

基于改进K-Means聚类算法和最小生成树融合模型的配电网规划研究

石万珍

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年9月22日; 录用日期: 2023年11月29日; 发布日期: 2023年12月7日

摘要

本文针对配电网在规划分区时各分区用户负荷差值较大而造成可靠性不足和经济损失等问题, 分别提出了一种基于负荷因子的平衡负荷初步分区方法和基于欧式距离权重的改进K-means聚类算法, 并将其应用于配电网规划研究。以我国某城市街道为例, 试验结果表明改进的K-means聚类在负荷分区的均匀分布上优于传统的K-means聚类方法, 改进的K-means聚类负荷差值的平均值由0.364MW下降到0.1675 MW。输出聚类结果后的最短路径以实现配电网规划线路图。

关键词

配电网规划, 改进K-Means聚类, 配电网分区

Research on Distribution Network Planning Based on Improved K-Means Clustering Algorithm and Minimum Spanning Tree Fusion Model

Wanzhen Shi

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Sep. 22nd, 2023; accepted: Nov. 29th, 2023; published: Dec. 7th, 2023

Abstract

This article proposes a preliminary balanced load partitioning method based on load factors and

an improved K-means clustering algorithm based on Euclidean distance weights to address the issues of insufficient reliability and economic losses caused by large differences in user loads between different zones in distribution network planning. These methods are applied to research on distribution network planning. Taking a certain urban street in China as an example, the experimental results show that the improved K-means clustering is superior to the traditional K-means clustering method in terms of uniform distribution of load zones. The average value of the improved K-means clustering load difference decreases from 0.364 MW to 0.1675 MW. The shortest path after clustering results is output to achieve distribution network planning circuit diagram.

Keywords

Distribution Network Planning, Improved K-Means Clustering, Distribution Network Zoning

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

配电网规划是指根据一段时期内该区域用户点的负荷预测和已有电网的基本情况,在满足当期内一定负荷增长和安全可靠供电的前提下,确定最优且可行的电网设备建设方案,使得配电网的投资建设成本和运行费用最小[1] [2] [3]。早期用相对传统的数学规划方法来解决配电网规划问题,例如动态规划法[4]、线性规划法[5]和层次分析法[6]等。随着计算机科学技术的发展,一些依赖于强大计算能力的解决方法被提出。如遗传算法[7]、粒子群算法[8]、蚁群算法[9]等,针对各种组合优化问题,该算法都已经证实了自身的优点。

但随着城镇化进程的加快,用电负荷的急剧增加,配电网的规模也越来越大,同时各变电站和电源用户数量庞大且分布杂乱无章。直接考虑配电网的所有负荷点会导致决策变量过多,计算难度较大等问题。因此,有必要在实施配电网规划前进行合理的分区规划。徐芮等[10]提出了采用 K-means 聚类算法对负荷点进行聚类分区的配电网规划技术,并将聚类中心作为等效负荷点,进行各分区的主次级网架和各分区内的联络线规划。董秋仙等[11]选取最大相异度参数值所对应的样本作为初始聚类中心。但这些方法都没有考虑到负荷分布不均匀对分区的影响,出现了一些分区负荷较重而一些分区负荷较轻的情况,不利于系统的经济运行。

为了避免上述问题,同时减少计算困难,本文分别提出了一种基于负荷因子的负荷平衡的初步分区方法和基于欧式距离权重的改进 K-means 聚类算法应用于配电网规划研究,以此来解决在规划分区时各分区用户负荷差值较大而造成可靠性不足和经济损失的问题。并结合最小生成树算法,输出聚类结果后的最短路径以实现配电网规划线路图,增强城市街道配电网规划研究可视化效果。

2. K-Means 聚类算法

聚类是针对给定的样本,依据它们特征的相似度或距离,将其归并到若干个“类”或“簇”的数据分析问题,其目的是通过得到的类或簇来发现数据的特点或对数据进行处理,在数据挖掘、模式识别等领域有着广泛应用[12] [13]。随着数据分析领域的不断发展,更多较好的数据分析方法不断被提出,例如 K-means、K-MS、ET-SSE 等[14] [15] [16]。其中 K-means 聚类算法又称 K-均值聚类算法,k-均值算法源于信号处理中的一种向量量化方法,现在则更多地作为一种聚类分析方法流行于数据挖掘领域,由于其

具有运算速度快、可解释性强等优点，因此在数据挖掘、图像处理、自然语言处理领域都有着重要应用 [17] [18] [19]。

K-means 聚类算法是基于原型目标函数的硬聚类算法。**K-means** 聚类算法定义观测集 (x_1, x_2, \dots, x_n) ，其中每个观测值都是一个 d -维实向量， k -均值聚类要把这 n 观测值划分到 k 个集合中 ($k \leq n$)，使得组内平方和(WCSS within-cluster sum of squares)最小。换句话说，它的目标是找到使得下式满足的聚类 S_i ：其中 μ_i 是 S_i 中所有点的均值。

$$\arg \min_s \sum_{i=1}^k \sum_{X \in S_i} \|X - \mu_i\|^2 \quad (1)$$

算法按照下面两个步骤交替进行：

1) 分配(Assignment)：将每个观测分配到聚类中，使得组内平方和(WCSS)达到最小。因为这一平方和就是平方后的欧式距离，所以很直观地把观测分配到离它最近的均值点即可：

$$S_i^{(t)} = \left\{ x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k \right\} \quad (2)$$

其中，每个 x_p 都只被分配到一个聚类 s' 中，尽管在理论上它可能被分配到 2 个或者更多的聚类。

2) 更新(Update)：对于上一步得到的每一个聚类，以聚类中观测值的质心(centroids)作为新的均值点：

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (3)$$

因为算术平均是最小二乘估计，所以这一步同样减少了目标函数组内平方和(WCSS)的值。这一算法将在对于观测的分配不再变化时收敛。由于交替进行的两个步骤都会减少目标函数 WCSS 的值，并且分配方案只有有限种，所以算法一定会收敛于某一最优解。

K-均值聚类算法的目标函数是组内平方和(WCSS)，而且按照“最小二乘和”来分配观测，确实是等价于按照最小欧氏距离来分配观测的。但是在配电网规划领域所要解决的不仅仅是简单的“距离问题”，还要将各个用户的负荷考虑进去以避免电源点的供电容量不足以支撑负荷点的需求总量等问题。因此，需要对现有的 **K-均值聚类**方法进行改进，才能更好地应用在配电网规划领域。

3. 最小生成树算法

最小生成树(Minimal Spanning Tree, MST)是基于图论的相关知识而发展起来的一种求解各点之间最小权重和、且两点之间只有一条通路的方法。具体来看，在一个给定的无向图 $G = (V, E)$ 中， V 表示顶点的集合， E 表示边的集合， (u, v) 代表连接顶点 u 与顶点 v 的边，而 $w(u, v)$ 表示此边的权重，若存在 T 为 E 的子集且为无循环图，使得联通所有结点的权重之和 $w(T)$ 最小，则此 T 即为无向图 G 的最小生成树。

常用的求解最小生成树问题的方法有 **Kruskal** 算法、**Prim** 算法，其基本步骤就是每一步都从未选择的边中选取一条边 e ，使之与已选边不构成“圈”，且 e 是未选边中权重最小的，直到选够 $n - 1$ 条边为止。具体操作如下：

- 1) 将所求规划问题模拟成二维无向图；
- 2) 把图上的每一条边存在一个数组中，数组中的每个元素均含有起点、终点、权重三个单元；
- 3) 将边数组按权重从小到大排序；
- 4) 依次按边的权重从小到大枚举每一条边，如果边的两个端点已经连通了，则跳过该边(通过并查集判断)；
- 5) 否则将该边加入“生成树”中，并将该边的两个端点也加入并查集中；

6) 返回第 4)步, 直到选够 $n - 1$ 条边为止。

可将本文的配电网规划简化成带权重的最短路径问题, 本文最小生成树主要应用于聚类之后寻求各个用户与电源点之间的最短路径问题, 得到配电网规划方案。

4. 基于改进 K-Means 聚类算法和最小生成树融合方法

4.1. 双供配电网初步分区

本文研究的是双供配电网规划, 当已知两个电源和一批用户的平面坐标、每个用户的用电需求。需要设计两个单供配电网, 本题初步考虑传统的 K-Means 聚类算法, 若通过欧式距离最小原则进行负荷点聚类分区, 会出现两个分区负载不均衡进而导致系统运行不经济的问题, 同时也会使得两个分区间的联络线的备用容量增大, 投资成本增加。所以针对此问题, 本文提出了一种基于负荷因子的负荷平衡的初步分区方法。

针对已知的用户坐标信息, 分别计算所有用户到两个电源点的欧氏距离, 根据距离最小原则初步把所有用户分给两个电源, 电源 1 所在分区记为分区一, 电源 2 所在分区记为分区二。在此基础上, 我们引入一种负荷因子, 取负荷因子为 0.01, $load_all$ 为用户总负荷值, 当 $|load_1 - load_2| > 0.01 * load_all$ 时, 进行迭代操作, 标签为 1 的改为 2, 直到此循环终止。此时可达到了分区负荷均衡的效果。

双供配电网初步分区方法的伪代码如下所示:

-
- 1) 输入: 用户总负荷 $load_all$, 分区一负荷 $load_1$, 分区二负荷 $load_2$, 负荷因子 λ
 - 2) repeat
 - 3) 计算分区一、分区二用户点差值绝对值;
 - 4) 当 $|load_1 - load_2| > 0.01 * load_all$ 时;
 - 5) $load_1 \leftarrow load_2$;
 - 6) until 分区不在发生变化
- 输出: 分区一、分区二各用户及负荷值
-

分区结束后可进行单个电源点的配电网规划, 传统的 K-means 聚类是通过各点到聚类中心点距离最小的原则进行负荷点的聚类分区, 当各个负荷点的负荷大小差异较大时, 传统的 K-means 聚类方法可能失效, 因为此时电源点的供电容量不足以支撑负荷点的需求总量, 因此会产生扩展供量提高投资成本。较好的解决办法是在对各个负荷点进行聚类分区时将各个负荷点的负荷大小考虑进对聚类产生的影响中, 可有效地解决负荷分布不均匀引起的负荷点聚类分区之间的负荷相差较大的问题。本文对 K-means 聚类进行改进, 结合现有数据特点将各个负荷点的负荷大小转变成相应的权重加入到聚类中心点欧式距离的判别中。第 j 个负荷点的负荷权重 α_j 计算公式如下:

$$\alpha_j = \frac{p_j}{\sum p_j} \quad (4)$$

式中, j 为负荷点的编号, α_j 是第 j 个负荷点的权重, p_j 为第 j 个负荷点的负荷大小。

在负荷点分布不均匀时, 引入权重变量可增大负荷密度较小区域的负荷点聚类分区的负荷点个数, 减小负荷密度较大区域的负荷点聚类分区的负荷点个数, 从而在一定程度上实现各负荷点聚类分区之间的负荷相对均匀, 可进一步节省联络线投资以提高网络架构规划的整体经济性。

4.2. 改进的 K-Means 聚类方法基本步骤

1) 确定负荷点聚类分区个数 k

本文采用 WGSS 的碎石图来寻找最优的 K ，目标是使得 WGSS 达到足够小。目标函数如下：

$$WGSS = \sum_{j=1}^p \sum_{l=1}^k \sum_{i \in G_l} (y_i - \bar{y}_j^{(l)})^2 = \sum_{l=1}^k \sum_{i \in G_l} (y_i - \bar{y}^{(l)})' (y_i - \bar{y}^{(l)}) \quad (5)$$

其中，WGSS (族群内方差)可理解为每个点到其簇(最终聚类)的质心的距离之和， G_j 为第 j 个族群， $\bar{y}^{(l)} = \sum_{i \in G_l} y_{ij} / n_i$ 是族群 G_j 中变量 j 的均值， $y_i = (y_{i1}, \dots, y_{ip})'$ ， $\bar{y}^{(l)} = (\bar{y}_1^{(l)}, \dots, \bar{y}_p^{(l)})'$ 。

根据上面的式子我们知道，当 $k = n$ (即每个族群只有一个点)的时候，WGSS = 0，即组内方差平方和最小。但是这个时候就是一个过拟合的现象。也就是说，我们的目标是选择一个使得 WGSS 足够小(但不是最小)的 k 值，通常选取拐点(knee point)为最优的 k 值。

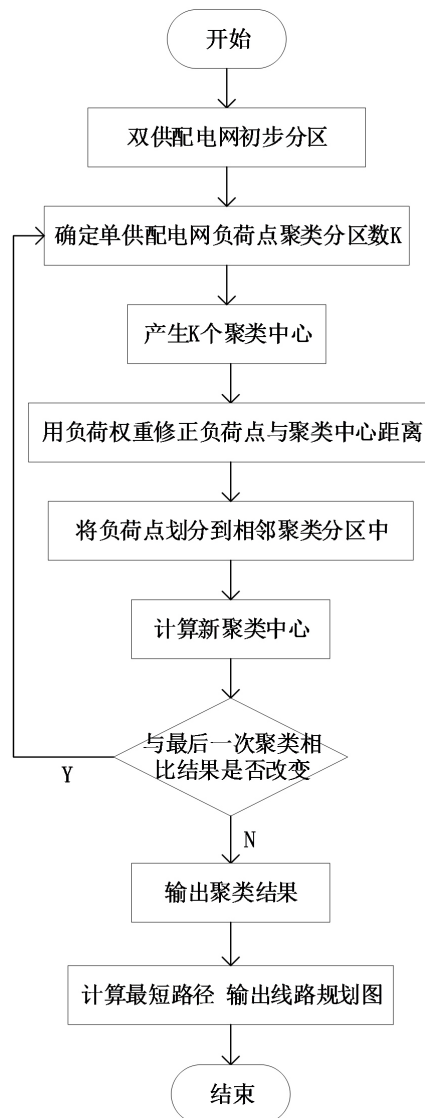


Figure 1. Flow chart of partition planning method

图 1. 分区规划方法流程图

2) 确定聚类中心

从包含 n 个样本的集合中随机挑选出 k 个样本, 作为 k 个初始的聚类中心。

3) 计算用户之间的修正加权欧氏距离

计算所有样本到各个初始聚类中心的修正的加权欧式距离, 并按照距离最小的原则划分给不同的类中心。其中修正的加权欧式距离定义如下:

$$\min d_{ij} = \alpha_j \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (6)$$

式中, d_{ij} 为第 i 个负荷点聚类分区的聚类中心和第 j 个负荷点之间的修正距离; 负荷点 j 的权重 α_j 由式(公式 4)计算; (x_i, y_i) 和 (x_j, y_j) 分别为第 i 个负荷点聚类分区的聚类中心和第 j 个负荷点的坐标。

4) 聚类完成

所有样本划分结束之后, 计算每个类划分的新的聚类中心, 然后重复步骤 3) 和步骤 4), 直到相邻两次的聚类中心不再发生变化或满足设定的终止条件。

于配电网的规划而言, 由于分布式电源的接入, 得以更好地应对用户负荷增长的需求。而根据负荷点之间的距离大小进行聚类并设置电源或变压器, 会更加节省线路的建造成本和后期的维护费用。因此, 本文考虑可以用改进的 K-means 算法先对用户点进行聚类, 并找到类中心。

综上, 基于改进 K-means 聚类算法的分区规划方法的流程如图 1 所示。

5. 实例分析

下表是我国某城市的实际街道, 包含 2 个电源(编号 1 和编号 28)、36 个用户点。各负荷点的大小及坐标见表 1。

Table 1. User information of dual power supply and single supply distribution network

表 1. 双电源单供配电网用户信息

用户编号	坐标 X (km)	坐标 Y (km)	负荷值 (kVA)	用户编号	坐标 X (km)	坐标 Y (km)	负荷值 (kVA)
1 (电源 1)	1.976	1.091	—	20	3.112	1.394	292
2	1.056	1.026	188	21	2.348	0.112	192
3	0.48	1.304	180	22	2.128	0.334	208
4	1.928	1.798	136	23	3.3	0.474	136
5	0.196	1.076	184	24	3.44	1.49	168
6	3.64	0.474	160	25	2.304	1.556	118
7	0.524	0.914	172	26	1.172	0.354	68
8	2.876	1.808	164	27	2.388	0.506	134
9	0.184	1.602	244	28 (电源 2)	2.944	1.196	—
10	1.008	1.586	252	29	3.616	0.718	68
11	0.664	1.822	180	30	1.605	1.611	227
12	3.36	0.904	204	31	1.334	0.321	299

Continued

13	0.548	0.43	248	32	1.31	0.992	153
14	0.916	0.182	160	33	1.877	0.505	248
15	3.424	1.192	196	34	1.704	0.308	182
16	2.856	0.182	144	35	1.643	1.299	249
17	2.488	0.272	172	36	1.564	0.91	216
18	3.272	1.738	188	37	1.463	1.516	228
19	2.876	1.56	172	38	1.846	1.631	283

首先利用本文提出了一种基于负荷因子的负荷平衡 K-均值聚类初步分区方法。将两个电源点分别分为两个分区，并使之达到分区负荷均衡的效果。得到各个分区的负荷值总和，如表 2 所示：

Table 2. Preliminary zoning results
表 2. 初步分区结果

负荷分区	最小欧式距离结果 (单位: MW)	基于负荷因子结果 (单位: MW)
分区一	4.1829	2.9155
分区二	1.6082	2.8755

对比分析可知，采用基于负荷因子的负荷平衡 K-均值聚类初步分区方法进行初步分区时会使得不同分区之间负荷分布更加均匀，使得配电网运行更加可靠。此时，1、3、5、7、9、10、11、13、14、26、31、34、2、30、32、35、36、37 号用户被划分到分区一；28、4、25、38、17、21、22、27、33、8、15、18、19、20、24、6、12、16、23、29 号用户被划分到分区二。

以分区一为例分别应用传统的 K-means 聚类和改进的 K-means 聚类算法进行聚类分析，输出总负荷和负荷差值，具体数值如表 3 和表 4 所示。采用改进的 K-means 聚类算法时各聚类分区之间的负荷差值最大为 0.286 MW，最小为 0.049；而采用传统的 K-means 聚类算法来看，各聚类分区之间的负荷差值最大为 0.639，最小为 0.089。与传统 K-means 聚类算法相比，改进的 K-means 聚类负荷差值的平均值由 0.364 MW 下降到 0.1675 MW。

Table 3. Comparison of total load results in cluster partitions
表 3. 聚类分区总负荷结果对比

聚类编号	K-means 聚类法总负荷 (单位: MW)	改进 K-means 聚类法总负荷 (单位: MW)
1	1.762	0.975
2	1.212	1.212
3	1.123	1.261

Table 4. Comparison of total load differences in cluster partitions
表 4. 聚类分区总负荷差值对比

负荷差值	K-means 聚类法总负荷 (单位: MW)	改进 K-means 聚类法总负荷 (单位: MW)
最大值	0.639	0.286
最小值	0.089	0.049
平均值	0.364	0.1675

由对比分析可知,改进的 K-means 聚类算法应用到配电网规划时,可有效地解决负荷分布不均匀引起的负荷点聚类分区之间的负荷相差较大的问题,进一步提高用户用电的稳定性和规划方案的经济性。

下面将改进的 K-means 聚类算法的分区一和分区二的聚类结果进行进一步的规划研究,利用最小生成树算法找到各用户之间的最短路径以获取分区规划效果图并输出各个分叉点的坐标。如图 2 所示,1 号和 28 号为电源点, m1, m2, m3, n1, n2, n3, n4 为分区一、二的聚类中心点。红色线段为配电网的主线,黑色线段为各分区的支线,各聚类中心的坐标如表 5 所示。

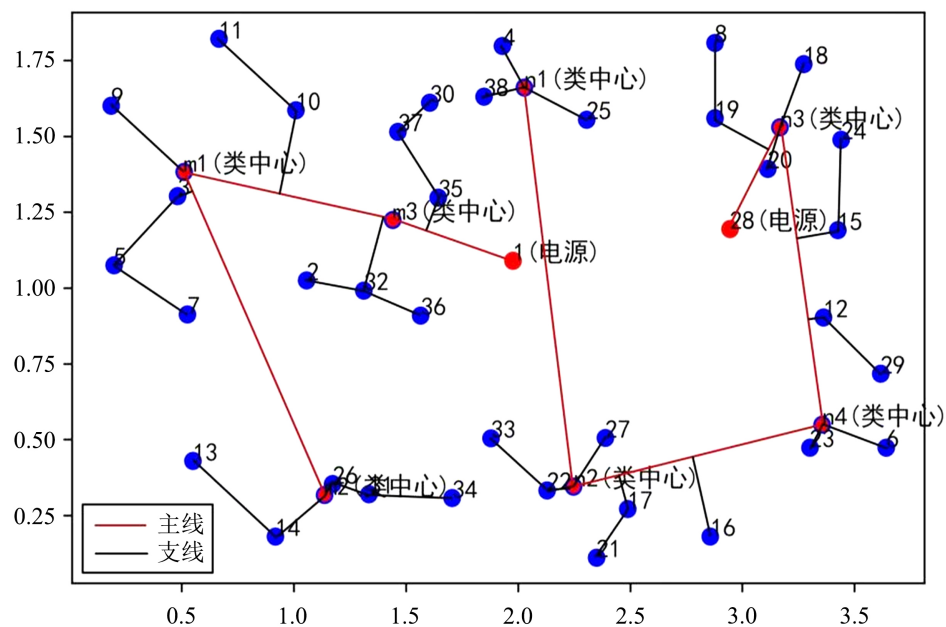


Figure 2. Partition planning effect
图 2. 分区规划效果图

Table 5. Bifurcation point coordinates for dual power supply and single supply distribution network planning
表 5. 双电源单供配电网规划的分叉点坐标

分区一	分叉点	m1	m2	m3	—
	坐标	(0.51, 1.38)	(1.13, 0.32)	(1.44, 1.23)	—
分区二	分叉点	m1	m2	m3	m4
	坐标	(2.03, 1.66)	(2.25, 0.35)	(3.17, 1.53)	(3.35, 0.55)

6. 结论

本文分别提出了一种基于负荷因子的负荷平衡的初步分区方法和基于欧式距离权重的改进 K-means 聚类算法应用于配电网规划研究, 以此来解决在规划分区时各分区用户负荷差值较大而造成可靠性不足和经济损失的问题, 试验结果表明改进的 K-means 聚类在负荷分区的均匀分布上优于传统的 K-means 聚类方法。将其应用于实际街道中, 结合最小生成树算法, 输出聚类结果后的最短路径以实现配电网规划线路图, 增强城市街道配电网规划研究的可视化作用, 使能够更清晰地理解配电线路与用户之间的拓扑连接关系、更加直观地理解配电网规划研究的过程。

参考文献

- [1] 陈根军, 唐国庆. 基于禁忌搜索与蚁群最优结合算法的配电网规划[J]. 电网技术, 2005, 29(2): 23-27.
- [2] 盛四清, 王浩. 用于配电网规划的改进遗传算法[J]. 电网技术, 2008, 32(17): 69-72+83.
- [3] 李亮, 唐巍, 白牧可. 考虑时序特性的多目标分布式电源选址定容规划[J]. 电力系统自动化, 2013, 33(3): 58-63+128.
- [4] 邓群, 孙才新, 周淑, 等. 采用动态规划技术实现配电网恢复供电[J]. 重庆大学学报(自然科学版), 2006(3): 40-44.
- [5] 苑津莎, 张铁峰, 刘建新, 等. 基于级别高于关系和线性规划的配电网规划辅助决策方法[J]. 中国电机工程学报, 2006(12): 106-110.
- [6] 张弛, 程浩忠, 奚珣, 夏夷, 沈晓岚, 奚增辉. 基于层次分析和模糊综合评价法的配电网供电模式选型[J]. 电网技术, 2006(22): 67-71.
- [7] 樊一娜. 一种改进型遗传算法在电网规划中的应用[J]. 青海大学学报(自然科学版), 2015, 33(2): 29-35+46.
- [8] 鲁晓秋, 叶影, 曹春, 等. 分布式光伏接入的配电网规划综合评价方法[J]. 华北电力大学学报(自然科学版), 2022, 27(11): 11-19.
- [9] 王志刚, 杨丽徒, 陈根永. 基于蚁群算法的配电网网架优化规划方法[J]. 电力系统及其自动化学报, 2002, 14(6): 73-76.
- [10] 徐芮, 刘俊勇, 刘友波, 等. 考虑负荷聚类分区与分布式发电接入的配电网主次网架规划方法[J]. 电力自动化设备, 2016, 36(6): 48-55.
- [11] 董秋仙, 朱赞生. 一种新的选取初始聚类中心的 K-means 算法[J]. 统计与决策, 2020, 36(16): 32-35.
- [12] 谭敏, 张宏源, 张海超. 基于弱监督深度学习的文本聚类算法及应用[J]. 计算机应用与软件, 2019, 36(4): 171-177.
- [13] 费贤举, 李虹, 田国忠. 基于特征加权理论的数据聚类算法[J]. 沈阳工业大学学报, 2018, 40(1): 77-81.
- [14] 毛秀, 冒纯丽, 丁岳伟. 基于密度和聚类指数改进的 K-means 算法[J]. 电子科技, 2015, 28(11): 47-50+64.
- [15] 王志华, 刘绍廷, 罗齐. 基于改进 K-modes 聚类的 KNN 分类算法[J]. 计算机工程与设计, 2019, 40(8): 2228-2234.
- [16] 汤深伟, 贾瑞玉. 基于改进粒子群算法的 k 均值聚类算法[J]. 计算机工程与应用, 2019, 55(18): 140-145.
- [17] 吴文静, 景鹏, 贾洪飞, 张铭航. 基于 K 均值聚类与随机森林算法的居民低碳出行意向数据挖掘[J]. 华南理工大学学报(自然科学版), 2019, 47(7): 105-111.
- [18] 曹帅帅, 陈雪鑫, 苗圃, 卜庆凯. 基于 PSO 与 K-均值聚类算法优化结合的图像分割方法[J]. 计算机与现代化, 2020(1): 22-27.
- [19] 于莉佳, 汪涛. 基于模糊 K 均值聚类的高校网络用户行为分析[J]. 智能计算机与应用, 2022, 12(10): 200-202.