

DOI: 10.19783/j.cnki.pspc.240173

基于频谱注意力和无交叉联合分位数回归的海上风电 功率超短期概率预测

苏向敬^{1,2}, 朱敏轩^{1,3}, 宇海波⁴, 李超杰⁵, 符杨^{1,2}, 米阳³

(1. 教育部海上风电技术工程研究中心(上海电力大学), 上海 200090; 2. 上海电力大学海上风电研究院, 上海 200090; 3. 上海电力大学电气工程学院, 上海 200090; 4. 国网四川省电力公司甘孜供电公司, 四川 甘孜 626000; 5. 新南威尔士大学电气工程与通信学院, 悉尼 2052, 澳大利亚)

摘要: 海上风电功率波动性和随机性强, 其时序数据既存在时间上的全局性关联, 又包含大量局部的短期噪声; 同时, 现有概率预测模型普遍采用的分位数回归方法存在分位数交叉问题, 严重影响了预测结果的精度与合理性。针对上述挑战, 提出了一种基于频谱注意力和无交叉联合分位数回归的海上风电功率超短期概率预测方法。首先, 结合长短期记忆网络(long short-term memory, LSTM)模型和频谱注意力模型, 挖掘风电功率时序数据的全局长期依赖, 并滤除局部噪声。其次, 在此基础上构建了一种基于联合分位数回归的损失函数模型, 通过共享各分位数预测任务间的信息, 自主动态调节各分位数损失的权重, 避免分位数交叉问题。最后通过真实算例分析表明, 相比于现有概率预测模型, 所提模型的锐度指标提升了6%以上, 连续等级概率分数提高了10%以上, 故具有更高的预测精度, 并有效解决了分位数交叉问题。

关键词: 海上风电; 概率预测; 频谱注意力机制; 分位数回归

Ultra-short-term probabilistic forecasting of offshore wind power based on spectral attention and non-crossing joint quantile regression

SU Xiangjing^{1,2}, ZHU Minxuan^{1,3}, YU Haibo⁴, LI Chaojie⁵, FU Yang^{1,2}, MI Yang³

(1. Engineering Research Center of Offshore Wind Technology Ministry of Education (Shanghai University of Electric Power), Shanghai 200090, China; 2. Offshore Wind Power Research Institute, Shanghai University of Electric Power, Shanghai 200090, China; 3. College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China; 4. State Grid Sichuan Ganzi Electric Power Supply Company, Ganzi 626000, China; 5. School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney 2052, Australia)

Abstract: Because of the high fluctuation and random nature of offshore wind power, time-series data contains not only global correlations but also a large amount of short-term local noise. In addition, existing probabilistic forecasting studies are commonly based on quantile regression methods which suffer from the quantile crossing problem and unreliable forecasting results. Therefore, this paper proposes an ultra-short-term offshore wind power probabilistic forecasting method based on a spectral attention and a crossing free joint quantile regression. First, combined with the long short-term memory (LSTM) model and spectral attention model, the global long-term dependence is mined and local noises are filtered in the time-series data. Then, a joint quantile regression loss function model is built to avoid the quantile crossing problem by dynamically adjusting the loss weights and sharing information among all the quantile regression tasks. Real case studies demonstrate that, compared to LSTM and TCN, the proposed model has improved the sharpness of quantile forecasting by over 6% and the continuously ranking probability score by over 10%, with superior prediction accuracy and the quantile crossing problem well addressed.

This work is supported by the Scientific Research and Innovation Program of Shanghai Education Commission (No. 2021-01-07-00-07-E00122).

Key words: offshore wind power; probabilistic prediction; spectral attention mechanism; quantile regression

基金项目: 上海市教育委员会科研创新计划项目资助(2021-01-07-00-07-E00122)

0 引言

风电场建设正在从陆地走向近海乃至深远海,到 2031 年底全球海上风电装机容量将达到 370 GW^[1]。对于负荷重、可调容量不足的沿海地区配网而言,如何消纳高随机性的大规模海上风电成为一大难题^[2-4]。海上风电超短期功率预测通过预测未来较短时间(低于 4 h)内的出力,可有效支撑风电场的运行优化调整、促进海上风电积极可靠消纳^[5-6]。

目前风电出力超短期预测研究主要包括基于物理模型和基于统计模型^[7-8]。其中,海上风电的物理模型在预测时缺乏足够的灵活性和泛化能力。随着风电机组的尺寸和功能不断增大和复杂化,难以对风电场海域的风能分布特征有全面而精确的认识,物理模型的构建也变得更加困难^[9]。相比之下,统计模型利用大量历史数据进行模型训练,可有效避免多维输入和输出间的复杂物理建模过程。早期统计模型预测多采用差分自回归移动平均(autoregressive integrated moving average, ARIMA)等传统方法,这类方法根据历史时间序列的统计特征来推导未来时刻的出力,但应用于分析海上风电这种具有高度不确定性的时间序列数据时,对特征间深层复杂关系挖掘能力不足^[10]。近年来风电功率预测逐步转向深度学习类算法,如人工神经网络(artificial neural network, ANN)^[11]、径向基网络(radial basis function, RBF)、卷积神经网络(convolutional neural network, CNN)^[12]以及递归神经网络(recurrent neural network, RNN)等。深度学习类方法通过神经网络建立输入特征与输出功率间的非线性映射,在应对高随机性新能源出力预测上,精度得到显著提升。其中基于 RNN 改进得到的长短期记忆(long short-term memory, LSTM)与门控循环单元(gate recurrent unit, GRU)模型,可充分挖掘时间序列间的深度依赖关系,在风电预测领域得到了极大关注与应用^[10]。例如:文献[13]利用 LSTM 模型进行风功率预测,相比早期统计模型和机器学习模型,在预测精度和稳定性方面均得到显著提升;文献[14]将时序分解技术与 LSTM 模型相结合,通过时-频域间的转换实现了对预测高频噪声的有效剔除。文献[15]将 LSTM 与注意力机制相结合,提出双重注意力 LSTM(dual-stage attention LSTM, DALSTM)网络,不仅提高了海上风电功率预测精度,而且提升了预测可信度。文献[16]采用时域卷积网络(temporal convolutional network, TCN)模型,并考虑机组状态和尾流效应的影响,提升了海上风电功率预测的精度。

需指出的是,上述研究均为确定性点预测且多

面向陆上风电,而海上风电出力受大气混沌系统的影响具有更显著的不确定性,导致确定性预测误差不可避免。相比之下,概率预测通过提供更全面的预测信息,可更有效支撑运行调度和风险防控方案的提前制定。传统风电概率预测多集中于参数类方法,即将预测误差通过某种概率分布模型进行近似,进而构建出力预测区间。例如:文献[17]假设风速误差服从正态分布,并据此构建了风电出力误差分布的预测模型;文献[18]利用高斯混合模型对风电预测误差进行了概率分析。但风电预测误差具有高度复杂的统计特性,仅靠假定参数模型来描述风电预测的不确定性是非常困难的。

为此,近年来风电概率预测逐渐转向非参数类方法,以降低对先验知识、统计推断或误差分布假设的依赖。典型非参数类概率预测方法包括核密度估计法^[19]和分位数回归法^[20]。其中分位数回归法以预测分位数为目标,通过一组分位数来描述未来时刻功率的概率密度分布。相比于核密度估计法需要估计整个概率密度函数,分位数预测所需的计算资源较少,并可以根据具体需求选择合适的分位点。因此在量化海上风电功率的不确定性时,尤其是在面对经常出现极端值的情况,分位数预测更有优势。例如:文献[21]将 ANN 与分位数回归相结合,提出了分位数回归神经网络(quantile regression neural networks, QRNN),以量化不确定性对风电出力概率预测的影响;文献[22]在分位数回归的基础上,通过深度置信网络来增强概率预测模型对不确定性的量化能力,可以更好地学习风速序列中的高阶非线性和非平稳特征,从而获得更好的性能。

以上模型虽然可以同时预测多个分位数,但容易出现分位数交叉问题。分位数交叉是指模型在同时预测多个分位数时,输出的较小分位点的预测值比较大分位点的预测值大,打破了累积分布函数单调不递减的基本性质,从而导致分位数估计极不合理^[23]。针对上述问题:文献[23]基于极限学习机与分位数回归,提出了直接分位数风电概率预测模型(direct quantile regression, DQR),将复杂的基于人工神经网络的非参数概率预测问题转化为线性规划问题,可以精确地逼近各种比例递增的分位数;同时通过约束可以有效缓解分位数的交叉问题。文献[24]提出了一种基于约束并行 LSTM 分位数回归模型,通过添加人工约束来缓解分位数交叉问题。上述研究虽然可以得到分位数预测结果,但无法保证在无交叉的情况下同时预测一组分位数^[25],而添加人工约束来解决分位数交叉会降低预测结果的合理性。

综上所述,现有海上风电功率概率预测虽取得

一定的效果, 但仍普遍存在以下不足: 1) 由于风电功率的波动性和随机性很强, 与陆上风电场相比, 海上风电场通常具有更稳定且更高强度的风速, 且更易遭受突发极端天气、高盐高湿度等恶劣环境的影响, 从而造成测量数据的误差表现出不稳定的特性, 其时序数据既存在全局性长期历史关联, 又包含大量局部噪声。现有模型虽有好的时序特征提取能力, 但其难以挖掘时间样本中全局长期依赖与局部噪声之间的关系; 2) 现有分位数预测模型易受不同目标任务损失量级差异的影响, 造成各任务损失收敛速度不一, 部分任务欠拟合而部分过拟合, 导致分位数交叉问题, 降低了分位数预测的可信度与精度。

针对上述挑战, 本文提出了一种基于频谱注意力和联合分位数回归(spectral attention and joint quantile regression, SA-JQR)的超短期海上风电概率预测模型。具体来说, 该方法在 LSTM 模型的基础上添加了频谱注意力模型, 通过引入频域上的注意力机制有效挖掘时间样本中全局长期历史信息与局部短时噪声的权重。在此基础上, 提出了基于多任务学习的联合分位数损失, 使模型可同时预测一组分位数, 并通过动态调整各损失的权重来平衡各任务损失的下降速度, 解决了分位数交叉问题。最后, 基于东海大桥海上风电场真实数据进行了算例验证与分析。

1 基于 SA-JQR 的海上风电概率预测模型

为了弥补现有模型对于输入时间窗口内的全局信息提取和局部噪声信息剔除能力的不足, 本文提出一种频域内的注意力模型——频谱注意力模型(SA-Model), 其结合注意力机制和 LSTM 编码单元隐藏状态向量的频域信息, 将输入样本特征的全局和局部频域关键信息纳入分位数预测过程。同时, 为预测多个互不交叉的分位数, 本文提出一种联合分位数损失模型(JQR-Model), 其接收频谱注意力模型输出的多个预测分位数, 计算带有权重参数的联合分位数损失函数, 进而通过梯度下降算法更新模型参数与权重。

1.1 频谱注意力模型 SA-Model

与使用多头自注意力机制对原时序信号的全局依赖模式进行捕捉的传统时序注意力模型相比, 所提 SA-Model 通过两个具有不同权重全局、局部信息提取模块来确定预测结果的全局模式, 并消除局部背景噪声, 其计算均在频域中进行。

1.1.1 注意力机制

作为一种资源分配策略, 注意力机制可自适应地为模型输入信息分配注意力权重, 以评估各输入信息对输出的贡献度, 并突出重要输入信息的表达。

鉴于注意力机制在机器翻译、语音识别等领域取得的显著效果, 本文将其引入以关注有效的频域信息, 并提高预测精度和效率。

本文旨在频域内提取关键信息, 故提出一种频域内的注意力机制, 其计算过程如图 1 所示。

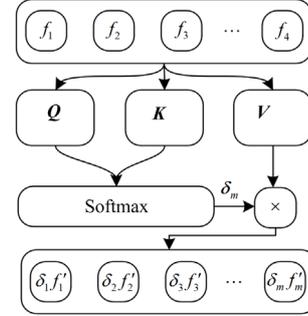


图 1 注意力计算过程

Fig. 1 Attention calculation process

输入为频域信息 $F_m = \{f_1, f_2, f_3, \dots, f_m\}$, m 为频域分量个数, f_i 为第 i 个初始频域分量。各分量对应的注意力权重 δ_m 为

$$\delta_m = V \times \text{Softmax}(Q \times K) \quad (1)$$

$$Q = \text{MLP}_Q(F_m) \quad (2)$$

$$K = \text{MLP}_K(F_m) \quad (3)$$

$$V = \text{MLP}_V(F_m) \quad (4)$$

式中: Q 、 K 、 V 是由三个多层感知机(multilayer perceptron, MLP) MLP_Q 、 MLP_K 、 MLP_V 训练得出的矩阵; $\text{Softmax}(\cdot)$ 为激活函数。借助式(1)得出的注意力权重 δ_m 对各频域分量进行加权, 可得到综合频域信息, 如式(5)所示。

$$F'_m = \text{Att}(F_m) = \delta_m \odot F_m = \{\delta_1 f'_1, \delta_2 f'_2, \delta_3 f'_3, \dots, \delta_m f'_m\} \quad (5)$$

式中: $\text{Att}(\cdot)$ 为注意力网络; F'_m 为综合频域信息; f'_i 为经过注意力网络处理过的第 i 个频域分量; \odot 表示哈达马乘积运算。通过对输入信息各频域分量分配注意力权重并做动态调整, 有助于增强关键信息表达。

1.1.2 频谱注意力模型

频谱注意力模型由特征输入与编码层、信息提取层、信息融合层和预测输出层组成, 如图 2 所示。

1) 特征输入与编码层

特征输入与编码层由输入模块与 LSTM 编码模块组成。输入模块中, 采用滑动窗口采样法获取输入样本。设训练集多维特征的时序数据为 $X = \{x_1^i, x_2^i, \dots, x_t^i, \dots, x_T^i\}_{i=1}^N$, T 为序列长度, N 为特征数, x_t^i 为 t 时刻第 i 个特征的值, $t \in \{1, 2, \dots, T\}$, $i \in \{1,$

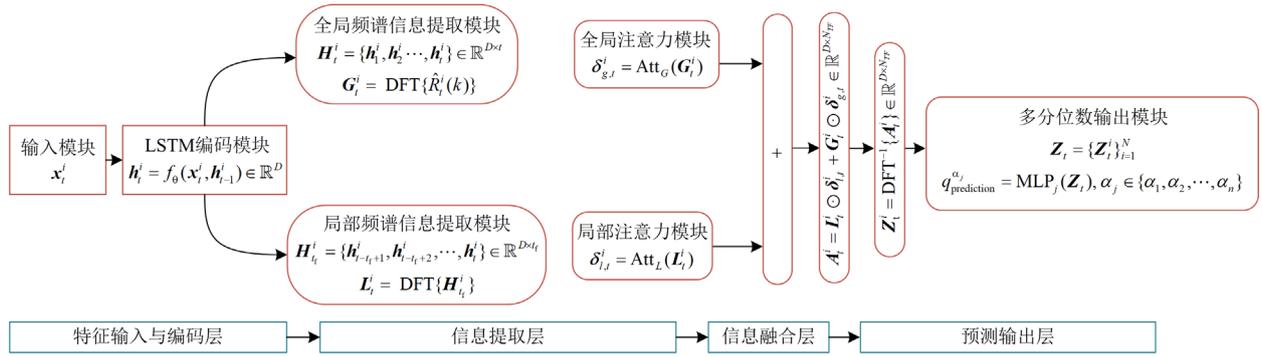


图2 频谱注意力模型结构

Fig. 2 Spectral attention model structure

$2, \dots, N\}$ 。设历史时间窗口长度为 t_0 ，滑动步长为 d ，输入样本可以表示为 $\mathbf{S}_1 = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{t_0}^i\}$ ， $\mathbf{S}_2 = \{\mathbf{x}_{1+d}^i, \mathbf{x}_{2+d}^i, \dots, \mathbf{x}_{t_0+d}^i\}$ ，以此类推。

LSTM 编码模块首先初始化隐藏状态 \mathbf{h}_0^i ，然后使用一个参数为 θ 的嵌入函数 $f_\theta(\cdot)$ 接收输入数据，并编码所有时间步长上的隐藏状态信息，将其转换到嵌入空间中，具体计算过程为

$$\mathbf{h}_t^i = f_\theta(\mathbf{x}_t^i, \mathbf{h}_{t-1}^i) \in \mathbb{R}^D, t = 1, 2, \dots, t_0 \quad (6)$$

式中： \mathbf{h}_{t-1}^i 与 \mathbf{h}_t^i 分别表示上一时刻与当前时刻的隐藏状态； D 表示 \mathbf{h}_t^i 的维度。在输入时间窗口内的每一个时刻计算式(6)，即可得到窗口内所有时间步长上的隐藏状态 $\{\mathbf{h}_1^i, \mathbf{h}_2^i, \dots, \mathbf{h}_{t_0}^i\}_{i=1}^N$ 。

2) 信息提取层

现有模型难以挖掘时序样本全局长期依赖与局部噪声之间的关系，为此所提模型在编码后的隐藏状态基础上进行信息提取。信息提取层由全局和局部频谱信息提取模块和相对应的注意力模块组成。

全局频谱信息提取模块使用隐藏状态的功率谱密度(power spectral density, PSD)表示全局频谱信息 \mathbf{G}_t^i ，功率谱密度用于描述信号在频域上的能量分布情况，它可以作为信号的全局表征^[25]。对于隐藏状态的平均自相关函数作离散傅里叶变换，即可得到当前时刻功率谱密度 $P_{\text{SD}_t}^i$ 的计算公式。

$$\hat{R}_t^i(k) = \frac{1}{B} \sum_{b=0}^{B-1} \left[\frac{1}{D} \sum_{d=1}^D \left(\frac{1}{t_0-1} \sum_{i=1}^{t_0-1} h_t^i(d) \cdot h_{t+k}^i(d) \right) \right] \quad (7)$$

$$\mathbf{G}_t^i = P_{\text{SD}_t}^i = \text{DFT}\{\hat{R}_t^i(k)\} \quad (8)$$

式中： $\hat{R}_t^i(k)$ 为当前时刻第 i 个时间序列的平均自相关函数； $h_t^i(d)$ 与 $h_{t+k}^i(d)$ 分别为当前时刻与 k 个时间步后隐藏状态的 d 维数据，其中， k 为滞后长度， $d \in D$ ； B 为训练样本批次大小； $\text{DFT}\{\cdot\}$ 表示离散

傅里叶变换函数。

局部频谱信息提取模块对最后 t_f 个时刻的隐藏状态 $\mathbf{H}_{t_f}^i$ 进行离散傅里叶变换，即可得到局部频谱信息 $\mathbf{L}_{t_f}^i$ ，其中： $t_f = N_{\text{TF}}$ ， $\mathbf{H}_{t_f}^i = \{\mathbf{h}_{t-t_f+1}^i, \mathbf{h}_{t-t_f+2}^i, \dots, \mathbf{h}_t^i\} \in \mathbb{R}^{D \times t_f}$ 。

$$\mathbf{L}_{t_f}^i = \text{DFT}\{\mathbf{H}_{t_f}^i\} \quad (9)$$

为提取长期历史关联、减少短期冗余噪声干扰，所提模型采用注意力机制筛选关键频域分量，以上获得的全局和局部频谱信息经两个对应的注意力模块进行信息处理。其中，全局注意力模块使用全局注意力网络 $\text{Att}_G(\cdot)$ 选择 \mathbf{G}_t^i 中的全局信息；局部注意力模块使用局部注意力网络 $\text{Att}_L(\cdot)$ 选择 $\mathbf{L}_{t_f}^i$ 中与预测目标相关和不相关的局部频率分量，各模块的注意力系数计算分别如式(10)、式(11)所示。

$$\delta_{g,t}^i = \text{Att}_G(\mathbf{G}_t^i) \quad (10)$$

$$\delta_{l,t}^i = \text{Att}_L(\mathbf{L}_{t_f}^i) \quad (11)$$

式中： $\delta_{g,t}^i$ 为全局注意力系数矩阵； $\delta_{l,t}^i$ 为局部注意力系数矩阵，其元素取值范围均为 $[0,1]$ 。所提模型通过动态调整全局、局部注意力系数来整合全局与局部频谱信息，可有效挖掘时序样本全局长期依赖与局部噪声间的关系，降低了冗余噪声在预测中的占比和影响。

3) 信息融合层

信息融合层使用式(12)和式(13)整合全局与局部频谱信息，再通过离散傅里叶逆变换将其转换到时域内。

$$\mathbf{A}_t^i = \mathbf{L}_{t_f}^i \odot \delta_{l,t}^i + \mathbf{G}_t^i \odot \delta_{g,t}^i \in \mathbb{R}^{D \times N_{\text{TF}}} \quad (12)$$

$$\mathbf{Z}_t^i = \text{DFT}^{-1}\{\mathbf{A}_t^i\} \in \mathbb{R}^{D \times N_{\text{TF}}} \quad (13)$$

式中： \mathbf{A}_t^i 与 \mathbf{Z}_t^i 分别为全局与局部融合信息的频域和时域形式； $\text{DFT}^{-1}\{\cdot\}$ 表示离散傅里叶逆变换函数。

4) 预测输出层

为了使模型能够抓取到更多的非线性特征和组合特征的信息, 预测输出层使用多个多层感知机并行预测多个分位数。

$$q_{\text{prediction}}^{\alpha_j} = \text{MLP}_j(\mathbf{Z}_t), \alpha_j \in \{\alpha_1, \alpha_2, \dots, \alpha_n\} \quad (14)$$

式中: $q_{\text{prediction}}^{\alpha_j}$ 为第 j 个预测分位数; $\mathbf{Z}_t = \{\mathbf{Z}_t^i\}_{i=1}^N$ 为综合频域信息; n 为预测分位数的个数; α_j 为第 j 个预测分位点; $\text{MLP}_j(\cdot)$ 为第 j 个多层感知机。

1.2 多分位数损失模型 JQR-Model

针对当前分位数预测模型普遍存在的分位数交叉问题, 本文提出一种基于多任务学习的联合分位数损失模型, 通过训练数据自主学习最优权重组合, 平衡各任务损失下降速度, 避免分位数交叉问题。

本文定义一个高斯似然模型^[26]来估计 t 时刻 α 分位数点的分位数 $y_{t(\alpha)}$, 其均值为 $\hat{y}_{t(\alpha)}$, 方差为 σ 。

$$p(y_{t(\alpha)} | \hat{y}_{t(\alpha)}) = N(\hat{y}_{t(\alpha)}, \sigma^2) \quad (15)$$

根据分位数回归的定义预测不同标准分位点下的预测分位数, 需要使用不同的损失函数, 故频谱注意力模型输出的多个预测分位数需要输入一个损失模型计算各个预测的总损失, 继而使用梯度下降算法更新频谱注意力模型的参数。本文把模型预测输出多个分位数看成一个多任务学习的过程, 传统计算多任务总损失 L_{sum} 的方法是对每个任务的损失进行加权线性求和。

$$L_{\text{sum}} = \sum_{e=1}^E \omega_e L_e \quad (16)$$

式中: E 为任务的个数; ω_e 为第 e 个任务的权重; L_e 为第 e 个任务的损失。基于上述理论框架, 对多个分位数 $\alpha_1, \alpha_2, \dots, \alpha_n$ 的似然估计取负对数, 即等价于联合分位数回归损失 L_{JQR} 。

$$\begin{aligned} & -\log p(y_{t(\alpha_1)}, y_{t(\alpha_2)}, \dots, y_{t(\alpha_n)} | \hat{y}_{t(\alpha_1)}, \hat{y}_{t(\alpha_2)}, \dots, \hat{y}_{t(\alpha_n)}) \propto \\ & \sum_{j=1}^n (1/(2\sigma_j^2)) \cdot \sum_{t=1}^T \rho_{\alpha_j} (y_t - \hat{y}_{t(\alpha_j)}) = \\ & \sum_{j=1}^n (1/(2\sigma_j^2)) \cdot L_{\text{QR}(j)} = L_{\text{JQR}} \end{aligned} \quad (17)$$

式中: $j \in 1, 2, \dots, n$; $L_{\text{QR}(j)}$ 为第 j 个分位数的损失; σ_j 为衡量各分位数预测任务的权重, 用于权衡单任务在总任务中的重要性。本文将权重 σ_j 设为网络模型参数之一, 参与模型训练和迭代更新。在分位数联合估计的过程, JQR-Model 通过训练数据自主学

习最优权重组合, 实现不同任务间的信息共享, 避免分位数交叉问题, 改善多分位数预测的精度。

2 风电功率概率预测流程与评价指标

2.1 基于 SA-JQR 模型的风电功率概率预测流程

基于所提频谱注意力和联合分位数损失模型的功率概率预测流程如图 3 所示, 其流程分为数据准备、模型训练与模型测试三个阶段。

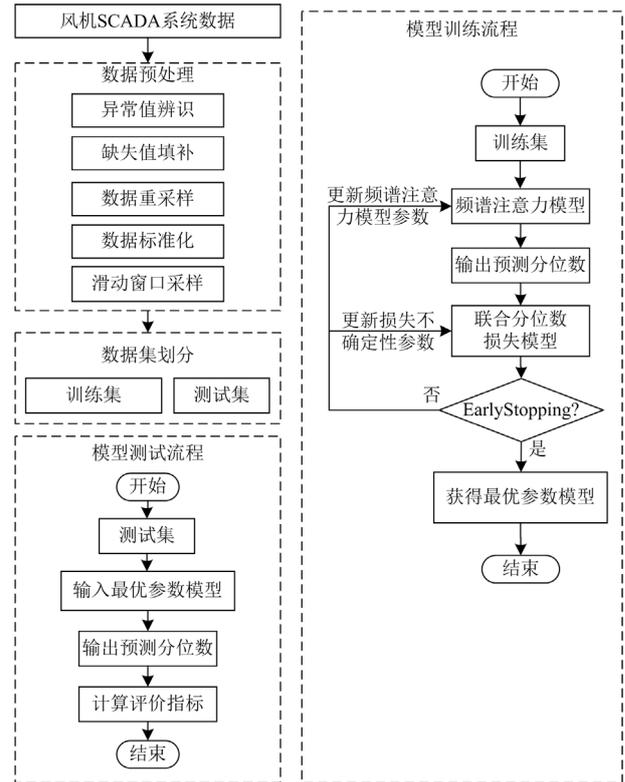


图 3 所提概率预测模型功率预测流程

Fig. 3 Flowchart of the proposed probabilistic prediction model

2.1.1 数据准备阶段

1) 异常值辨识与缺失值填补

海上风电场气象复杂、环境恶劣, 风机 SCADA 系统易出现传感器故障或通信网络时延等问题, 导致数据出现异常值和缺失值。本文通过 K-means 聚类方法结合风速-功率曲线对不满足风速功率曲线的数据进行识别。风速功率曲线是以风速为横坐标, 功率为纵坐标所拟合的一条近似曲线, 是衡量机组出力性能的重要指标之一^[27-29]。本文基于临近风电机组间的空间相关性对目标风电机组缺失值进行填充, 即通过探查除某时刻目标受损风机外其余风机的数据是否存在, 接着选择数据缺失时数据存在且与目标风机距离最近的风机数据进行填充; 若此时刻所有风机的数据都缺失, 则采用去年此时、上月

此时或上周此时的数据进行填充。

2) 数据重采样与标准化

风机 SCADA 系统收集数据的时间间隔与预测步长不匹配, 故需要进行数据重采样; 同时为避免梯度爆炸和消失, 加快网络训练收敛速度, 需要进行数据归一化。为避免数据异常值的影响, 采用 RobustScaler 方法对其进行标准化, 转化函数为

$$x^* = \frac{x - m_{\text{median}}}{I_{\text{QR}}} \quad (18)$$

式中: x 、 x^* 分别为标准化前后的数据; m_{median} 为样本中位数; I_{QR} 为样本四分位距即第 1 个四分位数(25%分位数)和第 3 个四分位数(75%分位数)之间的范围。

3) 滑动窗口采样与数据集划分

本文使用的样本来自 SCADA 系统多维特征的时序数据, 因而采用滑动窗口采样法制取样本集, 滑动窗口采样法详见本文第 1.1.2 节, 不再赘述。

2.1.2 模型训练阶段

模型训练阶段先采用 Xavier 初始化各层网络参数。Xavier 初始化方法旨在保持激活函数方差在前向和反向传播过程中大致相同, 避免梯度消失或梯度爆炸, 某层网络权重 w 的初始方差为

$$\text{Var}(w) = \frac{2}{C_{\text{in}} + C_{\text{out}}} \quad (19)$$

式中: C_{in} 、 C_{out} 分别为该层网络输入与输出神经元个数; w 为网络权重, 其服从均值为 0、方差为 $\text{Var}(w)$ 的高斯分布中随机抽取。

接着将训练集样本输入频谱注意力模型, 计算多个预测分位数, 再将预测分位数输入联合分位数损失模型, 得到联合分位数损失。以联合分位数损失最小为训练目标, 利用 Adam 优化器不断更新频谱注意力模型和损失模型参数。为获得更好的泛化性能, 流程使用 EarlyStopping 机制作为判断训练停止的条件, 当模型在训练集上表现开始下降时停止训练, 从而避免持续训练导致过拟合问题。

2.1.3 模型测试阶段

模型测试阶段将测试集样本输入最优参数的频谱注意力模型计算多个预测分位数, 接着计算概率预测评价指标以评估所提模型的精确度和性能。

2.2 预测性能评价指标

概率预测常用性能评价指标, 包括可靠性指标、锐度指标和总体性能指标^[30]。为了量化模型在预测多个分位数时出现分位数交叉现象的概率, 本文提出了一种分位数交叉率指标。

1) 可靠性指标

可靠性指标是概率预测的主要评估指标, 衡量的是预测概率分布与预测对象实际分布的偏差^[31], 可靠性低将会增加电网运行决策风险。当以分位数形式预测概率分布时, 通常采用平均覆盖率偏差(average coverage deviation, ACD)来衡量预测结果可靠性, 其定义为

$$E_{\text{ACD}}^{(\alpha)} = \alpha - b^{(\alpha)} \quad (20)$$

$$b^{(\alpha)} = \frac{1}{T} \sum_{t=1}^T c_t^{(\alpha)} \quad (21)$$

$$c_t^{(\alpha)} = \begin{cases} 1, & y_t \leq \hat{y}_{t(\alpha)} \\ 0, & y_t > \hat{y}_{t(\alpha)} \end{cases} \quad (22)$$

式中: $E_{\text{ACD}}^{(\alpha)}$ 为分位点 α 上的平均覆盖率偏差; $c_t^{(\alpha)}$ 为 t 时刻的示性函数; $b^{(\alpha)}$ 为示性函数在时间样本长度上的均值。在给定分位点上的 $E_{\text{ACD}}^{(\alpha)}$ 越接近 0, 预测可靠性越高。

2) 锐度指标

锐度指标利用预测区间的平均宽度(average width, AW)反映预测概率分布的集中程度^[32]。在相同的前提下, 过宽预测区间将增加应对不确定性所带来的决策成本, 降低运行经济性^[33]。给定概率预测区间的平均宽度 E_{AW} 可由式(23)计算得到。

$$E_{\text{AW}} = \frac{1}{T} \sum_{s=1}^T (q_{s(\alpha_{\text{up}})} - q_{s(\alpha_{\text{down}})}) \quad (23)$$

式中: $q_{s(\alpha_{\text{up}})}$ 和 $q_{s(\alpha_{\text{down}})}$ 分别表示预测区间的上界和下界; α_{up} 与 α_{down} 分别表示给定预测区间上下边界的分位点。 E_{AW} 越小说明预测结果的锐度性能越好。

3) 综合性能指标

综合性能指标用于评价概率预测整体性能。连续等级概率分数(continuous ranked probability score, CRPS)综合考虑了可靠性和锐度两个指标, 可有效评价概率预测整体性能, 在风电概率预测中得到了广泛应用^[34]。 E_{CRPS} 定义为

$$E_{\text{CRPS}} = \frac{1}{T} \sum_{s=1}^T \left(E(|Y - z|) - \frac{1}{2} E(|Y - Y'|) \right) \quad (24)$$

式中: z 为服从 $F_s(z)$ 的随机变量, 其中 $F_s(z)$ 为 t 时刻的预测概率分布; Y 和 Y' 为累积分布函数 $F_t(z)$ 中采样的两个独立随机变量; $E(\cdot)$ 为求期望函数。

E_{CRPS} 越小说明模型的综合性能越好。

4) 分位数交叉率指标

分位数交叉率指标反映了多个预测概率区间的交叉程度。在同时预测多个分位数时, 输出的较小分位点的预测值比较大分位点的预测值大, 不符合累积分布函数单调不递减的基本性质, 从而导致

分位数估计的极不合理^[23]。分位数交叉率 P 统计了模型预测分位数出现交叉的次数占测试集时间样本上的比例, 可由式(25)计算得到。

$$E_p = \frac{1}{T} \sum_{t=1}^T p(t) \quad (25)$$

$$p(t) = \begin{cases} 1, & \hat{y}_{t(\alpha_u)} < \hat{y}_{t(\alpha_v)} \\ 0, & \hat{y}_{t(\alpha_u)} \geq \hat{y}_{t(\alpha_v)} \end{cases} \quad (26)$$

式中: $p(t)$ 为 t 时刻的示性函数; $u, v \in \{1, 2, \dots, n\}$; $\hat{y}_{t(\alpha_u)}$ 与 $\hat{y}_{t(\alpha_v)}$ 分别为 t 时刻模型在分位点 α_u 与 α_v 上的预测分位数, 且 $\alpha_u > \alpha_v$ 。当 t 时刻出现分位数交叉现象时, $p(t)$ 值为 1, 否则为 0。 E_p 为 0 表示模型预测结果无分位数交叉问题。

3 案例分析

3.1 验证案例与数据准备

为验证本文所提概率预测模型的有效性与优越性, 基于国内某海上风电场真实数据开展算例分析。该风电场项目一期共安装了 34 台单机容量为 3 MW 的离岸型风机, 风电场经 4 回 35 kV 海底电缆接入岸上 110 kV 升压变电站, 进而接入城市电网。海上风电 SCADA 系统数据采集频率为 5 s, 集合了气象(如 90 m 高度风速、风向、温度)、机组状态(如机舱温度、电机绕组温度、齿轮箱温度、故障代码)等内外部特征。风电场常年风速为 3.5~25 m/s, 一期风电机组额定风速为 13 m/s, 由于实际工况超出额定风速较多, 所以本文特别考虑了风电机组的桨距角等变桨系统特征。本文通过相关性分析进行特征筛选, 选取气象、机组状态和变桨系统等 18 个特征共同构成了模型训练所需输入特征, 如表 1 所示。

表 1 模型训练输入特征集

Table 1 Input feature set for model training

特征	单位	特征	单位
齿轮箱油温、轴温	°C	环境温度	°C
发电机、叶轮转速	r/s	风速	m/s
有功	kW	无功	kvar
变桨力矩 1、2、3	N·m	变桨角度 1、2、3	(°)
机舱温度	°C	风向 1、2、3	—

在模型参数方面, 输入序列的历史时间窗口大小为 48, 输出为 15 min 后单风机有功功率的 10 个分位点。本文采用 Adam 优化器进行模型参数优化, 模型超参数随机搜索次数设置为 100, 其最优值如表 2 所示, 模型各层数据维度如表 3 所示。实验基于 Intel i9-11900k 处理器、64 GB 内存、Nvidia

Geforce RTX3090 硬件平台, 运行环境为基于 Python 的 Pytorch1.11。

表 2 模型超参数设置

Table 2 Model hyperparameters configuration

模型超参数	数值
LSTM 编码模块隐藏层神经元个数 hidden_dim	51
LSTM 网络层数 num_layers	1
LSTM 网络 dropout_rate	0.429
学习率 lr	0.0056
训练批次 Batch_size	64

表 3 模型各层数据维度

Table 3 Data dimensions of each layer

模型层	输入维度	输出维度
输入层	(64, 48, 18)	(48, 64, 18)
LSTM 编码层	(48, 64, 18)	(1, 64, 51)
信息提取层	(1, 64, 51)	(1, 64, 51)
信息融合层	(1, 64, 51)	(1, 64, 51)
预测输出层	(1, 64, 51)	(64, 10, 1)

在数据准备方面, 采用 2019 年 2 月—2020 年 12 月的海上风电单风机 SCADA 数据。为兼容国家能源局关于风电功率超短期预测技术规定, 对数据以 15 min 步长进行了重采样。训练集与测试集比例为 7:3, 训练集样本数为 6874 个, 测试集样本数为 2946 个。在异常值辨识与缺失值填补方面, 图 4 为该风电场 1 号风机风速功率曲线, 红色点代表异常数据, 蓝色代表正常数据, 低于切入或超出切出风速的数据也展示为红色。对上述识别出的红色数据不是正常发电情况下的数据进行了剔除, 并基于临近风电机组间的空间相关性对目标风电机组缺失值进行了填充。

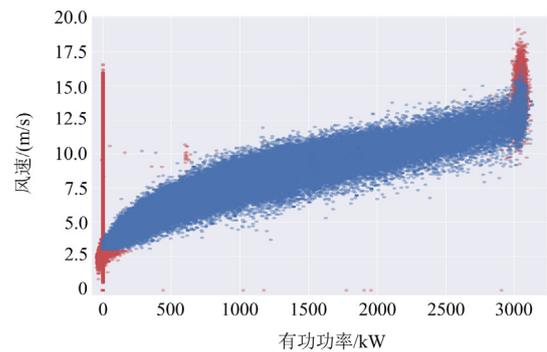


图 4 该风电场 1 号风机风速功率曲线

Fig. 4 The No.1 wind turbine wind speed-power curve

3.2 概率预测结果分析

在模型设置方面, 基于相同数据集与现有预测模型进行了对比分析: 如 LSTM^[14]、GRU^[10]、TCN^[16]、

双重注意力长短期记忆神经网络(DALSTM)^[15]和DQR^[23]模型。DALSTM网络是在LSTM神经网络的基础上引入特征和时序双重注意力机制,同时对输入特征间的潜在关联和历史时刻蕴含的有效信息进行挖掘。

在模型输出方面,使用各对比模型对测试集上0.05、0.1、0.2、0.3、0.4、0.6、0.7、0.8、0.9、0.95分位点的单步(15 min)风电有功功率进行预测,然后通过[0.4,0.6]、[0.3,0.7]、[0.2,0.8]、[0.1,0.9]、[0.05,0.95]构建置信度分别为20%、40%、60%、80%和90%的置信区间,并计算评估指标。

3.2.1 验证所提频谱注意力机制的有效性

为了验证所提频谱注意力机制的有效性,将基于LSTM编码模块的LSTM-SA-QR模型与考虑时序注意力机制的DALSTM-QR模型进行对比,同时对比分析了LSTM-QR、GRU-QR、TCN-QR模型的预测精度。

表4—表6展示了各模型概率预测结果。由表4可知,各模型在各分位数上的可靠性指标均接近0,故其均具有良好的预测可靠性。由表5可知,LSTM-SA-QR在不同置信区间下的性能均最佳,以置信度为80%的置信区间为例,LSTM-SA-QR模型的锐度指标相比LSTM-QR、GRU-QR、TCN-QR、DALSTM-QR分别降低了23.27%、21.31%、10.86%和7.06%。由表6可知:LSTM-SA-QR模型的综合性能指标最小,相较于LSTM-QR、GRU-QR、TCN-QR、DALSTM-QR分别降低了32.62%、26.88%、16.07%、10.14%;各模型的分位数交叉率均不为0,表示各模型均出现了分位数交叉现象。

表4 各模型概率预测结果可靠性对比

Table 4 ACD of probabilistic predicting models

模型	$E_{ACD}^{(0.2)}$	$E_{ACD}^{(0.4)}$	$E_{ACD}^{(0.6)}$	$E_{ACD}^{(0.8)}$	$E_{ACD}^{(0.95)}$
LSTM-QR	0.132	0.159	-0.173	-0.085	0.142
GRU-QR	0.038	-0.051	0.057	-0.089	-0.094
TCN-QR	0.032	-0.068	0.051	-0.054	0.097
DALSTM-QR	0.074	0.062	-0.097	0.083	0.061
LSTM-SA-QR	-0.038	0.054	0.092	-0.058	0.076

表5 各模型概率预测结果锐度对比

Table 5AW of probabilistic predicting models

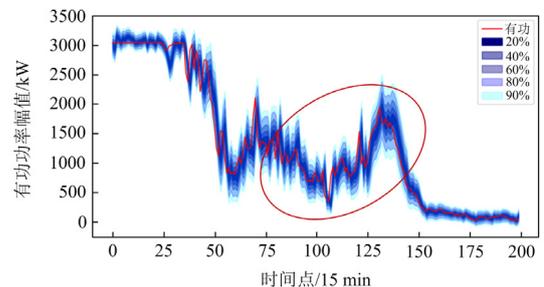
模型	E_{AW}				
	20%	40%	60%	80%	90%
LSTM-QR	238.36	358.41	457.38	485.35	569.35
GRU-QR	216.84	326.29	428.37	473.23	546.86
TCN-QR	219.46	317.36	385.42	417.74	473.89
DALSTM-QR	194.26	305.95	374.32	400.68	442.57
LSTM-SA-QR	152.62	268.84	328.98	372.37	415.69

表6 各模型概率预测结果综合性能与分位数交叉率对比

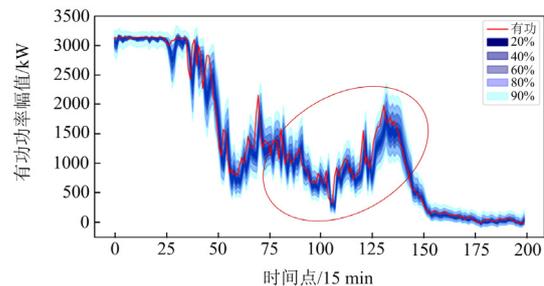
Table 6 CRPS and P of probabilistic predicting models

模型	E_{CRPS}	E_p
LSTM-QR	93.92	0.081
GRU-QR	86.54	0.055
TCN-QR	75.40	0.103
DALSTM-QR	70.42	0.117
LSTM-SA-QR	63.28	0.098

为了对比本文所提频谱注意力机制和时序上的注意力机制,图5展示了DALSTM-QR模型和LSTM-SA-QR模型在测试集中有功功率幅值与频率波动较为剧烈的200个时间样本上的预测结果。由图5可以看出,在各个置信度下,LSTM-SA-QR模型给出的预测区间均较窄。在风功率波动程度较为剧烈时,如图5中红圈处所示,DALSTM-QR模型各预测区间较宽。具体来说,在置信度分布为20%、40%、60%、80%和90%的置信区间上,LSTM-SA-QR模型相比于DALSTM-QR模型的AW分别降低了27.28%、13.80%、13.78%、7.60%和6.47%,且CRPS也降低了10.14%。这表明频谱注意力机制对于频率波动识别更加精确,不仅预测区间锐度高,而且综合性能更强。



(a) DALSTM-QR模型



(b) LSTM-SA-QR模型

图5 两种注意力机制的预测区间对比

Fig. 5 Comparison of prediction intervals between two attention mechanisms

综上,LSTM-SA-QR模型不仅相比传统时序预测模型具有更高的性能,时序上的注意力机制还可

以更精确地捕捉序列中的全局与局部频率信息, 从而得出更为准确的预测区间。但其在同时预测多个分位数时, 仍无法避免出现分位数交叉现象。

3.2.2 验证所提联合分位数损失模型的有效性

为了验证所提联合分位数损失模型对于分位数预测精度的提升和分位数交叉的抑制作用, 分别在 LSTM-QR、GRU-QR、TCN-QR、LSTM-SA-QR 模型中添加 JQR 模型, 将其与未加入 JQR 模型和 DQR 模型进行对比。

表 7—表 9 为各模型概率预测结果。由表 7 可知, 所有模型在各分位数上的 ACD 均接近 0, 故各模型均具有良好的预测可靠性。由表 5 和表 7 可知, 添加了 JQR 模型后, LSTM-JQR、GRU-JQR、TCN-JQR、LSTM-SA-JQR 模型的 AW 比原始模型均有不同程度的降低。以置信度为 80% 的置信区间为例, LSTM-JQR、GRU-JQR、TCN-JQR、LSTM-SA-JQR 模型相比 LSTM-QR、GRU-QR、TCN-QR、LSTM-SA-QR 的 AW 分别降低了 2.65%、4.08%、3.09% 和 4.4%。由表 6 和表 9 可知, 添加了 JQR 模型后, LSTM-JQR、GRU-JQR、TCN-JQR、LSTM-SA-JQR 模型的 CRPS 相比 LSTM-QR、GRU-QR、TCN-QR、LSTM-SA-QR 模型分别降低了 9.22%、14.23%、15.93% 和 11.78%, 分位数交叉率均降为 0。这与锐度指标共同说明了本文所提联合分位数损失模型对于预测精度的提升作用, 同时避免了分位数交叉现象。

表 7 各模型概率预测结果可靠性对比

Table 7 ACD of probabilistic predicting models

模型	$E_{ACD}^{(0.2)}$	$E_{ACD}^{(0.4)}$	$E_{ACD}^{(0.6)}$	$E_{ACD}^{(0.8)}$	$E_{ACD}^{(0.95)}$
LSTM-JQR	0.028	-0.094	0.081	0.064	0.011
GRU-JQR	0.015	-0.096	0.058	-0.046	0.084
TCN-JQR	-0.077	-0.086	0.037	0.083	-0.037
LSTM-SA-JQR	0.046	-0.079	0.025	-0.061	-0.092
DQR	0.023	0.075	-0.065	0.018	0.087

表 8 各模型概率预测结果锐度对比

Table 8 AW of probabilistic predicting models

模型	E_{AW}				
	20%	40%	60%	80%	90%
LSTM-JQR	179.53	317.48	437.93	472.47	529.84
GRU-JQR	174.36	304.62	426.46	453.91	512.17
TCN-JQR	164.31	284.79	373.75	404.83	463.63
LSTM-SA-JQR	152.84	247.32	317.97	355.97	406.10
DQR	173.80	289.95	350.67	377.69	426.31

表 9 各模型概率预测结果综合性能与分位数交叉率对比

Table 9 CRPS and P of probabilistic predicting models

模型	E_{CRPS}	E_p
LSTM-JQR	85.26	0
GRU-JQR	74.23	0
TCN-JQR	63.39	0
LSTM-SA-JQR	55.83	0
DQR	82.75	0.005

为了直观展示 JQR 模型对于分位数交叉问题的影响, 以 LSTM-QR 与 LSTM-JQR 模型为例, 图 6 是上述模型在测试集中 25 个随机的时间样本上对于 95%、90% 和 80% 分位数的预测结果。如图 6(a) 红圈处所示, 在没有使用 JQR 模型的 LSTM-QR 模型预测结果中出现 90% 分位点的预测值比 95% 分位点的预测值大, 即分位数交叉现象, 但在使用 JQR 模型的 LSTM-JQR 模型中未出现。

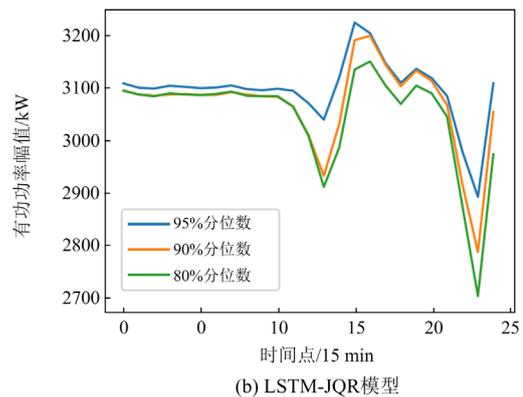
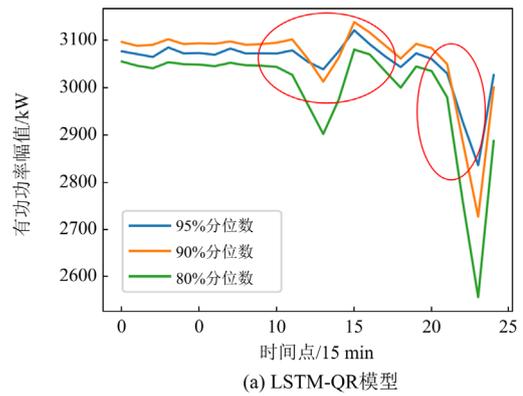


图 6 JQR 模型对于分位数交叉问题的影响

Fig. 6 Impact of quantile crossing problem by JQR model

图 7 为 JQR 模型对于分位数损失的影响, 同样以上述两个模型的 95%、90% 和 80% 分位数损失为例, LSTM-QR 模型受不同任务损失量级差异的影响, 各任务损失收敛速度不一, 导致分位数交叉问

题，而 LSTM-JQR 模型未出现。这是由于 JQR 模型在分位数联合估计过程中，通过不同任务之间的信息共享，保持各任务损失的平衡，从而避免了分位数交叉现象。在预测精度方面，LSTM-JQR 模型相比 LSTM-QR 模型的 CRPS 降低了 9.22%，说明所提 JQR 模型在预测多分位数时，提升了分位数预测精度。

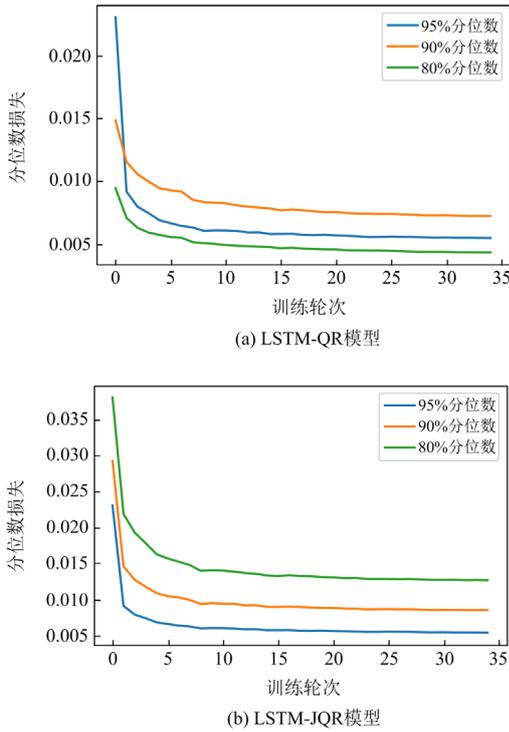


图 7 JQR 模型对于分位数损失的影响
 Fig. 7 Impact on quantile loss by JQR model

为了验证所提 JQR 模型通过自动调整各分位数损失权重实现任务间的信息共享，进而解决分位数交叉问题。图 8 展示了 LSTM-JQR 模型训练在第 1 至 30 轮时，各分位数预测任务权重参数的热力图，横轴为训练轮次，纵轴参数 1~10 为预测 10 个分位数任务的损失权重。

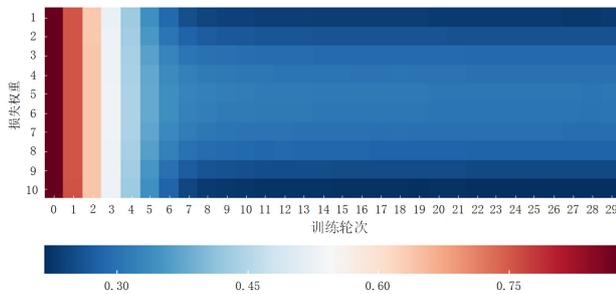


图 8 各分位数损失权重热力图
 Fig. 8 Heatmap of every quantile loss weights

由图 8 不难看出，在前 5 轮训练时，各权重参数是相等的；随着训练迭代次数的增加，JQR 模型给不同的任务参数赋予了不同的权重，结合图 7(b) 可以看出，此时权重组合有效平衡了各分位数损失的下降速度，解决了分位数交叉问题。

3.2.3 采用不同编码模块对于预测结果的影响

为了研究 SA-JQR 模型采用不同编码模块对于预测结果的影响，分别使用 LSTM、GRU、CNN、TCN 作为编码模块，并构成 LSTM-SA-JQR、GRU-SA-JQR、CNN-SA-JQR、TCN-SA-JQR 模型进行对比。

表 10 为各模型概率预测结果。可靠性 ACD 方面，对于所有模型而言，其在各分位数上均接近 0，故各模型均具有良好的预测可靠性。锐度指标 AW 方面，在 20%、40%、60%和 90%置信区间上，LSTM-SA-JQR 模型的 AW 最小，但在 80%置信区间上，TCN-SA-JQR 模型的 AW 最小。综合性能指标 CRPS 与分位数交叉率方面，可以观察到 LSTM-SA-JQR 模型的 CRPS 最小，相较于 GRU-SA-JQR、CNN-SA-JQR、TCN-SA-JQR 模型分别降低了 6.02%、20.17%、2.4%。各模型的分位数交叉率均为 0。

表 10 各模型概率预测结果可靠性对比

Table 10 ACD of probabilistic predicting models

模型	$E_{ACD}^{(0.2)}$	$E_{ACD}^{(0.4)}$	$E_{ACD}^{(0.6)}$	$E_{ACD}^{(0.8)}$	$E_{ACD}^{(0.95)}$
LSTM-SA-JQR	0.046	-0.079	0.025	-0.061	-0.092
GRU-SA-JQR	-0.083	0.017	0.088	0.071	-0.028
CNN-SA-JQR	0.019	0.049	-0.082	0.084	-0.093
TCN-SA-JQR	-0.006	0.021	-0.072	0.019	0.038
LSTM-SA-JQR	0.046	-0.079	0.025	-0.061	-0.092

综合表 10—表 12 可以得出：在大部分置信区间上，使用 LSTM 作为编码模块的 LSTM-SA-JQR 模型，其锐度指标更低，且综合性能更好。在少数情况下，使用 TCN 作为编码模块的 TCN-SA-JQR 模型，其锐度指标更低，这是由于 TCN 编码模块使用了一系列膨胀卷积层，能够更好地保留输入序列的整体结构信息；但本文所提 SA 模型需要提取时序上全局与局部关系，在使 TCN、CNN 作为编码模块时，由于其局部感知和共享权重的特性，会造成这部分信息的随机丢失，造成模型精度下降。类似地，GRU 单元由于简化了结构，合并了 LSTM 中的遗忘门和输入门成为单一的更新门，虽然单独使用时在某些特定任务上能够更有效地学习时间序列数据中的依赖关系，但并不适用于作为 SA 模型的

编码模块, 相比之下, LSTM 网络具有更复杂的记忆单元, 能够更好地处理长短期依赖关系, 更适用于作为本文所提 SA 模型的编码模块。

表 11 各模型概率预测结果锐度对比

Table 11 AW of probabilistic predicting models

模型	E_{AW}				
	20%	40%	60%	80%	90%
LSTM-SA-JQR	152.84	247.32	317.97	355.97	406.10
GRU-SA-JQR	155.72	274.42	331.70	351.36	442.75
CNN-SA-JQR	172.10	259.89	327.83	373.60	439.38
TCN-SA-JQR	168.42	254.50	324.98	349.42	417.42

表 12 各模型概率预测结果综合性能与分位数交叉率对比

Table 12 CRPS and P of probabilistic predicting models

模型	E_{CRPS}	E_p
LSTM-SA-JQR	55.83	0
GRU-SA-JQR	59.41	0
CNN-SA-JQR	69.94	0
TCN-SA-JQR	57.20	0

3.2.4 消融实验

为了量化 SA 模型中全局与局部频谱信息提取模块的有效性, 本节在 LSTM-SA-QR 模型的基础上进行消融实验。考虑到 SA 模型通过计算 A_i^j 从而整合全局与局部频谱信息, 本文分别将全局注意力系数矩阵置为 0 构成 No-local-SA 模型、局部注意力系数矩阵置为 0 构成 No-globe-SA 模型, 并与原模型进行对比。其中, No-local-SA 表示模型不进行局部噪声信息过滤, No-globe-SA 表示模型无法将全局长期信息纳入到预测结果中, 其参数如表 13 所示。

为直接对比各模型性能, 表 14 展示了消融实验结果。可以观察到在各预测区间上 LSTM-SA-QR 模型的 AW 最小, 其 CRPS 相较于 No-local-SA、No-globe-SA 模型分别降低了 33.76%、24.22%。综上, 禁用全局或局部频谱信息提取模块后, SA 模型的性能全面退化, 说明模型进行局部噪声信息过滤, 并将全局长期历史信息纳入到预测结果的必要性。

表 13 消融实验模型参数

Table 13 Parameters of ablation experiment model

模型	全局注意力系数矩阵	局部注意力系数矩阵
No-local-SA	$\delta_{g,t}^i$	恒为 0
No-globe-SA	恒为 0	$\delta_{l,t}^i$
Fall-SA	$\delta_{g,t}^i$	$\delta_{l,t}^i$

表 14 消融实验结果

Table 14 Ablation experiment results

模型	E_{AW}					E_{CRPS}
	20%	40%	60%	80%	90%	
No-local-SA	169.37	286.19	331.86	395.47	428.96	95.52
No-globe-SA	173.83	270.42	349.34	406.41	439.52	83.50
LSTM-SA-QR	152.62	268.84	328.98	372.37	415.69	63.28

3.2.5 模型多步预测性能

为对比所提 LSTM-SA-JQR 模型在进行多步预测时的性能, 分别将预测步长设置为单步(15 min)、两步(30 min)、四步(1 h)、八步(2 h)、十六步(4 h)重复实验。表 15 展示了其结果锐度指标与综合性能指标。可以观察到, 以八步与十六步预测结果为例, 其 80%置信区间 AW、CRPS 较单步分别升高了 6.94%、5.21%与 11.45%、13.73%。综合表 15 可以得出, 随着预测步长增加, 模型性能有不同程度下降, 但均不超过 15%, 仍在可用范围之内。

表 15 模型多步预测结果

Table 15 Multi-step prediction results

预测步长	E_{AW}					E_{CRPS}
	20%	40%	60%	80%	90%	
单步(15 min)	152.84	247.32	317.97	355.97	406.10	55.83
两步(30 min)	157.36	259.48	325.31	358.30	422.93	55.79
四步(1 h)	162.90	263.17	342.07	364.97	429.41	57.31
八步(2 h)	169.42	280.42	347.53	380.67	442.07	58.74
十六步(4 h)	171.89	314.59	360.80	396.72	456.42	63.50

3.3 模型鲁棒性测试

为评估模型在面对数据中的噪声、异常值、分布变化等不确定因素时, 能否保持稳定且可靠的性能, 本节进行所提模型的输入扰动测试。具体来说, 在测试集上加入随机噪声 $v \sim N(0, \sigma_v^2)$, 其中 σ_v^2 设置为 0.25、0.5、0.75 或 1, 并表示噪声多或少, 接着重复模型测试流程, 得到不同噪声下的模型评价指标。

为展示所提 LSTM-SA-JQR 模型的抗输入干扰能力, 将其与 LSTM-QR 模型在不同噪声下进行对比, 如表 16 所示。以 CRPS 与 80%置信区间上的锐度为例, 当随机噪声增加时, 模型精确度性能均有劣化, 但所提模型的性能衰减随噪声变化较小, 当噪声最大时, 所提模型性能衰减分别为 9.21%、1.62%, 而 LSTM 模型性能衰减分别为 31.69%、9.04%, 即所提模型具有较高的抗输入干扰能力。

3.4 模型训练时间对比

训练时间是评估模型性能和效率的重要指标。训练时间较长的模型意味着模型更为复杂或数据量更大, 但并不一定代表性能更好。通过对比不同模

表 16 不同噪声下的模型精确度性能

Table 16 Model accuracy performance under different noise levels

σ_v^2	LSTM-SA-JQR		LSTM-QR	
	$E_{AW} = 80\%$	E_{CRPS}	$E_{AW} = 80\%$	E_{CRPS}
0	355.97	55.83	485.35	93.92
0.25	357.31	59.02	492.60	113.48
0.5	359.20	61.38	529.37	119.31
0.75	361.84	61.50	533.38	137.50

型的训练时间,可以更全面地评估模型的性能表现。表 17 为本文模型与 LSTM-QR、GRU-QR、TCN-QR、DALSTM-QR 模型训练时间比较,训练轮次均设置为 30。比较各模型单轮最小训练时间与总训练时间可知:相比于现有模型,本文所提模型通过频域注意力机制去除了短期噪声冗余信息,使模型训练时间最短,单轮最小训练时间与平均每轮训练时间可分别缩短 27.42%、27.83%以上。

表 17 模型训练性能

Table 17 Model training performance

模型	单轮最小训练时间	平均每轮训练时间
LSTM-QR	25.76	26.24
GRU-QR	19.30	21.7
TCN-QR	23.72	27.08
DALSTM-QR	20.51	21.8
所提模型	14.83	15.66

3.5 模型超参数敏感性分析

为探究超参数对所提 LSTM-SA-JQR 模型性能的影响,分别设置模型超参数 hidden_dim、num_layers、dropout_rate、lr 参数的搜索空间为[16, 256]、[1, 3]、[0.1, 0.5]、[0.000 01, 0.1],搜索次数设置为 100,可视化搜索过程如图 9 所示,超参数重要性如图 10 所示。可以发现,学习率 lr 最为重要,在后续实验时可以考虑将重要的超参数范围加大以取得更好的搜索结果。

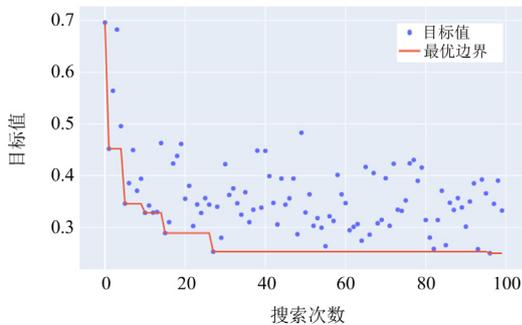


图 9 超参数优化历史

Fig. 9 Hyperparameter optimization history

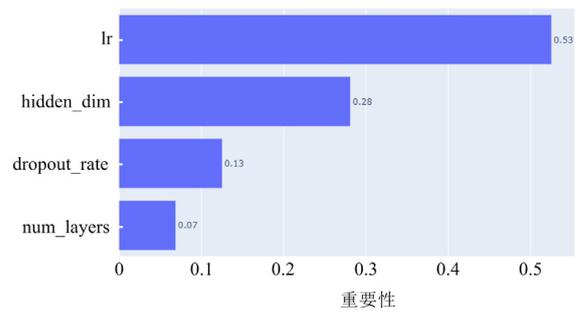


图 10 超参数重要性

Fig. 10 Hyperparameter importance

4 结论

针对现有海上风电概率预测存在的时序信息挖掘不足和分位数交叉问题,本文提出一种基于频谱注意力和联合分位数回归的海上风电功率超短期概率预测方法,基于海上风电场的真实案例分析表明:

1) 相比已有统计模型和机器学习模型,本文所提频谱注意力模型可以更好地捕捉时间序列中的全局与局部依赖关系,相比考虑传统时域上注意力机制的模型,其具有更优越的性能。分位数预测时的锐度指标提升 6%以上,连续等级概率分数提高 10%以上,具有优越的预测精度。

2) 所提联合分位数回归损失模型通过动态调整各分位数损失权重从而实现了任务间信息共享,不需要人工调参。相比传统分位数回归模型,其精度更优且解决了分位数交叉问题。

除了应对两处挑战,本文其余特色创新如下:

1) 通过对比不同编码器的实验,说明当所提频谱注意力模型采用 LSTM 单元作为编码模块时的综合性能较优。

2) 通过消融实验,说明频谱注意力模型进行局部短时噪声过滤,并将全局长期历史信息纳入到预测结果的必要性。

3) 所提模型在进行多步预测时性能下降较小,其八步(2 h)、十六步(4 h)的预测结果较单步的综合性能指标分别下降了 5.21%与 13.73%。

海上风力资源通常比陆地更为丰富和稳定,但也存在突发的极端天气事件,对功率预测的准确性提出了更高要求。本文所提模型通过调整训练数据维度,可扩展至风电场整体出力预测,从而为受端电网规划调度、改善电力系统运行安全性和经济性提供依据。但所提模型并未考虑不同风机出力间的空间关联性,后续工作可开展风机空间相关性研究。

参考文献

- [1] Global wind report 2022[R]. Global Wind Energy Council, 2022: 84-85.
- [2] 胡宇晗, 朱利鹏, 李佳勇, 等. 融合深度误差反馈学习和注意力机制的短期风电功率预测[J]. 电力系统保护与控制, 2024, 52(4): 100-108.
HU Yuhan, ZHU Lipeng, LI Jiayong, et al. Short-term wind power forecasting with the integration of a deep error feedback learning and attention mechanism[J]. Power System Protection and Control, 2024, 52(4): 100-108.
- [3] LIN C, BIE Z, CHEN C. Operational probabilistic power flow analysis for hybrid AC-DC interconnected power systems with high penetration of offshore wind energy[J]. IEEE Transactions on Power Systems, 2023, 38(4): 3016-3028.
- [4] 宋延涛, 范雪峰, 吴庆范, 等. 基于 MMC 的海上风电柔性直流送出控制策略研究[J]. 高压电器, 2023, 59(12): 63-74.
SONG Yantao, FAN Xuefeng, WU Qingfan, et al. Research on flexible dc output control strategy for offshore wind power based on MMC[J]. High Voltage Apparatus, 2023, 59(12): 63-74.
- [5] 刘新宇, 蒲欣雨, 李继方, 等. 基于贝叶斯优化的 VMD-GRU 短期风电功率预测[J]. 电力系统保护与控制, 2023, 51(21): 158-165.
LIU Xinyu, PU Xinyu, LI Jifang, et al. Short-term wind power prediction of a VMD-GRU based on Bayesian optimization[J]. Power System Protection and Control, 2023, 51(21): 158-165.
- [6] 林铮, 刘可真, 沈赋, 等. 考虑海上风电多机组时空特性的超短期功率预测模型[J]. 电力系统自动化, 2022, 46(23): 59-66.
LIN Zheng, LIU Kezhen, SHEN Fu, et al. Ultra-short-term power prediction model considering spatial-temporal characteristics of offshore wind turbines[J]. Automation of Electric Power Systems, 2022, 46(23): 59-66.
- [7] GONG M, YAN C, XU W, et al. Short-term wind power forecasting model based on temporal convolutional network and Informer[J]. Energy, 2023, 283.
- [8] 唐冬来, 周强, 宋卫平, 等. 基于网格聚类的峡谷风电短期功率预测方法[J]. 供用电, 2023, 40(1): 80-87.
TANG Donglai, ZHOU Qiang, SONG Weiping, et al. Short term power prediction method of canyon wind power based on grid clustering[J]. Distribution & Utilization, 2023, 40(1): 80-87.
- [9] 迟永宁, 梁伟, 张占奎, 等. 大规模海上风电输电与并网关键技术研究综述[J]. 中国电机工程学报, 2016, 36(14): 3758-3771.
CHI Yongning, LIANG Wei, ZHANG Zhankui, et al. An overview on key technologies regarding power transmission and grid integration of large scale offshore wind power[J]. Proceedings of the CSEE, 2016, 36(14): 3758-3771.
- [10] 赵凌云, 刘友波, 沈晓东, 等. 基于 CEEMDAN 和改进时间卷积网络的短期风电功率预测模型[J]. 电力系统保护与控制, 2022, 50(1): 42-50.
ZHAO Lingyun, LIU Youbo, SHEN Xiaodong, et al. Short-term wind power prediction model based on CEEMDAN and an improved time convolutional network[J]. Power System Protection and Control, 2022, 50(1): 42-50.
- [11] 王海燕, 刘佳康, 邓亚平. 基于预估-校正综合 BP 神经网络的短期光伏功率预测[J]. 智慧电力, 2023, 51(3): 46-52.
WANG Haiyan, LIU Jiakang, DENG Yaping. Short-term photovoltaic power forecasting based on predict-correct combination BP neural network[J]. Smart Power, 2023, 51(3): 46-52.
- [12] 杨建, 常学军, 姚帅, 等. 基于 WT-CNN-BiLSTM 模型的日前光伏功率预测[J]. 南方电网技术, 2024, 18(8): 61-69, 79.
YANG Jian, CHANG Xuejun, YAO Shuai, et al. Day-ahead photovoltaic power forecasting based on WT-CNN-BiLSTM model[J]. Southern Power System Technology, 2024, 18(8): 61-69, 79.
- [13] 张群, 唐振浩, 王恭, 等. 基于长短时记忆网络的超短期风电功率预测模型[J]. 太阳能学报, 2021, 42(10): 275-281.
ZHANG Qun, TANG Zhenhao, WANG Gong, et al. Ultra-short-term wind power prediction model based on long and short term memory network[J]. Acta Energiae Solaris Sinica, 2021, 42(10): 275-281.
- [14] 金秀章, 刘岳, 于静, 等. 基于变量选择和 EMD-LSTM 网络的出口 SO₂ 浓度预测[J]. 中国电机工程学报, 2021, 41(24): 8475-8484.
JIN Xiuzhang, LIU Yue, YU Jing, et al. Prediction of outlet SO₂ concentration based on variable selection and EMD-LSTM network[J]. Proceedings of the CSEE, 2021, 41(24): 8475-8484.
- [15] 苏向敬, 周汶鑫, 李超杰, 等. 基于双重注意力 LSTM 神经网络的可解释海上风电出力预测[J]. 电力系统自动化, 2022, 46(7): 141-151.
SU Xiangjing, ZHOU Wenxin, LI Chaojie, et al. Interpretable offshore wind power output forecasting based on dual-stage attentional long short-term memory[J]. Automation of Electric Power Systems, 2022, 46(7): 141-151.
- [16] 符杨, 任子旭, 魏书荣, 等. 基于改进 LSTM-TCN 模型的海上风电超短期功率预测[J]. 中国电机工程学报, 2022, 42(12): 4292-4303.
FU Yang, REN Zixu, WEI Shurong, et al. Ultra-short-term power prediction of offshore wind power based on

- improved LSTM-TCN model[J]. Proceedings of the CSEE, 2022, 42(12): 4292-4303.
- [17] YUAN K, ZHANG K, ZHENG Y, et al. Irregular distribution of wind power prediction[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(6): 1172-1180.
- [18] WANG Y, HU Q, MENG D, et al. Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model[J]. Applied Energy, 2017, 208: 1097-1112.
- [19] PATEL R D, NAZARIPOUYA H, AKHAVAN-H H. Non-parametric probabilistic demand forecasting in distribution grids: kernel density estimation and mixture density networks[C] // 2020 52nd North American Power Symposium (NAPS), April 11-13, 2021, Tempe, AZ, USA: 1-5.
- [20] CUI W, WAN C, SONG Y. Ensemble deep learning-based non-crossing quantile regression for nonparametric probabilistic forecasting of wind power generation[J]. IEEE Transactions on Power Systems, 2023, 38(4): 3163-3178.
- [21] CANNON A J. Quantile regression neural networks: Implementation in R and application to precipitation downscaling[J]. Computers & Geosciences, 2011, 37(9): 1277-1284.
- [22] WANG H, WANG G, LI G, et al. Deep belief network based deterministic and probabilistic wind speed forecasting approach[J]. Applied Energy, 2016, 182: 80-93.
- [23] WAN C, LIN J, WANG J, et al. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation[J]. IEEE Transactions on Power Systems, 2016, 32(4): 2767-2778.
- [24] 李丹, 张远航, 杨保华, 等. 基于约束并行 LSTM 分位数回归的短期电力负荷概率预测方法[J]. 电网技术, 2021, 45(4): 1356-1364.
LI Dan, ZHANG Yuanhang, YANG Baohua, et al. Short time power load probabilistic forecasting based on constrained parallel-LSTM[J]. Power System Technology, 2021, 45(4): 1356-1364.
- [25] MORENO-PINO F, OLMOSP M, ARTES-RODRIGUEZ A. Deep autoregressive models with spectral attention[J]. Pattern Recognition, 2023, 133.
- [26] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA: 7482-7491.
- [27] 国家能源局. 风电功率预测技术规定: NB/T 10205—2019[S]. 北京: 国家能源局, 2019.
- [28] 陈子含, 滕伟, 胥学峰, 等. 基于图卷积网络和风速差分拟合的中长期风功率预测[J]. 中国电力, 2023, 56(10): 96-105.
CHEN Zihan, TENG Wei, XU Xuefeng, et al. Medium and long term wind power prediction based on graph convolutional network and wind velocity differential fitting[J]. Electric Power, 2023, 56(10): 96-105.
- [29] 赵龙, 孟祥飞, 田鑫, 等. 基于模型-数据混合驱动的风电场暂态过程等值方法[J]. 电力建设, 2023, 44(10): 22-32.
ZHAO Long, MENG Xiangfei, TIAN Xin, et al. Equivalence method for wind farm transient process based on model-data hybrid drive[J]. Electric Power Construction, 2023, 44(10): 22-32.
- [30] 万灿, 崔文康, 宋永华. 新能源电力系统概率预测: 基本概念与数学原理[J]. 中国电机工程学报, 2021, 41(19): 6493-6509.
WAN Can, CUI Wenkang, SONG Yonghua. Probabilistic forecasting for power systems with renewable energy sources: basic concepts and mathematical principles[J]. Proceedings of the CSEE, 2021, 41(19): 6493-6509.
- [31] 林帆, 张耀, 东琦, 等. 基于分位数插值和深度自回归网络的光伏出力概率预测[J]. 电力系统自动化, 2023, 47(9): 79-87.
LIN Fan, ZHANG Yao, DONG Qi, et al. Probability prediction of photovoltaic output based on quantile interpolation and deep autoregressive network[J]. Automation of Electric Power Systems, 2023, 47(9): 79-87.
- [32] XU C, SUN Y, DU A, et al. Quantile regression based probabilistic forecasting of renewable energy generation and building electrical load: a state of the art review[J]. Journal of Building Engineering, 2023: 107772.
- [33] FAN H, ZHEN Z, LIU N, et al. Fluctuation pattern recognition based ultra-short-term wind power probabilistic forecasting method[J]. Energy, 2023, 266.
- [34] FERNANDEZ-JIMENEZ L A, MONTEIRO C, RAMIREZ-ROSADO I J. Short-term probabilistic forecasting models using Beta distributions for photovoltaic plants[J]. Energy Reports, 2023, 9: 495-502.

收稿日期: 2024-02-10; 修回日期: 2024-05-07

作者简介:

苏向敬(1984—), 男, 博士, 光明学者特聘教授, 研究方向为海上风电大数据技术、主动配电网优化规划运行;
E-mail: xiangjing_su@126.com

朱敏轩(1999—), 男, 硕士研究生, 研究方向为基于人工智能的海上风电功率概率预测技术;

符杨(1968—), 男, 通信作者, 博士, 教授, 博士生导师, 研究方向为风力发电与并网技术、变压器状态监测与故障诊断。E-mail: mfudong@126.com

(编辑 姜新丽)