

DOI: 10.13718/j.cnki.xdzk.2025.03.018

罗兴, 胡子健, 马洲俊, 等. 基于混合深度学习的光伏集群发电预测 [J]. 西南大学学报(自然科学版), 2025, 47(3): 200-210.

# 基于混合深度学习的光伏集群发电预测

罗兴, 胡子健, 马洲俊, 吕湛

国网江苏省电力有限公司南京供电分公司, 南京 210024

**摘要:** 光伏发电作为可再生能源在实现能源转型和减缓气候变化方面具有重要意义, 然而光伏发电预测面临诸多挑战, 包括天气不确定性、数据质量问题、复杂的影响因素以及模型复杂度和计算成本等, 并且往往只考虑单个站点的相关因素, 并没有考虑多个站点之间的相互影响。传统预测方法泛化能力有限, 不能很好地捕捉复杂的非线性关系, 缺乏足够的灵活性。为解决这些问题, 提出基于混合深度学习的光伏集群发电预测框架, 利用站点历史发电数据计算互信息并形成集群网络结构, 然后构建具有不同特征的经典深度学习模型进行预测, 包括长短期记忆网络(LSTM)、卷积神经网络(CNN)、时序卷积网络(TCN)以及极端梯度提升(XGBOOST), 最后通过自适应权重计算实现模型集成学习并完成预测。对国家可再生能源实验室(NREL)真实光伏发电量数据进行预测, 实验结果表明: 集成模型具有更好的泛化能力并且在预测精度上相较单一模型也有显著提升。

**关键词:** 混合深度学习; 光伏集群发电预测; 自适应权重;

互信息

中图分类号: TM615; TP183

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2025)03-0200-11

## Photovoltaic Cluster Power Generation Prediction Based on Hybrid Deep Learning

LUO Xing, HU Zijian, MA Zhoujun, LYU Zhan

Nanjing Power Supply Company, State Grid Jiangsu Electric Power Co. Ltd., Nanjing 210024, China

**Abstract:** Photovoltaic (PV) power generation, as a renewable energy source, is considered significant in achieving energy transformation and mitigating the impacts of climate change. Despite its potential, the prediction of photovoltaic power generation faces several challenges, including weather uncertainty, data quality issues, complex influencing factors, model complexity and computational costs. Moreover, existing prediction methods often only consider relevant factors at a single site, neglecting the interactions between multiple sites. These traditional approaches have limited the generalization ability and struggled

收稿日期: 2024-04-03

基金项目: 国家重点研发计划项目(2022YFB2404200); 国网江苏省电力有限公司科技项目(J2023017)。

作者简介: 罗兴, 高级工程师, 主要从事电网调度及其自动化技术研究。

to capture the complex nonlinear relationships effectively, lacking the necessary flexibility. To address these issues, a hybrid deep learning-based photovoltaic cluster power generation prediction framework was proposed in this paper. Firstly, historical power generation data from stations were utilized to calculate mutual information for forming a cluster network structure. Subsequently, classic deep learning models with different features were constructed for prediction, including the Long Short-Term Memory Network (LSTM), Convolutional Neural Network (CNN), Time Series Convolutional Network (TCN), and Extreme Gradient Boosting (XGBOOST). Finally, model ensemble learning was achieved through adaptive weight calculation to complete the prediction. Real photovoltaic power generation data from the National Renewable Energy Laboratory (NREL) were utilized in this study for prediction. The experimental results demonstrate that the integrated model exhibits superior generalization ability and significantly improves the prediction accuracy compared to individual model.

**Key words:** hybrid deep learning; photovoltaic cluster power generation prediction; adaptive weight; mutual information

近几十年来,由于化石燃料的过度消耗,人类面临着全球变暖和能源危机的问题,这些问题对全球经济政策、气候状况以及能源安全都产生了巨大影响。然而随着工业化和城市化的加速发展,能源需求持续增长。传统的化石燃料发电方式不仅昂贵,而且对环境造成了严重的污染和破坏,加剧了全球变暖的问题,导致气候变化引发的自然灾害频频发生<sup>[1]</sup>。为了应对这一挑战,全球各地开始寻求清洁、可持续的替代能源,其中包括太阳能、风能、生物质能、水力能以及地热能等。在这些替代能源中,太阳能光伏发电被认为是最具前景和可行的选择之一<sup>[2]</sup>,并且对我国 2030 年前实现“碳达峰”,2060 年前实现“碳中和”的目标具有重大意义<sup>[3]</sup>。太阳能光伏发电具有诸多优势,例如可再生性和环保性。然而光伏发电的一个关键挑战在于其发电量受到天气等多种因素的影响,这使得光伏发电量难以准确预测。同时传统光伏发电预测很少考虑多站点之间的关联性,为了更好地规划能源供应、电网管理和降低能源成本,对光伏集群发电预测的研究和应用具有重要的现实意义。

光伏发电预测方法主要有统计方法和机器学习方法。其中统计方法包括自回归移动平均模型(Auto Regressive and Moving Average, ARMA)、自回归积分移动平均模型(Auto Regressive Integrated Moving Average, ARIMA)和多元线性回归(Multiple Linear Regression, MLR)。ARMA 模型在时间序列分析领域应用广泛,常用于处理平稳序列,进行短期负荷预测<sup>[4-5]</sup>,但针对非线性数据等复杂情况时很难取得较好效果。MLR 模型能够有效地应对非线性数据,并能通过多因素考虑来进行预测<sup>[6]</sup>,但是存在过拟合风险并且不能捕捉时序数据的时间特征。而机器学习方法能够更好地处理复杂的非线性关系以及大量数据,能够更好地适应不同的环境变化和特征,具有更强的泛化能力和灵活性。通过机器学习方法进行光伏发电预测的准确率也高于利用统计方法所得的准确率<sup>[7]</sup>。常用的机器学习方法包括深度神经网络(Deep Neural Network, DNN)、支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)<sup>[8]</sup>等。在文献[9-10]中,分别使用了支持向量回归模型(Support Vector Regression, SVR)与 RF 模型、LSTM 模型与 RF 模型来完成光伏发电预测,说明了机器学习模型在此方面能够有效提高预测的准确率。但单一的机器学习模型存在过拟合问题,对数据分布也比较敏感,并且存在较多不确定性,导致不能在较多数据集情况中表现良好,达不到较好的综合性能。集成模型能够降低过拟合的风险,对数据分布也不敏感,具有较强的抗噪能力,以此来获得更好的性能。

因此,本文提出集成 LSTM、CNN、TCN 和 XGBOOST 单一模型的集成模型,集成模型可以聚合不同单一模型的各自特点,并结合互信息(Mutual Information, MI)集群网络结构,实现更精准的光伏集群发电预测。所提出的方法利用国家可再生能源实验室(National Renewable Energy Laboratory, NREL)的

光伏发电数据集<sup>[11]</sup>作为模型输入,利用历史数据计算站点之间的互信息,形成互信息集群网络结构,再构建具有不同特点的4个经典机器学习模型,使用自适应加权策略来对单一模型进行集成,从而实现更加精准的预测以及获得更强的泛化能力。

## 1 相关研究工作

在统计方法方面,Chen等<sup>[4]</sup>利用ARMA模型来预测平稳的残差序列,以提高预测精度。Colak等<sup>[12]</sup>使用ARMA模型和ARIMA模型进行多周期预测,对2个模型结果进行比较,作为确定多周期预测中最佳统计模型的结果。然而传统的统计方法对于具有复杂非线性结构的光伏发电数据不适用,于是Chen等<sup>[13]</sup>将ARIMA模型和LSTM模型进行结合,实现线性和非线性预测,可以更全面地从历史数据中提取模式,有效地减少了预测误差。

在机器学习方法方面,VanDeventer等<sup>[14]</sup>提出了一种基于遗传算法的支持向量机(GASVM)模型,用于住宅规模光伏系统的短期功率预测,先利用SVM模型对历史数据进行分类,再利用遗传算法进行优化。Pan等<sup>[15]</sup>提出了一种改进的蚁群优化(Ant Colony Optimization, ACO)算法,使用ACO算法的全局优化函数来优化SVM模型的参数,完成光伏发电预测。但SVM模型处理大规模数据集和高维特征空间的能力较弱,并且对核函数参数或惩罚参数的选择十分敏感,需要较大计算量才能确定对应参数。RF模型在文献<sup>[16]</sup>中被用来提高光伏发电预测的稳健性,研究者运用了该算法来降低特征空间的维度,并通过多个决策树的集成来提高预测的准确性,在缺乏数据的情况下也能取得良好的效果。Abuella等<sup>[9]</sup>和Roy等<sup>[10]</sup>采用RF模型进行预测,其预测输出是单个树的预测平均值,因此提供了具有较小误差的预测结果。Izzatillaev等<sup>[17]</sup>使用人工神经网络模型(Artificial Neural Network, ANN)进行预测,并得到了更高的精度。但上述机器学习方法不能有效地探索和捕捉与预测目标密切相关的各种特征之间的深层潜在关联。

深度学习技术通过扩展神经网络的深度,能够更有效地挖掘复杂数据中的非线性结构,并从中学到更具性能的特征表示。Kong等<sup>[18]</sup>使用递归神经网络模型(Recurrent Neural Network, RNN)来减轻负荷预测的不确定性。但由于梯度消失或爆炸问题,RNN模型难以维持中长期依赖性。针对此问题,LSTM模型通过引入门控机制来解决,利用存储单元来保持时间相关性,其能够有效地捕捉长期依赖关系,并且对序列数据的建模能力更加强。在文献<sup>[10, 13, 19-22]</sup>中都使用了LSTM模型。此外,Deng等<sup>[23]</sup>提出了一种具有时间认知的多尺度卷积神经网络模型(Multi-Scale Convolutional Neural Network with Time-Cognition, TCMS-CNN),该模型基于多尺度卷积的深度卷积神经网络模型(Multi-Scale Convolutional Neural Network, MS-CNN)提取不同层次的特征,并将其融合到网络中。同时还设计了一种创新的时间编码策略,有效地增强了序列模型的时间认知能力<sup>[24]</sup>。

然而,单一机器学习模型泛化能力有限,不能很好地捕捉复杂的非线性关系,缺乏足够的灵活性。针对此问题,利用集成学习的方法,能够有效实现更加稳健、精准的预测。在电力方面,Wang等<sup>[25]</sup>使用了集成学习方法来提高用电量预测的精准度。集成学习方法能够增强模型的泛化能力,面对复杂的非线性关系数据也能较好地捕捉其中的特征,面对噪声和异常值也有较强的容忍性、良好的稳健性和更高的预测精度。同时,传统的电力预测模型一般仅考虑单个站点的相关影响因素,例如发电量、天气及地理位置等,并未考虑多个发电站之间的相关性。多站点间发电量的相关性揭示了不同发电站之间发电量变化的相互影响和依赖关系,这种相关性不仅可以用于预测特定站点的发电量,还可以优化发电站的运营策略,降低风险,为新能源发电系统的规划和建设提供参考依据,对提高发电系统的效率和可靠性具有重要意义。

## 2 模型方法

图1对本文集成模型总体结构进行了说明。该模型将发电量数据作为输入,首先分析发电站站点之间的网络结构,利用历史发电量来计算发电站之间的互信息,衡量站点之间的相关性,借此构成具有不同集

群的网络结构,然后再构建具有不同特征的4个单一深度学习模型进行预测,根据预测结果完成权重自适应计算并实现模型集成,最后输出最终预测结果。接下来将先对真实数据集进行分析,再介绍4种单一模型以及集成模型。

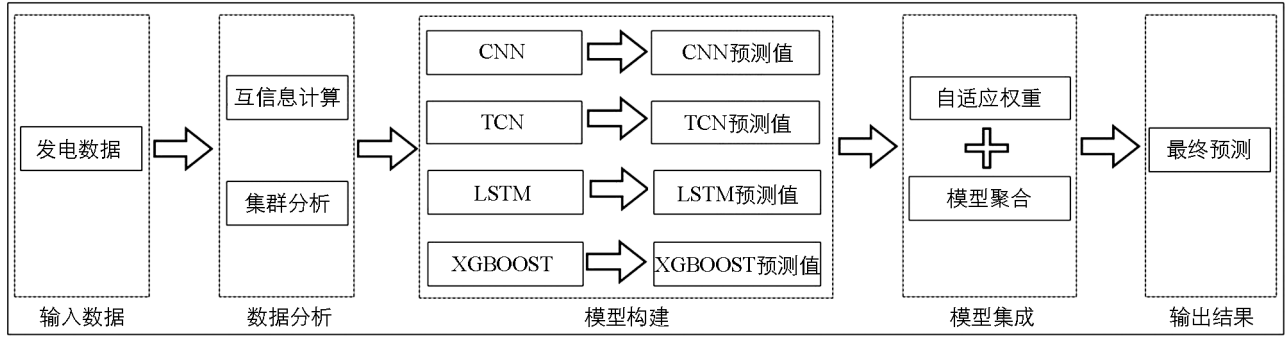


图 1 本文集成模型的总体结构

## 2.1 数据分析

数据来源于美国国家可再生能源实验室网站(www.nrel.gov)所提供的光伏发电量数据,选取美国东部阿拉巴马州137个站点在2006年的光伏发电量作为实验数据。

传统发电预测一般考虑单个发电站的相关影响因素,未考虑多个发电站之间的相关性。因此,本文利用互信息构建网络结构。所有发电站之间的互信息网络结构如图2所示,节点代表发电站,利用历史发电量数据计算互信息来衡量发电站之间的相关性并作为边的权重。互信息的计算公式为:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log \left( \frac{P(x, y)}{P(x)P(y)} \right) \quad (1)$$

式中:  $X$  和  $Y$  为 2 个离散的随机变量;  $P(x, y)$  为  $X$  和  $Y$  的联合概率分布;  $P(x)$  和  $P(y)$  分别为对应的边缘概率分布。

得到所有发电站之间的互信息网络结构后,通过分析对应互信息热力图(图3),能够根据互信息将发电站分为多个集群(如图2中虚线圈代表一个集群),并从中选取站点,利用与其相关性较高的站点发电量作为特征来进行集群发电预测。

## 2.2 卷积神经网络(CNN)

CNN 已经在诸如图像和视频识别、图像分类、图像处理、医学图像分析和自然语言处理等领域取得了巨大成功。除了这些领域之外,CNN 在基于时间序列数据领域也日益受到青

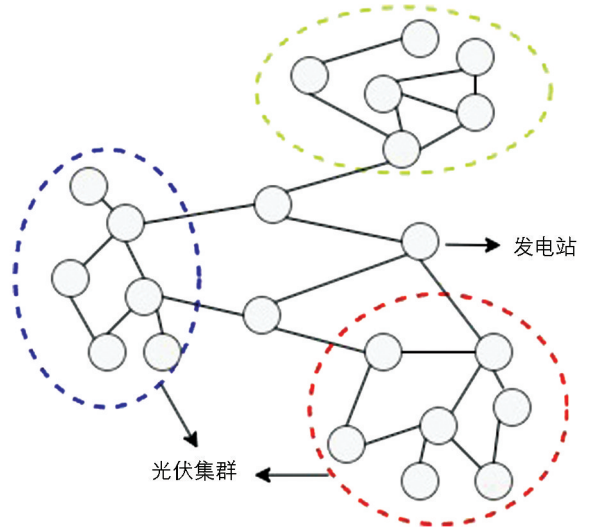


图 2 发电站网络结构示意图

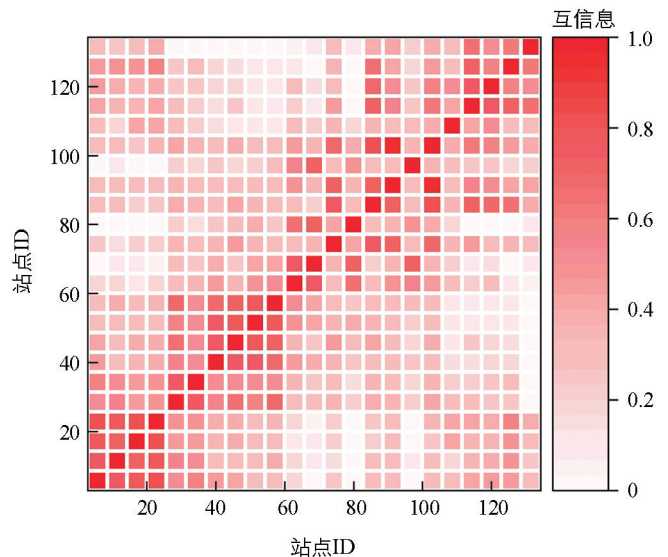


图 3 互信息热力图

睐,这主要归因于它具备处理多通道输入数据的能力<sup>[26]</sup>。其中输入数据  $f$  和卷积核  $g$  的卷积运算数学表达式如下:

$$S(i, j) = (g * f)(i, j) = \sum_m \sum_n g(m, n) \cdot f(i - m, j - n) \quad (2)$$

式中:  $S(i, j)$  为卷积运算的结果;  $m, n$  为卷积核的尺寸。通过卷积操作能够实现对输入数据的特征提取,从而在时间序列分析、行为识别、交通流量预测等任务中展现出优越的性能。

### 2.3 极端梯度提升(XGBOOST)

XGBOOST<sup>[27-28]</sup>是一种迭代训练多个决策树的梯度提升树算法。在每一轮迭代中 XGBOOST 训练一棵新的决策树,该决策树校正前一棵树的残差,然后将它们的输出进行聚合以进行预测。模型的最终输出是多个决策树输出的累积和,每个决策树的贡献由学习率控制。这种策略使其能够通过集成多个决策树来提高预测性能,并且具有较高的泛化能力。对于第  $i$  个样本的预测值,它可以表示为所有树的输出累积:

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

式中:  $K$  为树的数量;  $f_k(x_i)$  为样本  $x_i$  的第  $k$  棵树的输出。在训练模型方面,定义了损失函数和正则化项,旨在使得损失函数最小化,以下是目标函数表达式:

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (4)$$

式中:  $l$  为损失函数,计算预测值和真实值之间的误差;  $\Omega(f_k)$  为正则化项,控制每棵树的复杂度。

### 2.4 长短期记忆网络(LSTM)

作为 RNN 的一种变体, LSTM 于 1997 年被提出<sup>[29]</sup>。LSTM 的结构包括 1 个输入门、1 个遗忘门和 1 个输出门,通过控制这 3 个门的开闭程度, LSTM 能够学习需要存储、丢弃和读取的信息。输入门负责确定新信息的重要性,遗忘门决定旧信息的保留程度,而输出门则控制当前单元状态的输出。通过这种机制, LSTM 能够有效地长期记忆序列中的重要信息,并同时提取短期特征,使得网络更加适应序列数据的处理需求<sup>[30]</sup>。LSTM 的数学表达式如下:

$$f(t) = \sigma(\mathbf{w}_f^T \cdot [h(t-1), x(t)] + b_f) \quad (5)$$

$$i(t) = \sigma(\mathbf{w}_i^T \cdot [h(t-1), x(t)] + b_i) \quad (6)$$

$$g(t) = \tanh(\mathbf{w}_g^T \cdot [h(t-1), x(t)] + b_g) \quad (7)$$

$$c(t) = f(t) \cdot c(t-1) + i_t \cdot g(t) \quad (8)$$

$$o(t) = \sigma(\mathbf{w}_o^T \cdot [h(t-1), x(t)] + b_o) \quad (9)$$

其中:  $f(t)$  为遗忘门的输出;  $i(t)$  为输入门的输出;  $o(t)$  为输出门的输出;  $g(t)$  为候选记忆细胞的输出;  $c(t)$  为该时刻的细胞状态,其根据输入门的输出和遗忘门的输出,更新细胞状态,新的细胞状态由当前输入和上一时刻的细胞状态加权求和得到;  $\sigma$  为激活函数;  $\mathbf{w}, b$  为相应门的权重矩阵和偏置。

### 2.5 时序卷积网络(TCN)

TCN<sup>[31]</sup>是一种专为序列建模而设计的深度学习架构。不同于传统的 RNN, TCN 采用 CNN 来捕捉序列中的长期依赖关系。每层卷积层包含激活函数以及残差连接,以下是其运算表达式:

$$Y_i = H(X_i) + X_i \quad (10)$$

式中:  $X_i = (x_1, x_2, \dots, x_T)$  为输入序列;  $Y_i = (y_1, y_2, \dots, y_T)$  为输出序列;  $H(X_i)$  为卷积层输出。TCN 中使用的残差连接机制有助于缓解梯度消失和爆炸问题,提高了模型的稳定性和收敛速度。

### 2.6 集成模型

自适应聚合集成学习<sup>[31]</sup>通过自适应地调整每个基础模型的权重,可以有效地提高模型的泛化能力和灵活性,并能够很好地平衡各个模型的性能,在降低过拟合风险的同时也能提升预测效果。自适应聚合权

重的计算表达式为:

$$Al^t(n) = \sum_{h=1}^n Er^t(h) \quad (11)$$

$$\epsilon^t(n) = \frac{e^{-\delta Al^t(n)}}{\sum_{t=1}^4 e^{-\delta Al^t(n)}} \quad (12)$$

其中:  $Er^t(h)$  为第  $n$  次迭代时第  $t$  个基本模型的单独损失;  $Al^t(n)$  为到第  $n$  次训练迭代时第  $t$  个基本模型的累积损失;  $\epsilon^t(n)$  为第  $t$  个基本模型的集成权重;  $\delta = \sqrt{1/\ln T}$  为在训练过程中控制集合的聚合权重的平衡因子;  $T$  为最大训练迭代次数。

## 3 实验部分

### 3.1 数据集处理

根据发电站互信息集群网络结构,从 4 个集群中选取 4 个发电站进行光伏集群发电预测,将与被选择站点相关性高的站点发电数据作为其特征来完成预测。首先对数据集进行归一化处理,再利用滑动窗口方法将数据集构造成相应的训练数据和对应标签,最后将数据集按照 8:1:1 的比例划分成训练集、验证集和测试集。

### 3.2 模型参数

选择 CNN、TCN、LSTM 和 XGBOOST 模型来进行集成学习,所有模型参数如表 1 所示。

表 1 模型参数设置

模型	参数	
CNN	学习率: $1 \times 10^{-6}$	卷积核大小: 3
	通道数: 64	填充大小: 1
TCN	学习率: $1 \times 10^{-5}$	卷积核大小: 3
	丢弃率: 0.1	通道数: 64
LSTM	学习率: $1 \times 10^{-6}$	隐藏层维度: 36
XGBOOST	学习率: 0.01	最大深度: 3

### 3.3 实验设置

#### 3.3.1 实验设计

为了评估所提出的集成方法,设计了以下 2 组实验:

- 1) 单一模型与集成模型比较。
- 2) 预测模型的超参数分析。

#### 3.3.2 评价指标

实验利用均方误差(Mean Square Error, MSE)和平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)评估模型的预测能力,这些评价指标被广泛应用于时间序列预测方面,都具有较为直观的解释性,以下是其计算公式:

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

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (14)$$

其中:  $n$  为样本总数;  $y_i$  为第  $i$  个真实值;  $\hat{y}_i$  为第  $i$  个预测值。  $E_{MSE}$  和  $E_{MAPE}$  的值越小,说明预测值

与真实值越接近, 对应模型预测精度越高。

### 3.4 单一模型与集成模型的比较

在此实验环节, 将单一模型即 CNN、XGBOOST、LSTM 和 TCN 与集成模型的预测能力进行对比, 实验结果如表 2 和表 3 所示。其中表 2 是基于评价指标  $E_{MSE}$  得到的, 表 3 是基于评价指标  $E_{MAPE}$  得到的。根据实验结果, 本文还通过 Wilcoxon 符号秩检验和 Friedman 检验进行了统计分析<sup>[32-33]</sup>, 以下是分析结果:

1) 集成模型针对 2 个指标都表现出较好的预测结果, 在 4 个站点的单步预测指标都优于其余 4 个单一模型的指标。

2) 集成模型在 4 个站点的预测中, 指标  $E_{MSE}$  始终低于 10.00, 最低可至 0.000 2, 指标  $E_{MAPE}$  始终低于 112.06, 最低可至 0.19。

集成模型在预测精度上相较单一模型有明显提升, 针对不同光伏电站的发电量预测也有较强的泛化能力。

表 2 基于  $E_{MSE}$  的对比实验结果

	日期	CNN	TCN	LSTM	XGBOOST	集成模型
站点 1	第 1 天	6.41	8.00	7.21	6.98	6.24
	第 2 天	4.61	3.33	4.31	3.98	2.82
	第 3 天	1.17	1.91	0.20	0.50	0.04
	第 4 天	10.93	16.69	13.41	12.42	9.86
	第 5 天	11.71	13.25	10.72	9.84	7.99
站点 2	第 1 天	7.55	7.68	11.92	9.10	3.01
	第 2 天	0.37	0.45	0.63	0.76	0.20
	第 3 天	0.92	0.38	1.69	0.01	0.000 2
	第 4 天	0.84	0.86	0.61	0.93	0.08
	第 5 天	2.54	4.02	4.47	3.12	0.92
站点 3	第 1 天	7.94	10.34	10.29	10.50	7.57
	第 2 天	1.35	1.18	3.22	1.68	0.90
	第 3 天	0.79	0.12	0.34	0.04	0.002 7
	第 4 天	2.75	3.82	6.63	3.15	1.45
	第 5 天	8.24	10.89	11.77	8.93	7.01
站点 4	第 1 天	0.18	0.23	0.08	0.52	0.01
	第 2 天	1.78	1.21	0.92	0.37	0.14
	第 3 天	0.92	0.88	0.94	1.15	0.42
	第 4 天	9.26	19.50	12.08	12.22	9.06
	第 5 天	11.56	19.90	11.97	10.54	10.00
统计分析	平均值	4.59	6.23	5.67	4.84	3.39
	$p$ 值	$4.76 \times 10^{-5}$	$4.78 \times 10^{-5}$	$4.78 \times 10^{-5}$	$4.78 \times 10^{-5}$	
	$F$ 值	3.0	3.8	3.8	3.4	1.0

注: 如果  $p$  值低于 0.05, 本文提出的集成模型显著优于比较模型;  $F$  值越低表示性能越好。下同。

表 3 基于  $E_{MAPE}$  的对比实验结果

	日期	CNN	TCN	LSTM	XGBOOST	集成模型
站点 1	第 1 天	32.81	36.67	34.82	34.27	32.38
	第 2 天	30.40	25.83	29.36	28.24	23.74
	第 3 天	25.65	32.77	10.60	16.91	4.90
	第 4 天	39.14	48.36	43.35	41.72	37.17
	第 5 天	41.69	44.35	39.89	38.23	34.44
站点 2	第 1 天	55.02	55.46	69.11	60.39	34.71
	第 2 天	6.74	7.47	8.83	9.79	4.96
	第 3 天	12.64	8.11	17.13	1.84	0.19
	第 4 天	10.23	10.31	8.69	10.77	3.19
	第 5 天	25.77	32.45	34.20	28.58	15.54
站点 3	第 1 天	33.25	37.94	37.84	38.24	32.45
	第 2 天	17.23	16.14	26.65	19.26	14.07
	第 3 天	18.43	7.30	12.12	4.63	1.07
	第 4 天	51.25	60.43	79.60	54.92	37.17
	第 5 天	35.24	40.50	42.12	36.70	32.51
站点 4	第 1 天	7.01	7.99	4.77	11.95	1.83
	第 2 天	17.97	14.81	12.91	8.24	5.06
	第 3 天	16.99	16.68	17.25	19.05	11.47
	第 4 天	113.31	164.40	129.42	130.16	112.06
	第 5 天	34.40	45.14	35.01	32.85	31.99
统计分析	平均值	31.26	35.66	34.68	31.34	23.55
	$p$ 值	$4.78 \times 10^{-5}$	$4.78 \times 10^{-5}$	$4.78 \times 10^{-5}$	$4.78 \times 10^{-5}$	
	$F$ 值	3.0	3.8	3.8	3.4	1.0

### 3.5 集成模型分析

本实验环节对集成模型的集成过程和性能进行了可视化,图 4 和图 5 分别是集成模型对 4 个站点发电量预测的收敛曲线图和权重图,根据以下可视化信息,可以得出以下结论:

- 1) 所有模型在不同站点的数据下均能够随着训练次数的增加逐渐收敛,并得到最优预测结果,但单一模型不能在所有情况下都具有较好的预测结果。
- 2) 集成模型能够根据误差自适应调整权重,在 4 个发电站的发电量预测中获得良好的预测结果。
- 3) 集成模型相较单一模型,通过自适应权重计算,不仅在多种情况下能够拥有较好的性能,并且具备良好的灵活性。

## 4 结论

本文提出了一种基于混合深度学习的光伏集群发电预测模型。该模型首先根据历史数据构建互信息集群网络结构,基于此对经典单一深度学习模型 CNN、TCN、LSTM 和 XGBOOST 进行构建并完成初步预

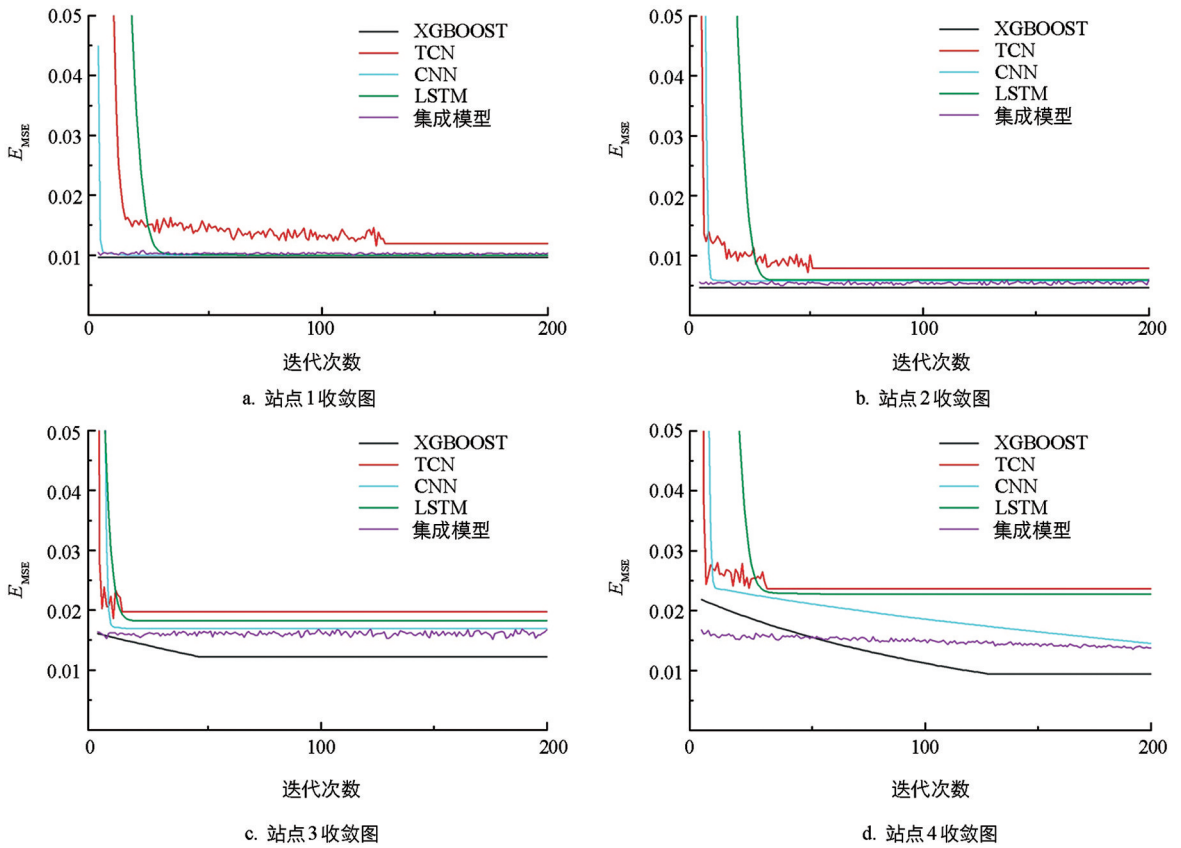


图 4 收敛曲线图

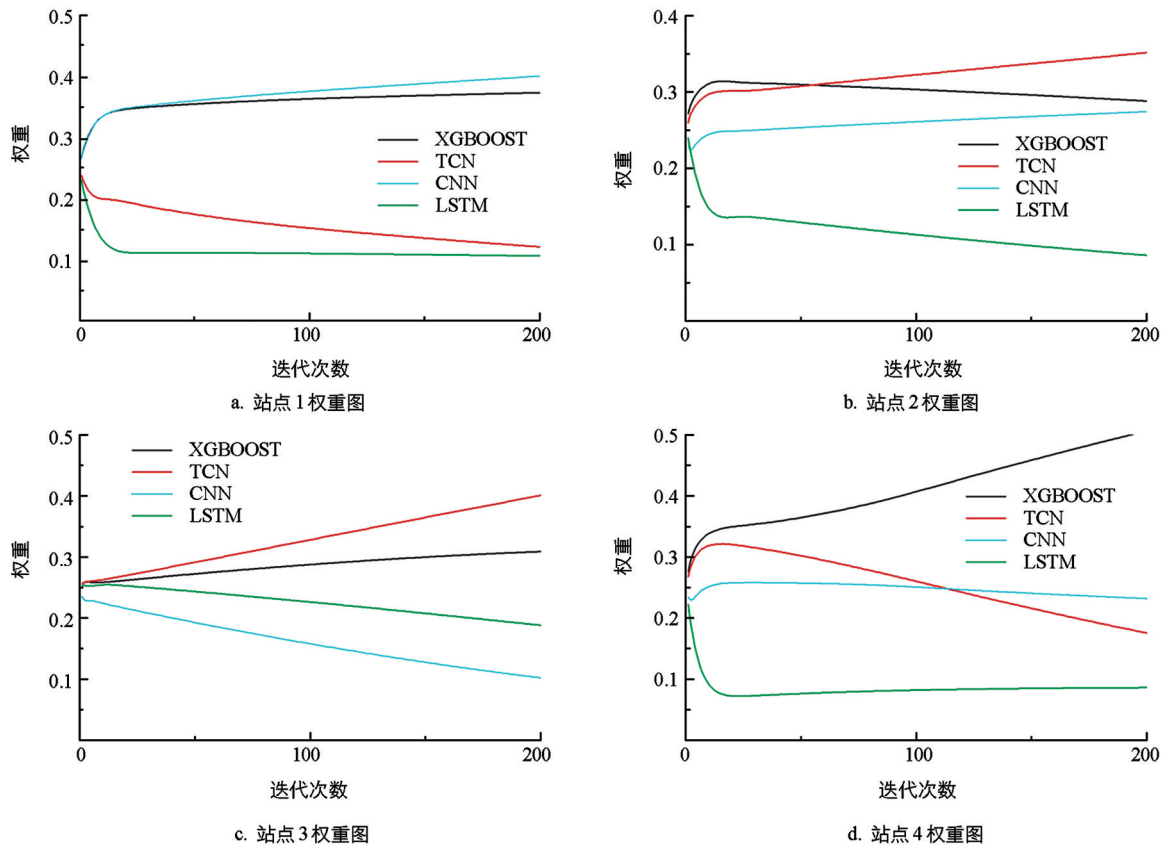


图 5 权重图

测。然后再根据其预测结果实现自适应权重计算,完成集成学习得到最终预测结果。集成模型相较于单一模型具有良好的灵活性和泛化能力。通过实验证明,该集成模型在不同光伏电站的发电量预测中,其预测精度较单一模型都有显著提升。本文主要集中在利用多个发电站的发电量数据进行预测,并没有考虑气象等因素对发电量的影响。在之后的研究中,将考虑地理位置或气象等因素对发电量的影响,以提高模型在各种复杂环境条件下的准确性和灵活性,为实际应用提供更可靠的预测结果。

### 参考文献:

- [1] WANG G C, SU Y, SHU L J. One-Day-Ahead Daily Power Forecasting of Photovoltaic Systems Based on Partial Functional Linear Regression Models [J]. *Renewable Energy*, 2016, 96: 469-478.
- [2] LI Y T, HE Y, SU Y, et al. Forecasting the Daily Power Output of a Grid-Connected Photovoltaic System Based on Multivariate Adaptive Regression Splines [J]. *Applied Energy*, 2016, 180: 392-401.
- [3] 孙凡, 李兴隆, 肖强, 等. 渝东南山地风能资源及风切变研究 [J]. *西南大学学报(自然科学版)*, 2023, 45(4): 226-230.
- [4] CHEN J F, WANG W M, HUANG C M. Analysis of an Adaptive Time-Series Autoregressive Moving-Average (ARMA) Model for Short-Term Load Forecasting [J]. *Electric Power Systems Research*, 1995, 34(3): 187-196.
- [5] JI W, CHEE K C. Prediction of Hourly Solar Radiation Using a Novel Hybrid Model of ARMA and TDNN [J]. *Solar Energy*, 2011, 85(5): 808-817.
- [6] SABER A Y, ALAM A K M R. Short Term Load Forecasting Using Multiple Linear Regression for Big Data [C] // 2017 IEEE Symposium Series on Computational Intelligence (SSCI), November 27-December 1, 2017, Honolulu, HI, USA. IEEE, 2017: 1-6.
- [7] LEVA S, DOLARA A, GRIMACCIA F, et al. Analysis and Validation of 24 Hours Ahead Neural Network Forecasting of Photovoltaic Output Power [J]. *Mathematics and Computers in Simulation*, 2017, 131: 88-100.
- [8] GABOITAOLELWE J, ZUNGERU A M, YAHYA A, et al. Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison [J]. *IEEE Access*, 2023, 11: 40820-40845.
- [9] ABUELLA M, CHOWDHURY B. Random Forest Ensemble of Support Vector Regression Models for Solar Power Forecasting [C] // 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), April 23-26, 2017, Washington, DC, USA. IEEE, 2017: 1-5.
- [10] ROY N, TRIPATHY P, DE S C, et al. Day-Ahead Solar Power Generation Forecasting Using LSTM and Random Forest Methods for North Eastern Region of India [C] // 2022 22nd National Power Systems Conference (NPSC), December 17-19, 2022, New Delhi, India. IEEE, 2022: 854-859.
- [11] U. S. National Renewable Energy Laboratory. Solar Power Data for Integration Studies [DB/OL]. (2006-12-31) [2024-03-15]. <https://www.nrel.gov/grid/solar-power-data.html>.
- [12] COLAK I, YESILBUDAK M, GENÇ N, et al. Multi-Period Prediction of Solar Radiation Using ARMA and ARIMA Models [C] // 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), December 9-11, 2015, Miami, FL, USA. IEEE, 2015: 1045-1049.
- [13] CHEN X B, XIE B, ZHANG P F, et al. Research on Wind and Solar Power Generation Forecasting Based on SARIMA-LSTM Model [C] // 2023 3rd International Conference on New Energy and Power Engineering (ICNEPE), November 24-26, 2023, Huzhou, China. IEEE, 2023: 695-699.
- [14] VANDEVENTER W, JAMEI E, THIRUNAVUKKARASU G S, et al. Short-Term PV Power Forecasting Using Hybrid GASVM Technique [J]. *Renewable Energy*, 2019, 140: 367-379.
- [15] PAN M Z, LI C, GAO R, et al. Photovoltaic Power Forecasting Based on a Support Vector Machine with Improved Ant Colony Optimization [J]. *Journal of Cleaner Production*, 2020, 277: 123948.
- [16] ZHOU M J, LI X F, ZHONG Z Y. Research on Photovoltaic Power Generation Power Prediction Based on RF-XGBoost

- Model [C] //2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICIS-CAE), September 23-25, 2023, Dalian, China. IEEE, 2023: 544-549.
- [17] IZZATILLAEV J, YUSUPOV Z. Short-Term Load Forecasting in Grid-Connected Microgrid [C] //2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), April 25-26, 2019, Istanbul, Turkey. IEEE, 2019: 71-75.
- [18] KONG W C, DONG Z Y, JIA Y W, et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network [J]. IEEE Transactions on Smart Grid, 2019, 10(1): 841-851.
- [19] GAO M M, LI J J, HONG F, et al. Day-Ahead Power Forecasting in a Large-Scale Photovoltaic Plant Based on Weather Classification Using LSTM [J]. Energy, 2019, 187: 115838.
- [20] TAN M, YUAN S P, LI S H, et al. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning [J]. IEEE Transactions on Power Systems, 2020, 35(4): 2937-2948.
- [21] SONG H Y, HUANG X G, LI D Y, et al. Short-Term Prediction Method of Distributed Photovoltaic Power Generation Based on LSTM Network [C] //2023 3rd International Conference on Intelligent Power and Systems (ICIPS), October 20-22, 2023, Shenzhen, China. IEEE, 2023: 348-353.
- [22] 谈玲, 康瑞星, 夏景明, 等. 融合多源异构气象数据的光伏功率预测模型 [J]. 电子与信息学报, 2024, 46(2): 503-517.
- [23] DENG Z F, WANG B B, XU Y L, et al. Multi-Scale Convolutional Neural Network with Time-Cognition for Multi-Step Short-Term Load Forecasting [J]. IEEE Access, 2019, 7: 88058-88071.
- [24] TZIOLIS G, LIVERA A, MONTES-ROMERO J, et al. Direct Short-Term Net Load Forecasting Based on Machine Learning Principles for Solar-Integrated Microgrids [J]. IEEE Access, 2023, 11: 102038-102049.
- [25] WANG Z Y, WANG Y R, SRINIVASAN R S. A Novel Ensemble Learning Approach to Support Building Energy Use Prediction [J]. Energy and Buildings, 2018, 159: 109-122.
- [26] RUBASINGHE O, ZHANG X N, CHAU T K, et al. A Novel Sequence to Sequence Data Modelling Based CNN-LSTM Algorithm for Three Years Ahead Monthly Peak Load Forecasting [J]. IEEE Transactions on Power Systems, 2024, 39(1): 1932-1947.
- [27] CHEN T Q, GUESTRIN C. XGBoost: A Scalable Tree Boosting System [C] //Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2016: 785-794.
- [28] 刘翠玲, 胡聪, 王鹏, 等. 基于营销大数据的电力客户多维度信用评价模型研究 [J]. 西南大学学报(自然科学版), 2022, 44(6): 198-208.
- [29] ZHEN Z, LIU J M, ZHANG Z Y, et al. Deep Learning Based Surface Irradiance Mapping Model for Solar PV Power Forecasting Using Sky Image [J]. IEEE Transactions on Industry Applications, 2020, 56(4): 3385-3396.
- [30] 宋毅, 张晗奕, 孙丰, 等. PpNet: 基于预先预测的降雨短时预测模型 [J]. 电子与信息学报, 2024, 46(2): 492-502.
- [31] HEWAGE P, BEHERA A, TROVATI M, et al. Temporal Convolutional Neural (TCN) Network for an Effective Weather Forecasting Using Time-Series Data from the Local Weather Station [J]. Soft Computing, 2020, 24(21): 16453-16482.
- [32] LIANG C, WU D, HE Y, et al. MMA: Multi-Metric-Autoencoder for Analyzing High-Dimensional and Incomplete Data [M] //Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023: 3-19.
- [33] WU D, SUN B, SHANG M S. Hyperparameter Learning for Deep Learning-Based Recommender Systems [J]. IEEE Transactions on Services Computing, 2023, 16(4): 2699-2712.