

引文格式:肖春,郝俊博,杨晓霞,等.基于改进高斯过程回归模型的光伏发电功率预测[J].太原理工大学学报,2026,57(2):287-295.

XIAO Chun, HAO Junbo, YANG Xiaoxia, et al. Power prediction of photovoltaic power generation based on improved gaussian process regression modeling[J]. Journal of Taiyuan University of Technology, 2026, 57(2): 287-295.

基于改进高斯过程回归模型的 光伏发电功率预测

肖春^{1,2}, 郝俊博¹, 杨晓霞¹, 韩肖清²

(1. 国网山西省电力有限公司营销服务中心, 山西 太原; 2. 太原理工大学 机械工程学院, 山西 太原)

摘要:【目的】随着我国光伏发电占比不断提高,光伏发电受气象因素的影响较大,其输出功率因气象特征的复杂多变表现出强烈的间歇性和波动性,对未来光伏发电功率预测的准确度将直接影响电网的稳定安全运行。针对光伏发电功率的预测精度提升问题,提出了一种融合局部离群因子算法、遗传算法与高斯过程回归(GA-LOF-GPR)的预测模型。【方法】首先,本文挖掘发电功率与气象特征的关系,并采用特征权重 K 均值聚类对气象类型进行分类;其次,将局部离群因子算法与高斯过程回归模型结合,构建局部异常因子加权的高斯过程回归(LOF-GPR)的预测模型;最后,通过运用遗传算法优化加权高斯过程回归模型的超参数。【结果】通过对2018年澳大利亚的光伏发电数据的仿真预测,验证了该预测模型的有效性和准确性。

关键词:功率预测;特征权重 K 均值聚类;局部离群因子;高斯过程回归;遗传算法

中图分类号:TM615 **文献标识码:**A

DOI: 10.16355/j.tyut.1007-9432.20240582

文章编号: 1007-9432(2026)02-0287-09

Power Prediction of Photovoltaic Power Generation Based on Improved Gaussian Process Regression Modeling

XIAO Chun^{1,2}, HAO Junbo¹, YANG Xiaoxia¹, HAN Xiaoqing²

(1. Marketing Service Center of State Grid Shanxi Electric Power Co. Ltd., Taiyuan, Shanxi, China; 2. College of Mechanical Engineering, Taiyuan University of Technology, Taiyuan, Shanxi, China)

Abstract: 【Purposes】The proportion of photovoltaic(PV) power generation has been increasing in China. PV power generation is greatly affected by meteorological factors and its output power shows strong intermittency and volatility because of the complexity and variability of meteorological characteristics. Thus, the accuracy of future PV power prediction will directly affect the stable and safe operation of the power grid. 【Methods】For improving the prediction accuracy of PV power, in this paper, a prediction model that integrates the local outlier algorithm, genetic algorithm and Gaussian process regression (GA-LOF-GPR) was proposed. First, the relationships between power generation and meteorological features were explored, and feature weight K -means clustering was used to classify meteorological types. Second, the local outlier factor algorithm was combined with the Gauss-

收稿日期:2024-07-22;修回日期:2024-10-15

基金项目:国网山西省电力公司科技项目(52051L230103)

第一作者:肖春,博士研究生,高级工程师,主要从事电力营销技术、电力系统运行与控制研究等,(E-mail)tyutxiaochun@163.com

通信作者:韩肖清,博士,教授,博士生导师,主要从事电力系统运行与控制、微电网及新能源技术研究,(E-mail)hanxiaqing@tyut.edu.cn

ian process regression model to construct the prediction model of local outlier factor-weighted Gaussian process regression (LOF-GPR). Finally, the hyper-parameters of the weighted Gaussian process regression model were optimized by applying the genetic algorithm. **【Results】**The effectiveness and accuracy of the prediction model are verified by simulation prediction of the PV power generation data in Australia in 2018.

Keywords: power prediction; feature weight K -mean clustering; local outliers factor; Gaussian process regression; genetic algorithm

随着传统化石能源储备日益下降以及其带来的污染问题,清洁能源日渐成为能源结构中的中坚力量。在我国“碳达峰、碳中和”的目标下,太阳能作为最易获取的可再生能源,已经得到了电力行业的广泛认同和应用^[1]。光伏发电受气象因素影响较大,导致光伏功率表现出强烈的间歇性和波动性^[2]。这些特性在光伏大规模并网运行时,带来系统稳定性和管理等方面的挑战,可靠的光伏功率预测可以使电力调度部门提前制定调度计划,确保电力系统安全运行。因此,为保障电网稳定运行,加强对光伏功率精准预测的研究具有重要意义^[3]。

目前已有众多研究方法应用于实际光伏功率预测。文献[4]通过将地基云图和气象数据分析与光伏功率预测相结合,实现了预测优化,但该类方法需要大量的资金和设备。实际工程中,由于技术和资金等条件的限制,很难在大规模地区进行配置。在此背景下,考虑气象数据特征的分类预测成为一大研究方向,分类预测模型将气象数据进行聚类划分天气类型,再调用预测模型进行分别预测,以减小天气变化对功率预测的影响。文献[5]为保留天气类型动态变化性质,采用面板数据层次聚类方法对历史天气类型进行划分后,建立风电功率预测模型;文献[6,7]分别用特征加权模糊聚类方法和 K -means算法对环境数据进行聚类,并验证了样本分类特征后应用于风电预测;文献[8]将多种聚类算法结合成基聚类器,获取准确的相似日集合来用于光伏跟踪预测;文献[9]运用模糊 C 均值聚类对降维后的数据划分天气类型,提高光伏短期预测精度;文献[10]和文献[11]通过 K 均值聚类将历史数据划分相似日,来解决光伏功率预测模型训练数据敏感的问题。上述聚类方法中,层次聚类方法在处理大规模数据集计算复杂度较高,且对初始样本顺序敏感;模糊聚类方法虽然可以处理不确定性和模糊性,但计算量大,且聚类结果的解释性不强;相比之下, K -means算法简单高效,但对初始聚类中心的选择敏感,需要预先确定 K 值。同时,大量的相

关文献已经确定了天气类型数量^[6-11],这为我们选择合适的 K 值提供了参考。

高斯过程回归(Gaussian process regression, GPR)是一种利用贝叶斯统计和概率模型来预测数据的非参数模型方法^[12]。与传统的预测方法相比,GPR的优势在于它不仅能够估计预测值,还能提供预测的不确定性度量。因此,其广泛应用于电力系统的相关预测,如居民用电量预测^[13],电池SOC估计^[14]和发电功率预测。文献[15]考虑设备老化和环境变化等因素,利用高斯过程回归模型和分块缓存相结合,实现了短期风电功率概率预测;文献[16]通过长短期记忆网络分析气象因素和输出功率的关联性,并结合高斯过程回归实现准确的光伏功率预测;文献[17]运用纵横交叉算法进行优化高斯过程回归模型,减小了光伏预测的误差;文献[18]采用粒子群优化算法对高斯过程回归模型进行优化,可以帮助模型更好地拟合训练数据。上述文献中,尽管已有研究探讨了气象特征对发电功率的影响,但在进行预测前的准备工作中,许多文献未能体现气象特征的影响。此外,高斯过程回归模型的基础数据直接影响预测结果,因此,对异常数据的检测和处理十分关键。

基于以上现状,本文提出一种基于改进高斯过程回归模型的光伏发电功率预测模型,采用GPR模型融合特征权重 K 均值聚类、局部离群因子算法和遗传算法,旨在更好地训练模型,提高预测效果。第一章,对气象特征进行相关性分析获取特征权重矩阵,采用特征权重 K 均值聚类划分天气类型;第二章,将基于异常离群值识别的融合局部离群因子算法与GPR模型结合,并利用遗传算法预测模型超参数;第三章,进行仿真对比分析,验证所提组合预测模型预测性能的优越性。

1 特征权重 K 均值聚类划分气象数据

本节首先通过特征权重 K 均值聚类对气象数据进行分类,旨在识别不同的气象类型,更准确地

捕捉气象因素对光伏发电功率的影响,并为光伏发电功率预测提供更可靠的基础。

1.1 相关性分析

皮尔逊相关系数^[19]是一种广泛使用的统计量,它能够量化两个变量之间线性关系的强度和方向。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (1)$$

其中, r 表示相关性系数; x_i 表示影响光伏出力的一系列气象参数,如温度、湿度、辐射强度和降雨量等, $\bar{x}$ 表示这些气象参数的平均值; y_i 表示光伏发电的功率输出, $\bar{y}$ 表示 y_i 的平均值。

通过计算皮尔逊相关系数,我们可以评估两个变量之间的关联程度。这种相关系数的值在 $[-1, 1]$ 之间,符号为正表示两个变量为正相关,反之表示负相关。 r 的绝对值越大,表明变量之间的线性关系越强; r 的绝对值越小,表明线性关系越弱。

1.2 特征权重 K 均值聚类划分天气类型

K 均值聚类^[20]是一种基于样本集合划分的聚类算法,它的基本思想是将数据点划分为 K 个聚类,使得每个数据点与其所属聚类中心的距离之和最小。对于处理大数据集合,其具有高效性和良好的伸缩性。传统的 K 均值聚类算法对所有特征赋予相同的权重,这可能导致某些特征对聚类结果的影响被低估或高估。本文将1.1的相关性分析与 K 均值聚类相结合,形成特征权重 K 均值聚类。该方法在传统的 K 均值聚类算法的基础上增加了特征权重的概念,以处理不同特征对聚类结果可能具有不同重要性的情况。这种方法适用于特征之间重要性不同的场景。具体来说,通过对传统的欧式距离进行加权,形成加权欧式距离。

首先将聚类数据集中到同一矩阵中,并将矩阵划分为 k 类(聚类的数量),随机选择 k 个数据行作为初始聚类中心。之后,计算矩阵的每一行与类中心的距离。这里采用特征权重调整后的欧式距离:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^n w_l (x_{il} - x_{jl})^2}. \quad (2)$$

其中, w 表示不同气象参数的特征权重,其值与表1中对应的相关性系数相等; x_{il} 表示第 i 个气象特征向量的第 l 个属性, x_{jl} 表示第 j 个聚类中心的第 l 个属性; n 表示数据维度。之后根据所有数据行和每个聚类中心的加权欧式距离,将其聚集到最近的聚类中心。根据各聚类的数据行,计算数据均值来更

新每个聚类中心的位置。之后重复计算加权欧氏距离和更新聚类中心,直到结果不再发生改变,即得到气象类型划分的结果。

2 基于GA-LOF-GPR的光伏出力预测模型

本节首先获取每个数据点的局部离群因子,然后采用加权的方式优化高斯过程回归模型,并使用遗传算法优化超参数,旨在不影响数据量的基础上,减小异常值的权重,提高预测精度。

2.1 基于LOF-GPR的光伏出力预测模型

2.1.1 基于LOF算法的异常值处理

局部离群因子(Local Outlier Factor, LOF)算法^[21]是一种基于数据点密度的无监督异常检测算法。LOF算法通过比较数据点与周围邻居的局部密度差异来识别异常值,将局部密度明显低于其邻近点的样本视为异常。算法为每个数据点计算一个局部离群因子,观察LOF值是否明显偏离1来判断一个点是否为异常点。

由LOF算法原理可以总结出,检测数据值是否异常的关键是计算LOF值。因此,需要首先计算局部样本数据的可达距离,根据可达距离计算可达密度,最后求解出局部离群因子LOF。

1)根据欧氏距离公式即式(2)可求解数据集任意两点之间的距离,计算点 p 与距离该点第 k 远的点 x (不包含 p 本身)之间的欧式距离 $d_k(p) = d(p, x)$ 。

2)点 p 的第 k 个集群点 x 与点 p 的可达距离 $\text{dist}_k(p, x)$ 为:

$$\text{dist}_k(p, x) = \max\{\text{dist}_k(p), \text{dist}(p, x)\}. \quad (3)$$

点 p 的 k 个集群样本点的可达距离被定义为点 p 与第 k 个集群点之间的距离的最大值,这些点到 p 的距离都等于 $\text{dist}_k(p)$ 。

3)局部可达密度

局部可达密度(Local Reachability Density, LRD)衡量点 p 在其邻域内与其他点之间的距离的倒数的平均值。

$$\text{lrd}_k(p) = \frac{|N_k(p)|}{\sum_{x \in N_k(p)} \text{dist}_k(p, x)}. \quad (4)$$

具体来说,这个度量考虑了点 p 与它所在群组中所有邻近点的第 k 个可达距离。因此,如果点 p 与其邻域内的点之间的距离较短,那么其局部可达密度就会较高。反之,如果这些距离较长,点 p 的局部

可达密度较低,这可能表明点 p 是一个异常值。

4) 局部离群因子

$$\text{LOF}_k(p) = \frac{\sum_{x \in N_k(p)} \frac{\text{lrd}_k(x)}{\text{lrd}_k(p)}}{|N_k(p)|}. \quad (5)$$

当 LOF 值接近 1 时,表明点 p 与其周围邻点 x 的密度相似,可能与这些邻点属于同一集群。如果 LOF 值小于 1,这通常意味着点 p 位于一个较为密集的区域,从而更可能是一个正常的内部点,应该设置较大的权重。相反,如果 LOF 值大于 1,通常表明点 p 可能是一个异常点,应该设置较小的权重。

进而获取权重公式为:

$$w_i = \exp(-A|\text{LOF} - 1|). \quad (6)$$

其中 A 表示系数,通过调整 A 的值,我们可以将局部异常因子映射到一个特定的数值范围。这种方法允许我们更精确地衡量点 p 与周围邻点的相对密度,从而更有效地识别异常点并分配权重。

2.1.2 基于 LOF-GPR 的光伏出力预测模型

GPR 模型利用先验信息在贝叶斯统计框架中进行状态预测,并能够估计后验分布。这种方法不仅能够提供预测的均值,还能输出预测的方差和置信区间,从而使得预测结果能够表达出相应的不确定性。

局部异常因子加权的高斯过程回归模型融合局部离群因子(Local Outlier Factor, LOF)算法与高斯过程回归模型。该模型通过分析历史数据集

中的每个样本点,评估其异常程度。样本点越异常,其权重就会越低,从而在预测过程中减少这些点对整体结果的影响。训练数据中每个数据都将被设置一个权重 w_i ,设 N 个数据的 n 阶对角线权重矩阵为 Ω :

$$\Omega = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}. \quad (7)$$

加权高斯过程是贝叶斯线性回归的改进,类似传统的高斯过程回归,在构建光伏功率预测模型时,可以首先设定模型的基本形式如下:

$$f(x) = \phi(x)^T \beta. \quad (8)$$

其中, $\phi(x)$ 是非线性映射函数, β 表示权重向量。若光伏输出功率 y 存在噪声 ε ,则权重估计模型可进一步变换:

$$w_i y_i = w_i f(x) + \varepsilon. \quad (9)$$

结合权重估计函数以及贝叶斯线性回归模型中后验分布推导过程,其后验积分得到预测结果的概率分布可表示为:

$$\begin{aligned} p(f_* | X_*, y) &= \int p(f_* | X_*, w) p(w | X, y) dw \\ &= N(\sigma_n^{-2} \phi(X_*)^T B^{-1} \phi(X) \Omega^2 y, \phi(X_*)^T B^{-1} \phi(X_*)). \end{aligned} \quad (10)$$

式中, $B = \sigma_n^{-2} \phi(X) \Omega^2 \phi(X)^T + \sum_p^{-1}$, 可得预测结果分布概率表达式为:

$$p(f_* | X_*, X, y) \sim N \left(\begin{pmatrix} K(X_*, X) (\Omega^2 K + \sigma^2 I_n)^{-1} \Omega^2 y, \\ K(X_*, X_*) - K(X_*, X) \cdot \\ \left((K(X, X) + \Omega^{-1} \sigma^2 I \Omega^{-1})^{-1} K(X, X_*) \right) \end{pmatrix} \right). \quad (11)$$

则训练集光伏发电功率数据和测试集的预测数据概率分布 f_* 的联合分布为:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left(0, \begin{bmatrix} K(X, X) + \Omega^{-2} \sigma_n^2 I \Omega K(X, X)^T \\ K(X, X) & K(X, X_*) \end{bmatrix} \right). \quad (12)$$

最终可获得预测结果的概率分布如下式:

$$p(f_* | X_*, X, y) \sim N \left(\begin{pmatrix} K(X_*, X) C^{-1} y, \\ K(X_*, X_*) - K(X_*, X) \cdot C^{-1} K(X, X_*) \end{pmatrix} \right). \quad (13)$$

其中, $C = K(X, X) + \Omega^{-2} \sigma_n^2 I$ 。

2.2 基于 GA-LOF-GPR 的光伏出力预测模型

在高斯过程回归模型中,核函数的选择直接影响模型预测的准确度,而作为核函数的重要组成部分,超参数的变化同样会显著改变预测结果。因此,优化超参数是提升高斯过程回归模型性能的关键步骤。在以往的研究中,常见的超参数优化方法

包括最大似然估计(Maximum Likelihood Estimation, MLE)和共轭梯度法。最大似然估计是一种基于经验风险最小化原理的方法。通过构造一个包含样本和未知参数的似然函数,研究者可以实现对该函数的最大化,从而使得样本出现的概率达到最大。这一过程能够有效地获得与模型相对应的超参数,进而提高模型的预测能力。另一方面,共

轭梯度法在最速下降方法的基础上引入共轭性,创建一组共轭方向,使得搜索过程更加高效,加快收敛速度,消除最速下降法中常见的抖动现象。然而,共轭梯度法作为一种确定性寻优算法,对初始值的选择极为敏感,容易陷入局部最优解,从而影响最终的优化结果。

鉴于以上方法的局限性,本文提出采用遗传算法(Genetic Algorithm, GA)对超参数进行优化,建立 GA-LOF-GPR 模型,实现对光伏出力的精准预测。遗传算法是一种基于自然选择和遗传学原理的全局优化方法,具有较强的全局搜索能力,能够有效避免局部最优解的问题。通过遗传算法的优化,期望找到更优的超参数组合,从而提升高斯过程回归模型在光伏出力预测中的表现。

本文算法的整体流程图如图 1 所示。

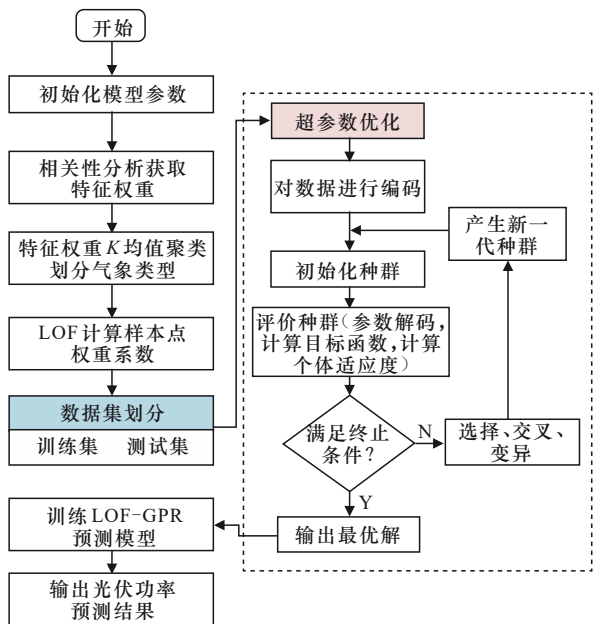


图 1 算法流程图

Fig.1 Algorithm flow chart

3 算例分析

3.1 参数设置与模型评价函数

本文算例采用澳大利亚某一年的实测光伏发电功率数据和其对应的天气特征数据,采样分辨率为 15 min,每日有 96 个数据,将其作为训练集和测试集的研究时段。模型的输入特征为:温度、相对湿度、太阳总辐射强度、散射水平辐射强度、风向(度)、日降雨量。本文采用的加权高斯过程回归模型包含两个超参数,分别是长度尺度和方差尺度。本文采用的加权高斯过程回归模型包含两个超参数,分别是长度尺度和方差尺度。GA 算法中,种群

规模 $N=200$,交叉概率 $P_c=0.8$,变异概率 $P_m=0.05$,迭代初始值 $g=0$,最大迭代次数 $G=100$ 。

为验证所提预测模型的适用性,本文考虑晴天、多云和雨天 3 种不同天气类型对光伏发电功率进行预测。同时,本文各预测方法的预测精度通过点预测误差来评估和对比,采用的评价指标分别是:

均方根误差(Root Mean Squared Error, RMSE,公式中表示为 E_{RMS}):

$$E_{\text{RMS}} = \sqrt{\frac{1}{T} \sum_{t=1}^T |p_t^{\text{true}} - p_t^{\text{forecast}}|^2} \quad (14)$$

平均绝对误差(Mean Absolute Error, MAE,式中表示为 E_{MA}):

$$E_{\text{MA}} = \frac{1}{T} \sum_{t=1}^T |p_t^{\text{true}} - p_t^{\text{forecast}}| \quad (15)$$

其中, p_t^{true} 表示 t 时刻光伏发电功率的真实值; p_t^{forecast} 表示 t 时刻光伏发电功率的预测值。

3.2 相关性分析与气象数据划分结果

本文探究各项气象数据与光伏发电功率之间的关系,通过曲线图展示它们之间的变化关系,如图 2 所示。

进一步根据皮尔逊相关性系数公式计算各气象特征与光伏发电功率的相关系数,结果如表 1 所示。

表 1 气象特征与光伏输出功率的相关系数

Table 1 Correlation coefficients between meteorological characteristics and PV output power

气象特征	温度	相对湿度	辐射强度	散射水平	风向	日降雨量
相关系数	0.33	-0.38	0.91	0.53	0.027	-0.062

由图 2 以及表 1 的相关系数可以看出,总辐射水平、散射水平以及温度与光伏发电功率的变化关系更为紧密。根据表 1 的相关性系数,设置各气象特征的特征权重,以此进行特征权重 K 均值聚类,聚类结果如图 3 所示。

由图 3 可知,三类曲线的波动特性明显不同。具体分析:第一类表现出稳定且平滑的上升和下降趋势,类似抛物线,反映了晴天光照条件稳定,光伏发电效率高;第二类曲线围绕抛物线波动,波动性较大,代表多云天气下云层遮挡导致的太阳辐射不稳定;第三类曲线波动幅度大且整体水平较低,反映雨天光照条件差,光伏发电效率显著降低。由此可将全年气象数据划分为晴天、多云和雨天三种类型。

3.3 功率预测结果与分析

本文在考虑上述三种不同气象类型的情况下,

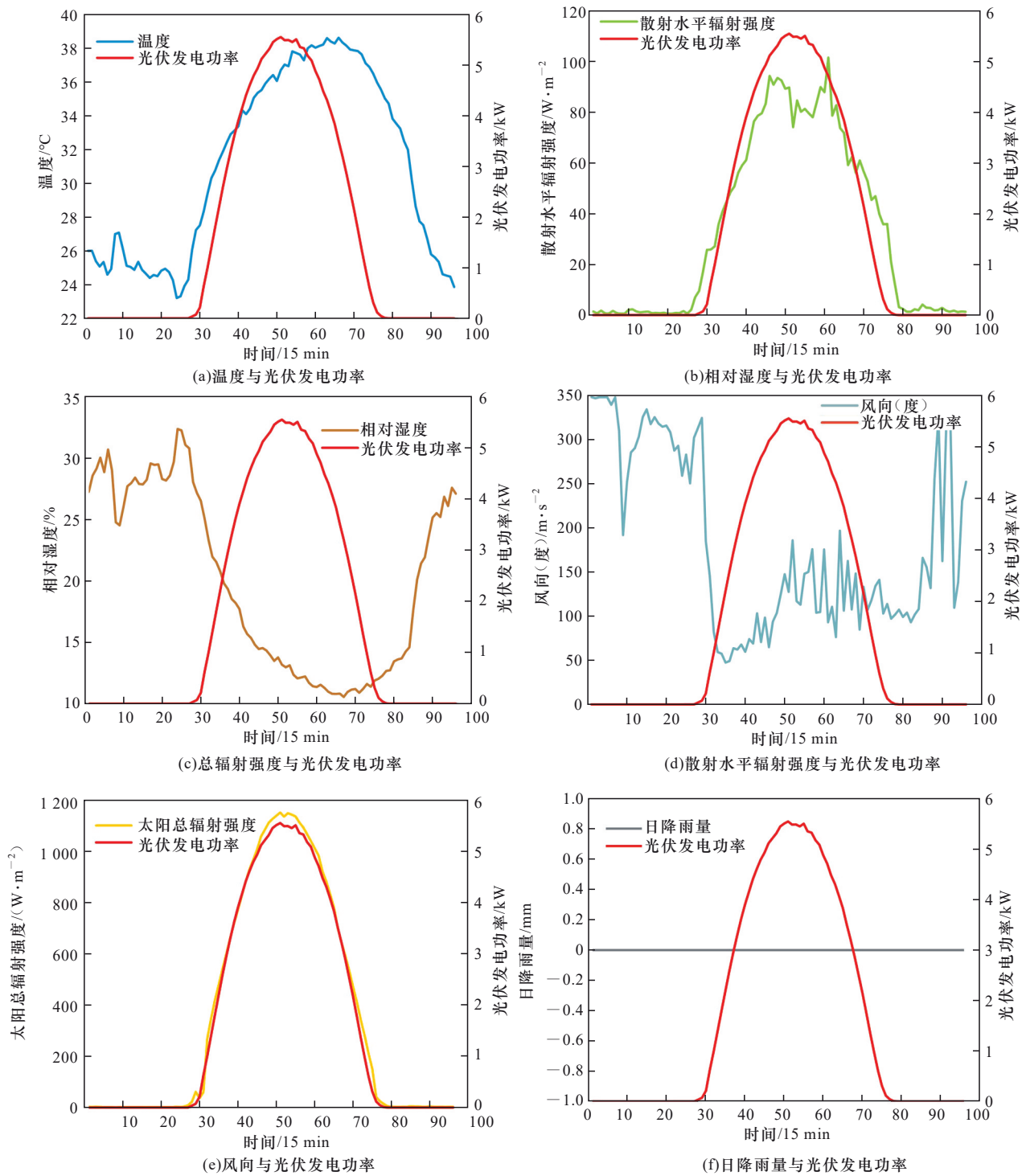


图 2 主要气象特征与光伏发电功率的变化关系

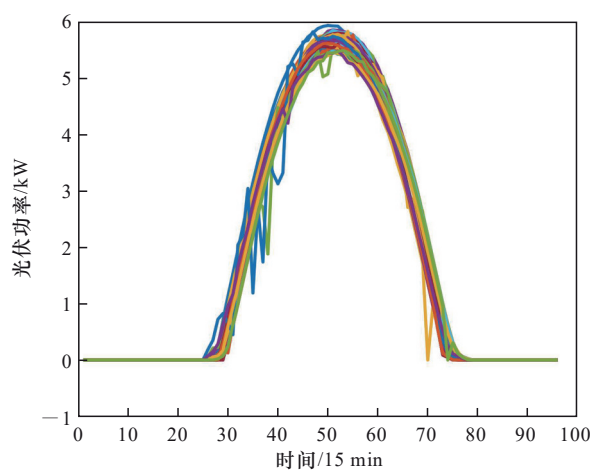
Fig.2 Variations of major meteorological features with PV power generation

分别使用 GPR、LOF-GPR、PSO-LOF-GPR、GA-LOF-GPR 等预测模型对光伏出力数据进行预测。预测结果如图 4 所示。

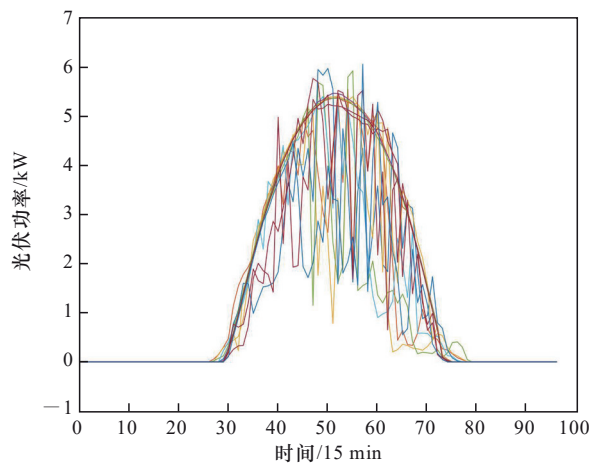
由图 4 可以看出,在晴天、多云和雨天这三种气象类型下,各预测模型生成的预测值均与真实值保持高度一致。这一结果表明,所采用的预测模型对于不同气象状况均具有较好的泛化能力和准确性,

能够为实际应用提供可靠的预测结果。具体的预测精度通过预测误差的大小来体现。表 2 为不同气象类型下,各模型的预测误差对比。

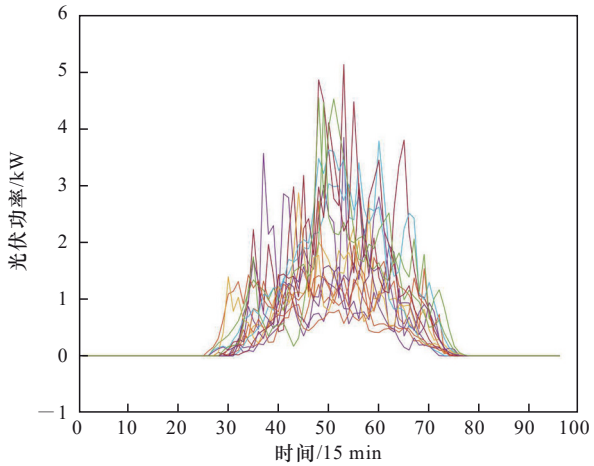
由表 2 可知,晴天、多云和雨天这三种典型气象类型下,本文所提的 GA-LOF-GPR 模型的预测误差均为最小,这表明本文所提模型在这三种场景中有着更高的预测精度。具体来说,晴天的 LOF-



(a)第一类聚类曲线



(b)第二类聚类曲线

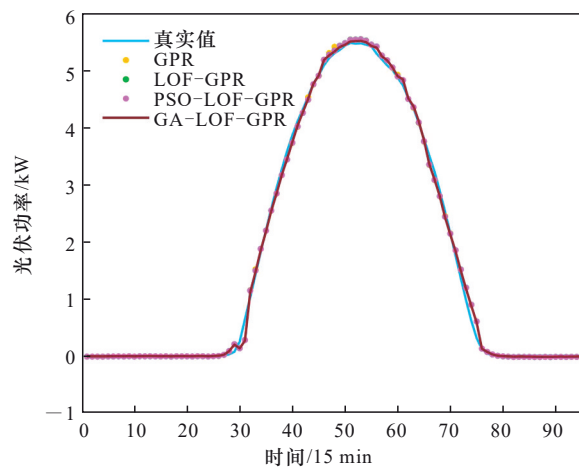


(c)第三类聚类曲线

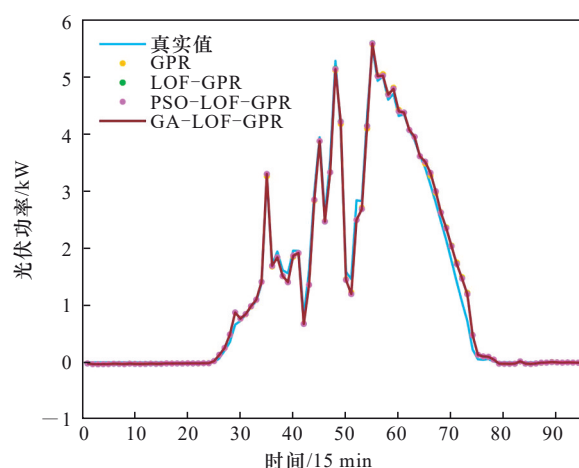
图3 特征权重K均值聚类曲线

Fig.3 Feature weight K-means clustering curves

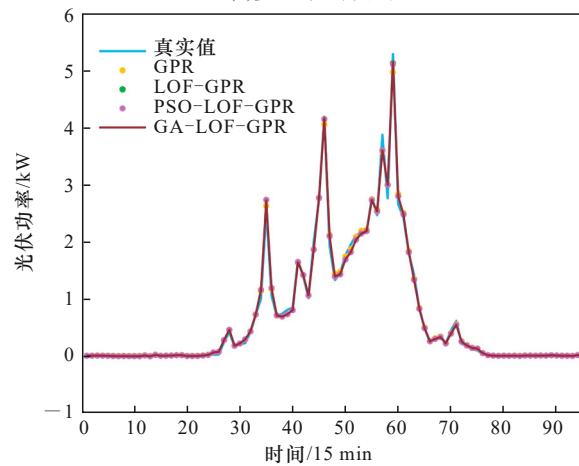
GPR模型相比于GPR模型,其误差明显减小, RMSE和MAE分别减小了1.2%和1.5%;GA-LOF-GPR模型在LOF-GPR模型的基础上, RMSE和MAE又分别减少了7.3%和14.5%。多云时, LOF-GPR模型相比于GPR模型, RMSE和MAE分别减小了1.8%和0.6%;GA-LOF-GPR



(a)晴天的预测结果



(b)多云的预测结果



(c)雨天的预测结果

图4 三种气象类型的预测结果

Fig.4 Prediction results for three weather types

模型在LOF-GPR模型的基础上, RMSE和MAE又分别减少了0.5%和0.6%。雨天时, LOF-GPR模型相比于GPR模型,其误差明显减小, RMSE和MAE分别减小了8.5%和0.7%;GA-LOF-GPR模型在LOF-GPR模型的基础上, RMSE和MAE又分别减少了12.1%和0.7%。

表 2 各模型预测误差
Table 2 Prediction errors

气象类型	预测模型	RMSE	MAE
晴天	GPR	0.085 0	0.053 8
	LOF-GPR	0.084 0	0.053 0
	PSO-LOF-GPR	0.083 2	0.052 1
	GA-LOF-GPR	0.077 9	0.046 0
多云	GPR	0.125 2	0.078 9
	LOF-GPR	0.122 9	0.078 4
	PSO-LOF-GPR	0.122 5	0.078 1
	GA-LOF-GPR	0.122 3	0.077 9
雨天	GPR	0.072 3	0.042 3
	LOF-GPR	0.066 1	0.042 0
	PSO-LOF-GPR	0.066 6	0.042 1
	GA-LOF-GPR	0.065 3	0.041 7

以上数据分析表明,结合了局部离群因子算法的加权高斯过程回归预测模型,能够减小异常值对预测结果的影响,提高预测精度。其中,对晴天的预测精度提高最为明显,原因是晴天的气象特征与光伏发电功率的变化关系规律性最强,对其异常数据的处理更能保证其规律性。由于多云气象特征与晴天和雨雪天气气象特征有部分相似,部分不同,其气象特征复杂多变,模型的预测难度更高,因此基于本文选择超参数优化的预测模型相对于原始模型的误差减小幅度较少。另外,本文所提的 GA-

LOF-GPR 模型与用共轭梯度法优化 GA-LOF-GPR 模型的超参数相比,误差更小,且不依赖于初值的设置;与 PSO-LOF-GPR 模型相比,其确保了更小的误差,即本文所提模型能够更容易找到更优的超参数,得到更好的预测结果。

4 结 语

本文提出了一种光伏功率预测的组合模型,该模型结合了相关性系数和特征 K 均值聚类,并构建了基于遗传算法、局部离群因子算法和高斯过程回归的 GA-LOF-GPR 模型。通过算例分析,得出以下结论:

- 1) 基于特征权重的 K 均值聚类能够很好地根据气象特征划分气象类型,有利于提高预测精度。
- 2) 采用局部离群因子算法生成权值后构建的加权高斯过程回归模型能够有效减小异常值对预测结果影响,尤其是对于晴天这种气象和功率数据规律变化的情况,该模型能显著提升预测精度。
- 3) 利用遗传算法优化超参数,能够很好地降低梯度优化法初值设置不合理导致的预测精度下降问题,避免模型陷入局部最优。同时,通过与其他超参数优化算法的对比,证明了遗传算法优化 LOF-GPR 模型的有效性。

参考文献:

[1] KHARE V, CHATURVEDI P, MISHRA M. Solar energy system concept change from trending technology: A comprehensive review[J]. E-Prime-Advances in Electrical Engineering, Electronics and Energy, 2023, 4: 100183.

[2] 吴硕. 光伏发电系统功率预测方法研究综述[J]. 热能动力工程, 2021, 36(8): 1-7.

WU S. Review of power forecasting methods of photovoltaic power generation system[J]. Journal of Engineering for Thermal Energy and Power, 2021, 36(8): 1-7.

[3] 刘洁, 林舜江, 梁炜焜, 等. 基于高阶马尔可夫链和高斯混合模型的光伏出力短期概率预测[J]. 电网技术, 2023, 47(1): 266-274.

LIU J, LIN S J, LIANG W K, et al. Short-term probabilistic forecast for power output of photovoltaic station based on high order Markov chain and Gaussian mixture model[J]. Power System Technology, 2023, 47(1): 266-274.

[4] 朱想, 周海, 朱婷婷, 等. 光伏系统中地基云图的预处理[J]. 电力系统自动化, 2018, 42(6): 140-145.

ZHU X, ZHOU H, ZHU T T, et al. Pre-processing of ground-based cloud images in photovoltaic system[J]. Automation of Electric Power Systems, 2018, 42(6): 140-145.

[5] 熊音笛, 刘开培, 秦亮, 等. 基于时序数据动态天气划分的短期风电功率预测方法[J]. 电网技术, 2019, 43(9): 3353-3359.

XIONG Y D, LIU K P, QIN L, et al. Short-term wind power prediction method based on dynamic wind power weather division of time sequence data[J]. Power System Technology, 2019, 43(9): 3353-3359.

[6] XU A R, YANG T, JI J W, et al. Application of fuzzy clustering algorithm in the evaluation of abandoned wind power [C]// Chinese Control and Decision Conference (CCDC), Shenyang, China, 2018: 5256-5259.

[7] XU Q Y, HE D W, ZHANG N, et al. A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining[J]. IEEE Transactions on Sustainable Energy, 2015, 6(4): 1283-1291.

[8] 杨锡运, 李艳军, 柏永华, 等. 基于聚类集成的 LSSVM-Adaboost 模型的短期光伏功率预测[J/OL]. 华北电力大学学报(自然科学版), 1-12[2024-07-12].

YANG X Y, LI Y J, BAI Y H, et al. Short-term Photovoltaic power prediction based on LSSVM-Adaboost model with clus-

- tering ensemble[J/OL]. Journal of North China Electric Power University (Natural Science Edition), 1-12[2024-07-12].
- [9] 周育才,肖添,谢七月,等. 基于聚类的HPO-BILSTM光伏功率短期预测[J]. 太阳能学报, 2024, 45(4): 512-518.
ZHOU Y C, XIAO T, XIE Q Y, et al. Clustering-based HPO-BILSTM short-term prediction of PV power [J]. Acta Energi-
ae Solaris Sinica, 2024, 45(4): 512-518.
- [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 Z, ZHANG J, XU R S, et al. Short term photovoltaic power prediction based on similar day clustering and PCC-VMD-
SSA-KELM model[J]. Acta Energi-ae Solaris Sinica, 2024, 45(2): 460-468.
- [11] 黄牧涛,邢芳菲,陈兴邦,等. 基于K-means 聚类和极限学习机组合算法的短期光伏功率预测[J]. 水电能源科学, 2024, 42
(2): 217-220, 216.
HUANG M T, XING F F, CHEN X B, et al. Short-term PV power prediction based on K-means clustering and extreme learn-
ing machine combination algorithm[J]. Water Resources and Power, 2024, 42(2): 217-220, 216.
- [12] 何志昆,刘光斌,赵曦晶,等. 高斯过程回归方法综述[J]. 控制与决策, 2013, 28(8): 1121-1129, 1137.
HE Z K, LIU G B, ZHAO X J, et al. Overview of Gaussian process regression[J]. Control and Decision, 2013, 28(8): 1121-
1129, 1137.
- [13] 王宏伟,黄元生,姜雨晴,等. 基于套索算法和高斯过程回归的中长期居民用电量概率预测[J]. 华北电力大学学报(自然科
学版), 2019, 46(6): 27-35.
WANG H W, HUANG Y S, JIANG Y Q, et al. Middle-and-long-term residential probabilistic load forecasting based on LAS-
SO and Gaussian process regression[J]. Journal of North China Electric Power University (Natural Science Edition), 2019, 46
(6): 27-35.
- [14] 李嘉波,魏孟,叶敏,等. 基于高斯过程回归的锂离子电池SOC估计[J]. 储能科学与技术, 2020, 9(1): 131-137.
LI J B, WEI M, YE M, et al. SOC estimation of lithium-ion battery based on Gaussian process regression[J]. Energy Storage
Science and Technology, 2020, 9(1): 131-137.
- [15] 王耀健,顾洁,温洪林,等. 基于在线高斯过程回归的短期风电功率概率预测[J]. 电力系统自动化, 2024, 48(11): 197-207.
WANG Y J, GU J, WEN H L, et al. Short-term wind power probability prediction based on online Gaussian process regression
[J]. Automation of Electric Power Systems, 2024, 48(11): 197-207.
- [16] 邓惟绩,肖辉,李金泽,等. 基于改进长短期记忆网络和高斯过程回归的光伏功率预测方法[J]. 电器与能效管理技术, 2021
(8): 51-57.
DENG W J, XIAO H, LI J Z, et al. Photovoltaic power prediction method based on improved long short term memory network
and Gaussian process regression[J]. Electrical & Energy Management Technology, 2021(8): 51-57.
- [17] 刘哲. 基于模糊聚类和高斯过程回归的光伏功率预测[D]. 广州: 广东工业大学, 2020.
- [18] 钱敏,桂林,连泉轩,等. 面向光纤光栅传感线性拟合度的PSO-GPR算法[J]. 光通信研究, 2024(4): 66-71.
QIAN M, GUI L, LIAN X X, et al. PSO-GPR for linear fit of fiber grating sensing[J]. Study on Optical Communications, 2024
(4): 66-71.
- [19] 纪德洋,金锋,冬雷,等. 基于皮尔逊相关系数的光伏电站数据修复[J]. 中国电机工程学报, 2022, 42(4): 1514-1523.
JI D Y, JIN F, DONG L, et al. Data repairing of photovoltaic power plant based on Pearson correlation coefficient[J]. Proceed-
ings of the CSEE, 2022, 42(4): 1514-1523.
- [20] 黄晓辉,王成,熊李艳,等. 一种集成簇内和簇间距离的加权K-means聚类方法[J]. 计算机学报, 2019, 42(12): 2836-2848.
HUANG X H, WANG C, XIONG L Y, et al. A weighting K-means clustering approach by integrating intra-cluster and inter-
cluster distances[J]. Chinese Journal of Computers, 2019, 42(12): 2836-2848.
- [21] 彭鹏,林达,汪湘晋,等. 基于局部异常因子的锂离子电池储能系统故障诊断[J]. 浙江电力, 2023, 42(5): 11-17.
PENG P, LIN D, WANG X J, et al. Fault diagnosis of lithium-ion battery energy storage systems based on local outlier factor
[J]. Zhejiang Electric Power, 2023, 42(5): 11-17.

(编辑:杨 鹏)