

基于气象因素充分挖掘的 BiLSTM 光伏发电短期功率预测^①



徐先峰, 刘阿慧, 陈雨露, 蔡路路

(长安大学 电子与控制工程学院, 西安 710064)

通讯作者: 徐先峰, E-mail: xxf_chd@163.com

摘要: 传统光伏发电功率预测存在因气象因素特征提取不综合不精确而导致预测精度不高的问题。为了充分挖掘气象因素对光伏出力的影响, 并有效利用深度学习技术在非线性拟合方面的优势, 本文提出了一种基于气象因素充分挖掘的双向长短期记忆 (Bi-directional Long Short Term Memory, BiLSTM) 网络光伏发电短期功率预测方法。在对原始数据进行异常值及标准化处理的基础上, 采用 K 近邻算法 (K-Nearest Neighbor, KNN) 在外界温度、湿度、压强等诸多气象因素中充分挖掘影响光伏出力的关键因素, 重构多元数据序列, 并在探索输入层时间步长、模型层数及每层维数等超参数的合理设置方案的基础上, 构建 BiLSTM 网络模型, 实现光伏发电短期功率的高精度预测。仿真结果表明, 与 KNN、深度信念网络 (DBN)、BiLSTM、PCA-LSTM 等经典方法比较, 所提 KNN-BiLSTM 方法具有更高的预测精度。

关键词: 光伏发电预测; 气象因素; 特征提取; K 近邻算法; 双向长短期记忆网络

引用格式: 徐先峰, 刘阿慧, 陈雨露, 蔡路路. 基于气象因素充分挖掘的 BiLSTM 光伏发电短期功率预测. 计算机系统应用, 2020, 29(7): 205–211.
<http://www.c-s-a.org.cn/1003-3254/7529.html>

BiLSTM Short-Term Forecasting Method for Photovoltaic Power Generation Based on Fully Exploiting Meteorological Factors

XU Xian-Feng, LIU A-Hui, CHEN Yu-Lu, CAI Lu-Lu

(School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China)

Abstract: The traditional PV power generation prediction has the problem that the prediction accuracy is not high due to incomplete and inaccurate feature extraction of meteorological factors. In order to fully explore the influence of meteorological factors on PV output and effectively utilize the advantage of deep learning technology in non-linear fitting, this study proposes a short-term forecasting method for PV power that is based on the full mining of meteorological factors and is realized through the BiLSTM network. Based on outliers and standardized processing of the original data, KNN is used to fully explore the key factors affecting PV output among meteorological factors such as external temperature, humidity, and pressure. And then multivariate data sequences are reconstructed. On the basis of exploring the reasonable setting scheme of the hyper parameters such as the time steps of the input layer, the number of model layers, and the dimensions of each layer, a BiLSTM network model is built to realize the high-precision prediction of short-term power of PV power generation. Simulation results show that the proposed KNN-BiLSTM method has higher prediction accuracy than the classical methods such as KNN, DBN, BiLSTM, and PCA-LSTM.

Key words: photovoltaic generation prediction; meteorological factors; feature extraction; K-Nearest Neighbor (KNN); bi-directional long short term memory network

① 基金项目: 长安大学研究生科研创新实践项目 (300103703041)

Foundation item: Graduate Research Innovation Practice Project of Chang'an University (300103703041)

收稿时间: 2019-12-24; 修改时间: 2020-01-20; 采用时间: 2020-02-11; csa 在线出版时间: 2020-07-03

引言

太阳能作为环境友好型的可再生能源,应用前景一片明朗^[1].因此,利用太阳能发电的光伏技术已成为我国应对世界能源危机和经济发展新形势的战略新兴产业.据统计,2018年,光伏发电跃居成为我国第四大主力电源,共发电1775亿千瓦时,同比增长50%,居世界首位^[2].然而,光伏发电的输出功率受多种气象因素影响,表现出明显的间歇性、随机性和波动性,进而严重影响与其连接的电网稳定性^[3].为确保光伏功率预测系统的预测精度,有效的光伏发电功率预测方法在电力生产方面占有举足轻重的地位^[4].

现存的光伏发电功率预测方法,总体可归纳为两类^[5]:基于物理模型的间接预测法和基于历史数据的直接预测法.后者因无需预测光照幅度,建模简单、预测成本较低,而得到大量应用.主要包括多元线性回归^[6]、支持向量机^[7](Support Vector Machine, SVM)和人工神经网络法(Artificial Neural Network, ANN)^[8].相对于传统的预测算法,人工神经网络具有良好的容错性,泛化能力好,适于拟合复杂的非线性关系,应用领域广泛.其中,文献[9]运用LSTM单一模型预测,模型简单,精度一般;文献[10]建立基于DBN的短期光伏出力预测模型,达到较高精度的预测效果,但其输入变量的遴选是直接选取的,即未考虑除光照强度、温度等相关性较强因素以外的气象因子对输出功率的影响;文献[11]对不同发电站的输出功率进行预测时采用主成分分析法,预测准确度因气象因素的进一步利用得到了提高,但是只是主观上突出主要矛盾,规避次要因素的影响,相对缺乏客观实验结果支持.可见,现有方法抑或不考虑气象因素对光伏出力的影响;抑或主观采用仅反映

预测对象某一或某几方面显著特征的方法进行预测,导致未能对多元气象因素进行充分挖掘、有效利用;再者因气象因素充分挖掘工作复杂等原因而将所有影响因素未加选择地全部作为预测模型的输入,未剔除相关性极小、甚至不相关因素,使得训练时间过长、预测精度不高.

为了充分挖掘气象因素对光伏出力的影响,有效利用深度学习技术在非线性拟合方面的优势,并稳定发挥复合模型强大的建模能力和显著的预测能力,本文提出一种基于气象因素充分挖掘的BiLSTM网络的光伏发电短期功率预测方法.在KNN算法的遴选机制下,采用马氏距离衡量各因素与发电数据之间的相关性,筛选出与输出相关性最强的K个气象因素,按序排列,优化样本.将重新构建的气象因素数据序列与历史发电功率作为BiLSTM预测模型的输入,训练并建立与出力值之间的映射关系,并以误差最小时对应的K值的预测结果作为最终的预测值.从而达到气象因素的充分挖掘,高效利用,进而提高预测精度的目的.基于实测数据的实验结果及对比结果表明,该方法具有高效的气象因素利用率、优渥的预测准确率,是一种有效的光伏发电功率预测途径.

1 KNN-BiLSTM 光伏发电短期功率预测方法

本文所提的KNN-BiLSTM光伏发电短期功率预测方法,先利用KNN算法对原始气象数据进行相关性筛选,生成按与光伏发电功率相关性强弱排列的全新数据集;然后,将此数据集与历史发电功率一同作为BiLSTM的输入,依据评价指标,得出最佳K值的预测结果.图1为模型图.其中,拟采用的理论知识介绍如下.

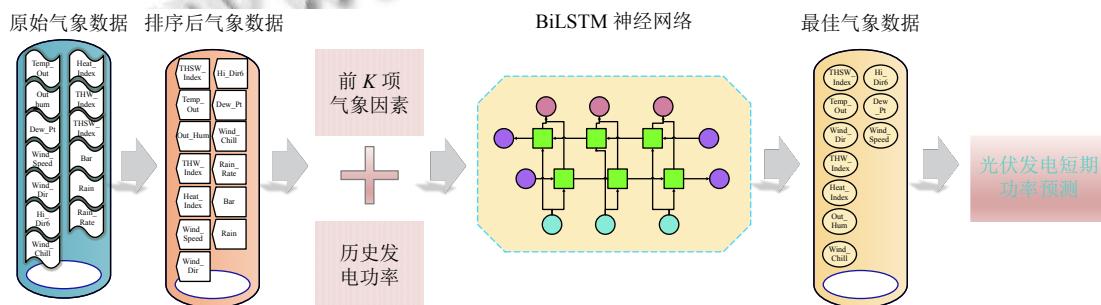


图1 预测方法模型图

1.1 基于KNN的关键气象因素遴选模块

为充分挖掘气象因素对光伏功率的影响程度,因

KNN算法具备强大的相似度问题处理能力,故采用KNN构建关键气象因素遴选模块.其核心思想为:对

于由气象数据及历史发电量构成的测试样本,从中选取目标光伏发电功率为参照样本,基于距离度量找出训练集中与参照样本最靠近的 k 个样本,然后综合参照样本及这 k 个“邻居”信息来进行预测。结合影响光伏发电气象因素的特点,在众多距离度量方法中,鉴于马氏距离具有独立于测量尺度,不受坐标之间量纲影响的特性,采用马氏距离来衡量待选取特征与发电功率之间的相关程度。其具体计算公式为:

$$dM(x_i, x_0) = \sqrt{(x_i - x_0)^T S^{-1} (x_i - x_0)} \quad (1)$$

式中, x_0 为光伏发电输出功率值, x_i 为数据集中第 i 个气象因素的检测值。

1.2 BiLSTM 模块

BiLSTM, 即双向长短期记忆神经网络, 用其构建

的预测模型具有良好的适应性, 可以高效快速利用气象因素和历史发电功率数据, 表现出卓越的光伏发电功率预测能力^[12]。

BiLSTM 由正、反向 LSTM 结合形成, 可对时间序列实现向前和向后两次 LSTM 训练, 进一步提高特征提取的全局性和完整性。LSTM, 即长短记忆网络, 在普通循环神经网络的基础上, 通过加入的“门”结构(即遗忘门、输入门和输出门), 显著改善了传统 RNN 存在的梯度消失、梯度爆炸以及长期记忆不足等问题, 从而有效控制信息的流动量, 学习到准确的长期依赖关系。因此, 在处理时间序列的预测和分类问题中被广泛应用^[13]。其结构如图 2 所示。

LSTM 中信息的传递须经过以下 3 个过程。

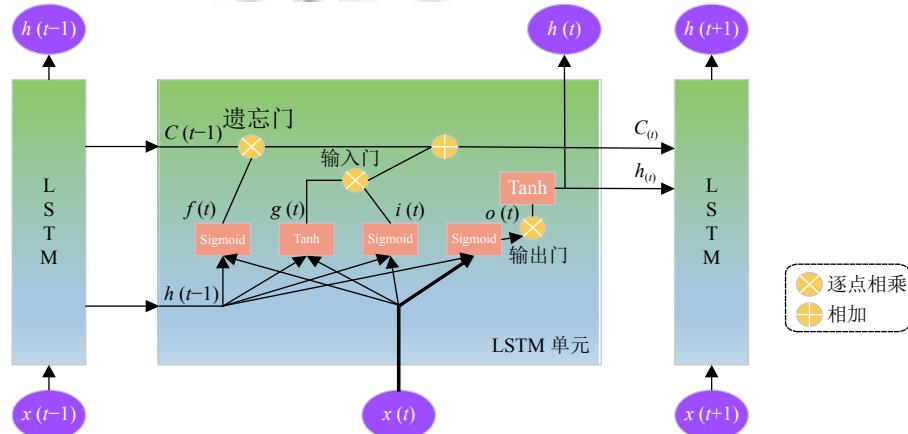


图 2 LSTM 模型结构

(1) 信号的取舍。遗忘门将 LSTM 在上一时刻隐藏层的输出 $h(t-1)$ 和当前的输入 $x(t)$ 连接, 在激励函数 Sigmoid 的激励下使输出值在 $[0, 1]$ 之间, 决定是否让上一时刻所学信息 $C_{(t-1)}$ 通过或部分通过。0 表示全部舍弃, 1 表示全部保留。具体公式如下:

$$f_t = \text{Sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

式中, W_f 为权重, f_t 为隐含层的输出。

(2) 信号的更新。该环节包括 3 部分: 首先是输入门通过 Sigmoid 层决定用于更新的信息 $o(t)$, 然后利用 tanh 生成新的候选值 $i(t)$, 最后将这两部分产生的值结合后进行信息的更新。具体公式如下:

$$g_t = \text{sigmoid}(W_g \cdot [h_{t-1}, x_t] + b_g) \quad (3)$$

$$i_t = \tanh(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$C_t = f_t * C_{t-1} + g_t * i_t \quad (5)$$

(3) 信号的输出。首先通过 Sigmoid 层得到一个初始输出 $o(t)$, 然后使用 tanh 将其值缩放, 再将处理后的值与 Sigmoid 层的输出逐对相乘, 最终得到模型的输出 $h(t)$ 。具体表达式如下:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

若记 t 时刻正向 LSTM 输出的隐藏状态为 h_{t1} , 反向 LSTM 输出的隐藏状态为 h_{t2} , 则 BiLSTM 输出的隐藏状态 h_t : $h_t = h_{t1} \oplus h_{t2}$ 。

1.3 可能影响光伏输出功率的气象因素

光伏输出功率与太阳辐射强度、温度、湿度和风况等因素息息相关, 但影响程度存在差异^[14]。然而实际中影响光伏发电功率的气象因素不止于此。列举本文中研究的可能影响光伏出力的气象因素如表 1 所示。

表1 气象因素列表

Temp_Out	Out_Hum	Dew_Pt.	Wind_Speed	Wind_Dir	Hi_Dir6	Wind_Chill	Heat_Index	THW_Index	THSW_Index	Bar	Rain	Rain_Rate
温度	湿度	露点	风速	风向	最大风速的风向	风寒指数	热指数	温湿度及风速指数	温湿度辐射、风综合指数	压强	降雨	降雨量

1.4 预测算法实现流程

依据 BiLSTM 神经网络在时序数据分析中强大的适应性,与 KNN 算法相结合,构建光伏发电短期预测模型,确定参数并进行训练测试,得出预测结果。预测流程图如图 3 所示。

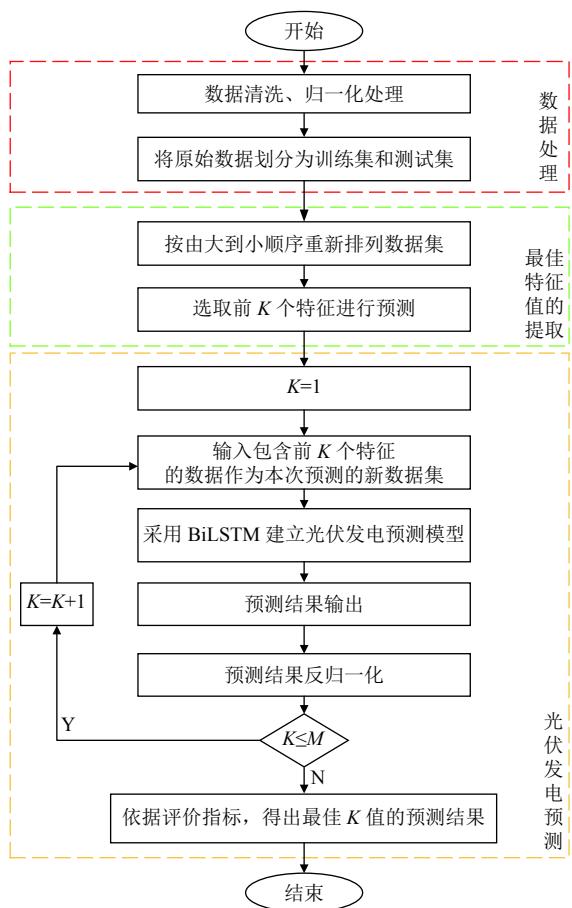


图 3 预测模型流程图

2 算例分析

为验证本算法的科学性和实用性,本文实验在 Tensorflow1.8.0、Python3.6 的运行环境下,并以某地区 2010 年 1 月 1 日到 2010 年 12 月 31 日的 525 576 光伏发电功率和气象数据为依托进行验证。数据采集周期为 1 分钟,除记录发电功率值外,每条历史数据还

包括时刻、气温、湿度等信息,最后与经典的 KNN 预测模型^[15]、BiLSTM 网络模型、DBN 预测模型^[10]、以及 PCA-LSTM 预测结果^[11]进行对比分析。

2.1 训练集和测试集划分

以 2010 年 1 月 1 日至 2010 年 11 月 30 日光伏发电数据及气象数据为训练集,以 2010 年 12 月 1 日至 31 日数据作为测试数据集用于验证模型的有效性。

2.2 性能评测指标

为了准确评估各类模型的预测效果,本文选用了 3 个评价指标,分别为:均方误差 (MSE)、均方根误差 ($RMSE$)、决定系数 R^2 。各评价指标的表达式如下:

$$MSE = \frac{\sum_{t=1}^{t=N} (y_i - \hat{y}_i)^2}{N} \quad (8)$$

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

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (10)$$

式中, y_i 实际光伏发电功率, \hat{y}_i 为预测发电功率, \bar{y} 为真实发光功率的平均值, N 为测试集的大小。其中, MSE 、 $RMSE$ 的值越小, R^2 越接近 1, 表征模型的预测能力越好。

2.3 BiLSTM 超参数确定

本文所提预测方法依附于 BiLSTM 模型强大的“记忆”功能,故该模型参数的设置是实现高预测准确度的前提。文献[16]指出, LSTM 主要由以下参数确定:输入层维数、输出层维数、输入层时间步长、模型的层数以及每层维数, BiLSTM 亦然。其中, 输入层维数即输入变量的个数,则该值等于每次输入 BiLSTM 模型的气象因素个数加 1;输出层数值为 1。输入层时间步长、模型的层数以及每层维数 3 项参数将通过控制变量法进行逐步调优,得出相应数值。

(1) 输入层时间步长

输入层时间步长即用于预测当前时刻光伏出力的输入参数的长度。若该值过小，则不能充分发挥BiLSTM在解决时间序列长期依赖问题中的优势；反之，则因隐含层信息容量有限导致训练时间过长，准确率降低。图4为输入层时间步长对预测准确度的影响。由RMSE及 R^2 结果可得，当输入时间步长为6时，预测效果最好。其含义为该分钟的光伏出力由前6分钟的气象数据和发电数据预测得出。

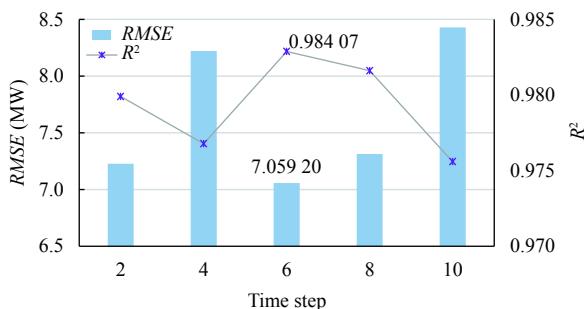


图4 不同时间步长下的预测结果

(2) 模型的层数以及每层维数

模型的层数以及每层维数的选取对于最终的预测精度有直接影响。模型的层数越多，非线性拟合能力越强，但复杂度也会随之增加。确定模型的层数及其维数时，按以下实验步骤进行：第1步，确定单层模型时维数的最佳值并将其固定；第2步，增加一层隐藏层，并在第一步基础之上确定该层最佳维数；以此类推，预测误差最小时所对应的层数及其维数为模型最终参数。实验结果如图5所示。

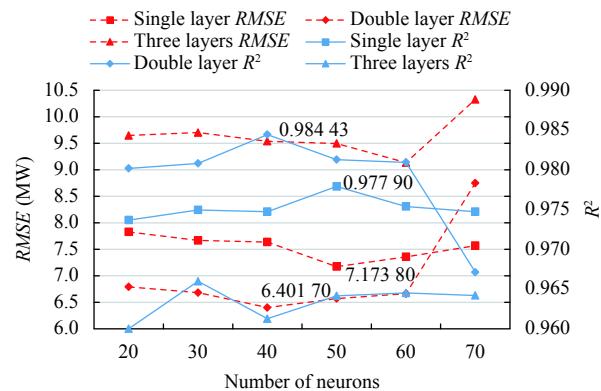


图5 不同层参数下的预测结果

实验表明：模型的层数为2，维数分别为50、40时，BiLSTM模型预测性能最佳。最终BiLSTM模型相关参数设置如表2所示。

表2 相关参数设置

参数	取值
Hidden_size_I	50
Hidden_size_II	40
Timesteps	6
Epoch	100
Batch_size	72
Learning_rate	0.01
Training_steps	50 000

2.4 实验结果及分析

采用KNN方法实现影响光伏出力气象因素的遴选，达到高效利用相关因子提高预测准确度的目的。 K 值不同即选取的相关气象因素个数不同，将不同 K 值下选取的因子及训练结果记录如下。表3为不同 K 值遴选出的气象因素，表4给出了不同 K 值时的预测误差。

表3 不同 K 值遴选出的气象因素

K 取值	THSW_Index	Temp_Out	Out_Hum	THW_Index	Heat_Index	Wind_Speed	Wind_Dir	Hi_Dir6	Dew_Pt	Wind_Chill	Rain_Bar	Rain_Rate
1	√											
2	√	√										
3	√	√	√									
4	√	√	√	√								
5	√	√	√	√	√							
6	√	√	√	√	√	√						
7	√	√	√	√	√	√	√					
8	√	√	√	√	√	√	√	√				
9	√	√	√	√	√	√	√	√	√			
10	√	√	√	√	√	√	√	√	√	√		
11	√	√	√	√	√	√	√	√	√	√	√	
12	√	√	√	√	√	√	√	√	√	√	√	√
13	√	√	√	√	√	√	√	√	√	√	√	√

表4 不同K值所得实验结果

K取值	RMSE	MSE	R ²
1	6.7839	46.02	0.980 23
2	6.8554	47.00	0.972 08
3	6.7729	45.87	0.980 30
4	6.7467	45.52	0.980 45
5	7.4224	55.09	0.976 34
6	6.7156	45.10	0.980 63
7	7.1352	50.91	0.978 13
8	6.6815	44.64	0.980 83
9	7.3658	54.26	0.976 70
10	6.5855	43.37	0.981 37
11	7.1050	50.48	0.978 32
12	6.9521	48.33	0.979 24
13	6.9503	48.31	0.979 25

由以上结果可得,当K=1时,选取的气象因素为THSW指数,即温度、湿度、辐射、风4个参数综合起来的参数。在该数据集所有的气象因素中,仅此特征含有与光伏发电功率相关性最强的因素——太阳辐射强度,则实验结果与1.3中进行的理论分析完全吻合,再次验证该算法的准确性;当K=2时,选取的气象因素为THSW指数和Temp_Out(外界温度)项,可见温度是影响光伏发电的另一个重要因素;当K=10时,预测效果最佳,此时R²的值可高达0.981 37,说明合理高效利用气象因素对光伏出力的影响,有助于预测精度的提高:遴选出影响光伏发电功率相关性较强的气象因素作为输入,剔除相关性较弱的因子,实现光伏发电功率与气象影响因子关联关系研究的基础上,预测准确度也因此得以提高。结果显示,是否降雨、降雨量大小、以及气压与光伏出力相关性较弱,加入上述因素作为输入结果未升反降,为得到最佳预测结果,最终确定除此之外的10项气象因素为影响光伏出力的关键因素。

同时,本文利用KNN-BiLSTM方法得到的预测值与其实际值进行对比,如图6所示。图中,红色实线为预测功率,蓝色实线为实际功率。

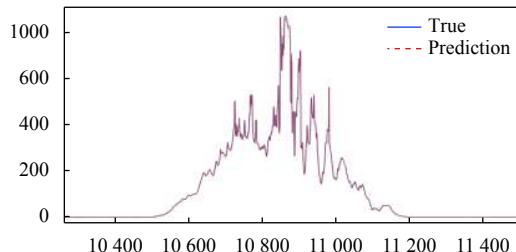


图6 KNN-BiLSTM仿真结果

2.5 对比试验

为了验证KNN-BiLSTM的有效性,与KNN、DBN、BiLSTM、PCA-LSTM相比,以完成对模型的评估工作,性能指标对比如表5,预测结果如图7。可见,采用任一指标衡量,本文提出的算法均具有最佳性能。再结合图7,可以看出DBN、KNN预测精度一般,在发电量峰值谷值均有较大偏差。而BiLSTM、PCA-LSTM虽预测精度有所提高,但输出功率出现波动时偏差较大。所以,各项误差均有明显改善,优于其他模型。

表5 各算法性能指标对比表

算法	RMSE	MSE	R ²
KNN	9.4959	90.17	0.961 27
DBN	10.3726	107.591	0.948 25
BiLSTM	7.8285	61.29	0.971 68
PCA-LSTM	6.7715	45.85	0.974 95
KNN-BiLSTM	6.5855	43.37	0.981 37

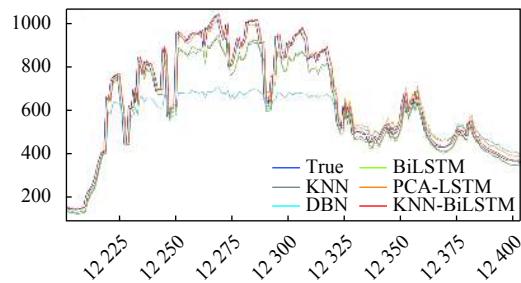


图7 各算法仿真结果对比图

3 结论

本文采用KNN-BiLSTM方法通过历史光伏发电功率、气象因素作为输入,预测光伏出力,可得结论:

1) 该方法有效解决了气象因素挖掘不深入、利用率不高等问题,实现了充分挖掘气象因素对光伏出力的影响,达到提高预测精度的目的。2) 该方法考虑发电数据的时序性,选择BiLSTM作为模型的核心算法,结果表明,与KNN、BiLSTM预测模型相比,预测效果有较大提升;与DBN及PCA-LSTM相比,具有更好的预测性能,预测误差更小,准确度更高。

参考文献

- Zhang JL, Tan ZF, Wei YM. An adaptive hybrid model for day-ahead photovoltaic output power prediction. Journal of Cleaner Production, 2020, 244: 118858. [doi: 10.1016/]

j.jclepro.2019.118858]

- 2 赛迪智库光伏产业形势分析课题组.2019年中国光伏产业发展形势展望.电器工业, 2019, 5: 51–55.
- 3 Liu J, Fang WL, Zhang XD, *et al.* An improved photovoltaic power forecasting model with the assistance of aerosol index data. IEEE Transactions on Sustainable Energy, 2015, 6(2): 434–442. [doi: [10.1109/TSTE.2014.2381224](https://doi.org/10.1109/TSTE.2014.2381224)]
- 4 Zhou ZY, Xiong F, Huang BY, *et al.* Game-theoretical energy management for energy internet with big data-based renewable power forecasting. IEEE Access, 2017, 5: 5731–5746. [doi: [10.1109/ACCESS.2017.2658952](https://doi.org/10.1109/ACCESS.2017.2658952)]
- 5 张一平, 宋红, 罗敏, 等. 人工神经网络算法在光伏发电短期功率预测中的应用. 太阳能, 2019, (4): 47–53. [doi: [10.3969/j.issn.1003-0417.2019.04.012](https://doi.org/10.3969/j.issn.1003-0417.2019.04.012)]
- 6 Tang YC, Zhao SD, Ten CW, *et al.* Enhancement of distribution load modeling using statistical hybrid regression. Proceedings of 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference. Washington, WA, USA. 2017. 1–5.
- 7 Yang XY, Jiang FF, Liu H. Short-term solar radiation prediction based on SVM with similar data. Proceedings of the 2nd IET Renewable Power Generation Conference. Beijing, China. 2013. 1–4.
- 8 Khan I, Zhu HL, Yao JX, *et al.* Hybrid power forecasting model for photovoltaic plants based on neural network with air quality index. International Journal of Photoenergy, 2017, 2017: 6938713.
- 9 Qing XY, Niu YG. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy, 2018, 148: 461–468. [doi: [10.1016/j.energy.2018.01.177](https://doi.org/10.1016/j.energy.2018.01.177)]
- 10 Neo YQ, Teo TT, Woo WL, *et al.* Forecasting of photovoltaic power using deep belief network. Proceedings of IEEE Region 10 Conference. Malaysia. 2017. 1189–1194.
- 11 Zhang JX, Chi YY, Xiao LP. Solar power generation forecast based on LSTM. Proceedings of the 2018 IEEE 9th International Conference on Software Engineering and Service Science. Beijing, China. 2018. 869–872.
- 12 Chen K, Huo Q. Training deep bidirectional LSTM acoustic model for LVCSR by a context-sensitive-chunk BPTT approach. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(7): 1185–1193. [doi: [10.1109/TASLP.2016.2539499](https://doi.org/10.1109/TASLP.2016.2539499)]
- 13 Fazle K, Majumdar S, Darabi H, *et al.* LSTM fully convolutional networks for time series classification. IEEE Access, 2017, 6: 1662–1669.
- 14 Antonanzas J, Osorio N, Escobar R, *et al.* Review of photovoltaic power forecasting. Solar Energy, 2016, 136: 78–111. [doi: [10.1016/j.solener.2016.06.069](https://doi.org/10.1016/j.solener.2016.06.069)]
- 15 Yesilbudak M, Sagiroglu S, Colak I. A novel implementation of kNN classifier based on multi-tupled meteorological input data for wind power prediction. Energy Conversion and Management, 2017, 135: 434–444. [doi: [10.1016/j.enconman.2016.12.094](https://doi.org/10.1016/j.enconman.2016.12.094)]
- 16 王志远, 王守相, 陈海文, 等. 考虑空间相关性采用LSTM神经网络的光伏出力短期预测方法. 电力系统及其自动化学报, 2020, 32(5): 78–85.