

DOI: 10.19783/j.cnki.pspc.241315

小样本条件下风电功率预测方法的研究

张志艳¹, 兰龙¹, 白景升¹, 杨唐毅格², 孔汉¹, 刘华³

(1. 郑州轻工业大学电气信息工程学院, 河南 郑州 450002; 2. 中广核新能源巢湖观湖风电场, 安徽 巢湖 238007;
3. 河南省电力公司建设部, 河南 郑州 450007)

摘要: 针对风电功率预测通常需要大量历史数据和复杂模型支持的问题, 提出了一种小样本条件下风电功率预测模型。首先, 针对样本不足问题, 提出基于时序马尔可夫链蒙特卡洛(time-Markov chain Monte Carlo, Time-MCMC)的数据扩充方法, 采用统计、插值、拟合和自适应识别方法进行数据清洗, 提升样本的多样性和质量。其次, 融合多种优化策略, 构建基于改进型混合天鹰优化与非洲秃鹫优化算法(improved hybrid aquila optimization and African vulture optimization algorithm, IHAOAVOA)的反向传播(backpropagation, BP)神经网络风电功率预测模型。最后, 以实际算例验证了数据扩充方法的有效性, 同时对扩充后的样本集分别采用BP神经网络、天鹰优化器(aquila optimizer, AO)优化BP神经网络、非洲秃鹫优化算法(African vultures optimization algorithm, AVOA)优化BP神经网络和IHAOAVOA-BP神经网络4种模型进行功率预测。预测结果表明, 与AVOA-BP模型相比, 所提模型平均绝对值误差(mean absolute error, MAE)和均方误差(mean squared error, MSE)分别至少降低了0.45 MW和21.48%。

关键词: 功率预测; 小样本; BP神经网络; 马尔可夫链蒙特卡洛; 混合优化策略

Research on wind power forecasting methods under small-sample conditions

ZHANG Zhiyan¹, LAN Long¹, BAI Jingsheng¹, YANG Tangyige², KONG Han¹, LIU Hua³

(1. College of Electrical and Information Engineering, Zhengzhou Light Industry University, Zhengzhou 450002, China;
2. Chaohu Guanhu Wind Farm, China General Nuclear New Energy, Chaohu 238007, China;
3. Construction Department of Henan Electric Power Company, Zhengzhou 450007, China)

Abstract: To address the issue that wind power forecasting typically requires substantial historical data and support from complex models, a wind power forecasting model under small-sample conditions is proposed. First, aiming at the problem of insufficient samples, a data augmentation method based on time-Markov chain Monte Carlo (Time-MCMC) is developed. Statistical analysis, interpolation, fitting, and adaptive recognition methods are employed for data cleaning to improve the diversity and quality of the samples. Second, by integrating multiple optimization strategies, a wind power forecasting model is constructed, where the backpropagation (BP) neural network is optimized using the improved hybrid aquila optimization and African vulture optimization algorithm (IHAOAVOA). Finally, the effectiveness of the data augmentation method is verified using actual case studies. Meanwhile, power forecasting is conducted on the augmented sample set using four models: the BP neural network, the BP neural network optimized by the aquila optimizer (AO), the BP neural network optimized by the African vulture optimization algorithm (AVOA), and the IHAOAVOA-BP neural network. The forecasting results show that, compared with the AVOA-BP model, the mean absolute error (MAE) and mean squared error (MSE) of the proposed model are reduced by at least 0.45 MW and 21.48% respectively.

This work is supported by the Science and Technology Key Project of Henan Province (No. 242102241030).

Key words: power forecasting; small sample; BP neural network; Markov chain Monte Carlo; hybrid optimization strategy

0 引言

风能因具有清洁环保、资源丰富且可再生的特

点, 日益受到广泛关注。然而风电机组受气候、地形、设备状态等多种因素影响, 输出功率存在较大的波动性和不确定性^[1-3], 给电网稳定运行带来了极大挑战, 因此如何精确预测风电功率已成为加快风电消纳研究的关键问题。近年来, 风电功率预测技

基金项目: 河南省科技攻关项目资助(242102241030)

术取得了显著进步,但多数预测方法依赖于大量的历史数据和复杂的物理模型,在一定程度上影响了预测精度和实时性。同时,实际应用中由于风电场建设时间短或数据收集不完善等原因,难以获得足够的历史数据。小样本条件下,传统的预测技术常常达不到预期效果,难以满足实际预测需求,促使针对小样本条件下风电功率预测模型的研究成为当前风电领域的研究热点。

目前,风电功率预测方法主要分为统计学方法、物理模型方法和人工智能方法^[4-6]。文献[7]利用回归和时间序列法对风速和功率数据进行分析,建立了基于统计学的风电功率预测模型。文献[8]依据历史数据建立了自回归移动平均模型和广义自回归条件异方差模型,消除了风电功率预测误差的条件异方差特性。同时针对预测误差尖峰轻尾的统计特性,采用改进广义误差分布模型对风电功率预测误差进行分层,并对处于不同情况下的预测误差给予对应的补偿方案,进而提高预测精度。文献[7-10]利用历史风电功率数据,采用基于统计学的预测方法,实现了较为准确的风电功率预测。然而,在处理复杂的非线性关系和面临历史风电样本较少时,依赖统计学方法进行预测可能无法达到预期效果。文献[11]基于风速与功率的物理模型,分析了风速波动特征,提出了一种引入风速信息的切换输出机制,削弱了风能波动性对功率预测的影响。文献[12]基于风电场尾流效应和功率曲线两种物理模型构建了物理深度学习网络,并利用物理模型产生预训练样本,进一步提升预测精度。文献[11-13]提出的基于物理模型的预测方法,对气象和风电数据依赖性较强,在小样本条件下应用受限。

随着人工智能技术在样本学习生成领域的快速发展,神经网络、深度学习等技术在风电功率预测领域得到了较为广泛的应用,在一定程度上为解决小样本条件下功率预测建模困难问题提供了更多的思路。文献[14]基于迁移卷积时空网络提出了样本数据迁移学习方法,实现了少样本条件下的数据特征迁移学习。文献[15]实施了随机欠采样与随机过采样、随机欠采样与合成少数过采样以及随机欠采样与误差自举过采样3种数据混合抽样策略,解决了类别不平衡问题。文献[16]采用双层生成器、双判别器循环生成对抗网络,丰富了故障标签样本,提升了机组故障预警的时效性和故障类型辨识准确度。文献[17]使用卷积神经网络构建生成模型和判别模型,以影响因素为条件生成预测数据,通过影响因素生成对抗网络的博弈训练。文献[14-17]应用

基于人工智能的小样本技术,在一定程度上缓解了样本量稀缺带来的难题。小样本条件下风电功率预测面临的突出问题是样本量不足,对抗生成网络(generative adversarial network, GAN)现有研究在生成模型和数据扩充方面已取得了一定成果,但对抗生成网络在训练过程中需要大量的样本数据优化生成器和判别器。如果训练样本较少,GAN无法有效学习数据分布,且上述文献未深入考虑数据的概率分布特性和时序特性,而生成样本的概率分布与真实样本的分布差异是生成数据质量的重要判断标准,据此提出基于时序马尔可夫链蒙特卡洛(time-Markov chain Monte Carlo, Time-MCMC)的数据扩充方法,为预测模型提供良好的数据基础,解决预测模型建模困难问题。

BP神经网络技术为复杂非线性问题建模提供了可行路径^[18-19],但其自身存在权值调整易陷入局部最优、收敛速度缓慢的固有缺陷,需依赖优化算法提升其性能^[20]。然而,单一优化算法在改进BP神经网络时存在一定局限,天鹰优化器(aquila optimizer, AO)算法虽具备较强的全局探索能力,能够在较广的解空间内搜索潜在最优解,但在局部开发阶段稳定性不足,难以精准捕捉变量间细微的非线性关联,导致对局部特征的挖掘精度受限^[21]。非洲秃鹫优化算法(African vultures optimization algorithm, AVOA)擅长局部精细搜索,可对局部解空间进行深度挖掘,但全局探索范围有限^[22]。上述单一优化算法改进的BP模型,因无法有效平衡全局探索与局部开发的关系,在处理复杂非线性系统建模时,对变量间深层关联和数据整体分布特征的捕捉精度受限^[23]。基于此,本文提出基于改进型混合天鹰优化与非洲秃鹫优化算法(improved hybrid aquila optimization and African vulture optimization algorithm, IHAOAVOA)改进的BP神经网络预测模型,通过结合不同算法的优点与策略增强,提升预测模型的性能。

为解决风电功率预测因样本量不足而难以精准预测的问题,提出一种小样本条件下的风电功率预测方法。为丰富风电功率历史样本,提高预测模型的泛化能力和稳定性,利用Time-MCMC方法进行数据扩充,采用统计、插值、拟合和自适应识别4种方法对扩充后数据集进行清洗。在此基础上,构建IHAOAVOA-BP预测模型,以实际风电场数据为算例样本,分别采用BP、AO-BP、AVOA-BP和IHAOAVOA-BP4种模型进行预测,对所提方法的有效性进行验证。

1 样本数据预处理

1.1 样本扩充

风电场建设时间短或数据收集不完整可能导致样本不足,难以依赖神经网络建立预测模型。因此,提出基于 Time-MCMC 样本扩充技术生成多样化样本,该方法在 MCMC 基础上同时考虑了风电功率数据之间的依赖关系,使扩充后的数据保留真实数据的时序特性。

MCMC 是一种基于概率统计的数值模拟方法,将随机过程中的马尔可夫链引入蒙特卡洛模拟,通过动态调整抽样分布,实现更加高效的模拟^[24-25]。基于 MCMC 算法扩充后的功率数据处于随机排序状态,为保留真实数据的时序特性,在 MCMC 方法的基础上构造时序特性提取函数和时序特性赋加函数。

时序特性提取函数为

$$H_a = \text{randi}(n_{\text{um}}) \quad (1)$$

式中: H_a 表示第 a 次迭代抽样数据的时间点; randi 为提取函数; n_{um} 表示原始数据的个数。

时序特性赋加函数为

$$L = [H_a, P_{\text{new}}(a)] \quad (2)$$

$$[R_z, i_{\text{dx}}] = \text{sortrows}(L, 1) \quad (3)$$

式中: L 为赋加时间点的数据集; $P_{\text{new}}(a)$ 为第 a 次迭代生成的样本; R_z 为经过时序排序后的数据集,即具有真实数据时序特性的数据集; i_{dx} 为排序的索引; sortrows 为时序排序函数。

基于 Time-MCMC 扩充技术生成样本的具体过程如下。

1) 设定要扩充的风电功率样本数量 M 和目标分布。

2) 遍历原始风电功率数据集,利用式(1)将数据的时序特性提取,并随机选择一个样本作为初始状态,根据数据集初步确定调整参数。

3) 根据试验确定最终的调整参数,并通过式(4)生成风电功率新数据。

$$P_{\text{new}} = P + \text{randn}(1, h + 1) \times \delta \quad (4)$$

式中: P_{new} 为生成的风电功率样本; P 为原始风电功率样本,包含风速特征; randn 为生成标准正态分布随机数的函数; h 为样本特征数; δ 为调整参数。

4) 根据生成功率数据与原始功率数据的差异,利用式(5)计算接受率 α 。同时生成一个均匀分布的随机数 $u \sim U(0, 1)$, 如果 $u > \alpha$, 则接受候选样本; 否则,拒绝候选样本,保持当前样本不变。

$$\alpha = \min(1, |P_{\text{new}} - P|) \quad (5)$$

5) 确定生成的新状态后,利用式(2)将所提取的时序索引赋加在生成的新状态前。

6) 重复步骤 3)~步骤 5), 迭代抽样后,为保证生成功率数据收敛,舍弃生成数据的初始部分,得到风电功率生成样本集。

7) 对扩充后的功率数据利用式(3)进行时序排序处理,形成具有原始时序特性的扩充数据集。

1.2 数据清洗

风电场原始数据受气象、设备故障及电网限制等影响,存在异常数据,直接扩充样本会导致异常数据保留。为确保数据的准确性,结合统计、插值、拟合和自适应识别方法处理样本集,提高数据分析和可信度。

对于扩充后的风电功率数据集,使用统计方法删除风电特征数据和功率数据相同的异常值,采用前向填充法处理缺失值,并通过滑动线性插值平滑每组风电数据,去除离群值。

采用随机抽样一致性拟合算法(random sample consensus, RANSAC)对数据集中风电特征与功率值相关性差的数据进行清洗。设置抽样个数及内点到模型最大距离参数,提取内点,得到相关性较高的风电样本。

采用孤立森林算法(isolation forest, IF)对得到的内点数据与 RANSAC 模型理论值的残差进行分析,设置异常比例,将识别出的异常值用 RANSAC 模型值替换。

2 风电功率预测模型的提出

2.1 AO 算法

AO 算法模拟了天鹰在捕猎过程中的行为^[26-27],通过分阶段行为模拟,能够在复杂的搜索空间中有效地平衡探索与开发,提高找到全局最优解的概率。

首先,在给定问题的搜索空间中利用式(6)初始化种群总体。

$$x_i = r_{\text{and}} \times (u_b - l_b) + l_b, i = 1, 2, 3, \dots, N' \quad (6)$$

式中: x_i 表示种群中第 i 个天鹰个体的位置; r_{and} 表示 $[0, 1]$ 内的随机数; u_b 和 l_b 分别表示搜索域的上限和下限; N' 表示天鹰个体的总数。

该算法的探索阶段能够有效进行全局搜索,在风电功率预测中,该阶段能遍历优化 BP 预测模型中对应风速特征的权重参数,提高对不同风速特征与功率关联规律的探索,该阶段的更新规则如下。

1) 扩展探索

$$x_i(t+1) = x_b(t) \times \left(1 - \frac{t}{T}\right) + x_o(t) - x_b(t) \times r_{\text{and}} \quad (7)$$

$$x_o(t) = \frac{1}{N'} \sum_{i=1}^{N'} x_i(t) \quad (8)$$

式中: $x_i(t+1)$ 为天鹰下一次迭代位置; $x_o(t)$ 为天鹰当前找到的最优解; t 和 T 分别为 AO 算法当前迭代次数和最大迭代次数; $x_o(t)$ 为种群中所有天鹰的平均位置。

2) 狭窄探索

$$x_i(t+1) = x_o(t) \times \text{Levy}(D) + x_r(t) \times (\omega - \beta) \times r_{\text{and}} \quad (9)$$

$$\text{Levy}(D) = 0.01 \frac{\mu \times \sigma}{|v|^{\frac{1}{3}}} \quad (10)$$

$$\sigma = \left(\frac{\tau(2.5) \times \sin\left(\frac{3\pi}{4}\right)}{\tau(2.5) \times 1.5 \times 2^{\frac{1}{4}}} \right)^{\frac{2}{3}} \quad (11)$$

$$\beta = (r + 0.00565 \times D_1) \times \sin\left(-0.005D_1 + \frac{3\pi}{2}\right) \quad (12)$$

$$\omega = (r + 0.00565 \times D_1) \times \cos\left(-0.005D_1 + \frac{3\pi}{2}\right) \quad (13)$$

式中: Levy 为莱维飞行函数; D 为问题维度; $x_r(t)$ 为从种群中随机选择的一个天鹰位置; μ 和 v 为两个不同的 $[0, 1]$ 内的随机数; τ 为伽马函数; r 为搜索周期数; D_1 为 $1 \sim D$ 的整数。

2.2 AVOA 算法

AVOA 算法通过模拟秃鹫在自然界中的领导者和追随者模式, 根据饥饿水平 F 进行全局探索和局部搜索, 寻找问题最优解^[28]。

算法的开发阶段能够进行精细的局部探索, 该局部开发特性与 BP 模型对风电功率的拟合需求高度适配, 利用位置更新规则微调参数, 使 BP 模型能更敏锐地识别风速微小变化引发功率跳变的局部特征, 位置更新规则如下。

在每个内部阶段, 每个机制的选择与否由参数 P_2 和 P_3 决定。参数 P_2 和 P_3 分别用于选择第一阶段和第二阶段中可用的机制, 两个参数的取值范围均为 $[0, 1]$ 。

当 $0.5 \leq |F| < 1$ 且 $r_{p_2} \leq P_2$ 时

$$X_j(n+1) = D_j(n) \times (F \times R_{\text{and1}}) - d_j(n) \quad (14)$$

$$D_j(n) = |C \times X_B(n) - X_j(n)| \quad (15)$$

$$d_j(n) = X_B(n) - X_j(n) \quad (16)$$

式中: r_{p_2} 、 R_{and1} 为不同的 $[0, 1]$ 内的随机数; $X_j(n+1)$ 为秃鹫个体下一次迭代位置; $X_j(n)$ 为秃鹫个体当前的位置; $X_B(n)$ 为秃鹫当前找到的最优解; n 为

AVOA 算法当前迭代次数; C 为 $[0, 2]$ 内的随机值。

当 $0.5 \leq |F| < 1$ 且 $r_{p_2} > P_2$ 时

$$X_j(n+1) = X_B(n) - (S_1(n) + S_2(n)) \quad (17)$$

$$S_1(n) = X_B(n) \times \left(\frac{R_{\text{and2}} \times X_j(n)}{2\pi} \right) \times \cos(X_j(n)) \quad (18)$$

$$S_2(n) = X_B(n) \times \left(\frac{R_{\text{and3}} \times X_j(n)}{2\pi} \right) \times \sin(X_j(n)) \quad (19)$$

式中: R_{and2} 、 R_{and3} 为不同的 $[0, 1]$ 内的随机数。

当 $|F| < 0.5$ 且 $r_{p_3} \leq P_3$ 时

$$X_j(n+1) = \frac{A_1(n) + A_2(n)}{2} \quad (20)$$

$$A_1(n) = B_{v_1}(n) - \frac{B_{v_1}(n) \times X_j(n)}{B_{v_1}(n) - X_j(n)^2} \times F \quad (21)$$

$$A_2(n) = B_{v_2}(n) - \frac{B_{v_2}(n) \times X_j(n)}{B_{v_2}(n) - X_j(n)^2} \times F \quad (22)$$

式中: r_{p_3} 为 $[0, 1]$ 内的随机数; $B_{v_1}(n)$ 和 $B_{v_2}(n)$ 分别为当前迭代第一组和第二组中的最优秃鹫。

当 $|F| < 0.5$ 且 $r_{p_3} > P_3$ 时

$$X_j(n+1) = X_B(n) - |d_j(n)| \times F \times \text{Levy}(D) \quad (23)$$

2.3 IHAAVOA 算法

AO 算法具有强大的勘探能力, 但其局部开发阶段不够稳定^[29], AVOA 算法具有良好的开发能力, 但勘探机制不足^[30-31]。为解决单一优化算法的不足, 以 AVOA 算法为核心框架, 将 AO 算法和 AVOA 算法相结合, 并对混合算法进行改进, 得到 IHAAVOA 算法。

1) 在探索阶段引入 AO 算法的拓展机制与狭窄探索机制, 取代 AVOA 算法原来的位置更新规则, 如式(7)一式(13)所示。

2) 为增加种群的多样性, 帮助混合算法避免局部最优解, 通过整合基于镜头对立的学习(Lens opposition-based learning, LOBL)策略和基于随机对立的学习(random opposition-based learning, ROBL)策略, 设计基于复合对立的学习(composite opposition-based learning, COBL)策略。

LOBL 策略和 ROBL 策略提高了避免陷入局部最优的能力, 同时 LOBL 策略也提高了算法收敛速度, ROBL 策略丰富了种群多样性。为充分利用两种策略的特性, 在优化过程中设置每个策略被选择的概率均为 50%, 得到 COBL 策略的表达式如式(24)和式(25)所示。

当 $q < 0.5$ 时, 有

$$\tilde{X}_g(m) = l_b^m + u_b^m - r_{\text{and4}} \times X'_g(m) \quad (24)$$

当 $q \geq 0.5$ 时, 有

$$\tilde{X}_g(m) = \frac{(l_b^m + u_b^m)}{2} + \frac{(l_b^m + u_b^m)}{2\eta} - \frac{X'_g(m)}{\eta} \quad (25)$$

式中: q 为 $[0,1]$ 内的随机值; $X'_g(m)$ 为第 g 个解 X'_g 的第 m 维分量; $\tilde{X}_g(m)$ 为解 $X'_g(m)$ 的对立解; l_b^m 和 u_b^m 分别为第 m 维搜索空间的下限和上限; r_{and4} 为 $[0,1]$ 内的随机值; η 为距离系数。

3) 为提高搜索效率并平衡混合算法的探索与利用, 引入了适应度-距离平衡 (fitness-distance balance, FDB) 选择策略^[32]。此策略综合考虑候选解的适应度值及与最优解的距离, 更新种群时, 既选适应度高的解, 也考虑距离远的解, 避免算法过早收敛于局部最优解, 保持种群在解空间的广泛分布。

FDB 策略通过调整核心位置更新式(9), 用候选解 $x_{\text{FDB}}(t)$ 替换 $x_r(t)$, 如式(26)所示, 实现位置更新,

确保每次更新都能朝着更优解的方向进行。

$$x_i(t+1) = x_b(t) \times \text{Levy}(D) + x_{\text{FDB}}(t) + (\omega - \beta) \times r_{\text{and}} \quad (26)$$

式中: $x_{\text{FDB}}(t)$ 为经过 FDB 策略识别出的候选解。

2.4 预测模型的提出

BP 神经网络算法具备处理复杂非线性关系的能力, 已在风电功率预测领域广泛应用^[33-34], 然而其权值通过沿局部改善方向逐渐调整, 使得局部搜索能力大于全局搜索能力, 易陷入局部最优, 且学习过程较长, 收敛速度较慢, 导致预测精度下降。因此, 利用 BP 神经网络预测风电功率, 需对其模型进行优化。利用改进后的混合优化算法 IHAOAVOA 对 BP 预测模型中映射风速与功率之间关系的权重和偏置进行优化, 提高 BP 神经网络的风电功率预测水平, 增强全局和局部搜索能力, 优化流程如图 1 所示, 具体优化步骤如下。

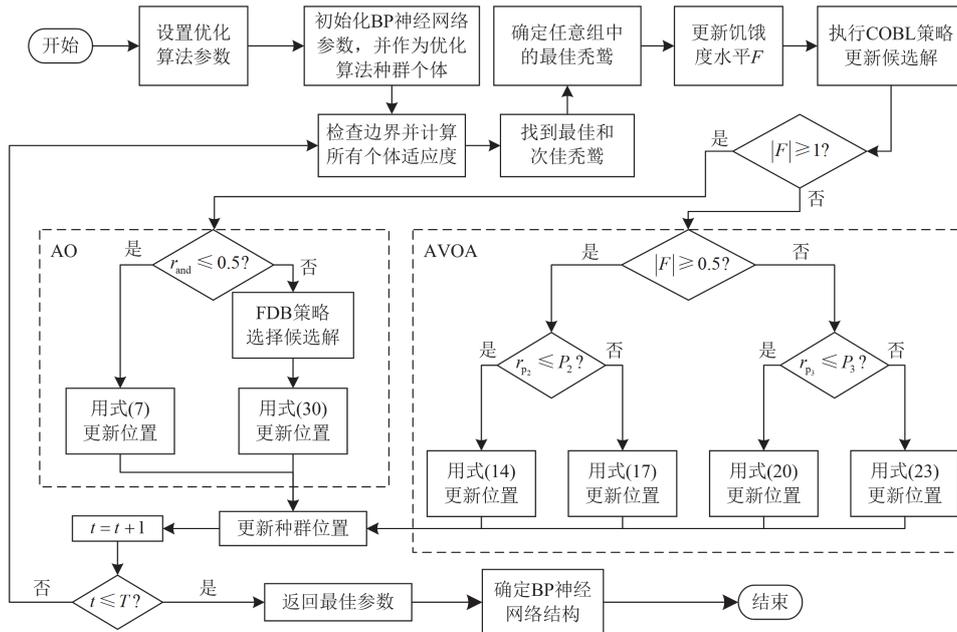


图 1 混合优化算法优化 BP 神经网络

Fig. 1 BP neural network optimized by hybrid optimization algorithms

步骤 1 初始化 BP 神经网络参数, 并作为优化算法种群个体, 计算种群个体适应度。

步骤 2 确定最佳秃鹫, 更新饥饿度水平, 并执行 COBL 策略更新候选解。

步骤 3 根据饥饿度水平选择进入 AO 探索阶段或 AVOA 开发阶段, 分别利用相应的位置更新公式进行更新。

步骤 4 判断当前迭代次数是否达到最大迭代次数 T , 若达到则输出最佳解, 确定优化后的网络结

构, 否则重复执行步骤 1—步骤 3。

3 风电功率预测模型的建立

基于 IHAOAVOA-BP 神经网络模型的风电功率预测流程如图 2 所示, 具体步骤如下。

1) 将扩充后的风电功率样本集划分为训练集和测试集, 并进行归一化处理。

2) 构建 BP 神经网络结构, 设置 BP 神经网络超参数。

3) 初始化 BP 神经网络权值和偏置, 作为优化算法的种群个体。

4) 设置混合优化算法参数, 包括种群大小、个体数量和迭代次数。

5) 计算种群个体适应度, 确定全局最优适应度值和位置。

6) 利用改进混合算法对种群个体位置进行更新, 确保向最优位置靠近。

7) 重新计算适应度值, 当满足精度要求或达到迭代次数后停止迭代, 否则返回步骤 5), 继续迭代。

8) 将全局最优个体解码作为网络参数, 对风电功率数据集进行训练和测试, 得到风电功率预测结果。

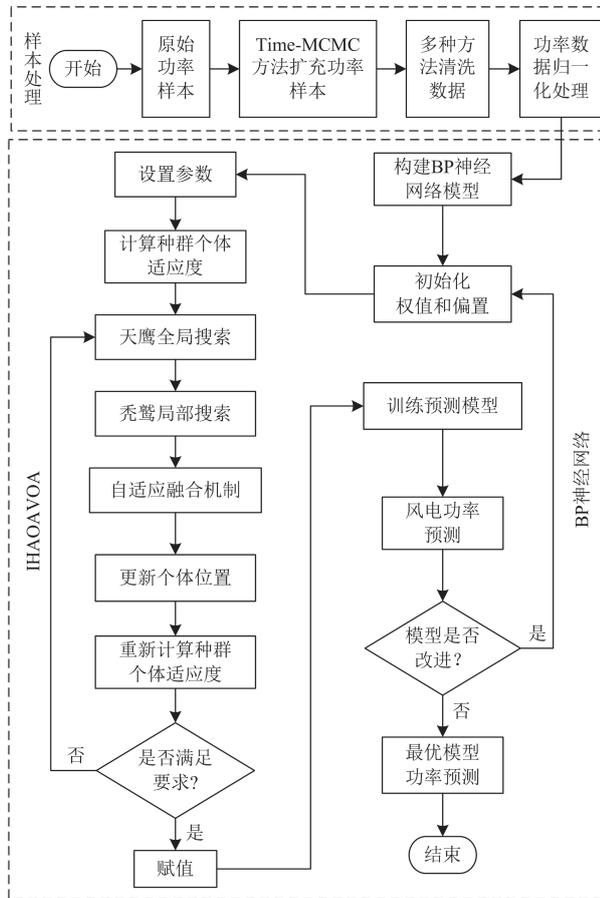


图 2 风电功率预测流程图

Fig. 2 Flow chart of wind power prediction

4 算例分析

4.1 算例概况及参数配置

为充分验证所提方法的有效性, 采用两个算例样本集, 算例 1 为德国某风电场 2019 年 01 月 21 日至 2019 年 01 月 24 日的风电功率数据, 算例 2 为河南省某风电场 2019 年 04 月 15 日至 2019 年 04 月 18 日的风电功率数据, 每个算例共 300 个样本,

每个样本包含 4 个特征, 依次为 10 m、30 m 和 50 m 处风速以及轮轴处的风速。

采用 MAE、均方根误差(root mean square error, RMSE)和决定系数(coefficient of determination, R^2) 3 个指标对数据扩充效果进行评价, 采用皮尔逊相关系数对数据清洗方法进行评价; 采用 MAE、平均绝对百分比误差(mean absolute percentage error, MAPE)、MSE 和 RMSE 4 个指标对预测效果进行分析, 各指标计算分别如式(27)一式(32)所示。

平均绝对值误差为

$$e_{MAE} = \frac{1}{N_c} \sum_{f=1}^{N_c} |P'_f - P_f^s| \quad (27)$$

均方根误差为

$$e_{RMSE} = \sqrt{\frac{1}{N_c} \sum_{f=1}^{N_c} (P'_f - P_f^s)^2} \quad (28)$$

均方误差为

$$e_{MSE} = \frac{1}{N_c} \sum_{f=1}^{N_c} (P'_f - P_f^s)^2 \quad (29)$$

平均绝对百分比误差为

$$e_{MAPE} = \frac{1}{N_c} \sum_{f=1}^{N_c} \left| \frac{P'_f - P_f^s}{P_f^s} \right| \times 100\% \quad (30)$$

决定系数为

$$R^2 = \frac{\sum_{f=1}^{N_c} (P'_f - P_f^s)^2}{\sum_{f=1}^{N_c} (P_f^s - \bar{P}^s)^2} \quad (31)$$

皮尔逊相关系数为

$$r' = \frac{\sum_{f=1}^N (y_f^k - \bar{Y}^k)(P_f^s - \bar{P}^s)}{\sqrt{\sum_{f=1}^N (y_f^k - \bar{Y}^k)^2} \sqrt{\sum_{f=1}^N (P_f^s - \bar{P}^s)^2}} \quad (32)$$

式中: N_c 为测试集样本总数; N 为数据集样本总数; P'_f 和 P_f^s 分别为功率预测值和实际值, 不包含风速特征值; \bar{P}^s 为风电功率实际值的均值; y_f^k 为第 k 个样本风速特征的值; \bar{Y}^k 为第 k 个样本风速特征的平均值。

样本处理有关参数设置: 每个算例原始风电功率小样本集的样本个数为 300, Time-MCMC 迭代次数为 8000, 生成风电功率样本个数为 5000; 滑动窗线性插值窗口大小为 15, RANSAC 直线拟合采样频率为 30 个/次, 内点到模型最大距离为 400; 孤立森立算法异常比例为 0.05。

功率预测参数设置：BP 神经网络输入层为 1 层，输入层神经元为 4 个，隐藏层为 1 层，隐藏层神经元 2 个；输出层为 1 层，输出层神经元为 1 个，学习率为 0.01；迭代次数为 1000，最小误差为 0.0001，动量因子为 0.01，延迟步骤为 15，预测间隔为 1；改进混合优化算法个体数量为 30，迭代次数为 300。

4.2 样本处理结果分析

4.2.1 生成样本质量分析

基于 Time-MCMC 算法对风电场数据进行扩充，生成样本数量为 5000，算例 1 样本扩充后的数据质量验证部分特征可视化结果如图 3 所示。

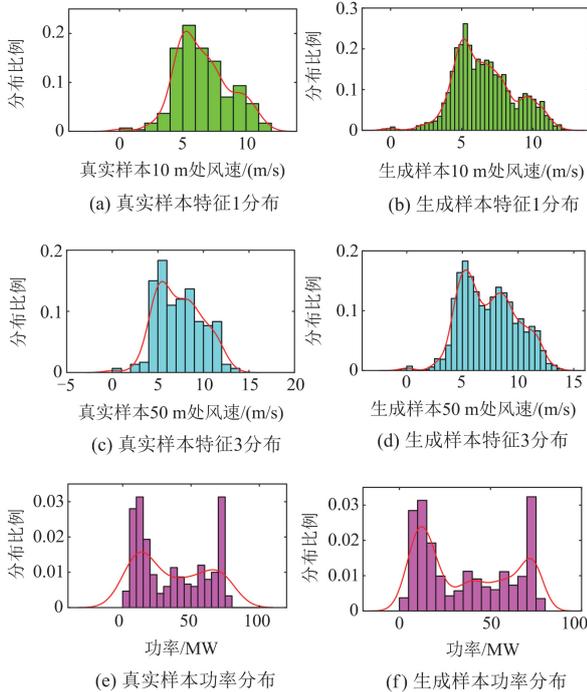


图 3 原始样本与生成样本特征分布

Fig. 3 Feature distribution of original sample and generated sample

图 3 中纵轴代表不同大小区间的样本所占总体的分布比例，可以看出，算例 1 真实样本和生成样本的特征 1、特征 3 和功率的分布大致相同，表明 Time-MCMC 算法较好地学习了原始样本的功率特征，生成了带有真实样本特征的新样本。

为进一步验证采用 Time-MCMC 技术生成数据的可靠性，采用 BP 神经网络预测模型进行测试。以原始小样本集后 20% 作为测试集，原始小样本集前 80%、MCMC 算法生成的风电功率样本和 Time-MCMC 算法生成的功率样本分别作为训练集，算例生成样本质量验证方法的训练集和测试集具体设置如表 1 所示，生成数据质量验证结果如表 2 所示。

表 1 样本质量验证方法设置

方法	训练集	测试集
不处理	原始小样本集(80%)	原始小样本集(20%)
MCMC	生成样本(5000)	原始小样本集(20%)
Time-MCMC	生成样本(5000)	原始小样本集(20%)

表 2 不同训练集下算例预测结果

算例	方法	e_{MAE}/MW	e_{MSE}/MW^2	R^2
算例 1	不处理	5.35	7.65	0.92
	MCMC	4.03	5.23	0.95
	Time-MCMC	3.83	4.92	0.96
算例 2	不处理	8.28	10.10	0.93
	MCMC	5.96	7.05	0.95
	Time-MCMC	5.47	6.79	0.96

由表 2 可以看出，对于相同的测试集，与原始小样本集训练的预测结果相比，算例 1 中 MCMC 方法生成样本训练的预测结果 MAE 和 MSE 分别降低了 1.32 MW 和 2.42 MW²，R² 提高了 3%，算例 2 中 MCMC 方法生成样本训练的预测结果 MAE 和 MSE 分别降低了 2.32 MW 和 3.05 MW²，R² 提高了 2%；与 MCMC 方法生成样本训练的预测结果相比，算例 1 中 Time-MCMC 方法生成样本训练的预测结果 MAE 和 MSE 分别降低了 0.20 MW 和 0.31 MW²，R² 提高了 1%，算例 2 中 Time-MCMC 方法生成样本训练的预测结果 MAE 和 MSE 分别降低了 0.49 MW 和 0.26 MW²，R² 提高了 1%。

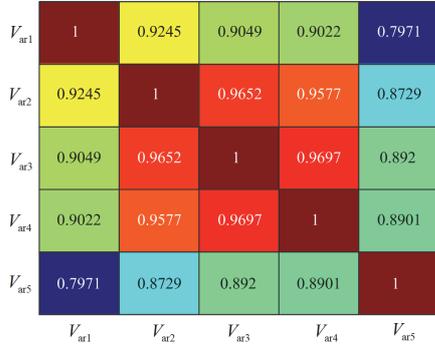
预测结果表明，与 MCMC 方法相比，Time-MCMC 方法处理小样本集的效果更好，能够更好地避免样本量不足导致预测精度低的问题，也能较好地学习风电样本特征，生成质量良好的数据。

4.2.2 数据清洗结果分析

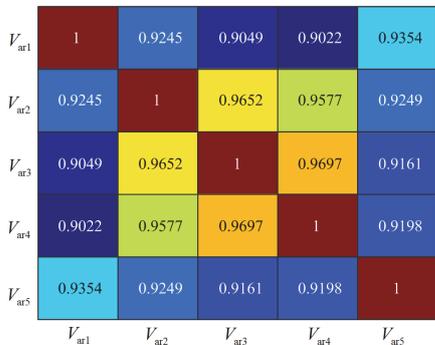
为验证数据清洗效果，运用皮尔逊相关系数衡量风速与功率之间的相关性，算例 1 扩充后的风电功率数据集清洗前后风速与功率之间的相关性热力图如图 4 所示。图中， V_{ar1} — V_{ar4} 依次表示 4 个样本的风速特征， V_{ar5} 表示风电功率。算例 1 和算例 2 清洗前后 4 个特征与功率之间的相关性如表 3 所示。

由图 4 可以看出，算例 1 数据清洗前风速与风电功率之间相关性不强，经数据清洗后，各风速特征与功率之间的相关性均得到了增强。由表 3 可以看出，算例 1 数据清洗后，各风速特征与功率之间的相关性分别提高了 0.1383、0.0520、0.0241 和 0.0297；算例 2 数据清洗后，各风速特征与功率之间的相关性也得到了增强，分别提高了 0.1466、0.0988、

0.0582 和 0.0458。两个算例的清洗结果表明, 数据集的相关性较差的点已被剔除或修正, 所提出的数据清洗方法能够提升风电功率数据集的整体质量, 为后续风电功率预测提供更可靠的样本数据。



(a) 数据清洗前相关性热力图



(b) 数据清洗后相关性热力图

图 4 算例 1 清洗前后相关性热力图

Fig. 4 Correlation heatmap before and after cleaning in case 1

表 3 两种算例数据清洗结果

Table 3 Data cleaning results of two cases

算例	相关性	特征 1	特征 2	特征 3	特征 4
算例 1	清洗前	0.7971	0.8729	0.8920	0.8901
	清洗后	0.9354	0.9249	0.9161	0.9198
算例 2	清洗前	0.7768	0.8273	0.8617	0.8658
	清洗后	0.9234	0.9261	0.9199	0.9116

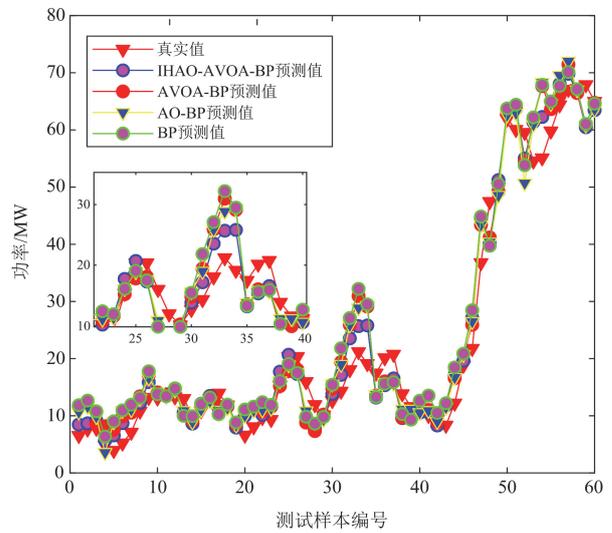
4.3 预测结果分析

为对 IHAOAVOA-BP 预测模型的有效性进行验证, 分别采用 BP、AO-BP、AVOA-BP 和 IHAOAVOA-BP 这 4 种模型对两个扩充后的算例进行风电功率预测, 其预测结果如图 5 所示, 两个算例 4 种模型误差指标 MAE、MAPE、MSE 和 RMSE 的计算结果如表 4 所示。

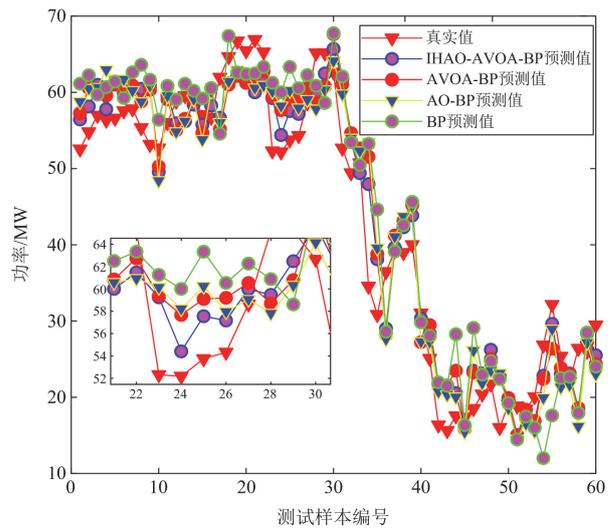
由图 5 可以看出, BP 模型预测结果与真实值偏差较大; 利用 AO 和 AVOA 分别对 BP 模型进行改进, AO-BP 模型和 AVOA-BP 模型预测精度均得到了提升, 两种模型预测结果更加接近真实值; 利

用 IHAOAVOA 对 BP 模型进行改进, IHAOAVOA-BP 预测偏差进一步减小, 与 AO-BP 模型和 AVOA-BP 模型相比, 预测准确性提升更多。

由表 4 可以看出, BP 模型 4 个误差指标均偏大, AO-BP 模型和 AVOA-BP 模型误差指标与 BP 模型相比均有下降, 但效果弱于 IHAOAVOA-BP 模型。与 BP 模型相比, AVOA-BP 模型算例 1 的 4 个指标 MAE、MAPE、MSE 和 RMSE 分别降低了 0.46 MW、3.64%、19.24% 和 0.48 MW, 算例 2 分别降低了 0.88 MW、2.95%、41.06% 和 1.54 MW; 与 AVOA-BP 模型相比, IHAOAVOA-BP 模型算例 1 的 MAE、MAPE、MSE 和 RMSE 分别降低了 0.45 MW、4.62%、



(a) 算例 1 预测结果



(b) 算例 2 预测结果

图 5 算例 4 种模型下的功率预测结果

Fig. 5 Power prediction results of the cases under four models

表 4 算例 4 种模型下的预测指标
Table 4 Prediction indicators of the cases under four models

算例	预测模型	e_{MAE}/MW	$e_{MAPE}/\%$	e_{MSE}/MW^2	e_{RMAE}/MW	计算时间/s
算例 1	BP	3.81	26.57	22.87	4.78	393.60
	AO-BP	3.36	22.10	18.63	4.32	557.16
	AVOA-BP	3.35	22.93	18.47	4.30	563.74
	IHAOAVOA-BP	2.90	18.31	13.01	3.61	652.32
算例 2	BP	5.29	15.42	44.15	6.64	382.35
	AO-BP	4.53	12.75	29.41	5.42	543.62
	AVOA-BP	4.41	12.47	26.02	5.10	560.53
	IHAOAVOA-BP	3.85	11.13	20.43	4.52	641.59

29.56%和 0.69 MW，算例 2 分别降低了 0.56 MW、1.34%、21.48%和 0.58 MW。因此，IHAOAVOA-BP 模型预测精度更高，性能更优。

5 结论

本文提出了一种小样本条件下的风电功率预测模型，对比了算例采用不同模型的预测结果，验证了所提模型在小样本条件下具有高精度和低误差的特点，同时得到了以下结论。

1) 采用 Time-MCMC 算法对小样本数据集进行扩充，利用 BP 模型对生成数据质量进行验证，表明了 Time-MCMC 扩充方法效果良好，丰富了样本数据。

2) 多种数据方法相结合对数据集进行清洗，相关性热力图表明数据处理后的特征与功率之间的关联性更加紧密，清洗后有效去除了对模型预测性能产生负面影响的异常值，增强了数据集的质量。

3) 分别采用 BP、AO-BP、AVOA-BP 和 IHAOAVOA-BP 这 4 种模型对算例进行训练和测试，结果表明 IHAOAVOA-BP 模型预测准确率较高，验证了该模型在风电功率预测方面的优越性。

本研究虽验证了所提方法在小样本条件下的有效性，但仍存在一定局限：Time-MCMC 生成数据依赖原始样本的统计特性，其在极端工况或数据存在系统性偏差场景下的泛化能力有待进一步验证；同时，混合优化算法因集成多重策略，计算成本相对较高，在预测实时性要求高的场景中应用受限，需探索轻量化策略以提升实时性。

参考文献

[1] 黄南天, 唐立鹏, 戴千斌. 计及 NWP 风速横纵向误差的循环确定性策略梯度风速修正模型[J]. 电力系统保护与控制, 2024, 52(23): 10-21.
HUANG Nantian, TANG Lipeng, DAI Qianbin. Wind speed correction method for numerical weather prediction based on a recurrent deterministic policy gradient algorithm

considering lateral and longitudinal error[J]. Power System Protection and Control, 2024, 52(23): 10-21.
[2] 潘超, 王超, 孙惠, 等. 基于超参数优化和误差修正的 STAGN 超短期风电功率预测[J]. 电力系统保护与控制, 2025, 53(8): 117-129.
PAN Chao, WANG Chao, SUN Hui, et al. STAGN ultra-short-term wind power forecasting based on hyperparameter optimization and error correction[J]. Power System Protection and Control, 2025, 53(8): 117-129.
[3] 张亚丽, 王聪, 张宏立, 等. 基于非平稳 Transformer 的超短期风电功率多步预测[J]. 智慧电力, 2024, 52(1): 108-115.
ZHANG Yali, WANG Cong, ZHANG Hongli, et al. Multi-step prediction of ultra-short-term wind power based on non-stationary transformer[J]. Smart Power, 2024, 52(1): 108-115.
[4] 李薇, 龚免彰, 时利香, 等. 基于区域气候模式的内蒙古地区风能资源预测评估[J]. 电网技术, 2020, 44(11): 4318-4327.
LI Wei, GONG Huanzhang, SHI Lixiang, et al. Prediction and evaluation of wind energy resource in inner Mongolia based on regional climate models[J]. Power System Technology, 2020, 44(11): 4318-4327.
[5] 杨国华, 祁鑫, 贾睿, 等. 基于 CEEMD-SE 的 CNN&LSTM-GRU 短期风电功率预测[J]. 中国电力, 2024, 57(2): 55-61.
YANG Guohua, QI Xin, JIA Rui, et al. Short-term wind power forecast based on CNN&LSTM-GRU model integrated with CEEMD-SE Algorithm[J]. Electric Power, 2024, 57(2): 55-61.
[6] 翟保豫, 张龙, 徐志, 等. 基于 WRF 模式和风速误差修正的中期风电功率预测方法[J]. 智慧电力, 2023, 51(7): 31-38.
ZHAI Baoyu, ZHANG Long, XU Zhi, et al. Mid-term wind power forecasting based on WRF mode and wind speed error correction[J]. Smart Power, 2023, 51(7): 31-38.
[7] 陈昊, 张建忠, 许超, 等. 基于多重离群点平滑转换自

- 回归模型的短期风电功率预测[J]. 电力系统保护与控制, 2019, 47(1): 73-79.
- CHEN Hao, ZHANG Jianzhong, XU Chao, et al. Short-term wind power forecast based on MOSTAR model[J]. Power System Protection and Control, 2019, 47(1): 73-79.
- [8] 赵永宁, 李卓, 叶林, 等. 基于时空相关性的风电功率超短期自适应预测方法[J]. 电力系统保护与控制, 2023, 51(6): 94-105.
- ZHAO Yongning, LI Zhuo, YE Lin, et al. A very short-term adaptive wind power forecasting method based on spatio-temporal correlation[J]. Power System Protection and Control, 2023, 51(6): 94-105.
- [9] 屈尹鹏, 徐箭, 姜尚光, 等. 基于频繁模式挖掘的风电爬坡事件统计特性建模及预测[J]. 电力系统自动化, 2021, 45(1): 36-43.
- QU Yinpeng, XU Jian, JIANG Shangguang, et al. Frequent pattern mining based modeling and forecasting for statistical characteristics of wind power ramp events[J]. Automation of Electric Power Systems, 2021, 45(1): 36-43.
- [10] 石立贤, 金怀平, 杨彪, 等. 基于局部学习和多目标优化的选择性异质集成超短期风电功率预测方法[J]. 电网技术, 2022, 46(2): 568-577.
- SHI Lixian, JIN Huaiping, YANG Biao, et al. Selective heterogeneous ensemble for ultra-short-term wind power forecasting based on local learning and multi-objective optimization[J]. Power System Technology, 2022, 46(2): 568-577.
- [11] 苗长新, 王霞, 李昊, 等. 基于数值天气预报风速误差修正的风电功率日前预测[J]. 电网技术, 2022, 46(9): 3455-3464.
- MIAO Changxin, WANG Xia, LI Hao, et al. Day-ahead prediction of wind power based on NWP wind speed error correction[J]. Power System Technology, 2022, 46(9): 3455-3464.
- [12] 邬永, 王冰, 陈玉全, 等. 融合精细化气象因素与物理约束的深度学习模型在短期风电功率预测中的应用[J]. 电网技术, 2024, 48(4): 1455-1468.
- WU Yong, WANG Bing, CHEN Yuquan, et al. Application of deep learning model integrating refined meteorological factors and physical constraints in short-term wind power prediction[J]. Power System Technology, 2024, 48(4): 1455-1468.
- [13] 牛东晓, 纪会争. 风电功率物理预测模型引入误差量化分析方法[J]. 电力系统自动化, 2020, 44(8): 57-65.
- NIU Dongxiao, JI Hui Zheng. Quantitative analysis method for errors introduced by physical prediction model of wind power[J]. Automation of Electric Power Systems, 2020, 44(8): 57-65.
- [14] 魏泽涛, 刘友波, 沈晓东, 等. 基于样本数据迁移学习的贫资料地区小水电超短期出力建模及发电预测[J]. 中国电机工程学报, 2023, 43(7): 2652-2666.
- WEI Zetao, LIU Youbo, SHEN Xiaodong, et al. Ultra-short-term power generation modeling and prediction for small hydropower in data-scarce areas based on sample data transfer learning[J]. Proceedings of the CSEE, 2023, 43(7): 2652-2666.
- [15] FUJIMOTO Y, TAKAHASHI Y, HAYASHI Y. Alerting to rare large-scale ramp events in wind power generation[J]. IEEE Transactions on Sustainable Energy, 2019, 10(1): 55-65.
- [16] 符杨, 周全, 贾锋, 等. 基于 SCADA 数据图形化的海上风电机组故障预测[J]. 中国电机工程学报, 2022, 42(2): 7465-7475.
- FU Yang, ZHOU Quan, JIA Feng, et al. Fault prediction of offshore wind turbines based on graphical processing of SCADA data[J]. Proceedings of the CSEE, 2022, 42(2): 7465-7475.
- [17] 程凯, 彭小圣, 徐其友, 等. 基于特征选择与多层级深度迁移学习的风电场短期功率预测[J]. 高电压技术, 2022, 48(2): 497-503.
- CHENG Kai, PENG Xiaosheng, XU Qiyu, et al. Short-term wind power prediction based on feature selection and multi-level deep transfer learning[J]. High Voltage Engineering, 2022, 48(2): 497-503.
- [18] 吴青峰, 杨艺涛, 刘立群, 等. 基于 GA-SA-BP 神经网络的锂电池健康状态估算方法[J]. 电力系统保护与控制, 2024, 52(19): 74-84.
- WU Qingfeng, YANG Yitao, LIU Liqun, et al. Lithium battery state of health estimation method based on a GA-SA-BP neural network[J]. Power System Protection and Control, 2024, 52(19): 74-84.
- [19] 王海燕, 刘佳康, 邓亚平. 基于预估-校正综合 BP 神经网络的短期光伏功率预测[J]. 智慧电力, 2023, 51(3): 46-52.
- WANG Haiyan, LIU Jiakang, DENG Yaping. Short-term photovoltaic power forecasting based on predict-correct combination BP neural network[J]. Smart Power, 2023, 51(3): 46-52.
- [20] CHEN Jinbao, HE Gang, WANG Yunhe, et al. Adaptive PID control for hydraulic turbine regulation systems based on INGWO and BPNN[J]. Protection and Control of Modern Power Systems, 2024, 9(4): 126-146.
- [21] 王妍, 王新发, 王延峰, 等. 基于改进天鹰优化算法优化 LSTM 的滚动轴承故障诊断方法[J]. 振动与冲击, 2024, 43(23): 144-154.
- WANG Yan, WANG Xinfu, WANG Yanfeng, et al. A rolling bearing fault diagnosis method based on improved aquila optimization algorithm to optimize LSTM[J].

- Journal of Vibration and Shock, 2024, 43(23): 144-154.
- [22] ABUELROUB A, AWWAD B, ALMASRI K M H. Solving wind-integrated unit commitment problem by a modified African vultures optimization algorithm[J]. IET Generation, Transmission & Distribution, 2023, 17(16): 3678-3691.
- [23] 李军毅, 汪兴兴, 陈祥, 等. 基于 AO-AVOA-BP 神经网络模型的锂电池 SOH 预测[J]. 电子测量技术, 2025, 48(4): 71-79.
LI Juyin, WANG Xingxing, CHEN Xiang, et al. SOH prediction of lithium battery based on AO-AVOA-BP neural network modeling[J]. Electronic Measurement Technology, 2025, 48(4): 71-79.
- [24] 叶林, 李稼辰, 路朋, 等. 基于近邻传播聚类与 MCMC 算法的风电时序数据聚合方法[J]. 中国电机工程学报, 2020, 40(12): 3744-3754.
YE Lin, LI Jiachen, LU Peng, et al. Wind power time series aggregation approach based on affinity propagation clustering and MCMC algorithm[J]. Proceedings of the CSEE, 2020, 40(12): 3744-3754.
- [25] 张柏林, 李希德, 魏博, 等. 基于改进的场景分类和去粗粒化 MCMC 的风电出力模拟方法[J]. 电测与仪表, 2024, 61(7): 41-49, 122.
ZHANG Bolin, LI Xide, WEI Bo, et al. Wind power output simulation method based on improved scene classification algorithm and coarse-grained MCMC[J]. Electrical Measurement & Instrumentation, 2024, 61(7): 41-49, 122.
- [26] LAITH A, YOUSRI D, ABD EILAZIZ M, et al. Aquila optimizer: a novel meta-heuristic optimization algorithm[J]. Computers & Industrial Engineering, 2021, 157.
- [27] 叶育林, 刘森, 黄松, 等. 基于 GJO 特征量优选的 AO-RF 的变压器故障诊断模型[J]. 高压电器, 2024, 60(5): 99-107.
YE Yulin, LIU Sen, HUANG Song, et al. Fault diagnosis model of transformer based on GJO feature optimization and AO-RF[J]. High Voltage Apparatus, 2024, 60(5): 99-107.
- [28] ABDOLLAHZADEH B, GHAREHCHOPOGH F S, MIRJALILI S. African vultures optimization algorithm: a new nature-inspired metaheuristic algorithm for global optimization problems[J]. Computers & Industrial Engineering, 2021, 158.
- [29] 刘倩, 陶文化, 王智聪, 等. 基于改进天鹰优化算法的隐式广义预测控制[J]. 国外电子测量技术, 2023, 42(3): 52-58.
LIU Qian, TAO Wenhua, WANG Zhicong, et al. Implicit generalized predictive control based on improved aquila optimization algorithm[J]. Foreign Electronic Measurement Technology, 2023, 42(3): 52-58.
- [30] 王彦海, 李书炀, 邓德慧, 等. 基于改进 IAVOA-BP 算法的 GFRP 布加固角钢极限承载力预测模型研究[J]. 国外电子测量技术, 2023, 42(11): 65-73.
WANG Yanhai, LI Shuyang, DENG Dehui, et al. Prediction algorithm of ultimate bearing capacity of angle steel reinforced by GFRP sheet based on improved IAVOA-BP algorithm[J]. Foreign Electronic Measurement Technology, 2023, 42(11): 65-73.
- [31] RENE E A, FOKUI W S T. Modeling and control of automatic voltage regulation for a hydropower plant using advanced model predictive control[J]. Global Energy Interconnection, 2025, 8(2): 269-285.
- [32] 金冰霖, 晁璞璞, 刘志辉, 等. 全功率型风电机组调频控制器的实用化辨识方法[J]. 电力系统自动化, 2024, 48(8): 91-100.
JIN Yonglin, CHAO Pupu, LIU Zhihui, et al. Practical identification method of frequency regulation controller for full-power wind turbines[J]. Automation of Electric Power Systems, 2024, 48(8): 91-100.
- [33] 孙国良, 伊力哈木·亚尔买买提, 张宽, 等. 基于小波变换与 IAGA-BP 神经网络的短期风电功率预测[J]. 电测与仪表, 2024, 61(5): 126-134, 145.
SUN Guoliang, YILIHAMU·Yaermainaiti, ZHANG Kuan, et al. Short-term prediction of wind power based on wavelet transform and IAGA-BP neural network[J]. Electrical Measurement & Instrumentation, 2024, 61(5): 126-134, 145.
- [34] 梁恩豪, 孙军伟, 王延峰. 基于自适应樽海鞘算法优化 BP 的风光互补并网发电功率预测[J]. 电力系统保护与控制, 2021, 49(24): 114-120.
LIANG Enhao, SUN Junwei, WANG Yanfeng. Wind and solar complementary grid-connected power generation prediction based on BP optimized by a swarm intelligence algorithm[J]. Power System Protection and Control, 2021, 49(24): 114-120.

收稿日期: 2024-09-28; 修回日期: 2025-04-20

作者简介:

张志艳(1975—), 女, 博士, 副教授, 研究方向为新能源发电与并网; E-mail: 2004074@zzuli.edu.cn

兰龙(2004—), 男, 通信作者, 本科生, 研究方向为新能源发电与并网; E-mail: 2532914630@qq.com

白景升(2004—), 男, 本科生, 研究方向为新能源发电与并网. E-mail: 857529718@qq.com

(编辑 周金梅)