

DOI: 10.19783/j.cnki.pspc.231097

基于因果正则化极限学习机的风电功率短期预测方法

杨茂¹, 张书天¹, 王勃²

(1. 现代电力系统仿真控制与绿色电能新技术教育部重点实验室(东北电力大学), 吉林 吉林 132012;
2. 新能源与储能运行控制国家重点实验室(中国电力科学研究院有限公司), 北京 100192)

摘要: 随着风电并网比例的逐年提高, 电力系统对风电功率预测的准确性和稳定性提出了更高要求。对于同一风电场而言, 为了避免不同特征选择方法所选择的风电场特征子集不同, 从因果关系的角度出发, 提出了一种基于因果正则化极限学习机(causal regularized extreme learning machine, CRELM)的风电功率短期预测方法。首先将极限学习机(extreme learning machine, ELM)建模为结构因果模型(structural causal model, SCM), 在此基础上计算隐藏层神经元与输出层神经元之间的平均因果效应向量。然后将该平均因果效应向量与输出层权重相结合构成因果正则化项, 在最小化训练误差的同时最大化网络的因果关系, 以进一步提升模型的预测准确性和预测稳定性。最后, 以国内蒙西某风电场数据为例, 与采用特征选择或不采用特征选择的预测模型相对比, 验证了所提方法的有效性和适用性。

关键词: 特征选择; 因果正则化; 结构因果模型; 平均因果效应向量; 极限学习机

Short-term wind power forecasting method based on a causal regularized extreme learning machine

YANG Mao¹, ZHANG Shutian¹, WANG Bo²

(1. Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China; 2. State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, Beijing 100192, China)

Abstract: With the increasing proportion of wind power connected to the grid year by year, the power system has higher requirements for the accuracy and stability of wind power forecasting. For the same wind farm, to avoid different feature subsets of wind farms selected by different feature selection methods, this paper proposes a short-term wind power forecasting method based on a causal regularization extreme learning machine from the perspective of causality. First, the extreme learning machine (ELM) model is modeled as a structural causal model, and then the average causal effect vector between hidden layer neurons and output layer neurons is calculated. Then the average causal effect vector is combined with the weight of the output layer to form a causal regularization term. This minimizes the training error and maximizes the causal relationship of the network, further improving the forecasting accuracy and stability of the model. Finally, taking the data of a wind farm in Mengxi, China as an example, and comparing with the forecasting model with or without the feature selection method, the effectiveness and applicability of the proposed method are verified.

This work is supported by the National Key Research and Development Program of China "Prediction Technology for Large-scale Wind Power/PV Multi-time Scale Capability" (No. 2022YFB2403000).

Key words: feature selection; causal regularization; structure causal model; average causal effect vector; extreme learning machine

0 引言

由于全球化石燃料消耗会带来环境恶化、能源

枯竭等问题, 可再生能源日益受到人们的青睐。风能作为一种绿色清洁能源, 在全球范围内得到了快速发展, 在现代电力系统中的占比不断提高^[1-4]。然而, 由于风能本身的波动性、间歇性和随机性^[5-8], 大规模风电整合往往会对电网的安全性和稳定性产生负面影响。风电功率预测有利于减少风电并入电

基金项目: 国家重点研发计划项目资助“大规模风电/光伏多时间尺度供电能力预测技术”(2022YFB2403000)

力系统的不利影响,降低电网运行成本,提高电力系统的运行可靠性,有效确保电网安全^[9-12]。因此,对风电功率预测的研究具有重要的理论和工程实践意义。

目前针对风电功率预测的方法可以按照驱动模式分为机理驱动和数据驱动两种。机理驱动方法通常建立在对大气运动刻画的基础之上,具有明确的风电转化机理^[13],虽然不需要使用大量历史数据,但计算复杂,适用于中长期预测。数据驱动方法以数据为导向,传统的统计学方法(如自回归移动平均^[14],卡尔曼滤波^[15]等)则基于大量历史数据以及其他外生因素,通过挖掘数据背后隐藏的模式和规律来预测未来风电功率。统计学方法虽然具有模型简单、可解释性强等优点,但预测精度随着时间的增加而降低^[16]。近年来,随着高维信息和多源异构数据的出现,基于人工智能模型的数据驱动方法在风电功率预测领域越来越流行,主要包括支持向量机^[17]、神经网络^[18]、极限学习机(extreme learning machine, ELM)^[19-20]、长短期记忆神经网络^[21]等,这些模型通过提取输入与输出之间的非线性映射关系来实现风电功率预测。随着风电场数据来源的增多,输入特征会对基于数据驱动的风电功率预测模型的预测精度造成严重影响^[22-23]。因此,通常需要在模型训练之前采用特征选择对特征进行筛选,来降低模型训练时间和减少预测误差。

特征选择方法可以分为过滤式、包裹式和嵌入式3种^[24]。过滤式方法通常基于相关性指标来对原始特征进行筛选,以筛选出与输出相关性较强的特征子集作为输入参数^[25]。常见的指标有 Pearson 相关系数^[26]、互信息和最大信息系数^[27]、灰色关联度^[28]等。该方法由于计算复杂度低、操作简单等优点,在风电功率预测领域被广泛使用。然而,当将多种特征选择方法应用于同一数据集时,会出现不同方法筛选出不同特征子集的情况。为了避免这种情况,目前已开发出多种基于稳定性度量的特征选择方法,但这些方法无法保证稳定性度量本身是稳定的^[29]。

因果关系作为一种不同于相关性的数据关系,可以通过从多源异构数据中进行因果推断,在保证预测稳定性的同时提升预测精度。文献[30]认为由于因果关系自身的不变性和鲁棒性等特点,在风险函数中考虑因果关系有利于预测稳定性。文献[31]提出,若要将因果推断应用于神经网络预测系统,需要先将系统本身建模为结构因果模型(structural causal model, SCM)。文献[32]证明了可以将前馈神经网络建模为结构因果模型,并通过因果归因方法计算输入神经元对输出神经元的因果效应。文献

[33]通过将计算出的因果效应作为正则化项来提升预测精度。目前绝大多数文献集中于将前馈神经网络建模为结构因果模型。

ELM 作为一种单隐层前馈神经网络,相比于前馈神经网络易陷入局部极小值和收敛速度慢等问题,它计算效率高并且倾向于实现全局最优^[34]。因此,本文从因果关系的角度出发,提出了一种基于因果正则化极限学习机(causal regularized extreme learning machine, CRELM)的风电功率预测模型,具体步骤如下:首先将整个 ELM 模型建模为 SCM,并计算输入神经元对输出神经元的平均因果效应。然后对网络进行切片,切除输入层,保留隐藏层和输出层。接着将网络剩余部分建模为结构因果模型,并在此基础上计算每个隐藏层神经元对输出神经元的平均因果效应向量。最后将计算出的平均因果效应向量与输出层权重相结合作为因果正则化项,通过最大化网络的因果关系和最小化训练误差,来实现风电功率预测稳定性和预测精度之间的权衡。以蒙西某风电场为例,与采用特征选择或不采用特征选择的算法相对比,验证了本文所提方法的有效性和适用性。

1 结构因果模型与平均因果效应

1.1 前馈神经网络与结构因果模型

为了对网络实施干预操作以求出神经元之间的平均因果效应,首要前提是将神经网络建模为 SCM^[35]。SCM 是由一个 4 元组 (X, U, f, P_U) 构成。其中 X 表示内生变量即系统可观测变量; U 表示外生变量即系统随机噪声或未观测变量; f 表示一系列因果函数,用来确定模型内部的因果机制; P_U 表示外生变量的概率分布。一个 SCM 可以被一个有向无环图 $G = (V, E)$ 所替代,其中 V 表示内生变量 X , 每个顶点 v_i 表示一个内生变量; E 表示边,一条边代表一个因果函数 f , 表示顶点之间的因果关系。经文献[32]证明,前馈神经网络可以被建模成 SCM,并且可以进一步简化为只含有输入层和输出层的结构因果模型,如图 1 所示。

1.2 平均因果效应

当将网络建模成 SCM 后,计算神经元之间的平均因果效应需要借助因果推断中关于平均因果效应(average causal effect, ACE)的定义^[36]。然而,不同于传统的离散型随机变量之间的 ACE,神经网络函数通常是连续的。所以给定一个输入层 I_1 和一个输出层 I_n , 将输入神经元 $x_i \in I_1$ 干预为 α_i 时,对输出 $y \in I_n$ 的 ACE 的定义如式(1)所示。

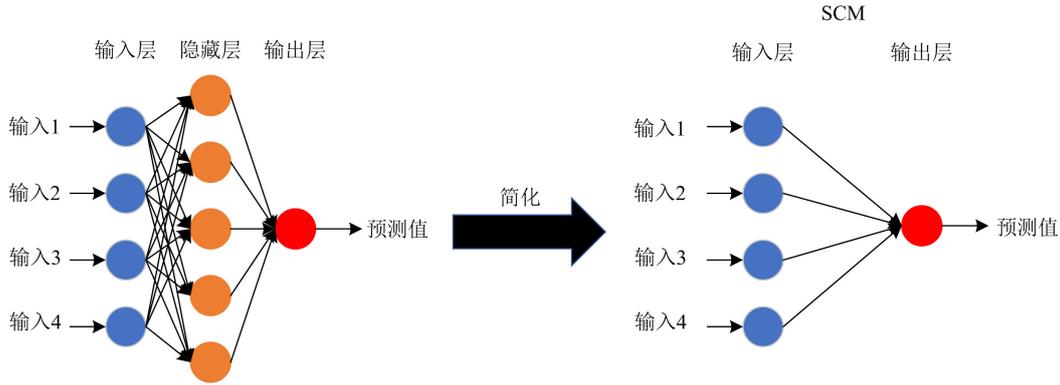


图1 前馈神经网络与结构因果模型

Fig. 1 Feedforward neural network and structural causal model

$$A = E[y | d(x_i = \alpha_i)] - b_{x_i} \quad (1)$$

式中: A 为平均因果效应; E 为期望; $d(\cdot)$ 为干预算子; $E[y | d(x_i = \alpha_i)]$ 为干预期望; b_{x_i} 为基线。

1.3 平均因果效应计算

由式(1)可知, ACE 的关键在于如何计算干预期望和选择合适的基线。传统的干预期望计算方法的局限性在于一次只干预一个特征值, 计算成本随着特征数量的增加而激增, 并且由于维数诅咒等问题, 使得最后的计算结果受到抽样偏差的影响。因此, 本文采用泰勒展开式框架作为一种替代机制来计算干预期望。

给定前馈神经网络及其简化后的结构因果模型 SCM: $M([l_1, l_n], f', U, P_U)$, 对应输出记作 $y = f(x_1, x_2, \dots, x_t)$, 其中 x_i 为输入层的第 i 个神经元; t 为输入神经元数量。在输入层 l_1 的干预均值向量处对因果函数 f'_y 进行二阶泰勒公式展开, 然后对方程两边同时取干预期望, 具体公式如式(2)和式(3)所示。

$$f'_y(l_1) = f'_y(\mu) + \nabla(f'_y(\mu)(l_1 - \mu))^T + \frac{1}{2}(l_1 - \mu)^T \nabla^2 f'_y(\mu)(l_1 - \mu) \quad (2)$$

$$E(f'_y(l_1) | d(x_i = \alpha_i)) \approx f'_y(\mu) + \frac{1}{2} T_r(\nabla^2 f'_y(\mu) E[(l_1 - \mu)(l_1 - \mu)^T | d(x_i = \alpha_i)]) \quad (3)$$

式中: $l_1 = [x_1, x_2, \dots, x_n]$ 为输入神经元向量; $\mu = [\mu_1, \mu_2, \dots, \mu_n]^T$ 为均值向量, 其中 $\mu_j = E[x_j | d(x_i = \alpha_i)]$, $\forall x_j \in l_1$; ∇, ∇^2 为求导算子; T_r 为矩阵迹算子; $E[f'_y(l_1) | d(x_i = \alpha_i)]$ 为干预期望; $E[(l_1 - \mu)(l_1 - \mu)^T | d(x_i = \alpha_i)]$ 为干预协方差。

在该近似计算中, 干预均值是指对另外一个不同于自身神经元的 x_i 执行干预操作时自身神经元

$x_j (j \neq i)$ 对应的观测均值。选择在干预均值处进行泰勒展开的原因在于, 当干预神经网络中的一个神经元时, 其他神经元的概率分布不会发生改变。这样就可以在没有其他神经元干扰的情况下进行干预从而计算出对应输入神经元对输出神经元的 ACE。

对于一个连续变量来说, 基线的存在用于识别神经元之间真正的因果效应。常见的基线选择方式有多种^[37], 可能是固定不变的值, 也有可能依赖于特定输入。本文选择依赖于特定特征 x_j 的基线, 如式(4)所示。

$$b_{x_i} = E_{x_i}[E[y | d(x_i = \alpha)]] \quad (4)$$

式中, E_{x_i} 为输入神经元 x_i 的期望。

式(4)意味着取所有干预期望的平均值作为基线。通过对所有干预的因果效应取平均值, 可以将干预期间网络内发生的副作用计算在内, 进而避免 ACE 计算过程中因特征之间存在较大差异而导致的错误。

2 因果正则化极限学习机

2.1 极限学习机

ELM 作为一种单隐藏层前馈神经网络, 主要通过两个阶段来进行训练^[34]: 第 1 阶段是采用随机特征映射, 使得隐藏层参数 (w, b) 根据任何连续的概率分布随机生成, 且独立于训练数据。第 2 阶段是通过最小化平方误差意义上的近似误差来求解连接隐藏层和输出层的权重, 如式(5)所示。

$$\min_{\beta} \|H\beta - T\|^2 \quad (5)$$

式中: β 为输出层权重; H 为隐藏层输出矩阵(随机化矩阵); T 为训练数据目标矩阵。

在给定训练样本的情况下, 隐藏层输出矩阵 H 是已知的, 并且保持不变。所以式(5)转化为求解线

性系统 $\mathbf{H}\beta = \mathbf{T}$ 的最小范数最小二乘解 $\hat{\beta}$ ，如式(6)所示。

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

式中， \mathbf{H}^\dagger 表示隐藏层输出矩阵 \mathbf{H} 的 Moore-Penrose 广义逆矩阵。

2.2 因果正则化极限学习机

CRELM 首先计算隐藏层神经元与输出层神经元之间的 ACE，并将其与输出层权重相结合，然后将该项作为因果正则化项并引入负号，目的是为了在最小化训练误差的同时最大化网络的因果关系。

$$\arg \min_{\beta} \lambda \|\mathbf{y} - \mathbf{H}\beta\|^2 - \|\mathbf{C}\beta\|^2 \quad (7)$$

式中： λ 为惩罚参数； \mathbf{y} 为目标输出矩阵； \mathbf{C} 为每个隐藏层神经元与输出层神经元之间的平均因果效应向量； $\mathbf{C}\beta$ 为因果正则化项。

方程(7)等价于式(8)所示的约束优化问题。

$$\arg \min_{\beta} \lambda \|\mathbf{e}\|^2 - \|\mathbf{C}\beta\|^2, \text{ s.t. } \mathbf{y} - \mathbf{H}\beta = \mathbf{e} \quad (8)$$

式中， $\mathbf{e} = [e_1, e_2, \dots, e_N]^\top$ 表示 N 个训练误差。对应的拉格朗日目标函数为

$$L(\beta, \mathbf{e}, \xi) = \lambda \|\mathbf{e}\|^2 - \|\mathbf{C}\beta\|^2 + \xi^\top (\mathbf{y} - \mathbf{H}\beta - \mathbf{e}) \quad (9)$$

式中， L 为拉格朗日目标函数； ξ 为拉格朗日乘子的列向量。根据 KKT 条件，最优性条件如式(10)所示。

$$\begin{cases} \frac{\partial L}{\partial \beta} = 0 \Rightarrow -2\mathbf{C}^\top \mathbf{C}\beta - \mathbf{H}^\top \xi = 0 \\ \frac{\partial L}{\partial \mathbf{e}} = 0 \Rightarrow 2\lambda \mathbf{e} - \xi = 0 \\ \frac{\partial L}{\partial \xi} = 0 \Rightarrow \mathbf{y} - \mathbf{H}\beta - \mathbf{e} = 0 \end{cases} \quad (10)$$

通过式(11)求出因果正则化下的最优权重 β' 。

$$\beta' = \left(\mathbf{H}^\top \mathbf{H} - \frac{\mathbf{C}^\top \mathbf{C}}{\lambda} \right)^{-1} \mathbf{H}^\top \mathbf{T} \quad (11)$$

式(11)表明计算最优权重需要计算网络内部每个隐藏层神经元与输出神经元之间的平均因果效应向量 $\mathbf{C} = [C_1, C_2, \dots, C_m]^\top$ ，其中 m 为隐藏层神经元数量。所以首先对神经网络进行切片，切除输入层，保留网络剩余部分来创建以隐藏层为输入层的 SCM，然后在 SCM 基础上，使用隐藏层的每个神经元的干预期望和基线计算对应 ACE。然而，当采用式(1)计算单个神经元对输出的 ACE 时，会随着干预值的不同出现或正或负的情况。正 ACE 意味着神经元以一种正因果方式影响输出，负 ACE 意

味着神经元以一种负因果方式影响输出。为了使网络的因果关系最大化，应该兼顾正向和负向的 ACE^[38]，所以本文取由单个神经元 x_i 计算得出的 A 的中位数绝对值作为总因果效应 C_i ， $i = 1, 2, \dots, m$ ，如式(12)所示。

$$C_i = |M(A)| \quad (12)$$

式中， $M(\cdot)$ 为中位数。

由于中位数不受异常值的影响，并且可以代表或正或负的 ACE，所以通过对中位数取绝对值来同时考虑两种不同效应，以此来获得输入神经元对输出神经元的总因果效应。

3 基于 CRELM 的风电功率预测流程

本文提出的基于 CRELM 模型的风电功率预测技术路线如图 2 所示。总体分为两个阶段：第 1 阶段将 ELM 建模为 SCM 以求出隐藏层神经元和输出层神经元之间的平均因果效应向量，并与输出层权重相结合构成因果正则化项；第 2 阶段求出最优权重，然后用训练好的 ELM 模型预测未来风电功率，具体步骤如下。

1) 由式(1)可知，干预期望计算需要对样本集进行干预，即将每个特征值设置成 α_i 。本文选择从每个特征的概率分布入手，以生成对应干预后的样本集。由于不同的天气特征以不同的方式分布^[39]，例如，风速通常服从威布尔分布，而风向通常遵循多模态分布，因此，对于每个特征，无论分布是否已知，都采用非参数概率建模采样方法^[40]来生成干预样本集。

2) 首先将 ELM 建模为 SCM，在 SCM 上输入训练样本及由 1) 生成的干预训练样本集来计算随机化矩阵 \mathbf{H} 和干预随机化矩阵 \mathbf{H}' 。然后对网络进行切片，切除输入层，保留隐藏层和输出层，并将这两层重新建模为 SCM。接着将 \mathbf{H} 和 \mathbf{H}' 作为输入，通过式(1)来计算每个隐藏层神经元对输出层神经元的 ACE。

3) 随后根据 2) 中计算得到的隐藏层神经元对输出层神经元的 ACE 和式(12)来求出两层之间的总因果效应 C_i 及其构成的平均因果效应向量 \mathbf{C} ，并与输出层权重 β 相结合构成因果正则化项。紧接着将训练误差与因果正则化项共同作为损失函数，由式(11)求出网络的最优权重 β' 。最后将测试样本输入至训练好的 CRELM 模型中来预测未来的风电功率。

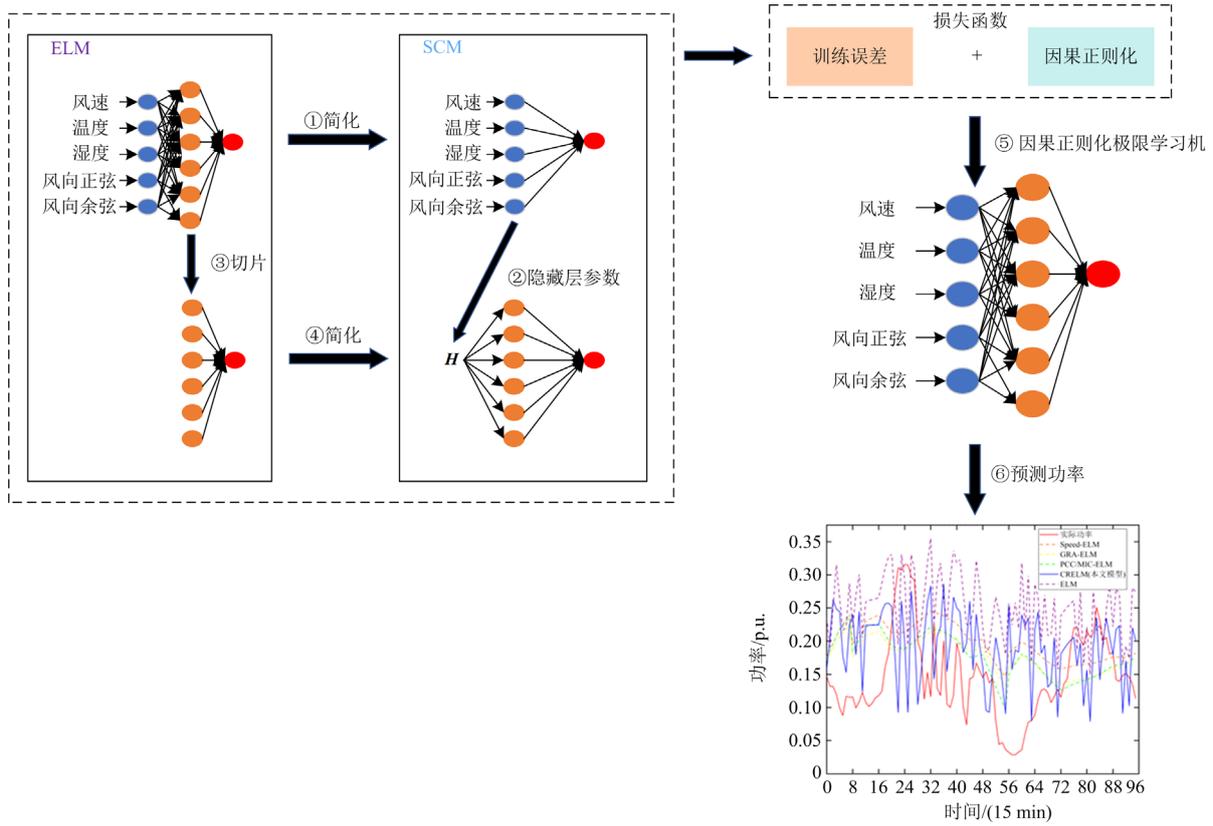


图 2 基于 CRELM 的风电功率预测技术路线图

Fig. 2 Wind power forecasting technology roadmap based on CRELM

4 算例分析

4.1 相关性指标

本文选择皮尔逊相关系数(Pearson correlation coefficient, PCC)、最大信息系数(maximal information coefficient, MIC)和灰色关联度分析(grey relational analysis, GRA)作为特征选择方法。

$$\rho_i = \frac{\text{cov}(X_0, X_i)}{\sqrt{D(X_0) \times D(X_i)}} \quad (13)$$

$$r_i = \frac{1}{m} \sum_{k=1}^m \xi_i(k) \quad (14)$$

$$\left\{ \begin{aligned} \xi_i(k) &= \frac{\min_i \min_k |X_0(k) - X_i(k)| + U}{|X_0(k) - X_i(k)| + U} \\ U &= \eta \times \max_i \max_k |X_0(k) - X_i(k)| \end{aligned} \right. \quad (15)$$

$$L(X_i, X_0) = \max_{a \times b < B} \frac{I(X_i, X_0)}{\log_2 \min(a, b)} \quad (16)$$

$$I(X_i, X_0) = H(X_0) - H(X_0 | X_i) \quad (17)$$

$$H(X_0) = - \sum_{x_0 \in X_0} P_{X_0}(x_0) \log P_{X_0}(x_0) \quad (18)$$

$$\left\{ \begin{aligned} H(X_0 | X_i) &= - \sum_{x_i \in X_i} P_{X_i}(x_i) \left(\sum_{x_0 \in X_0} B \log B \right) \\ B &= P_{X_0 | X_i}(X_0 | X_i) \end{aligned} \right. \quad (19)$$

式中: ρ_i 代表第 i 个特征与目标序列之间的 PCC; cov 为协方差; D 为方差; X_0 为目标序列; X_i 为第 i 个特征; $X_0(k)$ 代表特征序列 X_0 中的第 k 个样本, $k=1, 2, \dots, n$; η 为分辨系数, $\eta \in (0, 1)$, η 取值越小, 则 GRA 差异越大, 其区分能力越强, 本文 η 取 0.5; $L(X_i, X_0)$ 为 X_0 与 X_i 之间的最大信息系数; a 、 b 分别为沿 x 、 y 轴方向上的网络数量; $P_{X_0}(x_0)$ 为 $X_0 = x_0$ 时的概率。

4.2 数据描述与评价指标

采用国内蒙西某风电场的运行数据进行实例验证。数据包括 2020 年 3 月 1 日到 2020 年 9 月 1 日的数值天气预报(numerical weather prediction, NWP)数据和实测功率数据, 其中 NWP 数据包括风速、温度、湿度和风向。数据分辨率为 15 min, 风电场装机容量为 130.5 MW。选择 3—7 月的数据作为训练集, 8 月的数据作为测试集。采用含有 8 个隐藏层神经元的 ELM 模型作为预测模型, 惩罚参数 λ 值

越小,越能反映网络的因果关系,因此本文 λ 取1.2。以风速、温度、湿度归一化后的值和风向正弦与风向余弦这5个特征作为输入,预测未来一天24h的风电功率。

本文从预测精度指标和稳定性指标(stability accuracy, STA)^[41]两方面,对提出的风电功率短期预测模型的性能进行评估,其中预测精度的指标包括归一化均方根误差(normalized root mean square error, NRMSE)、归一化平均绝对误差(normalized mean absolute error, NMAE)。

NRMSE 指标用于评估误差的离散程度,其值越接近0越好,计算公式如式(20)所示。

$$N_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{nP_{cap}^2}} \quad (20)$$

式中: y_i 为*i*时刻的实际功率; \hat{y}_i 为*i*时刻的预测功率; n 为测试集长度; P_{cap} 为实际风电场的开机容量,由于实际风电场开机容量难以获取,所以本文以装机容量代替。

NMAE 指标用于评估误差平均分布情况,其值越接近0越好,计算公式如式(21)所示。

$$N_{MAE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i| \times 100\%}{P_{cap}} \quad (21)$$

STA 用于评估模型稳定性,值越接近0代表模型越稳定,计算公式如式(22)所示。

$$S_{TA} = \frac{D(|y_i - \hat{y}_i|)}{P_{cap}} \quad (22)$$

4.3 NWP 重要性分析

表1显示了采用PCC、MIC和GRA3种特征选择方法后各个NWP特征与功率的相关性程度。

表1 NWP 特征与功率相关程度

Table 1 Degree of correlation between NWP features and power

NWP 特征	PCC	MIC	GRA
风速	0.57	0.3119	0.6642
温度	-0.15	0.1865	0.6381
湿度	0.041	0.1064	0.6501
风向正弦	-0.038	0.0725	0.4118
风向余弦	0.01	0.0724	0.4153

由表1可以看出,3种特征选择方法下与功率最相关的特征都是风速,相关性程度分别是0.57、0.3119和0.6642。当选择第2个相关特征时出现分歧,PCC和MIC表明温度与功率相关性仅次于风速,相关性程度分别是-0.15和0.1865,而GRA表明湿度与功率相关性仅次于风速,相关性程度为0.6501。为了兼顾各种情况,本文将选择的特征分为以下5种类型:(1)仅考虑风速;(2)考虑风速和

温度;(3)考虑风速和湿度;(4)考虑风速、温度和湿度;(5)考虑所有特征。

4.4 预测性能分析

4.4.1 预测曲线分析

图3为基于上述5种特征类型的ELM模型以及本文所提模型对风电场未来一天(96个数据点)的预测曲线对比图。Speed-ELM模型代表选择的特征仅含有风速,GRA-ELM模型代表选择的特征是风速和湿度,PCC/MIC-ELM模型代表选择的特征是风速和温度,ELM和CRELM模型代表考虑全部特征。由图3可以看出,CRELM相较于其他各种特征下的ELM模型,整体上能够更好地追踪实际功率曲线的走势。

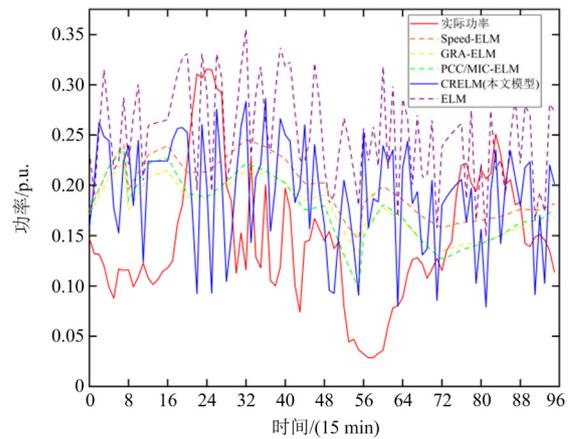


图3 不同模型的风电功率预测曲线

Fig. 3 Wind power forecasting curves of different models

4.4.2 预测指标分析

为了进一步验证本文方法的有效性,将5种特征类型分别与ELM和CRELM模型相结合进行预测指标对比。由于机器学习模型本身具有随机性,所以对最终预测结果运行5次并取平均值。具体结果如表2和图4所示。

表2 不同模型预测指标

Table 2 Evaluation index of different forecasting models

预测模型	NRMSE	NMAE	STA
Speed-ELM	0.1366	0.1089	0.0068
GRA-ELM	0.1347	0.1082	0.0064
PCC/MIC-ELM	0.1322	0.1083	0.0068
STH-ELM	0.1415	0.1103	0.0076
ELM	0.1571	0.1327	0.0074
CRELM(本文模型)	0.1290	0.1080	0.0050
Speed-CRELM	0.1247	0.1042	0.0047
GRA-CRELM	0.1297	0.1071	0.0053
PCC/MIC-CRELM	0.1287	0.1074	0.0048
STH-CRELM	0.1305	0.1070	0.0056

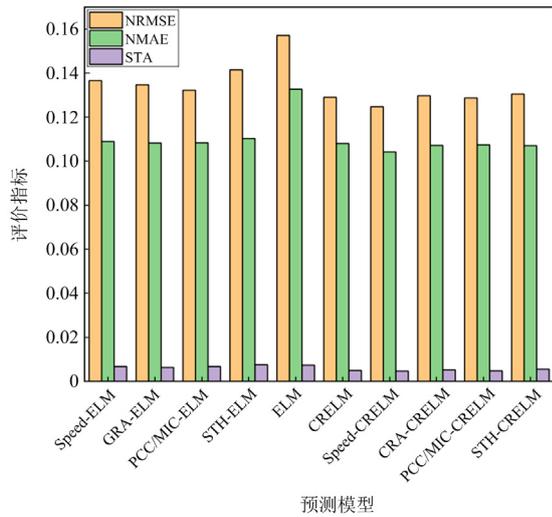


图 4 不同预测模型评价指标对比图

Fig. 4 Comparison chart of evaluation indexes of different forecasting models

预测精度指标即 NRMSE 和 NMAE 越小代表预测性能越好。由图 4 和表 2 可以看出, 若仅考虑风速, 则对应 ELM 模型的 NRMSE 和 NMAE 分别是 0.1366 和 0.1089。当考虑第 2 个相关特征时, 若考虑温度, 则对应 NRMSE 和 NMAE 分别是 0.1322 和 0.1083, 较 ELM 模型分别降低了 0.0249 和 0.0244; 若考虑湿度, 则对应 NMRSE 和 NMAE 分别是 0.1347 和 0.1082, 较 ELM 模型分别降低了 0.0224 和 0.0245。STH-ELM 模型代表选择的特征是风速、温度和湿度, 当考虑第 3 个相关特征时即同时考虑温度和湿度, 则对应 NMRSE 和 MMAE 较 ELM 模型

分别降低了 0.0156 和 0.0224。以上结果表明, 经特征选择后的预测性能虽然有所提高, 但选择不同的特征对应性能提升的幅度不同, 说明不同的特征选择方法对预测精度的影响不同, 容易产生分歧。从整体上看, 本文所提模型 CRELM 选择全部特征时对应 NRMSE 和 NMAE 分别是 0.1290 和 0.1080, NRMSE 较其余 ELM 模型分别降低了 0.0076、0.0057、0.0032、0.0125 和 0.0281, NMAE 分别降低了 0.0009、0.0002、0.0003、0.0023 和 0.0247。从局部上看, 当两种模型选择的特征一致时, CRELM 相较于 ELM, NRMSE 分别降低了 0.0119、0.005、0.0035 和 0.011; NMAE 分别降低了 0.0047、0.0011、0.0009 和 0.0033。结果表明 CRELM 能够在避免选择不同特征子集的同时提升风电功率预测精度。

接着讨论稳定性指标即 STA, 该指标越小代表模型稳定性越好。由表 2 可以看出, 不同特征对应 STA 相差不大甚至相同。从整体上看, 本文所提模型对应 STA 相较于其余 ELM 模型分别降低了 0.0018、0.0014、0.0026 和 0.0024。这是因为 CRELM 模型考虑因果关系, 由于其自身鲁棒性、不变性的特点, 与其他基于相关性的预测模型相比, CRELM 模型在运行的过程中更稳定。

为了进一步验证方法的通用性, 表 3 和表 4 分别记录了其余 6 个风电场下不同预测模型的评价指标。由表 3 和表 4 可以看出, 相较于不同特征下的 ELM 预测模型, CRELM 模型在不同风电场下均具有较高的预测精度和较好的预测稳定性。总体来说, CRELM 模型对不同风电场具有较强的适用性。

表 3 1、2、3 号风电场下模型评价指标

Table 3 Evaluation index of forecasting model under wind farms 1, 2 and 3

预测模型	1 号风电场			2 号风电场			3 号风电场		
	NRMSE	NMAE	STA	NRMSE	NMAE	STA	NRMSE	NMAE	STA
Speed-ELM	0.1382	0.1093	0.0071	0.1366	0.1089	0.0068	0.1357	0.1086	0.0066
GRA-ELM	0.1379	0.1094	0.0072	0.1380	0.1091	0.0071	0.1375	0.1094	0.0070
PCC/MIC-ELM	0.1386	0.1096	0.0071	0.1363	0.1107	0.0068	0.1402	0.1145	0.0066
STH-ELM	0.1356	0.1106	0.0062	0.1382	0.1105	0.0069	0.1342	0.1107	0.0067
ELM	0.1355	0.1214	0.0067	0.1460	0.1214	0.0067	0.1429	0.1278	0.0075
CRELM(本文模型)	0.1313	0.1086	0.0055	0.1346	0.1085	0.0061	0.1333	0.1083	0.0060
Speed-CRELM	0.1290	0.1080	0.0050	0.1301	0.1081	0.0052	0.1296	0.1063	0.0055
GRA-CRELM	0.1321	0.1081	0.0058	0.1311	0.1085	0.0057	0.1307	0.1070	0.0056
PCC/MIC-CRELM	0.1329	0.1098	0.0056	0.1330	0.1102	0.0055	0.1373	0.1147	0.0057
STH-CRELM	0.1340	0.1086	0.0061	0.1339	0.1085	0.0059	0.1334	0.1103	0.0056

表 4 4、5、6 号风电场下模型评价指标
Table 4 Evaluation index of forecasting model under wind farms 4, 5 and 6

预测模型	4 号风电场			5 号风电场			6 号风电场		
	NRMSE	NMAE	STA	NRMSE	NMAE	STA	NRMSE	NMAE	STA
Speed-ELM	0.1498	0.1234	0.0072	0.1355	0.1157	0.0050	0.1463	0.1180	0.0075
GRA-ELM	0.1579	0.1324	0.0074	0.1402	0.1217	0.0049	0.1506	0.1229	0.0076
PCC/MIC-ELM	0.1563	0.1301	0.0075	0.1523	0.1257	0.0074	0.1520	0.1251	0.0074
STH-ELM	0.1589	0.1298	0.0084	0.1544	0.1273	0.0076	0.1532	0.1253	0.0078
ELM	0.1662	0.1406	0.0079	0.1525	0.1245	0.0077	0.1538	0.1292	0.0069
CRELM(本文模型)	0.1483	0.1218	0.0070	0.1277	0.1081	0.0046	0.1454	0.1212	0.0065
Speed-CRELM	0.1317	0.1101	0.0052	0.1214	0.1027	0.0042	0.1347	0.1086	0.0063
GRA-CRELM	0.1378	0.1159	0.0056	0.1255	0.1049	0.0048	0.1389	0.1166	0.0057
PCC/MIC-CRELM	0.1386	0.1121	0.0066	0.1241	0.1070	0.0053	0.1351	0.1103	0.0061
STH-CRELM	0.1456	0.1198	0.0068	0.1277	0.1081	0.0046	0.1364	0.1110	0.0063

5 结论

特征选择方法虽然可以降低模型训练时间和减少预测误差，但不同方法选择的特征子集不同容易产生分歧。因此，本文提出了一种基于因果正则化的极限学习机模型，将 ELM 模型建模为 SCM，在此基础上计算不同神经元之间的平均因果效应向量，并与输出层权重相结合构成因果正则化项，在最小化训练误差的同时最大化网络的因果关系，以预测未来一天的风电功率。得出的结论如下。

1) 从预测精度和预测稳定性两个角度出发，与采用常见特征选择方法的 ELM 模型和传统 ELM 模型相比，本文所提模型对应 NRMSE、NMAE 和 STA 分别是 0.1290、0.1080 和 0.0050，较传统 ELM 模型分别降低了 0.0281、0.0247 和 0.0024；与采用特征选择方法后的 ELM 模型相比，平均降低了 0.0055、0.0046 和 0.0016，预测精度和预测稳定性均有所提升。

2) 由于目前神经网络建模为 SCM 的数学基础仅存在于前馈神经网络，下一步将讨论将因果正则化思想应用于循环神经网络中以进一步提升预测精度和预测稳定性。

参考文献

[1] 杨茂, 张书天, 王天硕, 等. 基于 IKLIEP-四分位模型的风电场异常数据识别算法[J]. 高电压技术, 2023, 49(7): 2952-2960.
YANG Mao, ZHANG Shutian, WANG Tianshuo, et al. Identification algorithm of wind farm abnormal data based on IKLIEP-quartile model[J]. High Voltage Engineering, 2023, 49(7): 2952-2960.

[2] VERMA N, KUMAR N, GUPTA S, et al. Review of sub-synchronous interaction in wind integrated power systems: classification, challenges, and mitigation techniques[J]. Protection and Control of Modern Power Systems, 2023, 8(2): 277-302.

[3] WANG Cong, LI Yan, FAN Yiwen, et al. Development of wind-energy modeling technology and standards[J]. Global Energy Interconnection, 2022, 5(2): 206-216.

[4] 葛维春, 张诗钊, 崔岱, 等. 海上风电送出与就地消纳技术差异综述[J]. 电测与仪表, 2022, 59(5): 23-32.
GE Weichun, ZHANG Shitan, CUI Dai, et al. Summary of differences between offshore wind power transmission and local consumption technology[J]. Electrical Measurement & Instrumentation, 2022, 59(5): 23-32.

[5] 戴千斌, 黄南天. 短期风电功率预测误差修正研究综述[J]. 东北电力大学学报, 2023, 43(2): 1-7.
DAI Qianbin, HUANG Nantian. Review of research on error correction of short-term wind power forecasting[J]. Journal of Northeast Electric Power University, 2023, 43(2): 1-7.

[6] 鲍颜红, 张金龙, 衣立东, 等. 含大规模风电电力系统的安全稳定风险预防控制方法[J]. 电力系统自动化, 2022, 46(13): 187-194.
BAO Yanhong, ZHANG Jinlong, YI Lidong, et al. Prevention and control method of security and stability risk for power system with large-scale wind power integration[J]. Automation of Electric Power Systems, 2022, 46(13): 187-194.

[7] 杨贤东, 袁旭峰, 熊炜, 等. 考虑源荷不确定性的风光火储系统低碳经济调度[J]. 智慧电力, 2022, 50(8): 22-29, 53.
YANG Xiandong, YUAN Xufeng, XIONG Wei, et al. Low-carbon economic dispatch of wind-solar-fired-storage

- system considering source-load uncertainty[J]. *Smart Power*, 2022, 50(8): 22-29, 53.
- [8] ZHAO Lingyun, WANG Zhuoyu, CHEN Tingxi, et al. Missing interpolation model for wind power data based on the improved CEEMDAN method and generative adversarial interpolation network[J]. *Global Energy Interconnection*, 2023, 6(5): 517-529.
- [9] 杨茂, 于欣楠. 考虑风电场功率爬坡的超短期组合预测[J]. *东北电力大学学报*, 2022, 42(1): 63-70.
YANG Mao, YU Xinnan. An ultra-short term combined prediction considering wind farm power climbing[J]. *Journal of Northeast Electric Power University*, 2022, 42(1): 63-70.
- [10] 崔昊杨, 孙昊宇, 杨程, 等. 计及空间相关性的轻量级风电功率预测方法[J]. *智慧电力*, 2022, 50(8): 7-13, 117.
CUI Haoyang, SUN Haoyu, YANG Cheng, et al. Lightweight wind power prediction method considering spatial correlation[J]. *Smart Power*, 2022, 50(8): 7-13, 117.
- [11] 翟保豫, 张龙, 徐志, 等. 基于 WRF 模式和风速误差修正的中期风电功率预测方法[J]. *智慧电力*, 2023, 51(7): 31-38.
ZHAI Baoyu, ZHANG Long, XU Zhi, et al. Mid-term wind power forecasting based on WRF mode and wind speed error correction[J]. *Smart Power*, 2023, 51(7): 31-38.
- [12] 孙蓉, 李强, 罗海峰, 等. 基于自适应移动平滑与时间卷积网络误差修正的风电功率预测[J]. *全球能源互联网*, 2022, 5(1): 11-22.
SUN Rong, LI Qiang, LUO Haifeng, et al. Wind power forecasting based on error correction using adaptive moving smoothing and time convolution network[J]. *Journal of Global Energy Interconnection*, 2022, 5(1): 11-22.
- [13] 杨茂, 张罗宾. 基于数据驱动的超短期风电功率预测综述[J]. *电力系统保护与控制*, 2019, 47(13): 171-186.
YANG Mao, ZHANG Luobin. Review on ultra-short term wind power forecasting based on data-driven approach[J]. *Power System Protection and Control*, 2019, 47(13): 171-186.
- [14] 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.
- [15] 杨茂, 刘慧宇, 孙勇, 等. 基于原子稀疏分解和支持向量机的风电功率实时预测研究[J]. *东北电力大学学报*, 2020, 40(3): 1-7.
YANG Mao, LIU Huiyu, SUN Yong, et al. A study on real-time prediction of wind power based on atomic sparse decomposition and support vector machine[J]. *Journal of Northeast Electric Power University*, 2020, 40(3): 1-7.
- [16] 薛禹胜, 郁琛, 赵俊华, 等. 关于短期及超短期风电功率预测的评述[J]. *电力系统自动化*, 2015, 39(6): 141-151.
XUE Yusheng, YU Chen, ZHAO Junhua, et al. A review on short-term and ultra-short-term wind power prediction[J]. *Automation of Electric Power Systems*, 2015, 39(6): 141-151.
- [17] 赵永宁, 李卓, 叶林, 等. 基于时空相关性的风电功率超短期自适应预测方法[J]. *电力系统保护与控制*, 2023, 51(6): 94-105.
ZHAO Yongning, LI Zhuo, YE Lin, et al. A very short-term adaptive wind power forecasting method based on spatio-temporal correlation[J]. *Power System Protection and Control*, 2023, 51(6): 94-105.
- [18] 范高锋, 王伟胜, 刘纯, 等. 基于神经网络的风电功率预测[J]. *中国电机工程学报*, 2008, 28(34): 118-123.
FAN Gaofeng, WANG Weisheng, LIU Chun, et al. Wind power prediction based on artificial neural network[J]. *Proceedings of the CSEE*, 2008, 28(34): 118-123.
- [19] 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(增刊): 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(S1): 146-153.
- [20] 韦权, 汤占军. 基于 SSA-VMD-SE-KELM 结合蒙特卡洛法的风电功率区间预测[J]. *智慧电力*, 2022, 50(9): 59-66.
WEI Quan, TANG Zhanjun. Wind power range prediction based on SSA-VMD-SE-KELM combined with Monte Carlo method[J]. *Smart Power*, 2022, 50(9): 59-66.
- [21] 吉兴全, 曾若梅, 张玉敏, 等. 基于注意力机制的 CNN-LSTM 短期电价预测[J]. *电力系统保护与控制*, 2022, 50(17): 125-132.
JI Xingquan, ZENG Ruomei, ZHANG Yumin, et al. CNN-LSTM short-term electricity price prediction based on an attention mechanism[J]. *Power System Protection and Control*, 2022, 50(17): 125-132.
- [22] LIU Hui, CHEN Chao. Data processing strategies in wind energy forecasting models and applications: a comprehensive review[J]. *Applied Energy*, 2019, 249: 392-408.
- [23] 潘霄, 张明理, 刘德宝, 等. 基于鲁棒多标签生成对抗的风电场日前出力区间预测[J]. *电力系统自动化*, 2022, 46(10): 216-223.
PAN Xiao, ZHANG Mingli, LIU Debao, et al. Interval

- prediction of wind farm day-ahead output based on robust multi-label generative adversarial[J]. *Automation of Electric Power Systems*, 2022, 46(10): 216-223.
- [24] 朱琼锋, 李家腾, 乔骥, 等. 人工智能技术在新能源功率预测的应用及展望[J]. *中国电机工程学报*, 2023, 43(8): 3027-3048.
- ZHU Qiongfeng, LI Jiateng, QIAO Ji, et al. Application and prospect of artificial intelligence technology in renewable energy forecasting[J]. *Proceedings of the CSEE*, 2023, 43(8): 3027-3048.
- [25] 唐新姿, 顾能伟, 黄轩晴, 等. 风电功率短期预测技术研究进展[J]. *机械工程学报*, 2022, 58(12): 213-236.
- TANG Xinzi, GU Nengwei, HUANG Xuanqing, et al. Progress on short term wind power forecasting technology[J]. *Journal of Mechanical Engineering*, 2022, 58(12): 213-236.
- [26] ZHANG Zhendong, QIN Hui, LIU Yongqi, et al. Long short-term memory network based on neighborhood gates for processing complex causality in wind speed forecasting[J]. *Energy Conversion and Management*, 2019, 192: 37-51.
- [27] LIU Xiangjie, ZHANG Hao, KONG Xiaobing, et al. Wind speed forecasting using deep neural network with feature selection[J]. *Neurocomputing*, 2020, 397: 393-403.
- [28] 彭晨宇, 陈宁, 高丙团. 结合多重聚类和分层聚类的超短期风电功率预测方法[J]. *电力系统自动化*, 2020, 44(2): 173-180.
- PENG Chenyu, CHEN Ning, GAO Bingtuan. Ultra-short-term wind power forecasting method combining multiple clustering and hierarchical clustering[J]. *Automation of Electric Power Systems*, 2020, 44(2): 173-180.
- [29] KHAIRE U M, DHANALAKSHMI R. Stability of feature selection algorithm: a review[J]. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(4): 1060-1073.
- [30] BÜHLMANN P. Invariance, causality and robustness[J]. *Statistical Science*, 2020, 35(3): 404-426.
- [31] GEIGER A, WU Z, LU H, et al. Inducing causal structure for interpretable neural networks[C]// *International Conference on Machine Learning*, June 17-23, 2022, Baltimore, Maryland, USA: 7324-7338.
- [32] CHATTOPADHYAY A, MANUPRIYA P, SARKER A, et al. Neural network attributions: a causal perspective[C]// *International Conference on Machine Learning*, June 9-15, 2019, California, USA: 981-990.
- [33] BAHADORIM T, CHALUPKA K, CHOI E, et al. Causal regularization[J]. *arXiv preprint arXiv: 1702.02604*, 2017.
- [34] HUANG Gao, HUANG Guangbin, SONG Shiji, et al. Trends in extreme learning machines: a review[J]. *Neural Networks*, 2015, 61: 32-48.
- [35] PERAL J. Causal inference in statistics: an overview[J]. *Statistics Surveys*, 2009, 3: 96-146.
- [36] YAO Liuyi, CHU Zhixuan, LI Sheng, et al. A survey on causal inference[J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2021, 15(5): 1-46.
- [37] HAUG J, ZURM S, EL-JIZ P, et al. On baselines for local feature attributions[J]. *arXiv preprint arXiv: 2101.00905*, 2021.
- [38] CHU Zhixuan, HUANG Jianming, LI Ruopeng, et al. Causal effect estimation: recent advances, challenges, and opportunities[J]. *arXiv preprint arXiv: 2302.00848*, 2023.
- [39] OPTIS M, PERR-SAUER J. The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production[J]. *Renewable and Sustainable Energy Reviews*, 2019, 112: 27-41.
- [40] 杨茂, 董昊. 基于数值天气预报风速和蒙特卡洛法的短期风电功率区间预测[J]. *电力系统自动化*, 2021, 45(5): 79-85.
- YANG Mao, DONG Hao. Short-term wind power interval prediction based on wind speed of numerical weather prediction and Monte Carlo method[J]. *Automation of Electric Power Systems*, 2021, 45(5): 79-85.
- [41] TIAN Chaonan, NIU Tong, WEI Wei. Developing a wind power forecasting system based on deep learning with attention mechanism[J]. *Energy*, 2022, 257: 124750.

收稿日期: 2023-08-25; 修回日期: 2023-12-29

作者简介:

杨茂(1982—), 男, 通信作者, 博士, 教授, 研究方向为电力系统稳定分析与新能源并网等; E-mail: yangmao820@163.com

张书天(1999—), 男, 硕士研究生, 研究方向为风电功率预测; E-mail: 2913302499@qq.com

王勃(1983—), 男, 硕士, 高级工程师(教授级), 主要从事电力气象、新能源发电功率预测技术研究与应用工作。E-mail: wangbo@epri.sgcc.com.cn

(编辑 周金梅)