

基于 K-means 的超短期光伏功率组合预测

张 帅,张继红,李兆泽

(内蒙古科技大学 自动化与电气工程学院,内蒙古 包头 014010)

摘要:针对光伏出力的间歇性、随机性导致传统方法难以准确预测的问题,提出一种融合斑翠鸟优化算法(Pied Kingfisher Optimizer, PKO)与核极限学习机(Kernel Based Extreme Learning Machine, KELM)的光伏短期功率预测模型。首先,基于皮尔逊相关系数选取影响光伏发电功率的主要特征因素,并采用相似日理论对天气类型进行聚类。其次,采用变分模态分解法将分类后的原始数据分解为多个子序列,有效降低随机波动分量对数据的影响;进一步利用斑翠鸟在觅食过程中的自然行为优化 KELM 模型的核函数参数和正则化系数。最终,通过 KELM 模型建立历史数据之间的时间特征关系,得到光伏发电功率的预测结果。结果表明:所提 PCC-VMD-PKO-KELM 模型的决定系数 R^2 能够达到 0.99,较其他算法具有更好的预测效果。

关键词:光伏功率预测;相似日理论;变分模态分解;斑翠鸟优化算法

中图分类号:TM615 文献标识码:A DOI:10.16146/j.cnki.rndlgc.2025.03.016

[引用本文格式]张 帅,张继红,李兆泽.基于 K-means 的超短期光伏功率组合预测[J].热能动力工程,2025,40(3):144-152.
ZHANG Shuai,ZHANG Jihong,LI Zhaoze. Combined prediction of ultra short-term photovoltaic power based on K-means[J]. Journal of Engineering for Thermal Energy and Power,2025,40(3):144-152.

Combined Prediction of Ultra Short-term Photovoltaic Power based on K-means

ZHANG Shuai, ZHANG Jihong, LI Zhaoze

(School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou, China, Post Code: 014010)

Abstract: Aiming at the intermittent and stochastic nature of PV power and the difficulty of accurate prediction by traditional methods, a PV short-term power prediction model based on kernel based extreme learning machine (KELM) optimized by pied kingfisher optimizer (PKO) was proposed. Firstly, the Pearson correlation coefficient was used to select the main characteristic factors affecting the PV power and the weather types were clustered based on the similar day theory. Then, the variational modal decomposition (VMD) was used to decompose the classified raw data into multiple sub-sequences, so as to reduce the influence of random fluctuation components on the data effectively. The natural behavior of the pied kingfisher in the foraging process was used to optimize the kernel function parameters and regularization coefficients of the KELM model. Finally, the KELM model was used to establish the temporal characteristic relationship between the historical data to obtain the prediction results of photovoltaic power generation. The results show that the proposed PCC-VMD-PKO-KELM model can achieve a coefficient of determination of 0.99, which has a better fitting effect than other algorithms.

收稿日期:2024-08-23; 修订日期:2024-12-03

基金项目:内蒙古自治区重点研发与成果转化项目(2022YFHH0019);内蒙古自治区新型重要能源综合利用技术集成攻关大平台(2023PTXM001)

Fund-supported Project: Key R&D and Achievement Transformation Project of Inner Mongolia Autonomous Region (2022YFHH0019); New Important Energy Comprehensive Utilization Technology Integration and Research Platform of Inner Mongolia Autonomous Region (2023PTXM001)

作者简介:张 帅(1997-),男,内蒙古科技大学硕士研究生。

通信作者:张继红(1975-),男,内蒙古科技大学教授。

Key words: photovoltaic power prediction, similar day theory, variational modal decomposition, pied kingfisher optimizer (PKO)

引言

随着全球对可再生能源需求量的不断攀升,光伏发电作为清洁、可再生的能源形式之一,其战略地位日益凸显^[1-2]。然而,光伏发电系统高度依赖于外部环境条件,如太阳辐射强度、大气可降水量、相对湿度等因素,导致光伏输出功率呈现显著的间歇性和难以预测性。这一特性加剧了电力系统的调度难度,对电网的稳定运行与维护提出了更高的要求,尤其是在实现智能电网和微电网的精细化管理中,对光伏功率的准确预测显得尤为重要^[3]。

在光伏功率预测领域,早期的研究大多依赖于数值天气预报(Numerical Weather Prediction, NWP)的物理预测方法^[4]。这种方法通过利用风速、辐照度等关键气象数据,直接关联环境因素与光伏输出功率,展示物理预测方法的高度直观性和可解释性。然而,物理预测模型的复杂性和对大量 NWP 数据的依赖,以及对光伏电站地理因素的细致考虑,限制了其中短期预测上的表现^[5]。

为了克服物理预测模型的局限性,已有众多学者开始探索将物理法与统计法相结合。随着人工智能技术的深入融合,统计法不再局限于传统时间序列分析的框架^[6],而是扩展到了包括极限学习机^[7]、支持向量机^[8]、深度学习^[9]等在内的多种先进模型。结合后显著增强统计法在处理大规模数据集和复杂非线性关系上的能力,更为短期光伏功率预测开辟新的途径,极大地提升了预测精度和实用性^[10],本文在此基础上进行深入研究。

然而,面对光伏数据持续增长的复杂性和预测需求的不断提升,如何进一步提高预测精度和效率仍然是一个挑战。在此背景下,提出一种基于 K-均值算法(K-means)的超短期光伏功率预测模型。首先采用皮尔逊相关系数法(Pearson Correlation Coefficient, PCC)选取主要影响因素作为输入特征向量,通过 K-means 算法进行相似日聚类,以实现针对不同类别数据的针对性模型训练;随后结合变分模态分解(Variational Mode Decomposition, VMD)与

斑翠鸟优化算法(Pied Kingfisher Optimizer, PKO)优化 KELM 模型,进一步提升预测性能。通过仿真分析,结合均方根误差 RMSE、平均绝对误差 MAE、决定系数 R^2 等评价指标,对比提出模型与其他主流预测方法的实验结果,验证其有效性和预测精度。

1 气象因素分析

光伏发电功率受多种气象因素影响,如辐照度、温度、湿度和风速等,这些因素的变化导致光伏发电不稳定。在构建预测模型前,需分析气象因素与光伏发电功率的相关性。利用皮尔逊相关系数法计算各气象因素与光伏输出功率的相关性,筛选主要影响因素作为模型输入,以提高预测精度。皮尔逊相关系数 r 的计算如下:

$$r = \frac{\sum_{p=1}^q (x_p - \bar{x})(y_p - \bar{y})}{\sqrt{\sum_{p=1}^q (x_p - \bar{x})^2} \cdot \sqrt{\sum_{p=1}^q (y_p - \bar{y})^2}} \quad (1)$$

式中: r —相关系数,取值范围在 $[-1, 1]$ 之间,该值接近 0,表明其相关性弱,该值接近 1 或 -1,则表明其相关度强^[11]; x_p 和 y_p —第 p 个数据点的两个因素值; \bar{x} 和 \bar{y} —两个因素的均值; q —数据点的个数; x —气象因素数据; y —光伏功率数据。

影响光伏功率输出的主要因素如表 1 所示。由表 1 可知,太阳总水平辐射、固定倾角辐射是影响光伏发电功率的主要气象因素,呈现正相关性,将上述两个要素作为最终输入。

表 1 影响光伏功率输出的主要影响因素分析

Tab. 1 Analysis of main factors affecting photovoltaic power output

环境因素	r
太阳散射辐射指数	0.343 1
太阳总水平辐射	1.000 0
太阳直接辐射指数	0.790 3
固定倾角辐射	0.971 0
大气可降水量	-0.186 3
相对湿度	-0.319 1

2 理论基础分析

2.1 聚类算法

分析气象因素与光伏发电功率之间的皮尔逊相关性后,为深化预测模型的精准度与适应性,引入 K-means 聚类算法对历史天气数据进行精细化分类。该算法作为一种无监督学习方法,能够基于数据点的相似性和差异性自动将数据划分为多个簇,确保簇内数据相似度高,而簇间差异显著。

在光伏发电功率预测场景中采用 K-means 聚类算法对历史天气数据进行分类,分为晴天、多云和雨天等具有明确特征类别。随后,针对每个类别独立构建预测模型,以充分利用同类数据在光伏发电功率变化规律上的相似性。该算法不仅提升了模型在特定天气条件下的预测准确性,还增强了模型在不同天气状况下的泛化能力,使预测结果更加可靠和稳定。

2.2 变分模态分解

在光伏电站运营中,数据的采集系统是研究光伏系统优化发电策略的关键。然而,在数据的测量、传输及存储等关键环节,均可能受到多种噪声源的干扰,导致数据质量下降,为后续研究带来不确定性^[12]。

为克服该难题,在对历史天气数据进行精细化分类的基础上,K-means 聚类引入 VMD 对原始信号分解,对分类后的光伏发电功率时间序列进行分析,该方法能够将原始信号 $f(t)$ 分解为具有不同中心频率和带宽的本征模态分量 (Intrinsic Mode Functions, IMF)。通过约束条件,确保每个模态具有有限的带宽和特定的中心频率分量,将总带宽之和作为目标函数。在该模型中,要求所有模态总和等于原始信号,并据此构造出具体的约束变分模型:

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \quad (2)$$

$$\text{s. t. } \sum_{k=1}^K u_k = f(t)$$

式中: t —时间; u_k —各层模态分量的函数; w_k —各层模态对应的中心频率; ∂_t —偏导的运算符; $\delta(t)$ —单位脉冲信号; $*$ —卷积运算符; K —模态分解层的数量, $k=1$ 表示第一个模态分量; $f(t)$ —原始序列信号的函数。

本文通过 VMD 将复杂的光伏历史发电功率分解为多个能够体现出光伏数据随机性和波动性的本征模态分量,降低数据复杂度,为后续高精度预测提供基础^[13]。

2.3 核极限学习机

传统的极限学习机 (Extreme Learning Machine, ELM) 是一种单隐藏层的前神经网络,只需要设置网络中隐含层节点数量就可以生成局部最优解,具有较快的收敛速度,且学习效率较高^[14]。在 ELM 算法中,模型输出为:

$$\begin{cases} y(x) = h(x)\boldsymbol{\eta} = \mathbf{H}\boldsymbol{\eta} \\ \boldsymbol{\eta} = \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} \end{cases} \quad (3)$$

式中: $y(x)$ —输出函数; x —训练样本输入; $h(x)$ —隐藏层输出函数; $\boldsymbol{\eta}$ —连接权重向量; \mathbf{I} —单位矩阵; C —正则化系数; \mathbf{H} —网络隐藏层矩阵; \mathbf{T} —目标输出矩阵。

KELM 通过引入核函数解决 ELM 随机隐藏层节点和输出层权值导致的预测波动,提升模拟的拟合能力以及预测精度。核矩阵 \mathbf{f}_{ELM} 的定义式为:

$$\begin{cases} \mathbf{f}_{\text{ELM}} = \mathbf{H}\mathbf{H}^T \\ \mathbf{f}_{\text{KELM}_{g,v}} = h(x_g)h(x_v) = U(x_g, x_v) \end{cases} \quad (4)$$

式中: $\mathbf{f}_{\text{KELM}_{g,v}}$ —通过输入样本 x_g 和 x_v 的核函数 U 计算得到的 KELM 核矩阵; $U(x_g, x_v)$ —核函数,本文核函数采用 RBF 函数。

所以, KELM 模型的输出可表示为:

$$y(x) = \begin{bmatrix} U(x, x_1) \\ \vdots \\ U(x, x_n) \end{bmatrix} \left(\frac{\mathbf{I}}{C} + \mathbf{f}_{\text{KELM}} \right)^{-1} \mathbf{T} \quad (5)$$

2.4 斑翠鸟优化算法

斑翠鸟优化算法由 Bouaouda 等人^[15]提出,是一种基于群体的新型元启发式算法。该算法受斑翠鸟狩猎与共生启发,围绕三大行为阶段:悬停搜索、潜水捕食以及共生协作构建关系促进全局搜索。

2.4.1 初始化阶段

与许多其他基于种群的方法类似,PKO 通过从搜索空间随机生成一组起始解作为第一次尝试来启动搜索过程:

$$X_{i,j} = \text{LB} + (\text{UB} - \text{LB}) \times \text{rand} \quad (6)$$

式中: $X_{i,j}$ —第 i 个个体在第 j 维上的位置 ($i=1, 2, 3 \dots N, j=1, 2, 3 \dots \text{Dim}$); rand — $[0, 1]$ 内均匀分布的

随机值;UB, LB—搜索范围的上、下限。

2.4.2 悬停搜索(搜索多样性)

PKO 的探索阶段受斑翠鸟栖息和盘旋行为的启发。在算法中,搜索种群模拟花斑翠鸟的觅食习性,在栖息地和悬停位置之间交替出击以更新其位置。斑翠鸟的位置根据以下等式进行更新:

$$X_i(t+1) = X_i(t) + \alpha T(X_j(t) - X_i(t)),$$

$$i, j = 1, 2, \dots, N \quad (7)$$

在迭代过程中,当前迭代的位置用 $X_i(t)$ 表示,下一次迭代的解用 $X_i(t+1)$ 表示, $X_i(t)$ 和 $X_j(t)$ 分别表示第 i 个和第 j 个个体在 t 时刻的位置,且 $i \neq j$ 。控制参数 α 计算为 $2 \times \text{randn}(1, \text{Dim}) - 1$, 其中 randn 表示正态分布中的随机数。 N 表示总体规模, Dim 表示所考虑问题的维度。参数 T 值根据当前策略动态确定,可以是“栖息”或“悬停”。 T 的计算针对每种策略进行定制,以确保在不同的操作模式下实现最佳性能。

栖息策略中参数 T 计算如下:

$$T = (\exp(1) - \exp\left(\frac{t-1}{\text{Max_Iter}}\right)^{\frac{1}{\text{BF}}}) \cdot \cos(\text{Crest_angles}) \quad (8)$$

$$\text{Crest_angles} = 2\pi \cdot \text{rand} \quad (9)$$

最大迭代次数由 Max_Iter 指定。恒定值 BF (跳动因子)被设置为 8,并且 rand 为 0~1 之间的随机值,用于决定是否执行局部逃逸; Crest_angles 为羽冠角度。

悬停策略中参数 T 计算如下:

$$T = \text{beating_rate} \left(\frac{t^{\frac{1}{\text{BF}}}}{\text{Max_Iter}^{\frac{1}{\text{BF}}}} \right) \quad (10)$$

$$\text{beating_rate} = \text{rand} \left(\frac{\text{PKO_Fitness}(j)}{\text{PKO_Fitness}(i)} \right) \quad (11)$$

式中: $\text{PKO_Fitness}(i)$, $\text{PKO_Fitness}(j)$ —第 i 只和第 j 只斑翠鸟的适应度; beating_rate —悬停时的拍打速度,与个体适应度有关。

2.4.3 潜水捕食(开发阶段)

根据斑翠鸟身体特性,捕鱼数学模型为:

$$X_i(t+1) = X_i(t) + HA \cdot o \cdot \alpha \cdot (b - X_{\text{best}}(t)) \quad (12)$$

$$HA = \text{rand} \left(\frac{\text{PKO_Fitness}(i)}{\text{Best_Fitness}} \right) \quad (13)$$

$$o = \exp \left(\frac{-t}{\text{Max_Iter}} \right)^2 \quad (14)$$

$$b = X_i(t) + o^2 \cdot \text{randn} \cdot X_{\text{best}}(t) \quad (15)$$

式中: Best_Fitness —所有迭代得到的最佳适应度值; HA —狩猎能力; o —动态衰减因子; b —基于当前位置和最佳位置的随机扰动点。

2.4.4 共生协作(局部逃逸阶段)

斑翠鸟与许多水獭物种有共生关系,这意味着两个物种都从对方身上受益。因为鸟能把鱼从躲藏的地方引出,水獭能捕到更多的鱼,反之,鸟通过捕捉被水獭打扰的鱼而从中受益。这种行为在数学上表现如下:

$$X_i(t+1) = \begin{cases} X_i(t) + o \cdot \alpha \cdot \text{abs}(X_m(t) - X_n(t)), \\ \text{rand} > (1 - PE) \\ X_n(t), \text{ 其他} \end{cases} \quad (16)$$

$$PE = PE_{\text{max}} - (PE_{\text{max}} - PE_{\text{min}}) \left(\frac{t}{\text{Max_Iter}} \right) \quad (17)$$

式中: $\text{abs}(X_m(t) - X_n(t))$ —群体中随机选择两个个体 $X_m(t)$ 和 $X_n(t)$ 在 t 时刻位置差的绝对值; PE —斑翠鸟的捕食效率, PE_{max} 和 PE_{min} 的常数值分别为 0.5 和 0。

3 组合预测模型及评价指标

3.1 组合预测模型

本文提出的组合模型主要包括皮尔逊相关性分析、K-means 聚类、VMD 分解和 PKO 参数优化等环节,具体步骤如下:

(1) 特征筛选与聚类:利用 PCC 筛选关键气象特征,以减少数据分析的复杂性,实现有效降维。随后通过 K-means 聚类将历史数据分为晴天、多云、雨天三类;

(2) 时间序列分解:对每类天气数据应用 VMD 分解,提取不同频率的子序列;

(3) 模型构建与优化:建立 KELM 模型,并利用 KPO 算法优化其正则化系数和核参数;

(4) 训练与预测:将 VMD 分解后的子序列输入优化后的 KELM 模型进行训练,并分别预测各类天气下的光伏功率;

(5) 结果集成与评估:集成各类天气的预测结

果,通过反归一化得到最终预测值,并采用均方根误差 RMSE、平均绝对误差 MAE、决定系数 R^2 等指标评估模型性能。

预测模型流程如图 1 所示。

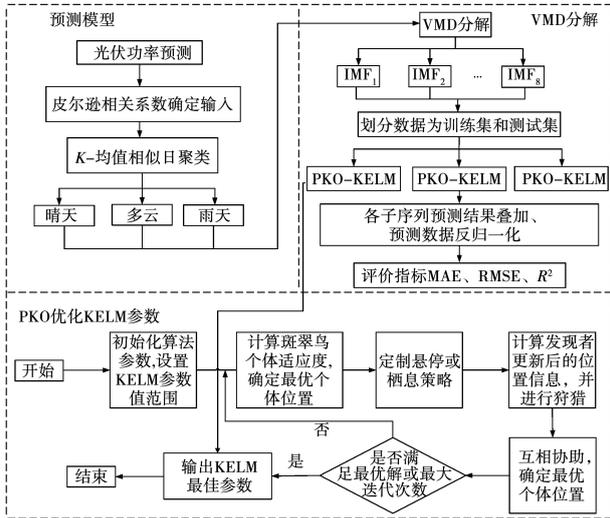


图 1 预测模型流程图

Fig. 1 Flowchart of prediction model

3.2 评价指标

平均绝对误差 MAE 是指预测值与真实值之间的平均绝对误差,范围为 $[0, +\infty]$,当经过模型预测后的光伏发电功率与真实值一致时,MAE 等于 0,代表该模型是理想模型。其表达式为:

$$MAE = \sum_{i=1}^n |y_1(i) - y_2(i)| / n \quad (18)$$

式中: $y_1(i)$ —实际值; $y_2(i)$ —预测值。

预测值与真实值之间的均方根误差 RMSE 的特点与 MAE 相近,表达式为:

$$RMSE = \sqrt{\sum_{i=1}^n [y_1(i) - y_2(i)]^2 / n} \quad (19)$$

决定系数 R^2 表示原始离散程度与拟合模型的关系,该指标正常取值区间为 $[0, 1]$ 。该值越接近 1,表明功率预测值对真实值的解释能力越强,模型拟合得越好。就实际经验而言, R^2 只要大于 0.4,则表明模型拟合效果较好^[16-18]。其表达式为:

$$R^2 = 1 - \sum_i [y_1(i) - y_2(i)]^2 / \sum_i [\bar{y}_1(i) - y_2(i)]^2 \quad (20)$$

式中: $\bar{y}_1(i)$ —实际值的平均值。

4 算例仿真分析

本文数据集源自浙江省某 120 kW 光伏电站单个分布式站点,选取 2020 年 10 月每日 10:00 - 18:45 时段,数据采集间隔为 15 min 并记录发电功率及同步气象数据,共 1 116 个样本点,用于构建与验证组合预测模型。

4.1 K-means 聚类结果

K-means 聚类算法得到的结果如图 2 所示。根据总水平辐射变量的数据特征,将天气划分为晴天、多云及雨天。以晴天聚类结果图为例,图中数据点紧密聚集,形成一个相对独立的簇,这表明数据点之间具有高度相似性。本文中簇的特征主要体现为光伏发电功率较高且平均值较接近,基于此特征,将数据点归为晴天类型。多云类型及雨天类型同理。

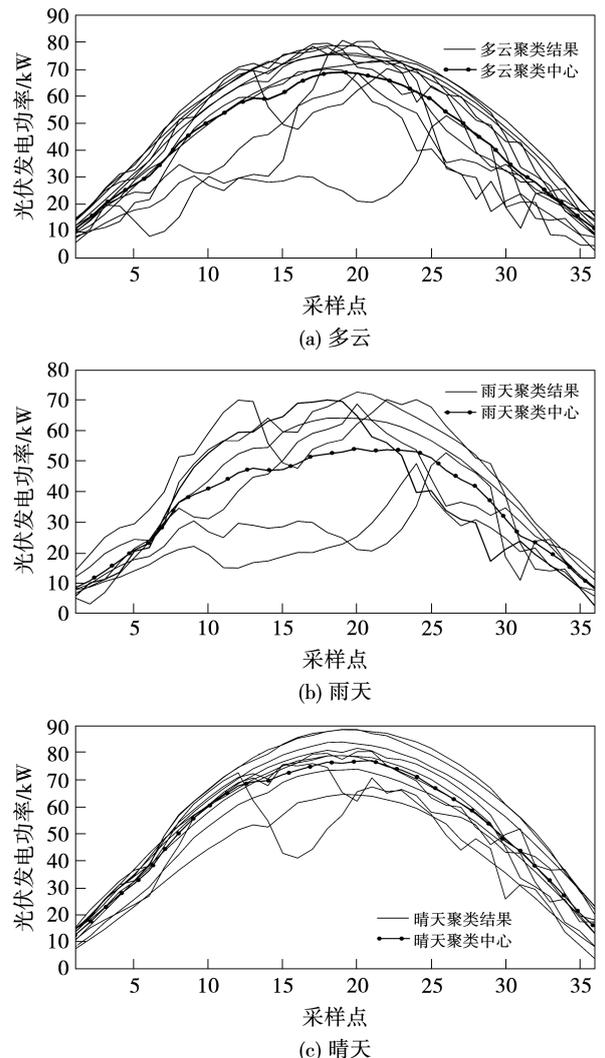


图 2 K-means 聚类结果

Fig. 2 Clustering results of K-means

4.2 VMD 参数设定

VMD 分解的精度高度依赖于分量数量 k 的恰当选择。 k 值过小易导致模态信息缺失,过大则可能引入冗余或噪声成分。选择中心频率法,针对晴天、多云、雨天等不同天气类型,通过系统实验优化 k 值选取^[19-20]。以多云天气太阳总水平辐射指数数据为例,确定最优 k 值。表 2 展示了多云天气下辐照度 VMD 分解的模态中心频率分布。从表 2 可看出,太阳总水平辐射模式中心频率在 $k=6$ 和 7 时最为相近,即在 $k=6$ 时模式中心频率达到稳定,分解得到的各子序列不存在明显的模态混叠现象,表明此模态分解数满足精度要求,因此分量数量确定为 6,初始中心频率 $\omega=1$,收敛依据 $r=1 \times 10^{-7}$ 。

表 2 多云天气模态中心频率

Tab.2 Center frequency of cloudy mode

k	太阳总水平辐射中心频率/kHz						
2	0.000 2	0.027 3					
3	0.000 2	0.027 0	0.056 1				
4	0.000 2	0.027 0	0.056 1	0.349 1			
5	0.000 2	0.027 0	0.056 0	0.196 9	0.351 9		
6	0.000 2	0.027 0	0.054 2	0.087 7	0.200 1	0.433 7	
7	0.000 2	0.027 0	0.054 2	0.087 7	0.199 8	0.349 5	0.435 5

图 3 展示了以多云天辐照度为例,经 VMD 分解后得到 IMF 分量。

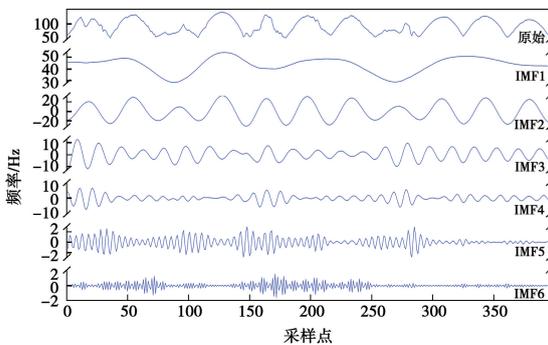
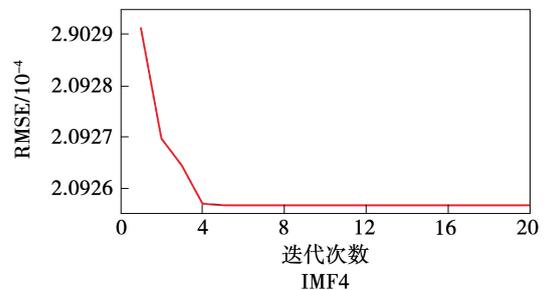
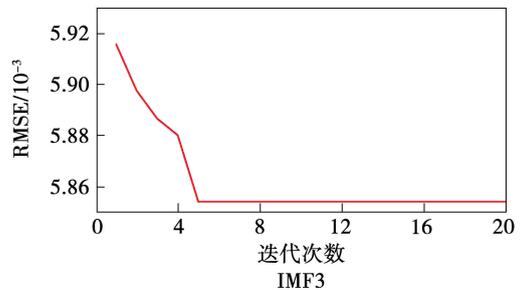
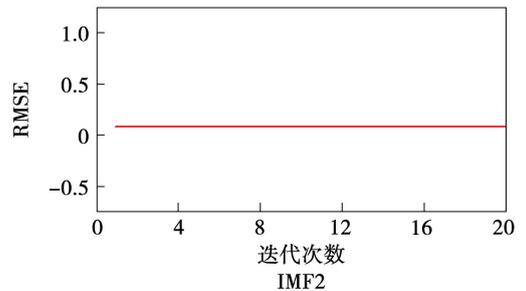
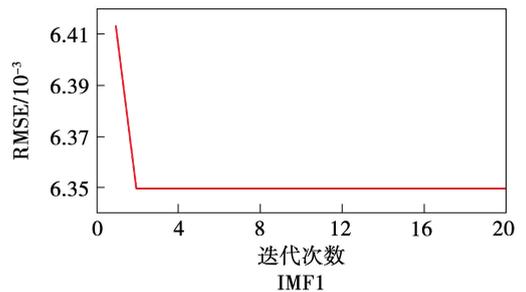


图 3 多云天气太阳总水平辐射 VMD 分解

Fig.3 VMD decomposition of total horizontal solar radiation in cloudy weather

4.3 模型参数设定

KELM 模型采用径向基函数作为核函数,并设定隐含层神经元数为 190。为了进一步提高光伏发电功率的预测精度,利用 PKO 算法对 KELM 模型的正则化系数 C 和核函数参数 S 进行优化。PKO 算法的参数设置为种群数 20,最大迭代次数为 20,维度为 2。在优化过程中,以太阳直接辐射指数、太阳总水平辐射及训练集光伏发电功率作为输入,针对多云、晴天、雨天 3 种不同天气类型下的各分量进行了细致的寻优,分别在每类天气中选取 1 天作为测试集,其余作为训练集。以多云天气类型为例进行了详细展示,具体的寻优过程如图 4 所示。可以看出,该算法在迭代时可快速达到最优值,具有更快的收敛速度。



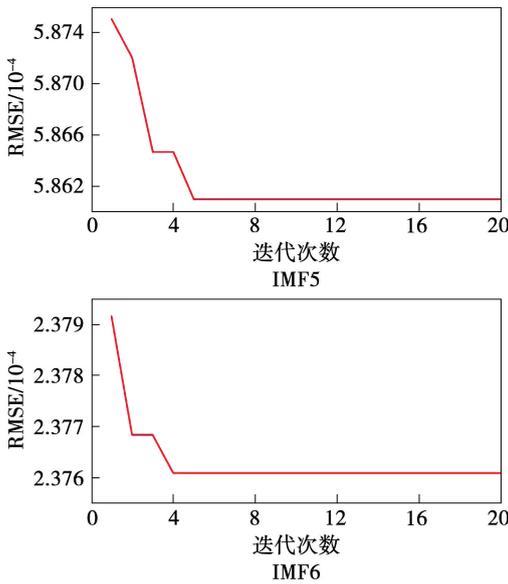


图 4 多云天气分量 1-6 寻优过程

Fig. 4 IMF1 - IMF6 optimization processes in cloudy weather

4.4 算例分析

为验证本文提出的预测模型在短期光伏功率预测中的有效性,将其应用于实际光伏电站数据,并与反向传播神经网络(Back Propagation Neural Network, BPNN),ELM,KELM,VMD-KELM 和 PKO-KELM 等基准模型进行了对比实验。由于同一算法对所有天气类型预测功能均有效,因此以多云天气为例,展示了各模型的预测性能以及采样次数与对应预测结果的关系,实验结果如图 5 所示。

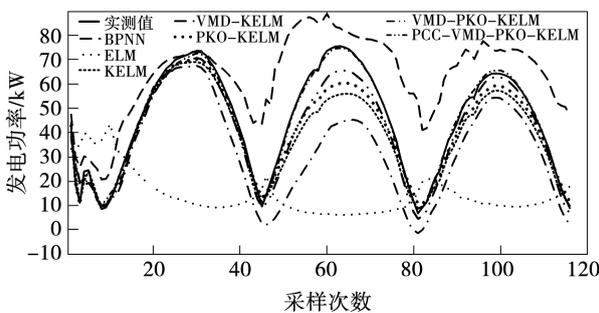


图 5 多云预测结果图

Fig. 5 Forecast result diagram of cloudy days

由图 5 可以看出,单一 ELM 模型的预测结果与实际值之间的偏差最为显著;相比之下,BPNN 模型虽有所改善,但仍显不足;引入 KELM 模型,预测结

果优于 BPNN,进一步提升了预测精度,但与实际光伏功率曲线之间仍存在较大偏差。为缩小这一差距,构建 VMD-KELM 模型,但预测结果并未优于基础 KELM 模型,反而较原 KELM 模型有所降低。引入 PKO 算法对 KELM 模型进行优化,构建 PKO-KELM 模型,结果证明,PKO-KELM 模型的预测结果得到了改善,表现结果优于 KELM 模型,证明其有效性。为进一步探索 VMD 在预测模型中的潜在优化效果,提出 VMD-PKO-KELM 模型,实验结果显示,预测结果优于 PKO-KELM 模型,与实际值偏差较小。聚类后 PCC-VMD-PKO-KELM 预测结果优于 VMD-PKO-KELM,且与实际功率曲线吻合度最高,预测结果更好,说明所提模型提高了短期光伏功率的预测精度。

图 6 为各模型的功率预测误差图。可以看出,本文所提 PCC-VMD-PKO-KELM 模型的功率预测误差最小。

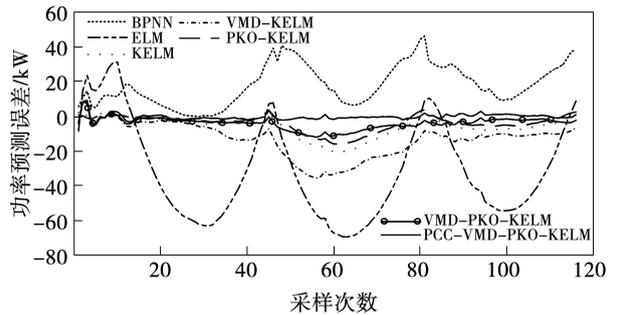


图 6 多云功率预测误差图

Fig. 6 Power forecast error diagram of cloudy days

同时,不同预测模型所需预测时间存在一定的差异性,为了更全面地评价各模型性能,对各模型的预测时间进行了对比实验,结果如表 3 所示。

表 3 各模型预测耗时

Tab. 3 Forecast time of each model

模 型	耗时/s
BPNN	3.03
ELM	0.80
KELM	0.06
VMD-KELM	0.22
PKO-KELM	10.94
VMD-PKO-KELM	65.08
PCC-VMD-PKO-KELM	62.49

分析表3可知,随着模型复杂度的增加,模型的预测时间也在逐渐增加,单个模型的耗时短于组合模型的耗时,但组合模型的预测精度较单个模型有很大提升,具体评价指标如表4所示。以雨天为例进行分析,在所有模型中,ELM预测效果最差,PCC-VMD-PKO-KELM模型预测效果最好。ELM模型的RMSE为18.5995,MAE为9.5771, R^2 为0.1799,而PCC-VMD-PKO-KELM的RMSE为1.4103,MAE为1.0559, R^2 为0.9953,相较于对比模型中效果最差的ELM,PCC-VMD-PKO-KELM模型的RMSE、MAE分别降低94.32%和88.97%, R^2 提升81.54%。同理,其他类型天气同样表明本文所提PCC-VMD-PKO-KELM组合模型的预测精度高于其他模型。

表4 评价指标计算

Tab.4 Calculation of evaluation indicators

天气	模型	RMSE	MAE	$R^2/\%$
晴天	BPNN	2.1789	1.5629	0.9904
	ELM	20.5005	16.8406	0.1555
	KELM	25.7564	19.2658	0.3300
	VMD-KELM	10.2415	8.558	0.7892
	PKO-KELM	3.6068	2.5431	0.9739
	VMD-PKO-KELM	2.9649	2.5149	0.9823
	PCC-VMD-PKO-KELM	1.1632	0.9126	0.9972
多云	BPNN	21.6183	18.1706	0.9846
	DELM	41.3941	35.7868	0.2836
	KELM	8.2298	6.1244	0.8484
	VMD-KELM	16.3843	13.3982	0.3990
	PKO-KELM	6.0854	4.3589	0.9171
	VMD-PKO-KELM	5.2144	4.2793	0.9391
	PCC-VMD-PKO-KELM	1.8082	1.2402	0.9927
雨天	BPNN	2.6916	2.4618	0.9828
	ELM	18.5995	9.5771	0.1799
	KELM	6.0081	4.8542	0.9144
	VMD-KELM	4.6159	4.1944	0.9495
	PKO-KELM	2.5581	2.3314	0.9845
	VMD-PKO-KELM	2.1621	1.7224	0.9889
	PCC-VMD-PKO-KELM	1.4103	1.0559	0.9953

5 结论

本文提出了一种基于PCC特征选择和PKO优化KELM的短期光伏功率预测方法,结合实际算例验证,得出具体结论如下:

(1) 利用K-means算法对历史数据进行相似日聚类,分为晴天、多云、雨天三类进行预测,降低不同天气条件对预测结果的影响;PCC确定关键输入因素,避免由相关性较低因素引起的预测误差且可以降低模型复杂度,减少计算量。

(2) 在3种典型天气下,PCC-VMD-PKO-KELM组合模型的 R^2 都超过0.99,其他天气情况下所提模型的RMSE和MAE也都有显著提升,说明组合模型预测效果更好且模型的预测时长仅1min,预测效率有大幅提升。

参考文献:

- [1] IRENA. Renewable capacity statistics 2023 [R]. 978-92-9260-5254, Abu Dhabi: International Renewable Energy Agency, 2023.
- [2] 张 姗,冬 雷,纪德洋,等. 基于 NWP 相似性分析的超短期光伏发电功率预测[J]. 太阳能学报,2022,43(4):142-147.
ZHANG Shan, DONG Lei, JI Deyang, et al. Power forecasting of ultra-short-term photovoltaic station based on NWP similarity analysis [J]. Acta Energetica Sinica, 2022, 43(4): 142-147.
- [3] RAHIMI N, PARK S, CHOI W, et al. A comprehensive review on ensemble solar power forecasting algorithms [J]. Journal of Electrical Engineering & Technology, 2023, 18(2): 719-733.
- [4] PAZIKADIN A R, RFAID, ALI K, et al. Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ANN): A review of five years research trend [J]. Science of the Total Environment, 2020, 715: 136848.
- [5] 商立群,黄辰浩,侯亚东,等. 采用特征优选和优化深层核极限学习机的短期风电功率预测[J]. 西安交通大学学报,2023, 57(1): 66-77.
SHANG Liqun, HUANG Chenhao, HOU Yadong, et al. Short-term wind power prediction by using the deep kernel extreme learning machine with well-selected and optimized features [J]. Journal of Xi'an Jiaotong University, 2023, 57(1): 66-77.
- [6] 高寒旭,袁祖晴,张淑婷,等. 基于 LSTM 模型的短期光伏功率预测[J]. 太阳能学报,2024,45(6):376-381.
GAO Hanxu, YUAN Zuqing, ZHANG Shuting, et al. Short-term photovoltaic power prediction based on LSTM model [J]. Acta Energetica Sinica, 2024, 45(6): 376-381.

- [7] 李 斌,张一凡,颜世焯,等.基于改进极限学习机(ELM)的光伏发电预测方法研究[J].热能动力工程,2022,37(10):207-214.
LI Bin,ZHANG Yifan,YAN Shiye,et al. Research on photovoltaic power generation prediction method based on improved extreme learning machine (ELM)[J]. Journal of Engineering for Thermal Energy and Power,2022,37(10):207-214.
- [8] SCOTT C,AHSAN M,ALBARBAR A. Machine learning for forecasting a photovoltaic (PV) generation system[J]. Energy,2023,278:127807.
- [9] 李立刚,孙立群,戴永寿,等.基于时空图卷积神经网络的光伏发电功率超短期预测方法[J].热能动力工程,2023,38(9):152-157.
LI Ligang,SUN Liqun,DAI Yongshou,et al. Ultra-short-term prediction method of photovoltaic power based on spatio-temporal graph convolutional neural network[J]. Journal of Engineering for Thermal Energy and Power,2023,38(9):152-157.
- [10] 李 争,张 杰,徐若思,等.基于相似日聚类 and PCC-VMD-SSA-KELM 模型的短期光伏功率预测[J].太阳能学报,2024,45(2):460-468.
LI Zheng,ZHANG Jie,XU Ruosi,et al. Short term photovoltaic power prediction based on similar day clustering and PCC-VMD-SSA-KELM model [J]. Acta Energetica Solaris Sinica,2024,45(2):460-468.
- [11] 宋 煜,郭军红,袁 荔,等.基于逐步聚类分析的短期光伏发电预测方法[J].热能动力工程,2023,38(10):158-166.
SONG Yu,GUO Junhong,YUAN Li,et al. Short-term photovoltaic power generation prediction method based on stepwise clustering analysis[J]. Journal of Engineering for Thermal Energy and Power,2023,38(10):158-166.
- [12] 杨锡运,马文兵,彭 琰,等.基于组合神经网络的分布式光伏超短期功率预测方法[J].热力发电,2023,52(8):162-171.
YANG Xiyun,MA Wenbing,PENG Yan,et al Distributed photovoltaic ultra-short-term power prediction method based on combined neural network [J]. Thermal Power Generation,2023,52(8):162-171.
- [13] 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.
- [14] 杨荔强,崔双喜.基于VMD-SSA-HKELM的短期光伏功率预测[J].电源技术,2024,48(6):1154-1159.
YANG Liqiang,CUI Shuangxi. Short-term PV power prediction based on VMD-SSA-HKELM [J]. Chinese Journal of Power Sources,2024,48(6):1154-1159.
- [15] BOUAOUA A,HASHIM F A,SAYOUTI Y,et al. Pied kingfisher optimizer: A new bio-inspired algorithm for solving numerical optimization and industrial engineering problems [J]. Neural Computing & Applications,2024,36:15455-15513.
- [16] 姜建国,杨效岩,毕洪波.基于VMD-FE-CNN-BiLSTM的短期光伏发电功率预测[J].太阳能学报,2024,45(7):462-473.
JIANG Jianguo,YANG Xiaoyan,BI Hongbo. Photovoltaic power forecasting method based on VMD-FE-CNN-BiLSTM [J]. Acta Energetica Solaris Sinica,2024,45(7):462-473.
- [17] 刘丽桑,郭凯琪,徐哲壮,等.基于数据挖掘的双模式组合光伏功率日前预测[J].武汉大学学报(工学版),2024,57(10):1459-1468.
LIU Lisang,GUO Kaiqi,XU Zhezhuang,et al. Day-ahead prediction of dual-mode combined photovoltaic power based on data mining [J]. Engineering Journal of Wuhan University,2024,57(10):1459-1468.
- [18] XIA Xin,WANG Xiaolu. A novel hybrid model for short-term wind speed forecasting based on twice decomposition, PSR, and IMVO-ELM[J]. Complexity,2022,2022:4014048.
- [19] BOUAOUA A,HASHIM F A,SAYOUTI Y,et al. Pied kingfisher optimizer: A new bio-inspired algorithm for solving numerical optimization and industrial engineering problems [J]. Neural Computing & Application,2024,36:15455-15513.
- [20] 李桂兰,杨 杰,周满国.基于改进时间卷积网络的光伏发电功率预测[J].激光与光电子学进展,2022,59(8):480-489.
LI Guilan,YANG Jie,ZHOU Manguo. Power prediction of photovoltaic generation based on improved temporal convolutional network [J]. Laser & Optoelectronics Progress,2022,59(8):480-489.

(刘 颖 编辑)