

基于逐步回归与深度学习的风电功率预测

周全民¹, 李润^{2*}, 叶征灯¹, 高飞¹, 周杨丰³

¹国家电投集团江西水电检修安装工程有限公司, 江西 南昌

²北京东润环能科技股份有限公司, 北京

³国家电投集团江西电力有限公司高新清洁能源分公司, 江西 南昌

收稿日期: 2021年10月18日; 录用日期: 2021年11月2日; 发布日期: 2021年12月13日

摘要

根据各因素对风电场输出功率的影响, 本文建立了逐步回归与深度学习超短期风电场输出功率的预测模型。对比分析了风电场实际输出功率、不同高度的风速、风向数据对于预测精度的影响。最终通过逐步回归挑选出对结果有重要影响的变量作为模型的输入数据, 提高预测精度。

关键词

风电场, 功率预测, LSTM模型, 逐步回归

Wind Power Prediction Based on Stepwise Regression and Deep Learning Model

Quanmin Zhou¹, Run Li^{2*}, Zhengdeng Ye¹, Fei Gao¹, Yangfeng Zhou³

¹Spic Jiangxi Hydropower Maintenance & Installation Engineering Co., Ltd., Nanchang Jiangxi

²Beijing East Environment Energy Technology Co., Ltd., Beijing

³High-Tech Clean Energy Branch, Spic Jiangxi Electric Power Co., Ltd., Nanchang Jiangxi

Received: Oct. 18th, 2021; accepted: Nov. 2nd, 2021; published: Dec. 13th, 2021

Abstract

According to the influence of various factors on the output power of wind farms, this paper establishes a stepwise regression and deep learning ultra-short-term wind farm output power prediction model. The effect of wind farm actual output power, wind speed and wind direction data at different heights to the prediction accuracy is comparatively analyzed. Finally, through stepwise regression, the variables that have an important impact on the results are selected as the input

*通讯作者。

文章引用: 周全民, 李润, 叶征灯, 高飞, 周杨丰. 基于逐步回归与深度学习的风电功率预测[J]. 电力与能源进展, 2021, 9(6): 310-317. DOI: 10.12677/aepe.2021.96033

data of the model to improve the prediction accuracy.

Keywords

Wind Farm, Wind Power Prediction, LSTM Model, Step-Wise Regression

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

能源危机和气候变化使新能源得到了迅速发展和广泛应用, 全国各地的风电场数目正在急剧增加, 风的波动性、间歇性导致风电场出力具有不确定性, 使电网公司在制定调度计划和风电场在制定风电机组检修计划时存在困难, 因此有必要对风电场进行功率预测, 这其中就包括超短期功率预测[1]。

王仁忠[2]提出了影响风功率预测精度的因素, 其中包括预测系统的功能, 要注意不同区域工作特点以及发展方向的差异性, 用针对性的问题来完善; 付浩和王圣达等人[3]提出了一种自回归移动平均-长短期记忆网络(ARMA-LSTM)光功率预测方法, 该方法通过“db5”小波基函数对原始光功率数据进行一层分解获得光功率数据趋势项部分和细节项部分。采用 LSTM 对趋势项进行建模预测, 采用 ARMA 对细节项进行建模预测并将两种预测结果相加得到最终的预测结果, 提高了预测精度; 颜晓娟等人[4][5]基于结构风险最小化的支持向量机进行建模。针对支持向量机参数难确定的问题, 利用小生境遗传算法对支持向量机惩罚参数 C 和核函数 δ 进行寻优, 并对算法中小生境半径 L 值难确定的问题, 引入自适应的思想, 使 L 并不取某个固定值, 而是随着最优个体的变化而变化, 参数更优, 预测精度更高。孟岩峰等人[6]对预测结果采用误差评价指标进行了评价分析, 提出通过预测模型修正逐步减小风电功率预测误差的方法。常鹏和高亚静等人[7]针对风电场功率时间序列的非线性和非平稳性, 分别将 EMD 和 EEMD 方法与时间序列的方法相结合应用于风电场功率预测中, 提出基于 EMS-ARMA 和 EEMD-ARMA 的风功率预测方法。上述方法均需要对原始数据进行分解, 过程较为复杂, 而且在分解过程中原始数据所包含的部分信息丢失, 对预测的精确度产生影响。

本文分析了影响风电场实际功率输出的因素, 设计了基于数值天气预报的逐步回归与深度学习神经网络预测模型, 对比了实测功率、不同高度的气象数据对于预测结果精度的影响, 在众多影响因素当中, 利用逐步回归的方法, 挑选出对预测变量的变化解释能力贡献大的变量, 作为 LSTM 模型的输入变量, 对输出功率进行预测。提高了预测精度, 简化了预测过程。

2. 基于逐步回归的 LSTM 神经网络算法

2.1. LSTM 神经网络

LSTM 是长短期记忆网络, 是一种时间递归神经网络, 适合于处理和预测时间序列中间隔和延迟相对较长的重要事件。LSTM 区别于 RNN 的地方, 主要就在于它在算法中加入了一个判断信息有用与否的“处理器”, 把有用的信息保存在 LSTM 中, 这个处理器作用的结构被称为 cell。一个 cell 当中被放置了三扇门, 分别叫做输入门、遗忘门和输出门, 其拓扑结构图如图 1。当一连串信息进入 LSTM 的网络当中, 可以根据规则来判断是否有用, 只有符合算法认证的信息才会留下, 不符的信息则通过遗忘门被遗忘。

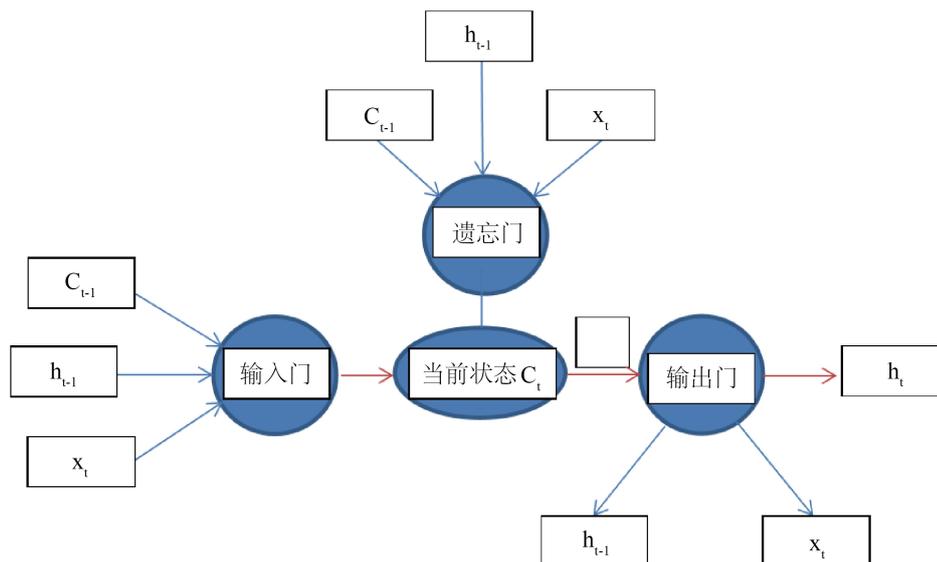


Figure 1. The topology structure of LSTM model

图 1. LSTM 模型拓扑结构

LSTM 的原理: LSTM 通过精心设计的被称作“门”的结构来去除或增加信息到细胞状态的能力。门是一种让信息选择式通过的方法。它们包含一个 sigmoid 神经网络层和一个 pointwise 乘法操作。LSTM 拥有三个门，来保护和控制细胞状态。忘记门：忘记门中训练的是一个 w_f 的权值，而且上一时刻的输出和当前时刻的输入是一个 concat 操作。忘记门决定我们会从细胞状态中丢弃什么信息，因为 sigmoid 函数的输出是一个小于 1 的值，相当于对每个维度上的值做一个衰减。信息增加门：决定更新到细胞状态中的信息。其中 sigmoid 决定了要更新的值，tanh 创建一个新的细胞状态的候选向量 C_t ，该过程训练两个权值 W_i 和 W_c 。经过第一个和第二个门后，可以确定传递信息的删除和增加，即可以进行“细胞状态”的更新。信息输出：通过 sigmoid 确定细胞状态那个部分将输出，tanh 处理细胞状态得到一个 -1 到 1 之间的值，再将它和 sigmoid 门的输出相乘，输出程序确定输出的部分。

2.2. 逐步回归

逐步回归的基本思想是将变量逐个引入模型，每引入一个解释变量后都要进行 F 检验，并对已经选入的解释变量逐个进行 t 检验，当原来引入的解释变量由于后面解释变量的引入变得不再显著时，则将其删除。以确保每次引入新的变量之前回归方程中只包含显著性变量。这是一个反复的过程，直到既没有显著的解释变量选入回归方程，也没有不显著的解释变量从回归方程中剔除为止。以保证最后所得到的解释变量集是最优的。

先用被解释变量对每一个所考虑的解释变量做简单回归，然后以对被解释变量贡献最大的解释变量所对应的回归方程为基础，再逐步引入其余解释变量。经过逐步回归，使得最后保留在模型中的解释变量既是重要的，又没有严重多重共线性。

2.3. 输入数据对预测结果的影响

对于超短期风电功率预测，前一时刻风电场的输出功率，表征了预测的起始值，可以作为风电场输出功率的一个重要参考量[8]。本文研究了分别将前 15 min、30 min、45 min 的实际输出功率作为输入变量对实际功率进行预测。计算得到各种情况下的误差。

2.4. LSTM 模型训练步骤

1) 构建网络：隐匿层选择 50 个神经元，输入层根据实际情况设置神经元个数，输入层选择 1 个神经元，设置初始学习率为 0.006，模型迭代次数为 200 次。

2) 对训练集和测试集数据进行归一化处理：

$$X_g = \frac{X_t - \bar{X}}{\sigma}$$

3) 设置输入层和输出层权重、偏置。

4) 训练模型，定义损失函数

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_i - Y_i \right)^2$$

5) 对数据进行反归一化处理，得到预测值。

3. 算例分析

3.1. 风电场的的数据

风电场提供了 2018 年 9 月份 30 天，时间间隔为 15 分钟、风塔高度为 30 m、70 m、90 m 的气象数据观测值(风速、风向、温度、湿度、气压、空气密度)，以及风电场的实际功率。训练数据为 2018 年 9 月 1 号到 9 月 21 号数值天气预报数据和风电场的实际功率，选取 2018 年 9 月 22 号到 9 月 30 号的数据用于测试。

3.2. 影响 LSTM 模型精度的因素

学习率对精度的影响

为了能够使得梯度下降法有较好的性能，我们需要把学习率的值设定在合适的范围内。太大的学习速率导致学习的不稳定，太小值又导致极长的训练时间。一个合适的学习速率通过保证稳定训练的前提下，达到了合理的高速率，可以减少训练时间。

由表 1 和图 2 可以看出，学习率为 0.05 时，损失函数下降速度很快，结合损失函数的方差可以发现，学习率较大时，会导致学习的不稳定。综合损失函数的平均数以及方差这两个数字特征，学习率为 0.01 时，既可以保证较快的训练速度，又可以保证学习的稳定以及预测结果的准确性。

3.3. 输入数据对精度的影响：

3.3.1. 实测功率对超短期预测的影响

将前一刻的实际功率测量值作为输入变量，经过计算得到不同情况下的误差，如表 2 所示。

Table 1. The effect of learning rates on forecasting accuracy

表 1. 学习率对预测精度的影响

LR	MSE	STDEV.MSE
0.001	0.048139	0.623379
0.01	0.037062	0.287312
0.05	0.068692	0.334516

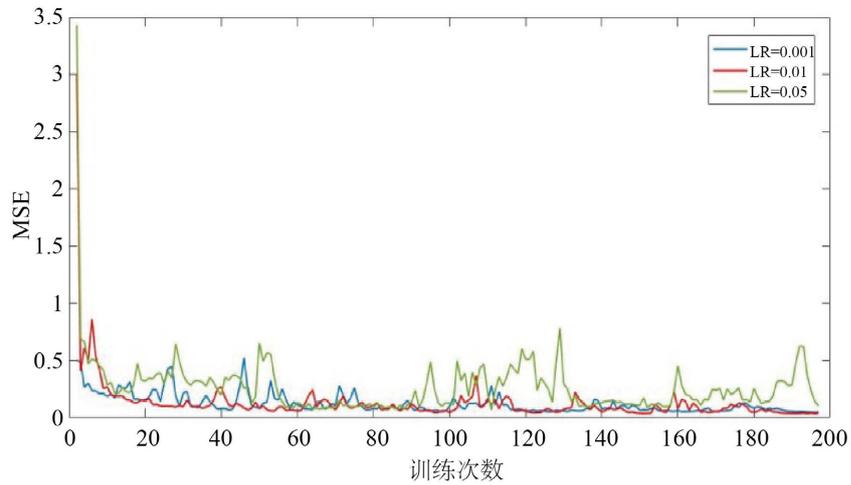


Figure 2. The loss function under different learning rates
图 2. 不同学习率下的损失函数

Table 2. The influence of the real power on the loss function
表 2. 实测功率对损失函数的影响

实测功率	MSE	STDEV
不考虑前一刻测量值	0.21901	0.1018
考虑前一刻测量值	0.1464	0.1360

由图 3 预测结果可以看出，把上一刻风电场的实际输出功率作为神经网络的输入，能够提高预测精度。15 min 内，风速的变化不会那么剧烈，前一刻风速与此刻风速相关性较大，所以将前一刻实际输出功率考虑进来，可以提高预测精度

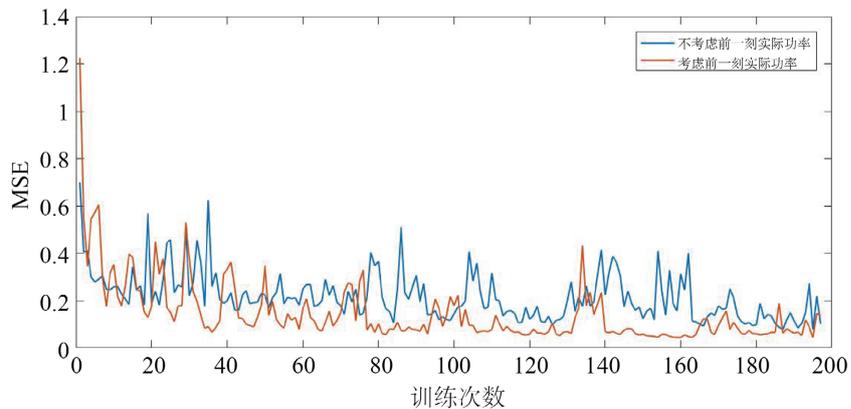


Figure 3. The influence of the real power on the loss function
图 3. 实测功率对损失函数的影响

3.3.2. 不同高度的风速对实际功率的影响

风轮扫风的最低点在 14 m~50 m 之间，风轮扫风的最高点在 71.6 m~150 m 之间，因此这段空间内的风速分布对风电机组的输出功率有较大影响。风电场提供的数据中，风速的测量高度有 30 m、70 m、90

m。下列情况考虑的分别是只考虑 30 m 高度的风速数据、只考虑 30 m 和 70 m 高度的风速数据、将 30 m、70 m、90 m 高度的风速全都纳入考虑。经分析得到下列预测结果的误差。

由表 3 结果可得, 第一种情况的均方误差为 13.66%, 第二种情况的均方误差为 7.84%, 第三种情况的均方误差为 8.64%, 因此将不同高度的数据纳入考虑, 可以提高预测的精度, 但要将变量合理纳入考虑, 否则会导致精度下降。

Table 3. The influence of wind speed at different heights on MSE

表 3. 不同高度的风速对均方误差的影响

不同高度的风速	MSE	STDEV.MSE
30 m	0.1365973	0.204171
30 m, 70 m	0.07839403	0.614722
30 m, 70 m, 90 m	0.08642108	0.841246

3.3.3. 最终预测模型的输入变量

由上述分析可知, 对模型预测结果产生影响的一共有两类因素, 一类是模型自身的参数设置, 比如学习率和隐层神经元的个数; 另一类是输入数据对于预测结果的影响, 考虑实测数据以及不同高度的风速。由于 3 种不同的高度对应的历史气象数据(风速、风向、温度、湿度、气压、空气密度)不同, 为了加快训练速度、减小计算量, 利用逐步回归的方法选出对预测变量解释能力较强的因素, 将这些因素作为 LSTM 的输入变量, 逐步回归结果如表 4 所示。

Table 4. The results of stepwise regression

表 4. 逐步回归结果

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.418	0.21		1.993	0.046
15 min	1.077	0.019	1.077	58.158	0
30 min	-0.125	0.018	-0.125	-6.794	0
5 风速(m/s) (70 m)	2.546	0.872	0.388	2.918	0.004
风向 90	-0.004	0.001	-0.014	-3.837	0
风速(m/s) (90 m)	-2.283	0.844	-0.357	-2.704	0.007

其中, 15 min、30 min 代表将该时刻提前 15 min、30 min 的实际功率, 风速(m/s) (70 m)、(m/s) (90 m) 代表 70 m, 90 m 高度风速的测量值; 风向 90 代表 90 m 高度风向的测量值。

将这些因素作为 LSTM 的输入变量, 设置学习率为 0.01, 隐层神经元个数为 50, 输入层神经元个数为 5, 输出层神经元个数为 1, 迭代次数为 200, 得到均方误差为 0.069059, 与 BP 神经网络、ARMA、LSTM 相比, 均方误差较小, 预测精度较高。经过和上面损失函数的对比, 由图 4 和表 5 可以得出经逐步回归挑选出来的输入变量预测出来的结果精度更高, 且更加稳定, 模型训练速度更快。

Table 5. Comparative analysis of prediction accuracy of different prediction models
表 5. 不同预测模型预测精度对比分析

方法	MSE
BP 神经网络	0.1366
ARMA	0.0784
LSTM	0.0864
Stepwise Regression-LSTM	0.0691

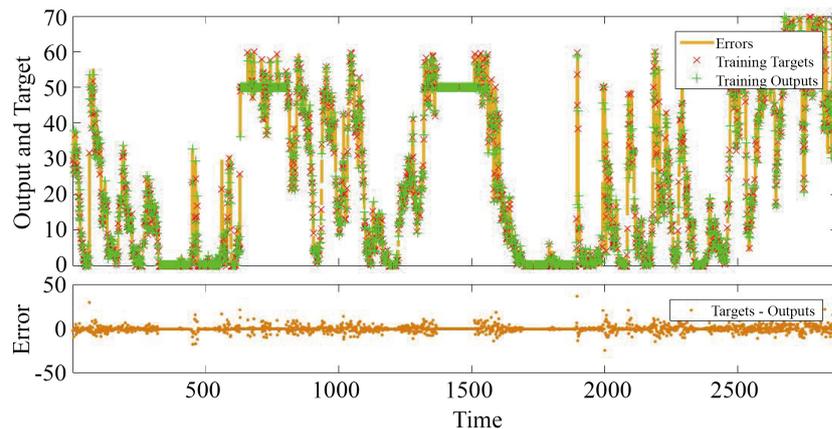


Figure 4. The forecast results and error curves based on stepwise regression and deep learning methods

图 4. 基于逐步回归与深度学习方法的预测结果及误差曲线

4. 结论

风电场输出功率预测对于保持电力系统的功率平衡和经济运行有重要意义。基于本文的研究结果有以下结论，模型本身对预测精度有重要影响。在 LSTM 神经网络模型当中，学习率、隐层神经元的个数都可以影响模型训练的时间以及精度，需要根据实际情况，选择合适的学习率以及隐层神经元个数，才能使模型的预测精度和训练速度提高且较为稳定。预测模型输入变量的选择对预测精度有重要影响。对于任何一种预测模型，选择不同的输入数据对精度有较大影响。在本文当中，输入变量包括不同高度的风速、风向、温度、湿度、气压、空气密度以及历史功率，变量选择过多会导致信息冗余、运算效率降低。风速、风向对风电场的实际输出功率有影响。实测功率作为 LSTM 的输入可以提高预测精度，不同高度的风速、风向数据作为输入变量可以提高预测精度。

参考文献

- [1] 陶玉波, 陈昊, 秦晓辉, 孟昭军. 短期风电功率预测概念和模型与方法[J]. 电力工程技术, 2018, 37(5): 7-13.
- [2] 王仁忠. 风电场风功率预测系统及预测精度[J]. 电子技术与软件工程, 2018(20): 144.
- [3] 付浩, 王圣达, 姜山, 康爱民. 基于 ARMA-LSTM 的光功率预测方法[J]. 光通信技术, 2019, 43(1): 20-23.
- [4] 颜晓娟, 龚仁喜. 基于改进遗传算法寻优的 SVM 风能短期功率预测[J]. 电测与仪表, 2014, 51(8): 38-41+46.
- [5] 戚双斌, 王维庆, 张新燕. 基于支持向量机的风速与风功率预测方法研究[J]. 华东电力, 2009, 37(9): 1600-1603.
- [6] 孟岩峰, 胡书举, 邓雅, 许洪华. 风电功率预测误差分析及预测误差评价方法[J]. 电力建设, 2013, 34(7): 6-9.
- [7] 常鹏, 高亚静, 张琳, 李均强. 基于 EEMD 与时间序列法的短期风电场功率预测[J]. 电力科学与工程, 2012,

28(3): 33-39.

- [8] 范高锋, 王伟胜, 刘纯, 戴慧珠. 基于神经网络的风电功率预测[J]. 中国电机工程学报, 2008, 28(34): 118-123.