

王 俊, 邱 爽, 鞠丹阳, 等. 基于改进 INFO-CNN-QRGRU 模型的农村分布式光伏发电短期概率预测[J]. 沈阳农业大学学报, 2024, 55(4): 490-502.

WANG Jun, QIU Shuang, JU Danyang, et al. Short-term probabilistic prediction of rural distributed photovoltaic power generation based on improved INFO-CNN-QRGRU model[J]. Journal of Shenyang Agricultural University, 2024, 55(4): 490-502.

基于改进 INFO-CNN-QRGRU 模型的农村分布式光伏发电短期概率预测

王 俊¹, 邱 爽¹, 鞠丹阳², 谢易澎³, 张楠楠¹, 王 慧¹

(1. 沈阳农业大学 信息与电气工程学院, 沈阳 110161; 2. 国网内蒙古东部电力有限公司 呼伦贝尔供电公司, 内蒙古 呼伦贝尔 021000; 3. 国网辽宁省电力有限公司 辽阳供电公司, 辽宁 辽阳 111000)

摘 要: 随着“双碳”目标的推进, 清洁能源所占比重大幅度增加, 分布式光伏发电在我国农村地区快速发展, 但其随机性、间歇性的特点给新能源消纳和电网稳定带来很大的挑战。光伏发电预测可以在一定程度上改善新能源消纳问题, 减少光伏发电的不稳定性对电网的冲击。因此, 为提高光伏发电功率预测精度, 提出一种基于改进向量加权平均算法优化 CNN-QRGRU 网络的光伏发电概率预测方法。首先采用 ReliefF 算法对特征变量进行选择, 在此基础上利用高斯混合模型(Gaussian mixture model, GMM)聚类方法将天气分为晴天、晴转多云和阴雨天 3 种类型, 将处理好的数据输入到 CNN-GRU 模型中, 并利用向量加权平均(weighted mean of vectors algorithm, INFO)优化算法对模型超参数进行调参, 将分位数回归模型(quantile regression, QR)与 INFO-CNN-GRU 模型相结合得到光伏功率条件分布, 结合核密度估计法从条件分布中获得概率密度函数, 完成概率预测。以实际光伏电站数据作为基础, 将提出的 INFO 优化算法与其他几种传统的优化算法进行对比, 结果表明 INFO 的优化效果更好, 在此基础上进行概率预测, 得到的概率预测结果相较于点预测能提供更多有效信息, 更具有应用价值。

关键词: 光伏出力; 高斯混合模型聚类; 门控循环单元; 向量加权平均算法; 分位数回归; 概率预测

中图分类号: TM615

文章编号: 1000-1700(2024)04-0490-13

文献标识码: A

开放科学(资源服务)标识码(OSID):



Short-term Probabilistic Prediction of Rural Distributed Photovoltaic Power Generation Based on Improved INFO-CNN-QRGRU Model

WANG Jun¹, QIU Shuang¹, JU Danyang², XIE Yipeng³, ZHANG Nannan¹, WANG Hui¹

(1. College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang 110161, China;

2. Hulunbeier Power Supply Company of State Grid Inner Mongolia East Power Co., Ltd., Hulunbeier Inner Mongolia 021000, China;

3. Liaoyang Power Supply Company of State Grid Liaoning Province Electric Power Co., Ltd., Liaoyang Liaoning 111000, China)

Abstract: With the advancement of the "dual carbon" goal, the proportion of clean energy has increased significantly, and distributed photovoltaic power generation has developed rapidly in rural areas of China, but its randomness and intermittent characteristics have brought great challenges to the consumption of new energy and the stability of the power grid. Photovoltaic power generation prediction can improve the new energy consumption and reduce the impact of its instability on the power grid. Therefore, in order to improve the accuracy of photovoltaic power generation power prediction, this paper proposes a photovoltaic power generation probability prediction method based on the improved vector weighted

收稿日期: 2024-02-10

基金项目: 国家自然科学基金项目(61903264); 辽宁省科学研究经费项目(LJKZ0681)

第一作者: 王 俊(1979-), 男, 博士, 教授, 博士生导师, 从事智能电网、可再生能源电力并网运行与控制研究, E-mail: wddream2016@syau.edu.cn

average algorithm to optimize the CNN-QRGRU network. First, the ReliefF algorithm is used to select the feature variables, based on which the Gaussian mixture model (GMM) clustering method is used to classify the weather into three types: sunny, sunny to cloudy and rainy, the processed data are input into the CNN-GRU model, and the INFO optimization algorithm is used to tune the model hyperparameters, and the quantile regression (QR) model is combined with the INFO-CNN-GRU model to obtain the PV power conditional distribution, and the kernel density estimation method is combined to obtain the probability density function from the conditional distribution to complete the probability prediction. Using the actual PV plant data as the research basis, the optimization effect of the newly proposed INFO optimization algorithm is compared with several other traditional optimization algorithms, and the results show that the optimization effect of INFO is better, and the probability prediction results obtained on this basis can provide more effective information compared with the point prediction, which is more valuable for application.

Key words: photovoltaic power output; Gaussian mixture model clustering; gated recurrent unit; weighted mean of vectors algorithm; quantile regression; probabilistic forecast

近年来我国农村地区光伏建设迅速发展,光伏的并入有效缓解了环境污染、农村低电压等问题,因此光伏发电已成为新能源发电的重要选择。但光伏发电的不稳定和间歇性对电压上升、保护安排和电能质量等造成不利影响^[1],需要通过对光伏发电的预测结果选择相应的调控措施,因此光伏发电预测技术至关重要^[2-3]。目前,随机森林、支持向量回归、极限学习机、人工神经网络等方法已广泛应用于光伏发电预测,其中典型的深度学习网络^[4]包括卷积神经网络、深度置信网络和循环神经网络等^[5-6]。循环神经网络(recurrent neural network, RNN)相较于前馈神经网络具有更强大的处理时间序列的能力,但同时也会出现梯度消失等现象^[7-8]。针对RNN在实际应用中存在的问题提出长短期记忆神经网络(long short-term memory, LSTM),它对RNN中的循环单元进行改进从而解决了梯度消失问题^[9]。由于LSTM结构由3个不同的门构成,导致参数较多,训练起来比较困难,所以在LSTM的基础上又进一步提出门控循环单元(gate recurrent unit, GRU),GRU只含有两个门控结构,它可以在光伏发电预测中探索必要的长期信息,缩短训练时间^[10-11]。现有研究针对卷积神经网络与循环神经网络的组合使用已经取得了一些成果,王开艳等^[12]将CNN与双向长短期神经网络(bidirectional long short-term memory, BiLSTM)相结合进行光伏发电预测,BiLSTM网络通过前后传播提高了模型的预测精度,同时训练时间也随之增加。由于模型参数的选择是决定模型预测效果的关键因素,且神经网络中超参数的选取非常困难,手动调参费时且不够精确,会出现过拟合或者欠拟合等现象,此时通过选择合适的优化算法对网络超参数进行优化是非常有必要的^[13-15]。然而多数文献研究仍停留在传统优化算法上,对于新提出的优化算法并未进行对比和研究。邹智等^[16]利用贝叶斯优化算法对CNN-BiGRU网络超参数进行优化,并与网格搜索算法、粒子群优化算法进行对比,得出贝叶斯优化效果更好。吴勇等^[17]通过布谷鸟搜索(cuckoo search, CS)算法优化变分模态分解(variational mode decomp, VMD)的主要参数,把光伏功率分解成若干趋于稳定的模态分量,随后将分解量送入联合卷积神经网络(convolutional neural network, CNN)与LSTM的组合预测模型进行逐一预测并将功率预测值进行叠加评估。以上方法均局限于对光伏发电点预测的研究,点预测只能得知某时刻的功率值,概率预测可以描述光伏发电功率变化趋势、波动范围、不确定性及面临的风险,所以更有研究前景^[18-19]。董雷等^[20]提出了应用动态贝叶斯网络(dynamic bayesian network, DBN)理论,建立光伏发电预测的DBN模型,该模型考虑影响光伏发电量的多种因素,及各因素之间的相互联系,以此进行更有效的预测,得到预测时刻短期光伏发电量的概率分布。在WU等^[21]的研究中,基于QR和ELM的直接分位数回归(DQR)模型被描述为具有较高计算效率的线性规划问题。许彪等^[22]提出了一种基于藤copula分位数回归的光伏功率日前概率预测模型,利用藤copula对光伏功率及其条件变量间的相依结构进行解析化表达,在此基础上建立起光伏功率条件分位数回归模型。

基于以上讨论,本研究在采用高斯混合聚类(GMM)方法进行相似日聚类的前提下,将相似日聚类结果和历史光伏功率数据输入到CNN-QRGRU网络中,并运用最新提出的向量加权平均算法(INFO)优化CNN-QRGRU区间预测模型,基于核密度估计方法生成光伏功率概率预测结果,最后通过与

其他优化算法优化的CNN-QRGRU模型的预测精度进行比较,证明了本研究所提出的模型对光伏功率概率预测具有有效性和创造性。

1 数据与研究方法

1.1 ReliefF算法

本研究数据来自于2017-2019年澳大利亚沙漠知识太阳能中心(DKASC)Alice Springs 站点某光伏阵列,每间隔5 min记录1次光伏功率,考虑到光照时间的因素,只保留每天6:00-19:30发电时间段的数据,以此数据对待测日进行提前24 h光伏功率预测。选择ReliefF算法进行特征选择,从而降低样本特征维度,以提高预测精度。通过计算数据不同特征与光伏功率的相关性,从而得出每种特征的权重,权重越大,则对分类贡献度越高,如果权重小于某个阈值时将会被剔除。图1为根据ReliefF算法对相关因素进行特征选择后的权重比例图,可知不同天气因素的权重大小,选择前6个气象影响因素作为后续光伏功率预测模型的输入数据。

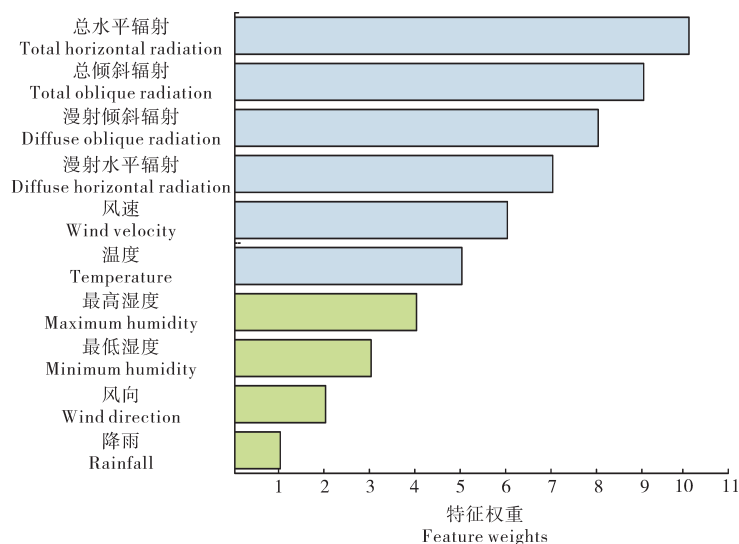


图1 特征权重比例

Figure 1 Feature weight scale

由光伏组件材料特性可知,光照强度越高,光伏电池板中的太阳能电子材料就能够吸收更多的光能,从而产生更多的电子流,光伏电池板的电能转换效率就越高。由图1可知,在众多相关因素中总水平辐射与光伏功率具有极强相关性,漫射水平辐射、总倾斜辐射以及漫射倾斜辐射与光伏功率均存在较强相关性,所以本研究选取光照强度作为光伏预测的主要依据。

1.2 高斯混合聚类算法

由于天气类型多种多样,不区分天气类型直接进行预测会大大降低预测精度,通过将天气类型划分使模型的训练更具有针对性,提高预测的效率和准确度,所以本研究选择高斯混合模型进行相似日聚类。GMM算法得到的是其属于各个类的概率,求解高斯混合模型参数一般采用期望极大算法(expectation maximization algorithm, EM),其步骤如下。

(1)初始化 k 个多元高斯分布的参数,即均值 μ_i 和方差 ε_i ;

(2)计算所有数据点 $x_i (i = 1, 2, \dots, m)$ 的响应度,即每个高斯分布的概率 $\gamma_{i,j}$;

$$\gamma_{i,j} = p(x_i | z_i = j) = \frac{1}{(2\pi)^{\frac{d}{2}} |\varepsilon_j|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (x_i - \mu_j)^T \varepsilon_j^{-1} (x_i - \mu_j) \right] \quad (1)$$

式中: $p(x_i | z_i = j)$ 为概率函数; z_i 为 x_i 所属的类; d 为 x_i 的维度;

(3)更新每个高斯分布的参数;

$$\mu'_j = \frac{\sum_{i=1}^m \gamma_{i,j} x_i}{\sum_{i=1}^m \gamma_{i,j}} \quad (2)$$

$$\varepsilon'_j = \frac{\sum_{i=1}^m \gamma_{i,j} (x_i - \mu'_j)(x_i - \mu'_j)^T}{\sum_{i=1}^m \gamma_{i,j}} \quad (3)$$

(4)若高斯参数不收敛,则重复(2)和(3)步操作直至收敛;

(5)用最终得到的高斯参数遍历所有样本,将样本归于概率 $\gamma_{i,j}$ 最大的一类。

由图1可知,光伏功率与总水平辐射相关性最强,所以将总辐射水平变量的标准值、平均值、最大值、峰波谷数、变异系数作为GMM聚类算法的聚类特征。

1.3 CNN-QRGRU模型

在GRU模型前加入CNN网络进行特征提取,利用其卷积层和池化层对原始数据进行更抽象和更高层次的处理来获取内部特征。GRU是在LSTM的基础上提出的,其结构更简单,参数更少,计算速度更快,GRU计算当前时刻状态 h_t 的公式如下。

$$\begin{cases} Z_t = \sigma(W_z \circ [h_{t-1}, x_t] + b_z) \\ r_t = \sigma(W_r \circ [h_{t-1}, h_t] + b_r) \\ \tilde{h}_t = \tanh(W_h \circ [r_t h_{t-1}, x_t] + b_h) \\ h_t = (1 - Z_t)h_{t-1} + Z_t \tilde{h}_t \end{cases} \quad (4)$$

式中: Z_t 为更新门; r_t 为复位门; W_z, b_z 为更新门的参数矩阵; W_r, b_r 为重置门的参数矩阵; W_h, b_h 为计算候选状态过程的参数矩阵;将GRU网络与QR算法相结合对光伏功率进行区间预测,QR回归模型公式如下。

$$Q_{P_i}(\tau|x_i) = \beta(\tau) x_i, i = 1, 2, \dots, n \quad (5)$$

式中: τ 的取值范围为(0,1), $Q_{P_i}(\tau|x_i)$ 是因变量 P_i 的第 τ 个条件分位数; x_i 是第 i 个自变量; $\beta(\tau)$ 是回归系数的向量。

本研究构建的QRGRU模型实质是将求解回归系数向量问题转化为求解QRGRU最小损失函数问题,最小损失函数如下。

$$L = \sum_{i=1}^n \gamma_\tau (P_i - x_i \beta(\tau)) \quad (6)$$

$$\gamma_\tau(s) = \begin{cases} \tau s, s \geq 0 \\ (\tau - 1)s, s < 0 \end{cases} \quad (7)$$

式中: γ_τ 为不对称函数。

最终,当 τ 在0~1间连续变化时,可通过线性QR模型估算第 τ 个条件分位数。

$$\hat{Q}_{P_i}(\tau|x_i) = \hat{\beta}(\tau) x_i, i = 1, 2, \dots, n \quad (8)$$

式中: $\hat{\beta}(\tau)$ 为 $\beta(\tau)$ 的估计值; $\hat{Q}_{P_i}(\tau|x_i)$ 为 $Q_{P_i}(\tau|x_i)$ 的估计值,实际上等价于未来光伏功率的条件分布。

概率预测结果利用核密度估计法从条件分布中获得,其中核函数的不同选择会影响概率密度分布结果。 N 个分位数在0~1之间等距分布,即 $\tau = [\tau_1, \tau_2, \dots, \tau_N]$,在每个分位数 τ_j 处, $\hat{Q}_{P_i}(\tau_j|x_i)$ 可以通过QR的融合模型来获取,概率密度分布公式如下。

$$f(x) = \frac{1}{N} \sum_{j=1}^N K\left(\frac{P_i - \hat{Q}_{P_i}(\tau_j|x_i)}{H}\right) \quad (9)$$

式中: N 为样本总数; H 为带宽,且 $H > 0$,可以由经验法选择,如式(10),其中 m 为样本数; K 为核函数,本研究使用高斯核函数,高斯核函数的计算如式(11)。

$$H \approx 1.06 \hat{\sigma}_m - 1/5 \quad (10)$$

$$K(\alpha) = (1/\sqrt{2\pi}) \cdot \exp(-\alpha^2/2) \quad (11)$$

1.4 改进向量加权平均优化算法

向量加权平均算法是于2022年新提出的一种基于群体的新型智能优化算法,该算法通过改进加权平均方法和更新向量的位置来帮助形成更强的鲁棒结构。其中,规则更新、向量合并和局部搜索是INFO的3个核心过程。下面介绍基于Circle映射改进的向量加权平均算法的优化步骤。

(1)种群个体在搜索空间内初始化。鉴于INFO算法在搜索空间内随机初始化,本研究采用Circle混沌映射初始化种群来改善初始解容易出现聚集等问题。Circle混沌映射在 $[0.2, 0.6]$ 之间取值较为密集,其公式如下。

$$x_{i+1} = \text{mod}\left(3.85x_i + 0.4 - \frac{0.7}{3.85\pi} \sin(3.85\pi x_i), 1\right) \quad (12)$$

式中: i 为解的维度。

(2)更新规则。INFO使用基于均值的规则更新向量的位置,这是从一组随机向量的加权均值中提取的。另外,为了提高全局搜索能力,在更新规则算子中加入了收敛加速部分(CA),更新规则的主要公式如下。

$$Z1_l^g = \begin{cases} x_{a1}^g + \sigma \times \text{MeanRules} + \text{randn} \times \frac{x_{bs} - x_{a1}^g}{f(x_{bs}) - f(x_{a1}^g) + 1}, \text{rand} < 0.5 \\ x_{bs} + \sigma \times \text{MeanRules} + \text{randn} \times \frac{x_{a2}^g - x_{a3}^g}{f(x_{a2}^g) - f(x_{a3}^g) + 1}, \text{rand} \geq 0.5 \end{cases} \quad (13)$$

$$Z2_l^g = \begin{cases} x_{bs} + \sigma \times \text{MeanRules} + \text{randn} \times \frac{x_{a1}^g - x_b^g}{f(x_{a1}^g) - f(x_{a2}^g) + 1}, \text{rand} < 0.5 \\ x_{bt} + \sigma \times \text{MeanRules} + \text{randn} \times \frac{x_{a1}^g - x_{a2}^g}{f(x_{a1}^g) - f(x_{a2}^g) + 1}, \text{rand} \geq 0.5 \end{cases} \quad (14)$$

$$\sigma = 2\alpha \times \text{rand} - \alpha \quad (15)$$

式中: $Z1_l^g, Z2_l^g$ 为第 g 次迭代的新位置向量; σ 为向量缩放率,通过式(15)计算所得, $a1 \neq a2 \neq a3 \neq 1$ 是从中随机选择的不同整数, randn 是一个标准正态分布随机值。

$$\text{MeanRule} = r \times \text{WM1}_l^g + (1 - r) \times \text{WM2}_l^g, l = 1, 2, \dots, NP \quad (16)$$

式(16)为MeanRule定义,其中 r 是 $[0, 0.5]$ 之间的随机数, $\text{WM1}_l^g, \text{WM2}_l^g$ 定义如式(17)和式(18)。

$$\text{WM1}_l^g = \delta \times \frac{w_1(x_{a1} - x_{a2}) + w_2(x_{a1} - x_{a3}) + w_3(x_{a2} - x_{a3})}{w_1 + w_2 + w_3 + \varepsilon} + \varepsilon \times \text{rand}, l = 1, 2, \dots, NP \quad (17)$$

$$\text{WM2}_l^g = \delta \times \frac{w_1(x_{bs} - x_{bt}) + w_2(x_{bs} - x_{ws}) + w_3(x_{bt} - x_{ws})}{w_1 + w_2 + w_3 + \varepsilon} + \varepsilon \times \text{rand}, l = 1, 2, \dots, NP \quad (18)$$

式中: w_1, w_2, w_3 为3个加权函数,用于计算向量的加权平均值,有利于算法在解空间中全局搜索; x_{bs}, x_{bt}, x_{ws} 为第 g 代种群中最优、次优和最差的解向量,这些解向量是在每次迭代时对种群向量进行排序后确定的。

(3)向量合并阶段。根据式INFO将前一阶段计算的两个向量($Z1_l^g, Z2_l^g$)与条件 $\text{rand} < 0.5$ 的向量(u_l^g)相结合,生成新向量。该算子用于提升局部搜索能力,以提供一个更好的向量。

$$u_l^g = \begin{cases} Z1_l^g + \mu |Z1_l^g - Z2_l^g|, \text{rand1} < 0.5 \text{rand2} < 0.5 \\ Z2_l^g + \mu |Z1_l^g - Z2_l^g|, \text{rand1} < 0.5 \text{rand2} \geq 0.5 \\ x_l^g, \text{rand1} < 0.5 \end{cases} \quad (19)$$

式中: u_l^g 为第 g 代中的向量合并得到的新向量; $\mu = 0.05 \times randn$ 。

(4)局部搜索阶段。INFO使用局部搜索阶段来防止陷入局部最优解。根据该算子,如果 $rand < 0.5$,则可以围绕生成一个新向量,为 $[0, 1]$ 中的随机值。

$$u_l^g = \begin{cases} x_{bs} + randn \times [MeanRule + randn \times (x_{bs}^g - x_{a1}^g)], & rand1 < 0.5rand2 < 0.5 \\ x_{rnd} + randn \times [MeanRule + randn \times (v_1 \times x_{bs} - v_1 \times x_{rnd})], & rand1 < 0.5rand2 \geq 0.5 \end{cases} \quad (20)$$

其中:

$$x_{rnd} = \phi \times x_{avg} + (1 - \phi) \times [\phi \times x_{bt} + (1 - \phi) \times x_{bs}] \quad (21)$$

$$x_{avg} = \frac{x_a + x_b + x_c}{3} \quad (22)$$

式中: ϕ 为 $(0, 1)$ 的随机数; x_{rnd} , x_{bt} 和 x_{bs} 组成的新解,这增加了所提出算法的随机性,以更好地在解空间中搜索; v_1, v_2 为两个随机数,定义如下:

$$v_1 = \begin{cases} 2 \times rand, p > 0.5 \\ 1, p \leq 0.5 \end{cases} \quad (23)$$

$$v_2 = \begin{cases} rand, p > 0.5 \\ 1, p \leq 0.5 \end{cases} \quad (24)$$

式中: p 是 $[0, 1]$ 之间的随机数。

2 模型的建立与评价

2.1 模型的建立

(1)首先进行数据预处理,填补缺失值、剔除异常值,最后将清洗好的数据进行归一化。

(2)通过ReliefF算法进行特征选择并提取出与光伏功率具有较强相关性的6个气象因素,并以其中具有强相关性的总辐射水平因素进行高斯混合聚类,得到3种天气类型。将80%的样本作为测试集,20%的样本作为训练集,训练集由相似日数据构成,测试集由待测日数据构成。

(3)初始化CNN-GRU模型,设置基本参数,卷积核大小等5种参数作为寻优参数,并规定寻优范围。

(4)混沌初始INFO算法向量种群,设置种群数量、最大迭代次数、维度,将参数及向量位置进行初始化。

(5)设置训练集的均方根误差为适应度函数,计算适应度值,公式如下。

$$\theta(\bullet) = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_i - T_i)^2} \quad (25)$$

式中: N 为预测光伏功率的总样本数; R_i 为功率实际值; T_i 为预测值。

(6)计算向量此时最佳位置,并判断是否有超出给定搜索边界的向量,并进行调整。

(7)计算更新后所对应的适应度值,并与前一次计算出的适应度值相比较,保留最佳向量位置。

(8)若达到终止条件,则为参数最优解,否则重复以上2个步骤。

(9)将测试集代入训练好的INFO-CNN-QRGRU模型中,完成光伏功率高精度预测,并结合分位数回归方法生成预测区间,最后通过核密度估计方法生成高质量的概率预测结果。

2.2 模型评价指标

2.2.1 点预测模型评价指标 点预测评价指标包括决定系数(r-squared, R^2),均方根误差(root mean square error, RMSE) I_{RMSE} ,平均绝对误差(mean absolute error, MAE) I_{MAE} 和平均绝对百分误差(mean absolute percentile error, MAPE) I_{MAPE} 。 I_{MAE} 、 I_{RMSE} 和 I_{MAPE} 值越小, I_{R^2} 值越接近1,说明该模型的预测准确度越高,其评价指标的计算方法公式如下。

$$I_{R^2} = \frac{\sum_{i=1}^N \left(T_i - \frac{1}{N} \sum_{i=1}^N T_i \right)^2}{\sum_{i=1}^N \left(R_i - \frac{1}{N} \sum_{i=1}^N R_i \right)^2} \quad (26)$$

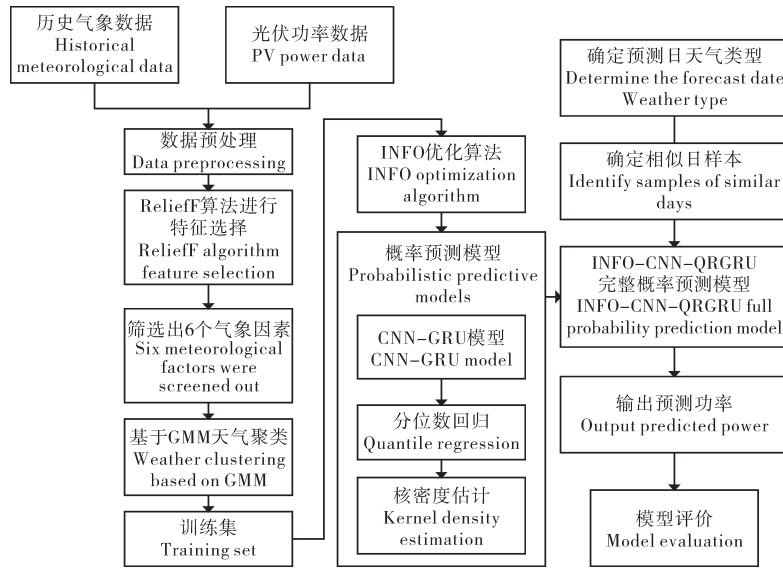


图2 INFO-CNN-QRGRU模型整体框架

Figure 2 INFO-CNN-QRGRU overall framework

$$I_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_i - T_i)^2} \tag{27}$$

$$I_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{R_i - T_i}{R_i} \right| \times 100\% \tag{28}$$

$$I_{MAE} = \frac{1}{N} \sum_{i=1}^N |R_i - T_i| \tag{29}$$

2.2.2 区间预测模型评价指标 区间评价指标包括区间综合评价指标(width-based criterion, WC) e_{WC} , 区间覆盖率(prediction interval coverage probability, PICP) e_{PICP} 和区间平均宽度(prediction interval normalized average, PINAW) e_{PINAW} 。当不同模型 e_{PICP} 值相等时, e_{PINAW} 越小预测效果越好。 e_{WC} 对区间预测结果评价最为全面,值越小测效果越好,计算公式如下。

$$e_{PICP} = \frac{1}{N} \sum_{n=1}^N S_n \tag{30}$$

$$e_{PINAW} = \frac{1}{N \cdot E} \sum_{i=1}^N (T_{upi} - T_{downi}) \tag{31}$$

$$e_{WC} = e_{PINAW} / e_{PICP} \tag{32}$$

式中: S_n 可取0或1值,当实际值落在预测区间范围内, S_n 取1,反之取0; E 为最大和最小实际功率之差; T_{upi}, T_{downi} 为预测区间的上下界。

2.2.3 概率预测模型评价指标 概率预测评价指标采用连续分级概率评分(continuous ranked probability score, CRPS) e_{CRPS} ,值越小预测效果越好,公式如下。

$$e_{CRPS} = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{+\infty} (F(T_i) - H(T_i - R_i))^2 dT_i \tag{33}$$

其中:

$$F(T_i) = \int_{-\infty}^{T_i} p(x) dx \tag{34}$$

$$H(T_i - R_i) = \begin{cases} 0, & T_i < R_i \\ 1, & T_i \geq R_i \end{cases} \tag{35}$$

式中: $F(T_i)$ 为 T_i 的累积分布函数; $p(x)$ 为概率密度函数; $H(T_i - R_i)$ 为阶跃函数。

3 结果与分析

3.1 相似日聚类

通过GMM算法得到的聚类结果如图3,数据经过聚类被分为3种天气类型,3种天气类型聚类效果明显,证明此数据适合做本次预测研究。

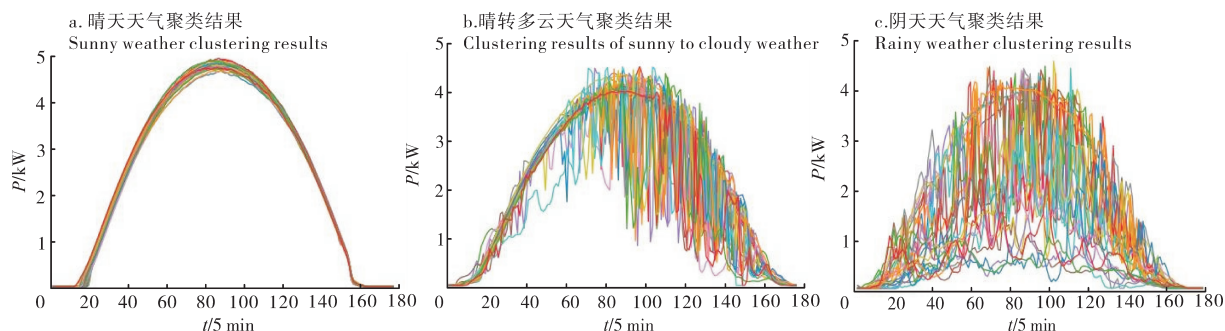


图3 GMM算法聚类结果
Figure 3 Results of GMM clustering

3.2 超参数优化及模型构架的确定

经过Relieff算法进行特征选择后得到6个气象影响因素,由此构造出6×6的特征矩阵,将n个6×6的向量矩阵作为输入向量输入到CNN-GRU网络中。根据输入向量的特点,本研究所采用的一维CNN网络由1个卷积层(Conv2D)、1个池化层(MaxPooling2D)和1个全连接层(Dense)组成,边界处理方式采用'same',激活函数为'relu',并通过优化算法对卷积核大小及数量进行寻优;GRU网络采用两层GRU结构,其中隐含层层数通过优化参数寻优确定,同时为验证实验有效性,4种优化算法的CNN-QRGRU模型除经过寻优的参数值不同外,其他参数设置均相同。INFO、WOA、GWO、SSA共4种优化方法设定种群数量均为10,最大迭代次数均为15次,优化变量个数为5个,并分别规定变量上下边界即寻优范围,通过训练适应度函数分别对CNN-GRU网络中的参数进行寻优,得到的寻优结果如表1。

表1 超参数寻优结果

Table 1 Hyperparameter optimization results

模型 Model	寻优参数 Optimization parameters	INFO	WOA	GWO	SSA
CNN-GRU	初始学习率 Initial learning rate	0.005	0.005	0.001	0.001
	卷积核数量 Number of convolution kernels	64	32	16	16
	卷积核大小 Convolution kernel size	4	6	10	20
	GRU 隐含层层数 GRU implied layers	22	16	10	10
	正则化系数 Regularization coefficient	0.001	0.005	0.003	0.001

3.3 点预测结果仿真分析

表2所展示的依次为CNN-GRU,以及CNN-GRU算法分别由SSA、GWO、WOA和改进的INFO算法进行超参数调优后模型的精度比较。经过优化算法优化的网络较CNN-GRU模型的预测精度均有所提高,相较之下,本研究所提出的INFO-CNN-GRU模型预测精度最高,在晴天、阴雨天及晴转多云的天气情况下,决定系数 R^2 分别达到0.997 7,0.990 1,0.993 3,相比于其他4种方法更接近1;MAE、

表2 模型精度比较

Table 2 Model accuracy comparison

模型 Model	晴天 Sunny weather				阴雨天 Rainy weather				晴转多云 Sunny to cloudy weather			
	R^2	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE
CNN-GRU	0.894 3	0.144 1	0.148 8	0.181 2	0.768 4	0.199 1	0.140 2	0.199 2	0.799 3	0.217 8	0.279 2	0.298 7
SSA-CNN-GRU	0.952 1	0.085 1	0.142 3	0.171 1	0.823 2	0.133 4	0.137 4	0.196 4	0.899 3	0.140 4	0.261 1	0.288 5
GWO-CNN-GRU	0.993 2	0.143 1	0.161 5	0.262 8	0.964 3	0.125 6	0.116 7	0.137 6	0.962 1	0.098 3	0.237 8	0.256 8
WOA-CNN-GRU	0.994 0	0.080 5	0.100 6	0.168 5	0.894 5	0.082 1	0.102 6	0.109 8	0.863 7	0.130 1	0.230 1	0.206 7
INFO-CNN-GRU	0.997 7	0.072 4	0.100 1	0.149 7	0.990 1	0.078 3	0.097 2	0.092 8	0.993 3	0.088 5	0.127 6	0.100 8

RMSE及MAPE值相较于其他4种方法都有所降低。

在3种天气情况下不同预测模型的结果对比如图4。相较于未经优化的CNN-GRU模型,其他4种优化模型的拟合程度及预测精度均有所提升,INFO-CNN-GRU相比较于其他3种预测模型在光伏发电预测方面的效果更贴近真实值,预测精度更高。晴天天气状况下曲线波动较小,4种预测模型预测精度相对较高,而在预测功率最大处INFO-CNN-GRU显示出更高的拟合度,其预测精度明显优于其他模型,阴雨天和晴转多云天气下曲线波动较大,两种天气情况下CNN-GRU、GWO-CNN-GRU及SSA-CNN-GRU模型的预测结果产生明显波动,与真实光伏功率值偏离程度较大,预测精度不足,而INFO-CNN-GRU预测出的曲线相比其余3种模型在波峰波谷处的曲线更加平滑,较为贴合真实光伏功率值,拟合效果最好。

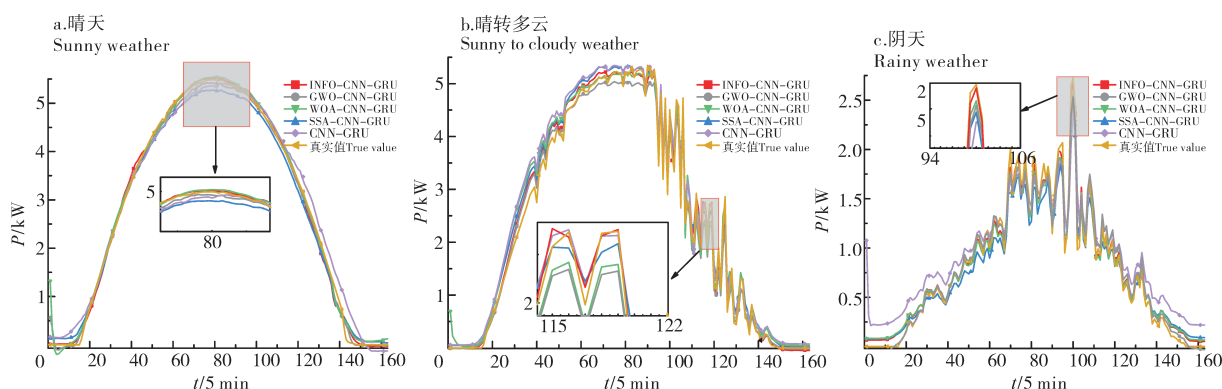


图4 不同天气状况下点预测对比

Figure 4 Comparison of point forecasts under different weather conditions

3.4 区间预测和概率预测结果仿真分析

将本研究提出的INFO-CNN-QRGRU模型与其他优化算法的分位数回归模型的预测效果分别在3种天气类型下进行了比较。选取95%置信水平分别建立不同天气情况典型样本的预测模型,并将INFO-CNN-QRGRU模型与其他算法的区间预测结果精度进行对比,区间预测评价指标WC、PICP、PINAW及CRPS结果如表3。

表3 模型区间概率预测精度比较

Table 3 Comparison of model interval probability prediction accuracy

置信区间 Confidence interval	优化算法 Optimization algorithms	晴天 Sunny weather				阴雨天 Rainy weather				晴转多云 Sunny to cloudy weather			
		WC	PICP	PINAW	CRPS	WC	PICP	PINAW	CRPS	WC	PICP	PINAW	CRPS
95%	SSA	0.070 7	0.986 7	0.059 6	0.080 2	0.134 4	0.958 7	0.150 4	0.083 0	0.216 3	0.950 9	0.201 3	0.138 7
	WOA	0.047 6	1.000 0	0.034 5	0.060 4	0.124 5	0.963 4	0.129 9	0.105 6	0.168 5	0.967 4	0.198 6	0.119 8
	GWO	0.063 1	0.992 7	0.050 8	0.066 5	0.129 8	0.967 3	0.132 1	0.107 6	0.199 7	0.958 4	0.200 4	0.120 6
	INFO	0.019 2	1.000 0	0.017 8	0.067 8	0.116 7	0.972 1	0.119 7	0.090 3	0.150 9	0.977 5	0.155 8	0.114 5

4种不同优化算法的预测模型在晴天天气状况下的功率区间预测结果如图5。由图5可知,在95%置信水平下,各个模型的预测区间覆盖率较大,且INFO-CNN-QRGRU预测区间相对较窄,意味着在晴天天气状况下进行光伏发电预测效果最佳。WC指标较SSA-CNN-QRGRU降低5.15%,较WOA-CNN-QRGRU降低2.84%,较GWO-CNN-QRGRU降低4.39%;同时INFO-CNN-QRGRU模型的区间平均宽度评价指标值PINAW也最小,PINAW指标较SSA-CNN-QRGRU降低了4.18%,较WOA-CNN-QRGRU降低1.67%,较GWO-CNN-QRGRU降低3.3%,因此其区间预测性能最佳。图6为INFO-CNN-QRGRU模型概率预测结果,选择9个采样点进行分析,可以观察到一部分采样点实际值靠近中心,也有一些采样点实际值偏离中心,表明概率预测是可靠的,如果所有实际值都位于中心或远离中心,则这些概率预测结果可能没有说服力。核密度曲线形状平滑饱满,说明选取了合适的带宽,通过观察CRPS值可知,INFO-CNN-QRGRU模型的概率评价指标相比于其他模型最小,由此说明IN-

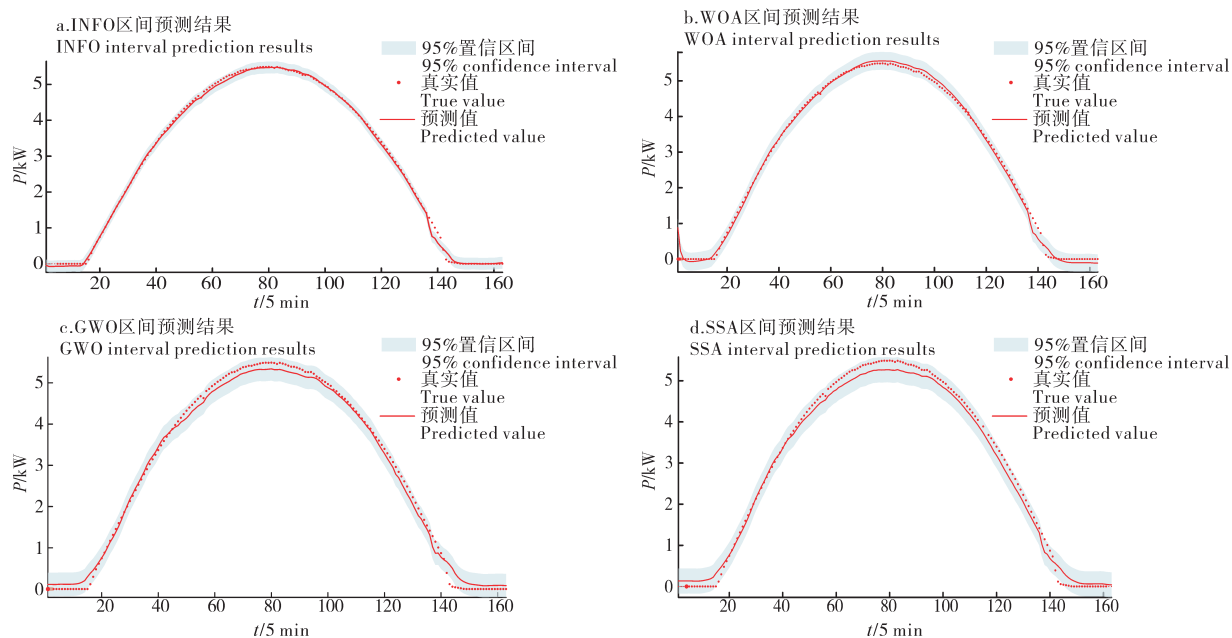


图5 晴天天气状况下区间预测结果

Figure 5 Interval forecast results under sunny weather conditions

FO-CNN-QRGRU 模型预测性能更好。

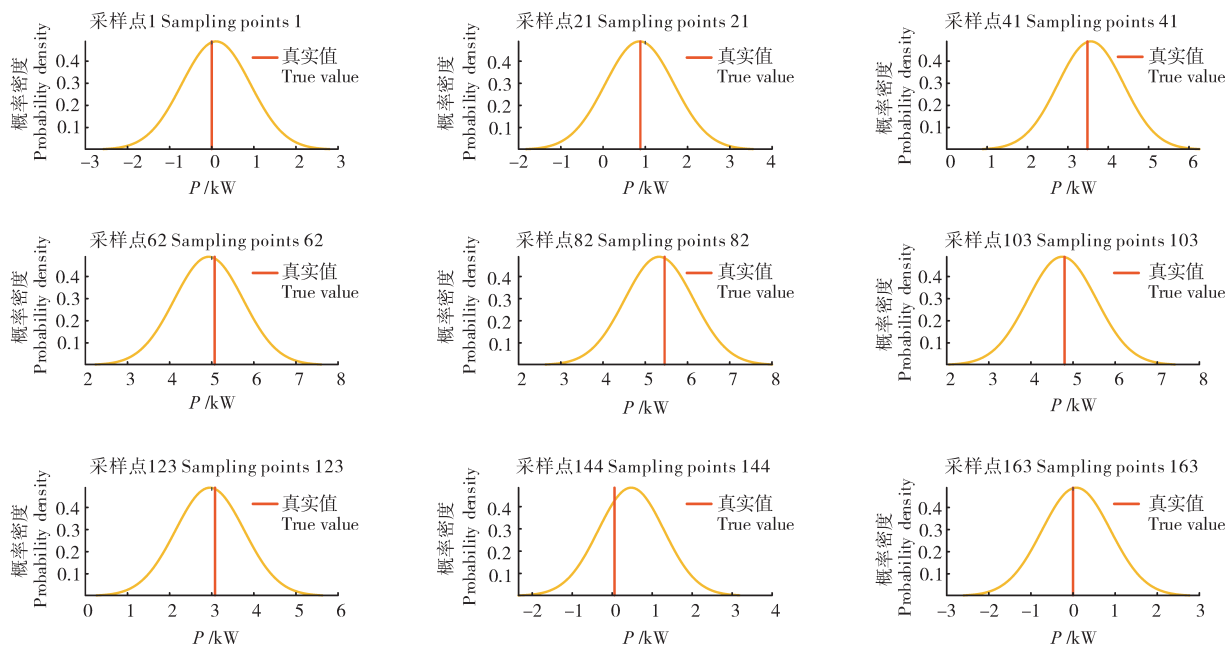


图6 晴天天气状况下概率预测结果

Figure 6 Probability prediction results under sunny weather conditions

在95%置信区间水平下,在晴转多云天气状况下4种优化算法的区间和概率预测结果如图7和图8。由图7和图8可知,INFO-CNN-QRGRU模型的预测宽度较其他3种模型差别不大,但预测区间覆盖率优于其他3种模型,几乎所有实际值均可落在预测区间内,参考模型评价数据可知,INFO-CNN-QRGRU模型的区间综合评价指标值WC最小,WC指标较SSA-CNN-QRGRU降低6.54%,较WOA-CNN-QRGRU降低1.76%,较GWO-CNN-QRGRU降低4.88%,因此其区间预测性能最佳。INFO-CNN-QRGRU模型CRPS值最小,表明其概率预测性能更好。

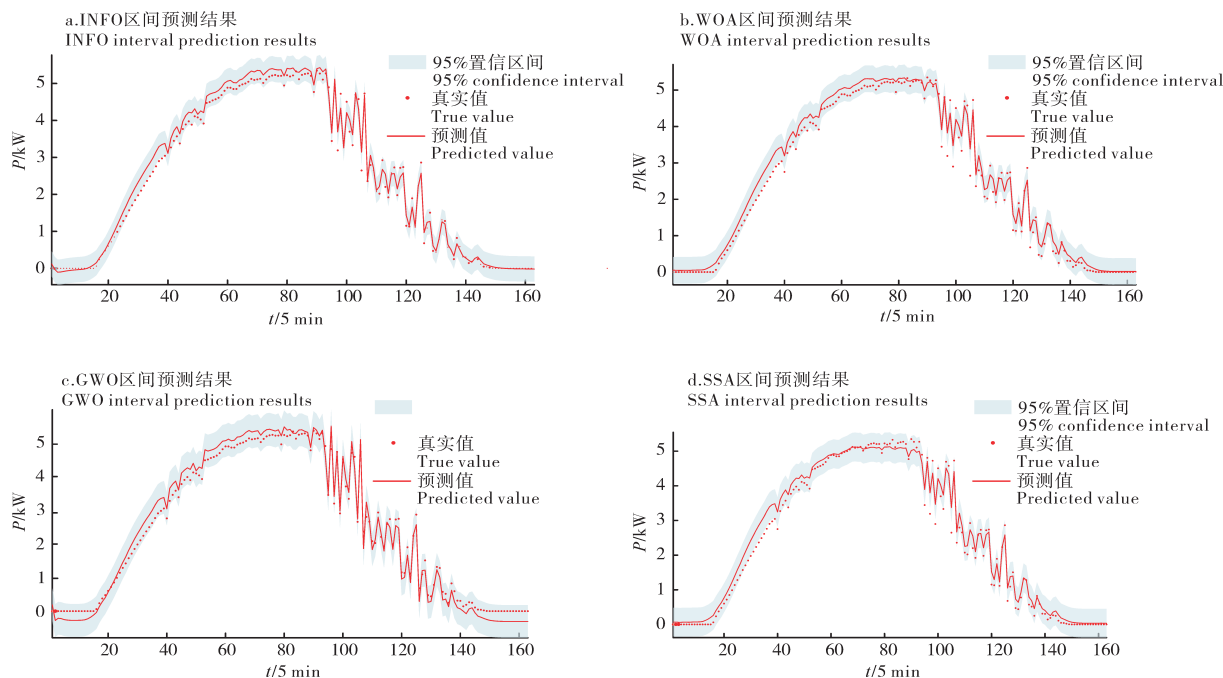


图7 晴转多云天气状况下区间预测结果

Figure 7 Interval forecast results under sunny to cloudy weather conditions

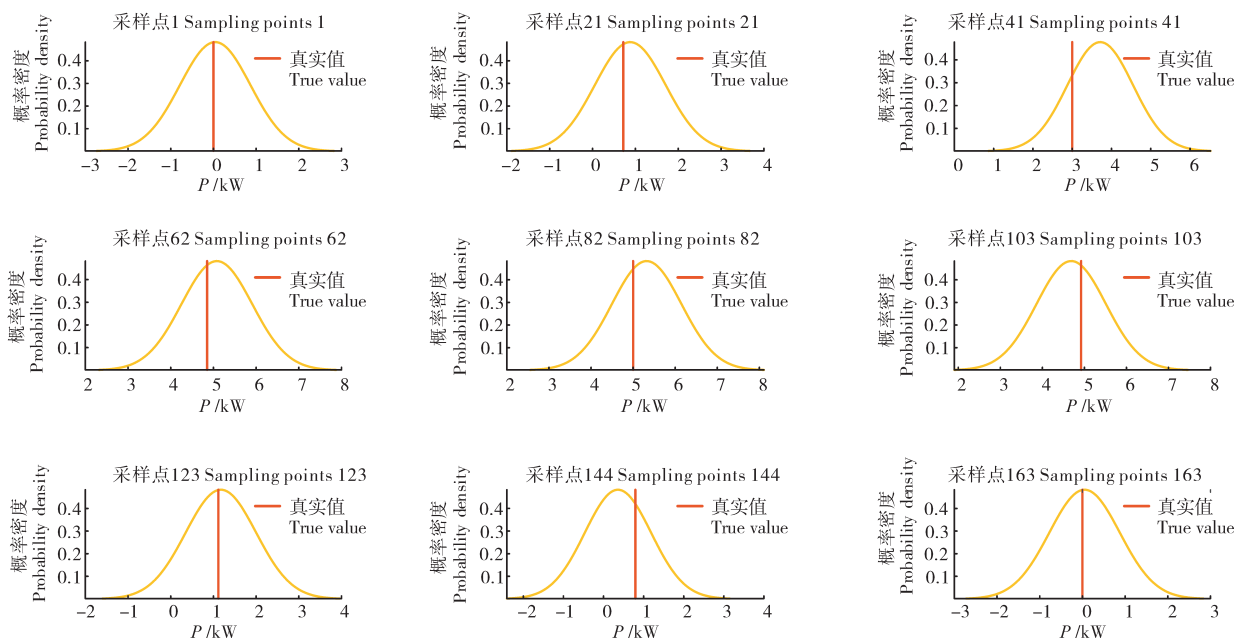


图8 晴转多云天气状况下概率预测结果

Figure 8 Probability prediction results under sunny to cloudy weather conditions

在95%置信区间水平下,4种预测模型在阴天天气状况下的区间和概率预测结果图如图9和图10。由图9和图10可知,在恶劣天气情况下,INFO-CNN-QRGRU模型表现出了更好的预测性能,预测区间宽度明显窄于其他预测模型,在波动剧烈时段区间的预测性能也显著优于其他模型,PINAW指标较SSA-CNN-QRGRU降低3.07%,较WOA-CNN-QRGRU降低1.02%,较GWO-CNN-QRGRU降低1.24%。且INFO-CNN-QRGRU模型CRPS值最小,不同采样点的实际值均在中心附近摆动,综合比较可以得出此模型具有良好的预测性能,且面对预测难度最大的阴雨天气时,预测效果更为突出。

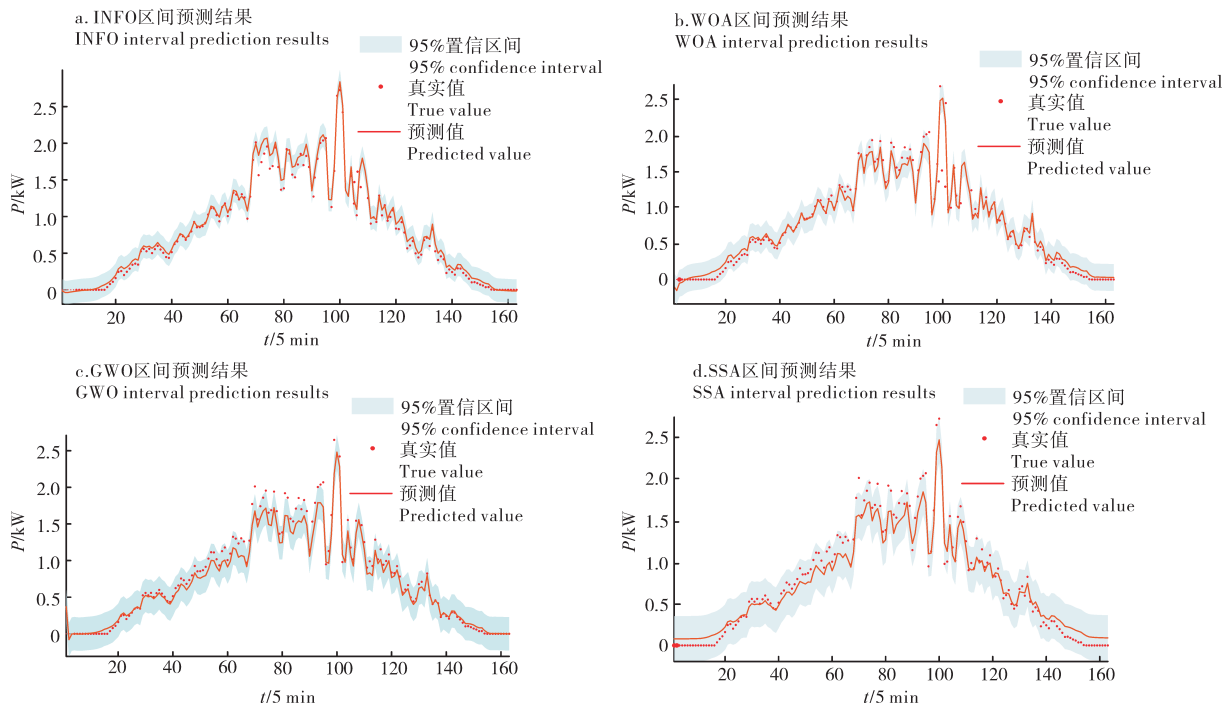


图9 阴天天气状况下区间预测结果

Figure 9 Interval forecast results under rainy weather conditions

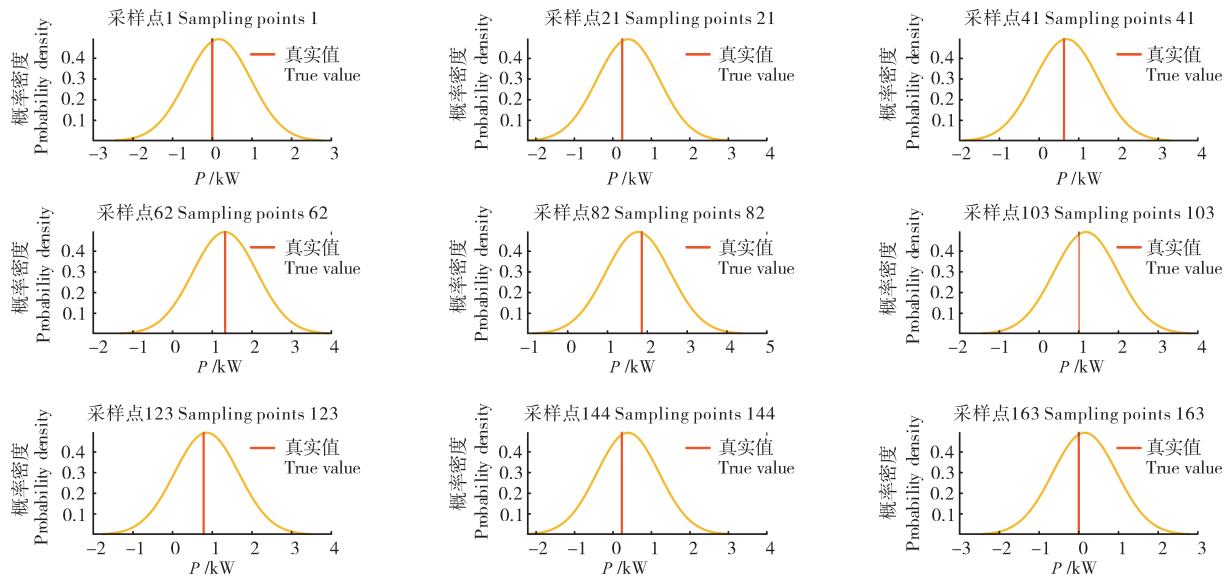


图10 阴天天气状况下概率预测结果

Figure 10 Probability prediction results under rainy weather conditions

4 讨论与结论

目前我国农村地区已大量投入光伏,农村光伏具有安装地区广,分散性强等特点,对并网的要求更加严格,因此光伏功率预测技术对光伏的顺利并网至关重要。在该背景下,本研究提出了基于改进的INFO算法优化CNN-QRGRU网络的概率预测模型,并通过实际算例进行验证,依赖光照强度这一因素进行预测,对于不同天气环境的农村地区进行快速准确的预测。INFO优化算法是2022年最新提出的优化算法,且在光伏预测领域并未应用,本研究首次将INFO优化算法与CNN-QRGRU模型相结合,并利用混沌映射对初始化种群进行改进得出点预测结果,在此基础上利用核密度函数完成概率预测,通过与其他模型对比发现此模型在准确性、不确定性和可靠性方面具有较好的预测效果。

未来的光伏预测系统将更加注重多源数据的融合,不仅限于历史气象数据、实时天气信息,还将融合卫星图像、地理空间数据、电网运行状态以及光伏电站的具体运营数据等信息,结合更先进的机器学习模型、物理模型和数据驱动模型,进一步提升预测的准确性。光伏发电功率预测技术将在未来不断发展和完善,为光伏电站的高效运行和可再生能源的可持续发展提供有力支持。

参考文献:

- [1] 吴 硕.光伏发电系统功率预测方法研究综述[J].热能动力工程,2021,36(8):1-7.
- [2] 龚莺飞,鲁宗相,乔 颖,等.光伏功率预测技术[J].电力系统自动化,2016,40(4):140-151.
- [3] SOBRI S,KOOHI-KAMALI S,RAHIM N A.Solar photovoltaic generation forecasting methods:A review[J].Energy Conversion and Management,2018,156:459-497.
- [4] 赵玉林,张冬梅,马文川,等.局部阴影条件下光伏电池多峰值最大功率点控制策略[J].东北农业大学学报,2015,46(5):89-94.
- [5] 王 俊,王继焯,程 坤,等.基于双层优化VMD-LSTM的农村超短期电力负荷预测[J].沈阳农业大学学报,2024,55(1):92-102.
- [6] OZCANLI A K,YAPRAKDAL F,BAYSAL M.Deep learning methods and applications for electrical power systems: A comprehensive review[J].International Journal of Energy Research,2020,44(9):7136-7157.
- [7] 赖昌伟,黎静华,陈 博,等.光伏发电出力预测技术研究综述[J].电工技术学报,2019,34(6):1201-1217.
- [8] SHI H,XU M H,LI R.Deep learning for household load forecasting—a novel pooling deep RNN[J].IEEE Transactions on Smart Grid,2018,9(5):5271-5280.
- [9] 孟 巍,郭腾炫,刘跌娟,等.基于长短记忆神经网络优化的短期光伏发电预测方法[J].电网与清洁能源,2022,38(5):129-134.
- [10] 孟安波,许炫淙,陈嘉铭,等.基于强化学习和组合式深度学习模型的超短期光伏功率预测[J].电网技术,2021,45(12):4721-4728.
- [11] 刘国海,孙文卿,吴振飞,等.基于Attention-GRU的短期光伏发电功率预测[J].太阳能学报,2022,43(2):226-232.
- [12] 王开艳,杜浩东,贾 嵘,等.基于相似日聚类和QR-CNN-BiLSTM模型的光伏功率短期区间概率预测[J].高电压技术,2022,48(11):4372-4388.
- [13] 李崎勇,张伟斌,赵新哲,等.改进鲸鱼算法优化支持向量回归的光伏最大功率点跟踪[J].电工技术学报,2021,36(9):1771-1781.
- [14] 马昕霞,林玉杰,宋明中,等.风光储发电系统容量配置优化算法综述[J].上海电力大学学报,2022,38(6):589-593.
- [15] 高岳林,杨钦文,王晓峰,等.新型群体智能优化算法综述[J].郑州大学学报(工学版),2022,43(3):21-30.
- [16] 邹 智,吴铁洲,张晓星,等.基于贝叶斯优化CNN-BiGRU混合神经网络的短期负荷预测[J].高电压技术,2022,48(10):3935-3945.
- [17] 吴 勇,高 昕,郭灏阳,等.基于优化的VMD-CNN-LSTM模型的光伏功率预测[J].邵阳学院学报(自然科学版),2022,19(6):9-17.
- [18] CHENG Z,LIU Q,ZHANG W.Improved probability prediction method research for photovoltaic power output[J].Applied Sciences,2019,9(10):2043.
- [19] BRACALE A,CARPINELLI G,DE FALCO P.A probabilistic competitive ensemble method for short-term photovoltaic power forecasting[J].IEEE Transactions on Sustainable Energy,2017,8(2):551-560.
- [20] 董 雷,周文萍,张 沛,等.基于动态贝叶斯网络的光伏发电短期概率预测[J].中国电机工程学报,2013,33(增刊1):38-45.
- [21] WU X M,LAI C S,BAI C C,et al.Optimal kernel ELM and variational mode decomposition for probabilistic PV power prediction[J].Energies,2020,13(14):3592.
- [22] 许 彪,徐青山,黄 煜,等.基于藤copula分位数回归的光伏功率日前概率预测[J].电网技术,2021,45(11):4426-4435.

[责任编辑 马迎杰]