

引用格式: 高盛, 许沛华, 陈正洪, 等. 基于改进 LSTM 神经网络的风电功率短期预报算法 [J]. 南方能源建设, 2024, 11(1): 112-121. GAO Sheng, XU Peihua, CHEN Zhenghong, et al. A short-term algorithm based on improved LSTM neural network [J]. Southern energy construction, 2024, 11(1): 112-121. DOI: 10.16516/j.ceec.2024.1.12.

# 基于改进 LSTM 神经网络的风电功率 短期预报算法

高盛, 许沛华<sup>✉</sup>, 陈正洪, 成驰  
(湖北省气象服务中心, 湖北 武汉 430205)

**摘要:** [目的] 风能的波动性和间歇性给大规模的风电并网提出了挑战, 解决这一问题的有效途径是能够提供准确的风电功率预报。[方法] 针对这一挑战, 提出了一种新的基于改进 LSTM (长短期记忆) 架构的深度学习神经网络的风功率预报模型, 包含自主研发的数据异常检测与处理、风速特征提取、超参数调优于一体的风电功率预报方法。为了使神经网络模型能更加准确地学习风速特征对风电功率的影响, 还定义了一种使用特征筛选以及特征倍增相结合的特征工程方法。[结果] 仿真结果表明: 所提出的数据清洗及数据增强算法在各种机器学习算法上可以将准确率提升约 5%。提出的改进 LSTM 神经网络模型在数据清洗后与传统算法以及业内主流神经网络算法进行对比, 可以将准确率提升 2.5%。[结论] 改进的方法不但具有较好的噪声数据清洗能力, 而且在所有的试验中, 改进模型在预报准确性方面优于其他所有算法, 可以为实际应用提供指导。

**关键词:** 风功率日前预报; LSTM 神经网络; 深度学习; 异常检测与处理; 风速波动特征

**中图分类号:** TK89; TM614

**文献标志码:** A

**文章编号:** 2095-8676(2024)01-0112-10

**DOI:** 10.16516/j.ceec.2024.1.12

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



论文二维码

## A Short-Term Algorithm Based on Improved LSTM Neural Network

GAO Sheng, XU Peihua<sup>✉</sup>, CHEN Zhenghong, CHENG Chi

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

**Abstract:** [Introduction] The volatility and intermittency of wind energy pose significant challenges for large-scale wind power integration. An effective approach to address this issue is to provide accurate wind power forecasting. [Method] In response to this challenge, this study proposed a wind power forecasting model for deep learning neural networks based on an improved LSTM (Long Short-Term Memory) architecture. The model incorporated a wind power forecasting approach that included independently developed data anomaly detection and processing, wind speed feature extraction and hyperparameter tuning. To enhance the neural network model's ability to accurately learn the impact of wind speed features on wind power, a feature engineering method combining feature screening and feature augmentation was also defined. [Result] The simulation results demonstrate that the proposed data cleaning and data augmentation algorithm can enhance the accuracy of various machine learning algorithms by approximately 5%. Furthermore, the proposed improved LSTM neural network model, after data cleaning, outperforms traditional algorithms and state-of-the-art neural network algorithms in the industry, achieving a 2.5% increase in accuracy. [Conclusion] The improved approach not only exhibits robust capability in cleaning noisy data but also consistently outperforms other algorithms in terms of forecasting accuracy across all experiments. This model provides valuable guidance for practical applications in the field of wind power forecasting.

**Key words:** wind power day-ahead forecasting; LSTM neural network; deep learning; anomaly detection and handling; wind speed fluctuation features

收稿日期: 2023-11-30 修回日期: 2023-12-15

基金项目: 湖北省自然科学基金资助项目“风光水能互补协同机理及耦合预报方法研究”(2022CFD017); 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)

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/>).

## 0 引言

推进能源革命并构建清洁低碳、安全高效的现代能源供给体系, 最重要的途径是大力发展清洁能源, 实施化石能源清洁替代。中国提出力争于 2030 年前达到 CO<sub>2</sub> 排放峰值, 努力争取 2060 年前实现碳中和, 到 2030 年非化石能源占一次能源消费的比重达到 25% 左右<sup>[1]</sup>。风力发电是实现这一计划的重要组成部分。截至 2021 年底, 中国风电累计装机容量达 328.48 GW。然而, 大规模风电并网给电力系统的正常运行和调度带来了巨大的挑战。由于风力资源间歇性、随机性, 造成电流频率波动, 降低电力系统的可靠性。显然, 准确的日前风电功率预报在风电并网中起着主导作用, 有利于优化调度方案, 降低系统备用容量, 提高经济效益和社会效益<sup>[2-3]</sup>。

准确的风电功率预测具有以下好处: (1) 减少风电随机波动对电力系统的负面影响, 提高调度能力; (2) 降低风电系统运行风险; (3) 维护风电系统供需平衡, 促进风电交易<sup>[4-5]</sup>。近年来, 许多研究都致力于风电功率预测, 促进了风电发展。风力发电预测领域按照预测方法不同分为物理模型<sup>[6]</sup>、传统统计模型<sup>[7]</sup>和人工神经网络等。物理模型物理方法一般依靠计算流体动力学在空间和时间尺度上模拟气象现象的大气演化和物理过程<sup>[8]</sup>。目前与物理方法相关的主流模式有全球同化和预报系统等。理论上, 这些方法具有较好的时空连续性、长期预测能力和较高的时空分辨率<sup>[9]</sup>。但这些方法计算量大, 局部预报性能不足。另外, 现实世界中, 风机的理论功率曲线与实际功率曲线存在一定偏差, 使精确的物理预报模型构建变得困难, 主要有以下几个方面影响:

1) 因为涡轮机不同程度的老化, 给定风速产生的理论功率将随着时间的推移而变化。

2) 不同海拔高度气压、温度会发生变化, 导致风功率密度变化, 从而使风机因不同地形、气候条件下出力曲线发生变化<sup>[10]</sup>。

3) 风机叶片受三维空间风速湍流影响, 实际功率呈高度非线性散射点分布<sup>[11]</sup>。

传统统计模型依赖于相关的历史数据来预测未来的发电量, 主要采用历史数据、模式识别、参数估计和模型检验来建立问题的数学模型。如自回归模型(AR)或自回归移动平均模型(ARMA)等<sup>[4]</sup>。这些算法虽然计算时间复杂度小, 但存在无法捕捉非线性的风机出力规律, 往往不能产生良好的预报性能<sup>[3]</sup>。

近年来, 人工智能算法已逐渐成为风力发电预报的主流方法, 其独特的能力在特征提取和数据挖掘方面具有显著优势, 相比传统的物理模型和统计方法, 其在预报性能上表现出更高的效率<sup>[12]</sup>。这些算法能够捕获非线性关系、时间依赖性、递归和非平稳事件等重要特征。在众多技术中, 支持向量机(SVM)、神经网络、极限学习机、贝叶斯方法、马尔可夫切换模型、集成系统、模糊逻辑系统和遗传算法等都被广泛应用于风力发电预报领域<sup>[3]</sup>, 效果显著。

在深度学习领域, 卷积神经网络(CNN)被广泛应用于提取风电时间序列中的隐藏特征。许沛华等<sup>[12]</sup>认为长-短记忆网络(LSTMs)、循环神经网络(RNN)及其改进版本对处理时间序列数据具有较强的能力, 已被广泛应用于风电功率预报, 且效果良好。然而, RNN 相比传统的 ANN 虽更具优势, 但有时会出现梯度消失或爆炸现象, 导致在训练期间权重振荡或长期依赖关系的丧失。为解决此问题, 一些引入门控机制来控制层之间的信息流的方法被提出。长短期记忆(LSTM)和门控循环单元(GRU)是比较典型的例子<sup>[14-15]</sup>, 这 2 种方法的应用可以提高预报精度, 并减少训练时间。同时, 也有利用卷积神经网络进行时序预报的例子, 其中被广泛认可的模型是因果扩张时域卷积网络(TCN)<sup>[16]</sup>, 它除了具备 RNN 网络所具备的高精度特征外, 由于采用扩张卷积, 在处理长时间序列上相比其他网络也具备比较好的效果。过去的研究表明, 注意力机制在处理长序列时可以显著提升网络推理效果, 其在解决图像识别、计算机视觉等领域的问题上已经取得了显著的成功。一种自注意力模型改进 LSTM<sup>[17]</sup>也被提出。有研究证明, 具有独特结构的改进 LSTM 网络在处理长时间步长的时间序列时相比 RNN 网络表现更好<sup>[18]</sup>。因此, 改

进 LSTM 可以有效地利用注意力机制序列数据的特征,非常适合于短期风电功率的预报<sup>[19]</sup>。

本研究旨在探索并应用最先进的机器学习技术,以构建和优化基于改进 LSTM 的深度学习模型。我们通过先进的数据过滤、特征工程和模型优化技术来提高预报模型的预报精度、鲁棒性和计算性能。

## 1 数据处理方法

### 1.1 数据集描述

本研究收集了位于中国东北和华中地区的 2 个不同风电场数据验证模型有效性,其中风电场 1 位于湖北省黄冈市麻城县,经度为 115.113°,纬度为 31.566°,海拔高度为 740 m,装机容量为 80 MW,数据从 2021 年 7 月 1 日至 2022 年 6 月 30 日,时间长度为 12 个月。风电场 2 为位于内蒙古通辽市,经度为 113.432°,纬度为 41.212°,海拔高度为 1 373 m,装机容量为 49.5 MW,数据从 2021 年 2 月 1 日至 2022 年 9 月 30 日时间间隔为 15 min 的数据,时间长度为 20 个月。数据集的划分及数据量如表 1 所示。

表 1 实验数据集  
Tab. 1 Experimental data set

风电场编号及地点	数据集类型	时间跨度	样本数量/条
1号 湖北省黄冈市	全部数据	2021年7月1日~ 2022年6月30日	34 944
	训练集	2021年7月1日~ 2022年4月30日	29 088
	验证集	2022年4月30日~ 2022年5月30日	2 976
	测试集	2022年5月30日~ 2022年6月30日	2 880
2号 内蒙古自治区通辽市	全部数据	2021年2月1日~ 2022年9月30日	58 176
	训练集	2021年2月1日~ 2022年7月30日	52 320
	验证集	2022年8月1日~ 2022年8月31日	2 976
	测试集	2022年9月1日~ 2022年9月30日	2 880

### 1.2 异常数据清洗算法

风力发电离群点和异常值严重影响了风电功率预报模型的准确率,一些有代表性的研究方法被提出,郑蕾等<sup>[20]</sup>提出了局部离群点风速处理算法(Local Outlier Factor, LOF);曹立新等<sup>[21]</sup>结合滑差四分位数方法和基于多项式曲线回归方法对数据进行清理。尽管这些方法都取得了一定效果,但基于聚类的算法会将部分异常数据识别为正常数据,给应用带来不便。本研究将风速和风力发电厂整场发电功率作为变量绘制散点图,如图 1 所示。通过观察图 1 中风机出力数据围绕以靠近中心附近的区域分布,可以通过 RBF 核函数有效地将数据分布平面进行划分。本工作以此为基础,提出了一种高效的基于 OneSVM 函数的风力发电离群点数据的清洗算法。

#### 1.2.1 OneSVM 函数

OneSVM(One-Class Support Vector Machine)是一种基于支持向量机的异常检测算法。该算法旨在通过仅使用正常数据进行训练,构建一个边界来描述正常数据的特征空间,并通过检测与该边界显著偏离的样本来识别异常数据。OneSVM 作为一种无监督学习方法,能够有效地处理高维数据和非线性数据,并且具有较好的鲁棒性。OneSVM 算法的核心决策过程为式(1),通过式(1)计算决策结果以进行数据二分类划分。

$$\min_{w, \rho, \xi} \frac{1}{2} \|w\|^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i - \rho \quad (1)$$

式中:

$w$  ——决策函数的权重向量;

$\rho$  ——决策函数的截距;

$\xi_i$  ——松弛变量;

$n$  ——训练样本的数量;

$\nu$  ——1 个用于控制异常样本比例的超参数。

在本研究中选取 RBF(Radial Basis Function)核函数来构建分类平面,RBF 是径向基函数,它可以将数据从输入空间映射到一个高维特征空间,从而更好地处理非线性问题。

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (2)$$

式中:

$x, x'$  ——输入样本点;

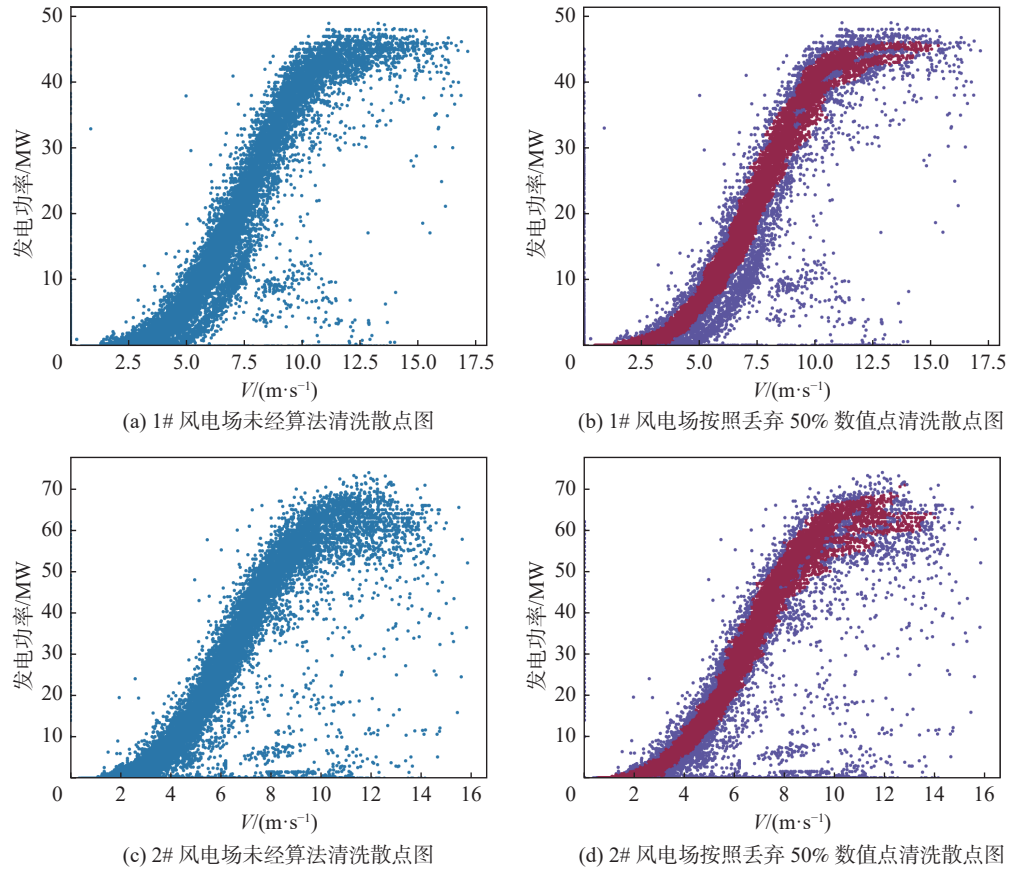


图 1 变量绘制散点图

Fig. 1 Draw a scatter plot of variables

$\gamma$  —— 1 个控制函数曲线陡峭程度的参数;  
 $\|x-x'\|^2$  —— 输入样本点之间的欧氏距离的平方。

在确定数据清洗分类时, 对于识别数据边界影响较大的参数为核函数及核函数参数。对于 RBF 核需要选择合适的参数, 过于陡峭的 RBF 函数曲线会对部分正确样本进行错误映射从而导致一些正常的风电曲线特征丢失。通常的选取方式为将  $\gamma$  设置为 1/分类数量, 对于二分类问题该值通常设为 0.5。在进行工程应用时式(1)中的决策边界一般由训练样本误差百分比进行确定。

## 2 数据特征工程方法

### 2.1 数据特征相关性

过去的研究表明在训练集中存在大量不相关数据会显著增加模型的复杂度, 也会影响模型推理的稳定性。因此通过特征工程方法筛选掉一些对模型学习有影响的特征显得尤为重要<sup>[22]</sup>。在本研究中采用了一种基于数据相关性矩阵进行特征筛选的方法。

相关性矩阵通过计算各个特征之间的相关系数, 反映特征间的依赖性。在相关性矩阵中采用皮尔逊相关系数进行元素计算, 相关系数的计算公式为式(3), 其中  $X_{i,k}$ 、 $X_{j,k}$  是 2 个变量,  $\bar{X}_i$ 、 $\bar{X}_j$  分别是它们的均值,  $n$  是样本数量。

$$r_{i,j} = \frac{\sum_{k=1}^n (X_{i,k} - \bar{X}_i)(X_{j,k} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{i,k} - \bar{X}_i)^2} \sqrt{\sum_{k=1}^n (X_{j,k} - \bar{X}_j)^2}} \quad (3)$$

进行数据相关性筛选的具体步骤如下:

1) 构建特征数据集的相关性矩阵  $R$ , 其中  $R[i, j]$  表示第  $i$  个和第  $j$  个特征的相关系数。

2) 设置相关性系数阈值  $\rho$ , 在本研究中该值设置为 0.5。

3) 对  $R$  进行检查, 如果存在  $|R[i, j]| > \rho$ , 则认为第  $i$  个特征与第  $j$  个特征高度相关。

4) 在每对高度相关的特征中, 删除相关性较小的特征。此时需比较各特征与目标值的相关性, 删



除与目标值相关性较小者。

5) 重复步骤 3) 和步骤 4), 直到不存在高度相关的特征对。

6) 最终获得的特征子集中, 各特征间具有较低的冗余性。

通过以上方法, 在本文中选择了风机出力功率影响最大的 13 个指标, 包括: 10 m 高度的风速、风向、湿度、气压、温度以及 30 m、50 m、70 m、轮毂高度的风速和风向。其中, 对风电出力影响最大的因素是风速, 为了使得模型可以更好地学习风速的变化规律以及风速与发电功率之间的关系, 需要进一步针对风速特征进行数据增强。

## 2.2 数据特征倍增方法

由于风速具有随机性强的特点, 因此研究者通常不会采用对风速进行单纯的数值处理的方式进行特征倍增。结合风电场的地形特征, 通常在特定的季节内, 特定测风点风速会存在一些时延重复特征。这种特征通常表现为在临近时间范围附近可以观察到相似的风速波动规律片段。传统的数据处理方法很少考虑特征片段时移带来的增益。因此, 本研究中采用一种创新的数据倍增方式即时延滑动窗口特征处理方法。这种方法通过在数据集上滑动一个固定长度的窗口, 依次截取窗口中的子集进行统计计算, 从而得到一组移动平均或者其他滚动指标。

设时间序列为  $X = x_1, x_2, \dots, x_n$ , 窗口大小为  $w$ , 窗口起点  $t$  的数据集为  $X_t = x_t, \dots, x_{t+w-1}$ 。

在  $t$  时刻, 根据指定统计函数  $f$ , 可以计算得到:  $y_t = f(X_t)$ 。通过在原始序列上滑动窗口, 重复该过程, 最终得到输出移动平均序列:

$$Y = y_1, y_2, \dots, y_{n-w+1} \quad (4)$$

滑动窗口方法支持多种统计函数  $f$ , 如求平均、最大值、相关系数等。通过调整窗口大小  $w$ , 可以观察不同时间尺度上的趋势和模式。本文采用的滑动窗口大小分别为 4 和 24, 统计函数方法选择了最大值和平均值统计法。经过实验这组参数可以最有效地提升模型效果。

## 2.3 风速特征重建

由于风的波动性、随机性和受湍流的影响, 导致风击中掠叶转子的速度和方向迅速变化, 使观测风速与撞击转子叶片的风速不匹配, 导致风速与输出

功率之间的不匹配<sup>[23]</sup>。大量研究表明准确地刻画风速的波动特征有利于提高预报模型的准确率, 尤其某一时刻风速  $v_0$  与邻近时间段  $\Delta t = \{15, 30, 45, 60\}$  (单位为 min) 内风速的变化特征与发电功率存在潜在映射关系。通过对风速数据进行分析, 并参考许沛华等<sup>[12, 24]</sup>提出的风速数据处理方法, 在此基础上结合第 1 章中 2 个风电场的实测数据将对风速变化波动特征影响较大的因子进行筛选以及检验, 得到了 4 个影响比较明显的特征因子。通过数据分析及验证, 在我们的模型中引入了邻近的 1 h 内风速波动特征的表示, 这些特征分别为风速绝对波动幅度  $V_A$ 、风速相对波动幅度  $V_R$ 、风速标准差  $V_\varepsilon$ 、最大风速  $V_{\max}$ 、最小风速  $V_{\min}$ 、风速爬坡次数  $R_{\text{cnt}}$  6 个量表示风速的波动特征。

定义 1: 风速标准差  $V_\varepsilon$  反映  $\Delta t$  时间内风速偏离风速均值  $\bar{v}$  的离散程度, 该值越大波动越大, 否则波动较小, 如式(5):

$$V_\varepsilon = \sqrt{\frac{\sum_{i \in \Delta t} (v_i - \bar{v})^2}{n}} \quad (5)$$

式中:

$n$ ——样本总数(个)。

定义 2: 最大风速  $V_{\max}$  反映  $\Delta t$  时间内风速的最大值, 如式(6):

$$V_{\max} = \max_{i \in \Delta t} (v_i) \quad (6)$$

定义 3: 最小风速  $V_{\min}$  反映  $\Delta t$  时间内风速的最小值, 如式(7):

$$V_{\min} = \min_{i \in \Delta t} (v_i) \quad (7)$$

定义 4: 风速爬坡次数  $R_{\text{cnt}}$  是  $\Delta t$  时间内风速变化超过阈值  $\delta$  的次数, 该值越大代表波动越频繁, 否则波动不频繁, 如式(9):

$$C_i = \begin{cases} 1 & |v_0 - v_i| \geq \delta \quad (i \in \Delta t) \\ 0 & |v_0 - v_i| < \delta \quad (i \in \Delta t) \end{cases} \quad (8)$$

$$R_{\text{cnt}} = \sum_{i \in \Delta t} C_i \quad (9)$$

## 3 预报模型

### 3.1 机器学习回归模型

在风电发电功率预报领域, 机器学习技术被广泛应用, 其中被认为效果比较好的算法包括随机森

林算法 Random Forest(RF), 梯度提升树集成算法 LightGBM 等。目前被更多使用的方法是 LightGBM, LightGBM 采用了梯度提升树原理, 其中包括叶子分裂策略、直方图优化、基于梯度的学习和正则化, 使其能够高效处理风能数据。以下是 LightGBM 的关键原理:

**叶子分裂策略:** LightGBM 采用了叶子分裂 (Leaf-wise) 的生长方式, 与传统的深度优先生长 (Depth-wise) 不同。这种策略选择具有最大梯度的叶子来分裂, 以最大程度地减小损失函数。这导致树的深度相对较小, 有助于捕获风能数据中的非线性关系和季节性模式。

**直方图优化:** LightGBM 使用直方图算法, 将数据集分成多个直方图, 每个直方图代表一个特征的取值范围。这减少了对数据的排序和遍历, 提高了训练速度, 尤其适用于大规模数据集。

**基于梯度的学习:** LightGBM 在每次迭代中计算损失函数对模型预报的梯度, 并根据梯度拟合一个新的决策树。这使模型逐步优化, 不断减小预报误差。

**正则化:** LightGBM 引入了正则化项, 包括叶子结点的最大深度和叶子结点的最小数据数。这有助于防止过拟合, 提高模型的泛化性能。

### 3.2 时域卷积神经网络 TCN

神经网络方法也经常被用在为风电功率预报领域, 其中比较典型的算法有基于 RNN 的 GRU,

LSTM 等算法, 基于卷积神经网络的 TCN 算法等。其中 TCN 算法是近期被认为具有潜力的算法。由于风电具备长时间序列依赖特征, 因此为了捕获具有较长历史的时间序列, TCN 使用了因果空洞卷积, 采用指数级的放大方式扩大接受野。 $t$  时刻的输出取决于  $t$  时刻及之前的输入, 说明历史信息没有被忽略, 通过膨胀系数  $d$  实现了对卷积层的间隔采样, 它扩大了接受野。总之, 扩张因果卷积在处理数据时可以考虑更多的历史信息, 获得更大的一维序列的特征视界。针对这种网络结构特征的定义如式 (10) 所示。

$$F(s) = (x_d^* f)(s) = \sum_{i=0}^{n-1} f(i) \cdot x_{s-d \cdot i} \quad (10)$$

式中:

$d$  ——膨胀系数, 代表每间隔多少数据点进行一次采样;

$x_{s-d \cdot i}$  ——经过按膨胀系数  $d$  间隔选取的第  $i$  个数据点。

### 3.3 改进 LSTM 方法

受到近期被学术界广泛研究的编码器-解码器神经网络结构的启发, 本研究中采用了一种新型的基于 LSTM 单元堆叠的编码器与解码器结构。相比传统 LSTM, 这种形式可以通过引入时间变量编码、解码时间段的数值预报信息来提升预报的准确性。改进的 LSTM 神经网络结构如图 2 所示, 对于网络结构描述具体如下。

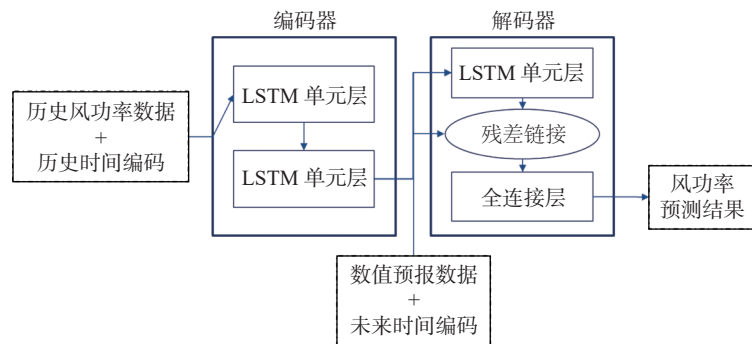


图 2 改进的 LSTM 网络结构

Fig. 2 Improved LSTM network structure

编码器部分由多个编码器层堆叠而成, 在本研究中每个编码器层包含 2 个子层。编码器输入层之前包含 1 个归一化层, 用于平稳风电时间序列数据

的波动情况, 将所有值采用标准归一化方法归一化到  $[-1, 1]$  的区间内。编码器层的输入包含对时间序列的编码, 编码器的输出为堆叠 LSTM 的输出。解

码器部分为单个 LSTM 子层,其后包括 1 个残差连接层和 1 个全连接层。残差连接将编码器的输出与解码器的输出进行加和连接,残差层后经过单个全连接层进行最终结果输出。

总体来说,改进 LSTM 模型的网络结构除了引入编码器-解码器机制外还采用了残差连接和归一化的设计,以增强模型的训练稳定性和表达能力。

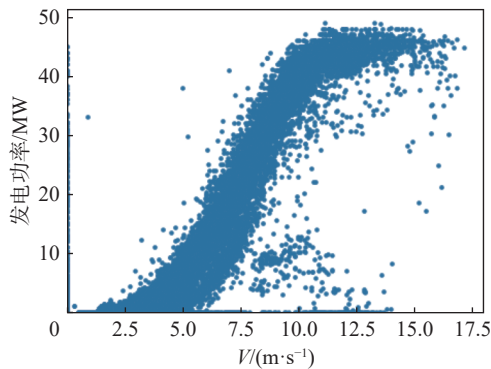
## 4 实验结果与讨论

### 4.1 评估指标

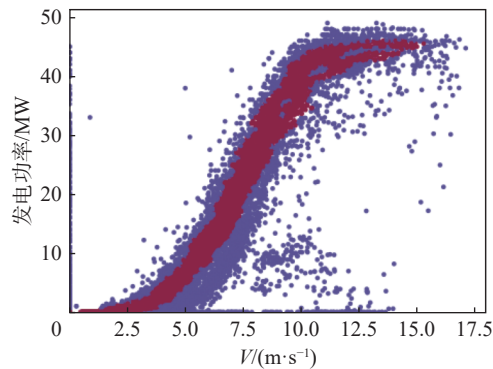
由于平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Squared Error, RMSE)可以反映预报值与真实值之间的距离,本文使用 MAE、RMSE 来评估预报误差。一般越小的值代表算法表现越好。同时本文也引入了被广泛风电场采用的准确率评估公式进行准确率评估对比,该公式可以表示实际发电功率与预报功率之间的偏差百分比。通常准确率越接近 100% 表示算法性能越好。它们的公式见式(11)~(13)。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$



(a) 1# 风电场未经算法清洗散点图



(b) 1# 风电场按照丢弃 50% 数值点清洗散点图

图 3 风电场 1 散点图异常数据点清洗效果

Fig. 3 Cleaning effect of abnormal data points in scatter diagram of wind farm 1

### 4.3 对比实验结果

本文对比了不同算法在数据清洗前和数据清洗后的预报误差,风电场 1 的结果数据如表 2 所示,风电场 2 的结果数据如表 3 所示。

$$\text{准确率} = \left\{ 1 - \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{y_{\text{cap}}} \right)^2} \right\} \times 100\% \quad (13)$$

式中:

$y_i$  ——真实值(MW);

$\hat{y}_i$  ——预报值(MW);

$y_{\text{cap}}$  ——电厂的装机容量(MW)。

### 4.2 数据清洗结果

风电场 1 和风电场 2 的数据清洗结果可以分别参见图 3 和图 4 所示。在图 3 中,左侧图为使用未经清洗的风速-风力发电功率绘制的散点图,右侧图的红色点状内容为经过使用 RBF 核的 SVM 进行边界划分数据清洗后的结果,在本实验中所选取的核参数为其中使用训练误差分数的上界划分参数 nu 值为 0.05, RBF 核曲线斜率  $\gamma$  选取 0.5。图 4 与图 3 的呈现方式相同。从图中可以看出经过 OneSVM 算法清洗后的部分数据可以保留风电场的有效风功率特征,诸如功率记录异常数据、弃风时段的风功率数据等均可以被有效清除。图 3 与图 4 中散点右下方区域的点可能是由于部分风机故障检修或者部分风机迎风角度不正确导致,这种情况在日常发电过程中不属于常发事件,但这部分数据会对机器学习的结果产生较大的影响。经过 OneSVM 清洗后的数据不会包含这部分数据,因此可以整体提升机器学习算法的准确率。

表 2 和表 3 分别对应风电场 1 和风电场 2 的实验结果,从表中可以看到风电场 1 和风电场 2 所使用的方法相比其他方法具有比较优势。从风电场 1 的数据表中横向比较进行数据清洗及数据增强前后

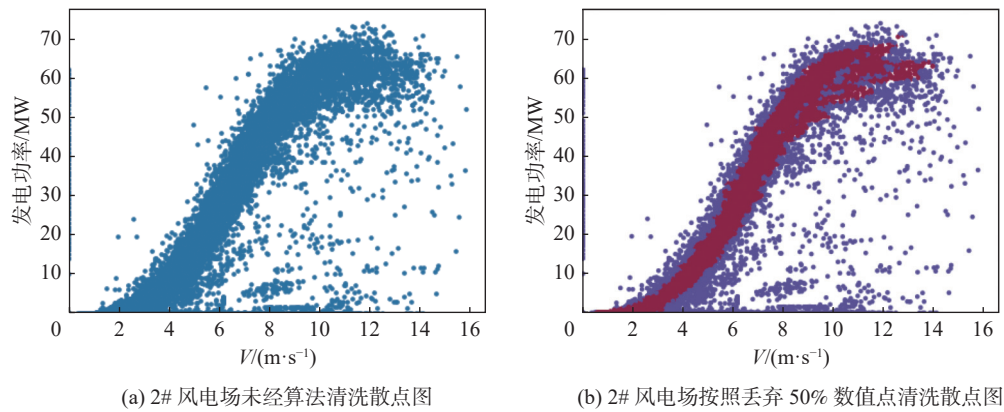


图 4 风电场 2 散点图异常数据点清洗效果

Fig. 4 Cleaning effect of abnormal data points in scatter diagram of wind farm 2

同样算法下的 MAE 指标相比未清洗情况平均误差减小了约 2 MW, 约占场站装机容量的 2.5%, RMSE 指标相比之前减少了约 3 MW, 约占场站装机容量的 3.8%, 平均准确率提升约 5%。风电场 2 的提升数据也呈类似情况。比较表中使用原始数据集以及使用经过特征工程处理后的数据相比特征工程处理之前在所有指标上都有显著提升。

使用改进的 LSTM 算法在 MAE、RMSE 及准确率指标相比 LightGBM 算法在风电场 1 提升约 0.94 MW、2.141 MW 及 4.21%, 相比传统 LSTM 及 TCN 平均提升 0.333 MW、0.313 MW 及 0.62%。风电场 2 的提升情况与风电场 1 类似。综合来看, 平均 2 个场站的结果本文提出的改进算法在数据清洗后进行对比平均可以提升约 2.5% 的准确率, 对比数据清洗前准

确率平均可以提升约 5.9%。表明文中所提出的改进 LSTM 算法针对风电功率预报模型的误差减小以及准确率提升是有效的。

#### 4.4 结果分析与总结

通过上文分析结论, 本文主要有以下 3 个贡献:

1) 与现有的神经网络模型对比, 文章中提出的改进 LSTM 模型在已知数据集上的表现优于 TCN 等流行的神经网络, 它表现出了显著的性能。到目前为止, 在风力发电日前预报领域还缺乏这样的比较。因此, 本文将一种新的基于改进 LSTM 的深度学习方法应用于风电功率的预报, 并通过与其他传统模型以及神经网络模型比较, 综合评估了该方法在风电功率预报中的有效性。

2) 风机出力受一段时间内风速相对或绝对波动

表 2 风电场 1 结果数据

Tab. 2 Result data of wind farm 1

指标	原始数据集				数据清洗&数据增强后数据集			
	LightGBM	LSTM	TCN	改进LSTM	LightGBM	LSTM	TCN	改进LSTM
MAE/MW	4.882	4.318	5.617	4.738	3.644	2.830	3.244	2.704
RMSE/MW	7.729	5.922	7.094	6.066	5.648	3.559	4.081	3.507
准确率/%	84.54	88.16	85.81	87.87	88.77	92.88	91.84	92.98

表 3 风电场 2 结果数据

Tab. 3 Result data of wind farm 2

指标	原始数据集				数据清洗&数据增强后数据集			
	LightGBM	LSTM	TCN	改进LSTM	LightGBM	LSTM	TCN	改进LSTM
MAE/MW	1.273	0.532	0.898	0.569	0.994	0.372	0.726	0.369
RMSE/MW	2.028	0.907	1.336	0.903	1.597	0.540	1.244	0.526
准确率/%	88.73	94.96	92.58	94.98	91.13	97.00	93.09	97.08



幅度、最大风速、最小风速、风速爬坡的频率等因子影响,因此准确的风速波动特征提取非常重要。虽然 LSTM 和 TCN 深度学习模型对历史信息具有记忆功能,但对风力发电的这些特征提取仍然不够。本研究中创新性地提出了一种时延滚动数据拼接的方法,并定义了 4 个对风机出力影响较大的风速特征变量输入到模型中,这种数据处理大大增强了模型的学习能力。

3)在这项研究中,一种基于 OneSVM 函数的检测和消除风力发电数据集中异常值的方法被提出,然后将其输入到 TCN,改进 LSTM 等深度学习模型中。提出的基于 OneSVM 函数的异常值检测算法可以根据数据分布密度划分识别区域,识别异常值更加精准,增强了预报模型的准确率。

## 5 结论

本研究针对风电功率预报问题,提出一种基于改进 LSTM 的预报模型。该模型集成了数据异常检测、风速特征提取、超参数优化等模块,形成了一套端到端的预报解决方案。研究表明,该模型可以有效检测和处理异常数据,准确学习风速特征,并利用改进 LSTM 网络进行预报,相较 TCN 等算法具有明显优势。本研究的 3 点创新与贡献包括:利用改进 LSTM 网络进行风电功率预报,与 TCN 等算法进行比较,验证其在风功率预报任务上的有效性;提出时延滚动数据拼接方法,定义风速特征变量,增强模型学习能力;应用基于 OneSVM 函数的异常值检测算法,提高预报准确率。

本研究为风电功率预报提供了一种新的基于深度学习的解决方案。相较于传统算法,该方案可以更好地建模时间依赖关系,提取关键特征,处理异常数据。研究结果表明,该模型可以显著提高风功率预报的精度。

### 参考文献:

- [1] 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. DOI: 10.3969/j.issn.1003-2355.2021.04.001.
- [2] 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.
- [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.
- [4] 冯泽深, 赵增海, 郭雁珩, 等. 2021 年中国风电发展现状与展望 [J]. 水力发电, 2022, 48(10): 1-3,8. DOI: 10.3969/j.issn.0559-9342.2022.10.001.
- [5] 雷旭, 马鹏飞, 宋智帅, 等. 计及风电预测误差的柔性负荷日内调度模型 [J]. 发电技术, 2022, 43(3): 485-491. DOI: 10.12096/j.2096-4528.pgt.20083.
- [6] KLEIST D T, PARRISH D F, DERBER J C, et al. Introduction of the GSI into the NCEP global data assimilation system [J]. *Weather and forecasting*, 2009, 24(6): 1691-1705. DOI: 10.1175/2009WAF2222201.1.
- [7] JUNG J, BROADWATER R P. Current status and future advances for wind speed and power forecasting [J]. *Renewable and sustainable energy reviews*, 2014, 31: 762-777. DOI: 10.1016/j.rser.2013.12.054.
- [8] LANGE M, FOCKEN U. Physical approach to short-term wind power prediction [M]. Berlin: Springer, 2006. DOI: 10.1007/3-540-31106-8.
- [9] SHCHEPETKIN A F, MCWILLIAMS J C. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model [J]. *Ocean modelling*, 2005, 9(4): 347-404. DOI: 10.1016/j.ocemod.2004.08.002.
- [10] LIU H, CHEN C, LÜ X W, et al. Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods [J]. *Energy conversion and management*, 2019, 195: 328-345. DOI: 10.1016/j.enconman.2019.05.020.

- [11] JIA X D, JIN C, BUZZA M, et al. Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves [J]. *Renewable energy*, 2016, 99: 1191-1201. DOI: [10.1016/j.renene.2016.08.018](https://doi.org/10.1016/j.renene.2016.08.018).
- [12] 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](https://doi.org/10.3390/app13064042).
- [13] MA L, LUAN S Y, JIANG C W, et al. A review on the forecasting of wind speed and generated power [J]. *Renewable and sustainable energy reviews*, 2009, 13(4): 915-920. DOI: [10.1016/J.RSER.2008.02.002](https://doi.org/10.1016/J.RSER.2008.02.002).
- [14] 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, Arlington, TX, USA, September 26-28, 2010. Arlington: IEEE, 2010: 1-8. DOI: [10.1109/NAPS.2010.5619586](https://doi.org/10.1109/NAPS.2010.5619586).
- [15] 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).
- [16] LEA C, VIDAL R, REITER A, et al. Temporal convolutional networks: a unified approach to action segmentation [C]//Anon. European Conference on Computer Vision, Amsterdam, The Netherlands, October 8-10 and 15-16, 2016. Amsterdam: Springer, 2016: 47-54. DOI: [10.1007/978-3-319-49409-8\\_7](https://doi.org/10.1007/978-3-319-49409-8_7).
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Anon. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, December 4-9, 2017. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [18] SHAFIEE M, FINKELSTEIN M, BÉRENGUER C. An opportunistic condition-based maintenance policy for offshore wind turbine blades subjected to degradation and environmental shocks [J]. *Reliability engineering & system safety*, 2015, 142: 463-471. DOI: [10.1016/j.ress.2015.05.001](https://doi.org/10.1016/j.ress.2015.05.001).
- [19] LIM B, ARIK S O, LOEFF N, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting [EB/OL]. (2020-09-27) [2023-11-04]. <http://arxiv.org/abs/1912.09363>.
- [20] ZHENG L, HU W, MIN Y. Raw wind data preprocessing: a data-mining approach [J]. *IEEE transactions on sustainable energy*, 2015, 6(1): 11-19. DOI: [10.1109/TSSTE.2014.2355837](https://doi.org/10.1109/TSSTE.2014.2355837).
- [21] 曹立新, 刘伟民, 郭虎全. 风电场功率曲线异常数据的清洗与建模 [J]. *兰州理工大学学报*, 2022, 48(4): 64-70. CAO L X, LIU W M, GUO H Q. Cleaning and modeling of abnormal data of wind farm power curve [J]. *Journal of Lanzhou University of Technology*, 2022, 48(4): 64-70.
- [22] ALI M, PRASAD R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition [J]. *Renewable and sustainable energy reviews*, 2019, 104: 281-295. DOI: [10.1016/j.rser.2019.01.014](https://doi.org/10.1016/j.rser.2019.01.014).
- [23] SEVLIAN R, RAJAGOPAL R. Detection and statistics of wind power ramps [J]. *IEEE transactions on power systems*, 2013, 28(4): 3610-3620. DOI: [10.1109/TPWRS.2013.2266378](https://doi.org/10.1109/TPWRS.2013.2266378).
- [24] 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](https://doi.org/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](https://doi.org/10.11755/j.issn.1006-7639(2021)-03-0524).

---

 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](mailto: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](mailto: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](mailto:447183633@qq.com)。

成驰

1981-, 男, 高级工程师, 硕士, 主要从事气象科学研究与新能源电力方向管理及研究工作 (e-mail) [244564224@qq.com](mailto:244564224@qq.com)。

(编辑 孙舒)