

# 基于贝叶斯参数优化 Transformer 时间段分割模型的超短期风电功率预测方法

江善和, 徐小艳, 涂亮, 陈文胜

(安庆师范大学电子工程与智能制造学院, 安庆 246011)

**摘要:** 针对风电场数据具有复杂的非线性关系,且难以捕捉长距离依赖关系的问题,引入时间段分割策略,提出一种新的基于贝叶斯参数优化 Transformer 时间段分割模型的超短期风电功率预测方法(PBY-Trans)。该方法采用时间段分割技术分割风电场数据为子序列,将其作为 Transformer 模型编码器的输入,从而更好地适应时间序列的非线性特征;使用贝叶斯算法搜索 Transformer 模型参数的最优配置,实现模型性能提升,进一步提高预测准确度。采用 Bengaluru 某风电场数据集对所提方法的预测性能进行比较验证,相较于 SVM、RNN、Informer、LSTM、GRU 和 TCN 模型,所提 PBY-Trans 方法在平均绝对误差(MAE)分别下降 33.56%、59.75%、47.27%、32.34%、40.46% 和 27.71%,均方根误差(RMSE)分别下降 68.99%、37.05%、27.60%、14.43%、16.42% 和 12.29%。结果表明,该文提出的 Transform 时间段分割模型的风电功率预测方法能够进一步提高预测精度,鲁棒性能得到有效提升。

**关键词:** 风电功率; 预测模型; 贝叶斯优化; Transformer 模型; 时间段分割

**中图分类号:** TM614

**文献标志码:** A

## 0 引言

随着可再生能源在能源行业的广泛应用,风力发电作为一项重要的清洁能源,其发电量的准确预测对于电力系统的稳定运行至关重要。在现代电力系统中,超短期风电功率预测的准确性直接影响着电力供需平衡的实现以及电力市场的稳定运行<sup>[1]</sup>。然而,与传统能源相比,风力发电的不确定性和波动性使得其功率预测工作更加具有挑战性<sup>[2]</sup>。风速的突然变化、地理位置的差异以及气象条件的复杂性都增加了风电功率预测的复杂度。

传统的风电功率预测方法往往依赖于数值天气预报(numerical weather prediction, NWP)数据<sup>[3-4]</sup>,结合统计模型或机器学习技术进行预测,如差分自回归移动平均(autoregressive integrated moving average, ARIMA)<sup>[5]</sup>、支持向量机(support vector machine, SVM)<sup>[6]</sup>等。然而,由于风力发电具有显著的间歇性和不稳定性,这些方法在面对复杂气象条件和多变的风电场景时,预测性能仍有待提高,且在处理大规模数据和捕捉时间序列的长期依赖关系方面存在局限性。

近年来,深度学习技术在时间序列分析领域取得突破性进展<sup>[7-8]</sup>,其中循环神经网络(recurrent neural network, RNN)<sup>[9]</sup>、长短期记忆(long short-term memory, LSTM)神经网络<sup>[10]</sup>、门控

循环单元(gated recurrent unit, GRU)<sup>[11]</sup>、深度回波网络<sup>[12]</sup>、生成对抗网络(generative adversarial networks, GAN)<sup>[13]</sup>、时间卷积网络(temporal convolutional network, TCN)<sup>[14]</sup>和 Temporal Fusion Transformers(TFT)<sup>[15]</sup>被广泛应用于风电功率预测。特别是 Transformer 模型因其出色的特征提取和长距离依赖关系建模能力而备受关注<sup>[16]</sup>,在自然语言处理<sup>[17]</sup>、计算机视觉处理<sup>[18]</sup>、语音模式识别<sup>[19]</sup>等众多领域得到广泛应用,尤其在时序序列领域取得巨大成功<sup>[20]</sup>,由于其能够从序列数据中提取出长距离关联信息,因而非常适合用于时序建模。同时,在处理时间序列建模时,Informer<sup>[21]</sup>、Autoformer<sup>[22]</sup>和 FEDformer<sup>[23]</sup>也是一种较好的变体。然而,最新研究表明,无论 Transformer 的设计多么复杂,仅是一种极其简单的线性模型,但其性能比以往任何一种模型都要好<sup>[24]</sup>。文献[25]探讨了 Transformer 模型在风电功率预测中的应用,特别是在多风场相关性分析中的应用,指出 Transformer 模型在捕捉长依赖关系方面表现出色,但需要与其他模型结合,以更好地处理局部关系和提高预测精度。文献[26]介绍了深度学习模型在风电功率预测中的应用,包括 Transformer 模型。这些研究表明,Transformer 模型通过其自注意力机制能够有效捕捉长时间序列数据中的依赖关系,但可能忽略局部关系。虽然 Transformer 架构在时序序列预测中能够在长距离信息依赖

收稿日期: 2024-09-02

基金项目: 国家自然科学基金(51607004); 安徽省自然科学基金(2008085MF197)

通信作者: 江善和(1975—),男,博士、教授,主要从事智能计算理论、风电预测方面的研究。jshxlw@163.com

方面表现出卓越的捕捉能力,但其在风电功率预测中的应用仍面临两大瓶颈:其一,原始时序的全局输入策略难以辨识局部气象事件(如阵风突变)导致的非线性模式切换;其二,模型超参数组合的搜索空间随网络复杂度呈指数增长,传统的试错法调参易陷入局部最优。为此,本文提出一种基于贝叶斯参数优化的 Transformer 时间段分割模型 PBV-Trans。该模型采用基本的 Transformer 模型,融合时间段分割技术分割子序列,将其作为 Transformer 模型编码器的输入,具有能够更精细地捕捉风电功率时间序列的内在规律。此外,为进一步提升模型的泛化能力和鲁棒性,引入贝叶斯参数优化方法,通过后验分布采样实现模型参数的不确定性估计和优化。

## 1 Transformer 模型

Transformer 是基于自注意力机制替代传统循环网络的深度学习架构,由 Vaswani 等<sup>[16]</sup>于 2017 年提出,被广泛应用于自然语言处理领域。相比于传统的循环神经网络(RNN)<sup>[9]</sup>和卷积神经网络(convolutional neural network, CNN)<sup>[27]</sup>,Transformer 在处理序列数据时能够更好地捕捉长距离依赖关系,避免了信息在远距离传递时产生梯度消失或爆炸的问题。

Transformer 模型的核心是注意力机制,通过对输入序列中每个位置的信息进行加权求和,得到上下文表示,这使得模型能够实现全局性的信息交互,无需考虑输入序列的顺序。此外,Transformer 模型利用编码器和解码器分别用于提取输入序列的特征和生成输出序列。

### 1.1 多头自注意力机制

自注意力机制的优势是能够让模型针对不同位置的输入信息之间建立起长距离连接的依赖关系。在传统的循环神经网络中,信息只能从前向后或从后向前传递,无法进行全局的信息交互。自注意力机制通过计算每个位置之间的注意力权重,可实现全连接的位置关联,这种机制使得模型预测时能够更好地理解输入数据之间的关系,并有效地捕捉到时间段分割的特征,其注意力机制网络结构如图 1 所示。自注意力机制是将模型输入中的所有 token 均与其他 token 建立起关联,并将输入数据序列中的每个矩阵  $X$  分别输入到 3 个完全不同的全连接层,以此创建 3 个向量,即查询(Query,  $Q$ )向量:  $Q = XW^Q$ ; 键(Key,  $K$ )向量:  $K = XW^K$ ; 值(Value,  $V$ )向量:  $V = XW^V$ ,其中  $W^Q$ 、 $W^K$ 、 $W^V$  分别是用于线性变换的权重矩阵。

查询和键经过点积矩阵乘法产生一个分数矩阵,分数矩阵相当于确定了一个 token 是如何关注其他的 token。因此,每个 token 都会有一个与时间步长中其他 token 相对应的分数,分数越高,关注度越高,即查询如何映射到键的。然后对缩放后的得分进行 Softmax 计算,得到注意力权重,从而获得 0 到 1 之间的概率值,较高的得分会得到增强,而较低的得分

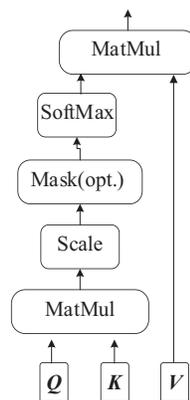


图 1 注意力机制网络结构

Fig. 1 Structure of scaled dot-product attention

就会被抑制。最后,将注意力权重与值向量相乘,得到输出向量。注意力机制的计算公式如式(1)所示。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

式中:  $d_k$ ——Key 的维度,通过查询和键的维度开平方并加以缩放,这样可以使梯度更加稳定,避免乘法产生爆炸效果。

较高的 Softmax 得分会保留模型认为更重要的 token 值,较低的得分将淹没不相关的 token。但在某些情况下,单个注意力头可能无法捕捉到所有必要的信息,尤其是当序列非常长或任务非常复杂时,就需要引入多头自注意力机制来改善。

多头自注意力模块如图 2 所示,多头自注意力机制通过引入多个注意力头,每个头都能够关注序列中不同位置的信息,为使每个注意力头能够完成多头自注意力计算,因此应用自注意力计算之前,需要将每个查询、键和值分解成  $R$  个向量,这样分割完成的  $R$  个向量能够实现各自相同的自注意力计算过程。于是,每个自注意力的计算过程称之为一个头,且每个头都会自动产生一个输出向量,这些输出向量再通

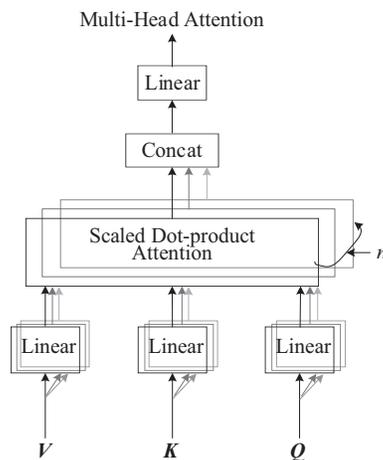


图 2 多头自注意力模块

Fig. 2 Multi-head self-attention module

过线性层完成拼接形成一个向量。至此,每个头均会学习到不同头的连接信息,从而为 Transformer 模型的编码器提供更多的信息表达能力。多头自注意力机制的表达式如式(2)所示。

$$\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O \quad (2)$$

式中:  $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ 。

多头自注意力的输出通过输出权重矩阵  $W^O$  映射到期望的输出维度上以产生最终的多头自注意力表示。多头自注意力机制能够同时考虑到输入序列中不同位置的信息,并通过使用多个注意力头来捕捉不同的序列特征。

### 1.2 编码-解码结构

在 Transformer 模型中,编码-解码结构是实现序列到序列建模的核心组件。编码器将输入数据序列能够映射到一组抽象的连续表示,且包括了该输入序列中的所有学习过程信息,解码器接着获取这个连续表示,并逐步生成单一输出,同时还会输入前一个输出。在短期风电功率预测任务中,编码器接收历史风电功率数据作为输入序列,而解码器则负责生成未来时刻的预测风电功率。为实现编码器和解码器之间的信息交互,自注意力机制允许模型在同一序列中在不同位置之间建立关联,从而捕捉到序列内部依赖的长距离关系。同时借鉴多头注意力机制,模型可同时获得多组不同学习机制的注意力权重,以提高模型的表达能力和泛化性能。

编码器和解码器分别如图 3 的左半部分和右半部分所示。编码器和解码器具有多个相同的层组成。编码器负责将

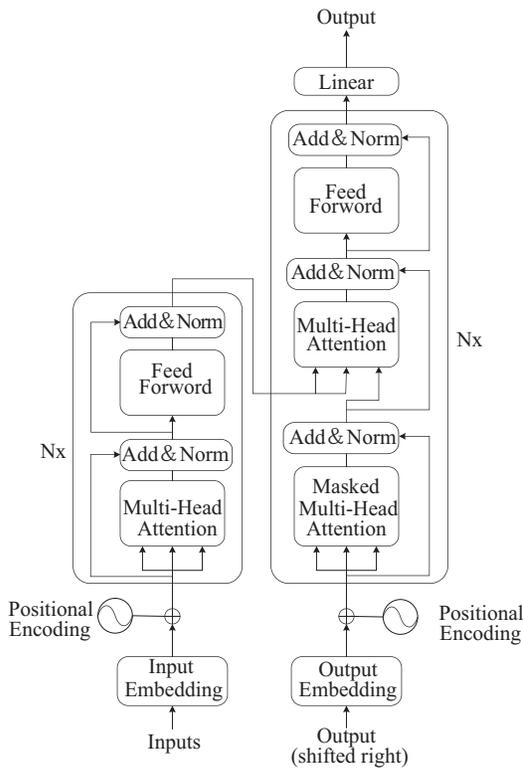


图3 编码-解码结构

Fig. 3 Structure of encoding-decoding

输入序列转换为上下文表示,其中包含了输入序列的信息。它包括多头自注意力层和前馈全连接层,以捕捉输入序列内部的依赖关系并进行非线性变换。残差连接和层归一化用于稳定训练和防止梯度消失。解码器则利用编码器生成的上下文表示逐步生成目标序列,通过掩码自注意力层和编码器-解码器注意力层来实现。分割结构使得模型能够更好地适应时间序列数据的局部特征变化,提高预测的准确性和稳定性。在时间段分割的 Transformer 模型中,编码器和解码器根据时间段的特征进行分割,并分别对每个时间段内的数据进行建模。这种分割结构使得模型更好地适应时间序列数据的局部特征变化,并提高风电功率预测的准确性和稳定性。

## 2 时间段分割技术

时间序列进行分析,旨在揭示各阶段间的相互关系。但是,相对于词语,单一的时间步无独立的意义,如何从文本中获取局部的语义信息是理解时序关系的关键。已有研究多是基于单点输入标记的方法<sup>[28]</sup>,或手动抽取数据进行分析<sup>[22]</sup>,而将时序数据融合在子时序层次上,能够更好地捕捉到单个节点所不能覆盖的完整语义信息。

时间段分割<sup>[29]</sup>是一种将连续的时间序列数据分割成多个时间段的策略,它可帮助模型更好地捕捉到数据的局部特征和变化趋势。在风电功率预测中,由于风速和风向的变化受到多种因素的影响,如季节、天气条件等,因此,时间序列数据往往呈现出强的非平稳性。通过时间段分割,可将复杂数据分解为更易于分析和预测的子序列(patch),以帮助模型更好地捕捉到数据的局部特征和变化趋势。分割数据为子序列(patch)的过程如图 4 所示。

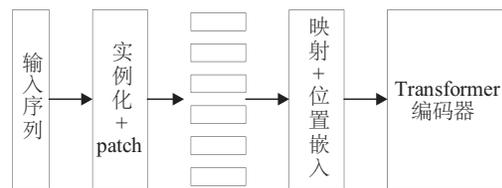


图4 分割 patch

Fig. 4 Patch splitting

本文采用基于滑动窗口的时间段分割方法,通过设定固定长度的窗口,在时间序列上按照一定的步长滑动,将原始序列分割成一系列不重叠的时间段。每个时间段内的数据被视为一个独立的子序列用于模型的训练。

给定一个时间序列  $\hat{X} = (x_1, x_2, \dots, x_n)$ , 设定窗口长度为  $L$ , 滑动步长为  $S$ , 则有:

$$p_i = \{x_{(i-1) \cdot S + 1}, x_{(i-1) \cdot S + 2}, \dots, x_{(i-1) \cdot S + L}\} \quad (3)$$

式(3)生成子序列  $P = (p_1, p_2, \dots, p_m)$ , 其中  $m = \lfloor \frac{n-L}{S} \rfloor + 1$

为子序列的数量。通过将输入数据按照时间段分割成多个

子序列,并将这些子序列作为 Transformer 模型编码器的输入,使得模型能够更好地利用时间序列数据的时序信息。然后预测  $T$  步的未来值,即 Transformer 将提供预测结果  $(x_{L+1}, x_{L+2}, \dots, x_{L+T})$ 。具体步骤如下:

- 1) 对数据集进行归一化和分割训练集。
- 2) 通过滑动窗口的方式截取一定长度的历史时间步数据,形成输入窗口。
- 3) 利用滑动窗口的方式,每次移动一定的步长,构建连续的历史窗口,形成一个 patch。每个 patch 包含多个历史窗口,而每个历史窗口则包含了一定数量的历史时间步数据。
- 4) 根据参数确定目标标签预测未来若干时间步的值。

### 3 贝叶斯优化算法

贝叶斯优化是一种通过构建目标函数的后验分布来指导搜索过程的全局优化算法<sup>[30]</sup>。在贝叶斯优化中,由于其对目标函数的具体形式要求较低,它能够适应不同数据集和模型的调优需求,具有较好的通用性,且目标函数可被视为一个黑盒函数,即其具体形式未知,只能通过输入输出观测值来进行评估。

传统的网格搜索<sup>[31]</sup>或随机搜索算法<sup>[32]</sup>在调参过程中效率低下,而贝叶斯优化则通过构建参数优化的概率模型,根据先验信息和观测结果动态地更新参数的选择,在有限的迭代次数内逐步逼近最优解<sup>[33]</sup>。具体来说,贝叶斯优化通过不断实验不同的参数配置,并根据先前实验的结果来选择下一个实验的参数组合,从而在较少的尝试次数内找到最优参数配置,提高模型的泛化性能。

在超短期风电功率预测中,贝叶斯优化可用于优化 Transformer 模型的参数,以提高预测的准确性和效率。通过构建目标函数,利用贝叶斯优化可搜索最优的参数组合,获得更加准确和可靠的预测结果。

#### 3.1 贝叶斯优化原理

贝叶斯优化的核心思想是将目标函数视为一个随机过程,并在已知的观测值的基础上,利用贝叶斯定理来更新目标函数的后验分布。通过不断地观测和更新,贝叶斯优化可找到全局最优解或接近最优解的解。高维参数空间搜索问题是在进行贝叶斯参数优化过程中经常面临的挑战之一。在短期风电功率预测中,需要优化多个参数,这将导致参数空间的维度增加,使得搜索变得更加困难。

为解决高维参数空间搜索问题,这里采用高斯过程<sup>[34]</sup>作为贝叶斯优化的概率模型。然而,贝叶斯优化过程中,易陷入局部最优解空间上进行不断采样,需要权衡优化算法的开发和探索性能。开发(exploitation)能力指的是在最大概率获得全局最优解的空间区域上进行采样,因此开发深度越高,其均值越高;而探索(exploration)指的是在未取样的空间区域上新获取采样点,则探索程度越高,其方差就会越大。

如何在开发和探索过程中进行高效采样,需要寻找下一个采样点  $x$  的采取函数,本文选取期望函数(expected improvement, EI)作为有效平衡开发和探索的采取函数<sup>[35]</sup>,从而获取下一个采样点的计算值,如式(4)和式(5)所示。

$$H(x) = \begin{cases} Z\sigma(x)\phi(Z) + \sigma(x)\varphi(x), & \sigma(x) > 0 \\ 0, & \sigma(x) = 0 \end{cases} \quad (4)$$

$$Z = \begin{cases} \frac{\mu(x) - f(x^*) - \xi}{\sigma(x)}, & \sigma(x) > 0 \\ 0, & \sigma(x) = 0 \end{cases} \quad (5)$$

式中: $\sigma(x)$ ——在超参数配置  $x$  下模型预测的目标函数值的标准差; $\phi, \varphi$ ——标准正态分布的分布函数和概率密度函数; $\mu(x)$ ——超参数配置  $x$  下模型预测的目标函数值的均值; $x^*, f(x^*)$ ——当前已知的最优样本点和高斯过程后验给出的最优样本点的预测值; $\xi$ ——决定开发已知设计空间和探索未知设计空间之间的权重。

当 EI 值较大时,意味着在超参数配置下的探索可能会带来更大的性能提升,因此超参数配置更有可能成为下一个实验点。通过采取函数,优化算法可在每次迭代中根据当前模型的预测和观察到的目标函数值来决定下一个实验点,从而有效地探索超参数空间,找到最佳的超参数配置。

#### 3.2 贝叶斯优化流程

贝叶斯寻优参数流程如图 5 所示。该流程始于创建一个 Study 对象,其旨在追踪实验结果和超参数配置的优化过程。具体步骤如下:

- 1) 初始化用于贝叶斯优化的 Study 对象。检查是否已有初始化点,若有则根据上一轮的验证损失,选出表现较好的点集合作为新的初始点,这些点将用于指导下一轮的搜索方向。否则随机生成初始点。
- 2) 定义目标函数用于最小化验证集损失。根据当前的超参数配置训练模型,并计算验证集上的损失。
- 3) 计算高斯过程建模验证集损失。基于高斯过程模型,计算每个候选超参数配置的期望改进(EI)值,并选择最大 EI 值的超参数配置。
- 4) 检查是否达到迭代次数。若未达到,返回步骤 3)高斯过程建模,选择新的超参数配置并继续迭代优化过程。每次迭代完成后记录超参数配置和验证集损失。
- 5) 从 Study 对象中提取最佳超参数配置,使用该配置重新训练模型。最后在测试集上评估模型性能并计算均方误差(mean squared error, MSE),输出最终结果,结束优化流程。

#### 3.3 时间复杂性分析

贝叶斯优化算法的计算时间效率受到多个因素的影响,尤其是 Transformer 模型的训练过程和贝叶斯优化本身的迭代过程,故贝叶斯优化的计算时间复杂性主要由这两个部分组成。Transformer 的训练过程通常由自注意力机制主导,其

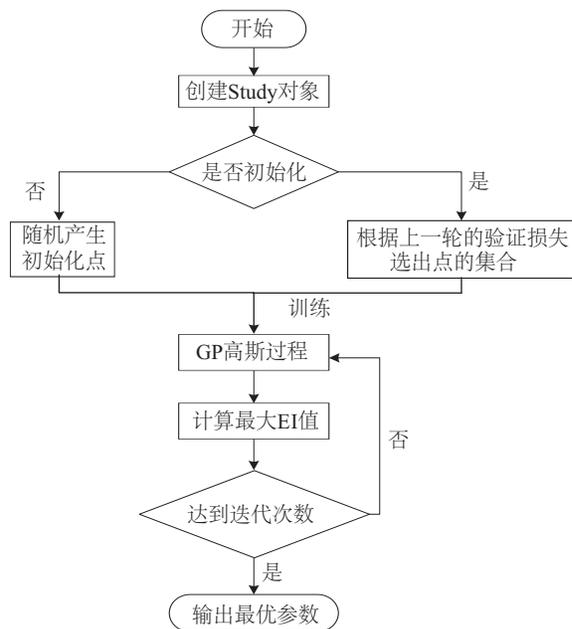


图5 贝叶斯优化流程

Fig. 5 Bayesian optimization process

时间复杂度为  $O(n^2)$ , 其中  $n$  是输入序列的长度。本方法中, 数据经过时间段分割后被处理为多个子序列, 每个子序列的长度为  $l$ , 则每个子序列的时间复杂度为  $O(l^2)$ 。对于  $m$  个子序列, 整体训练的时间复杂度为  $O(ml^2)$ , 而  $m \times l = n$ 。贝叶斯优化的每次迭代都需要训练模型以评估目标函数, 因此每次迭代的时间复杂度为  $O(\ln)$ , 而贝叶斯优化进行了  $k$  次迭代, 即贝叶斯优化的总时间复杂度为  $O(k \ln)$ 。通过贝叶斯优化, 所提 PBV-Trans 方法能够在较少的迭代次数内找到最优参数配置, 避免了传统网格搜索和随机搜索的低效率, 尤其是在高维参数空间中, 贝叶斯优化能够显著减少搜索次数, 从而提高模型调优的效率。

## 4 实验设置及结果分析

### 4.1 实验设置

本文采用 Bengaluru 某风电场 2007 年 1 月—2021 年 7 月的风电数据。该数据集采集频率为每小时 1 次, 记录的内容除风电功率外还包含气温、露点温度、相对湿度、风速、风向、气压等气象数据。所提的模型和算法采用 Python3.6 实验平台进行验证。在进行实验之前, 首先对样本数据完成清洗, 删除异常数据, 补充缺项数据, 从而保证样本的可靠和质量。同时, 进行数据的归一化处理, 将数据统一缩放到相同的尺度, 以便模型更好地学习特征之间的关系。将处理后的样本划分成 3 个部分, 即 40% 样本为训练集, 10% 为验证集, 50% 为测试集。

本文模型中用于建模序列的历史信息 look\_back 步数为 60, 预测未来的时间步  $T$  为 1, 数据被分为 96 个 patch 序列。

Transformer 模型中关键参数包括嵌入维度、学习率、失活率、层数、隐藏单元数等, 这些超参数直接影响着模型的训练速度、收敛性能以及预测准确度。为获取最佳的超参数配置, 根据贝叶斯优化算法, 对 Transformer 模型中的超参数进行多轮实验, 实验结果如图 6 所示, 其中第 14 轮的均方误差 (MSE) 最小, 即该轮实验寻到最优的超参数配置。

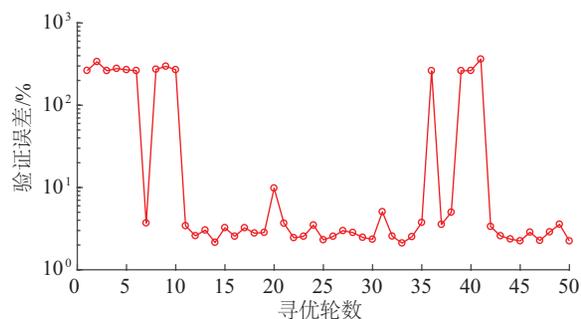


图6 Transformer模型超参数优化结果

Fig. 6 Results of Transformer model hyperparameter optimization

表 1 为达到迭代次数后得到的各超参数的最优值。Transformer 模型的超参数即采用表 1 贝叶斯寻优的结果。通过将高斯过程应用于贝叶斯优化中, 在高维参数空间中高效地搜索最优解, 避免传统优化算法在高维空间中需要大量计算资源和时间的问题。

表1 贝叶斯优化 Transformer 超参数结果

Table 1 Transformer hyperparameter results optimized by Bayesian

超参数	Dense_ dim	num_ heads	embed_ dim	dropout_ rate	num_ blocks	learn_ rate
最优值	32	16	128	0.01	4	0.0001

### 4.2 结果分析

在对比实验中, 选择 SVM<sup>[6]</sup>、RNN<sup>[9]</sup>、Informer<sup>[21]</sup>、LSTM<sup>[10]</sup>、GRU<sup>[11]</sup>和 TCN<sup>[14]</sup>预测模型, 与本文所提 PBV-Trans 模型进行比较。通过计算 3 个评价指标决定系数 ( $R^2$ )、平均绝对误差 (MAE) 和均方根误差 (RMSE) 3 个评价指标, 对各模型的预测性能进行评估, 如表 2 所示。

表2 7种预测方法的性能指标结果

Table 2 Results of performance indicators of seven methods

预测方法	$R^2$	MAE	RMSE
SVM	0.9440	50.75	113.87
RNN	0.9505	83.77	107.03
Informer	0.9610	63.96	94.98
LSTM	0.9732	49.84	78.75
GRU	0.9719	56.64	80.62
TCN	0.9753	46.65	75.67
PBV-Trans	0.9804	33.72	67.38

从表 2 所示 7 种预测方法的性能指标结果可知, SVM、RNN、LSTM、GRU、TCN 模型均展现出较为满意的预测性能,但在这些方法中, PBY-Trans 模型表现最为突出, 显著优于其他模型。PBY-Trans 的  $R^2$  值达到 0.9804, 显著高于 SVM (0.9440)、RNN (0.9505)、Informer (0.9610)、LSTM (0.9732)、GRU (0.9719) 和 TCN (0.9753)。此外, PBY-Trans 在 MAE 和 RMSE 指标上也表现优异, 分别为 33.72 和 67.38, 远低于其他对比模型。特别地, PBY-Trans 在 MAE 指标方面较 SVM、RNN、Informer、LSTM、GRU 和 TCN 模型分别降低 33.56%、59.75%、47.27%、32.34%、40.46% 和 27.71%; 在 RMSE 指标方面较其他模型分别下降 68.99%、37.05%、27.60%、14.43%、16.42% 和 12.29%。上述分析结果表明, PBY-Trans 模型不仅提高了预测精度, 且在 MAE 和 RMSE 方面展现出显著优势。

同时, TCN 和 LSTM 模型在  $R^2$ 、MAE 和 RMSE 3 个指标上均表现良好, 特别是在  $R^2$  和 RMSE 上仅次于 PBY-Trans。GRU 模型在  $R^2$  和 RMSE 方面的表现也较为优秀, 虽然不如 TCN 和 PBY-Trans, 但在 MAE 方面仍具有较强的竞争力, 而 SVM、RNN 和 Informer 模型的表现相对较差。图 7 和图 8 分别给出了各模型 MAE、RMSE 的变化趋势。

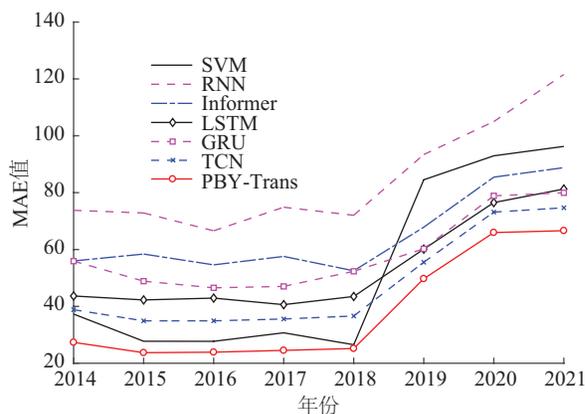


图 7 7种模型的MAE比较

Fig. 7 Comparison results of MAE for seven models

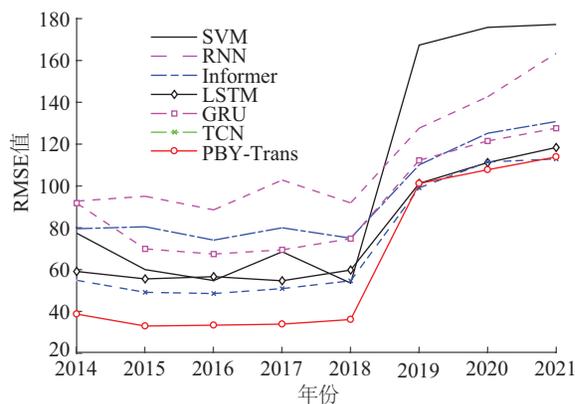


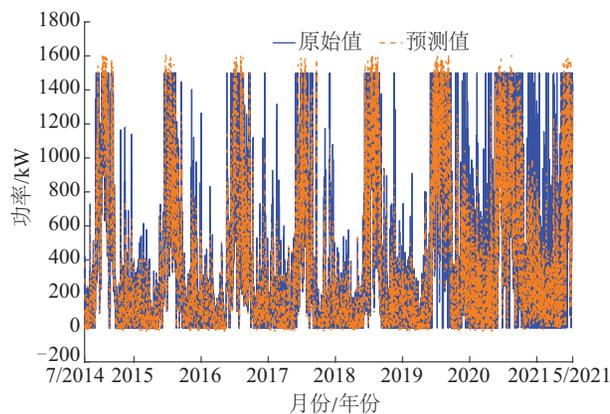
图 8 7种模型的RMSE比较

Fig. 8 Comparison results of RMSE for seven models

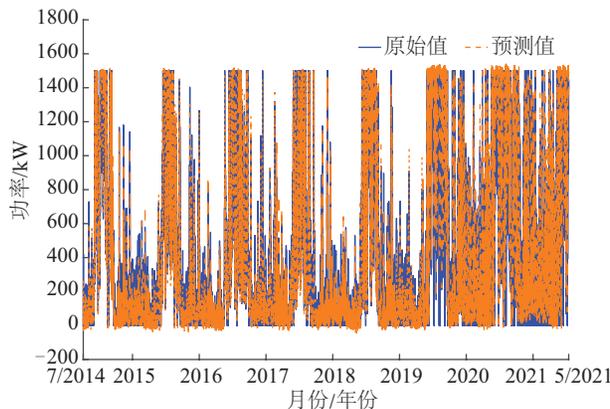
从图 7 和图 8 可看出, 各模型的 MAE 和 RMSE 值在 2014—2021 年间的变化趋势有所不同。SVM 模型的 MAE 在多个年份呈现较大波动, 尤其在 2016—2017 年间表现较差, 表明其在某些年份的适应性较差。RNN 模型的 MAE 在 2014—2016 年间有所下降, 但随后出现波动, 特别是在 2017—2018 年, 预测精度受到限制。相比之下, LSTM 模型表现较为稳定, 在 2014—2018 年期间, 其 MAE 较低, 尽管在 2020—2021 年间略有上升。GRU 和 TCN 模型的 MAE 较低且稳定, 显示出较强的鲁棒性。总体而言, PBY-Trans 模型始终表现出优异的预测能力, MAE 在所有年份中均明显低于其他对比模型, 展示了其在超短期风电功率预测中的卓越性能。对于图 8 中的 RMSE 数据, 其变化趋势与 MAE 类似, PBY-Trans 模型在所有年份中的 RMSE 表现均为最低。

为进一步观察 7 种不同方法的预测性能, 针对 2014 年 7 月—2021 年 5 月的数据进行模型测试, 图 9 给出了 SVM、RNN、Informer、LSTM、GRU、TCN 和 PBY-Trans 7 种模型的预测结果对比。

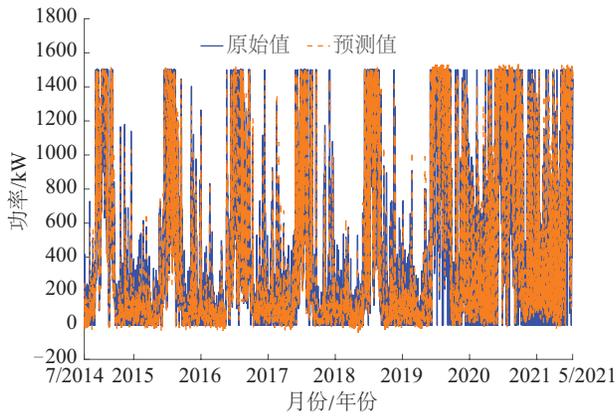
通过图 9 可看出, 各模型在超短期风电功率预测中表现出不同的趋势和波动。SVM 的表现在初期相对较好, 但随着时间的推移, 预测结果偏离真实值的程度逐渐增大; RNN 模型的预测结果显示与真实值存在较大偏差的情况, 尤其是在风电功率波动较大的时刻; Informer 模型后期同样呈现明显



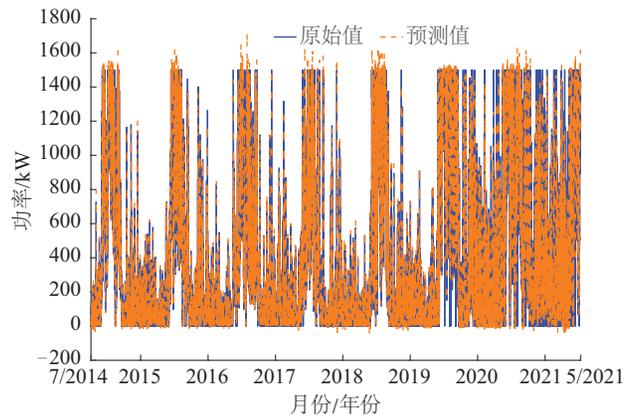
a. SVM模型



b. RNN模型



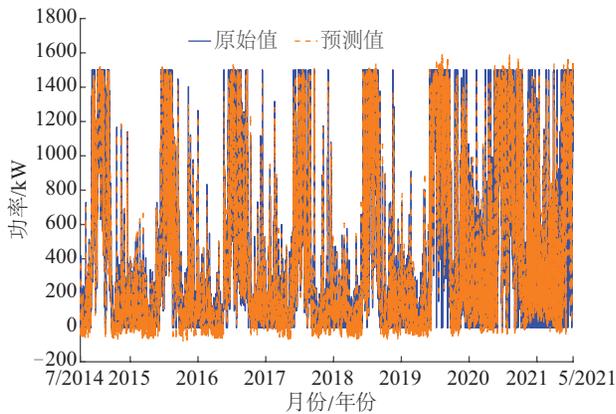
c. Informer 模型



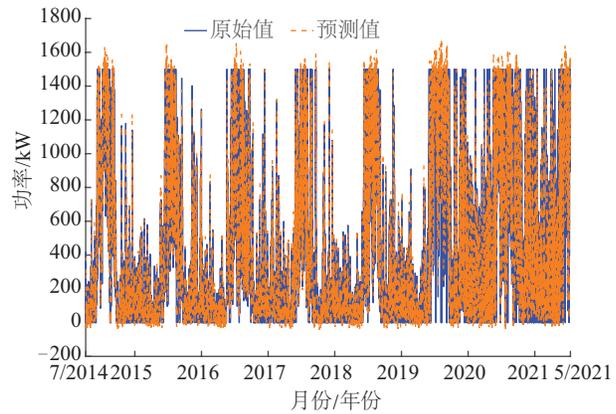
g. PBY-Trans 模型

图9 不同模型方法的预测性能结果

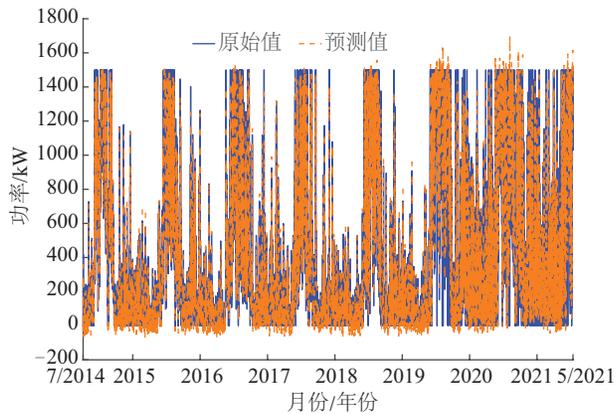
Fig. 9 Prediction results of different methods



d. LSTM 模型



e. GRU 模型



f. TCN 模型

增长趋势;LSTM 模型在预测风电功率方面表现出更高的准确性和稳定性,其预测结果与真实值之间的差异相对较小;GRU 模型相较于 RNN 模型表现更为优异,其预测结果与真实值更为接近,但在某些时刻仍存在预测值偏离真实值较大的情况;TCN 的预测结果表现得更加平稳;相比之下,PBY-Trans 模型的预测结果与真实值更为接近,呈现出较小的偏差,特别是高波动性的风速数据,PBY-Trans 模型展现出更强的鲁棒性和适应性,能够更加准确地捕捉序列中的全局依赖关系,具有更好的预测能力。

为更直观地验证本文提出的模型在复杂局部信息提取方面的预测性能,选取覆盖典型季节波动周期的 2016 年 9 月 10 日—13 日共 96 个连续时间步的预测样本进行测试,图 10 给出了 PBY-Trans 与其他模型在典型季节波动周期内的预测性能对比结果。

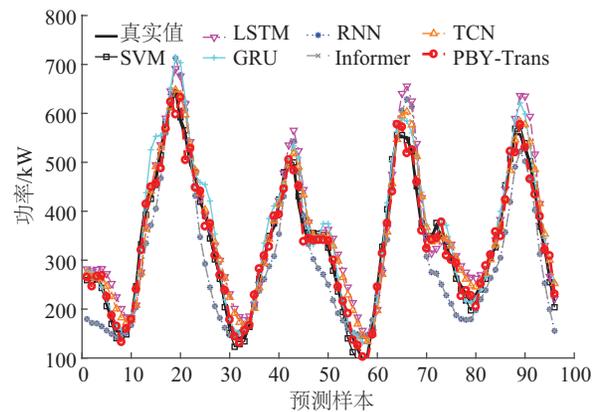


图10 典型季节波动周期的预测结果

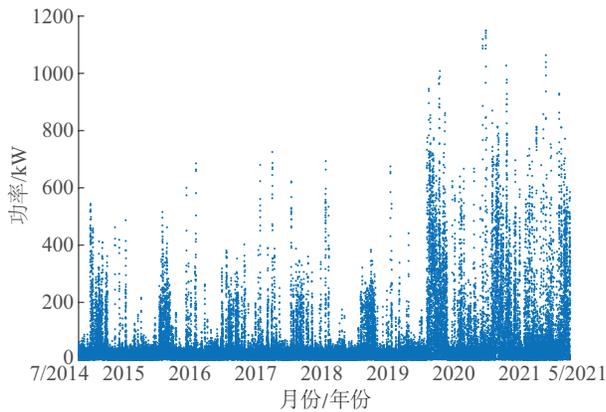
Fig. 10 Prediction results of typical seasonal fluctuation cycle

通过图 10 密集采样区间的预测结果对比可知,SVM 在局部表现出分段线性拟合特性;RNN、LSTM、GRU 在极值区间出现显著滞后与平滑效应;Informer 因全局注意力机制在长期依赖建模上的优势,部分平稳区间预测精度较高,但对

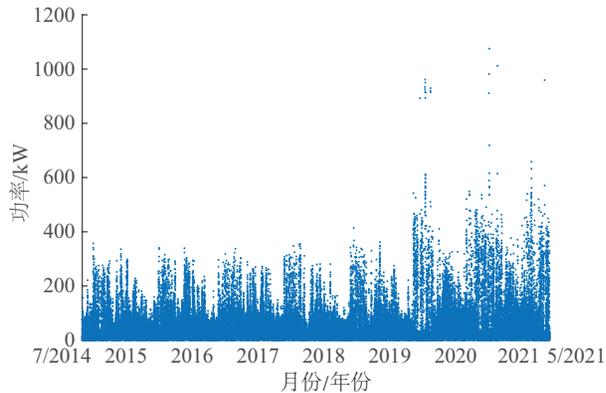
突变相位仍存在明显偏差;TCN 因空洞卷积的时序覆盖能力,对周期性波动的捕捉能力较优,但 TCN 与 SVM 模型二者均难以平衡长期趋势与短期扰动的关系。本文所提 PBY-Trans 模型对  $t=55(558.22)$ 、 $t=85(577.12)$  等关键峰值的预测误差较基准模型降低约 12.7%;在突变相位中,预测曲线与真实值的均方误差(MSE)较 Informer 降低 19.3%,在全周期范围内未出现显著累积误差,其拟合优度( $R^2$ )达 0.963。

为更加准确地观察不同方法的预测误差结果,图 11 给出了 SVM、RNN、Informer、LSTM、GRU、TCN 和 PBY-Trans 模型预测方法获得的误差散点图。

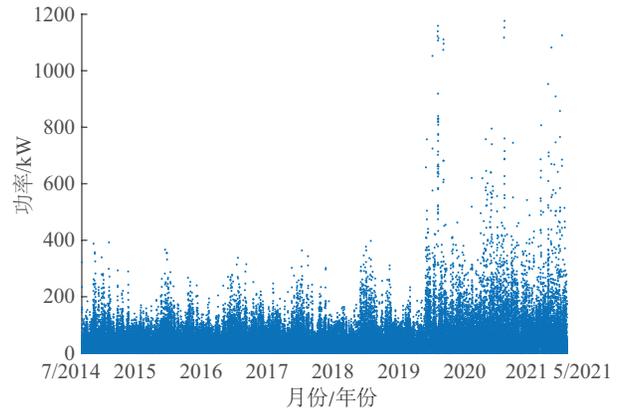
从图 11 可清晰地观察到各模型的预测误差情况,本文提出的 PBY-Trans 模型相对于 SVM、RNN、Informer、LSTM、GRU 和 TCN 模型,展现出更高的预测精度和可靠性,其预测误差显著减小,使得模型能够更加准确地拟合真实数据。此外,通过时间段分割的 Transformer 模型与贝叶斯参数优化方法相结合,使得突变值处的预测最大误差显著降低。这一结果表明了时间序列分割策略的有效性,归因于该模型在处理时间序列数据时能够更好地捕捉到变化的趋势和特征,更加适应超短期风电功率预测过程中突变风电数据的预测场景,而贝叶斯参数优化方法的引入则进一步提高了模型的性能,在选择最佳超参数的同时减小了模型过拟合的风险。



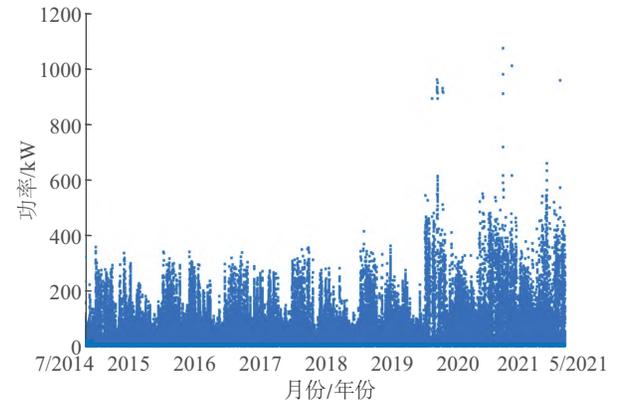
a. SVM 模型



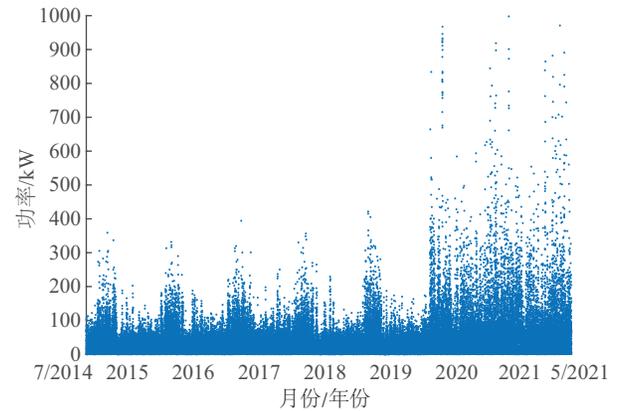
b. RNN 模型



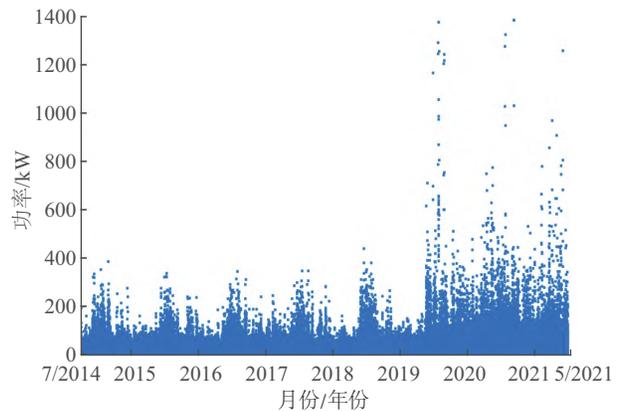
c. Informer 模型



d. LSTM 模型



e. GRU 模型



f. TCN 模型

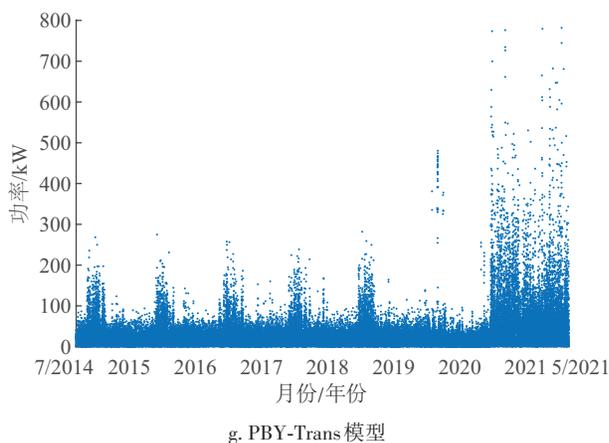


图 11 不同模型方法的预测误差结果

Fig. 11 Prediction error results among different methods

## 5 结 论

针对风电场风电功率具有高变异性 and 长距离依赖关系等复杂特征,对其进行超短期准确预测十分困难,本文提出一种结合时间段分割策略和贝叶斯参数优化的 Transformer 模型(PBY-Trans)。实验结果表明,采用时间段分割策略能够显著提升 Transformer 模型对风电功率时间序列复杂特征的捕捉能力,提高模型的泛化能力。同时,贝叶斯算法可高效优化模型参数,提高模型的预测效率和精度,进而更好地适应实际数据的复杂特征,以提升预测性能。与其他方法相比,本文方法在超短期风电功率预测方面取得了更高的  $R^2$  值,表现出优越的预测性能,不仅能够提高预测精度,还可降低突变处的误差,使预测结果更加稳定可靠。

### [参考文献]

- [1] 钟雅珊,付聪,钱峰,等.考虑广义储能和条件风险价值的综合能源系统经济调度[J].电力系统保护与控制,2022,50(9):54-63.  
ZHONG Y S, FU C, QIAN F, et al. Economic dispatch model of an integrated energy system considering generalized energy storage and conditional value at risk [J]. Power system protection and control, 2022, 50(9): 54-63.
- [2] 陈海鹏,李赫,阚天洋,等.考虑风电时序特性的深度小波-时序卷积网络超短期风功率预测[J].电网技术,2023,47(4):1653-1665.  
CHEN H P, LI H, KAN T Y, et al. DWT-DTCNA ultra-short-term wind power prediction considering wind power timing characteristics[J]. Power system technology, 2023, 47(4): 1653-1665.
- [3] JUNG J, BROADWATER R P. Current status and future advances for wind speed and power forecasting[J]. Renewable and sustainable energy reviews, 2014, 31: 762-777.
- [4] 向阳,刘亚娟,孙志伟,等.基于帝王蝶算法的CNN-GRU-LightGBM模型短期风电功率预测[J].太阳能学报,2025,46(1):105-114.  
XIANG Y, LIU Y J, SUN Z W, et al. Short term wind power prediction using CNN-GRU-LightGBM model based on emperor butterfly algorithm [J]. Acta energiae solaris sinica, 2025, 46(1): 105-114.
- [5] HAN Q K, MENG F M, HU T, et al. Non-parametric hybrid models for wind speed forecasting [J]. Energy conversion and management, 2017, 148: 554-568.
- [6] CORTES C, VAPNIK V. Support-vector networks [J]. Machine learning, 1995, 20(3): 273-297.
- [7] 刘擎龙,张宏立,王聪,等.基于序列到序列和注意力机制的超短期风速预测[J].太阳能学报,2021,42(9):286-294.  
LIU B L, ZHANG H L, WANG C, et al. Ultra-short-term wind speed prediction based on sequence-to-sequence and attention mechanism [J]. Acta energiae solaris sinica, 2021, 42(9): 286-294.
- [8] 王颀,刘兴杰,梁英,等.一种基于MGWO-Informer的超短期风电功率预测方法[J].太阳能学报,2024,45(11):477-485.  
WANG J, LIU X J, LIANG Y, et al. An ultra-short-term wind power prediction method based on MGWO-Informer [J]. Acta energiae solaris sinica, 2024, 45(11): 477-485.
- [9] LIPTON Z C, BERKOWITZ J, ELKAN C. A critical review of recurrent neural networks for sequence learning [EB/OL]. 2015: arXiv preprint arXiv: 1506.00019. <https://arxiv.org/abs/1506.00019>.
- [10] LIU H, MI X W, LI Y F. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM [J]. Energy conversion and management, 2018, 159: 54-64.
- [11] NIU Z W, YU Z Y, TANG W H, et al. Wind power forecasting using attention-based gated recurrent unit network[J]. Energy, 2020, 196: 117081.
- [12] HU H L, WANG L, LYU S X. Forecasting energy consumption and wind power generation using deep echo state network[J]. Renewable energy, 2020, 154: 598-613.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Advances in Neural Information Processing Systems. Montreal, Canada, 2014, 27: 2672-2680.
- [14] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. 2018: arXiv: 1803.01271. <https://arxiv.org/abs/1803.01271>.

- [15] LIM B, ARIK S Ö, LOEFF N, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting[J]. *International journal of forecasting*, 2021, 37(4): 1748-1764.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*Advances in Neural Information Processing Systems*. Longbeach, America, 2017, 30: 5998-6008.
- [17] KALYAN K S, RAJASEKHARAN A, SANGEETHA S. AMMUS: a survey of transformer-based pretrained models in natural language processing [EB/OL]. 2021: arXiv: 2108.05542. <https://arxiv.org/abs/2108.05542>.
- [18] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey [J]. *ACM computing surveys*, 2022, 54(10s): 1-41.
- [19] KARITA S, CHEN N X, HAYASHI T, et al. A comparative study on transformer vs RNN in speech applications[C]//2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). SG, Singapore, 2019: 449-456.
- [20] WEN Q S, ZHOU T, ZHANG C L, et al. Transformers in time series: a survey[EB/OL]. 2022: arXiv: 2202.07125. <https://arxiv.org/abs/2202.07125>.
- [21] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[J]. *Proceedings of the AAAI conference on artificial intelligence*, 2021, 35(12): 11106-11115.
- [22] WU H X, XU J H, WANG J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting [C]//*Advances in Neural Information Processing Systems*, 2021, 34: 22419-22430.
- [23] ZHOU T, MA Z Q, WEN Q S, et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting [EB/OL]. 2022: arXiv preprint arXiv: 2201.12740. <https://arxiv.org/abs/2201.12740>.
- [24] ZENG A L, CHEN M X, ZHANG L, et al. Are transformers effective for time series forecasting? [J]. *Proceedings of the AAAI conference on artificial intelligence*, 2023, 37(9): 11121-11128.
- [25] BASHIR T, WANG H F, TAHIR M, et al. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models [J]. *Renewable energy*, 2025, 239: 122055.
- [26] WANG D, XU M, ZHU G M, et al. Enhancing wind power forecasting accuracy through LSTM with adaptive wind speed calibration (C-LSTM) [J]. *Scientific reports*, 2025, 15: 5352.
- [27] YU Y X, HAN X S, YANG M, et al. Probabilistic prediction of regional wind power based on spatiotemporal quantile regression [C]//2019 IEEE Industry Applications Society Annual Meeting. Baltimore, MD, USA, 2019: 1-16.
- [28] ZHAO H S, ZHANG Y, LIU S, et al. PSANet: point-wise spatial attention network for scene parsing [C]//*Computer Vision-ECCV 2018*. Munich, Germany, 2018: 270-286.
- [29] NIE Y Q, NGUYEN N H, SINHTONG P, et al. A time series is worth 64 words: long-term forecasting with transformers [EB/OL]. 2022: arXiv: 2211.14730. <https://arxiv.org/abs/2211.14730>.
- [30] SNOEK J, LAROCHELLE H, ADAMS R P. Practical Bayesian optimization of machine learning algorithms [C]//*Neural Information Processing Systems*. Nevada, USA, 2012
- [31] 王健峰, 张磊, 陈国兴, 等. 基于改进的网格搜索法的 SVM 参数优化[J]. *应用科技*, 2012, 39(3): 28-31.
- WANG J F, ZHANG L, CHEN G X, et al. A parameter optimization method for an SVM based on improved grid search algorithm [J]. *Applied science and technology*, 2012, 39(3): 28-31.
- [32] BERGSTRA J, BENGIO Y. Random search for hyperparameter optimization [J]. *The journal of machine learning research*, 2012, 13: 281-305.
- [33] 陈峰, 余轶, 徐敬友, 等. 基于 Bayes-LSTM 网络的风电出力预测方法[J]. *电力系统保护与控制*, 2023, 51(6): 170-178.
- CHEN F, YU Y, XU J Y, et al. Prediction method of wind power output based on a Bayes-LSTM network [J]. *Power system protection and control*, 2023, 51(6): 170-178.
- [34] 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 X Y, PU X Y, LI J F, 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.
- [35] 季廷炜, 莫邵昌, 谢芳芳, 等. 基于高斯过程回归的机翼/短舱一体化气动优化[J]. *浙江大学学报(工学版)*, 2023, 57(3): 632-642.
- JI T W, MO S C, XIE F F, et al. Integrated aerodynamic optimization of wing/nacelle based on Gaussian process regression [J]. *Journal of Zhejiang University (engineering science)*, 2023, 57(3): 632-642.

## ULTRA-SHORT-TERM WIND POWER PREDICTION METHOD BASED ON BAYESIAN PARAMETER-OPTIMIZED TRANSFORMER TIME SEGMENTATION MODEL

Jiang Shanhe, Xu Xiaoyan, Tu Liang, Chen Wensheng

(School of Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, Anqing 246011, China)

**Abstract:** Aiming at the complex nonlinear relationship and the difficulty in capturing long-range dependencies of wind farm data, a new ultra-short-term wind power prediction method (PBY-Trans) based on a Bayesian parameter-optimized Transformer time segmenting model is proposed by introducing a time segmentation strategy. This method employs the time segmentation technique to divide the wind farm data into subsequences, which are then utilized as the inputs to the Transformer model encoder to better adapt to the nonlinear characteristics of the time series. Furthermore, a Bayesian algorithm is employed to search for the optimal configuration of the Transformer model parameters, thereby enhancing the model performance and improving the prediction accuracy. The prediction performance of the proposed method is compared and verified using the data set of a wind farm in Bengaluru. Compared with the SVM, RNN, Informer, LSTM, GRU and TCN models, the mean absolute error(MAE) metric of the proposed PBY-Trans method achieved a decrease of 33.56%, 59.75%, 47.27%, 32.34%, 40.46% and 27.71%, and the root mean square error(RMSE) was also reduced by 68.99%, 37.05%, 27.60%, 14.43%, 16.42% and 12.29%, respectively. The results indicate that the proposed PBY-Trans prediction model can further enhance prediction accuracy and the robustness can be effectively enhanced.

**Keywords:** wind power; prediction model; Bayesian optimization; Transformer model; time segmentation