

DOI: 10.19666/j.rlfed.202301023

基于相似日聚类的超短期光伏功率 组合预测模型

常青松¹, 杨昭², 杨熠辉², 雷阳², 何信林²

(1. 华能吉林发电有限公司九台电厂, 吉林 长春 130500;

2. 西安热工研究院有限公司, 陕西 西安 710054)

[摘要] 针对功率预测模型受光伏功率波动性影响导致预测精度低的问题, 提出一种基于相似日聚类的光伏功率预测组合模型。首先, 采取 k -means 聚类算法将原始功率数据按不同天气类型划分为晴天、雨天和多云 3 种相似日样本集, 并利用变分模态分解 (VMD) 对相似日样本进行分解; 其次, 采用卷积神经网络优化支持向量机 (CNN-SVM) 和双向长短时记忆 (BiLSTM) 神经网络 2 个单模型分别对分解后的功率数据进行预测叠加并将预测结果进行加权组合, 利用网格搜索 (GS) 算法寻找最优组合权重, 提升组合预测模型性能; 最后, 以澳大利亚某光伏电站 1 年实测数据为例, 验证所提出光伏功率预测模型的有效性。实验结果表明: 无论何种天气类型, 所提出模型均能很好地对光伏功率实现预测, 具有较强的适应性。

[关键词] 光伏功率预测; 卷积神经网络; 支持向量机; 长短时记忆神经网络; 网格搜索算法

[引用本文格式] 常青松, 杨昭, 杨熠辉, 等. 基于相似日聚类的超短期光伏功率组合预测模型[J]. 热力发电, 2023, 52(11): 123-131. CHANG Qingsong, YANG Zhao, YANG Yihun, et al. Ultrashort term photovoltaic power combinatorial forecasting model based on similar day clustering[J]. Thermal Power Generation, 2023, 52(11): 123-131.

Ultrashort term photovoltaic power combinatorial forecasting model based on similar day clustering

CHANG Qingsong¹, YANG Zhao², YANG Yihun², LEI Yang², HE Xinlin²

(1. Jiutai Power Plant, Huaneng Changchun Power Generation Co., Ltd., Changchun 130500, China;

2. Xi'an Thermal Power Research Institute Co., Ltd., Xi'an 710054, China)

Abstract: Aiming at the problem of low prediction accuracy of single power prediction model due to the impact of photovoltaic power fluctuation, a combined photovoltaic power prediction model based on similar day clustering is proposed. Firstly, k -means clustering is selected to divide the original power data into three similar day sample sets of sunny, rainy and cloudy according to different weather types, and the variational mode decomposition (VMD) is used to decompose the similar day samples; Secondly, the convolution neural network is used to optimize the support vector machine (CNN-SVM) and bidirectional short-term and short-term memory (BiLSTM) neural network, respectively, to predict and superimpose the decomposed power data and combine the prediction results with weights, and the grid search algorithm (GS) is used to find the optimal combination weight to improve the performance of the combination prediction model. Finally, the validity of the PV power prediction model proposed in this paper is verified by the one-year measured data of a photovoltaic power station in Australia. The experimental results show that the model proposed in this paper can predict the photovoltaic power well and has strong adaptability no matter what weather type.

Key words: PV power prediction; CNN; SVM; LSTM neural network; grid search algorithm

收稿日期: 2023-01-11

基金项目: 中国华能集团有限公司总部科技项目 (HNKJ22-H36)

Supported by: Science and Technology Project of China (HNKJ22-H36)

第一作者简介: 常青松 (1983), 男, 硕士, 高级工程师, 主要研究方向为电力系统运行与规划, 13904318697@163.com.

围绕“十四五”规划和“双碳”目标，新型电力系统结构不断发展，新能源发电量占比在2060年预计达到60%以上。光伏发电是最主要的新能源发电技术之一，已经得到迅猛发展。但是，其间歇性、多变性、随机性的特点又给新能源消纳和电网稳定性带来极大的挑战。光伏功率预测能够在一定程度上缓解新能源消纳的问题，降低弃光率，同时在电网管理和调度中发挥积极作用，进一步提供可靠的电力成本分析，提高经济效益^[1]。

目前在光伏功率预测方面已经有很多学者做了研究工作，大致可分为物理法和统计法2种：物理法是根据光伏发电原理，基于天气预报数据和光伏系统设计参数来建立光伏功率预测模型，但该方法对电站组件参数和气象数据依赖性较高，且建模过程复杂，预测精度较低，商业化推广的难度较大；统计法^[2-3]主要基于以往累计的功率数据和气象数据，通过特定算法建立功率输入、输出的映射关系实现光伏功率预测，统计法由于建模简单、信息易获取等优点受到研究者的青睐。

现在常用光伏功率预测技术主要分为模型输入、模型构建以及参数优化3部分。模型输入是通过聚类、降维及分解等方法对原始功率数据进行预处理。文献[4]通过经验模态分解(EMD)对环境因素序列分解，然后利用主成分分析法(PCA)对特征输入降维。文献[5]通过集合经验模态(EEMD)分解对光伏功率数据进行分解。文献[6]通过经验小波变换(EWT)对原始功率序列进行分解。文献[7]首先使用 k -means算法对历史数据按天气分类，然后使用变分模态分解(VMD)对数据序列进行分解。文献[8]通过模糊 C 均值聚类对初始数据按不同的季节和天气进行划分。模型构建主要采用传统学习算法(如支持向量机、随机森林等)和深度学习算法(如卷积神经网络、循环神经网络)等智能算法的预测技术。参数优化主要涉及模型训练和模型超参数优化，如采用灰狼、蚁群等寻优算法进行参数调优。文献[9]使用了基于改进的麻雀搜索算法(SSA)优化极限学习机(ELM)来预测短期光伏功率。文献[10]使用基于自适应遗传算法优化Elman神经网络来预测光伏功率。文献[11]使用基于粒子群优化极限梯度提升树(PSO-XGBoost)来预测光伏功率。文献[12]采用长短时记忆(LSTM)神经网络来预测光伏功率。文献[13]基于模拟退火粒子群算法优化BP神经网络(SAPSO-BP)来预测光伏功

率。然而，上述研究都是采用单一模型来预测光伏功率，组合预测能够结合各单一模型的优势，进一步提升光伏功率预测精度。已有不少文献验证了组合模型的有效性：文献[14]采用熵权法组合3种单一模型来预测光伏功率；文献[15]采用遗传算法组合单一模型来预测光伏功率；文献[16]采用强化学习的Q学习算法组合单一模型来预测光伏功率。这些组合模型有效提升了光伏功率预测精度。

基于此，本文提出一种基于 k -means聚类、变分模态分解、卷积神经网络优化支持向量机、双向长短时记忆神经网络和网格搜索算法的组合光伏功率预测模型(GS+VMD+CNN-SVM+BiLSTM)。首先，使用 k -means进行相似日聚类，划分为晴天、雨天和多云3种天气类型的数据集；其次，使用VMD分解法将相似日数据集分解为多个子序列；然后，分别使用CNN-SVM和BiLSTM对各子序列预测并叠加得到2个单一模型各自的光伏功率预测结果；最后，使用网格搜索算法对单个模型预测结果进行最优权重组合，得到最终的光伏预测结果。

1 组合预测模型框架

本文所提出的基于相似日聚类和网格搜索算法优化的组合式光伏功率预测模型如图1所示。

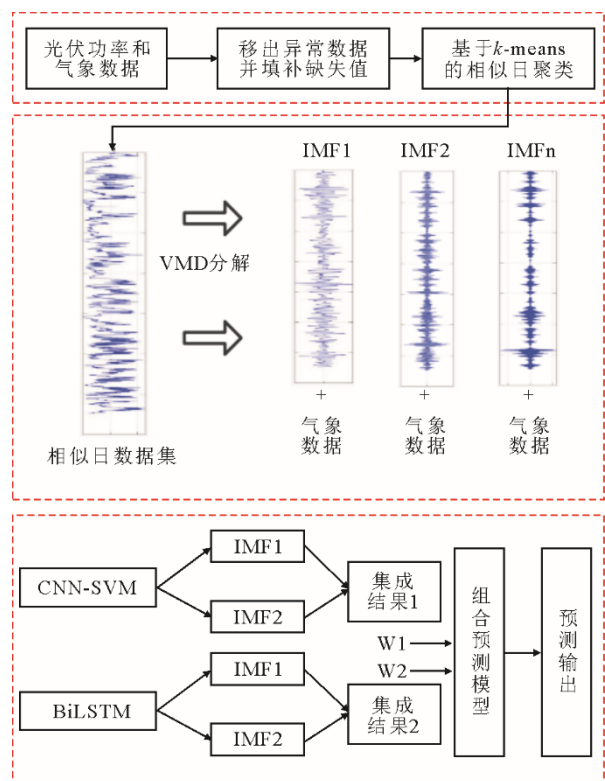


图1 组合式光伏功率预测模型

Fig.1 Combined model of photovoltaic power prediction

2 k -means 聚类 和 VMD 分解

2.1 基于 k -means 的相似日聚类

通过对天气分型, 相似日进行聚类, 对不同天气类型建立相应的样本和预测模型, 能够有效提升预测模型的精度。 k -means 聚类算法原理简单, 适用性强, 且计算大规模数据集时比较高效。其将数据集划分成事先设定好的 K 种类别, 每类中的数据间保持有较大的相似度, 具有类内紧凑、类间疏远的特点^[17-18]。 k -means 算法步骤为:

1) 设定输入数据集 $X=\{x_{1i}, x_{2i}, \dots, x_{mi}\}$, $i=1, 2, \dots, n$, n 为样本数据量, m 为每个样本数据的特征维度, 输出簇划分为 $C=\{c_1, c_2, \dots, c_K\}$, K 为设定所需聚类数量;

2) 从输入样本集 X 中随机抽取 K 个数据向量作为初始聚类中心 $C_j=\{c_{1j}, c_{2j}, \dots, c_{mj}\}$;

3) 分别计算每个数据点到 K 个聚类中心的欧氏距离, 并将该点归入与聚类中心距离最小的那一类中, 形成 K 个簇;

4) 对新的簇重新计算聚类中心;

5) 重复步骤 3)、步骤 4), 若均值不再变化或达到预设迭代次数, 则终止算法。

本文进行聚类时所选取的特征向量有每日风速的最大值和最小值, 每日风向平均值的正弦值和余弦值, 每日温度的最大值、最小值及平均值, 每日湿度的最大值、最小值及平均值, 每日辐射度的最大值、最小值和平均值。

2.2 VMD 分解

变分模态分解 (VMD) 能够依据信号的频域特性把原始数据序列分解成多个准正交模态函数序列, 其本质是变分问题的求解过程。以变分问题为架构, 利用交替方向乘子法持续更新各模态及模态中心频率, 并分离各模态及其中心频率, 最终使各模态的估计带宽之和最小。VMD 具有自适应好的特点, 且能够很好地避免模态混叠现象^[19]。

约束变分模型为:

$$\begin{cases} \min \left\{ \sum_{k=1}^K \left\| \hat{\partial}_t \left[\left\langle \delta_t + \frac{j}{\pi_i} \right\rangle * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^K u_k = f(t) \end{cases} \quad (1)$$

式中: u_k 为分解后的第 k 个模态分量; w_k 为对应模态分量的中心频率; δ_t 为 Dirac 分布函数; $*$ 为卷积运算符; K 为预设的模态数目。

引入二次惩罚因子 α 和 Lagrange 乘子 λ , 使变分问题不受约束, Lagrange 函数为:

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \hat{\partial}_t \left[\left\langle \delta_t + \frac{j}{\pi_i} \right\rangle u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \\ & \|f(t) - [u_k(t)]\|_2^2 + \left[\lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right] \end{aligned} \quad (2)$$

利用交替方向乘子法对式(2)求解, 得到数据分解后的 K 个模态分量。

模态分量 u_k 及中心频率 w_k 的更新表达式为:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \hat{u}(w) / 2}{1 + 2\alpha(w - w_k)^2} \quad (3)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \quad (4)$$

式中: $\hat{f}(w)$ 、 $\hat{u}_k(w)$ 、 $\hat{u}(w)$ 为分别 $f(t)$ 、 $u_i(t)$ 、 $u(t)$ 的傅里叶变换; w 为频率; n 为迭代次数。

通过式(3)、式(4)循环迭代, 更新 u_k 及 w_k , 代入式(5)更新 λ :

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left[\hat{f}(w) - \sum_{k=1}^K \hat{u}_k^{n+1}(w) \right] \quad (5)$$

式中: τ 为 Lagrange 乘子更新系数。

对于已知判别精度 $\xi > 0$, 有停止迭代条件:

$$\frac{\sum_{k=1}^K \left\| \hat{u}_k^{n+1}(w) - \hat{u}_k^n(w) \right\|_2^2}{\left\| \hat{u}_k^n(w) \right\|_2^2} < \xi \quad (6)$$

倘若 ξ 满足迭代停止要求, 则迭代停止, 输出 K 个模态序列; 若 ξ 不满足迭代停止要求, 将迭代结果重新返回式(3)、式(4), 进行新一轮迭代。 K 值选取直接影响分解结果, 一般通过频率中心法来确定。

3 预测模型

3.1 CNN-SVM 模型

时间序列数据可视为 1 组一维网络数据, 卷积神经网络具有很强地处理网络拓补特征数据的能力, 能够很好地提取数据序列的空间特征。将提取的数据特征输入支持向量机中, 采取 CNN-SVM 模型实现功率预测。

3.1.1 卷积神经网络

CNN 模型通过局部连接和权重共享的形式, 将原始数据进行高维映射处理, 有效提取数据特征。CNN 一般由卷积层、池化层和全连接层 3 部分构

成。卷积层主要用来提取输入数据的特征，用奇数卷积核对输入进行多深度卷积运算，再利用 Relu 等激活函数进行非线性映射；池化层主要负责数据特征降维，通过对卷积层获得的数据特征进行汇总，采取最大值或平均值处理，实现特征降维；全连接层嵌入网络的底层，对池化层处理后的数据特征实现整合，计算回归的结果^[20-22]。

由于影响光伏功率的温度、辐射度等因素为一维数据，本文选择一维卷积神经网络构建光伏功率预测输入输出模型，一维卷积神经网络预测结构如图 2 所示。

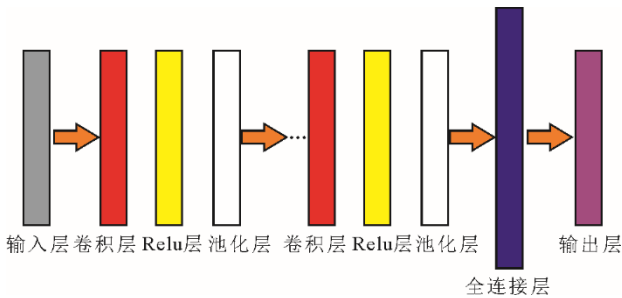


图 2 一维卷积神经网络预测模型结构
Fig.2 Prediction model structure of one-dimensional convolution neural network

3.1.2 支持向量机

SVM 是基于统计学理论的一种机器学习方法，根据结构风险最小化原则提出，具有泛化能力强及计算速度快等特点^[23-24]。

对 SVM 引入核函数实现非线性变化，目标函数为：

$$\min \phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} \langle w \cdot w \rangle + c \sum_{i=1}^n \xi_i^2 \quad (7)$$

$$\text{s.t. } y_i [\langle w \cdot x_i \rangle + b] - 1 + \xi_i \geq 0 \quad (8)$$

式中： w 为权重向量； c 为惩罚因子； ξ 为误差； x_i 为样本向量； b 为偏置量。

引入核函数 $K(x_i, x_j)$ 后，函数转化为：

$$\min Q(\alpha) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i \quad (9)$$

$$\text{s.t. } \alpha_i \geq 0, i = 1, 2, \dots, n$$

$$\sum_{i=1}^n y_i \alpha_i = 0$$

分类函数为：

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right\} \quad (10)$$

径向基 (RBF) 函数为局部核函数，可以直观体现数据间的距离，并且针对数据噪声具有显著的

抗干扰能力，分类性能优于其他核函数，选用 RBF 函数作为 SVM 核函数。

核函数表达式为：

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (11)$$

式中： γ 为 RBF 的核参数，决定其镜像作用范围。

3.1.3 基于 CNN-SVM 的预测模型

本文基于 CNN-SVM 的光伏功率预测模型通过卷积神经网络中卷积和池化操作对原始数据进行特征学习和提取，随后将提取出的特征即全连接层的结果作为 SVM 输入，从而实现光伏功率的预测。

构建 CNN-SVM 模型的具体步骤如下：

- 1) 将光伏功率数据划分为训练集、验证集及测试集，并进行归一化处理；
- 2) 建立 CNN-SVM 预测模型，对模型参数进行初始化；
- 3) 将训练集样本输入 CNN 模型中训练。训练时采用梯度下降算法进行传播并更新模型参数，利用验证集数据实时验证模型的误差和损失值，在预设的迭代次数内，若误差和损失值满足要求，则保存模型的最优超参数和权重；
- 4) 将最优超参数迁移至 CNN-SVM 模型中的 CNN 部分，利用 Flatten 层的输出对 SVM 进行训练；
- 5) 将训练好的 CNN-SVM 模型用测试集进行验证，输出预测结果。

基于 CNN-SVM 模型光伏功率预测流程如图 3 所示。

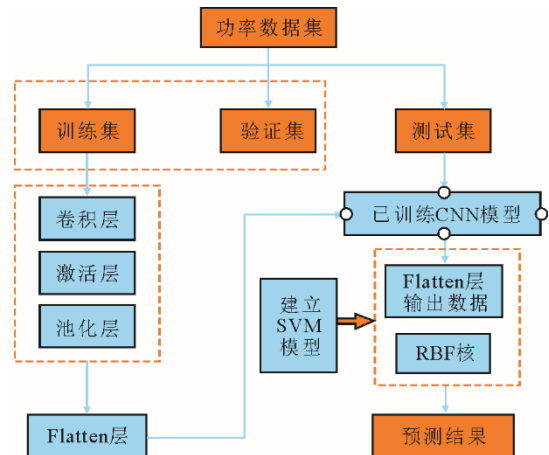


图 3 基于 CNN-SVM 模型光伏功率预测流程
Fig.3 Photovoltaic power prediction flow based on CNN-SVM model

3.2 双向 LSTM 模型

光伏功率数据序列具有很强的时间相关性。CNN 模型注重的是提取输入数据的局部空间特征，

而 LSTM 模型因其特殊的网络结构可以提取时间序列数据之间的时间相关性, 非常适合解决时间序列预测问题。LSTM 模型相比于传统循环神经网络, 加入了遗忘门、输入门以及输出门, 模型会更加有选择性地筛选有效信息, 剔除无效信息, 很好地避免了 RNN 梯度爆炸及容易消失的问题。LSTM 模型长期依赖信息的学习, 能够利用前时刻的状态信息推导下一时刻的状态信息, 形成“记忆”的功能, 从而显著提升模型的预测性能。LSTM 模型单元结构如图 4 所示。

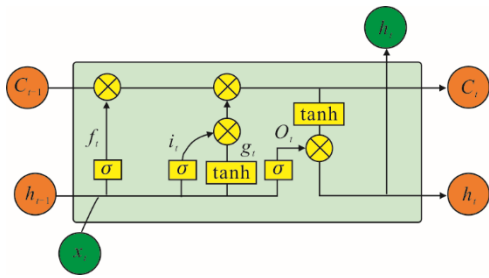


图 4 LSTM 模型单元结构
Fig.4 LSTM model unit structure

LSTM 模型单元结构计算过程为:

$$\begin{cases} i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ g_t = \tanh(W_{gx} x_t + U_{gh} h_{t-1} + b_g) \\ C_t = g_t i_t + C_{t-1} f_t \\ h_t = \tanh(C_t) O_t \end{cases} \quad (12)$$

式中: σ 为 sigmoid 函数; W 为权值; b 为偏置; C_{t-1} 为前一时刻单元状态; h_{t-1} 为前一时刻中间状态; x_t 为当前时刻输入; f_t 为遗忘门输出; i_t 和 O_t 分别为输入门和输出门的输出; g_t 为输入节点的输出; C_t 和 h_t 分别为输出变量的单元状态和中间状态。

双向 LSTM(BiLSTM)模型是在 LSTM 模型的基础上, 将单相的 LSTM 神经网络层变为双向的 Backward 层和 Forward 层。Forward 层从初始时刻到 t 时刻正向计算一遍, 得到并保存每个时刻向前隐含层的输出。Backward 层则沿着时刻 t 到初始时刻反向计算, 得到并保存每个时刻向后隐含层的输出。最后结合 Forward 层和 Backward 层的相应时刻输出的结果得到最终输出^[25-27]。图 5 为 BiLSTM 模型单元结构。

3.3 组合预测模型

由于机理不同, 单一模型均有自身的局限性。CNN 模型能够很好地提取功率数据集的空间特征, LSTM 模型能够很好地提取功率数据集的时间特

征, 采用组合模型将 2 种不同的单一模型结合起来, 综合利用 2 种单一模型所提供的信息, 可以提升预测精度。

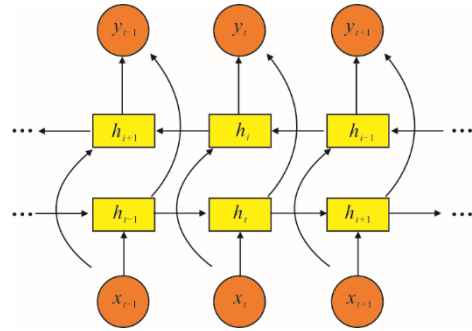


图 5 BiLSTM 模型单元结构
Fig.5 BiLSTM model unit structure

3.3.1 基于熵权法的定权重模型

基于熵权法 (entropy method, EM) 的光伏功率组合预测模型, 通过信息熵理论得到各单一模型的组合权值。将单一模型预测值与真实值的误差视作评价指标, 每个预测点视作评价对象。

1) 计算单一模型预测值与真实值的误差绝对值, 并以每个模型预测点的误差绝对值形成评价矩阵, 记作 $A=[a_{ij}]_{m \times n}$, 其中 $m=120$, $n=2$ 。

2) 标准化处理评价矩阵, 得到标准化矩阵 $R=[r_{ij}]_{m \times n}$:

$$r_{ij} = \frac{a_{ij} - \min(a_{ij})}{\max(a_{ij}) - \min(a_{ij})} \quad (13)$$

3) 计算第 j 个模型中第 i 个预测点的误差绝对值占比:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^m r_{ij}} \quad (14)$$

式中: $i=1, 2, \dots, m$; $j=1, 2, \dots, n$ 。

4) 计算 2 个模型相应的熵值大小:

$$H_j = -\frac{1}{\ln m} \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (15)$$

5) 计算 2 个模型的组合权值:

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^n H_j} \quad (16)$$

6) 计算组合模型预测值, P_1 、 P_2 分别为 2 个单一模型预测点对应的预测值, P_3 为组合预测值:

$$P_3 = w_1 P_1 + w_2 P_2 \quad (17)$$

3.3.2 基于网格搜索法的变权重模型

采用网格搜索法获取变权重系数的目的是在

任意时刻使得组合模型的误差绝对值最小, 设其适应度函数为第 k 时刻组合模型的均方根误差 (root mean square error, δ_{RMSE}), 即:

$$\delta_{\text{RMSE}}(k) = \sqrt{\frac{1}{n} \sum_{n} [w_{\text{cs}}(k)y_{\text{cs}}(k) + w_{\text{bl}}(k)y_{\text{bl}}(k) - y(k)]^2} \quad (18)$$

式中: $w_{\text{cs}}(k)$ 、 $y_{\text{cs}}(k)$ 分别为第 k 时刻 CNN-SVM 模型的权重与预测值; $w_{\text{bl}}(k)$ 、 $y_{\text{bl}}(k)$ 分别为第 k 时刻 BiLSTM 模型的权重与预测值; $w_{\text{cs}}(k) + w_{\text{bl}}(k) = 1$; $y(k)$ 为第 k 时刻的真实值。

变权重组合模型在 k 时刻的表达式为:

$$P_k = w_{\text{cs}}(k)y_{\text{cs}}(k) + w_{\text{bl}}(k)y_{\text{bl}}(k) \quad (19)$$

4 算例分析

4.1 实验数据说明

本文采用澳大利亚 (DKASC) 某光伏电站在 2016.04.01—2017.03.31 的实测光伏功率和气象数据 (全球水平辐射、漫射水平辐射、温度、相对湿度、风速、风向) 进行仿真实验。本文数据采样间隔为 5 min, 每天取 08:00—17:55 时段的功率值, 每天有 120 个数据点, 全年共有 43 800 个数据点。根据当地气候状况, 将原始功率集按季节划分为春 (9 月—11 月)、夏 (12 月—2 月)、秋 (3 月—5 月)、冬 (6 月—8 月) 4 个数据集, 每个数据集进行相似日聚类, 聚类为 3 种代表性的相似日类型 (晴天、雨天、多云) 数据集, 聚类后的数据集为训练样本, 选取各相似日数据集下的最后一天为测试样本, 验证模型在不同季节、不同天气下的预测性能。

4.2 误差指标

误差指标是衡量预测精度的统一标准, 单一误差指标不能全面分析模型预测效果。本文采用 3 种误差指标对功率预测模型的结果进行评价, 依次为 δ_{RMSE} 、平均绝对误差 (mean absolute error, δ_{MAE}) 和平均绝对百分比误差 (mean absolute square error, δ_{MAPE}):

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

$$\delta_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (21)$$

$$\delta_{\text{MAPE}} = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times \frac{100}{n} \quad (22)$$

式中: y_i 为第 i 个数据点的功率真实值; \hat{y}_i 为第 i 个数据点的功率预测值; n 为光伏功率点的总数。

4.3 VMD 有效性验证

本文采用 VMD 对光伏功率原始数据集进行分

解, 为了验证 VMD 分解的有效性, 选取 CNN-SVM、BiLSTM、VMD+CNN-SVM 和 VMD+BiLSTM 4 个预测模型进行对比, 分别对晴天、雨天和多云 3 种天气类型的相似日样本进行预测。经 k -means 聚类后 3 种天气历史功率数据集 VMD 分解结果 (以雨天为例) 如图 6 所示, 预测结果误差见表 1。

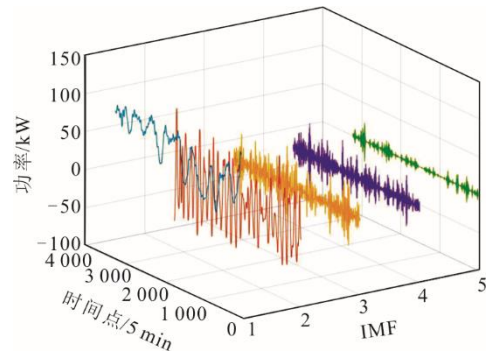


图 6 雨天光伏功率原始数据 VMD 分解结果
Fig.6 VMD decomposition results of raw photovoltaic power data on rainy days

表 1 光伏功率 VMD 分解前后误差对比
Tab.1 Error comparison before and after VMD decomposition of photovoltaic power

天气	预测模型	$\delta_{\text{RMSE}}/\text{kW}$	$\delta_{\text{MAE}}/\text{kW}$	δ_{MAPE}
晴天	CNN-SVM	3.667	2.776	2.599
	BiLSTM	3.701	2.408	2.332
	VMD+CNN-SVM	2.923	1.957	2.105
	VMD+BiLSTM	3.022	1.951	2.279
雨天	CNN-SVM	21.369	13.662	13.420
	BiLSTM	22.266	14.069	13.510
	VMD+CNN-SVM	6.871	5.127	5.971
	VMD+BiLSTM	6.706	4.990	5.828
多云	CNN-SVM	24.328	15.728	34.697
	BiLSTM	23.971	15.591	34.605
	VMD+CNN-SVM	6.818	5.440	17.996
	VMD+BiLSTM	6.873	5.510	25.259

从表 1 可以得出, 经 VMD 分解后的模型在 3 种不同天气类型下预测精度均比未采用 VMD 分解模型的预测精度高。晴天天气下, 历史功率曲线趋势明显, 各模型均能取得较好的预测结果, 预测精度提升较小。对于雨天和多云天气而言, 功率曲线趋势多变, 规律性差, 经 VMD 分解后的预测模型的预测精度显著提升。以雨天为例, 对于 δ_{RMSE} 、 δ_{MAE} 和 δ_{MAPE} , VMD+CNN-SVM 模型比 CNN-SVM 模型分别降低了 67.84%、62.47% 和 55.51%; VMD+BiLSTM 模型比 BiLSTM 模型分别降低了 69.88%、64.53% 和 56.86%。实验结果证明, 光伏功率数据集经 VMD 分解后再进行预测可以显著提高预测精度。

4.4 组合预测模型验证

为验证基于网格搜索法变权重模型(GS+VMD+CNN-SVM+BiLSTM)对 CNN-SVM 和 BiLSTM 2 个单一模型的预测结果加权求和性能,引入熵权法定权重模型(EM+VMD+CNN-SVM+BiLSTM)进行误差对比,功率样本集不变,预测结果误差见表 2。3 种测试日光伏功率预测曲线和误差对比如图 7—图 12 所示。

表 2 组合预测模型误差对比

Tab.2 Error comparison of combined forecasting model

天气	预测模型	δ_{RMSE}/kW	δ_{MAE}/kW	δ_{MAPE}
晴天	EM+VMD+CNN-SVM+BiLSTM	2.680	1.826	2.048 3
	GS+VMD+CNN-SVM+BiLSTM	2.606	1.876	1.885 0
雨天	EM+VMD+CNN-SVM+BiLSTM	5.805	4.328	5.050 0
	GS+VMD+CNN-SVM+BiLSTM	5.477	3.957	4.767 0
多云	EM+VMD+CNN-SVM+BiLSTM	6.068	4.809	18.221 0
	GS+VMD+CNN-SVM+BiLSTM	5.796	4.445	16.937 0

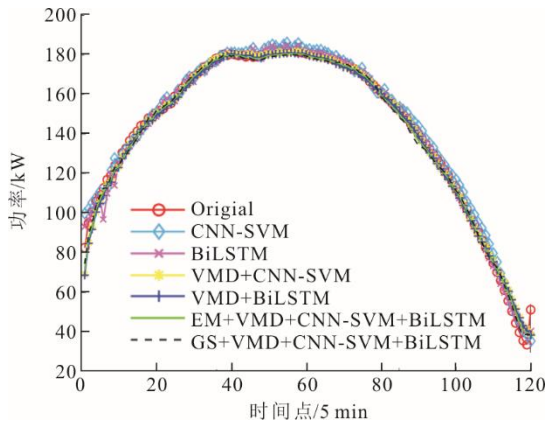


图 7 晴天光伏功率预测曲线

Fig.7 Photovoltaic power forecast curve on sunny days

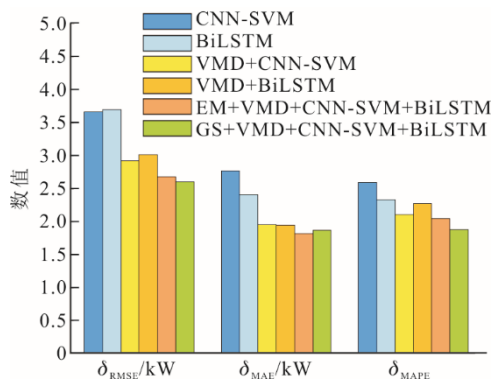


图 8 晴天光伏功率预测误差对比

Fig.8 Comparison of photovoltaic power forecast error on sunny days

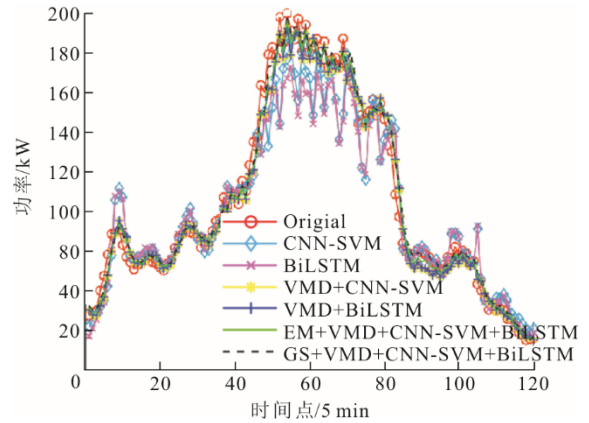


图 9 雨天光伏功率预测曲线

Fig.9 Photovoltaic power forecast curve on rainy days

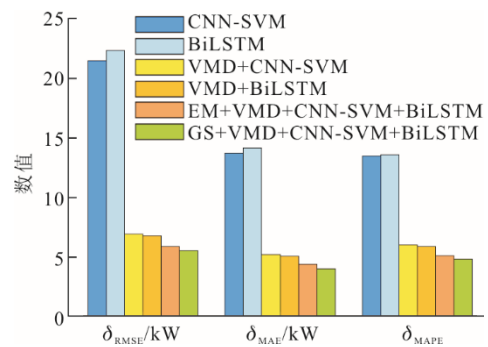


图 10 雨天光伏功率预测误差对比

Fig.10 Comparison of photovoltaic power forecast error on rainy days

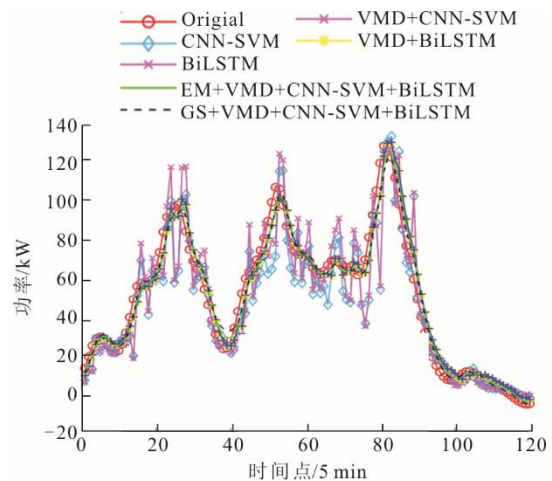


图 11 多云光伏功率预测曲线

Fig.11 Photovoltaic power forecast curve on cloudy days

由表 1—表 2 和图 7—图 12 可以得到如下结论:

- 1) 单个模型实现光伏功率预测时精度偏低。以多云天气为例,对于 δ_{RMSE} 、 δ_{MAE} 和 δ_{MAPE} , VMD+CNN-SVM+BiLSTM 模型比 VMD+CNN-SVM 模型分别降低了 14.99%、18.29%和 5.88%;比 VMD+BiLSTM 模型分别降低了 15.67%、19.33%和 32.95%。

2) 组合预测模型精度优于单一预测模型, 时变权重模型优于定权重模型。以雨天、多云天气的测试日为例, 相比于熵权法定权重组合 (EM+VMD+CNN-SVM+BiLSTM) 预测模型, 基于网格搜索算法的时变权重预测模型 (GS+VMD+CNN-SVM+BiLSTM) 的 δ_{RMSE} 在 2 个测试日分别降低了 5.65% 和 4.48%; δ_{MAE} 分别降低了 8.57% 和 7.57%; δ_{MAPE} 分别降低了 5.60% 和 7.05%。

3) GS+VMD+CNN-SVM+BiLSTM 组合模型在 3 种不同天气情况下, 模型误差均是最低的, 表明该组合模型预测性能优越。

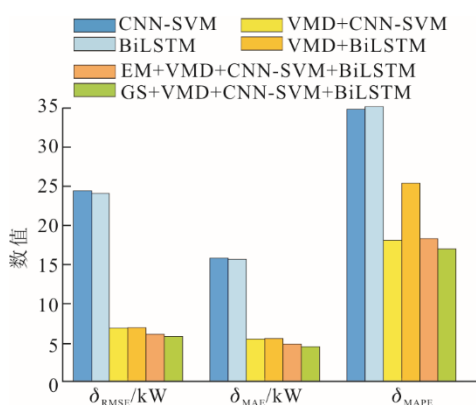


图 12 多云光伏功率预测误差对比

Fig.12 Comparison of photovoltaic power forecast error on cloudy days

5 结 语

本文提出了一种基于相似日聚类 and 网格搜索算法优化的超短期光伏功率组合预测模型。所提的组合预测模型中, 其思想是“聚类-分解-集成-加权-预测”。考虑到光伏功率数据季节性的特点, 首先将原始功率数据划分 4 个季节, 并利用 k -means 聚类算法将每个季节功率数据分为 3 类相似日天气集; 然后采用 VMD 方法将各相似日数据集分解为多个子序列; 接着分别采用 CNN-SVM、BiLSTM 2 个预测模型对子序列功率进行预测并叠加获得实际光伏功率, 实验验证了 VMD 分解的可行性; 最后采用网格搜索法对 2 个单一模型的预测结果进行加权组合得到最优的组合预测模型, 并与熵权法加权组合模型进行对比。实验结果表明 GS+VMD+CNN-SVM+BiLSTM 模型能够很好地预测光伏功率, 相较于其他混合模型, 本文所提模型预测性能最优, 对实际工程中光伏功率预测模型的选择提供了参考价值, 在其他领域也具有一定应用前景和意义。

[参 考 文 献]

- [1] 朱琼峰, 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.
- [2] BHAVSAR S, PITCHUMANI R, ORTEGA-VAZQUEZ M A. Machine learning enabled reduced-order scenario generation for stochastic analysis of solar power forecasts[J]. Applied Energy, 2021, 293: 116964.
- [3] 杨锡运, 杨岩, 孟令卓越, 等. 计及 NWP 信息缺失的数据共享与 GRA 权重优化的分布式光伏电站功率预测[J/OL]. 电测与仪表: 1-10. (2022-07-05)[2023-05-16]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20220704.1020.004.html>.
YANG Xiyun, YANG Yan, MENG LingzhuoChao, et al. Data sharing and GRA weight optimization for power prediction of distributed photovoltaic power plant considering missing NWP information[J/OL]. Electrical Measurement and Instrumentation: 1-10. (2022-07-05) [2023-05-16]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20220704.1020.004.html>.
- [4] 张云钦, 程起泽, 蒋文杰, 等. 基于 EMD-PCA-LSTM 的光伏功率预测模型[J]. 太阳能学报, 2021, 42(9): 62-69.
ZHANG Yunqin, CHENG Qize, JIANG Wenjie, et al. Photovoltaic power prediction model based on EMD-PCA-LSTM[J]. Acta Energetica Solaris Sinica, 2021, 42(9): 62-69.
- [5] 王振浩, 王翀, 成龙, 等. 基于集合经验模态分解和深度学习的光伏功率组合预测[J]. 高电压技术, 2022, 48(10): 4133-4142.
WANG Zhenhao, WANG Chong, CHENG Long, et al. Photovoltaic power combination prediction based on ensemble empirical mode decomposition and deep learning[J]. High Voltage Engineering, 2022, 48(10): 4133-4142.
- [6] 陈涛, 王艳, 纪志成. 基于经验小波变换的光伏功率组合预测模型[J]. 系统仿真学报, 2021, 33(11): 2627-2635.
CHEN Tao, WANG Yan, JI Zhicheng. Combination forecasting model of photovoltaic power based on empirical wavelet transform[J]. Journal of System Simulation, 2021, 33(11): 2627-2635.
- [7] 张娜, 任强, 刘广忱, 等. 基于 VMD-GWO-ELMAN 的光伏功率短期预测方法[J]. 中国电力, 2022, 55(5): 57-65.
ZHANG Na, REN Qiang, LIU Guangchen, et al. Short-term forecasting method of photovoltaic power based on VMD-GWO-ELMAN[J]. Electric Power, 2022, 55(5): 57-65.
- [8] 杨国华, 张鸿皓, 郑豪丰, 等. 基于相似日聚类和 IHGWO-WNN-AdaBoost 模型的短期光伏功率预测[J]. 高电压技术, 2021, 47(4): 1185-1194.
YANG Guohua, ZHANG Honghao, ZHENG Haofeng, et al. Short-term photovoltaic power forecasting based on similar weather clustering and IHGWO-WNN-AdaBoost model[J]. High Voltage Engineering, 2021, 47(4): 1185-1194.
- [9] 陈骏豪, 张娜, 刘广忱, 等. 基于 EOSSA-ELM 的光伏短期输出功率预测[J]. 可再生能源, 2022, 40(7): 890-898.
CHEN Junhao, ZHANG Na, LIU Guangchen, et al. Short-term photovoltaic output power prediction based on EOSSA-ELM[J]. Renewable Energy Resources, 2022, 40(7): 890-898.

- [10] 孙子涵, 姜志海, 刘延龙, 等. 基于小波变换和优化的 Elman 神经网络的光伏功率预测[J]. 电网与清洁能源, 2022, 38(6): 98-103.
SUN Zihan, JIANG Zhihai, LIU Yanlong, et al. Photovoltaic power prediction based on wavelet transform and optimized Elman neural network[J]. Power System and Clean Energy, 2022, 38(6): 98-103.
- [11] 吴春华, 董阿龙, 李智华, 等. 基于图相似日和 PSO-XGBoost 的光伏功率预测[J]. 高电压技术, 2022, 48(8): 3250-3259.
WU Chunhua, DONG Along, LI Zhihua, et al. Photovoltaic power prediction based on graph similarity day and PSO-XGBoost[J]. High Voltage Engineering, 2022, 48(8): 3250-3259.
- [12] 李丰君, 王磊, 赵健, 等. 基于天气融合和 LSTM 网络的分布式光伏短期功率预测方法[J]. 中国电力, 2022, 55(11): 149-154.
LI Fengjun, WANG Lei, ZHAO Jian, et al. Research on distributed photovoltaic short-term power prediction method based on weather fusion and LSTM-net[J]. Electric Power, 2022, 55(11): 149-154.
- [13] 贾德香, 吕干云, 林芬, 等. 基于 SAPSO-BP 和分位数回归的光伏功率区间预测[J]. 电力系统保护与控制, 2021, 49(10): 20-26.
JIA Dexiang, LYU Ganyun, LIN Fen, et al. Photovoltaic power interval prediction based on SAPSO-BP and quantile regression[J]. Power System Protection and Control, 2021, 49(10): 20-26.
- [14] 杨锡运, 刘欢, 张彬, 等. 基于熵权法的光伏输出功率组合预测模型[J]. 太阳能学报, 2014, 35(5): 744-749.
YANG Xiyun, LIU Huan, ZHANG Bin, et al. Photovoltaic output power combination prediction model based on entropy weight method[J]. Acta Energetica Solaris Sinica, 2014, 35(5): 744-749.
- [15] 谭小钰, 刘芳, 马俊杰, 等. 基于 DBN 与 T-S 时变权重组合的光伏功率超短期预测模型[J]. 太阳能学报, 2021, 42(10): 42-48.
TAN Xiaoyu, LIU Fang, MA Junjie, et al. Ultra-short-term forecasting model of photovoltaic power based on DBN and T-S time-varying weight combination[J]. Acta Energetica Solaris Sinica, 2021, 42(10): 42-48.
- [16] 孟安波, 许炫淙, 陈嘉铭, 等. 基于强化学习和组合式深度学习模型的超短期光伏功率预测[J]. 电网技术, 2021, 45(12): 4721-4728.
MENG Anbo, XU Xuancong, CHEN Jiaming, et al. Ultra short-term photovoltaic power prediction based on reinforcement learning and combined deep learning model[J]. Power System Technology, 2021, 45(12): 4721-4728.
- [17] 陈瑶, 陈晓宁. 基于自适应 Kmeans 和 LSTM 的短期光伏发电预测[J]. 电测与仪表, 2023, 60(7): 94-99.
CHEN Yao, CHEN Xiaoning. Prediction of short-term photovoltaic power generation based on adaptive Kmeans and LSTM[J]. Electrical Measurement and Instrumentation, 2023, 60(7): 94-99.
- [18] 张雨金, 杨凌帆, 葛双冶, 等. 基于 Kmeans-SVM 的短期光伏发电功率预测[J]. 电力系统保护与控制, 2018, 46(21): 118-124.
ZHANG Yujin, YANG Lingfan, GE Shuangye, et al. Short-term photovoltaic power forecasting based on kmeans algorithm and support vector machine[J]. Power System Protection and Control, 2018, 46(21): 118-124.
- [19] 王俊, 李霞, 周昔东, 等. 基于 VMD 和 LSTM 的超短期风速预测[J]. 电力系统保护与控制, 2020, 48(11): 45-52.
WANG Jun, LI Xia, ZHOU Xidong, et al. Ultra short-term wind speed prediction based on VMD-LSTM[J]. Power System Protection and Control, 2020, 48(11): 45-52.
- [20] ZHANG G, LIU H C, LI P L, et al. Load prediction based on hybrid model of VMD-m RMR-BPNN-LSSVM[J]. Complexity, 2020, 4: 1-20.
- [21] 杨晶显, 张帅, 刘继春, 等. 基于 VMD 和双重注意力机制 LSTM 的短期光伏功率预测[J]. 电力系统自动化, 2021, 45(3): 174-182.
YANG Jingxian, ZHANG Shuai, LIU Jichun, et al. Short-term photovoltaic power prediction based on VMD and dual attention mechanism LSTM[J]. Automation of Electric Power Systems, 2021, 45(3): 174-182.
- [22] 刘亚辉, 赵倩. 基于聚类经验模态分解的 CNN-LSTM 超短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4444-4451.
LIU Yahui, ZHAO Qian. Ultra-short-term power load forecasting based on cluster empirical mode decomposition of CNN-LSTM[J]. Power System Technology, 2021, 45(11): 4444-4451.
- [23] 姚程文, 杨苹, 刘泽健. 基于 CNN-GRU 混合神经网络的负荷预测方法[J]. 电网技术, 2020, 44(9): 3416-3424.
YAO Chengwen, YANG Ping, LIU Zejian. Load forecasting method based on CNN-GRU hybrid neural network[J]. Power System Technology, 2020, 44(9): 3416-3424.
- [24] 魏鹏飞, 樊小朝, 史瑞静, 等. 基于改进麻雀搜索算法优化支持向量机的短期光伏发电功率预测[J]. 热力发电, 2021, 50(12): 74-79.
WEI Pengfei, FAN Xiaochao, SHI Ruijing, et al. Short-term photovoltaic power generation forecast based on improved sparrow search algorithm optimized support vector machine[J]. Thermal Power Generation, 2021, 50(12): 74-79.
- [25] 马骏, 江锐, 丁倩, 等. 基于多元宇宙优化支持向量机的短期光伏发电功率预测[J]. 热力发电, 2020, 49(4): 87-92.
MA Jun, JIANG Rui, DING Qian, et al. Prediction of short-term photovoltaic generation power based on HIMVO-SVM[J]. Thermal Power Generation, 2020, 49(4): 87-92.
- [26] 林顺富, 詹银枫, 李毅, 等. 基于 CNN-BiLSTM 与 DTW 的非侵入式住宅负荷监测方法[J]. 电网技术, 2022, 46(5): 1973-1981.
LIN Shunfu, ZHAN Yinfeng, LI Yi, et al. Non-intrusive residential load monitoring method based on CNN-BiLSTM and DTW[J]. Power System Technology, 2022, 46(5): 1973-1981.
- [27] 朱凌建, 荀子涵, 王裕鑫, 等. 基于 CNN-BiLSTM 的短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4532-4539.
ZHU Lingjian, XUN Zihan, WANG Yuxin, et al. Short-term power load forecasting based on CNN-BiLSTM[J]. Power System Technology, 2021, 45(11): 4532-4539.

(责任编辑 杜亚勤)