

文章编号: 1006-2467(2024)03-0285-10

DOI: 10.16183/j.cnki.jsjtu.2022.338

基于极限学习机模型参数优化的 光伏功率区间预测技术

何之倬¹, 张颖¹, 郑刚¹, 郑芳¹, 黄琬迪², 张沈习², 程浩忠²

(1. 国网上海市电力公司青浦供电公司, 上海 201700;

2. 上海交通大学 电力传输与功率变换控制教育部重点实验室, 上海 200240)

摘要: 提出一种基于极限学习机(ELM)模型参数优化的光伏功率区间预测技术. 首先, 提出加权欧氏距离作为光伏功率预测区间评估指标, 筛选历史样本单元并优化 ELM 训练集; 然后, 提出 ELM 参数混合寻优算法, 利用精英保留策略遗传算法与分位数回归优化 ELM 模型隐层输入及输出权重与偏置参数, 并采用训练后的模型预测光伏功率区间; 最后, 基于光伏电站与气象站历史数据构建实际算例, 预测光伏功率区间, 并与其他方法得到的结果进行对比. 算例结果表明: 所提方法在增加区间预测可信度的同时, 能较大程度提高区间预测准确度.

关键词: 光伏功率; 区间预测; 极限学习机; 参数优化; 加权欧氏距离指标

中图分类号: TM615

文献标志码: A

Interval Prediction Technology of Photovoltaic Power Based on Parameter Optimization of Extreme Learning Machine

HE Zhizhuo¹, ZHANG Ying¹, ZHENG Gang¹, ZHENG Fang¹

HUANG Wandu², ZHANG Shenxi², CHENG Haozhong²

(1. State Grid Shanghai Qingpu Electric Power Supply Company, Shanghai 201700, China;

2. Key Laboratory of Control of Power Transmission and Conversion of the Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: This paper proposes an interval prediction technology of photovoltaic (PV) power based on parameter optimization of extreme learning machine (ELM) model. First, the weighted Euclidean distance is proposed as the evaluation index of PV power prediction interval. The historical sample units are screened and the ELM training set is optimized. Then, a hybrid optimization algorithm for ELM parameters is proposed. The hidden layer input and output weights and biases parameters of the ELM model are optimized by using the elitist strategy genetic algorithm and quantile regression, and the trained model is used to predict the PV power range. Finally, an actual calculation example is constructed based on the historical data of PV power plants and weather stations. The PV power interval is predicted, and the results are compared with those obtained by other methods. The results of the calculation example show that the method proposed can greatly improve the accuracy of interval prediction while increasing the reliability of interval prediction.

收稿日期: 2022-08-30 修回日期: 2022-11-20 录用日期: 2022-12-08

基金项目: 国网上海市电力公司科技项目(52093421N001), 国家自然科学基金(51907123)

作者简介: 何之倬(1990-), 硕士, 工程师, 从事分布式光伏并网相关研究.

通信作者: 张颖, 博士, 高级工程师; E-mail: zhyingt@163.com.

Keywords: photovoltaic (PV) power; interval prediction; extreme learning machine (ELM); parameter optimization; weighted Euclidean distance index

为缓解日益突出的能源和环境问题,太阳能光伏(Photovoltaic, PV)作为一种资源丰富、安全可靠、环境友好的可再生能源,近年来得到大力发展^[1]. 由于光伏发电系统仅在白天有出力,且受到气象因素变化的影响,所以光伏出力具有一定波动性和随机性^[2]. 光伏在电网中渗透率的增加可能会对电力系统的潮流分布、暂态特性、电能质量等产生不利影响,使电网的运行与控制面临挑战^[3]. 对光伏功率进行准确预测有助于调度部门调整含 PV 电力系统的运行方式,以维持电力系统的安全稳定^[4].

目前,基于数据驱动的方法是使用最广泛的光伏功率预测方法,它可避免复杂的物理建模过程,从大量历史数据中获取光伏功率出力与外界因素之间的联系. 时间序列法^[5-6]、回归分析法^[7-8]、人工智能法^[9-11]等都属于基于数据驱动的预测方法.

依据预测结果的形式,光伏功率预测分为点预测和区间预测. 点预测是一种确定性预测,结果较为直观,但是难以表征光伏出力的不确定性;区间预测可以得到一定置信度水平下光伏功率的上下限,对于含光伏电力系统的风险评估、不确定性评估具有重要参考价值. Delta 法^[12-13]是一种构造预测区间的方法,需假定数据噪声同质且满足标准正态分布;事实上,噪声在很多情况下难以满足这个假设,因此 Delta 法得到的预测区间与实际情况可能有较大差距. 文献[14-15]中使用贝叶斯法构造预测区间,但是 Hessian 矩阵的求取使得此方法的计算负担过重. 文献[16-17]中使用 Bootstrap 法进行区间预测,得到认知不确定性和偶然不确定性分别对应的预测区间后进行叠加,其结果可以表征总体不确定性;Bootstrap 法虽易于实现,但当数据样本较多时,计算效率低. 边界估值(Lower Upper Bound Estimation, LUBE)理论^[18]根据预测区间评估指标,利用启发式算法对神经网络参数进行寻优,可以得到满足可信度和准确度要求的神经网络区间预测模型. 文献[18]中通过多个算例分析说明:与 Delta 法、贝叶斯法、Bootstrap 法相比,LUBE 法得到的区间预测模型性能更为稳定,且预测区间可信度较高. 传统前馈神经网络多采用梯度下降法,训练时间较长,容易陷入局部最优,且学习率具有选择敏感的特点. 极限学习机(Extreme Learning Machine, ELM)可克服上述缺陷,具有更快的学习速度和更优的泛化能

力,被广泛应用于预测领域^[19]. 文献[20]中使用 ELM 模型来进行光伏电站功率区间预测,ELM 模型隐层输入权重与偏置可以随机生成,而隐层输出权重通过求解最优化问题确定,进而得到预测区间. 然而,ELM 模型也存在缺点^[21]:当数据样本自变量过多时,模型的稳定性与泛化能力会受到不利影响;原始数据集中若存在离群点,可能会导致模型预测性能不佳;此外,模型的隐层输入权重与偏置参数随机生成,其预测精度仍有提升空间.

针对现有研究存在的问题,提出一种考虑 ELM 训练集优化与参数寻优的光伏功率区间预测技术. 利用相关性分析对 ELM 的输入参数进行筛选,剔除无关历史信息,仅保留与因变量具有较高相关性的自变量;提出基于加权欧氏距离指标的 ELM 训练集选取方法,在去除异常离群点、提高训练效率的同时,使得训练集样本和待预测样本具有较高的相似度,从而避免过拟合,提升预测的可信度和准确度;提出一种 ELM 参数混合寻优算法,在根据预测区间评估与优化准则设定适应度函数后,采用精英保留策略遗传算法(Elitist Strategy Genetic Algorithm, ESGA)优化 ELM 隐层的输入权重与偏置取值,同时使用分位数回归方法优化 ELM 的隐层输出权重,生成评价指标最优的预测区间,有效降低模型预测随机性. 将提出的预测方法应用于实际算例,并与其他方法进行比较,以证明本文方法在光伏功率区间预测上的优越性.

1 光伏功率区间预测

1.1 光伏功率预测区间定义

光伏功率是一个随机变量,利用区间预测可以得到其在一定条件下的取值区间,在工程实际中比点预测具有更高参考价值.

预测区间是表征随机变量在未来时刻不确定性的取值区间^[18],与置信度水平相关^[22]. 光伏功率概率密度函数如图 1 所示. 给定显著性水平 α ,则光伏功率预测区间额定置信水平(Prediction Interval Nominal Confidence, PINC)为 $100(1-\alpha)\%$. 分别用 $q_i^{\bar{}}$ 和 $q_i^{\underline{}}$ 表示预测区间上下限,在理想情况下,预测区间 $I_i = [q_i^{\underline{}}, q_i^{\bar{}}]$ 需要满足以下条件:

$$P_r(q_i^{\underline{}} \leq f_i \leq q_i^{\bar{}}) = 100(1-\alpha) \quad (1)$$

式中: $P_r(\cdot)$ 为概率; f_i 为预测的变量值.

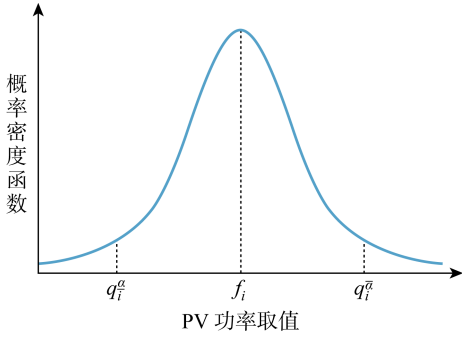


图1 PV 功率概率密度函数

Fig. 1 Probability density function of PV power

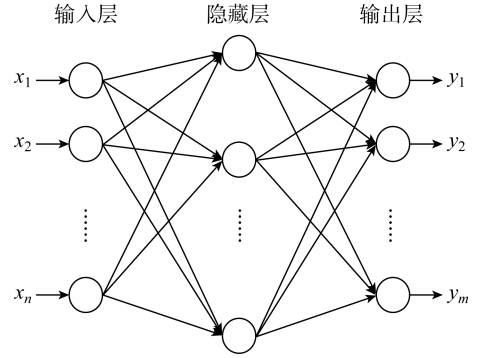


图2 ELM 结构

Fig. 2 Structure of ELM

1.2 光伏功率预测区间评估指标

分别采用可信度与准确度衡量区间预测性能。首先,预测区间覆盖率(Prediction Interval Coverage Probability, PICP)可以表征预测区间可信度,即

$$p_{\text{PICP}} = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} \delta_i^{\alpha} \quad (2)$$

式中: N_{tot} 为预测样本总数; δ_i^{α} 为指示函数,此时为布尔类型,当光伏功率目标值位于预测区间上下限之间时, $\delta_i^{\alpha} = 1$, 否则 $\delta_i^{\alpha} = 0$ 。有效预测区间的 PICP 值应当不低于预置的 PINC 值 (p_{PINC}), 若 PICP 远小于 PINC, 则说明预测区间无效。

此外,如果 PICP 非常高,但是预测区间宽度非常大,则区间预测结果没有参考价值。预测区间归一化平均带宽(Prediction Interval Normalized Average Width, PINAW)可以表征预测区间准确度,即

$$p_{\text{PINAW}} = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} (q_i^a - q_i^b) \quad (3)$$

在 PINC 一定的情况下, PINAW 越小,说明预测区间平均宽度越窄,准确度越高。

2 ELM

2.1 ELM 原理

ELM 是一种单层前馈神经网络,于 2004 年由 Huang 等^[23]提出。ELM 的隐藏层输入权重和偏置可以随机生成,唯一需要确定的是隐藏层输出权重。与传统前馈神经网络相比,ELM 运算效率非常高,且具有优越的泛化性能^[24]。

具有 n 个输入层神经元、 L 个隐藏层神经元、 m 个输出层神经元的 ELM 结构如图 2 所示。对于 N 个离散训练样本 $\{\mathbf{x}_i, \mathbf{t}_i\}_{i=1}^N$, 输入数据为 $\mathbf{x}_i = [x_{i1} \ x_{i2} \ \cdots \ x_{in}]^T \in \mathbf{R}^n$, 目标输出数据为 $\mathbf{t}_i = [t_{i1} \ t_{i2} \ \cdots \ t_{im}]^T \in \mathbf{R}^m$, ELM 可以通过训练得到如下模型:

$$\mathbf{y}_i = \sum_{j=1}^L \beta_j g(\omega_j \cdot \mathbf{x}_i + b_j) \quad (4)$$

式中: $\mathbf{y}_i = [y_{i1} \ y_{i2} \ \cdots \ y_{im}]^T \in \mathbf{R}^m$ 为第 i 个训练样本的输入 \mathbf{x}_i 经 ELM 处理后的输出; $\beta_j = [\beta_{j1} \ \beta_{j2} \ \cdots \ \beta_{jm}]^T \in \mathbf{R}^m$ 为第 j 个隐藏层神经元的输出权重; $g(\cdot)$ 为激励函数; $\omega_j = [\omega_{j1} \ \omega_{j2} \ \cdots \ \omega_{jn}]^T \in \mathbf{R}^n$ 为第 j 个隐藏层神经元的输入权重; b_j 为第 j 个隐藏层神经元的偏置。

由于 ELM 能够以极小误差逼近训练样本,所以训练样本的目标输出可表示为

$$\mathbf{t}_i = \sum_{j=1}^L \beta_j g(\omega_j \cdot \mathbf{x}_i + b_j) \quad (5)$$

相应的矩阵形式为

$$\mathbf{T} = \mathbf{H}\beta \quad (6)$$

式中: $\mathbf{T} \in \mathbf{R}^{N \times m}$ 为目标输出矩阵; $\mathbf{H} \in \mathbf{R}^{N \times L}$ 为隐藏层输出矩阵; $\beta \in \mathbf{R}^{L \times m}$ 为隐藏层输出权重矩阵。

当隐藏层输入权重和偏置生成后, \mathbf{H} 为常数矩阵。因此, β 的求解可以视为求解线性系统的最小二乘特解问题,即寻找 β 的最优值使代价函数, ELM 模型输出和目标输出之差的模最小,如下式所示:

$$\|\mathbf{H}\beta^* - \mathbf{T}\| = \min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\| \quad (7)$$

由广义逆理论可得, β 的最小二乘特解可以表示为

$$\beta^* = \mathbf{H}^+ \mathbf{T} \quad (8)$$

式中: \mathbf{H}^+ 为 \mathbf{H} 的 Moore-Penrose 广义逆。

ELM 被广泛应用于基于数据驱动的变量预测,并能取得较为理想的预测结果。在训练前,通常将样本数据归一化,并使用 ELM 预测光伏电站功率。

2.2 ELM 参数优化与预测区间生成

为了使 ELM 区间预测获得更优结果,从两个方面对 ELM 进行改进。第一,提出加权欧氏距离指标,对历史样本进行筛选以获得 ELM 训练集,使得

训练集样本和待预测日各时刻的样本具有较高度度的相似性;第二,提出 ELM 参数混合寻优算法,使用 ESGA 对 ELM 隐藏层输入权重 ω_{hid} 与偏置 b_{hid} 进行寻优,在每次迭代中,对于给定的 ω_{hid} 和 b_{hid} ,选取分位数回归方法优化 ELM 隐藏层输出权重的参数值和对应预测区间,并计算个体适应度,最终确定使 ELM 预测性能最优的隐藏层输入权重与偏置值.

2.2.1 ELM 训练集选取 提出加权欧氏距离指标来衡量待预测样本与历史样本自变量数值之间的相似度.加权欧氏距离越小,表示相似度越高.为充分考虑不同自变量对因变量影响程度的差异性,首先应进行相关性分析.考虑光伏出力数据具有周期性等特征,采用 Spearman 相关系数描述数据间相关程度.根据定义,变量 x 与 y 间的 Spearman 相关系数计算如下:

$$r(x, y) = \frac{\sum_{i=1}^t (R_{x_i} - \bar{R}_x)(R_{y_i} - \bar{R}_y)}{\sqrt{\sum_{i=1}^t (R_{x_i} - \bar{R}_x)^2} \sqrt{\sum_{i=1}^t (R_{y_i} - \bar{R}_y)^2}} \quad (9)$$

式中: x_i, y_i 分别为变量 x 与 y 的第 i 项样本数据,即气象因素与 PV 功率时间序列的第 i 项历史数据; t 为变量中的样本总数; R_{x_i}, R_{y_i} 分别为 x_i 与 y_i 在相应变量中的秩次; \bar{R}_x, \bar{R}_y 分别为变量 x 与 y 的秩次平均值. Spearman 相关系数 $r(x, y)$ 取值范围为 $[-1, 1]$,其正负代表变量间的正相关或负相关关系,而绝对值则表征变量相关性大小.

变量间距离度量方面,传统欧氏距离主要计算变量间的真实距离,描述样本间不相似程度.对应元素较多时,为充分考虑各元素对累积相似性的影响,可采用加权欧氏距离描述气象特征对光伏出力的影响情况.设 x_1, x_2, \dots, x_{n_s} 是 n_s 个对因变量影响较大的自变量, Spearman 相关系数分别为 r_1, r_2, \dots, r_{n_s} , 对于气象因素与光伏功率时间序列,采用 Spearman 相关系数确定权值,并定义样本之间的加权欧氏距离如下:

$$d(h_i, x_p) = \sqrt{\sum_{k=1}^{n_s} \frac{1}{\omega_k} (h_{ik} - x_{pk})^2} \quad (10)$$

式中: h_i 为历史数据中某样本单元的自变量数值向量; x_p 为某待预测样本单元的自变量预报值向量; $1/\omega_k$ 为自变量 k 的权重, Spearman 相关系数绝对值越大, ω_k 越大,距离权重赋值越小. ω_k 的计算公式如下:

$$\omega_k = \frac{|r_k|}{\sum_{l=1}^n |r_l|}, \quad k \in \{1, 2, \dots, n\} \quad (11)$$

加权欧氏距离可充分考虑各因素对累积相似性的重要程度,通过各因素相关性决定变量权值系数,使变量间的欧氏距离标准化.对于待预测的各样本单元,在历史数据中筛选与其加权欧氏距离最小的若干个样本单元,构成最终的 ELM 训练集.

2.2.2 ELM 隐藏层输入权重与偏置混合寻优 由于 ELM 的参数会对其预测性能产生影响,而随机生成的 ELM 隐藏层输入权重与偏置可能导致预测模型无法获得最优的预测区间,所以有必要确定 ELM 参数的最优值.生成 ELM 训练集后,将 ESGA 与分位数回归相结合,对 ELM 的隐藏层输入权重与偏置进行混合寻优.

预测区间的优化需要构建相应准则,以判断个体的优劣.综合考虑 PICP 和 PINAW,定义适应度函数为

$$f_{fit}(\omega_{hid}, b_{hid}) = -(\beta(\omega_{hid}, b_{hid})M \times |d_{ACE}(\omega_{hid}, b_{hid})| + p_{PINAW}(\omega_{hid}, b_{hid})) \quad (12)$$

式中: $\beta(\omega_{hid}, b_{hid})$ 为布尔类型指示函数,其具体数值由隐藏层输入权重及偏置矩阵决定,当 $p_{PINC} > p_{PICP}$ 时, $\beta(\omega_{hid}, b_{hid}) = 0$, 否则, $\beta(\omega_{hid}, b_{hid}) = 1$; M 为惩罚系数,此时取较大值;平均覆盖率误差是 PINC 和 PICP 的差值,用 $d_{ACE}(\omega_{hid}, b_{hid})$ 表示.

ESGA 优化 ELM 隐藏层输入权重与偏置具体流程可概括如下:

- (1) 生成 m_c 个待优化隐藏层输入权重及偏置组合 (ω_{hid}, b_{hid}) , 构成初始种群.
- (2) 根据适应度函数 $f_{fit}(\omega_{hid}, b_{hid})$ 计算结果进行个体评价,其中适应度最高的 m_{cfit} 个个体 (ω_{hid}, b_{hid}) 保留为精英个体.
- (3) 除精英个体外,通过选择、交叉、变异等遗传操作构建新的子代种群,并计算子代个体适应度.
- (4) 用保留的精英个体替换子代种群中适应度最低的个体,并将精英个体更新为子代种群中适应度最高的 m_{cfit} 个个体.
- (5) 若达到最大迭代次数,则退出循环,输出最优个体对应的隐藏层权重与偏置参数;否则,返回步骤(3).

在 ESGA 的每一次迭代中,对于每个个体对应的 ELM 隐藏层输入权重与偏置,利用分位数回归法得到 ELM 隐藏层输出权重最优值与预测区间,然后计算个体适应度.个体适应度越大,说明相应

ω_{hid} 和 b_{hid} 的取值越优, ELM 的区间预测性能越优。

2.2.3 预测区间生成 通常情况下, 历史数据中仅包含随机变量的观测值, 而不含取值区间上下限, 无法通过直接训练 ELM 对变量的取值区间进行预测。因此, 采用分位数回归理论^[25]生成光伏出力预测区间。

对于随机变量 y_{rand} , 分布函数用 $F(y_{\text{rand}})$ 表示。 y_{rand} 的第 τ 分位数定义为

$$F^{-1}(\tau) = \inf\{y_{\text{rand}} : F(y_{\text{rand}}) \geq \tau\} \quad (13)$$

式中: τ 是位于 $0 \sim 1$ 之间的数; $\inf\{\cdot\}$ 表示变量的下确界。

令 $q_i^{\bar{\alpha}}$ 和 $q_i^{\underline{\alpha}}$ 分别为随机变量的第 $\bar{\alpha}$ 和第 $\underline{\alpha}$ 分位数, 则预测区间的求解相当于在给定 p_{PINC} 的情况下确定分位数 $q_i^{\bar{\alpha}}$ 和 $q_i^{\underline{\alpha}}$, 可通过最小化分位数损失函数来实现。引入分位数损失函数后, 求解线性系统最小二乘问题可以转化为求解加权线性最小二乘问题, 目标函数和约束条件如下所示:

$$\begin{aligned} \min_{\beta_{\text{hid}, \bar{\alpha}}, \beta_{\text{hid}, \underline{\alpha}}} &= \sum_{i=1}^N [\rho_{\bar{\alpha}}(t_i - f(x_i, \beta_{\text{hid}, \bar{\alpha}})) + \rho_{\underline{\alpha}}(t_i - f(x_i, \beta_{\text{hid}, \underline{\alpha}}))] \\ \text{s. t.} & 0 \leq f(x_i, \beta_{\text{hid}, \underline{\alpha}}) \leq f(x_i, \beta_{\text{hid}, \bar{\alpha}}) \leq 1 \end{aligned} \quad (14)$$

式中: $\beta_{\text{hid}, \bar{\alpha}}, \beta_{\text{hid}, \underline{\alpha}}$ 分别为输出预测区间上下限对应的 ELM 模型隐藏层输出权重矩阵, 作为待优化变量; $\rho_{\bar{\alpha}}, \rho_{\underline{\alpha}}$ 分别为第 $\bar{\alpha}$ 和第 $\underline{\alpha}$ 分位数时的损失函数; $f(x_i, \beta_{\text{hid}, \bar{\alpha}}), f(x_i, \beta_{\text{hid}, \underline{\alpha}})$ 分别为光伏出力预测区间上下限; 归一化后, 变量取值范围为 $[0, 1]$ 。当 ELM 的隐藏层输入权重和偏置给定时, 隐藏层输出权重决定 ELM 模型输出。此外, 预测区间上下限应位于变量值归一化后的上下限之间。

在优化问题中引入辅助变量, 可以将原问题转化为线性规划问题^[25]。令 $\{u_i^{\bar{\alpha}}\}_{i=1}^N$ 和 $\{v_i^{\bar{\alpha}}\}_{i=1}^N$ 分别表示 $\{t_i - f(x_i, \beta_{\text{hid}, \bar{\alpha}})\}_{i=1}^N$ 的正、负残差, 同时令 $\{u_i^{\underline{\alpha}}\}_{i=1}^N$ 和 $\{v_i^{\underline{\alpha}}\}_{i=1}^N$ 分别表示 $\{t_i - f(x_i, \beta_{\text{hid}, \underline{\alpha}})\}_{i=1}^N$ 的正、负残差, 原问题等效为以下线性规划问题, 可使用 Mosek、Gurobi 等商业求解器求解。

$$\begin{aligned} \min_{\substack{u_i^{\bar{\alpha}}, v_i^{\bar{\alpha}}, u_i^{\underline{\alpha}}, v_i^{\underline{\alpha}}, \\ \beta_{\text{hid}, \bar{\alpha}}, \beta_{\text{hid}, \underline{\alpha}}}} &= \sum_{i=1}^N [\alpha u_i^{\bar{\alpha}} + (1 - \alpha) v_i^{\bar{\alpha}} + \bar{\alpha} u_i^{\underline{\alpha}} + (1 - \bar{\alpha}) v_i^{\underline{\alpha}}] \\ \text{s. t.} & y_i - f(x_i, \beta_{\text{hid}, \underline{\alpha}}) = u_i^{\underline{\alpha}} - v_i^{\underline{\alpha}} \\ & y_i - f(x_i, \beta_{\text{hid}, \bar{\alpha}}) = u_i^{\bar{\alpha}} - v_i^{\bar{\alpha}} \\ & u_i^{\underline{\alpha}}, v_i^{\underline{\alpha}}, u_i^{\bar{\alpha}}, v_i^{\bar{\alpha}} \geq 0 \\ & 0 \leq f(x_i, \beta_{\text{hid}, \underline{\alpha}}) \leq f(x_i, \beta_{\text{hid}, \bar{\alpha}}) \leq \hat{P} \end{aligned} \quad (15)$$

式中: \hat{P} 为光伏出力预测区间上限约束值。

3 光伏功率区间预测流程

3.1 数据获取

基于数据驱动的光伏功率预测需要分析大量历史数据, 通过模型训练来构建光伏功率与外部因素之间的关系。已有研究表明光伏功率特性存在季节性变化特征, 即不同季节对应的光伏功率预测模型有差异^[26]。确定待预测日的日期后, 需要分别从光伏电站和气象站获取此季节各个时刻的光伏功率数据与气象历史数据, 构成历史样本单元。

3.2 数据预处理

获得历史样本单元后, 需要对数据进行预处理, 得到可供模型训练的数据集。第一, 剔除存在缺失或异常数据点的历史样本单元, 以免对模型训练造成不利影响。第二, 使用 Spearman 相关系数量化气象因素对光伏功率的影响程度, 从中筛选出与光伏出力相关性较高的气象因素作为 ELM 的输入。第三, 基于 Spearman 相关系数值, 根据式(10)~(11)计算待预测日样本单元与历史样本单元间的加权欧氏距离; 对于待预测日的每个时刻, 选取与其相距最近的若干个历史样本单元, 共同构成 ELM 的训练集。

3.3 模型训练

确定训练集后, 需要进行模型训练。首先, 应确定 ELM 的结构。通过相关性分析确定 n 个与光伏出力具有较高相关程度的气象因素后, 相应地, ELM 输入层的神经元个数为 n ; 因 ELM 的输出是光伏功率上下限, 故输出层神经元个数为 2; 隐藏层神经元个数的选取需要保证预测模型的回归性能稳定, 不能过少, 同时也应当避免过多, 否则会加重计算负担, 且对提升 ELM 预测性能无益。交叉验证法^[27]是一种常用的确定神经网络结构的方法, 采用此法确定 ELM 的隐藏层神经元个数: 将 ELM 训练集样本随机均分为 5 个部分, 对同一隐藏层神经元个数, 依次取其中 4 个部分进行训练以确定 ELM 区间预测模型, 并将最后一部分评估最终区间预测性能, 然后将 5 次预测区间评估指标取均值。分析不同隐藏层神经元数量下预测区间评估指标均值, 便可以确定隐藏层神经元数目优化结果。优化 ELM 结构之后, 为了获得使模型预测性能最佳的 ELM 参数, 使用 ESGA 和分位数回归对 ELM 进行参数寻优, 确定 ELM 隐藏层输入权重与偏置的最优值, 形成 ELM 预测模型。

3.4 区间预测

得到 ELM 预测模型后, 输入待预测日各时刻

经相关性分析筛选后的气象因素预报值,通过分位数回归,获得相应置信水平下光伏功率的预测区间。

4 算例分析

4.1 数据介绍

为验证所提光伏功率区间预测方法的适用性,选取 2018 年澳大利亚昆士兰大学露西亚校区的光伏出力数据与气象站数据进行算例分析。光伏装机容量为 433 kW,监测的气象因素包括风向角、风速、温度、相对湿度、海平面气压、降雨量和太阳辐照度。光伏功率和气象数据的采样间隔均为 1 min。光伏在夜晚出力恒为 0,仅对白天光伏出力大于 0 的时刻进行光伏功率区间预测,选取时间步长为 15 min^[27]。在指定光伏功率待预测日后,通过对之前若干邻近日光光伏出力起止时刻的拟合分析,可确定待预测日中光伏出力大于 0 的时间段。

剔除历史数据中数据缺失或异常的样本后,分别计算四季光伏功率与各项因素的 Spearman 相关系数。根据相关系数绝对值的计算结果,太阳辐照度与光伏功率相关系数绝对值为最高水平,始终维持于 0.96 以上;四季相对湿度与光伏功率相关系数绝对值分布于 0.46~0.66;温度与光伏功率相关系数绝对值分布于 0.39~0.60,略低于相对湿度;其他因素如风向角、风速等与光伏功率的相关系数绝对值总体维持较低水平,低于 0.2,因此该类因素的影响忽略不计。

由于光伏出力特性与天气类型相关^[28],为测试所提预测方法在各种天气类型下是否均能实现高性能的光伏功率区间预测,选择 2018 年 7 月 1 日(阴雨天)、8 月 3 日(多云天)和 9 月 10 日(晴天)的数据,分别在 PINC 为 95%、90%、85% 和 80% 的情况下预测光伏功率区间,并计算指标值评估预测结果。2018 年夏季(6~8 月)与秋季(9~11 月)数据中分别剔除 7 月 1 日、8 月 3 日与 9 月 10 日数据后,作为 3 次区间预测中 ELM 训练集的待选集。对于待预测日的每一个样本单元,计算其与待选训练集中每一个样本单元的加权欧氏距离,选取其中相似度最高的 50 个样本单元,组成 ELM 训练集。

4.2 预测结果

考虑到光照时间,对于 7 月 1 日、8 月 3 日、9 月 10 日 3 个典型日,确定 ELM 训练集后,需要通过交叉验证选取 ELM 隐藏层神经元数量。交叉验证结果如图 3(a)~3(c)所示。由于交叉验证时 ELM 隐藏层输入权重与偏置的生成是随机的,所以得到的预测区间并非最优区间,但评估指标随 ELM 隐藏

层神经元个数增加所呈现的变化趋势可以为隐藏层神经元个数的选取提供重要依据。由图 3 可知,在各情况下,当隐藏层神经元个数达到 20 后,ELM 的预测性能趋于稳定。因此,在算例分析中选用的隐藏层神经元节点数为 20。

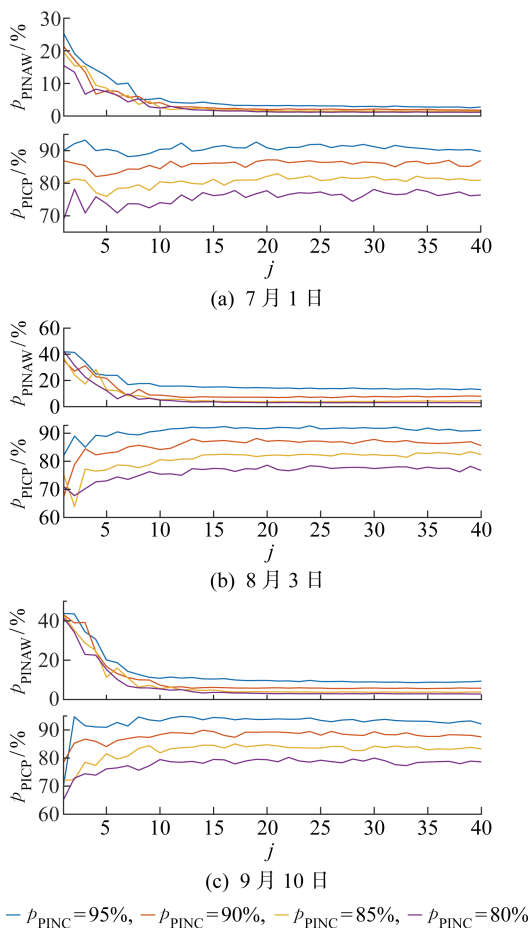


图 3 交叉验证结果

Fig. 3 Results of cross validation

4.2.1 阴雨天区间预测结果 7 月 1 日 6:30—17:00 在 PINC 分别为 95%、90%、85%、80% 的 4 种场景下的光伏电站功率区间预测结果如图 4 所示。7 月 1 日的预测区间评估指标值如表 1 所示。

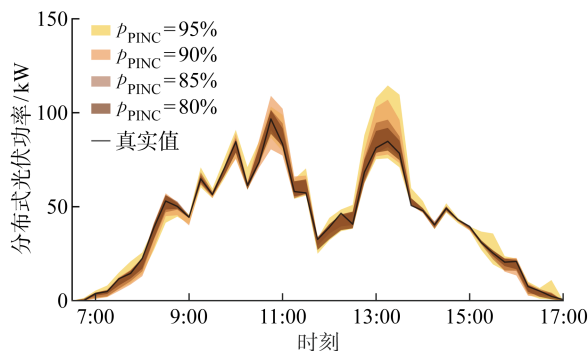


图 4 不同 PINC 下 7 月 1 日预测区间

Fig. 4 Prediction intervals on July 1 at different PINCs

表 1 7 月 1 日预测区间评估指标

Tab. 1 Assessment of prediction intervals on July 1

$p_{PINC}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$
95	95.12	2.69
90	95.12	1.93
85	92.68	1.48
80	87.80	1.18

4.2.2 多云天区间预测结果 8 月 3 日 6:45—17:15 在 PINC 分别为 95%、90%、85%、80% 的 4 种场景下的光伏电站功率区间预测结果如图 5 所示。8 月 3 日的预测区间评估指标值如表 2 所示。

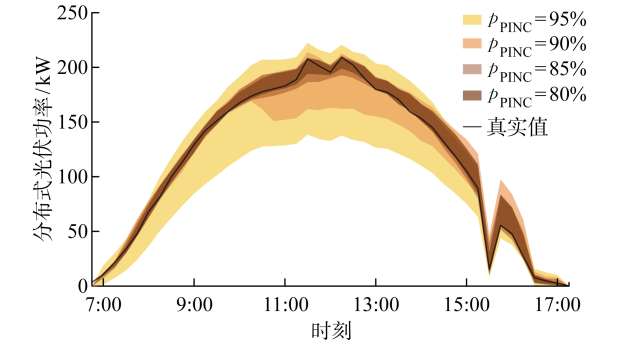


图 5 不同 PINC 下 8 月 3 日预测区间

Fig. 5 Prediction intervals on August 3 at different PINCs

表 2 8 月 3 日预测区间评估指标

Tab. 2 Assessment of prediction intervals on August 3

$p_{PINC}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$
95	100	12.87
90	100	6.23
85	97.56	3.67
80	90.24	3.03

4.2.3 晴天区间预测结果 9 月 10 日 6:15—17:30 在 PINC 分别为 95%、90%、85%、80% 的 4 种场景下的光伏电站功率区间预测结果如图 6 所示。

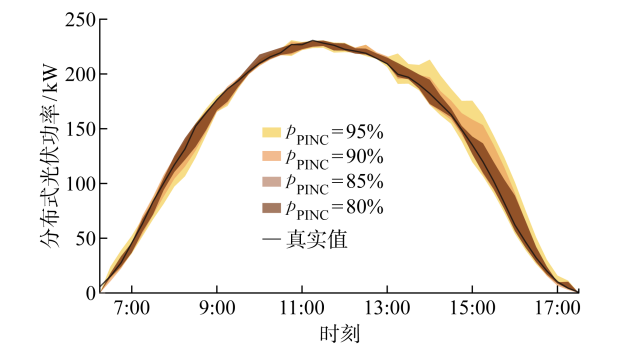


图 6 不同 PINC 下 9 月 10 日预测区间

Fig. 6 Prediction intervals on September 10 at different PINCs

示。9 月 10 日的预测区间评估指标值如表 3 所示。

表 3 9 月 10 日预测区间评估指标

Tab. 3 Assessment of prediction intervals on September 10

$p_{PINC}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$
95	97.73	7.87
90	97.73	4.85
85	95.45	3.36
80	90.91	2.67

4.2.4 光伏功率区间预测结果分析 从图 4~6 可以看出,光伏功率在不同天气类型下呈现出不同特性。阴雨天太阳辐射度较小,光伏功率也较小;多云天的光伏功率也可能具有较大峰值,但由于云层移动和遮挡,光伏功率曲线存在剧烈波动;晴天的光伏功率曲线较为平滑,且峰值较大。在 3 种天气类型下,预测区间上下限的变化趋势和实际光伏功率变化趋势均能保持一致。随着 PINC 减小,图 4~6 中预测区间覆盖率降低,且预测区间宽度也变小。表 1~3 中 PICP 与 PINAW 的数值更加直观地描述了这一现象。数据显示,在 3 种天气类型下,本文方法得到的光伏功率区间预测结果均能满足可信度要求,即 PICP 均高于 PINC。以 8 月 3 日为例, PINC 为 95%及 90%对应的预测区间能够完全覆盖预测时间段内实际的光伏功率点, PICP 高达 100%,而 95%预测区间是 4 种置信水平下最宽的, PINAW 达到 12.87%,意味着预测结果可信度高,但较为保守。将 PINC 由 95%减小为 80%,实际的 PICP 也在逐渐减小;在牺牲预测区间可信度的同时,预测区间准确度提高, PINAW 最小可以低至 3.03%。在 PINC 取值不同的情况下,预测区间的可信度均满足置信水平要求, PICP 明显高于 PINC。

4.3 不同区间预测方法结果对比

为验证本文光伏功率区间预测方法的优越性,使用其他预测方法对同一算例进行光伏功率区间预测。各方法说明如下:

- (1) 方法 I 为本文所提光伏功率区间预测法。
 - (2) 方法 II 使用普通欧氏距离指标对 ELM 训练集进行选取,即式(10)中距离权重恒为 1,其他步骤均与方法 I 相同。
 - (3) 方法 III 不含基于加权欧氏距离指标的 ELM 训练集选取这一步骤,其他步骤均与方法 I 相同。
 - (4) 方法 IV 为 ELM 模型的隐藏层输入权重与偏置随机生成,其他步骤均与方法 I 相同。
- 方法 I 与方法 II、方法 III、方法 IV 的对比可以

分别体现出加权欧氏距离指标、ELM 训练集选取以及隐藏层输入权重与偏置优化对预测区间性能的影响. 考虑到方法 IV 中 ELM 隐藏层输入权重与偏置的随机性,使用方法 IV 进行 10 次 ELM 训练,并取 10 次预测区间评估指标的期望值与其他方法的结果进行对比.

由于历史数据样本庞大,为避免训练时间过长,同时为了控制训练集相同,以排除无关因素的影响,方法 IV 的训练集由本文提出的加权欧氏距离指标确定. 方法 III 的训练集在历史样本单元中随机抽样生成,且训练集样本规模与其他方法相同. 不同方法区间预测结果对比如表 4 所示.

表 4 不同方法区间预测结果对比
Tab. 4 Comparison of interval prediction results obtained from different methods

日期	$p_{PINC}/\%$	方法 I		方法 II		方法 III		方法 IV	
		$p_{PICP}/\%$	$p_{PINAW}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$	$p_{PICP}/\%$	$p_{PINAW}/\%$
7 月 1 日	95	95.12	2.69	95.12	2.78	97.56	7.82	93.90	3.00
7 月 1 日	90	95.12	1.93	92.68	1.95	97.56	4.50	88.78	2.01
7 月 1 日	85	92.68	1.48	92.68	1.60	95.12	4.42	86.83	1.51
7 月 1 日	80	87.80	1.18	87.80	1.21	90.24	3.91	85.37	1.23
8 月 3 日	95	100.00	12.87	100.00	14.31	100.00	14.35	99.76	14.27
8 月 3 日	90	100.00	6.23	100.00	6.95	97.56	7.11	99.27	6.93
8 月 3 日	85	97.56	3.67	95.12	3.71	95.12	4.66	97.07	3.81
8 月 3 日	80	90.24	3.03	90.24	3.11	95.12	3.99	92.44	3.12
9 月 10 日	95	97.73	7.87	100.00	8.52	100.00	22.30	97.50	9.28
9 月 10 日	90	97.73	4.85	100.00	4.97	100.00	8.92	96.14	5.47
9 月 10 日	85	95.45	3.36	93.18	3.64	95.45	8.46	95.68	3.64
9 月 10 日	80	90.91	2.67	88.64	2.77	95.45	5.37	92.56	2.91

方法 I 和方法 II 在各种情况下的对比结果显示,两种方法得到的预测区间 PICP 较为接近,但加权欧氏距离指标与普通欧氏距离指标相比,采用前者选择出的 ELM 训练集可获得更窄的 PINAW,即可有效提高光伏功率区间预测的准确度. 这是因为权重的设置充分考虑不同气象因素与光伏功率之间的相关性大小,选择出的 ELM 训练集与待预测样本相似程度更高.

从方法 I 和方法 III 的对比可得出,两者的 PICP 均能满足置信水平要求,某些情况下方法 III 结果的 PICP 甚至略优于方法 I. 但从准确度角度来看,方法 I 结果的 PINAW 在不同情况下均明显小于方法 III. 此对比结果说明,ELM 训练集筛选可以大大提高预测准确度.

从表 4 数据中可以看出,在可信度与准确度两方面,方法 I 得到的预测区间评估指标在各个情况下均优于方法 IV 结果相应指标的期望值,说明 ELM 隐藏层输入权重与偏置的寻优可以降低随机生成参数给预测结果带来的不稳定性,有效提高预测区间的可信度与准确度.

5 结论

光伏功率区间预测相较于点预测而言,可以提供更加丰富的信息. 所提考虑 ELM 模型优化的光伏功率区间预测技术可对光伏出力区间进行高可信度和准确度的预测.

(1) 加权欧氏距离指标充分考虑光伏功率与气象因素的相关性,在数量庞大的历史数据中筛选出和待预测样本气象因素有较高相似度的样本,减少 ELM 训练时间的同时,可大大提高准确度.

(2) 采用 ESGA 优化 ELM 隐藏层输入权重与偏置参数,消除 ELM 参数生成的随机性给预测结果带来的不确定影响,提高光伏功率区间预测准确性,使得预测模型性能更加稳定.

参考文献:

[1] REN21 Community. Renewables 2021 global status report[R]. Paris: REN21 Secretariat, 2021.
[2] 方晓涛, 严正, 王晗, 等. 考虑概率电压不平衡度越限风险的共享储能优化运行方法[J]. 上海交通大学

- 学报, 2022, 56(7): 827-839.
- FANG Xiaotao, YAN Zheng, WANG Han, *et al.* A shared energy storage optimal operation method considering the risk of probabilistic voltage unbalance factor limit violation[J]. **Journal of Shanghai Jiao Tong University**, 2022, 56(7): 827-839.
- [3] 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.
- [4] 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.
- LAI Changwei, LI Jinghua, CHEN Bo, *et al.* Review of photovoltaic power output prediction technology[J]. **Transactions of China Electrotechnical Society**, 2019, 34(6): 1201-1217.
- [5] LI Y T, SU Y, SHU L J. An ARMAX model for forecasting the power output of a grid connected photovoltaic system[J]. **Renewable Energy**, 2014, 66: 78-89.
- [6] 黄磊, 舒杰, 姜桂秀, 等. 基于多维时间序列局部支持向量回归的微网光伏发电预测[J]. **电力系统自动化**, 2014, 38(5): 19-24.
- HUANG Lei, SHU Jie, JIANG Guixiu, *et al.* Photovoltaic generation forecast based on multidimensional time-series and local support vector regression in microgrids[J]. **Automation of Electric Power Systems**, 2014, 38(5): 19-24.
- [7] PERSSON C, BACHER P, SHIGA T, *et al.* Multi-site solar power forecasting using gradient boosted regression trees[J]. **Solar Energy**, 2017, 150: 423-436.
- [8] SHANG C F, WEI P C. Enhanced support vector regression based forecast engine to predict solar power output[J]. **Renewable Energy**, 2018, 127: 269-283.
- [9] BEHERA M K, NAYAK N. A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm[J]. **Engineering Science & Technology**, 2020, 23(1): 156-167.
- [10] 张倩, 马愿, 李国丽, 等. 频域分解和深度学习算法在短期负荷及光伏功率预测中的应用[J]. **中国电机工程学报**, 2019, 39(8): 2221-2230.
- ZHANG Qian, MA Yuan, LI Guoli, *et al.* Application of frequency domain decomposition and deep learning algorithm in short-term load and photovoltaic power prediction[J]. **Proceedings of the CSEE**, 2019, 39(8): 2221-2230.
- [11] 王岩, 陈耀然, 韩兆龙, 等. 基于互信息理论与递归神经网络的短期风速预测模型[J]. **上海交通大学学报**, 2021, 55(9): 1080-1086.
- WANG Yan, CHEN Yaoran, HAN Zhaolong, *et al.* Short-term wind speed forecasting model based on mutual information and recursive neural network[J]. **Journal of Shanghai Jiao Tong University**, 2021, 55(9): 1080-1086.
- [12] KHOSRAVI A, MAZLOUMI E, NAHAVANDI S, *et al.* Prediction intervals to account for uncertainties in travel time prediction[J]. **IEEE Transactions on Intelligent Transportation Systems**, 2011, 12(2): 537-547.
- [13] KHOSRAVI A, NAHAVANDI S, CREIGHTON D, *et al.* Comprehensive review of neural network-based prediction intervals and new advances[J]. **IEEE Transactions on Neural Networks**, 2011, 22(9): 1341-1356.
- [14] 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(Sup.1): 38-45.
- DONG Lei, ZHOU Wenping, ZHANG Pei, *et al.* Short-term photovoltaic output forecast based on dynamic Bayesian network theory[J]. **Proceedings of the CSEE**, 2013, 33(Sup.1): 38-45.
- [15] VOYANT C, DARRAS C, MUSELLI M, *et al.* Bayesian rules and stochastic models for high accuracy prediction of solar radiation[J]. **Applied Energy**, 2014, 114: 218-226.
- [16] KHOSRAVI A, NAHAVANDI S, SRINIVASAN D, *et al.* Constructing optimal prediction intervals by using neural networks and bootstrap method[J]. **IEEE Transactions on Neural Networks & Learning Systems**, 2015, 26(8): 1810-1815.
- [17] WAN C, XU Z, PINSON P, *et al.* Probabilistic forecasting of wind power generation using extreme learning machine[J]. **IEEE Transactions on Power Systems**, 2014, 29(3): 1033-1044.
- [18] KHOSRAVI A, NAHAVANDI S, CREIGHTON D, *et al.* Lower upper bound estimation method for construction of neural network-based prediction intervals[J]. **IEEE Transactions on Neural Networks**, 2011, 22(3): 337-346.
- [19] 叶林, 马明顺, 靳晶新, 等. 考虑风电-光伏功率相关性的因子分析-极限学习机聚合方法[J]. **电力系统自动化**, 2021, 45(23): 31-40.

- YE Lin, MA Mingshun, JIN Jingxin, *et al.* Factor analysis-extreme learning machine aggregation method considering correlation of wind power and photovoltaic power[J]. **Automation of Electric Power Systems**, 2021, 45(23): 31-40.
- [20] WAN C, LIN J, SONG Y H, *et al.* Probabilistic forecasting of photovoltaic generation: An efficient statistical approach[J]. **IEEE Transactions on Power Systems**, 2017, 32(3): 2471-2472.
- [21] LI S, WANG P, GOEL L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection[J]. **IEEE Transactions on Power Systems**, 2016, 31(3): 1788-1798.
- [22] 杨锡运, 关文渊, 刘玉奇, 等. 基于粒子群优化的核极限学习机模型的风电功率区间预测方法[J]. **中国电机工程学报**, 2015, 35(Sup.1): 146-153.
YANG Xiyun, GUAN Wenyuan, LIU Yuqi, *et al.* Prediction intervals forecasts of wind power based on PSO-KELM[J]. **Proceedings of the CSEE**, 2015, 35(Sup.1): 146-153.
- [23] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: A new learning scheme of feedforward neural networks[C]// **2004 IEEE International Joint Conference on Neural Networks**. Budapest, Hungary: IEEE, 2004: 985-990.
- [24] HUANG G, ZHU Q, SIEW C. Extreme learning machine: Theory and applications[J]. **Neurocomputing**, 2006, 70(1): 489-501.
- [25] KOENKER R. Quantile regression[M]. Cambridge: Cambridge University Press, 2005.
- [26] 丁明, 鲍玉莹, 毕锐. 应用改进马尔科夫链的光伏出力时间序列模拟[J]. **电网技术**, 2016, 40(2): 459-464.
DING Ming, BAO Yuying, BI Rui. Time series simulation of photovoltaic output using improved Markov chain[J]. **Power System Technology**, 2016, 40(2): 459-464.
- [27] 黎敏, 林湘宁, 张哲原, 等. 超短期光伏出力区间预测算法及其应用[J]. **电力系统自动化**, 2019, 43(3): 10-18.
LI Min, LIN Xiangning, ZHANG Zheyuan, *et al.* Interval prediction algorithm for ultra-short-term photovoltaic output and its application[J]. **Automation of Electric Power Systems**, 2019, 43(3): 10-18.
- [28] 代倩, 段善旭, 蔡涛, 等. 基于天气类型聚类识别的光伏系统短期无辐照度发电预测模型研究[J]. **中国电机工程学报**, 2011, 31(34): 28-35.
DAI Qian, DUAN Shanxu, CAI Tao, *et al.* Short-term PV generation system forecasting model without irradiation based on weather type clustering[J]. **Proceedings of the CSEE**, 2011, 31(34): 28-35.

(本文编辑:王历历)