

DOI: 10.19666/j.rlfed.202502015

# 基于聚类算法与贝叶斯优化的BP神经网络 实时碳排放量预测模型

姜宇恒<sup>1</sup>, 乔宗良<sup>1</sup>, 李逗<sup>2</sup>, 任少君<sup>1</sup>, 司凤琪<sup>1</sup>

(1.东南大学能源与环境学院, 江苏 南京 210096;

2.江苏方天电力有限公司, 江苏 南京 211006)

[摘要] 为了构建燃煤电厂碳排放量预测模型, 针对燃煤机组普遍缺少入炉煤实时元素分析的问题, 依据某百万千瓦机组2023年入炉煤质信息, 以低位发热量、挥发分、硫分作为煤质划分依据, 采用K-means++算法进行聚类分析, 通过相关性分析筛选碳排放量预测模型的输入参数, 基于贝叶斯优化的BP神经网络对聚类后各簇数据分别构建碳排放量预测模型, 并对升降负荷等工况进行模型测试。结果显示, 经煤质聚类后的模型在预测碳排放量时准确性显著提高, 与未聚类的模型相比, 平均均方根误差和平均相对误差最优情况降低约53.4%、49.2%, 特别是在变负荷工况下, 预测结果较准确。该方法不仅能有效预测燃煤电厂的碳排放量, 还能在煤质复杂多变的情况下保持较高精度。

[关键词] 碳排放量预测; 聚类算法; 贝叶斯优化; BP神经网络; 煤质

[引用本文格式] 姜宇恒, 乔宗良, 李逗, 等. 基于聚类算法与贝叶斯优化的BP神经网络实时碳排放量预测模型[J]. 热力发电, 2025, 54(11): 126-135. JIANG Yuheng, QIAO Zongliang, LI Dou, et al. Real-time prediction model for carbon emission using BP neural network based on clustering algorithm and Bayesian optimization[J]. Thermal Power Generation, 2025, 54(11): 126-135.

## Real-time prediction model for carbon emission using BP neural network based on clustering algorithm and Bayesian optimization

JIANG Yuheng<sup>1</sup>, QIAO Zongliang<sup>1</sup>, LI Dou<sup>2</sup>, REN Shaojun<sup>1</sup>, SI Fengqi<sup>1</sup>

(1.School of Energy and Environment, Southeast University, Nanjing 210096, China;

2.Jiangsu Frontier Electric Technology Co., Ltd., Nanjing 211006, China)

**Abstract:** To construct a prediction model for carbon emission from coal-fired power plants and address the problem of general lack of real-time elemental analysis for coal entering the furnace of coal-fired units, according to the in-furnace coal quality information of a million kilowatt unit in 2023, the low calorific value, volatile matter, and sulfur content were used as the basis for coal quality classification, K-means++ algorithm was used for clustering analysis, and correlation analysis was used to screen the input parameters of the carbon emission prediction model. The BP neural network suffered Bayesian optimization was used to construct carbon emission prediction models for each cluster data after clustering, and the models were tested for working conditions such as load increase and decrease. The results show that, the accuracy of the coal quality clustering model in predicting carbon emissions increases significantly. Compared with the non clustered model, the optimal cases of average root mean square error and average relative error reduce by about 53.4% and 49.2%, respectively. Especially under variable load conditions, the predicted results are more in line with the actual values. This indicates that the proposed method can not only effectively predict the carbon emissions of coal-fired power plants, but also maintain high accuracy in the case of complex and variable coal quality.

**Key words:** carbon emission prediction; clustering algorithm; Bayesian optimization; BP neural network; coal quality

收稿日期: 2025-02-15

第一作者简介: 姜宇恒 (2001), 男, 硕士研究生, 主要研究方向为燃煤机组特性建模及碳排放优化, 220234979@seu.edu.cn.

通信作者简介: 乔宗良 (1981), 男, 博士, 高级工程师, 主要研究方向为发电机组特性建模及优化运行, qiaozongliang@seu.edu.cn.

随着“碳达峰、碳中和”战略的实施,我国能源行业的结构体系也在向清洁化、低碳化迈进。尽管可再生能源的发电技术正高速发展,燃煤发电仍凭借自身稳定的电力输出及更低的成本,作为保障性、调节性电源在全球电力市场中占据一席之地<sup>[1]</sup>。燃煤电厂作为中国电力供应的主体,排放了大量的二氧化碳以及其他污染物<sup>[2]</sup>,为减少燃煤发电过程中产生的高碳排放量,且改善目前严重的环境问题<sup>[3]</sup>,对燃煤电厂的碳排放量进行准确预测十分重要。

煤作为燃煤机组主要燃料,在燃烧过程中会释放大量的二氧化碳等温室气体,成为全球碳排放的重要来源之一<sup>[4]</sup>。煤炭的使用量和燃烧效率直接决定了燃煤机组的碳排放量,且煤质也影响着碳排放强度。近年来火电燃煤机组面临煤炭供应紧张和煤价波动的挑战,导致煤源复杂多变,煤质特性如发热量和挥发分差异显著<sup>[5]</sup>。混煤燃烧对二氧化碳排放量的影响取决于煤种差异及配煤比例<sup>[6]</sup>。但煤质元素无在线测量方法,可以通过聚类方式探讨煤质对于燃煤机组碳排放的影响。安吉振等<sup>[7]</sup>利用  $k$  均值聚类算法聚类历史煤质信息,将煤质相似的来煤划分为同一类,在数字化煤场平台中设计并实现智能堆煤模块,提高煤场空间利用率和管理效率,并提供电厂配煤燃烧的改进方法。曹歌瀚等<sup>[8]</sup>采用  $k$  均值聚类算法对预处理后的静态数据集进行工况划分,依据误差平方和曲线确定聚类数,按负荷和相对煤质系数分类锅炉工况,并利用神经网络模型对各工况分别预测。文雯等<sup>[9]</sup>采用 K-means 聚类算法对不同工况进行划分,基于火电机组运行特性,通过分析影响供电煤耗的关键参数并利用机组历史运行数据,建立了基于改进集成学习算法的煤耗预测模型。

为评估燃煤电站的碳排放量,碳表计量技术得以开发并投入使用。该技术通过不间断地监测和分析燃煤电站从发电起始到结束的每个环节,能够确定其碳排放的准确数值<sup>[10]</sup>。鉴于碳监测体系的数据采集范围不够广泛,无法涵盖所有碳排放源,且对测点和设备的依赖性过高,当烟气测点发生堵塞、烟气涡流等情况时,监测数据与实际值存在较大误差<sup>[11]</sup>,科研人员转而探索更加经济高效的解决方案——火电厂碳排放预测模型<sup>[12]</sup>。这类模型基于大数据分析和机器学习算法,通过挖掘燃煤电站历史运行数据、燃料消耗、燃烧效率及烟气排放等多维度信息,在不依赖实时碳含量监测设备的情况下,实现对电厂碳排放量的可靠预测。例如,袁鹏

等<sup>[13]</sup>设计火电机组碳排放数据自动采集系统,结合迁移深度学习和强化学习技术,提出了一种实时碳排放预测方法,旨在提高火电厂碳排放预测的准确性。此外,为了进一步提高预测精度,还可以将多个模型进行集成学习,通过组合不同算法的优点来弥补单一模型的不足<sup>[14]</sup>。目前,已有许多学者进行了研究。Chen 等人<sup>[15]</sup>提出的 CNN-LSTM-Attention 模型为燃煤电厂的 CO<sub>2</sub> 排放预测提供了一种高精度的方法,不仅利用了电厂的历史运行数据,还通过贝叶斯优化方法优化了模型参数,提高了预测性能。Lu 等人<sup>[16]</sup>采取粒子群优化 (particle swarm optimization, PSO) 算法结合传统 BP 神经网络的措施,对重化工业的碳排放量进行预测,并预测了碳排放在不同情景下的峰值,为制定有效的减排政策提供了科学依据。

本文以某百万千瓦机组实际运行数据为基础,研究燃煤机组的碳排放量预测方法。首先,基于实际测量数据进行数据处理,并计算碳排放量;随后,依据煤质进行数据聚类,通过运用聚类算法提取煤质特征;之后,通过数据相关性分析确定碳排放量预测模型输入参数,同时采取贝叶斯网络优化神经网络参数。本文提出的预测模型融合了多种方法,旨在考虑煤质影响的同时实现碳排放量预测。

## 1 煤质聚类

### 1.1 K-means++ 聚类算法

K-means 是一种无监督聚类算法,采用聚类分析方法将样本集合分割为  $k$  个互不重叠的子集,其优化目标在于最大化簇内样本的关联度,同时确保不同子集间的样本具有显著的可区分性。每个簇由其质心表示,数据点与所属簇质心的距离越近,表明其相似度越高<sup>[17]</sup>。

K-means++ 算法是在经典 K-means 算法基础上进行优化的变体,其核心改进在于提出了更为科学的初始质心选取策略,从而提升了聚类效果。在聚类分析中,传统 K-means 方法通过随机采样确定  $k$  个初始质心,而改进后的 K-means++ 算法则引入了一种基于概率分布的优化策略来初始化聚类质心,以提高算法的收敛速度和聚类质量<sup>[18]</sup>。K-means++ 聚类算法中,数据点到其所属簇中心点用欧氏距离来表示其样本相似性度量,若 2 个数据点的欧式距离越近,则表示这 2 个点的相似度越高,对于 2 个点  $A(x_1, y_1, \dots, z_1)$  和  $B(x_2, y_2, \dots, z_2)$  在  $n$  维空间中的欧式距离  $d$  计算公式为<sup>[19]</sup>:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

式中： $d(x_i, x_j)$ 表示样本点  $x_i, x_j$  间的欧式距离； $x_i=(x_{i1}, x_{i2}, \dots, x_{ip})^T$  和  $x_j=(x_{j1}, x_{j2}, \dots, x_{jp})^T$  均表示  $p$  维数据集中的任意 2 个样本点。

肘部法则 (elbow method) 用于确定 K-means 算法的最佳聚类数<sup>[20]</sup>。基于肘部法则的聚类数确定方法即，通过分析不同聚类数  $k$  对应的误差平方和 (sum of squared errors, SSE) 变化规律来选取最优  $k$  值，随着  $k$  值的增大，SSE 呈现递减趋势。其中存在 1 个明显拐点：当  $k$  小于该临界值时，SSE 随  $k$  增加而快速下降；当  $k$  超过该阈值后，SSE 的下降幅度显著减小。这个拐点所对应的  $k$  值即为最优聚类数，它反映了在聚类效果和计算复杂度之间的最佳平衡点。SSE 作为选取算法  $k$  值的准则函数，其表达式为<sup>[21]</sup>：

$$S_{SE} = \sum_{i=1}^k \sum_{x \in X_i} \|x - \mu_i\|^2 \quad (2)$$

式中： $S_{SE}$  为该样本点的聚集程度； $k$  为聚类数； $x$  为样本集  $X_i$  的样本对象； $\mu_i$  为聚类中心点。

若使用肘部法则时发现  $k$  值较模糊，无法完全确定，可配合 CH 分数 (Calinski Harabasz Index) 进行判断。CH 分数，也称为方差比则，是一种用于评估聚类算法效果的指标，通过衡量聚类结果的紧密度和分离度来评估聚类的质量。CH 分数越高，表示聚类效果越好，其计算公式为<sup>[22]</sup>：

$$CH(k) = \frac{B_k / (k - 1)}{W_k / (N - k)} \quad (3)$$

$$B_k = \sum_{i=1}^k n_i \|\mu_i - \mu\|^2 \quad (4)$$

$$W_k = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (5)$$

式中： $W_k$  为簇内离散度； $B_k$  为簇间离散度； $N$  为所有样本总数； $k$  为簇的数量； $\mu_i$  为第  $i$  个簇的中心 (均值)； $n_i$  为第  $i$  个簇中的样本数； $\mu$  为所有数据点的全局中心 (总均值)； $C_i$  为第  $i$  个簇包含的所有样本； $x$  为簇  $C_i$  中的 1 个数据点。

## 1.2 搭建 K-means++ 聚类模型

由于该厂接收的煤炭品质随时间变化显著且煤质数据差异较大，若使用单一预测模型处理所有类型煤炭，可能导致较大误差。为提高预测准确性，先根据煤质特征进行聚类分析，将性质相近的煤炭归为一类，再针对每个簇分别进行碳排放量预测。

基于该电厂 2023 年度燃煤质量数据，本研究

选取关键煤质特征变量，运用改进型 K-means 算法对样本进行聚类分析。选择关键煤质信息时需结合实际需求<sup>[23]</sup>。低位发热量影响机组负荷和锅炉稳定运行；高硫分会导致低温腐蚀和积灰堵塞，降低热交换效率，增加燃料消耗和碳排放；高挥发分虽能提升燃烧效率，但过快释放可能引发燃烧不稳定和不完全燃烧，影响碳排放。

K-means++ 算法在使用时往往需要选定  $k$  值，而随机的选择  $k$  值会使聚类效果变差<sup>[24]</sup>。所以本文选择采取肘部法则与 CH 分数来判断合适的聚类数  $k$ 。

### 1) 肘部法则

本文取  $k=2 \sim 10$  的整数，观察 SSE 的变化情况。图 1 为 SSE 随  $k$  值的变化。分析确定了斜率变化显著的点为潜在的肘部点。在  $k=3, 4$  时，两侧的线段斜率均发生了较大变化，因此，3 和 4 都有可能成为最佳的聚类数目。为了确定最合理的聚类数目，需要进一步评估。

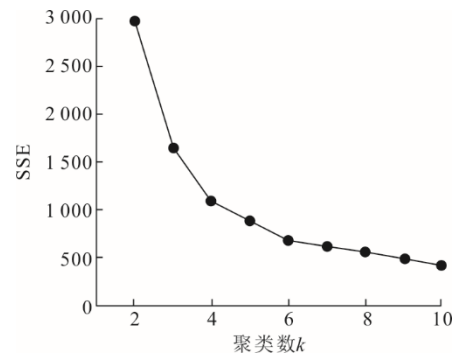


图 1 不同聚类数下的误差平方和 SSE  
Fig.1 Sum of squared errors under different numbers of clusters

### 2) CH 分数

计算采取不同聚类数下的 CH 分数值，结果如图 2。如图 2 所示， $k=4$  时的 CH 分数值大于  $k=3$  时的 CH 分数值，所以选择  $k=4$  较为合理，且聚类轮廓系数为 0.51，说明聚类效果较好。

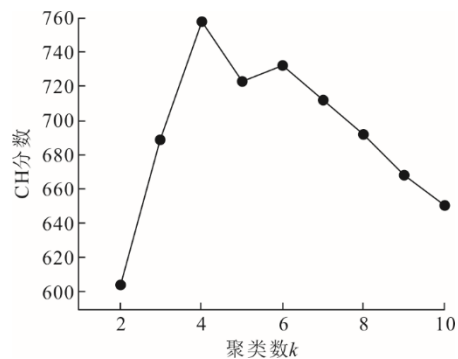


图 2 不同聚类数下的 CH 分数  
Fig.2 CH scores under different numbers of clusters

3) 聚类结果

根据肘部法则以及 CH 分数综合判断, 将该厂来煤数据聚类成 4 类效果较好, 聚类后的数据集分布三维散点图如图 3 所示。为方便区分煤质, 且为模型预测服务, 表 1 为各簇三维范围, 同时若有各簇范围重合点, 则计算该点到各簇中心点三维距离, 将该点归类为距离最小的簇。表 2 为各簇中心值。

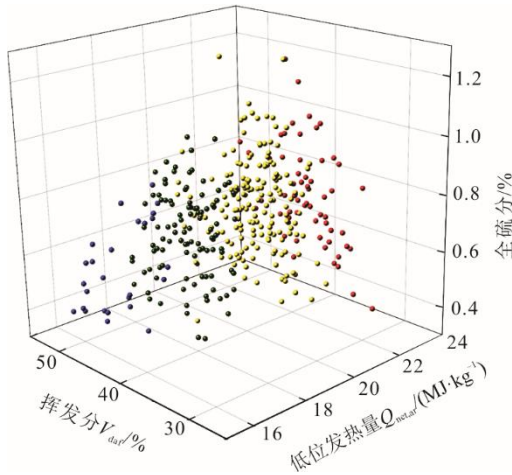


图 3 聚类后的数据集分布三维散点图  
Fig.3 3D scatter plot of clustered dataset distribution

表 1 样本聚类三维范围

Tab.1 Three dimensional range of sample clustering

簇	挥发分/%	低位发热量/(MJ·kg <sup>-1</sup> )	全硫分/%
1	[27.31,36.13]	[18.34,22.61]	[0.46,1.06]
2	[36.25,40.41]	[17.75,22.20]	[0.34,1.26]
3	[40.08,44.87]	[16.27,21.94]	[0.32,1.27]
4	[44.82,50.74]	[15.69,20.84]	[0.35,0.90]

表 2 样本聚类各簇中心值

Tab.2 Center values of each cluster in sample clustering

簇	挥发分/%	低位发热量/(MJ·kg <sup>-1</sup> )	全硫分/%
1	33.881	20.859	0.712
2	38.512	20.065	0.790
3	42.674	18.945	0.729
4	47.615	17.947	0.601

2 贝叶斯优化的 BP 神经网络预测模型

2.1 BP 神经网络模型

BP (back propagation) 神经网络由输入层、隐藏层和输出层组成, 其工作流程可分为以下几个阶段。

1) 首先, 原始数据从输入层进入系统, 经过多个隐含层的非线性变换和特征学习, 最终在输出层产生预测值。模型训练初期, 系统会对连接权重和偏置项进行随机赋值。在前向计算阶段, 输入信号通过加权求和和激活函数, 逐层传递至输出端。

2) 随后, 系统会评估预测输出与目标值之间的差异, 通常采用 L2 范数或对数似然函数作为损失函数。误差反向传播阶段则通过链式求导法则, 将输出层的误差信号逐层回传, 并利用梯度下降策略迭代调整网络参数。上述过程反复执行, 直至模型收敛或达到设定的终止条件。网络结构如图 4 所示。

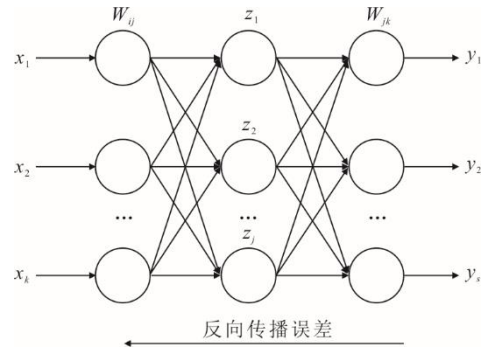


图 4 BP 神经网络结构  
Fig.4 Architecture diagram of BP neural network

BP 神经网络的输出预测结果为<sup>[25]</sup>:

$$y_n = \sum_{n=1}^s f(\sum_{i=1}^l w_{ij} - b_j) n_{jk} - a_c \quad (6)$$

式中:  $y_n$  为神经网络的输出结果;  $f$  为神经网络的激活函数;  $w_{ij}$  为第  $i$  个神经元到第  $j$  个神经元的连接权重;  $x_i$  为输入层第  $i$  个神经元的输出结果;  $b_j$  为第  $j$  个神经元的偏置;  $n_{jk}$  为隐含层第  $j$  个神经元与输出层第  $k$  个神经元的连接权值;  $a_c$  为输出层第  $c$  个神经元的阈值;  $l$  为隐藏层神经元个数;  $s$  为输出层神经元个数。

2.2 贝叶斯 (Bayesian) 优化理论

贝叶斯优化是一种高效的全局优化方法, 适用于目标函数计算成本高且无明确解析表达式的场景。其核心是通过构建概率模型 (如高斯过程) 模拟目标函数, 并智能选择采样点, 通过迭代逐步逼近全局最优解, 该优化方法在处理计算密集型或目标函数评估成本较高的复杂系统优化问题时展现出独特优势。贝叶斯优化的核心机制主要依赖于基于概率统计的高斯过程回归模型和智能采样策略 2 部分<sup>[26]</sup>。

1) 高斯过程回归

一般使用高斯过程回归作为贝叶斯优化代理模型的核心组件。作为一种非参数的贝叶斯建模方法, 高斯过程能够有效地对目标函数进行建模与预测。该方法基于一个关键假设, 即在概率统计框架下, 任意选取的有限个观测值均满足多元正态分布特性。作为高斯过程的核心要素, 协方差函数在模

型构建过程中起着决定性作用，它负责量化不同数据点之间的相似程度，从而影响模型的预测性能。高斯过程的公式为：

$$f(x) \sim GP(m(x), k(x, x')) \quad (7)$$

式中： $m(x)$ 为均值函数； $k(x, x')$ 为协方差函数(核函数)。

2) 采集函数

贝叶斯优化是一种寻找最佳参数的方法，采用 1 个代理模型（通常是高斯过程）来预测目标函数的表现，并决定下 1 个要测试的参数点。这个过程中，它尝试平衡探索（尝试新参数）和利用（改进已知好参数）来找到最好的参数组合，让目标函数达到最优状态。每次迭代，贝叶斯优化都会根据之前的测试结果更新对目标函数的理解，逐渐接近最佳的参数设置。采集函数的数学表达式为：

$$x^* = \arg \max_{x \in X} f(x) \quad (8)$$

式中： $f(x)$ 为有关于超参数  $x$  的目标函数，该目标函数可被视为 1 个黑箱函数； $x^*$ 为最优的参数集。

### 3 实时预测模型搭建及结果分析

图 5 为燃煤机组碳排放量预测模型的构建流程。首先，进行数据预处理，确保数据质量；然后识别关键参数作为输入特征，采用机器学习方法进行回归分析，建立预测模型；接着调整模型参数以提高准确性，并评估模型精度，若精度满足要求，则用于新数据预测，否则返回参数调整步骤继续优化。通过这一流程，最终建立有效的碳排放预测模型。

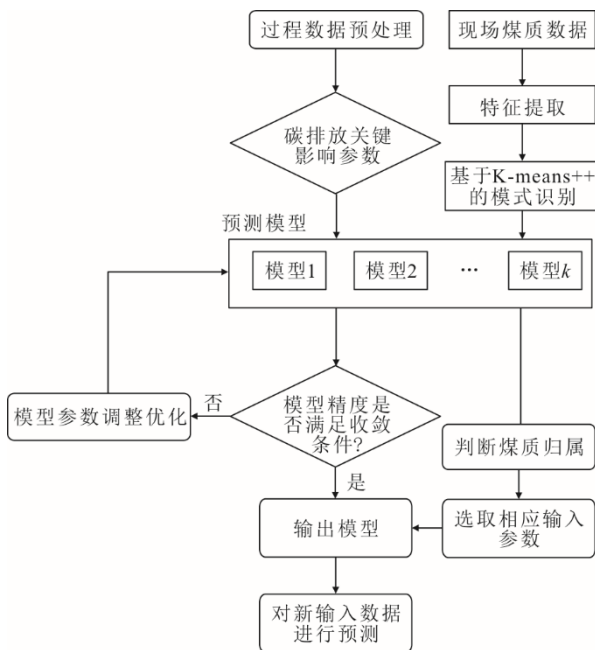


图 5 技术路线示意

Fig.5 Schematic diagram of the technical roadmap

### 3.1 碳排放量计算

鉴于该厂监测设备的数据采集频率为 1 次/min，因此碳排放量的计算时间单位亦相应设定为 1 min。

燃煤电厂标准状态下干烟气量  $Q_{gas,vent,N}$  为：

$$Q_{gas,vent,N} = t \times A \times V_{gas,vent} \times p \times T \times X \quad (9)$$

$$p = \frac{101\ 325 + p_{gas,vent}}{101\ 325} \quad (10)$$

$$T = \frac{273}{273 + T_{gas,vent}} \quad (11)$$

$$X = \frac{100 - X_{gas,vent}}{100} \quad (12)$$

式中： $Q_{gas,vent,N}$ 为标准状态下干状态下烟气排放量，万  $m^3$ （标准工况）； $t$ 为机组运行时间，s； $A$ 为烟囱排口的截面积， $m^2$ ； $V_{gas,vent}$ 、 $p_{gas,vent}$ 、 $T_{gas,vent}$ 、 $X_{gas,vent}$ 分别为烟囱排口处烟气的流速（m/s）、压力（Pa）、温度（ $^{\circ}C$ ）和湿度（%）。

根据机组实时采集的二氧化碳排放浓度均值和标干烟气量，按照式(13)计算机组每分钟二氧化碳排放量。

$$E_{CO_2} = Q_{gas,vent,N} \times 10^4 \times C_{CO_2} \times 10^4 \times 44 / 22.4 \times 10^{-9} \quad (13)$$

式中： $E_{CO_2}$ 为机组二氧化碳排放量，t/min； $C_{CO_2}$ 为二氧化碳排放体积分数均值，%； $C_{CO_2} \times 10^4$ 为二氧化碳排放质量浓度的转换系数。

### 3.2 相关性分析

为了选定燃煤电站碳排放量预测模型的输入参数，根据机理分析及相关文献选定碳排放量的主要影响因素，再采用斯皮尔曼相关系数法对该机组 2023 年实际生产运行数据进行相关性分析，量化其线性关系强度，最终识别出包括煤质特性参数和运行调控参数共计 13 个，包括表征煤质特征的低位发热量、挥发分和全硫分，以及机组负荷、主蒸汽温度、压力等运行参数。输入参数相关性结果如图 6 所示。所选参数对碳排放量的相关系数绝对值都在 0.5 以上，筛选出的参数对碳排放量有显著影响，表明其与碳排放量有较高相关性。

### 3.3 预测结果分析

为了验证该方法的性能，采集该机组 2023 年的碳排放数据（数据基本完整），其中有些数据远离正常值，需要使用四分位距法进行野点剔除异常值，所得数据保留作为样本数据。在模型训练与测试阶段，将火电机组历史数据按 8:2 划分，其中 80%

用于训练，20%用于评估模型性能。针对 BP 神经网络的超参数优化采用了贝叶斯方法，优化的参数包括初始学习率（0.000 1~0.010）以及训练迭代次

数。在此过程中选用高斯过程回归作为代理模型，并利用期望改进（expected improvement, EI）作为采集函数，迭代 50 次后收敛至最优参数组合。

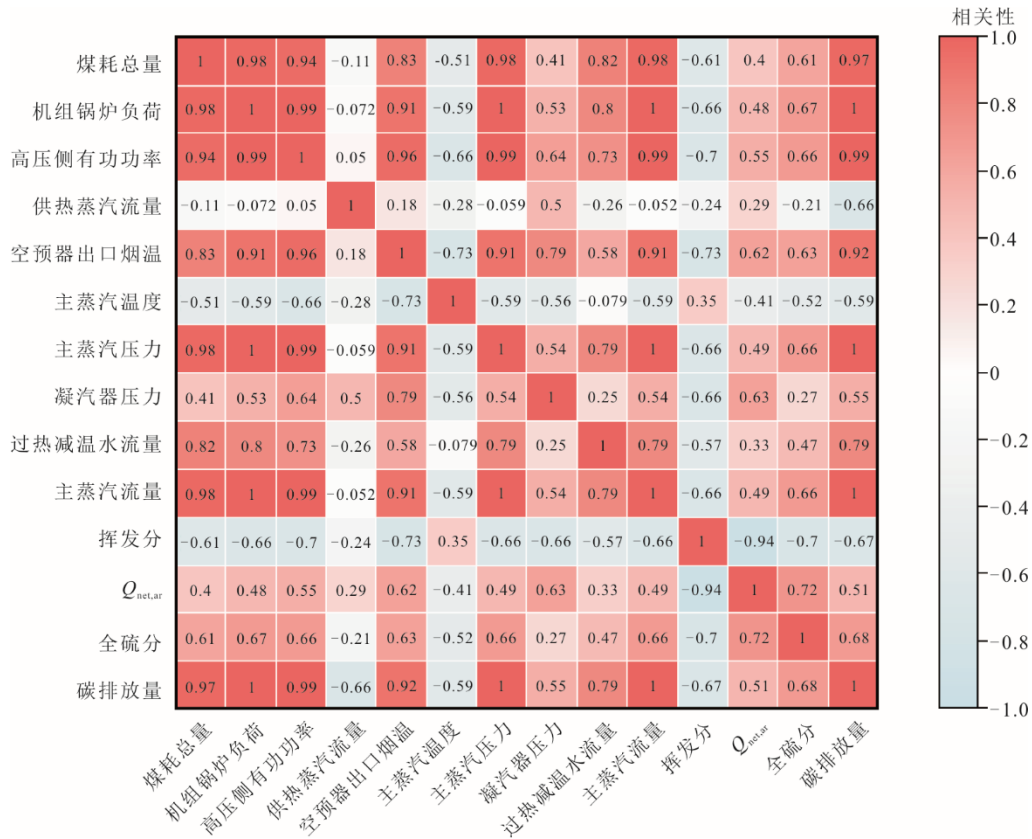


图 6 输入参数相关性结果  
Fig.6 Input parameter correlation results

为验证煤质聚类对预测结果的影响，文中随机选取 150 个数据点进行可视化展示，图 7 展示了模

型预测值与实际值的对比，反映了模型的拟合效果和预测精度。

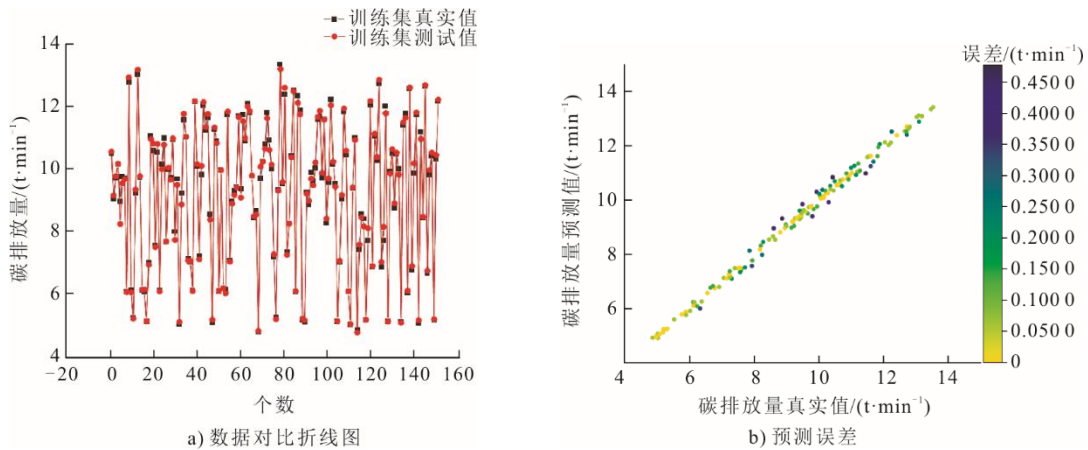


图 7 原始数据预测结果  
Fig.7 Raw data prediction results

通过贝叶斯优化搜索学习率参数空间，以验证集均方根误差（ $\delta_{RMSE}$ ）最小化为目标，确定最优学习率为 0.002 9 时，模型取得了最优性能。训练集的

相关系数  $R^2$  达到 0.983， $\delta_{RMSE}$  为 0.204 86 t/min；测试集相关系数  $R^2$  为 0.981， $\delta_{RMSE}$  为 0.218 73 t/min。模型整体预测效果良好，但分析发现其预测精度仍

受限于未考虑煤质因素而存在一定误差。可视化结果显示,未引入聚类分析的模型虽能较好地反映碳排放整体变化趋势,但在局部细节预测上仍存在一定偏差。

1) 聚类预测结果

聚类分析以全硫分、挥发分和低位发热量这三个

关键指标作为聚类维度,通过聚类算法将数据划分为4个簇,每个簇代表了具有相似燃煤特性的样本群体。聚类完成后,针对每个簇分别构建了碳排放预测模型以确保模型能更精准地捕捉不同燃料特性下的碳排放规律,同时各选用了约150个具有代表性的数据点展示预测结果,具体预测结果如图8所示。

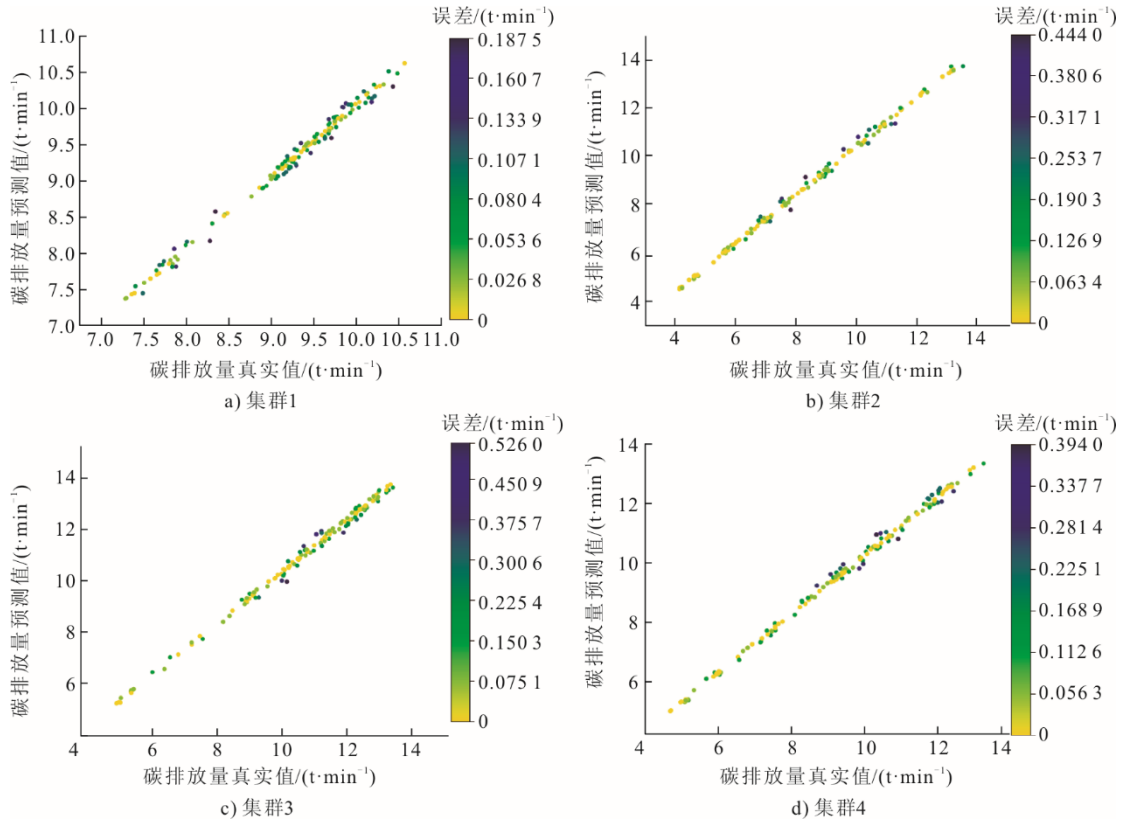


图8 聚类预测结果  
Fig.8 Cluster prediction results

为衡量提出的预测方法在火电厂机组实时碳排放预测方面的准确性,采用火电机组碳排放总量作为评价标准,并在实验过程引入 $\delta_{RMSE}$ 和平均相对误差( $\delta_{MRE}$ )函数评估碳排放预测的性能指标。评价指标对比结果见表3。

表3 预测评价指标对比

Tab.3 Comparison of predictive evaluation indicators

簇	最佳学习率	$\delta_{RMSE}/(t \cdot min^{-1})$		$\delta_{MRE}$	
		训练集	测试集	训练集	测试集
未聚类	0.002 9	0.204 9	0.218 7	0.013 5	0.014 2
1	0.000 1	0.097 2	0.100 1	0.006 1	0.008 0
2	0.000 4	0.158 1	0.153 6	0.011 1	0.011 8
3	0.000 3	0.189 9	0.183 7	0.010 2	0.011 0
4	0.001 0	0.163 4	0.166 1	0.010 9	0.012 1

由表3数据可见,采用煤质聚类方法对火电厂机组碳排放量进行预测后,得到的 $\delta_{RMSE}$ 、 $\delta_{MRE}$ 均显

著低于未采用煤质聚类的对比模型预测方法。最优的结果为簇1, $\delta_{RMSE}$ 、 $\delta_{MRE}$ 最优情况较未采用煤质聚类降低约53.4%、49.2%;较差的结果为簇3, $\delta_{RMSE}$ 、 $\delta_{MRE}$ 最优情况较未采用煤质聚类降低约11.7%、23.5%。这表明,基于煤质聚类的预测方法在降低预测误差方面表现优异,其预测结果的 $\delta_{RMSE}$ 、 $\delta_{MRE}$ 均较小,说明该方法能够更准确地捕捉燃煤机组碳排放的变化规律。煤质聚类方法通过将煤质特性相近的样本归类并分别建模,有效减少了煤质差异对预测结果的干扰,从而提升了模型的整体预测精度。因此,该方法在燃煤机组碳排放预测中具有较高的可行性。

2) 升降负荷工况预测精度比对

为进一步探究煤质聚类对碳排放预测精度的影响机制,分别选取了机组升、降负荷工况下各150min的运行数据作为分析样本。对比分析聚类前后的碳

排放量预测效果如图 9 所示, Case1 代表基于原始未聚类数据的预测结果, Case 2 则为经过煤质聚类处理后的预测值。从图 9 可以观察到, Case2 的数据与实际碳排放量值更吻合。变负荷工况的数据来源于第 1 簇, 其  $\delta_{RMSE}$  为 0.098 t/min, 相比之下未经聚类处理的预测数据  $\delta_{RMSE}$  为 0.212 t/min。这表明, 在负荷波动的情况下, 采用聚类方法处理后的预测模型对碳排放量预测的准确性高于直接使用原始数据的预测模型。

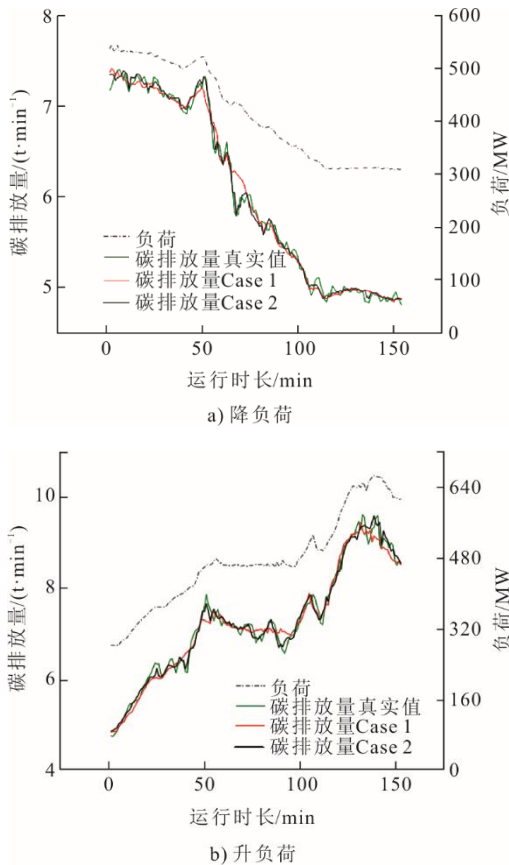


图 9 变负荷工况下电厂碳排放预测  
Fig.9 Prediction results of carbon emissions from power plants under variable load conditions

### 3.4 算例分析

为了验证本文所提方法的实际应用效果, 选取该厂 2024 年 2 月的数据进行检验, 经判断煤质分别属于簇 3、簇 4, 预测结果如图 10 所示。

从图 10 可见, Case1 和 Case2 的碳排放量预测值均呈现出与负荷变化趋势一致的波动特征, 但由于 Case 1 未考虑煤质特性的影响, 所以在负荷波动较小的工况下, 其预测结果难以准确反映碳排放量的细微变化。定量分析表明, 未进行煤质聚类的碳排放预测  $\delta_{RMSE}$  为 0.186 t/min, 而簇 3、簇 4 经过煤

质聚类处理后的预测  $\delta_{RMSE}$  分别降低至 0.151、0.165 t/min, 较未聚类模型分别降低了 18.8%和 11.3%, 预测值在负荷波动时仍能紧密跟踪实际值, 显著优于未聚类情况。这一结果不仅验证了本文所提方法对燃煤电厂新数据具有良好的适应性和预测精度, 同时也表明基于煤质聚类的神经网络模型能够更准确地捕捉碳排放的真实变化趋势。此外, K-means++算法在优化聚类中心点选取方面的改进提升了煤质聚类的质量, 从而进一步提高了碳排放量预测模型的性能。

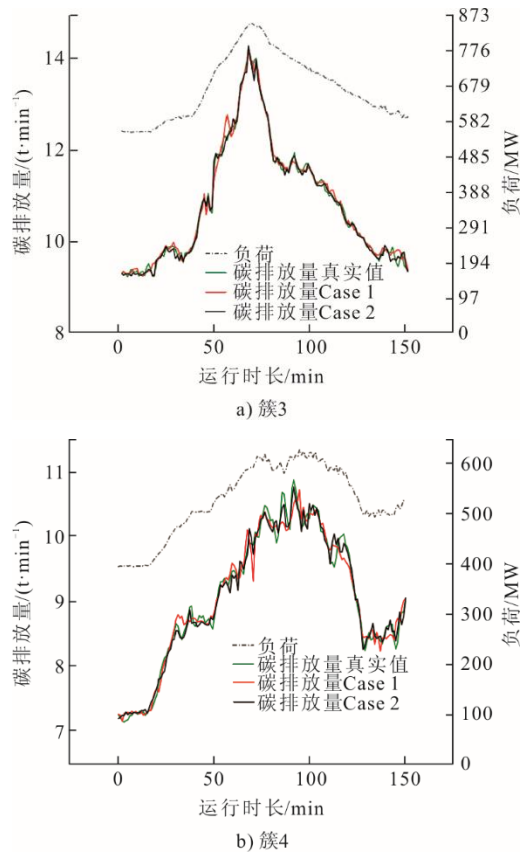


图 10 2024 年 2 月碳排放预测结果  
Fig.10 Forecast results of carbon emission in Feb. 2024

## 4 结 论

本文在考虑到燃煤电厂煤质复杂多变, 对该厂来煤数据进行聚类, 并利用神经网络算法对相应的碳排放量数据进行预测, 获得结论如下。

1) 根据煤质特征通过聚类将样本数据划分为 4 个类别, 各类别的挥发分中心值分别为 33.881%、38.512%、42.674%和 47.615%, 呈现出明显的梯度分布特征, 且各中心值间的显著差异 (最大差值达 13.73 百分点) 充分证实了样本数据存在明显的差异性。这一结果从量化角度证实了数据分类的必要

性和合理性。

2) 基于聚类结果将数据集划分为 4 个簇, 并对每个簇分别进行建模。相较于整体数据集统一建模的方法, 分工况建模后的预测误差有所减少, 模型 RMSE 也有所降低, 表明聚类方法在碳排放量预测中展现出了其有效性和优势。

3) 对某百万千瓦机组 2024 年 2 月实际运行数据的预测实验表明, 采用聚类方法的模型在新数据上的预测误差明显低于未聚类的模型, 进一步验证了本文所提方法的有效性和实际应用价值。未来研究可进一步探索更多影响碳排放的因素, 并结合其他先进的机器学习算法, 不断提升预测模型性能。

### [参 考 文 献]

- [1] WANG S, LIN B. Performance linkage in renewable energy supply chain: a comparative analysis with coal power and the entire industry[J]. *Energy Policy*, 2024, 193: 114274.
- [2] 张小丽, 崔学勤, 王克, 等. 中国煤电锁定碳排放及其对减排目标的影响[J]. *中国人口·资源与环境*, 2020, 30(8): 31-41.  
ZHANG Xiaoli, CUI Xueqin, WANG Ke, et al. China's coal power locks in carbon emissions and its impact on emission reduction targets[J]. *China's Population, Resources and Environment*, 2020, 30(8): 31-41
- [3] DU L, ZHAO H, TANG H, et al. Analysis of the synergistic effects of air pollutant emission reduction and carbon emissions at coal-fired power plants in China[J]. *Environmental Progress & Sustainable Energy*, 2021, 40(5): e13630.
- [4] 陈怡, 田川, 曹颖, 等. 中国电力行业碳排放达峰及减排潜力分析[J]. *气候变化研究进展*, 2020, 16(5): 632-640.  
CHEN Yi, TIAN Chuan, CAO Ying, et al. Analysis of peak carbon emissions and emission reduction potential in China's power industry[J]. *Progress in Climate Change Research*, 2020, 16(5): 632-640
- [5] 张丽. 火电厂碳排放的计算方法及降低碳排放的有效调整方法[J]. *中外企业家*, 2016(15): 228.  
ZHANG Li. Calculation methods for carbon emissions in thermal power plants and effective adjustment methods for reducing carbon emissions[J]. *Chinese and Foreign Entrepreneurs*, 2016(15): 228.
- [6] 傅文明. 火电厂锅炉混煤掺烧技术及节能运行措施研究[J]. *自动化应用*, 2024, 65(18): 25-27.  
FU Wenming. Research on coal mixing and combustion technology and energy saving operation measures for thermal power plant boilers[J]. *Automation Application*, 2024, 65(18): 25-27.
- [7] 安吉振, 陈衡, 乔世超, 等. 基于 K-means 聚类算法的电站煤场来煤堆放优化研究[J]. *热力发电*, 2023, 52(4): 135-143.  
AN Jizhen, CHEN Heng, QIAO Shichao, et al. Research on optimization of coal storage in power plant coal yard based on K-means clustering algorithm[J]. *Thermal Power Generation*, 2023, 52(4): 135-143.
- [8] 曹歌瀚, 黄亚继, 徐文韬, 等. 基于机器学习的燃煤锅炉分工况建模与燃烧优化[J]. *锅炉技术*, 2023, 54(5): 41-47.  
CAO Gehan, HUANG Yaji, XU Wentao, et al. Modeling and combustion optimization of coal-fired boilers based on machine learning for different operating conditions[J]. *Boiler Technology*, 2023, 54(5): 41-47
- [9] 文雯, 刘文哲, 肖祥武, 等. 基于大数据和并行随机森林算法火电机组供电煤耗计算模型[J]. *热力发电*, 2018, 47(9): 9-14.  
WEN Wen, LIU Wenzhe, XIAO Xiangwu, et al. A coal consumption calculation model for power supply of thermal power units based on big data and parallel random forest algorithm[J]. *Thermal Power Generation*, 2018, 47(9): 9-14.
- [10] 郭魁星, 李姚旺, 何晓宜, 等. 面向用电碳计量的备用碳表系统优化配置方法[J]. *电力系统自动化*, 2025, 49(6): 1-13.  
GUO Kuixing, LI Yaowang, HE Xiaoyi, et al. Optimization configuration method for backup carbon meter system for electricity carbon measurement[J]. *Power System Automation*, 2025, 49(6): 1-13.
- [11] 董锐锋, 陈浩军, 曹振谦, 等. 燃煤锅炉碳排放核查方法及对比[J]. *工业炉*, 2021, 43(2): 41-44.  
DONG Ruifeng, CHEN Haojun, CAO Zhenqian, et al. Methods and comparison of carbon emission verification for coal fired boilers[J]. *Industrial Furnace*, 2021, 43(2): 41-44.
- [12] PROKHORSKII G, RUDRA S, PREIBINGER M, et al. A data-driven regression model for predicting thermal plant performance under load fluctuations[J]. *Carbon Neutrality*, 2024, 3(1): 1-15.
- [13] 袁鹏, 谭琛, 李锋, 等. 基于迁移深度强化学习的火电机组实时碳排放预测方法[J]. *工业加热*, 2024, 53(7): 65-69.  
YUAN Peng, TAN Chen, LI Feng, et al. Real time carbon emission prediction method for thermal power units based on transfer deep reinforcement learning[J]. *Industrial Heating*, 2024, 53(7): 65-69.
- [14] WANG X, YAN C, LIU W, et al. Research on carbon emissions prediction model of thermal power plant based on SSA-LSTM algorithm with boiler feed water influencing factors[J]. *Sustainability*, 2022, 14: 15988.
- [15] CHEN J, ZHENG L, CHE W, et al. A method for measuring carbon emissions from power plants using a CNN-LSTM-Attention model with Bayesian optimization[J]. *Case Studies in Thermal Engineering*, 2024, 63: 105334.
- [16] LU C, LI W, GAO S. Driving determinants and prospective prediction simulations on carbon emissions peak for China's heavy chemical industry[J]. *Journal of Cleaner Production*, 2020, 251: 119642.
- [17] 卢青. 基于改进的 K-means 聚类算法的火电厂锅炉燃烧优化研究[D]. 秦皇岛: 燕山大学, 2015: 1.  
LU Qing. Research on combustion optimization of thermal power plant boilers based on improved k-means clustering algorithm[D]. Qinhuangdao: Yanshan University, 2015: 1.
- [18] ARTHUR D, VASSILVITSKII S. K-Means++: the advantages of careful seeding[C]// *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2007: 1027-1035.
- [19] 孙林, 刘梦含, 徐久成. 基于优化初始聚类中心和轮廓系数的 K-means 聚类算法[J]. *模糊系统与数学*, 2022, 36(1): 47-65.  
SUN Lin, LIU Menghan, XU Jiucheng. K-means

- clustering algorithm based on optimizing initial clustering centers and contour coefficients[J]. *Fuzzy Systems and Mathematics*, 2022, 36(1): 47-65.
- [20] 何选森, 何帆, 于海澜. K-均值算法的初始化改进与聚类质量评估[J]. *西安工程大学学报*, 2024, 38(6): 114-123.  
HE Xuansen, HE Fan, YU Hailan. Initialization improvement and clustering quality evaluation of K-means algorithm [J]. *Journal of Xi'an Engineering University*, 2024, 38(6): 114-123.
- [21] 王仲, 顾煜炯, 韩旭东, 等. 基于历史数据聚类的火电机组工况划分[J]. *仪器仪表学报*, 2019, 40(2): 90-95.  
WANG Zhong, GU Yujiong, HAN Xudong, et al. Classification of thermal power unit operating conditions based on historical data clustering[J]. *Journal of Instrumentation and Design*, 2019, 40(2): 90-95.
- [22] 安吉振. 火电机组燃料系统运行优化及故障预警方法研究[D]. 北京: 华北电力大学, 2023: 1.  
AN Jizhen. Research on optimization of fuel system operation and fault warning methods for thermal power units[D]. Beijing: North China Electric Power University (Beijing), 2023: 1.
- [23] 雷萌, 李翠, 王鑫, 等. 基于模糊聚类径向基神经网络的煤质分析模型[J]. *选煤技术*, 2015(4): 10-14.  
LEI Meng, LI Cui, WANG Xin, et al. Coal quality analysis model based on fuzzy clustering radial basis function neural network[J]. *Coal Preparation Technology*, 2015(4): 10-14.
- [24] 贺思云, 高建瓴, 陈岚. 基于改进人工蜂群算法的 K-means 聚类算法[J]. *贵州大学学报(自然科学版)*, 2017, 34(5): 83-87.  
HE Siyun, GAO Jianling, CHEN Lan. K-means clustering algorithm based on improved artificial bee colony algorithm[J]. *Journal of Guizhou University (Natural Science Edition)*, 2017, 34(5): 83-87.
- [25] 龚广京, 周光, 郑涛, 等. 基于线性回归与 BP 神经网络的火电厂燃煤碳排放计算研究[J]. *热能动力工程*, 2024, 39(3): 73-81.  
GONG Guangjing, ZHOU Guang, ZHENG Tao, et al. Research on carbon emission calculation of coal fired power plants based on linear regression and bp neural network[J]. *Thermal Power Engineering*, 2024, 39(3): 73-81.
- [26] WANG G, JIA R, LIU J, et al. A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning[J]. *Renewable Energy*, 2020, 145: 2426-2434.

(责任编辑 杨嘉蕾)