

基于 VMD_LSSVM 的日前光伏出力预测

冯耀冬 刘 闯 史 上

华北水利水电大学电气工程学院 河南郑州 450011

摘要: 准确预测光伏电站输出功率是促进光伏并网发电的主要途径。本文提出一种基于变分模态分解的日前光伏功率预测方法。首先利用变分模态分解算法提取功率时间序列中的有效信息。利用 LSSVM 模型对 NWP 数据与分解后的序列进行模型训练, 将预测出的每一条子序列重构得到最终模型的输出。与 VMD_LSTM、VMD_GRU、VMD_ELM 模型的预测结果进行对比分析, 结果表明本文提出的 VMD_LSSVM 模型的预测精度高于上述模型。

关键词: 变分模态分解; 最小二乘支持向量机, 日前光伏功率预测

引言

随着人类能源需求的增长和化石燃料的枯竭, 太阳能作为新能源中的一种, 得到了大规模的开发和利用。由于光伏发电的随机性、波动性和不确定性, 大规模光伏并网给电网的稳定运行带来了挑战^[1]。准确预测光伏输出功率, 减少光伏功率对电网稳定性的影响, 促进电网对光伏功率的消纳, 已成为光伏发电领域的重要研究内容。随着人工智能的快速发展, 当下光伏功率预测方法根据时间尺度可分为: 中长期预测、短期预测、超短期预测^[2]。

最小二乘支持向量机是一种基于统计学习理论和结构风险最小化原则的机器学习方法。它是传统支持向量机 (SVM) 的一个变体, 近年来在处理大数据、深度学习和人工智能领域的问题方面得到了广泛的应用。谢等人提出改进优化觅食算法 - 支持向量机预测模型, 与基础模型对比, 该模型预测误差指标最小^[3]。周等人利用 kmeans 算法等对气象特征进行聚类, 然后利用 SSA 算法优化后的 LSSVM 模型对各类天气进行功率预测, 结果表明, 该模型优于基础模型的预测精度^[4]。

为了获得更高的预测精度, 近年来大多数人员采用混合方法来实现预测光伏发电。其中, 基于时间序列的分解技术得到了快速发展。小波变换、经验模态分解和变分模态分解是最具有代表性的三种分解算法。刘等人将小波变换用于提取光伏出力的细节分量与近似分量, 利用 CNN 和 LSTM 分别提取时频域和长期依赖的特征^[5]。黄等人采用变分模态分解算法对实际功率数据进行分解后, 采用最小二乘支持向量机对每一个分量进行预测, 测试结果表明, 该算法预测误

差比其他算法更小^[6]。陈等人采用变分模态分解分解光伏功率数据, 利用射影算法优化的增强极限学习机对分解后的分量进行预测, 降低了光伏出力预测的非平稳性^[7]。

基于当前研究现状, 本文提出一种基于 VMD_LSSVM 的日期光伏功率预测方法。该方法首先用四分位法对数值天气预报与实际发电数据进行异常值检测与替换; 其次利用 VMD 分解后的子序列用于模型训练。利用 LSSVM 深度挖掘 NWP 数据与分解后的功率时间序列之间的关系, 并将预测后的子序列重构得到最终模型的输出。

1. 所用方法介绍

收集到的 NWP 数据和实际光伏发电功率通常包含着意外误差以及系统误差, 这些误差可能是由于传感器的故障以及其他因素造成的。这时需要对数据进行预处理, 来提高预测精度。本文采用四分位数法来识别功率时间序列中的异常值并进行填充。

1.1 异常值检测与替换

四分位算法又称为箱型法。该方法可以将一组数据从小到大进行排列并进行四等分, 每份序列占整组序列的 1/4。其中将 Q_1 称为下四分位数, Q_2 为中位数, Q_3 为上四分位数。通过上下四分位数可以求得四分位数之差^[8]。

$$I_{OR} = Q_3 - Q_1 \quad (1)$$

通过式 (1) 可以获得异常值的范围, 即上下限。当数据超出这个界限我们可以将其作为异常值处理, 界限可以通过上个公式进行计算。

$$[Q_L, Q_U] = [Q_1 - 1.5 * I_{OR}, Q_3 + 1.5 * I_{OR}] \quad (2)$$

四分位法根据实测的数值天气预报数据进行离群值分

析处理, 任何位于这些定义边界之外的数据都可以被认为是一个离群值, 且不需要事先假定数据服从某种特定的分布形式。因此, 四分位法在识别异常数据时有一定的优越性。考虑到光伏发电数据的高不确定性, 本文用四分位法来识别异常值。

1.2 变分模态分解

VMD 的本质是将一个实值输入信号 f 分解为 k 个分量 (模态), 保证分解序列为具有中心频率的有限带宽的模式分量。VMD 算法分解的具体过程如下:

- 1) 初始化模态 $\mu_k = \{\mu_1, \mu_2, \dots, \mu_k\}$, 各模态中心频率 $w_k = \{w_1, w_2, \dots, w_k\}$, Lagrange 乘子 λ , 迭代次数 n 赋值为 0
- 2) 令 $n=n+1$, 进入循环, 依据式 (3) 和 (4) 对 μ_k 和 w_k 进行更新, 直至分解个数达到 K 时停止迭代。

$$\hat{\mu}_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} \hat{\mu}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{\mu}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{\mu}_k(\omega) \right|^2 d\omega} \quad (4)$$

- 3) 依据式 (5) 对 λ 进行更新。

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \gamma \left(f(\omega) - \sum_{k=1}^K \hat{\mu}_k(\omega) \right) \quad (5)$$

- 4) 给定评价精度 ε , 若满足条件式 (6) 条件则停止迭代, 否则继续循环。

$$\sum_{k=1}^K \left\| \hat{\mu}_k^{n+1} - \hat{\mu}_k^n \right\|_2^2 / \left\| \hat{\mu}_k^n \right\|_2^2 < \varepsilon \quad (6)$$

其中, μ_k 为分解后的单分量调幅调频信号; w_k 为每个单分量调幅调频信号的中心频率; λ 为拉格朗日乘子, n 为迭代次数。 $u_k^{n+1}(w)$, $u_i(w)$, $f(w)$ 和 $\lambda(w)$ 分别是 $u_k^{n+1}(t)$, $u_i(t)$, $f(t)$ 和 $\lambda(t)$ 的傅里叶变换, γ 表述噪声容限。

1.3 最小二乘支持向量机

最小二乘支持向量机 (LSSVM) 是 Suykens 等人在标准向量机的基础上, 将最小二乘线性系统理论引入到模型中, 同时将约束条件中的不等式约束改为等式约束, 采用二次规划方法解决分类和函数估计问题。

假设样本为包含 N 种数据特征, 则 N 个样本组成的样本集数据表示为

$$S\{(x_k, y_k) | k = 1, 2, \dots, N\} \quad (7)$$

其中, x_k 表示输入数据, $x_k \in R^n$, y_k 表示输出数据, $y_k \in R$ 。

然后根据结构风险最小化的原则, 构造 LSSVM 回归方程如下:

$$\begin{cases} \min_{\omega, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \\ s.t. \gamma_k = \omega' \varphi(x_k) + b + e_k \quad k = 1, \dots, n \end{cases} \quad (8)$$

其中, $\varphi(x_k)$ 为输入空间到高维特征空间的映射函数; γ 为正则化参数; e_k 为实际值与回归函数间的误差; ω 为权重向量; b 为偏置向量。

采用拉格朗日法求解该回归方程, 引入拉格朗日乘子 α_k , 函数如下:

$$L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{k=1}^N \alpha_k \times \{ \omega^T \varphi(x_k) + b + e_k - \gamma_k \} \quad (9)$$

得到 LSSVM 的回归估计函数为

$$f(x) = \sum_{k=1}^N \alpha_k K(x_k, x_i) + b \quad (10)$$

其中, $K(x_k, x_i)$ 为满足 Mercer 条件的核函数, 核函数选用经向基 (RBF) 核函数, 求解得到最终的回归模型如下:

$$f(x) = \sum_{k=1}^N \alpha_k \exp\left(-\frac{\|x_k - x_i\|^2}{2\sigma^2}\right) + b \quad (11)$$

其中, σ 为经向基核函数宽度。

2. 案例分析

2.1 测试指标

为了验证预测模型的精度好坏, 本文将平均绝对误差 (MAE)、均方根误差 (RMSE) 作为模型评价指标, 具体公式如下。

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_{Pi} - P_{Mi}}{C_i} \right| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{P_{Pi} - P_{Mi}}{C_i} \right)^2} \quad (13)$$

其中, n 为测试集样本个数, P_{Pi} 为模型预测值, P_{Mi} 为功率实际值, C_i 为光伏电站装机容量。

2.2 预测结果分析

数据来自新疆某光伏电站, 该数据集为 2019 年全年的数据, 包含 8 种特征, 分别为组件温度、环境温度、气压、相对湿度、总辐射、直射辐射、散射辐射和实际发电功率。时间分辨率为 15min, 每个特征包含 35040 个数据点。

首先将数据集进行四分位法异常值检测, 为了保证数据的完整性, 将异常值点使用相邻的非离群值的线性插值进行填充。

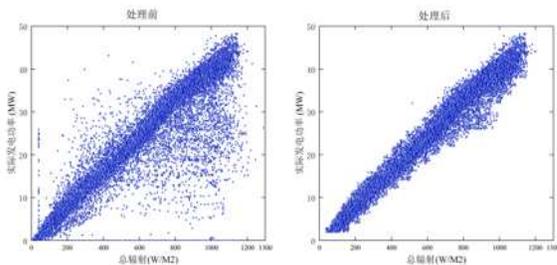


图 3 处理前后总辐射 - 实际发电功率二维散点图

由上图可以看出, 处理后的数据符合实际中光伏电站的运行状态。

由于实际功率与光照强度密切相关, 本文将每天光照强度小于 0 的数据点剔除, 只将白天有光照强度的数据输入模型中训练。剔除每个特征包含 15218 个数据点。初步按月份分成四季, 为了验证模型在不同季节的预测效果, 将每个季度的最后一天作为测试集。将每个季度训练集的功率时间序列进行变分模态分解, 得到不同的子序列后输入到 LSSVM 模型中, 将子序列预测出来的结果进行重构得到本文混合模型最终的输出。

表 1 各预测模型在不同季节下的预测指标值

预测天	模型	MAE	RMSE
3 月 31 日	VMD-LSSVM	0.0158	0.0213
	VMD-LSTM	0.0537	0.0619
	VMD-GRU	0.0086	0.0106
	VMD-ELM	0.3228	0.4604
6 月 30 日	VMD-LSSVM	0.0187	0.0255
	VMD-LSTM	0.0424	0.0505
	VMD-GRU	0.0309	0.0351
	VMD-ELM	0.0895	0.1140
9 月 30 日	VMD-LSSVM	0.0187	0.0255
	VMD-LSTM	0.0770	0.0893
	VMD-GRU	0.0457	0.0545
	VMD-ELM	0.1727	0.2041
12 月 31 日	VMD-LSSVM	0.0293	0.0383
	VMD-LSTM	0.0480	0.0667
	VMD-GRU	0.0597	0.0725
	VMD-ELM	0.1475	0.1782

由表 1 可以看出, 本文提出的混合模型预测效果最好。

下图从左向右依次为 3 月 31 日、6 月 30 日、9 月 30 日和 12 月 31 日, 本文所提出模型的预测效果与其他模型的对比图。其中, 黑色实线为实际功率, 红色实线为本文所提出的混合模型 NGO-VMD-LSSVM 的预测结果, 蓝色划线为 VMD-LSTM 的预测结果, 绿色划线为 VMD-GRU 的预测结果, 粉色划线为 VMD-ELM 的预测结果, 可以看出本文提出的模型与实际功率曲线的拟合效果最好。

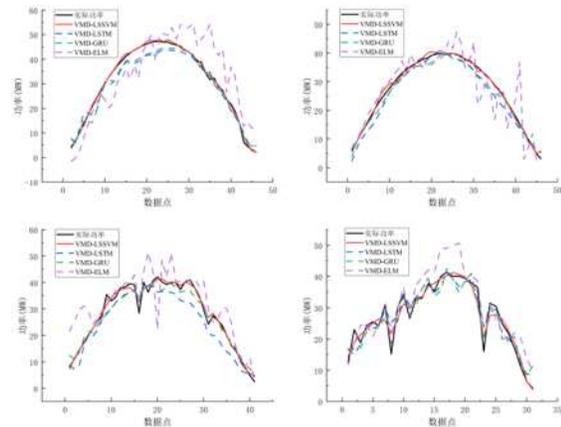


图 4 不同季节下不同模型的预测结果

3. 结论

本文提出了一种基于变分模态分解的日前光伏出力预测方法。主要贡献如下:

1. 为了提高模型的预测精度, 对光伏数据集进行数据预处理, 使其符合正常工作状态。
2. 利用最小二乘支持向量机深度挖掘输入特征与分解后时间序列中的关系, 并将重构数据为最终模型的输出, 提高了预测的精度。

参考文献

[1] 邹鹏辉, 张治, 张显立. 高渗透率分布式光伏系统谐波与电压控制 [J]. 智慧电力, 2020,48(12):40-45.

[2] 蔡源, 吴浩, 唐丹. 光伏发电功率预测方法综述 [J/OL]. 四川电力技术, 1-8[2024-01-12]

[3] 谢波, 高建宇, 张惠娟等. 短期风电功率预测中的 IOFA-SVM 算法实现 [J]. 电子测量技术, 2021,44(12):63-69.

[4] 谢波, 高建宇, 张惠娟等. 短期风电功率预测中的 IOFA-SVM 算法实现 [J]. 电子测量技术, 2021,44(12):63-69.

[5] 刘旭丽, 莫毓昌, 吴哲等. 基于 DWT-CNN-LSTM 的超短期光伏发电功率预测 [J]. 郑州大学学报 (理学版), 2022,

54(04):86-94.

[6] 黄峰, 王睿, 陈晓雷等. 基于 EMD 的 LS-SVM 短期风电功率预测研究 [J]. 湖南工程学院学报 (自然科学版), 2022, 32(4):1-5.

[7] 陈龙, 张菁, 张昊立等. 基于 VMD 和射箭算法优化改进 ELM 的短期光伏发电预测 [J]. 太阳能学报, 2023, 44(10): 135-141.

[8] 马良玉, 耿妍竹, 袁乃正等. 基于四分位与 CFSFDP 的风电机组异常数据清洗方法 [J]. 电力科学与工程, 2023, 39(06):9-16.

作者简介:

冯耀冬, 男, 汉族, 河南新乡人, 硕士研究生在读, 研究方向: 风电功率预测。