

DOI: 10.12158/j.2096-3203.2024.05.023

# 基于空间相关性与 Stacking 集成学习的风电功率预测方法

王小明<sup>1</sup>, 徐斌<sup>1</sup>, 尹元亚<sup>2</sup>, 潘文虎<sup>2</sup>, 吴红斌<sup>3</sup>, 韩屹<sup>3</sup>

(1. 国网安徽省电力有限公司电力科学研究院, 安徽 合肥 230601;

2. 国网安徽省电力有限公司芜湖供电公司, 安徽 芜湖 241000;

3. 合肥工业大学电气与自动化工程学院, 安徽 合肥 230009)

**摘要:**针对目标气象预报数据缺失导致风电预测精度不足的问题,提出一种基于空间相关性和 Stacking 集成学习的风电功率预测方法。首先,分析目标风电场与相邻气象站点之间的空间相关性,根据相关系数极值点确定延迟时间,构建风速时移数据集;其次,利用 Stacking 集成方法融合多元算法,从多个数据观测角度预测目标风电场的风电功率,实现不同算法的优势互补,提升整体泛化能力,并采用粒子群优化算法搜索模型超参数,较好地平衡搜索时间与模型效果;最后,采用华东地区某风电场的实测数据验证了文中所提方法的有效性和准确性。结果表明,通过考虑不同位置的信息偏差,从数据输入和预测模型两方面可有效提高数据缺失情况下的风电预测精度。

**关键词:**风电功率预测;空间相关性;Stacking 集成学习;风速时移;多元算法融合;粒子群优化

中图分类号: TM614

文献标志码: A

文章编号: 2096-3203(2024)05-0224-09

## 0 引言

2022年,全球新增77.6 GW 风电并入电网,风电总装机容量达到906 GW,与2021年相比增长9%<sup>[1]</sup>。然而,风能的随机性和波动性可能导致供电稳定性及储备容量管理等方面的问题<sup>[2]</sup>。因此,准确的风电功率预测对于保障电网调度运行和高效消纳至关重要<sup>[3]</sup>。

目前,人工智能法<sup>[4-5]</sup>通过大量历史样本挖掘输入特征与输出结果之间的非线性映射关系,具有较高的预测精度,其主要有卷积神经网络(convolutional neural network, CNN)<sup>[6-7]</sup>、支持向量机(support vector machine, SVM)<sup>[8-9]</sup>、极限学习机<sup>[10-11]</sup>等方法。结合数据挖掘的方法可以提高人工智能法的预测精度,分析数据的深层特性,如聚类方法<sup>[12]</sup>、信号分解<sup>[13]</sup>、关联度分析<sup>[14]</sup>等。文献[12]综合考虑功率轨迹和气象特征,采用多重聚类和分层聚类的方法获得更加完善的样本子集;文献[13]采用小波包分解方法处理原始风速数据,剔除随机扰动,凸显数据的细节特性;文献[14]采用灰色关联分析选取与目标电站关联度较高的周边电站,为后续模型提取空间特征奠定基础。

然而,人工智能法往往需要借助数值天气预报(numerical weather prediction, NWP)数据进行预测,但出于实际成本考虑,气象预报服务难以做到精确全覆盖。当风电场缺乏 NWP 数据时,该方法难以

发挥实际作用。但是,由于大气运动的连续性,相邻风电场或者气象站点的气象数据具有相似性,可以为目标风电场提供数据参考。因此,考虑空间相关性的建模方法能够获取完善的气象变化规律与风速分布信息,在一定程度上解决了上述问题。文献[15]利用 Copula 函数建立了一种多风电场空间相关性模型。文献[16]提出了一种计及空间相关的多风电场短期功率预测模型。文献[17]考虑整体对局部的影响,通过 CNN 进行多尺度空间特征提取,进行局部特征提取时都注入了全局特征。然而,现有方法没有考虑利用相邻风电场之间数据的延迟时间来优化风场数据。

此外,集成学习通过融合多种学习器或者算法,采用集体智慧提高整体泛化能力,因此成为数据竞赛中的常用方法。文献[18]考虑不同时间尺度下预测方法的性能差异,采用自适应动态加权的方法代替固定权值,取得了更好的预测效果;文献[19]利用随机森林(random forest, RF)算法融合不同网络层数的门控循环单元(gated recurrent unit, GRU),结果表明集成模型具有更强的适用性;文献[20]采用模型融合与集成修剪的方法,有效提升了点预测与区间预测的预测精度;文献[21]设计了一种卷积循环神经网络模型,准确捕捉了风速的动态分布,同时获得了多个点位处的风速结果,明显提升了风速预测效果。

综上,文中对整体风电场的输出功率进行研究,提出一种基于空间相关性和 Stacking 集成学习的风电功率预测方法。首先,计算目标风电场与相

收稿日期:2024-04-14;修回日期:2024-06-25

基金项目:国家自然科学基金资助项目(U19A20106)

邻气象站点的延迟时间,构建风速时移数据集,弥补目标风电场气象数据的缺失;其次,基于风速时移数据集,选用粒子群优化的 SVM、RF、极端梯度上升(extreme gradient boosting, XGBoost)算法、CNN、GRU 作为基学习器,采用 Stacking 集成方法实现多模型融合,完成目标风电场的功率预测。

## 1 空间相关性分析

风速的空间相关性是指不同空间位置处的风电场或气象站点的风速数据在空间上的相关程度<sup>[22]</sup>。由于大气运动的连续性,相邻位置的风速序列具有一定的相似性和延时性。合理运用空间相关性,尤其是上游风电场,可以追溯目标风电场未来的气象信息,因此,可以直接沿用相邻位置的数据。而对于随机波动较强的风速数据而言,需要考虑不同地理位置导致的延时,对风速数据进行修正,从而提升目标风电场的预测精度。

要准确计算不同位置之间的延迟时间应综合考虑实际地形、地貌、风速以及风向变化,涉及复杂的时空变化过程<sup>[23]</sup>。已有研究利用 Pearson 相关系数计算风速和功率之间的相关性<sup>[24-27]</sup>,文中也采用 Pearson 相关系数计算相邻气象站点风速数据与目标风电场功率数据之间的相关性,以确定最佳延迟时间。计算步骤如下:

(1) 将相邻风电场或气象站点的原始风速数据相对于目标风电场功率数据进行移动,移动的时间间隔为采样周期 15 min 的整数倍,时间间隔的正、负分别表示气象站点数据和风站数据之间的超前、滞后关系。

(2) 分别计算移动不同时间间隔下风速数据与目标风电场功率数据之间的 Pearson 相关系数。Pearson 相关系数作为两组数据相似程度的判据,其取值范围为 $[-1, 1]$ ,当相关系数大于零时,两变量正相关,当相关系数小于零时则负相关,相关系数的绝对值越大,变量间的相关程度越高。令相邻气象站点风速序列为  $x$ ,目标风电场功率序列为  $y$ , Pearson 相关系数的计算如式(1)所示。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中: $r$  为 Pearson 相关系数; $x_i$  为相邻气象站点第  $i$  个风速数据; $y_i$  为目标风电场第  $i$  个功率数据; $\bar{x}$ 、 $\bar{y}$  分别为序列  $x$ 、 $y$  中数据的平均值; $n$  为序列的样本总数。

(3) 在不同时间间隔下计算得到的 Pearson 相关系数中选择极大值作为延迟时间,调整原始风速序列,获取目标风电场精确的功率数据。

以华东地区某风电场及其相邻气象站点为例,一段时间内的风电场功率数据与气象站点风速数据如图 1(a)、(b)所示。分别计算不同时间间隔下的相关系数,具体计算结果如图 1(c)所示。

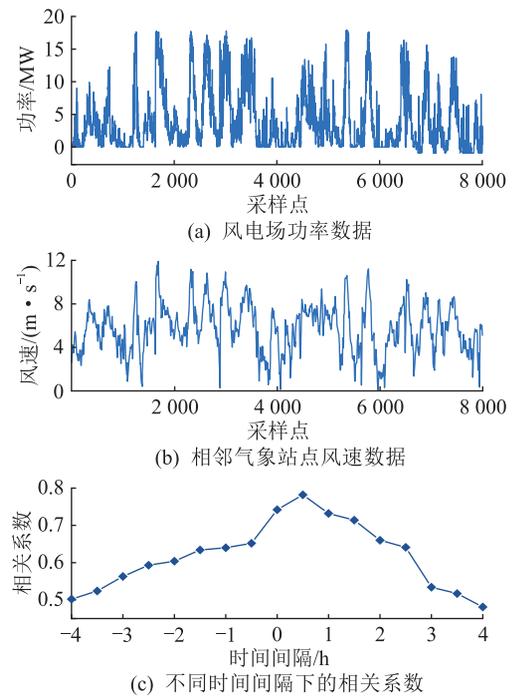


图1 风电场与气象站点的数据及其相关系数

Fig.1 Wind farm and meteorological station data and their correlation coefficients

由图 1 可知,由于地理位置相邻,目标风电场的功率数据与气象站点的风速数据之间存在较高的相关性,这表明相邻气象站点具有为信息缺失风电场提供数据参考的可行性。当时间间隔为 30 min 时,两者的相关系数达到最高,为 0.782,且在 $-4 \sim 4$  h 的变化范围内,仅存在唯一极值点,证实了采取相关系数确定延迟时间的合理性,避免了复杂的时空建模过程。

## 2 风电功率预测模型

### 2.1 Stacking 集成策略

集成方法结合多元算法优势,能够适应多种预测场景,可以解决单一算法对于短时间内急剧波动的突发性天气预测精度不足的问题,获得更加稳定的预测结果。集成学习主要包含三大类:Bagging、Boosting 以及 Stacking。Stacking 集成学习具有严谨的集成规则和强大的学习能力,其初步预测的个体学习器被称为基学习器,用于结合的学习器被称为

元学习器。先用全部数据训练基学习器,每个基模型都对训练样本进行预测,将其预测值作为训练样本的特征值,然后得到新的训练样本,最后基于新的训练样本数据训练元学习器,得到更准确的预测结果。Stacking 集成学习通过结合不同个体学习器的预测结果,可以规避陷入局部极值的风险,在大规模训练数据上具有优越表现。Stacking 集成学习框架如图 2 所示。

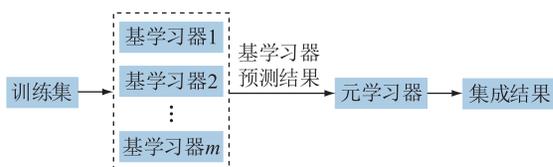


图 2 Stacking 集成学习框架

Fig.2 Stacking ensemble learning framework

在训练过程中,若直接使用基学习器的训练集来生成次级训练集,将会增加过拟合的风险,大大降低集成模型在测试集上的表现<sup>[28]</sup>。为避免重复学习,采用  $K$  折交叉验证的方式生成次级训练集<sup>[29]</sup>。具体地,将原始数据集划分成  $K$  个大小相等且不重合的子集,记为  $D = \{D_1, D_2, \dots, D_j, \dots, D_k\}$ ,其中  $D_j$  为第  $j$  次交叉验证的验证集,其余子集  $D_{-j} = D \setminus D_j$  为第  $j$  次交叉验证的训练集。对于每个基学习器而言,在训练集上训练模型,在验证集上预测结果,经过  $K$  轮训练和测试后,将预测结果纵向堆叠产生次级训练集,且次级训练集的样本量与原始数据的样本量相同。该方式在减小过拟合风险的同时,可以确保元学习器有充足的训练样本。

## 2.2 基学习器

基学习器是集成模型中的个体学习器,通常为弱学习器,一般采用相对简单或性能略低的模型。但 Stacking 集成策略中的基学习器为强学习器,以期在个体学习阶段获得较为准确的预测值。基学习器是保证集成效果与稳健决策的关键,可采用常见的机器学习算法如 SVM、RF 和 XGBoost,此外, CNN、GRU 等神经网络也常被作为基学习器使用。

### 2.2.1 基于机器学习算法的基学习器

文中将 Bagging 流派集成学习算法中广泛使用的 RF 算法以及 Boosting 流派集成学习算法中典型的 XGBoost 算法作为 Stacking 集成学习的基模型。利用 Stacking 集成策略可以并行训练多个基学习器的特点,将 Bagging 和 Boosting 两个派系融入到集成学习算法中,综合发挥集成学习算法的优势。此外,文中的基学习器还包含 SVM。SVM 通过核函数将低维数据映射到高维特征空间,在高维特征空间内寻找线性可分的超平面,将非线性回归问题转化

为线性回归问题,其凭借完善的数学逻辑,在现实任务中具有良好表现;RF 作为 Bagging 集成方法的典型代表,构建多个相互独立的决策树,将多个决策树的预测结果均值作为最终结果,此外,在构建单个学习器的过程中,使用随机样本和随机特征,增加集成模型的泛化能力;XGBoost 对梯度提升决策树算法进行改进,在优化目标中除了考虑损失函数外,还加入限制模型复杂度的正则项,对于原始目标函数难以直接优化的问题,采用二阶泰勒展开的方式精确近似目标函数,具有良好的防止过拟合效果和预测精度。

### 2.2.2 基于神经网络的基学习器

文中的 Stacking 集成学习框架中还包含 CNN、GRU 两种神经网络。CNN 具有局部连接和权值共享的特点,可以在大幅减少参数计算量的同时提取输入特征,因此常被用于图像识别、故障诊断等领域。CNN 一般由卷积层、池化层和全连接层构成,其典型结构如图 3 所示。卷积层通过设置不同尺寸的卷积核来获取相应的感受野,从中挖掘关联信息。针对卷积层提取的局部非线性特征,池化层可以缩减特征图的尺寸,起到数据降维的作用。

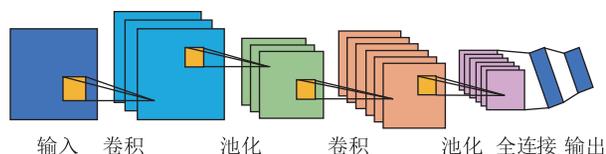


图 3 CNN 结构

Fig.3 Structure of CNN

由于风电场实测信息的时序特性,文中通过堆叠多层一维卷积层来挖掘风电场内的耦合联系,一维卷积的运算过程如图 4 所示,其中设置卷积核的数量为 1,尺寸为 3,步长为 1。卷积核沿着时间维度对特征矩阵进行扫描,最终得到  $4 \times 1$  的特征向量。考虑到风电功率的随机波动性,一般采用多个卷积核进行扫描,挖掘原始数据中的深层特性。

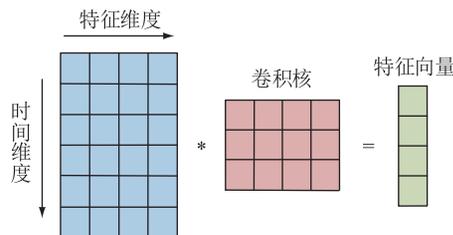


图 4 一维卷积运算过程

Fig.4 One-dimensional convolution operation process

GRU 作为循环神经网络的典型代表,在长短期记忆(long short-term memory, LSTM)网络的基础上进行改进,将 LSTM 网络中的输入门和遗忘门合并

为 GRU 中的更新门,实现信息输入与遗忘之间的平衡。GRU 通过更新门和重置门来控制时序信息的动态传递过程,因此,模型结构得到简化,且在相同的预测任务下,可以在保证预测效果的同时有效降低模型计算量与训练时间。

### 2.2.3 基于粒子群优化算法的超参数优化

集成学习中有众多超参数,因此精准高效的参数优化方法至关重要。一般采用随机搜索方法确定模型超参数,从参数范围中随机采样参数点,但该方法可能会错过最优超参数结果,未能充分发挥基学习器性能。因此,文中利用粒子群优化算法在参数空间中寻找最优的参数集,改善预测性能<sup>[30]</sup>。具体过程如下。

在  $D$  维搜索空间内初始化一群随机粒子,设置每个粒子的初始速度和初始位置,并计算每个粒子的适应度。对于超参数优化,将模型的预测误差作为粒子适应度。在迭代过程中,粒子通过追踪个体极值和全局极值来更新自己的速度和位置,其中,个体极值是粒子本身搜寻的最优解,全局极值是整个种群目前搜索到的最优解,具体公式如下:

$$v_{i,t+1} = \omega v_{i,t} + c_1 r_1 (p_{is,t} - x_{i,t}) + c_2 r_2 (p_{gs,t} - x_{i,t}) \quad (2)$$

$$x_{i,t+1} = x_{i,t} + v_{i,t+1} \quad (3)$$

式中:  $v_{i,t+1}$ 、 $x_{i,t+1}$  分别为更新后  $t+1$  时刻粒子  $i$  的速度和位置;  $v_{i,t}$ 、 $x_{i,t}$  分别为粒子  $i$  当前  $t$  时刻的速度和位置;  $\omega$  为惯性因子,表示粒子维持自己先前速度的趋势,一般设置为非负;  $p_{is,t}$  为粒子  $i$  在  $t$  时刻搜索到的最优位置;  $p_{gs,t}$  为群体搜索到的最优位置;  $c_1$ 、 $c_2$  为学习因子;  $r_1$ 、 $r_2$  为均匀随机数,取值为  $[0,1]$ 。

最后,判断是否满足终止条件或者迭代精度要求,若不满足,则继续迭代;反之,则返回优化问题的最优解,并基于最优参数训练预测模型,从而输出最终结果。

## 2.3 元学习器

基学习器是对同一数据集进行观测分析,其预测结果存在相似性,元学习器则用于组合基学习器

的策略或算法,通过对基学习器的输出进行组合,生成最终集成模型的预测结果。为防止过度学习,Stacking 集成策略中的元学习器一般选用弱学习器。元学习器的训练数据为各基学习器在训练集上预测结果的连接,若某个基学习器错误地学习了特征空间的某个区域,那么元学习器通过结合其基学习器的学习行为,可以对错误进行适当纠正,以提高最终预测结果的精度。

文中基学习器选择了 5 种不同的算法,5 种算法的预测结果之间异质化明显,使得总模型具备较强的泛化能力。因此,元学习器不需要选择太复杂的模型,以保持总模型的泛化能力。对于文中研究的回归问题,元学习器主要起到对集成算法的加权集成作用,采用复杂的线性回归器作为元学习器也只能发挥加权集成作用,因此简单的线性回归器可以满足 Stacking 模型中元学习器的要求,且有较快的计算速度。

贝叶斯回归模型在一般的误差函数基础上添加了正则化项,预测过程简单且速度快,可以有效避免产生过拟合现象<sup>[31]</sup>。此外,贝叶斯回归模型适合处理多分类任务,其表现稳定,能够在缺失数据的情况下进行预测,适用于文中研究的集成学习环境,对基学习器的预测结果进行汇总并实现风速预测。因此,文中采用贝叶斯回归模型作为元学习器。

## 2.4 基于空间相关性与 Stacking 集成学习的风电功率预测方法

针对目标风电场气象预报数据缺失的状况,采取相邻气象站点的风速时移作为数据参考,构建融合多元算法的集成模型以提升预测精度。预测模型训练过程如图 5 所示。

具体训练流程如下:

(1) 对数据集进行预处理,在此基础上,采用 Pearson 相关系数分析目标风电场功率数据与相邻气象站点风速数据之间的延迟时间,采用风速时移数据集作为目标风电场的气象数据。

(2) 对风速时移数据集进行 5 折交叉验证,避免学习器重复学习而产生过拟合现象。



图 5 基于空间相关性与 Stacking 集成学习的预测模型训练过程

Fig.5 Predictive model training process based on spatial correlation and Stacking ensemble learning

(3) 在交叉验证风速时移数据集的基础上,利用 RF、XGBoost、SVM、CNN 及 GRU 共 5 种基学习器进行训练,并通过粒子群优化算法优化基学习器的超参数,提升预测精度。

(4) 将基学习器的预测结果组合成一个新的矩阵,其中每一列代表一个基学习器的预测结果,称之为次级训练集。

(5) 在次级训练集上,采用贝叶斯回归模型作为元学习器,训练学习如何最好地组合这些基学习器的预测结果。元学习器的输出是最终的集成模型结合基学习器的预测结果,输出最终风电出力,完成基于空间相关性与 Stacking 集成学习的预测模型训练过程。

(6) 在测试集上进行对比分析,并通过评估指标验证文中所提方法的准确性与有效性。

### 3 算例分析

文中实验数据采用华东地区某风电场 2019 年 6 月 1 日至 2020 年 1 月 31 日的实测风电功率以及气象站点的气象预报数据,该风电场的装机容量为 18 MW。单一气象预报结果难以准确反映气象运动规律,因此选取目标风电场周围 2 个气象站点的气象预报数据作为风电场的数据参考,分别记为气象站点 1、气象站点 2。两气象站点与风电场的距离分别为 24.8、19.4 km;气象预报数据包括风速、风向、温度、湿度、气压;所有数据的采样周期均为 15 min。为避免预测结果的偶然性,以及确保集成模型具有足够的训练样本,将 2019 年 6 月 1 日至 8 月 31 日的数据作为训练集 1,2019 年 9 月 1 日至 30 日的数据作为测试集 1;将 2019 年 10 月 1 日至 12 月 31 日的数据作为训练集 2,2020 年 1 月 1 日至 31 日的数据作为测试集 2。

针对不同方法的预测结果,选取均方根误差 (root mean squared error, RMSE)  $E_{\text{RMSE}}$  和平均绝对误差 (mean absolute error, MAE)  $E_{\text{MAE}}$  作为评估指标以验证文中方法的准确性,具体公式如下:

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$E_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

式中:  $y_i$  为  $t$  时刻的风电功率实际值;  $\hat{y}_i$  为  $t$  时刻的风电功率预测值。

#### 3.1 数据预处理

由于传感器故障、设备检修等诸多原因,采集的数据中往往会存在异常和缺失的状况,为降低不

良数据的干扰,需要对数据进行清洗。对于大规模缺失的数据予以删除,对于少量缺失或者异常的数据采用线性插值法进行填补。

由于不同特征间的量纲差异,需要对实验数据进行归一化处理,从而提升模型训练效率与预测精度。其中,对于气象特征采用 min-max 归一化,将原始数据  $z$  变换到  $[0, 1]$  之间,即:

$$z' = \frac{z - z_{\min}}{z_{\max} - z_{\min}} \quad (6)$$

式中:  $z'$  为归一化结果;  $z_{\max}$ 、 $z_{\min}$  分别为原始数据中的最大值和最小值。

特别地,风向变化范围在  $0^\circ \sim 360^\circ$  之间,  $0^\circ$  附近的风向数据 (如  $1^\circ$  和  $359^\circ$ ) 差异较小,但 min-max 归一化会割裂这种内在联系。因此,归一化方法既要区分风向数据的差异性,又要关注实际变化过程中的关联性,故采用如式 (7) 所示的正余弦化方法进行处理,将其作为 2 个气象因素。

$$\begin{cases} z_\theta = \sin \theta \\ z'_\theta = \cos \theta \end{cases} \quad (7)$$

式中:  $z_\theta$ 、 $z'_\theta$  为风向数据的归一化结果;  $\theta$  为风向数据。

将归一化后的气象特征输入到预测模型中,得到的功率预测结果仍为归一化数据。因此,对预测输出进行反归一化操作,使其具有明确的物理含义,并与真实数据进行对比分析,具体公式如下:

$$z = z'(z_{\max} - z_{\min}) + z_{\min} \quad (8)$$

#### 3.2 延迟时间的计算

通过时间序列平移的方式,依次计算不同时间间隔下气象站点历史风速数据与风电场历史功率数据之间的相关系数,根据相关系数极值点确定延迟时间。由于地理位置的分布以及风向的改变,延迟时间有正负之分,当延迟时间为正时,表示原始气象站点的风速数据超前于风电场的功率数据,需要将风速数据相较于功率数据后移;当延迟时间为负时,表示原始气象站点的风速数据滞后于风电场的功率数据,需要将风速数据相较于功率数据前移。对数据集 1、数据集 2 利用第 1 章中的方法进行空间相关性分析,计算各个气象站点与目标风场之间的延迟时间,计算结果如表 1 所示。

由表 1 可知,对于数据集 1 而言,气象站点 1 的风速数据超前于目标风电场的功率数据 75 min,气象站点 2 的风速数据滞后于目标风电场的功率数据 60 min;对于数据集 2 而言,由于风向发生变化,气象站点 1 的风速数据滞后于目标风电场的功率数据 30 min,气象站点 2 的风速数据超前于目标风电场

表 1 相邻气象站点与风电场的延迟时间  
Table 1 Delay time between adjacent meteorological stations and wind farms

数据集	气象站点	延迟时间/min	相关系数极值
数据集 1	气象站点 1	75	0.749
	气象站点 2	-60	0.721
数据集 2	气象站点 1	-30	0.771
	气象站点 2	30	0.782

的功率数据 30 min。由于数据集 1 内气象站点 1 的风速数据与风电场的功率数据之间的相关系数极值较大,选择气象站点 1 的气象数据作为参考;同理,数据集 2 内选择气象站点 2 的气象数据作为参考。

### 3.3 Stacking 集成学习的预测效果

为展示 Stacking 集成学习的优越性能,选取基学习器 SVM、RF、XGboost、CNN 和 GRU 在测试集上的预测误差与 Stacking 集成方法的预测误差进行对比,结果如表 2 所示。

表 2 不同预测模型的预测误差对比  
Table 2 Comparison of prediction errors of different prediction models MW

预测模型	测试集 1		测试集 2	
	$E_{RMSE}$	$E_{MAE}$	$E_{RMSE}$	$E_{MAE}$
SVM	0.952	0.693	1.104	0.718
RF	0.941	0.682	0.944	0.657
XGBoost	0.936	0.679	0.946	0.652
CNN	0.882	0.654	0.915	0.646
GRU	0.923	0.667	0.927	0.652
Stacking	0.872	0.641	0.921	0.650

由表 2 可知,对于基学习器而言,数学逻辑严谨的 SVM 模型通过选取合适的核函数和相关参数,可以较好地拟合风电功率;基于 Bagging 和 Boosting 原理的集成算法通过构建多个学习器提升整体泛化能力,在测试集上取得了良好的预测效果,特别是 XGBoost 模型凭借精确的目标函数优化和防止过拟合的优势获得了较好的预测效果;CNN 擅长处理高维数据中的耦合联系,通过卷积核获取局部关联特征,可以较好地拟合实际数据;GRU 通过历史数据挖掘气象变化规律,从而推演风电功率波动;Stacking 集成学习融合多元算法的优势推演风电功率。在测试集 1 上,较基学习器而言,Stacking 集成方法的 RMSE 分别下降 8.40%、7.33%、6.84%、1.13%、5.53%, MAE 分别下降 7.50%、6.01%、5.60%、1.99%、3.90%,在单一算法的基础上进一步提升了预测精度。然而,在测试集 2 上,相较于 CNN 的预测结果,Stacking 集成方法的 RMSE 和

MAE 均有所上升。这是由于基学习器的选取是保证集成效果与稳健决策的关键,测试集 2 上集成模型存在过拟合的状况。因此,重新选取 SVM、XGBoost、CNN 作为基学习器进行集成,其 RMSE 和 MAE 分别为 0.908 MW 和 0.642 MW。综上所述,选取合适策略的集成方法较单一算法而言,具有更好的准确性与适用性。

### 3.4 风电功率预测方法对比

考虑空间距离带来的信息偏差,可以构造更加准确的气象数据集。此外,集成学习可以融合不同算法的优势来弥补单一学习器的局限性,面对随机波动的风电出力,其预测结果更精确,适应性更强。为验证文中所提方法的有效性,采取以下 3 种方法进行风电功率预测,并与实际功率对比。

方法 1:忽略风电场与相邻气象站点间的信息偏差,直接采用气象站点的气象预报数据作为模型输入,利用 Stacking 集成方法输出最终风电出力。通过与方法 3 对比,验证基于时间间隔的数据修正方法对风电功率预测精度的提升效果。

方法 2:考虑风电场与相邻气象站点间的信息偏差,采用风速时移数据集作为模型输入,利用平均法集成基学习器的预测结果作为最终风电出力。通过与方法 3 对比,验证 Stacking 集成方法对风电功率预测精度的提升效果。

方法 3:考虑风电场与相邻气象站点间的信息偏差,采用风速时移数据集作为模型输入,利用 Stacking 集成方法确定最终风电出力,模型超参数采用粒子群优化算法搜索确定。方法 3 为文中提出的基于空间相关性与 Stacking 集成学习的风电功率预测方法,通过与方法 1、方法 2 进行对比以验证文中所提方法的有效性。

为对比 3 种方法的性能,选取测试集中的某个典型日的风电预测结果进行展示,以 15 min 为采样间隔、8 个采样点为时间窗口,滚动预测生成的 96 个采样点,预测结果如图 6 所示。图 6 展示了测试集上 3 种预测方法的预测结果,不同预测方法在测试集上的预测误差如表 3 所示。相较于其他方法,方法 3 凭借更加准确的气象信息和严谨的集成方法,得到更加贴合实际曲线的预测曲线,验证了风速时移的重要性和 Stacking 集成学习的强大学习能力。在不同的预测场景下,该方法都能更好地追踪功率起伏。

由表 3 分析可知,方法 1 直接采用相邻气象站点的气象数据,未考虑地理位置分布造成的信息偏差,集成模型推演的功率结果存在相应的误差;方法

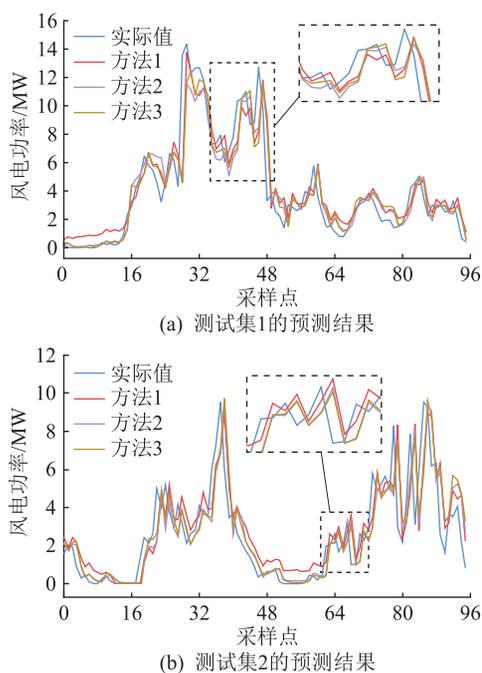


图6 风电功率预测结果对比

Fig.6 Comparison of wind power prediction results

表3 不同预测方法的预测误差对比

Table 3 Comparison of prediction errors of different prediction methods MW

预测方法	测试集 1		测试集 2	
	$E_{RMSE}$	$E_{MAE}$	$E_{RMSE}$	$E_{MAE}$
方法 1	0.921	0.663	0.962	0.670
方法 2	0.885	0.645	0.936	0.651
方法 3	0.872	0.641	0.921	0.650

2 采用平均法集成基学习器的预测结果,未能凸显不同学习器的性能差异,提升效果有限;方法 3 在 Stacking 集成学习的基础上,构建风速时移数据集,更加准确地挖掘目标风电场的气象特征,同时采用粒子群优化算法搜索超参数,提升了整体预测精度。以测试集 1 为例,相较于方法 1、2,方法 3 的 RMSE 分别下降 5.32%、1.47%,MAE 分别下降 3.32%、0.62%。实际算例表明,方法 3 在测试集上都取得了最佳的预测结果,证明了文中所提方法的优越性。文中方法在融合多元算法优势的同时,能够更好地应对气象信息缺失的状况。

#### 4 结论

在对风电场功率预测的过程中,考虑大气运动的连续性和集成模型的优越表现,提出了一种基于空间相关性与 Stacking 集成学习的风电功率预测模型,通过算例对比,得到如下结论:

(1) 利用相关系数的极值来确定相邻气象站点与目标风电场的延迟时间,构建风速时移数据集,

挖掘气象运动规律,完善目标风电场的气象数据。

(2) 采用 Stacking 集成方法融合多样化的基学习器,实现多元算法的优势互补,提高集成模型的泛化能力,同时,采用交叉验证的方式产生次级训练集,可以减小过拟合的风险。

(3) 针对集成学习中的众多超参数,采用粒子群优化算法进行寻优,在保证模型预测性能的同时,降低了搜索时间,提升了预测精度。

需要说明的是,文中的研究对象是整体风电场的输出功率,并不是风电场内部不同风电机组的功率,没有更细致地考虑风电场内部风电机组之间的影响,且风速的空间相关性也具有时变性能。因此,后续将综合考虑风速、风向等多种气象因素,通过动态划分不同风电场子集,采取分类建模提升预测效果,并在此基础上,引入不确定性分析,实现对风电功率的区间预测。

#### 致 谢

本文得到国网安徽省电力有限公司科技项目(B312C0230007)资助,谨此致谢!

#### 参考文献:

[1] Global wind report 2023 [EB/OL]. [2023-03-27]. <https://gwec.net/globalwindreport2023>.

[2] 马迪,高红均,张江林,等. 考虑风电不确定性的电-气综合能源系统协调优化[J]. 智慧电力,2022,50(11):33-40.  
MA Di,GAO Hongjun,ZHANG Jianglin, et al. Coordinated optimization of power-gas integrated energy system considering wind power uncertainty[J]. Smart Power,2022,50(11):33-40.

[3] MA Z J,MEI G. A hybrid attention-based deep learning approach for wind power prediction [J]. Applied Energy, 2022, 323: 119608.

[4] 王伟胜,王铮,董存,等. 中国短期风电功率预测技术现状与误差分析[J]. 电力系统自动化,2021,45(1):17-27.  
WANG Weisheng,WANG Zheng,DONG Cun, et al. Status and error analysis of short-term forecasting technology of wind power in China[J]. Automation of Electric Power Systems, 2021, 45(1):17-27.

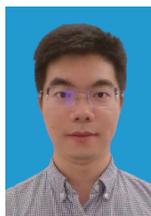
[5] 彭小圣,熊磊,文劲宇,等. 风电集群短期及超短期功率预测精度改进方法综述[J]. 中国电机工程学报,2016,36(23):6315-6326,6596.  
PENG Xiaosheng,XIONG Lei,WEN Jinyu, et al. A summary of the state of the art for short-term and ultra-short-term wind power prediction of regions[J]. Proceedings of the CSEE,2016,36(23):6315-6326,6596.

[6] 季玉琦,严亚帮,和萍,等. 基于 K-Medoids 聚类与栅格法提取负荷曲线特征的 CNN-LSTM 短期负荷预测[J]. 电力系统保护与控制,2023,51(18):81-93.  
JI Yuqi,YAN Yabang,HE Ping, et al. CNN-LSTM short-term load forecasting based on the K-Medoids clustering and grid method to extract load curve features[J]. Power System Protec-

- tion and Control, 2023, 51(18):81-93.
- [7] 赵洋,王瀚墨,康丽,等. 基于时间卷积网络的短期电力负荷预测[J]. 电工技术学报, 2022, 37(5):1242-1251.  
ZHAO Yang, WANG Hanmo, KANG Li, et al. Temporal convolution network-based short-term electrical load forecasting[J]. Transactions of China Electrotechnical Society, 2022, 37(5):1242-1251.
- [8] 陈逸枫,张大海,李宇欣,等. 基于 DWT-MOSMA-SVM 的多目标优化短期母线负荷预测[J]. 电力建设, 2023, 44(3):49-55.  
CHEN Yicong, ZHANG Dahai, LI Yuxin, et al. Multi-objective optimization based on DWT-MOSMA-SVM for short-term bus load forecasting[J]. Electric Power Construction, 2023, 44(3):49-55.
- [9] 杨光雨,李晓航. 基于最大信息挖掘宽度学习系统短期电力负荷预测研究[J]. 电测与仪表, 2022, 59(3):38-45.  
YANG Guangyu, LI Xiaohang. The short-term power load forecast in wide area learning system based on maximum information mining[J]. Electrical Measurement & Instrumentation, 2022, 59(3):38-45.
- [10] 龙干,黄媚,方力谦,等. 基于改进多元宇宙算法优化 ELM 的短期电力负荷预测[J]. 电力系统保护与控制, 2022, 50(19):99-106.  
LONG Gan, HUANG Mei, FANG Liqian, et al. Short-term power load forecasting based on an improved multi-verse optimizer algorithm optimized extreme learning machine[J]. Power System Protection and Control, 2022, 50(19):99-106.
- [11] WAN C, ZHAO C F, SONG Y H. Chance constrained extreme learning machine for nonparametric prediction intervals of wind power generation[J]. IEEE Transactions on Power Systems, 2020, 35(5):3869-3884.
- [12] 彭晨宇,陈宁,高丙团. 结合多重聚类和分层聚类的超短期风电功率预测方法[J]. 电力系统自动化, 2020, 44(2):173-180.  
PENG Chenyu, CHEN Ning, GAO Bingtuan. Ultra-short-term wind power forecasting method combining multiple clustering and hierarchical clustering[J]. Automation of Electric Power Systems, 2020, 44(2):173-180.
- [13] 叶瑞丽,郭志忠,刘瑞叶,等. 基于小波包分解和改进 Elman 神经网络的风电场风速和风电功率预测[J]. 电工技术学报, 2017, 32(21):103-111.  
YE Ruili, GUO Zhizhong, LIU Ruiye, et al. Wind speed and wind power forecasting method based on wavelet packet decomposition and improved Elman neural network[J]. Transactions of China Electrotechnical Society, 2017, 32(21):103-111.
- [14] 时珉,许可,王珏,等. 基于灰色关联分析和 GeoMAN 模型的光伏发电功率短期预测[J]. 电工技术学报, 2021, 36(11):2298-2305.  
SHI Min, XU Ke, WANG Jue, et al. Short-term photovoltaic power forecast based on grey relational analysis and GeoMAN model[J]. Transactions of China Electrotechnical Society, 2021, 36(11):2298-2305.
- [15] 唐锦,张书怡,吴秋伟,等. 基于 Copula 函数与等概率逆变换的风电出力场景生成方法[J]. 电力工程技术, 2021, 40(6):86-94.  
TANG Jin, ZHANG Shuyi, WU Qiwei, et al. Wind power output scenario generation method based on Copula function and equal probability inverse transformation[J]. Electric Power Engineering Technology, 2021, 40(6):86-94.
- [16] 李丹,甘月琳,缪书唯,等. 计及时间演变和空间相关的多风电场短期功率预测[J]. 电网技术, 2023, 47(3):1117-1128.  
LI Dan, GAN Yuelin, MIAO Shuwei, et al. Short-term power prediction for multiple wind farms considering temporal evolution and spatial correlation[J]. Power System Technology, 2023, 47(3):1117-1128.
- [17] 王晨,寇鹏,王若谷,等. 利用多空间尺度下时空相关性的点云分布多风机风速预测[J]. 电力系统自动化, 2021, 45(22):65-73.  
WANG Chen, KOU Peng, WANG Ruogu, et al. Wind speed forecasting for multiple wind turbines with point cloud distribution using spatio-temporal correlation on multiple spatial scale[J]. Automation of Electric Power Systems, 2021, 45(22):65-73.
- [18] 王铮, Rui Pestana, 冯双磊,等. 基于加权系数动态修正的短期风电功率组合预测方法[J]. 电网技术, 2017, 41(2):500-507.  
WANG Zheng, PESTANA R, FENG Shuanglei, et al. Short-term wind power combination forecasting method based on dynamic coefficient updating[J]. Power System Technology, 2017, 41(2):500-507.
- [19] 陈海文,王守相,王绍敏,等. 基于门控循环单元网络与模型融合的负荷聚合体预测方法[J]. 电力系统自动化, 2019, 43(1):65-72.  
CHEN Haiwen, WANG Shouxiang, WANG Shaomin, et al. Aggregated load forecasting method based on gated recurrent unit networks and model fusion[J]. Automation of Electric Power Systems, 2019, 43(1):65-72.
- [20] SU H Y, HUANG C R. Enhanced wind generation forecast using robust ensemble learning[J]. IEEE Transactions on Smart Grid, 2021, 12(1):912-915.
- [21] 梁超,刘永前,周家慷,等. 基于卷积循环神经网络的风电场多点位风速预测方法[J]. 电网技术, 2021, 45(2):534-542.  
LIANG Chao, LIU Yongqian, ZHOU Jiakang, et al. Wind speed prediction at multi-locations based on combination of recurrent and convolutional neural networks[J]. Power System Technology, 2021, 45(2):534-542.
- [22] 叶林,赵永宁. 基于空间相关性的风电功率预测研究综述[J]. 电力系统自动化, 2014, 38(14):126-135.  
YE Lin, ZHAO Yongning. A review on wind power prediction based on spatial correlation approach[J]. Automation of Electric Power Systems, 2014, 38(14):126-135.

- [23] 薛禹胜,陈宁,王树民,等. 关于利用空间相关性预测风速的评述[J]. 电力系统自动化,2017,41(10):161-169.  
XUE Yusheng, CHEN Ning, WANG Shumin, et al. Review on wind speed prediction based on spatial correlation[J]. Automation of Electric Power Systems, 2017, 41(10):161-169.
- [24] BOWDEN G J, BARKER P R, SHESTOPAL V O, et al. The Weibull distribution function and wind power statistics [J]. Wind Engineering, 1983, 7(2):85-98.
- [25] APT J. The spectrum of power from wind turbines[J]. Journal of Power Sources, 2007, 169(2):369-374.
- [26] 张凯锋,杨国强,陈汉一,等. 基于数据特征提取的风电功率预测误差估计方法[J]. 电力系统自动化,2014,38(16):22-27,34.  
ZHANG Kaifeng, YANG Guoqiang, CHEN Hanyi, et al. An estimation method for wind power forecast errors based on numerical feature extraction [J]. Automation of Electric Power Systems, 2014, 38(16):22-27, 34.
- [27] 赵铁成,谢丽蓉,叶家豪. 基于误差修正的 NNA-ILSTM 短期风电功率预测[J]. 智慧电力,2022,50(1):29-36.  
ZHAO Tiecheng, XIE Lirong, YE Jiahao. NNA-ILSTM short term wind power prediction based on error correction [J]. Smart Power, 2022, 50(1):29-36.
- [28] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:33-35,183-190.  
ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016:33-35, 183-190.
- [29] 史佳琪,张建华. 基于多模型融合 Stacking 集成学习方式的负荷预测方法[J]. 中国电机工程学报, 2019, 39(14):4032-4042.  
SHI Jiaqi, ZHANG Jianhua. Load forecasting based on multi-model by Stacking ensemble learning [J]. Proceedings of the CSEE, 2019, 39(14):4032-4042.
- [30] 王振浩,王翀,成龙,等. 基于集合经验模态分解和深度学习的光伏功率组合预测[J]. 高电压技术, 2022, 48(10):4133-4142.  
WANG Zhenhao, WANG Chong, CHENG Long, et al. Photovoltaic power combined prediction based on ensemble empirical mode decomposition and deep learning [J]. High Voltage Engineering, 2022, 48(10):4133-4142.
- [31] 苏晨博,刘崇茹,徐诗甜,等. 利用贝叶斯线性回归结合混合 Copula 函数分析风电功率的相关性[J]. 中国电力, 2021, 54(8):182-189.  
SU Chenbo, LIU Chongru, XU Shitian, et al. Mix Copula function based wind power correlation analysis: a Bayesian linear regression approach [J]. Electric Power, 2021, 54(8):182-189.

作者简介:



王小明

王小明(1987),男,硕士,高级工程师,从事电能质量及综合能源技术相关工作(E-mail:316855097@qq.com);

徐斌(1980),男,硕士,高级工程师,从事配电网系统规划与设计相关工作;

尹元亚(1972),男,硕士,高级工程师,从事调度运行、继电保护相关工作。

## Wind power prediction method based on spatial correlation and Stacking ensemble learning

WANG Xiaoming<sup>1</sup>, XU Bin<sup>1</sup>, YIN Yuanya<sup>2</sup>, PAN Wenhui<sup>2</sup>, WU Hongbin<sup>3</sup>, HAN Yi<sup>3</sup>

(1. Electric Power Research Institute, State Grid Anhui Electric Power Co., Ltd., Hefei 230601, China;

2. Wuhu Power Supply Company, State Grid Anhui Electric Power Co., Ltd., Wuhu 241000, China;

3. School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China)

**Abstract:** Aiming at the problem of insufficient accuracy in wind power prediction caused by the lack of target meteorological forecast data, a wind power prediction method based on spatial correlation and Stacking ensemble learning is proposed in this paper. Firstly, the spatial correlation between the target wind farm and adjacent meteorological stations is analyzed. The delay time is determined based on the extreme points of the correlation coefficient, and a wind speed time-shift dataset is constructed. Secondly, the Stacking ensemble method is used to integrate multiple algorithms to predict the wind power of the target wind farm from multiple data observation perspectives. It leverages the complementary advantages of different algorithms, enhancing overall generalization ability. Additionally, it can effectively balance search time and model performance by adopting the particle swarm optimization algorithm to search for model hyperparameters. Finally, the effectiveness and accuracy of the proposed method have been verified by the measured data from a wind farm in East China. The results show that it can effectively improve the wind power prediction accuracy in the case of missing data by considering the information bias at different locations with the data input and prediction models.

**Keywords:** wind power prediction; spatial correlation; Stacking ensemble learning; wind speed time-shift; multiple algorithms fusion; particle swarm optimization

(编辑 钱悦 陆海霞)