

文章编号: 2095-2163(2024)01-0112-07

中图分类号: TP183;TM714

文献标志码: A

基于改进 CEEMDAN-TCN 模型的风电功率预测研究

李望月, 樊重俊

(上海理工大学 管理学院, 上海 200093)

摘要: 针对风力发电时间序列数据随机性大、单一算法难以获得准确预测结果的问题, 本文采用改进的完全自适应噪声集合经验模态分解 (ICEEMDAN) 与时域卷积网络 (TCN) 相结合的模型预测风电功率。首先, 针对集成模态分解的残余噪声和杂散模式问题, 采用改进的 CEEMDAN 将原始序列数据分解为多个复杂度差异明显的子序列; 其次, 利用样本熵评估各分量复杂度, 采用注意力机制的 TCN, 分别对低复杂度子序列和高复杂度子序列进行预测; 最后将各子序列的预测结果叠加, 得到最终的预测结果。经在弗兰德伦地区的数据集上测试结果表明, 所提的 ICEEMDAN-TCN 模型的 MAPE 为 1.74%, RMSE 为 127.36, 优于其它对比模型, 预测效果表现最优。

关键词: 时域卷积神经网络; 自适应噪声集合经验模态分解; 风电功率; 预测

Research on wind power prediction based on improved CEEMDAN-TCN model

LI Wangyue, FAN Chongjun

(School of Business, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Aiming at the problem that wind power time series data is highly randomized and it is difficult for a single algorithm to obtain accurate prediction results, this paper uses an improved model combining fully adaptive noise ensemble empirical mode decomposition (ICEEMDAN) and time-domain convolutional network (TCN) to predict wind power power. Firstly, aiming at the residual noise and spurious mode problems of integrated modal decomposition, the improved CEEMDAN is used to decompose the original sequence data into multiple subsequences with obvious complexity differences. Secondly, the sample entropy is used to evaluate the complexity of each component, and the TCN of the attention mechanism is used to predict the low complexity subsequence and the high complexity subsequence respectively. Finally, the prediction results of each subseries are superimposed to obtain the final prediction results. The results of the dataset in Flandren show that the proposed ICEEMDAN-TCN model has a MAPE of 1.74% and an RMSE of 127.36, which is better than other comparison models and has the best prediction effect.

Key words: time-domain convolutional neural networks; adaptive noise ensemble empirical mode decomposition; wind power; forecast

0 引言

近年来, 燃煤发电引发了资源消耗和全球各地环境问题, 已经促使能源行业和相关政策发生了重大变化。风能作为化石燃料替代品, 是一种环保和清洁的能源。根据《全球风能报告 2023》数据显示, 至 2022 年底, 全球新增的风电装机容量已达到 77.6 GW, 其中陆上风电装机容量占 68.8 GW, 海上风电则占 8.8 GW。然而, 风力发电由于其高度不确定的特性变得难以预测, 与之伴随的是运营和规划风险。因此, 探索准确而稳定的风力发电预测技术, 对于降

低运营成本、提高发电系统的可靠性具有重要意义^[1]。

预测是将风能和太阳能等可变、可再生能源资源整合到电力系统^[2]中的一种关键和经济有效的方法。目前的风力发电预测大致可分为物理技术^[3]、统计技术^[4]和人工智能技术等 3 类。物理技术主要利用测量得到的物理信息进行预测, 但具有较高的信息收集成本, 较适用于长期的风力发电预测。统计技术可以捕获风力发电时间序列数据的历史信息, 由于统计模型是建立在线性数据的前提下, 会影响模型的准确度。随着人工智能技术的发展,

基金项目: 2020 教育部哲学社会科学重大课题攻关项目 (20JZD010)。

作者简介: 李望月 (1998-), 女, 硕士研究生, 主要研究方向: 人工智能理论及应用。

通讯作者: 樊重俊 (1963-), 男, 博士, 教授, 主要研究方向: 人工智能、电子商务。Email: fanchongjun2022@163.com

收稿日期: 2023-10-23

机器学习及深度学习模型可以进一步描述复杂和非线性的风力发电数据信息。常见的人工智能技术如卷积神经网络 (CNNs) 和支持向量机 (SVM) 等^[5]。时间卷积网络^[6]使用多种技术手段 (如: 膨胀卷积、时序卷积), 提高了时序预测性能, 其在时序预测方面相比于其他网络预测的精度较高, 在传统网络中加入注意力机制^[7]后, 预测精度得到进一步提升^[8]。但单一的预测模型不能进行有效的风力发电预测, 由于混合模型基于基准预测模型、数据预处理技术、智能优化算法等技术构建, 因此目前风电预测的方向已转变为对混合模型的研究。

由于风电功率时序特性具有明显的随机性、非线性, 因此使用分解算法一定程度上可以增加数据量, 降低原始序列的复杂度。分解算法是从时间序列数据中提取动态特征的有效策略^[9], 广泛应用在各个领域。李婧琦^[10]提出一种基于鲸鱼算法 (WOA) 优化长短期记忆网络 (LSTM) 的股票价格预测模型, 有效降低了时序数据间的冗余性。康文豪等^[11]为了解决风电功率间歇性带来的功率预测难度, 提出了一种基于 CEEMDAN-精细复合多尺度熵和 Stacking 集成学习的风电功率预测方法, 拥有更高的预测性能。庞博文等^[12]采用 CEEMDAN 对清洗后的原始功率数据进行分解, 得到若干个子序列, 通过对渭南某风电场机组实测数据进行实例仿真, 结果表明所提方法相较于其他模型预测精度得到了有效提升。

本文在上述研究的基础上, 将改进的 CEEMDAN-TCN 模型应用于风电功率预测。该方法利用 ICEEMDAN 算法充分挖掘了原始功率数据中的信息, 并利用时域卷积神经网络和基于注意力机制的时域卷积神经网络模型构建了一个耦合预测模型。本文提出的耦合预测模型在弗兰德伦地区实际风力发电功率数据集上进行了预测, 并显示出较高的准确性。

1 基于改进 CEEMDAN-TCN 预测模型构建

1.1 改进的 CEEMDAN 算法

1.1.1 集成经验模态分解

经验模态分解算法可将非线性时间序列数据分解为一系列具有不同振幅和残差的内涵模态分量 (Intrinsic Mode Functions, IMF)。在此基础上, 集成经验模态分解算法 (Ensemble Empirical Mode Decomposition, EEMD) 在原始信号中加入正态分布的白噪声, 以减少模式混合效应。

自适应噪声集合经验模态分解算法 (Complete

EEMD with Adaptive Noise, CEEMDAN) 通过在 EEMD 中加入有限自适应白噪声, 克服了 EEMD 不完整和加入白噪声后的重建误差问题。其可以将具有强波动性的原始数据, 完全分解为具有不同频率特性的几个本征模函数 (IMF) 分量, 从而降低数据的波动性并提高预测精度^[13]。算法实现步骤如下:

步骤 1 添加白噪声 $v^i(t)$ 与原始信号 $S(t)$ 呈标准正态分布。其中, 信号 i 可表示为 $S^i(t) = S(t) + v^i(t)$, $i = 1, 2, \dots, I$ 。相应的, 第一个分量可以表示为 $IMF_1 = \frac{1}{I} \sum_{i=1}^I IMF_1^i$, 残余分量为 $r_1(t) = S(t) - \frac{1}{I} \sum_{i=1}^I IMF_1^i$ 。

步骤 2 添加白噪声 $v^i(t)$ 到残余 $r_1(t)$, 执行 i 次实验 ($i = 1, 2, \dots, I$)。每个实验采用 EMD 进行分解 $r_1^i(t) = x(t) - v^i(t)$, 获得其一阶分量 IMF_1^i 。 $IMF_2 = \frac{1}{I} \sum_{i=1}^I E_1(r_1(t) + \varepsilon_1 E_1(v^i(t)))$, 残余分量为 $r_2(t) = S(t) - IMF_2$ 。

步骤 3 重复上述分解过程, 得到符合条件的 IMF 分量和相应的残差。当残差为单调函数且无法通过 EMD 分解时, 程序终止。原始信号可以表示为 $S(t) = \sum_{i=1}^n IMF_i + r_n(t)$ 。

1.1.2 改进的集成经验模态分解

改进的完全集成经验模态分解 (ICEEMDAN) 作为一种有效的数据处理策略, 可将原始风力发电时间序列分解为高频和低频时间序列, 从而解决 CEEMDAN 的残余噪声和杂散模态问题。ICEEMDAN 的实现过程如下:

(1) 将一组白噪声添加到原始序列中以获得新序列, 式(1):

$$x^{(i)} = x + \beta_0 E_1[\omega^{(i)}] \quad (1)$$

其中, x 是原始信号; $E_1[\omega^{(i)}]$ 是特殊噪声; $\omega^{(i)}$ 是第 i 个添加的白噪声 $i = 1, 2, \dots, N$; β_0 是第一个分解信号的预期信噪比。

(2) 对新序列进行 EMD 分解, 以获得第一组残差, 式(2)和剩余分量, 式(3):

$$r_1 = \frac{1}{N} \sum_{i=1}^N M[x^{(i)}] \quad (2)$$

$$\tilde{c}_1 = x - r_1 \quad (3)$$

其中, r_1 是第一次分解的残差; $[\cdot]$ 是计算局部均值的运算符; $M[\cdot]$ \tilde{c}_1 是第一个 IMF 的值。

计算第二个 IMF 的值见式(4), 残余分量, 见式(5):

$$r_2 = \frac{1}{N} \sum_{i=1}^N M\{r_1 + \beta_1 E_2[\omega^{(i)}]\} \quad (4)$$

$$\tilde{c}_2 = r_1 - r_2 \quad (5)$$

其中, r_2 是第二次分解的残余, \tilde{c}_2 是第二次风电功率时间序列的值。

(3) 根据上述计算步骤继续使用方程, 找到第 k 个 IMF 的值, 式(6)、式(7):

$$r_k = \frac{1}{N} \sum_{i=1}^N M\{r_{k-1} + \beta_{k-1} E_k[\omega^{(i)}]\}, k = 2, 3, \dots, N \quad (6)$$

$$\tilde{c}_k = r_{k-1} - r_k \quad (7)$$

其中, r_k 是第 k 个分解的残差, \tilde{c}_k 是第 k 个 IMF 的值。原始信号的 IMF 可以通过对该方程的迭代计算进行精确分解。

1.2 样本熵

样本熵(Sample entropy, SE)可以测量序列的复杂性, 具有更好的精度、更强的抗噪能力。样本熵代表时间序列复杂度的表征参数, 其值可以直接反映序列的复杂性。样本熵值越大, 序列越复杂, 反之亦然。样本熵用 $SE(n, \alpha, M)$ 表示, 描述如下

$$SE(n, \alpha) = \lim_{M \rightarrow \infty} \left\{ -\ln \left(\frac{R^{n+1}(\alpha)}{R^n(\alpha)} \right) \right\} \quad (8)$$

其中, N 是时间序列的长度; n 是维度; α 是相似的容差; $R^{n+1}(\alpha)$ 和 $R^n(\alpha)$ 是两个时间序列在阈值 α 处匹配 n 或 $n+1$ 点的概率。当 M 取有限值时, 由式(8)得到的估计样本熵计算, 式(9):

$$SE(n, \alpha, M) = -\ln \left(\frac{R^{n+1}(\alpha)}{R^n(\alpha)} \right) \quad (9)$$

1.3 注意力机制

注意力机制可以在稀疏数据中提取出有用的特征, 其优越性已逐渐在不同领域得到证明。注意力机制的主要思想是将更多的注意力分配给重要元素, 而不是平等地考虑所有元素。模型可以在不同时间为输入序列的元素赋予不同的权重。

Attention 权重计算见式(10)~式(12)。首先, 计算每个输入元素和时间点 t 时输出元素之间的相关性, $j = 1, 2, \dots, J$, 并由注意力分数式(10)表示, 其中, We 和 be 分别为注意力权重矩阵和偏置向量。其次, 通过 softmax 函数将相关性转换为概率形式, 计算在时间 t 处第 j 个元素输入序列的注意力权重表示为式(11); 最后, 将获得的概率和相应输入元素的隐式表示相乘, 以表示元素对预测值的贡献, 并将所有输入元素的贡献相加, 公式见式(12), 将其作为输入部分预测下一个值。

$$e_{ij} = \tanh(We x_{tj} + be) \quad (10)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^J \exp(e_{ij})} \quad (11)$$

$$\tilde{x}_{tj} = \alpha_{ij} x_{tj} \quad (12)$$

1.4 TCN 神经网络

TCN 是在一维卷积的基础上优化的一种新型时间序列模型^[14]。该模型具有以下突出特征:

(1) 使用扩张卷积。采样率由膨胀系数控制, 膨胀系数随着层次结构的增加呈指数增长。扩张卷积的计算见式(13):

$$F(s) = (x * f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (13)$$

其中: $*$ 是卷积运算; d 是膨胀系数; k 是卷积核的大小; $f(i)$ 代表卷积核的第 i 个元素; $s-d \cdot i$ 表示扩展卷积输入序列中参与计算的神经元的下标。

(2) 使用因果卷积。因果卷积如图 1 所示, 其结构在历史上是单向可追溯的, 是一个严格受时间约束的模型。

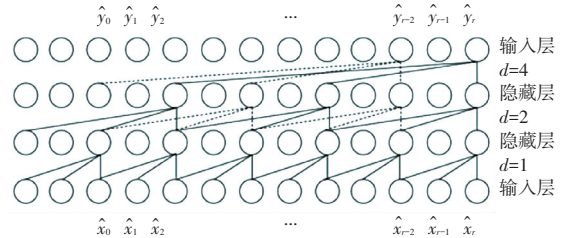


图 1 因果膨胀卷积结构

Fig. 1 The structure of causal expansion convolution

(3) 使用残差模块。残差模块涉及残差连接、权重归一化和辇差正则化, 以实现模型的效率。残差连接可以有效解决梯度消失问题, 从而有效地将浅层网络转变为深层网络。残差模块结构如图 2 所示。

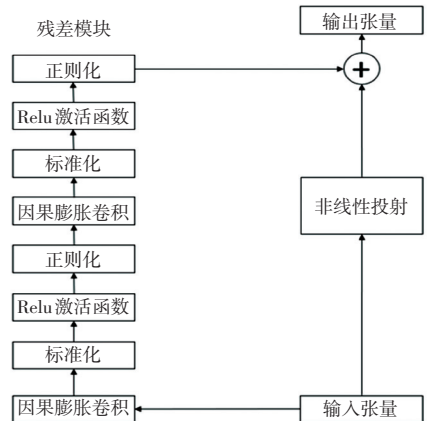


图 2 残差模块结构

Fig. 2 The structure of residual module

通过对样本数据的持续训练和权重的优化,当损失函数达到其最小值时,此时的输出数据被视为最优估计值^[15]。假设输入数据为: x_0, \dots, x_T , 输出数据为 y_0, \dots, y_T , 即 $y_0, \dots, y_T = f(x_0, \dots, x_T)$, 则 TCN 模型处理的新输出结果可以表示为 $y_0, \dots, y_T = g(x_0, \dots, x_T)$ 。

1.5 基于改进 CEEMDAN-TCN 模型构建

由于风电功率时间序列数据非线性、非平稳的特点。本文将数据分解算法与注意力机制以及 TCN 模型相结合,提出了一种基于 ICEEMDAN-TCN 的组合预测模型。模型采用 CEEMDAN 算法,将非线性非平稳时间序列预测问题转化为若干个稳态时间序列预测问题,将复杂的风电功率时间序列数据分解为相对简单的子序列。同时,利用样本熵分析分解的稳态子序列的复杂性,并重新组合形成一些新的子序列,从而减少了计算量和模型的复杂度。对于复杂度较高的子序列,采用基于注意力机制的 TCN 模型进行预测,可以获得更准确的预测结果;对于复杂度较低子序列,其周期性较强,因此采用结构简单的 TCN 进行预测,从而减少了训练时间,避免消耗更多的资源。具体实现步骤如下:

- (1) 利用 ICEEMDAN 及样本熵将风电功率时间序列数据分解为一系列复杂度差异明显的子序列。
- (2) 利用 TCN 和基于注意力机制的 TCN 模型,分别对低复杂度和高复杂度的子序列进行建模和预测。
- (3) 将序列预测结果叠加,得到最终的功率时序数据预测结果。
- (4) 分析预测值和实际风电功率的误差。

2 实例分析

2.1 数据集

以弗兰德伦地区陆上风电场实际风力发电功率作为实验数据,时间段选取 2022 年 10 月 16 日 00 时 00 分至 2023 年 2 月 16 日 23 时 45 分。其中间隔时间为 15 min,共计 11 908 条数据。如图 3 所示,该数据序列具有强随机性的特点和明显的非平稳性特征。

由于原始数据集有几处空缺,首先对缺失的数据进行补全。本文以 3 次求平均值的方式补全数据,式(14):

$$x_t = \frac{x_{t-1} + x_t + x_{t+1}}{3} \quad (14)$$

为了使预测模型更好的提取到功率数据的特征,采用归一化对数据进行处理,归一化公式见式(15):

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

其中, x 为归一化处理后的数据; x_{\min} 为功率序列数据的最小值; x_{\max} 为功率序列数据的最大值。

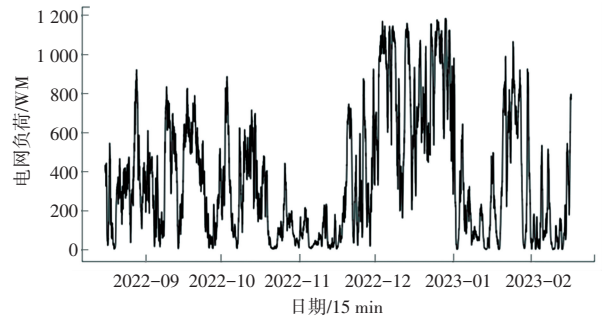


图 3 风电场风功率时间序列

Fig. 3 Wind power time series of wind farms

在实验分析阶段,所有实验的运行环境在 python3.7。本文将 11 908 条的原始风电功率序列数据按照 8 : 2 的比例来划分训练集和测试集,其中 9 526 条风力发电功率数据作为训练集,2 382 条风力发电功率数据作为测试集。

2.2 对比模型及评价指标

在对比模型上,本文选取 ICEEMDAN_SE_SVM、ICEEMDAN_SE_XGBoost、ICEEMDAN_SE_TCIN、ICEEMDAN_SE_注意力 TCN、ICEEMDAN_SE_LSTM 5 种模型进行对比实验。本文所提模型为基于 ICEEMDAN_SE 模型,高复杂度序列,即形成新序列 1、2,用注意力机制的 TCN 进行预测;低复杂度的序列,即形成新序列 3、4、5,用 TCN 进行预测;高复杂度的序列。

为了更具代表性,SVM 选用 RBF 核函数,通过网格搜索法确定惩罚因子和 RBF 核参数。XGBoost 通过正则化参数调优,选择树的深度为 6,每次生成树时随机抽样特征的比例为 1,最大迭代次数为 300 次。LSTM 参数选择序列输入量为 10,hidden_size = 64,num_layers = 2,output_size = 1,num_epochs = 300,batch_size = 64。TCN 的参数选择序列输入量为 10,输出量为 1,隐藏层数为 4,学习率为 0.000 1,dropout 率为 0.02,kernelsize 为 3,最大迭代次数为 300。此外,为验证模型的性能,对本文所提模型进行了不同时间步长以及峰值预测的对比实验,优化器使用 Adam,采用学习率递减方式进行训练。

本文采用均方根误差 (RMSE)、平均绝对误差 (MAE) 以及平均绝对百分比误差 (MAPE) 为评价指标,更好的评估了所用模型的有效性,见式(16)~式(18):

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (16)$$

$$\sqrt{MAE} = \sqrt{\frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|} \quad (17)$$

$$E_{MAPE} = \frac{1}{N} \sum_{t=1}^N \left(\frac{y_t - \hat{y}_t}{y_t} \right) \times 100\% \quad (18)$$

其中, y_t, \hat{y}_t 分别为 t 时刻的预测值和真实值, N 为样本总数。

2.3 数据分解及复杂度分析

采用 ICEEMDAN 算法对归一化的风电功率数据进行分析, 高斯白噪声标准差为 0.2。

风力发电功率时序序列数据被分解为 10 个子 IMF_s 分量和 1 个残余项, 如图 4 所示。由图中可见, IMF1-IMF6 具有高频特征, 而 IMF7-IMF10 具有低频特性, 故模型的超参数应不同。同时, 为对比 ICEEMDAN 分解的有效性, 使用 CEEMDAN 分解原始风电功率序列数据如图 5 所示。

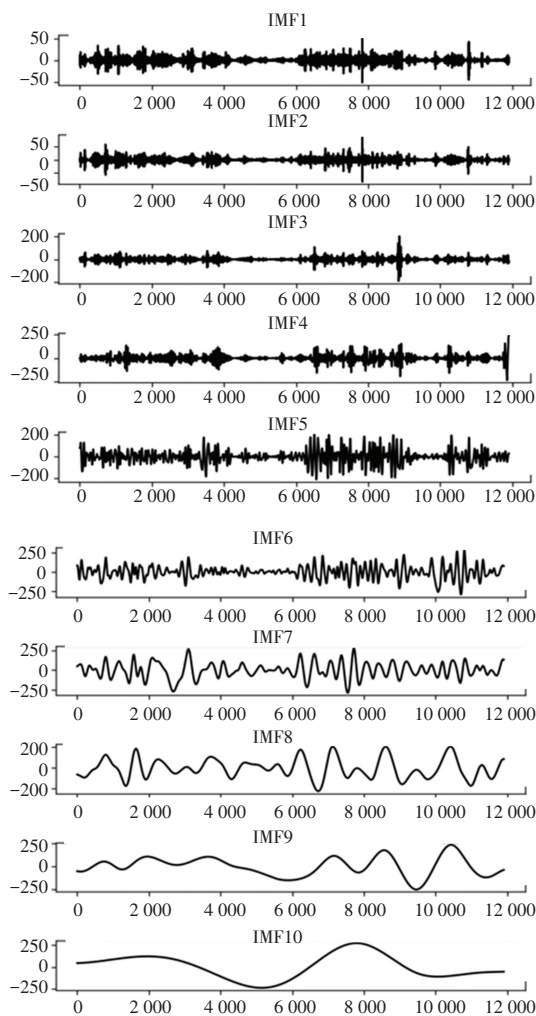


图4 ICEEMDAN 分解的各分量图

Fig. 4 Plot of each component of the ICEEMDAN decomposition

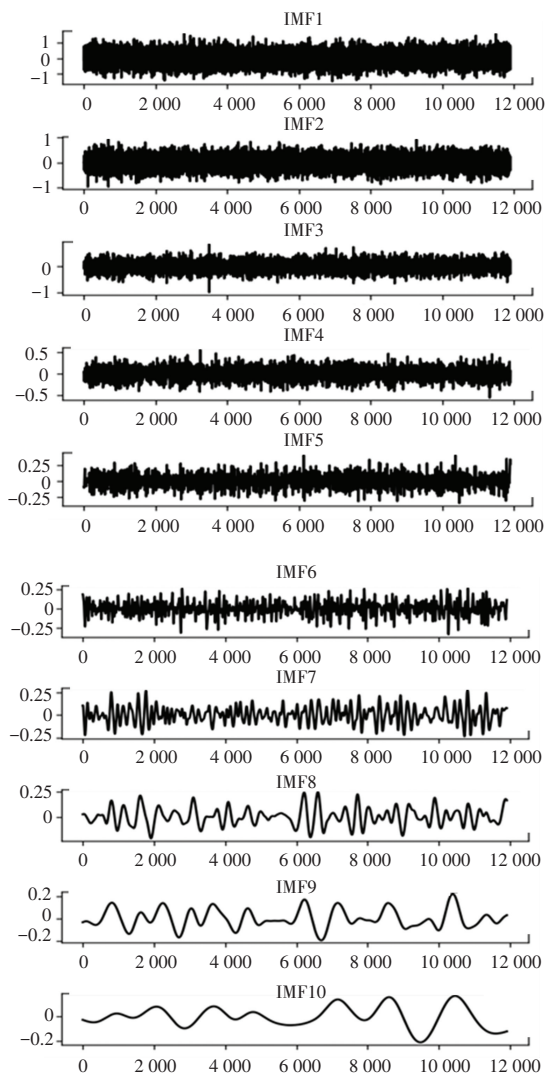


图5 CEEMDAN 分解的各分量图

Fig. 5 Plot of each component of the CEEMDAN decomposition

对比图 4 和图 5 可知, 使用 CEEMDAN 分解的分量在 IMF1-IMF8 都具有高频特征, 且序列复杂度明显比 ICEEMDAN 高很多。只有 IMF9-IMF10 具有低频特性, 且 ICEEMDAN 的低频序列复杂度更低。因此, 采用 ICEEMDAN 分解更有效性。

由于获得的 IMF 多达 10 个, 若直接对每个分量进行建模和预测, 计算量将大大增加。且过多分量误差的累积会导致最终预测结果出现过多的误差。因此, 本文利用样本熵来评估分解分量的复杂性, 并根据样本熵的值对每个分量进行重组, 然后对重组的子序列进行建模和预测。

由图 6 可知, IMF1、IMF2 子序列的熵值相差较小, 表明这两个 IMF 产生新模式的概率大致相同, 因此将这两个分量叠加形成一个新的子序列。由于 IMF1、IMF2 子序列样本熵均大于 1, 具有较强的随机性, 因此形成新的序列用基于注意力机制的 TCN

模型进行预测。同样,IMF3、IMF4、IMF5 3 个分量叠加形成一个新的子序列,并使用基于注意力机制的 TCN 模型进行预测。IMF6、IMF7 两个分量的熵值比较接近且低于 1,叠加为一个新的子序列熵值,相对熵值较低,因此采用 TCN 模型进行预测。同样,IMF8、IMF9 两个分量也是如此。IMF10 分量的熵值较低,因为单独作为一个分量采用 TCN 模型预测。各分量重组结果见表 1。

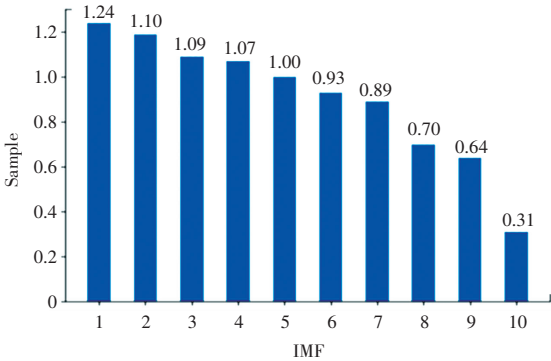


图 6 各分量样本熵

Fig. 6 Entropy of each component sample

表 1 功率子序列重组结果

Table 1 Power subsequence recombination results

原 CIMF 分量序号	新序号	预测模型
1,2	1	基于注意力机制的 TCN
3,4,5	2	基于注意力机制的 TCN
6,7	3	TCN
8,9	4	TCN
10	5	TCN

2.4 预测结果及分析

在得到重组子序列的预测结果后,可通过将预测结果相加得到最终的功率时间序列数据的预测值。为了更好的观察混合 TCN 模型的使用,采用 ICEEMDAN_SE_SVM、ICEEMDAN_SE_XGBoost、ICEEMDAN_SE_TCN、ICEEMDAN_SE_注意力 TCN、ICEEMDAN_SE_LSTM 5 种模型对各分量进行预测,预测结果见表 2。

表 2 分量预测评价指标对比

Table 2 Comparison of component prediction evaluation indicators

算法	新建序列 1		新建序列 2		新建序列 3		新建序列 4		新建序列 5	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ICEEMDAN_SE_SVM	9.10	2.29	13.85	2.81	9.33	2.31	9.46	2.33	4.81	1.67
ICEEMDAN_SE_XGBoost	9.57	2.33	17.72	2.89	9.74	2.35	10.10	2.40	5.19	1.70
ICEEMDAN_SE_TCN	8.12	2.11	11.75	2.59	7.91	2.07	8.05	2.12	4.12	1.50
ICEEMDAN_SE_注意力 TCN	8.01	2.04	11.21	2.48	8.31	2.09	8.35	2.14	4.24	1.51
ICEEMDAN_SE_LSTM	9.34	2.33	13.69	2.85	8.79	2.18	8.91	2.23	4.51	1.59

根据表 2 可以发现,在各分量预测的 5 种模型评价指标中,ICEEMDAN_SE_TCN 与 ICEEMDAN_SE_注意力 TCN 表现最优。新建序列 1、2 属于高频分量,使用 ICEEMDAN_SE_注意力 TCN 模型预测的各类指标都表现最优。新建序列 3、4、5 属于低频分量,使用 ICEEMDAN_SE_TCN 模型预测的各类指标都表现最优。因此,根据熵值可以表明,本文模型将高频稍复杂分量由基于注意力机制的 TCN 预测;低频较平稳分量由 TCN 预测,可以更好地挖掘各分量数据的信息。

将各模型预测分量数据结果叠加后,与原始功率时间序列数据对比的误差见表 3。

根据表 3,本文所提的模型在 RMSE 上,分别比 ICEEMDAN_SE_SVM、ICEEMDAN_SE_XGboost、ICEEMDAN_SE_LSTM 降低 40.79%、38.07%、43.57%。在 MAPE 上,分别比 ICEEMDAN_SE_SVM、

ICEEMDAN_SE_XGboost、ICEEMDAN_SE_LSTM 降低 25%、29.83%、41.41%。在 MAE 上,分别比 ICEEMDAN_SE_SVM、ICEEMDAN_SE_XGboost、ICEEMDAN_SE_LSTM 降低 41.26%、44.20%、58.43%。ICEEMDAN-TCN 模型表现出最佳的预测结果。如图 7,本文模型预测数据为红色线条、原始风电功率序列曲线为黑色数据,相较于其它模型,所提预测模型在尖峰处也有较好的预测效果。

表 3 总体预测指标误差表

Table 3 Table of indicators of overall forecast error

模型	RMSE	MAPE	MAE
ICEEMDAN_SE_SVM	291.12	2.32	197.42
ICEEMDAN_SE_XGboost	278.4	2.48	207.83
本文模型	172.36	1.74	115.96
ICEEMDAN_SE_LSTM	305.45	2.97	279.01

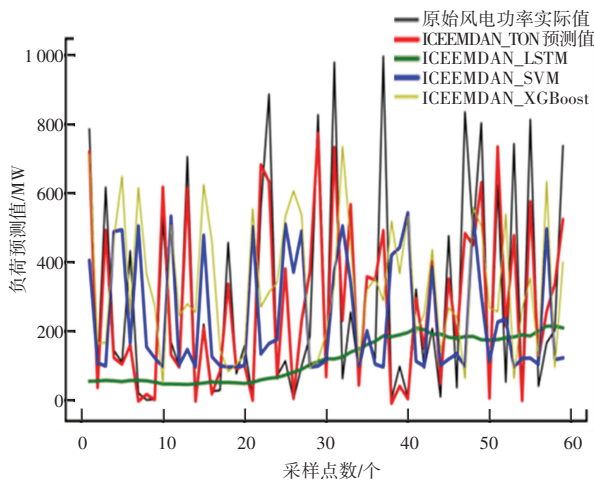


图7 各模型与原始数据对比图

Fig. 7 Comparison of each model with the original data

为验证模型的性能,对本文所提模型进行了不同时间步长的预测。根据表4,时间步长分别为单步、2步、4步、6步,4种步长的指标差别不大。在4种步长中4步时间步长的 $RMSE$ 为 169.96,相较于单步、4步、6步在 $RMSE$ 上分别降低 1.3%、0.5%、3.9%。

表4 不同时间步长预测评价指标

Table 4 Prediction and evaluation indicators for different time steps

时间步长	$RMSE$	$MAPE$	MAE
单步	172.36	1.74	115.96
2步	170.60	1.13	112.64
4步	169.96	1.19	114.33
6步	176.91	1.81	117.56

3 结束语

针对风电功率时间序列的预测,本文提出一种基于改进的 CEEMDAN-TCN 风电功率模型。在弗兰德伦地区陆上风电场的真实运营数据集的实验中,所提方法不仅克服了单一模型不能有效提取荷载数据特征的缺点,而且有效提高了预测精度。

(1)通过 ICEEMDAN 算法将随机性强的非平稳序列预测问题转化为多个相对稳定的时间序列预测问题,可以充分挖掘了原始功率时间序列数据中的信息,降低了风电功率非线性、复杂性对风电功率预测的难度。

(2)采用样本熵评估分解子序列的信息复杂度,对相近的子序列进行合并,利用分解集成的方法,对两种复杂度的子序列采取不同的预测方法有助于提高整体预测精度。

(3)使用时域卷积神经网络的方法,一方面减少了网络堆叠带来的过拟合问题,一方面能更好识别数据特征。

参考文献

- [1] YANG M, SHI C, LIU H. Day-ahead wind power forecasting based on the clustering of equivalent power curves [J]. Energy, 2021, 218(3): 119515.
- [2] NIE Y, JIANG P, ZHANG H. A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting [J]. Applied Soft Computing, 2020, 97(10): 106809.
- [3] HE Y, WANG Y. Short-term wind power prediction based on EEMD-LASSO-QRNN model [J]. Applied Soft Computing, 2021, 105(7): 107288.
- [4] ZHOU Min, WANG Bo, GUO Shudong, et al. Multi-objective prediction intervals for wind power forecast based on deep neural networks [J]. Information sciences, 2021, 550(3): 207-220.
- [5] WANG Jianzhou, HU Jianming. A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model [J]. Energy, 2015, 93(12): 41-56.
- [6] LUO H, DOU X, SUN R, et al. A multi-step prediction method for wind power based on improved TCN to correct cumulative error [J]. Frontiers in Energy Research, 2021, 9: 723319.
- [7] JAHANGIR H, TAYARANI H, GOUGHERI S, et al. Deep learning-based forecasting approach in smart grids with microclustering and bidirectional LSTM network [J]. IEEE transactions on industrial electronics, 2021, 68(9): 8298-8309.
- [8] KO M S, HONG C W, HUR K. Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting [J]. IEEE transactions on sustainable energy, 2021, 12(2): 15.
- [9] LEI B, WANG, Z. Research on short-term load forecasting method based on EEMD-CS-LSSVM [C] // Proc. CSU-EPSA, 2019: 117-122.
- [10] 李婧琦. 基于鲸鱼算法优化 LSTM 的股票价格预测模型 [J]. 智能计算机与应用, 2023, 13(2): 35-40.
- [11] 康文豪, 徐天奇, 王阳光, 等. 基于 CEEMDAN-精细复合多尺度熵和 Stacking 集成学习的短期风电功率预测 [J]. 水利水电技术(中英文), 2022, 53(2): 163-172.
- [12] 庞博文, 丁月明, 杜善慧, 等. 基于 CEEMDAN-BO-LSTNet 的风电出力短期预测 [J]. 电测与仪表, 2023, 60(9): 109-116, 170.
- [13] BAN W, SHEN L. PM2.5 prediction based on the CEEMDAN algorithm and a machine learning hybrid model [J]. Sustainability, 2022, 14(23): 16128.
- [14] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J]. arxiv preprint arxiv:1803.01271, 2018.
- [15] HUANG Ru, DU Wenjuan, WANG Haifeng. Short-term prediction of wind power considering turbulence intensity [J]. Power Syst Technol, 2019, 43(6): 1970-1913.