DOI:10.14188/j.1671-8844.2023.0414 文章编号:1671-8844(XXXX)XX-0001-08

基于组合相似日的变权多模型融合光伏预测方法

黄杨珏[1,2],贾梦麒[3],沈开程[1,2],朱远哲[1,2],汪进锋[1,2],张晓珂[3]

(1. 广东电网有限责任公司电力科学研究院,广东 广州 510080; 2. 广东省电力装备可靠性企业重点实验室,广东 广州 510080; 3. 武汉大学电气与自动化学院,湖北 武汉 430072)

摘要:针对光伏功率预测,首先改进传统的相似日理论,计算不同时刻对应的相似时刻数据构成组合相似日数据,以保障每一时段的相似性;然后构建分位数回归-长短期记忆网络预测模型,将相似日数据和多步功率数据分别作为输入,获取2组不同分位数下的光伏功率预测值;最后利用核密度估计,并根据相似日与待预测日的相似程度赋予2组输出不同的权重,得到最终的光伏预测功率值和概率预测区间。仿真案例表明,该预测方法的均方根误差为0.017 0,90%概率预测覆盖率为97.08%,平均带宽为0.087 8,与传统模型相比兼顾了预测的可靠性和敏锐性。

关键词:光伏发电;组合相似日;时变权重;概率预测

中图分类号:TM 615 **文献标志码:**A

The variable-weight multi-model fusion PV forecasting method based on combined similar days

HUANG Yangjue[1,2], JIA Mengqi[3], SHEN Kaicheng[1,2], ZHU Yuanzhe[1,2], WANG Jinfeng[1,2], ZHANG Xiaoke[3]

(1. Electric Power Scientific Research Institute ,Guangdong Power Grid Co. , Ltd. , Guangzhou 510080, China;
2. Guangdong Provincial Key Laboratory of Electric Power Equipment Reliability, Guangzhou 510080, China;
3. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China)

Abstract: For photovoltaic (PV) power prediction, firstly, the traditional similar day theory is improved by calculating the similar day data corresponding to different moments to constitute the combined similar day data to guarantee the similarity of each moment; secondly, a quantile regression-long short-term memory prediction model is constructed, and the similar day data and multi-step power data are used as inputs respectively to obtain the PV power prediction values under two different quartiles; finally, using kernel density estimation and assigning different weights to the two sets of outputs according to the degree of similarity between the similar day and the day to be predicted, the final PV predicted power values and probabilistic prediction intervals are obtained. The simulation case shows that the root mean square error of the proposed prediction method is 0.017 0, the 90% probability prediction coverage is 97.08%, and the average bandwidth is 0.087 8. Compared with the traditional model, the prediction reliability and sensitivity are taken into account.

Key words: photovoltaic power generation; combinatorial similarity day; time-varying weights; probabilistic prediction

收稿日期:2024-09-27

作者简介:黄杨珏(1988-),女,高级工程师,主要从事配网设备研究,E-mail: hyjourney@126.com。
通信作者:贾梦麒(1997-),女,博士研究生,主要从事微电网控制研究,E-mail: mengqijia@whu.edu.cn。

基金项目:南方电网公司科技项目(编号:GDKJXM20210194(036100KK52210078))。

本文引用格式:黄杨珏,贾梦麒,沈开程,等.基于组合相似日的变权多模型融合光伏预测方法[J].武汉大学学报(工学版),XXXX,XX(XX):1-8.

Citation:HUANG Yangjue, JIA Mengqi, SHEN Kaicheng, et al. The variable-weight multi-model fusion PV forecasting method based on combined similar days[J]. Engineering Journal of Wuhan University, XXXX, XX(XX):1-8.

随着“碳达峰、碳中和”目标的推进,近几年光伏发电的装机容量逐步增加[1-3]。然而,光伏发电功率与气象波动密切相关,呈现出波动性、间歇性和随机性[4-6],光伏发电大规模并网对电网的稳定运行造成巨大的冲击,精准的光伏功率预测有助于电力相关部门及时作出规划和调度,对电力系统安全稳定运行具有十分重要的意义。

光伏功率预测方法可以分为物理法、统计法、机器学习法和混合法[7]。人工神经网络自学习、泛化能力强,在光伏预测领域获得了广泛的应用[8-10]。文献[11]对多种深度神经网络(deep neural network,DNN)对未来多步提前的预测效果进行评估,结果表明基于DNN的预测模型提前8步的预测结果相关系数高于96.9%。为提升预测模型对光伏发电系统非线性响应的预测精度,文献[12]提出一种基于混合小波的长短期记忆(long short-term memory, LSTM)-DNN预测模型,该模型显著提高了预测准确性,同时减少了预测所需的数据量。然而,文献[11-12]中的预测模型都难以避免冗余数据的影响,存在模型训练速度慢、预测精度低等问题,与光伏功率输出直接相关的气象因素也没有被充分分析及利用。

为进一步提升光伏预测模型的预测精度和计算速度,相似日理论被引入光伏预测研究中[13-16]。针对光伏功率输出与气象因素密切相关的特点,可以选择与待预测日气象数据相似的日期,将相似日的气象数据和功率数据输入预测模型,以得到精准的预测数据。文献[13]通过加权模糊聚类选取与预测日天气类型相同的相似日样本作为模型训练集,利用结合纵横交叉算法与改进高斯过程回归算法进行光伏预测。为降低模型复杂性,并克服递归神经网络(recurrent neural network,RNN)的梯度消失问题,文献[14]采用k均值聚类法选取相似日数据作为训练样本实现光伏预测,该模型预测时间增加,但满足了预测准确性的要求。为提升所选取的相似日气象条件与待测日气象条件的吻合度,文献[15]提出一种基于熵权法的相似日光伏预测模型,该模型有效保证了相似日与待测日气象条件的相似性。然而,上述研究的预测输出为单点功率,也称点预测,难以反映光伏发电的概率范围和波动特征。

作为一种区间预测的机器学习方法,分位数回归(quantile regression,QR)-LSTM网络被用于水环境、交通、负荷等领域的预测。文献[17]通过QR-LSTM预测模型验证了时间序列生成对抗网络合成数据的有效性。文献[18]基于深度学习模型及分位数回归模型建立了行程时间概率预测模型,可为出行者和交通管理机构提供参考。文献[19]提出一种基于QR-LSTM模型的建筑能耗概率预测,以此设计了建筑节能潜力评价方案。为实现对光伏发电系统输出功率的概率预测,文献[20]采用模式发现方法区分不同类别的光伏系统运行条件,将其集成到高阶马尔可夫链中获取不同运行条件下光伏系统发电功率概率分布函数,最终预测结果遵循了不同天气条件下光伏发电的真实概率分布。文献[21]提出一种集成改进的卷积神经网络(convolutional neural network ,CNN)与非线性分位数回归的区域光伏发电概率预测方法,结果表明集成改进的CNN-QR模型更适合区域光伏发电的概率预测。文献[20-23]所提出的模型没有与相似日理论结合,在模型预测精度和计算速度方面仍有待改进。

传统相似日选取算法以天为单位,无法保障每一时段相似性,难以准确追踪光伏功率剧烈波动场景;为获得准确的光伏发电功率概率区间,大量概率预测的改进方法尚未考虑光伏发电预测时间的关联性对输出功率的影响;单一机器学习的预测模型虽然取得了良好的预测效果,然而由于其自身局限性,时序预测结果存在显著的滞后。

本文通过定量分析待预测时刻和相似时刻的相似程度,提出改进的相似日理论,克服了传统相似日理论无法保障每一时段相似性的弊端,且计算所得改进后的相似日数据可以有效提高预测输入的准确性。考虑到光伏数据的时间关联性对输出功率的影响,本文对2组QR-LSTM预测模型分别输入计算所得相似日光伏数据和利用滑动窗口法所得传统历史光伏数据,通过改进的相似日算法所得待预测日与相似日之间的相似程度,为2组输出分配相应的权重,可以融合得到精准光伏预测功率值和区间。

1 改进的相似日方法

为了保证数据的有效性,首先需要对数据进行相关性分析,然后结合改进的相似日理论计算获取相似日数据,以其作为后续预测模型的输入。

1.1 影响因素与相关性分析

光伏发电功率受多种气象因素的影响。在时间$t\in[1,T]$内,为了衡量某种单一光伏气象数据序

列$X=[X(1),X(2),\cdots,X(T)]$和对应的光伏功率序列$Y=[Y(1),Y(2),\cdots,Y(T)]$之间的非线性关系，可采用Kendall秩相关系数进行实际光伏数据的相关性分析，并在此基础上进行相似日计算。如果$(X(j)-X(i))(Y(j)-Y(i))=0$，$i=1,2,\cdots,T$，$j=1,2,\cdots,T$，$i\neq j$，则称2个数据对$\{X(i),Y(i)\}$、$\{X(j),Y(j)\}$为一个结，Kendall秩相关系数τ_b的计算公式为

$$\tau_b=\frac{\sum_i\sum_j \mathrm{sign}\left((X(j)-X(i))(Y(j)-Y(i))\right)}{\sqrt{\left[\frac{1}{2}T(T-1)-n_X\right]\left[\frac{1}{2}T(T-1)-n_Y\right]}} \tag{1}$$

式中：n_X、n_Y分别为由序列X、Y产生的结的个数。

相关系数是变量间相关程度的度量，取值范围为[−1.0,1.0]，正值表示正相关，负值表示负相关。其绝对值越接近1.0，表示2个变量之间关系越密切；越接近0，表示2个变量之间关系越不密切。相关系数对应的相关强度如表1所示。

表1 相关系数对应的相关强度

Table 1 The correlation coefficient corresponds to the correlation strength

$\lvert\tau_b\rvert$	相关强度
[0.8,1.0]	极强相关
[0.6,0.8)	强相关
[0.4,0.6)	中等程度相关
[0.2,0.4)	弱相关
[0.0,0.2)	极弱相关或不相关

在后续计算中，可以舍去极弱相关或不相关的数据，并对其余变量的相关系数进行归一化处理。某变量x归一化的相关性系数τ'_{bx}计算式如下：

$$\tau'_{bx}=\lvert\tau_{bx}\rvert\Big/\sum_y\lvert\tau_{by}\rvert \tag{2}$$

式中：τ_{bx}、τ_{by}分别为变量x、y的相关性系数。

1.2 相似日的选取

在历史光伏发电数据中搜索与预测日具有相似的气候特征的历史日，通过计算影响光伏发电的主要因素的相似程度，进而选取相似日。为了有效追踪剧烈波动场景，保障每一时段的相似性，对传统的相似日理论进行了改进，分时段地进行相似性分析。改进的相似日理论通过计算不同时刻的气象数据之间的带有权重因子的欧氏距离，来评估预测时刻与过去历史时刻光伏数据的相似度，本文通过计算待预测时刻t_1与不同时刻历史气象数据之间的欧式距离，选取最小欧式距离$d(t_1,t_2)$对应的时刻t_2作为t_1的相似时刻：

$$d(t_1,t_2)=\min_{t_2}\sqrt{\sum_{x\in k}\tau_x\left(w_x(t_1)-w_x(t_2)\right)^2} \tag{3}$$

式中：τ_x为气象特征x与光伏功率之间的归一化Kendall秩相关系数；$w_x(t_1)$和$w_x(t_2)$分别为时刻t_1和t_2的光伏气象特征x的归一化值；k为气象特征的集合。欧氏距离$d(t_1,t_2)$越小，相似度越高。

2 多模型融合预测方法

上文分析得到的相似日数据可以作为光伏功率预测模型的输入。

2.1 QR-LSTM预测模型

LSTM有效解决了训练过程中梯度消失和梯度爆炸问题，其计算式如下：

$$\begin{cases}\boldsymbol{f}(t)=\mathrm{sigmoid}(\boldsymbol{W}_{\mathrm{fh}}\boldsymbol{h}(t-1)+\boldsymbol{W}_{\mathrm{fx}}\boldsymbol{x}(t)+\boldsymbol{b}_{\mathrm{f}})\\ \boldsymbol{i}(t)=\mathrm{sigmoid}(\boldsymbol{W}_{\mathrm{ih}}\boldsymbol{h}(t-1)+\boldsymbol{W}_{\mathrm{ix}}\boldsymbol{x}(t)+\boldsymbol{b}_{\mathrm{i}})\\ \boldsymbol{c}'(t)=\tanh(\boldsymbol{W}_{\mathrm{ch}}\boldsymbol{h}(t-1)+\boldsymbol{W}_{\mathrm{cx}}\boldsymbol{x}(t)+\boldsymbol{b}_{\mathrm{c}})\\ \boldsymbol{o}(t)=\mathrm{sigmoid}(\boldsymbol{W}_{\mathrm{oh}}\boldsymbol{h}(t-1)+\boldsymbol{W}_{\mathrm{ox}}\boldsymbol{x}(t)+\boldsymbol{b}_{\mathrm{o}})\\ \boldsymbol{c}(t)=\boldsymbol{i}(t)\boldsymbol{c}'(t)+\boldsymbol{f}(t)\boldsymbol{c}(t-1)\\ \boldsymbol{h}(t)=\boldsymbol{o}(t)\tanh(\boldsymbol{c}(t))\end{cases} \tag{4}$$

式中：$\boldsymbol{x}(t)$为时刻t的输入；$\boldsymbol{f}(t)$、$\boldsymbol{i}(t)$、$\boldsymbol{o}(t)$、$\boldsymbol{c}(t)$分别为遗忘门、输入门、输出门、记忆环节时刻t的输出；$\boldsymbol{h}(t-1)$、$\boldsymbol{c}(t-1)$分别为时刻$t-1$的输出、单元状态；$\boldsymbol{c}'(t)$为更新的单元状态；$\boldsymbol{W}_{\mathrm{fh}}$、$\boldsymbol{W}_{\mathrm{fx}}$、$\boldsymbol{W}_{\mathrm{ih}}$、$\boldsymbol{W}_{\mathrm{ix}}$、$\boldsymbol{W}_{\mathrm{oh}}$、$\boldsymbol{W}_{\mathrm{ox}}$、$\boldsymbol{W}_{\mathrm{ch}}$、$\boldsymbol{W}_{\mathrm{cx}}$分别为对应环节的权重矩阵；$\boldsymbol{b}_{\mathrm{f}}$、$\boldsymbol{b}_{\mathrm{i}}$、$\boldsymbol{b}_{\mathrm{o}}$、$\boldsymbol{b}_{\mathrm{c}}$分别为对应的偏置矩阵；$\boldsymbol{h}(t)$为时刻$t$的输出，也是下一时刻的输入之一。

在分位数回归模型中，分位数τ对应的参数向量$\boldsymbol{\beta}(\tau)$可以由下式计算得到：

$$\min_{\boldsymbol{\beta}(\tau)}\sum_{Y\geqslant\boldsymbol{X}'\boldsymbol{\beta}(\tau)}\tau\left|Y-\boldsymbol{X}'\boldsymbol{\beta}(\tau)\right|+\sum_{Y<\boldsymbol{X}'\boldsymbol{\beta}(\tau)}(1-\tau)\left|Y-\boldsymbol{X}'\boldsymbol{\beta}(\tau)\right| \tag{5}$$

式中：$\boldsymbol{X}'$为输入变量的转置矩阵；Y为响应变量。

将分位数回归引入长短期记忆网络，分位数τ对应的模型的损失函数$L(\tau)$为

$$L(\tau)=\min_{\boldsymbol{W}(\tau),\boldsymbol{b}(\tau)}\sum_{t=1}^{T}\rho_\tau\left(Y_t-f\left(X_t,\boldsymbol{W}(\tau),\boldsymbol{b}(\tau)\right)\right) \tag{6}$$

式中：Y_t、X_t分别为t时刻的实际光伏功率、气象特征；$\boldsymbol{W}(\tau)$、$\boldsymbol{b}(\tau)$分别为分位数τ下长短期记忆网络相应门的权重矩阵、偏置矩阵，$\boldsymbol{W}(\tau)=(\boldsymbol{W}_{\mathrm{f}}(\tau),\boldsymbol{W}_{\mathrm{i}}(\tau),\boldsymbol{W}_{\mathrm{c}}(\tau),\boldsymbol{W}_{\mathrm{o}}(\tau))$，$\boldsymbol{b}(\tau)=(\boldsymbol{b}_{\mathrm{f}}(\tau),\boldsymbol{b}_{\mathrm{i}}(\tau),b_{\mathrm{c}}(\tau),$

$\boldsymbol{b}_0(\tau))$;$f(\,)$为τ分位数下的预测功率;ρ_τ为检验函数。

$$\rho_\tau(x)=\begin{cases}\tau x, & x\geqslant 0\\ (\tau-1)x, & x<0\end{cases} \tag{7}$$

通过最小化损失函数确定QR-LSTM模型在不同分位点下的权值和偏置,分位数τ在(0,1)区间等间距取值,并对不同分位数下的模型进行训练,得到不同分位数下的光伏预测值。

2.2 核密度估计

QR-LSTM预测模型可以获得光伏功率的条件分位数,但无法获得预测功率的概率密度函数。核密度估计是一种非参数估计方法,无需对先验分布进行假设便可得到随机变量的概率密度分布。因此,可以通过核密度估计得到光伏功率的概率密度函数$g(x)$:

$$g(x)=\frac{1}{Nh}\sum_{n=1}^{N}K\left(\frac{P_n-x}{h}\right) \tag{8}$$

式中:P_n为第n个功率样本;N为输出样本总数;h为带宽;$K(x)$为核函数。本文选择高斯核函数进行研究,假设存在随机变量$\alpha\in R$,那么关于α的高斯核函数计算式如下:

$$K(\alpha)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-\alpha^2}{2}\right) \tag{9}$$

这样就可以通过概率密度函数$g(x)$计算得到不同置信区间下的光伏功率的概率预测结果。

2.3 变权多模型融合预测方法

预测模型的输入可以是传统的历史光伏功率数据,也可以是相似日数据。考虑到历史光伏功率趋势对预测时刻的影响,采用滑动窗口法确定模型的输入和输出。滑动窗口法如图1所示。

2种不同输入的输出结果结合可以综合输入的方法特点,因此本文采用时变权重对二者进行融合。预测框架模型如图2所示。传统预测和改进的相似日预测方法在t_1时刻对应的权重$\omega_{\mathrm{tra}}(t_1)$、$\omega_{\mathrm{sim}}(t_1)$计算如下:

$$\omega_{\mathrm{tra}}(t_1)=d(t_1,t_2)\Big/\big(a+d(t_1,t_2)\big) \tag{10}$$

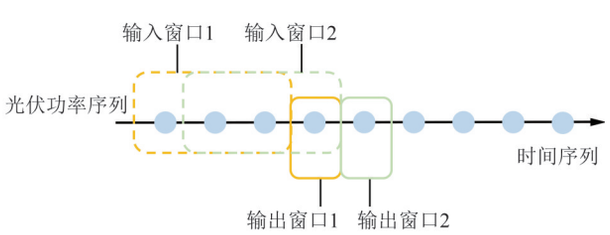


图1 滑动窗口法

Fig.1 Sliding window method

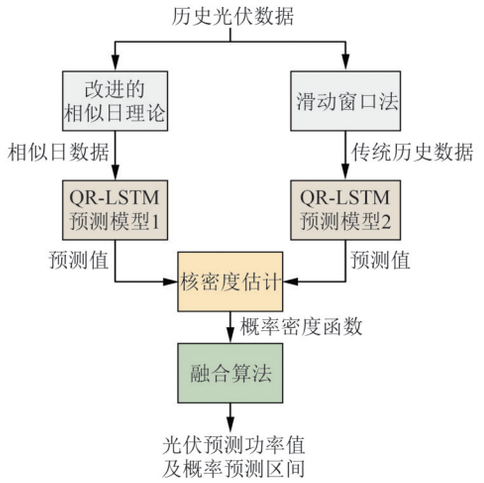


图2 预测模型框架

Fig.2 The frame diagram of the prediction model

$$\omega_{\mathrm{sim}}(t_1)=a\Big/\big(a+d(t_1,t_2)\big) \tag{11}$$

式中:a为常数。

因此,得到t_1时刻最终预测功率$P_{\mathrm{total}}(t_1)$为

$$P_{\mathrm{total}}(t_1)=\omega_{\mathrm{tra}}(t_1)P_{\mathrm{tra}}(t_1)+\omega_{\mathrm{sim}}(t_1)P_{\mathrm{sim}}(t_1) \tag{12}$$

式中:$P_{\mathrm{tra}}(t_1)$和$P_{\mathrm{sim}}(t_1)$分别为传统预测和改进的相似日预测方法在t_1时刻的预测功率。

2.4 评价指标

2.4.1 点预测

点预测的评价指标有2种,即平均均方根误差E_{RMSE}和平均绝对百分比误差E_{MAPE}:

$$E_{\mathrm{RMSE}}=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\big(P(t)-P_0(t)\big)^2} \tag{13}$$

$$E_{\mathrm{MAPE}}=\frac{1}{T}\sum_{t=1}^{T}\frac{\big|P(t)-P_0(t)\big|}{P(t)}\times 100\% \tag{14}$$

式中:$P(t)$、$P_0(t)$分别为时刻t的预测、实际功率。

2.4.2 概率区间预测

概率区间预测的评价指标通常包括可靠性指标和敏锐性指标。可靠性指标通常采用区间覆盖率I_{PICP}:

$$I_{\mathrm{PICP}}=\frac{1}{T}\sum_{t=1}^{T}S_t \tag{15}$$

$$S_t=\begin{cases}0, P_0(t)<P(t)^{\min}\text{或}P_0(t)>P(t)^{\max}\\ 1, P(t)^{\min}\leqslant P_0(t)\leqslant P(t)^{\max}\end{cases} \tag{16}$$

式中:$P(t)^{\max}$和$P(t)^{\min}$分别为时刻t概率预测区间的上下限。I_{PICP}越大,预测效果越好。

敏锐性指标通常采用区间平均宽度I_{PINAW}:

$$I_{\mathrm{PINAW}}=\sum_{t=1}^{T}\Big(P(t)^{\max}-P(t)^{\min}\Big)\Big/T \tag{17}$$

I_{PINAW}越小,预测敏锐性越高。

由 I_{PICP}、I_{PINAW} 可得综合指标 I_{WC}：

$$I_{WC}=\frac{I_{PICP}}{I_{PINAW}} \tag{18}$$

I_{WC} 值越大，概率区间的综合预测效果越好。

3　算例分析

以欧洲中期天气中心采集的武汉市光伏发电数据[24]为例，该数据包括时间及对应的光伏发电输出功率、历史气象及数值气象预报数据等，气象数据包括太阳辐照度、温度、湿度、气压、降水量等。

3.1　相似日计算

以2022年1月1日0:00—4月10日23:00、时间戳为1 h的光伏数据为例。依时间将数据按照9:1划分为训练集和测试集，分别计算对应的相似日数据。首先进行Kendall秩相关性分析，相关系数矩阵如图3所示。由图3可知，太阳辐照度与输出功率之间的相关系数最大，为0.842 1，二者极强相关；风向、气压、地表风速、降水量与输出功率之间极弱相关或不相关，在后续计算中舍去。对保留变量的权重进行归一化处理，结果如表2所示。

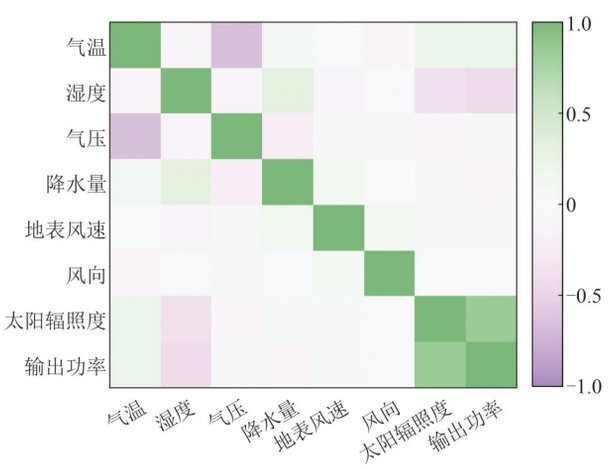


图3　光伏数据Kendall秩相关系数矩阵

Fig.3　Results of Kendall rank analysis for PV data

表2　保留变量及其归一化权重

Table 2　PV data and the normalised weights

保留变量	归一化权重
太阳辐照度	0.572 3
湿度	0.284 7
气温	0.143 0

以4月3日光伏数据为例，改进相似日和传统相似日算法的计算结果如图4所示。显然，改进相似日算法可以更有效跟踪光伏波动场景，其平均绝对百分比误差为0.003 4，而传统相似日算法平均绝对百分比误差为0.070 2。

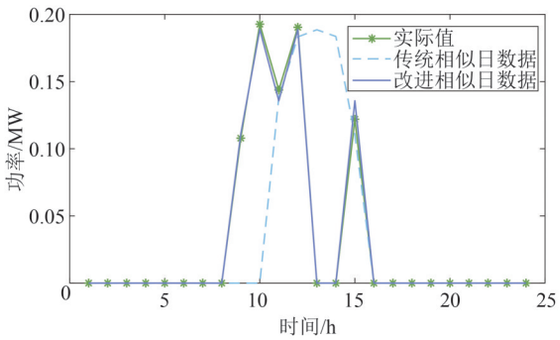


图4　2022年4月3日的相似日结果对比

Fig.4　Comparison of similar day data on April 3, 2022

3.2　预测效果及分析

基于前文分析，首先确定2组预测模型的输入数据集，即相似日数据和基于滑动窗口法所得传统历史光伏功率数据；其次分位数在区间(0,1)均匀取值，基于2组训练集和测试集数据对不同分位数下的QR-LSTM模型进行训练并预测。QR-LSTM模型网络隐藏层为数2，LSTM层节点数分别为50和20，进行30轮训练。初始学习率为0.001，在2轮训练后通过乘以因子0.5以降低学习率。滑动窗口窗宽为5，常数 a 取为0.3。

对不同分位数下的预测结果进行核密度估计得到的概率密度函数如图5所示。由图5可知，光伏功率的概率分布不平衡，表现出明显的偏态，而非标准的正态分布。另外，实际功率点和概率密度曲线峰值点对应的功率之间的平均均方根误差为0.017 0，没有表现出较大偏离。

本文将概率密度曲线峰值点对应的功率作为光伏功率的点预测值，基于此进行概率区间分析，

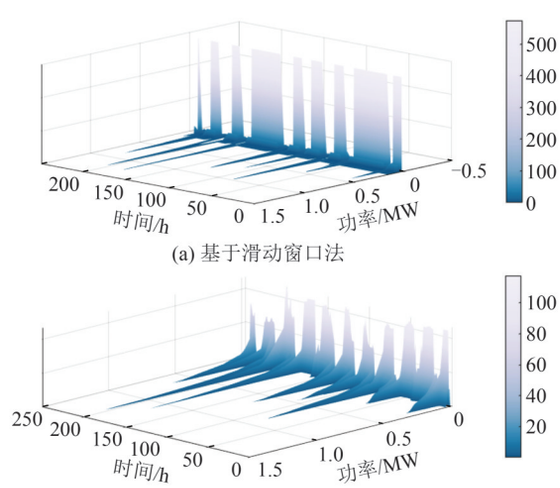


图5　光伏功率的概率分布

Fig.5　Probability distribution of PV power

即对2组概率密度函数进行积分并根据变权算法融合2组结果,多模型融合预测结果如图6所示。

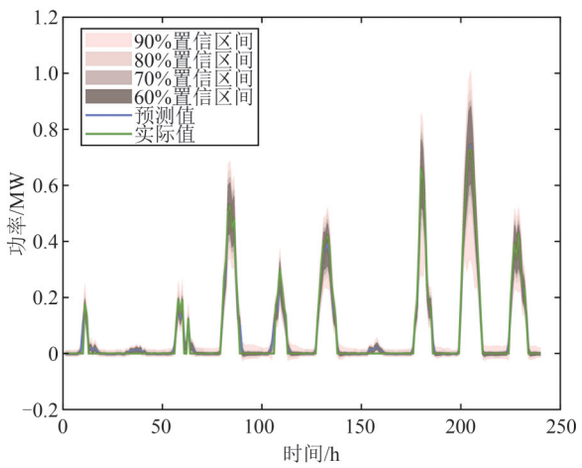


图6 多模型融合预测结果

Fig.6 Probability distribution of PV power

将本文所提融合模型与基于传统相似日的融合模型、基于组合相似日的单一预测模型、基于组合相似日的单一预测模型及QR-DNN、QR-CNN、QR-GRU(gated recurrent unit)模型的预测结果进行对比,各模型预测效果的评价指标如表3所示。其中:预测模型QR-DNN全连接层a节点个数为50,b节点个数为20;预测模型QR-CNN激活函数为Tanh(),卷积核为[2,1],激活函数为RELU函数;预测模型QR-GRU的GRU层a节点个数为50,b节点个数为20。其他参数与基于组合相似日的融合预测模型一致。

由表3可知,本文提出的基于组合相似日的融合预测模型的点预测误差E_{RMSE}和E_{MAPE}比对照模型都低;90%置信区间下覆盖率准确率比对照模型分别提高了2.08%、6.66%、3.75%、1.66%、2.08%、1.66%和4.58%,平均带宽比对照模型都小,综合评价指标值最高,具有最优的可靠性和敏锐性指标。因此,本文所提预测方法兼顾了预测的准确性和概率区间的清晰度,可为决策者提供有效参考。

表3 预测效果比较

Table 3 Comparison of prediction effect

预测模型	点预测		90%置信区间概率预测		
	E_{RMSE}	E_{RAPE}/%	I_{PICP}/%	I_{PINAW}	I_{WC}
基于组合相似日的融合预测模型	0.017 0	0.88	97.08	0.087 8	11.057 3
基于传统相似日的融合预测模型	0.049 2	2.23	95.00	0.129 1	7.358 6
基于组合相似日的单一预测模型	0.020 3	1.10	90.42	0.088 3	10.235 3
基于传统数据的单一预测模型	0.062 4	3.00	93.33	0.121 0	7.713 5
QR-LSTM	0.048 3	2.43	95.42	0.170 1	5.610 5
QR-DNN	0.066 4	3.38	95.00	0.212 2	4.477 2
QR-CNN	0.063 7	3.07	95.42	0.164 4	5.804 0
QR-GRU	0.051 7	2.76	92.50	0.135 6	6.819 3

4 结论

本文首先改进了传统的相似日理论,然后将改进的相似日数据和滑动窗口法所得数据输入QR-LSTM光伏功率预测模型,并对输出进行核密度分析,根据选取相似日得到的欧氏距离赋予2组输出不同的权重,得到最终的光伏功率预测值和概率预测区间。验证结果表明:

1)改进的相似日理论通过Kendall秩相关性分析选取太阳辐照度、湿度、气温作为计算依据,并计算相似时刻的光伏数据构成组合相似日数据,保障了每一时段的相似性。

2)通过时变权重算法融合改进相似日算法和滑动窗口法所得数据的光伏功率预测的均方根误差为0.017 0,90%概率预测覆盖率为97.08%,平均带宽为0.087 8,与传统的单一模型相比,具有更高的准确性和概率区间清晰度。

本文为多模型驱动实现光伏功率概率预测提供了一种有效的思路,可以为新能源的调度决策提供更有价值的决策信息。然而本研究仍存在一些开放性的问题亟待解决,例如:如何增强基于机器学习的预测方法的可解释性、如何设计寻优算法来实现多模型融合算法中相关参数的最优取值等,这是下一步的研究方向。

参考文献:

[1] 丁明,王伟胜,王秀丽,等.大规模光伏发电对电力系统影响综述[J].中国电机工程学报,2014,34(1):1-14.
DING Ming, WANG Weisheng, WANG Xiuli, et al. A review on the effect of large-scale PV generation on power systems[J]. Proceedings of the CSEE, 2014, 34(1): 1-14.

[2] YAO T C, WANG J, WU H Y, et al. Intra-hour photovoltaic generation forecasting based on multi-source data and deep learning methods[J]. IEEE Transactions on Sustainable Energy, 2022, 13(1): 607-618.

[3] 王函.风光发电功率与用电负荷联合预测方法研究[D].北京:华北电力大学,2021.
WANG Han. Research on integrated forecasting

method for wind power, solar power, and system load [D]. Beijing: North China Electric Power University, 2021.

[4] SANJARI M J, GOOI H B, NAIR N C. Power generation forecast of hybrid PV-wind system[J]. IEEE Transactions on Sustainable Energy, 2020, 11(2): 703-712.

[5] 李建林, 方知进, 谭宇良, 等. 电化学储能系统在整县制屋顶光伏中应用前景分析[J]. 太阳能学报, 2022, 43(4): 1-12.
LI Jianlin, FANG Zhijin, TAN Yuliang, et al. Application prospect analysis of electrochemical energy storage technology in county-wide rooftop photovoltaic system[J]. Acta Energiae Solaris Sinica, 2022, 43(4): 1-12.

[6] 吴鹏飞, 刘博, 杨鹏程, 等. 考虑天气和运行状态的光伏一次调频控制及评估[J]. 武汉大学学报(工学版), 2022, 55(3): 275-282.
WU Pengfei, LIU Bo, YANG Pengcheng, et al. PV primary frequency regulation and evaluation considering weather characteristics and operating conditions[J]. Engineering Journal of Wuhan University, 2022, 55(3): 275-282.

[7] 戚晓侠. 基于深度学习的光伏/风电功率预测研究[D]. 哈尔滨: 哈尔滨工程大学, 2020.
QI Xiaoxia. Research on photovoltaic/wind power forecasting based on deep learning[D]. Harbin: Harbin Engineering University, 2020.

[8] 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.

[9] 张展耀. 基于深度学习理论的光伏功率短期预测研究[D]. 北京: 华北电力大学, 2019.
ZHANG Zhanyao. The short-term solar power forecasting research based on deep learning theory[D]. Beijing: North China Electric Power University, 2019.

[10] ESEYE A T, ZHANG J H, ZHENG D H. Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information[J]. Renewable Energy, 2018, 118: 357-367.

[11] MELLIT A, PAVAN A M, LUGHI V. Deep learning neural networks for short-term photovoltaic power forecasting[J]. Renewable Energy, 2021, 172 (C): 276-288.

[12] OSPINA J, NEWAZ A, FARUQUE M O. Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model[J]. IET Renewable Power Generation, 2019, 13(7): 1087-1095.

[13] SANJARI M J, GOOI H B. Probabilistic forecast of PV power generation based on higher order Markov chain[J]. IEEE Transactions on Power Systems, 2017, 32(4): 2942-2952.

[14] YU Y X, WANG M X, YAN F Q, et al. Improved convolutional neural network-based quantile regression for regional photovoltaic generation probabilistic forecast[J]. IET Renewable Power Generation, 2020, 14(14): 2712-2719.

[15] 孟安波, 陈嘉铭, 黎湛联, 等. 基于相似日理论和CSO-WGPR的短期光伏发电功率预测[J]. 高电压技术, 2021, 47(4): 1176-1184.
MENG Anbo, CHEN Jiaming, LI Zhanlian, et al. Short-term photovoltaic power generation prediction based on similar day theory and CSO-WGPR[J]. High Voltage Engineering, 2021, 47(4): 1176-1184.

[16] 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.

[17] WENG P Y, TIAN Y, LIU Y F, et al. Time-series generative adversarial networks for flood forecasting[J]. Journal of Hydrology, 2023, 622: 129702.

[18] 张航, 张竞宇, 吕能超, 等. 基于可靠度的公路隧道进出口段行车安全分析[J]. 公路, 2023, 68(4): 238-245.
ZHANG Hang, ZHANG Jingyu, LÜ Nengchao, et al. Analysis on driving safety at entrance and exit of the highway tunnel based on reliability[J]. Highway, 2023, 68(4): 238-245.

[19] 李昱瑾. 分位数回归在办公建筑节能潜力评估上的应用研究[D]. 武汉: 华中科技大学, 2021.
LI Yujin. Office building energy efficiency benchmarking method based on quantile regression[D]. Wuhan: Huazhong University of Science and Technology, 2021.

[20] ZHANG N, REN Q, LIU G C, et al. Short-term PV output power forecasting based on CEEMDAN-AE-GRU[J]. Journal of Electrical Engineering & Technology, 2022, 17(2): 1183-1194.

[21] 葛乐, 陆文伟, 袁晓冬, 等. 基于改进相似日和ABC-SVM的光伏电站功率预测[J]. 太阳能学报, 2018, 39 (3): 775-782.
GE Le, LU Wenwei, YUAN Xiaodong, et al. Power

forecasting of photovoltaic plant based on improved similar day and ABC-SVM[J]. Acta Energiae Solaris Sinica, 2018, 39(3): 775-782.

[22] 杨皓然, 杨茂, 苏欣. 计及预测误差时空-条件相依特性的日内光伏出力区间预测方法[J]. 南方电网技术, 2023, 17(2): 128-136.
YANG Haoran, YANG Mao, SU Xin. Intraday photovoltaic output interval prediction method considering the spatiotemporal-conditional dependence of prediction error[J]. Southern Power System Technology, 2023, 17(2): 128-136.

[23] 万灿, 宋永华. 新能源电力系统概率预测理论与方法及其应用[J]. 电力系统自动化, 2021, 45(1): 2-16.
WAN Can, SONG Yonghua. Theories, methodologies and applications of probabilistic forecasting for power systems with renewable energy sources[J]. Automation of Electric Power Systems, 2021, 45(1): 2-16.

[24] 羲和能源气象大数据平台[N/OL].https://www.xihe-energy.com.
Xihe energy meteorological big data platform[N/OL]. https://www.xihe-energy.com.