



基于改进 Stacking 与误差修正的短期太阳辐照度预测

摘要

为提高光伏电站辐照度预测的准确性和可靠性,提出一种基于改进 Stacking 集成学习与误差修正的短期辐照度预测模型.首先使用梯度提升决策树(Gradient Boosting Decision Tree, GBDT)对原始数据集进行特征筛选,清除冗余特征,提高预测精度和运算效率;然后建立改进 Stacking 辐照度预测模型,根据 K 折交叉验证下初级层不同预测模型所表现出预测准确度的差异性,对预测结果进行赋权,并对 Stacking 第一层输入第二层的训练集数据采用 Box-Cox 变换处理,以提高预测的正态性和同方差性;最后提取历史预测误差数据,采用随机森林(Random Forest, RF)构造误差模型,进一步提高预测精度.实验结果表明,该模型相比传统模型和经典 Stacking 模型其预测性能有了较大的提升.

关键词

太阳辐照度;光伏发电;Stacking 算法;回归预测算法;交叉验证

中图分类号 TP391;TM615

文献标志码 A

收稿日期 2023-02-10

资助项目 国家重点研发计划(2018YFC0116100);湖北省重点研发计划(2020BAB114);湖北省教育厅科学研究计划重点项目(D20211402)

作者简介

王珊珊,女,博士,副教授,研究方向为机器学习与深度学习.wangshanshan@hbut.edu.cn

¹ 湖北工业大学 电气与电子工程学院/太阳能高效利用湖北省协同创新中心,武汉,430068

0 引言

太阳能因其典型的波动性与间歇性^[1],造成光伏发电系统输出功率的不稳定性,对光伏发电并网与电网的安全稳定运行构成巨大挑战^[2-3],同时也阻碍了大规模光伏发电并网.在光伏功率预测的众多影响因素中,太阳辐照度的影响是最直接、最显著的,因此准确的辐照度预测能够提高光伏发电系统输出功率的预测精度,有着重要的理论与应用价值^[4].

近年来,随着机器学习技术的兴起,国内外许多学者将以 SVR(Support Vector Regression)^[5]、神经网络和随机森林为代表的机器学习算法用于辐照度预测问题中^[6-8].文献[9]通过对9项气象参数的不同组合作为输入,对模型的预测精度进行分析,提出一种基于非线性自回归神经网络的辐照度预测模型,有效提高了预测精度.文献[10]利用 EMD(Empirical Mode Decomposition, EMD)和 LMD(Local Mean Decomposition)将原始数据分解为多个分量序列,然后对各分量分别进行 LSSVM(Least Square Support Vector Machine)预测,最后将各分量的预测结果进行叠加得到最终预测值,相比 LSSVM 单独预测,精度有了明显提升.单一的预测模型都是对特定假设空间进行预测,所以用单一模型来预测辐照度不可避免会存在预测误差.而集成模型相比单一模型能够集成多个模型的不同特点,对各个模型取长补短,从而提高预测性能,Stacking 模型便是目前最热门的集成模型之一.文献[11]通过构建一种基于同质 SVM(Support Vector Machine)弱学习器的 Stacking 模型,得到了比单一模型更精确的预测效果.文献[12]提出一种新的向量表示法来稳定模型数据维度,并根据预测精度来调整 Stacking 基模型赋权,减少了输出的噪声和时间开销.目前大多研究专注于提升预测模型的精度来提高辐照度预测效果,却忽略了模型的预测误差中也存在非常多的有用信息,这导致现有研究难以进一步提升短期辐照度预测的准确度.

为了解决以上问题,本文提出一种基于改进 Stacking 集成学习与误差修正的短期辐照度预测模型.在数据优化层面,利用梯度提升决策树(Gradient Boosting Decision Tree, GBDT)模型计算辐照度数据集各分量的重要度并排序,清除其中的冗余特征,提高预测精度和运算效率.在算法创新方面,预测模型采用 Stacking 集成模型将4种差异性较大的算法融合,通过异质集成得到优于个体学习器的预测精度

和泛化能力.针对 Stacking 模型平均处理测试集预测结果而带来的掩盖学习器优劣的问题,提出对初级层的输出根据各模型的预测精度进行加权处理.同时,将 Box-Cox 变换嵌入 Stacking,此举可有效提高数据的正态性、可加性和同方差性.在误差修正层面,针对 Stacking 模型的预测误差构建了基于随机森林(Random Forest, RF)的辐照度误差预测模型,并最终将改进 Stacking 模型的预测结果与辐照度误差预测结果进行叠加以获得最终预测结果.实验结果表明,该集成模型具有优于单一模型和传统 Stacking 模型的预测精度.

1 相关理论和方法

1.1 GBDT 算法

GBDT 是一种由多个弱学习器组成的迭代决策树算法.该方法利用梯度增强算法来最小化损失函数,达到逼近真实值的目的,具有灵活性高、鲁棒性强、预测精度高等特点.

将 GBDT 应用到特征重要度计算中,通过对决策树中分裂节点增益进行计算并积累求和,从而对某个特征进行重要度评价.其中,特征 j 的全局重要性是以特征重要度平均值来度量的,其计算公式为

$$\hat{I}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{I}_j^2(T_m), \quad (1)$$

其中, $\{T_m\}_1^M$ 表示决策树的集合, M 是树的数量.在单棵树中,特征 j 的重要度如下:

$$\hat{I}_j^2(T) = \sum_{i=1}^{J-1} \hat{I}_i^2(v_i = j), \quad (2)$$

其中, J 代表叶子节点的数量, v_i 代表与 j 关联的特征, \hat{I}_i^2 代表节点分裂后平方损失的减少值.

1.2 RF 随机森林

随机森林^[13]是一种基于决策树的集成式机器学习算法,它在每一棵回归树建立随机采样样本空间与特征空间,随机属性的引入减少了回归树模型间的相关关系,通过结合大量回归树来提高模型泛化能力,从而使算法具有效率高、精度高的特点.

1.3 Box-Cox 变换

Box-Cox 变换是一种基于极大似然估计的数据变换技术,计算过程简单且无需先验信息,能够有效提升观测的同方差性、正态性和可加性^[14].

Box-Cox 变换一般形式如下:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log y, & \lambda = 0, \end{cases} \quad (3)$$

式中, λ 为变换参数, y 为原始因变量, $y(\lambda)$ 为新变量, Box-Cox 逆变换为

$$y = \begin{cases} ((\lambda y(\lambda) + 1)^\lambda)^{\frac{1}{\lambda}}, & \lambda \neq 0, \\ \exp(y(\lambda)), & \lambda = 0. \end{cases} \quad (4)$$

参数 λ 采用最大似然估计进行计算,构造似然函数 $L^*(\lambda)$ 如下:

$$L^*(\lambda) = -\frac{n}{2} \log_{10}(e^2) + (\lambda - 1) \sum_{i=1}^n \log_{10}(y_i), \quad (5)$$

式中, n 表示采样次数, e^2 表示 $y(\lambda)$ 方差的极大似然估计值,通过式(5)求解出使得 $L^*(\lambda)$ 取最大值的最优 λ , 利用该参数进行 Box-Cox 变换可以很大程度上提升数据的正态性和数据间的相关性.

1.4 其他模型

1) K 最邻近算法(KNN)原理

K 最邻近算法(k -Nearest Neighbor, KNN)作为一种经典的机器学习算法在理论上十分完善,具有算法简单、容易实现、训练效率高、对异常值不敏感等特点.该算法核心思想是对不同特征值之间的距离进行度量,通常采用欧式距离和曼哈顿距离.

2) XGBoost 算法

XGBoost(eXtreme Gradient Boosting)算法是基于 GBDT 算法改进而来的,它通过增加正则项控制模型计算复杂度,并利用二阶泰勒展开式并行计算特征分裂增益以提高模型预测精度、减少计算时间.

3) 支持向量回归机(SVR)原理

支持向量机应用于回归问题形成了 SVR,输入向量按照既定非线性映射映射至高维特征空间进行线性回归以获得空间内非线性回归的效果.在回归运算时,将不敏感损失函数引入到构造中,以搜索到最优的分类面使得训练样本和该分类面之间的综合误差达到最小值.支持向量回归在小样本、高维、复杂数据上进行非线性回归预测时展现出了优异的性能.

4) 岭回归原理

岭回归(Ridge Regression)是一种正则化方法,通过舍弃最小二乘的无偏性和部分精确度,获得了效果更好且回归系数更符合实际的回归过程.岭回归通过对损失函数增加惩罚项以控制线性模型复杂程度,增强了模型的稳健性.

2 基于 Box-Cox 变换的改进 Stacking 短期辐照度预测方法

2.1 Stacking 集成学习模型

Stacking 算法是一种分层集成的方法.不同于

Bagging 和 Boosting 算法整合同类型模型, Stacking 算法能够集成异质模型^[15]. 在 Stacking 集成模型(图1)的训练过程中, 通常使用 K 折交叉验证法来划分数据集和进行模型训练, 以减少过拟合的风险. 首先将原始数据集以 8:2 的比例划分为训练集和测试集, 接着将训练集等分地划分为 k 个子集, 分别选择其中 $k-1$ 个子集的并集作为训练集, 余下的 1 个子集作为验证集, 由此可获得 k 组训练集和验证集. 对每个基模型都采用这 k 组训练集和验证集进行学习器的训练和验证, 并将预测结果配合样本真实值标签构造为新的训练集输入第二层元学习器训练, 最终得到的预测结果即为 Stacking 模型的最终预测结果. Stacking 算法最突出的优势是集不同模型之长, 能够对原始数据进行多角度的分析, 使得模型获得相比基模型更好的预测性能. 所以基学习器应选择性能优越且原理各异的模型, 元学习器则应选择泛化能力强的算法, 以融合各基学习器预测的优点, 达到最优的预测效果. 本文选择 SVR、XGBoost、KNN、岭回归作为基学习器, 元学习器则使用泛化能力较强的岭回归算法.

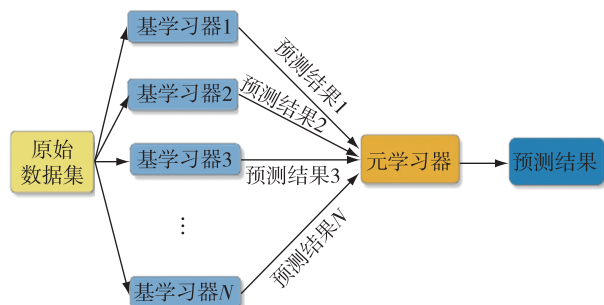


图1 Stacking 模型结构

Fig. 1 Stacking model structure

2.2 改进 Stacking

2.2.1 基于 GBDT 算法的特征筛选

利用 GBDT 对辐照度数据进行特征选择, 对各个特征进行重要度排序. 各特征与实际辐照度之间的相关系数如表 1 所示. 由表 1 的相关系数数据, 剔除掉风向、气压 2 个相关性弱的特征, 最终选择温度、时刻、风速、湿度作为模型输入, 降低计算复杂度, 提升模型训练效率.

2.2.2 基模型赋权与 Box-Cox 变换

Stacking 算法中元学习器输入向量来自基学习器的输出, 使得训练数据被各层学习器重复学习, 造成模型的严重过拟合. 因而需要在模型建立前对数据进行交叉验证划分, 提升模型的泛化能力^[16]. 如果

表 1 相关系数

Table 1 Correlation coefficients	
变量	与辐照度的相关系数
辐照度	1
温度	0.512
时刻	0.455
湿度	0.209
风速	0.131
风向	0.006
气压	0.007

训练得到的基学习器模型预测效果比较好, 那么该模型的预测结果也会更接近真实值. 但传统的 Stacking 模型初级层对测试集预测结果的处理方式为直接平均处理, 使得优秀模型的优越性被其他模型掩盖. 同理, 效果较差的模型也会因与其他模型平均而掩盖其预测性能的不足. 因此, 本文针对同一基学习器在每折上训练得到的不同预测模型, 依据其验证集预测值和真实值之间的误差求得权值, 再给测试集对同一种基模型的不同预测结果赋权. 设学习器所训练出的模型预测误差为 $e_{ii}(i=1, \dots, k)$, 以此误差在 $e_{1i}, e_{2i}, \dots, e_{ki}$ 精度总和中所占比例来分别确定权值 $\rho_1, \rho_2, \dots, \rho_k$, 然后对测试集的输出结果赋权, 可以使精度更高的预测模型发挥其优越性, 增大其对最终结果的影响, 并降低精度较低的模型对最终输出的影响.

针对数据特征与辐照度数值相关程度不高的问题, 本文对第一层输入第二层的训练集进行 Box-Cox 变换处理, 此举可提高训练集的正态性和可预测性, 提升各个特征与辐照度数据的内在联系, 进一步减小预测误差. 经过 Box-Cox 变换后的训练集输入元学习器, 训练完成后用此学习器输入经加权处理后的测试集进行预测, 得到预测结果. 精度加权改进和 Box-Cox 预处理流程如图 2 所示.

精度加权和 Box-Cox 改进后的 Stacking 预测模型的流程如算法 1 所示.

2.2.3 基于 RF 模型的误差修正方法

通过 RF 模型来寻找辐照度预测系统误差的变化规律, 有助于发现提高模型效果的有益规律, 达到提升预测精度的目的.

辐照度预测误差:

$$e = p' - p, \quad (6)$$

式中, e 表示辐照度预测误差, p' 表示辐照度预测值, p 表示辐照度真实值.

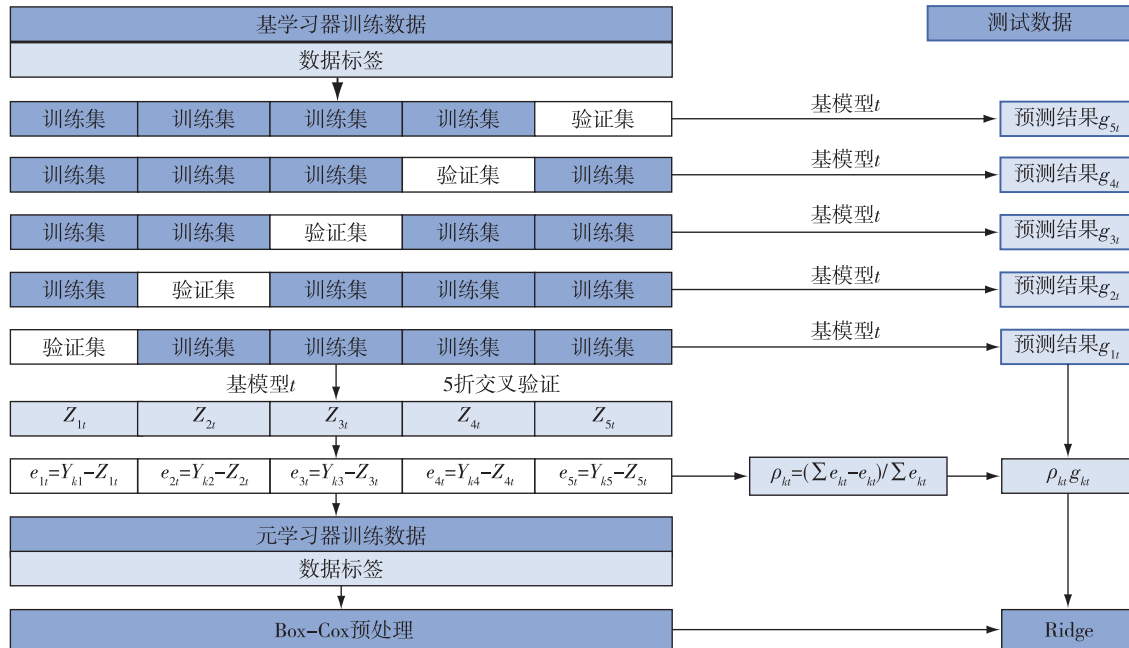


图2 精度加权改进和 Box-Cox 预处理

Fig. 2 Accuracy-weighted improvement and Box-Cox preprocessing

算法1: 基于 Box-Cox 的改进 Stacking 模型

```

Input: 训练集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}$ 
1 for  $t = 1, 2, \dots, T$  do
2   for  $i = 1, 2, \dots, k$  do
3     第零层第  $t$  个同质基学习器  $k$  次训练学习:
        $\{Z_{it} = h_i(x_m, y_m), i = 1, \dots, k, t = 1, \dots, T\} \rightarrow \{Z_{1t}, Z_{2t}, \dots, Z_{kt}\}$ 
4     第零层第  $T$  个异质基学习器学习训练并构成新的数据样本:
        $\{Z_{it} = h_i(x_m, y_m), i = 1, \dots, k, t = 1, \dots, T\} \rightarrow \{(Z_{11}, Z_{k1}), \dots, (Z_{1T}, \dots, Z_{kT})\}$ 
5     构成新的数据样本:  $D_{\text{new}} = \{((Z_{11}, \dots, Z_{k1}), \dots, (Z_{1T}, \dots, Z_{kT}), y_m)\}$ 
6     对新数据样本作 Box-Cox 变换处理后输入第一层
7     第一层预测算法  $\delta$  训练得到模型:  $H = \delta((Z_{11}, \dots, Z_{k1}), \dots, (Z_{1T}, \dots, Z_{kT}))$ 
8     计算第零层同质基学习器的误差:
        $\{[(y_1, y_2, \dots, y_k), \dots, (y_1, y_2, \dots, y_k)] - [(Z_{11}, Z_{21}, \dots, Z_{k1}), \dots, (Z_{1T}, Z_{2T}, \dots, Z_{kT})]\} =$ 
9      $\{(e_{11}, e_{21}, \dots, e_{k1}), \dots, (e_{1T}, e_{2T}, \dots, e_{kT})\}$ 
10    权重计算:
        $\rho_{11} = (e_{21} + \dots + e_{k1}) / (e_{11} + e_{21} + \dots + e_{k1})$ 
        $\vdots$ 
        $\rho_{k1} = (e_{11} + \dots + e_{k-1,1}) / (e_{11} + e_{21} + \dots + e_{k1})$ 
        $\vdots$ 
        $\rho_{kT} = (e_{1T} + \dots + e_{k-1,T}) / (e_{1T} + e_{2T} + \dots + e_{kT})$ 
11    对第零层测试集数据输出进行权重分配:
        $(\rho_{11}g_{11}, \rho_{21}g_{21}, \dots, \rho_{k1}g_{k1})$ 
        $(\rho_{12}g_{12}, \rho_{22}g_{22}, \dots, \rho_{k2}g_{k2})$ 
        $\vdots$ 
        $(\rho_{1T}g_{1T}, \rho_{2T}g_{2T}, \dots, \rho_{kT}g_{kT})$ 
12  end
13 end
    
```


计算辐照度初步预测值 p' 与实际值 p 误差 e , 将其作为校正因子, 将辐照度校正的补偿值 e 添加到预测值 p' 中, 获得校正后的辐照度预测值.

2.2.4 改进 Stacking 模型总体流程

基于上述的 GBDT 算法、基模型赋权与 Box-Cox 变换、RF 误差修正, 本文搭建了短期辐照度预测模型, 如图 3 所示.

具体思路如下:

1) 利用 GBDT 算法对辐照度原始数据进行特征筛选, 去除掉相关程度较低的冗余特征后, 其余数据输入改进 Stacking 模型.

2) 采用 XGBoost、SVR、岭回归、KNN 作为 Stacking 的基学习器, 岭回归作为元学习器. 对于同一基学习器用不同样本进行训练, 基学习器训练出的模型对验证集的预测误差为 $e_{i_k} (i = 1, \dots, k)$, 以此误差在 $e_{1i}, e_{2i}, \dots, e_{ki}$ 精度总和之中所占比值来确定权值 $\rho_1, \rho_2, \dots, \rho_k$, 然后对第一层输入第二层测试集的输出结果进行赋权.

3) 对第一层输入第二层的训练集进行 Box-Cox 变换来提高训练集的正态性和可预测性.

4) 经过 Box-Cox 变换处理的训练集输入元学习器, 训练完毕之后, 用此学习器对加权过后的测试集进行预测, 得到预测结果.

5) 采用上述模型进行初步预测后, 将初步预测的误差数据输入 RF 模型中训练, 由此获得误差分布模型. 将初步预测结果与误差预测值进行叠加得到

最终预测结果. RF 误差修正模型可以有效弥补初步预测模型本身存在的误差, 进一步提高预测的准确度.

3 算例分析

3.1 数据选取

为验证本文模型对于短期辐照度预测的有效性, 将中国北方某光伏电站作为具体研究对象. 选取 2017 年 2 月 1 日至 2018 年 1 月 31 日的辐照度数据作为原始数据集, 共 2 920 个样本数据. 辐照度数值采用太阳总辐射量, 采样间隔为 3 h, 每日采样 8 个点, 包括地表太阳辐照度和风速、风向、温度、湿度、气压共 5 项观测指标. 将数据集按 8:2 划分, 前 80% 的数据作为训练集, 后 20% 数据作为测试集, 进行回归预测.

3.2 模型评价指标

本文选取均方根误差 (E_{RMSE})、平均绝对误差 (E_{MAE}) 和决定系数 (R^2) 作为辐照度预测模型的效果评价指标, 其中: R^2 用来检验模型的拟合度, R^2 越大表明模型拟合程度越好; MAE、RMSE 则用来检验预测模型的精度, 它们值越小说明精度越高. 具体的计算公式如下:

$$E_{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|, \tag{7}$$

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \tag{8}$$

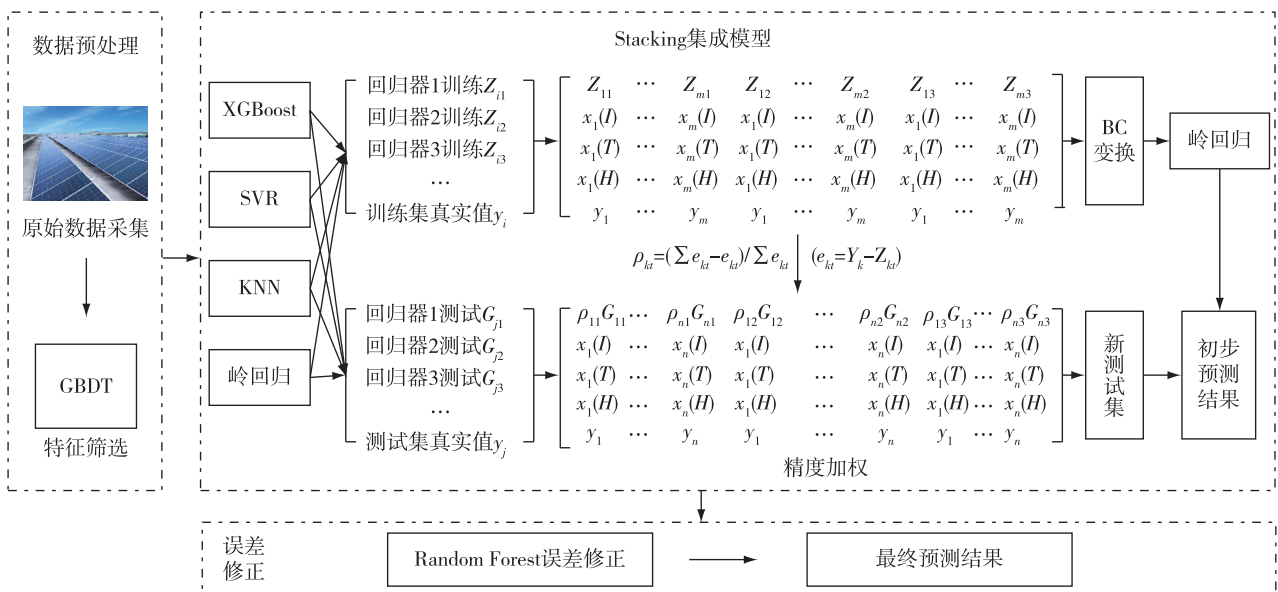


图 3 基于改进 Stacking 与误差修正预测模型流程

Fig. 3 Prediction process based on improved Stacking with error correction

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (9)$$

3.3 仿真实验及结果分析

首先,设计第一组实验,对比 Stacking 模型与其各个基模型在辐照度短期预测上的性能,各模型预测值与辐照度实际值的横向对比结果如图 4 所示,各模型的绝对误差如图 5 所示,各模型的评价指标值如表 2 所示。

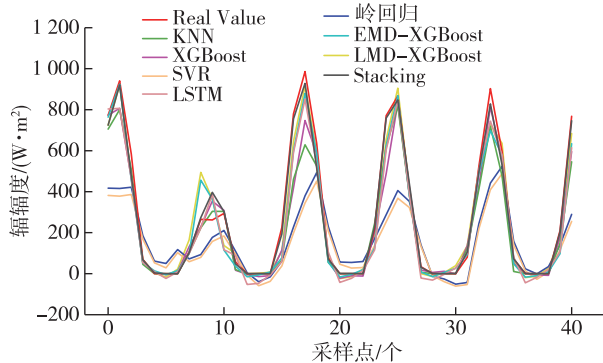


图 4 不同模型预测结果对比

Fig. 4 Prediction result comparison between Stacking method and traditional models

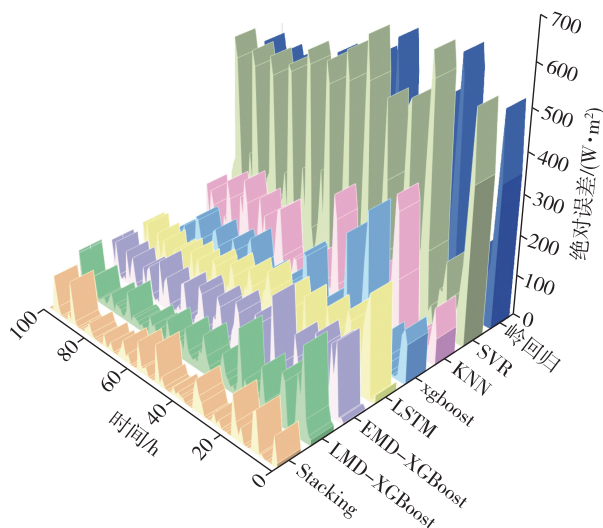


图 5 不同模型的预测误差对比

Fig. 5 Prediction error comparison between Stacking method and traditional models

由图 4、5 和表 2 分析可知:Stacking 模型相较于其基模型中的单一模型 KNN、SVR 和岭回归, R^2 分别提升 0.118、0.424 和 0.504, 与深度学习模型 LSTM 相比, R^2 指标提升 0.101; 将 EMD 和 LMD 分

表 2 不同模型的评价指标

Table 2 Performance comparison between Stacking method and traditional models

预测模型	RMSE/(W·m ²)	MAE/(W·m ²)	R ²
KNN	122.008	74.542	0.851
XGBoost	96.013	59.822	0.881
SVR	147.407	130.408	0.545
岭回归	159.719	142.733	0.465
LSTM	102.625	65.127	0.868
EMD-XGBoost	81.375	54.118	0.923
LMD-XGBoost	74.467	44.757	0.934
Stacking	55.135	29.906	0.969

解与集成模型 XGBoost 结合后, R^2 分别达到 0.923 和 0.934, 与 XGBoost 直接预测相比分别提升 0.042 和 0.053, 但仍低于 Stacking 的 0.969. 因此, 相较于其他传统机器学习、深度学习方法和集成学习模型 XGBoost, Stacking 模型在太阳辐照度的短期预测上有着更高的精度, 拟合能力更强。

为了显示本文模型的优越性和各改进点的有效性, 搭建了以下 4 种 Stacking 对比模型:

- 1) 模型 1: 经典 Stacking 模型 (Stacking).
- 2) 模型 2: 对传统 Stacking 模型进行 GBDT 特征筛选的模型 (GBDT-Stacking).
- 3) 模型 3: 对传统 Stacking 模型预测结果使用 RF 算法进行误差修正 (Stacking-RF).
- 4) 模型 4: 对传统 Stacking 模型进行权重分配与 Box-Cox 处理的模型 (赋权 Stacking).

由此设计了消融实验来分析各个改进点对 Stacking 模型预测性能的影响, 并将各模型的预测结果与辐照度实际值进行横向对比, 结果如图 6 所示, 各消融实验模型的绝对误差如图 7 所示, 消融实验模型的评价指标值如表 3 所示。

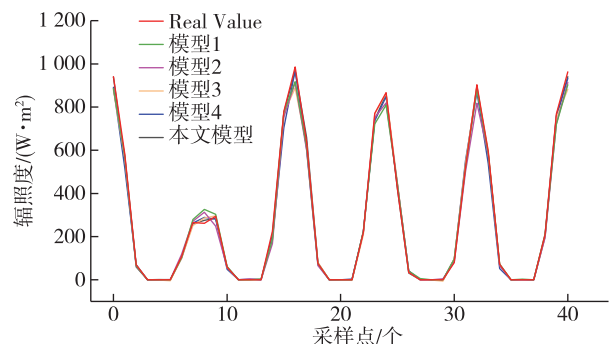


图 6 对比模型与本文所提模型的预测结果

Fig. 6 Prediction result comparison between the proposed method and classic or modified Stacking models

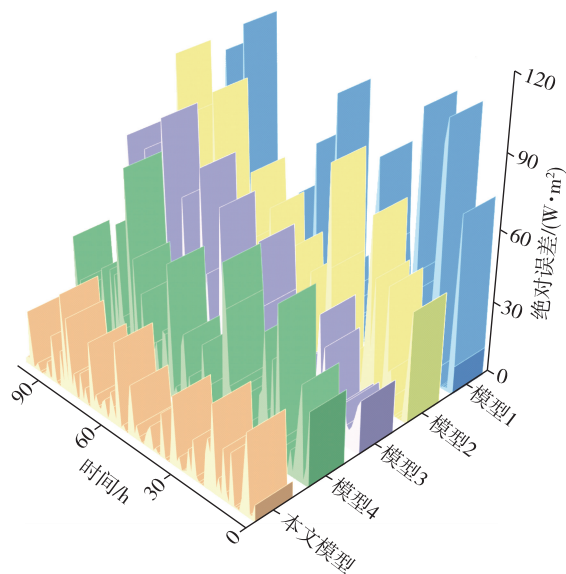


图7 对比模型与本文所提模型预测误差

Fig. 7 Prediction error comparison between the proposed method and classic or modified Stacking models

表3 不同模型的评价指标

Table 3 Performance comparison between the proposed method and classic or modified Stacking models

预测模型	RMSE/($W \cdot m^2$)	MAE/($W \cdot m^2$)	R^2
模型1	55.135	29.906	0.969
模型2	45.057	26.779	0.979
模型3	37.096	18.741	0.982
模型4	31.636	16.481	0.987
本文模型	23.205	10.367	0.993

由表3可知:本文模型相较于 Stacking、GBDT-Stacking、Stacking-RF 及赋权 Stacking 模型, R^2 分别提升 0.024、0.014、0.011 和 0.006, MAE 分别降低 65.3%、61.3%、44.7% 和 37.1%。另外还可以看出:模型4相较于拥有相同学习器的模型1预测效果提升明显且误差波动较小;模型2和模型3也可显著提升 Stacking 集成模型的预测精度,其 MAE 分别降低 10.4% 和 37.3%。因此,本文模型能准确预估不同时期的辐照度变化趋势,预测准确度较 Stacking 模型有较大提升。

4 结语

本文将基于 Box-Cox 变换和权值分配的改进 Stacking 模型应用于短期辐照度预测领域,使用4种相互异质的算法作为学习器,充分利用各算法在数据特征结构与特征空间上的不同视角,从而使 Stacking 集成模型的优越性得以充分发挥。同时,采

用 GBDT 算法进行特征选取和 RF 算法对误差进行修正,达到了简化计算复杂度和提升辐照度预测精度的目的。实验结果表明:

1) 通过对数据进行特征重要度分析,筛选掉相关度较弱的特征,达到了过滤冗余特征,构造出效率精度更高、复杂度更低的独立预测模型的目的。

2) 引入误差修正算法计算拟合 Stacking 集成模型预测结果的动态误差,获得余项预测值,通过加法模型融合预测值与余项得到最终辐照度预测值。结果表明,通过将余项融合进预测结果能降低预测误差。

3) 通过与单一预测模型以及传统 Stacking 模型相比,表明本文所提出的改进 Stacking 模型具有良好的稳定性和较高的预测精度。

参考文献

References

- [1] 陈中,车松阳.基于云变换的光伏出力预测模型[J].太阳能学报,2019,40(11):3054-3061
CHEN Zhong, CHE Songyang. Photovoltaic output forecast based on cloud models[J]. Acta Energetica Solaris Sinica, 2019, 40(11): 3054-3061
- [2] Cui S C, Wang Y W, Li C J, et al. Prosumer community: a risk aversion energy sharing model [J]. IEEE Transactions on Sustainable Energy, 2020, 11(2): 828-838
- [3] Inman R H, Edson J G, Coimbra C F M. Impact of local broadband turbidity estimation on forecasting of clear sky direct normal irradiance [J]. Solar Energy, 2015, 117: 125-138
- [4] Zhu T T, Zhou H, Wei H K, et al. Inter-hour direct normal irradiance forecast with multiple data types and time-series [J]. Journal of Modern Power Systems and Clean Energy, 2019, 7(5): 1319-1327
- [5] 丁宏,陶晓峰,陆春艳,等.基于 RWT-SVM 的台区配电网日前负荷预测研究[J].南京信息工程大学学报(自然科学版),2023,15(3):330-336
DING Hong, TAO Xiaofeng, LU Chunyan, et al. Day-ahead load forecasting of distributed power grids based on RWT-SVM [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2023, 15(3): 330-336
- [6] Bouzgou H, Gueymard C A. Minimum redundancy-maximum relevance with extreme learning machines for global solar radiation forecasting: toward an optimized dimensionality reduction for solar time series [J]. Solar Energy, 2017, 158: 595-609
- [7] 杨春熙,韩威,高志球.基于 GRU 神经网络的太阳辐照度短期预测[J].中国科技论文,2020,15(1):8-14
YANG Chunxi, HAN Wei, GAO Zhiqiu. Short-term forecasting for solar irradiance using GRU neural network [J]. China Science Paper, 2020, 15(1): 8-14

- [8] 顾德文,李文博,徐箭,等.基于相似日特征的雷电天气下辐照度预测[J].武汉大学学报(工学版),2021,54(10):952-957
GU Dewen, LI Wenbo, XU Jian, et al. Forecast of irradiance under lightning weather based on similar day characteristics[J]. Engineering Journal of Wuhan University, 2021, 54(10): 952-957
- [9] 周满国,黄艳国,段锦锋.基于 GRU-RF 模型的太阳辐照度短时预测[J].太阳能学报,2022,43(7):166-173
ZHOU Manguo, HUANG Yanguo, DUAN Jinfeng. Short term prediction of solar irradiance based on GRU-RF model[J]. Acta Energiae Solaris Sinica, 2022, 43(7): 166-173
- [10] 田翠霞,黄敏,朱启兵.基于 EMD-LMD-LSSVM 联合模型的逐时太阳辐照度预测[J].太阳能学报,2018,39(2):504-512
TIAN Cuixia, HUANG Min, ZHU Qibing. Hourly solar irradiance forecast based on EMD-LMD-LSSVM joint model[J]. Acta Energiae Solaris Sinica, 2018, 39(2): 504-512
- [11] Zhou H X, Zhang Y J, Yang L F, et al. Short-term photovoltaic power forecasting based on stacking-SVM[C]//2018 9th International Conference on Information Technology in Medicine and Education (ITME). October 19-21, 2018, Hangzhou, China. IEEE, 2018:994-998
- [12] 李鹏钦,张长胜,李英娜,等.改进 Stacking 算法的光伏发电功率预测[J].应用科学学报,2022,40(2):288-301
LI Pengqin, ZHANG Changsheng, LI Yingna, et al. Photovoltaic power forecast improved stacking algorithm[J]. Journal of Applied Sciences, 2022, 40(2): 288-301
- [13] 刘志超,袁三男,唐万成.基于 BLSTM-随机森林的短期光伏发电输出功率预测[J].电源技术,2021,45(4):495-498
LIU Zhichao, YUAN Sanman, TANG Wancheng. Output power prediction of short-term photovoltaic power generation based on BLSTM-random forest[J]. Chinese Journal of Power Sources, 2021, 45(4): 495-498
- [14] Zhang Y Z, Xiong R, He H W, et al. Lithium-ion battery remaining useful life prediction with Box-Cox transformation and Monte Carlo simulation[J]. IEEE Transactions on Industrial Electronics, 2019, 66(2): 1585-1597
- [15] Breiman L. Stacked regressions[J]. Machine Language, 1996, 24(1): 49-64
- [16] 李永刚,王月,刘丰瑞,等.基于 Stacking 融合的短期风速预测组合模型[J].电网技术,2020,44(8):2875-2882
LI Yonggang, WANG Yue, LIU Fengrui, et al. Combination model of short-term wind speed prediction based on Stacking fusion[J]. Power System Technology, 2020, 44(8): 2875-2882

Short-term solar irradiance prediction based on improved Stacking and error correction

WANG Shanshan¹ WU Ni¹ HE Jiawen¹ ZHU Wei¹

¹ School of Electrical and Electronic Engineering/Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Hubei University of Technology, Wuhan 430068, China

Abstract To improve the accuracy and reliability of solar irradiance prediction for photovoltaic power system, we propose a model to forecast short-term solar irradiance based on improved Stacking ensemble learning and error correction. First, the Gradient Boosting Decision Tree (GBDT) is used to perform feature selection and remove redundant characteristics of original data set, thus increase prediction accuracy and computing efficiency. Then, an improved Stacking irradiance prediction model is established. In accordance with the difference in prediction accuracy of prediction models in the primary layer under K -fold cross-validation, the prediction results are weighted, and the Box-Cox is employed to transform and process the training set data input from the first layer to the second layer of Stacking, so as to increase the normality and homoscedasticity of prediction. Finally, the historical prediction error data are extracted, and Random Forest is applied to construct an error model to further improve the prediction accuracy. The experimental results show that, compared with traditional models and classic Stacking models, the proposed method significantly improves the prediction performance on solar irradiance.

Key words solar radiation; photovoltaic power generation; Stacking; regression prediction algorithm; cross-validation