

DOI: 10.19666/j.rlfed.202206115

# 基于 $K$ -means 聚类算法的电站煤场 来煤堆放优化研究

安吉振, 陈 衡, 乔世超, 潘佩媛, 徐 钢

(华北电力大学热电生产过程污染物监测与控制北京市重点实验室, 北京 102206)

**[摘 要]** 为有效应对复杂的电力和煤炭市场, 加强燃料智慧管理成为火电厂管理的重要组成部分。针对某燃煤电厂的煤场占地面积小、来煤煤种复杂和燃煤堆放混乱的问题, 通过提取历史来煤的煤质信息, 使用  $K$ -means 和 DBSCAN 2 种聚类算法对来煤低位发热量、挥发分、硫分进行聚类分析, 从轮廓系数、聚类稳定性和样本划分精细度 3 个方面对 2 种聚类算法进行对比, 最终选择聚类效果更好的  $K$ -means 聚类算法作为煤质划分的计算方法。 $K$ -means 聚类算法将选取的历史煤质信息数据集划分为 4 类, 轮廓系数为 0.587, 且划分后的同一类别内煤质成分相近。统计不同聚类标签下的来煤频率和来煤质量比例, 对煤场进行了相应的比例划分, 每一分区堆放相同分类的来煤, 并以此为基础设计了数字化煤场平台中的来煤堆放指导及信息存储流程, 对提高堆场空间利用率和煤场管理效率具有重要的意义。

**[关 键 词]** 燃料智慧管理; 煤场分区; 来煤堆放;  $K$ -means 聚类; 最优方案

**[引用本文格式]** 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 incoming coal stacking in power station coal yard based on  $K$ -means clustering algorithm[J]. Thermal Power Generation, 2023, 52(4): 135-143.

## Research on optimization of incoming coal stacking in power station coal yard based on $K$ -means clustering algorithm

AN Jizhen, CHEN Heng, QIAO Shichao, PAN Peiyuan, XU Gang

(Beijing Key Laboratory of Emission Surveillance and Control for Thermal Power Generation, North China Electric Power University, Beijing 102206, China)

**Abstract:** In order to effectively deal with the complex electricity and coal market, strengthening the smart fuel management has become an important part of thermal power plant management. Aiming at solving the problems that the coal yard of a coal-fired power station occupies small area, the types of incoming coal are complex, and the coal-fired coal stacking is chaotic, by extracting the coal quality information of historical incoming coal,  $K$ -means and DBSCAN clustering algorithms are used to analyze the low-level coal. The calorific value, volatile matter and sulfur content are clustered and analyzed, and the two clustering algorithms are compared from the perspective of silhouette coefficient, cluster stability and sample division fineness, and finally  $K$ -means with better clustering effect is selected as the calculation method for coal quality division. The  $K$ -means algorithm divides the selected historical coal quality information data set into four categories, the contour coefficient is 0.587, and the coal quality components in each category are similar. The incoming coal frequency and the incoming coal weight ratio under different cluster labels are counted, and the coal yard is divided into corresponding proportions. The incoming coal of the same classification is stacked in each partition, and on this basis, the incoming coal in the digital coal yard platform is designed. Coal stacking guidance and information storage process are of great significance to improving the utilization of storage yard space and the efficiency of coal yard management.

**Key words:** smart fuel management; coal yard division; incoming coal stacking;  $K$ -means clustering; best plan

收稿日期: 2022-06-14 网络首发日期: 2023-01-12

基金项目: 国家自然科学基金青年项目(52106008); 国家自然科学基金创新研究群体项目(51821004)

Supported by: National Natural Science Foundation of China Youth Program (52106008); National Natural Science Foundation of China Innovative Research Group Project (51821004)

第一作者简介: 安吉振(1997), 男, 硕士研究生, 主要研究方向为电力大数据分析 & 智能优化、设备故障诊断及智能预警, ajz@ncepu.edu.cn。

通信作者简介: 陈衡(1989), 男, 博士, 副教授, 主要研究方向为热力系统优化、电力大数据分析 & 智能优化, heng@ncepu.edu.cn。

近年来,随着“碳达峰、碳中和”发展战略的实施,风力、光伏等可再生能源发电发展迅速,但是火电机组仍然承担着大部分的发电任务<sup>[1]</sup>。截至2021年末,全国发电装机总容量237 692万kW,其中火电机组装机容量129 678万kW,占总装机容量的50%以上<sup>[2]</sup>。在我国火电领域,发电与燃煤供应之间的矛盾较为突出,主要问题是燃烧煤质煤种不稳定且偏离设计值,从而在一定程度上影响火电机组运行的安全性、经济性和环保性<sup>[3]</sup>。加强燃料智慧管理是火电企业安全稳定生产、有效应对复杂严峻的电力和煤炭市场形势、提高火电企业盈利能力的手段之一<sup>[4-6]</sup>。截至目前,华能德州电厂、沁北电厂、辛店电厂等多家火电厂已经完成了智慧燃料系统改造,智慧燃料系统涵盖了斗轮机无人值守、精准配煤管理、输煤皮带DCS改造、煤场分区管理等板块,实现了燃煤从入厂到入炉全过程严格管控,提高了配煤掺烧准确率,最终降低了发电成本,提高了在电力市场的竞争力。

然而,目前国内仍有许多火电厂的煤场管理以“人治”为主,来煤堆放混乱,管理方式比较落后,传统来煤堆放主要是依靠煤场现场工程师的经验,操作斗轮机将煤堆放到一个区域,这样会造成煤场的堆放杂乱无章,性质差别大的煤有可能会堆放在一起<sup>[7]</sup>。以北方某电厂的3期煤场改造为对象,开发数字煤场平台中的来煤智能堆放板块。该煤场初为露天条形煤场,2020年4月底完成煤场封闭改造项目,主要为南北走向的条形煤棚,封闭煤棚跨度108 m,长度236 m,总高度38.76 m,煤场堆高上限为14 m,可储煤10.5万t。该电厂现有工程燃煤来自太原、阳泉、寿阳、朔州、榆林等处,来煤煤种复杂,而且煤场较小,对于煤种多且以火车、汽车运输方式为主的电厂,很难实现按煤种分区堆放来煤<sup>[8-9]</sup>。

本文采用合理的聚类算法对历史来煤的煤质信息进行聚类,然后统计不同种类来煤的频率以及来煤总量,从而对煤场进行比例分区,每个分区堆放煤质成分相差不大的来煤,煤场管理人员可以直观、清晰地了解煤场各区存煤情况。通过对火车来煤计划的采集,所设计的数字煤场平台读取到来煤信息后,自动对来煤进行分类,并指导现场工程师操作斗轮机将来煤堆放到分好的区域,提高来煤堆放效率,同时也可以为配煤环节的工作人员提供直观的数据指导,在一定程度上提高了煤场管理的科

学性,推动煤电协同高效、灵活发展。

## 1 聚类算法介绍

聚类算法是数据挖掘中应用最广泛的技术之一,在信息检索、图像识别、故障预警等领域取得了一定的成果<sup>[10]</sup>。它属于一种无监督学习方法,对样本数据集按某个标准或数据的内在性质及规律,将样本划分为若干个不相交的子集,每个子集称为一个簇,使得每一簇中的样本数据尽可能相似,不同簇中的数据信息差异尽可能大<sup>[11]</sup>。聚类算法通常有划分法、密度法、层次法、网格法等<sup>[12-13]</sup>。

### 1.1 K-means 聚类算法介绍

K-means 聚类算法是具有代表性的聚类方法,它在处理大规模数据集时可以保持良好的伸缩性和高效性<sup>[14]</sup>。该算法是一种基于距离的聚类算法,采用样本间的距离作为样本相似性度量,其主要思想是从数据集中随机选择 $k$ 个样本对象,作为初始聚类中心,计算每个样本点到聚类中心的距离,通过迭代过程逐次更新聚类中心的位置,直到更新的聚类中心不再变化或变化小于某个阈值时,迭代算法终止,最终将整个数据集划分为 $k$ 个不同聚类。它适用于处理凸数据集或球形簇。具体的算法流程框图如图1所示。

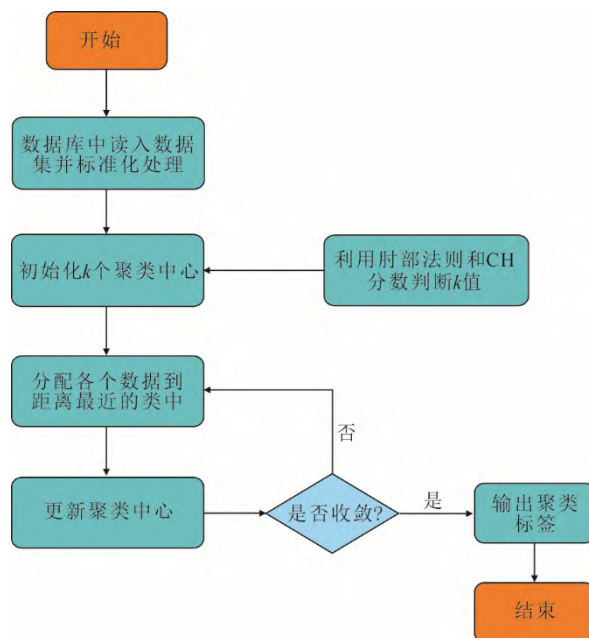


图1 K-means 聚类算法流程  
Fig.1 Flow chart of K-means clustering algorithm

K-means 聚类算法使用欧氏距离作为样本相似性度量,样本点的距离越近,表明这2个样本点的

相似性越大。样本点之间的欧式距离公式为<sup>[11]</sup>:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

式中:  $d(\mathbf{x}_i, \mathbf{x}_j)$  表示样本点  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间的欧式距离;  
 $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})^T$  和  $\mathbf{x}_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jp})^T$  均表示  $p$  维数据集中的任意 2 个样本点。

使用误差平方和作为 K-means 算法的准则函数。其表达式为<sup>[15]</sup>:

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

式中:  $k$  为聚类数;  $x$  为聚类集  $X_i$  中的样本对象;  $\mu_i$  为聚类中心;  $S_{SE}$  为样本点的密集程度, 理论上来说,  $S_{SE}$  的值越小, 聚类效果越好。

## 1.2 DBSCAN 聚类算法介绍

DBSCAN (density based spatial clustering of applications with noise) 聚类算法是一种典型的**基于密度的聚类方法**, 具有良好的抗干扰性, 对噪声点和离群点的敏感性较低<sup>[16-17]</sup>。该算法的关键是要设定合理的邻域半径  $E_{ps}$  和密度阈值  $M_{inPts}$ 。主要流程是先从数据集中任选 1 个未被处理过的样本点, 如果此点的  $E_{ps}$  邻域内数据点的数量大于等于设定的  $M_{inPts}$ , 则该点就被标记为核心点; 如果该点  $E_{ps}$  邻域内数据点的数量小于  $M_{inPts}$ , 但是落在核心点的邻域内, 则该对象为边界点, 否则该点即为噪声点<sup>[18]</sup>。第 1 个核心点将形成 1 个簇, 并将所有与此核心点直接密度可达的未归类点添加到此簇中, 直到没有符合要求的点可以添加。图 2 为 DBSCAN 聚类算法原理示意。图 2 中红色点为核心点, 黄色点为边界点, 蓝色点为噪声点。

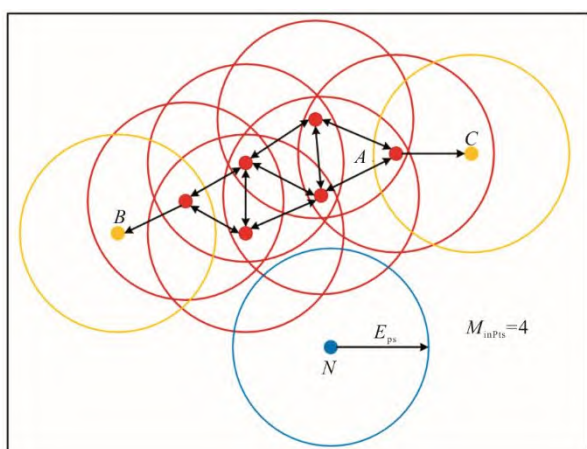


图 2 DBSCAN 聚类算法原理示意  
 Fig.2 Schematic diagram of the DBSCAN clustering principle

## 2 聚类模型建立

### 2.1 数据来源与处理

聚类模型是针对此火电厂的来煤智能堆放设计的, 历史煤质信息来源于燃料平台的数据库备份文件, 利用 SQL Server 软件将此备份文件还原, 并利用 Python 软件连接此数据库文件直接读取需要的数据。选择读取“化验管理”表格用于获取历史煤质化验信息, 主要包括煤批批号、全水分、收到基低位发热量、水分、挥发分、全硫分、灰分、灰熔点等测点数据; 读取“煤批管理”表格, 用于获取不同批次来煤的煤矿名称、矿发量、矿发时间等信息。

由于该煤场的来煤煤种复杂且无规律, 统计起来较为困难, 因此可以统计历史来煤的关键煤质信息, 并对其分类。所选择的关键煤质信息要结合实际需求, 如发热量是燃煤重要的指标, 它直接影响机组负荷, 保证锅炉炉膛安全稳定的运行; 挥发分的含量会影响燃煤的着火特性和燃烧特性, 挥发分含量相差较大的煤应分开堆放; 硫分是煤质评价的重要指标之一, 其含量会直接影响燃烧后排放的含硫污染物是否超标; 煤的灰熔点和灰分含量会直接影响锅炉结渣、结焦等<sup>[19]</sup>。

该火电厂在配煤掺烧时, 重点关注的是混煤发热量、挥发分和硫分, 因此本文选择这 3 个指标作为历史来煤分类统计的依据。以来煤批次号为索引, 选取 2021 年全年来煤的上述 3 个煤质指标作为聚类分析的数据集, 本年度 3 期煤场共进 858 批次来煤, 每一批来煤的煤质信息作为数据集中的 1 个样本, 数据集分布的三维散点图如图 3 所示。

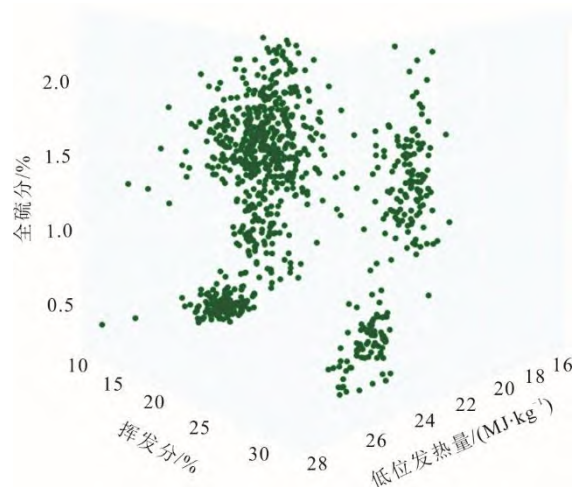


图 3 关键煤质信息数据集分布三维散点图  
 Fig.3 Scatter distribution of key coal quality information



聚类是一种距离分类方法,为了避免量纲对分类的影响,同时提高模型的收敛速度,常对建模数据进行线性处理到更小的变换区间。采用最大最小值归一化方法,将数据集的每一维数据缩放到 0~1 范围内。

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

式中:  $x$  是样本观测值;  $x_{\max}$  是数据的最大值;  $x_{\min}$  是数据最小值;  $x^*$  是归一化后的数据。

## 2.2 建立 K-means 聚类模型

由于 K-means 聚类算法要事先指定  $k$  值,但是  $k$  值很多时候靠人们的经验来确定,这就造成很多时候分类标准并不准确。若  $k$  值过小,则类与类之间的相似度过大;若  $k$  值过大,类与类之间的相似度过小。这 2 种情况都会造成聚类结果不准确<sup>[20-21]</sup>。为了选择合适的聚类数  $k$ ,本文选择肘部法则(K-SSE)和 CH 分数(Calinski-Harabasz)2 种方法综合判断。为了找到合适的聚类数  $k$ ,使  $k$  遍历 2~10 的整数值,计算出  $k$  取不同值的时候对应的聚类指标值,从而综合判断出  $k$  的合理取值。

### 1) 肘部法则判断

随着聚类数  $k$  值的增大,样本划分会更加精细,每个簇的聚合程度会逐渐提高,那么误差平方和  $S_{SE}$  会逐渐变小。当  $k$  小于最优聚类数时,由于  $k$  的增大会大幅增加每个簇的聚合程度,故  $S_{SE}$  的下降幅度会很大;而当  $k$  达到最优聚类数后,再增加  $k$  值  $S_{SE}$  的下降幅度会变缓,即  $S_{SE}$  和  $k$  的关系图类似一个手肘的形状,而这个肘部对应的  $k$  值就是样本数据集的最优聚类数<sup>[22]</sup>。图 4 为聚类数目  $k$  取不同值时对应的误差平方和  $S_{SE}$ 。

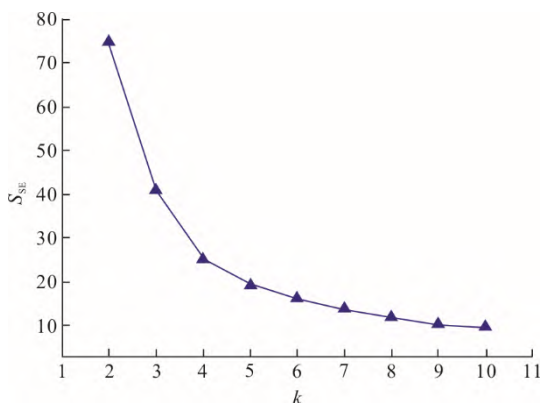


图 4 不同聚类数下的误差平方和  
Fig.4 Error sum of squares with different number of clusters

通过计算每一条折线段的斜率,找到两边斜率变化较大的点作为肘部点,经计算  $k=3$  和  $k=4$  点处两边的折线斜率变化都较大。因此 3 和 4 都有可能是最佳聚类数目,要确定最合理的聚类数目,还需要参考聚类后的 CH 分数值大小。

### 2) CH 分数判断

CH 分数是数据集的分离度与紧密度的比值,通过计算类中各点与类中心的距离平方和来度量类内的紧密度,通过计算各类中心点与数据集中心点距离平方和来度量数据集的分离度。CH 分数由分离度与紧密度的比值得到。CH 越大表示类自身越紧密,类与类之间越分散,即聚类结果更优<sup>[23]</sup>。其工作原理为:利用聚类算法将数据集  $E$  划分为  $k$  类,则其 CH 分数  $s$  被定义为簇间离散度与簇内离散度的比值。其表达式为:

$$s = \frac{\text{tr}(\mathbf{B}_k)}{\text{tr}(\mathbf{W}_k)} \times \frac{n_E - k}{k - 1} \quad (4)$$

$$\mathbf{W}_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T \quad (5)$$

$$\mathbf{B}_k = \sum_{q=1}^k n_q (c_q - c_E)(c_q - c_E)^T \quad (6)$$

式中:  $\mathbf{W}_k$  为簇内离散矩阵;  $\mathbf{B}_k$  为簇间离散矩阵;  $n_E$  为数据集  $E$  的样本点数;  $c_E$  为数据集  $E$  的中心点;  $C_q$  为聚类集  $q$  的点集;  $c_q$  为聚类集  $q$  的中心点;  $n_q$  为聚类集  $q$  的样本点数;  $\text{tr}$  表示矩阵的迹。

计算不同聚类数下的 CH 分数值(图 5),  $k=4$  时 CH 值最大,  $k=3$  时 CH 值还比较小,聚类效果不如  $k=4$ , 因此选择聚类数  $k=4$  更为合理。

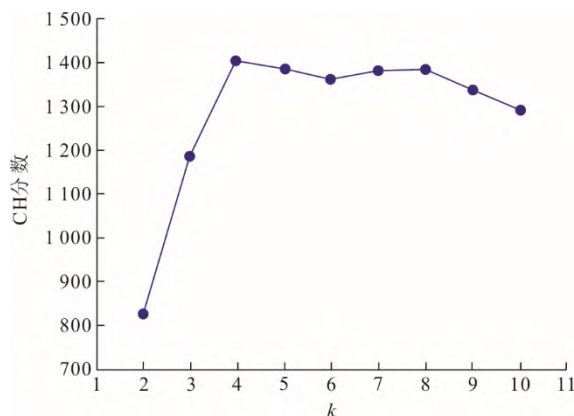


图 5 不同聚类数目下的 CH 分数值  
Fig.5 CH scores with different number of clusters

根据肘部法则以及 CH 分数综合判断,将历史来煤聚类成 4 类是比较合理的,因此利用 2021 年全年批次来煤的化验信息,以低位发热量、挥发分、

全硫分为聚类维度,利用 K-means 聚类算法进行聚类分析。为了直观展示聚类结果,利用 Python 软件的 Scatter 3D 功能作出聚类后的散点图,相同的聚类类别采用同颜色点示出,如图 6 所示。

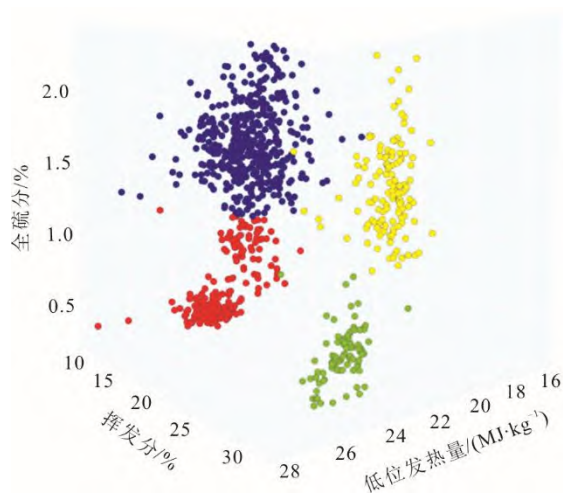


图 6 聚类后的煤质信息散点分布  
Fig.6 Scattered distribution of coal quality information after clustering

无监督聚类效果评价的指标有很多,如轮廓系数、邓恩指数、兰德指数等。轮廓系数是最常用评估聚类效果的指标,它结合了聚类样本的内聚度和分离度 2 种因素,可以在原数据的基础上评价聚类算法<sup>[24]</sup>。单个样本的轮廓系数计算式为:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (7)$$

式中:  $a(i)$  为该样本到同簇其他样本的平均距离,  $a(i)$  越小,说明该样本越应该被聚类到该簇;  $b(i)$  为该样本到其他所有簇的样本的平均距离中最小的一个,  $b(i)$  越大说明该样本越不属于其他簇。

轮廓系数的取值范围为-1~1, 0 附近的值表示重叠的聚类, 负值表示样本已被分配到错误的集群, 轮廓系数越接近 1, 其聚类效果越好。经计算本文 K-means 聚类模型的轮廓系数为 0.587, 聚类效果良好。

### 2.3 建立 DBSCAN 聚类模型

与 K-means 聚类算法不同, 基于密度聚类的 DBSCAN 聚类算法并不需要预先指定聚类数目和设置初始聚类中心, 它最大的优势是可以发现任意形状的聚类簇<sup>[25]</sup>。此聚类模型的关键是选取合理范围的邻域半径  $E_{ps}$  和密度阈值  $M_{inPts}$ , 将标准化后的煤质数据代入 DBSCAN 聚类模型中, 在合理范围内迭代不同的  $E_{ps}$  和  $M_{inPts}$  数值, 计算不同组合参数

下模型的轮廓系数值, 利用 Python 软件中的可视化图形包 Seaborn 作出上述轮廓系数值的散点分布。其中散点的大小代表数值的大小, 点越大代表轮廓系数越大。图 7 为不同组合参数下的轮廓系数。

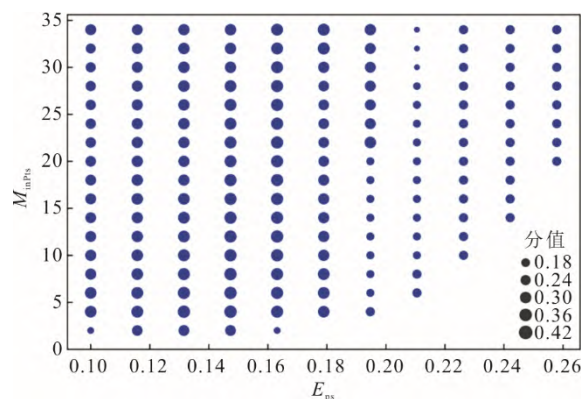


图 7 不同组合参数下的轮廓系数  
Fig.7 Silhouette coefficients under different combination parameters

噪声比是在 DBSCAN 聚类运算中, 数据集中被标记为噪声点的数目占总样本点的比例。利用同样的方法计算不同组合参数下的噪声比, 并用 Seaborn 作出噪声比的散点分布, 如图 8 所示。

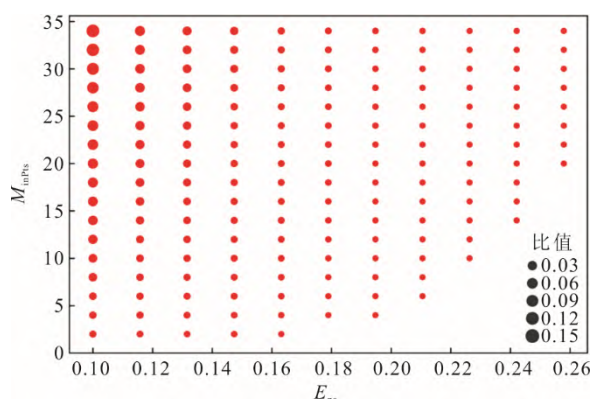


图 8 不同组合参数下的噪声比  
Fig.8 Noise ratios under different combinations of parameters

要选择合理范围的邻域半径  $E_{ps}$  和密度阈值  $M_{inPts}$ , 就要使得模型的轮廓系数尽可能的大, 噪声比尽可能的小。综合图 7 和图 8 来看,  $E_{ps}$  取值范围为 0.14~0.18,  $M_{inPts}$  取值范围为 5~25 时比较合适, 可以取  $E_{ps}=0.15$ ,  $M_{inPts}=20$ , 做出此组合参数下的密度聚类散点图, 如图 9 所示。图 9 中, 蓝色为噪声点, 共计 35 个, 噪声比为 4% 左右, 分类数目为 3, 计算出轮廓系数为 0.425。

### 2.4 2 种算法结果对比

把 2021 年全年来煤的低位发热量、挥发分、全硫分作为样本数据集, 分别采用基于划分的

$K$ -means 聚类算法和基于密度的 DBSCAN 聚类算法进行聚类运算,对二者的聚类结果进行对比分析得到如下结论:

1)  $K$ -means 聚类算法的轮廓系数为 0.587, DBSCAN 聚类算法的轮廓系数为 0.425,  $K$ -means 聚类的效果更好一些;

2) DBSCAN 聚类算法的噪声点过于分散且不稳定,不适宜单独作为一类或几类煤堆放;

3) DBSCAN 聚类算法将主体部分划分为 3 类,分类数目过少,会造成同类煤质信息差异过大,划分过于粗糙。

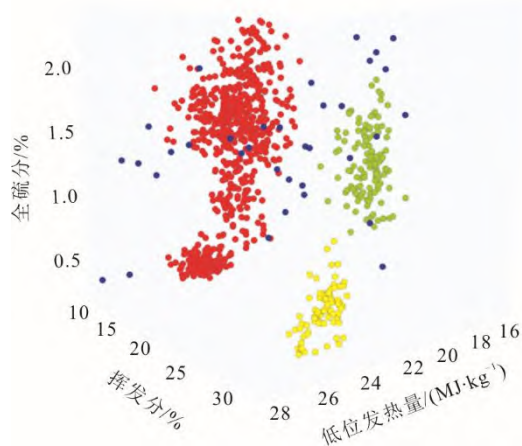


图 9 DBSCAN 聚类后的煤质信息散点分布  
Fig.9 Scatter distribution of coal quality information after DBSCAN clustering

综上对比,最终选择  $K$ -means 聚类算法对历史煤质信息进行聚类分析。虽然 2021 年的关键煤质信息散点分布不是严格的凸数据集,但是采用  $K$ -means 聚类算法对其进行聚类运算仍有良好的效果。统计  $K$ -means 聚类分析后的不同类别的煤质历史信息,见表 1。

表 1 不同类别的煤质信息范围  
Tab.1 Range of coal quality information for different categories

标签 (随机)	低位发热量/ (MJ·kg <sup>-1</sup> )	挥发分/%	全硫分/%	样本数
0	16.26~26.40	9.05~19.36	1.11~2.43	398
1	17.44~22.63	20.42~28.80	0.81~2.22	140
2	18.72~27.53	9.16~17.14	0.23~1.10	243
3	21.37~26.09	24.34~31.44	0.26~0.89	77

### 3 煤场分区设计

为了保证电厂稳定运行,煤场必须储备一定量的燃煤,存煤量要依据电厂负荷、机组运行状况、

锅炉性质、输煤调运状况等诸多因素来确定。据统计,燃料成本在火电总成本中占据 70%~80%,因此煤场的科学管理在整个火电机组安全、经济运行过程中都起到至关重要的作用<sup>[5,26-27]</sup>。煤场管理如同是仓库管理,要有条理和依据,分区堆放是煤场管理的基础,每个分区预设煤质情况,与预设煤质相同的煤就存放在该煤区中。煤场分区数目不能太多,避免增加堆取难度;同时分区也不能太少,这样无法满足煤场精细化管理要求<sup>[28]</sup>。

数据集在完成聚类运算后,每一个样本都会给定一个聚类标签,拥有同类标签的样本归为一个类别<sup>[29]</sup>。由于  $K$ -means 聚类算法是一种无监督聚类算法,同一样本在每次聚类运算中的聚类标签 label 是在 0、1、2、3 中随机分配的,如“label=0”在二次聚类运算中可能代表了 2 类煤。因此要统计同一类煤的煤质范围,就需要固定它的标签,可以通过统计聚类标签出现频率,由高到底依次替换为 a、b、c、d,这样同一类煤在每次  $K$ -means 聚类运算中都有相同的标签。依照此方法,统计 2021 年 858 批来煤,不同类别的来煤频率统计如图 10 所示。

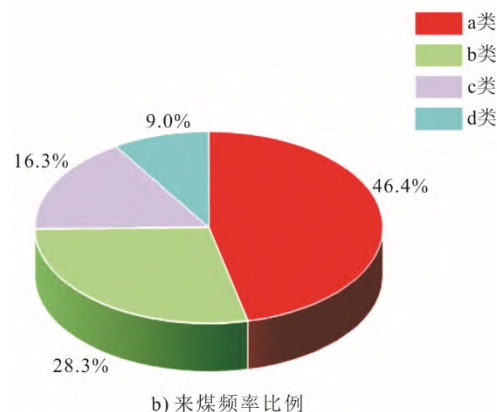
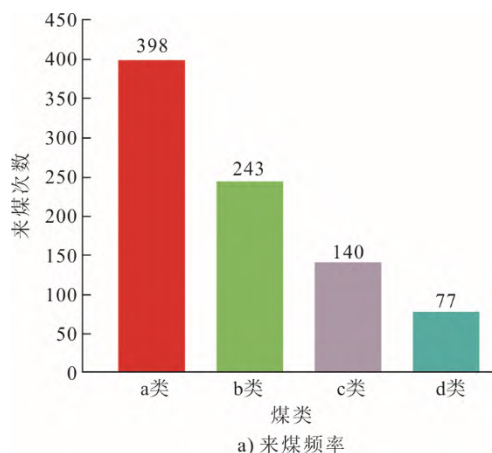


图 10 不同类别的来煤频率统计  
Fig.10 Statistics on the frequency of coal incoming in different categories



以煤批批号为索引,在“验收管理”表格中检索每批次来煤的验收量,然后统计以上不同分类的煤的历史验收量以及各自的来煤质量比例,如图 11 所示。由图 10、图 11 可以看出,4 类煤的来煤频率比例和来煤量比例相差不大,反映了每批煤的平均来煤量大致相同,且 4 类煤的来煤频率相差较大,

在一段时间内不会发生相对数量的反转。因此以 4 类煤的来煤量比例为依据,将煤场进行大致的比例划分。该项目基于煤场开发了一套盘煤系统软件,图 12 为某一时刻该煤场的真实扫描全貌图像,为了直观展示煤场分区,在煤场扫描图上按上述历史来煤量比例进行区域划分。

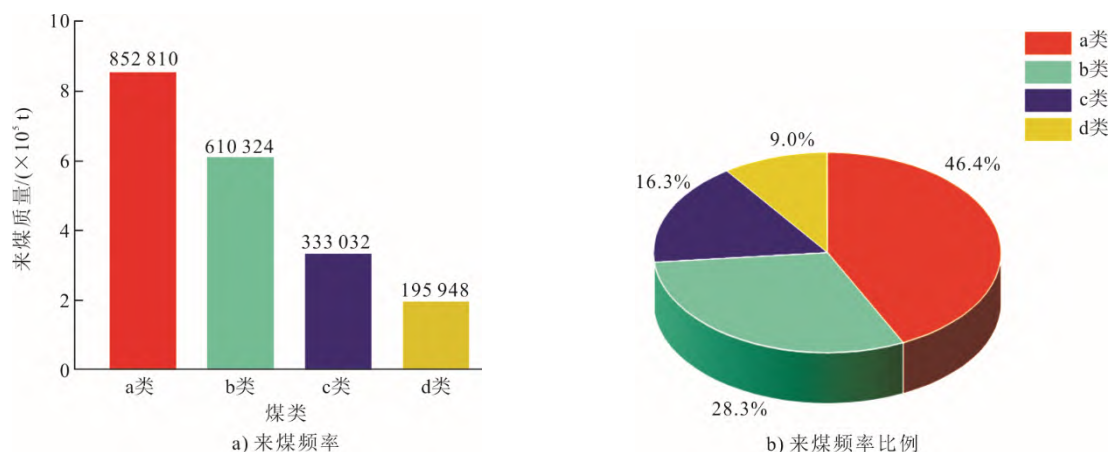


图 11 不同类别的来煤量统计  
Fig.11 Statistics on the frequency of coal incoming in different categories

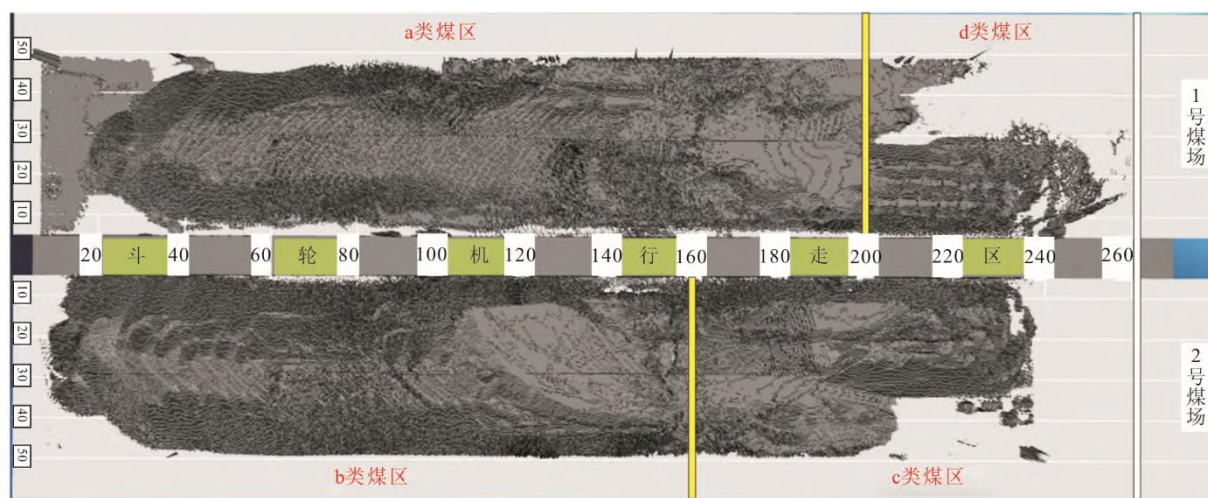


图 12 煤场分区示意  
Fig.12 Schematic diagram of coal yard zoning

## 4 存煤流程设计

在煤场分区之后,就要对来煤的分区堆放流程进行设计,在数字煤场平台后端接入基于 Python 编程语言的煤质信息的聚类算法。聚类数据集直接来源于燃料平台服务器,并且它可以利用煤批批号蕴含的时间信息,进行聚类数据的不断更新,使得每次聚类运算的煤质信息数据集都是 1 个自然年内。如 2022 年 5 月 1 日到厂一批煤,预估煤质信息上传至平台服务器后,Python 编程语言中的聚类程序

就调用 2021 年 5 月 1 日到此时的来煤信息作为样本集进行聚类运算,然后给出这一批煤的来煤分类。

来煤到场需要接卸时,真正的化验数据还未生成,这时就用矿发煤质表单或预采样结果作为预估煤质信息,结合来煤批次、矿发量、验收量、矿点名称等作为来煤信息,并将其上传到数字煤场平台服务器,平台将来煤进行分类,燃料管理人员看到来煤信息和分类建议后,如果同意此堆煤方案,则方案下发到煤场接卸人员,操作斗轮机将来煤堆放

到指定分区内,反之不同意此方案,管理人员可根据煤场实际存煤情况进行人工调整。

通过采集输煤程控系统的可编程逻辑控制器(PLC)及斗轮机堆料系统数据,数字煤场平台服务器记录下真实的来煤堆放信息<sup>[30]</sup>。来煤的真实化验结果出来后,煤场管理人员要及时将化验信息上传至服务器,平台以煤批批号为对应索引,实现对煤质信息的更新。图 13 为具体来煤堆放及存放流程。

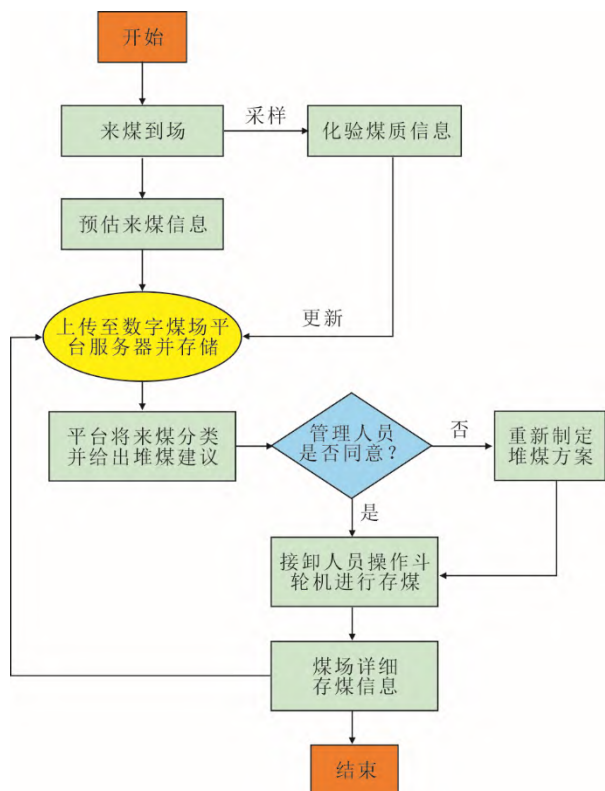


图 13 来煤堆放及存放流程

Fig.13 The process of coal stacking and information storage

## 5 结 论

1) 针对某电站煤场的来煤煤种复杂,难以按煤种将煤场分区堆放来煤的情况,本文提出利用  $K$ -means 聚类算法和 DBSCAN 聚类算法对历史来煤的关键煤质信息进行聚类分析,对比二者的聚类效果,最终选择  $K$ -means 聚类算法作为煤质信息分类的方法。

2) 利用  $K$ -means 聚类算法将历史来煤分为 4 类,使得每一类别内的煤质相差不大,统计不同类别的来煤频率和来煤质量比例,对煤场进行了相应的比例分区。

3) 所设计数字煤场平台通过读取过去 1 年的煤质数据并采集来煤信息,对即将到场的来煤进行

分类,结合输煤程控系统的 PLC 数据和斗轮机堆取料记录,设计了来煤自动化堆放及信息存储流程。

与传统的来煤堆放方式相比,利用聚类算法对来煤进行分类堆放,在一定程度上提升了燃料管理的科学性,提高了煤场空间的有效利用率,燃料数字化管理是火电厂未来发展的必然趋势。

## [参 考 文 献]

- [1] 赵兵, 景杰. “碳达峰、碳中和”目标下火力发电行业的转型与发展[J]. 节能与环保, 2021(5): 32-33.  
ZHAO Bing, JING Jie. Transformation and development of thermal power industry under the goal of “carbon peaking and carbon neutralization”[J]. Energy Conservation and Environmental Protection, 2021(5): 32-33.
- [2] 国家统计局. 中华人民共和国 2021 年国民经济和社会发展统计公报 [EB/OL]. (2022-02-28) [2022-05-18]. [http://www.stats.gov.cn/xgk/sjfb/zxfb2020/202202/t20220218\\_1827971.html](http://www.stats.gov.cn/xgk/sjfb/zxfb2020/202202/t20220218_1827971.html).  
National Bureau of Statistics. Statistical bulletin of the People's Republic of China on national economic and social development in 2021[EB/OL]. (2022-02-28) [2022-05-18]. [http://www.stats.gov.cn/xgk/sjfb/zxfb2020/202202/t20220218\\_1827971.html](http://www.stats.gov.cn/xgk/sjfb/zxfb2020/202202/t20220218_1827971.html).
- [3] 陈刚. 电站锅炉配煤掺烧及经济运行[M]. 北京: 中国电力出版社, 2013: 68-75.  
CHEN Gang. Coal blending and economic operation of power plant boilers[M]. Beijing: China Electric Power Press, 2013: 68-75.
- [4] 苏宏刚. 燃料数字化管理系统在火电厂燃料管理中的应用[J]. 河南科技, 2019(20): 125-127.  
SU Honggang. Application of digital fuel management system in fuel management of thermal power plant[J]. Henan Science and Technology, 2019(20): 125-127.
- [5] 阮光, 江学文, 周晓亮, 等. 燃煤电厂智能燃料系统整体解决方案探讨[J]. 中国设备工程, 2019(7): 140-143.  
RUAN Guang, JIANG Xuewen, ZHOU Xiaoliang, et al. Discussion on the overall solution of intelligent fuel system for coal-fired power plants[J]. China Plant Engineering, 2019(7): 140-143.
- [6] 秦丰姣. X 火电厂数字化煤场管理系统研究[D]. 长春: 吉林大学, 2017: 41.  
QIN Fengjiao. Research on digital coal yard management system of X thermal power plant[D]. Changchun: Jilin University, 2017: 41.
- [7] 张冬练, 成佳慧, 沈纳昊, 等. 火电厂智慧煤场管控系统技术研究[J]. 电力科技与环保, 2020, 36(5): 49-53.  
ZHANG Donglian, CHENG Jiahui, SHEN Nahao, et al. Research on intelligent management and control system of coal yard in thermal power plant[J]. Electric Power Technology and Environmental Protection, 2020, 36(5): 49-53.
- [8] 朱宇超, 柴成林, 卢子轩, 等. 煤堆分层技术在智能燃料系统中的应用[J]. 自动化应用, 2021(4): 134-136.  
ZHU Yuchao, CHAI Chenglin, LU Zixuan, et al. Application of coal stack stratification technology in intelligent fuel system[J]. Automation Application, 2021(4): 134-136.
- [9] 黄海东, 刘福国, 国钦光, 等. 配煤掺烧最小煤质偏差模型及煤场存放优化[J]. 热能动力工程, 2020, 35(5): 97-104.



- HUANG Haidong, LIU Fuguo, GUO Qinguang, et al. Modeling of minimum coal quality deviation for coal blending and optimization of coal yard storage[J]. Journal of Engineering for Thermal Energy and Power, 2020, 35(5): 97-104.
- [10] SUN L, QIN X Y, DING W P, et al. Density peaks clustering based on  $k$ -nearest neighbors and self-recommendation[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(7): 1913-1938.
- [11] 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 using optimizing initial cluster centers and contour coefficients[J]. Fuzzy Systems and Mathematics, 2022, 36(1): 47-65.
- [12] ABBAS M, EL-ZOGHABI A, SHOUKRY A. DenMune: density peak based clustering using mutual nearest neighbors[J]. Pattern Recognition, 2021, 109: 107589.
- [13] BAI L, CHENG X, LIANG J, et al. Fast density clustering strategies based on the  $k$ -means algorithm[J]. Pattern Recognition, 2017, 71: 375-386.
- [14] 赵庆. 基于 Hadoop 平台下的 Canopy-Kmeans 高效算法[J]. 电子科技, 2014, 27(2): 29-31.
- ZHAO Qing. Efficient algorithm of Canopy-Kmeans based on hadoop platform[J]. Electronic Science and Technology, 2014, 27(2): 29-31.
- [15] 史静, 南开辉, 周琪, 等. 大数据环境下基于 K-means 聚类算法的分组负荷预测研究[J]. 电力大数据, 2020, 23(6): 9-16.
- SHI Jing, NAN Kaihui, ZHOU Qi, et al. Research on classification load forecasting based on K-means clustering algorithm in big data environment[J]. Power Systems and Big Data, 2020, 23(6): 9-16.
- [16] 张文宇, 治瑜, 秦乐. 基于改进天牛群优化的 DBSCAN 聚类算法[J]. 统计与决策, 2022, 38(10): 20-25.
- ZHANG Wenyu, ZHI Yu, QIN Le. DBSCAN clustering algorithm based on improved beetle swarm optimization[J]. Statistics and Decision, 2022, 38(10): 20-25.
- [17] 程世超. K-Means 及 DBSCAN 聚类算法的研究与改进[D]. 赣州: 江西理工大学, 2021: 15.
- CHENG Shichao. Research and improvement of K-means and DBSCAN clustering algorithms[D]. Ganzhou: Jiangxi University of Science and Technology, 2021: 15.
- [18] 雷萌, 郭鹏, 刘博嵩. 基于自适应 DBSCAN 算法的风电机组异常数据识别研究[J]. 动力工程学报, 2021, 41(10): 859-865.
- LEI Meng, GUO Peng, LIU Bosong. Study on abnormal data recognition of wind turbines based on adaptive DBSCAN algorithm[J]. Journal of Chinese Society of Power Engineering, 2021, 41(10): 859-865.
- [19] 徐楠楠. 混煤掺配方案优化方法的研究[D]. 保定: 华北电力大学, 2015: 20.
- XU Nannan. Research on optimization method of blending scheme of coal blending[D]. Baoding: North China Electric Power University, 2015: 20.
- [20] 于佐军, 秦欢. 基于改进蜂群算法的 K-means 算法[J]. 控制与决策, 2018, 33(1): 181-185.
- YU Zuojun, QIN Huan. K-means algorithm based on improved artificial bee colony algorithm[J]. Control and Decision, 2018, 33(1): 181-185.
- [21] 王巧玲, 乔非, 蒋友好. 基于聚合距离参数的改进 K-means 算法[J]. 计算机应用, 2019, 39(9): 2586-2590.
- WANG Qiaoling, QIAO Fei, JIANG Youhao. Improved K-means algorithm with aggregation distance coefficient[J]. Journal of Computer Applications, 2019, 39(9): 2586-2590.
- [22] 钟志峰, 李明辉, 张艳. 机器学习中自适应  $k$  值的  $k$  均值算法改进[J]. 计算机工程与设计, 2021, 42(1): 136-141.
- ZHONG Zhifeng, LI Minghui, ZHANG Yan. Improved  $k$ -means clustering algorithm for adaptive  $k$  value in machine learning[J]. Computer Engineering and Design, 2021, 42(1): 136-141.
- [23] LIU Y, LI Z, HUI X, et al. Understanding of internal clustering validation measures[C]//ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010. IEEE, 2010.
- [24] 张朝, 郭秀娟, 张坤鹏. K-means 算法聚类中心选取[J]. 吉林大学学报(信息科学版), 2019, 37(4): 437-441.
- ZHANG Chao, GUO Xiujuan, ZHANG Kunpeng. Clustering center selection on K-means clustering algorithm[J]. Journal of Jilin University (Information Science Edition), 2019, 37(4): 437-441.
- [25] 代少升, 刘小兵, 赖智颖, 等. 网格化局部自适应 DBSCAN 聚类算法[J]. 重庆邮电大学学报(自然科学版), 2022, 34(2): 250-257.
- DAI Shaosheng, LIU Xiaobing, LAI Zhiying, et al. Gridded local adaptive DBSCAN clustering algorithm[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, 34(2): 250-257.
- [26] ZHANG Y, HAN X, ZHANG L, et al. Integrated generation consumption dispatch based on compensation mechanism considering demand response behavior[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(5): 1025-1041.
- [27] FU G, WU S. Study on the optimization of transregional transmission scale for coal and power electricity under the background of coal capacity reduction[C]. Proceedings of the 2018 First International Conference on Environment Prevention and Pollution Control Technology (EPPCT 2018), 2018: 1881-1886.
- [28] 韩云昕. 火力发电企业数字化煤场的设计与实现[D]. 武汉: 华中科技大学, 2019: 16.
- HAN Yunxin. Design and implementation of digital coal yard in thermal power generation enterprises[D]. Wuhan: Huazhong University of Science and Technology, 2019: 16.
- [29] 刘国华. 基于 Kmeans 算法的学生行为分析系统的设计与实现[D]. 石家庄: 河北科技大学, 2014: 17.
- LIU Guohua. Design and implementation of student behavior analysis system based on Kmeans algorithm[D]. Shijiazhuang: Hebei University of Science and Technology, 2014: 17.
- [30] 刘华炜, 黄伟, 许杰. 燃煤电厂燃料系统煤场数字化管理的设计与实现[J]. 自动化仪表, 2020, 41(9): 106-110.
- LIU Huawei, HUANG Wei, XU Jie. Design and implementation of digital management of fuel system coal yard in coal-fired power plant[J]. Process Automation Instrumentation, 2020, 41(9): 106-110.

(责任编辑 杜亚勤)