, KUMAR A, GARCÍA-DÍAZ V, et al. Study and analysis of SARIMA and LSTM in forecasting time series data[J]. *Sustainable Energy Technologies and Assessments*, 2021, 47: 101474.
- [15] 李庆生, 张裕, 龙家焕, 等. 基于 IMFO-LSTM 模型的光伏功率预测研究[J]. *电力科学与技术学报*, 2024, 39(3): 199-206.
LI Qingsheng, ZHANG Yu, LONG Jiahuan, et al. Photovoltaic power prediction based on IMFO-LSTM model[J]. *Journal of Electric Power Science and Technology*, 2024, 39(3): 199-206.
- [16] 王海军, 居蓉蓉, 董颖华. 基于时空关联特征与 B-LSTM 模型的分布式光伏功率区间预测[J]. *中国电力*, 2024, 57(7): 74-80.
WANG Haijun, JU Rongrong, DONG Yinghua. Distributed photovoltaic power interval prediction based on spatio-temporal correlation feature and B-LSTM model[J]. *Electric Power*, 2024, 57(7): 74-80.

- [17] 李滨, 高枫. 基于虚拟相似日与 DA-LSTPNet 的地区电网短期负荷预测[J]. 电力系统自动化, 2021, 45(22): 55-64.
LI Bin, GAO Feng. Short-term load forecasting of regional power grid based on virtual similar day and DA-LSTPNet[J]. Automation of Electric Power Systems, 2021, 45(22): 55-64.
- [18] 雷柯松, 吐松江·卡日, 伊力哈木·亚尔买买提, 等. 基于 WGAN-GP 和 CNN-LSTM-Attention 的短期光伏功率预测[J]. 电力系统保护与控制, 2023, 51(9): 108-118.
LEI Kesong, TUSONGJIANG·Kari, YILIHAMU·Yaermaimaiti, et al. Short-term photovoltaic power prediction based on WGAN-GP and CNN-LSTM-Attention[J]. Power System Protection and Control, 2023, 51(9): 108-118.
- [19] 张云钦, 程起泽, 蒋文杰, 等. 基于 EMD-PCA-LSTM 的光伏功率预测模型[J]. 太阳能学报, 2021, 42(9): 62-69.
ZHANG Yunqin, CHENG Qize, JIANG Wenjie, et al. Photovoltaic power prediction model based on EMD-PCA-LSTM[J]. Acta Energetica Solaris Sinica, 2021, 42(9): 62-69.
- [20] 滕陈源, 丁逸超, 张有兵, 等. 基于 VMD-Informer-BiLSTM 模型的超短期光伏功率预测[J]. 高电压技术, 2023, 49(7): 2961-2971.
TENG Chenyuan, DING Yichao, ZHANG Youbing, et al. Ultra-short-term photovoltaic power prediction based on VMD-Informer-BiLSTM model[J]. High Voltage Engineering, 2023, 49(7): 2961-2971.
- [21] QU J Q, QIAN Z, PEI Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern[J]. Energy, 2021, 232: 120996.
- [22] 朱继忠, 苗雨旺, 董朝阳, 等. 基于 Attention-LSTM 与多模型集成的短期负荷预测方法[J]. 电力工程技术, 2023, 42(5): 138-147.
ZHU Jizhong, MIAO Yuwang, DONG Zhaoyang, et al. Short-term load forecasting method based on Attention-LSTM and multi-model integration[J]. Electric Power Engineering Technology, 2023, 42(5): 138-147.
- [23] 吉铨格, 李慧, 刘思嘉, 等. 基于 MIE-LSTM 的短期光伏功率预测[J]. 电力系统保护与控制, 2020, 48(7): 50-57.
JI Xinge, LI Hui, LIU Sijia, et al. Short-term photovoltaic power prediction based on MIE-LSTM[J]. Power System Protection and Control, 2020, 48(7): 50-57.
- [24] 薛阳, 李金星, 杨江天, 等. 基于相似日分析和改进鲸鱼算法优化 LSTM 网络模型的光伏功率短期预测[J]. 南方电网技术, 2024, 18(11): 97-105.
XUE Yang, LI Jinxing, YANG Jiangtian, et al. Short-term forecasting of photovoltaic power based on similar day analysis and improved whale algorithm to optimize LSTM network model[J]. Southern Power System Technology, 2024, 18(11): 97-105.
- [25] ZHOU X, PANG C X, ZENG X H, et al. A short-term power prediction method based on temporal convolutional network in virtual power plant photovoltaic system[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 9003810.
- [26] 王开艳, 杜浩东, 贾嵘, 等. 基于相似日聚类和 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 day clustering and QR-CNN-BiLSTM model[J]. High Voltage Engineering, 2022, 48(11): 4372-4384.
- [27] 陈庆斌, 杨耿煌, 耿丽清, 等. 基于相似日选取和数据重构的短期光伏功率组合预测方法[J]. 中国电力, 2024, 57(12): 71-81.
CHEN Qingbin, YANG Genghuang, GENG Liqing, et al. Short term photovoltaic power combination prediction method based on similar day selection and data reconstruction[J]. Electric Power, 2024, 57(12): 71-81.
- [28] 吴家葆, 曾国辉, 张振华, 等. 基于 K-means 分层聚类的 TCN-GRU 和 LSTM 动态组合光伏短期功率预测[J]. 可再生能源, 2023, 41(8): 1015-1022.
WU Jiabao, ZENG Guohui, ZHANG Zhenhua, et al. Short-term power prediction of TCN-GRU and LSTM dynamic combination photovoltaic based on K-means hierarchical clustering[J]. Renewable Energy Resources, 2023, 41(8): 1015-1022.
- [29] 王晓霞, 俞敏, 冀明, 等. 基于气候相似性与 SSA-CNN-LSTM 的光伏功率组合预测[J]. 太阳能学报, 2023, 44(6): 275-283.
WANG Xiaoxia, YU Min, JI Ming, et al. Photovoltaic power combination forecasting based on climate similarity and SSA-CNN-LSTM[J]. Acta Energetica Solaris Sinica, 2023, 44(6): 275-283.
- [30] LUO X, ZHANG D X, ZHU X. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge[J]. Energy, 2021, 225: 120240.
- [31] 王涛, 王旭, 许野, 等. 计及相似日的 LSTM 光伏出力预测模型研究[J]. 太阳能学报, 2023, 44(8): 316-323.
WANG Tao, WANG Xu, XU Ye, et al. Study on LSTM photovoltaic output prediction model considering similar days[J]. Acta Energetica Solaris Sinica, 2023, 44(8): 316-323.
- [32] 唐冬来, 倪平波, 张捷, 等. 基于离散弗雷歇距离的户变关系识别方法[J]. 电力系统自动化, 2021, 45(6): 223-230.
TANG Donglai, NI Pingbo, ZHANG Jie, et al. Identification method of household variable relationship based on discrete Fréchet distance[J]. Automation of Electric Power Systems, 2021, 45(6): 223-230.
- [33] 周浩杰, 杨建卫, 王尊, 等. 基于 LSTM 光伏发电功率超短期预测模型研究[J]. 电源技术, 2023, 47(6): 785-789.
ZHOU Haojie, YANG Jianwei, WANG Zun, et al. Research on ultra-short-term forecasting model of photovoltaic power generation based on LSTM[J]. Chinese Journal of Power Sources, 2023, 47(6): 785-789.
- [34] Past weather in Alice Springs, Northern Territory, Australia:

- 2019-2020[EB/OL]. [2024-08-04]. <https://www.timeanddate.com>.
- [35] 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 photovoltaic power prediction based on MBI-PBI-ResNet[J]. *Power System Protection and Control*, 2024, 52(2): 165-176.
- [36] 汪繁荣, 梅涛, 卢璐. 基于相似日聚类 and VMD-LTWDBO-BiLSTM 的短期光伏功率预测[J]. 智慧电力, 2024, 52(10): 56-63,111.
WANG Fanrong, MEI Tao, LU Lu. Short-term PV power prediction based on similar day clustering with VMD-LTWDBO-BiLSTM[J]. *Smart Power*, 2024, 52(10): 56-63,111.
- [37] 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.

作者简介:



石宇龙

石宇龙(1999), 男, 硕士在读, 研究方向为新能源预测方法与控制技术(E-mail: yl_shi@stu.scu.edu.cn);

彭乔(1994), 女, 博士, 副研究员, 研究方向为新型电力系统频率稳定与控制、新能源与储能并网控制;

刘天琪(1962), 女, 博士, 教授, 研究方向为电力系统稳定与控制、高压直流输电、柔性直流电网。

Photovoltaic power short-term forecasting method based on similar days and bi-layer correction LSTM model

SHI Yulong¹, PENG Qiao¹, LIU Tianqi¹, CHEN Gang², ZENG Xueyang², LI Yan²

(1. College of Electrical Engineering, Sichuan University, Chengdu 610065, China;

2. State Grid Sichuan Electric Power Company Research Institute, Chengdu 610041, China)

Abstract: Accurate forecasting of photovoltaic (PV) power is essential for power system dispatch and decision-making. To enhance the prediction accuracy of PV power, a data-driven short-term forecasting method based on similar days and a bi-layer correction long short-term memory (LSTM) model is proposed. Firstly, both PV power and related meteorological data are normalized, and key factors influencing PV power are identified using the Pearson correlation coefficient, reducing the dimensions of the training dataset. Next, the Fréchet distance algorithm is applied to match similar days with the target prediction day, improving the quality of the training data. Then, based on numerical weather predictions, the initial PV power forecast is obtained through the baseline LSTM using feature learning. A correction LSTM, using a time series approach, predicts the error and adjusts the initial forecast to produce the final prediction. A case study uses real-world data under different weather conditions, i.e., sunny, cloudy, and rainy conditions. It shows that the proposed model consistently delivers accurate short-term PV power predictions for the next 24 hours. The model significantly improves accuracy compared to existing methods.

Keywords: photovoltaic system; short-term forecasting; similar days; long short-term memory (LSTM) network; bi-layer correction model; Fréchet distance algorithm

(编辑 方晶)