

DOI: 10.19783/j.cnki.pspc.200750

考虑机组动态特性的超短期风电功率预测及不确定性量化分析

黄慧^{1,2}, 贾嵘¹, 师小雨¹, 王颂凯¹

(1. 西安理工大学电气工程学院, 陕西 西安 710048; 2. 华北水利水电大学电力学院, 河南 郑州 450011)

摘要: 针对数据驱动的风电功率预测模型中, 高维异质大数据特征信息挖掘问题, 提出考虑机组动态特性的轻量梯度上升学习机(LGBM)预测模型和区间估计的不确定性量化方法。首先, 设置发电机转速、叶片角度为机组动态特性指标, 构建 LGBM 超短期风电功率预测模型。其次, 采用模糊 C 均值聚类对历史预测出力和预测误差样本进行区间划分; 考虑预测出力和预测误差条件相依性, 采用非参数估计拟合误差概率分布, 并以置信区间对风电功率预测区间进行了离散化表征。最后, 选取实际风电场数据进行验证。结果表明: 考虑机组动态特性的 LGBM 预测模型的精度和计算效率显著提升; 基于区间估计的不确定性量化方法解耦拟合过程与预测方法, 可靠性高, 灵活性强。

关键词: 特征选择; 轻量梯度上升学习机; 风电功率预测; 区间估计

Ultrashort-term wind power prediction considering the dynamic characteristics of a unit and uncertainty quantitative analysis

HUANG Hui^{1,2}, JIA Rong¹, SHI Xiaoyu¹, WANG Songkai¹

(1. College of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China; 2. College of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou 450011, China)

Abstract: To deeply mine the information of features in a data-driven model with high-dimensional heterogeneous data inputs, a Light Gradient Boosting Machine (LGBM) prediction model is proposed which considers the dynamic characteristics of the unit and the uncertainty quantitatively based on interval estimation. First, taking the generator rotor speed and pitch angle as the indicators of the dynamic status of the wind turbine generator, the LGBM algorithm is explored to build the prediction model. Secondly, the historical predicted power and error are divided into different groups using fuzzy C-means clustering. Then considering the conditional dependence of prediction output and error, the error probability distribution is established by nonparametric estimation, and the confidence interval is used to discretize the prediction interval of wind power. Last, a case study from an actual wind farm is conducted to test the proposed study framework. The results show that the prediction accuracy and calculation efficiency of the LGBM model considering the dynamic characteristics of unit are significantly improved, and the uncertainty quantification method based on interval estimation decouples the fitting process and prediction method. This has high reliability and flexibility.

This work is supported by the National Natural Science Foundation of China (No. 51779206) and Key Scientific and Technological Project in Henan Province (No. 162102210236).

Key words: feature selection; light gradient boosting machine; wind power forecasting; interval estimation

0 引言

风资源的随机性和波动性是造成风电机组出力

不确定性的主要因素, 进而限制大规模风电并网。风电功率预测方法按时间尺度可分为超短期(0~4 h)、短期(4~72 h)、中期和长期预测(> 72 h)^[1-2]。风电功率预测时间尺度越短, 对预测精度要求越高。精确的超短期风电功率预测对风电机组参与优化调频、旋转备用容量的优化配置和实时电力市场确定

基金项目: 国家自然科学基金项目资助(51779206); 河南省科技攻关项目资助(162102210236)

清算价格提供重要依据^[3]。

近年来,神经网络^[4-5]和回归模型^[6]的人工智能方法及非参数的概率预测方法^[7-8],在风电功率预测研究中被广泛关注,主要包括输入数据的预处理、预测模型的构建和预测不确定性的量化分析三个方面研究。特征选取作为数据预处理的方法之一,在保证预测精度前提下,降低输入变量的维度,以提高模型计算效率。风电功率预测的输入特征依据建模对象(风电场或风电机组)的不同有所侧重。文献[9]利用模糊神经网络处理采集数据的不确定性,湿度、温度、压力和风速作为输入特征,提出粒子群优化的模糊神经网络模型对整个风电场功率进行预测。文献[10]选取风向,偏航角和风速作为预测机组出力曲线的输入,以提高预测的精度,但忽略发电机运行的动态控制因素影响。

相比于传统的神经网络和回归模型训练高维大数据时计算成本较高和模型不稳定的问题,新兴的基于决策树的集成学习方法,通过构建和组合多个基础学习器来完成学习任务,能够提供比单个学习器更健壮的性能^[11]。基于 Boosting 思想的 Adaboost, GBRT 和 Xgboost 方法在风速预测^[12]、光伏发电预测^[13]和径流预测^[14]得到应用。轻量梯度上升学习机(LGBM)是基于上述 Boosting 方法进一步改进的算法,在不失去预测精度的前提下,其计算速度和内存消耗大大降低,针对于高维的大数据具有较高的预测精度、较快的计算速度和较强的最小化过拟合等优点^[15]。文献[16]提出利用卷积神经网络和 LGBM 构造超短期功率预测模型进一步提升预测精度和鲁棒性,模型最优参数采用传统的网格搜索获取,其计算时间较长。

目前,风电功率不确定性量化分析的研究主要有概率密度、分位数回归和区间估计三个方面^[17]。文献[18]采用分位数回归进行风电功率的概率预测,分位数回归不能描述联系连续的概率分布,难以描述风速等气象变量与风电功率间的非线性映射,实际预测效果受限。文献[19]对三个确定性预测值使用加权相加,求取风电功率预测值,基于预测值的方差,获取一定置信度下的预测区间,但仅根据预测方差获取区间估计,忽略了预测误差和机组出力的条件相依性。条件相依性指风电功率预测误差的分布与风速、风向和确定性预测结果等因素相关。条件相依性质的区间估计,即需要将历史预测功率值作为原始样本集进行区间划分,形成多个子样本集,提高分布函数统计学显著性,还需要保证子集内样本的数量,避免样本数量过少导致概率密度函数拟合不稳定^[20]。

上述研究针对超短期风电功率预测和区间估计有一定的指导意义,但在数据特征信息挖掘和模型性能方面还有进一步提升空间。因此,本文考虑发电机转速和叶片角度对发电机组实时运行动态的影响,建立贝叶斯超参数优化的轻量梯度提升机确定性风电功率预测模型。然后采用模糊 C 均值聚类对历史预测风电功率值和预测误差进行工况区段划分,从预测误差和预测出力的条件相依性出发,构建机组预测出力和预测误差耦合的非参数区间估计方法,并以置信区间对风电功率预测区间进行了离散化表征,实现风电机组出力预测的不确定性的量化。

1 超短期风电功率预测及区间估计的思路

本文研究思路如图 1 所示。采集机组运行的实时动态数据,基于皮尔逊相关性和模型特征重要度排序分析影响机组出力的重要特征。选取叶片角度、发电机平均转速和风速作输入,机组出力为输出,采用贝叶斯优化的轻量梯度上升学习机模型训练历史数据获取机组出力特性曲线,并在测试数据集上测试模型的预测精度和计算性能。其次,利用非参数估计建立预测出力和预测误差联合概率密度分布,分析预测出力和预测误差的条件相依性,采用模糊 C 均值聚类对预测出力进行功率区段划分,比较非参数估计和参数估计方法拟合的误差概率分布,选取不同出力特性下的最优的误差分布模型。判定未来时刻预测值归属的出力特性,通过相应的误差分布即可求取一定置信水平下的预测区间。

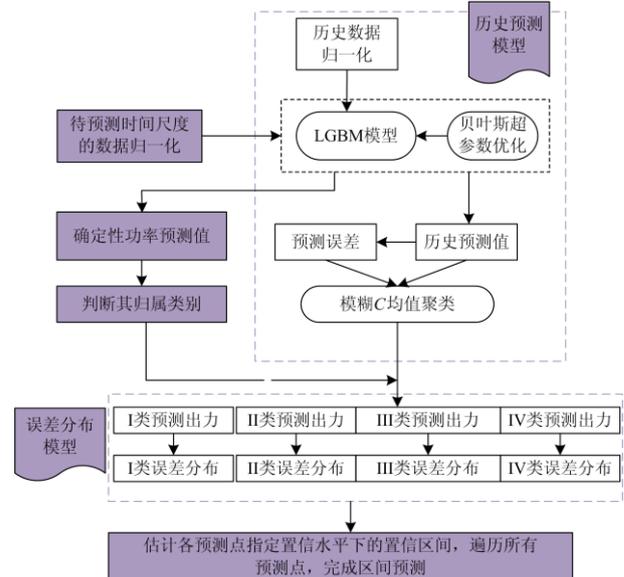


图 1 超短期风电功率预测及不确定性量化分析思路

Fig. 1 Framework of ultrashort-term wind power forecasting and uncertainty quantitative analysis

2 贝叶斯超参数优化的 LGBM 预测模型

2.1 LightGBM 模型(LGBM)

LightGBM(Light Gradient Boosting Machine)是基于直方图的决策树梯度提升算法,针对风电功率预测高维异质性复杂的输入特征,给定训练数据和损失函数形式后,学习模型定义为一个损失函数极小化的问题,具有较快的训练速度和更高的效率^[21-22]。

给定一组风电机组历史时间序列数据, $D = \{(x_i, y_i) : i = 1, \dots, N\}$, x_i 为输入特征时间序列, y_i 为风电机组出力功率序列, N 为样本个数。损失函数定义为 $L(y, f(x)) = (y - f(x))^2$, 优化的目标函数为 $\arg \min \sum_{i=1}^N L(y_i, f(x_i))$ 。算法实现步骤如下:

1) 输入训练数据 (x_i, y_i) ;

2) 构建提升树模型 $f_m(x) = \sum_{m=1}^M \beta_m T(x, \Theta_m)$;

3) 初始化 $f_0(x) = \arg \min_{\Theta} \sum_{i=1}^N L(y_i, \Theta)$;

For $m = 1$ to M

(1) 对于第 m 弱学习器, 首先计算梯度 $g_m(x_i) =$

$$\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} ;$$

(2) 根据梯度学习第 m 弱学习器 $\Theta_m =$

$$\arg \min_{\Theta, \beta} \sum_{i=1}^N [-g_m(x_i) - \beta T(x_i; \Theta)]^2 ;$$

(3) 通过 line search 求取最佳步长 $\beta_m =$

$$\arg \min_{\beta} \sum_{i=1}^N L[y_i, f_{m-1}(x_i) + \beta T(x_i; \Theta)] ;$$

(4) 更新模型, $f_m(x) = f_{m-1}(x) + \beta_m T(x; \Theta_m)$;

结束, 输出 $f_m(x)$ 。

2.2 贝叶斯超参数优化的 LGBM 模型

机器学习模型的超参数决定其模型的性能,提升机器学习的性能需要对模型参数进行优化。经典的优化算法如粒子群算法,模拟退火,网格搜索等因需要昂贵的计算成本已不再适用参数规模较大的复杂的机器学习方法^[23]。

本文采用计算效率较高且易实现的贝叶斯优化算法对 LGBM 模型的超参数进行优化。首先定义需要优化的超参数空间 X , 如树的叶子节点数,学习深度等。超参数 $x \in X$ 的函数 $f: X \rightarrow \mathfrak{R}$ 。超参数优化问题转化为寻求 $x^* \in \arg \min_{x \in X} f(x)$ 的过程。目标函数 $f(x)$ 未知, 可以从考虑误差的观察函数

$y(x) = f(x) + \varepsilon$, $\varepsilon \in N(0, \delta_{\text{noise}}^2)$ 中获得。已知观测点数据 $D = \{(x_0, y_0), \dots, (x_{i-1}, y_{i-1})\}$, 构造未知目标函数的状态假设分布 $P(f|D)$ 。然后基于当前模型的分布采用采集函数 $\alpha: X \rightarrow \mathfrak{R}$ 进一步对模型进行挖掘和评估^[24]。基于贝叶斯优化方法优化 LGBM 超参数的流程如图 2 所示。

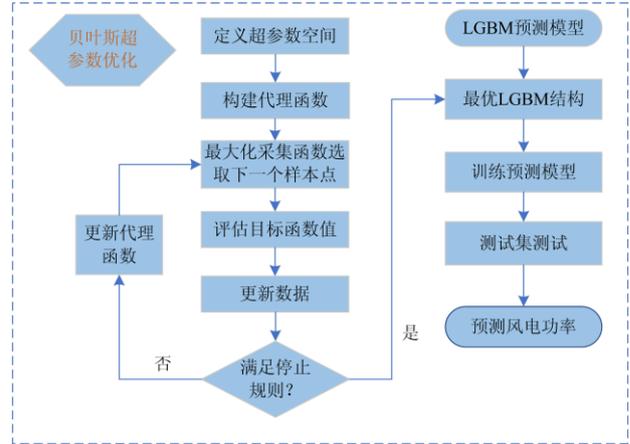


图 2 贝叶斯优化的 LGBM 预测模型流程图

Fig. 2 Flowchart of BO-LGBM model

2.3 评价指标

本文确定性预测采用具有普遍性的均方根误差 (RMSE) 和平均绝对误差 (MAE) 作为评价指标。评价指标值越小, 模型的预测效果越好。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \hat{P}_i)^2} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - \hat{P}_i| \quad (2)$$

式中: P_i 和 \hat{P}_i 分别为第 i 预测点的风电功率实际值和预测值; N 为测试样本点的个数。

3 基于区间估计的不确定性量化分析

3.1 模糊 C 均值聚类

考虑预测误差与预测功率条件相依性的区间估计, 需要将历史预测功率值作为样本集, 并将预测功率值离散划分为多个功率区段, 获取各样本子集误差分布函数, 提高误差分布函数统计学的显著性。本文采用模糊 C 均值聚类^[25]对历史预测风电功率进行聚类, 形成不同的样本子集。模糊 C 均值聚类考虑被划分对象属于每个类别的隶属度问题, 没有规定严格的划分界限, 相对于 K-MEANS 聚类的硬性划分规定, 其应用更为灵活。为评价模糊 C 均值聚类结果的好坏和确定最佳聚类数目, 引入轮廓系数 K_{PC} 和分类熵 K_{CE} 两个评价指标。 K_{PC} 用于评价不同

样本类分群间的分离程度, 取值越大越好; K_{CE} 用于评价样本聚类分群间的模糊程度, 取值越小越好。

$$\begin{cases} K_{PC} = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (u_{ij})^2 \\ K_{CE} = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (u_{ij}) \log(u_{ij}) \end{cases} \quad (3)$$

式中: u_{ij} 表示第 j 样本关于 i 聚类中心的隶属度; c 为第 i 类样本的个数; N 为每类的样本个数。

3.2 预测区间估计

已知不同功率区段内的样本子集, 非参数估计拟合误差直方图获取其概率密度函数(PDF)和预测误差的累积分布函数(CDF)。假设预测误差的概率密度函数为 $F(e)$, e 为随机预测误差值, 其逆函数为 $G(\varepsilon)$ 反映预测区间的边界值。即当 $P\{e \leq G(\varepsilon)\} = 1 - \alpha$ 成立, 则预测误差 $(1 - \alpha)$ 置信度水平下的预测功率的上限 $P_{f,\max}$ 和下限 $P_{f,\min}$ 。

$$\begin{cases} P_{f,\max} = P_f + G(\alpha_1) \\ P_{f,\min} = P_f + G(\alpha_2) \end{cases} \quad (4)$$

式中: $\alpha_2 - \alpha_1 = 1 - \alpha$; $\alpha_1 = \frac{\alpha}{2}$, $\alpha_2 = 1 - \frac{\alpha}{2}$; P_f 为预测功率值。

已知待预测时刻的风电功率点预测值, 预测误差未知, 需要判定待预测点风电功率所属功率区段, 依据相应的误差概率分布图, 求取不同置信水平下待预测点的区间估计, 完成功率预测不确定性的量化分析。

3.3 评价指标

区间估计效果采用表征预测区间覆盖率的可靠性、预测区间平均宽度、区间覆盖率三个指标^[26]。

1) 可靠性是评判置信区间的可信程度的评价指标。绝对值越小, 预测区间可信程度越高, 说明预测效果越好。

$$R^{(1-\alpha)} = \left[\frac{\omega^{(1-\alpha)}}{N} - (1-\alpha) \right] \times 100\% \quad (5)$$

式中: $R^{(1-\alpha)}$ 为置信度 $(1 - \alpha)$ 下的可靠性指标值; N 为预测测试样本点的个数; $\omega^{(1-\alpha)}$ 为置信度 $(1 - \alpha)$ 下实际功率值落入预测置信区间的个数。

2) 区间平均宽度是衡量预测有效性的评价指标, 反映预测结果中涵盖不确定信息的能力。保证可靠性的前提下, 区间宽度越小越好。

$$I^{(1-\alpha)} = \frac{1}{N} \sum_{i=1}^N \lambda_i^{(1-\alpha)} \quad (6)$$

式中: $I^{(1-\alpha)}$ 为置信度 $(1 - \alpha)$ 下的风电功率预测区间平均宽度; N 为样本的个数; $\lambda_i^{(1-\alpha)}$ 是第 i 样本在

置信度 $(1 - \alpha)$ 下功率预测区间上界和下界之差。

3) 区间覆盖率是描述区间预测覆盖实际值的情况。覆盖率大于指定置信度时, 预测达到预期效果, 且覆盖率越大, 预测效果越好。

$$\delta^{(1-\alpha)}_{PI} = \frac{1}{N} \sum_{i=1}^N C_i \quad (7)$$

式中: N 为样本的个数; C_i 为覆盖因子, 若第 i 实际功率落在预测区间内, 则 C_i 取 1, 否则取 0。

4 实例验证

实验采用中国山东某一风电场风电机组在 2017 年至 2018 年实测运行数据, 包括在线实测的机组风速、叶片角度, 发电机转子转速和风电功率等历史数据。数据采样时间间隔为 10 min, 单台机组额定功率为 2 MW。考虑机组运行过程中的损耗, 时间跨度过长的观测数据, 并不能真实反映机组的运行工况。最终选取某一台运行状态完好的风机运行一个月的完整数据。样本数据包含 4 464 个数据点, 选取样本数据集的 80% 作为训练集, 剩余 20% 作为测试集, 预测的时间跨度为 4 h。

4.1 机组出力特征分析

所选时间范围内, 机组没有出现故障停机和被限电停机的情况, 所以, 为保证算法仿真的真实有效性, 机组出力小于 0 的数据置 0, 但不剔除。数据按照最大最小化归至 $[0, 1]$ 。依据风力发电原理, 风电功率主要有风速决定, 而又受发电机转速制约。风速是不可控的, 为了保证风力发电机的正常和安全运行, 发电机转速需要控制在一定范围之内, 发电机的转速变化间接反映了机组出力的动态运行特性。

选取 SCADA 系统采集历史数据包括: 环境温度(f_0)、1 s 平均风速(f_1)、1 s 最大风速(f_2)、1 s 最小风速(f_3)、风向(f_4)、机舱位置(f_5)、扭矩(f_6)、叶片角度(f_7)、叶片最大角度(f_8)和叶片最小角度(f_9)发电机平均转速(f_{10})、发电机最大转速(f_{11})、发电机最小转速(f_{12})、风电功率(f_{13})共 14 个特征量, 分别采用皮尔逊相关系数和基于模型的特征重要度排序分析影响机组出力特性的特征。图 3(a)皮尔逊相关系数热力图中蓝色部分表示强相关, 风速、发电机转速和叶片角度与机组出力呈现强相关性。基于模型的特征排序图 3(b)中发电机平均转速 f_{10} 的 F-score 最高, 意味着该特征对预测功率模型相关性最强, 其次为叶片角度 f_7 和风速 f_1 特征。为提高模型的计算效率, 剔除相关性较弱和 F-score 分数较低的特征, 选取发电机平均转速、叶片角度和风速作为风电机组预测模型的输入特征。

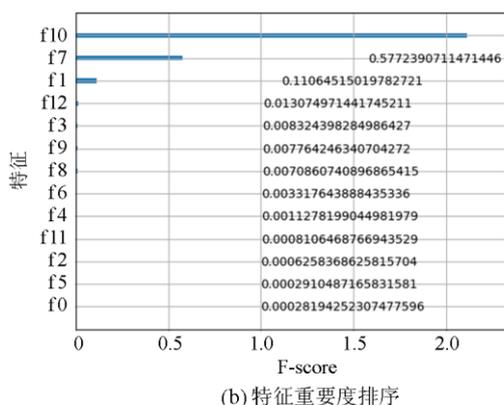
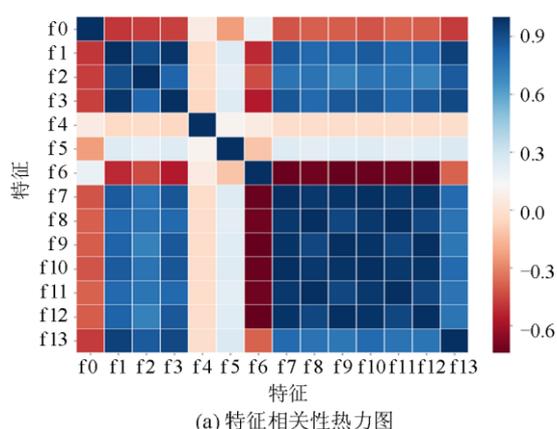


图 3 输入特征相关性热力图和重要度排序

Fig. 3 Correlation matrix between features and diagram of feature importance scores

4.2 历史数据确定性预测结果

超参数优化有利于提高预测模型的准确性和控制过拟合。原始数据输入贝叶斯优化模型, 定义 LGBM 模型的超参数如表 1 所示。选取 RMSE 为评估函数, 采用 5 折交叉验证训练模型, 共迭代 30 次, 选取负 RMSE 最大对应的参数, 即为最优。

表 1 LGBM 模型的最优参数

参数类别	参数
弱学习器数量	74
列采样比例	0.8
树的深度	7
叶子节点最小数据	10
叶子节点最小权重	3
叶子数量	99
正则化 L1	0.91
正则化 L2	0.98
样本采样比例	0.7
学习率	0.3

贝叶斯优化的超参数代入 LGBM 模型, 在训练数据集上进行模型训练, 测试集进行测试, 并与随机森林(RF)、支持向量机回归(SVR)、三层感知神经网络(MLP)预测结果进行比较。RF 和 SVR 的参数采用贝叶斯优化获取, MLP 隐藏层神经元个数经多次测试后选取为 6。选取控制用叶片角度、发电机平均转速和风速为 I 类输入特征, 风速输入为 II 类特征。针对所选取的 LGBM、RF、SVR 和 MLP 四种预测模型, 测试集上两类输入特征对应的预测误差对比结果如表 2 所示。模型输入 I 类特征时, LGBM 的 RMSE 和 MAE 值最小。RF 方法仅次于 LGBM 方法, 其次为 MLP 和 SVR 方法。LGBM 的训练时间最小为 0.045 s, SVR 的训练时间 0.057 s 比 RF 时间 0.088 s 短, 但其误差值比 RF 大。MLP 的预测误差比 SVR 方法大, 但其计算时间要比 SVR 方法增加 3 倍。相同输入特征, LGBM 预测误差和计算成本都要优于 RF、SVR 和 MLP 预测模型。相对于 II 类特征输入的 LGBM 预测模型, 其 RMSE 值减小了 53.7%, MAE 值减少了 58.9%, 计算时间增加了 0.006 s。因此随着输入数据的增加, LGBM 的预测性能和计算效率依然表现优越。

表 2 历史预测误差结果对比

输入特征	LGBM	RF	SVR	MLP	评价指标
I 类特征	31.6	40.2	149.5	62.1	RMSE/kW
	16.1	20.1	120.6	27	MAE/kW
	0.045	0.088	0.057	0.29	Time/s
II 类特征	68.8	80.3	116	189.4	RMSE/kW
	39.2	44.7	84.5	143.3	MAE/kW
	0.039	0.059	0.055	0.13	Time/s

4.3 预测误差分布模型

采用核密度估计建立预测出力 and 误差的联合概率密度分布如图 4 所示, 图 4(a) 中历史预测出力-误差联合概率呈现多峰分布, 单一的概率密度分布获取未来时刻预测的区间不够准确。图 4(b) 为历史预测出力数据的分布和预测误差的分布, 预测出力在两端较为集中, 中间较为分散, 机组出力较大时误差较小。机组出力较小时, 出现大误差概率较大。个别误差较大值可能是由于原始数据中存在异常点未剔除产生。根据风电机组的预测出力和预测误差条件相依性, 采用模糊 C 均值聚类对预测出力样本进行聚类。由聚类的评判准则, 聚类数目取 3 类时, K_{PC} 和 K_{CE} 值分别为 0.611 和 0.48。聚类为 4 类时 K_{PC} 和 K_{CE} 值分别为 0.79 和 0.37。因此, 风电机组预测出力功率划分为 4 个特性区段, 其范围分别为 [0, 203]、[203, 698]、[698, 1450] 和 [1450, 2023]。

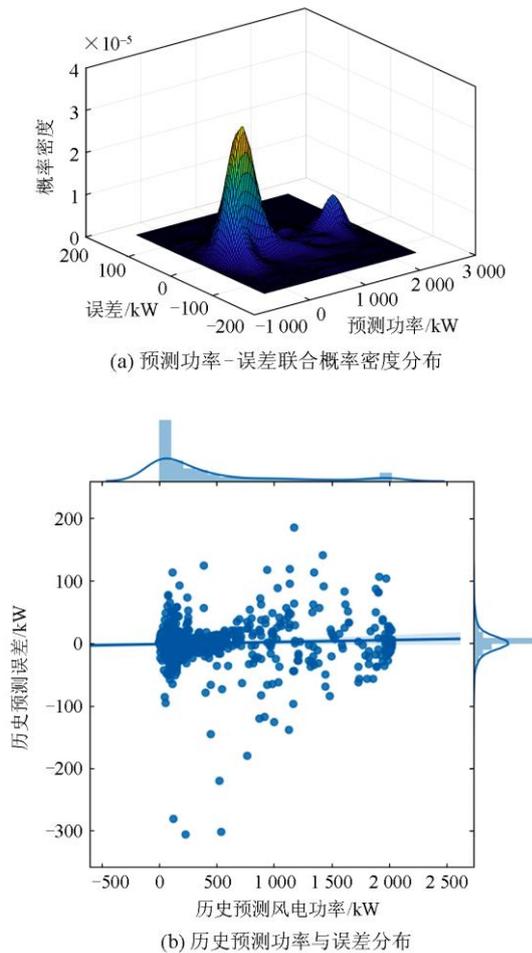


图4 预测功率-预测误差联合概率分布
Fig. 4 Wind power-prediction error joint distribution

预测误差的精确分布拟合是提高风电功率预测区间的可靠性的前提，有利于反映历史预测误差的变化。4类预测误差数据中，误差统计分布去除原始功率和预测功率为0的点。分别采用逻辑分布(Logistic)、非参数分布(Non-parametric)、正态分布(Normal)和 $t(t\text{-Location})$ 分布对4类误差分布直方图进行拟合，误差概率密度分布如图5所示。第I、II类功率区段内的误差波动范围集中在 ± 60 kW和 ± 50 kW，第III类功率区段内误差波动范围较 ± 150 kW，第IV类功率区段内分布明显呈现不对称。四个区段内的误差分布有个别误差较大点，因原始采集的数据中一些数据会出现误差，为保证基于数据驱动的模型的真实性和完整性，原始数据的异常点未被剔除。四类直方图拟合误差分布图可以反映出非参数估计的优越性，尤其是IV类误差呈现非正态分布时，采用非参数核密度估计方法更能精确表示真实误差分布。选取最优非参数估计拟合的误差分布图

进行区间预测。

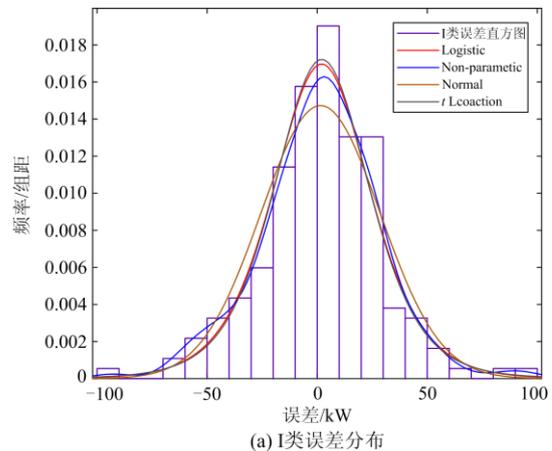
依据式(4)计算4类预测功率区段在指定置信水平下的误差波动范围，并与未进行功率区间划分的误差波动结果对比如表3。同一置信水平下，不同功率段内预测误差的波动范围不同。第III类功率段对应的误差波动范围最大，与第III类出力特性范围跨度最大对应。进行预测出力特性划分的误差波动范围更能反映实际情况。

表3 不同置信水平下误差区间
Table 3 Error interval under different confidence levels

出力特性	边界	95%置信水平	90%置信水平	80%置信水平	60%置信水平
I类	$G(\alpha_1)/\text{kW}$	56.6	45.4	34.1	22.2
	$G(\alpha_2)/\text{kW}$	-52.4	-41.2	-29.9	-18.1
II类	$G(\alpha_1)/\text{kW}$	61.6	41.9	32.2	27.6
	$G(\alpha_2)/\text{kW}$	-34.5	-24.8	-24.1	-12.5
III类	$G(\alpha_1)/\text{kW}$	133.2	116.4	96.2	76.9
	$G(\alpha_2)/\text{kW}$	-132.6	-107.2	-67.8	-49.3
IV类	$G(\alpha_1)/\text{kW}$	37.6	32.2	24.7	18.8
	$G(\alpha_2)/\text{kW}$	-57.4	-43.4	-31.2	-20.7
未分类	$G(\alpha_1)/\text{kW}$	66.1	42.3	28.7	19.5
	$G(\alpha_2)/\text{kW}$	-62.2	-49.3	-33.5	-20.5

4.4 区间估计结果对比分析

由4.2节LGBM、RF、SVR和MLP四类模型测试集的结果对比得知，LGBM的预测精度和计算效率相对突出，因此未来4h多步预测采用贝叶斯超参数优化的LGBM方法，其模型超参数与测试集所用一致，并与MLP、RF和SVR三种模型对比，提前4h LGBM多步预测的RMSE和MAE值为23.1 kW和18.5 kW，RF、SVR和MLP的RMSE分别为26.5 kW、43.5 kW和28.2 kW，MAE值分别为21.2 kW、37.9 kW和24.1 kW。贝叶斯优化超参数优化LGBM预测精度依然优于其他三种比较模型。



(a) I类误差分布

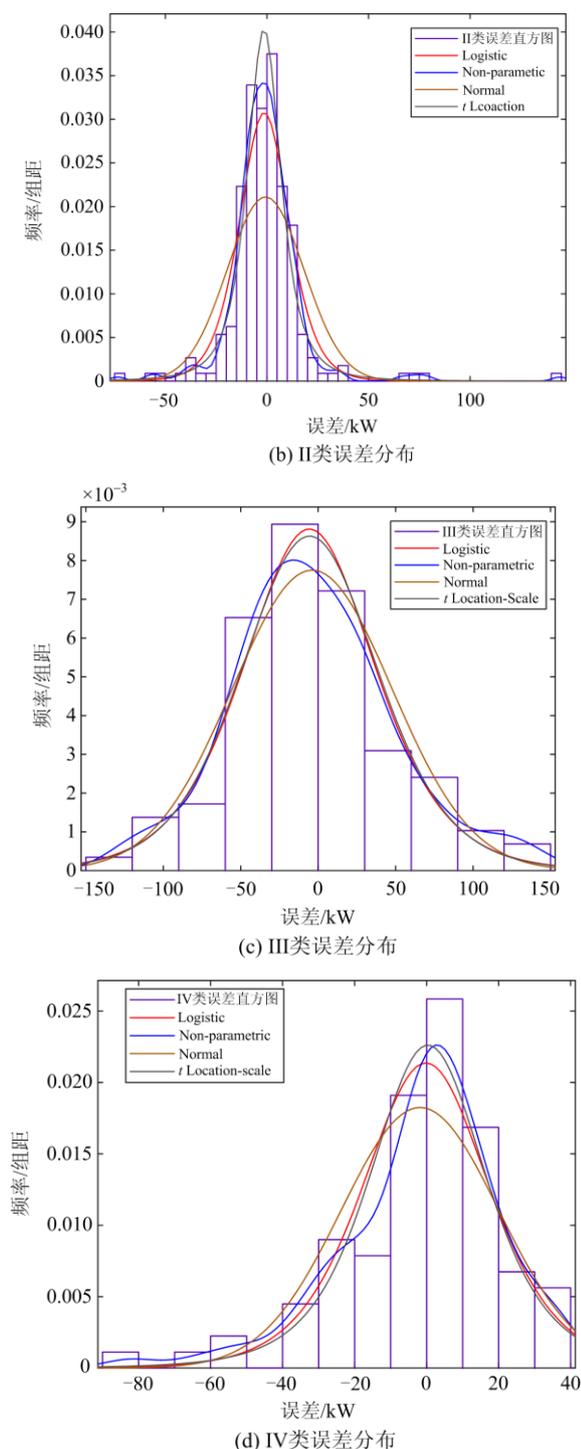


图 5 历史预测误差分布

Fig. 5 Error distribution of historical forecast values

判定待预测点的出力特性, 确定其误差波动范围, 遍历所有预测点值, 获取预测区间。提前 4 个小时的机组出力功率预测区间估计结果如图 6。机组出力较小时, 误差分类获取的区间宽度如图 6(a) 要小于未分类时的区间宽度图 6(b)。机组出力较大

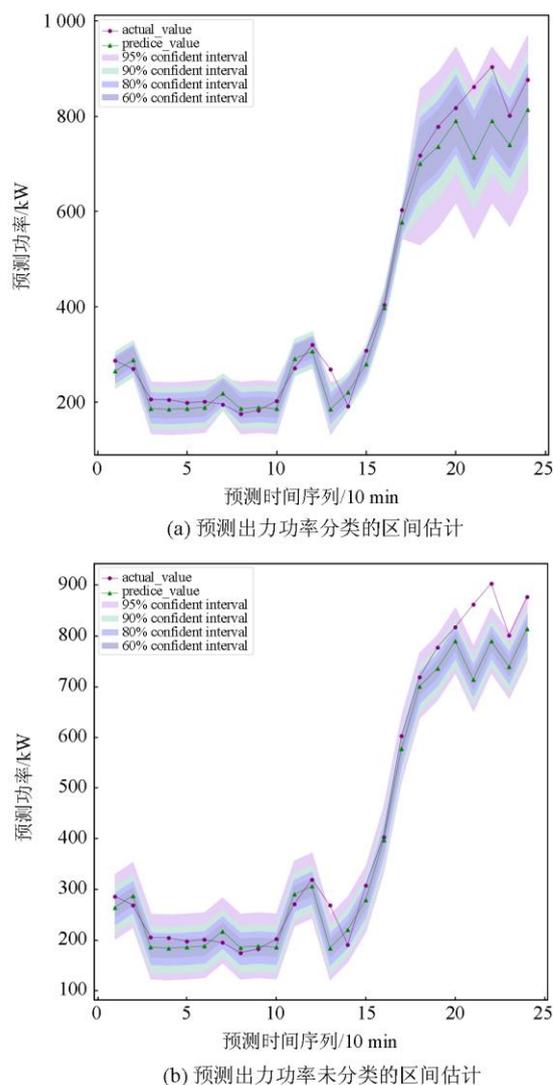


图 6 超短期风电功率预测区间估计结果对比

Fig. 6 Comparison of interval estimation for ultra-short-term wind power prediction

时, 误差分类的区间预测宽度较大, 且区间包络实际值较多。

区间预测评价指标如表 4 所示。指定 95% 置信水平下, 误差分类的预测覆盖率 0.968 大于误差未分类的预测覆盖率 0.951。同样, 其他指定置信度下, 误差分类的预测覆盖率要大于误差未分类的预测覆盖率, 且前者的可靠性指标较低, 说明误差分类的区间估计的可靠性程度较高。相同置信水平下, 误差分类的预测平均区间宽度要高于未分类的预测区间平均宽度, 说明可靠性和区间平均宽度之间的矛盾性。为了保证可靠性, 需要适当增加区间平均宽度。

表 4 风电功率预测区间估计评价指标

Table 4 Evaluation index of wind power interval forecast					
类别	95%置信水平	90%置信水平	80%置信水平	60%置信水平	评价指标
误差未分类	3.5%	2.8%	4.5%	8.81%	可靠性
	0.951	0.895	0.802	0.614	覆盖率
误差分类	128.3	91.6	62.2	40	区间平均宽度/kW
	0.8%	1.6%	3.33%	6.67%	可靠性
误差分类	0.968	0.917	0.835	0.662	覆盖率
	149.8	101.6	90.25	65.26	区间平均宽度/kW

5 结论

数据驱动的风电功率预测建模过程中, 高维异质大数据输入可能会引起模型预测的准确度和可靠性不足。为深度挖掘多源异质数据的特征信息和耦合特性, 寻求高效、稳定且易于实现的预测模型, 本文从影响机组出力的特征选取、确定性预测模型建立和基于区间估计的不确定性分析三个方面进行了分析和实例验证, 并与现有文献方法进行了对比, 得到如下结论:

1) 选取发电机转子转速、风机叶片角度和风速作为预测模型的输入, 深度挖掘了影响风电机组出力的因素。算例表明所选输入特征的有效性, 其预测精度均高于单一风速的输入模型。

2) 基于贝叶斯超参数优化的 LGBM 的预测误差确定性预测模型, 尤其是对于高维异质输入数据, 表现出其优越的计算效率和稳定性。

3) 依据预测误差和机组出力条件相依性, 提出了模糊 C 均值聚类的机组出力工况划分, 构建了非参数估计的误差概率分布模型, 相较于工况未划分的区间估计, 明显提升了预测的可靠性和预测的覆盖率。

本文研究考虑风电机组运行的动态特性及数据驱动下建模复杂化和计算成本问题, 提出了基于梯度上升学习的超短期风电功率预测及基于区间估计的不确定性量化分析方法, 并应用于实际风电场的案例学习, 对后续风电机组参与电网优化调频和维护调停计划具有一定的参考。由于缺乏更多其他区域风电场的实际测量数据, 在未来研究中, 会将所提出的预测框架模型结合其他地区风电场数据加以改进修正, 针对历史数据样本更大和预测时间跨度增加时, 探索 LGBM 确定性预测模型和基于区间估计的不确定性分析的普适性。

参考文献

[1] MA J, YANG M, LIN Y. Ultra-short-term probabilistic

wind turbine power forecast based on empirical dynamic modeling[J]. IEEE Transactions on Sustainable Energy, 2020, 11(2): 906-915.

- [2] 杨茂, 张罗宾. 基于数据驱动的超短期风电功率预测综述[J]. 电力系统保护与控制, 2019, 47(13): 171-186.
YANG Mao, ZHANG Luobin. Review on ultra-short term wind power forecasting based on data-driven approach[J]. Power System Protection and Control, 2019, 47(13): 171-186.
- [3] 薛禹胜, 郁琛, 赵俊华, 等. 关于短期及超短期风电功率预测的评述[J]. 电力系统自动化, 2015, 39(6): 141-151.
XUE Yusheng, YU Chen, ZHAO Junhua, et al. A review on short-term and ultra-short-term wind power prediction[J]. Automation of Electric Power Systems, 2015, 39(6): 141-151.
- [4] 琚珏, 祁林, 刘帅. 基于改进乌鸦算法和 ESN 神经网络的短期风电功率预测[J]. 电力系统保护与控制, 2019, 47(4): 58-64.
JU Yao, QI Lin, LIU Shuai. Short-term wind power forecasting based on improved crow search algorithm and ESN neural network[J]. Power System Protection and Control, 2019, 47(4): 58-64.
- [5] 张颖超, 王雅晨, 邓华, 等. 基于 IAFSA-BPNN 的短期风电功率预测[J]. 电力系统保护与控制, 2017, 45(7): 58-63.
ZHANG Yingchao, WANG Yachen, DENG Hua, et al. IAFSA-BPNN for wind power probabilistic forecasting[J]. Power System Protection and Control, 2017, 45(7): 58-63.
- [6] SUN G, JIANG C, CHENG P, et al. Short-term wind power forecasts by a synthetical similar time series data mining method[J]. Renewable Energy, 2018, 115: 575-584.
- [7] 吴问足, 乔颖, 鲁宗相, 等. 风电功率概率预测方法及展望[J]. 电力系统自动化, 2017, 41(18): 167-175.
WU Wenzu, QIAO Ying, LU Zongxiang, et al. Methods and prospects for probabilistic forecasting of wind power[J]. Automation of Electric Power Systems, 2017, 41(18): 167-175.
- [8] 杨楠, 黄禹, 叶迪, 等. 基于自适应多变量非参数核密度估计的多风电场出力相关性建模[J]. 中国电机工程学报, 2018, 38(13): 3805-3812, 4021.
YANG Nan, HUANG Yu, YE Di, et al. Modeling of output correlation of multiple wind farms based on adaptive multivariable nonparametric kernel density estimation[J]. Proceedings of the CSEE, 2018, 38(13): 3805-3812, 4021.
- [9] SHARIFIAN A, GHADI M J, GHAVIDEL S, et al. A new method based on type-2 fuzzy neural network for accurate wind power forecasting under uncertain data[J].

- Renewable Energy, 2018, 120: 220-230.
- [10] YAN J, ZHANG H, LIU Y, et al. Uncertainty estimation for wind energy conversion by probabilistic wind turbine power curve modeling[J]. Applied Energy, 2019, 239: 1356-1370.
- [11] TORRES-BARRÁN A, ALONSO Á, DORRONSORO J R. Regression tree ensembles for wind energy and solar radiation prediction[J]. Neurocomputing, 2019, 326-327: 151-160.
- [12] LANDRY M, ERLINGER T P, PATSCHKE D, et al. Probabilistic gradient boosting machines for GEFCom2014 wind forecasting[J]. International Journal of Forecasting, 2016, 32(3): 1061-1066.
- [13] PERSSON C, BACHER P, SHIGA T, et al. Multi-site solar power forecasting using gradient boosted regression trees[J]. Solar Energy, 2017, 150: 423-436.
- [14] YU X, WANG Y, WU L, et al. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day stream flow forecasting[J]. Journal of Hydrology, 2020, 58: 124-129.
- [15] HE Jiajun, YU Chuanjin, LI Yongle, et al. Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning[J]. Energy Conversion and Management, 2020, 205: 112-118.
- [16] JU Y, SUN G, CHEN Q, et al. A model combining convolutional neural network and light GBM algorithm for ultra-short-term wind power forecasting[J]. IEEE Access, 2019, 7: 28309-28318.
- [17] QUAN H, KHOSRAVI A, YANG D, et al. A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(11): 4582-4599.
- [18] WAN C, LIN J, WANG J, et al. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation[J]. IEEE Transactions on Power Systems, 2017, 32(4): 2767-2778.
- [19] KIM D, HUR J. Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method[J]. Energy, 2018, 157: 211-226.
- [20] 乔颖, 鲁宗相, 吴问足. 考虑误差时序 - 条件性质的短期风电功率概率预测[J]. 电网技术, 2020, 44(7): 2529-2537.
- QIAO Ying, LU Zongxiang, WU Wenzu. Day-ahead wind power probabilistic forecast considering conditional dependency and temporal correlation[J]. Power System Technology, 2020, 44(7): 2529-2537.
- [21] 李焱, 贾雅君, 李磊, 等. 基于随机森林算法的短期电力负荷预测[J]. 电力系统保护与控制, 2020, 48(21): 117-124.
- LI Yan, JIA Yajun, LI Lei, et al. Short term power load forecasting based on a stochastic forest algorithm[J]. Power System Protection and Control, 2020, 48(21): 117-124.
- [22] KE G, MENG Q, FINLEY T, et al. Light GBM: a highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems, 2017: 3146-3154.
- [23] RASMUSSEN C E, NICKISCH H. Gaussian processes for machine learning (GPML) toolbox[J]. Journal of Machine Learning Research, 2010, 11(6): 3011-3015.
- [24] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. 软件学报, 2018, 29(10): 3068-3090.
- CUI Jiaxu, YANG Bo. Survey on Bayesian optimization methodology and applications[J]. Journal of Software, 2018, 29(10): 3068-3090.
- [25] 刘长良, 曹威, 王梓齐. 基于模糊 C 均值软聚类工况辨识的风电功率短期区间预测[J]. 华北电力大学学报: 自然科学版, 2019, 46(5): 83-90.
- LIU Changliang, CAO Wei, WANG Ziqi. Short-term interval prediction of wind power based on fuzzy C-means soft clustering condition identification [J]. Journal of North China Electric Power University: Natural Science, 2019, 46(5): 83-90.
- [26] 李惠琴, 陈燕, 王岗红, 等. 计及功率预测误差的主动配电网实时经济调度方法[J]. 电力系统保护与控制, 2020, 48(16): 100-107.
- LI Huiqin, CHEN Yan, WANG Ganghong, et al. Real-time economic dispatching method for active distribution networks considering power prediction errors[J]. Power System Protection and Control, 2020, 48(16): 100-107.
-
- 收稿日期: 2020-06-29; 修回日期: 2020-12-07
- 作者简介:
- 黄 慧(1982—), 女, 通信作者, 博士研究生, 研究方向为电力系统优化规划与运行、新能源并网、智能电网技术; E-mail: cghuanghui2000@163.com
- 贾 嵘(1971—), 男, 教授, 博士生导师, 研究方向为电力系统运行与控制、电力设备状态监测与诊断、智能电网技术。E-mail: jiarong@xaut.edu.cn
- (编辑 张爱琴)