

双向数据扩充和 LSTNet 的户用光伏发电预测

王媛媛¹, 尹有鹏¹, 籍宏震¹, 张立志², 曹成军¹, 叶宇轩¹

(1.长沙理工大学 电网防灾减灾全国重点实验室, 湖南 长沙 410114; 2.国网湖南超高压变电公司, 湖南 长沙 410004)

摘要: 整县光伏政策促使小容量屋顶光伏急剧增长, 实现屋顶分布式光伏超短期发电功率的准确预测是分析海量细粒户用光伏电站对电力系统影响的前提。然而, 屋顶分布式光伏在原有波动性的基础上存在小容量、分散式、离线式经营的特点, 同时缺乏准确的气象数据, 使得光伏功率预测异常复杂。为此, 文章在有限数据下纵向地从光伏系统历史功率数据中搜索相似样本, 横向地收集相邻分布式光伏发电用户功率数据, 实现双向数据扩充, 在一定程度上克服了光伏发电预测对于一些关键输入特征的依赖; 在此基础上借助 LSTNet(Long-and-Short-term Time-series Network)神经网络的短期局部特征捕捉、长期时序信息强化、周期线性成分提取功能实现光伏功率预测。实验结果表明, 在缺乏重要辐照数据的情况下, 所提模型仍具有较好的预测精度。

关键词: 整县光伏; 光伏发电; 短期功率预测; 双向数据扩充; 神经网络

中图分类号: TK51 **文献标志码:** A **文章编号:** 1671-5292(2025)01-0045-09

0 引言

光伏作为清洁能源的主力军, 在光伏整县开发政策^[1]的大力推进下, 屋顶分布式光伏发电系统得到快速发展。然而, 小容量屋顶分布式光伏密集、大量接入后, 累计发电量不容忽视, 其发电的随机性和波动性将给电网的安全、稳定运行带来挑战。准确预测光伏发电功率的变化趋势有助于提高电网运行的稳定性, 促进新能源消纳。

现有光伏发电功率预测^[2]大致可分为物理模型、统计分析模型和人工智能模型。物理模型根据站内光伏倾角、方位角等参数, 结合详细气象特征数据, 建立映射模型进行预测^[3,4]。统计分析模型训练相对简易, 但依赖输入数据的质量和完整性, 抗干扰能力差^[5,6]。目前较多学者倾向于人工智能模型, 通过神经网络挖掘序列变化特性, 进而实现光伏功率预测, 研究重点集中在处理光伏数据与提升预测精度两方面^[7]。

为提高预测精度, 大量文献均进行了光伏数据扩充处理。文献[8]使用相似日搜索方法, 计算最小欧氏距离, 进行数据维度扩充。文献[9]为扩充样本规模, 考虑相关度权重, 选取待预测日对应的相似日。文献[10]采用统计性指标进行聚类分析。文献[11]则通过聚类算法修正相邻光伏电站

气象数据缺失。一些学者还更加细致地从大气环境、太阳辐射等方面展开深入研究。文献[12]通过提取云层图像特征, 结合云层运动轨迹与辐射数据进行发电功率预测。文献[13]考虑云团状态对太阳辐射的影响, 构建映射模型进行光伏功率预测。上述方法利用卫星数据, 有效提升了预测性能, 对数据测量提出了更高的要求。

筛选和改进预测算法是提升预测精度另一个研究方向。常用自回归、多元回归等模型进行时间序列预测^[14,15], 具有原理简单、快速建模的优点, 但当样本数据过大时预测精度有限。得益于光伏电站数据精确性的提高和测量站点的广泛铺设, 使用神经网络模型^[16,17]进行光伏发电预测成为热点。文献[18]将循环神经网络(Recurrent Neural Network, RNN)应用于光伏短期功率预测。文献[19]使用长、短期记忆网络(Long Short Term Memory, LSTM)进行预测, 相较于传统预测模型效果更好。文献[20]添加注意力机制, 使模型在长时间序列输入中关注重点信息。在此基础上, 文献[21]利用双向长、短期记忆网络(Bi-directional Long Short Term Memory, Bi-LSTM)挖掘时序数据的前后联系, 有效提高了光伏功率的预测精度。

综上, 大型光伏电站发电功率预测的研究方

收稿日期: 2023-10-25。

基金项目: 国家自然科学基金项目(52177069); 湖南省研究生科研创新项目(QL20220198)。

作者简介: 王媛媛(1980-), 女, 博士, 教授, 研究方向为电力系统继电保护、新能源发电。E-mail: wyy_1202@163.com

向是数据要求更多、模型构建更准、预测精度更高。而户用分布式光伏存在地理空间上的密集性、气象特征上的一致性、数据维度上的稀疏性;用户一般不愿增加投资而增设数据测量装置,无法为光伏功率预测提供有效、可靠的数据来源;并且户用分布式光伏发电还存在住户人为因素干扰,数据缺失问题更为严重。因此,须要针对户用分布式光伏数据集特性,改进功率预测方法,为其大规模接入电力系统提供技术支撑。本文考虑长时气象关联因素,纵向地进行相似日搜索,横向地收集相邻分布式光伏发电用户功率数据,实现双向数据扩充;在此基础上使用长、短时神经网络(Long-and Short-term Time-series Network, LSTNet)^[22]进行户用分布式光伏发电功率预测;采用维多利亚州拉筹伯大学屋顶光伏数据集进行实例验证。实验结果表明,所提模型不依赖详细的气象数据,预测效果较好。

1 历史光伏数据双向扩充方法

为解决住宅式光伏数据气象特征有限、数据缺失较多、稀疏特性明显的问题,本文从光伏历史数据相似日纵向扩充和光伏相邻用户发电数据横向扩充两个角度,在 2 个维度上扩充数据。

1.1 考虑气象关联因素的相似日纵向数据扩充

首先,分析历史数据中影响分布式光伏发电量的气象因素,根据全天气象数据的相似性选取历史相似日,将相似日的发电功率作为扩充数据,实现纵向历史数据扩展。构造气象特征矩阵 G 为

$$G = \begin{bmatrix} t_1 & e_1 & d_1 & h_1 & r_1 & w_1 \\ t_2 & e_2 & d_2 & h_2 & r_2 & w_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{96} & e_{96} & d_{96} & h_{96} & r_{96} & w_{96} \end{bmatrix} \quad (1)$$

式中: t, e, d, h, r, w 分别为历史气象数据全天的体感温度、空气温度、露点温度、相对湿度、风速、风向。

相似日搜索^[23]方法是搜索与待预测日的气象特征矩阵 G 欧氏距离最小的历史日作为相似日。利用两个特征矩阵之差的欧几里得范数衡量相似程度,计算式为

$$d_{ij} = \| G_i - G_j \|_2 = \left[\sum_{j=1}^N \sum_{t=1}^{96} (g_{it} - g_{jt})^2 \right]^{\frac{1}{2}} \quad (2)$$

式中: G_i 为待预测日气象特征矩阵; G_j 为历史气象特征矩阵; g_{it} 为待预测日 i 在 t 时刻的气象特

征; g_{jt} 为历史日 j 在 t 时刻的气象特征。

1.2 基于相邻分布式光伏发电用户的横向数据扩充

在整县光伏推进政策的助力下,海量细粒的小容量分布式光伏安装到学校、政府、居民屋顶,这些分布式光伏系统地理位置极其相近,发电特性受天气影响基本一致。

县级屋顶光伏的布设一般由少数几个光伏生产商承包,这就导致接入的小容量分布式光伏发电系统不仅容量区间相近,甚至品牌型号一致。根据先验知识,光伏发电与地面辐射存在高度相关性,相邻分布式光伏用户历史发电数据一定程度上也蕴含着该区域历史辐射数据。因此,本文依据容量、型号等铭牌数据,横向选择相邻分布式光伏发电用户的历史发电数据作为待预测分布光伏发电输入特征矩阵的一部分,实现输入特征矩阵数据扩充。

2 光伏出力影响因素分析

充分挖掘各种因素与光伏发电出力之间的关联性是准确预测光伏功率的前提。本文利用皮尔逊相关系数 $r_{x,y}$ ^[24]进行相关性分析,计算式为

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

式中: x_i 和 y_i 分别为第 i 个样本中特征 x 和 y 的数值; \bar{x} 和 \bar{y} 分别为 n 个样本中特征 x 和 y 的平均值。

为了深入探究相关性,分析气象因素与光伏发电出力之间的关系,并进一步提高光伏发电系统建模的精确度和有效性,本文利用澳大利亚维多利亚州拉筹伯大学屋顶光伏发电数据集进行分析。不同特征与光伏功率间的皮尔逊相关系数见表 1。

表 1 光伏功率与影响因素之间的皮尔逊相关系数
Table 1 Pearson correlation coefficient between PV power and influencing factors

影响因素	皮尔逊 相关系数	影响因素	皮尔逊 相关系数
表观温度	0.42	空气温度	0.52
风向	0.017	风速	0.31
相对湿度	0.57	露点温度	0.003 4
相似日光伏功率	0.81	相邻光伏功率	0.99

表 1 中的相关系数是进行了正向化处理后的结果,当相关系数的绝对值越趋近于 1 时,表示相关性越强,反之则越弱。由表 1 可知,风向和露点温度与光伏功率相关性较差,而相似日光伏功率、相邻光伏功率与光伏功率的相关性较高。由此也可以验证从历史数据中搜索得到相似日数据以及相邻用户数据作为模型的特征输入数据的合理性。本文将与光伏功率相关性较高的表观温度、空气温度、风速、相对湿度、相似日光伏功率、相邻光伏功率作为预测模型的多维输入特征。

3 LSTNet 光伏发电功率预测模型

针对住宅分布式光伏发电功率具有不确定性、周期性和自回归性等特点,本文使用 LSTNet 光伏发电预测模型,该模型网络结构如图 1 所示。

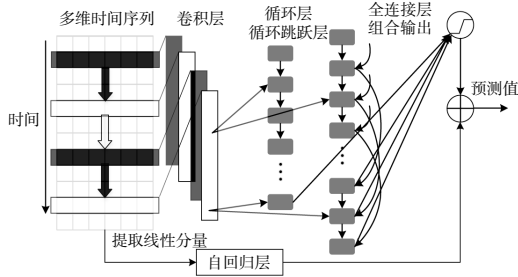


图 1 LSTNet 神经网络
Fig.1 LSTNet neural network

由图 1 可见,LSTNet 神经网络由卷积层、循环层、循环跳跃层以及自回归层组成。使用循环层捕捉气象数据短期局部特征;使用循环跳跃层把握长期时序信息,反映光伏发电功率随着太阳辐射季节性变化在长时间尺度上的变化特性;利用自回归层提取光伏发电功率序列的线性成分;最后结合线性和非线性模式运算,将全连接层输出与传统自回归模型的输出相加得到最终预测结果,解决了传统模式的尺度不敏感问题,模型预测性能更加准确可靠。

3.1 卷积单元

卷积神经网络 (Convolutional Neural Networks, CNN)作为深度学习的典型代表算法之一,是一种包含卷积运算和具有深度解析结构的前馈型神经网络。CNN 根据生物的视觉机制构建,核心特性是通过局部捕捉和权值共享大大减少网络参数的数量,并缓解模型过拟合问题。目前,CNN 也被广泛用于时间序列特征处理^[10],在提高模型时间序列的预测性能方面表现出强大优

势。LSTNet 使用一个没有池化层的卷积神经网络来获取短期时间序列和特征变量之间的局部依赖关系。卷积层中使用的卷积核会对多维时间序列进行计算,即:

$$h_k = \text{RELU}(w_k * X + b_k) \quad (4)$$

式中: h_k 为输出矢量;RELU(x)为激活函数; w_k 为卷积核权重矩阵;*表示卷积运算; X 为输入的多维时间序列特征矩阵; b_k 为偏置权重。

3.2 循环层

将卷积层的输出作为循环层和循环跳跃层输入,LSTNet 模型在循环层中使用 LSTM 作为循环单元,与传统的循环神经网络 RNN 相比,LSTM 设计了遗忘门、输入门、输出门和储存单元来实现对过去信息的去除和添加,解决了传统 RNN 的长期时间序列依赖性问题的。输入门可以选取输入到储存单元的信息,遗忘门可以从记忆单元中删除“遗忘”的信息,而输出门可以控制传递到下一个 LSTM 单元的隐藏状态信息。LSTM 的隐藏状态计算过程为

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (6)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (7)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (8)$$

$$h_t = o_t \tanh(c_t) \quad (9)$$

式中: i_t, f_t, c_t, o_t 分别为输入门、遗忘门、储存单元和输出门的值; b_i, b_f, b_c 和 b_o 分别为其相应的偏差; W_x 为输入节点和隐藏节点之间的权重; W_h 为隐藏层和记忆单元之间的权重; W_c 为记忆单元到输出节点之间的权重。

3.3 循环跳跃层

LSTM 可以通过“记忆”和“遗忘”来捕捉长期时间序列的依赖关系,但是随着时间序列长度的增加,梯度会逐渐减小甚至消失,LSTM 捕捉超长期时间序列依赖关系的能力将急剧下降甚至消失。因此须要加入循环跳跃层来提高网络捕捉超长期依赖关系的能力。这种循环跳跃层更适合于光伏发电功率等具有稳定周期和时间序列问题的预测。循环跳跃层的更新过程为

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-p} + W_{ci}c_{t-p} + b_i) \quad (10)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-p} + W_{cf}c_{t-p} + b_f) \quad (11)$$

$$c_t = f_t c_{t-p} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-p} + b_c) \quad (12)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-p} + W_{co}c_t + b_o) \quad (13)$$

$$h_t = o_t \tanh(c_t) \quad (14)$$

式中: 参数 p 为时间序列的平均周期长度; c_{t-p} 为记忆单元在 p 时间步长之前的值; h_{t-p} 为隐藏层在 p 时间步长之前的状态。

在 LSTNet 神经网络末尾需要一个全连接层来整合循环层和循环跳跃层的隐状态输出, 其计算式为

$$h_t^D = W^R h_t^R + \sum_{i=0}^{p-1} W_i^S h_{t-i}^S + b \quad (15)$$

式中: h_t^D 为输出; W^R 和 W_i^S 为待学习的权重; h_t^R 为循环层的隐状态; h_{t-i}^S 为循环跳跃层的隐状态; b 为待学习的偏置。

3.4 自回归层

卷积层和循环层都是完全非线性的深度神经网络, 对于输入特征的尺度敏感性较差, 降低了神经网络对时间序列中非周期性变化数据的预测精度, 因此在 LSTNet 中加入了一个自回归 (Auto Agressive, AR) 模型来预测序列中的线性部分, AR 可以被看作是一个多元线性回归模型, 计算式为

$$h_{t,i}^L = \sum_{k=0}^{q^{AR}-1} W_k^{AR} y_{t-k,i} + b^{AR} \quad (16)$$

式中: $h_{t,i}^L$ 为 AR 层的输出; $y_{t-k,i}$ 为 AR 层的输入; W_k^{AR} 和 b^{AR} 分别为 AR 层自回归系数和偏差。

LSTNet 的最终输出是全连接层和 AR 层的输出之和, 即:

$$\hat{Y}_t = h_t^D + h_t^L \quad (17)$$

式中: \hat{Y}_t 为 t 时刻的最终预测结果。

3.5 短期发电功率预测模型

本文提出的结合双向历史数据扩充和 LSTNet 的户用光伏发电功率预测流程如图 2 所示。

具体过程如下。

①对原始户用光伏数据集进行预处理。通过相似日算法搜索发电单元本身历史功率数据, 实现纵向数据扩充; 搜索相邻户用光伏历史功率数据, 实现横向数据扩充。构建双向扩充的户用光伏数据集。

②设置循环层、循环跳跃层、自回归层, 构建

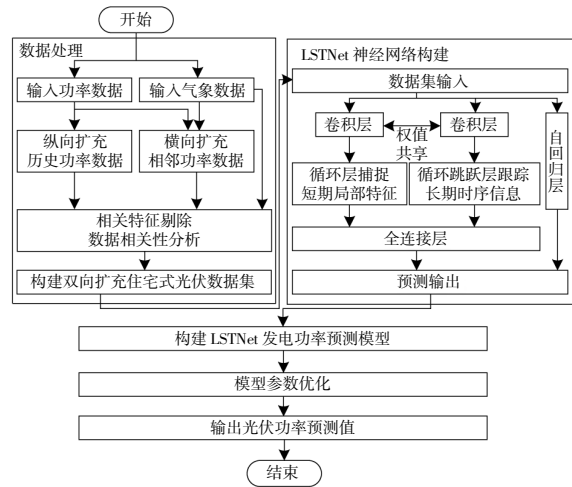


图 2 LSTNet 分布式光伏短期功率预测流程
Fig.2 LSTNet distributed PV short-term power prediction process

LSTNet 神经网络进行发电功率预测。对模型参数进行寻优, 确定最优户用光伏发电功率预测模型, 最后得到光伏功率预测值。

4 实例分析

本文使用澳大利亚维多利亚州拉筹伯大学屋顶光伏发电数据集作为研究对象, 选取了 2020 年 6 月 12 日-2021 年 6 月 29 日该校区屋顶分布式光伏发电系统的实测功率数据和气象数据, 数据采样周期为 15 min。

4.1 数据预处理

①数据清洗。采用滑动窗口对数据集中的奇异值和缺失值进行辨识, 使用 3 次样条插值填补时间序列中的缺失值, 修正奇异值。

②数据集划分。将双向扩充的发电功率数据集作为模型输入的训练集, 挑选春季、夏季、秋季、冬季 4 个季节各一个月的实测数据, 将其作为测试集。

③数据归一化。对原始数据 X 进行归一化处理, 将数据大小约束在 $[0, 1]$ 内, 归一化公式为

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (18)$$

式中: X^* 为归一化结果; X_{\max} 和 X_{\min} 分别为数据集中的最大值和最小值。

4.2 模型性能评价指标

考虑到光伏功率曲线具有明显的日周期性, 借助均方根误差 (Root Mean Square Error, RMSE) 反映预测误差的分散性, 借助平均绝对值

误差(Mean Absolute Error,MAE)反映实际预测误差的大小,借助平均绝对百分比误差(Mean Absolute Percentage Error,MAPE)反映预测值与实际值之间的差异。本文选择RMSE,MAE和MAPE作为光伏功率预测模型的评价指标,其公式分别为

$$\lambda_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (19)$$

$$\lambda_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (20)$$

$$\lambda_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{\hat{y}_i} \right| \times 100\% \quad (21)$$

式中: N 为样本总数; \hat{y}_i 和 y_i 分别为第 i 时刻的预测值和实际值。

4.3 不同季节预测结果及分析

为验证模型在长时间序列下的预测性能,对于相应季节内一个月的分布式光伏进行发电功率预测。本文选取了3种方法进行比较分析,包括纵向扩充+LSTNet、横向扩充+LSTNet、双向扩充+LSTNet。4个季节下模型性能评价指标RMSE,MAE和MAPE情况如表2所示。

根据表2可知,纵向扩充+LSTNet方法整体的预测误差都较大。这是因为户用光伏数据集本

表2 4个季节不同模型的预测误差

Table 2 Predictive performance of different models for the four seasons

季节	预测模型	RMSE/kW	MAE/kW	MAPE/%
春季	纵向扩充+LSTNet	3.023 3	2.295 1	88.67
	横向扩充+LSTNet	0.481 3	0.300 5	12.54
	双向扩充+LSTNet	0.493 1	0.298 4	9.52
夏季	纵向扩充+LSTNet	2.054 2	1.506 9	64.23
	横向扩充+LSTNet	0.481 8	0.314 8	9.54
	双向扩充+LSTNet	0.464 0	0.270 1	7.63
秋季	纵向扩充+LSTNet	1.749 6	1.772 2	78.73
	横向扩充+LSTNet	0.350 3	0.215 3	10.21
	双向扩充+LSTNet	0.362 2	0.212 0	7.32
冬季	纵向扩充+LSTNet	1.836 8	1.388 5	69.54
	横向扩充+LSTNet	0.441 2	0.235 4	12.12
	双向扩充+LSTNet	0.437 2	0.214 1	10.20

身缺少关键的地面辐射数据,仅依赖相似日搜索以选择历史相似发电数据进行光伏功率预测时,分布式光伏短期功率的预测精度有限。在此情况下,横向扩充+LSTNet方法、双向扩充+LSTNet方法在春、夏、秋、冬4个季节的长时间序列的预测性能较为稳定,其中双向扩充+LSTNet方法表现出更好的预测效果。

各方法在秋季某月的部分预测曲线(取白天采样点预测数据)如图3所示。

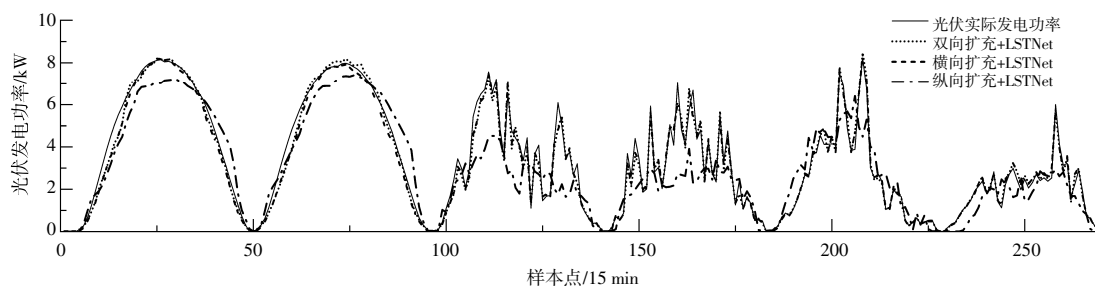


图3 各方法在秋季某月的预测情况

Fig.3 Forecast results between different methods in a certain autumn month

由图3可见:纵向扩充+LSTNet方法应用于户用分布式光伏功率预测精度较差,这表明常规光伏预测方法用于稀疏性较强、维度有限的户用光伏数据时存在局限,无法有效预测发电功率;横向扩充+LSTNet方法的预测值较实际值偏低,同时在接近日出、日落等时段预测精度较低;在实际光伏发电功率突变时段,双向扩充+LSTNet方法能较好预测发电功率“尖峰”,预测曲线与光伏实际发电功率曲线无论是在功率较为平缓还是急剧变化的情况下,两者整体上更为贴近。预测结果进

一步说明双向扩充+LSTNet方法在针对户用分布式光伏数据集的预测方面具有优越性。

4.4 典型天气预测结果及分析

将测试集按天气类型分为晴天、多云、阴天、雨天4种典型天气,从典型天气类型的角度分析双向扩充+LSTNet、双向扩充+LSTM、双向扩充+支持向量机(Support Vector Machine,SVM)方法的预测效果。

晴天预测结果如图4所示。

在晴天时,由于光照强烈,太阳辐射变化稳定,

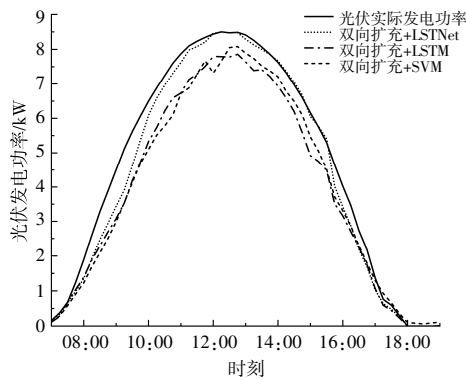


图4 晴天预测结果(2020年9月7日)

Fig.4 Sunny day forecast results (September 7, 2020)

光伏实际发电功率普遍较大且功率曲线波动性较小,曲线整体呈倒U字形。在08:00-10:00,3种模型预测结果都偏保守;双向扩充+LSTMNet方法和双向扩充+SVM方法的预测结果相近,但较实际光伏发电功率值偏小;双向扩充+LSTMNet方法整体更接近实际光伏发电功率,预测性能更好。

多云天预测结果如图5所示。

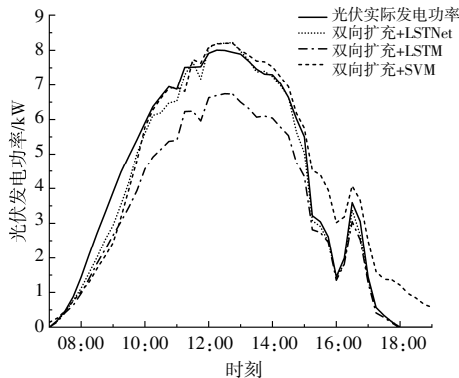


图5 多云天预测结果(2021年4月8日)

Fig.5 Cloudy day forecast results (April 8, 2021)

在多云时,虽然光照较强,但光伏实际发电功率可能会受到云层的干扰而出现一定程度上的波动。在08:00-15:00,双向扩充+LSTMNet方法略优于双向扩充+SVM方法;在之后的时间段,光伏实际发电功率出现急剧跌落,双向扩充+SVM方法无法很好地拟合实际发电功率曲线;而双向扩充+LSTM方法在日落时段预测性能较好,但整体预测效果较差,在10:00-15:00高发时段并不能准确预测光伏发电功率。总体而言,双向扩充+LSTMNet方法在整体预测效果和预测曲线某些局部突变时综合预测性能更好。

阴天预测结果如图6所示。

在阴天时,由于分布式光伏接收到的光照整

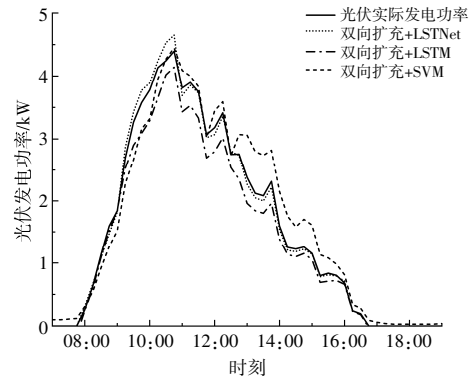


图6 阴天预测结果(2020年7月24日)

Fig.6 Overcast day forecast results (July 24, 2020)

体较弱,光伏实际发电功率曲线随时间变化较为稳定,某些时刻存在较小的功率曲线突变。双向扩充+LSTMNet方法整体的预测曲线能够较好贴合光伏实际发电功率曲线的“突变”点,10:00-11:00时刻对光伏发电功率峰值的预测有所欠缺;双向扩充+LSTM方法和双向扩充+SVM方法预测结果整体偏大或偏小;双向扩充+SVM方法预测日出和日落时刻发电功率还存在局限。

雨天预测结果如图7所示。

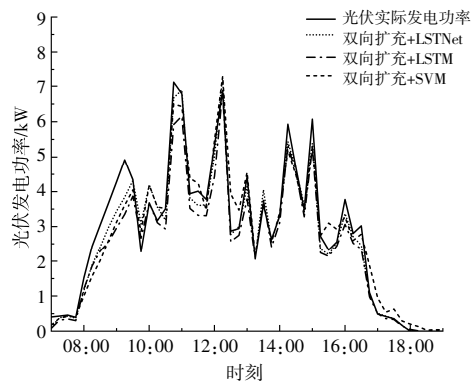


图7 雨天预测结果(2020年9月5日)

Fig.7 Rainy day forecast results (September 5, 2020)

在雨天时,由于受到阵雨影响,分布式光伏接收到的光照波动较大,因此光伏发电功率曲线呈现短时间内急剧突变的情况,整体发电功率较低。在雨天时,3种模型的预测效果比较接近,且针对光伏实际发电功率曲线的“尖峰”位置都不能较好地预测。整体来看,双向扩充+LSTMNet模型在雨天时更接近真实值,预测效果更好。

利用式(19)~(21)给出的3种评价指标对预测模型的综合预测误差进行分析。4种典型天气类型下3种模型预测误差见表3。

由表3可知,在4种典型天气类型下,双向扩

表 3 4 种典型天气类型下各模型预测误差情况
Table 3 Differences in prediction errors among different models under four typical weather types

天气类型	预测模型	RMSE/kW	MAE/kW	MAPE/%
晴天	双向扩充+LSTNet	0.385 8	0.255 1	7.81
	双向扩充+LSTM	0.855 0	0.698 2	15.40
	双向扩充+SVM	0.785 2	0.648 0	18.31
多云	双向扩充+LSTNet	0.520 6	0.315 8	9.10
	双向扩充+LSTM	0.867 7	0.612 7	14.18
	双向扩充+SVM	0.776 0	0.595 1	20.85
阴天	双向扩充+LSTNet	0.451 6	0.256 2	8.73
	双向扩充+LSTM	0.719 5	0.472 7	13.86
	双向扩充+SVM	0.683 8	0.506 0	22.90
雨天	双向扩充+LSTNet	0.430 0	0.236 2	8.79
	双向扩充+LSTM	0.626 5	0.388 3	13.89
	双向扩充+SVM	0.661 3	0.441 6	22.81

充+LSTNet 模型的预测效果最佳,其对晴天的功率预测精度最高、对多云的功率预测精度最低。在晴朗天气下,综合对比其他模型,双向扩充+LSTNet 模型在预测精度的提升效果上更为显著,双向扩充+LSTNet 模型与双向扩充+LSTM、双向扩充+SVM 对比, RMSE 分别减少了 0.469 2, 0.399 4 kW, MAE 分别减小了 0.443 1, 0.392 9 kW, MAPE 预测精度分别提高了 7.59%, 10.5%。在多云天气下,双向扩充+LSTNet 模型与双向扩充+LSTM、双向扩充+SVM 对比, RMSE 分别减少了 0.347 1, 0.255 4 kW, MAE 分别减小了 0.296 9, 0.279 3 kW, MAPE 预测精度分别提高了 5.08%, 11.75%。综合来看,无论在何种天气类型下,本文提出的双向扩充+LSTNet 模型均有着较好的预测结果和更高的预测精度。

5 结论

本文针对小容量户用光伏数据特征维度较少,稀疏性较强的特点,提出了一种基于双向扩充的 LSTNet 短期光伏功率预测方法。与其他方法相比,得到以下结论。

①仅扩充常规相似日数据集无法有效提升小容量分布式光伏预测效果,本文进一步增扩了相邻分布式光伏用户数据,构建了纵向+横向的双向功率预测数据集。所提出的双向扩充方法在样本维度上进一步丰富了特征数据集,一定程度克服了光伏发电预测对于总辐照度、散射辐照度等关键输入特征的依赖。皮尔逊相关因素分析结果表明,光伏实际发电功率与纵向的相似日光伏功

率和横向的相邻光伏功率相关系数较高,分别为 0.81 和 0.99。

②LSTNet 模型在处理户用光伏稀疏数据集时具有明显优势。循环层能够及时捕捉数据集短期局部特征,保证了户用光伏功率序列在某些突变时刻的预测精度;循环跳跃层强化了光伏功率序列随季节变化的长时间尺度变化特性;自回归层则提取了光伏功率序列的线性成分,使模型预测性能更加稳定可靠。在面临户用光伏数据集稀疏的情况下,所提模型仍具有较好的预测性能。

③在不同季节以及不同天气类型下,对不同数据扩充方法进行了比较。结果表明,双向扩充较横向、纵向扩充的数据处理方法对预测模型精度的提升更好。所提出的双向扩充+LSTNet 光伏功率预测方法,在不同季节下,针对长时间序列的预测效果也较为稳定;在不同天气类型下,也能够较好地拟合户用光伏在局部突变时刻以及多“尖峰”突变时刻的发电功率情况。在预测误差上,均优于长、短期记忆神经网络模型和支持向量机模型。本文方法尤其适用于在光伏数据存在明显稀疏特征的情况下,预测并网的海量细粒离散分布式光伏电站的发电功率。

参考文献:

- [1] 陈娟,高江梅.基于演化博弈的户用光伏整县开发协同机制及仿真[J].电网技术,2023,47(2):669-684.
- [2] 赖昌伟,黎静华,陈博,等.光伏发电出力预测技术研究综述[J].电工技术学报,2019,34(6):1201-1217.
- [3] 魏嫖嫖,崔承刚,杨宁,等.光伏发电功率预测斜面辐射组合模型的评估研究[J].可再生能源,2018,36(6):842-849.
- [4] Scolari E, Reyes-Chamorro L, Sossan F, et al. A comprehensive assessment of the short-term uncertainty of grid-connected PV systems [J].IEEE Transactions on Sustainable Energy, 2018, 9(3): 1458-1467.
- [5] 李国庆,李欣彤,边竞,等.基于每时晴空指数的大规模光伏电站出力多维时间序列模拟 [J]. 电网技术, 2020,44(9):3254-3262.
- [6] 路志英,任一墨,葛路琨.基于样条估计分位数回归的光伏功率回归模型[J].湖南大学学报(自然科学版), 2017,44(10):91-98.
- [7] 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):3027-3048.

- [8] 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.
- [9] 郑若楠,李国杰,韩蓓,等.基于加权扩展日特征矩阵的分布式光伏发电日前功率预测[J].电力自动化设备, 2022, 42(2): 99-105.
- [10] 吕伟杰,方一帆,程泽.基于模糊 C 均值聚类和样本加权卷积神经网络的日前光伏出力预测研究[J].电网技术, 2022, 46(1): 231-238.
- [11] 王晓霞,俞敏,霍泽健,等.基于近邻传播聚类与 LST-Net 的分布式光伏电站群短期功率预测[J].电力系统自动化, 2023, 47(6): 133-141.
- [12] 柴闵康,夏飞,张浩,等.基于云图特征自识别的光伏超短期预测模型 [J]. 电网技术, 2021, 45(3): 1023-1035.
- [13] Si Z, Yu Y, Yang M, et al. Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks [J]. IEEE Transactions on Industry Applications, 2020, 57(1): 5-16.
- [14] 许彪,徐青山,黄煜,等.基于藤 copula 分位数回归的光伏功率日前概率预测[J].电网技术, 2021, 45(11): 4426-4435.
- [15] Wen Y, AlHakeem D, Mandal P, et al. Performance evaluation of probabilistic methods based on bootstrap and quantile regression to quantify PV power point forecast uncertainty [J]. IEEE transactions on neural networks and learning systems, 2019, 31(4): 1134-1144.
- [16] 侯慧,王晴,赵波,等.关键信息缺失下基于相空间重构及机器学习的电力负荷预测[J].电力系统保护与控制, 2022, 50(4): 75-82.
- [17] Jia Z, Yang L, Zhang Z, et al. Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring [J]. International Journal of Electrical Power & Energy Systems, 2021, 129: 106837.
- [18] 姚宏民,杜欣慧,秦文萍.基于密度峰值聚类及 GRNN 神经网络的光伏发电功率预测方法[J].太阳能学报, 2020, 41(9): 184-190.
- [19] Wang Y, Shen Y, Mao S, et al. LASSO and LSTM integrated temporal model for short-term solar intensity forecasting [J]. IEEE Internet of Things Journal, 2018, 6(2): 2933-2944.
- [20] 马磊,黄伟,李克成,等.基于 Attention-LSTM 的光伏超短期功率预测模型 [J]. 电测与仪表, 2021, 58(2): 146-152.
- [21] 王金玉,金宏哲,王海生,等. ISSA 优化 Attention 双向 LSTM 的短期电力负荷预测[J].电力系统及其自动化学报, 2022, 34(5): 111-117.
- [22] Lai G, Chang W C, Yang Y, et al. Modeling long- and short-term temporal patterns with deep neural networks [A]. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Ann Arbor: Association for Computing Machinery, 2018.
- [23] 陈通,孙国强,卫志农,等.基于相似日和 CAPSO-SNN 的光伏发电功率预测 [J]. 电力自动化设备, 2017, 37(3): 66-71.
- [24] 纪德洋,金锋,冬雷,等.基于皮尔逊相关系数的光伏电站数据修复 [J]. 中国电机工程学报, 2022, 42(4): 1514-1523.

Two-dimensional data expansion and LSTNet for residential PV generation forecasting

Wang Yuanyuan¹, Yin Youpeng¹, Ji Hongzhen¹, Zhang Lizhi², Cao Chengjun¹, Ye Yuxuan¹
(1.State Key Laboratory of Disaster Prevention & Reduction for Power Grid, Changsha University of Science & Technology, Changsha 410114, China; 2.State Grid Hunan Extra High Voltage Substation Company, Changsha 410004, China)

Abstract: China's "Whole County PV" programme has been dramatically expanding the use of solar power in rural areas, by building on government, commercial, industrial and residential rooftops. However, a large number of dispersed residential PV will have an impact on the power system, and accurately predicting the short-term power generation of residential PV is a prerequisite for addressing the impact. However, in addition to its original volatility, residential rooftop PV also has the characteristics of small capacity, decentralized and offline operation, together with the lack of accurate meteorological data, making PV power prediction exceptionally complex. Therefore, under the limited data, this paper longitudinally detects similar samples from the previous power data of the residential PV to be predicted, and horizontally collects similar samples from the power data of neighboring residential PV, ultimately jointly realizing two-dimensional data expansion, which overcomes the dependence of PV power generation prediction on some key input features to a certain extent. And then a residential PV generation prediction method is proposed based on LSTNet neural network, which has the functions of short-term local features capture, long-term time series information reinforcement, and cyclical linear component extraction.

Keywords: whole county PV; PV generation; short-term power prediction; two-dimensional data expansion; neural network