

基于极端梯度提升法的某光伏电站 发电量预测

吴江江^{1,2} 兰智¹

(1. 西安理工大学省部共建西北旱区生态水利国家重点实验室 2. 中国能源建设集团投资有限公司西北分公司)

摘要:为解决复杂气象因素下光伏电站发电量预测结果不理想及波动较大的问题,针对陕西洛南某光伏厂区发电情况,基于XGBoost模型建立光伏电站发电量预测模型并对其进行相关评价及验证。研究表明,XGBoost模型在光伏电站发电量预测方面存在较好的精度与可靠性,相较于其他预测模型(支持向量机模型、决策树模型、随机森林模型、BP神经网络模型),XGBoost模型预测值与实际测定值的相对误差最小,相对误差可以控制在6%以内,决定系数最高,证实了XGBoost模型在光伏电站发电量预测方面存在较好的精度与可靠性。本研究可为此类光伏电站厂区的发电量预测及管理运行提供参考。

关键词:光伏电站;机器学习模型;发电量预测;气象因素;极端梯度提升法

0 引言

为改善升级我国能源消费结构,清洁能源的应用势在必行^[1]。作为一种方便高效的可再生能源,光伏能源已得到广泛关注。而作为光伏发电系统的核心环节,光伏电站发电量的精准预测对于电力系统的运行和管理显得尤为重要^[2-4]。截至目前,主流预测方法主要分为数学统计法和物理法两部分^[5]。数学统计法指利用电站历史数据预测发电功率;物理法指预测太阳辐照量,结合光电转换效率公式计算发电功率。诸多学者针对光伏电站发电量影响因素作用程度展开研究。周强^[5]等人研究发现太阳辐照度是影响电站发电效率最为关键的因素之一,其变化呈周期性,与电站发电效率呈正相关关系。何旻^[6]在研究中指出,日常照射温度、空气湿度和场区内部风速对电站发电

量也存在一定的影响,温度的变化会影响光伏组件电流,同时改变电压大小,导致发电功率改变,空气湿度的变化会影响太阳辐照度和温度,风速的改变会影响日常温度高低,各因素之间相互作用影响组件发电效率,致使电站发电量发生改变。综上所述,由于各影响因素所存在的不确定性,综合考虑各因素权重占比,进行光伏电站的发电量预测是非常有必要的。

目前,相关学者关于采用机器学习进行发电量预测已取得一定进展。周强^[5]等人采用基于小波神经网络预测方法,发现GA-WNN预测模型具有较高的精度,适用于光伏电站的发电量预测。宰红斌^[7]等人根据监测数据,基于GA-SVM算法,并结合GA-Elman神经网络算法估计输电线路的温度,达到了防治采空区灾害的目标。李军^[8]等人建立反向传播神经网络分

析相关气象数据进行发电量预测，发现传统的BP神经网络所得数据存在精度不够的弊端。

综上所述，为寻求一种合适算法进行数据建模以确保预测的可靠性，本文针对陕西商洛某光伏电站发电量预测偏差较大及气象影响因素易波动的问题，采用极端梯度提升法进行光伏电站某时段内发电量预测，并结合陕西商洛夏秋季该电站气象数据（太阳辐射度、温度、空气湿度和风速）及实际测定发电量进行分析，结果表明，极端梯度提升法（XGboost）具有对训练样本数目及特征数据类型需要较小，训练时间低、适用范围广、精度高的优点，能对相关电站的稳定运行提供帮助。

1 预测模型建立

XGBoost（极端梯度提升法）是采用梯度提升为一般原理设计的机器学习模型，采用强弱学习器相结合的方法。梯度提升树以顺序方式形成，对数据进行学习及改善迭代预测，并结合创建树，通过管理模型的复杂性和减少内置调节过拟合来提高预测精度^[9]。

设 x_1 为自变量，特征因子为太阳辐射度、温度、湿度和风速，因变量 y_1 为该光伏场区特征时间段（10:00~17:00）内光伏电站平均发电功率，建立过程如下^[9]：

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

式中， f_k 为决策树； F 为决策树所有相关函数； \hat{y}_i 为第 i 天场区特征时间段内光伏电站平均发电功率的预测值。训练过程中目标函数如下式：

$$obj(\theta) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

式中， l 为损失函数； Ω 为正则化项； y_i 为第 i 天特征时间段内光伏电站平均发电功率测定值。当第 k 次迭代时，第 i 天样本预测公式为：

$$\hat{y}_i^k = \hat{y}_i^{(k-1)} + f_k(x_i) \quad (3)$$

式中， $\hat{y}_i^{(k-1)}$ 为经历 $(k-1)$ 次迭代后来自先前累积模型的电站平均发电功率预测值。因此，改写目标函数如下：

$$obj(\theta) = \sum_{i=1}^n l(\hat{y}_i^{(k-1)} + f_k(x_i), y_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

$\Omega(f_k)$ 可表示为：

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w_i\|^2 \quad (5)$$

式中， T 为叶的总数； λ 和 γ 为惩罚系数。

提取第 k 棵分类回归树的电站平均发电功率预测值进行优化。次元二阶泰勒展开式近似表达：

$$L^{(k)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(k-1)} + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)) \right] + \Omega(f_k) \quad (6)$$

式中， g_i 与 h_i 分别为第 i 天电站平均发电功率预测值的一阶梯度与二阶梯度，计算公式如下：

$$g_i = \partial_{y^{(k-1)}} l(y_i, \hat{y}_i^{(k-1)}) \quad (7)$$

$$h_i = \partial^2_{y^{(k-1)}} l(y_i, \hat{y}_i^{(k-1)}) \quad (8)$$

通过代数计算得到第 k 次迭代中损失函数极值点的数值为：

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (9)$$

得到最优解目标函数为：

$$obj(\theta) = - \frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (10)$$

2 电站发电量预测

本次预测研究选用陕西洛南某光伏电站的日平均发电功率作为发电量预测的评价参数。使用该年度内秋夏两季，三种气象条件下（晴天、阴天、雨天）

光伏电站特征时间段内（10:00~17:00）日平均发电功率为基础数据，单次工况各取三组样本作为平行数据，不同工况下各天气类型预测参数如表 1 所示。

表 1 夏季和秋季天气类型预测参数

天气	温度 / °C	湿度 / (%)	风速 / (m·s ⁻¹)	太阳辐照度 / (W·m ⁻²)
晴天	30.3	62.7	0.99	236.4
阴天	28.8	69.7	1.4	170.9
雨天	25.9	85.5	0.47	132.5

在进行模型输入前，完成数据标准化处理，符合模型标准后作为初始数据分别输入 XGBoost、支持向量机、决策树、随机森林和 BP 神经网络等各预测模型，预测模型中自变量特征因子为太阳辐照度、温度、湿度和风速，因变量为该厂区特征时间段（10:00~17:00）内光伏电站平均发电功率。表 2 为模型参数设置，随机选用 70% 数据为训练集，剩余 30% 为测试集，用于验证模型精度。

表 2 模型最优参数设置表

参数名	XGBoost	支持向量机	决策树	随机森林	BP 神经网络
训练用时	0.814s	0.134s	0.651s	0.111s	0.348s
数据切分	0.7	0.7	0.7	0.7	0.7
数据洗牌	是	是	是	是	是
交叉验证	否	否	否	否	否
学习率	0.1	0.1	0.1	0.1	0.1
节点特征采样率	1	/	1	1	/
叶子节点中样本的最小权重	0	/	0	1	/
树的最大深度	10	/	10	10	/

各预测模型训练完毕后应用于测试集，可以得到相关特征集中不同特征的相关程度。结果如图 1 所示，发现日太阳辐照度权重占比可以达到 72.3%，电站厂区日常温度、湿度和风速权重占比基本维持在 10% 左右。日太阳辐照度对电站发电功率的影响显著大于其他因素，与相关学者研究结论一致^[5]。各预测模型预测日平均发电功率与实测值对比结果如图 2 所示。可以发现，各模型训练完成后，决定系数均接近

于 1，表明各模型的预测精度均较高。且与支持向量机、决策树、随机森林和 BP 神经网络等预测模型对比，XGBoost 模型训练所得预测值与光伏厂区的日平均发电功率实际测定值相差波动范围最小，决定性系数可达 0.99。

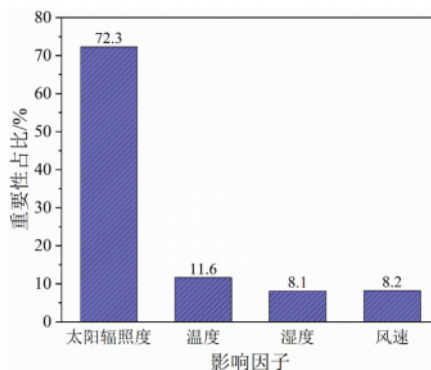
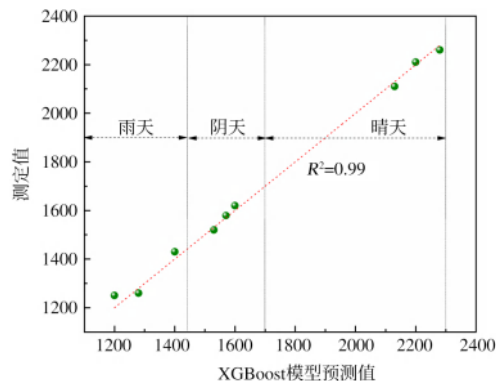
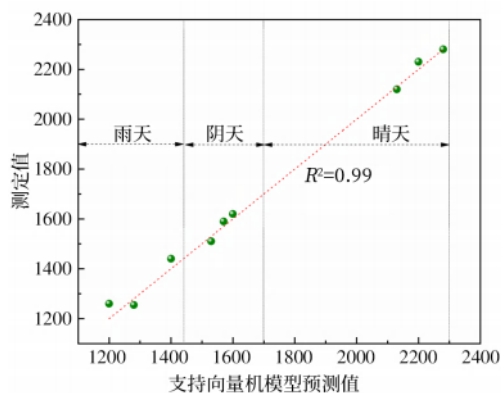


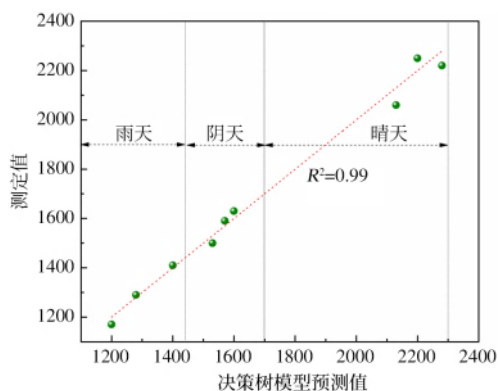
图 1 特征因子权重排序



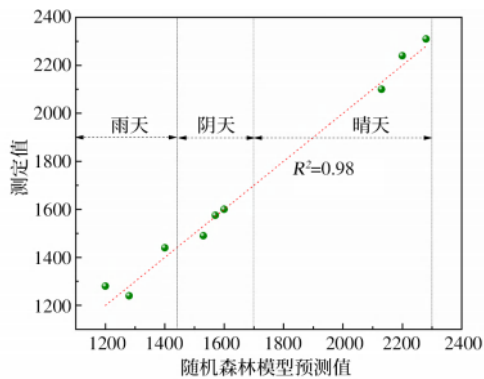
(a) XGBoost 模型预测



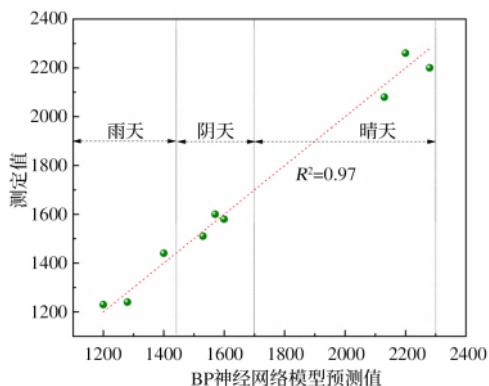
(b) 支持向量机模型预测



(c) 决策树模型预测



(d) 随机森林模型预测值



(e) BP神经网络模型预测

图2 预测结果对比图

采用训练完毕的各训练模型输入相应训练天数，可以得到不同工况下光伏电站的日发电功率，进而求得日发电量，并与实际测定值对比，所得结果如表3所示。

由表3可以发现，XGBoost模型预测值与实际测定值误差范围在6%以内。晴天工况下的日平均发电量相较于阴天和雨天两种工况存在显著提升，间接证明了日太阳辐照度是影响光伏电站发电效率的关键因素。且相较于阴天和雨天两种工况，厂区晴天环境下发电量增幅区间在35%~42%和68%~75%之间，差距尤为明显。对比实际测定值，支持向量机模型、决策树模型、随机森林模型、BP神经网络模型相对误差均处在7%~18%，误差区间明显高于XGBoost模型，证明XGBoost模型存在预测精度更高的优势。

表3 模型预测结果表

(单位: kWh)

天气类型	测定值	XGBoost 预测值	支持向量机预测值	决策树 预测值	随机森林 预测值	BP神经网络 预测值
晴天-1	14.77	15.56	17.49	13.58	17.15	16.49
晴天-2	15.47	15.40	17.31	14.23	16.97	16.32
晴天-3	15.82	14.91	16.76	14.55	16.43	15.80
阴天-1	10.64	11.20	12.59	9.78	12.34	11.87
阴天-2	11.05	10.99	12.35	10.16	12.11	11.64
阴天-3	11.34	10.71	12.04	10.43	11.80	11.35
雨天-1	8.75	8.80	9.89	8.05	9.70	9.32
雨天-2	8.82	8.96	10.07	8.11	9.87	9.49
雨天-3	8.61	8.40	9.44	7.92	9.26	8.90

3 XGboost 模型预测评价

为了验证XGBoost模型的预测合理性。采用平均绝对误差 (MAE)^[10]、均方根误差 (RMSE)^[11]、决定系数 (R^2)^[12] 等三个指标来对模型的泛化性、精确性、稳定性进行相关评价。各指标计算公式如下:

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

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

式中, y_i 和 \hat{y}_i 分别为日平均发电功率的实际测定

值和发电功率模型预测值； \bar{y}_i 为测定值的平均值； n 为测定值数量。

决定系数 R^2 越靠近 1，表示模型预测精度也高。均方根误差值越小说明误差越小，即模型预测精度较为可靠。为验证 XGBoost 模型合理性与优势，本文采用该厂区特征时间段（10:00~17:00）内光伏电站平均发电功率作测试数据集，将 XGBoost 模型与其他模型进行精度对比，模型参数见表 3。以上述三个指标作为评价标准，训练完毕后各模型对应评价指标结果如图 3 所示。

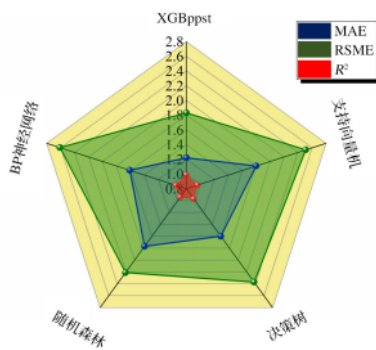


图 3 各模型评价参数对比图

可以发现，训练完成后各模型精度均符合精度要求（ > 0.9 ），其中，XGBoost 模型精度最高，决定系数达到 0.99。且 XGBoost 模型的均方根误差最小，相较于支持向量机模型、决策树模型、随机森林模型、BP 神经网络模型分别降低了 0.681、0.543、0.382、0.762。证明 XGBoost 模型预测的合理性，为相关光伏电站发电量的预测提供新的思路。

4 结束语

本文就光伏电站发电量预测精度不够及气象因素复杂等一系列问题，提出一种基于极端梯度提升法的光伏电站发电量预测方法。以特征时间段（10:00~17:00）日平均发电功率为评价参数，建立并通过 XGBoost 模型预测不同气象条件下厂区内的日发

电量，完成模型可靠性验证，预测结果精度较高。并采用相关评价指标（RSME、MAE、 R^2 ），与诸多模型对比确定了 XGBoost 模型的优势。

参考文献

- [1] 徐修衍. 光伏发电的发展现状及控制方法研究[J]. 光源与照明, 2022 (4): 96-98.
- [2] 王婧骅, 张娟, 赵婉茹, 等. 基于时间序列的分布式光伏电站发电数据采集方法[J]. 电网与清洁能源, 2022, 38 (6): 137-142.
- [3] 马胜红, 赵玉文, 王斯成, 等. 光伏发电在我国电力能源结构中的战略地位和未来发展方向[J]. 电器工业, 2005 (6): 27-33.
- [4] 华龙, 齐冲, 刘雪娇. 基于 RBF 神经网络的太阳能发电量预测[J]. 科技和产业, 2022, 22(7): 375-380.
- [5] 周强, 张晓忠, 陈久益, 等. 基于遗传算法小波神经网络的光伏电站发电量预测方法[J]. 智慧电力, 2024, 52 (4): 78-84.
- [6] 何旻. 分布式光伏电站中长期发电量影响因素分析及预测方法研究[D]. 杭州: 浙江工业大学, 2020.
- [7] 宰红斌, 吴浩林, 王昊, 等. 基于改进机器学习的输电线路弧垂温度估计方法[J]. 电力工程技术, 2022, 41 (2): 209-214, 223.
- [8] 李军, 许继生, 王生渊. 一种并网型光伏电站光功率及发电量预测的方法[J]. 青海电力, 2010, 29 (2): 18-20.
- [9] 覃源, 薛存, 李遥, 等. 盐冻耦合作用下水工混凝土耐久性及寿命预测[J]. 水力发电学报, 2024, 43 (2): 110-122.

(下转第 38 页)

(上接第 10 页)

- [10] 纪德洋, 金锋, 冬雷, 等. 基于气象相似性的光伏电站输出功率估计 [J]. 太阳能学报, 2022, 43 (5): 173-179.
- [11] 李遥, 李照荣, 王小勇, 等. 基于斜面辐射算法的短期光伏功率预测方法研究 [J]. 干旱气象, 2020, 38 (5): 869-877.
- [12] CHEN Lichun, TYLER L M, GHANTOUS R M, et al. Measuring damaging freeze-thaw cycles in the field[J]. Construction and Building Materials, 2023 (387) .
- [13] 张苏宁, 王芳, 朱燕, 等. 基于极端梯度提升的跨地区多种类电力需求预测 [J]. 计算机与现代化, 2022 (3): 18-22, 29.

(收稿日期: 2024-05-25)