

# 基于PEEEMD-BiLSTM-XGboost光伏发电功率预测方法研究

许时佳

长安大学理学院, 陕西 西安

收稿日期: 2023年11月11日; 录用日期: 2023年12月4日; 发布日期: 2023年12月15日

## 摘要

由于光伏功率数据的强不确定性, 单一模型预测精度受到限制, 提出多融合信号模态分解和双向长短期记忆网络(BiLSTM)、极端梯度提升(XGboost)组合模型的短期光伏发电功率预测方法。首先, 为降低光伏功率信号的复杂性, 通过自适应噪声完备集合经验模态分解(EEMD)、排列熵(PE)对光伏发电功率数据进行预处理, 得到各模态分量; 其次分析选取重要相关影响因素, 构建BiLSTM-XGboost组合模型对光伏发电历史功率数据进行预测。最后, 以某地光伏电站数据进行测试, 仿真结果表明所提出的集成预测模型能够有效提高短期光伏功率预测精度, 具有更少计算时间、较高的估计精度、算法稳定性高、鲁棒性强, 并带来较强的实用价值。

## 关键词

光伏发电功率预测, 排列熵, 多融合信号模态分解, 双向长短期记忆网络, 极端梯度提升

## Research on Power Prediction Method for Photovoltaic Power Generation Based on PEEEMD-BiLSTM-XGboost

Shijia Xu

School of Science, Chang'an University, Xi'an Shaanxi

Received: Nov. 11<sup>th</sup>, 2023; accepted: Dec. 4<sup>th</sup>, 2023; published: Dec. 15<sup>th</sup>, 2023

## Abstract

Due to the strong uncertainty of photovoltaic power data, the prediction accuracy of a single mod-

el is limited. A short-term photovoltaic power prediction method based on the combination model of multi fusion signal mode decomposition and bidirectional short-term memory network (BiLSTM) and extreme gradient boost (XGboost) is proposed. Firstly, in order to reduce the complexity of photovoltaic power signals, the photovoltaic power data is preprocessed using adaptive noise complete set empirical mode decomposition (EEMD) and permutation entropy (PE) to obtain various modal components; secondly, analyze and select important relevant influencing factors, and construct a BiLSTM-XGboost combination model to predict the historical power data of photovoltaic power generation. Finally, testing was conducted using data from a photovoltaic power plant in a certain area. The simulation results showed that the proposed integrated prediction model can effectively improve the accuracy of short-term photovoltaic power prediction, with less computational time, higher estimation accuracy, high algorithm stability, strong robustness, and strong practical value.

## Keywords

Photovoltaic Power Generation Power Prediction, Permutation Entropy, Multi Fusion Signal Modal Decomposition, Bidirectional Long Short Term Memory Network, Extreme Gradient Enhancement

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着全球能源需要的增长，新能源成为解决能源供应和环境问题的重要方向之一。由于光伏发电功率具有强不确定性、波动性，受太阳辐照度、温度、太阳能光伏板倾斜角度等多种因素的影响，给国家电力系统的安全运行带来了巨大的挑战。准确的光伏功率预测可以为电力系统规划、运行、优化调度等提供可靠的信息支撑，对促进新能源发展具有重要意义[1]。

目前光伏发电功率预测主要为物理建模或者统计方法等，物理建模预测基于各输入特征对光伏发电功率的影响关系建立数学模型，根据物理模型实现功率预测。文献[2]通过云图，提取出各特征作为输入建立径向基函数网络预测模型，预测太阳辐射并根据模型得出光伏功率预测结果，具有较高精确度。文献[3]提出基于卫星遥感数据的光伏发电功率预测方法，获得了较高精度的预测结果。文献[4]针对雾霾天气问题，通过对硬件实验的数据统计，并结合模型实现了较高精度的功率预测。除了基于物理建模方法，很多研究基于历史数据建立输入数据与预测目标之间的映射关系，文献[5]用同样一批样本数据训练四种不同结构和参数的 LSTM 模型，将结果通过邵凯积分进行聚合，具有更高的精度。文献[6]提出基于粒子群算法优化的前馈神经网络的组合模型对光伏发电量进行预测，预测方法精度显著提高。文献[7]针对光伏发电量受到不同天气条件的影响问题，将天气条件分类，通过支持向量机(SVM)建立不同天气条件下的光伏功率预测模型，具有更高的精度。文献[8]提出了一种用于分布式光伏电站发电量预测的四阶段时空混合预测方法，与其他模型相比具有较高精度。文献[9]提出混合的预测框架，利用小波分解、LSTM 和卷积神经网络(CNN)预测光伏发电功率的更多波动细节，该方法拥有较强的准确性和适应能力。文献[10]利用三次插值和 BiLSTM 模型提高天气预报数据的精度，最后用门控递归单元(GRU)与 CNN 模型预测。文献[11]基于改进的平均影响值算法计算气象因素与功率间相关程度，然后通过遗传算法改进的网络进行预测。文献[12]通过主成分分析法对原始数据降维，然后通过遗传算法对 Elman 网络的反馈因子寻优预测，该方法具有较高精度。文献[13]提出一种基于经验模态分解的光伏功率预测模型，利用 LSTM

网络实现光伏发电的预测, 该模型较传统预测方法具有较高精度。文献[14]提出一种基于变分模态分解(VMD)和双重注意力机制 LSTM 的光伏发电量预测模型, 该方法预测效果较为理想。

为了降低光伏发电功率的不稳定性, 提高预测精度, 本文首先基于集合经验模态分解(EEMD)对历史光伏发电功率信号分解, 参考 PE 值对模态分量进行重构, 简化模型的复杂度和模型的计算时间。其次基于 BiLSTM 网络对重构的模态分量分别进行预测, 同时基于 XGboost 对原始光伏发电功率进行预测, 使用组合预测模型得到最后的预测结果。

## 2. 基于信号分解的光伏功率数据预处理

本文提出 PEEEMD 组合模型对光伏发电功率数据进行分解为若干不同的模态分量, 降低原始光伏发电数据的非平稳性, 并通过皮尔逊相关系数分析光伏发电功率数据与相关影响因素, 选取合适的因素作为预测网络的输入。

### 2.1. 基于 EEMD 的光伏发电功率数据分解方法

EEMD 是一种自适应的信号分解方法, 根据信号的局部特征自动确定 IMFs 的数量和尺度, 更好的适应信号的复杂性; 通过在光伏发电功率信号中引入不同的白噪声实现分解, 对得到的 IMFs 进行平均, 有效减少模态混合的问题, 更好的保留信号的局部特征, 图 1、图 2 为光伏历史数据和经 EEMD 分解后的光伏发电功率数据。

### 2.2. 基于 PE 的光伏功率分量重构方法

EEMD 对原始光伏功率信号进行分解可以获得更多的分量细节, 显著增加预测模型的准确度, 但 IMFs 的数量越多, 预测模型复杂度越高、计算量越大、训练时间越长、若将相近的不同模态分量重构, 能够减少模型计算时间。通过 PE 指标将相近的模态分量聚合重构。PE 用于分析时间序列信号的复杂性、非线性特征; 熵越大表示时间序列具有较大的不规则性, 反之, 时间序列较为规则、有序、平稳。图 3 为 EEMD 分解后各个模态分量的排列熵。

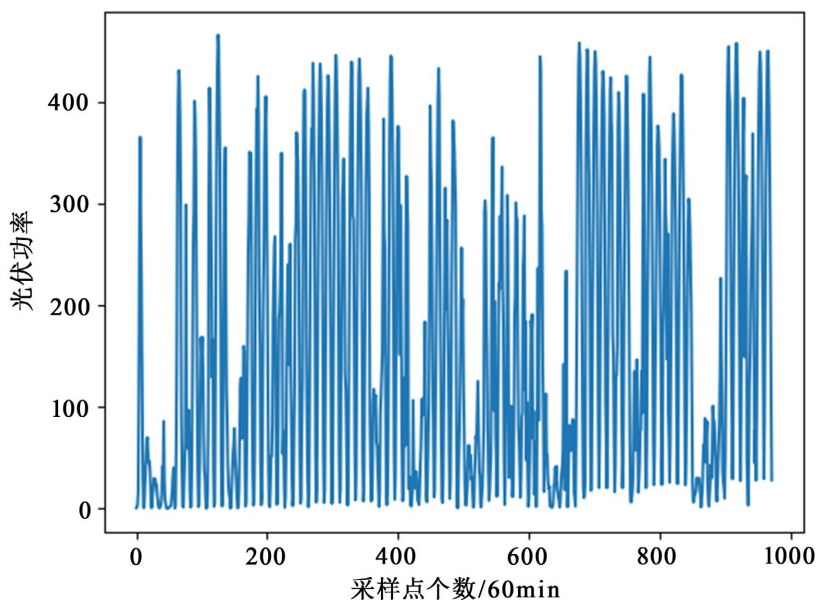


Figure 1. Raw photovoltaic power generation data

图 1. 原始光伏发电功率数据

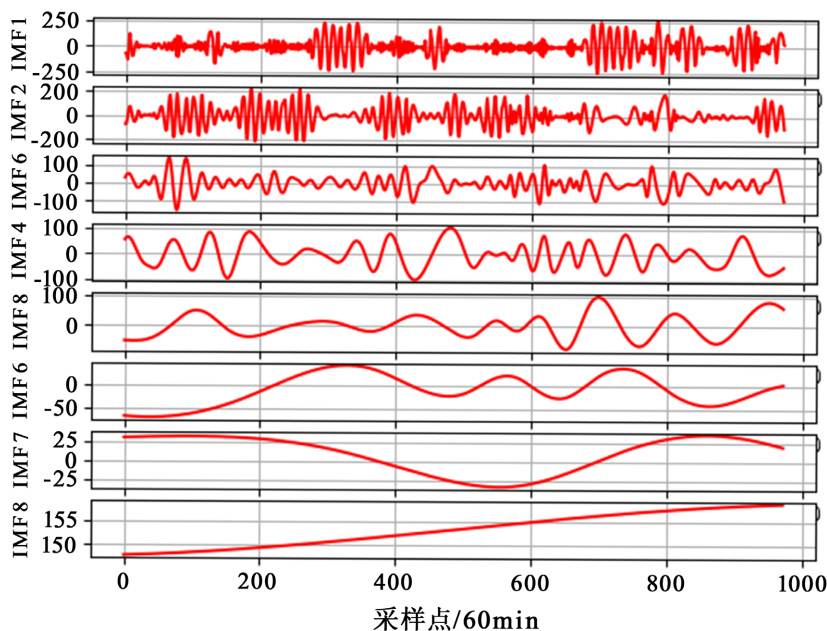


Figure 2. Photovoltaic power generation data decomposed by EEMD  
 图 2. 经 EEMD 分解的光伏发电功率数据

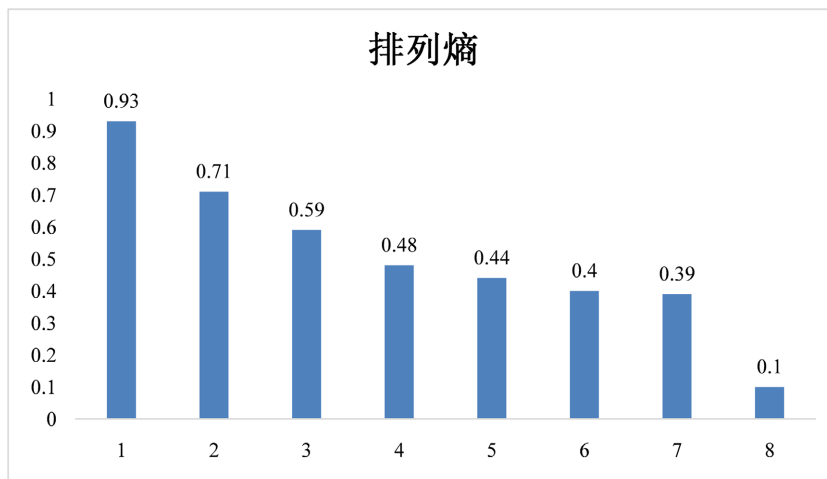


Figure 3. The arrangement entropy of each modal component after EEMD decomposition  
 图 3. 经 EEMD 分解后各个模态分量的排列熵

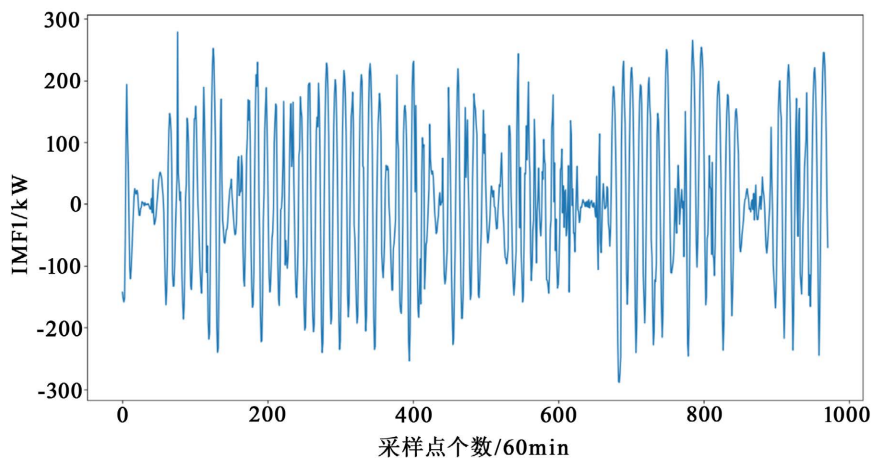
可将排列熵相近的模态分量重构，减少模型复杂度和计算时间，重构的模态分量为高频 IMF1、中频 IMF2、低频 IMF3；如图 4 所示，后续对重构的模态分量分别预测。

### 2.3. 光伏功率预测特征选择

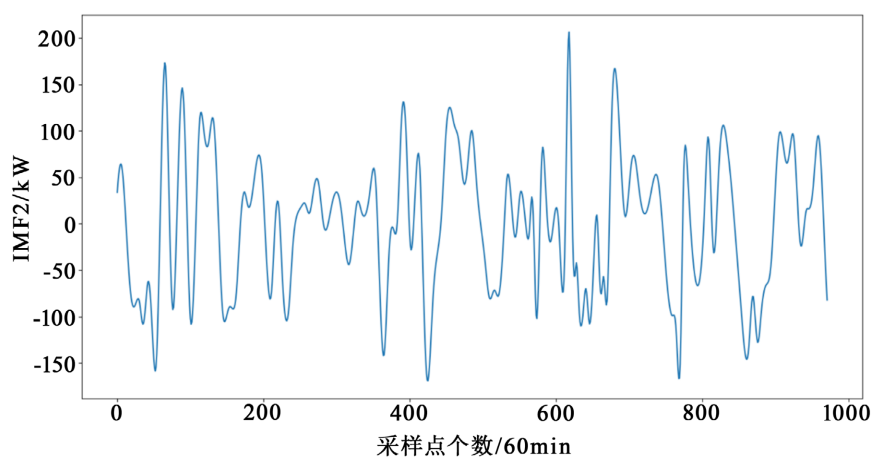
影响功率的因素主要为太阳辐照度、太阳天顶角、温度、湿度、降水、云量等，而太阳辐射主要分为直射辐射度(DNI)、散射辐射度(DHI)；已知太阳光伏板的最佳倾斜角为 30°，利用一天中太阳高度角的变化，计算出光伏板能够接收到的辐照值(GHI)，计算公式如下：

$$GHI = DHI + DNI * \cos(z) \tag{1}$$

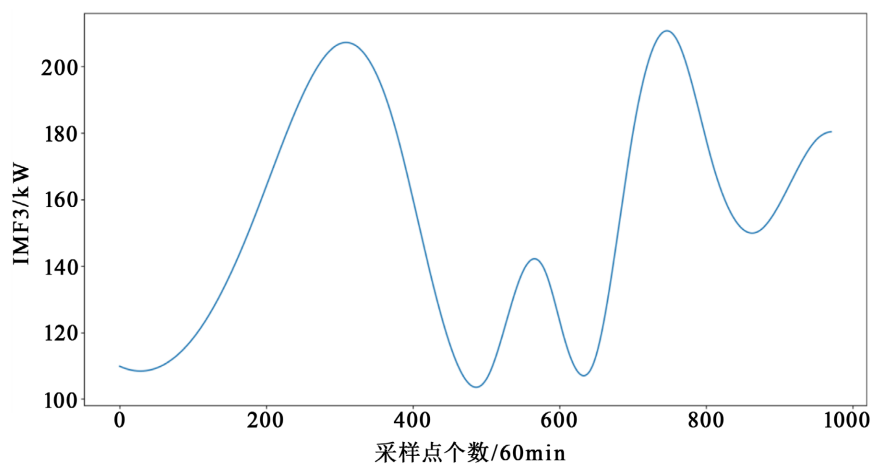
其中  $z$  是太阳天顶角。



(a) 重构后高频 IMF1



(b) 重构后中频 IMF2



(c) 重构后低频 IMF3

**Figure 4.** Reconstructed photovoltaic power generation data**图 4.** 重构后的光伏发电功率数据

皮尔逊相关系数是衡量两变量间的线性相关程度的重要指标，计算功率与其他影响因素的皮尔逊相关系数，如图 5 所示。

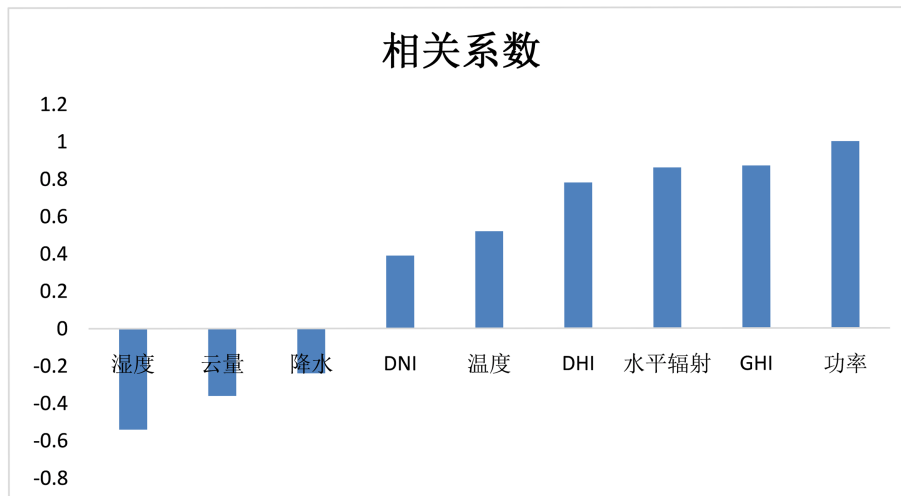


Figure 5. Pearson correlation coefficient graph of power and other influencing factors  
图 5. 功率与其他影响因素的皮尔逊相关系数图

由图可看出湿度、云量、降水均与功率呈负相关，其他影响因素呈正相关，本文选取 GHI、温度、湿度、降水、云量五个影响因素作为输入特征。

## 2.4. 光伏功率 BiLSTM-XGboost 预测模型

### 2.4.1. XGboost 概述

XGboost 是一种 Boosting 树模型的集成学习的框架，XGboost 通过集成多个决策树模型来预测功率，通过最小化损失函数梯度不断优化模型性能，加入了正则化控制模型复杂度，避免过拟合，能够自动处理缺失值，极大的提高了模型的训练速度。

### 2.4.2. BiLSTM 概述

BiLSTM 是 LSTM 的变体结构，它能同时考虑当前时刻的前后信息，能够较好的前后信息进行捕捉，充分挖掘光伏功率的时序特征，具有更高的预测精度；BiLSTM 通过双向传递，更好的保留重要的序列信息。

BiLSTM 模型参数计算如下：

$$\begin{cases} a^{(i)} = \tanh(W_a h^{(i-1)} + U_a x^{(i)} + b_a) \\ f^{(i)} = \sigma(W_f h^{(i-1)} + U_f x^{(i)} + b_f) \\ o^{(i)} = \sigma(W_o h^{(i-1)} + U_o x^{(i)} + b_o) \\ C^{(i)} = i^{(i)} \odot a^{(i)} + f^{(i)} \odot C^{(i-1)} \end{cases} \quad (2)$$

其中： $x_t$  表示当前时刻输入， $\sigma$  和  $\tanh$  分别为 sigmoid 和双曲正切激活函数； $h^{(i-1)}$  表示上一时刻的状态信息， $W_a$ 、 $W_f$ 、 $W_o$  分别表示为表示输入门、遗忘门、输出门的权重矩阵， $b_a$ 、 $b_f$ 、 $b_o$  分别表示输入门、遗忘门、输出门的偏置， $C^{(i-1)}$  表示上一时刻的隐藏状态， $f^{(i)} \odot C^{(i-1)}$  表示对上一时刻进行选择， $C^{(i)}$  为当前时刻更新后的隐藏状态信息， $\odot$  表示按位相乘。

### 2.4.3. PEEEMD-BiLSTM-XGboost 光伏功率预测方法

为了提高光伏发电功率预测的精度，本文提出 PEEEMD-BiLSTM-XGboost 组合光伏功率预测模型，预测流程图如图 6 所示。

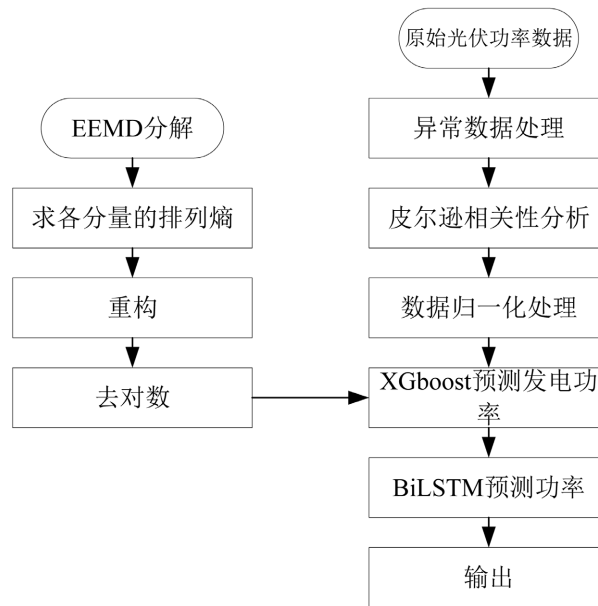


Figure 6. PEEEMD-BiLSTM-XGboost model photovoltaic power prediction flowchart  
图 6. PEEEMD-BiLSTM-XGboost 模型光伏功率预测流程

预测步骤如下：

- 1) 数据预处理。针对历史光伏功率与其他相关特征因素数据，对所有数据进行异常值检测，采用均值法对缺失值进行填补，确保数据无异常。
- 2) 特征选择。计算功率与所有相关影响因素的皮尔逊相关系数，选取合适的影响因素作为特征变量，减少模型复杂度，减少过拟合。对于处理后的光伏发电功率数据采用 EEMD 进行分解，并计算每个分量的 PE，将 PE 值临近的分量合成，作为光伏功率预测的输入变量。
- 3) 模型预测。首先使用 XGboost 对光伏发电功率进行初次预测，其次将其输出结果作为一个特征输入至 BiLSTM 进行二次预测，增加预测精度。
- 4) 通过与其他模型对比，充分说明本文提出的组合模型的有效性和鲁棒性。

### 3. 算例分析

#### 3.1. 实验配置及数据来源

本文使用 Python 编译器和 Tensorflow 深度学习框架构建短期光伏功率预测模型，选用某光伏场站实测数据进行实验。以光伏场站 2023 年 2 月 9 日~5 月 30 日的发电功率数据和气象数据进行仿真实验，验证本文提出模型和方法的有效性。考虑到晚间光伏出力值为零，选取每天 08:00~19:00 的数据进行实验，采样间隔为 5 min，但气象数据采样间隔为 60 min，故将光伏功率数据与气象数据统一量纲，均为 60 min。

#### 3.2. 评价指标选择

本文选取平均绝对误差(MAE)  $\varepsilon_{MAE}$ 、均方根误差(RMSE)  $\varepsilon_{RMSE}$ 、和平均绝对百分比误差  $\varepsilon_{MAPE}$  作为主要评价指标，表示光伏发电功率模型预测的精度。其公式如下：

$$\varepsilon_{MAE} = \frac{1}{n} \sum_i^n |\hat{y}_i - y_i| \quad (3)$$

$$\varepsilon_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (4)$$

$$\varepsilon_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (5)$$

式中： $\hat{y}_i$  和  $y_i$  分别为第  $i$  个光伏功率样本的预测值和真实值； $N$  为样本个数。

### 3.3. XGboost 模型参数选择

为了验证本文方法的有效性，首先通过 XGboost 对光伏发电功率进行预测，本文模型中 XGboost 部分参数设置情况如表 1 所示。

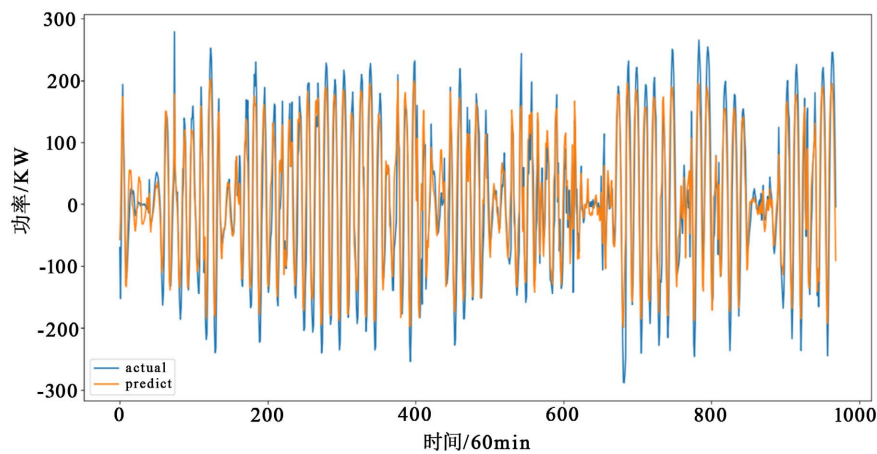
**Table 1.** XGboost partial parameter settings  
**表 1.** XGboost 部分参数设置

参数	设置数值
迭代次数	1000
决策树数量	50
学习率	0.05
树最大深度	10
采样率	0.6

表 1 中，迭代次数、决策树数量、数最大深度影响着模型对问题的求解能力，学习率设置是为了减少过拟合现象的发生。

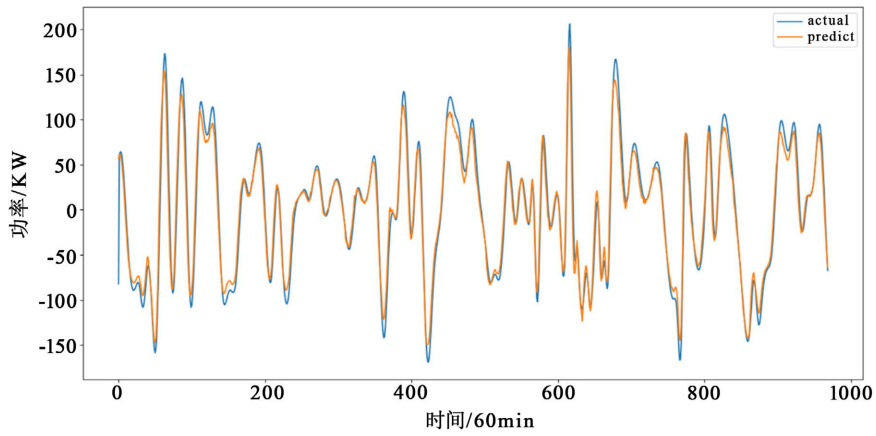
### 3.4. 预测结果对比分析

通过对重构的模态分量高频 IMF1、中频 IMF2、低频 IMF3 分别使用 BiLSTM 进行测，预测结果如图 7 所示。

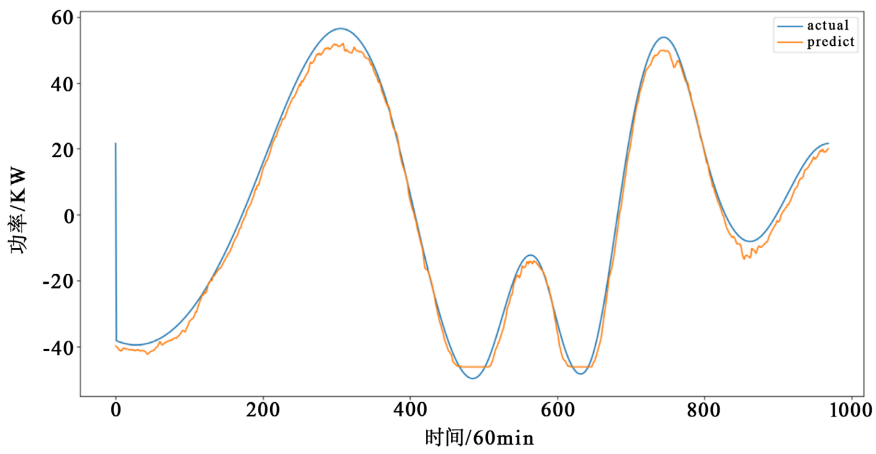


(a) 高频 IMF1 预测结果





(b) 中频 IMF2 预测结果



(c) 低频 IMF3 预测结果

Figure 7. BiLSTM prediction results

图 7. BiLSTM 预测结果

从上图可以发现中频 IMF2、低频 IMF3 预测的效果相较高频 IMF1 的预测效果更精确。

下图 8 为 XGboost 预测测试集效果图:

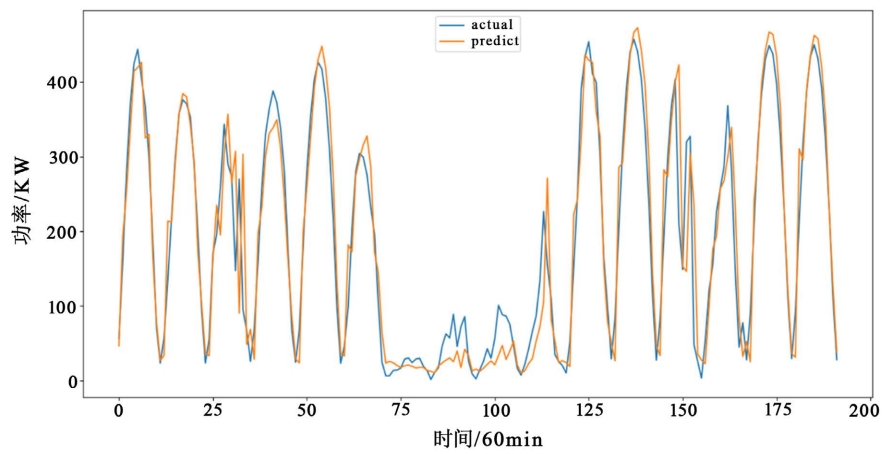
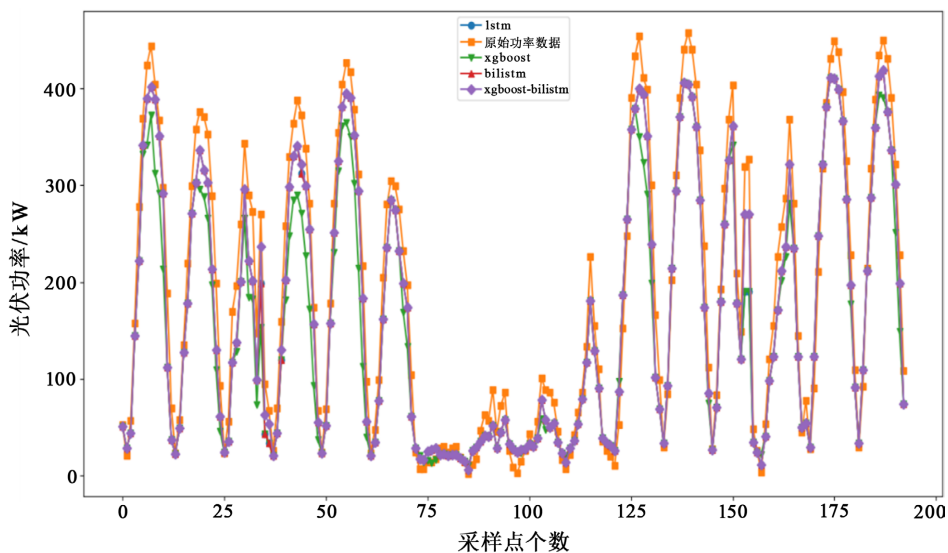


Figure 8. XGboost prediction test set rendering

图 8. XGboost 预测测试集效果图

为了增加模型预测精度,将 XGboost 模型预测结果作为 BiLSTM-XGboost 组合模型的一个特征输入。

本文使用了 BiLSTM-XGboost 组合模型预测、LSTM 预测模型、BiLSTM 预测模型以及 XGboost 预测模型的预测结果进行对比分析,验证本文模型的有效性、预测的精确性,本文以平均绝对误差、均方根误差、平均绝对百分比误差作为预测精度的判定标准,四种预测模型的输出功率与实际功率测试集对比曲线如图 9 所示。



**Figure 9.** Comparison curve of output power of four prediction models and actual power test set  
**图 9.** 四种预测模型输出功率与实际功率测试集对比曲线

上述不同模型的预测功率中,本文所提组合预测方法 BiLSTM-XGboost 的预测效果较好,尤其可见在阴天或非晴天时预测效果相较其他三种方法效果最好,精度最高。

预测结果评价指标对比:

**Table 2.** Error evaluation index values of various prediction models under different optimization algorithms  
**表 2.** 不同优化算法下各预测模型的误差评价指标值

预测模型	$\epsilon_{MAE}$	$\epsilon_{RMSE}$	$\epsilon_{MAPE}$
LSTM	5.3156	6.1732	0.1541
Xgboost	5.9712	7.5476	0.1587
BiLSTM	4.3121	5.6449	0.1255
BiLSTM-XGboost	3.1022	3.4649	0.1147

从表 2 可知,模型 LSTM 和模型 Xgboost 预测效果比较相近,模型 BiLSTM 能够较好的前后信息进行捕捉,具有更高的预测精度。模型 BiLSTM-XGboost 对于光伏发电功率的预测性能更好,也表明 BiLSTM-XGboost 模型在光伏发电功率预测领域具有更好的建模效果。

### 3.5. 结论

- 1) 提出的 PEEEMD 方法可有效避免网络过拟合及减少计算时间成本,从而提高预测精度。

2) 所构建的 BiLSTM-XGboost 模型能充分发挥 BiLSTM 特征提取能力, 通过 XGboost 模型预测结果作为 BiLSTM 的另一特征, 可最大程度提高时序预测精度。

3) 仿真结果表明, 所提 BiLSTM-XGboost 模型的平均绝对误差(MAE)  $\varepsilon_{MAE}$ 、均方根误差(RMSE)  $\varepsilon_{RMSE}$ 、和平均绝对百分比误差  $\varepsilon_{MAPE}$  均优于其他 3 种模型, 预测效果更好, 在短期光伏功率预测领域具有良好的应用前景。

目前大部分光伏功率预测研究大部分为神经网络或者组合模型预测, 如果能充分挖掘光伏发电功率的机理, 更多的从机理角度出发, 再结合神经网络或者其他统计方法, 此方向具有深远的研究意义, 将作为下一步研究方向。

## 基金项目

陕西省重点产业创新链项目(2020ZDLGY09-09)。

## 参考文献

- [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): 1144-1157.
- [2] 陈志宝, 丁杰, 周海, 程序, 朱想. 地基云图结合径向基函数人工神经网络的光伏功率超短期预测模型[J]. 中国电机工程学报, 2015, 35(3): 561-567.
- [3] 刘晓艳, 王珏, 姚铁锤, 等. 基于卫星遥感的超短期分布式光伏功率预测[J]. 电工技术学报, 2022, 37(7): 1800-1809.
- [4] 刘卫亮, 刘长良, 林永君, 等. 计及雾霾影响因素的光伏发电超短期功率预测[J]. 中国电机工程学报, 2018, 38(14): 4086-4095+4315.
- [5] Abdel-Nasser, M., Mahmoud, K. and Lehtonen, M. (2020) Reliable Solar Irradiance Forecasting Approach Based on Choquet Integral and Deep LSTMs. *IEEE Transactions on Industrial Informatics*, **17**, 1873-1881. <https://doi.org/10.1109/TII.2020.2996235>
- [6] Muhammad, Q.R., Nadarajah, M., Li, J.M., et al. (2019) An Ensemble Framework for Day-Ahead Forecast of PV Output Power in Smart Grids. *IEEE Transactions on Industrial Informatics*, **15**, 4624-4634. <https://doi.org/10.1109/TII.2018.2882598>
- [7] Shi, J., Li, W.J., Liu, Y.Q., et al. (2012) Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. *IEEE Transactions on Industry Applications*, **48**, 1064-1069. <https://doi.org/10.1109/TIA.2012.2190816>
- [8] Shi, J., Chen, Y.B., Cheng, X.G., et al. (2023) Four-Stage Space-Time Hybrid Model for Distributed Photovoltaic Power Forecasting. *IEEE Transactions on Industry Applications*, **57**, 1129-1138. <https://doi.org/10.1109/TIA.2022.3205570>
- [9] Li, J.J., Zhang, C.H., Sun, B., et al. (2023) Two-Stage Hybrid Deep Learning with Strong Adaptability for Detailed Day-Ahead Photovoltaic Power Forecasting. *IEEE Transactions on Sustainable Energy*, **14**, 193-205. <https://doi.org/10.1109/TSTE.2022.3206240>
- [10] Li, H., Ren, Z.Y., Xu, Y., et al. (2022) A Multi-Data Driven Hybrid Learning Method for Weekly Photovoltaic Power Scenario Forecast. *IEEE Transactions on Sustainable Energy*, **13**, 91-100. <https://doi.org/10.1109/TSTE.2021.3104656>
- [11] 王英立, 陶帅, 候晓晓, 齐宏. 基于 MIV 分析的 GA-BP 神经网络光伏短期发电预测[J]. 太阳能学报, 2020, 41(8): 236-242.
- [12] 胡兵, 詹仲强, 陈洁, 等. 基于 PCA-GA-Elman 的短期光伏出力预测研究[J]. 太阳能学报, 2020, 41(6): 256-263.
- [13] 张雲钦, 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.
- [14] 杨晶显, 张帅, 刘继春, 等. 基于 VMD 和双重注意力机制 LSTM 的短期光伏功率预测[J]. 电力系统自动化, 2021, 45(3): 174-182.