

基于MV-WC和门控循环单元的短期净负荷概率预测

刘蓉晖¹, 石炬烽¹, 孙改平¹, 殷昌智²

(1. 上海电力大学电气工程学院, 上海 200090; 2. 南京赫曦电气有限公司, 南京 211100)

摘要: 分布式光伏在用户侧的大量接入增加了净负荷预测的难度。为了提高预测精度, 提出了一种基于天气分型和神经网络的短期净负荷概率预测模型。首先, 为了更好地描述气象条件, 将对光伏出力影响最大的云量和环境温度进行加权求和, 构造出一种综合的气象影响因子, 并提出了一种基于气象影响因子波动率的天气分型方法。同时, 针对气象因子对净负荷的影响存在时间差异性, 采用最大信息系数帮助模型分段选择输入特征。最后, 将样本选择结果作为输入, 通过分位数回归和门控循环单元(quantile regression and gated loop unit, QR-GRU)组成的神经网络进行概率预测, 并利用核密度估计生成概率密度曲线。通过仿真验证了该模型具有良好的预测性能。

关键词: 净负荷; 气象影响因子; 波动率; 天气分型; 最大信息系数; 分位数回归; 概率预测

Short-Term Net Load Probability Forecasting Based on MV-WC and Gated Recurrent Unit

LIU Ronghui¹, SHI Jufeng¹, SUN Gaiping¹, YIN Changzhi²

(1. School of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China; 2. Nanjing Hexi Electric Co., Ltd., Nanjing 211100, China)

Abstract: The large number of distributed photovoltaic (PV) access on the user side increases the difficulty of net load forecasting. In order to improve the forecasting accuracy, a short-term net load probabilistic forecasting model based on weather classification and neural network is proposed. First of all, in order to better describe the meteorological conditions, the weighted sum of cloud cover and ambient temperature that have the greatest influence on PV output is carried out to construct a comprehensive meteorological impact factor, and a weather classification method based on the volatility of meteorological impact factor is proposed. At the same time, the maximum information coefficient is used to help the model select the input characteristics in different time frames in view of the time difference of the influence of meteorological factors on the net load. Finally, using the sample selection results as input, probability prediction is made by quantile regression and gated loop unit (QR-GRU) neural network, and probability density curve is generated by kernel density estimation. The simulation results show that the model has good prediction performance.

Key words: net load; meteorological impact factor; volatility; weather classification; maximum information coefficient; quantile regression; probabilistic prediction

0 引言

随着分布式光伏在用户侧的渗透率提高^[1], 净负荷逐渐形成。受间歇光伏发电功率和不稳定用户负荷需求影响, 电力系统运行被注入了不确定性^[2],

而短期净负荷预测凭借能够准确预测次日或连续几天的净负荷, 有效应对这种不确定性^[3], 逐步成为现代可再生能源电网运行管理的重要组成部分。

分布式光伏的发电量取决于接收的太阳辐照度。然而太阳辐照度会受到多种气象因素的影响, 使得不同天气类型下的光伏功率具有不同的波动特征^[4], 因此划分天气类型可以提高预测精度。如文献[5-6]根据太阳辐照度和历史光伏功率划分天气

基金项目: 国家自然科学基金资助项目(51977127)。

Foundation item: Supported by the National Natural Science Foundation of China (51977127).

类型。然而,这些研究只考虑了单一变量,限制了模型对天气变化的响应^[7],而云量和温度等因素的波动也直接影响净负荷输出。云量可以直接影响太阳辐照度,使其在短时间内出现剧烈波动^[8]。环境温度是影响电力消耗的主要因素之一,它与耗电量之间存在着复杂的非线性关系^[9]。因此,在划分天气类型时考虑它们的波动性,有助于更准确地对天气类型进行划分。

以往的研究忽略了分布式光伏的时间输出特性。如文献[10]直接将太阳辐照度和净负荷数据作为模型输入,没有考虑不同时间太阳辐照度的时间变化差异和各种气象因素对净负荷的影响,导致夜间预报精度较低。为了提高预测精度,需要对输入特征进行分段选择。

随着算法的进步,更复杂的预测方法得到了发展,包括神经网络^[11]、小波分析^[12]、支持向量机^[13]等。神经网络凭借优异的预测能力被广泛用于负荷预测^[14]。循环神经网络(recurrent neural network, RNN)^[15]常用于净负荷的时序预测,但RNN存在梯度消失和梯度爆炸等问题。长短时记忆网络(long short-term memory, LSTM)通过引入“门”结构可以改善这个问题^[16],而门控循环单元(gated recurrent unit, GRU)简化了LSTM的“门”结构,进一步提高了预测精度^[17]。文献[18]基于门控循环单元,引入社交网络搜索算法,构建光伏功率预测模型,取得了较好的预测精度。概率预测常用于评估预测的不确定性^[19],通过分位数回归(quantile regression, QR)和神经网络获得条件分位数,然后使用核密度估计(kernel density estimation, KDE)生成概率密度函数,这是概率预测的常用方法^[20]。为此,提出了一种结合QR和门控循环单元(QR-GRU)的深度学习模型,用于净负荷概率预测。该模型通过天气分型和最大信息系数(maximum information coefficient, MIC)选择输入参数,并通过提供预测时间点净负荷的预测区间来描述预测的不确定性。

综上所述,本文提出一种基于气象影响因子波动率的天气分型方法和一种基于MIC的特征选择方法。首先分析多种气象因子对净负荷影响的时间差异性,引入波动率对气象状况进行描述,提出利用云量和环境温度构造一种气象影响因子(meteorological impact factors, MIF),通过对MIF

的波动率进行K-means聚类进行天气分型。其次,针对气象影响具有的时间差异性,利用MIC量化各时段气象因子对净负荷的贡献,对输入特征进行分段选择。最后,基于MV-WC和QR-GRU构造短期净负荷概率预测模型,预测结果表明该模型具有良好的预测性能。

1 天气分型

1.1 净负荷建模

分布式光伏安装于表后电表只能测量和采集净负荷数据。净负荷受用户负荷需求和光伏发电量地共同影响,反映用户的实际负荷需求。任意给定时间 t 的净负荷 P_{net} 可表示为:

$$P_{\text{net}} = P_{\text{need}} - P_{\text{pv}} \quad (1)$$

式中 P_{need} 和 P_{pv} 分别为用户负荷需求和光伏发电量。

用户负荷需求 P_{need} 主要受工作日类型(工作日或周末)、每天工作小时数和环境温度的影响,如式(2)所示。

$$P_{\text{need}} = f(M_t, W_t, H_t, T_{A,t}) \quad (2)$$

式中: M_t 、 W_t 、 H_t 分别为月份、星期和小时; $T_{A,t}$ 为环境温度;函数 $f(\cdot)$ 为用户负荷需求模型。

分布式光伏的发电量 P_{pv} 与太阳辐照度、环境温度等变量相关。大多数研究使用经验模型^[21]来估计光伏发电的潜力,具体如下。

$$P_{\text{pv}} = G_{\text{rds}} \times \eta_c \quad (3)$$

$$\eta_c = \eta_{T_{\text{ref}}} [1 - \beta_{\text{ref}}(T_c - T_{\text{ref}}) + \gamma \log_{10} G_{\text{rds}}] \quad (4)$$

$$\beta_{\text{ref}} = \frac{1}{T_0 - T_{\text{ref}}} \quad (5)$$

$$T_c = T_a + (T_{\text{NOCT}} - T_{a,\text{NOCT}}) \times \frac{G_{\text{rds}}}{G_{\text{NOCT}}} \quad (6)$$

式中: G_{rds} 和 η_c 分别为地表太阳总辐射和发电效率; $\eta_{T_{\text{ref}}}$ 为参考温度; T_{ref} 为光伏电池的参考温度, $T_{\text{ref}} = 25^\circ\text{C}$,太阳辐照度为 1000 W/m^2 时的电池电效率; β_{ref} 为温度系数,通常为 $0.004/\text{K}$; T_c 为光伏电池的温度; T_a 为环境温度; T_0 为光伏电池的初始温度; T_{NOCT} 为标准电池温度; $T_{a,\text{NOCT}} = 20$; γ 为辐射系数,通常为 0.12 ; G_{NOCT} 为地表太阳标准辐照度。

1.2 气象影响的时间差异性分析

如式(1)所示,净负荷输出由光伏发电量和用户负荷需求决定。由于光伏发电极易受到天气条件的影响,当天气发生变化时,净负荷会出现显著波动。此外,由于太阳辐照度的影响,光伏发电表现

出明显的时间输出特性。为了准确分析多种气象因素的影响,将净负荷数据分为日间(08:00—18:45)和夜间(00:00—07:45、19:00—00:00)。将各变量归一化后,用 Pearson 系数量化气象因子对两个时段的净负荷影响。归一化及 Pearson 系数计算公式如式(7)—(8)所示。

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

$$\rho_{xy} = \frac{\text{COV}(x, y)}{\sigma_x \sigma_y} \quad (8)$$

式中: x 、 y 为两组时间序列; $\text{cov}(x, y)$ 为 x 、 y 间的协方差; σ_x 、 σ_y 为 x 、 y 间的标准协方差; $x_{\max}$ 、 $x_{\min}$ 分别为 x 的最大值和最小值; ρ_{xy} 取值范围为 $[-1, 1]$, $|\rho_{xy}|=1$ 相关性最高, $|\rho_{xy}|$ 越接近 1, 相关性越强。当 $|\rho_{xy}|>0.8$ 时, 两个变量可以视为高度相关; 当 $|\rho_{xy}|<0.3$ 时, 可认为两个变量不相关; 当 $0.3<|\rho_{xy}|<0.8$ 时, 两个变量为中度相关^[22]。结果如表 1 所示, 日间净负荷与太阳辐照度呈现高度相关; 夜间净负荷只与环境温度和相对湿度呈现中度相关。由此可以推断气象因子对净负荷的影响具有时间差异。

表 1 气象因子对净负荷的相关性

Tab. 1 Correlation of meteorological factors on net load

气象因子	日间	夜间
环境温度	-0.73	0.55
相对湿度	0.27	-0.42
气压	0.20	0.23
北向风速	-0.21	-0.10
东向风速	-0.28	-0.12
云量	0.44	0.16
太阳辐照度	-0.80	

1.3 气象影响因子

日间, 太阳辐照度和环境温度对净负荷的影响较大, 同时它们也是影响光伏发电潜力^[23]的关键气候因素, 说明净负荷的波动主要受光伏发电量的影响。在光伏发电过程中, 太阳辐照度的波动会引起环境温度的变化, 从而直接影响光伏转换效率。然而, 太阳辐照度变化对光伏输出的响应不能简单量化, 直接辐射与散热辐射的比值决定了太阳辐照度变化的影响, 但该比值也会受到云量变化^[24]的影响, 从而导致光伏发电量的波动。

为了量化云量和环境温度对白天太阳辐照度的影响, 采用 Pearson 系数来评价它们之间的关系,

结果为 $\{-0.49, 0.82\}$, 表明太阳辐照度与云量呈中度负相关, 而太阳辐照度与环境温度呈强正相关。在此基础上, 通过对云量和温度进行加权和来计算一个新的指数, 称为气象影响因子, 其中权重由各自的 Pearson 系数转换而来, 计算公式如式(9)—(11)所示。

$$\omega_1 = \frac{a}{a+b} \quad (9)$$

$$\omega_2 = \frac{b}{a+b} \quad (10)$$

$$\mu_{\text{MIF}} = \omega_1 C_{\text{TCC}} + \omega_2 T_{\text{A}} \quad (11)$$

式中: a 、 b 为变量与太阳辐照度之间的相关系数; ω_1 、 ω_2 分别为变量与太阳辐照度的占比, 两者的和为 1; μ_{MIF} 为气象影响因子值; C_{TCC} 和 T_{A} 分别为云量和环境温度。

1.4 波动率

波动率是衡量时间序列历史变化的重要指标, 通过计算特定时间段内时间序列变化的标准差, 能够反映特定时间窗口 T 内时间序列的活跃程度^[25]。要确定波动率, 第一步是计算时间序列在时间段 h 内的对数回报率。时间序列在时间段 h 内的对数回报率通常用 $R_{t,h}$ 表示, 通过式(12)进行计算。

$$R_{t,h} = \ln\left(\frac{L_t}{L_{t-h}}\right) = \ln(L_t) - \ln(L_{t-h}) \quad (12)$$

式中 L_t 、 L_{t-h} 分别为时间序列在时间 t 和 $t-h$ 的参数。通过式(13)可以计算出时间序列在特定时间窗口 T 内的对数回报率的平均值 R_m 。

$$R_m = \frac{1}{T} \sum_{t=1}^T R_{t,h} \quad (13)$$

波动率定义为时间序列在特定时间窗口 T 内的对数回报率的标准差, 用 $\sigma_{h,T}$ 表示, 其公式如下。

$$\sigma_{h,T} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (R_{t,h} - R_m)^2} \times 100\% \quad (14)$$

受光伏发电的随机性影响, 净负荷会在白天表现出显著波动。因此, 将时间段 h 设为 1 h, 通过计算 MIF 值每小时的对数回报率, 获得白天 MIF 值的波动率。根据气象影响因子的波动率, 采用基于 MV-WC 的天气分型方法对天气类型进行划分。

2 方法

2.1 最大信息系数

最大信息系数(maximal information coefficient,

MIC, 其值记作 C_{MIC}) 是 Reshef 于 2011 年在 Science 上提出的一种方法, 用于检测大型数据集中的数据相关性。该系数能够有效识别两个变量之间的线性或非线性函数关系, 不受异常值的影响, 具有很强的普适性^[26]。 C_{MIC} 的取值范围为 $[0, 1]$, C_{MIC} 越高表明两个变量之间的相关性越强。 C_{MIC} 的具体计算如下。

$$M_{G(x,y)} = \frac{\max \{f_{MI}(D, G(x, y))\}}{\log_2 \min \{x, y\}} \quad (15)$$

$$C_{MIC} = \max_{xy < B(\delta)} \{M_{G(x,y)}\} \quad (16)$$

式中: x 、 y 分别为自变量和因变量; D 为关于变量的数据集; $G(x, y)$ 为定义的网格; $f_{MI}(D, G(x, y))$ 为数据集 D 在 $G(x, y)$ 上的互信息; $B(\delta) = \delta^{0.6}$, δ 为数据集 D 的大小。

2.2 门控循环单元基本结构

门控循环单元(gated recurrent unit, GRU)通过将 LSTM 中的遗忘门和输入门集成为一个更新门, 简化了 LSTM 的结构^[27]。在保持预测精度^[28]的情况下, GRU 比 LSTM 更容易训练。GRU 的结构如图 1 所示。

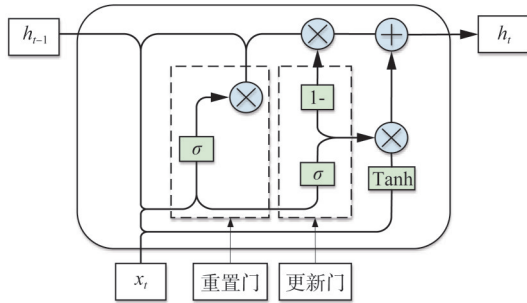


图1 GRU 的结构图

Fig. 1 Structure diagram of GRU

式(17)给出了 GRU 的数学描述。

$$\begin{cases} r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \\ \tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \\ z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \\ Y_t = \sigma(W_o \cdot h_t) \end{cases} \quad (17)$$

式中: r_t 和 z_t 分别为更新门和重置门的输出变量; $\tilde{h}_t$ 、 h_t 为遗忘门的输出变量; W_r 、 W_h 、 W_z 分别为重置门、隐藏单元和更新门的权重参数; x_t 、 h_{t-1} 为遗忘门的输入变量; W_o 为输出权值矩阵; b_r 、 b_h 、

b_z 分别为重置门、隐藏单元和更新门的偏置矩阵; σ 为激活函数; $\odot$ 为点积; Y_t 为输出。

2.3 概率预测模型

2.3.1 分位数回归

由 Koenker 等提出的分位数回归(quantile regression, QR)可用于探究自变量与因变量的条件分位数之间的关系。其中, 非线性 QR 比线性 QR 有更广阔的前景, 因为它不需要输入变量和输出变量之间存在线性关系。非线性 QR 模型可以描述为:

$$Q_{y_i}(\lambda | x_i) = f[x_i, w(\lambda), v(\lambda)], i = [1, 2, \dots, n] \quad (18)$$

式中: $Q_{y_i}(\lambda | x_i)$ 为输入变量 x 在条件分位数 λ 下的响应变量 y_i , λ 的取值范围为 $[0, 1]$; $w(\lambda)$ 和 $v(\lambda)$ 为权重向量; $f[x_i, w(\lambda), v(\lambda)]$ 为一个非线性函数; n 为样本数量。

通过将非线性 QR 模型中参数的权重预测转化为最小化损失函数的优化挑战, 可以构建神经网络的分位数回归(QR)模型, 损失函数如式(19)所示。

$$\begin{aligned} L(w, v) = \min_{w, v} \sum_{i=1}^n Q_{\lambda}(Y_i - f[x_i, w(\lambda), v(\lambda)]) = \\ \min_{w, v} \sum_{i|Y_i \geq f[x_i, w(\lambda), v(\lambda)]} \lambda |Y_i - f[x_i, w(\lambda), v(\lambda)]| + \\ \min_{w, v} \sum_{i|Y_i < f[x_i, w(\lambda), v(\lambda)]} (1 - \lambda) |Y_i - f[x_i, w(\lambda), v(\lambda)]| \end{aligned} \quad (19)$$

式中: Y_i 为实际值; f 为预测值。模型权重针对给定的分位数 λ 进行优化。在式(18)的不同分位数下可以得到净负荷预测值。在 $(0, 1)$ 区间内连续取分位数 λ , 可得到分布预测。

2.3.2 核密度估计

核密度估计(kernel density estimation, KDE)作为一种非参数估计方法不依赖于数据分布的先验知识, 概率密度估计结果完全由样本数据本身分布决定^[29]。KDE 的基本原理如下: 假设从连续分布集合 X 中提取独立同分布的样本数据 $x_i (i=1, 2, \dots, n)$, 则概率密度函数 $f(x)$ 在任意点 x 处的 KDE 为:

$$\hat{f}_d(x) = \frac{1}{Nd} \sum_{i=1}^N K\left(\frac{x - x_i}{d}\right) \quad (20)$$

式中: N 为样本数量; d 为带宽, $d > 0$; $K(\cdot)$ 为核函数, 本文采用高斯核函数, 如式(21)所示。

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \quad (21)$$

式中 u 为高斯核函数的输入变量。

2.4 模型框架

本文提出的净负荷混合预测模型如图 2 所示, 具体过程如下。

1) 收集净负荷和气象的历史数据。从原始数据中去除异常值, 然后进行归一化。

2) 利用 Pearson 系数量化云量和环境温度对太阳辐射度的影响, 并将其转化为权重, 通过对它们进行加权得到 MIF 值。

3) 采用 K-means 算法对 MIF 值的波动率进行聚类, 将数据分为 4 类。对于每个天气类型数据集, 90% 的数据用于训练预测模型, 10% 用于测试。

4) 利用 MIC 衡量昼夜天气因子对净负荷的贡献度, 帮助模型分段挖掘最优输入特征。

5) 对预测点的净负荷时间序列进行预测, 并利用 QR 和 KDE 拟合概率密度函数。采用 95% 置信区间作为预测区间。

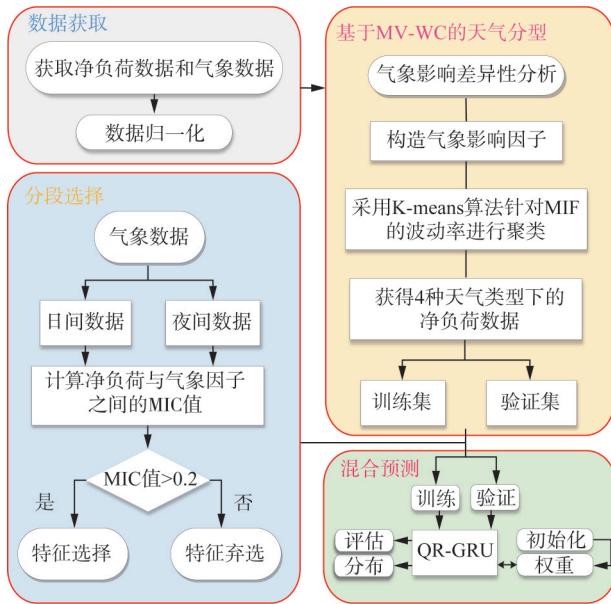


图 2 流程框架

Fig. 2 Process framework

3 评价指标

3.1 点预测评估

为了验证模型的性能, 本文使用的评价指标包括: 平均绝对百分比误差 (mean absolute percentage error, MAPE, 其值记作 E_{MAPE})、均方根误差 (root mean square error, RMSE, 其值记作 E_{RMSE}) 和决定系数 R^2 。

$$E_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - y_i}{Y_i} \right| \quad (22)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - y_i)^2} \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - y_i)^2}{\sum_{i=1}^n (Y_i - y_m)^2} \quad (24)$$

式中: n 为样本数量; Y_i 和 y_i 分别为实际值和预测值; y_m 为实际值的平均值。

3.2 区间预测评估

预测区间覆盖概率 (prediction interval coverage probability, PICP, 其值记作 P_{PICP}) 是概率预测区间评估的重要指标。公式定义如下。

$$P_{\text{PICP}} = \frac{1}{V} \sum_{i=1}^V c_i \quad (25)$$

$$c_i = \begin{cases} 1, & y_i \in [L_i, U_i] \\ 0, & y_i \notin [L_i, U_i] \end{cases} \quad (26)$$

式中: V 为验证集的样本数量; c_i 为预测区间内的样本数量; U_i 和 L_i 分别为预测区间的上下限。采用预测区间归一化平均宽度 (prediction interval normalized average width, PINAW, 其值记作 P_{PINAW}) 评估预测区间宽度, 确保区间有效性, 公式定义如下。

$$P_{\text{PINAW}} = \frac{1}{V(Y_{\max} - Y_{\min})} \sum_{i=1}^V U_i - L_i \quad (27)$$

式中 $Y_{\max}$ 、 $Y_{\min}$ 分别为实际最大值和最小值。

合适的预测区间具有较高的 PICP 值和较低的 PINAW 值, 故定义区间预测评价的综合测度值 M_{MC} , M_{MC} 越小则预测区间越合适。

$$M_{\text{MC}} = \frac{P_{\text{PINAW}}}{P_{\text{PICP}}} \quad (28)$$

4 仿真案例研究

4.1 数据获取

本文收集了中国某地区 2021 年 1 月 1 日—12 月 31 日的历史净负荷和气象数据 (包括环境温度、太阳辐射、相对湿度、大气压力、北向风速、东向风速和云量), 采样频率为 15 min, 每天共 96 个采样点, 全年共有 35 040 组数据, 并消除了极端异常值。

4.2 基于MV-WC的天气分型方法

本文提出了一种基于MV-WC的天气分型方法划分天气类型,根据MIF的波动率对不同时间对应的天气条件进行划分,具体划分标准如表2所示,其中 σ_{MIF} 表示气象影响因子的波动率。图3描述了MV-WC的划分结果,从图中可看出,4种天气类型的净负荷曲线呈现不同的趋势和波动。

4.3 MIC提取最佳特征集

表1结果显示,影响净负荷的因素在不同时段并不相同。为了选择最优特征作为模型的输入,利

表2 天气分型结果

Tab. 2 Weather classification results

天气类型	划分标准
1	$0.69\% < \sigma_{\text{MIF}} < 7.21\%$
2	$7.32\% < \sigma_{\text{MIF}} < 17.02\%$
3	$17.19\% < \sigma_{\text{MIF}} < 36.04\%$
4	$36.13\% < \sigma_{\text{MIF}} < 65.83\%$

用MIC量化了气象特征在不同时段对净负荷的贡献。候选特征集如表3所示。图4显示了MIC分析结果,其中X轴和Y轴的序号代表相应的气象特征,

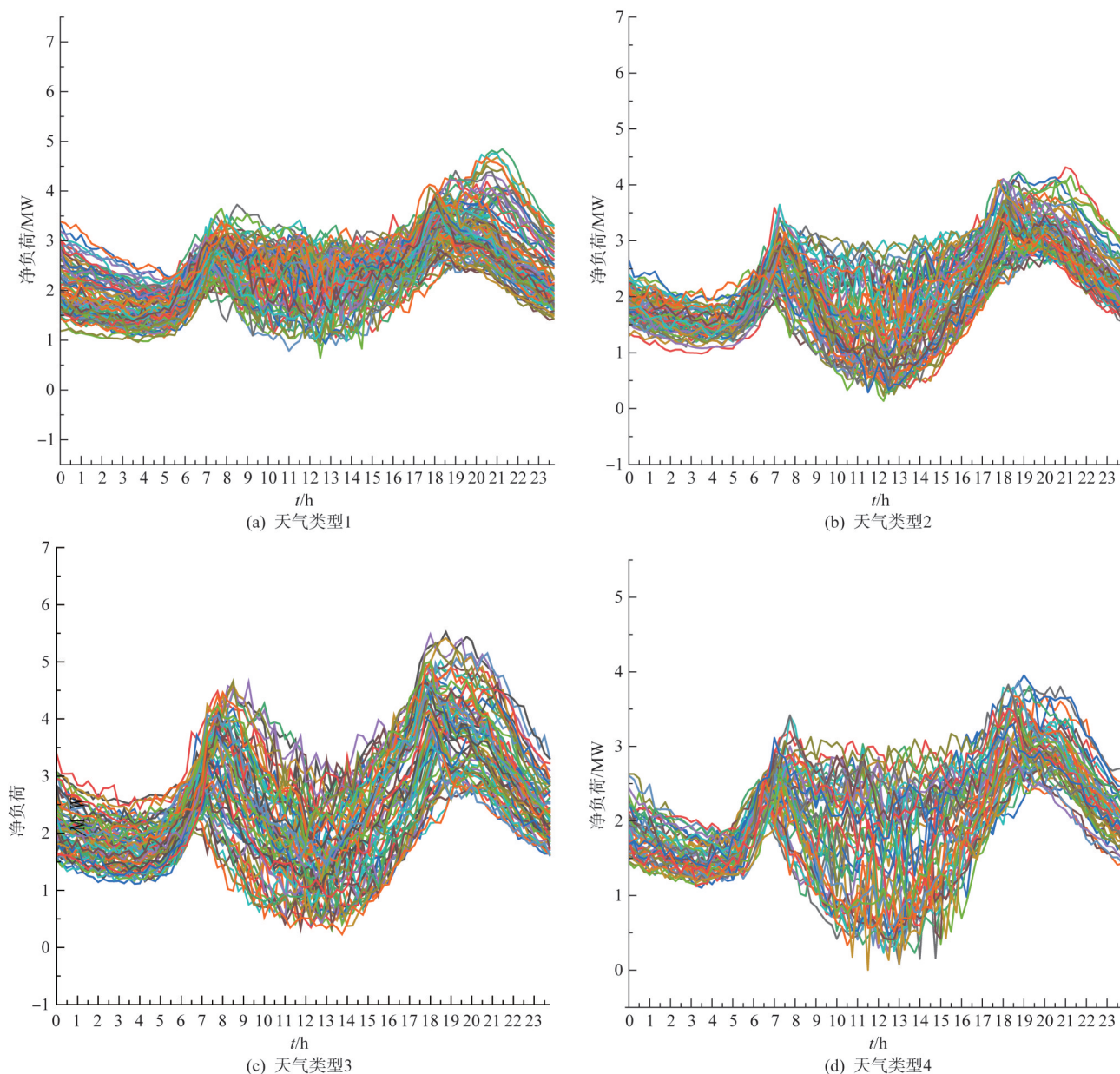


图3 4种天气类型的净负荷输出

Fig. 3 Net load outputs for 4 weather types

序号 7 代表净负荷数据。

表 3 候选气象特征集

Tab. 3 Candidate meteorological feature set

变量序列	气象特征
0	太阳辐射度
1	环境温度
2	相对湿度
3	气压
4	北向风速
5	东向风速
6	云量

在 MIC 理论中贡献度决定了影响的程度。选取贡献度大于 0.2 的气象特征作为模型输入。如图 4 所示, 日间预测可选择环境温度和太阳辐射作为输入, 夜间预测可选择环境温度和相对湿度。另外, 将由 MV-WC 筛选的净负荷数据作为一个关联度较强的时间序列, 作为模型的输入。

4.4 净负荷日前预测

为了验证所提模型的性能, 对 MIC-QR-GRU、QR-GRU、QR-RNN、QR-LSTM 进行了比较分析, 其中 MIC-QR-GRU 表明样本特征使用了特征分段选择方法。表 4 列举了各模型的训练及计算时间, 从表 4 可知, MIC-QR-GRU 所需的训练及计算时间均少于其他模型 (QR-GRU、QR-LSTM、QR-RNN), 具有较好的性能。

图 5 为 4 种天气类型下所提模型的区间预测结果, 预测时间尺度为 15 min, 能够从图 5 看出, 所提模型的预测曲线能够有效跟踪到预测曲线的变化情况。表 5 总结了 4 种天气类型下的净负荷预测误差。可以看出, 在天气类型 1 中所有模型均能够取得较好的预测精度。其中, 样本数据在经过 MV-WC 算法划分并通过 MIC 分段选择最优输入特征后, GRU 的预测精度得到了有效提升, MIC-QR-GRU 的 MAPE 值和 RMSE 值比起 QR-GRU, 分别降低了 44.39%、11.25%, R^2 从 0.993 达到了 0.995。同时, 随着 MIF 的波动率变大, 所有模型的预测精度均出现了下降, 在天气类型 4 中达到最低。但是, MIC-QR-GRU 的误差仍低于其他模型 (QR-GRU、QR-LSTM、QR-RNN)。可见 MV-WC 算法和 MIC 分段选择能够帮助 GRU 进一步提升特征信息的捕获能力, 从而提高预测精度。

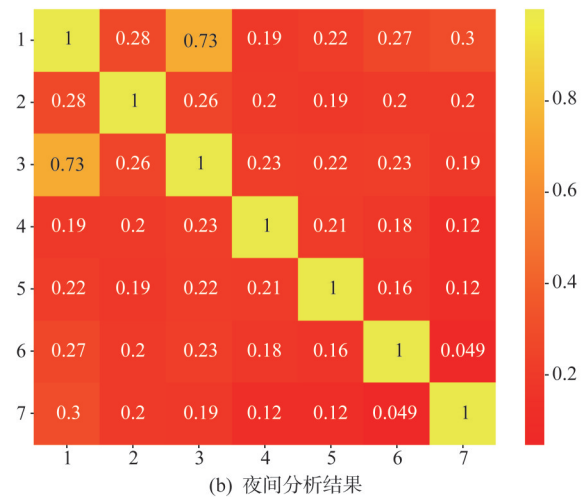
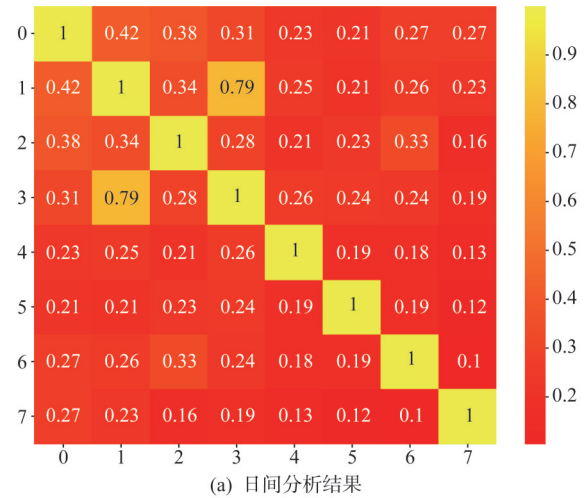


图 4 分析结果热力图

Fig. 4 Heat map of analysis results

表 4 模型训练及计算时间对比

Tab. 4 Comparison of model training and calculation time

模型	天气类型 1		天气类型 2	
	训练时间	计算时间	训练时间	计算时间
MIC-QR-GRU	31.890	1.603	32.508	1.663
QR-GRU	32.016	1.627	32.558	1.666
QR-LSTM	32.275	1.629	32.670	1.694
QR-RNN	32.432	1.637	32.807	1.708
模型	天气类型 1		天气类型 2	
	训练时间	计算时间	训练时间	计算时间
MIC-QR-GRU	33.831	1.710	35.874	2.165
QR-GRU	33.946	1.727	36.219	2.317
QR-LSTM	33.982	1.759	37.132	2.334
QR-RNN	34.030	1.763	38.064	2.380

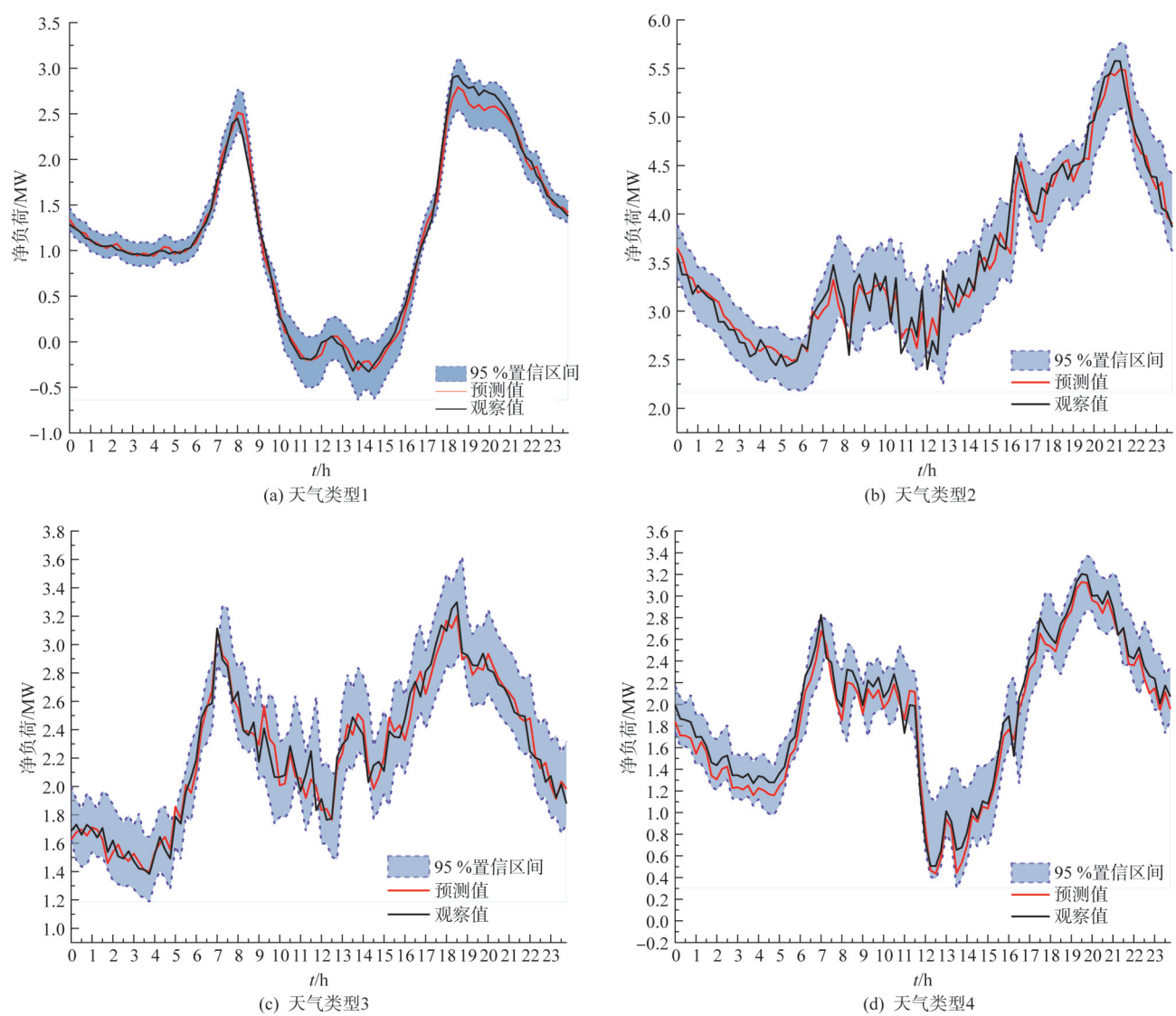


图 5 4 种天气类型的概率预测分布

Fig. 5 Probability forecast distribution for four kinds of weathers

表 5 点预测性能评价

Tab. 5 Point prediction performance evaluation						
模型	天气类型 1			天气类型 2		
	MAPE 值	RMSE 值	R ² 值	MAPE 值	RMSE 值	R ² 值
MIC-QR-GRU	2.843	0.071	0.995	3.452	0.079	0.991
QR-GRU	5.112	0.080	0.993	5.910	0.094	0.941
QR-LSTM	5.635	0.082	0.992	5.817	0.091	0.942
QR-RNN	5.977	0.079	0.993	6.352	0.096	0.939
模型	天气类型 3			天气类型 4		
	MAPE 值	RMSE 值	R ² 值	MAPE 值	RMSE 值	R ² 值
MIC-QR-GRU	4.882	0.119	0.978	6.342	0.107	0.975
QR-GRU	7.084	0.162	0.932	7.889	0.128	0.945
QR-LSTM	8.657	0.158	0.938	9.484	0.143	0.926
QR-RNN	9.749	0.116	0.957	10.381	0.155	0.929

4.5 概率密度预测

本节给出 4 种天气类型下 3 种模型(QR-GRU、QR-RNN、QR-LSTM)在预测日 12:00 时(此为预测日当天波动率较大时刻)的概率密度预测曲线,如图 6 所示。从图 6 中可以看出,3 种模型均在天气类型 1 中拥有较好的概率预测结果,预测概率范围都包含了实际值,并且实际值都接近于最高概率点,它表明获得接近实际值的预测值的概率很高。但随着波动率的增加,实际值开始偏离 QR-RNN 和 QR-LSTM 的概率密度曲线上的最高概率点,即这两种模型的预测精度有所下降,而在 QR-GRU 的概率密度曲线中,实际值仍然位于最高概率点附近。这表明,当天气条件变得复杂时,QR-GRU 仍然拥有良

好的预测性能。

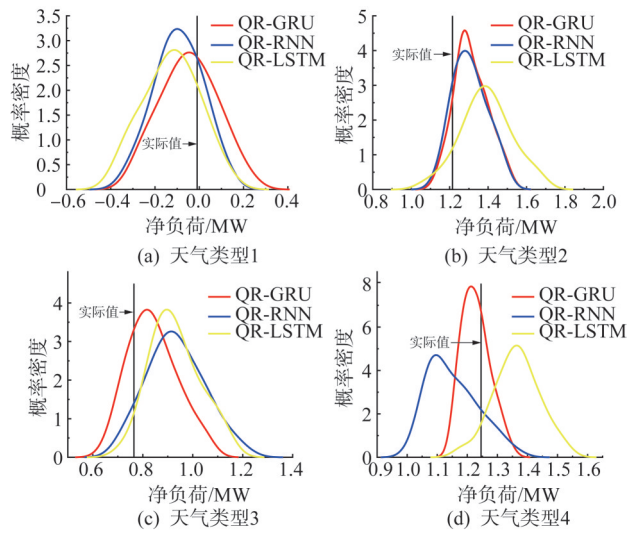


图6 各模型的概率密度曲线对比

Fig. 6 Comparison of probability density curves of each model

表6给出了区间预测评价结果,结果表明:QR-GRU模型在4种天气类型上的PICP均高于QR-RNN和QR-LSTM, PICP值分别达到0.990、0.988、0.945和0.895,表明QR-GRU的预报区间能够覆盖大部分实际值。此外,与QR-RNN和QR-LSTM相比,QR-GRU在4种天气类型中实现了较低的PINAW值,分别为0.128、0.152、0.166和0.180。这表明QR-GRU可以提供更合适的预测区间宽度。由于QR-GRU均拥有较高的PICP值和较低的PINAW值,使得MC值较低,因此QR-GRU的预报区间更合适。

表6 区间预测性能评价

Tab. 6 Interval prediction performance evaluation

模型	天气类型1			天气类型2		
	PICP值	PINAW值	MC值	PICP值	PINAW值	MC值
QR-GRU	0.990	0.128	0.129	0.988	0.152	0.154
QR-RNN	0.990	0.164	0.166	0.962	0.211	0.219
QR-LSTM	0.979	0.151	0.154	0.959	0.207	0.216
模型	天气类型3			天气类型4		
	PICP值	PINAW值	MC值	PICP值	PINAW值	MC值
QR-GRU	0.945	0.166	0.176	0.895	0.180	0.201
QR-RNN	0.862	0.165	0.191	0.861	0.177	0.206
QR-LSTM	0.914	0.179	0.196	0.852	0.179	0.210

5 结语

为了提高净负荷预测的精度,本文提出了一种

基于MV-WC和门控循环单元的短期净负荷概率预测模型。首先分析了云量和温度对净负荷的影响,并构建MIF。在此基础上提出了一种基于MV-WC的天气分型方法来识别不同波动水平的净负荷时间序列。此外,针对天气因素对净负荷影响的时间差异将MIC用于输入特征的分段选择。最后,建立QR-GRU模型生成概率密度曲线。实验结果表明,在最复杂的天气类型4中,与获得最低精度的QR-RNN模型相比, MIC-QR-GRU模型的MAPE值降低了38.91%, RMSE值降低了30.97%, R^2 提高了4.95%。QR-GRU在4种天气类型的MC值分别为0.129、0.154、0.176和0.201,均优于QR-LSTM和QR-RNN模型,表明其预测区间最优。综上所述,该模型具有较好的预测精度和不确定度评价。

参考文献

- [1] 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.
- [2] YI W, NING Z, QIXIN C, et al. Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV [J]. IEEE Transactions on Power Systems, 2018, 33(3): 3255-3264.
- [3] 康重庆,夏清,张伯明.电力系统负荷预测研究综述与发展方向的探讨[J].电力系统自动化,2004,28(17):1-11.
KANG Chongqing, XIA Qing, ZHANG Boming. Review of power system load forecasting and its development [J]. Automation of Electric Power Systems, 2004, 28(17): 1-11.
- [4] ISHII T, OTANI K, TAKASHIMA T, et al. Solar spectral influence on the performance of photovoltaic (PV) modules under fine weather and cloudy weather conditions [J]. Progress in Photovoltaics: Research and Applications, 2013, 21(4): 481-489.
- [5] WANG F, ZHEN Z, MI Z, et al. Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting [J]. Energy & Buildings, 2015(86): 427-438.
- [6] YANG H T, HUANG C M, HUANG Y C, et al. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output [J]. IEEE Transactions on Sustainable Energy, 1949, 5(3): 917-926.
- [7] OSAKA R, TINGZE Z, XINAN Z, et al. Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration [J]. Applied Energy, 2023(333): 110-121.
- [8] HARRISON E F, MINNIS P, BARKSTROM B R, et al. Seasonal variation of cloud radiative forcing derived from the

- earth radiation budget experiment[J]. *Journal of Geophysical Research: Atmospheres*, 1990, 95(11): 18687 – 18703.
- [9] IMANI M. Electrical load-temperature CNN for residential load forecasting[J]. *Energy*, 2021(10): 227 – 235.
- [10] FAUSTINE A, PEREIRA L, FPSEQ Q. Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates[J]. *IEEE Transactions on Smart Grid*, 2022, 13(3): 2440 – 2451.
- [11] SINGH P, DWIVEDI P. Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem[J]. *Applied Energy*, 2018(217): 537 – 549.
- [12] HU Y, YE C, DING Y, et al. Short-term load forecasting utilizing wavelet transform and time series considering accuracy feedback[J]. *International Transactions on Electrical Energy Systems*, 2020, 30(7): 131 – 142.
- [13] 王小君, 毕圣, 徐云鹏, 等. 基于数据挖掘技术和支持向量机的短期负荷预测[J]. *电测与仪表*, 2016, 53(10): 62 – 67.
- WANG Xiaojun, BI Sheng, XU Yunkun, et al. Short-term load forecasting based on data mining techniques and support vector machines[J]. *Electrical Measurement and Instrumentation*, 2016, 53(10): 62 – 67.
- [14] SUN M Y, ZHANG T Q, WANG Y, et al. Using Bayesian deep learning to capture uncertainty for residential net load forecasting[J]. *IEEE Transactions on Power Systems*, 2020, 35(1): 188 – 201.
- [15] RAHMAN A, SRIKUMAR V, SMITH A D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks[J]. *Applied Energy*, 2018(212): 372 – 385.
- [16] QIU X, REN Y, SUGANTHAN P N, et al. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting[J]. *Applied Soft Computing*, 2017(54): 246 – 255.
- [17] 吴慧军, 郭超雨, 苏承国, 等. 基于EEMD-GRU-MC的短期风功率组合预测方法[J]. *南方电网技术*, 2023, 17(2): 66 – 73.
- WU Huijun, GUO Chaoyu, SU Chengguo, et al. Combined prediction method for short-term wind power based on EEMD-GRU-MC[J]. *Southern Power System Technology*, 2023, 17(2): 66 – 73.
- [18] 李宏扬, 高丙朋. 基于改进VMD和SNS-Attention-GRU的短期光伏发电功率预测[J]. *太阳能学报*, 2023, 44(8): 292 – 300.
- LI Hongyang, GAO Bingpeng. Short-term PV power forecasting based on improved VMD and SNS-attention-GRU[J]. *Acta Energetica Solaris Sinica*, 2023, 44(8): 292 – 300.
- [19] 胡文慧, 苏欣, 姜林, 等. 考虑预测误差时空相依性的短期风电功率概率预测[J]. *南方电网技术*, 2023, 17(2): 137 – 144.
- HU Wenhui, SU Xin, JIANG Lin, et al. Probabilistic prediction of short-term wind power considering the temporal and spatial dependence of prediction errors[J]. *Southern Power System Technology*, 2023, 17(2): 137 – 144.
- [20] CAN W, JIN L, YONGHUA S, et al. Probabilistic forecasting of photovoltaic generation: an efficient statistical approach[J]. *IEEE Transactions on Power Systems*, 2017, 32(3): 2471 – 2472.
- [21] ANTONANZA N, OSORIO R, ESCOBAR R, et al. Review of photovoltaic power forecasting[J]. *Solar Energy*, 2016, 13(20): 78 – 111.
- [22] PENG S W, HAN W C, JIA G Z. Pearson correlation and transfer entropy in the Chinese stock market with time delay[J]. *Data Science and Management*, 2022, 5(3): 117 – 123.
- [23] ZHAO X H, HUANG G H, LU C, et al. Impacts of climate change on photovoltaic energy potential: a case study of China[J]. *Applied Energy*, 2020(280): 115888.1 – 115888.12.
- [24] MÜLLER J, FOLINI D, WILD M, et al. CMIP-5 models project photovoltaics are a no-regrets investment in Europe irrespective of climate change[J]. *Energy*, 2018(171): 135 – 148.
- [25] SHAKER H, ZAREIPOUR H, WOOD D. Impacts of large-scale wind and solar power integration on California's net electrical load[J]. *Renewable and Sustainable Energy Reviews*, 2016(58): 761 – 774.
- [26] RESHEF N D, RESHEF A Y, FINUCANE K H, et al. Detecting novel associations in large data sets[J]. *Science*, 2011, 334(6062): 1518 – 1524.
- [27] CHO K, MERRIENBOER V B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches[J]. *Computer Science*, 2014(140): 1246 – 1259.
- [28] WANG Y, LIAO W, CHANG Y. Gated recurrent unit network-based short-term photovoltaic forecasting[J]. *Energies*, 2018, 11(8): 1131 – 1143.
- [29] LU Z, SIYUE L, YIFENG D, et al. Probability prediction of short-term user-level load based on random forest and kernel density estimation[J]. *Energy Reports*, 2022, 8(S5): 1130 – 1138.

收稿日期: 2023-12-13; 网络首发日期: 2024-06-24

作者简介:

刘蓉晖(1975), 女, 教授, 硕士生导师, 博士, 研究方向为电能质量和电机磁场分析, liuronghui@126.com;

石炬烽(1998), 男, 硕士研究生, 研究方向为净负荷预测, shijufeng@126.com;

孙改平(1984), 女, 通信作者, 副教授, 硕士生导师, 博士, 研究方向为电力系统自动化和新能源预测, sunfrog2002@163.com。