

孟建军, 江相君, 孟高阳, 等. 基于 IBES - XGBoost 的高速铁路沿线风速预测模型[J]. 灾害学, 2024, 39(1): 57 - 63.
[MENG Jianjun, JIANG Xiangjun, MENG Gaoyang, et al. Wind Speed Prediction Model Along High - speed Railways Based on IBES - XGBoost[J]. Journal of Catastrophology, 2024, 39(1): 57 - 63. doi: 10.3969/j.issn.1000-811X.2024.01.010.]

基于 IBES - XGBoost 的高速铁路沿线风速预测模型^{*}

孟建军^{1,2,3}, 江相君¹, 孟高阳⁴, 李德仓^{1,2,3}

(1. 兰州交通大学 机电技术研究所, 甘肃 兰州 730070; 2. 甘肃省物流及运输装备信息化工程技术研究中心, 甘肃 兰州 730070; 3. 甘肃省物流与运输装备行业技术中心, 甘肃 兰州 730070;
4. 兰州交通大学 机电工程学院, 甘肃 兰州 730070)

摘 要: 为保证高速铁路沿线风灾预警信息具有较高时效性, 需要进行高速铁路沿线超短期风速的提前多步预测。针对众多预测模型在预测中可能存在较大误差的问题, 采用 Tent 混沌映射和 BFGS 拟牛顿法对秃鹰搜索算法进行改进, 并用改进的秃鹰搜索算法 (IBES) 对 XGBoost 的初始参数进行优化。在构建 IBES - XGBoost 模型时, 加入风速以外的其他气象特征, 以提高预测精度。实验结果表明: ①改进的秃鹰算法相比其他智能优化算法有更好的寻优能力, 与其他模型相比 IBES - XGBoost 在超短期风速的提前多步预测上有着较高的精度和较好的拟合效果。②Tent 混沌映射和 BFGS 拟牛顿法对秃鹰算法有着较好的改进效果。③IBES - XGBoost 能为高速铁路规范下的大风预警提供可靠的提前多步预测结果。

关键词: 高速铁路; 风灾; 风速预测; 机器学习; 秃鹰搜索算法

中图分类号: P425.6+2; X43; X915.5 **文献标志码:** A **文章编号:** 1000-811X(2024)01-0057-07

doi: 10.3969/j.issn.1000-811X.2024.01.010

当前我国高速铁路总里程数已突破 4 万 km, 高速铁路遍布全国各地。而在影响列车运行的诸多自然灾害中, 风灾的作用较为显著^[1]。我国幅员辽阔, 大部分地区处在季风区, 冬春季节受蒙古 - 西伯利亚高压影响盛行冬季风, 夏秋季节主要受东南季风影响, 特定的季节主导风向容易形成较强势气, 进而形成大风天气。我国许多高速铁路段都处在强风的威胁中。例如, 兰新铁路第二双线新疆段近 60% 线路需要穿越大风区^[2]。东部沿海地区遭受台风灾害的影响非常严重^[3]。由大风引起的行车事故在各国屡见不鲜, 在我国南疆线等强风区, 曾多次出现列车脱轨、倾覆事故^[4]。因此, 建立一套行之有效的风速预测方法对高速铁路沿线的大风预测预警尤为重要。

风速预测相关算法层出不穷, 随着人工智能相关理论日趋成熟, 许多相关算法逐渐应用到风速预测领域。XU 等^[5] 将长短期记忆神经网络 (Long short-term memory neural network, LSTM) 和卷积神经网络相结合的方式对风速进行预测。TUNERXUN 等^[6] 先对风速数据进行分解, 而后使用 LSTM 辅以改进的金枪鱼群优化算法进行风速预测。LI 等^[7] 先使用变分模态分解对风速数据进行分解, 并采用粒子群算法 (Particle Swarm Optimiza-

tion, PSO) 优化双向长短期记忆神经网络 (Bidirectional Long Short Term Memory, Bi-LSTM), 进而完成风速预测。SHEN 等^[8] 使用卷积神经网络 (Convolutional Neural Network, CNN) 和 LSTM 相结合的方式对风速进行预测。BOMMIDI 等^[9] 使用完全集成经验模态分解 (Complete Ensemble Empirical Mode Decomposition With Adaptive Noise, CEEM-DAN) 分解风速数据, 再使用 Informer 进行风速预测。向玲等^[10] 使用变分模态分解 (Variational Mode Decomposition, VMD) 对风速数据进行一次分解, 再使用 CEEMDAN 对风速数据进行二次分解, 最后使用 LSTM 进行风速预测。以上方法均为改进深度学习模型, 或将深度学习模型与分解算法、智能优化算法等方法相结合。这些方法能在一定程度上解决风速预测问题, 但还存在一系列问题。例如, 神经网络算法存在可解释性差、计算时间长等问题^[11], 部分算法未结合其他气象要素进行预测等。此外, 王瑞^[12] 指出, 高速铁路大风预警信息应至少提前 3 min 预测, 因此对逐分钟风速数据进行提前三步进行风速预测是高速铁路大风预警业务的最重要的组成部分。

针对以上存在的问题, 本文使用 XGBoost 融合改进的秃鹰搜索算法 (Improved Bald Eagle Search

^{*} 收稿日期: 2023-06-05 修回日期: 2023-07-28

基金项目: 国家自然科学基金项目“复杂环境下高速列车安全运行智能管控策略研究: 以川藏高铁为例”(7206121); 国家自然科学基金项目“高速车辆主动悬架超磁致伸缩作动器设计理论及控制方法”(62063013); 甘肃省科技计划项目铁路大风监测预警系统关键技术研究(21JR7RA284)

第一作者简介: 孟建军(1966-), 男, 汉族, 河南淮阳人, 博士, 教授, 博士生导师, 主要从事轨道交通装备检测与监控技术研究。E-mail: mengjj1966@163.com

Algorithm, IBES) 构建提前三步风速预测模型。最后, 利用逐分钟风速数据集进行验证, 使用多个评价指标对比 IBES - XGBoost 与其他模型的精度。结果表明, 所提模型相比其他众多模型具有较高预测精度。

1 理论基础

1.1 XGBoost 算法

XGBoost 对梯度提升回归数进行了改进^[13], 相比于传统的 Boosting 实现方法, XGBoost 在训练速度和预测精度上都有了提升。XGBoost 对损失函数进行了二阶泰勒展开, 同时引入两个正则化项求解整体最优, 并使用线性搜索方法寻找弱学习器。树的集成模型为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F. \quad (1)$$

式中: \hat{y}_i 为样本 i 的预测结果, K 为树的总数, f_k 对应树的结构 q 和叶子权重 ω , x_i 为特征向量, F 为回归树的空间。

XGBoost 的目标函数为:

$$X_{obj} = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k). \quad (2)$$

式中: $\sum_i l(\hat{y}_i, y_i)$ 用于衡量预测值和实际值的差异, \hat{y}_i 为预测值, y_i 为实际值。

式(2)中 $\Omega(f_k)$ 的表达式为:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2. \quad (3)$$

式中: γ 和 λ 为权重系数, T 为叶子节点的数量, ω 为各叶子节点的分数。

在 t 次迭代时, 目标函数为:

$$X_{obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) + C. \quad (4)$$

对式(4)进行二阶泰勒展开可得式(5):

$$X_{obj}^{(t)} \cong \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + C. \quad (5)$$

式中: g_i 和 h_i 为损失函数的一阶导数和二阶导数, $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$, $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ 。

1.2 秃鹰搜索算法

秃鹰搜索算法 (Bald Eagle Search Algorithm, BES) 是 ALSATTAR 等^[14] 提出的群智能优化算法。该算法模仿秃鹰的捕猎行为, 在给定空间内进行寻优。该算法分为三个阶段, 第一阶段为秃鹰选择最佳搜索空间, 第二阶段为秃鹰在选定的空间内进行搜索, 第三阶段为秃鹰从最佳位置俯冲向猎物。

在选择搜索空间阶段, 秃鹰根据猎物数量选择最佳搜索区域 (选择猎物量大的区域), 具体数学描述如式(6)。

$$P_{new,i} = P_{best} + \alpha \times r \times (P_{mean} - P_i). \quad (6)$$

式中: P_{best} 为秃鹰根据先前搜索情况确定的最佳位置, α 是区间 $[1.5, 2]$ 内取值的控制位置变化的参数, r 为区间 $(0, 1)$ 内产生的随机数, P_{mean} 为先前搜索结束后秃鹰的平均分布位置, P_i 为第 i 个秃鹰的位置。

在选定空间内搜索阶段, 秃鹰在选定的空间以螺旋状飞行, 寻找最佳俯冲位置, 具体数学表达式如式(7) ~ (11) 所示。

$$P_{i,new} = P_i + y(i) \times (P_i - P_{i+1}) + x(i) \times (P_i - P_{mean}); \quad (7)$$

$$x(i) = \frac{xr(i)}{\max(|xr|)}, y(i) = \frac{yr(i)}{\max(|yr|)}; \quad (8)$$

$$xr(i) = r(i) \times \sin(\theta(i)), yr(i) = r(i) \times \cos(\theta(i)); \quad (9)$$

$$\theta(i) = \alpha \times \pi \times \text{rand}; \quad (10)$$

$$r(i) = \theta(i) + R \times \text{rand}. \quad (11)$$

式中: $\alpha \in (5, 10)$ 为搜索点和中心点之间的夹角, $R \in (0.5, 2)$ 用于确定搜索周期数, $\theta(i)$ 和 $r(i)$ 分别为螺旋方差的极角与极径, $x(i)$ 和 $y(i)$ 取值在 $(-1, 1)$ 内用于表示极坐标下秃鹰的位置, rand 为 $[0, 1]$ 内的随机数。

在俯冲向猎物阶段, 秃鹰从搜索空间内最佳位置俯冲向目标猎物, 种群内其他个体也向最佳点移动, 具体数学表达式如式(12) ~ (18) 所示。

$$P_{i,new} = \text{rand} \times P_{best} + x1(i) \times (P_i - c1 \times P_{mean}) + y1(i) \times (P_i - c2 \times P_{best}); \quad (12)$$

$$x1(i) = \frac{xr(i)}{\max(|xr|)}; \quad (13)$$

$$y1(i) = \frac{yr(i)}{\max(|yr|)}; \quad (14)$$

$$xr(i) = r(i) \times \sinh(\theta(i)); \quad (15)$$

$$yr(i) = r(i) \times \cosh(\theta(i)); \quad (16)$$

$$\theta(i) = a \times \pi \times \text{rand}; \quad (17)$$

$$r(i) = \theta(i). \quad (18)$$

式中: $c1, c2 \in [1, 2]$, 且均用于增加秃鹰俯冲向最佳点和中心点移动的强度。

2 改进秃鹰搜索算法

BES 虽然收敛速度和精度优于大部分智能优化算法, 但是仍然存在陷入局部最优和部分问题求解精度不高等问题。本文从增强全局搜索能力和增强局部搜索能力两个角度出发, 提出了一种改进的秃鹰搜索算法 (IBES)。首先, 在种群初始化阶段使用 Tent 混沌映射替代 BES 的初始化策略, 以此增强种群的多样性, 提高全局搜索能力。其次, 每间隔一定的迭代次数, 若算法收敛情况较差, 则在 BES 的基础上进一步采用 BFGS 拟牛顿法进行搜索, 从而增强局部的搜索能力。

2.1 Tent 混沌映射

BES 的初代种群, 是在搜索空间中通过随机数生成的。通过随机数产生的种群, 其多样性可能较差, 进而导致寻优结果较差。而混沌映射具有随机性和遍历性的特点, 这些特性可以保持种群的多样性, 使智能优化算法摆脱局部陷阱, 提算法的全局探索能力。目前常用的混沌模型较多, FAN 等^[15] 指出, 大量实验表明, Tent 混沌映射在目前常用的混沌映射中性能最好。Tent 映射能在 $[0, 1]$ 内产生分布较为均匀的数值, 应用于智能优化算法后能够产生分布较为均匀的初始种群。

Tent 混沌映射的表达式为:

$$x_{i+1} = \begin{cases} \frac{x_i}{u}, & 0 \leq x_i \leq \frac{1}{2}; \\ \frac{1-x_i}{1-u}, & \frac{1}{2} \leq x_i \leq 1. \end{cases} \quad (19)$$

当 $u = 0.5$ 时, Tent 混沌映射的均匀性最好^[16]。因此, 本文所用的 Tent 混沌映射公式为:

$$x_{i+1} = \begin{cases} 2x_i, & 0 \leq x_i \leq 0.5; \\ 2(1-x_i), & 0.5 \leq x_i \leq 1. \end{cases} \quad (20)$$

2.2 BFGS 拟牛顿法

拟牛顿法是牛顿法的一种改进。BFGS 拟牛顿

法求解 Hessian 矩阵的近似矩阵, 避免每次迭代都要计算 Hessian 矩阵, 从而大大降低了计算的复杂度。BFGS 拟牛顿法是一种经典的局部搜索策略^[17], 部分学者研究表明在智能优化算法中加入 BFGS 拟牛顿法可以加快收敛速度^[18-19]。

若目标函数在定义域上二次连续可微, 则对函数 $f(x)$ 在 x_{k+1} 处作二阶泰勒展开可得:

$$f(x) \approx f(x_{k+1}) + g_{k+1}^T(x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^T G_{k+1}(x - x_{k+1}). \quad (21)$$

对式(21)进行求导可得:

$$g(x) \approx g_{k+1} + G_{k+1}(x - x_{k+1}). \quad (22)$$

式中: g_{k+1} 为梯度值, G_{k+1} 为 Hessian 矩阵。

令 $y_k = g_{k+1} - g_k$, $x = x_k$, $s_k = x_{k+1} - x_k$, 则有:

$$y_k \approx G_{k+1}s_k. \quad (23)$$

使用近似 Hessian 矩阵 B_{k+1} 替代 G_{k+1} 即可得到拟牛顿条件, 拟牛顿条件为:

$$y_k = B_{k+1}s_k. \quad (24)$$

式中: 近似 Hessian 矩阵 B_{k+1} 向 Hessian 矩阵做近似逼近的迭代公式为:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}. \quad (25)$$

在算法迭代过程中, 使用 Armijo 准则确定每次迭代的步长^[20]。若已知当前位置 x_k 和近似 Hessian 矩阵确定的搜索方向 d_k , 指定 $\beta \in (0, 1)$ 且 $\sigma \in (0, 0.5)$, 步长因子 $\alpha_k = \beta^{m_k}$, m_k 为满足式(26)的最小非负整数。

$$f(x_k + \beta^m d_k) \leq f(x_k) + \sigma \beta^m g_k^T d_k. \quad (26)$$

故可得下一个位置, 下一位置为:

$$x_{k+1} = x_k + \alpha_k d_k. \quad (27)$$

3 模型建立

3.1 基于 IBES - XGBoost 的提前多步风速预测模型

根据上文所述理论基础和业务背景, 构建基于 IBES - XGBoost 的提前多步风速预测模型。由于 XGBoost 的超参数对预测结果影响较大, 故采用改进的秃鹰搜索算法参数对 XGBoost 的学习率、树的最大深度和回归树数量进行寻优。

建立预测模型流程图(图 1), 具体步骤如下:

1) 输入原始数据, 根据皮尔逊系数筛选和风速数据正相关的特征。

2) 划分数据集, 将前 90% 的数据集作为训练集, 后 10% 的数据作为测试集。

3) 对数据集进行归一化处理, 并构建 XGBoost 模型。

4) 使用 IBES 对 XGBoost 的学习率、树的最大深度和回归树数量进行寻优。使用 Tent 混沌映射生成初始种群, 使用 XGBoost 模型作为目标函数, 采用 BES 原搜索策略的同时加入 BFGS 拟牛顿法加强局部搜索能力, 找到最优的初始参数组合。

5) 用步骤 4) 的寻优所得的初始参数进行训练, 训练得到 XGBoost 模型后即可进行测试。

6) 将数据反归一化, 并计算评价指标。

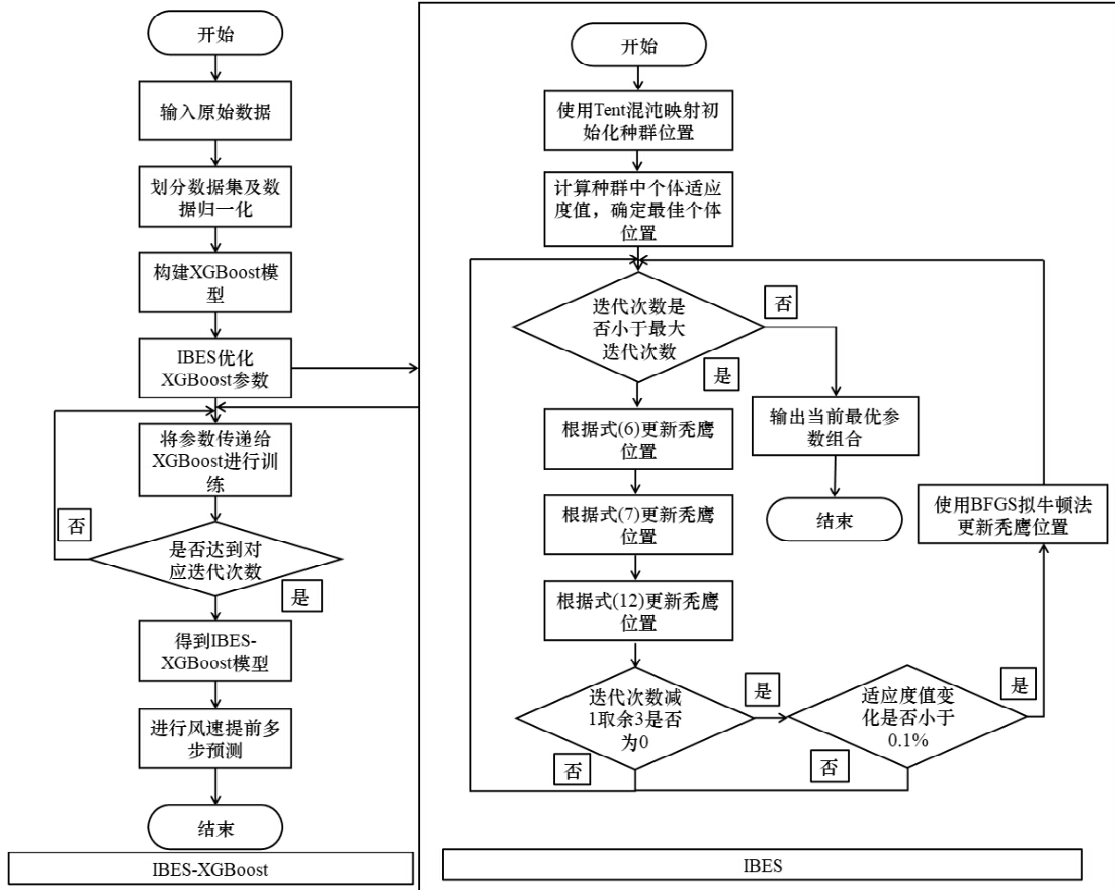


图 1 基于 IBES - XGBoost 的风速提前多步预测模型

表 1 基准测试函数

函数表达式	自变量范围	最优值
$F_1(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$[-30, 30]$	0
$F_2(x) = \sum_{i=1}^n ([x_i + 0.5])^2$	$[-100, 100]$	0
$F_3(x) = \frac{\pi}{n} \{ 10 \sin(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \sin^2(\pi y_{i+1})] + (y_n - 1)^2 \} + \sum_{i=1}^n u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i + 1}{4}$ $u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, & x_i > a; \\ 0; & \\ k(-x_i - a)^m, & x_i < a. \end{cases}$	$[-50, 50]$	0

3.2 模型评估指标

将测试集预测结果进行量化评价,用平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方根误差(RMSE)作为精度的评估指标。拟合优度(R^2)作为拟合程度的评价指标。具体公式为:

$$MAE = \frac{1}{n} \sum_{i=1}^m |y_i - \hat{y}_i|; \quad (28)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right|; \quad (29)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^m (y_i - \hat{y}_i)^2}; \quad (30)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2}. \quad (31)$$

式中: y_i 为真实值, \hat{y}_i 为预测值, \bar{y}_i 为真实数据的平均值。

4 实验分析与验证

4.1 IBES 性能测试

为了验证所提改进秃鹰搜索算法的有效性,选用三个基准测试函数进行测试(表 1)。表 1 中 $F_1 \sim F_2$ 为单峰测试函数, F_3 为多峰测试函数。单峰测试函数可以测试寻优精度,多峰测试函数可以测试全局寻优能力和收敛速度。

选择 IBES、BES、GWO 和 PSO 四种算法在 4 个测试函数上进行测试,种群初始数量均设置为 30,迭代次数均为 500,各个算法均独立运行 30 次。考虑到迭代次数较多,在进行性能测试时 IBES 间隔 50 次迭代判断一次是否使用 BFGS 拟牛顿法进行寻优。测试结果见表 2,表中加粗字体为最佳值。由表 2 可知,IBES 在多个基准测试函数上的性能明显优于其他三个算法,其精度大幅度优于其他算法。并且 IBES 的标准差也远小于其他算法,这说明 IBES 的寻优结果都集中在较小的区间内,算法的稳定性更优。各函数三维模型及算法收敛曲线见图 2~图 4。

表 2 智能优化算法在基准函数上测试结果

函数	算法	平均值	标准差
F_1	PSO	128.232	74.441
	GWO	27.191	0.779
	BES	19.532	1.219
	IBES	5.422×10^{-16}	3.158×10^{-16}
F_2	PSO	0.165	0.155
	GWO	0.632	0.359
	BES	1.074×10^{-18}	2.913×10^{-18}
	IBES	2.934×10^{-30}	9.599×10^{-30}
F_3	PSO	2.878×10^{-2}	6.068×10^{-2}
	GWO	4.299×10^{-2}	2.699×10^{-2}
	BES	2.091×10^{-20}	5.853×10^{-20}
	IBES	4.875×10^{-32}	7.778×10^{-32}

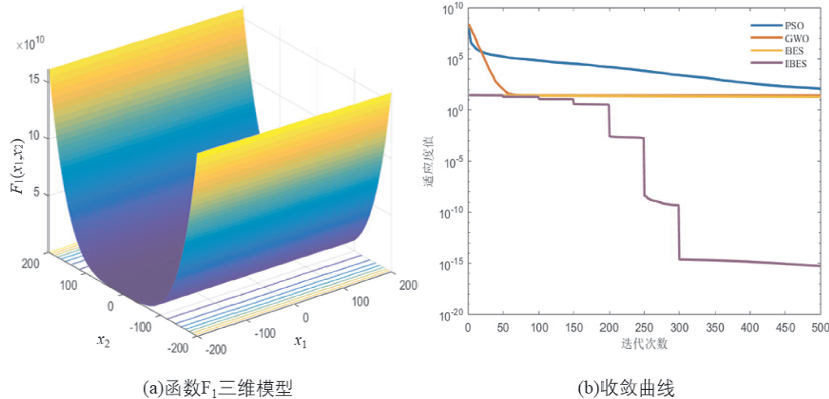
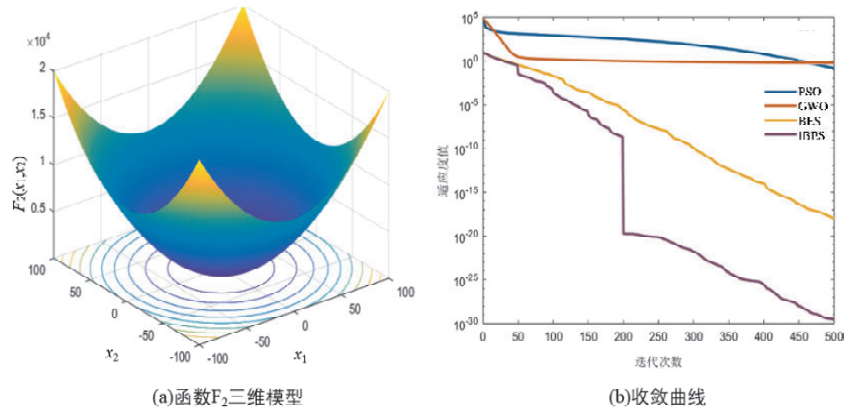
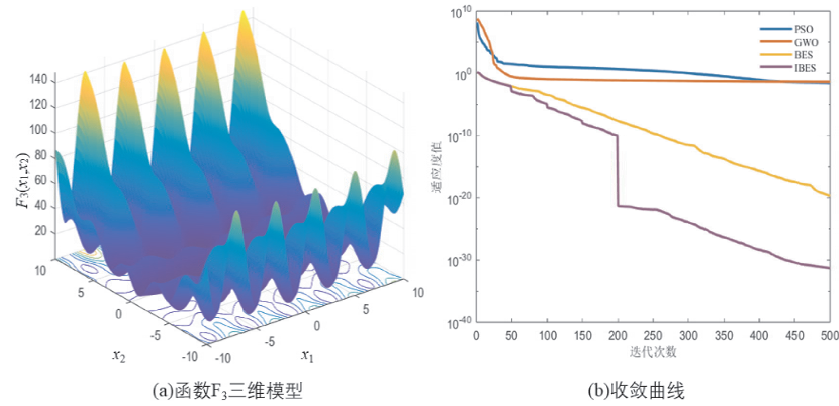


图 2 函数 F_1 三维模型及收敛曲线

图3 函数 F_2 三维模型及收敛曲线图4 函数 F_3 三维模型及收敛曲线

由图中可以看出, IBES 可以快速接近全局最优解, 而其他算法则会陷入局部最优。随着迭代次数的增加, IBES 借助 BFGS 拟牛顿法能进一步靠近最优值。

4.2 数据分析与预处理

风速容易受到其他气象因素变化的影响^[21], 因此考虑使用皮尔逊相关系数对比风速和温度、风寒温度、露点温度等其他气象因素的相关性, 从而选取相关性强的气象要素作为模型的特征输入。风寒温度指数衡量在一定气温和风速的情况下人体感受到寒冷的程度, 低(高)风寒温度代表低(高)温和强(弱)风^[22]。皮尔逊相关系数计算公式为:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (32)$$

数据集选用美国可再生能源实验某观测点的逐分钟观测数据^[23], 共计选取 22 257 条数据。选取前 90% 数据作为训练集, 后 10% 数据作为测试集。该数据集除风速数据外还包含逐分钟的部分气象信息。

表3 气象因素与风速之间的皮尔逊相关系数

气象因素	皮尔逊相关系数
温度/℃	0.282
露点温度/℃	-0.085
风寒温度/℃	0.072

由各气象要素与风速之间的皮尔逊相关系数(表3)可知, 在该数据集风速与温度相关性较强, 风速与风寒温度和露点温度相关性较弱。因此选用温度作为特征输入模型。

为提高模型的准确性和稳定性, 同时避免极端值对模型产生影响, 故对输入数据进行归一化, 对输出预测值进行反归一化得到最终风速数据。归一化公式为:

$$x_{\text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (33)$$

式中: x_i 为第 i 个原始数据, x_{\min} 为该数据的最小值, x_{\max} 为该数据的最大值, x_{norm} 为归一化后的数据。

4.3 预测结果

为验证 IBES 寻优预测模型时的优越性, 将 IBES 与 PSO、GWO 及 BES 进行对比, 对数据集进行逐分钟数据的提前 3 步预测。XGBoost 的学习率、树的最大深度和回归树数量寻优范围分别为 $[0.01, 0.3]$ 、 $[2, 100]$ 和 $[5, 1000]$ 。为确保实验的公平性, 各智能优化算法的初始种群数目均设为 20, 最大迭代次数均设为 15。从(图5)可以看出, IBES 具有较快的收敛速度, Tent 混沌映射策略能够生成更具有代表性的初始种群, BFGS 拟牛顿法能够加速收敛速度, 并提升迭代后期的寻优能力, 二者令 IBES 获得最优的适应度值。4 种算法寻优所得的最佳 XGBoost 超参数见表4。

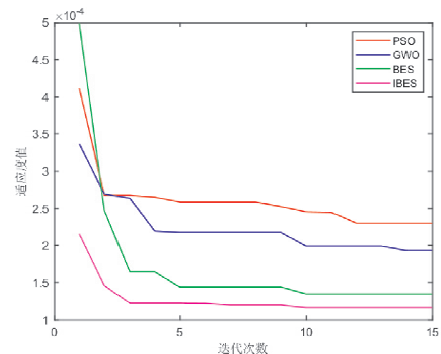


图5 各算法对 XGBoost 预测模型寻优时的适应度曲线

表 4 智能优化算法对 XGBoost 超参数寻优结果

算法	学习率	树的最大深度	回归树数量
IBES	0.258	3	997
BES	0.209	3	772
GWO	0.292	4	696
PSO	0.165	4	509

为全面评估模型性能，将 IBES - XGBoost、BES - XGBoost、GWO - XGBoost、PSO - XGBoost、XGBoost 和 LSTM 六种模型在测试集上的提前三步预测结果进行对比分析。单一 XGBoost 的学习率设置为 0.3，树的最大深度设置为 6，回归树的数量设置为 100。LSTM 的最大迭代次数为 500，隐藏层神经元个数为 200，学习率为 0.000 1。各模型的提前三步预测结果(图 6)可以看出，IBES - XGBoost 的拟合程度最好，其在波动较大的区间内也展现出了较好的预测效果。BES - XGBoost、GWO - XGBoost、PSO - XGBoost 和 XGBoost 在波动较大的采样点附近，预测精度不如 IBES - XGBoost。LSTM 在波动较大的区间内预测效果较差，拟合程度远不如其他模型。

表 5 各模型预测结果的评价指标

预测模型	RMSE(m/s)	MAE(m/s)	MAPE/%	R ²
IBES - XGBoost	0.257	0.178	16.644	0.996
BES - XGBoost	0.294	0.200	17.504	0.995
GWO - XGBoost	0.351	0.204	16.557	0.993
PSO - XGBoost	0.383	0.217	17.157	0.991
XGBoost	0.528	0.263	18.655	0.984
LSTM	1.003	0.597	38.083	0.941

从(表 5)各模型预测结果的评价指标可以看出，IBES - XGBoost 的各项评价指标均为最优。相比于 XGBoost，IBES - XGBoost 的 RMSE、MAE 和 MAPE 分别降低了 0.271 m/s、0.085 m/s 和 2.215%，拟合优度提升了 0.012。相比于 LSTM，IBES - XGBoost 的 RMSE、MAE 和 MAPE 分别降低了 0.746 m/s、0.419 m/s 和 21.439%，拟合优度提升了 0.055。相比于 BES - XGBoost、GWO - XG-

Boost 和 PSO - XGBoost，IBES - XGBoost 在各项评价指标上均有不同幅度的提升。

虽然 IBES - XGBoost 各项评价指标均为最优，但是我国高速铁路采用分级限速模式，对于不同的风速采取不同的限速措施。因此还需对预测结果分区间进行统计分析，判断预测结果与实际值是否处于同一区间。

表 6 高速列车在不同风速下的限速标准

风速/(m/s)	限速值/(km/h)
<15	正常速度运行
(15, 20]	≤300
(20, 25]	≤200
(25, 30]	≤120
>30	禁止运行

根据(表 6)我国高速列车在不同风速下的限速标准^[24]。标准，对各模型预测结果进行统计，判断预测值和真实值是否处于同一风速区间。由统计结果表 7 可知，IBES - XGBoost 的预测值与真实值处于同一风速区间的准确率最高，其准确率高达 99.728%，优于其余预测模型。主要原因在于，IBES - XGBoost 预测精度较高，对诸多接近风速区间边界值的数据进行预测时能有较高的效果。因此，IBES - XGBoost 根据高速铁路业务需求的逐分钟风速提前三步预测的预测结果，在高速铁路风速限速标准下能得到最高的准确率。

表 7 各模型预测值与真实值处于同一风速区间的准确率

预测模型	准确率/%
IBES - XGBoost	99.728
BES - XGBoost	98.69
GWO - XGBoost	98.51
PSO - XGBoost	98.555
XGBoost	98.554
LSTM	96.883

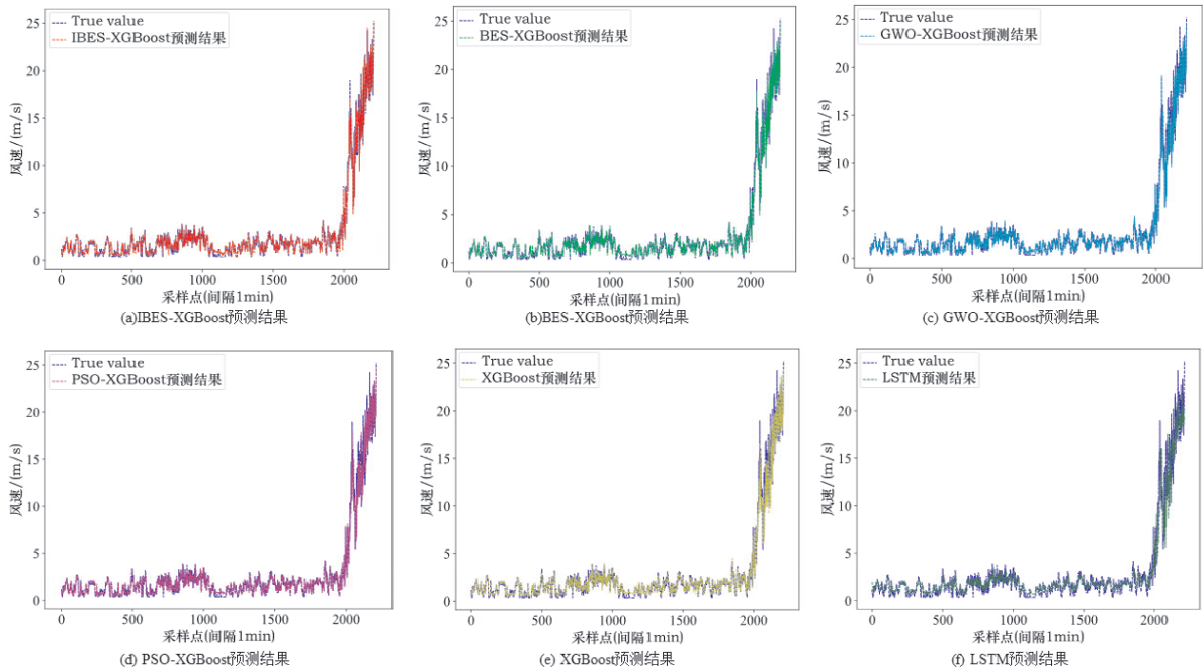


图 6 各模型提前 3 步预测结果对比

5 结论与讨论

本文从高速铁路风速预测的需求入手, 采用 IBES - XGBoost 对逐分钟风速数据进行提前三步预测。采用 Tent 混沌映射和 BFGS 拟牛顿法改进秃鹰搜索算法, 能够有效提升秃鹰搜索算法的性能, 进而提升对 XGBoost 超参数的优化能力。在本文的对比实验中, IBES - XGBoost 的各项评价指标均优于所对比模型, 具有更高的预测精度和拟合能力, 验证了所提模型的有效性。根据高速列车在不同风速下的限速标准对各模型预测结果分析可得, IBES - XGBoost 具有极高的准确率, 因此该方法能为高速铁路沿线风速预测提供参考。

参考文献:

- [1] 包云, 李亚群, 马祯, 等. 基于监测数据挖掘的高速铁路大风危险性评估研究[J]. 铁道运输与经济, 2021, 43(9): 82 - 87.
- [2] 潘新先. 兰新铁路第二双线大风环境下动车组“晃车”现象分析[J]. 铁道运输与经济, 2016, 38(12): 64 - 68.
- [3] 韩鹏, 郭桂祯, 孙宁, 等. 广东省台风灾害时空格局及影响因素研究[J]. 灾害学, 2022, 37(3): 112 - 117.
- [4] 李德仓, 陈晓强, 孟建军, 等. 恶劣环境下铁路行车安全研究综述[J]. 交通信息与安全, 2022, 40(4): 26 - 37.
- [5] XU X, WEI Y. An ultra - short - term wind speed prediction model using LSTM and CNN[J]. Multimedia Tools and Applications, 2022, 81(8): 10819 - 10837.
- [6] TUERXUN W, XU C, GUO H, et al. An ultra - short - term wind speed prediction model using LSTM based on modified tuna swarm optimization and successive variational mode decomposition[J]. Energy Science & Engineering, 2022, 10(8): 3001 - 3022.
- [7] LI J, SONG Z, WANG X, et al. A novel offshore wind farm typhoon wind speed prediction model based on PSO - Bi - LSTM improved by VMD[J]. Energy, 2022, 251: 123848.
- [8] SHEN Z, FAN X, ZHANG L, et al. Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network[J]. Ocean Engineering, 2022, 254: 111352.
- [9] BOMMIDI B S, KOSANA V, TEEPARTHI K, et al. A hybrid ap-

- proach to ultra short - term wind speed prediction using CEEM-DAN and Informer[C]//2022 22nd National Power Systems Conference (NPSC). IEEE, 2023: 207 - 212.
- [10] 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.
 - [11] 林秋婷, 刘建军, 逢辉, 等. 应用 MIPSO - XGBoost 算法预测汽油产率[J]. 石油学报(石油加工), 2023, 39(3): 659 - 667.
 - [12] 王瑞. 高速铁路灾害及超限监测关键技术研究[D]. 北京: 中国铁道科学研究院, 2022.
 - [13] CHEN T Q, GUESTIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 13 - 17, 2016, San Francisco, California, USA. New York: ACM, 2016: 785 - 794.
 - [14] ALSATTAR H A, ZAIDAN A A, ZAIDAN B B. Novel meta - heuristic bald eagle search optimisation algorithm[J]. Artificial Intelligence Review, 2020, 53(3): 2237 - 2264.
 - [15] FAN J H, LI Y, WANG T. An improved African vultures optimization algorithm based on tent chaotic mapping and time - varying mechanism[J]. PLoS One, 2021, 16(11): e0260725.
 - [16] CARRASCO - OLIVERA D, MORALES C, VILLAVICENCIO H. Stability and expansivity of tent map[J]. Proceedings of the American Mathematical Society, 2021, 149(2): 773 - 786.
 - [17] XIA X, GUI L, HE G, et al. A hybrid optimizer based on firefly algorithm and particle swarm optimization algorithm[J]. Journal of computational science, 2018, 26: 488 - 500.
 - [18] XIAO L, SHAO W, YU M, et al. Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting[J]. Applied energy, 2017, 198: 203 - 222.
 - [19] RAJAKUMAR R, SEKARAN K, HSU C H, et al. Accelerated grey wolf optimization for global optimization problems[J]. Technological Forecasting and Social Change, 2021, 169: 120824.
 - [20] NEZHAD A M, SHANDIZ R A, JAHROMI A E. A particle swarm - BFGS algorithm for nonlinear programming problems[J]. Computers & operations research, 2013, 40(4): 963 - 972.
 - [21] 赵振宇, 马旭, 包格日乐图. 基于风速预测模型的风电一次调频仿真研究[J]. 系统仿真学报, 2022, 34(10): 2233 - 2243.
 - [22] 姚吴秋实, 任雪娟, 邵典伟, 等. 东亚冬季风寒湿度的气候特征分析[J]. 气象科学, 2022, 42(6): 721 - 729.
 - [23] JAGER D, ANDREAS A. NREL national wind technology center (NWTC): M2 tower; boulder, Colorado (data) [R]. National Renewable Energy Lab. (NREL), Golden, CO (United States), 1996.
 - [24] 包云, 高歌, 李亚群. 基于监测数据挖掘的高铁气象灾害风险评估方法研究[J]. 灾害学, 2022, 37(2): 44 - 48, 53.

Wind Speed Prediction Model Along High - speed Railways Based on IBES - XGBoost

MENG Jianjun^{1,2,3}, JIANG Xiangjun¹, MENG Gaoyang⁴, LI Decang^{1,2,3}

(1. Institute of Mechanical and Electrical Technology, Lanzhou Jiaotong University, Lanzhou 730070, China;

2. Gansu Provincial Engineering Technology Center for Informatization of Logistics and Transport Equipment, Lanzhou 730070, China; 3. Gansu Provincial Industry Technology Center of Logistics and Transport Equipment, Lanzhou 730070, China;

4. School of Mechanical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: To ensure the high timely and effective wind disaster warning information along the high - speed railway, it is necessary to conduct advanced multi - step prediction of short - term wind speeds along the high - speed railway. To address the potential significant errors in numerous prediction models, the Tent chaos mapping and BFGS Quasi - Newton method are employed to improve the bald eagle search algorithm. The improved bald eagle search algorithm (IBES) is then used to optimize the initial parameters of XGBoost. When constructing the IBES - XGBoost model, additional meteorological features and apart from wind speed and are incorporated to enhance the prediction accuracy. Experimental results demonstrate that the improved bald eagle search algorithm exhibits its superior optimization capabilities compared to other intelligent optimization algorithms. In comparison to other models, IBES - XGBoost achieves higher accuracy and better fitting effects in the advanced multi - step prediction of short - term wind speeds. From the experimental findings, the following conclusions can be drawn: the Tent chaos mapping and BFGS Quasi - Newton method can effectively improve the bald eagle search algorithm, and IBES - XGBoost can provide reliable advanced multi - step prediction results for wind warnings under high - speed railway standards.

Keywords: high speed railway; wind disaster; wind speed forecast; machine learning; bald eagle search algorithm