引用格式:周灿,周育才,谭艳祥,等.基于聚类和RIME-LightGBM的光热电站太阳直接法向辐射预测[J].电力科学与技术学报,2025,40(6):241-249.
Citation: ZHOU Can, ZHOU Yucai, TAN Yanxiang, et al. Prediction of direct normal irradiance from photovoltaic power plants based on clustering and RIME-LightGBM[J]. Journal of Electric Power Science and Technology,2025,40(6):241-249.

基于聚类和RIME-LightGBM的光热电站太阳直接法向辐射预测

周　灿[1],周育才[1],谭艳祥[2],肖　添[1],谢七月[1],申忠利[1],付　强[1],秦缘恒[1]
(1.长沙理工大学电气与信息工程学院,湖南 长沙 410114;2.长沙理工大学数学与统计学院,湖南 长沙 410114)

摘　要:太阳直接法向辐射的间歇性与不确定性会影响光热电站电力输出的稳定性。针对该问题,提出一种基于聚类、霜冰优化算法(rime optimization algorithm,RIME)与优化轻梯度提升机(light gradient boosting machine,LightGBM)的光热电站太阳直接法向辐射预测模型。先通过皮尔逊相关系数筛选太阳直接法向辐射的强相关气象参数,并采用小批量K均值(mini batch K-means,MBK)聚类算法对历史气象数据进行分类;再利用RIME对LightGBM超参数寻优,建立不同类别历史气象数据的太阳直接法向辐射预测模型;然后,以预测日每小时与各聚类中心强相关气象参数数据的欧式距离为依据,选择相应预测模型,对太阳直接法向辐射进行预测;最后,采用美国加州某地光热电站2000—2019年的历史气象数据,对所提模型进行验证。研究结果表明:所提预测模型能较准确地预测太阳直接法向辐射的数值及变化趋势。

关　键　词:光热发电;太阳直接法向辐射;小批量K均值聚类;轻梯度提升机;霜冰优化算法
DOI:10.19781/j.issn.1673-9140.2025.06.024　　**中图分类号**:TM863　　**文章编号**:1673-9140(2025)06-0241-09

Prediction of direct normal irradiance from photovoltaic power plants based on clustering and RIME-LightGBM

ZHOU Can[1], ZHOU Yucai[1], TAN Yanxiang[2], XIAO Tian[1],
XIE Qiyue[1], SHEN Zhongli[1], FU Qiang[1], QIN Yuanheng[1]
(1. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China;
2. School of Mathematics and Statistics, Changsha University of Science & Technology, Changsha 410114, China)

Abstract: The stability of the power output of the photovoltaic power plants is affected by the intermittency and uncertainty of direct normal irradiance (DNI). In order to solve this problem, a prediction model of DNI based on the clustering and rime optimization algorithm (RIME) optimizing the light gradient boosting machine (LightGBM) is proposed. Firstly, the strongly correlated meteorological parameters of the DNI are determined by the Pearson correlation coefficient, and the historical meteorological data is classified by the mini batch K-means (MBK) clustering algorithm. Then, RIME is used to optimize the hyperparameters of the LightGBM and to establish the prediction model of DNI for different categories of historical meteorological data. The Euclidean distances between the hourly data of the forecast day and the strongly correlated meteorological parameters of each cluster center are used to select the corresponding prediction model for the prediction of the DNI. Finally, by using the historical meteorological data from 2000 to 2019 of a concentrating solar power (CSP) plant in California, USA, the proposed model is validated. The experimental results show that the proposed prediction model can accurately predict the value and variation trend of DNI.

Key words: concentrating solar power; direct normal irradiance; mini batch K-means clustering; light gradient boosting machine; rime optimization algorithm

收稿日期:2025-03-08;**修回日期**:2025-04-07
基金项目:国家自然科学基金(62373067);长春晟德科技有限公司科技项目(3040402-2483)
通信作者:周育才(1971—),男,博士,副教授,主要从事智能识别与控制等方面的研究;E-mail: zyucai@163.com

随着煤、石油、天然气等化石能源的日益枯竭，其所带来的环境污染问题日益突出，作为一种清洁能源的太阳能得到了前所未有的关注。在国内外众多学者的研究下，太阳能的利用率逐步提高[1-5]。光热发电（concentrating solar power，CSP）的基本原理是通过反光镜场将太阳辐射能聚焦在集热区，集热区通过加热介质产生高温蒸汽，高温蒸汽推动汽轮机发电机组工作，产生电能[6]。这是一种将太阳能转化成热能，再将热能转化成电能的发电技术。光热发电系统运行主要利用的是太阳直接法向辐射（direct normal irradiance，DNI）。但由于太阳直接法向辐射的间歇性与不稳定性，光热发电系统并网后会影响电网的稳定运行[7-9]。因此，为充分发挥光热电站的优势，方便电网的调度安排与储能系统的配置[10]，对光热电站的太阳直接法向辐射进行预测具有重要的实际意义。

在太阳直接法向辐射的预测方面，国内外学者主要研究的是气象参数变化与太阳辐射之间的联系。目前，其预测方法大致可以分为传统物理模型预测和机器学习算法预测。传统物理模型预测方法有数值天气预报模型法（numerical weather prediction，NWP）、统计回归模型法、全天空及卫星成像外推法等。文献[11]针对不同天气条件下NWP辐照度预测准确性的差异，总结出相应的错误特征，提出了不同场景下的NWP修正方法，提高了NWP辐照度预测的准确性。文献[12]在不同气候条件下使用自回归积分滑动平均模型（autoregressive integrated moving average，ARIMA）对太阳辐照度进行预测，建立了月数据和小时数据两个时间序列的预测模型。机器学习算法预测中常用的模型有支持向量机（support vector machine，SVM）[13-16]、随机森林（random forest，RF）[17-19]、梯度提升树（gradient boosting decision tree，GBDT）[20-21]、神经网络[22-25]等。SVM、RF等算法虽有一定的预测效果，但其预测精度仍有待提升。文献[26]使用多头CNN（convolutional neural networks，CNN）和多通道CNN模型，并引入每小时云指数气象参数对太阳辐照度进行了中期预测。文献[27]将改进的递归深度信念网络和静态模型结合，对CSP电站短期发电功率进行预测。文献[28]将扩张因果卷积（dilated causal convolution，DCC）和堆叠长短期记忆网络（stacking long short-term memory，LSTM）结合，形成扩张因果卷积堆叠长短期记忆网络（dilated causal convolution stacking long short-term memory，DSLSTM），先利用扩张因果卷积网络充分挖掘光伏历史数据的空间特征，再利用LSTM充分学习历史数据序列中的时间特征，有效提高了模型的预测精度。文献[29]引入残差注意力机制对原始气象特征进行重构，并利用LSTM进行了短期太阳辐照度的预测。文献[30]提出加权K均值聚类算法和双向长短期记忆网络（bidirectional long short-term memory，Bi-LSTM）太阳辐照度预测方法。文献[31]先将历史气象数据聚类，再利用改进的鲸鱼优化算法（improved whale optimization algorithm，IWOA）与优化变分模态分解（variational mode decomposition，VMD），分解历史数据，构建双向长短期记忆网络结合注意力机制（bi-directional long short-term memory-attention，Bi-LSTM-Attention）的预测混合模型，并用仿真证明了其有效性。

面对庞大数据集时，LSTM模型的训练速度较缓慢。此外，预测模型最佳超参数的确定对提高模型的精准性也极为重要。因此，本文提出一种基于聚类结合霜冰优化算法（rime optimization algorithm，RIME）优化轻梯度提升机（light gradient boosting machine，LightGBM）的光热电站DNI预测模型。先通过皮尔逊相关系数确定DNI的强相关气象参数，并将其作为聚类和预测模型的输入，采用小批量K均值（mini batch k-means，MBK）聚类算法对历史气象数据进行分类；再利用RIME[32]优化LightGBM超参数，建立不同类别历史气象数据的DNI预测模型；然后，计算预测日每小时与各聚类中心强相关气象数据的欧式距离，选择最小欧氏距离的聚类中心所对应的预测模型来进行预测，以此得到预测日每小时的DNI预测结果；最后，进行了系列试验来对比分析，验证该模型在对光热电站DNI预测上的准确性和鲁棒性。

1 数据处理与分析

1.1 数据处理

1.1.1 缺失值及异常值处理

原始数据中的缺失值和异常值会对模型的建立和预测的精度造成不利影响。因此，对原始数据进行处理是十分必要的。

缺失值常见的处理方式是利用均值、中位数或众数进行填充。处理数值型数据时，如果数据分布较对称且缺失值占比较少，可使用对异常值较为敏感的均值进行填充；当数据存在偏斜或者异常值时，因中位数是数据的中间值，其比均值更不易受

到异常值的影响，其鲁棒性比采用均值的方法的鲁棒性更好，故常使用中位数进行填充。处理离散型数据，尤其是分类变量时，可使用众数进行缺失值的填充，保持数据的离散性，避免缺失值的不恰当填充引入不符合数据分布的偏差。

异常值常采用统计学中的3σ原则来进行处理。根据此原则，大约99.7%的数据会落在$(\mu-3\sigma,\mu+3\sigma)$范围内（$\mu$为该随机变量分布的均值，$\sigma$为其标准差），故将不在此范围的数据视为异常值，将其进行剔除或视之为缺失值。

1.1.2　数据归一化

不同特征变量的单位和量纲均不尽相同，这些差异会导致模型在训练时出现偏差。为消除不同特征变量的量纲影响，加快模型的收敛速度，增强模型稳定性，本文使用归一化操作对数据进行处理，使数据落在[0,1]。归一化后数据x_{new}为

$$x_{new}=\frac{x-x_{min}}{x_{max}-x_{min}} \tag{1}$$

式中，x为原始数据；x_{min}为数据中的最小值；x_{max}为数据中的最大值。

1.2　数据分析

1.2.1　皮尔逊系数

皮尔逊相关系数是一种衡量2个连续变量之间线性相关程度的统计量。本文使用皮尔逊相关系数来衡量DNI与不同气象参数间的相关程度，找出与DNI相关性强的气象参数，皮尔逊相关系数P可表示为

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

式中，x_i和y_i分别为历史气象数据集中样本i的DNI和其余气象参数样本点；$\bar{x}$和$\bar{y}$分别为x_i和y_i的平均值；n为样本数。

$P\in[-1,1]$，若$P=0$，则这2个变量不相关；若$P>0$，则这2个变量正相关；若$P<0$，则这2个变量负相关。P的绝对值越大，这2个变量间的相关性越强。

1.2.2　热力图分析

本文所用数据集为美国国家太阳辐射数据库（national solar radiation database，NSRDB）发布的美国加州某地光热电站历史气象数据集，数据时间间隔为1 h。在进行聚类和预测模型建立之前，应找到与DNI相关性强的气象参数，剔除相关性弱的气象参数。这样可提高聚类的质量和预测模型的精准性。图1为基于皮尔逊相关系数的气象数据热力图。在图1中，横纵坐标轴上数字0～9依次代表DNI、露点、地表反射率、风速、湿度、温度、气压、云类型、太阳时角、风向。从图1中可以看出，温度(5)、风速(3)与DNI(0)呈正相关，温度与风速的皮尔逊相关系数分别为0.58、0.38；太阳时角(8)、相对湿度(4)、云类型(7)与DNI均呈负相关，这三者的皮尔逊相关系数依次为−0.79、−0.59、−0.33。这5个气象参数会对DNI产生较强的影响。风向(9)、露点(1)、地表反射率(2)、气压(6)这些气象参数与DNI皮尔逊相关系数的绝对值均低于0.2，故这4个气象参数对DNI的影响较小，可忽略不计。因此，本文筛选出的强相关气象参数为温度、风速、太阳时角、相对湿度、云类型。

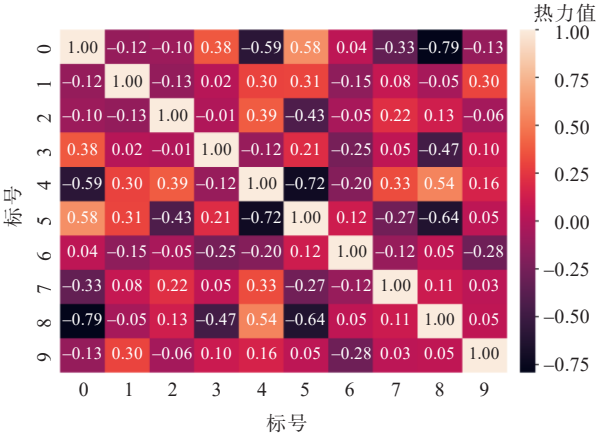

图1　基于皮尔逊相关系数的气象数据热力图

Figure 1　Heatmap of meteorological data based on Pearson correlation coefficients

2　聚类分析

2.1　小批量K均值聚类

MBK是在K-Means聚类的基础上优化改进得到的。其算法旨在通过减少每次更新所需的数据量来加速聚类过程。不同于K-Means，MBK在每次迭代中不使用整个数据集来更新聚类中心，而是采用随机选取的小批量数据。通过这种方式，MBK每次迭代时仅计算一个小子集的数据，显著降低了计算量，其聚类速度也得到了提升。MBK在保持聚类质量的同时，能显著提高计算效率，其在聚类精度和计算速度之间达到了良好的平衡。

MBK算法步骤共5步。

1）对样本容量为n的样本集$X=\{x_1,x_2,\cdots,x_n\}$

分簇，随机初始化K个聚类中心$C=\{c_1, c_2, \cdots, c_k\}$。

2）从样本集X中随机选取样本子集$Z=\{z_1, z_2, \cdots, z_f\}$，$f$为该子集的元素个数。

3）计算样本子集中每个样本点与各聚类中心的距离，将样本点分配给距离最近的聚类中心所对应的簇l_i，即

$$l_i=\arg\min_j \left\| z_i-c_j \right\|^2 \tag{3}$$

式中，z_i为样本子集Z中的第i个样本；c_j为第j个聚类中心。

4）对于每个簇，重新计算新的聚类中心，即

$$c_j'=\frac{\sum_{i=1}^{m} Q_i}{m} \tag{4}$$

式中，m为簇l_i中的样本数量；Q_i为簇l_i中的某一样本；c_j'为新聚类中心。

5）以新聚类中心为基准，重复执行步骤2）～4），直至簇中心不再发生变化或达到最大迭代次数。

2.2 聚类评价

MBK是一种无监督学习算法。由于无监督学习算法无法直接评判其聚类结果的优劣，本文采用轮廓系数对其聚类效果进行评判，第i个样本的轮廓系数$s(i)$为

$$s(i)=\frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{5}$$

式中，$a(i)$为凝聚度，即第i个样本与同簇其他样本的平均距离；$b(i)$为分离度，即第i个样本与其他簇中所有样本的平均距离。

所有样本轮廓系数的平均值即为最终的轮廓系数s，其取值范围为$[-1, 1]$。s越接近于1，簇内距离越小，簇间距离越大，聚类效果越好。

本文计算聚类肘部图拐点范围内的轮廓系数，以找到最佳聚类簇数。将MBK、K-Means、模糊C均值聚类（fuzzy C-means，FCM）和高斯混合聚类（Gaussian mixture model，GMM）的聚类效果进行对比，这4种聚类算法在不同簇数K下的轮廓系数见表1。

表1 不同聚类算法在不同簇数K下的轮廓系数

Table 1 Silhouette coefficients of different clustering algorithms at varying cluster numbers K

簇数K	MBK	K-Means	FCM	GMM
4	0.323	0.320	0.291	0.238
5	0.308	0.303	0.267	0.208
6	0.287	0.273	0.258	0.160

由表1可知，MBK聚类在不同簇数下，其轮廓系数比其他3种聚类算法的轮廓系数更高，这说明MBK的聚类效果更好，更适合对本文所用数据集进行聚类。随着簇数K的增加，MBK聚类的轮廓系数逐渐下降。因此，将MBK算法的簇数K定为4。MBK聚类气象数据聚类过程如图2所示。从图2中可以看出，MBK将历史气象数据分为4类。此后，可分别对每类气象数据进行训练，以提高预测模型精准度。

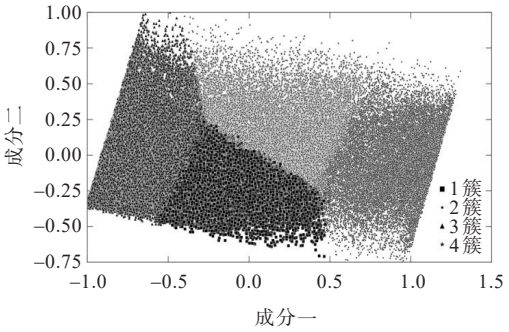

图2 MBK气象数据聚类过程

Figure 2 MBK meteorological data clustering process

3 预测模型构建

3.1 LightGBM

LightGBM是一种基于GBDT的梯度提升框架。与同样基于GBDT模型的XGBoost相比，LightGBM在XGBoost的基础上引入了直方图、带深度限制的Leaf-wise决策树生长策略、单边梯度采样、互斥特征捆绑（exclusive feature bundling，EFB）等，进行了优化改进，使其可进行高效的并行训练，训练速度更快、内存消耗更低、准确率更高，能支持分布式快速处理海量数据，非常适合处理本文数据集。LightGBM结构如图3所示。

直方图算法对特征的原始数据进行分桶操作，把数据划分至多个不同的离散区域内，将连续的特征取值转化成离散值。该算法降低了模型的复杂度，减少了候选分类。直方图算法示意图如图4所示。LightGBM采用带深度限制的Leaf-wise决策树生长策略，该策略在决策树分裂时，只对分裂增益最大的一个叶节点进行分裂，且增加了对最大深度的限制，有效防止过拟合现象的发生，该算法运行效率更高，误差更低。决策树生长策略对比示意图如图5所示。单边梯度采样通过对梯度大的样本进行保留，对梯度小的样本进行随机采样，减少了样本数量；EFB将互斥的特征进行捆绑，减少了样本特征数量。

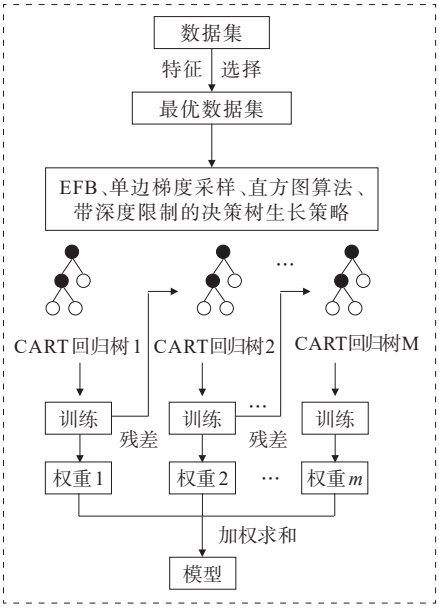

图 3　LightGBM 结构

Figure 3　LightGBM structure

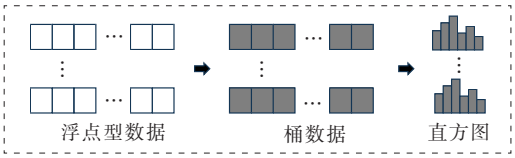

图 4　直方图算法示意图

Figure 4　Histogram algorithm

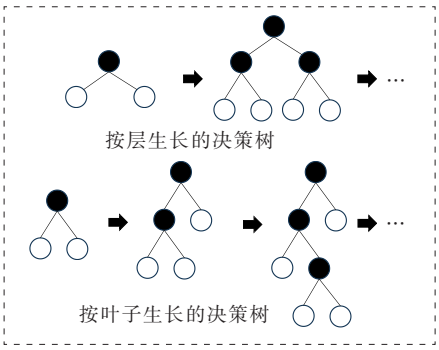

图 5　决策树生长策略对比示意图

Figure 5　Comparison of decision tree growth strategies

3.2　霜冰优化算法

RIME受自然界霜冰生长机制的启发，通过模拟冰霜的软霜和硬霜生长过程，构建软霜搜索策略和硬霜穿刺机制，实现优化方法中的探索和开发。RIME还提出了正贪婪选择机制，在选择最优解阶段更新种群，增强了算法的开发能力。本文采用RIME对LightGBM超参数进行寻优。

软霜搜索策略利用软霜粒子的强随机性和覆盖性，使算法在迭代初期能够快速覆盖整个搜索空间，避免陷入局部最优解。软霜搜索策略为

$$R_{i,j}^{\text{new}} = R_{\text{best},j} + r_1\beta\left[h\left(b_{i,j}^{\text{up}} - b_{i,j}^{\text{low}}\right) + Lb_{i,j}\right]\cos\theta, r_2 < E \tag{6}$$

式中，$R_{i,j}^{\text{new}}$ 为第 i 个霜冰粒子在第 j 个维度上更新后的新位置；$R_{\text{best},j}$ 为最佳霜冰种群中的第 j 个粒子；r_1 为范围在(−1, 1)的参数；r_2 为范围在(0, 1)的一个随机数；h 为黏附度；L 为约束霜冰粒子运动的参数；$b_{i,j}$ 为第 i 个霜冰粒子在第 j 个维度上的当前位置；$b_{i,j}^{\text{up}}$ 和 $b_{i,j}^{\text{low}}$ 分别为第 i 个霜冰粒子在第 j 个维度上逃逸空间的上界和下界，用来限制霜冰粒子的运动区域；β 为环境因素；E 为附着系数；θ 为角度参数。

式(6)中各参数分别为

$$\theta = \frac{\pi t}{10T} \tag{7}$$

$$\beta = 1 - \frac{\left\lceil \frac{wt}{T} \right\rceil}{w} \tag{8}$$

$$E = \sqrt{t/T} \tag{9}$$

式中，t 为当前迭代次数；T 为算法最大迭代次数；[·]为取整算子；w 是调节步进函数的段数，默认为5。

RIME硬霜穿刺机制可更新种群间算法，实现算法粒子的置换，改善算法的收敛性。粒子间置换表达式为

$$R_{i,j}^{\text{new}} = R_{\text{best},j}, r_3 < F^{\text{normr}}(S_i) \tag{10}$$

式中，$F^{\text{normr}}(S_i)$ 为第 i 个霜冰粒子当前代理适应度的归一化值；r_3 为范围在(−1, 1)的随机数。

RIME优化LightGBM超参数流程如图6所示。

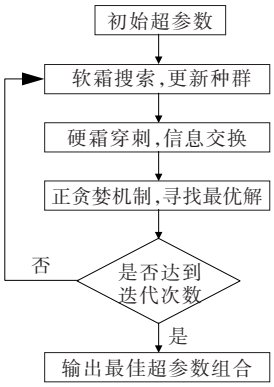

图 6　RIME优化LightGBM超参数流程

Figure 6　Hyperparameter optimization process by RIME-LightGBM

3.3　MBK-RIME-LightGBM的建立

本文构建了MBK-RIME-LightGBM模型，预测光热电站的DNI。MBK-RIME-LightGBM预测流程如图7所示，其具体步骤共4步。

1）对历史气象数据进行处理与分析，填充缺失值、剔除异常值并将数据归一化，并利用皮尔逊相关系数筛选与DNI强相关的气象参数。

2）利用MBK对处理后的数据进行聚类，使整个历史气象数据集分为4类不同气象特征的数据集，将竞争迭代后的聚类中心作为每类气象特征数据集的典型气象特征数据，并把每类气象特征数据集按8∶2的比例划分为训练集和验证集。

3）利用RIME寻找LightGBM的最佳超参数，建立4类不同气象特征数据对应的DNI预测模型。

4）计算日每小时预测参数数据与各聚类中心的强相关气象参数数据的欧式距离，选择欧氏距离最小的聚类中心对应的预测模型来进行DNI预测，得到预测日每小时的DNI预测结果。欧氏距离$D(x,y)$为

$$D(x,y)=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2} \tag{11}$$

式中，x_i为预测日每小时的强相关气象参数数据；y_i为各聚类中心的强相关气象参数数据；m为强相关气象参数的个数。

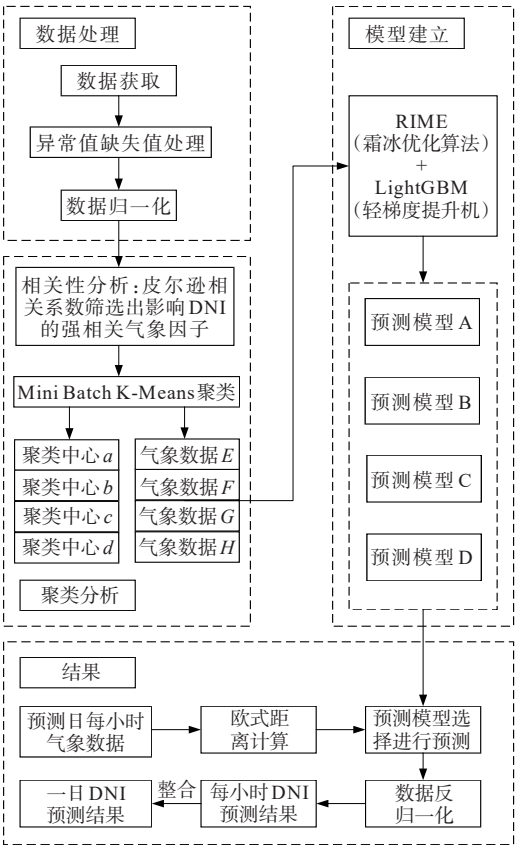

图7 MBK-RIME-LightGBM预测流程

Figure 7 Prediction process of MBK-RIME-LightGBM

4 预测试验分析

4.1 模型参数设置

本文选用均方根误差（root mean square error，RMSE）、决定系数R^2作为预测模型超参数选择的标准及评价指标，即

$$I_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-y_i'\right)^2} \tag{12}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-y_i'\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2} \tag{13}$$

式中，I_{RMSE}为均方根误差；y_i为DNI样本i的实际值；y_i'为DNI样本i的预测值；$\overline{y_i}$为DNI样本i实际值的平均值；n为样本总数。

针对这4类不同的气象特征数据，建立对应的4类DNI预测模型。由于LightGBM超参数众多，为快速准确地获得模型理想超参数，本文利用RIME对LightGBM超参数进行调参，先设定模型各超参数取值范围，将叶子数l、树最大深度d、学习率r、最小子节点样本数m、特征采样比例f、数据采样比例s、数据采样频率t的取值范围分别设为[20, 50]、[3, 15]、[0.001, 0.100]、[5, 100]、[0.1, 1.0]、[0.1, 1.0]、[1, 10]，以模型输出的五折交叉验证的RMSE作为目标函数，寻找模型的最佳超参数。这4类气象特征数据对应的DNI预测模型最佳超参数见表2。

表2 4类气象特征数据对应的DNI预测模型最佳超参数

Table 2 Optimal hyperparameters for DNI prediction models corresponding to four categories of meteorological feature data

预测模型	l	d	r	m	f	s	t
A	49	15	0.031 4	56	0.978 3	0.670 2	1
B	50	12	0.023 5	15	0.877 7	0.964 1	1
C	42	13	0.037 2	5	0.819 4	0.900 0	3
D	47	9	0.013 9	7	0.925 0	0.801 7	5

4.2 预测结果及分析

本文训练所用数据为美国加州某地光热电站2000—2019年历史气象数据。为实际检验预测模型、MBK和RIME提高LightGBM对DNI预测精准度上的有效性，以该光热电站2020年5月1日（多云转晴）、2020年7月13日（晴天）作为预测日，分别使用LightGBM、BOA-LightGBM、RIME-LightGBM、MBK-RIME-GBDT、MBK-RIME-LightGBM模型

对预测日中05:00—18:00的DNI进行预测，并对不同模型的预测结果进行对比分析。晴天预测结果对比如图8所示，多云转晴预测结果对比如图9所示。

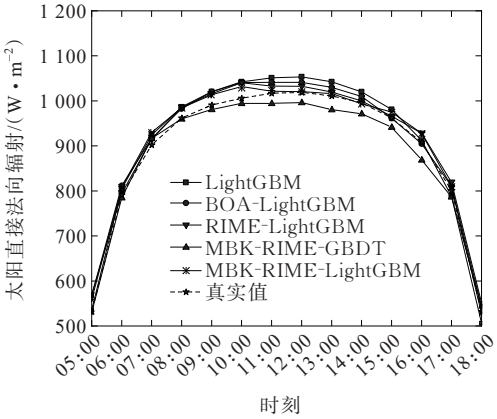

图8　晴天预测结果对比

Figure 8　Comparison of sunny weather forecast results

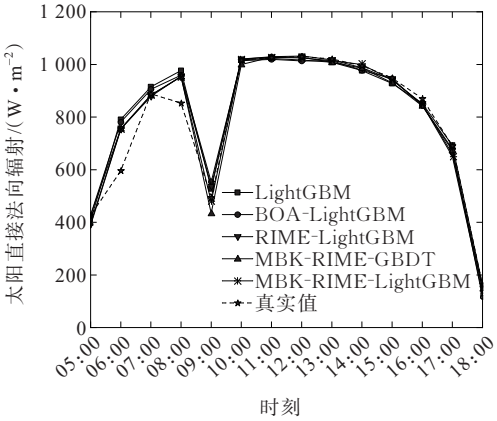

图9　多云转晴预测结果对比

Figure 9　Comparison of forecast results for cloudy-to-sunny conditions

从图8中可以看出，在晴天非突变天气类型下，DNI的真实值波动不大，但LightGBM模型的预测结果比其他模型的稍差一些，在09:00—14:00时间段，DNI预测值与真实值存在较大误差。引入贝叶斯优化算法（Bayesian optimization algorithm，BOA）、RIME优化算法对LightGBM超参数寻优。BOA-LightGBM、RIME-LightGBM这两种模型在该时间段对DNI的预测误差均明显下降，RIME-LightGBM模型的预测值更接近真实值；MBK-RIME-LightGBM模型对DNI的预测误差最小，且其整体预测趋势最贴近真实变化趋势。从图9中可以看出，在多云转晴突变天气类型下，DNI的真实值波动较大，这5种模型的预测误差比晴天时的误差有下降。在07:00—10:00天气突变时间段，引入RIME优化算法。RIME-LightGBM模型的DNI的预测误差比LightGBM和BOA-LightGBM模型的有所下降；MBK-RIME-LightGBM模型在09:00的DNI预测值最接近真实值，其预测误差最小。

各模型在不同天气类型下对DNI预测的I_{RMSE}、R^2见表3。

表3　各模型在不同天气类型下的I_{RMSE}，R^2

Table 3　RMSE and R^2 of each model under different weather types

天气	预测模型	I_{RMSE}/(W·m^{-2})	R^2
晴朗	LightGBM	24.817 6	0.976 6
晴朗	BOA-LightGBM	21.143 6	0.983 0
晴朗	RIME-LightGBM	20.385 7	0.984 2
晴朗	MBK-Means-RIME-GBDT	19.512 1	0.985 5
晴朗	MBK-Means-RIME-LightGBM	17.335 2	0.988 6
多云转晴	LightGBM	63.694 5	0.941 1
多云转晴	BOA-LightGBM	61.516 8	0.945 1
多云转晴	RIME-LightGBM	53.750 6	0.958 1
多云转晴	MBK-Means-RIME-GBDT	54.912 5	0.956 2
多云转晴	MBK-Means-RIME-LightGBM	52.629 4	0.959 8

由表3可知，在不采用聚类和优化算法的情况下，LightGBM模型在这两种天气类型下对DNI的预测误差均较高。在晴天天气下，BOA-LightGBM与RIME-LightGBM的I_{RMSE}比单纯的LightGBM模型的分别降低了14.80%和17.86%，BOA-LightGBM与RIME-LightGBM的I_{RMSE}比单纯的LightGBM分别提升了0.65%和0.78%；在多云转晴天气下，BOA-LightGBM与RIME-LightGBM的I_{RMSE}比单纯的LightGBM模型的分别降低了3.42%和15.61%，BOA-LightGBM与RIME-LightGBM的I_{RMSE}比单纯的LightGBM分别提升了0.43%和1.81%。这说明优化算法RIME、BOA能有效提高LightGBM模型的DNI预测精度，且其I_{RIME}的优化效果比BOA的更佳；在晴天和多云转晴天的天气下，MBK-RIME-LightGBM模型的I_{RMSE}分别比RIME-LightGBM降低了14.96%和2.09%，MBK-RIME-LightGBM模型的I_{RMSE}分别比RIME-LightGBM提升了0.45%和0.18%，这说明先利用MBK对历史气象数据分类，再分别进行预测模型的训练，可有效提高DNI的预测精度。在多云转晴天气下，DNI预测的各项评价指标均劣于晴朗天气的。这是因为突变天气

下各项气象参数变化明显且规律性较低。但从整体上看，不同天气类型下MBK-RIME-LightGBM模型对DNI预测的精度优于其他模型的精度，这验证了本文所提预测模型的有效性和鲁棒性。

5 结语

针对DNI的间歇性和不确定性对光热电站电力输出稳定性造成的影响，本文提出一种基于MBK-RIME-LightGBM的光热电站DNI预测模型，得出以下结论：

1）先处理分析历史气象数据，再利用聚类将其进行分类，建立每类气象数据对应的DNI预测模型。且预测时根据欧式距离选择每小时气象数据最匹配的DNI预测模型进行预测，可有效提高预测模型的精度。

2）利用RIME对LightGBM进行超参数寻优，可快速准确地确定模型最佳超参数，进一步提高了预测模型的准确性和鲁棒性。

3）试验对比验证了本文所提模型对太阳直接法向辐射预测的有效性和准确性，对电网提前制定调度计划有一定的参考价值。

参考文献：

[1] 臧海祥，程礼临，刘玲，等．基于数据驱动的太阳辐射估计和预测研究与展望[J]．电力系统自动化，2021，45(11): 170-183.
ZANG Haixiang，CHENG Lilin，LIU Ling，et al. Research and prospect for data-driven estimation and prediction of solar radiation[J]. Automation of Electric Power Systems, 2021,45(11): 170-183.

[2] HERRERA-CASANOVA R, CONDE A, SANTOS-PÉREZ C. Hour-ahead photovoltaic power prediction combining BiLSTM and Bayesian optimization algorithm, with bootstrap resampling for interval predictions[J]. Sensors, 2024,24(3): 882.

[3] 王光华，张纪欣，崔良，等．基于双重注意力变换模型的分布式屋顶光伏变电站级日前功率预测[J]．全球能源互联网，2024，7(4): 393-405.
WANG Guanghua, ZHANG Jixin, CUI Liang, et al. Substation-level distributed rooftop photovoltaic power day-ahead prediction based on double attention mechanism transformer model[J]. Journal of Global Energy Interconnection, 2024,7(4): 393-405.

[4] YAO L X, WANG Y, XIAO X Y. Concentrated solar power plant modeling for power system studies[J]. IEEE Transactions on Power Systems, 2024,39(2): 4252-4263.

[5] 王湘艳，陈宁，王维洲，等．计及置信容量的光热电站储热容量优化配置[J]．电力工程技术，2022，41(5): 103-109,115.
WANG Xiangyan, CHEN Ning, WANG Weizhou, et al. Optimal design for thermal energy storage capacity of CSP considering credible capacity[J]. Jiangsu Electrical Engineering, 2022,41(5): 103-109,115.

[6] MFETOUM I M, NGOH S K, MOLU R J J, et al. A multilayer perceptron neural network approach for optimizing solar irradiance forecasting in Central Africa with meteorological insights[J]. Scientific Reports, 2024, 14(1): 3572.

[7] GUERMOUI M, ARRIF T, BELAID A, et al. Enhancing direct Normal solar Irradiation forecasting for heliostat field applications through a novel hybrid model[J]. Energy Conversion and Management, 2024,304: 118189.

[8] HERRANDO M, WANG K, HUANG G, et al. A review of solar hybrid photovoltaic-thermal (PV-T) collectors and systems[J]. Progress in Energy and Combustion Science, 2023,97: 101072.

[9] 李宗隆，柏舸，胡剑，等．考虑光热光伏混合电站内外双重电压安全的两阶段无功优化控制[J]．电力系统保护与控制，2023，51(20): 170-179.
LI Zonglong, BAI Ge, HU Jian, et al. Two-stage optimal reactive power control considering internal-external dual voltage security of a PV-CSP hybrid power station[J]. Power System Protection and Control, 2023, 51(20): 170-179.

[10] 唐雅洁，林达，倪筹帷，等．基于XGBoost的双层协同实时校正超短期光伏预测[J]．电力系统自动化，2021，45(7): 18-27.
TANG Yajie, LIN Da, NI Chouwei, et al. XGBoost based bi-layer collaborative real-time calibration for ultra-short-term photovoltaic prediction[J]. Automation of Electric Power Systems, 2021,45(7): 18-27.

[11] DOU W J, WANG K, SHAN S, et al. Evaluation of performance for day-ahead solar irradiance forecast using numerical weather prediction[J]. Journal of Renewable and Sustainable Energy, 2024,16(4): 043703.

[12] CHODAKOWSKA E, NAZARKO J, NAZARKO Ł, et al. ARIMA models in solar radiation forecasting in different geographic locations[J]. Energies, 2023,16(13): 5029.

[13] GUERMOUI M, MELGANI F, GAIRAA K, et al. A comprehensive review of hybrid models for solar radiation forecasting[J]. Journal of Cleaner Production, 2020,258: 120357.

[14] 青灿，行舟，智勇，等．基于SVM的变电站母线日净负荷曲线大数据识别方法[J]．电力科学与技术学报，2022，37(6): 125-131.
QING Can, XING Zhou, ZHI Yong, et al. Big data recognition method for daily net load curve of substation bus based on SVM[J]. Journal of Electric Power Science

and Technology,2022,37(6): 125-131.

[15] ALZAHRANI A. Short-term solar irradiance prediction based on adaptive extreme learning machine and weather data[J].Sensors,2022,22(21): 8218.

[16] 汪自虎,王文天,惠慧,等.基于近邻成分分析与优化核极限学习机的光伏接入配电网漏电识别[J].高压电器,2024,60(6): 203-211.
WANG Zihu, WANG Wentian, HUI Hui, et al. Leakage recognition method of photovoltaic connected to distribution network based on neighborhood component analysis and optimized kernel extreme learning machine [J].High Voltage Apparatus,2024,60(6): 203-211.

[17] JOGUNURI S, JOSH F T, STONIER A A, et al. Random forest machine learning algorithm based seasonal multi-step ahead short-term solar photovoltaic power output forecasting[J].IET Renewable Power Generation,2025,19 (1): e12921.

[18] 封钰,宋佑斌,金晟,等.基于随机森林算法和粗糙集理论的改进型深度学习短期负荷预测模型[J].发电技术,2023,44(6): 889-895.
FENG Yu,SONG Youbin,JIN Sheng,et al.Improved deep learning model for forecasting short-term load based on random forest algorithm and rough set theory[J]. Power Generation Technology,2023,44(6):889-895.

[19] HISSOU H,BENKIRANE S,GUEZZAZ A,et al.A novel machine learning approach for solar radiation estimation [J].Sustainability,2023,15(13): 10609.

[20] 王俊杰,毕利,张凯,等.基于多特征融合和XGBoost-LightGBM-ConvLSTM的短期光伏发电量预测[J].太阳能学报,2023,44(7): 168-174.
WANG Junjie, BI Li, ZHANG Kai, et al. Short-term photovoltaic power generation prediction based on multi-feature fusion and XGBoost-lightgbm-convlstm[J]. Acta Energiae Solaris Sinica,2023,44(7): 168-174.

[21] XU W H, WANG Z K, WANG W S, et al. Short-term photovoltaic output prediction based on decomposition and reconstruction and XGBoost under two base learners [J].Energies,2024,17(4): 906.

[22] 乔楠,蒋波涛,郑雨,等.基于深度模糊神经网络的太阳总辐射预测研究[J].太阳能学报,2024,45(2): 59-64.
QIAO Nan,JIANG Botao,ZHENG Yu,et al.Research on global solar radiation forecast based on deep fuzzy neural network[J]. Acta Energiae Solaris Sinica, 2024, 45 (2): 59-64.

[23] 王海燕,刘佳康,邓亚平.基于预估-校正综合BP神经网络的短期光伏功率预测[J].智慧电力,2023,51(3): 46-52.
WANG Haiyan, LIU Jiakang, DENG Yaping. Short-term photovoltaic power forecasting based on predict-correct combination BP neural network[J].Smart Power,2023,51 (3): 46-52.

[24] 王东风,刘婧,黄宇,等.结合太阳辐射量计算与CNN-LSTM组合的光伏功率预测方法研究[J].太阳能学报,2024,45(2): 443-450.
WANG Dongfeng, LIU Jing, HUANG Yu, et al. Photovoltaic power prediction method combinating solar radiation calculation and cnn-lstm[J]. Acta Energiae Solaris Sinica,2024,45(2): 443-450.

[25] 王德文,焦天媛.基于二次分解的不同太阳辐射下光伏功率预测[J].太阳能学报,2024,45(9): 360-368.
WANG Dewen, JIAO Tianyuan. Prediction of photovoltaic power under different solar radiation based on secondary decomposition[J]. Acta Energiae Solaris Sinica,2024,45(9): 360-368.

[26] BLAZAKIS K, KATSIGIANNIS Y, STAVRAKAKIS G. One-day-ahead solar irradiation and windspeed forecasting with advanced deep learning techniques[J]. Energies,2022,15(12): 4361.

[27] 李锦键,王兴贵,杨维满,等.基于改进递归深度信念网络的CSP电站短期出力预测[J].太阳能学报,2022,43 (7): 225-232.
LI Jinjian, WANG Xinggui, YANG Weiman, et al. Csp station output power short-term forecast based on improved rnn-dbn[J]. Acta Energiae Solaris Sinica, 2022, 43(7): 225-232.

[28] TIAN C Y, LIN L L, YAN Y, et al. Photovoltaic power prediction based on dilated causal convolutional network and stacked LSTM[J]. Mathematical Biosciences and Engineering,2024,21(1): 1167-1185.

[29] 臧海祥,张越,程礼临,等.基于ICEEMDAN-LSTM和残差注意力的短期太阳辐照度预测[J].太阳能学报,2023, 44(12): 175-181.
ZANG Haixiang,ZHANG Yue,CHENG Lilin,et al.Short term solar radiation forecasting based on iceemdan-lstm and residual attention[J]. Acta Energiae Solaris Sinica, 2023,44(12): 175-181.

[30] 官松泽,唐钰本,蔡争,等.基于Kmeans++-Bi-LSTM的太阳辐照度超短期预测[J].太阳能学报,2023,44(12): 170-174.
GUAN Songze, TANG Yuben, CAI Zheng, et al. Ultra-short-term forecast of solar irradiance based on kmeans++-bi-lstm[J]. Acta Energiae Solaris Sinica, 2023, 44(12): 170-174.

[31] LI M, WANG W, HE Y, et al. Deep learning model for short-term photovoltaic power forecasting based on variational mode decomposition and similar day clustering[J].Computers and Electrical Engineering,2024, 115: 109116.

[32] SU H, ZHAO D, HEIDARI A A, et al. RIME: a physics-based optimization[J]. Neurocomputing, 2023, 532: 183-214.