

# 考虑多风电机组关联特性的超短期功率预测方法

朱童<sup>1</sup>,王彦泮<sup>1</sup>,叶希<sup>1</sup>,黄格超<sup>1</sup>,李甘<sup>1</sup>,朱琳俐<sup>2</sup>,张巍<sup>2</sup>,王渝红<sup>2</sup>

(1. 国网四川省电力公司,四川成都 610041;2. 四川大学电气工程学院,四川成都 610065)

**摘要:**由于邻近多风电机组间存在复杂的关联关系,深度挖掘多风电机组的空间特征有利于提高风电功率的预测精度。因此,提出了一种考虑多风电机组关联特性的超短期功率预测方法。首先,基于能量谷优化算法对变分模态分解的关键参数进行优化,将原始风电功率数据分解为多个利于预测的模式分量;随后,在双向门控循环单元时序预测网络中引入时序注意力机制,充分提取多风电机组间的复杂联系,从时空角度对各模式分量进行精准预测;最后,对各模式分量预测值进行重构得到多台风机的风电功率预测。实验结果表明,与其他预测模型相比,所提方法不仅能有效提高多风电机组的超短期功率预测精度,同时也能缩短训练时间。

**关键词:**能量谷优化算法;变分模态分解;双向门控循环单元;时序注意力机制;风电功率预测

**中图分类号:**TM 715 **文献标志码:**A **文章编号:**1003-6954(2025)01-0023-09

**DOI:**10.16527/j.issn.1003-6954.20250104

## Ultra-short-term Power Prediction Method Considering Correlation Characteristics of Multiple Wind Turbines

ZHU Tong<sup>1</sup>, WANG Yanfeng<sup>1</sup>, YE Xi<sup>1</sup>, HUANG Gechao<sup>1</sup>, LI Gan<sup>1</sup>, ZHU Linli<sup>2</sup>,  
ZHANG Wei<sup>2</sup>, WANG Yuhong<sup>2</sup>

(1. State Grid Sichuan Electric Power Company, Chengdu 610041, Sichuan, China;2. College of Electrical Engineering, Sichuan University, Chengdu 610065, Sichuan, China)

**Abstract:** Due to the complex correlations among neighboring wind turbines, deeply exploring the spatial features of multiple wind turbines is beneficial for improving wind power prediction accuracy. Therefore, an ultra-short-term power prediction method considering the correlation characteristics of multiple wind turbines is proposed. Firstly, the energy valley optimization (EVO) algorithm is employed to optimize the key parameters of variational mode decomposition (VMD), which decomposes the original wind power data into multiple mode components that are more conducive to prediction. And then, a temporal attention mechanism is introduced into the bidirectional gated recurrent unit (BiGRU) sequence prediction network to fully extract the complex relationships among multiple wind turbines and accurately predict each mode component from a spatiotemporal perspective. Finally, the predicted values of each mode component are reconstructed to obtain the wind power prediction for multiple turbines. Experimental results show that, compared to other prediction models, the proposed method not only effectively can improve the ultra-short-term power prediction accuracy for multiple wind turbines, but also can reduce the training time.

**Key words:** energy valley optimization algorithm; variational mode decomposition; bidirectional gated recurrent unit; temporal attention mechanism; wind power prediction

## 0 引言

在全球追求碳达峰和碳中和的“双碳”目标背景下,风能因其清洁和无污染的特性在可再生资源

基金项目:国家电网有限公司科技项目(52199723001G)

中占据了重要地位<sup>[1]</sup>。但风电因其间歇性和随机性特征,其功率难以准确预测,这构成了对风电高效利用以及电力系统稳定运行的挑战<sup>[2]</sup>。所以,提高风电功率预测的准确性不仅可以增加风电资源的利用效率,还有助于维护电网的稳定并推动全球能源系统向低碳化转型<sup>[3]</sup>。

当前的风电功率预测方法主要有三类:物理驱动法<sup>[4]</sup>、统计分析法<sup>[5]</sup>和组合预测法<sup>[6]</sup>。物理驱动方法利用数字天气预报和风电场数据预测风电功率,此方法依赖于收集大量的地理和历史气象数据<sup>[7]</sup>,建模复杂、成本高且计算也具有一定难度<sup>[8]</sup>。传统的统计分析法,如自回归积分移动平均法<sup>[9]</sup>和贝叶斯回归法<sup>[10]</sup>,则利用风电场历史数据来提取线性特征。然而,这些线性模型难以处理非线性和非平稳的数据变化,限制了其预测效果。随着信息技术和人工智能的进步,机器学习技术如人工神经网络<sup>[11]</sup>和支持向量机<sup>[12]</sup>,在风电预测领域展现出显著的优势,它们能够有效地处理复杂数据模式和提高预测的准确性。尽管机器学习技术在多方面超越了传统的统计方法,但在处理含有不规则空缺和噪声的实际风电功率数据时,这些技术在数据量不充分的情况下预测准确性仍受限<sup>[13]</sup>。为了增强预测的精度,研究者们提出了组合预测技术,主要包括两种策略:一种策略是结合多种算法,并通过分配不同的权重来强化关键信息的作用,如将注意力机制集成到双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)或双向门控循环单元(bi-directional gated recurrent unit, BiGRU)的隐藏状态以提高预测准确性,但这种方法可能存在泛化性能的问题;另一种策略是对风电功率序列进行分解后再进行预测,其中常用的分解技术包括小波变换、经验模态分解和变分模态分解等<sup>[14]</sup>。

文献[15]引入经验模态分解提取固有模式函数,实现降低神经网络的预测难度,但存在模态混叠问题<sup>[16]</sup>。文献[17]将原始功率信号通过变分模态分解(variational mode decomposition, VMD)算法分解为多个子序列,然后使用蝴蝶优化算法优化最小二乘支持向量机,对各子序列进行预测,但该模型存在关键参数难以确定的问题。因此,确定 VMD 算法中最优的固有模式函数(intrinsic mode function, IMF)分解数量  $K$  和惩罚因子  $\delta$  的参数组合是关键。文献[18]通过中心频率观察法选择  $K$  值,而  $\delta$  值则仍需人工设置。文献[19]通过粒子群算法优化 VMD 参数,并将其应用于磁瓦声振信号的分析,但粒子群算法可能会陷入局部最优问题。文献[20]中的能量谷优化算法(energy valley optimizer, EVO)受到物理学中粒子寻求最低能量状态的现象启发,提供了一种解决复杂系统优化问题的新策略。该算法能

够避免传统优化技术常见的局部最优陷阱,使得在更广阔的参数空间中找到到更有效的解决方案成为可能。文献[21]通过卷积神经网络(convolutional neural networks, CNN)进行特征提取,并使用 LSTM 处理多变量时间序列建模,相较于单一网络模型精确度有所提升,但是 LSTM 结构相对复杂,需要较长的训练时间。文献[22]采用时间卷积网络(temporal convolutional network, TCN)提取时间序列特征,并结合门控循环单元进行非线性动态建模,显著提升了预测精度。然而,TCN 在处理长时间依赖时存在特征稀疏问题,且模型结构需要精细调参,增加了计算复杂度。文献[23]使用注意力机制(attention mechanism, AM)以提高对关键信息的影响力,为门控循环单元神经网络的隐藏状态分配不同的权重。文献[24]采用时序注意力机制增强了模型对时间序列相关性的依赖,但传统注意力机制仅为单台风机的输出功率在时间序列中分配权重,未能捕捉到多台风机之间的关系。文献[25]介绍了时间模式注意力(temporal pattern attention, TPA)机制,它通过对隐藏状态进行卷积计算,可以探索多风机之间的复杂关联。此外,考虑到每台风机的特性不同,TPA 机制还能在不同时间步骤为它们分配差异化的权重<sup>[26]</sup>。

综上所述,当前风电功率预测方法尽管取得了一定进展,但仍存在未解决的问题。物理驱动方法依赖大量气象和地理数据,建模复杂且成本高;传统统计方法难以处理非线性和非平稳数据,预测效果受限;机器学习和深度学习技术虽然处理复杂数据模式表现优异,但在面对含有不规则空缺和噪声的数据时预测准确性不足。此外,组合预测技术虽然提高了预测精度,但对多风电机组关联特性的利用仍不足,且关键参数优化复杂,容易陷入局部最优解。下面提出一种风电功率预测方法:通过 EVO 优化 VMD 参数,增强对复杂数据的处理能力,避免了局部最优陷阱;利用 BiGRU 和 TPA 机制有效捕捉多风电机组之间的 IMF 时序关联,充分利用其关联特性;通过优化关键参数组合,提高了预测精度和可靠性。通过与其他 5 种模型的预测性能对比分析,验证了所提方法在实际应用中的优越性和有效性。

## 1 预测模型原理

### 1.1 VMD 原理

VMD 是一种创新的信号处理方法,专为解决复

杂信号的非线性和非平稳性问题而设计。这种方法基于一组自适应的维纳滤波器,通过构建变分问题并求最优解,从而实现信号的高效分解。与传统的经验模态分解和小波分解相比,VMD 在信号重构和抗噪性能方面表现更佳,能够将原信号  $f(t)$  精确分解为  $K$  个具有不同中心频率的稳定子信号  $u_k(t)$ ,每个子信号均表现出良好的相对正交性,无模态混叠问题,如式(1)所示。

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( F(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \sum_{k=1}^K u_k = f(t) \end{cases} \quad (1)$$

式中: $u_k$ 、 $\omega_k$  分别为第  $k$  个模态分量和其中心频率; $F(t)$  表示狄拉克分布;\* 为卷积运算符。

通过引入二次惩罚因子  $\delta$  和拉格朗日乘子  $\lambda$ , 可将式(1)初步的约束问题, 转换成一个更易于处理的无约束问题, 如式(2)所示。

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \\ & \delta \sum_{k=1}^K \left\| \partial_t \left[ \left( F(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \quad (2)$$

式中: $L(\cdot)$  为拉格朗日函数; $\langle \cdot, \cdot \rangle$  为内积运算符号。

为了获得最优解,使用交替方向乘子法来更新各个模态分量  $u_k$  及其对应的中心频率  $\omega_k$ , 以及拉格朗日乘子  $\lambda$ 。更新公式如式(3)所示。

$$\begin{cases} \hat{u}_{k,n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_{i,n}(\omega) + \hat{\lambda}_n(\omega)/2}{1 + 2\delta(\omega - \omega_{k,n})^2} \\ \omega_{k,n+1} = \frac{\int_0^\infty \omega |\hat{u}_{k,n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_{k,n+1}(\omega)|^2 d\omega} \\ \hat{\lambda}_{n+1}(\omega) = \hat{\lambda}_n(\omega) + \tau \left( \hat{f}(\omega) - \sum_{k=1}^K \hat{u}_{k,n+1}(\omega) \right) \end{cases} \quad (3)$$

式中: $n$  为迭代次数; $\tau$  为更新步长;上标 $\hat{\cdot}$ 表示该量为对应信号的傅里叶变换。

## 1.2 EVO 优化算法

鉴于实际信号的复杂性,手动设定 VMD 方法中的分解数量  $K$  和惩罚因子  $\delta$  不仅困难,而且可能

造成结果的不稳定性。为了解决这一问题,采用了 EVO 算法来自动优化这些参数,从而提高分解效果的稳定性。

EVO 算法受物理粒子寻找最低能量状态的启发,为解决复杂系统优化问题提供了一种新颖方法。该算法通过模拟粒子在能量势场中的运动来探索解空间,过程如下:

1) 初始化过程。在这一阶段,候选解  $X$  被假设为具有不同稳定性水平的粒子,粒子的初始化位置由式(4)给出:

$$\begin{cases} X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_b \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,c} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,c} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{b,1} & x_{b,2} & \cdots & x_{b,c} & \cdots & x_{b,d} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,c} & \cdots & x_{m,d} \end{bmatrix} \begin{cases} b = 1, 2, \dots, m \\ c = 1, 2, \dots, d \end{cases} \\ x_{b,c} = x_{b,c \min} + \xi \cdot (x_{b,c \max} - x_{b,c \min}) \end{cases} \quad (4)$$

式中: $m$  为粒子的总数; $d$  为所考虑问题的维度; $x_{b,c}$  为第  $b$  个粒子在第  $c$  个维度上的位置值; $x_{b,c \max}$  和  $x_{b,c \min}$  为第  $b$  个粒子中第  $c$  个维度的上、下限; $\xi$  为  $[0, 1]$  范围内均匀分布的随机数; $X$  为粒子群(候选解集合)的初始位置; $X_b$  为种群中第  $b$  个粒子的当前位置向量。

2) 确定粒子的富集边界,用于考虑富中子粒子和贫中子粒子之间的差异。将每个粒子的中子富集水平计算出来,用于区分富中子和贫中子粒子,进而确定富集边界,计算公式为

$$E_B = \frac{\sum_{b=1}^m N_{Lb}}{m}, \quad b = 1, 2, \dots, m \quad (5)$$

式中: $N_{Lb}$  为第  $b$  个粒子的中子富集水平; $E_B$  为粒子种群中粒子的富集边界。

3) 根据目标函数的评估结果,确定粒子的稳定性等级,如式(6)所示。

$$S_{Lb} = \frac{N_{Lb} - B_S}{W_S - B_S}, \quad b = 1, 2, \dots, m \quad (6)$$

式中: $S_{Lb}$  为第  $b$  个粒子的稳定度; $B_S$  和  $W_S$  分别为种群中稳定性最好和最差的粒子,相当于迄今发现的目标函数值的最小值和最大值。基于目标函数的评估,可计算得到粒子的稳定程度。

如果粒子的中子富集水平高于富集边界  $E_B$ , 则需要通过  $\alpha$ 、 $\beta$  或  $\gamma$  衰变来降低能量。如果粒子的稳定性水平高于稳定性界限, 则认为发生了  $\alpha$  和  $\gamma$  衰变, 因为对于稳定性水平较高的粒子来说, 这两种衰变是可能发生的。 $\alpha$  衰变是一种放射性衰变过程, 其中原子核释放  $\alpha$  粒子, 使得原子核转变为更稳定的状态。在数学上, 可将其表述为 EVO 中的一种位置更新策略。该策略涉及生成两个随机整数作为  $[1, d]$  范围内的  $\alpha$  指标  $\alpha_I$ , 表示发射光线的数量, 以及  $[1, \alpha_I]$  范围内的  $\alpha$  指标  $\alpha_{II}$ , 它定义了要发射的  $\alpha$  射线。这些  $\alpha$  射线代表候选解中的决策变量, 在更新中, 它们被从当前粒子或具有最佳稳定性水平的候选粒子中移除并替换。在数学上表述如下:

$$\mathbf{X}_b^{\text{New1}} = \mathbf{X}_b[\mathbf{X}_{B_s}(x_{b,c})], \begin{cases} b = 1, 2, \dots, m \\ c = \alpha_{II} \end{cases} \quad (7)$$

式中:  $\mathbf{X}_b^{\text{New1}}$  为种群中新生成的通过  $\alpha$  衰变位置更新得到的粒子;  $\mathbf{X}_{B_s}$  为稳定性水平最好的粒子的位置向量。

$\gamma$  衰变过程与  $\alpha$  衰变类似, 此处不再赘述。其数学表述如下:

$$\mathbf{X}_b^{\text{New2}} = \mathbf{X}_b[\mathbf{X}_{N_g}(x_{b,c})], \begin{cases} b = 1, 2, \dots, m \\ c = \gamma_{II} \end{cases} \quad (8)$$

式中:  $\mathbf{X}_b^{\text{New2}}$  为种群中新生成的通过  $\gamma$  衰变位置更新得到的粒子;  $\mathbf{X}_{N_g}$  为第  $b$  个粒子周围邻近粒子的位置向量。

如果粒子的稳定性水平低于稳定性界限, 则认为会发生  $\beta$  衰变, 因为这种类型的衰变发生在稳定性水平较低的更不稳定的粒子中, 表达式为:

$$\mathbf{X}_{CP} = \frac{\sum_{b=1}^m \mathbf{X}_b}{m}, \quad b = 1, 2, \dots, m \quad (9)$$

$$\mathbf{X}_b^{\text{New1}} = \mathbf{X}_b + \frac{(r_1 \times \mathbf{X}_{B_s} - r_2 \times \mathbf{X}_{CP})}{S_{Lb}}, \quad b = 1, 2, \dots, m \quad (10)$$

式中:  $\mathbf{X}_{CP}$  为粒子中心的位置向量;  $r_1$  和  $r_2$  为  $[0, 1]$  范围内的两个随机数, 决定粒子的移动量。

为了提高算法的开发和探索水平, 对采用  $\beta$  衰变的粒子进行了另一种位置更新过程。在此过程中, 向具有最佳稳定性水平的粒子 ( $\mathbf{X}_{B_s}$ ) 或候选粒子及其邻近的粒子 ( $\mathbf{X}_{N_g}$ ) 或候选粒子进行受控移动, 而粒子的稳定性水平不会影响移动过程。数学表达式为

$$\mathbf{X}_b^{\text{New2}} = \mathbf{X}_b + (r_3 \times \mathbf{X}_{B_s} - r_4 \times \mathbf{X}_{N_g}), \quad b = 1, 2, \dots, m \quad (11)$$

式中,  $r_3$  和  $r_4$  为  $[0, 1]$  范围内的两个随机数, 决定粒子的移动量。

如果粒子的中子富集水平低于富集界限, 则假定粒子具有较小的中子数与质子数之比, 因此粒子倾向于经历电子捕获或正电子发射以向稳定带移动。在这方面, 确定搜索空间中的随机移动以考虑这些类型的运动, 如式 (12) 所示。

$$\mathbf{X}_b^{\text{New}} = \mathbf{X}_b + r, \quad b = 1, 2, \dots, m \quad (12)$$

式中:  $\mathbf{X}_b^{\text{New}}$  为种群中第  $b$  个粒子 (候选解) 即将到来的位置向量;  $r$  是  $[0, 1]$  范围内的随机数, 它决定粒子的移动量。

在 EVO 主循环结束时, 如果粒子的富集水平高于富集边界, 则每个粒子只有两个新生成的位置向量  $\mathbf{X}_b^{\text{New1}}$  和  $\mathbf{X}_b^{\text{New2}}$ , 而对于具有较低的富集水平, 仅生成  $\mathbf{X}_b^{\text{New}}$  作为新的位置向量。这些新生成的向量与当前种群合并, 形成新的候选解集, 并由最佳粒子引导进入算法的下一轮搜索循环。对于超出预定上限和下限的决策变量确定边界违规标记, 同时可以利用目标函数评估的最大次数或迭代的最大次数作为终止准则。

最佳 VMD 参数  $K$  和  $\delta$  的求解过程如下:

- 1) 初始化 EVO 算法的参数以及种群。
- 2) 为每个粒子计算其中子富集水平  $N_{Lb}$ , 确定种群中粒子的富集边界  $E_B$ 。
- 3) 通过目标函数 (VMD 的残差范数) 计算每个粒子的稳定度, 找出稳定性最好 ( $B_s$ ) 和最差 ( $W_s$ ) 的粒子。
- 4) 根据粒子的中子富集水平和稳定度, 决定采用哪种衰变 ( $\alpha$ 、 $\beta$  或  $\gamma$ ) 来更新粒子的位置。将新位置向量与现有种群合并, 形成新一代种群。
- 5) 经过多次迭代, 从中选出具有最佳目标函数值的粒子, 该粒子的位置代表了最优的 VMD 参数  $K$  和  $\delta$ 。

### 1.3 BiGRU 原理

BiGRU 模型是一种循环神经网络, 它由两个独立的 GRU 单元组成, 一个按照时间序列正向处理数据, 另一个按照时间序列逆向处理数据。通过这种双向结构, BiGRU 模型能够同时捕捉序列数据的前向和后向信息, 从而更好地理解 and 预测序列中的模式。BiGRU 结构如图 1 所示。图 1 中,  $x_t$  为  $t$  时刻的



输入序列数据; $y_t$  为  $t$  时刻 GRU 单元输出的结果。

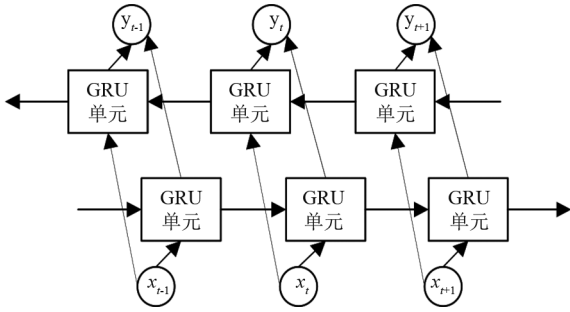


图 1 BiGRU 结构

BiGRU 通过输入正序列与其反序列到前向和后向 GRU 中,能够并行训练两个网络。前向 GRU 根据历史数据预测未来,而后向 GRU 则是依据未来数据来推断过去,最终的预测结果是综合这两个网络输出的结果。因为 BiGRU 能够有效处理对前后文信息都敏感的时间序列数据,所以可以结合风力发电功率数据使用 BiGRU 进行了双向预测。

#### 1.4 TPA 原理

注意力机制在自然语言处理、图像和语音识别等领域有着广泛的应用,近年来也被用于各种预测问题。传统的注意力机制主要考虑各时刻的权值分布,且该方法仅适用于每一时刻仅有一种情形。但是,考虑到区域内多台风机的功率预测,各阶段均含有多个不同的变量,且各变量间具有复杂的非线性关系,并且各变量具有自身的特性与周期性,难以将单个时间步长作为研究对象。为了解决这个问题,提出了 TPA 结构,该结构采用多个一维 CNN 滤波器,通过对 BiGRU 隐含的行向量进行特征抽取,从而实现多个变量在时间步长上的相关性学习。TPA 结构如图 2 所示。

原始时间序列经 BiGRU 处理后,得到的隐藏状态向量表示为  $y_{t-w}—y_t$ ,其中  $w$  为时间序列的长度,表示输入数据中包含的时间步数量。定义隐藏状态

矩阵  $Y = \{y_{t-w}, y_{t-w+1}, \dots, y_{t-1}\}$ ,其中每列向量代表了同一时间步下 BiGRU 内部门神经元参数的变化情况,每行向量则表示了单个变量在所有时间步下的状态。

图 2 中,不同颜色的矩阵表示了隐藏状态矩阵  $Y$  中的不同一维卷积核,对  $Y$  的  $r$  个特征进行一维卷积,可提取出可变信号模型的时间模式矩阵  $Y_{p,q}^C$ 。

$$Y_{p,q}^C = \sum_{l=1}^w Y_{p,(t-w-1+l)} * C_{q,T-w+l} \quad (13)$$

式中: $Y_{p,q}^C$  为特征一维积累后得到的时间模式矩阵  $Y^C$  中第  $p$  行、第  $q$  列的元素; $Y_{p,(t-w-1+l)}$  为隐藏状态矩阵  $Y$  中第  $p$  行、第  $t-w+l$  列的元素; $C_{q,T-w+l}$  为加权系数,对应时间窗口内第  $T-w+l$  个时间步的权重;符号  $*$  表示卷积运算。一维滤波器的卷积核有  $s$  个。

将以下函数用来计算  $Y_p^C$  与  $y_t$  相关性:

$$f(Y_p^C, y_t) = (Y_p^C)^T W_a y_t \quad (14)$$

$$\alpha_p = \sigma(f(Y_p^C, y_p)) \quad (15)$$

式中: $Y_p^C$  为  $Y^C$  的行向量; $W_a$  为  $r \times s$  的权重矩阵; $\alpha_p$  为注意力权重,表示  $Y_p^C$  和  $y_t$  的相关性; $\sigma(\cdot)$  为激活函数,用于对相关性进行归一化处理。利用  $\alpha_p$  与  $Y^C$  加权求和,得到注意力向量  $v_t$ 。

$$v_t = \sum_{p=1}^r \alpha_p Y_p^C \quad (16)$$

将注意力向量  $v_t$  与  $y_t$  按式(17)和式(18)处理后得到最终预测值。

$$y'_t = (W_y y_t + W_v v_t) \quad (17)$$

$$y_{t-1+\Delta} = W_y'(W_y y_t + W_v v_t) \quad (18)$$

式中: $y'_t$  为时间步  $t$  的最终预测值; $y_{t-1+\Delta}$  为预测未来时间步  $t+\Delta$  的结果; $W_p, W_y, W_y'$  分别为模型中对应的权重矩阵。

传统的注意力机制只能针对单个序列的时域特性,且没有考虑到时序间的相关性。隐藏状态矩阵具有多个时序序列之间的复杂联系,将 CNN 对隐藏

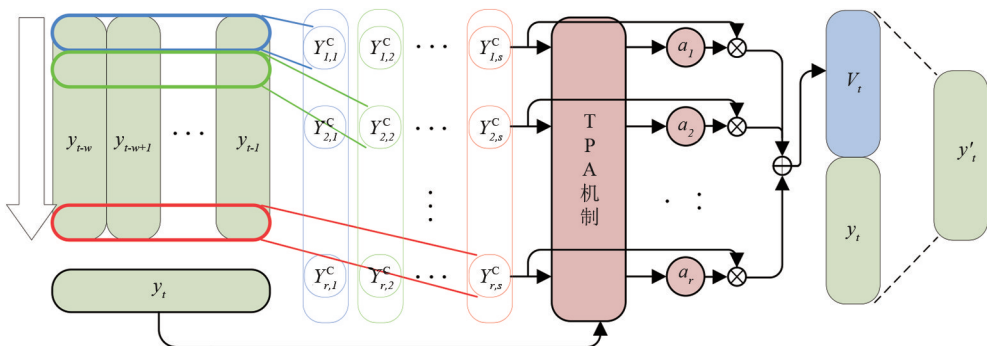


图 2 TPA 结构

状态矩阵行向量进行特征抽取,可以反映出时序序列的内在联系和各变量之间的复杂关系。由于  $v_i$  结合了时序信息,所以该模型可以在不同的时间步长中选取相关的信息。TPA 技术可以有效地解决多台风机组之间的耦合关系,在多台风机的超短期功率预测中具有明显的优越性。

## 2 多风电机组功率预测

针对目前风电超短期功率预测所存在的问题,提出了基于 TPA 机制的 EVO-VMD-BiGRU 预测模型。首先,通过 EVO 算法对 VMD 的关键参数  $K$  和  $\delta$  进行优化,以便利用 VMD 进行特征提取。BiGRU 模型凭借其双向信息流的特性,在时间序列分析领域展现出卓越的处理能力和鲁棒性。虽然传统注意力机制在分析单变量时间序列数据方面效果显著,但当面临具有多重特性和复杂相互作用的风力发电机组功率预测问题时,其性能显得不足。采用 TPA 机制有助于有效识别并利用风电机组之间的复杂动态关系,通过对各特性进行适当的加权处理,提高预测准确度。

### 2.1 预测模型概述

基于 TPA 机制的 EVO-VMD-BiGRU 多风电机组功率预测流程如图 3 所示。

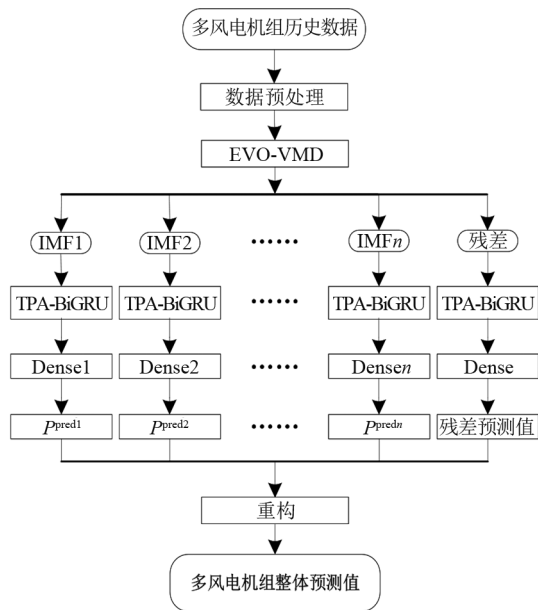


图 3 基于 TPA 机制的 EVO-VMD-BiGRU 多风电机组功率预测流程

为了加快神经网络训练的收敛速度,提高模型训练效率和预测准确性,同时避免数据中的量纲差

异导致的偏差,对时间序列长度为  $w$  的各风电机组功率按式(19)作如下归一化处理。

$$P'_o = \frac{P_o - P_{o\min}}{P_{o\max} - P_{o\min}} \quad (19)$$

式中: $P'_o$ 和  $P_o$  分别为第  $o$  台风电机组归一化功率输出和实际功率输出; $P_{o\max}$ 和  $P_{o\min}$  分别为第  $o$  台风电机组输出功率的最大值和最小值。

预处理后的数据首先经过 EVO-VMD 分解,从而获得各风机的 IMF 和残差,这些 IMF 随后被送入 TPA-BiGRU 网络进行分析。TPA 利用隐藏状态向量来生成注意力向量  $v_i$ ,将其与隐状态行向量进行线性映射相加,从而获得各风机各模态的预测值  $P^{\text{pred}}$  和残差预测值,最后经过重构得到各风机的预测值。

多台风机的第  $g$  个模态功率预测模型如图 4 所示。

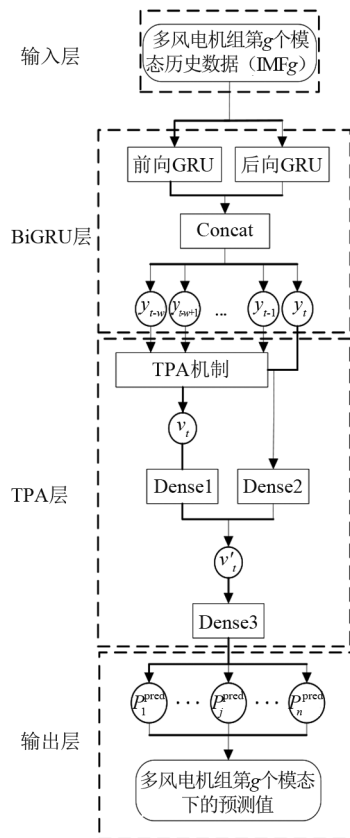


图 4 基于 TPA 机制的 EVO-VMD-BiGRU 多风电机组第  $g$  个模态功率预测模型

### 2.2 评估指标

为全面评估所提出预测模型的有效性,采用了均方根误 (root mean squared error, RMSE) 和平均绝对误差 (mean absolute error, MAE) 作为评价指标。

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{g=1}^N (P_g - P^{\text{pred}g})^2} \quad (20)$$

$$E_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |P_g - P^{\text{predg}}| \quad (21)$$

式中:  $N$  为预测序列的时间长度;  $P_g$  和  $P^{\text{predg}}$  分别为预测序列在第  $g$  个时刻的风电机组输出功率真实值和预测值。

### 3 算例分析

使用分散在某区域的 14 台风电机组的功率数据, 采样间隔为 15 min。数据集包括 4277 组训练样本和 1000 组测试样本, 每个样本输入不同风机各模态时间序列, 下一时间点的风机功率值作为输出, 时间序列长度设定为 24。所有实验均在 Intel (R) UHD Graphics 770 上进行。

#### 3.1 EVO-VMD 优化结果分析

##### 3.1.1 基于 EVO 算法的 VMD 参数优化

通过使用最小包络熵作为适应度函数, 并运用 EVO 算法优化, 得到 VMD 分解的最佳参数  $[K, \delta]$ 。种群数量设置为 20, 最大迭代 30 次, 分解层数  $K$  的范围是 2 至 10, 惩罚因子  $\delta$  的范围是 1000 至 3000。EVO 算法优化的 VMD 迭代过程如图 5 所示, 全局最优适应度值为 8.370 8, 且最优参数为  $K=6$ ,  $\delta=1241$ 。

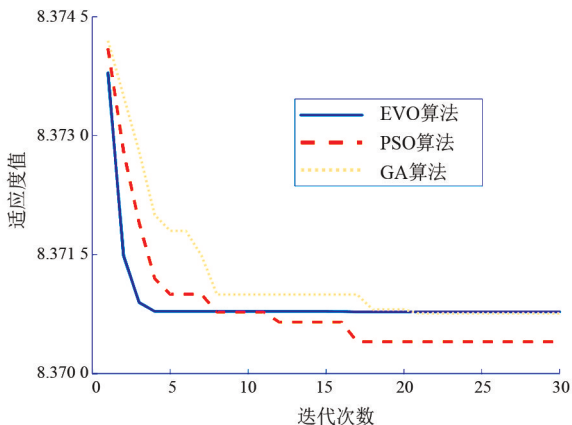


图 5 EVO、PSO 和 GA 算法的收敛曲线对比

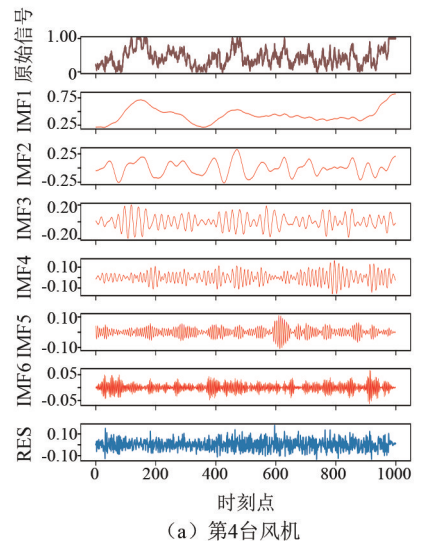
图 5 中的迭代曲线显示, EVO 算法在较少的迭代次数内 (约 5 次) 就达到了全局最优适应度值, 说明其具有快速收敛的特点。相比之下, 其他两种优化算法, 即粒子群优化算法 (particle swarm optimization, PSO) 和遗传算法 (genetic algorithm, GA), 在前期收敛速度较慢, 且适应度值波动较大。从图中可以看出, PSO 算法在约 10 次迭代后逐渐接近 EVO 算法的适应度值, 但收敛速度明显较慢。

而 GA 算法初期存在较大的波动性, 收敛速度最慢, 最终适应度值略高于 EVO 和 PSO。

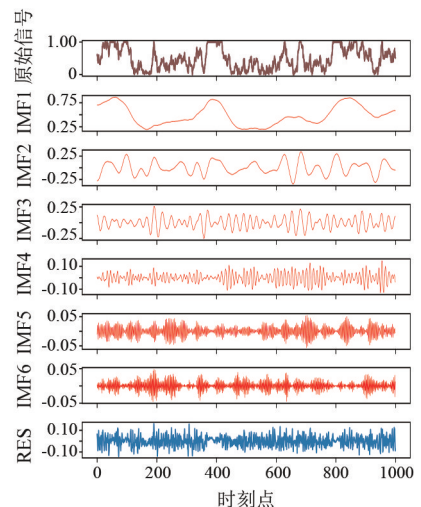
总体来看, EVO 算法通过模拟粒子在能量谷中的运动, 有效避免了陷入局部最优解的问题, 曲线平滑且稳定, 表明其在搜索过程中保持了较好的全局搜索能力, 能够在更广阔参数空间中找到更优的解。此外, 曲线没有出现明显的振荡和回退, 说明 EVO 算法有效避免了传统优化算法中常见的局部最优陷阱问题, 保持了连续的优化进程。

##### 3.1.2 风电数据 VMD 结果分析

根据前一节优化的参数结果, 设定生成的 IMF 数量为 6, 惩罚参数  $\delta$  设置为 1241。使用这些参数对一段时间内风机的输出功率进行 VMD 分解。图 6 分别展示第 4 台和第 11 台风机的 VMD 分解结果。



(a) 第4台风机



(b) 第11台风机

图 6 风电功率 VMD 分解结果

在图 6 中, 初始序列展示了风电功率的原始信

号,接下来的序列则呈现了 VMD 分解所得的各个模态 (IMF1 至 IMF6) 以及残差值 (residual, RES)。IMF1 和 IMF2 呈现出较为缓慢的变化和较弱的波动,反映了风机在某段时间内的基本发电情况。IMF3 和 IMF4 的波动频率相对较低,但波动周期性明显,有助于揭示风电功率的周期性波动情况。而 IMF5、IMF6 以及 RES 的频率较高,波动更为剧烈,这在一定程度上展现了风电功率的随机波动特征。

每个 IMF 均突出了风力发电独有的特点,尤其是其低频模态波动更小,使得预测效果更好。由上述分析可知,整体预测的误差主要由后两个高频模态的预测误差所决定。而 VMD 可以克服直接使用原始风机功率预测造成的负面影响,降低模型的计算复杂性,进而提高 GRU 神经网络的预测精度。

### 3.2 风电机组功率预测结果分析

#### 3.2.1 VMD 子序列预测效果

将第 6 台风机的各模态训练 20 代,输入时间序列长度为 300。损失函数选用 RMSE 和 MAE。表 1 列出了从 IMF1 到 IMF6 的预测误差。各模态经训练后,用于预测 300 个时间点的值。各模态的预测值如图 7 所示,将其进行叠加后,即可得到完整的风机功率预测序列。

表 1 不同模态下的预测误差

模态	RMSE	MAE
IMF1	0.004 07	0.003 35
IMF2	0.012 18	0.010 33
IMF3	0.011 31	0.009 15
IMF4	0.006 30	0.005 04
IMF5	0.006 82	0.005 48
IMF6	0.010 58	0.008 52
RES	0.039 26	0.030 90

#### 3.2.2 风电机组整体功率预测效果分析

为进一步验证所提模型的性能,分别将 VMD-TPA-BiLSTM 模型、VMD-TCN-BiLSTM 模型、VMD-AM-BiGRU 模型、TPA-BiGRU 模型和 CNN-GRU 模型与所提模型作为对比,结果如图 8 所示。

表 2 详细描述了不同模型的预测误差以及训练模型所用时间。

从表 2 中可知,各种模型在风电功率预测中的表现存在显著差异。首先,VMD-TPA-BiGRU 模型表现最佳,具有最低的 RMSE 和 MAE,表明其预测

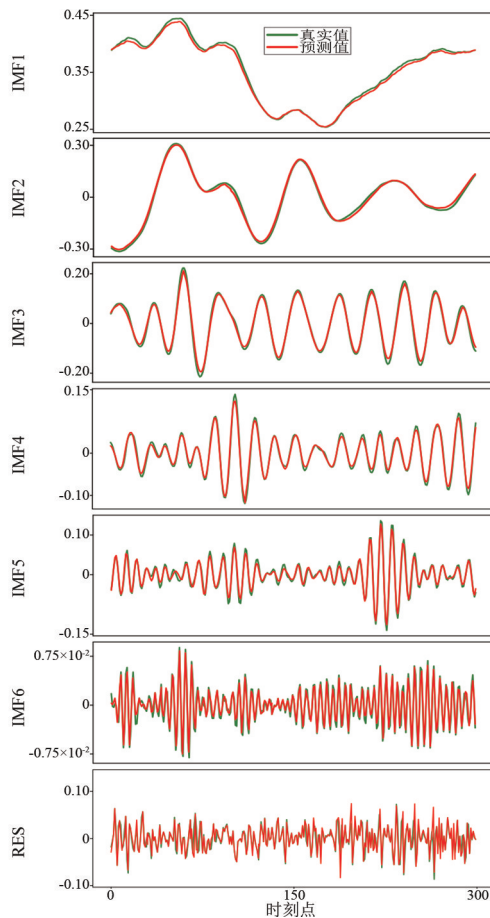


图 7 风机各模态分量预测

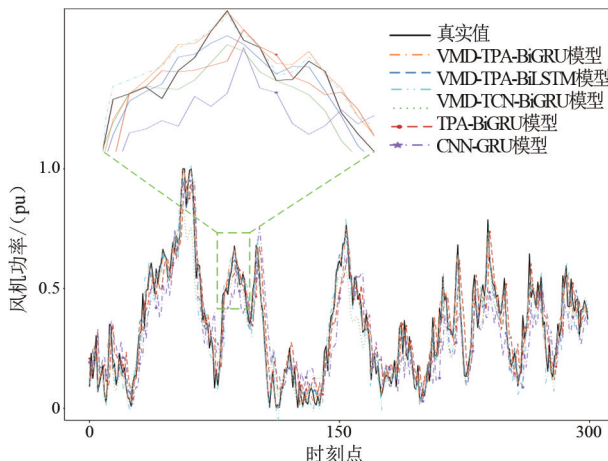


图 8 第 6 台风机功率预测结果

表 2 不同模型下的预测误差及训练用时

预测模型	RMSE	MAE	用时/s
VMD-TPA-BiGRU (所提)	0.041 31	0.033 06	0.481 40
VMD-TCN-BiGRU	0.063 95	0.047 96	0.463 72
VMD-TPA-BiLSTM	0.041 96	0.033 53	0.519 19
VMD-AM-BiGRU	0.096 91	0.077 28	0.355 11
TPA-BiGRU	0.078 10	0.056 54	0.380 59
CNN-GRU	0.131 51	0.104 22	0.536 58





误差最小。这得益于 VMD 分解减少了原始数据的噪声和非平稳性,TPA 机制捕捉关键时间步信息,BiGRU 处理双向依赖关系,三者结合显著提升了预测精度。尽管其训练时间相对较长,但预测准确性表现出色。相比之下,VMD-TCN-BiGRU 模型虽然预测误差略大,但训练时间更短,显示出 TCN 在并行处理方面的优势。然而,TCN 在捕捉多风机关联特性方面不及 TPA 机制,因此预测精度稍逊。VMD-TPA-BiLSTM 的 RMSE 和 MAE 接近于 VMD-TPA-BiGRU,但由于 BiLSTM 结构复杂、参数更多,其训练时间更长。VMD-AM-BiGRU 模型用时最短,但预测误差较高,说明 AM 机制在捕捉多风机关联特性方面不如 TPA。TPA-BiGRU 模型未使用 VMD 分解,训练时间较短,但预测误差更大,验证了 VMD 分解在提升预测精度中的重要作用。CNN-GRU 模型在所有对比模型中表现最差,其预测误差最大且训练时间最长,表明其不适合处理风电功率预测这一复杂时间序列任务。

尽管 VMD-TPA-BiGRU 在预测精度和可靠性方面表现出色,但其复杂性和计算成本不可忽视。EVO 算法用于优化 VMD 参数的搜索过程增加了复杂度,VMD 分解和 BiGRU 的双向依赖处理也提高了计算量。TPA 机制虽显著提升了预测精度,但也延长了训练时间。表 2 显示,该模型的训练时间略高于 VMD-TCN-BiGRU 和 TPA-BiGRU,但低于 VMD-TPA-BiLSTM 和 CNN-GRU。

因此,在实际应用中需要权衡预测精度与计算成本,以满足实时预测需求。未来研究可通过优化模型结构和提升算法效率来降低计算成本和复杂度,从而进一步提高实际应用性能。

## 4 结 论

针对多风电机组之间存在复杂联系的特性,上面提出了一种新的方法,利用 EVO 算法来优化 VMD 参数,并结合 TPA 机制的 BiGRU 网络进行超短期功率预测。主要结论如下:

1)通过采用 EVO 算法优化 VMD 参数,能够更有效地从原始信号中提取有用信息,提高信号处理的精度和效率。VMD 分解所获得的各频带内的 IMF 分布有规律可循,这为 BiGRU 神经网络的预测降低了困难。BiGRU 将正向和逆向信息相结合,

充分利用数据中的时间序列信息,提高了预测的准确性。

2)采用 TPA 机制,从隐含的 BiGRU 中抽取出多台风机间的复杂关系,并实现了不同特性的风机在不同时刻选取所需信息。与传统的注意力机制相比,该算法对多台风机的功率预测具有更高的准确性。

3)所提模型相较于其他 5 种模型,不仅具有较短的训练时间,而且能显著提高多风电机组功率超短期预测的精度。

## 参考文献

- [1] YANG J B, LIU Q Y, LI X, et al. Overview of wind power in China: Status and future [J]. Sustainability, 2017, 9(8):1-12.
- [2] 吴肃霜,蔡小路,李俊贤. 风电功率预测方法与最新技术发展研究[J]. 水电与新能源, 2024, 38(3):38-41.
- [3] 吕云龙,胡琴,熊俊杰,等. 基于时空注意力卷积模型的超短期风电功率预测[J]. 电网技术, 2024, 48(5): 2064-2073.
- [4] 杨茂,王达,王小海,等. 基于数据物理混合驱动的超短期风电功率预测模型[J]. 高电压技术, 2024, 50(11):5132-1541.
- [5] 崔杨,陈正洪,刘丽珺. 弃风限电条件下复杂地形风电场短期风功率预测对比分析[J]. 太阳能学报, 2017, 38(12):3376-3384.
- [6] 代佳琨,向月,刘俊勇,等. 基于数字孪生的区域气象关联风电预测模型[J]. 四川电力技术, 2023, 46(2):32-38.
- [7] ZHANG J H, YAN J, INFIELD D, et al. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model [J]. Applied Energy, 2019, 241:229-244.
- [8] YIN H, OU Z H, HUANG S Q, et al. A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition [J]. Energy, 2019, 189:116316.
- [9] 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.
- [10] WANG G, JIA R, LIU J H, et al. A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning [J]. Renewable energy, 2020, 145:2426-2434.