

基于 MPA-LSTM 模型和 Bootstrap 方法的 短期光伏功率区间预测

宋绍剑*, 罗世坚, 李国进, 刘斌

(广西大学 电气工程学院, 广西 南宁 530004)

摘要:光伏发电功率的波动性和间歇性为电力系统调度管理带来巨大的挑战,精确的光伏功率区间预测是解决上述问题的一种有效途径。为此,本文提出了一种基于 LSTM 网络的新型短期光伏功率区间预测模型。采用 MPA 对 LSTM 网络的隐含层神经元数和训练批次等超参数进行自动寻优,以克服随机选取 LSTM 模型参数过程中存在的盲目性、费时等问题;并将 MPA-LSTM 模型用于光伏功率点预测。然后,采用 Bootstrap 方法分析模型预测结果的误差分布,确定模型预测输出的区间范围。最后,通过对比仿真验证所提模型的有效性。结果表明:本文所提的 MPA-LSTM 模型均方误差的平均值为 1.09%,优于 SVM、LSTM、PSO-LSTM 和 MPA-SVM 模型;Bootstrap 方法能够准确地描述给定置信度水平下的光伏功率波动范围。

关键词:光伏;区间预测;长短期记忆网络;海洋捕食者算法;Bootstrap

中图分类号:TU473.1

文献标识码:A

文章编号:1001-7445(2022)04-0986-12

A short-term PV power interval forecasting based on MPA-LSTM network model and Bootstrap method

SONG Shao-jian*, LUO Shi-jian, LI Guo-jin, LIU Bin

(School of Electrical Engineering, Guangxi University, Nanning, 530004, China)

Abstract: The fluctuation and intermittence of photovoltaic (PV) power bring great challenges to the power system scheduling and management. Accurate PV power interval prediction is an effective way to address these challenges. Therefore, this paper proposed a new short-term PV power interval prediction model based on LSTM network. In order to avoid the blindness and time wasting during random selecting LSTM model parameters, MPA was used to optimize the number of hidden layer neuron and batch size of LSTM network, and MPA-LSTM model was used for PV power point prediction. Then, Bootstrap method was used to analyze the error distribution and determine the interval prediction range of proposed model. Finally, the effectiveness of the proposed method was

收稿日期:2021-12-20;**修订日期:**2022-04-28

基金资助:国家自然科学基金项目(62141103);广西自然科学基金项目(2016GXNSFAA380327);广西研究生教育创新计划资助项目(YCSW2021042)

通讯作者:宋绍剑(1970—),男,广西象州人,广西大学教授;E-mail:sjsong03@163.com。

引文格式:宋绍剑,罗世坚,李国进,等.基于 MPA-LSTM 模型和 Bootstrap 方法的短期光伏功率区间预测[J].广西大学学报(自然科学版),2022,47(4):986-997.

verified by comparison simulations. The results show that the proposed MPA-LSTM model has an average root mean square error of 1.09%, which is better than models established by SVM, LSTM, PSO-LSTM and MPA-SVM, respectively. Moreover, the Bootstrap method accurately describes the fluctuation range of PV power at given confidence levels.

Key words: photovoltaic power; interval prediction; long short-term memory network; marine predators algorithm

0 引言

太阳能是一种绿色能源,具有储量丰富、污染小和可再生等优点^[1]。据估计,到2030年全球光伏发电总装机容量将达到1700GW^[2];但由于受到云层、风力和温度等天气因素影响,因此光伏电站的输出功率具有很强的波动性和间歇性。大规模光伏电源接入电网将给电力的调度和安全稳定运行带来巨大的挑战^[3],精确的光伏功率预测是电力系统科学调度决策的前提,对电网的稳定运行具有重要意义。

光伏输出功率与影响因素之间存在很强的非线性关系,精确的解析模型难以获得。由于人工神经网络具有很强的非线性映射能力,因此被广泛用于光伏功率预测^[4-6]。长短期记忆网络(long short term memory, LSTM)是一种循环神经网络(recurrent neural network, RNN)的变体,具有从大量时序数据中挖掘数据间复杂关联特性的能力^[7]。与支持向量机(support vector machine, SVM)、极限学习机(extreme learning machine, ELM)等人工神经网络相比, LSTM 网络在时序数据建模方面具有明显优势,目前被广泛用于负荷预测、风力功率预测和光伏功率预测等领域。从文献[7-9]可知, LSTM 网络在预测光伏功率领域有良好的精确度和鲁棒性,但是, LSTM 网络模型性能与其结构和超参数选取紧密相关,其参数目前主要通过人工试凑法选取。这种方法不但耗时、工作量大,而且难以获得合理的参数,为此,将遗传算法(genetic algorithm, GA)和粒子群算法(particle swarm optimization, PSO)等群智能优化算法用于 LSTM 超参数优化已成为一种新趋势,通常可以取得更好的精度^[5,10],但是,大多数群智能优化算法仍面临结构复杂、搜索速度慢、易于陷入局部最优等问题。海洋捕食者算法(marine predators algorithm, MPA)是一种新型群智能优化算法^[11],与其他群智能算法相比,具有结构简单、算法灵活、易于实现、兼顾勘探与开发的协调能力等优点。目前 MPA 优化 LSTM 网络的模型用于光伏功率预测的案例还鲜有报道。

目前大多数光伏功率预测主要侧重于光伏功率的点预测,即给出待预测点的期望值,忽略了更有价值的区间信息,并且模型不可避免的存在误差。随着调度决策对预测精度的要求进一步提高,给出精确的点预测结果及其区间范围更具实际意义^[12]。光伏功率波动区间分析(即区间预测)方法主要有参数法和非参数法^[13]。其中,参数法需假设预测误差服从已知概率分布,如 Gaussian 分布、Gamma 分布和 Beta 分布等^[14]。预测误差实际分布往往难以事先确定,且不同的应用场景和分布函数以及区间预测结果差别很大,这些因素制约了参数法的应用。相反,非参数法是根据预测误差直接估算误差的分布,给出不同置信度水平下的预测输出区间,无需事先假设误差的分布形式。它能够更加准确地量化误差分布特征,有助于提高区间预测的准确性。常用的非参数法包括分位数回归、核密度估计和 Bootstrap 等^[15]。文献[12]用 Bootstrap 方法评估风电点预测误差,并计算置信区间,实验结果表明该方法具有较佳的性能。此外, Bootstrap 方法无需对样本分布作假设和新增样本,在小样本和数据分布规律未知情况下依然具有很好的鲁棒性和精度,还可以借助计算机模拟,具有很高的灵活性和实用性^[16]。

综上所述,如何进一步提高光伏功率预测模型的精度、鲁棒性及区间预测的可靠性仍然是一项具有挑战性的任务。为此,本文提出了一种基于 MPA-LSTM 网络模型的新型光伏功率区间预测方法。首先,针对 LSTM 网络的超参数难以确定的问题,采用 MPA 对 LSTM 网络的超参数(隐含层神经元数、训练批次数)自动寻优,并将优化后的 LSTM 模型用于光伏功率点预测;然后,采用 Bootstrap 方法分析

光伏功率预测误差,并结合点预测结果确定预测输出区间;最后,通过算例对比分析,验证所提出模型的有效性。

1 基于 MAP-LSTM 网络的区间预测方法

1.1 基于 MPA-LSTM 短期光伏功率点预测模型

为了提高基于 LSTM 网络光伏功率模型的预测精度,减少因传统人工试凑法设置 LSTM 超参数不合理对模型精度带来的不利影响,将一种新颖的群智能优化方法(MPA)用于 LSTM 网络的超参数寻优,从而提高光伏功率预测模型的精度和鲁棒性。

1.1.1 LSTM 网络

LSTM 网络是一种特殊结构的 RNN 网络,用具有记忆功能的 LSTM 单元替代 RNN 中隐含层神经元,解决了 RNN 的梯度消失和梯度爆炸问题。单层 LSTM 网络的结构图如图 1 所示。图 1 中 x_t 和 z_t 分别表示 t 时刻 LSTM 网络的输入和输出; x_{t-1} 和 z_{t-1} 分别表示 $t-1$ 时刻输入和输出; x_{t+1} 和 z_{t+1} 分别表示 $t+1$ 时刻输入和输出。一个 LSTM 单元主要由存储单元、遗忘门、输入门、输出门组成,LSTM 单元结构如图 2 所示。

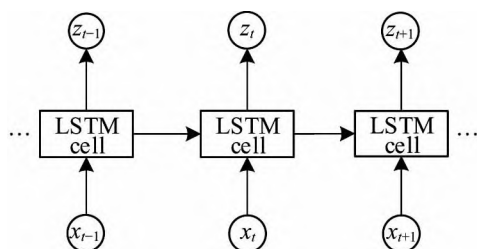


图 1 LSTM 网络结构模型

Fig 1 Structure of LSTM model

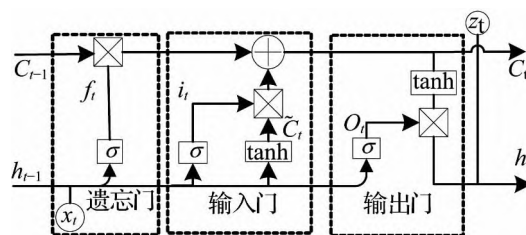


图 2 LSTM 单元结构

Fig 2 Structure of LSTM cell

LSTM 单元的内部运算过程可以通过以下公式来描述:

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f), \quad (1)$$

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i), \quad (2)$$

$$\tilde{C}_t = \tanh(U_c x_t + W_c h_{t-1} + b_c), \quad (3)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \quad (4)$$

$$O_t = \sigma(U_o x_t + W_o h_{t-1} + b_o), \quad (5)$$

$$h_t = O_t \otimes \tanh C_t, \quad (6)$$

式(1)~(6)中: x_t 表示当前时刻的输入; h_t 和 h_{t-1} 分别表示 t 时刻和 $t-1$ 时刻的状态; C_t 和 C_{t-1} 分别表示 t 和 $t-1$ 时刻存储单元的状态; f_t 为遗忘门; i_t 为输入门; O_t 代表输出门; U_f 、 W_f 、 U_i 、 W_i 、 U_c 、 W_c 、 U_o 、 W_o 为权重矩阵; b_f 、 b_i 、 b_c 和 b_o 为神经元的偏置量; $\sigma(\cdot)$ 和 $\tanh(\cdot)$ 为激活函数。

1.1.2 海洋捕食者算法(MPA)

MPA^[11]是一种新型群智能优化算法,灵感源于海洋生物适者生存理论,即模仿海洋捕食者通过 Lévy 游走或布朗游走选取最佳捕食策略过程。该算法可分为 3 个阶段,即初始阶段、优化阶段、涡流的形成与鱼类聚集装置(fish aggregating devices,FADs)影响。其具体的数学描述过程如下。

(1)初始阶段。MPA 通过随机生成初始解来启动算法。其数学表达式为:

$$X_{n,d} = X_{\min} + \text{rand}(X_{\max} - X_{\min}), \quad (7)$$

式中: X_{\max} 、 X_{\min} 分别是变量的上、下分布边界,rand()是一个 0~1 的随机变量; $X_{n,d}$ 为第 n 个猎物、第 d 维空间位置。

根据适者生存原理构建顶级捕食者的精英矩阵 E 和猎物矩阵 G ,它们的数学表达式为

$$\mathbf{E} = \begin{bmatrix} \mathbf{X}_{1,1}^1 & \mathbf{X}_{1,2}^1 & \cdots & \mathbf{X}_{1,D}^1 \\ \mathbf{X}_{2,1}^1 & \mathbf{X}_{2,2}^1 & \cdots & \mathbf{X}_{2,D}^1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{N,1}^1 & \mathbf{X}_{N,2}^1 & \cdots & \mathbf{X}_{N,D}^1 \end{bmatrix}, \quad (8)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{X}_{1,1} & \mathbf{X}_{1,2} & \cdots & \mathbf{X}_{1,D} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} & \cdots & \mathbf{X}_{2,D} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_{N,1} & \mathbf{X}_{N,2} & \cdots & \mathbf{X}_{N,D} \end{bmatrix}, \quad (9)$$

式中: \mathbf{X}^1 为顶级捕食者向量; \mathbf{E} 是由 \mathbf{X}^1 复制 N 次构成, 其中 N 为种群数; D 为搜索空间的维度。

(2) MPA 优化阶段。MPA 的优化阶段根据捕食者和猎物的速度变化, 分为 3 个阶段。

① 优化阶段 A ($k < \frac{1}{3}K$)。捕食者的速度比猎物速度要快, 算法主要执行勘探策略, 其数学表达式为

$$\begin{cases} S_n = \mathbf{R}_B \otimes (\mathbf{E}_n - \mathbf{R}_B \otimes \mathbf{G}_n), \\ \mathbf{G}_n = \mathbf{G}_n + P \cdot \mathbf{R} \otimes S_n, \end{cases} \quad n = 1, 2, \dots, N; \quad (10)$$

式中: S_n 为移动步长; \mathbf{R}_B 为呈正态分布的布朗随机向量; \otimes 为逐项乘法运算符; P 取值为 0.5; \mathbf{R} 为 $[0, 1]$ 内均匀随机向量; K 表示最大迭代次数; k 为当前迭代次数。

② 优化阶段 B ($\frac{1}{3}K < k < \frac{2}{3}K$)。在此阶段, 猎物与捕食者具有相同的速度, 捕食者基于布朗游走策略负责勘探, 猎物基于 Lévy 游走策略负责开发, 并逐渐由勘探策略转向开发策略。其数学描述为:

$$\begin{cases} S_n = \mathbf{R}_L \otimes (\mathbf{E}_n - \mathbf{R}_L \otimes \mathbf{G}_n), \\ \mathbf{G}_n = \mathbf{G}_n + P \cdot \mathbf{R} \otimes S_n, \end{cases} \quad n = 1, 2, \dots, N/2; \quad (11)$$

$$\begin{cases} S_n = \mathbf{R}_B \otimes (\mathbf{R}_B \otimes \mathbf{E}_n - \mathbf{G}_n), \\ \mathbf{G}_n = \mathbf{E}_n + P \cdot CF \otimes S_n, \end{cases} \quad n = \frac{N}{2} + 1, \dots, N; \quad (12)$$

式(11)~(12)中: \mathbf{R}_L 为呈 Lévy 分布的随机向量; CF 是控制捕食者移动步长的自适应参数, 它的表达式为 $CF = (1 - k/K)^{2k/K}$;

③ 优化阶段 C ($k > \frac{2}{3}K$)。在这一阶段中, 捕食者的速度将比猎物的要慢, 捕食者基于 Lévy 游走执行开发策略。其数学描述为

$$\begin{cases} S_n = \mathbf{R}_L \otimes (\mathbf{R}_L \otimes \mathbf{E}_n - \mathbf{G}_n), \\ \mathbf{G}_n = \mathbf{E}_n + P \cdot CF \otimes S_n, \end{cases} \quad n = 1, 2, \dots, N; \quad (13)$$

(3) 涡流效应与 FADs 影响。FADs 或者涡流效应会改变海洋捕食者的觅食行为, 这一策略有利于算法克服早熟问题并跳出局部最优, 从而实现全局优化。其数学表达式为

$$\mathbf{G}_n = \begin{cases} \mathbf{G}_n + CF[X_{\min} + \mathbf{R} \otimes (X_{\max} - X_{\min})] \otimes \mathbf{U}, & r \leq FAD \\ \mathbf{G}_n + [FAD(1 - r) + r](\mathbf{G}_{r_1} - \mathbf{G}_{r_2}), & r > FAD \end{cases} \quad (14)$$

式中: FAD 取值为 0.2; \mathbf{U} 为二进制向量, r 为 $[0, 1]$ 内的随机数, r_1, r_2 分别为猎物矩阵的随机索引。

1.1.3 基于 MPA 优化的 LSTM 网络模型

LSTM 网络的隐含层神经元数和训练批次数对模型的预测性能具有重要影响, 合理选择这 2 个参数有益于提高预测模型的性能, 因此, 将这 2 个参数作为 MPA 寻优的特征, 利用 MPA 对 LSTM 网络模型超参数寻优。MPA 方法对 LSTM 网络的关键超参数优化的流程框图如图 3 所示, 具体实现步骤如下:

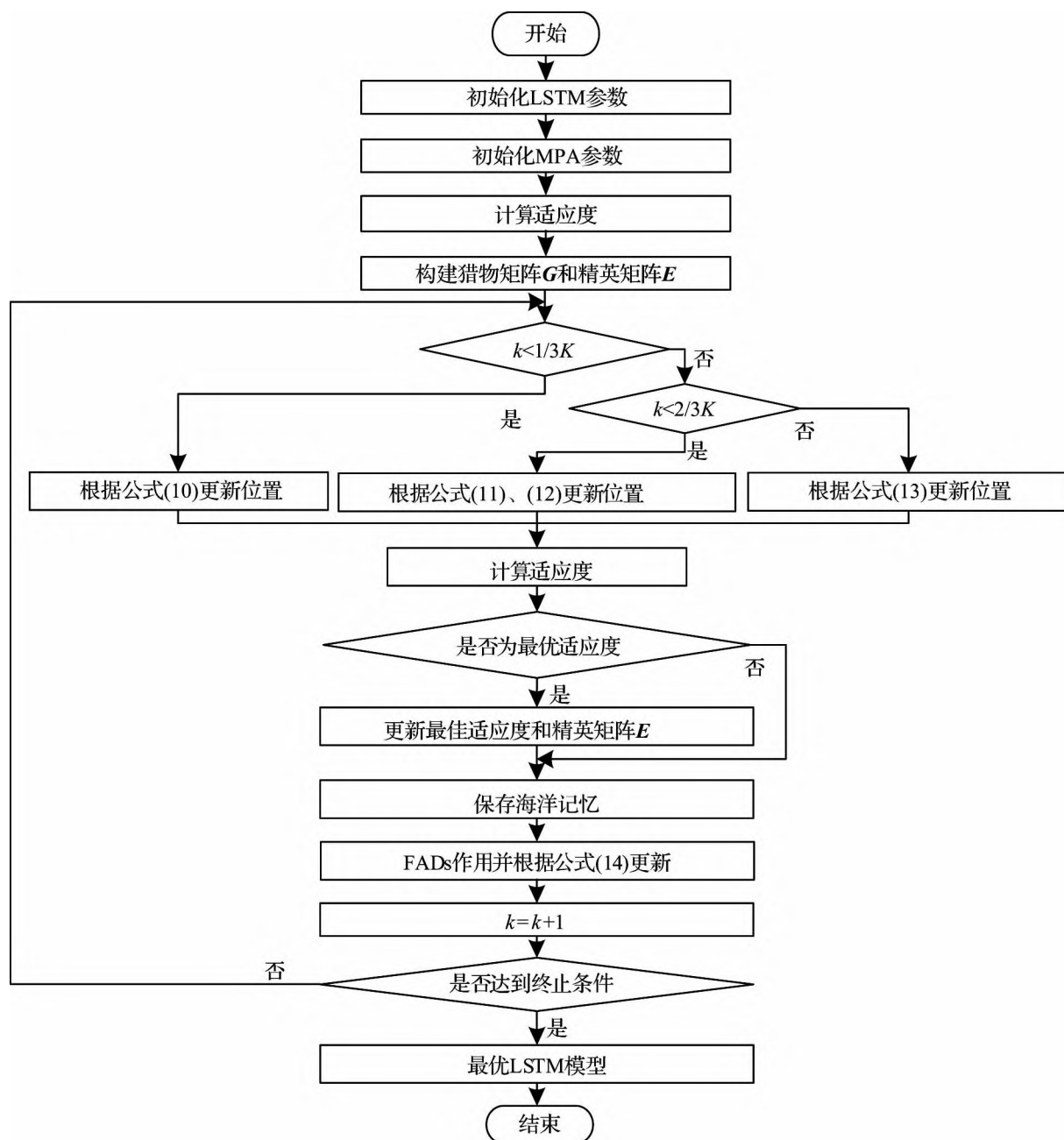


图 3 MPA-LSTM 的流程框图

Fig 3 Flow chart of MPA-LSTM

① 设置 LSTM 网络隐含层神经元个数和训练批次数的范围, 根据 LSTM 网络参数范围初始化 MPA 的参数。

② 根据捕食者的位置更新 LSTM 网络参数, 并计算训练平均绝对误差 (mean absolute error, MAE), 将 MAE 作为 MPA 中个体适应度函数, 判断是否要更新精英矩阵 E 。

③ 根据公式(10)–(14)更新参数, 重复步骤②。

④ 判断是否达到终止条件, 如果没有达到, 则重复步骤③。当达到终止条件时, 精英矩阵 E 的参数值即为 LSTM 网络的最优参数。

1.2 基于 Bootstrap 方法的置信区间估计

Bootstrap 方法是由美国统计学教授 Efron 提出, 具有计算方便、实现容易等优点, 在数据分析领域

得到广泛应用^[17]。该方法通过假设给定一组数据集服从一个未知分布,并且这个分布的总体方差也未知,通过对给定数据集进行反复抽样,根据抽样样本估计样本的分布区间。具体的实现过程如下:

① 假设从原始样本 y_1, y_2, \dots, y_m 进中行又放回抽取 M 次,得到 1 个 Bootstrap 样本集 Y_1 ;按照上述步骤重复 B 次,即可得到 B 个 Bootstrap 样本集, $Y^B = (Y_1, Y_2, \dots, Y_b)$, $b=1, 2, \dots, B$;

② 对于每个子样本 y^* , 由于 y^* 的分布与 y 的分布一致,因此都是 \hat{y} 的一个估计值。 \hat{y}_b^* 代表第 b 个 Bootstrap 样本,则样本估计均值和样本方差计算公式如下:

$$\Delta Y = \frac{1}{B} \sum_{b=1}^B \hat{y}_b^*, \quad (15)$$

$$\text{Var}^*(y) = \frac{1}{B-1} \sum_{b=1}^B (y_b^* - \Delta Y)^2, \quad (16)$$

式中: y_b^* 为第 b 个子样本的估计值; ΔY 和 $\text{Var}^*(y)$ 分别为样本均值和方差。

③ 计算样本在置信度 $1-\alpha$ 下的置信区间的上下限 Y_{up} 和 Y_{low} :

$$Y_{\text{low}} = \hat{y} - Z_{1-\alpha/2} \cdot \text{Var}^*(y), \quad (17)$$

$$Y_{\text{up}} = \hat{y} + Z_{1-\alpha/2} \cdot \text{Var}^*(y), \quad (18)$$

式中 $Z_{1-\alpha/2}$ 是标准正态分布 $100 \cdot (1-\alpha)\%$ 对应的临界值。

1.3 区间预测方法

本文选取两步法实现光伏功率区间预测。首先通过 MPA-LSTM 模型实现短期光伏功率的点预测;然后,采用 Bootstrap 方法对预测误差分析,获得给定置信水平下的误差分布区间,通过综合点预测的结果和误差分布区间,实现区间预测。本文所提方法的区间预测流程图如图 4 所示。

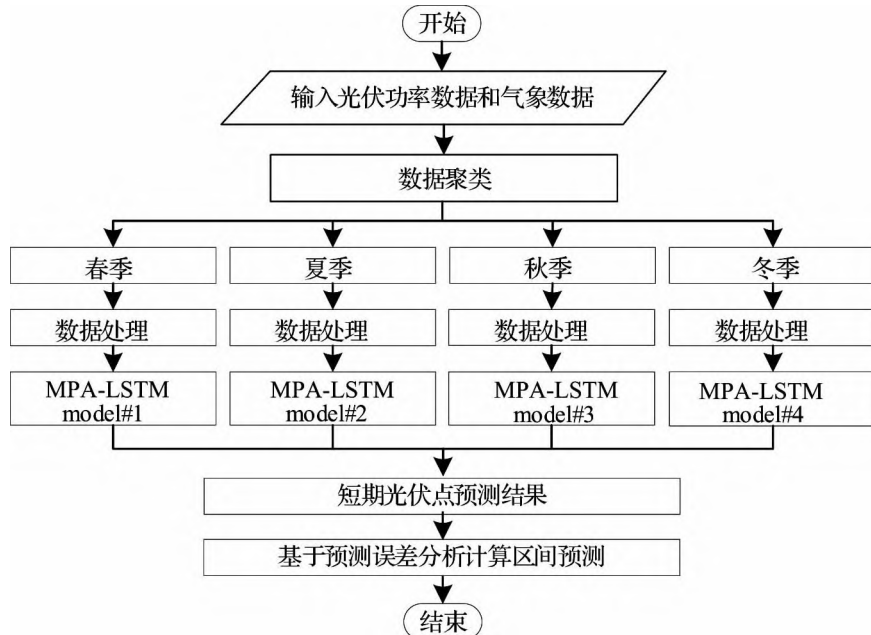


图4 区间预测流程图

Fig 4 Flow chart of interval prediction

其实现的具体过程可具体描述为:

① 数据集的划分。由于地球的公转影响,全球的大部分地区的辐照度和温度等因素都呈现季节性周期变化,因此将样本数据按季节划分为春季、夏季、秋季和冬季 4 个子集,以便于更好地训练模型。

② 数据预处理。将光伏功率数据及对应的辐照度等数据根据数据边界原则,将异常数据修正、对缺失数据进行插值补全。由于数据集中每一个输入、输出特征量的量纲并不一样,为了加快模型训练速

度和减小误差,利用最大最小值原理对数据进行归一化处理。

③ 将子数据集分别分为训练集和测试集,用训练集用于模型训练,并获取最优的 MPA-LSTM 网络;利用测试集验证优化后 MPA-LSTM 网络模型,实现光伏功率点预测。

④ 采用 Bootstrap 方法分析预测误差样本,计算给定置信区间下的预测误差的分布边界值。

⑤ 通过对 MPA-LSTM 模型的点预测结果和误差分布区间综合分析,计算光伏功率的区间预测结果。

2 算例分析

为了验证本文所提出方法的有效性,本文选取美国标准工程实验室设在马里兰校区内的光伏示范电站的地面光伏阵列作为研究对象,其安装容量为 271 kW^[18]。该电站提供大量不同时间尺度的光伏电站数据及气象数据,特别适合光伏电站功率预测的精细建模。本文选取了时间跨度为 2015 年 12 月到 2016 年 12 月的历史数据。为了使模型更具普适性、降低模型的复杂性,仅选取辐照度、温度、风速和风向作为光伏输出功率的主要影响因素。由于原始历史数据的时间分辨率为 1 min,因此首先通过预处理,获得时间分辨率为 15 min 的光伏功率数据和气象数据。然后,依据季节特性,将数据划分为春、夏、秋和冬 4 个子数据集。此外,为了加快训练速度、降低输入变量不同量纲对模型的影响,对每个子数据的数据进行归一化处理。最后,分别从归一化后子数据集随机选取 10% 的数据作为测试数据集,其余数据作为训练集。模型的输入为辐照度、温度、风速和风向,输出为光伏功率。

本文实验的实验环境如下: Win10 操作系统,处理器为 Intel (R) Core (TM) i5-9400F CPU @ 2.90 GHz, 16 GB RAM,无 GPU 卡;所有仿真结果均在基于 Python3.7.1 的 Tensorflow 2.1.0 环境下完成。

2.1 评价指标

为了量化分析所建光伏功率预测模型的精度,本文选取均方根误差(root mean square error, RMSE)、平均绝对误差(MAE)作为点预测模型的典型评价指标^[15]。对于区间预测结果,本文选取预测区间覆盖率(prediction interval coverage probability, PICP)、预测区间平均宽度(prediction interval normalized average, PINAW)和温克勒系数(Score)等指标评价模型^[13]。上述评价指标在有关光伏功率点预测与区间预测上得到广泛的应用,限于篇幅,对上述评价指标的详细定义不再赘述。

2.2 光伏功率点预测结果及对比分析

为了验证本文所提出模型的性能的优越性,本文还建立了基于 SVM、LSTM、PSO-LSTM、MPA-SVM 网络的类似预测模型作为对照模型。为了公平起见,PSO 和 MPA 的种群数也都设置为 30,迭代次数均为 500, LSTM 和 SVM 模型的训练迭代次数为 200。此外,本文所建的不同模型均在相同的训练集和测试集中完成了训练和测试。

为了更加直观地了解实际值与预测值之间的偏差,本文从 4 个子测试数据集中分别随机选取 1 d 的预测结果作为展示。光伏功率点预测结果如图 5 所示,其中(a)、(b)、(c)和(d)分别是 4 个季节的测试集中某一天不同方法的预测结果。

从图 5 可知,早上 7:00 时,春、夏两季已经有功率输出,而秋、冬季功率为 0。夏季的输出值最大且稳定性好,秋季的输出功率小于春、夏两季的,但大于冬季的;从功率曲线可知光伏功率存在季节性差异。图中带实心圆的曲线为实际功率曲线,含倒三角形的黑色曲线为本文所提模型的预测曲线,其他曲线分别为对比模型预测曲线。从图中可知,在功率波动平缓的区域,各个方法的结果差异不大,但是在波动剧烈的部分,结果差异较大。LSTM 和 SVM 模型的输出与实际输出功率的差距较大,预测性能较差。PSO-LSTM 和 MPA-SVM 模型的预测曲线与实际功率曲线的拟合程度要好于 SVM 和 LSTM 模型的,MPA-LSTM 模型的预测曲线与实际曲线差距最小,具有比较优异的预测性能,表明超参数优化后的预测模型精度更高。单纯从曲线的形状往往难以判别实际模型性能,因此,本文还分别统计了所提模型与对比模型在整个测试集上预测结果的 RMSE 和 MSE,MPA-LSTM 及对比模型的点预测指标如表 1 所示。

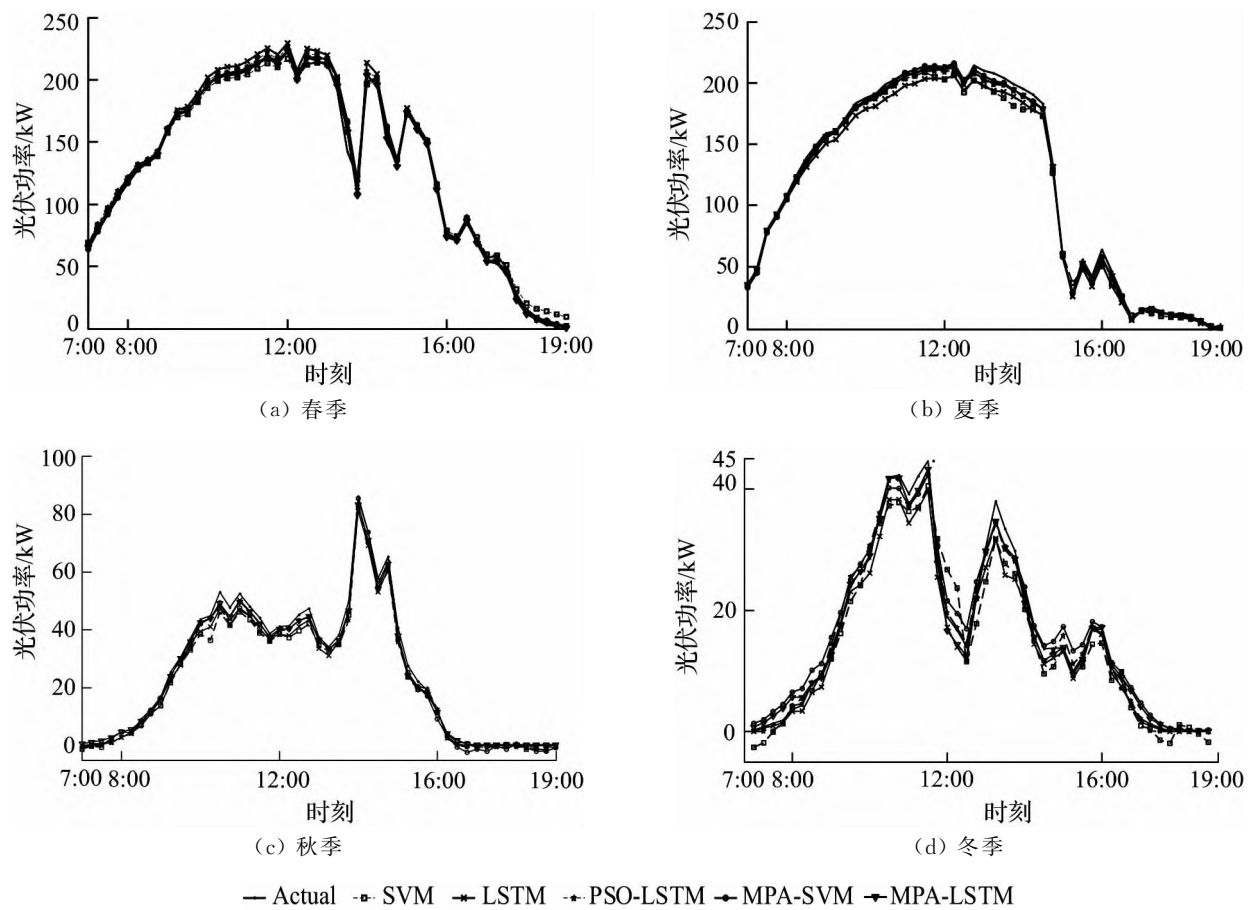


图 5 光伏功率点预测结果

Fig 5 Results of PV power point forecasting

表 1 点预测指标

Tah 1 Indices for point forecasting

季节	误差类型	SVM	LSTM	PSO-LSTM	MPA-SVM	MPA-LSTM
春季	RMSE/%	2.24	1.97	1.73	1.65	1.51
	MAE/kW	5.05	2.77	2.31	2.27	1.86
夏季	RMSE/%	2.10	1.94	1.24	1.23	1.16
	MAE/kW	3.17	3.02	2.14	2.59	1.89
秋季	RMSE/%	1.59	1.00	0.95	1.16	0.92
	MAE/kW	2.86	1.23	1.38	1.97	1.32
冬季	RMSE/%	1.30	1.10	1.02	0.82	0.77
	MAE/kW	2.82	1.54	1.27	1.23	1.02
平均	RMSE/%	1.81	1.51	1.24	1.22	1.09
	MAE/kW	3.48	2.14	1.78	2.02	1.52

从表 1 可知,相比于其他模型,基准 SVM 模型在 4 个季节中 RMSE 和 MAE 的值最大,由此可知 SVM 模型的预测效果最差;LSTM 模型的 RMSE 和 MAE 值都小于 SVM 模型的。MPA-SVM 模型的 RMSE 和 MAE 平均值分别比 SVM 模型的小 0.59% 和 1.46 kW,但是与 PSO-LSTM 模型的结果相近。MPA-LSTM 模型在春、夏、秋和冬季 4 个季节中, RMSE 和 MAE 的值都小于 SVM、LSTM、PSO-LSTM 和 MPA-SVM 模型的。MPA-LSTM 模型的 RMSE 和 MAE 的平均值为 1.09% 和 1.52 kW;与 SVM 等对比模型相比, RMSE 分别降低了 0.72%、0.42%、0.15% 和 0.13%。从结果可知, MPA-

LSTM 模型对于不同的季节都有较好的预测效果,LSTM 模型的预测精度比 SVM 模型的更高。同时通过 MPA 改进的 LSTM 模型精度比基准 LSTM 模型和 PSO-LSTM 模型有了提高,说明相比于 PSO 模型,MPA 具有更好的勘探和开发能力。

2.3 区间预测结果及对比分析

由于光伏电站上空云层分布和周围的天气变化具有很大的随机性,分析光伏输出功率的区间范围更具有意义,因此本文采用 Bootstrap 方法分析 MPA-LSTM 模型的预测误差分布,然后根据误差分布结果与点预测结果,求解光伏功率概率区间。分别讨论置信度为 80%、90%和 95%条件下光伏功率区间的分布情况,并与参数法中的 Gaussian 分布和 Gamma 分布(分别假设预测误差分别服从 Gaussian 分布和 Gamma 分布)作对比。为了方便结果展示,区间结果与点预测选取同一天。Bootstrap 方法区间预测结果如图 6 所示,图 6 中描述了置信度为 80%、90%和 95%的 MPA-LSTM 模型的预测区间,以及实际功率值落在区间内的情况。从图中可知,随着置信度增大,预测区间宽度变宽;与此同时,落在预测区间内实际光伏功率的点数也随之增多,但是从春季的区间结果可知,仍然存在少部分实际功率值会落在预测区间以外,这可能是由光伏功率的预测偏差或所采集的数据有误造成。

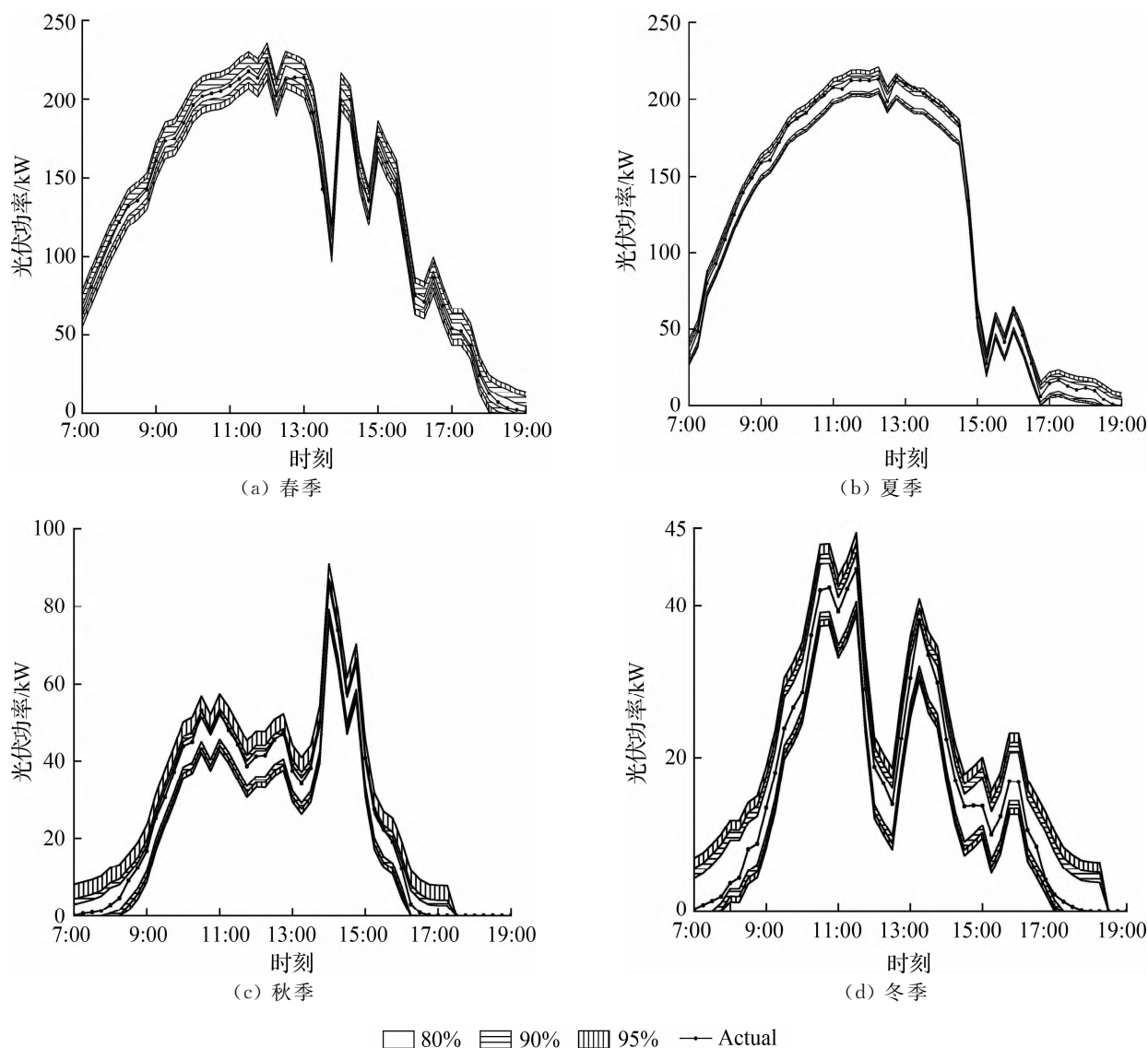


图 6 区间预测结果

Fig 6 Results of interval prediction

分别计算 Bootstrap 方法和 Gamma 分布、Gaussian 分布法的 PICP, PINAW 和 Score 指标, Bootstrap 方法及对比方法的区间预测指标如表 2 所示。

表 2 区间预测指标
Tab 2 Indices of interval prediction

季节	指标	方法	置信水平		
			80%	90%	95%
春季	PICP/%	Gamma	91.9	93.1	94.0
		Gaussian	90.7	92.5	93.1
		Bootstrap	86.9	92.6	96.4
	PINAW	Gamma	0.029	0.037	0.044
		Gaussian	0.025	0.032	0.038
		Bootstrap	0.015	0.036	0.052
	Score	Gamma	-5.140	-3.464	-2.235
		Gaussian	-5.007	-3.517	-2.435
		Bootstrap	-4.892	-3.340	-2.029
夏季	PICP/%	Gamma	89.8	93.3	95.5
		Gaussian	86.6	90.9	93.6
		Bootstrap	87.6	93.6	96.6
	PINAW	Gamma	0.022	0.028	0.033
		Gaussian	0.019	0.024	0.028
		Bootstrap	0.019	0.027	0.036
	Score	Gamma	-3.495	-2.148	-1.249
		Gaussian	-3.506	-2.245	-1.383
		Bootstrap	-3.494	-2.144	-1.204
秋季	PICP/%	Gamma	93.8	95.8	97.0
		Gaussian	92.8	94.7	95.8
		Bootstrap	91.2	95.5	97.8
	PINAW	Gamma	0.014	0.018	0.021
		Gaussian	0.012	0.015	0.018
		Bootstrap	0.010	0.014	0.021
	Score	Gamma	-2.294	-1.496	-0.971
		Gaussian	-2.236	-1.504	-1.003
		Bootstrap	-2.208	-1.515	-0.944
冬季	PICP/%	Gamma	92.7	96.5	97.4
		Gaussian	89.5	93.9	96.6
		Bootstrap	90.0	94.6	97.5
	PINAW	Gamma	0.013	0.016	0.019
		Gaussian	0.011	0.014	0.017
		Bootstrap	0.010	0.015	0.018
	Score	Gamma	-2.010	-1.284	-0.828
		Gaussian	-1.999	-1.292	-0.841
		Bootstrap	-1.997	-1.270	-0.809

从表 2 中可知,随着置信水平增大,PICP 和 PINAW 的值都会增大。在 4 个季节中,当置信度为 80%和 90%时,Gamma 分布方法的 PICP 值都大于 Bootstrap 方法的,但是 PINAW 的值也比 Bootstrap 方法的大;Gaussian 分布方法的 PICP 值在春季和秋季要大于 Bootstrap 方法,而在夏、冬季则小于 Bootstrap 方法。当置信度为 95%时,Bootstrap 方法在 4 个季节中的 PICP 值分别为 96.4%、96.6%、97.8%和 97.5%,均大于 95%,但是在 Gamma 分布方法春季时分别只有 94.0%;Gaussian 分布方法在春季和夏季中,PICP 分别只有 93.1%和 93.6%,这 2 种方法并不满足给定需求,因此,Gamma 和 Gaussian 分布方法并不能很好描述光伏功率的误差分布。同时,从表中可知,Bootstrap 方法的 Score 的值在不同的季节和置信度下,大部分时候都是最大的,仅有在秋季的置信水平为 90%时,它的值略小于 Gamma 分布方法的。Score 是一个负数,值越大,则代表预测效果越好,因此,总的来说 Bootstrap 方法在不同的季节和不同置信度之下都能很好地描述光伏功率地误差分布,表现出优异的区间预测性能,能够提供全面的预测信息。

综合点预测和区间预测结果可知,本文所提出的 MPA-LSTM 模型对光伏功率点预测有较高的精度,Bootstrap 方法能准确描述误差分布,验证了所提出的方法有效性和实用性。

3 结论

针对目前光伏功率模型预测精度仍有待提高、以及因缺乏光伏电站功率输出误差分布先验知识而难以对光伏输出功率波动区间进行准确估计等问题,本文提出一种基于 MPA-LSTM 网络的光伏功率短期点预测模型,首先利用 MPA 对 LSTM 网络的关键参数隐含层神经元数和训练批次数优化,然后将优化后的 MPA-LSTM 模型用于预测光伏功率;然后,采用 Bootstrap 方法分析点预测的预测分布,并结合点预测结果实现其区间预测。实验结果表明:

① 与其他群智能优化算法相比,MPA 具有更好的局部开发和全局搜索能力。

② 采用 MPA 对 LSTM 的参数寻优后,可以减少人为因素对模型效果的影响,能够提高基于 LSTM 网络模型预测精度。

③ 与参数法中的 Gamma 和 Gaussian 分布相比,Bootstrap 方法可以直观评估不同置信度水平下光伏功率预测误差的分布,并且满足实际需求,而且该方法无需事先假设误差分布功率及其参数,更有利于工程应用。

参考文献:

- [1] TAKILALTE A, HARROUNI S, YAICHE M R, et al. New approach to estimate 5-min global solar irradiation data on tilted planes from horizontal measurement[J]. Renewable Energy, 2020,145:2477-2488.
- [2] ZANG H, CHENG L, DING T, et al. Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning [J]. International Journal of Electrical Power & Energy Systems, 2020,118:105790.
- [3] 余光正,陆柳,汤波,等. 基于云图特征提取的改进混合神经网络超短期光伏功率预测方法[J]. 中国电机工程学报, 2021,41(20):6989-7003.
- [4] AHMED R, SREERAM V, MISHRA Y, et al. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization[J]. Renewable and Sustainable Energy Reviews, 2020,124:109792.
- [5] 王晨阳,段倩倩,周凯,等. 基于遗传算法优化卷积长短记忆神经网络模型的光伏发电功率预测[J]. 物理学报, 2020,69(10):149-155.
- [6] CHENG G, SONG S, LIN Y, et al. Enhanced state estimation and bad data identification in active power distribution networks using photovoltaic power forecasting[J]. Electric Power Systems Research, 2019,177:105974.
- [7] WANG K, QI X, LIU H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning

- neural network[J]. Applied Energy, 2019,251:113315.
- [8] 黄亚峰,何威,吴光琴,等. 基于K-means++和LSTM网络的光伏功率预测研究[J]. 电气自动化, 2020,42(5): 25-27.
- [9] 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.
- [10] ZHENG J, ZHANG H, DAI Y, et al. Time series prediction for output of multi-region solar power plants[J]. Applied Energy, 2020,257:114001.
- [11] FARAMARZI A, HEIDARINEJAD M, MIRJALILI S, et al. Marine predators algorithm: a nature-inspired meta-heuristic[J]. Expert Systems with Applications, 2020,152:113377.
- [12] 王焱,汪震,黄民翔,等. 基于OS-ELM和Bootstrap方法的超短期风电功率预测[J]. 电力系统自动化, 2014, 38(6):14-19.
- [13] PAN C, TAN J, FEND D. Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation[J]. Neurocomputing, 2021,453:552-562.
- [14] 赵康宁,蒲天骄,王新迎,等. 基于改进贝叶斯神经网络的光伏出力概率预测[J]. 电网技术, 2019,43(12): 4377-4386.
- [15] LIU L, ZHAO Y, CHANG D, et al. Prediction of short-term PV power output and uncertainty analysis[J]. Applied Energy, 2018,228:700-711.
- [16] WANG X, GHIDAoui M S, LIN J. Confidence interval localization of pipeline leakage via the bootstrap method [J]. Mechanical Systems and Signal Processing, 2022,167:108580.
- [17] 薛阳,张宁,俞志程,等. 基于BiLSTM和Bootstrap方法的风电功率区间预测[J]. 可再生能源, 2020,38(8): 1059-1064.
- [18] BOYD M. Performance data from the nist photovoltaic arrays and weather station[J]. Journal of Research of the National Institute of Standards and Technology, 2017,122:1-4.

(责任编辑 刘炜)