

一种新型分布式光伏出力区间预测方法

杨康¹, 李蓝青¹, 李艺丰¹, 宋东阔², 王博仑¹, 陈金², 周霞^{3*}, 单宇³

(1. 国网江苏省电力有限公司, 江苏省 南京市 210008; 2. 国电南瑞科技股份有限公司, 江苏省 南京市 211000; 3. 南京邮电大学自动化学院/人工智能学院, 江苏省 南京市 210023)

A Novel Distributed Photovoltaic Output Interval Prediction Method

YANG Kang¹, LI Lanqing¹, LI Yifeng¹, SONG Dongkuo², WANG Bolun¹, CHEN Jin², ZHOU Xia^{3*}, SHAN Yu³
(1. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210008, Jiangsu Province, China; 2. Guodian NARI Technology Co., Ltd., Nanjing 211000, Jiangsu Province, China; 3. College of Automation/College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu Province, China)

摘要:【目的】分布式光伏功率预测对光伏电站运行和调度具有重要意义, 针对点预测方法难以全面描绘分布式光伏功率不确定性的问题, 提出了一种基于自适应噪声完备集合经验模态分解 (complete ensemble empirical mode decomposition with adaptive noise, CEEMDAN) 和麻雀搜索算法优化的最小二乘支持向量机 (sparrow search algorithm optimized least square support vector machine, SSA-LSSVM) 分布式光伏功率区间预测模型。【方法】首先, 通过 CEEMDAN 将光伏功率序列分解为多个模态分量, 再对一次分解得到的高频非平稳分量进行二次分解; 其次, 采用样本熵 (sample entropy, SE) 将所有分量重构为趋势分量和振荡分量; 然后, 通过 SSA-LSSVM 得到 2 个分量的点预测值; 最后, 对振荡分量的点预测误差进行概率密度估计, 叠加点预测值得到总体的预测区间结果。【结果】所提区间预测模型具有更高的区间覆盖率且区间平均宽度更窄。【结论】在分布式光伏功率数据处理中加入二次模态分解, 再结合样本熵对其子序列进行重构, 可有效降低原始预测分量的复杂程度, 同时提升模型预测准确性。

关键词: 分布式光伏; 功率预测; 核密度估计; 区间预测
ABSTRACT: [Objectives] Distributed photovoltaic power prediction is of great significance for the operation and scheduling of photovoltaic power plants. Point prediction methods are difficult to comprehensively describe the uncertainty of distributed photovoltaic power. This article proposed a distributed photovoltaic power interval prediction model based on complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and sparrow

search algorithm optimized least squares support vector machine (SSA-LSSVM). [Methods] Firstly, the photovoltaic sequence was broken down into multimodal components through CEEMDAN, and then the high-frequency non-stationary components obtained from the first decomposition were decomposed twice. Secondly, sample entropy (SE) was used to reconstruct all components into trend and oscillation components. Then, the point prediction values of the two components were obtained through SSA-LSSVM. Finally, the probability density estimation was performed on the point prediction error of the oscillation component, and the stacked point prediction value was used to obtain the overall prediction interval result. [Results] The interval prediction model proposed in this paper has higher interval coverage and narrower average interval width. [Conclusions] Adding secondary modal decomposition to distributed photovoltaic power data processing and combining sample entropy to reconstruct its sub-sequences can effectively reduce the complexity of the original prediction components and improve the accuracy of model prediction.

KEY WORDS: distributed photovoltaic; power prediction; kernel density estimation; interval prediction

0 引言

分布式光伏发电的出力预测是光伏电站运行和调度的重要组成部分^[1-6], 对于提高光伏电站的发电量、降低发电成本具有重要意义。然而, 分布式光伏发电因受天气和环境等因素的影响, 导致光伏发电出力存在一定的波动性和不确定性^[7-8]。基于此, 为了确保电网运行的安全性, 提高电网的经济性, 国内外学者已经提出了多种将不确定

基金项目: 国家自然科学基金项目(61933005)。

Project Supported by National Natural Science Foundation of China (61933005).

性因素纳入其中的电网规划和调度方法^[9-10]。近年来,光伏发电的点预测模型成为研究的热点。从预测方法上讲,点预测方法主要分为物理研究法^[11]和统计研究法。物理法主要是将数值天气预报信息和分布式光伏电站详细的地理地貌信息相结合,预测精确度主要受天气预报准确性的影响。文献[12]将气象信息的时间变化纳入到时间序列中,提出了一种新的混合预测方法。统计法是建立在对光伏发电的历史统计数据的基础上,通过数据训练、曲线拟合等方法构建光伏发电系统和影响因素的关系。常见的方法有神经网络法^[13]、支持向量机^[14]、时间序列预测法^[15]等。然而,由于光伏功率受气象因素影响导致的波动性和不平稳性,点预测模型难以全面描绘功率变化特征。

为了对电网调度、可靠性评估等决策提供理论依据,同时为光伏出力的波动性提供新的定量分析,研究者们日益关注区间预测方法。区间预测是一种对光伏发电不确定性造成的预测结果偏差进行量化的强大工具,它在给出点预测结果的同时,还勾勒出了光伏发电功率波动的区间。文献[16]对时间序列数据二次分解处理后再进行预测,提升了一次分解中高频部分的预测准确率,但是所产生的子分量数目太多,单独进行预测所需的时间比较长。文献[17]提出了一种利用极限学习机与蒙特卡洛方法相结合的风力发电功率区间预测方法,取得了较好的预测效果。文献[18]提出一种基于信息熵变权区间组合和边界逼近的短期光伏功率区间预测方法。然而在不同的情形下,光伏出力的真实概率分布形状会有很大的差别,因此仅仅使用固定形状的先验分布很难达到精确建模。文献[19]提出一种将二次模态分解和深度学习相结合的短期风电功率区间预测方法,通过预测结果对比,说明完备集合经验模态分解比变分模态分解在功率信号分解上更有优势。文献[20-22]从不同的研究角度,对光伏功率区间预测的方法进行了研究,但是鲜有针对降低原始光伏功率序列的波动性、随机性来提升光伏功率区间预测的精准度的方法。

鉴于以上方法的不足,本文提出了一种基于自适应噪声完备集合经验模态分解(complete

ensemble empirical mode decomposition with adaptive noise, CEEMDAN)和麻雀搜索算法优化的最小二乘支持向量机(sparrow search algorithm-least square support vector machine, SSA-LSSVM)分布式光伏功率区间预测模型。通过实例验证,本文所提区间预测模型相比其他多个方法具有更高的预测准确性。

1 光伏功率分解和重构

1.1 光伏历史出力数据相关性分析

为了从历史数据中抽取出所需的输入特性,在进行预测之前,必须先对这些数据进行相关性分析。文中所讨论的相关系数主要指在某个时间点上的输出功率与前若干天的同一时间点上的输出功率和当天前若干时间点上的输出功率之间的关系。因为本区域一整年的电力有效输出时段为08:00—18:00,共计11 h。所以,在分析数据时,只讨论这一时期的功率关系。本文选择了2018年6月1日到2018年6月31日的光伏功率历史数据,在每天的08:00—18:00,以15 min为间隔,对功率数据进行采样,获得了41个样本的功率值。

1) 水平相关性

水平相关指要预测的功率点与前若干天同一时间的功率点之间的关系。将第*i*天的功率值记为 W_i ,第*j*天的功率值记为 W_j ,采用皮尔逊相关系数法,则任意2列的功率值的相关性计算公式为

$$P(W_i, W_j) = \frac{\text{Cov}(W_i, W_j)}{\sigma_{W_i} \sigma_{W_j}} \quad (1)$$

式中: $\text{Cov}(W_i, W_j)$ 表示第*i*天与第*j*天功率的协方差; σ_{W_i} 和 σ_{W_j} 分别为各自功率的方差。

需要计算的水平相关性流程如下:

$$\begin{cases} W = (W_1, W_2, \dots, W_i, \dots, W_n)_{41 \times 365} \\ r_i = (p_{i-s,i}, \dots, p_{i-1,i}), i = 1, \dots, 365 \\ R = (r_1, r_2, \dots, r_i, \dots, r_{n-s})_{(365-s) \times 5} \\ \bar{R}_j = \sum_{i=1}^{365-s} \frac{r_{ij}}{365-s}, j = 1, \dots, s \end{cases} \quad (2)$$

式中: W_i 为 41×1 的向量; n 为一年的天数,即365天; $p_{i,j}$ 为任意2列数据的相关性大小; r_i 为第*i*天与前*s*天同一时刻功率的相关性向量。

2) 垂直相关性

垂直相关分析指的是被预测的功率点与当天前若干时间点上的输出功率之间的关系，一天中一共有41个样本。按照天数排序，总共365天。第*i*列即为41个采样点中第*i*个采样点共计365天的数据，将 W 矩阵进行转置即可得到所需分析的数列 W_{TRS} 。则垂直相关性的计算公式如下：

$$\begin{cases} W_{TRS} = (W_1, W_2, \dots, W_i, \dots, W_n)_{365 \times 41}^T \\ m_i = (o_{i-s,i}, \dots, o_{i-1,i}), i = 1, \dots, 41 \\ R_c = (m_1, m_2, \dots, m_i, \dots, m_{41-s})_{(41-s) \times s}^T \\ \bar{R}_j = \sum_{i=1}^{n-s} \frac{m_{ij}}{41-s}, j = 1, \dots, s \end{cases} \quad (3)$$

式中 m_i 为第*i*时刻与当天前*s*个时刻功率的相关性向量。

1.2 基于CEEMDAN和变分模态分解的光伏功率序列模态分解

1.2.1 CEEMDAN分解算法

考虑到经验模态分解(empirical mode decomposition, EMD)存在模态混叠的因素，CEEMDAN是在EMD法的基础上对其进行改进，并借鉴了EEMD法中加入高斯型噪声及采用多次迭加求平均的方法消除噪声的思路。相比集合经验模态分解(ensemble empirical mode decomposition, EEMD)算法，自适应噪声完备集合经验模态分解有效地减少了信号重建的误差并使分解的效率得到改善。CEEMDAN具体流程如下：

1) 将待分解信号 $X(t)$ 添加*k*次均值为0的高斯白噪声，构造共*k*次实验的待分解序列 $X_i(t)$ ，其中*i*=1,2,3,...,*k*。

$$X_i(t) = X(t) + \varepsilon \delta_i(t) \quad (4)$$

式中： ε 为高斯白噪声权值系数； $\delta_i(t)$ 为第*i*次处理时产生的高斯白噪声。

2) 对上述序列 $X_i(t)$ 进行EMD分解，分解得到第1个模态分量(intrinsic mode functions, IMF)并取其均值作为CEEMDAN分解得到的第1个IMF。

$$\begin{cases} I_{MF1}(t) = \frac{1}{k} \sum_{i=1}^k I_{MF1}^i(t) \\ r_1(t) = X(t) - I_{MF1}(t) \end{cases} \quad (5)$$

式中： $I_{MF1}(t)$ 为CEEMDAN分解得到的第1个模态分量； $r_1(t)$ 为第1次分解后的余量信号。

3) 将分解后得到的第*j*阶段余量信号添加特定噪声后，继续进行EMD分解。

$$\begin{cases} I_{MFj}(t) = \frac{1}{k} \sum_{j=1}^k E_1 \{r_{j-1}(t) + \varepsilon_{j-1} E_{j-1}[\delta_1(t)]\} \\ r_j(t) = r_{j-1}(t) - I_{MFj}(t) \end{cases} \quad (6)$$

中： $E_{j-1}(\cdot)$ 表示对序列进行EMD分解后的第*j*-1个IMF分量； ε_{j-1} 表示CEEMDAN对第*j*-1阶段余量信号加入噪声的权值系数。

4) 如果满足EMD停止条件，第*n*次分解的余量信号 $r_n(t)$ 为单调信号，则迭代停止，CEEMDAN算法分解结束。

1.2.2 变分模态分解算法

由于CEEMDAN分解得到的IMF分量中存在高频非平稳分量，若对此分量直接进行预测会产生较大的误差，二次分解将一次分解得到的高频非平稳分量分解为具有周期性、较为平稳的分量，有助于改善其平稳度，提升预测精度。各分量的样本熵越大，表示信号越不平稳，对熵值最大的分量进行二次分解。

在变分模态分解(variational mode decomposition, VMD)中，泛函指的是VMD约束变分模型，而要求的极值就是每个模态分量中心频率的带宽之和最小。

构造出来的VMD约束变分模型如下：

$$\begin{cases} \min_{\{u_k\}, \{w_k\}} \left\{ \sum_k \left\| \partial_t \left\{ [\delta(t) + \frac{j}{\pi t}] * u_k(t) \right\} e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_k u_k = f \end{cases} \quad (7)$$

式中： $\{u_k\}$ 与 $\{w_k\}$ 分别为VMD分解出来的*k*个子模态分量集合和其对应的中心频率集合； ∂_t 为偏导运算符； $*$ 为卷积运算符； $\delta(t)$ 为单位脉冲函数； f 为待分解的信号。

加入惩罚项 α 、拉格朗日乘法算子 λ ，转变为非约束变分问题：

$$L = (\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 +$$

$$\left\| f(t) - \sum_k u_k(t)^2 \right\|_2^2 + [\lambda(t), f(t) - \sum_k u_k(t)] \quad (8)$$

采用交替方向乘子法对 $\{u_k\}$ 与 $\{w_k\}$ 寻优，达到所需精度则输出分解分量。

1.3 基于SE样本熵的光伏功率分量重构

样本熵是在近似熵的基础上改进得到的一种用于分析时间序列数据的统计量，用来描述时间序列的不规则度和复杂度，样本熵值越大，表示时间序列自身相似度越大，其复杂性随之越小。其计算公式为

$$S_E(m, r, N) = -\ln \left[\frac{C_m(r, N)}{C_{m+1}(r, N)} \right] \quad (9)$$

式中： m 为嵌入维数，一般取1或者2； r 为相似容限，通常取待计算数据方差的0.1~0.25倍，本文选择0.25倍； $C_m(r, N)$ 和 $C_{m+1}(r, N)$ 分别为序列在相似容限 r 下匹配 m 和 $m+1$ 个点的概率。

计算各分量的样本熵，将熵值大小划定范围并分类为趋势分量和振荡分量2类，最后分别对2类分量进行预测再相加得到最终预测值。趋势分量和原始分量波动规律十分贴近，包含了原始分量的绝大多数信息。振荡分量为一系列非平稳高频信号的叠加。通过分类能减少预测运行时间，提升预测精度。

2 基于SSA-LSSVM的优化预测模型

2.1 LSSVM模型

LSSVM是一种基于SVM的方法，它用等式约束来代替最优问题中的不等式约束，通过解决等式约束和最小二乘问题，可以进一步简化问题，提高问题的求解效率。

给定一组输入输出测试样本 $(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^N, i=1, 2, \dots, N$ ，其中 $\mathbf{x}_i \in \mathbf{R}$ 为输入， $\mathbf{y}_i \in \mathbf{R}$ 为输出。其回归函数为

$$\mathbf{y}(\mathbf{x}) = \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}) + \mathbf{b} \quad (10)$$

式中： $\boldsymbol{\omega}$ 为权值向量； \mathbf{b} 为偏差； $\boldsymbol{\varphi}(\mathbf{x})$ 为对训练样本的非线性变换。

针对局部特异点的问题，对每个样本引入一个误差变量 e_i ，并将该误差变量的正则项添加到原函数中。于是，关于LSSVM的最优解问题变为

$$\begin{cases} \min_{\boldsymbol{\omega}, \mathbf{b}, \mathbf{e}} J(\boldsymbol{\omega}, \mathbf{e}) = \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{\lambda}{2} \sum_{i=1}^N e_i^2 \\ \text{s.t. } \mathbf{y}_i = \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} + e_i \end{cases} \quad (11)$$

式中 λ 为正则化系数。

列出拉格朗日函数，则原问题可以转化为对拉格朗日乘数 α_i 求极值的问题：

$$L(\boldsymbol{\omega}, \mathbf{b}, \mathbf{e}, \boldsymbol{\alpha}) = J(\boldsymbol{\omega}, \mathbf{e}) - \sum_{i=1}^N \alpha_i [\boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} + e_i - \mathbf{y}_i] \quad (12)$$

对式(12)中参数 $\boldsymbol{\omega}, \mathbf{b}, \mathbf{e}, \alpha_i$ 分别进行偏微分处理，使其为0，解方程组可得其回归函数为

$$\mathbf{y} = \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b} \quad (13)$$

其中 $k(\mathbf{x}_i, \mathbf{x})$ 为核函数，其公式为

$$k(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}\|^2}{2\sigma^2}\right) \quad (14)$$

核函数 σ 和正则化系数 λ 是LSSVM的重要参数，关系到模型的抗干扰和泛化能力，本文采用SSA来对这2个参数进行寻优。

2.2 麻雀搜索算法

在SSA算法中，一般会把种群中的麻雀分成2类：一类是探索者，另一类是追随者，这两类麻雀的地位是不断改变的。另外，麻雀会根据初始位置和适应度值在迭代中更新自己的位置和身份。

2.2.1 探索者位置更新

在觅食的同时，所有个体均会对周围环境保持警惕以防天敌的到来。探索者会根据预警值 R_2 和报警值 S_T 来改变自己的位置。

$$\mathbf{X}_{i,D}^{t+1} = \begin{cases} \mathbf{X}_{i,D}^t \cdot \exp\left(\frac{-i}{\alpha \cdot N}\right), & R_2 < S_T \\ \mathbf{X}_{i,D}^t + Q \cdot \mathbf{L}, & R_2 \geq S_T \end{cases} \quad (15)$$

式中： $\mathbf{X}_{i,D}^{t+1}$ 为第 i 只麻雀在 D 维中的位置； t 为目前迭代次数； N 为最大迭代次数； $\alpha \in [0, 1]$ ； $R_2 \in [0, 1]$ ； $S_T \in [0, 1]$ ； Q 为服从正态分布的随机数； \mathbf{L} 为一个 $1 \times D$ 的矩阵，其中每个元素均为1。当 $R_2 < S_T$ 时，意味着周围没有天敌，探索者可以进行全局搜索；当 $R_2 \geq S_T$ 时，说明有天敌在附近，他们必须去其他地方寻找食物。

2.2.2 追随者位置更新

在觅食过程中，一些追随者会时刻监视着探索者，它们会与监视者争夺食物。其位置更新方

式如下：

$$\mathbf{X}_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{\mathbf{X}_{\text{worst}}^t - \mathbf{X}_{i,j}^t}{t^2}\right), & i > \frac{n}{2} \\ \mathbf{X}_p^{t+1} + |\mathbf{X}_{i,j}^t - \mathbf{X}_p^{t+1}| \cdot \mathbf{A}^+ \cdot \mathbf{L}, & \text{其他} \end{cases} \quad (16)$$

式中： \mathbf{X}_p^{t+1} 为第 $t+1$ 次迭代时最优探索者的位置； $\mathbf{X}_{\text{worst}}^t$ 为第 t 次迭代时全局最差位置； n 为种群规模； \mathbf{A}^+ 为一个 $1 \times D$ 的矩阵，每个元素随机幅值为1或-1。当 $i > \frac{n}{2}$ 时，表明适应度值较低的第 i 个追随者状态较差，需要飞往其他地方觅食。

2.2.3 警戒者位置更新

警戒者时刻关注着周围情况，其位置更新方式

$$\mathbf{X}_{i,j}^{t+1} = \begin{cases} \mathbf{X}_{\text{best}}^t + \beta \cdot |\mathbf{X}_{i,j}^t - \mathbf{X}_{\text{best}}^t|, & f_i > f_g \\ \mathbf{X}_{i,j}^t + K \left(\frac{|\mathbf{X}_{i,j}^t - \mathbf{X}_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon} \right), & f_i = f_g \end{cases} \quad (17)$$

式中： $\mathbf{X}_{\text{best}}^t$ 为第 t 次迭代时全局最优位置； β 为步长控制参数，其值为服从均值为0、方差为1的正态分布的随机数； K 为 $[-1, 1]$ 内一随机数； f_i 为麻雀个体适应度值， f_g 、 f_w 分别为当前最优、最差适应度值；当 $f_i > f_g$ 时，表示此麻雀处于群体边缘，为躲避危险会向种群中心集中；当 $f_i = f_g$ 时，说明种群中的麻雀对威胁有所警觉，为了躲避天敌，它们必须和同伴靠近以避免危险。

采用SSA对LSSVM的超参数进行优化，提升模型运行速度，具体预测步骤如下：

1) 初始化SSA算法的参数：麻雀种群数量、探索者和追随者之间比例、优化变量大小的上下限、最大迭代次数等。

2) 输入数据集，划分训练集和测试集，计算SSA算法个体适应度，保存最优个体及最优位置。

3) 若在满足了终止迭代条件的情况下，可以确定最优解，退出程序；否则返回步骤2)。

4) 将寻优得到的参数重新赋值给LSSVM，对网络进行训练，得到优化后的结果。

3 基于核密度估计的分布式光伏出力区间预测

3.1 核密度估计

核密度估计(kernel density estimation, KDE)

作为一种非参数估计方法，无需使用历史数据已知的分布，并对数据的分布进行假设，可以直接从数据自身取样获取对应的预估周期的分布。核密度估计的计算公式为

$$\hat{f}_d(x) = \frac{1}{Nd} \sum_{i=1}^N K\left(\frac{T_i - x}{d}\right) \quad (18)$$

式中： d 为带宽； N 为分位数总数； T_i 为由条件分位数构成的数据集； $K(\cdot)$ 为核函数； $\hat{f}_d(x)$ 为带宽为 d 时的核密度表达式。

由于核函数的选择会对估计结果产生一定的影响，本文采用gaussian核函数，此函数均方误差较小，预测精度较高，公式为

$$K(x; h) = \exp\left(-\frac{x^2}{2h^2}\right) \quad (19)$$

通过核密度估计方法计算得到振荡分量的误差概率密度函数，对其求积分，得到 $\widehat{F}_d(x)$ 概率分布函数。在显著性水平为 α 下，若误差样本落入误差区间 $[\varepsilon_{\text{low}}, \varepsilon_{\text{high}}]$ 的概率不低于 $1 - \alpha$ ，则把区间 $[\varepsilon_{\text{low}}, \varepsilon_{\text{high}}]$ 称为置信度为 $1 - \alpha$ 的置信区间。将振荡分量的误差置信区间与趋势分量和振荡分量的点预测值进行叠加，得到最终的预测区间，计算公式如下：

$$p_{\text{ub}} = \begin{cases} p_{\text{pre}} + \varepsilon_{\text{high}}, & p_{\text{pre}} + \varepsilon_{\text{high}} < p_e \\ p_e, & p_{\text{pre}} + \varepsilon_{\text{high}} \geq p_e \end{cases} \quad (20)$$

$$p_{\text{lb}} = \begin{cases} p_{\text{pre}} + \varepsilon_{\text{low}}, & p_{\text{pre}} + \varepsilon_{\text{low}} > 0 \\ 0, & p_{\text{pre}} + \varepsilon_{\text{low}} \leq 0 \end{cases} \quad (21)$$

式中： p_{ub} 为预测区间上限； p_{lb} 为预测区间下限； p_{pre} 为趋势分量和振荡分量叠加的点预测值； p_e 为光伏额定功率。

3.2 分布式光伏出力点预测模型

传统的点预测方法总会与真实值之间产生差异，然而分布式光伏发电系统的功率输出有着明显的概率分布特征。光照强度对光伏发电功率有直接的影响，比如在晴天、多云、阴天、雨雪4种典型的天气状态下，光伏出力有着显著的差别。

本文将CEEMDAN模态分解算法与LSSVM模型相结合，提出了一种基于CEEMDAN-SSA-LSSVM的分布式光伏功率点预测模型。具体预测流程如图1所示。

1) 功率数据预处理，首先提取08:00—18:00

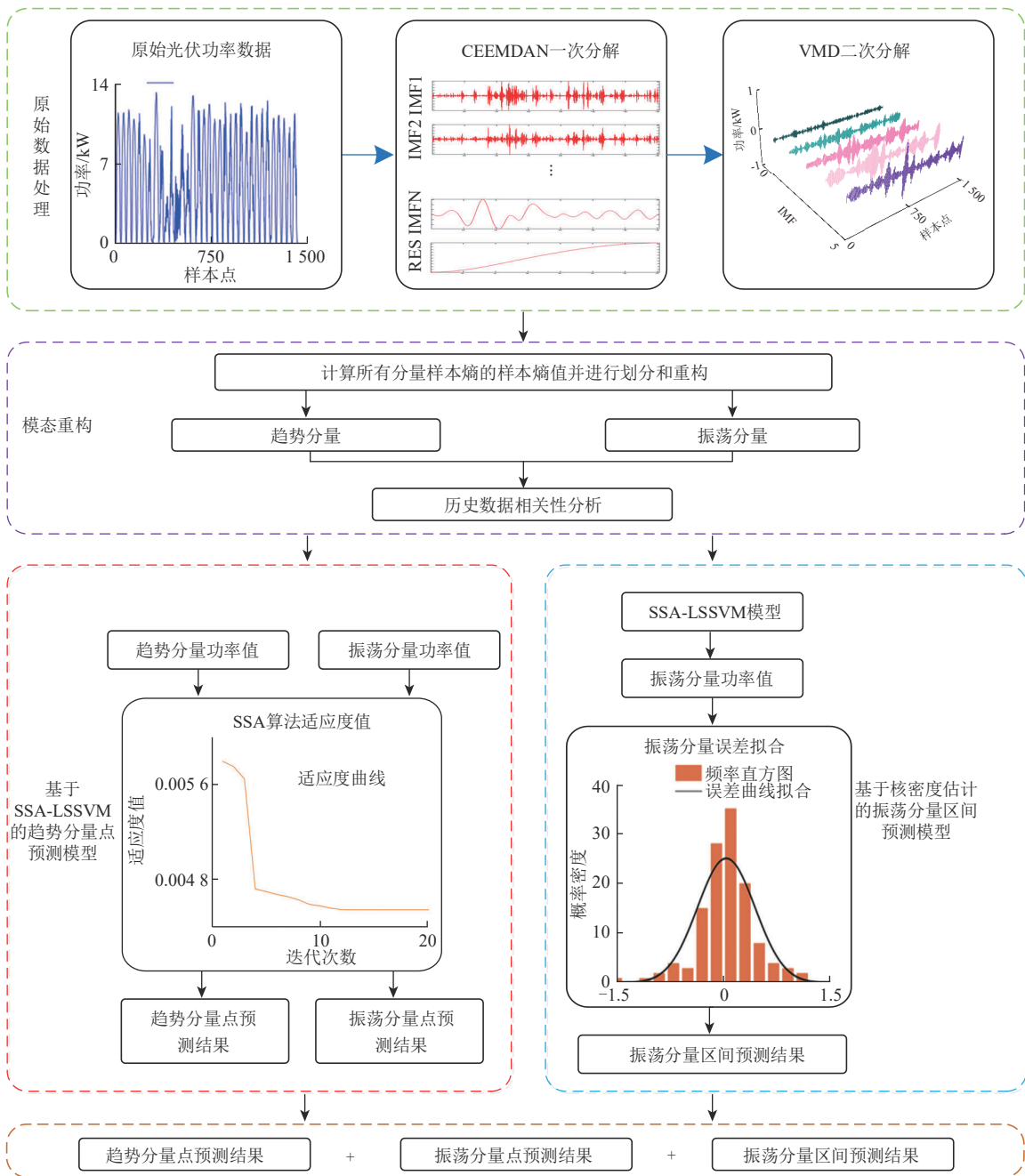


图1 区间预测方法流程图

Fig. 1 Flow chart of interval prediction method

有出力功率时段的数据，然后再排除不正常的数
据，补全缺失的数据等，最后对历史数据进行相
关性分析，得到模型的输入特征。

2) 采用 CEEMDAN 算法将光伏功率序列进
行分解，得到共计 K 个具有不同特征、频率互异
的 IMF 分量。

3) 通过计算各 IMF 分量的样本熵值，将样本
熵值最大，即频率最高、最不稳定分量二次分解

为 N 个子分量。

4) 根据样本熵值的大小，熵值大小相近的聚
为一类，将多个 IMF 分量重构为振荡分量和趋势
分量。

5) 整理 LSSVM 预测模型的输入输出序列，
并划分训练集和测试集；初始化 SSA 算法的参
数，包括种群个数、探索者与追随者之间比例、
迭代次数。对 LSSVM 的 2 个参数(核函数和正则

化系数)进行寻优,并重新赋值给LSSVM模型。

6) 利用寻优好的LSSVM模型分别对趋势分量和振荡分量进行预测,2组预测结果叠加得到最终点预测结果。

3.3 分布式光伏出力区间预测模型结构

针对不同天气因素影响所带来的功率波动以及单一的点功率预测模型不能反映功率波动的局限性,本文在点预测模型的基础上加入核密度估计方法,提出一种基于CEEMDAN-SSA-LSSVM的分布式光伏功率区间预测模型。此模型可有效提升预测的可靠性,反映预测算法的全面性,在一定程度上能提升功率预测精度,减少人工寻优时间。

首先,在分布式光伏出力点预测阶段,采用寻优能力强、收敛速度快的SSA对LSSVM模型进行优化,得到趋势分量和振荡分量点预测值;然后,在分布式光伏出力区间预测阶段,采用KDE对振荡分量的点预测误差进行概率密度估计,得到误差置信区间。振荡分量区间预测结果与趋势分量和振荡分量点预测值叠加,得到整体的区间预测结果。

4 算例分析

4.1 数据处理与误差评价指标

4.1.1 输入特征选取

本文基于江苏省某地2018年6月1日到6月31日的历史光伏数据进行算例分析。考虑到全年光伏有明显出力时段为08:00—18:00,以15 min为间隔,对功率数据进行采样,并对异常数据进行剔除以及补充缺失数据。

通过水平相关性和垂直相关性计算公式,分别得到某时刻的输出功率与前6天同一时间点上的输出功率和当天前6个时间点上的输出功率之间的相关性,如图2所示。可以看出,某时刻的功率与历史时刻的功率之间存在着很强的关联度,相关性系数均在0.86以上,且在水平和垂直方向上与最近的3个时间点关联程度较大。根据综合考虑,添加前3天的 $t-3$ 、 $t-2$ 、 $t-1$ 、 t 时刻功率,当天 $t-3$ 、 $t-2$ 、 $t-1$ 时刻功率共计15个特征作为输入。

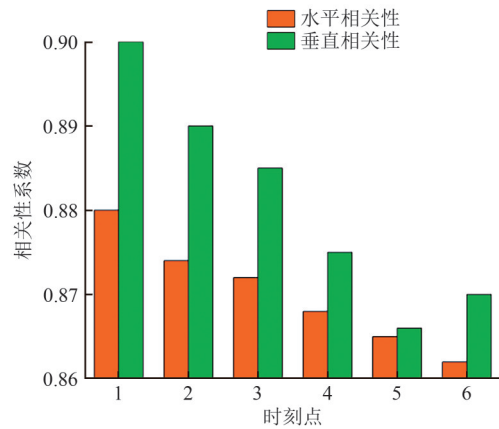


图2 水平和垂直相关性系数

Fig. 2 Horizontal and vertical correlation coefficient

4.1.2 点预测误差评价指标

本文采用不同指标对模型性能从多方面进行评价,包括均方根误差(root mean square error, RMSE)、均方误差(mean square error, MSE)、平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)、拟合优度 R^2 。前4个变量越接近0,第5个变量越接近1,表示预测效果越好。具体公式如下:

$$\left\{ \begin{array}{l} e_{\text{RMSE}} = \sqrt{\frac{1}{m} \sum_{i=1}^m [h(x_i) - y_i]^2} \\ e_{\text{MSE}} = \frac{1}{m} \sum_{i=1}^m [(h(x_i) - y_i)]^2 \\ e_{\text{MAE}} = \frac{1}{m} \sum_{i=1}^m |h(x_i) - y_i| \\ e_{\text{MAPE}} = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{h(x_i) - y_i}{y_N} \right| \\ R^2 = 1 - \frac{\sum_{i=1}^m [h(x_i) - y_i]^2}{\sum_{i=1}^m (y_i - y_m)^2} \end{array} \right. \quad (22)$$

式中: $h(x_i)$ 为模型预测值; y_i 为真实值; y_N 为额定装机容量; y_m 为测试集数据平均值; m 为测试集样本数量。

4.1.3 区间预测误差评价指标

预测区间覆盖率 $\varepsilon_{\text{PICP}}$ 和预测区间平均宽度 $\varepsilon_{\text{PINAW}}$ 是评估区间预测方法的重要指标。预测区间覆盖率表现为预测区间所能包含实际功率点的个数。当预测区间宽度足够宽时,区间覆盖率会产生全覆盖的情况,这样并不能对分布式光伏功率

进行合理的预测。通过引入预测区间平均宽度能够保证模型的有效性。具体公式如下：

$$\varepsilon_{PICP} = \frac{f}{m} \times 100\% \quad (23)$$

$$\varepsilon_{PINAW} = \frac{1}{m} \sum_{i=1}^m \frac{p_{ub} - p_{lb}}{y_{\max} - y_{\min}} \quad (24)$$

式中： f 为实际值在预测区间内的数目； y_{\max} 为测试集数据最大值； y_{\min} 为测试集数据最小值。

4.2 功率序列模态分解与重构结果

利用CEEMDAN对原有的分布式光伏出力数据进行分解，得到11个固定模式的子成分和1个残差成分，分解结果如图3所示。通过对这12个子项进行样本熵的计算，分析它们的复杂度。采用VMD对熵值最大的分量3(IMF3)进行二次分解，得到5个固定模态子分量和1个残差分量，5个固定模态子分量分解结果如图4所示。

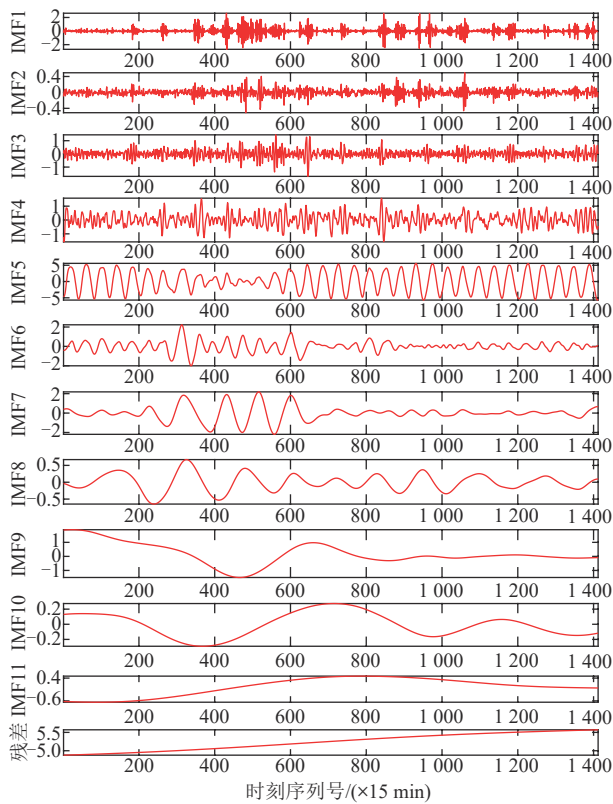


图3 CEEMDAN一次分解结果图

Fig. 3 CEEMDAN single decomposition result chart

2次分解得到的16个分量的样本熵值如图5所示，可知，子分量4—11的样本熵值较为接近，且熵值均分布在0~0.5，分量较为平稳，可以将其重构为一个分量来进行预测。这些分量重构得到

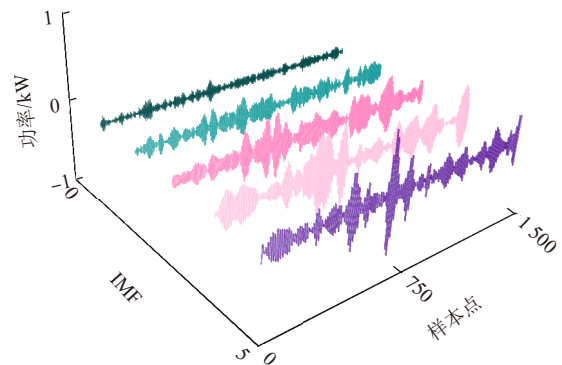


图4 VMD二次分解结果图

Fig. 4 VMD secondary decomposition result diagram

的分量能大致反映原始序列的波动趋势，故称此重构的分量为趋势分量。同理，子分量1—3、12—16样本熵值较为接近，将其重构为振荡分量，重构结果如图5所示。

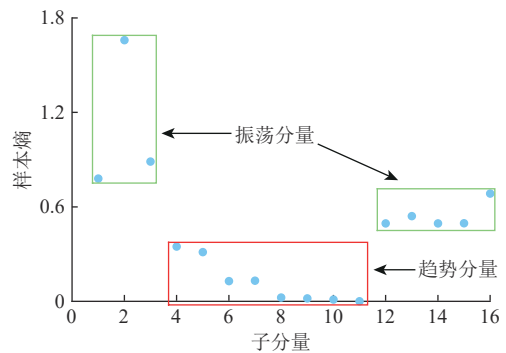


图5 样本熵重构结果

Fig. 5 Sample entropy reconstruction results

4.3 点功率预测结果

对趋势分量和振荡分量分别进行预测，最后叠加得到模型预测结果。实验中，SSA-LSSVM模型的最优超参数设置如下：麻雀种群数量为10，最大迭代次数为50次，核函数 σ 和正则化系数 λ 的寻优范围均设置为[1, 10 000]，适应度目标函数为均方误差最小值。采用SSA优化，最终得到LSSVM的最优参数为 $\sigma=360$ ， $\lambda=1 286$ 。

分别选取粒子群优化 (particle swarm optimization, PSO)算法和SSA算法对LSSVM参数进行寻优，2种算法分别运行50次后的平均适应度曲线如图6所示。

从图6可知，SSA算法的平均最优适应度值收敛速度更快，模型运行时间更短。

本文以前21天、前26天、前29天和前30天

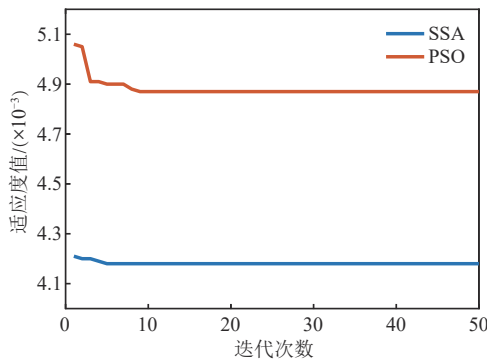


图6 PSO与SSA迭代收敛对比图

Fig. 6 Comparison chart of PSO and SSA iteration convergence

的数据作为训练集，将后10天、后5天、后3天和后1天的数据作为测试集进行仿真模拟。为凸显本文所提方法的预测效果，将LSSVM、SSA-LSSVM和CEEMDAN-LSSVM这3种方法与本文方法进行对比，利用RMSE、MSE、MAE、MAPE、 R^2 这5个指标进行综合评价。因为机器学习算法具有随机性，为了提高模型可信度，分别对以上4种模型进行5次预测仿真，取5次实验的平均值作为最终结果，4种模型的预测结果对比如图7、8所示。

从图7可以看出，与其他预测模型相比，CEEMDAN-SSA-LSSVM模型的预测效果更好，预测曲线与真实值最为接近。由图8可以更直观地发现，相较于其他的预测模型，CEEMDAN-SSA-LSSVM模型的各项误差评价指标都优于其他模型。以测试集为1天的结果为例，相较于CEEMDAN-LSSVM、SSA-LSSVM、LSSVM方法，本文方法的RMSE分别降低了20.4%、66.69%、76.15%；MSE分别降低了36.63%、88.9%和94.3%；MAE分别降低了26.1%、69.49%

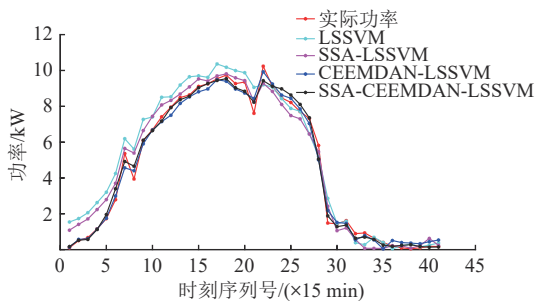


图7 不同模型的点预测结果

Fig. 7 Point prediction results of different models

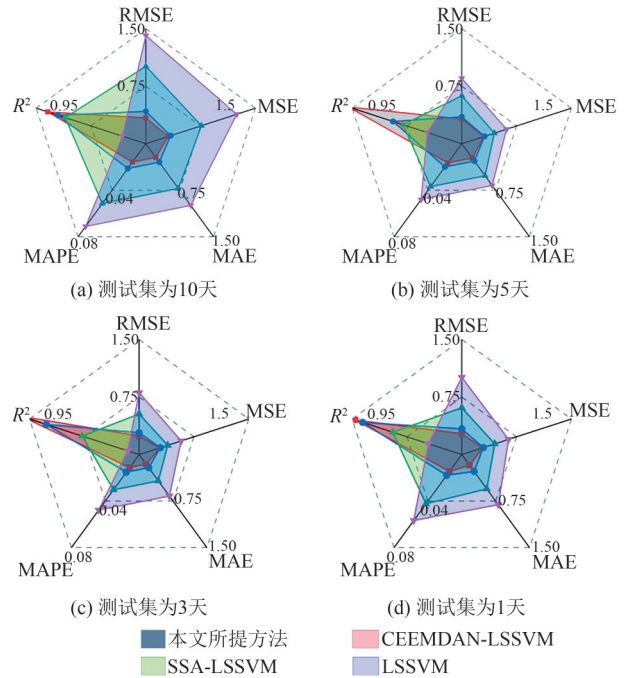


图8 不同测试集下各类模型预测指标雷达图

Fig. 8 Radar chart of various model prediction indicators under different test sets

和77.89%；MAPE分别降低了0.556%、3.571%和5.536%。本文方法拟合优度 R^2 高达0.9935，均优于其他3个模型。实验结果表明，本文所提方法预测效果极佳，具有较高的预测精度。

4.4 区间功率预测结果

利用CEEMDAN-SSA-LSSVM的点预测结果，将振荡分量的预测值与真实值作差得到预测误差功率序列。通过核密度估计方法拟合预测误差序列，得到误差累计概率密度分布函数，给定置信度，与点预测结果结合得到区间预测结果。不同置信水平下的预测结果如图9所示。

由图9的结果可以得出，在置信度为95%时，

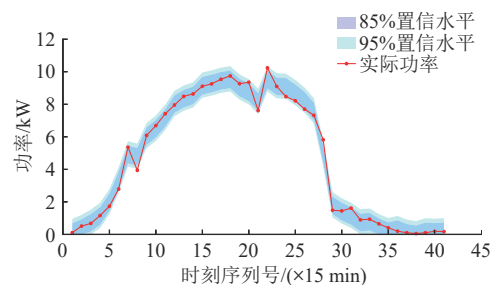


图9 2种置信水平下光伏功率区间预测结果

Fig. 9 Prediction results of photovoltaic power interval under two confidence levels

光伏功率实际值有39个点落入置信区间内,说明本文所提方法在一定程度上可以提高区间预测的可靠性。

为了检验核密度估计对误差分布拟合的准确性,分别计算出CEEMDAN-SSA-LSSVM模型与蒙特卡洛法、Bootstrap重采样法结合后的预测区间覆盖率和预测区间平均宽度。表1为3种方法区间预测指标对比结果。

表1 不同方法区间预测指标对比

Tab.1 Comparison of interval prediction indicators using different methods

算法	置信度/%	$\epsilon_{PICP}/\%$	ϵ_{PINAW}
Bootstrap	85	82.93	1.33
	95	87.80	1.71
蒙特卡洛法	85	85.36	1.231
	95	90.24	1.667
本文方法	85	87.80	1.092 8
	95	95.12	1.554 0

从表1可以看出,在相同置信度下,相较于其他2种方法,本文所提CEEMDAN-SSA-LSSVM的区间预测模型,能够密切跟踪光伏功率序列的变化,并能够获得更高的预测区间覆盖率和更窄的预测区间平均宽度。在置信度为95%时,本文方法的预测区间覆盖率相较于其他2种方法分别提升了7.32%和4.88%;预测区间平均宽度也为最小,相较于其他2种方法分别降低了0.156和0.113。

为了进一步证明本文提出模型的有效性,将本文区间预测方法与LSSVM、SSA-LSSVM和CEEMDAN-LSSVM方法进行对比,结果如图10所示。

由图10可以发现,本文方法的在不同置信度

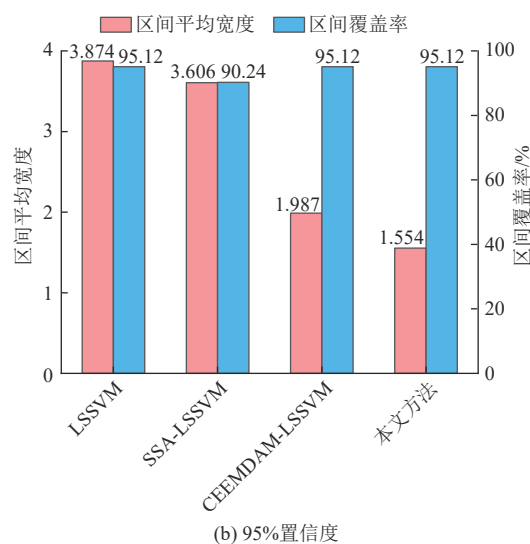
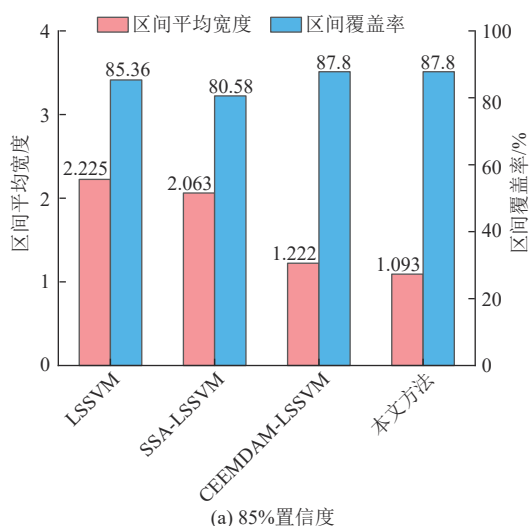


图10 不同算法区间预测指标对比

Fig. 10 Comparison of interval prediction indicators for different algorithms

下的准确度、可靠度都较为出色,预测区间均能得到较窄的区间宽度和较高的区间覆盖率。

5 结论

1) 在分布式光伏功率数据处理中加入二次模态分解,再结合样本熵对其子序列进行重构,可有效降低原始预测分量的复杂程度的同时提升模型预测准确性。

2) 所提核密度估计区间预测方法可以有效实现光伏功率区间预测,相较于其他传统预测模型,能够提升预测区间覆盖率的同时降低预测区间平均宽度。在置信度为95%时,所提方法的预测区间覆盖率相较于其他2种方法分别提升了7.32%和4.88%,预测区间分别降低了0.156和0.113。

参考文献

[1] 张汀荃, 谢明成, 王蓓蓓, 等. 分布式光伏的共享价值及其对配电网影响的系统动力学仿真[J]. 电力系统自动化, 2021, 45(18): 35-44.
ZHANG T H, XIE M C, WANG B B, et al. System dynamics simulation of shared value of distributed photovoltaic and its impact on distribution network[J]. Automation of Electric Power Systems, 2021, 45(18): 35-44.

[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.
- LAI C W, LI J H, CHEN B, et al. Review of photovoltaic power output prediction technology[J]. Transactions of China Electrotechnical Society, 2019, 34(6): 1201-1217.
- [3] 李博彤, 李明睿, 刘梦晴. 基于通径分析和相空间重构的光伏发电预测模型[J]. 电测与仪表, 2022, 59(11): 79-87.
- LI B T, LI M R, LIU M Q. PV power generation forecast model based on path analysis and phase space reconstruction[J]. Electrical Measurement & Instrumentation, 2022, 59(11): 79-87.
- [4] 商立群, 李洪波, 侯亚东, 等. 基于VMD-ISSA-KELM的短期光伏发电功率预测[J]. 电力系统保护与控制, 2022, 50(21): 138-148.
- SHANG L Q, LI H B, HOU Y D, et al. Short-term photovoltaic power generation prediction based on VMD-ISSA-KELM[J]. Power System Protection and Control, 2022, 50(21): 138-148.
- [5] 李琰, 吕南君, 刘雪涛, 等. 考虑新能源消纳和网损的分布式光伏集群出力评估方法[J]. 电力建设, 2022, 43(10): 136-146.
- LI Y, LÜ N J, LIU X T, et al. Output evaluation method of distributed photovoltaic cluster considering renewable energy accommodation and power loss of network[J]. Electric Power Construction, 2022, 43(10): 136-146.
- [6] 吕清泉, 张珍珍, 马彦宏, 等. 区域光伏发电出力特性分析研究[J]. 发电技术, 2022, 43(3): 413-420.
- LÜ Q Q, ZHANG Z Z, MA Y H, et al. Analysis and research on output characteristics of regional photovoltaic power generation[J]. Power Generation Technology, 2022, 43(3): 413-420.
- [7] 叶畅, 柳丹, 曹侃. 基于云图特征的超短期光伏发电功率预测模型[J]. 电网与清洁能源, 2023, 39(10): 70-79.
- YE C, LIU D, CAO K. An ultra-short-term photovoltaic power forecasting model based on cloud features[J]. Power System and Clean Energy, 2023, 39(10): 70-79.
- [8] 张津源, 蒲天骄, 李焯, 等. 基于多智能体深度强化学习的分布式电源优化调度策略[J]. 电网技术, 2022, 46(9): 3496-3503.
- ZHANG J Y, PU T J, LI Y, et al. Multi-agent deep reinforcement learning based optimal dispatch of distributed generators[J]. Power System Technology, 2022, 46(9): 3496-3503.
- [9] 张雪松, 李鹏, 周亦尧, 等. 基于贝叶斯概率的光伏出力组合预测方法[J]. 太阳能学报, 2021, 42(10): 80-86.
- ZHANG X S, LI P, ZHOU Y Y, et al. Photovoltaic output combination forecasting method based on Bayesian probability[J]. Acta Energetica Sinica, 2021, 42(10): 80-86.
- [10] 王勇, 王婷, 岳园园, 等. 提升分布式光伏承载力的灵活资源协同规划[J]. 电网与清洁能源, 2023, 39(8): 116-126.
- WANG Y, WANG T, YUE Y Y, et al. Flexible resource collaborative planning to enhance the capacity of distributed photovoltaic[J]. Power System and Clean Energy, 2023, 39(8): 116-126.
- [11] 叶林, 程文丁, 李卓, 等. 光伏集群有功功率分层预测控制策略[J]. 电力系统自动化, 2023, 47(2): 42-52.
- YE L, CHENG W D, LI Z, et al. Hierarchical prediction control strategy of active power for photovoltaic cluster[J]. Automation of Electric Power Systems, 2023, 47(2): 42-52.
- [12] 王彪, 吕洋, 陈中, 等. 考虑信息时移的分布式光伏机理-数据混合驱动短期功率预测[J]. 电力系统自动化, 2022, 46(11): 67-74.
- WANG B, LÜ Y, CHEN Z, et al. Hybrid mechanism-data-driven short-term power forecasting of distributed photovoltaic considering information time shift[J]. Automation of Electric Power Systems, 2022, 46(11): 67-74.
- [13] 李丰君, 王磊, 赵健, 等. 基于天气融合和LSTM网络的分布式光伏短期功率预测方法[J]. 中国电力, 2022, 55(11): 149-154.
- LI F J, WANG L, ZHAO J, et al. Research on distributed photovoltaic short-term power prediction method based on weather fusion and LSTM-net[J]. Electric Power, 2022, 55(11): 149-154.
- [14] 卿会, 李薇, 刘文娇, 等. 基于极点对称模态分解-支持向量机的短期光伏发电预测方法[J]. 热能动力工程, 2022, 37(10): 189-197.
- QING H, LI W, LIU W J, et al. Research on short-term photovoltaic power forecasting technology based on ESMD-SVM[J]. Journal of Engineering for Thermal Energy and Power, 2022, 37(10): 189-197.
- [15] 丁明, 虞海彪, 刘练, 等. 基于多变量相空间重构和RBF神经网络的光伏功率预测方法[J]. 电子测量与仪器学报, 2020, 34(8): 1-7.
- DING M, YU H B, LIU L, et al. Power prediction method of photovoltaic generation based on multivariable phase space reconstruction and RBF neural

- network[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(8): 1-7.
- [16] 向玲, 刘佳宁, 苏浩, 等. 基于CEEMDAN二次分解和LSTM的风速多步预测研究[J]. 太阳能学报, 2022, 43(8): 334-339.
- XIANG L, LIU J N, SU H, et al. Research on multi-step wind speed forecast based on ceemdan secondary decomposition and LSTM[J]. Acta Energetica Sinica, 2022, 43(8): 334-339.
- [17] 韦权, 汤占军. 基于SSA-VMD-SE-KELM结合蒙特卡洛法的风电功率区间预测[J]. 智慧电力, 2022, 50(9): 59-66.
- WEI Q, TANG Z J. Wind power range prediction based on SSA-VMD-SE-KELM combined with Monte Carlo method[J]. Smart Power, 2022, 50(9): 59-66.
- [18] 杨国清, 李建基, 王德意, 等. 基于信息熵变权区间组合和边界逼近的短期光伏功率区间预测[J]. 太阳能学报, 2023, 44(2): 381-390.
- YANG G Q, LI J J, WANG D Y, et al. Short-term photovoltaic power interval prediction based on information entropy variable weight interval combination and boundary approximation[J]. Acta Energetica Sinica, 2023, 44(2): 381-390.
- [19] 肖白, 张博, 王辛玮, 等. 基于组合模态分解和深度学习的短期风电功率区间预测[J]. 电力系统自动化, 2023, 47(17): 110-117.
- XIAO B, ZHANG B, WANG X W, et al. Short-term wind power interval prediction based on combined mode decomposition and deep learning[J]. Automation of Electric Power Systems, 2023, 47(17): 110-117.
- [20] 杨海柱, 石剑, 江昭阳, 等. 基于CEEMD-SSA-LSSVM短期电力负荷预测模型[J]. 武汉大学学报(工学版), 2022, 55(6): 609-616.
- YANG H Z, SHI J, JIANG Z Y, et al. Short term power load prediction model based on CEEMD-SSA-LSSVM [J]. Journal of Wuhan University (Engineering Edition), 2022, 55(6): 609-616.
- [21] 王鹏翔, 沈娟, 王菁昀, 等. 基于PCA-LMD-WOA-ELM的短期光伏功率预测[J]. 智慧电力, 2022, 50(6): 72-78.
- WANG P X, SHEN J, WANG J Y, et al. Short term photovoltaic power prediction based on PCA-LMD-WOA-ELM[J]. Smart Power, 2022, 50(6): 72-78
- [22] 赵超, 王斌, 孙志新, 等. 基于改进灰狼算法的独立微电网容量优化配置[J]. 太阳能学报, 2022, 43(1): 256-262.
- ZHAO C, WANG B, SUN Z X, et al. Optimal configuration optimization of islanded microgrid using improved grey wolf optimizer algorithm[J]. Acta Energetica Sinica, 2022, 43(1): 256-262.

收稿日期: 2023-04-12。

修回日期: 2023-08-25。

作者简介:



杨康

杨康(1990), 男, 硕士, 工程师, 主要研究方向为电力电子, yangkangtc@163.com;



周霞

周霞(1978), 女, 博士, 副教授, 从事电力通信、电力系统分析与控制研究, 本文通信作者, zhouxia@njsu.edu.cn。

(责任编辑 辛培裕)