

采用级联回归残差校正麻雀搜索算法优化 长短期记忆网络的光伏发电功率预测方法

高淑萍^{1,2}, 张昊洋^{1,2}, 宋国兵³, 张警行^{1,2}, 赵萌雨^{1,2}

(1. 西安科技大学电气与控制工程学院, 710054, 西安; 2. 西安科技大学西安市电气设备状态监测与
供电安全重点实验室, 710054, 西安; 3. 西安交通大学电气工程学院, 710049, 西安)

摘要: 针对光伏发电功率序列波动性强、非线性特征显著以及预测精度易受噪声干扰等问题, 提出一种基于改进自适应噪声完备集合经验模态分解(ICEEMDAN)与级联回归(CR)残差校正麻雀搜索算法优化长短期记忆网络(SSA-LSTM)的光伏发电功率预测方法。首先, 采用 ICEEMDAN 分解对原始光伏发电功率序列进行分解, 得到多个固有模式函数(IMF), 以提取不同频率尺度下的时序特征并削弱噪声干扰; 其次, 引入麻雀搜索算法(SSA)对长短期记忆网络(LSTM)模型的关键参数进行寻优, 并利用优化后的 LSTM 模型对各 IMF 分量分别进行预测; 最后, 采用级联回归对各分量重构后的预测结果进行残差校正, 进一步减小模型输出偏差, 提高整体预测精度。仿真结果表明: 按四季测试样本数加权平均计算, 相较于传统 LSTM 模型, 所提方法平均绝对误差(MAE)降低了 69.97%, 均方根误差(RMSE)降低了 67.49%; 相较于未校正模型, MAE 降低了 34.01%, RMSE 降低了 27.78%, 决定系数 R^2 大于 0.96。研究结果表明, 该方法能够有效降低预测偏差, 提高模型在不同季节条件下的稳定性和适应性, 为光伏发电功率预测提供了高精度、高鲁棒性的技术手段。

关键词: 光伏发电功率预测; 级联回归残差校正; 麻雀搜索算法; 自适应噪声完备集合经验模态分解; 长短期记忆网络

中图分类号: TM615 文献标志码: A

DOI: 10.7652/xjtuxb202609003 文章编号: 0253-987X(2026)09-0019-11

Photovoltaic Power Forecasting Method Based on a Long Short-Term Memory Network Optimized by Cascade Regression

GAO Shuping^{1,2}, ZHANG Haoyang^{1,2}, SONG Guobing³,
ZHANG Jingxing^{1,2}, ZHAO Mengyu^{1,2}

(1. School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;
2. Xi'an Key Laboratory of Electrical Equipment Condition Monitoring and Power Supply Security,
Xi'an University of Science and Technology, Xi'an 710054, China;
3. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710054, China)

Abstract: To address the strong volatility and significant nonlinear characteristics of photovoltaic power sequences, as well as the susceptibility of forecasting accuracy to noise interference, a photovoltaic power forecasting method integrating improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), a sparrow search algorithm-optimized long

收稿日期: 2026-02-23。 作者简介: 高淑萍(1970—), 女, 副教授, 硕士生导师。 基金项目: 国家自然科学基金资助项目(52577138)。

网络出版时间: 2026-06-24

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20260623.1506.008>

short-term memory network (SSA-LSTM), and cascade regression (CR) error correction is proposed. The original power sequence is decomposed using ICEEMDAN to obtain multiple intrinsic mode functions (IMFs), thereby effectively separating multiscale temporal features and suppressing noise interference. The sparrow search algorithm (SSA) is introduced to adaptively optimize the key parameters of the LSTM model. Cascade regression is used to perform parallel error correction on the preliminary prediction results obtained after reconstruction of the components, thereby further reducing the model output bias and improving the overall prediction accuracy. It is shown by simulation results that, based on a weighted average according to the numbers of test samples in the four seasons, the mean absolute error (MAE) and root mean square error (RMSE) are reduced by 69.97% and 67.49%, respectively, compared with those of the conventional LSTM model. Compared with the uncorrected model, the e_{MAE} and e_{RMSE} are reduced by 34.01% and 27.78%, respectively, and all coefficients of determination (R^2) exceed 0.96. The results show that the forecasting bias can be effectively reduced and the stability and adaptability of the model under different seasonal conditions can be improved by the proposed method, providing a high-accuracy and highly robust technical approach for photovoltaic power forecasting.

Keywords: photovoltaic power forecasting; cascade regression; sparrow search algorithm; complete ensemble empirical mode decomposition with adaptive noise; long short-term memory network

近年来,随着光伏发电装机规模的持续增长,其在电力系统中的作用不断增强^[1-2]。由于光伏发电功率易受太阳辐射强度、云量变化、环境温度及季节因素等影响,输出功率呈现出较强的波动性和随机性。随着光伏并网规模的扩大,这种不确定性给电网调度运行、功率平衡及新能源消纳带来了较大压力。因此,提高光伏发电功率预测精度,对于保障电力系统安全稳定运行具有重要意义^[3-4]。

光伏发电功率预测与气象条件密切相关,而气象数据本身具有非线性、时变性和波动性等特点,这在一定程度上增加了预测建模的难度。现有研究中,极限学习机^[5]、支持向量回归^[6]、神经网络^[7]等方法已被广泛应用于光伏发电功率预测任务,但在复杂气象条件下对非线性波动特征和随机扰动的刻画能力存在一定的局限,因此,对传统预测模型进行优化改进具有重要意义。文献[8]提出一种基于改进灰狼算法优化长短期记忆网络算法(LSTM)的光伏功率预测模型。文献[9]结合变分模态分解(VMD)和改进灰狼算法,搭建了基于VMD结合样本熵(SE)和优化支持向量机的光伏发电功率预测模型,但其所用的灰狼算法的寻优能力一般,且只有当误差变量符合高斯分布时才能达到最优。文献[10]提出了改进的自适应噪声完备集合经验模态分解(ICEEMDAN)算法,能够有效降低分解过程中产生的残余噪声及虚假分量。文献[11]使用粒子群优化算法(PSO)对反向传播网络进行优化,得到了更加精

确的预测结果,但粒子群算法难以保证得到最优解,同时会消耗大量运算资源。文献[12]构建了利用麻雀搜索算法(SSA)和泄漏积分型回声状态网络(LIESN)的短期光伏发电功率预测模型。文献[13]使用改进的麻雀搜索算法优化极限梯度提升模型,不仅提升了模型精度,还降低了模型计算时间。文献[14-15]提出了一种使用级联回归(CR)模型优化图像关键点的技术。

基于此,本文提出了一种基于LSTM的深度学习预测模型,并结合级联回归策略、麻雀搜索算法(SSA)和改进的ICEEMDAN方法进行优化。LSTM网络能够有效捕捉光伏发电输出中的时序特征,特别适合处理非线性和时变的动态数据。为进一步减小预测残差并提升连续时间段内预测结果的稳定性,本文引入CR残差校正模块对预测结果进行校正。针对数据中的噪声干扰和模态混叠问题,本文采用ICEEMDAN方法对原始功率数据进行预处理,将数据分解为多个本征模态函数(IMF),有效地减少了噪声影响,同时保留了关键的信号特征。此外,利用麻雀搜索算法对LSTM模型超参数进行优化,解决了传统方法中存在的经验依赖性和收敛性问题,进一步提升了模型的预测精度和收敛效率。

1 理论分析

1.1 ICEEMDAN 分解算法

ICEEMDAN 是在自适应噪声完备集合经验模

态分解(CEEMDAN)算法基础上发展而来的一种改进型信号处理方法。该方法通过引入互补噪声扰动机制与自适应集成策略,在缓解传统经验模态分解(EMD)类算法中常见的模态混叠问题方面具有显著优势^[16]。通过对原始信号进行多尺度分解,ICEEMDAN能够提取出一组具有物理意义的固有模态函数分量,各分量分别对应于信号在不同频率带宽下的局部振荡特性。ICEEMDAN的基本思路是在多次噪声扰动下构造信号样本集,并对每组扰动信号分别执行EMD分解,随后对同阶模态分量进行集成平均,从而获得稳定而具有代表性的本征模态分量。与CEEMDAN及其他变体相比,ICEEMDAN在残差更新策略和趋势信息提取方面进行了改进,增强了算法对突变点处理的敏感性,并有效抑制了边界效应对分解质量的影响^[17]。因此,ICEEMDAN在复杂时序信号建模与分析中表现出更高的分解精度和更强的适应性,其算法分解过程如下。

步骤1 设原始信号为 x ,第 i 次加入的高斯白噪声为 $\omega^{(i)}$,其中 $i=1,2,\dots,I$, I 为集合规模,且 $\omega^{(i)}$ 的均值为0、方差为1。 $E_k(\cdot)$ 表示经EMD分解得到的第 k 个本征模态分量; $\langle \cdot \rangle$ 表示对集合规模为 I 的噪声实现对应的同阶段结果取集合平均; β_k 表示第 $k+1$ 阶段分解时加入白噪声的幅值系数。构造的第1阶段待分解信号计算式为

$$Z_1^{(i)} = x + \beta_0 E_1(\omega^{(i)}) \quad (1)$$

式中: $Z_1^{(i)}$ 为第 i 次实验构造的第1阶段待分解信号; β_0 为首次分解时加入的白噪声系数。

步骤2 对 $Z_1^{(i)}$ 进行EMD分解,并与EMD分解的分量做差取集合平均得到残差信号

$$r_1 = \langle Z_1^{(i)} - E_1(Z_1^{(i)}) \rangle \quad (2)$$

x 与 r_1 之差即为ICEEMDAN的第1个本征模态分量 A_{IMF1} ,表达式为

$$A_{IMF1} = x - r_1 \quad (3)$$

步骤3 以 r_1 为基础构造第2阶段辅助信号,加入第2阶段噪声模态并取集合平均,得到第2个残差信号 r_2 ,其与 r_1 之差即为ICEEMDAN的第2个本征模态分量 A_{IMF2} ,表达式为

$$Z_2^{(i)} = r_1 + \beta_1 E_2(\omega^{(i)}) \quad (4)$$

$$r_2 = \langle Z_2^{(i)} - E_2(Z_2^{(i)}) \rangle \quad (5)$$

$$A_{IMF2} = r_1 - r_2 \quad (6)$$

步骤4 重复步骤3,则第 k 个本征模态分量为

$$A_{IMFk} = r_{k-1} - r_k \quad (7)$$

步骤5 重复上述操作,直至当前残余信号为单调函数或其极值点数不足以继续提取IMF分量。最

终,原始信号可重构为

$$x = \sum_{k=1}^K A_{IMFk} + r_K \quad (8)$$

式中: K 为提取的IMF个数; r_K 为最终残差信号。

1.2 SSA算法

麻雀搜索算法是一种受麻雀群体觅食与防捕食行为启发的智能优化算法,具有实现简单、全局搜索能力较强、参数较少等特点,在部分复杂优化问题中表现出较好的收敛性能。本文针对LSTM模型参数选取问题,引入SSA算法对其关键参数进行自适应优化^[18],并与粒子群优化算法、模拟退火算法(SA)、遗传算法(GA)进行对比试验,验证其适用性。

在SSA算法中,种群主要包括生产者 and 追随者,并从中选取一定比例的预警个体。生产者通常对应适应度较优的个体,负责搜索潜在优质区域并引导种群更新,其算法原理如下。

步骤1 生产者的位置更新计算式为

$$X_{a,d}^{g+1} = \begin{cases} X_{a,d}^g \exp\left(\frac{-i}{aG_{\max}}\right), & R < \theta \\ X_{a,d}^g + QD, & R \geq \theta \end{cases} \quad (9)$$

式中: g 为当前迭代次数; $G_{\max}$ 为设定最大迭代次数; $X_{a,d}^g$ 为第 g 次迭代、第 a 个麻雀个体的第 d 维取值; a 为 $[0,1]$ 内的随机数; θ 为取值范围在 $[0.5,1.0]$ 的安全阈值; Q 为服从正态分布的随机数; D 为所有元素为1的 $1 \times d$ 维行向量; R 表示安全判断随机数,当 $R < \theta$ 时,表示生产者按指数衰减方式搜索;当 $R \geq \theta$ 时,种群进入警戒状态并采用随机扰动更新位置。

步骤2 跟随者的位置更新计算式为

$$X_{a,d}^{g+1} = \begin{cases} Q \exp\left(\frac{X_{\text{worst}}^g - X_{a,d}^g}{i^2}\right), & a > \frac{N_p}{2} \\ X_p^{g+1} + |X_{a,d}^g - X_p^{g+1}|AD, & a \leq \frac{N_p}{2} \end{cases} \quad (10)$$

式中: X_p^{g+1} 为第 $g+1$ 次迭代中生产者的最优位置; X_{worst}^g 为第 g 次迭代中生产者的全局最差位置; N_p 为种群大小; A 为每个元素是1或-1的 $1 \times d$ 维矩阵。

步骤3 当预警个体检测到潜在风险时,触发预警位置更新机制,预警者的位置更新计算式为

$$X_{a,d}^{g+1} = \begin{cases} X_{\text{best}}^g + \lambda(X_{a,d}^g - X_{\text{best}}^g), & F_\zeta > F_{\text{best}} \\ X_{a,d}^g + H \frac{(X_{a,d}^g - X_{\text{worst}}^g)}{(F_\zeta - F_{\text{worst}}) + \delta}, & F_\zeta = F_{\text{best}} \end{cases} \quad (11)$$

式中: λ 作为步长控制参数,是均值为0、方差为1的正态分布随机数; H 是麻雀移动方向,范围为 $[-1,1]$; F_ζ 为麻雀搜索算法种群第 ζ 的适应度; F_{best} 和 F_{worst} 分别为麻雀搜索算法种群的全局最优和最差适应

度; $\mathbf{X}_{\text{best}}^g$ 为第 g 次迭代时种群的最优位置; δ 是为避免分母为 0 而设置的较小正数。

1.3 级联回归算法

级联回归算法是一种训练后能够快速定位目标关键点的算法,回归的目标并非只有关键点本身,同时包括回归过程中关键点的形状增量集合。其原理是通过训练得到一组形状回归器 $\mathbf{B}$,回归过程从初始形状 $\mathbf{S}^0$ 开始,经过 M 次级联迭代逐步逼近最终预测形状 $\mathbf{S}^M$ 。每一次迭代均由不同的回归器完成,并且每一个回归器的输入都依赖前一个回归器的输出,所有的级联回归器是通过人工标注关键点位置 $\hat{\mathbf{S}}$ 与当前回归形状的对齐误差学习到的^[19-20]。级联回归算法提取特征能够直接将关注点全部放在目标上,且无需对整体作特征处理,其训练思路流程如下。

步骤 1 首先输入图像数据 $\mathbf{A}$ 及其人工标注关键点位置 $\hat{\mathbf{S}}$,在这个过程中,关键点位置 $\hat{\mathbf{S}}$ 的表现形式为 $\{u_0, \dots, u_L, v_0, \dots, v_L\}$, L 为事先规定的关键点数, u_L, v_L 分别为关键点位置的横、纵坐标。

步骤 2 以预测形状与真实标注形状之间的平方误差为优化目标,表达式为

$$\min \epsilon = \sum_{q=1}^N \|\mathbf{S}_q - \hat{\mathbf{S}}_q\|_2^2 \quad (12)$$

式中: $\mathbf{S}_q$ 为第 q 张图像的预测形状向量; $\hat{\mathbf{S}}_q$ 为第 q 张图像的人工标注关键点位置形状向量; N 为输入图像的样本数。

步骤 3 关键点位置按照上一阶段的估计结果和本阶段预测的形状增量进行更新,表达式为

$$\mathbf{S}_q^m = \mathbf{S}_q^{m-1} + \Delta \mathbf{S}_q^{m-1} \quad (13)$$

式中: $\mathbf{S}_q^{m-1}$ 为第 q 张图像的第 $m-1$ 次迭代结果; $\mathbf{S}_q^m$ 为第 m 次迭代结果, $m = 1, 2, \dots, M$, M 为级联回归算法的迭代次数最大值; $\Delta \mathbf{S}_q^{m-1}$ 为第 $m-1$ 次迭代的形状增量。

步骤 4 第 m 级回归器根据图像及当前估计形状预测形状增量,其表达式为

$$\Delta \mathbf{S}_q^m = \mathbf{B}^m(\mathbf{A}_q, \mathbf{S}_q^{m-1}) \quad (14)$$

式中: $\Delta \mathbf{S}^m$ 为第 m 次迭代的形状增量; $\mathbf{B}^m$ 为第 m 次迭代所用的回归器; $\mathbf{A}_q$ 为第 q 张输入图像。

虽然级联回归最初主要用于图像关键点定位,但其“逐步修正预测误差”的思想同样适用于时序预测任务。本文借鉴其误差逐级校正机制,将其扩展用于光伏功率预测结果的残差建模与修正。

1.4 长短期记忆网络

在时间序列建模中,传统循环神经网络(RNN)在处理长序列时易产生梯度消失或梯度爆炸现象,导

致长期依赖特征无法有效保留。为克服此问题,研究者提出了长短期记忆网络结构,通过引入门控机制与细胞状态,实现对时序信息的长期记忆与动态更新^[21]。

LSTM 网络结构如图 1 所示,遗忘门和输入门共同更新细胞状态,输出门据此生成当前隐藏状态,当前隐藏状态可传递至下一时刻,并作为后续网络层的输入,实现对时间序列的建模与预测,其计算过程如下。

步骤 1 设输入的时序序列 $\mathbf{X}_t = [x_1, x_2, \dots, x_T] \in \mathbf{R}^{T \times v}$, T 为序列长度, v 为特征维度。遗忘门的计算式为

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (15)$$

式中: $\mathbf{f}_t$ 为 t 时刻遗忘门的激活向量; σ 为 Sigmoid 激活函数; $\mathbf{W}_f$ 为遗忘门的权重矩阵; $\mathbf{b}_f$ 为遗忘门的偏置向量; $\mathbf{h}_{t-1}$ 为上一时刻隐藏层状态输出; $\mathbf{x}_t$ 为当前时刻输入量; $[\cdot, \cdot]$ 表示两个向量作向量拼接,合并成 1 个新的联合输入向量。

步骤 2 输入门的计算式为

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (16)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \quad (17)$$

式中: $\mathbf{i}_t$ 为 t 时刻输入门的激活向量; $\mathbf{W}_i$ 和 $\mathbf{b}_i$ 分别为输入门权重矩阵与偏置向量; $\tilde{\mathbf{C}}_t$ 为 t 时刻的候选细胞状态; $\mathbf{W}_c$ 为候选细胞状态输入权重矩阵; $\mathbf{b}_c$ 为候选细胞状态输入偏置向量。

步骤 3 根据遗忘门激活值、输入门激活值和候选细胞状态更新当前细胞状态,计算式为

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \quad (18)$$

式中: $\mathbf{C}_t$ 为 t 时刻细胞状态; $\mathbf{C}_{t-1}$ 为 $t-1$ 时刻细胞状态; $\odot$ 为逐元素乘法运算。

步骤 4 输出门的计算式为

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad (19)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \quad (20)$$

式中: $\mathbf{o}_t$ 为 t 时刻输出门的激活向量; $\mathbf{W}_o$ 为输出门权重矩阵; $\mathbf{b}_o$ 为输出门偏置向量; $\mathbf{h}_t$ 为 t 时刻隐藏层状态输出。

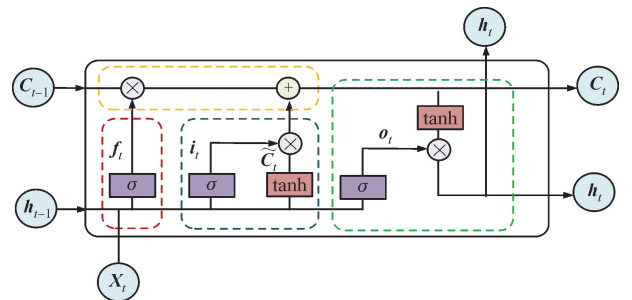


图 1 LSTM 网络结构

Fig. 1 LSTM structure diagram

2 ICEEMDAN-SSA-LSTM-CR 预测模型构建

鉴于光伏发电功率序列具有较强的复杂性和不确定性,本文首先采用 ICEEMDAN 对原始功率序列进行分解,获得不同频率尺度下的固有模态函数,以分离噪声成分、变化趋势及周期性信息,为后续建模提供更清晰的数据基础。然后,针对各 IMF 分量,利用 SSA 算法对 LSTM 模型关键参数进行优化,以提升模型对复杂时序特征的学习能力和预测精度。在此基础上,将 SSA 算法优化后的 LSTM 模型分别用于各 IMF 分量的建模与预测,并将各分量预测结果进行重构,得到预测结果。最后,引入级联回归对预测结果进行残差校正,以减小预测偏差并提高模型整体预测精度,模型预测流程如图 2 所示。

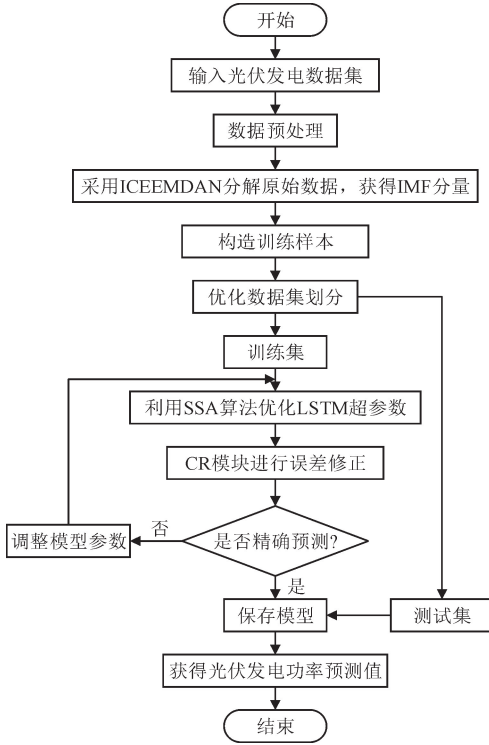


图2 ICEEMDAN-SSA-LSTM-CR 模型整体预测流程

Fig. 2 Model overall prediction process of the ICEEMDAN-SSA-LSTM-CR model

考虑到单一评价指标难以全面反映模型性能,为更全面地评价模型预测性能,本文采用平均绝对误差 e_{MAE} 、均方根误差 e_{RMSE} 、决定系数 R^2 、平均绝对百分比误差 e_{MAPE} 和平均偏差误差 e_{MBE} 来评价模型预测性能,其数学表达式如下

$$e_{MAE} = \frac{1}{\eta} \sum_{\beta=1}^{\eta} |y_{\beta} - \hat{y}_{\beta}| \quad (21)$$

$$e_{RMSE} = \sqrt{\frac{1}{n} \sum_{\beta=1}^{\eta} (y_{\beta} - \hat{y}_{\beta})^2} \quad (22)$$

$$R^2 = 1 - \frac{\sum_{\beta=1}^{\eta} (y_{\beta} - \hat{y}_{\beta})^2}{\sum_{\beta=1}^{\eta} (y_{\beta} - \bar{y})^2} \quad (23)$$

$$e_{MAPE} = \frac{1}{\eta} \sum_{\beta=1}^{\eta} \left| \frac{y_{\beta} - \hat{y}_{\beta}}{y_{\beta}} \right| \times 100\% \quad (24)$$

$$e_{MBE} = \frac{1}{\eta} \sum_{\beta=1}^{\eta} (y_{\beta} - \hat{y}_{\beta}) \quad (25)$$

式中: η 为测试样本数; y_{β} 为第 β 个样本的真实值; $\hat{y}_{\beta}$ 为相应的预测值; $\bar{y}$ 为真实值的平均值。

3 算例分析

3.1 样本数据预处理

本文选取某地区光伏发电厂 2024 年全年光伏出力数据作为研究对象,数据采样间隔为 15 min,全年共获得 $366 \times 96 = 35\,136$ 个原始数据点。由于夜间无太阳辐照,光伏发电功率通常接近 0,对模型构建影响有限。因此,首先对原始数据进行清洗,以剔除夜间无效样本。

本文结合研究区域四季昼夜时长差异,采用基于每日时间窗口的筛选方法。考虑该地区四季分明,不同季节日照时长及起止时间差异显著,因此本文分别设定春季、秋季数据的有效采样时段为 06:15—18:30,夏季为 05:00—20:30,冬季为 07:30—17:45。对于不在对应季节有效时间窗口内的样本,均视为无效数据并予以剔除。经清洗后,共保留 17 493 个有效数据点,并按春、夏、秋、冬四季进行分类。在完成季节分类后可以发现,各季节内部的出力变化趋势整体较为一致。因此,春、夏、秋、冬四季分别选取 4 月、7 月、10 月和 1 月作为典型月份样本,用于后续的分解、建模与预测分析。

为避免时间序列预测中的信息泄漏,归一化参数仅由训练集估计,并用于测试集验证;SSA 参数寻优和残差校正器训练均只使用训练集,训练集占 70%,测试集占 30%。ICEEMDAN 分解对训练集和测试集整体作混合处理。测试阶段在每个预测时刻仅利用该时刻之前可观测的历史序列进行分解和预测,测试样本不参与模型训练、参数寻优或残差校正器权重确定。

由于各数据特征的量纲和取值范围不同,为了便于后续模型训练,本文采用最小-最大归一化方法将数据映射到 $[0, 1]$ 区间,其计算式为

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (26)$$

式中: x' 为归一化后的数据; $x_{\max}$ 为数据中的最大值; $x_{\min}$ 为数据中的最小值。

图 3 展示了某光伏发电厂春季典型月份预处理

后的发电功率数据集与该数据集经 ICEEMDAN 分解所得结果。设定 ICEEMDAN 的噪声标准差 N_{std} 为 0.4; 噪声重叠数 N_R 设置为 100; 最大迭代次数设置为 4 000, 且信噪比调节功能已启用。经过 ICEEMDAN 分解后, 得到 9 组不同频率尺度的 IMF 分量。从图 3 中可以看出, 随着 IMF 序号的增加, 各分量的频率逐渐降低, 呈现出从高频到低频的变化趋势。

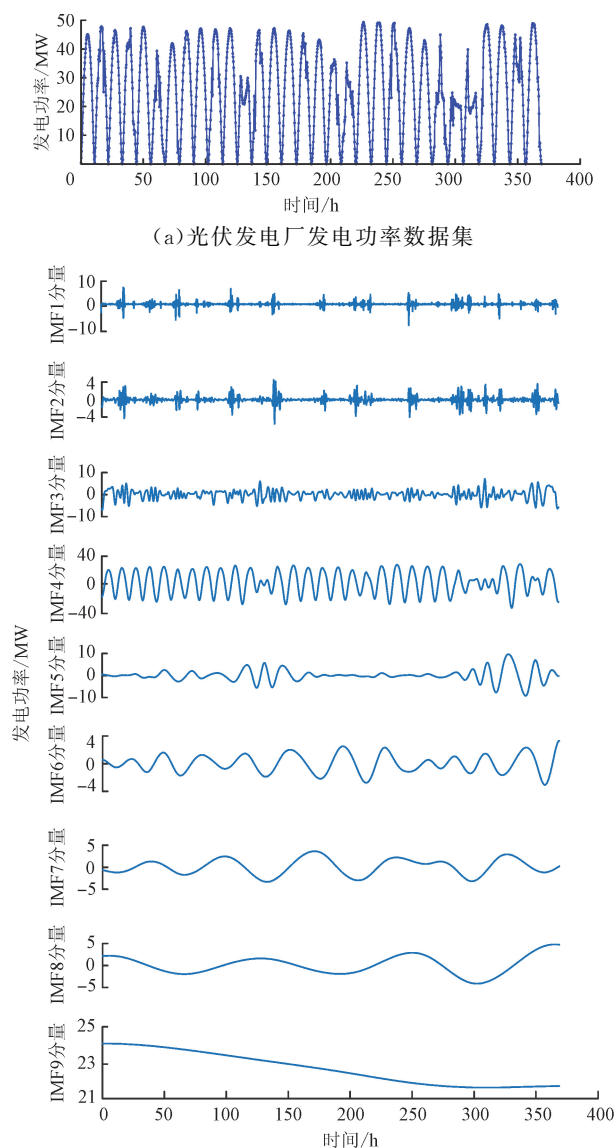


图 3 原始信号经 ICEEMDAN 分解得到的 9 个 IMF 分量
Fig. 3 ICEEMDAN-decomposed IMF components of the sample sequence

由图 3 可知, 高频 IMF 分量 (IMF1 至 IMF3) 主要包含高频成分, 表现为快速波动, 可能与天气变化、设备振动或测量噪声有关; 中频 IMF 分量 (IMF4 至 IMF6) 包含相对中等频率的变化, 具有一定周期性; 低频 IMF 分量 (IMF7 至 IMF9) 变化较

为缓慢, 反映了发电过程中的长期趋势或季节性因素。其中, 最后一个 IMF 分量已接近平稳, 几乎不再包含明显的周期成分。

总体来看, ICEEMDAN 能够有效地将不同时间尺度下的信息分离开来。高频分量帮助揭示信号中的快速扰动, 而低频和趋势部分则更适合后续预测建模。这种分解方式为光伏电站发电数据的深入分析提供了更清晰的结构基础^[22-23]。

3.2 SSA 优化

为提高模型性能并加速收敛, 本文采用麻雀搜索算法 SSA 对 LSTM 模型关键参数进行优化。其中, 预警值 S_T 设定为 0.6, 用于控制麻雀是否进行全局搜索, 较高的 S_T 有助于提高搜索多样性并避免陷入局部最优。生产者比例 P_D 设定为 0.9, 以增强种群的全局探索能力; 预警者比例 S_D 设定为 0.4, 以兼顾局部搜索效率和算法稳定性。种群规模设定为 20, 以在保证搜索能力的同时降低计算负担。边界控制机制用于保证麻雀个体不超出问题解空间; 惩罚因子的最小值 a_{min} 和最大值 a_{max} 分别设定为 500 和 6 000, 用于对越界个体进行适应度惩罚, 从而确保搜索结果始终位于合理范围内。

图 4 展示了 SSA 优化算法适应度随迭代次数的变化。预实验设置最大迭代次数为 50, 观察到当最大迭代次数超过 25 后, 适应度曲线趋于平稳。因此, 本文将 SSA 算法的最大迭代次数设为 30, 以在较少迭代次数下获得较优的优化结果^[24-25]。

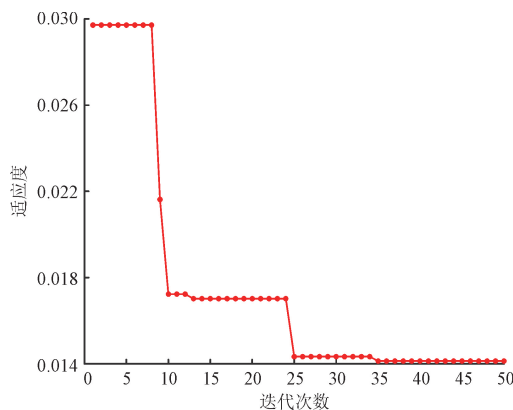


图 4 SSA 优化算法适应度

Fig. 4 Fitness in SSA optimization algorithm

3.3 ICEEMDAN-SSA-LSTM 预测输出结果

原始光伏发电功率序列经 ICEEMDAN 分解后得到多尺度 IMF 分量, 并据此构建 LSTM 预测模型。为进一步提升模型性能, 利用 SSA 对 LSTM 参数 (最大训练轮次、正则化参数 L2、学习率调整因子) 进行自适应优化, 以提高模型训练效率和预测精度。最终得到的 ICEEMDAN-SSA-LSTM 模型最优参数如表 1 所示。

表1 LSTM 预测模型参数

Table 1 LSTM prediction model parameters

参数	数值	参数	数值
隐藏层数	2	初始学习率	0.004
隐藏层神经元数	64	学习率衰减系数	85%
最大训练轮次	137	学习率调整因子	0.18
梯度阈值	1	正则化参数 L_2	0.001 4

图5给出了LSTM、ICEEMDAN-LSTM和ICEEMDAN-SSA-LSTM这3种模型的预测结果对比,表2列出了对应的评价指标。

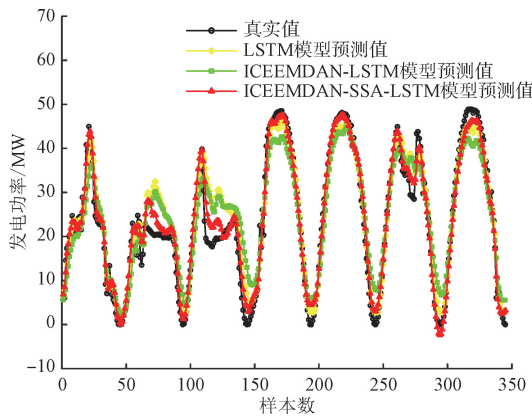


图5 3种模型预测结果

Fig. 5 Prediction results of three models

表2 3种模型的预测性能对比

Table 2 Comparison of forecasting performance among the three models

模型	R^2	$e_{MAE}/$ MW	$e_{MBE}/$ MW	$e_{MAPE}/$ %	$e_{RMSE}/$ MW
LSTM	0.874	3.517	1.488	10.116	4.752
ICEEMDAN-LSTM	0.868	3.125	1.479	9.709	4.475
ICEEMDAN-SSA-LSTM	0.970	1.584	0.859	4.688	2.123

从图5和表2可以看出,经ICEEMDAN分解和SSA算法参数优化后,LSTM模型对光伏发电数据的预测性能得到明显提升。与单一LSTM模型相比,ICEEMDAN-LSTM模型在误差指标上已有一定改善,而ICEEMDAN-SSA-LSTM模型进一步降低了 e_{MAE} 、 e_{MBE} 、 e_{MAPE} 和 e_{RMSE} 等误差指标,同时提高了决定系数,表明其对复杂非线性时序特征的表征能力更强,预测精度更高。综合比较3种模型,ICEEMDAN-SSA-LSTM模型表现出最优的预测效果,说明ICEEMDAN分解能够有效提取多尺度特征,SSA优化能够进一步提升LSTM模型参数配置的合理性,从而实现对光伏功率序列更准确的预测^[26]。

3.4 ICEEMDAN-SSA-LSTM-CR 模型预测输出结果

在构建的ICEEMDAN-SSA-LSTM-CR组合预测模型中,CR残差校正模块由卷积神经网络(CNN)、最小绝对收缩与选择算子(LASSO)、极限梯度提升(XGBoost)、K近邻(KNN)和支持向量回归(SVR)5个残差校正回归器并联组成。首先,ICEEMDAN-SSA-LSTM模型完成初步预测,并得到相应的预测残差;随后,将残差同时输入5个回归器中,由各回归器分别学习残差变化规律并输出对应的残差修正项,提升模型的整体预测性能。

为融合5个回归器的修正结果,本文采用5折交叉验证误差倒数的加权策略确定各回归器的权重。具体而言,在训练集上对每个回归器进行五折交叉验证,计算其在各验证折上的均方根误差 e_{RMSE} ,再将各自平均RMSE的倒数作为权重依据,并进行归一化处理,得到5个回归器的固定权重(权重之和为1),具体加权验证结果如表3所示。该方法能够使验证误差小、预测性能好的回归器在集成中占据更大比例,从而提高残差校正结果的可靠性。

表3 5折交叉验证下各回归器权重及集成 e_{RMSE} Table 3 Weights of individual regressors and ensemble e_{RMSE} under five-fold cross-validation

折数	5折交叉验证加权系数					集成 e_{RMSE}/MW
	CNN	LASSO	XGBoost	KNN	SVR	
1	0	0.260	0.649	0	0.091	1.468 5
2	0.119	0.206	0.654	0	0.020	1.285 3
3	0	0.125	0.669	0.067	0.139	1.214 1
4	0.282	0.119	0.470	0	0.129	1.024 9
5	0.066	0.168	0.370	0	0.396	0.970 2
最终 集成权重	0.093	0.175	0.562	0.013	0.155	1.192 6

在测试阶段,将5个回归器输出的残差校正值按照上述权重进行线性加权,得到最终残差校正项,再将最终残差校正项与ICEEMDAN-SSA-LSTM的初步预测结果相加,形成最终预测值。相比单一回归器校正方式,该集成策略能够更充分地利用不同算法之间的互补优势。同时,该方法无需进行复杂的迭代寻优,权重计算过程简单、清晰,具有较好的实用性和解释性。

为进一步验证所提方法的预测效果,本文引入SA、PSO和GA作为对照优化算法,并与本文采用的SSA算法进行比较,SA、PSO和GA均采用相同的LSTM参数优化方法,并对春、夏、秋、冬四季均进行预测,得到了图6展示的预测结果图与表4的各项评价指标数值。

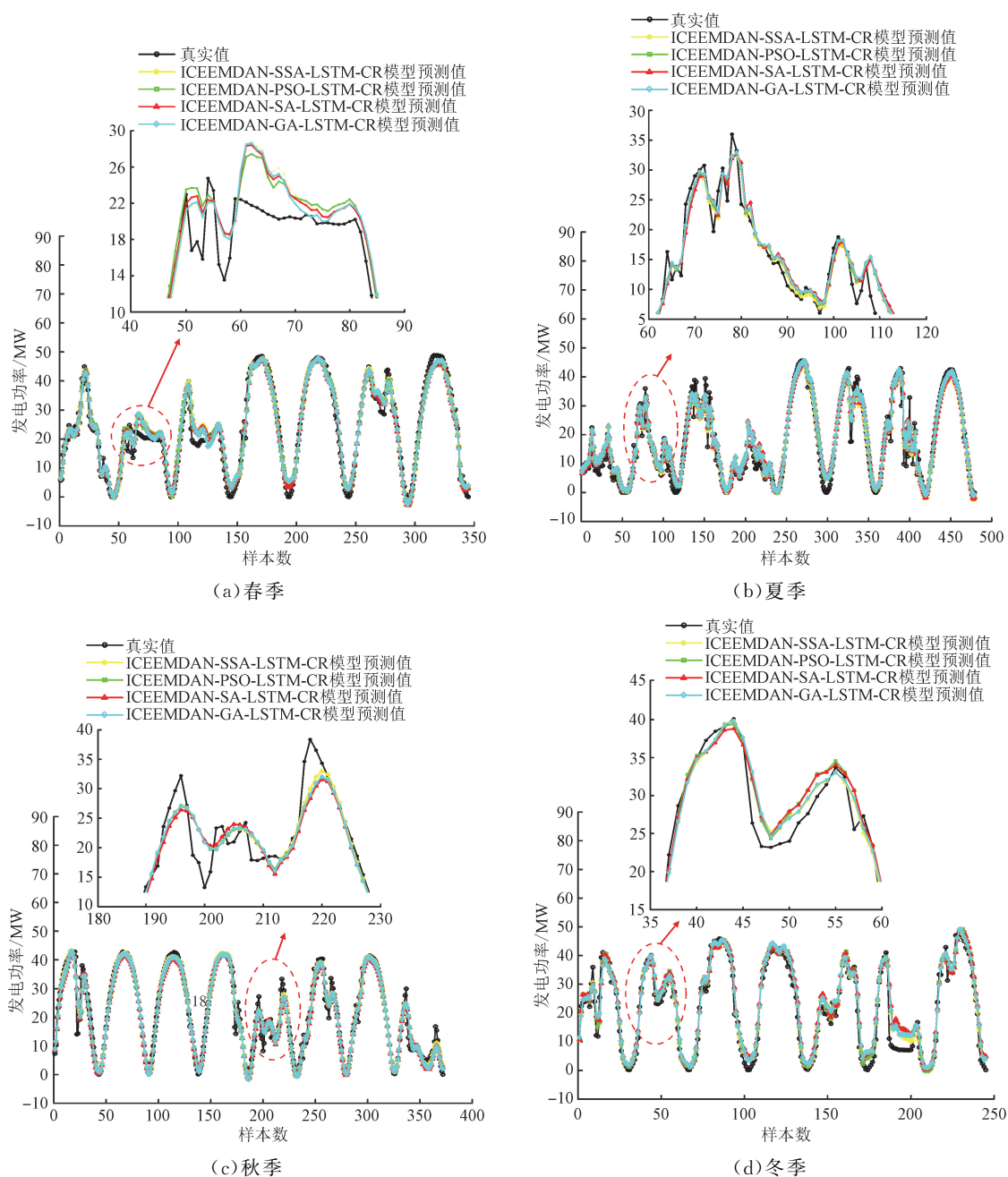


图 6 不同季节下 4 种算法预测效果对比

Fig. 6 Model prediction comparison of four algorithms across different seasons

表 4 不同季节下各模型的评价指标对比

Table 4 Comparison of performance metrics of the model in different seasons

季节	模型	R^2	e_{MAE}/MW	e_{MBE}/MW	$e_{MAPE}/\%$	e_{RMSE}/MW
春季	ICEEMDAN-SSA-LSTM-CR	0.989	1.111	0.153 1	2.124	1.370
	ICEEMDAN-GA-LSTM-CR	0.986	1.247	0.159 7	3.238	1.596
	ICEEMDAN-PSO-LSTM-CR	0.983	1.376	0.439 4	4.089	1.738
	ICEEMDAN-SA-LSTM-CR	0.988	1.347	0.341 7	3.234	1.536
	ICEEMDAN-SSA-LSTM	0.970	1.584	0.859 4	4.688	2.123
	ICEEMDAN-LSTM	0.868	3.125	1.479 7	9.709	4.475
	LSTM	0.874	3.517	1.488 4	10.116	4.752

续表

季节	模型	R^2	e_{MAE}/MW	e_{MBE}/MW	$e_{MAPE}/\%$	e_{RMSE}/MW
夏季	ICEEMDAN-SSA-LSTM-CR	0.975	1.128	0.364 1	4.638	1.615
	ICEEMDAN-GA-LSTM-CR	0.970	1.464	0.673 2	4.997	1.888
	ICEEMDAN-PSO-LSTM-CR	0.967	1.731	0.850 6	5.281	2.139
	ICEEMDAN-SA-LSTM-CR	0.966	1.702	0.562 9	5.158	1.948
	ICEEMDAN-SSA-LSTM	0.962	1.828	0.712 8	5.995	2.010
	ICEEMDAN-LSTM	0.822	4.844	1.571 7	14.451	4.763
	LSTM	0.896	3.487	-0.155 2	15.139	4.396
秋季	ICEEMDAN-SSA-LSTM-CR	0.976	1.091	0.381 4	3.949	1.741
	ICEEMDAN-GA-LSTM-CR	0.972	1.218	0.321 3	4.153	2.014
	ICEEMDAN-PSO-LSTM-CR	0.969	1.503	0.431 8	4.944	2.137
	ICEEMDAN-SA-LSTM-CR	0.963	1.835	0.812 8	5.144	2.632
	ICEEMDAN-SSA-LSTM	0.961	1.845	0.855 8	5.372	2.314
	ICEEMDAN-LSTM	0.882	3.633	-0.209 3	10.802	4.461
	LSTM	0.874	3.743	2.911 4	11.246	4.910
冬季	ICEEMDAN-SSA-LSTM-CR	0.969	1.264	0.253 4	4.154	1.981
	ICEEMDAN-GA-LSTM-CR	0.962	1.314	0.652 7	4.214	2.681
	ICEEMDAN-PSO-LSTM-CR	0.968	1.518	0.149 1	4.892	2.462
	ICEEMDAN-SA-LSTM-CR	0.967	1.629	0.390 1	4.398	2.509
	ICEEMDAN-SSA-LSTM	0.959	1.732	0.897 3	5.585	2.881
	ICEEMDAN-LSTM	0.818	4.792	3.086 0	10.973	6.387
	LSTM	0.797	4.698	3.092 7	11.811	6.837

从图 5、图 6 及表 4 可知,各模型在光伏发电功率预测中均取得了较好的预测效果,但本文提出的 ICEEMDAN-SSA-LSTM-CR 模型整体表现优于对照模型。在四季分明地区,不同季节下该模型的各项预测指标均有不同程度提升,且决定系数 R^2 均大于 0.96,说明该模型具有较好的拟合能力和预测性能。相比之下,冬季受极端天气、云层遮挡以及太阳辐射波动较大等因素影响,各模型的预测精度均出现了一定程度下降。

从表 4 可知,与 SA、PSO 和 GA 优化模型相比,本文所提模型在春、夏、秋、冬 4 个季节下的平均绝对误差分别降低了 16.05%、30.90%、28.16%和 15.00%;均方根误差分别降低了 15.61%、18.91%、23.00%和 22.33%,这表明所提模型具有较好的误差修正能力,能够有效提升预测性能。由于 CR 仅对模型最终预测结果进行残差校正,因此引入 CR 校正后的模型计算时间较引入前仅增加 4%,在保证预测精度提升的同时,未显著增加计算开销。

此外,所提模型在不同季节下均表现出较好的稳定性。与未引入 CR 的 ICEEMDAN-SSA-LSTM 模型相比,经过残差校正后,所提模型的最终预测值更加接近真实值,尤其在数据波动较大的季节中,优化效果更为明显。总体来看,CR 残差校正不仅提高了模型预测精度,还兼顾了计算效率,具有较好的工程应用价值。

4 结 论

针对光伏发电功率预测中序列波动性强、非线性明显及噪声干扰较大的问题,本文提出一种 ICEEMDAN-SSA-LSTM-CR 组合预测模型。首先,利用 ICEEMDAN 对原始功率序列进行分解,以削弱噪声干扰并提取多尺度特征;其次,采用 SSA 对 LSTM 关键参数进行优化,提升模型的时序建模能力;最后,引入 CR 对初步预测结果进行残差校正,从而进一步提高预测精度与稳定性。仿真结果有以下几项:

(1) SSA 算法能够在搜索过程中不断调整 LSTM 模型的关键参数,使参数配置更加合理,同时提升模型的全局搜索能力与收敛表现,从而提升对复杂非线性时序特征的特征能力;

(2) CR 在预测结果输出阶段对预测残差进行进一步校正,在未显著增加计算开销的条件下,有效降低了预测偏差,提高了模型整体预测精度;

(3) 与 SA、PSO、GA 等对照模型相比,所提模型在不同季节条件下均取得了较好的预测效果,在误差指标、预测稳定性及抗干扰能力等方面表现更优,表明该模型具有较好的适用性。

尽管如此,该方法在实际应用中仍存在一定局限。一方面,模型对气象数据质量和样本完整性较为敏感,当输入数据存在缺失、异常或气象预测偏差较大时,模型性能可能下降;另一方面,本文仅基于单一区域光伏电站数据开展验证,其在不同地理环境、气候条件及运行模式下的适用性仍有待进一步评估。后续研究可考虑引入多源气象信息、扩展跨区域数据样本,并结合在线更新机制,以提升模型的泛化能力和工程应用价值。

参考文献:

- [1] Li Zhengmao, Wu Lei, Xu Yan, et al. Distributed tri-layer risk-averse stochastic game approach for energy trading among multi-energy microgrids [J]. *Applied Energy*, 2023, 331: 120282.
- [2] Sun Wei, Wang Yuwei. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network [J]. *Energy Conversion and Management*, 2018, 157: 1-12.
- [3] 文劲宇, 周博, 魏利岫. 中国未来电力系统储电网初探 [J]. *电力系统保护与控制*, 2022, 50(7): 1-10.
Wen Jinyu, Zhou Bo, Wei Lishen. Preliminary study on an energy storage grid for future power system in China [J]. *Power System Protection and Control*, 2022, 50(7): 1-10.
- [4] 康重庆, 夏清, 张伯明. 电力系统负荷预测研究综述与发展方向的探讨 [J]. *电力系统自动化*, 2004, 28(17): 1-11.
Kang Zhongqing, Xia Qing, Zhang Boming. Review of power system load forecasting and its development [J]. *Automation of Electric Power Systems*, 2004, 28(17): 1-11.
- [5] Ramkumar G, Sahoo S, Amirthalakshmi T M, et al. A short-term solar photovoltaic power optimized prediction interval model based on FOS-ELM algorithm [J]. *International Journal of Photoenergy*, 2021, 2021(1): 3981456.
- [6] Pan Mingzhang, Li Chao, Gao Ran, et al. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization [J]. *Journal of Cleaner Production*, 2020, 277: 123948.
- [7] 马伟, 乔颖, 谢丽蓉, 等. 考虑气象特征波动划分阈值的双目标短期风功率预测 [J]. *高电压技术*, 2022, 48(10): 4154-4162.
Ma Wei, Qiao Ying, Xie Lirong, et al. Short-term wind power prediction with dual targets considering the threshold of meteorological characteristic fluctuation partition [J]. *High Voltage Engineering*, 2022, 48(10): 4154-4162.
- [8] 薛阳, 燕宇斌, 贾巍, 等. 基于改进灰狼算法优化长短期记忆网络的光伏功率预测 [J]. *太阳能学报*, 2023, 44(7): 207-213.
Xue Yang, Yan Yucheng, Jia Wei, et al. Photovoltaic power prediction model based on IGWO-LSTM [J]. *Acta Energeticae Solaris Sinica*, 2023, 44(7): 207-213.
- [9] 武小梅, 张琦, 田明正. 基于 VMD-SE 和优化支持向量机的光伏预测方法 [J]. *电力科学与工程*, 2017, 33(9): 29-36.
Wu Xiaomei, Zhang Qi, Tian Mingzheng. Photovoltaic prediction method based on VMD-SE and optimized support vector machine [J]. *Electric Power Science and Engineering*, 2017, 33(9): 29-36.
- [10] 郑少凯. 基于深度学习与前端信号处理的 MEMS 陀螺仪温漂补偿研究 [D]. 山东泰安: 山东农业大学, 2024.
- [11] 袁建华, 谢斌斌, 何宝林, 等. 基于 DTW-VMD-PSO-BP 的光伏发电功率短期预测方法 [J]. *太阳能学报*, 2022, 43(8): 58-66.
Yuan Jianhua, Xie Binbin, He Baolin, et al. Short term forecasting method of photovoltaic output based on DTW-VMD-PSO-BP [J]. *Acta Energeticae Solaris Sinica*, 2022, 43(8): 58-66.
- [12] 杨宁宁, 王怡昕, 吴朝俊, 等. 基于 SSA-VMD-LIESN 的短期风电功率预测方法研究 [J]. *太阳能学报*, 2025, 46(5): 440-447.
Yang Ningning, Wang Yixin, Wu Chaojun, et al. Research on short-term wind power prediction method based on SSA-VMD-LIESN [J]. *Acta Energeticae Solaris Sinica*, 2025, 46(5): 440-447.
- [13] 岳有军, 吴明沅, 王红君, 等. 基于 CNN-GRU-ISSA-XGBoost 的短期光伏功率预测 [J]. *南京信息工程大*

- 学学报, 2024, 16(2): 231-238.
- Yue Youjun, Wu Mingyuan, Wang Hongjun, et al. Short term photovoltaic power prediction based on CNN-GRU-ISSA-XGBoost [J]. Journal of Nanjing University of Information Science & Technology, 2024, 16(2): 231-238.
- [14] 葛泉波, 李凯, 张兴国. 基于多关键点检测加权融合的无人机相对位姿估计算法 [J]. 自动化学报, 2024, 50(7): 1402-1416.
- Ge Quanbo, Li Kai, Zhang Xingguo. Relative pose estimation algorithm for unmanned aerial vehicles based on weighted fusion of multiple keypoint detection [J]. Acta Automatica Sinica, 2024, 50(7): 1402-1416.
- [15] 代少升, 熊昆, 吴云铎, 等. 多视角约束级联回归的视频人脸特征点跟踪 [J]. 计算机应用, 2022, 42(8): 2415-2422.
- Dai Shaosheng, Xiong Kun, Wu Yunduo, et al. Video facial landmark tracking by multi-view constrained cascade regression [J]. Journal of Computer Applications, 2022, 42(8): 2415-2422.
- [16] 曹瑞钧, 郭其一. 基于 ICEEMDAN-FE 和支持向量机的三电平逆变器的故障诊断技术 [J]. 机车电传动, 2023(1): 97-103.
- Cao Ruijun, Guo Qiyi. Fault diagnosis technology for three-level inverter based on ICEEMDAN-FE and SVM [J]. Electric Drive for Locomotives, 2023(1): 97-103.
- [17] 杨洋, 郭兴明, 郑伊能, 等. 基于 ICEEMDAN-MSE 的左室舒张功能障碍心音信号的识别研究 [J]. 仪器仪表学报, 2022, 43(1): 274-281.
- Yang Yang, Guo Xingming, Zheng Yineng, et al. Study on left ventricular diastolic dysfunction heart sound signals identification based on ICEEMDAN-MSE [J]. Chinese Journal of Scientific Instrument, 2022, 43(1): 274-281.
- [18] Xue Jiankai, Shen Bo. A novel swarm intelligence optimization approach: sparrow search algorithm [J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.
- [19] 任俊, 魏霞, 黄德启, 等. 基于眼睛状态多特征融合的疲劳驾驶检测 [J]. 计算机工程与设计, 2022, 43(11): 3187-3194.
- Ren Jun, Wei Xia, Huang Deqi, et al. Fatigue driving detection based on multi-feature fusion of eye state [J]. Computer Engineering and Design, 2022, 43(11): 3187-3194.
- [20] 石启萌. 融合深度学习与级联回归的头影测量分析系统 [D]. 天津: 天津大学, 2021.
- [21] 陈庆明, 廖鸿飞, 孙颖楷, 等. 改进的蜣螂优化算法及光伏发电功率预测应用 [J]. 太阳能学报, 2025, 46(9): 445-454.
- Chen Qingming, Liao Hongfei, Sun Yingkai, et al. Improved dung beetle optimization algorithm and application to photovoltaic power generation prediction [J]. Acta Energetica Solaris Sinica, 2025, 46(9): 445-454.
- [22] 徐昌, 许野, 王晓晖, 等. 基于 KPCA 和 HC 的 IPSO-LSTM 光伏出力预测模型研究 [J]. 太阳能学报, 2025, 46(5): 362-374.
- Xu Chang, Xu Ye, Wang Xiaohui, et al. Research on combined photovoltaic output prediction model based on KPCA, HC and IPSO-LSTM [J]. Acta Energetica Solaris Sinica, 2025, 46(5): 362-374.
- [23] 姚奇才, 向文国, 陈时熠, 等. 基于 ICEEMDAN-KPCA-ICPA-LSTM 的光伏发电功率预测 [J]. 动力工程学报, 2025, 45(3): 374-382.
- Yao Qincai, Xiang Wenguo, Chen Shiyi, et al. Photovoltaic power forecasting based on ICEEMDAN-KPCA-ICPA-LSTM [J]. Journal of Chinese Society of Power Engineering, 2025, 45(3): 374-382.
- [24] Wang Tianlei, Gu Shaowei, Liu Renju, et al. Cuckoo catfish optimizer: a new meta-heuristic optimization algorithm [J]. Artificial Intelligence Review, 2025, 58(10): 326.
- [25] Dragomiretskiy K, Zosso D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [26] 高岩, 吴汉斌, 张纪欣, 等. 基于组合深度学习的光伏功率日前概率预测模型 [J]. 中国电力, 2024, 57(4): 100-110.
- Gao Yan, Wu Hanbin, Zhang Jixin, et al. Day-ahead probabilistic prediction model for photovoltaic power based on combined deep learning [J]. Electric Power, 2024, 57(4): 100-110.

(编辑 武红江)