

引用格式: 高盛, 许沛华, 陈正洪. 基于机器学习的风电场风速多模式集合预报 [J]. 南方能源建设, 2024, 11(1): 85-95. GAO Sheng, XU Peihua, CHEN Zhenghong. Wind speed multi-mode ensemble forecasting for wind farms based on machine learning [J]. Southern energy construction, 2024, 11(1): 85-95. DOI: 10.16516/j.ceec.2024.1.09.

基于机器学习的风电场风速多模式集合预报

高盛, 许沛华[✉], 陈正洪

(湖北省气象服务中心, 湖北 武汉 430205)

摘要: [目的] 随着大量风电场的兴建, 组合研究不同的机器学习算法和气象预报模式已成为研究焦点。[方法] 文章以湖北省风能资源的空间分布特征为基础, 通过选取代表站点结合实验数据分析对结果进行深入探讨。[结果] 在湖北省, 已建和在建的风电场主要集中在“三带一区”的区域, 具体包括: 位于湖北省中部, 从荆门至荆州的南北向风带; 位于鄂北, 从枣阳至英山的东西向风带; 部分湖岛和沿湖地带; 以及鄂西南和鄂东南的部分高山地区。该研究采用4种不同的数值预报产品, 包括CMA-WSP、CMA-GD、WHMM和EC, 与实测风速对比深入探究这些数值模式的适用范围。[结论] 通过分析基于机器学习的5种集合预报方法及均值法在湖北省各地区的表现确定了适合的算法和预报模式组合, 为提高集合预报的准确性提供了参考。

关键词: 风功率预测; 机器学习算法; 随机森林; LightGBM; AdaBoost; GRU; LSTM; 集合预报

中图分类号: TK89; TM614

文献标志码: A

文章编号: 2095-8676(2024)01-0085-11

DOI: 10.16516/j.ceec.2024.1.09

OA: <https://www.energychina.press/>



论文二维码

Wind Speed Multi-Mode Ensemble Forecasting for Wind Farms Based on Machine Learning

GAO Sheng, XU Peihua[✉], CHEN Zhenghong

(Hubei Provincial Meteorological Service Center, Wuhan 430205, Hubei, China)

Abstract: [Introduction] With the extensive construction of wind farms, the combination of researches on different machine learning algorithms and meteorological forecasting modes has received widespread attention. [Method] This paper was based on the spatial distribution characteristics of wind energy resources in Hubei Province, and utilized representative stations in combination with experimental data analysis to conduct in-depth discussions on the results. [Result] The wind farms in operation and under construction in Hubei Province are all located in the "Three Zones and One Area", including the north-south wind zone from Jingmen to Jingzhou in the central part of Hubei Province, the east-west wind zone from Zaoyang to Yingshan in the north of Hubei Province, certain lake islands and zones along the lake, as well as some high mountainous areas in the southwest and southeast of Hubei Province. This research uses four different numerical forecasting products, namely CMA-WSP, CMA-GD, WHMM, and EC, to compare with the measured wind speeds and investigated the applicable range of these four numerical modes. [Conclusion] By analyzing the performance of five ensemble forecasting methods based on machine learning and the mean method, we identified suitable algorithm and forecasting model combinations, providing references for improving the accuracy of ensemble forecasting.

Key words: wind power day-ahead forecasting; machine learning algorithms; random forest; LightGBM; AdaBoost; GRU; LSTM; ensemble forecasting.

2095-8676 © 2024 Energy China GEDI. Publishing services by Energy Observer Magazine Co., Ltd. on behalf of Energy China GEDI. This is an open access article under the CC BY-NC license (<https://creativecommons.org/licenses/by-nc/4.0/>).

收稿日期: 2023-12-01 修回日期: 2023-12-20

基金项目: 湖北省气象局科研基金资助项目“基于多种深度学习组合模型的光伏发电功率超短期预测准确率提升研究”(2023Q13)

0 引言

中国致力于发展可再生能源、实施化石能源清洁替代,到 2030 年使非化石能源占一次能源消费的比重达到 25% 左右,其中风力发电是重要组成部分,但大规模风电并网对电力系统的运行和调度带来挑战。风电的随机性、波动性和间歇性等固有特点严重影响了其并网利用。为了保证电网的安全稳定运行,华中能源局华中监管局 2019 年发布了《关于印发华中区域“两个细则”的通知》,该规定要求风电场日前短期预报准确率要达到 80% 以上^[1]。不能满足上述预报准确率要求的电站的售电经济收益将受到影响。部分电站因功率预报不准确,全年发电量的 3%~5% 无法对电站产生经济价值,严重影响了电站的经济效益^[2]。甚至有电站因预报准确率不高出现亏损的严重情况。因此准确稳定的日前风电功率预报在风力发电运行过程中起着主导作用。

1) 国内外研究现状

国内外传统短期风电功率预报方法主要包括基于数值天气预报数据的物理方法和统计方法^[3-4]。物理方法通过数值天气预报数据计算风电场的预计风速,再带入风电场的风功率曲线进行拟合从而得到最终发电功率;而统计方法则基于历史数据和实时数据进行统计计算预报,比较常见的包括动力统计法,它在数值预报结果和风电场的风电功率之间建立一种映射关系,包括线性以及非线性方法来获得预报结果。在实际应用中,这些方法存在数据不完备、自动化通讯设备故障、风电出力限制等问题,这些问题都会影响预报结果^[4]。

随着机器学习技术的普及,越来越多的研究开始探讨如何将机器学习普及。机器学习算法能够从大规模的气象数据中学习规律,并能够适应不同的气象条件,因此具有潜在的优势^[5]。随着机器学习在各个领域的广泛应用,预测算法的性能和效果成为近期研究的关注点。本文旨在深入研究和对比多种基于机器学习的预测算法,包括随机森林算法(Random Forest, RF)^[6]、基于决策树的集成算法(Light Gradient Boosting Machine, LightGBM)^[7]、自适应增强算法 AdaBoost^[8]、机器学习算法以及基于深度神经网络的(Gated Recurrent Unit, GRU)^[9]、(Bidirectional Long Short Term Memory, Bi-LSTM)^[10]等

方法。在过去的研究中,这些算法在不同任务和数据集上表现出色,但其性能差异和适用场景仍需深入了解^[4,11]。

2) 本文研究内容

本文着重选取位于湖北省内多个典型风电场,针对这些场站数据开展集合预报算法的比较与研究,得出适用于湖北省各个区域的数值预报模式以及在湖北省表现比较好的预报方法。在文章中,通过选择重点试验场站并检验对比各种集合预报方法在试验场站的结果,通过仔细分析结果数据从而得到本文的研究结论数据。最终由预报风速的对比分析得出适用于不同地区场站的集合预报方法。

1 研究数据选取

本研究采用了陈正洪等^[12-13]研究中被证明在湖北省较为有效的 4 种不同的数值预报产品,包括 CMA-WSPV1.0、CMA-GD、WHMMV2.0 和 EC 数值预报^[14]。其中, CMA-WSP V1.0 是由中国气象局风能太阳能中心下发的面向风能和太阳能的数值天气预报^[12],水平分辨率为 3 km × 3 km(下简称为 3 km),预报时长为未来 72 h。CMA-GD 是由广东省区域数值天气预报重点实验室基于 GRAPES 非静力模式^[15]开发的华南区域中尺度模式,水平分辨率为 3 km,预报时长为未来 72 h。WHMM V2.0 是由中国气象局武汉暴雨所提供,华中区域中尺度数值天气预报系统以广泛使用的中尺度数值天气模式 WRF 为基础。模式区域中心位于(114.133°E,30.617°N),采用 3 层嵌套包括全国、华中、湖北省 3 个区域,其分辨率分别为 27 km、9 km 和 3 km。本项目中采用湖北区域的预报,预报时效为 84 h。EC 预报采用数据来自欧洲气象中心的再分析资料,资料水平分辨率为 1 km,每天模拟时效为 72 h。在本研究中通过网格间隔选取数据点的方式得到水平分辨率为 3 km 的数据进行使用。此外,本研究还采用了基于机器学习的 5 种集合预报方法,在研究中对比以上 5 种方法和直接采用数值预报模式及通过选取多种数值预报模式在预报地点的预报结果取平均值的方法(简称为均值法)进行对比。

为了方便后续对比,本文选取的 4 种模式均选取相同的模式水平分辨率和预报时效。本文统一使用的模式分辨率为 3 km,时间分辨率为 15 min。对

于分辨率不同的预报模式, 如果该预报模式有水平分辨率为 3 km 的数据源则采用该数据源, 如果该预报模式的水平分辨率比 3 km 更高则采用间隔取网格数据点的方式将水平分辨率降为 3 km 后进行对比。所有数据每日 08:00(北京时间, 下同)起报, 仅采用未来 72 h 的预报结果进行使用。在根据模式计算场站实际格点数据的时候采用线性插值的方式计算从网格数据到格点的具体数值。

本文使用的资料年限为 2021 ~ 2022 年两年的数据, 主要种类如下: 7 个风电场的测风塔观测数据, 时间分辨率为 15 min, 7 个风电场基本情况如表 1 所示。

所选风电场在湖北省的地图位置如图 1 所示。图中颜色代表所在地区 70 m 高度层的年平均风速,

表 1 7 个代表性风电场基本信息表

Tab. 1 Basic information of seven representative wind farms

编号	风电场名称	轮毂高度/m	海拔高度/m	所属地区
1	板桥风电场	140	1510	恩施板桥
2	周楼风电场	140	540	襄阳枣阳
3	象河风电场	140	247	荆门东宝
4	黄龙山风电场	140	793	咸宁通山
5	大金中部风电场	140	172	黄冈武穴
6	天河口风电场	140	850	随州随县
7	蔡家寨风电场	140	500	黄冈麻城

越接近红色代表年, 平均风速越高, 越接近蓝色代表年, 平均风速越低。

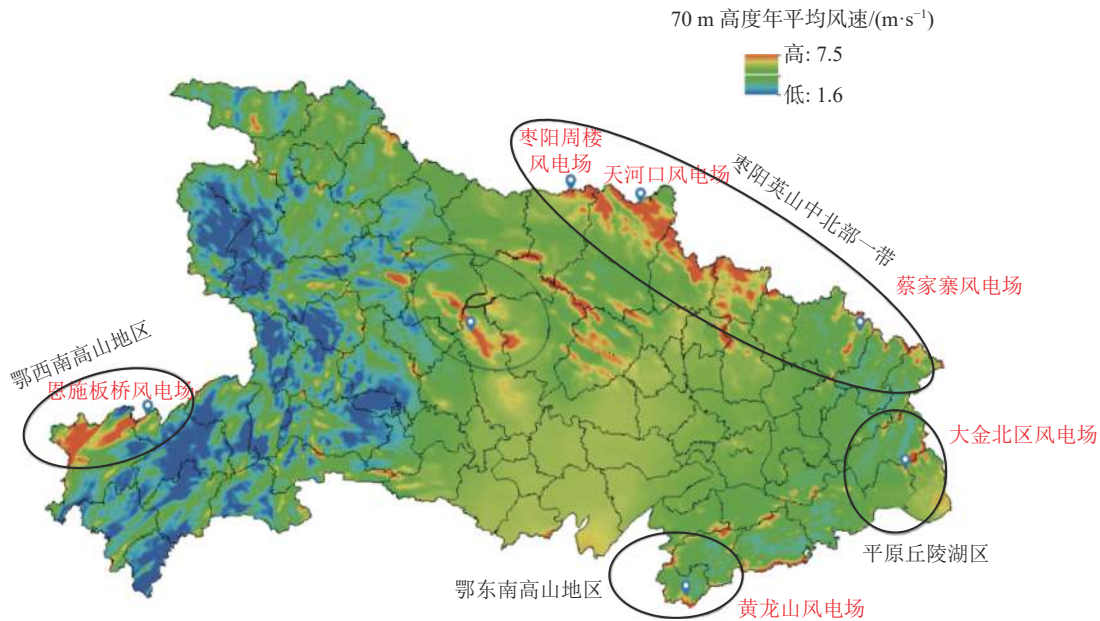


图 1 湖北省典型风电场选取分布示意图

Fig. 1 Distribution diagram of typical wind farms selected in Hubei Province

2 机器学习及集合预报方法

2.1 RF 算法

RF 算法是随机训练决策树的集合, 被广泛应用于解决分类和回归问题, 随机森林回归模型是一种集成方法, 它结合了各种不相关的回归树, 减轻了每棵树的不稳定性问题, 每棵树基于随机抽样的训练数据和随机选择的特征进行构建。在分类任务中,

每棵树对样本进行分类, 最后通过投票机制确定最终的分类结果。对于回归任务, 各树的输出取平均值。该算法在数据和特征的选择上体现出随机性, 使得每棵树都是独立且略有不同的, 提高了整个模型的泛化能力。

2.2 Adaboost 算法

Adaboost 是一种集成学习算法, 旨在通过组合多个弱分类器来构建一个更强大的分类器。其核心

思想是通过迭代训练,每一轮关注被前一轮分类错误的样本,为其分配更高的权重,从而集中处理难以分类的样本。新的分类器以加权投票的方式与之前的分类器进行组合。AdaBoost 算法的主要思想是将每次迭代产生的弱学习者结合起来,形成一个强学习分类器。该算法通过迭代,每次迭代都关注之前模型分类错误的样本,提高其权重。在每个迭代中,构建一个基础分类器,最终通过加权投票将这些基础分类器组合成一个强分类器。样本的权重会被调整,以便后续模型更加关注先前分类错误的样本,从而逐步改善整体性能。

2.3 LightGBM 算法

LightGBM(也简称为 LGBM)是于 2016 年提出的一种基于决策树 GBDT(Gradient Boosting Decision Tree)的梯度提升算法^[7]。它通过迭代地训练决策树,每次迭代都试图纠正上一轮迭代的错误。采用直方图算法,将连续的特征值分桶,然后在桶上进行分裂,大幅提高了训练速度。

LightGBM 的核心思想是基于直方图的决策树算法,将样本中连续的浮点特征值离散化成 K 个整数并构造与之长度相等的直方图。遍历时,将离散化后的值作为索引在直方图中累计统计量,然后根据直方图的离散值,遍历寻找最优的分割点。这样可以有效地降低内存消耗,同时达到降低时间复杂度的目的。LightGBM 在处理大规模数据集时具有较好的性能,并且能够处理高维稀疏特征,还支持并行化训练。

基于 LightGBM 算法能够并行处理海量数据的特性,将该算法用于对时间序列的残差和风速、温度、湿度、气压等多种参数进行多特征并行处理,能够更好地降低模型计算的时间复杂度,提高预测的效率和精度。

2.4 GRU 算法

GRU 是一种门控循环神经网络。它包括更新门、重置门和候选隐藏状态。这些门控制着信息的选择性传递和遗忘。GRU 旨在解决传统 RNN 中的梯度消失问题,能够更好地捕捉序列数据中的长期依赖关系。由于其门控机制,特别适用于处理序列数据,如时间序列数据^[6]。GRU 算法的核心公式见式(1):

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\ h_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * h_t \end{aligned} \quad (1)$$

式中:

z_t ——代表更新门;

r_t ——代表重置门;

h_t ——代表候选隐藏层细胞状态;

h_t ——最终的隐藏状态。

这个公式描述了 GRU 网络的控机制和细胞状态更新方式,通过控制信息的流动从而可以生成合理的时间序列处理输出结果。

2.5 Bi-LSTM 算法

双向 LSTM(Bi-LSTM)由两个 LSTM 单元组成,它考虑了过去和未来的输入特性。相比 GRU,它通过引入门控机制和双向 LSTM 单元更有效地解决了长期依赖性的问题^[7]。LSTM 的核心原理是通过一系列的门控单元来控制信息的流动和遗忘。具体而言,LSTM 包括输入门、遗忘门和输出门 3 个关键组件。输入门决定了当前输入信息的重要程度,遗忘门控制了前一时刻的记忆是否被保留,输出门决定了当前记忆的输出。LSTM 算法的核心计算公式见式(2):

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (2)$$

式中:

i_t ——代表输入门;

f_t ——代表遗忘门;

c_t ——代表隐藏层细胞状态;

o_t ——输出门的输出;

h_t ——最终的隐藏状态。

Bi-LSTM 通过将数据双向输入可以进一步提升长期依赖数据的利用效果,使用 Bi-LSTM 可以捕捉过去和未来状态的影响。Bi-LSTM 的计算公式见式(3):

$$\begin{aligned} \vec{h}_t &= \text{LSTM}(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t &= \text{LSTM}(x_t, \overleftarrow{h}_{t-1}) \\ h_t &= \vec{h}_t + \overleftarrow{h}_t \end{aligned} \quad (3)$$

式中:

\vec{h} ——正向隐藏状态;

\overleftarrow{h} ——反向的隐藏状态。

这种受控存储是长短期记忆网络和门控循环单元的基础,可以缓解梯度爆炸和消失等问题。长短期记忆神经网络在时序序列数据中具有良好的性能。

2.6 集合预报算法

针对复杂转折性天气,单一数值模式预报产品无法有效提高预报准确率^[18],采用多种集合预报方法进行对比检验是一种比较好的方法^[19]。集合预报算法是一种通过将多种数值预报模式数据进行对比优选,再输入给机器学习方法或传统预报方法进行检验,最终输出表现最优的预报结果的方法。其主要特征在于针对每种预报算法,在集合预报过程中都会通过对比优选数值预报数据源进行预报检验得到最终输出。许杨等^[3,11,13]在过去的研究中提出了基本的集合预报思想,本研究采用的集合预报方法核心思想与之前的研究保持一致,依然是通过多种预报源与算法进行对比的方式得到较好的结果。本研究在对比过程中通过数据组合计算,并讨论分析集合预报的表现来探讨适合于湖北省的集合预报算法。本研究采用的集合预报系统流程如图 2 所示。

为确保实验的全面性和可信度,本研究采用了多个数据集和场景,并综合考虑了算法的准确性、泛化能力和计算效率等方面的指标。在本研究中,所采用的评价指标包括主观评价指标与客观评价指标两种。其中主观评价指标主要是针对预报结果绘制成的预报曲线进行人工交叉检验对比,客观评价指标主要采用平均均方根误差计算、算法运行时间计算等指标进行集合预报的算法评价。

3 实验设计

本研究实验使用 RF、LightGBM、AdaBoost、GRU、Bi-LSTM 共 5 种先进的机器学习进行集合预报,并与数值预报模式方法及根据场站优选数值预报模式取均值的方法进行对比检验,根据对比检验结果优

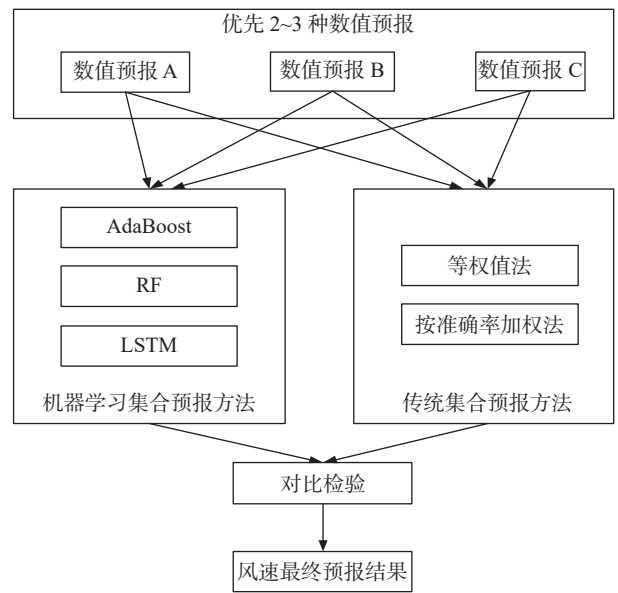


图 2 集合预报流程图

Fig. 2 Flow chart of ensemble forecasting

选准确率较高的方法。

3.1 实验平台

为公平对比,实验平台统一使用同一台计算机。实验计算机配置为 1 颗英特尔 Xeno Gold 6330 CPU,内存大小为 128 GB,所用操作系统为银河麒麟 V10 Linux 操作系统,Python 版本为 Python 3.7,深度学习算法使用 Pytorch 1.8 框架。

3.2 实验训练方法

所有基于机器学习的集合预报实验均采用被经常采用的 10 折交叉验证^[20]来验证效果。具体而言,研究中将整个数据集分为 10 个子集,其中 8 个用作训练数据,1 个用作测试数据,1 个用作验证数据。我们重复这个过程 10 次,每次选择不同的子集作为测试及验证数据,并计算模型在每次验证集上的性能指标,如平均误差、均方根误差。最终,我们对这 10 次评估结果进行平均,得到模型在整个数据集上的性能评估。

对于机器学习算法 RF、LightGBM、AdaBoost,使用网格搜索方式来确定最优超参数组合,它的特征在于首先确定每个对结果影响较大的超参数的值域空间,在该空间内等间隔生成超参数值并与其他超参数进行组合形成超参数空间进行实验搜索。对于 RF 方法最核心的超参数为最大深度(max_depth),和弱学习器个数(n_estimators);对于 Adaboost 最核

心的超参数为学习率(lr)及弱学习器个数($n_estimators$);对于LightGBM最核心的超参数包括学习率(lr),最大深度(max_depth)、弱学习器个数($n_estimators$)、列采样比例(feature_fraction);在实验过程中,通过搜索超参数组合确定对当前机器学习算法最优的学习配置。

对于深度学习算法,使用动态学习率调整方式根据训练曲线动态降低学习率避免过拟合,同时引入早停训练机制,当连续5个epoch在测试集上的效果不再下降的时候停止训练,避免过拟合。二者组合使用解决了epoch数量需要手动设置或设置不合理的问题,同时可以保证针对各种算法均在一定范围内可以体现该算法的最优效果。

对于均值法采用同一时刻的多个数值预报模式通过线性插值算法将格点预报转换到站点预报的风速值求平均值计算的算法进行比较。

在以上算法中,为了公平评估各个集合预报方法的性能。由于均值法是直接使用数值预报的结果因此直接输出的结果只有风速,其他机器学习算法都可以通过配置学习数据直接生成功率预报结果。为了公平比较,所有预报方法都只预报风速,而不是对比预报功率准确性。因为从风速转换到发电功率可以使用不同的公式计算,不同的公式选择会对结果产生微小的影响,这对于评估均值法和其他算法的差异情况有不利影响。本研究通过对比预报风速的偏差来比较不同算法的实际效果,对于后续应用评估功率预报的准确性也具有指导意义。

3.3 实验结果

为有效评估实验结果,实验采用平均均方根误差RMSE指标确定不同集合预报方法的偏差, RMSE计算公式见式(4):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

式中:

y ——预报风速(m/s);

\hat{y}_i —— i 时刻的真实风速(m/s)。

表2是各集合预报算法均方根误差逐月对比结果,在此表中集合预报的时候使用了全部预报成员进行预报,控制算法变量不同从而对比不同的预报算法的优劣。表3是各集合预报算法均方根误差按

年对比的情况,其中列“集合成员1”和列“集合成员2”是分别优选2个不同的集合成员(见电场名称列,列中名称从左到右分别对应集合成员1和集合成员2)进行预报后得到的预报结果。

通过表2~表3、图3进行对比分析发现,在所有风电场中,传统均值法预报误差最大。在象河风电场,Adaboost集合预报算法的均方根误差最小,在天河口风电场、周楼风电场、蔡家寨风电场深度学习算法GRU和Bi-LSTM算法优于传统机器学习算法,在周楼风电场和蔡家寨风电场GRU算法优于Bi-LSTM,在天河口风电场Bi-LSTM优于GRU算法。

在天河口风电场集合预报算法RF、AdaBoost、LightGBM、GRU和Bi-LSTM、均值法的均方根误差分别为2.44 m/s、2.25 m/s、2.41 m/s、1.94 m/s、1.90 m/s,在5种算法中,Bi-LSTM算法最优,较均值法误差降低0.73 m/s。Bi-LSTM集合算法相较于集合成员EC预报均方根误差3.24 m/s下降了1.34 m/s,CMA-WSP预报均方根误差3.03 m/s下降了1.13 m/s。分析发现EC预报与CMA-WSP均方根误差较大,但是从图形检验来看CMA-WSP预报方法的偏差相对较稳定,因此误差下降较大。

在周楼风电场GRU算法最优,均值法预报误差最大,较均值法均方根误差降低了0.33 m/s,Bi-LSTM与GRU算法误差接近。其次是RF算法较均值法下降0.1 m/s,AdaBoost算法较均值法下降了0.06 m/s。最优的GRU集合预报算法相较于集合成员EC预报均方根误差1.8 m/s下降了0.06 m/s,相较于CMA-GD预报均方根误差2.42 m/s下降了0.68 m/s。

在象河风电场AdaBoost最优,较均值法均方根误差降低了0.52 m/s,传统均值法预报误差最大。其次是,RF较均值法误差降低了0.46 m/s,LightGBM算法较均值法误差降低了0.42 m/s。深度学习算法GRU和Bi-LSTM的误差相同,均为1.52 m/s。集合预报算法AdaBoost相较于集合成员CMA-WSP均方根误差3.14 m/s下降了1.71 m/s,EC预报均方根误差1.67 m/s下降了0.24 m/s。

在蔡家寨风电场GRU算法误差最小,较均值均方根误差降低了0.36 m/s,Bi-LSTM比GRU误差略大,其次AdaBoost误差较均值法降低了0.14 m/s,LightGBM误差较均值法降低了0.11 m/s,RF较均值

表 2 各集合预报算法均方根误差逐月对比

Tab. 2 Monthly comparison of root mean square errors of ensemble forecasting algorithms

m/s

电站名称	年	月	RF	AdaBoost	LightGBM	GRU	Bi-LSTM	均值法
天河口风电场	2022	4	3.11	2.85	2.69	1.95	1.98	2.93
天河口风电场	2022	5	2.47	2.08	2.20	1.89	1.84	2.84
天河口风电场	2022	6	2.71	2.61	2.99	2.04	2.19	3.27
天河口风电场	2022	7	2.11	2.09	2.28	1.73	1.83	2.78
天河口风电场	2022	8	3.05	3.09	3.04	2.15	2.13	2.19
天河口风电场	2022	9	2.28	1.90	2.24	1.54	1.58	1.89
天河口风电场	2022	10	2.30	2.20	2.31	2.11	2.24	2.65
天河口风电场	2022	11	1.85	1.23	1.88	1.47	1.40	2.07
天河口风电场	2022	12	2.03	1.67	2.01	1.67	1.89	2.61
天河口风电场	2023	1	2.16	2.09	2.12	1.93	1.88	2.71
天河口风电场	2023	2	2.37	2.23	2.30	2.27	2.22	2.74
周楼风电场	2022	4	2.14	2.09	2.19	1.99	1.92	2.05
周楼风电场	2022	5	1.75	1.63	1.83	1.57	1.49	1.8
周楼风电场	2022	6	1.95	1.83	1.88	1.68	1.64	1.93
周楼风电场	2022	7	1.92	1.87	1.65	1.52	1.56	1.98
周楼风电场	2022	8	2.00	2.00	1.89	1.72	1.73	2.12
周楼风电场	2022	9	1.79	1.73	1.75	1.48	1.51	2.08
周楼风电场	2022	10	1.92	2.09	1.94	2.00	2.00	2.57
周楼风电场	2022	11	1.75	1.95	1.68	1.95	1.86	2.1
周楼风电场	2022	12	2.05	2.04	2.02	1.96	1.91	2.15
周楼风电场	2023	1	2.24	2.48	2.13	1.74	1.75	1.98
周楼风电场	2023	2	2.14	2.39	2.10	1.62	1.68	1.87
象河风电场	2022	4	1.51	1.50	1.52	1.58	1.56	2.03
象河风电场	2022	5	1.30	1.38	1.45	1.43	1.42	1.78
象河风电场	2022	6	1.50	1.52	1.64	1.64	1.67	1.95
象河风电场	2022	7	1.32	1.37	1.52	1.51	1.56	1.82
象河风电场	2022	8	1.37	1.35	1.49	1.58	1.64	2.06
象河风电场	2022	9	1.34	1.34	1.59	1.51	1.61	1.78
象河风电场	2022	10	1.37	1.38	1.58	1.56	1.54	2.01
象河风电场	2022	11	1.22	1.28	1.37	1.44	1.38	1.91
象河风电场	2022	12	1.49	1.45	1.64	1.64	1.58	2.14
象河风电场	2023	1	1.46	1.50	1.61	1.40	1.39	1.96
象河风电场	2023	2	1.47	1.53	1.46	1.44	1.43	1.97
蔡家寨风电场	2022	4	1.89	1.79	1.85	1.75	1.71	2.17
蔡家寨风电场	2022	5	1.59	1.73	1.65	1.63	1.60	1.91
蔡家寨风电场	2022	6	2.46	2.40	2.36	2.33	2.27	2.23
蔡家寨风电场	2022	7	2.02	1.92	1.91	1.65	1.65	1.76
蔡家寨风电场	2022	8	1.62	1.54	1.66	1.38	1.39	1.74
蔡家寨风电场	2022	9	1.66	1.54	1.56	1.40	1.41	1.68
蔡家寨风电场	2022	10	1.73	1.82	1.71	1.26	1.27	1.94

表 2 (续) 各集合预报算法均方根误差逐月对比

Tab. 2(Cont.) Monthly comparison of root mean square errors of ensemble forecasting algorithms m/s

电站名称	年	月	RF	AdaBoost	LightGBM	GRU	Bi-LSTM	均值法
蔡家寨风电场	2022	11	2.10	2.19	2.00	1.69	1.67	2.19
蔡家寨风电场	2022	12	2.60	2.23	2.62	2.35	2.22	2.87
蔡家寨风电场	2023	1	1.83	1.73	1.84	1.32	1.30	2.02
蔡家寨风电场	2023	2	1.87	1.75	1.81	1.32	1.31	1.89

表 3 各集合预报算法均方根误差按年对比

Tab. 3 Yearly comparison of root mean square errors of various ensemble forecasting algorithms m/s

电站名称	RF	AdaBoost	LightGBM	GRU	Bi-LSTM	集成员1	集成员2	均值法
天河口风电场 (EC,CMA-GD)	2.44	2.25	2.41	1.94	1.90	3.24	2.45	2.63
周楼风电场 (EC,CMA-GD)	1.97	2.01	1.92	1.74	1.76	1.80	2.42	2.07
象河风电场 (EC,CMA-WSP)	1.49	1.43	1.53	1.52	1.52	1.67	3.14	1.95
蔡家寨风电场 (EC,CMA-GD)	1.94	1.86	1.89	1.63	1.66	1.86	2.49	2.00

法均降低了 0.06 m/s。最优集合预报算法 GRU 相较于集合预报成员 CMA-GD 的均方根误差 2.49 m/s 下降了 0.86 m/s, 相较于 EC 均方根误差 1.86 m/s 下降了 0.23 m/s。

综上所述可以看出,在集合预报算法中,一般情况下深度学习算法 GRU 和 Bi-LSTM 优于传统机器学习算法,传统机器学习算法又优于均值法,其中 Bi-LSTM 在天河口风电场提升最为显著,较集成员 EC 预报均方根误差 3.24 m/s 下降了 1.34 m/s,较均值法均方根误差全年下降了 0.73 m/s。

3.4 实验曲线检验

为了进一步验证不同方法的实验结果,本研究进一步抽取象河风电场 2022 年 8 月作为典型月份,通过曲线比较验证不同集合预报方法的效果。绘制结果如图 4 所示,图中图例 RF_WS 代表使用 RF 方法预测的风速波动曲线,其他图例依次类推,图例 OBS_WS 代表实际风速。

从图中可以观察均值法(绿色曲线)在大部分时间对比实际风速曲线都存在相对最大的误差,AdaBoost 的预报走势在大部分时间与实际风速差值不大。这与数值分析结论基本一致。

数值分析均值法集合预报误差为 2.06 m/s, RF 集合预报误差为 1.37 m/s, LightGBM 为 1.49 m/s, 二者预报误差较接近。GRU 为 1.58 m/s, Bi-LSTM 为

1.64 m/s, AdaBoost 为 1.35 m/s, 预报性能最优,较传统均值法集合预报误差降低 0.71 m/s。这些数值与实验曲线观察到的差距基本一致。因此通过以上分析,针对象河风电场选取 AdaBoost 算法可以获得最佳的集合预报效果。

3.5 实验结果讨论

本文旨在提供对比不同机器学习算法在集合预报中性能的深入理解,为未来预测算法的选择和优化提供有益的参考。通过对比实验结果,确定了适用于湖北省不同地区的合适算法和预报模式的组合,为提高集合预报的准确性和效率提供了实际参考。通过数据分析及实验研究,本文主要有以下两个贡献:

1)通过研究对比确定了适合湖北省内“三带一区”中各个地域最适合的数值预报模式。

2)通过开展集合预报方法的对比研究,确定使用单一数值模式预报产品无法有效提高预报准确率。通过采用多种集合预报方法,使用先进的机器学习、人工智能方法进行集合预报,如使用 AdaBoost、RF、Bi-LSTM 等机器学习方法,并与传统的均值法、加权法以及熵值法等传统集合预报进行对比检验,根据对比检验结果优选一种准确率较高的方法。

4 结论

本文实验项目研制完成的不同数值预报在湖北

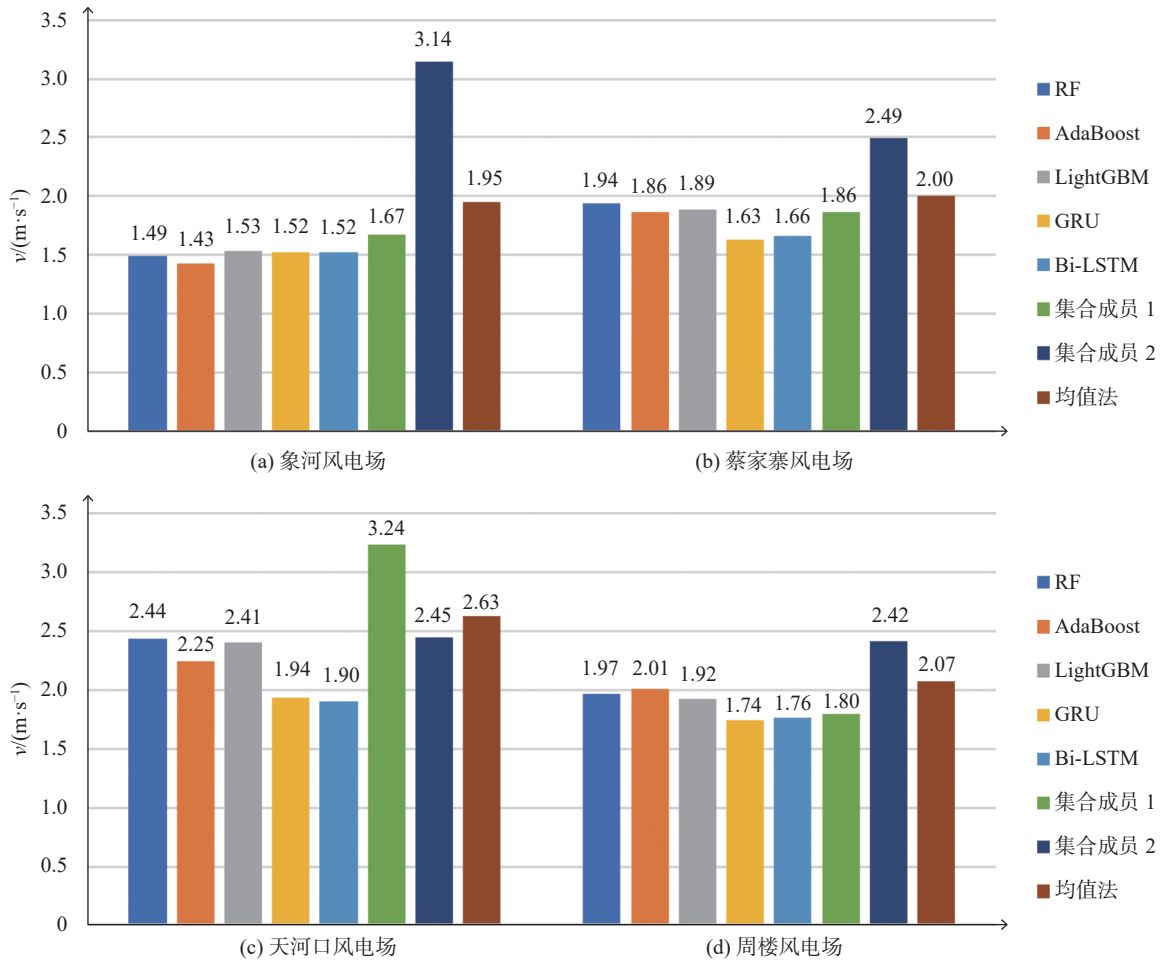


图 3 集合预报算法 RMSE 按年对比分析

Fig. 3 Yearly comparative analysis of ensemble forecasting algorithm RMSE

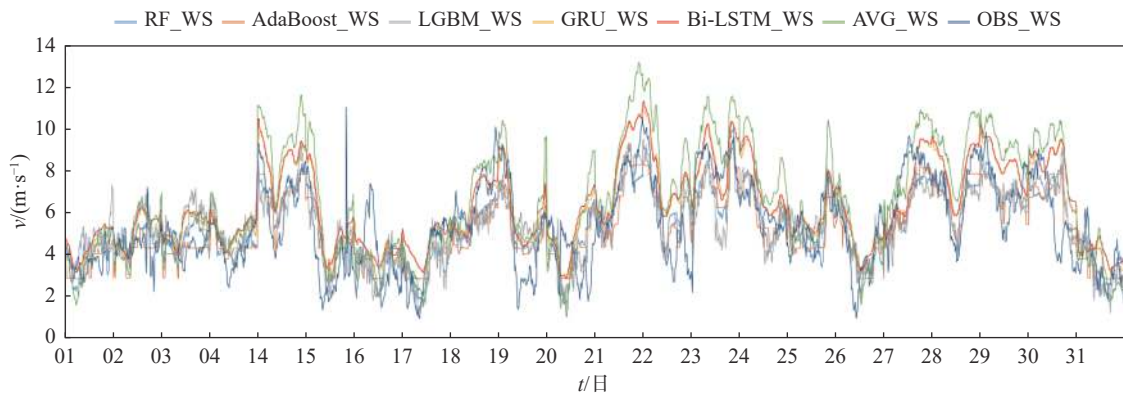


图 4 象河风电场 2022 年 8 月不同预报风速随时间波动曲线

Fig. 4 Fluctuation curve of different forecast wind speed with time in Xianghe wind farm in August 2022

省内不同区域的适用范围, 以及研究了集合预报算法技术。具体结论如下:

1) 在枣阳周楼和蔡家寨风电场 GRU 集合预报效果最优, 在天河口水风电场 Bi-LSTM 集合预报效果

最优, GRU 集合预报算法接近 Bi-LSTM, 在象河风电场 AdaBoost 集合预报效果最优。

2) GRU 和 Bi-LSTM 集合预报算法相对于 CMA-WSP 单一预报月均方根误差最大降低了 2.41 m/s,

相对于 EC 单一预报 GRU 集合预报算法月平均方根误差最大降低了 1.91 m/s, Bi-LSTM 集合预报算法月平均降低了 1.85 m/s。年平均 GRU 和 Bi-LSTM 相对于 CMA-WSP 误差降低了 1.45 m/s, 相对于 EC 误差降低了 0.42 m/s。

3) GRU 在 4 个电站的预报性能比 Bi-LSTM 和 AdaBoost 两种集合预报算法表现更为稳定, 均值法集合预报误差作为对比参照算法, 在所有集合预报算法中误差最大。

该项研究对于指导风电场功率预报有指导意义, 因此具有广阔的市场前景。如果能获得准确的风速预报, 通过风速到功率的计算公式可以计算出适用于风电场的功率预报结果。本文所使用的基于机器学习的集合预报方法通过调整训练数据的方式在未来也可以直接输出较为准确的风力发电功率预报。通过提高功率预报的准确率, 一方面可以帮助风电场减轻电网对其考核压力, 提高发电并网比例, 提升风电场的经济效益; 另一方面也有利于提高电网运行的稳定性, 有着明显的社会效益。此研究的结论在未来也可以通过类比推广到湖北省外其他地域, 对于全国风力发电多模式集合预报的准确率改进也具有指导意义。

参考文献:

- [1] 本刊编辑部. 新时代可再生能源绘出“新航标” [J]. 中国能源, 2021, 43(4): 1-2.
Editorial Department. Editorial department renewable energy in the new era draws a "new navigation mark" [J]. Energy of China, 2021, 43(4): 1-2.
- [2] 蔡绍宽. 双碳目标的挑战与电力结构调整趋势展望 [J]. 南方能源建设, 2021, 8(3): 8-17. DOI: 10.16516/j.gedi.issn2095-8676.2021.03.002.
CAI S K. Challenges and prospects for the trends of power structure adjustment under the goal of carbon peak and neutrality [J]. Southern energy construction, 2021, 8(3): 8-17. DOI: 10.16516/j.gedi.issn2095-8676.2021.03.002.
- [3] 许杨, 陈正洪, 杨宏青, 等. 风电场风电功率短期预报方法比较 [J]. 应用气象学报, 2013, 24(5): 625-630. DOI: 10.3969/j.issn.1001-7313.2013.05.012.
XU Y, CHEN Z H, YANG H Q, et al. Comparison of short-term forecast method of wind power in wind farm [J]. Journal of applied meteorological science, 2013, 24(5): 625-630. DOI: 10.3969/j.issn.1001-7313.2013.05.012.
- [4] XU P H, ZHANG M Y, CHEN Z H, et al. A deep learning framework for day ahead wind power short-term prediction [J]. Applied sciences, 2023, 13(6): 4042. DOI: 10.3390/app13064042.
- [5] 雷茂, 施炎, 刘川文, 等. 风速和发电功率预测综述 [J]. 可再生与可持续能源评论, 2009, 13: 915-920. DOI: 10.1016/J.RSER.2008.02.002.
LEI M, SHI Y, LIU C W, et al. Overview of wind speed and power generation forecast [J]. Review of renewable and sustainable energy, 2009, 13: 915-920. DOI: 10.1016/J.RSER.2008.02.002.
- [6] BREIMAN L. Random forests [J]. Machine learning, 2001, 45(1): 5-32. DOI: 10.1023/A:1010933404324.
- [7] KE G L, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [C]//Anon. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, USA, December 4-9, 2017. Long Beach: ACM, 2017: 3149-3157. DOI: 10.5555/3294996.3295074.
- [8] FREUND Y. Experiment with a new boosting algorithm [J/OL]. (1996-01-06) [2023-12-01]. http://www.researchgate.net/publication/244498683_Experiment_With_a_New_Boosting_Algorithm.
- [9] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J/OL]. (2014-12-11) [2023-12-01]. <https://arxiv.org/abs/1412.3555>.
- [10] SOMAN S S, ZAREIPOUR H, MALIK O, et al. A review of wind power and wind speed forecasting methods with different time horizons [C]//Anon. North American Power Symposium 2010, Arlington, USA, September 26-28, 2010. Arlington: IEEE, 2010: 1-8. DOI: 10.1109/NAPS.2010.5619586.
- [11] 许沛华, 陈正洪, 孙延维, 等. 湖北山区复杂地形条件下风电功率预报算法研究 [J]. 干旱气象, 2021, 39(3): 524-532. DOI: 10.11755/j.issn.1006-7639(2021)-03-0524.
XU P H, CHEN Z H, SUN Y W, et al. Research on wind power prediction algorithm under complicated terrain in mountainous area of Hubei Province [J]. Journal of arid meteorology, 2021, 39(3): 524-532. DOI: 10.11755/j.issn.1006-7639(2021)-03-0524.
- [12] 中国气象局. 太阳能光伏发电功率短期预报方法: QX/T 244—2014 [S]. 北京: 中国标准出版社, 2015.
China Meteorological Administration. Short term forecasting methods for solar photovoltaic power: QX/T 244—2014 [S]. Beijing: Standards Press of China, 2015.
- [13] 陈正洪, 许杨, 许沛华, 等. 风电功率预测预报技术原理及其业务系统 [M]. 北京: 气象出版社, 2013.
CHEN Z H, XU Y, XU P H, et al. The principle and business system of wind power prediction and forecasting technology [M]. Beijing: China Meteorological Press, 2013.

- [14] SCHMITZ J T, MULLEN S L. Water vapor transport associated with the summertime north american monsoon as depicted by ECMWF analyses [J]. *Journal of climate*, 1996, 9(7): 1621-1634. DOI: [10.1175/1520-0442\(1996\)009<1621:WVTAWT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1621:WVTAWT>2.0.CO;2).
- [15] 王光辉, 沈学顺, 胡江林, 等. GRAPES 全球非静力大气模式的正规模分析 [J]. *地球物理学报*, 2009, 52(12): 2964-2972. DOI: [10.3969/j.issn.0001-5733.2009.12.005](https://doi.org/10.3969/j.issn.0001-5733.2009.12.005).
WANG G H, SHEN X S, HU J L, et al. Normal mode analysis of the GRAPES global non-hydrostatic atmospheric model [J]. *Chinese journal of geophysics*, 2009, 52(12): 2964-2972. DOI: [10.3969/j.issn.0001-5733.2009.12.005](https://doi.org/10.3969/j.issn.0001-5733.2009.12.005).
- [16] KISVARI A, LIN Z, LIU X L. Wind power forecasting-a data-driven method along with gated recurrent neural network [J]. *Renewable energy*, 2021, 163: 1895-1909. DOI: [10.1016/j.renene.2020.10.119](https://doi.org/10.1016/j.renene.2020.10.119).
- [17] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J/OL]. (2018-04-19) [2023-12-02]. <https://arxiv.org/abs/1803.01271>. DOI: [10.48550/arXiv.1803.01271](https://doi.org/10.48550/arXiv.1803.01271).
- [18] 许沛华, 陈正洪, 谷春, 等. 风电功率预测预报系统的设计与开发 [J]. *水电能源科学*, 2013, 31(3): 166-168.
XU P H, CHEN Z H, GU C, et al. Design and development of wind power prediction system [J]. *Water resources and power*, 2013, 31(3): 166-168.
- [19] 王颖, 魏云军. 风电场风速及风功率预测方法研究综述 [J]. *陕西电力*, 2011, 39(11): 18-21,30. DOI: [10.3969/j.issn.1673-7598.2011.11.006](https://doi.org/10.3969/j.issn.1673-7598.2011.11.006).
- WANG Y, WEI Y J. Review on forecasting model of wind speed and wind power [J]. *Shaanxi electric power*, 2011, 39(11): 18-21, 30. DOI: [10.3969/j.issn.1673-7598.2011.11.006](https://doi.org/10.3969/j.issn.1673-7598.2011.11.006).
- [20] WONG T T, YE H P Y. Reliable accuracy estimates from k -fold cross validation [J]. *IEEE transactions on knowledge and data engineering*, 2020, 32(8): 1586-1594. DOI: [10.1109/TKDE.2019.2912815](https://doi.org/10.1109/TKDE.2019.2912815).

作者简介:



高盛

高盛 (第一作者)

1991-, 男, 工程师, 硕士, 主要从事新能源功率预报算法研究工作 (e-mail) gordonx8@163.com。

许沛华

许沛华 (通信作者)

1980-, 男, 高级工程师, 博士, 主要从事新能源功率预报工作 (e-mail) xphboxky@126.com。

陈正洪

1964-, 男, 正研级高级工程师, 硕士, 主要从事气象科学研究工作 (e-mail) 447183633@qq.com。

(编辑 孙舒)