

DOI: 10.13382/j.jemi.B2407829

基于 PCHIP-VMD 数据分析与 SSA-LSTM 模型的 短期风电功率预测*

张登攀¹ 兰征¹ 杜怡衡²

(1. 湖南工业大学电气与信息工程学院 株洲 412007; 2. 郑州轻工业大学电气与信息工程学院 郑州 450002)

摘要:短期风电功率预测对电力系统调度与安全运行具有重要的支撑作用,然而风电数据具有强随机性和非平稳性,现有预测方法存在数据预处理保形性不足、模态混叠、预测模型参数优化效率低等问题,严重影响短期风电功率预测的准确性。为此,提出分段三次 Hermite 插值法(PCHIP)与变分模态分解(VMD)相结合的数据预处理方法,以及麻雀搜索算法(SSA)与长短期记忆网络(LSTM)相结合的风电功率预测模型。首先,对风电原始数据异常值进行检测,针对异常值导致的时序数据保形性缺失问题,采用 PCHIP 法进行数据修复;其次,结合 VMD 法将预处理后的风电功率数据分解为 4 个内在模态分量,得到不同时间尺度上变化的数据信号;然后,将分解后的稳定项序列输入 SSA-LSTM 风电功率预测模型,得到风电功率预测结果。以某风电场 21 d 的实测功率数据为样本数据进行分析验证,所提模型的预测结果与真实值拟合程度可达到 0.989 1,较当前 LSTM 模型预测精度提升 5.558%,证明了所提方法的有效性和优越性。

关键词:功率预测;分段三次 Hermite 插值;变分模态分解;麻雀搜索算法;长短期神经网络

中图分类号: TN911; TP183 **文献标识码:** A **国家标准学科分类代码:** 470.40

Short-term wind power forecasting based on PCHIP-VMD data analysis and SSA-LSTM model

Zhang Dengpan¹ Lan Zheng¹ Du Yiheng²

(1. School of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, China;

2. School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China)

Abstract: Short-term wind power prediction is crucial for power system scheduling and operational security. However, the accuracy of such predictions is severely compromised by the inherent strong randomness and non-stationarity of wind power data, as well as limitations in existing methods, including insufficient shape-preserving capability in data preprocessing, modal aliasing, and inefficient parameter optimization in prediction models. To address these issues, this paper proposes a novel hybrid framework combining a piecewise cubic hermite interpolating polynomial (PCHIP) with variational mode decomposition (VMD) for data preprocessing and a sparrow search algorithm (SSA)-optimized long short-term memory (LSTM) network for prediction. First, abnormal values in raw wind power data are identified and repaired using PCHIP, which preserves the local monotonicity and curvature of the original sequence through Hermite interpolation. Second, the preprocessed data are decomposed into four intrinsic mode components (IMFs) via VMD to capture multi-scale temporal features. Finally, the stabilized IMF sequences are input into the SSA-LSTM wind power forecasting model to yield prediction outcomes. Experimental validation using 21-day measured power data from a wind farm demonstrates that the proposed model achieves a fitting degree of 0.989 1 with actual values, improving prediction accuracy by 5.558% compared to conventional LSTM, thereby verifying the effectiveness and superiority of the method.

Keywords: power prediction; piecewise cubic Hermite interpolating polynomial; variational mode decomposition; sparrow search algorithm; long and short term neural networks

收稿日期: 2024-09-18 Received Date: 2024-09-18

* 基金项目: 湖南省自然科学基金面上项目(2025JJ50232)、湖南省教育厅科学研究重点项目(24A0404)资助

0 引言

随着“双碳”目标的提出,我国将加快构建适应新能源占比逐渐提高的新型电力系统,作为绿色清洁能源的代表风力发电比重将大幅提升^[1]。风电具有储量大,分布广泛,对环境友好,成本具备规模效应,运行成本和人力成本较低,具有较高经济性等优点^[2],但风电也具有强烈的波动性,给电力系统的调度增加了不确定因素^[3]。因此,对风电输出功率进行短期预测,并提升风电功率的预测精度,将不确定的风电转变为可调度的友好型电源,将有助于降低电力系统弃风量、缓解调度压力,同时能有效提高清洁能源利用率、减少碳排放^[4]。

对于风电功率预测模型而言,早期预测模型主要以传统机器学习为主,如随机森林、支持向量机、极限学习机算法等^[5-6]。这类方法广泛应用于人工智能、机器学习、数据预测等领域,但用于数据预测时,往往存在较大预测误差^[7-8]。为了提升预测精度,现在一些研究机构和学者通过改进长短记忆(long short-term memory, LSTM)算法,引入门控机制、残差连接等技术,还通过专注于风机级与站级的数据清洗工作,确保数据的精确性和完整性,同时引入多源数值天气预报数据作为关键输入,以进一步提升模型的预测能力。

然而,现有方法仍受限于单一技术路径优化,存在算法适用性受限与不足的共性问题。王永生等^[9]提出一种改进熵权法与样本熵-互补集合经验模态分解(sample entropy complementary ensemble empirical mode decomposition, SECEEMD)分解相结合的预测框架,通过改进熵权法对特征进行加权,有效提升了非线性特征的提取能力,提高了预测精度,但其 SECEEMD 算法在风速突变时仍然可能存在模态混叠的现象,将会导致高频分量预测滞后显著。此类改进方法难以兼顾多时间尺度的特征耦合关系,与之形成对比的是,刘金朋等^[10]结合变分模态分解(variational mode decomposition, VMD)、改进海洋捕食算法(improved marine predator algorithm, IMPA)和支持向量机(support vector machine, SVM)提出一种超短期风电功率预测方法,利用改进 MPA 算法优化支持向量机参数,在超短期预测中能够取得较高精度,然而该方法在处理风电功率数据的非平稳性和复杂波动时仍具有一定局限性,一方面,SVM 模型对数据的非线性特征捕捉能力有限,另一方面,该方法未考虑分解残差项的累积误差带来的影响。不同的预测方法其特点不同,每种预测方法反映出的结果往往波动性大,因此需要通过组合预测来提高预测精度^[11]。在实际应用中,要根据风电场的特殊性灵活选择与之相适应的预测方法。李俊卿

等^[12]提出一种基于 Kriging 和 LSTM 网络的风电功率预测模型。首先,基于偏互信息对风电功率的重要影响变量加权处理后作为 Kriging 预测模型的输入,预测风电功率的线性分量;然后,使用 LSTM 网络预测功率的非线性残差分量;最后线性和非线性分量预测结果叠加,得到风电功率预测结果。肖烈禧等^[13]提出一种基于改进算术优化算法(improve arithmetic optimization algorithm, IAOA)、VMD 和 LSTM 相结合的超短期风电功率预测模型,利用 IAOA 对 VMD 的关键分解参数和 LSTM 网络参数进行优化,提升 LSTM 的预测精度。毕贵红等^[14]提出一种基于多模式分解、多通道输入、并联卷积神经网络和双向长短期记忆网络相结合的组合预测方法,用于不同天气类型的超短期发电功率预测,建立了历史数据之间的时间特征关系,提高预测结果的准确性和稳定性。

除了模型结构创新,数据预处理环节的挑战仍然不可忽视。风力发电机组在启动过程中可能出现发电机故障、传感器故障等一系列事故,同时风速和风向的突然变化以及极端温度和湿度等环境因素都会影响风电场的正常运行,从而产生异常数据,因此需要对复杂多变的原始风电功率数据进行预处理^[15-16]。刘景森等^[17]和宫婷等^[18]利用三次样条插值法拟合,得到一条插值函数和被插值函数密合程度较高的光滑曲线,但不能够保证插值节点处的函数值和对应的导数值相等,且数据点对其影响较大,降低了预测精度。

现有研究在数据预处理时采用传统线性插值方法易破坏原始功率序列的统计特性,加剧后续分解过程的模态混叠,同时采用单一优化算法在 LSTM 超参数寻优时易陷入局部最优,难以平衡网络深度与训练效率,而且现有混合模型大多采用固定结构处理所有模态分解,未充分考虑高频噪声与低频趋势分量的差异化特性。基于此,为解决插值数据结果不够平滑、数据点影响大、插值后整体数据非平稳以及简单预测模型预测效果不理想的问题,考虑了多种环境因素对风电的影响,提出了分段三次 Hermite 插值法(piecewise cubic hermite interpolating polynomial, PCHIP)与 VMD 相结合的 PCHIP-VMD 数据处理方法,麻雀搜索算法(sparrow search algorithm, SSA)与长短期神经网络相组合的 SSA-LSTM 风电功率预测模型。利用 PCHIP 方法对异常数据进行插值和替换,采用 VMD 分解后的模态分量,降低数据的非线性和非平稳性,运用 SSA-LSTM 组合模型进行风电功率预测,进一步提高了预测精度。该模型的关键在于实现数据预处理阶段保留序列统计特性,特征分解阶段区分高频噪声与低频趋势以及参数优化阶段平衡网络深度与训练效率的 3 个层级的协同优化。算例分析结果表明,PCHIP-VMD-SSA-LSTM 组合预测模型在预测结果精确性更高。

1 基于 PCHIP-VMD 的数据分析

1.1 数据的异常检测

风电场数据中存在众多缺失值和部分异常值,这些不完整或不正常的数 据可能会对预测结果的准确性产生直接影响,因此需要对风电功率的原始数据进行预处理。

1) 采用广义极端学习偏差 (generalized extreme studentized deviate, GESD) 离群值法对数据进行判断,检测是否具有异常值,并将异常数据剔除。

(1) 将每个数据点与均值的差除以标准差,得到标准化分数(z 分数);

(2) 计算标准化分数的累积分布函数 (cumulative distribution function, CDF);

(3) 算经验累积分布函数 (empirical cumulative distribution function, ECDF), 得到 Q 统计量, 确定离群值;

(4) 如果 Q 统计量的绝对值大于某个阈值, 则认为该数据点是离群值;

(5) 再次执行步骤(1)~(4), 直到数据集中的所有异常值均被识别为止。

2) 在找到异常点和空缺点之后, 采用 PCHIP 填补缺失值, 保证数据的完整程度, 通过在每个插值节点处给定函数值和导数值的条件进行高次多项式构造, 进而对相关数据进行填充。

1.2 基于保形插值的缺失值处理

为填补风电功率原始数据的缺失值, 保证数据的完整程度, 可以将离散的数据点 $(x_1, y_1), \dots, (x_n, y_n)$ 近似为一个变量 $f(x)$ 的连续函数, 而保形分段三次样条插值法能够在保证曲线的斜率在数据点之间连续的同时, 平滑地经过数据点。

采用基函数方法构建 $H_3(x)$, 如式(1)所示。

$$H_3(x) = y_0\alpha_0(x) + y_1\alpha_1(x) + m_0\beta_0(x) + m_1\beta_1(x) \quad (1)$$

式中: $\alpha_0(x), \alpha_1(x), \beta_0(x), \beta_1(x)$ 为插值函数, 且均次数不超过 3 的多项式。为满足插值条件, 应满足:

$$\alpha_i(x_j) = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, \alpha'_i(x_j) = 0$$

$$\beta_i(x_j) = 0, \beta'_i = \delta_{ij} (i, j = 0, 1)$$

$$H_3(x_i) = y_i, H'_3(x_i) = m_i (i = 0, 1)$$

式中: δ_{ij} 代表狄利克雷函数; $\alpha_i(x_j)$ 指 x_j 代入多项式 $\alpha_i(x)$ 后的值; $\beta_i(x_j)$ 指 x_j 代入多项式 $\beta_i(x)$ 后的值; $H_3(x_i)$ 指 x_i 代入 $H_3(x)$ 后得的值。由于 $\alpha_0(x_1) = \alpha'_0(x_1) = 0, \beta_0(x_0) = \beta_0(x_1) = \beta'_0(x_1) = 0$, 可得:

$$\alpha_0(x) = \left(1 + 2 \frac{x - x_0}{x_1 - x_0}\right) \left(\frac{x - x_1}{x_0 - x_1}\right)^2 \quad (2)$$

$$\alpha_1(x) = \left(1 + 2 \frac{x - x_1}{x_0 - x_1}\right) \left(\frac{x - x_0}{x_1 - x_0}\right)^2 \quad (3)$$

$$\beta_0(x) = (x - x_0) \left(\frac{x - x_1}{x_0 - x_1}\right)^2 \quad (4)$$

$$\beta_1(x) = (x - x_1) \left(\frac{x - x_0}{x_1 - x_0}\right)^2 \quad (5)$$

将 $\alpha_0(x), \alpha_1(x), \beta_0(x), \beta_1(x)$ 代入式(1)可以得出 $H_3(x)$ 。

1.3 多尺度信号分解与模态提取

VMD 是一种信号分解方法, 将多分量信号一次性分解成多个具有不同频率和振幅的单分量信号, 实现对非平稳信号的分解和处理, 降低复杂度且非线性强的时间序列的非平稳性^[19], 分解的过程如下。

1) 初始化 $\{u_k^1\}, \{\omega_k^1\}, \{\lambda_k^1\}$ 和 $n=0$ 。

2) 令 $n=n+1$, 进行循环迭代。

3) 按照式(6)和(7)不断更新 u_k 和 ω_k , 直到分解个数达到 k 时停止内循环。

$$u_k^{n+1} = \frac{f(\omega) - \sum_{i \neq k} u_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (6)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |u_k^{n+1}(\omega)|^2 d\omega} \quad (7)$$

4) 根据式(8)更新 λ 。

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \gamma(f(\omega) - \sum_k u_k^{n+1}(\omega)) \quad (8)$$

给定精度 ε , 若满足式(9)则停止循环; 若不满足则重复步骤 3), 继续进行循环。

$$\sum_k \|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 < \varepsilon \quad (9)$$

式中: u_k 为分解后的单分量调幅调频信号; ω_k 为每一个单分量调幅调频信号的中心频率; α 为二次惩罚因子; γ 表示噪声, 当信号含有强噪声时, 设定 $\gamma=0$ 达到更好的去噪效果; λ 为拉格朗日乘数; n 为迭代次数。

然而, VMD 算法直接作用于原始功率序列时, 易因数据缺失或噪声干扰导致模态混叠问题。所以本文将原始信号先采用 PCHIP 插值来修复损失以及平滑噪声得到优化后的信号, 再将此信号进行 VMD 自适应分解得到 IMF 分量。这种将 PCHIP 与 VMD 的联合处理方法相较于传统线性插值或三次样条插值来说保形性更加优化, PCHIP 在插值过程中严格保持局部数据单调性, 能避免三次样条插值可能产生的虚假振荡, 从而减少高频噪声对 VMD 分解的干扰, 同时计算效率提升, PCHIP 插值仅需线性时间复杂度 $O(n)$, 而传统插值方法需 $O(n^3)$, 可显著降低大规模风电数据预处理的耗时。

2 基于 SSA-LSTM 模型的短期预测模型

2.1 群体智能优化 SSA 算法设计

SSA 根据麻雀群体的觅食行为及反捕食行为,实现找到最优目标函数的目的^[20-22]。

在整个过程中,发现者和参与者的身份会随着彼此行为的不同而发生相应变化,但两类在整个群体中所占比例固定。如果参与者的适应性不强,无法发现食物,它们就会逃离这个地区,去找到更能依靠的发现者,从而寻找到最优解^[23]。SSA 通过模拟发现者-参与者机制实现全局搜索,发现者和参与者位置更新策略如式(10)和(11)所示。

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \exp(-\frac{i}{attter_{max}}) \\ X_{i,j}^t + QL, R_2 < ST \end{cases} \quad (10)$$

式中: $X_{i,j}^{t+1}$ 代表第 i 只麻雀在第 t 次迭代中的第 j 维位置信息; α 代表 $[0,1]$ 的随机数; Q 代表服从正态分布的随机数; L 表示全 1 的 $1 \times d$ 矩阵; R_2 表示 $[0,1]$ 的随机数,代表预警; ST 表示 $[0.5,1]$ 的常数,代表安全值。

当 $R_2 < ST$ 时,意味着此时的觅食环境周围没有捕食者,发现者可以进行食物搜索;当 $R_2 \geq ST$ 时,表示种群中的一部分麻雀发现了捕食者,并发出警报,此时所有麻雀都迅速飞到其他安全的地方去觅食^[24]。

$$X_{i,j}^{t+1} = \begin{cases} Q \exp(\frac{x_{worst} - X_i^t}{i^2}), & i > n/2 \\ X_{i,j}^{t+1} + |X_{i,j}^t - X_p^{t+1}| A^+ \cdot L, & \text{其他} \end{cases} \quad (11)$$

当 $i > n/2$ 时,说明适应能力较弱且搜寻食物能力较差的个体未能获得食物,处于极度饥饿状态,因此它们需要飞往别处寻找食物,以获取所需的能量^[25]。

2.2 时序网络架构设计

LSTM 是一种特殊的循环神经网络(recurrent neural network, RNN),其显著特征在于保留信息的内在能力^[26-27]。不同 LSTM 的重复单元构成各不相同,但都采用链式设计,其基本结构如图 1 所示。

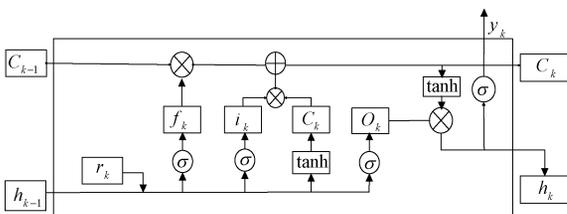


图 1 LSTM 的结构

Fig. 1 Structure diagram of LSTM

从图 1 可以看出,LSTM 由输入门、遗忘门和输出门 3 部分组成。其中,输入门用以控制输入以及当前计算

状态更新到记忆单元的程度;遗忘门主要决定在当前的单元状态中应该被保留到下一个时刻的部分;输出门则是主要决定当前时刻的隐藏状态中应该被传递到下一个时刻的部分^[28]。

首先,通过输入门决定了当前时刻的所有信息以何种量进入网络中,同时也防止一些无关紧要的信息进入网络,其数学模型如式(12)~(14)所示。

$$i_k = \sigma(W_i[h_{k-1}, r_k] + b_i) \quad (12)$$

$$C'_k = \tanh(W_c[h_{k-1}, r_k] + b_c) \quad (13)$$

$$C_k = f_k \cdot C_{k-1} + i_k \cdot C'_k \quad (14)$$

式中: W_i 和 b_i 分别表示输入门的权重矩阵和偏置项; σ 表示 S 型函数; W_c 和 b_c 分别表示输入门的权重矩阵和偏置项。

然后,通过遗忘门决定了当前时刻保存的最后时刻单元状态 C_{k-1} 个数 C_k ,如式(15)所示。

$$f_k = \sigma(W_f[h_{k-1}, r_k] + b_f) \quad (15)$$

式中: W_f 和 b_f 分别表示遗忘门的权重矩阵和偏置项。

最后,通过输出门决定控制最终单元状态 $c(t)$ 以何种程度输出到 LSTM 神经网络最后的输出 $h(t)$ 中,其数学模型如式(16)和(17)所示。

$$O_k = \sigma(W_o[h_{k-1}, r_k] + b_o) \quad (16)$$

$$h_k = O_k \times \tanh(C_k) \quad (17)$$

式中: W_o 和 b_o 分别表示输出门的权重矩阵和偏置项。

2.3 数据驱动的超参数动态寻优策略

将 LSTM 关键的超参数编码为 SSA 中的麻雀个体位置向量 $X_i = [x_1, x_2, x_3]$,其中, x_1 指学习率, x_2 指隐藏层节点数, x_3 指训练轮数。

之后定义适应度函数 $F(X_i)$ 衡量超参数组合的性能,融合预测精度与训练效率,其数学模型如式(18)所示。

$$F(X_i) = \alpha \cdot RMSE_{norm} + \beta \cdot \frac{T_{train}}{T_{max}} \quad (18)$$

式中: $RMSE_{norm}$ 指归一化后的均方根误差,反映预测精度; T_{train} 指当前参数下的单轮训练时间; T_{max} 指最大允许训练时间; α 和 β 为权重系数。

在整个过程中先解码麻雀位置为超参数组合,之后构建 LSTM 模型,使用训练集数据训练,利用式(10)调整参数进行全局搜索,利用式(11)进行局部开发,向最优解靠拢,在验证集中计算预测结果的均方根误差(RMSE)作为适应度值 $f(X)$,最后引入早停机制,若适应度持续无改善,终止当前训练以节省计算资源。

2.4 模型评价指标

为对模型预测结果进行评价,选取决定系数(R-squared, R^2)、RMSE、平均绝对误差(mean absolute error, MAE)、平均相对百分比误差(mean absolute percentage

error, MAPE) 作为评价指标^[10]。

R^2 表示变量之间的相关程度, 如式 (19) 所示。

$$R^2 = \frac{\sum_{i=1}^m (h(x_i) - \bar{y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (19)$$

式中: m 是数据点的数量; $h(x_i)$ 是第 i 个数据点的实际值; y_i 是模型对第 i 个数据点的预测值; \bar{y} 是 i 个数据点的实际值的平均值。

RMSE 表示预测值和观测值之间差异的样本标准差, 如式 (20) 所示。

$$RMSE(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2} \quad (20)$$

MAE 表示预测值和观测值之间绝对误差的平均值, 如式 (21) 所示。

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x_i) - y_i| \quad (21)$$

MAPE 表示预测值相对于真实值的相对误差大小, 如式 (22) 所示。

$$MAPE(X, h) = \frac{1}{m} \sum_{i=1}^m \frac{h(x_i) - y_i}{h(x_i)} \quad (22)$$

在以上 4 个评价指标中, R^2 是评估模型拟合程度的指标, 其数值越接近 1 表示模型对数据的拟合效果越好。对于其他指标来说, 其数值越小, 说明预测模型的拟合效果越好, 预测准确性越高, 相对误差越小。

2.5 输入特征分析与模型构建

实验样本为实测风电功率数据以及风电场对应的环境监测仪获取的风速、风向、气温和空气密度 4 种环境数据。为分析上述 4 种因素对风电功率的影响, 采用式 (23) Pearson 相关系数进行计算。

$$r_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (23)$$

式中: $|r| \leq 1$ 为相关系数; X, Y 分别表示风电场的环境数据与其时序功率; X_i, Y_i 表示第 i 序列组的值; r_{XY} 表示 Pearson 系数。

通过式 (23) 计算得到各因素与风电功率的相关系数, 如表 1 所示。

表 1 风电功率与环境因素的 Pearson 系数

Table 1 Pearson correlation coefficients between wind power and environmental factors

环境因素	Pearson 系数
风速	0.761
风向	0.124
气温	0.148
空气密度	-0.161

由表 1 可以看出, 风速对风电功率的影响最大, 风向

和气温对其影响较小, 而空气密度与风电功率呈负相关, 所以主要考虑风速对风电功率的影响。

针对插值数据结果的波动以及单一预测模型预测结果不精准的问题, 提出了一种基于 PCHIP-VMD 数据分析与 SSA-LSTM 模型的风电功率预测方法。首先, 采用 PCHIP-VMD 组合算法在剔除异常数据后填充比较拟合的数据, 并保证在不丢失原始数据信息的情况下对填充后的非平稳数据进行特征提取, 将稳定项序列作为预测模型的输入, 从而有效减小了数据的波动; 其次, 针对 LSTM 网络预测精确度较低的问题, 采用 SSA 算法优化 LSTM 网络, 模拟麻雀在寻求食物和躲避天敌的行为, 寻找 LSTM 网络参数的最优配置, 使 LSTM 充分利用原始数据得到最佳解; 最后, 将融合了数据处理模块的 PCHIP-VMD 网络与预测模型 SSA-LSTM 相组合进行风电功率预测。该组合模型通过 PCHIP-VMD 对数据进行预处理, 降低了原始数据中异常数据以及数据不稳定性对预测结果的影响。同时, SSA 和 LSTM 的结合能够充分解决最优参数配置问题, 提高了预测精度。

整个模型的预测流程如图 2 所示。

1) 异常数据处理。采用 GESD 对异常数据进行识别并剔除, 并用 PCHIP 算法对原始数据进行填充。

2) VMD 变分模态分解。使用 VMD 对数据进行分解, 得到一系列有不同频率和振幅的单分量信号。

3) 最优模型参数确定。利用训练集数据对 SSA-LSTM 组合预测模型进行学习训练, 对比预测结果, 确定达到目标准确率的最优参数。

4) 风电功率预测。采用 SSA 优化后得到的最佳模型对测试集数据进行测试, 得到待预测的风电功率, 计算预测指标对预测效果进行评价。

3 算例分析

为验证 PCHIP-VMD-SSA-LSTM 组合预测模型对风电功率预测的可行性与精准性, 本文以华中地区某风电场为算例进行分析。风电功率为 2023 年 10 月 1 日 0:00~10 月 22 号 0:00 的实测功率, 采样时间间隔为 15 min, 按照 7:3 的比例划分为训练集和测试集。模型输入为变模态分解后的风电功率数据, 使用 SSA 算法优化 LSTM 模型中的各个参数值以达到最佳预测, 模型输出为待预测日风电功率, 分别采用 PCHIP-LSTM、PCHIP-SSA-LSTM、PCHIP-VMD-LSTM 和 PCHIP-VMD-SSA-LSTM 4 种模型进行预测。

3.1 风电功率数据分析与处理实验

首先, 对风电功率数据进行 GESD 法找到相关的异常数据并进行剔除, 异常数据剔除前后风电功率散点图如图 3 和 4 所示。

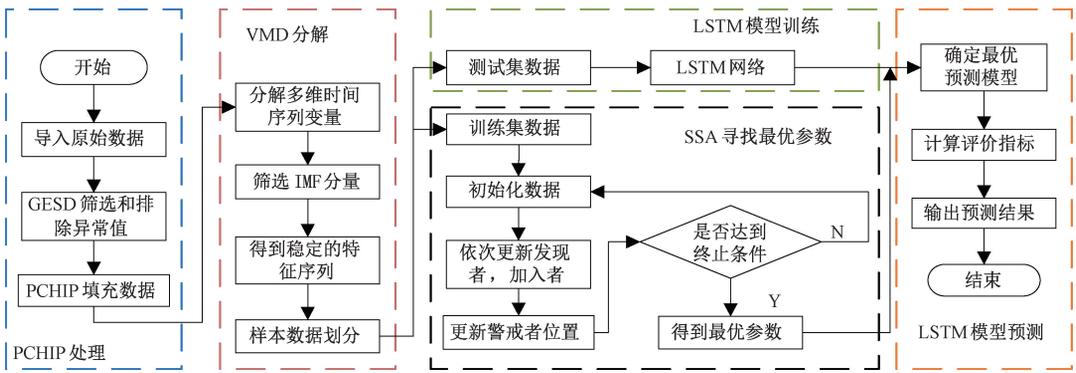


图 2 PCHIP-VMD-SSA-LSTM 预测流程

Fig. 2 Flowchart of PCHIP-VMD-SSA-LSTM prediction process

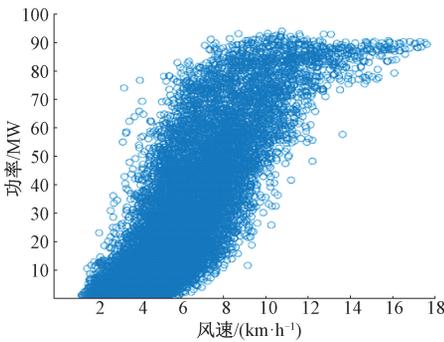


图 3 剔除前风速-风电功率散点图

Fig. 3 Scatter plot of wind speed versus wind power before outlier removal

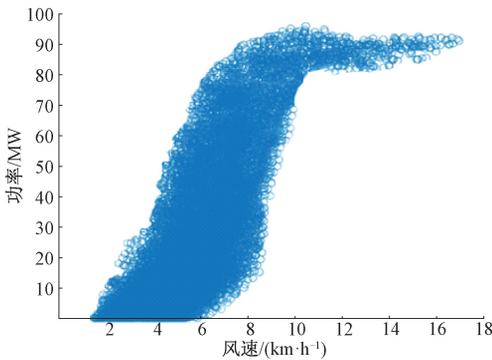


图 4 剔除后风速-风电功率散点图

Fig. 4 Scatter plot of wind speed versus wind power after outlier removal

受环境因素影响,复杂度高且非线性强,采用 VMD 算法对输入数据进行分解,得到多个单分量调幅调频信号,可在丰富输入变量的同时,根据得到的 IMF 分量,突出环境序列在不同时间尺度下的局部特性,反映出原始环境序列的波动性、周期性和趋势变化,分解得到原信号更为有效的特征信息。图 5 所示为风速特征序列 VMD 分解过程。

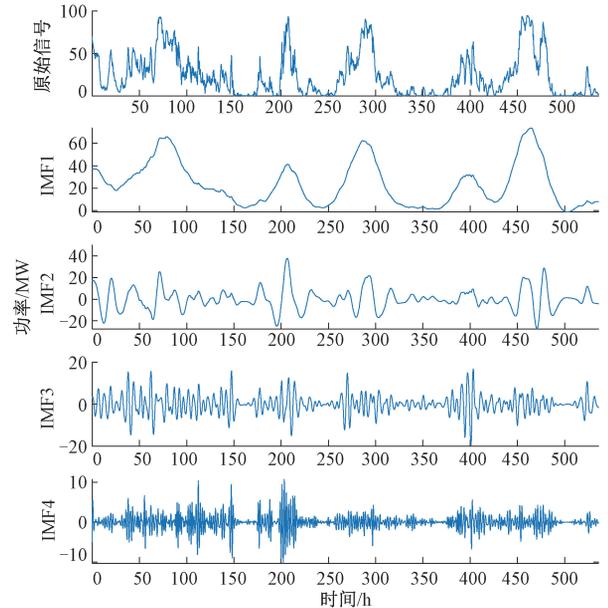


图 5 风速序列 VMD 分解过程

Fig. 5 VMD decomposition process of wind speed series

由图 3 和 4 可以看出,剔除后的数据大部分集中分布在理想的功率特性曲线周围,因此,利用 GESD 法能够有效对数据进行剔除。之后对剔除异常数据后以及空缺数据采用 PCHIP 法进行填充,确保数据的完整性。

整个实验数据中的风电功率数据为非平稳信号,

由于环境时间序列具有一定的随机性和间断性,经过 VMD 分解得到的数据序列充实了特征序列的数量,但是输入变量的维数也随之增多,由图 5 可以看出,IMF1 呈现频率降低、周期性的趋势,属于趋势项,而 IMF2~IMF4 呈现不平稳和振荡特性,属于随机项,因此将 IMF1 的特征序列作为预测模型的输入数据。

3.2 PCHIP-VMD-SSA-LSTM 模型参数设置

PCHIP-VMD-SSA-LSTM 网络预测模型参数设置如下:延时步长为 5,跨 1 个时间点进行预测。PCHIP-VMD-SSA-LSTM 具体参数如表 2 所示。

3.3 预测结果与对比分析

为验证 PCHIP-VMD-SSA-LSTM 组合模型的有效性和预测的高精度,采用 SSA 算法来优化 LSTM 模型参数,得到训练集和测试集的预测结果如图 6 和 7 所示。

由图 6 和 7 可以看出,训练集的预测结果良好,训练集的预测值与真实值相比相关性较高。从预测效果来看,测试集的预测效果整体低于训练集的预测效果好,但仍处于误差范围内,其拟合程度较高。对算例分别采用 PCHIP-LSTM、PCHIP-SSA-LSTM、PCHIP-VMD-LSTM 和

PCHIP-VMD-SSA-LSTM 4 种模型进行预测,预测结果如图 8 所示。

表 2 PCHIP-VMD-SSA-LSTM 参数设置

Table 2 Parameter settings of PCHIP-VMD-SSA-LSTM model

模型	模型参数	初始值
PCHIP	最大训练次数	1 000
-	梯度阈值	1
VMD	初始学习率	0.01
-	开始调整学习率次数	850
SSA	学习率调整因子	0.2
-	正则化参数	0.01
LSTM		

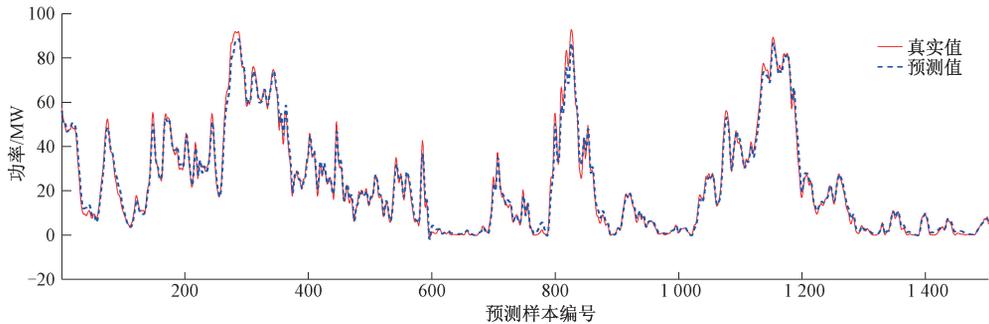


图 6 基于 PCHIP-VMD-SSA-LSTM 模型的风电集群功率训练集预测结果

Fig. 6 Prediction results of wind power cluster training set based on PCHIP-VMD-SSA-LSTM model

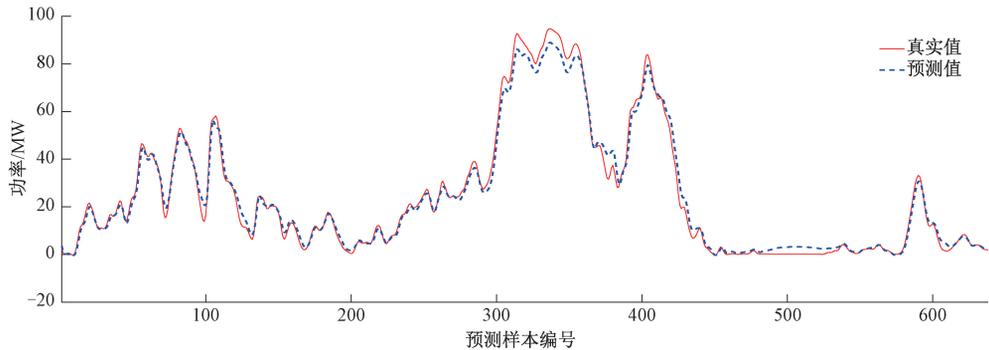


图 7 基于 PCHIP-VMD-SSA-LSTM 模型的风电集群功率测试集预测结果

Fig. 7 Prediction results of wind power cluster test set based on PCHIP-VMD-SSA-LSTM model

由图 8 可以看出,利用 PCHIP-VMD-SSA-LSTM 模型预测后的结果更接近于真实值,拟合了绝大部份时间序列功率值。PCHIP-LSTM、PCHIP-SSA-LSTM、PCHIP-

VMD-LSTM 和 PCHIP-VMD-SSA-LSTM 4 种模型功率预测结果误差如图 9 所示,所用各预测方法的误差指标如表 3 所示。

表 3 不同模型风电功率预测误差指标

Table 3 Error metrics of wind power prediction for five models

预测方法	R^2	RMSE	MAE	MAPE/%	预测精度/%
PCHIP-LSTM	0.937 02	6.693 1	4.775 2	14.799 2	基准
PCHIP-SSA-LSTM	0.976 64	4.076 4	2.714 4	8.105 7	+4.228
PCHIP-VMD-LSTM	0.943 95	6.260 8	4.705 2	14.331 1	+0.739 6
PCHIP-VMD-SSA-LSTM	0.989 10	2.761 5	1.994 8	5.904 5	+5.558

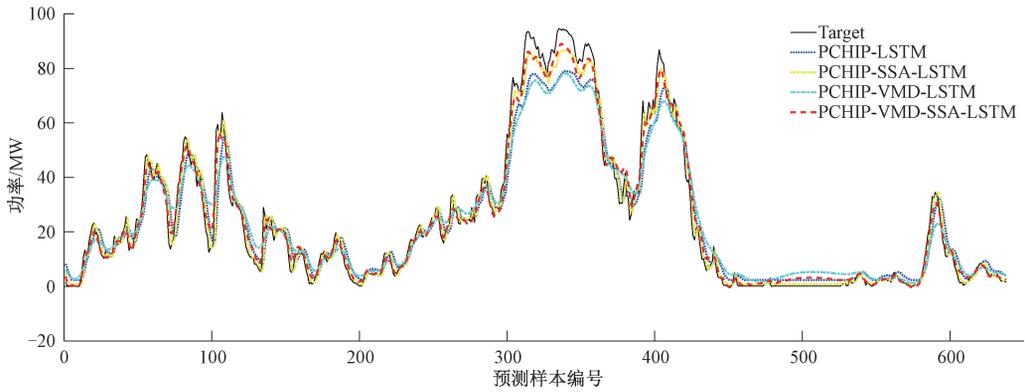


图 8 4 种模型功率预测结果

Fig. 8 Power prediction results of four models

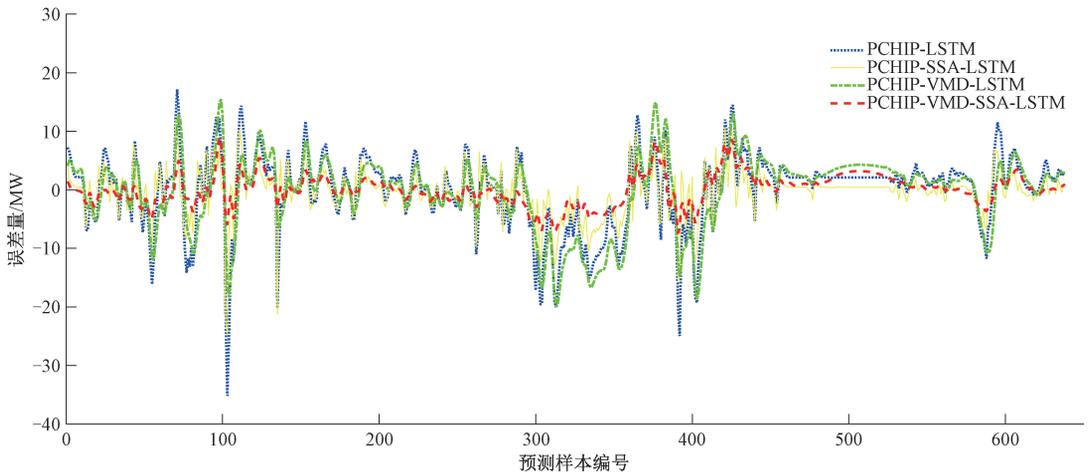


图 9 4 种模型风电功率预测结果误差

Fig. 9 Prediction error of wind power for four models

由表 3 可以看出, PCHIP-VMD-SSA-LSTM 模型的 RMSE、MAPE、MAE 指标均小于其他 3 种算法, 该模型的 RMSE、MAE 和 MAPE 与 PCHIP-LSTM 模型、PCHIP-SSA-LSTM 模型和 PCHIP-VMD-LSTM 模型相比, 分别下降了 58.74%、58.22%、60.10%, 32.25%、26.51%、27.16% 和 55.89%、57.60%、58.80%, 且 R^2 更接近于 1, 由此表明, 采用 VMD 算法预测精度提高明显。同时在输入完全相同的情况下, 与其他 3 种方法相比, PCHIP-VMD-SSA-LSTM 方法的预测精度有明显提升, 表明 PCHIP-VMD-SSA-LSTM 模型相较于其他模型预测效果更好。

为进一步评估模型的效果, 将本文所用方法与现有先进方法进行对比分析, 结果如表 4 所示。

通过对当前不同预测方法的对比分析可以看出, 本文提出的 PCHIP-VMD-SSA-LSTM 模型在风电功率预测中表现出显著优势。文献[10]采用的 VMD-IMPA-SVM 方法虽然通过改进海洋捕食算法优化支持向量机参数, 但其 RMSE 和 MAPE 均明显高于本文方法, 主要是由于

SVM 模型对非平稳时序数据的非线性特征捕捉能力较弱, 且未考虑分解残差项的累积误差所带来的影响。文献[13]提出的 IAOA-VMD-LSTM 模型虽引入算术优化算法进行参数调优, 但其 RMSE 和 MAE 指标显著劣化, 表明当前优化算法在 LSTM 超参数寻优时易陷入局部最优的缺陷。文献[19]的 VMD-SSA-LSSVM 方法虽通过麻雀算法优化支持向量机取得较高 R^2 , 但其 MAPE 仍高于本文方法, 表明单一的分解-预测模型对高频噪声抑制不足。

表 4 与当前方法对比分析

Table 4 Comparative analysis with state-of-the-art methods

预测方法	R^2	RMSE	MAE	MAPE/%
文献[10]	0.977 48	11.464	8.521	14.181
文献[13]	0.978 1	16.59	12.37	1.34
文献[19]	0.986 4	2.819 6	2.088 9	7.955
本文方法	0.989 1	2.761 5	1.994 8	5.904 5

实验数据显示,本文方法 R^2 达到 0.989 1,较文献[10,13,19]分别提升 1.18%、1.1%和 0.27%,且 RMSE 和 MAE 较对比方法均降低,验证了 PCHIP 插值保持数据导数连续性、SSA 算法全局搜索参数优化的协同效应。本文方法的 MAPE 为 5.904 5%,仅次于文献[13]方法,表示其在极端天气等复杂工况下也能保持预测的稳定性。

本文将寻找数据最优解的 SSA 模型与 LSTM 时序特性模型结合起来,充分考虑影响风电功率预测结果的各项因素的相关特性,相较当前参数优化方法,该机制使 RMSE 降低 58.74%(从 6.69 降至 2.76),验证了 SSA 在参数空间探索能力上的优越性。同时采用 PCHIP-VMD 联合数据预处理,在保留数据统计特性的同时将非平稳信号分解为 4 个本征模态分量,使 SSA-LSTM 模型能针对频率分量建立预测策略。相较于单一 LSTM 模型,融合 SSA 优化的混合架构使 R^2 提升 4.428%;而 PCHIP-VMD 预处理进一步提高 1.33%精度增益,证实了多技术协同的倍增效应。

4 结 论

本文提出了风电功率 PCHIP-VMD-SSA-LSTM 组合预测模型,成功解决插值后整体数据非平稳、预处理保形性不足、模态混叠、预测模型参数优化效率低等问题。利用 PCHIP 对相关数据进行填充以及 VMD 模型对风电历史功率数据进行分解和处理来有效降低具有显著变异性、非平稳性及动态结构特性的时间序列,得到不同时间尺度上变化的数据信号。同时为了进一步提高预测精度,解决 LSTM 模型因局部最优导致的预测偏差累积问题,将 SSA 模型与 LSTM 时序特性模型结合起来进行模型参数优化,从而形成了数据重构-特征分解-参数调优的 3 层级优化架构的创新设计。实验结果表明,利用 PCHIP-VMD 对数据的深度处理能力与 SSA-LSTM 模型的强学习能力形成“数据清洗→特征分解→参数优化→时序预测”的解决方案,可以具有更好的预测精度,为高比例新能源电力系统调度提供了更可靠的技术支撑。未来工作将致力于将该框架拓展至风电-光伏-储能等多能源协同预测与调度优化,构建多能互补动态模型,提升高比例新能源电网的稳定性与消纳能力。

参考文献

[1] 魏旭,刘东,高飞,等. 双碳目标下考虑源网荷储协同优化运行的新型电力系统发电规划[J]. 电网技术, 2023,47(9):3648-3661.
WEI X, LIU D, GAO F, et al. Generation planning of a new-type power system considering the collaborative optimal operation of source-grid-load-storage under the

dual carbon objectives [J]. Power System Technology, 2023, 47(9): 3648-3661.

- [2] 高强,秦晓辉,张媛媛,等. 碳减排需求下面向多元交互策略设计的微电网分布式能源发展规划研究[J]. 电网技术, 2024,48(9):3574-3583.
GAO Q, QIN X H, ZHANG Y Y, et al. Research on the development planning of distributed energy in microgrids oriented to multi-interactive strategy design under carbon emission reduction demands [J]. Power System Technology, 2024,48(9):3574-3583.
- [3] 王慧莹,吴亮红,梅盼盼,等. 果蝇优化广义神经网络的风电功率短期预测[J]. 电子测量与仪器学报, 2019,33(6):177-183.
WANG H Y, WU L H, MEI P P, et al. Short-term wind power prediction based on fruit fly optimization algorithm and generalized neural network [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(6): 177-183.
- [4] 任轩,汪庆年,尚宝,等. 基于混合神经网络的短期电力负荷预测方法[J]. 电子测量技术, 2022, 45(14): 71-77.
REN X, WANG Q N, SHANG B, et al. Short-term power load forecasting method based on hybrid neural network [J]. Electronic Measurement Technology, 2022, 45(14): 71-77.
- [5] SARANGI S, DASH K P, B S R. Short-term prediction of wind power using an improved kernel based optimized deep belief network [J]. Energy Conversion and Management, 2024,316(4):118821.
- [6] LIN Y, HU G, WANG L, et al. A multi-AGV routing planning method based on deep reinforcement learning and recurrent neural network[J]. IEEE/CAA Journal of Automatica Sinica, 2024,11(7):1720-1722.
- [7] 马志侠,张林鎔,巴音塔娜,等. 基于自适应二次分解与 CNN-BiLSTM 的超短期风电功率预测[J]. 太阳能学报, 2024,45(6):429-435.
MA ZH X, ZHANG L L, BAYINTANA, et al. Ultra short term wind power forecasting based on adaptive quadratic decomposition and CNN-BiLSTM [J]. Acta Energetica Solaris Sinica, 2024, 45(6): 429-435.
- [8] ZHAO Y, LIAO H, PAN S, et al. Interpretable multi-graph convolution network integrating spatio-temporal attention and dynamic combination for wind power forecasting[J]. Expert Systems with Applications, 2024, 255(2): 124766.
- [9] 王永生,张哲,刘利民,等. 基于改进熵权法和 SECEEMD 的短期风电功率预测[J]. 科学技术与工程, 2023,23(27):11688-11697.

- WANG Y SH, ZHANG ZH, LIU L M, et al. Short-term wind power forecasting based on improved entropy weight method and SECEEMD [J]. *Science Technology and Engineering*, 2023, 23(27): 11688-11697.
- [10] 刘金朋, 邓嘉明, 高鹏宇, 等. 基于 VMD-IMPA-SVM 的超短期风电功率预测[J]. *智慧电力*, 2024, 52(7): 24-31, 79.
- LIU J P, DENG J M, GAO P Y, et al. Ultra-short-term wind power forecasting based on VMD-IMPA-SVM [J]. *Smart Power*, 2024, 52(7): 24-31, 79.
- [11] 赵一鸣, 吉月辉, 刘俊杰, 等. 基于 EMD-IPSO-LSTM 模型的短期电力负荷预测[J]. *国外电子测量技术*, 2023, 42(1): 132-137.
- ZHAO Y M, JI Y H, LIU J J, et al. Short-term power load forecasting based on the EMD-IPSO-LSTM model [J]. *Foreign Electronic Measurement Technology*, 2023, 42(1): 132-137.
- [12] 李俊卿, 李秋佳. 基于 Kriging 和长短期记忆网络的风电功率预测方法[J]. *太阳能学报*, 2020, 41(11): 241-247.
- LI J Q, LI Q J. Wind power forecasting method based on kriging and long short-term memory network [J]. *Acta Energeticae Solaris Sinica*, 2020, 41(11): 241-247.
- [13] 肖烈禧, 张玉, 周辉, 等. 基于 IAOA-VMD-LSTM 的超短期风电功率预测[J]. *太阳能学报*, 2023, 44(11): 239-246.
- XIAO L X, ZHANG Y, ZHOU H, et al. Ultra-short-term wind power forecasting based on IAOA-VMD-LSTM [J]. *Acta Energeticae Solaris Sinica*, 2023, 44(11): 239-246.
- [14] 毕贵红, 赵鑫, 陈臣鹏, 等. 基于多通道输入和 PCNN-BiLSTM 的光伏发电功率超短期预测[J]. *电网技术*, 2022, 46(9): 3463-3476.
- BI G H, ZHAO X, CHEN CH P, et al. Ultra-short-term forecasting of photovoltaic power generation based on multi-channel input and PCNN-BiLSTM [J]. *Power System Technology*, 2022, 46(9): 3463-3476.
- [15] 李丹, 唐建, 缪书唯, 等. 考虑时序特征缺失值动态插补的超短期风电功率预测[J/OL]. *中国电机工程学报*, 1-15 [2025-04-16]. <http://kns.cnki.net/kcms/detail/11.2107.tm.20240425.1905.005.html>.
- LI D, TANG J, MIAO SH W, et al. Ultra-short-term wind power forecasting considering dynamic imputation of missing values in temporal characteristics [J/OL]. *Proceedings of the CSEE*, 1-15 [2025-04-16]. <http://kns.cnki.net/kcms/detail/11.2107.tm.20240425.1905.005.html>.
- [16] 张淑清, 宿新爽, 陈荣飞, 等. 基于变分模态分解和 FABP 的短期电力负荷预测[J]. *仪器仪表学报*, 2018, 39(4): 67-73.
- ZHANG SH Q, SU X SH, CHEN R F, et al. Short-term power load forecasting based on variational mode decomposition and FABP [J]. *Chinese Journal of Scientific Instrument*, 2018, 39(4): 67-73.
- [17] 刘景森, 吉宏远, 李煜. 基于改进蝙蝠算法和三次样条插值的机器人路径规划[J]. *自动化学报*, 2021, 47(7): 1710-1719.
- LIU J S, JI H Y, LI Y. Robot path planning based on improved bat algorithm and cubic spline interpolation [J]. *Acta Automatica Sinica*, 2021, 47(7): 1710-1719.
- [18] 宫婷, 车建峰, 王勃, 等. 考虑误差概率分布及波动特性的短期风电功率预测修正方法[J]. *高压技术*, 2025, 51(1): 379-389.
- GONG T, CHE J F, WANG B, et al. Correction method for short-term wind power forecasting considering error probability distribution and fluctuation characteristics [J]. *High Voltage Engineering*, 2025, 51(1): 379-389.
- [19] 王维高, 魏云冰, 滕旭东. 基于 VMD-SSA-LSSVM 的短期风电预测[J]. *太阳能学报*, 2023, 44(3): 204-211.
- WANG W G, WEI Y B, TENG X D. Short-term wind power forecasting based on VMD-SSA-LSSVM [J]. *Acta Energeticae Solaris Sinica*, 2023, 44(3): 204-211.
- [20] 郭建帅, 崔双喜, 郭建斌, 等. 基于 VMD-SSA-HKELM 的超短期负荷预测[J]. *国外电子测量技术*, 2022, 41(6): 105-111.
- GUO J SH, CUI SH X, GUO J B, et al. Ultra-short-term load forecasting based on VMD-SSA-HKELM [J]. *Foreign Electronic Measurement Technology*, 2022, 41(6): 105-111.
- [21] WANG Z, YING Y, KOU L, et al. Ultra-short-term offshore wind power prediction based on PCA-SSA-VMD and BiLSTM [J]. *Sensors*, 2024, 24(2): 444.
- [22] 章志晃, 徐启峰, 林穿. 基于 SSA-BiLSTM-AM 的短期风电功率预测[J]. *国外电子测量技术*, 2023, 42(3): 46-51.
- ZHANG ZH H, XU Q F, LIN CH. Short-term wind power forecasting based on SSA-BiLSTM-AM [J]. *Foreign Electronic Measurement Technology*, 2023, 42(3): 46-51.
- [23] 马宏忠, 肖雨松, 颜锦, 等. 基于多特征提取和麻雀搜索算法优化 XGBoost 的变压器绕组松动诊断方法[J]. *电机与控制学报*, 2024, 28(6): 87-97.
- MA H ZH, XIAO Y S, YAN J, et al. Diagnosis method of transformer winding looseness based on multi-feature extraction and sparrow search algorithm optimized XGBoost [J]. *Electric Machines and Control*, 2024, 28(6): 87-97.

- [24] 张俊潇,高崇,劳永钊,等. 基于混沌二进制麻雀搜索的配电网故障恢复方法[J]. 高电压技术, 2023, 49(S1): 247-253.
ZHANG J X, GAO CH, LAO Y ZH, et al. Distribution network fault recovery method based on chaotic binary sparrow search algorithm [J]. High Voltage Engineering, 2023, 49(S1): 247-253.
- [25] 张群,唐振浩,王恭,等. 基于长短时记忆网络的超短期风功率预测模型[J]. 太阳能学报, 2021, 42(10): 275-281.
ZHANG Q, TANG ZH H, WANG G, et al. Ultra-short-term wind power forecasting model based on long short-term memory network [J]. Acta Energiæ Solaris Sinica, 2021, 42(10): 275-281.
- [26] 王贺,陈燕峰,熊敏,等. 融合 CEEMDAN 和 ICS-LSTM 的短期风速预测建模[J]. 电子测量与仪器学报, 2022, 36(4): 17-23.
WANG H, CHEN H F, XIONG M, et al. Short-term wind speed forecasting modeling based on CEEMDAN and ICS-LSTM [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(4): 17-23.
- [27] 魏玮,吕游,齐欣宇,等. 基于 CNN-LSTM-AM 动态集成模型的电站风机状态预测方法[J]. 仪器仪表学报, 2023, 44(4): 19-27.
WEI W, LYU Y, QI X Y, et al. Power station fan condition prediction method based on CNN-LSTM-AM dynamic ensemble model [J]. Chinese Journal of Scientific Instrument, 2023, 44(4): 19-27.
- [28] 杨生鹏,文中,丁剑,等. 二次分解组合 LSTM 的短期风电功率预测模型[J]. 国外电子测量技术, 2024, 43(1):

87-93.

YANG SH P, WEN ZH, DING J, et al. Short-term wind power forecasting model based on secondary decomposition and combined LSTM [J]. Foreign Electronic Measurement Technology, 2024, 43(1): 87-93.

作者简介



张登攀, 2024 年于郑州轻工业大学获得学士学位, 现为湖南工业大学硕士研究生, 主要研究方向为电力系统运行优化、新能源发电与微电网。

E-mail: 3518421865@qq.com

Zhang Dengpan received his B. Sc. degree from Zhengzhou University of Light Industry in 2024. Now he is a M. Sc. candidate in Hunan University of Technology. His main research interests include power system operation optimization, new energy generation, and microgrids.



兰征(通信作者), 分别在 2008 年于湖南科技大学获得学士学位, 2011 年和 2017 年于湖南大学获得硕士学位和博士学位, 现为湖南工业大学教授, 研究方向为电能质量、电力电子技术及其在电力系统中的应用。

E-mail: lan@hut.edu.cn

Lan Zheng (Corresponding author) received B. Sc. from Hunan University of Science and Technology in 2008, M. Sc. and Ph. D. both from Hunan University in 2011 and 2017, respectively. Now he is a professor in Hunan University of Science and Technology. His main research interests include power quality, power electronics technology, and their applications in power systems.