

改进的密度峰值聚类算法在岩体结构面 优势分组中的应用

王述红, 高晨翔, 侯钦宽
(东北大学 资源与土木工程学院, 辽宁 沈阳 110819)

摘要: 岩体稳定性评价依赖于合理的结构面分组,但传统方法存在易受边缘点与异常点影响的弊端。为此,提出一种改进的密度峰值聚类算法用于结构面优势分组。首先,将结构面产状转换为空间坐标,并以单位法向量夹角正弦值的平方作为相似性度量。随后,基于有效性评价指标构建目标函数,并利用乌鸦算法优化截断距离以获取最佳分组结果。通过模拟数据集验证了该算法能够有效减少人为干预,避免异常点干扰,确保聚类结果更加可靠和合理。结果表明,所提方法不仅与传统方法一致性良好,还具有更高的适用性,为工程中结构面优势分组提供了可靠的参考。

关键词: 密度峰值聚类;乌鸦算法;有效性评价指标;结构面;优势分组

中图分类号: TU 45 文献标志码: A 文章编号: 1005-3026(2025)03-0130-08

Application of Improved Density Peak Clustering Algorithm in Dominant Grouping of Rock Discontinuities

WANG Shu-hong, GAO Chen-xiang, HOU Qin-kuan

(School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: GAO Chen-xiang, E-mail: 1062802882@qq.com)

Abstract: The stability evaluation of rock mass relies on reasonable rock discontinuities grouping. However, traditional methods are susceptible to boundary and outlier points. To address this issue, an improved density peak clustering algorithm was proposed for rock discontinuities grouping. Firstly, the rock discontinuities orientations were converted into spatial coordinates, and the squared sine of the angle between unit normal vectors was used as a similarity metric. Then, an objective function was constructed based on validity evaluation indices, and the cutoff distance was optimized using the crow algorithm to obtain the optimal grouping results. Validation with simulated datasets demonstrates that the proposed algorithm effectively reduces human intervention, avoids interference from outliers, and ensures more reliable and reasonable clustering outcomes. The results show that the proposed method not only maintains good consistency with traditional methods but also exhibits higher applicability, providing a reliable reference for dominant joint grouping in engineering applications.

Key words: density peak clustering; crow algorithm; validity evaluation indices; rock discontinuities; dominant grouping

自然界中的岩体存在非连续、非均匀的特性,随着地质活动与人为干扰的不断影响,这些岩体中逐渐形成了大量的具有不同的规模以及形态的结构面。边坡岩体变形破坏模式在很大程

度上取决于这些大量随机分布的构造结构面的性质。因此结构面是控制工程岩体稳定性的关键因素^[1]。结构面在岩体中的分布通常成组存在,识别随机结构面的优势组别对实际工程具有重要

收稿日期: 2023-09-04

基金项目: 中国-中东欧国家高校联合教育项目(2021111); 国家自然科学基金资助项目(U1602232, 52374157); 辽宁省重点科技计划项目(2019JH2-10100035); 中央高校基本科研业务费(N2301005, N2301006)。

作者简介: 王述红(1969-),男,江苏泰州人,东北大学教授,博士生导师。

意义,这些随机结构面的优势组别是描述结构面分布规律的基础,也是进行结构面网络模拟和岩体稳定性评价的前提条件.因此,在选择岩体强度参数、研究力学特性和评估工程稳定性方面,分析随机结构面的特征具有重要意义^[2-3].

工程中最常使用的图形分析方法是分析者采用玫瑰花图、极点图等进行目测分析,这类方法主要由分析者依靠经验进行分组,但在结构面分布离散的情况下,难以获得符合工程需要的结果^[4-5].Shanley等^[6]首次提出了结构面产状的聚类方法,之后结构面聚类研究取得快速发展.聚类分析是统计学上研究分类问题的一种方法^[7],当前的聚类算法大体可以分为6类,分别是基于划分的聚类算法,例如K均值聚类^[8]算法;基于层次的聚类算法,例如平衡迭代削减聚类^[9]算法;基于网格的聚类算法,例如集群查询^[10]算法;基于模糊数学的聚类算法,例如模糊C均值聚类^[11]算法;基于模型的聚类算法,例如高斯混合模型^[12];基于密度的聚类算法,最典型的是密度峰值聚类(DPC)^[13]算法.K均值聚类算法是结构面优势分组中应用最广泛的算法,其本质是局部寻优,对于边界不清晰的结构面数据容易陷入局部最优解.许多学者在此基础上引用人工智能对K均值聚类算法进行改进.Li等^[14]和王述红等^[15]分别采用蚁群算法和鱼群算法对K均值聚类算法进行改进,以解决其对初始值敏感的问题.Hammah等^[11]引入隶属度函数,提出模糊C均值聚类算法(FCM),并将其应用到结构面优势产状分组中,该算法属于模糊聚类算法的一种.此外,FCM与K均值聚类算法都存在人工确定初始质心数目和位置的问题.为了解决这一问题,许多学者另辟蹊径,将其他聚类算法应用到结构面分组中.蔡美峰等^[16]将遗传算法引入结构面的聚类分析当中,但该算法易陷入早熟收敛;Li等^[17]将粒子算法引入结构面优势分组中,用于搜索全局最优解,这种算法原理较为复杂,输入参数较多,不利于实际工程应用;Jimenez等^[18]引入谱聚类算法,通过对图的特征值分解和特征向量聚合进行分类;侯钦宽等^[19]使用编网算法对结构面产状进行划分,通过生成领域解和选择最优解进行聚类,这两类分组方法均无需预先确定聚类中心,分组结果良好,且较传统的分组方法有较快的收敛速度,但该算法仍需人为确定最优分组数,容易受到边缘点以及异常点的影响;张奇等^[20]采用凝聚层次聚类法划分结构面产状,有效解决了分组结

果受孤立数据影响的问题,但对初始聚类中心较为敏感.刘铁新等^[21]引入基于模型的聚类方法,考虑结构面产状与间距2个因素,通过调整隶属度进行结构面分组,该方法分组结果良好但是迭代过程较长,不利于应用.为解决上述问题,本文引入DPC算法,该算法具有非参数性以及自适应性的优势,在不需预先指定分组数以及分组中心的前提下根据数据的密度分布确定聚类的数量以及形状;DPC算法具有较强的鲁棒性,在异常值以及噪声数据下,仍可以给出较好的聚类效果.因此将DPC算法应用于结构面分组可以有效解决以上问题.

本文提出一种改进的DPC算法用于结构面优势分组.该方法在结构面产状完成空间坐标转换的基础上,以单位法向量夹角正弦值的平方作为相似性度量标准.为了避免分组结果受边缘点和异常点的影响,以有效性评价指标 V_d ^[22]建立目标函数.利用乌鸦搜索算法(CSA)^[23]对目标函数进行优化,以寻找DPC算法中的最佳截断距离.最后,利用计算机模拟数据对算法进行了验证.该研究可为工程结构面的优势分组提供借鉴.

1 基于CSA-DPC算法的结构面优势分组

1.1 结构面表征

结构面有2个最基本的属性,即倾向 α 和倾角 β ,将结构面假设为无厚度的二维平面,每个结构面有唯一的法向量,在实际工程中以法向量对结构面进行表示.对于任意一个结构面,其法向量可以表示为 $N=[n_1, n_2, n_3]$,单位法向量与倾向 α 以及倾角 β 的转化公式为

$$\left. \begin{aligned} n_1 &= \cos \alpha \sin \beta, \\ n_2 &= \sin \alpha \sin \beta, \\ n_3 &= \cos \beta, \\ n_1^2 + n_2^2 + n_3^2 &= 1. \end{aligned} \right\} \quad (1)$$

结构面优势分组的目的是将具有相似空间方向的结构面进行分组,以便进一步分析与应用.因此,必须选取合适的度量方式来判断两结构面之间的相似性,为了避免将倾向相差 180° 的结构面误分成一组,本文选取结构面单位法向量夹角正弦值平方作为两结构面相似性的度量^[24],假设结构面1的单位法向量为 p_1 ,结构面2的单位法向量为 p_2 ,则两结构面之间的距离 d 可以表示为

$$d(\mathbf{p}_1, \mathbf{p}_2) = (\sin \theta)^2 = 1 - (\mathbf{p}_1 \cdot \mathbf{p}_2^T)^2, \left\{ \begin{array}{l} \theta = \arccos |\mathbf{p}_1 \cdot \mathbf{p}_2^T|. \end{array} \right. \quad (2)$$

式中, θ 为两单位法向量之间的锐角夹角.

1.2 算法基本原理

密度峰值聚类算法由 Rodriguez 等^[13]首次提出. 该算法可以识别各种形状类簇, 并且参数很容易确定. 该算法对于数据集有 2 个基本假设: 一是数据在空间分布并不均匀, 数据中局部高密度点被一些局部低密度点包围; 二是数据集中局部高密度点之间的相对距离较大. 在算法实现过程中仅需要人工确定并输入一个值, 即截断距离 d_c . 密度峰值聚类算法主要基于局部密度和相对距离进行聚类, 下面给出参数定义以及该算法的具体流程.

1) 局部密度. 假设存在样本数据集, 其中 N 为样本个数, M 为样本维数, 则局部密度有 2 种计算方式: 离散值采用截断核的计算方式; 连续值则采用高斯核的计算方式.

截断核的计算公式为

$$\rho_i = \sum_{i \neq j} X(d_{ij} - d_c). \quad (4)$$

式中: ρ_i 为第 i 个数据点的局部密度; d_{ij} 为第 i 个数据点与第 j 个数据点之间的欧氏距离; $X(\cdot)$ 为阶跃函数, 其定义为

$$\left\{ \begin{array}{l} X(x) = 1, x < 0; \\ X(x) = 0, x \geq 0. \end{array} \right. \quad (5)$$

在图形算法的具体计算中, $X(d_{ij} - d_c)$ 在 $d_{ij} < d_c$ (即 $d_{ij} - d_c < 0$) 时取值为 1, 否则取值为 0, 该算法可以用来判断第 j 个样本与第 i 个样本之间的距离是否小于截断距离 d_c . 这样就可以对在给定截断距离 d_c 内的样本数量进行计算, 从而得到相应的密度估计.

高斯核的计算公式为

$$\rho_i = \sum_{i \neq j} \exp[-(\frac{d_{ij}}{d_c})^2]. \quad (6)$$

截断距离 d_c 通过式 (7) 计算得到. 该式先对所有样本间的欧氏距离 d_{ij} 进行升序排列, 然后选取位于前 2% 位置的距离^[13]作为截断距离 d_c .

$$d_c = d_{[0.02 \times \frac{N(N-1)}{2}]} \quad (7)$$

2) 相对距离. 相对距离 δ_i 为第 i 个数据点与其他局部密度更大数据点之间的最小距离.

$$\delta_i = \min_{\rho_j > \rho_i} (d_{ij}). \quad (8)$$

对于密度最高点, 相对距离被定义为

$$\delta_i = \max_{i \neq j} (d_{ij}). \quad (9)$$

在 DPC 算法中, 根据式 (4) 和式 (6) 可以计算得到每个点的局部密度, 再通过式 (8) 可以计算得到每个点的相对距离. 根据局部密度与相对距离绘制二维决策图, 通过对二维决策图观察, 选取同时拥有较大 δ_i 和较大 ρ_i 的点作为聚类中心, 最后将其余的点分配到密度比它高的最近类簇中. 以下为 DPC 算法的具体流程.

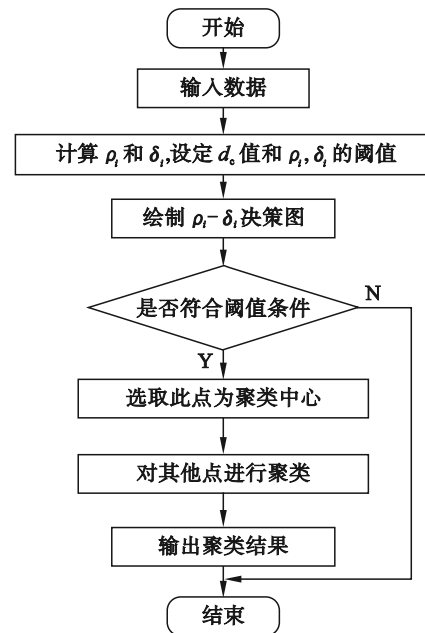


图 1 DPC 算法流程

Fig. 1 DPC algorithm flow

CSA 是由 Askarzadeh^[23]提出的一种新型元启发式优化器. CSA 受乌鸦行为的启发, 其基本思路为乌鸦将多余的食物储存在藏身之处, 并在需要时将其取回. 假设在 g 维空间中包含 h 个随机分布的乌鸦. 每个乌鸦表示 1 个问题的可行解决方案, g 是决策变量的数量.

$$C_r = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{g,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{g,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,h} & x_{2,h} & \cdots & x_{g,h} \end{bmatrix}. \quad (10)$$

式中, C_r 为乌鸦算法中某只乌鸦的位置矩阵.

每只乌鸦在每 1 轮位置决策时会选择 1 只乌鸦进行跟踪, 此时有 2 种情况:

情况 1 乌鸦 C_j 未发现乌鸦 C_i , 则乌鸦 C_j 继续前往食物隐藏地, 由式 (11) 可以得到乌鸦 C_i 的新位置.

$$x_{i,t+1} = x_{i,t} + P_{r,t} \cdot l_{i,t} \cdot (m_{i,t} - x_{i,t}). \quad (11)$$

式中: $x_{i,t}$ 为乌鸦 C_i 第 t 轮的位置; $m_{i,t}$ 表示乌鸦 C_i 第 t 轮的最优位置; $P_{r,t}$ 表示乌鸦 C_i 被发现的随机概率, 大小分布在 (0, 1) 之间; $l_{i,t}$ 表示乌鸦 C_i 第 t 轮的跟

随步长.

情况2 乌鸦 C_j 发现了乌鸦 C_i ,则乌鸦 C_j 决定进行误导,乌鸦 C_i 的位置变为随机位置.

综合以上2种情况,乌鸦 C_i 在跟随乌鸦 C_j 时得到的新位置如下:

$$x_{i,t+1} = \begin{cases} x_{i,t} + P_{r,i} \cdot l_{i,t} \cdot (m_{i,t} - x_{i,t}), & P_{r,j} > P_{j,t}; \\ a_i, & P_{r,j} < P_{j,t}. \end{cases} \quad (12)$$

式中: $P_{j,t}$ 表示乌鸦 C_j 第 t 轮的警觉概率; a_i 为随机位置.

1.3 改进的密度峰值聚类算法

在DPC算法中,截断距离需要人为设定且取值较为敏感^[25-26].对于同一组数据,取不同的截断距离将会得到不同的聚类结果,人为凭借经验对截断距离取值将影响最终的聚类效果,因此对截断距离 d_c 进行优化就变得很有必要.

聚类过程是一种无监督的学习过程,因此选择公正客观的评价方法去评判聚类结果的有效性十分重要^[27].为了得到更加准确以及分离度更好的聚类结果,引入有效性评价指标 V_{zi} 作为CSA的目标函数, V_{zi} 是由Xie等^[22]基于紧凑性和分离措施提出的有效性指标,该指标的紧凑性由基于黄金分割点的模糊隶属函数定义,分离性由单调线性函数描述,该方法的鲁棒性较其他指标更好,并且可以很好地处理一些异常数据.因此本文使用 V_{zi} 作为CSA的目标函数,对DPC算法进行优化.

假设样本数据集包含 n 个结构面,并且这 n 个结构面可以被聚为 k 类,则

$$V_{zi} = s \sum_{j=1}^n \sum_{i=1}^c (\omega_{ij})^2 + (1-s) \sum_{i=1}^{c-1} \sum_{k=i+1}^c (m_{ik})^2. \quad (13)$$

其中,

$$\omega_{ij} = \begin{cases} 1, & d \leq 0.372d_{\min}; \\ 1 - d/d_{\max}, & 0.372d_{\min} < d \leq 0.628d_{\min}; \\ 0, & d > 0.628d_{\min}. \end{cases} \quad (14)$$

$$m_{ik} = \| \mathbf{v}_i - \mathbf{v}_k \| / d_{\max}, \quad (15)$$

$$d_{\min} = \max_{i \neq k} (\| \mathbf{v}_i - \mathbf{v}_k \|), \quad (i, k = 1, 2, \dots, c); \quad (16)$$

$$d_{\max} = \max_{i \neq k} (\| \mathbf{v}_i - \mathbf{v}_k \|), \quad (i, k = 1, 2, \dots, c). \quad (17)$$

式中: s 为权重参数,其范围为 $[0, 1]$ ^[22],一般取值由经验判断; ω_{ij} 为隶属度函数,用于描述总体紧凑性; m_{ik} 为单调性函数,用于描述数据整体的分离度; d_{\max} 为数据集中两结构面间最大距离; d_{\min} 为数据集中两结构面间最小距离; \mathbf{v}_i 和 \mathbf{v}_k 均为各个分组中心的法向量.

本文通过CSA寻优确定 s 值,即CSA同时对

参数 s 以及截断距离 d_c 进行优化.Xie等^[22]根据聚类有效性试验发现, V_{zi} 越大,聚类结果越精确.因此,本文通过CSA多次迭代寻优,获得DPC算法中最优截断距离,进而通过DPC算法实现岩体结构面的聚类.

基于CSA-DPC算法的基本思想,下面介绍该结构面优势分组方法的基本流程:

步骤1 利用式(1)对初始数据进行预处理,计算每个样本数据的法向量;

步骤2 利用式(2)计算样本间相似距离;

步骤3 初始化截断距离 d_c ,利用截断距离 d_c 及对应公式分别计算出所有点的局部密度 ρ_i 及相对距离 δ_i ,利用DPC算法对数据进行第1次聚类,将结果作为CSA初始输入值;

步骤4 设置CSA的初始参数,包括:种群数量、最大迭代次数 T 、当前迭代次数 t ;

步骤5 引入 V_{zi} 指标,并初始化 V_{zi} 指标中的参数 s ;

步骤6 将 V_{zi} 指标作为CSA的目标函数,并更新该指标最大时对应的 d_c 值及参数 s ;

步骤7 利用CSA同时更新 d_c 及参数 s ;

步骤8 判断CSA是否满足迭代终止条件,即判断结果是否收敛,若满足条件,则转到步骤9,若不满足,则转到步骤4重复进行;

步骤9 将最优截断距离 d_c 、及参数 s 所对应的聚类结果进行输出.

1.4 聚类有效性评价

为了避免重复使用 V_{zi} 指标和对聚类结果评价的单一性,本文采用轮廓系数 S ^[14]和Xie等^[28]提出的 V_{xb} 指标进行聚类有效性评价,这2类指标在聚类的有效性评价中被广泛应用,其计算公式分别为

$$\left. \begin{aligned} S &= \frac{1}{N} \sum_{i=1}^N s_p(i), \\ s_p(i) &= \frac{b(i) - a(i)}{\max[a(i), b(i)]}. \end{aligned} \right\} \quad (18)$$

$$\left. \begin{aligned} V_{xb} &= \frac{\sum_{k=1}^c \sum_{i=1}^p \mu_{ij}^m d(\mathbf{p}_i, \mathbf{c}_k)}{i \min d(\mathbf{c}_j, \mathbf{c}_k)}, \\ \mu_{ij} &= \frac{d(\mathbf{p}_i, \mathbf{c}_k)^{-2(m-1)}}{\sum_{j=1}^c d(\mathbf{p}_i, \mathbf{c}_j)^{-2(m-1)}}, \\ \mathbf{P} &= \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_i\}, \\ \mathbf{C} &= \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}. \end{aligned} \right\} \quad (19)$$

式中: $s_p(i)$ 为数据集中某个样本的轮廓系数; S 为

数据结果的轮廓系数, S 取值范围为 $[-1, 1]$, S 值越大, 聚类结果越好^[28]; μ_{ij} 为隶属度, 代表第 i 个结构面属于第 j 个聚类中心 c_j 的程度; c_j 为各个分组中心的法向量; m 为模糊加权系数, 其取值范围一般为 $[1.5, 2.5]$ ^[28], 本文取值为2; $a(i)$ 为结构面 i 与其他结构面的平均距离; $b(i)$ 为结构面 i 与其他结构面的最小距离.

2 算法检验

为了验证改进的密度峰值聚类算法的有效性, 使用计算机程序随机生成4组界限不明显的结构面数据集, 假定结构面产状服从二维正态分布. 表1为生成的结构面数据. 利用dips软件绘制结构面的极点图, 如图2所示.

表1 结构面产状的二维正态分布参数

Table 1 Two-dimensional normal distribution parameters of rock discontinuities orientation

组号	倾向		倾角		结构面 条数
	平均值/(°)	方差	平均值/(°)	方差	
1	166	30	56	8	80
2	326	50	41	8	60
3	31	10	65	10	30
4	90	12	76	10	40

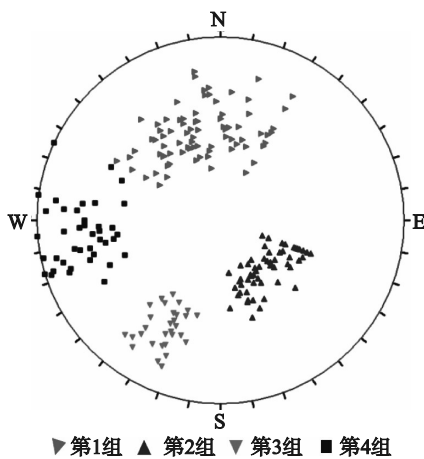


图2 模拟极点图

Fig. 2 Simulated pole plot

分别使用 K 均值聚类、DPC聚类、FCM聚类和改进的密度峰值聚类算法对4组结构面进行对比分析. 其中, K 均值聚类与FCM聚类的初始分组数设置为4组. 图3为4种算法的分组结果.

表2为4种算法聚类后各个分组中心结果对比. 表3为利用 S 指标和 V_{sb} 指标对4种算法聚类结果的有效性评价. 结果显示, 改进的密度峰值聚类算法得到的截断距离 d_c 为0.198, 权重参数 s 为1.

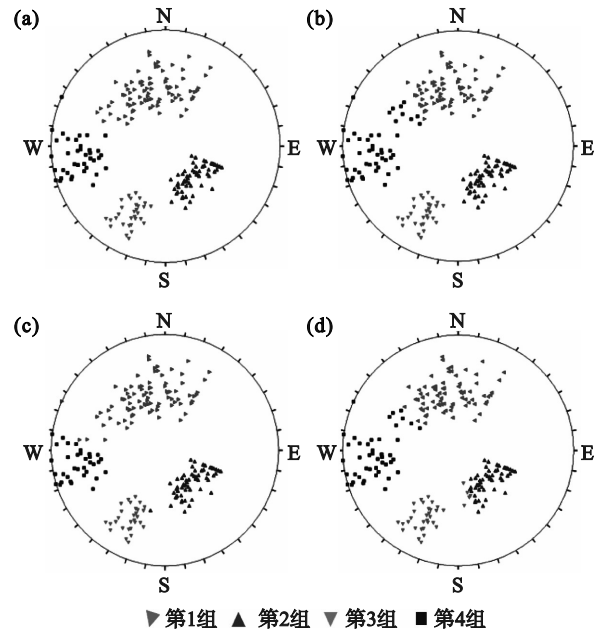


图3 结构面聚类结果对比

Fig. 3 Comparison of rock discontinuities clustering results

(a)—本文方法; (b)— K 均值聚类;
(c)—DPC聚类; (d)—FCM聚类.

对比以上4种算法的分组结果, 其中 K 均值聚类的初始中心是任意选取的, 具有较大的主观因素; 由图3a和图3c可知, DPC算法对截断距离 d_c 的取值十分敏感; 由图3d可知, FCM聚类算法在2类簇边缘存在重叠部分时可能导致同一组的结构面被分到另外一组, 从而影响分组结果; 对比图2和图3a可知, 本文方法所得的结构面聚类结果更加接近计算机模拟的结构面分组情况, 对于边界重叠的类簇也可以得到令人满意的结果, 并且使用本文方法不需要人工输入初始分组数以及分组中心. 因此, 本文方法具有更加精确的聚类结果, 验证了本方法的有效性.

3 工程应用

京西露天矿区位于科古尔琴山南麓北部, 与伊尔曼德矿区南部相接, 处于克孜勒库拉河、恰特尔特尔河、琼阿希河之间的一个近南北向延伸展布的河间地块式山岭地带, 为剥蚀中低山地形. 地势西北高东南低, 海拔标高1 650~1 875 m. 山体两侧为河谷段, 西伊尔曼得河经矿区由北向南流过, 东侧京西河由北向南穿过矿区, 地理坐标为 $81^{\circ}31'8.58''E, 44^{\circ}20'3.47''N$. 京西露天矿边坡坡高115 m, 一共有11级台阶, 坡上发育凝灰岩. 本文以从野外获得的118个不连续面的产状

信息为例.利用这些结构面数据进一步验证所提出的改进DPC算法的合理性和工程实用性.图4

为实测结构面产状极点图;图5为实测结构面等值线图;图6为实测结构面走向玫瑰花图.

表2 结构面聚类结果对比
Table 2 Comparison of rock discontinuities clustering results

组号	本文方法			K均值聚类		
	倾向/(°)	倾角/(°)	样本个数	倾向/(°)	倾角/(°)	样本个数
1	165.87	55.08	82	171.53	54.54	73
2	316.05	42.14	60	316.05	42.14	60
3	27.87	66.33	30	27.87	66.33	30
4	83.58	75.00	38	90.77	71.50	47
组号	DPC聚类			FCM聚类		
	倾向/(°)	倾角/(°)	样本个数	倾向/(°)	倾角/(°)	样本个数
1	162.93	55.73	86	170.22	54.72	75
2	311.11	42.37	61	313.01	42.14	55
3	32.74	66.74	29	30.51	62.87	35
4	81.31	75.39	34	91.97	72.50	45

表3 各聚类结果评价指标
Table 3 Each clustering result evaluation index

算法类型	S	V_{xb}
本文方法	0.711	26.41
K均值聚类	0.700	26.88
DPC聚类	0.679	27.44
FCM聚类	0.694	27.50

用S指标、 V_{xb} 指标对分组情况评价.表5为4种方法聚类结果的有效性检验.

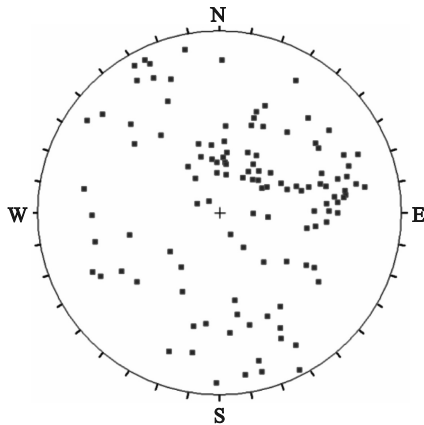


图4 实测结构面产状极点图

Fig. 4 Pole map of the measured rock discontinuity orientation

由图5可以看出,各组结构面之间的界限并不清晰,人为无法确定大致分组数范围,因此对结构面数据采用改进的DPC算法进行分组,并用S指标以及 V_{xb} 指标进行有效性评价.

经过算法寻优得到最佳截断距离 d_c 为0.102, s 为0.987.图7为使用3种常规方法以及改进的DPC算法对实测结构面分组情况.表4为使用改进方法对实测结构面数据分组信息以及使

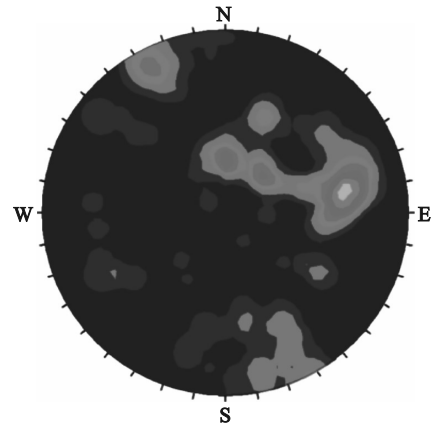


图5 实测结构面等值线图

Fig. 5 Contour plot of measured rock discontinuity

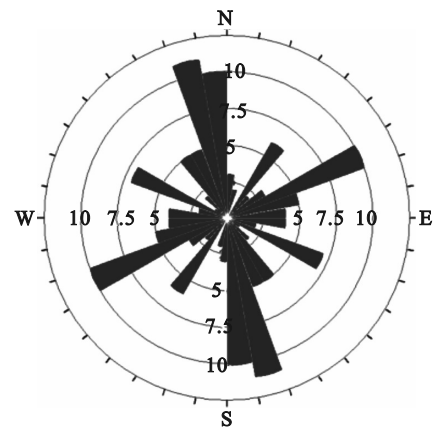


图6 实测结构面走向玫瑰花图

Fig. 6 Rose diagram of measured rock discontinuity trend

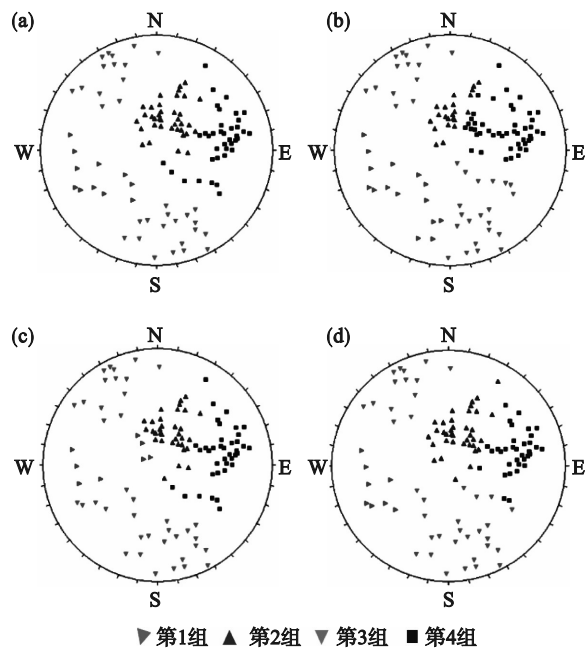


图 7 实测结构面聚类结果对比

Fig. 7 Comparison of measured rock discontinuities clustering results

(a)—本文方法; (b)— K 均值聚类;
(c)—DPC 聚类; (d)—FCM 聚类.

表 4 实测结构面聚类结果
Table 4 Clustering results of measured rock discontinuities

组号	倾向/(°)	倾角/(°)	样本个数	S	V_{xb}
1	265.29	61.21	38	0.52	24.09
2	198.41	36.78	37		
3	205.34	71.22	32		
4	62.73	60.91	11		

表 5 实测结构面聚类有效性检验
Table 5 Test for the effectiveness of clustering of measured rock discontinuities

分组方法	评价指标	
	S	V_{xb}
本文方法	0.52	24.09
K 均值聚类	0.46	34.11
DPC 聚类	0.36	29.76
FCM 聚类	0.50	29.61

对比图 5 和图 7a 可以发现,聚类结果符合实际情况.使用该方法不需要人为指定初始聚类组数以及初始聚类中心就可以完成聚类,在聚类过程中可以避免受到边缘点以及异常点的影响,得到组间边界清晰的聚类结果,克服了常见结构面优势分组方法存在的 2 个难点问题.对比图 5 和图 7 可知,本文方法的分组结果更加符合现场实

测结构面分布的等值线图,传统的 K 均值聚类简单易实现,但需要人为预先指定分组数,对于类似该工程分组不明显的情况,很难确定初始分组数;DPC 算法不需要预先指定分组数以及分组中心,但受截断距离影响较大,不同的截断距离会导致出现不同的分组情况;模糊 C 均值聚类引入隶属度概念,聚类结果更加灵活,但当分组边界不明确,容易陷入局部最小值;本文方法在 DPC 算法基础上利用 CSA 对截断距离进行寻优,使 DPC 算法可以取得更加精确的结果.由表 5 可以看出,对于工程实测结构面数据,在利用 2 种有效性指标对 4 种方法的分组结果进行评价中,本文方法的聚类结果具有更高的 S 值以及更低的 V_{xb} 值,进一步说明了改进 DPC 算法的有效性以及优势.

4 结 论

1) 传统的模糊聚类算法需要人为指定初始聚类组数以及初始聚类中心,聚类结果受到人为主观因素影响较大.DPC 算法具有对数据异常点不敏感、鲁棒性强、计算速度快等优势,且不需要人为指定初始聚类组数以及初始聚类中心,聚类结果具有唯一性,优越性明显.

2) 利用 CSA 强大的寻优能力对 DPC 算法中的截断距离 d_c 进行优化.以 V_{zi} 指标建立目标函数,对 V_{zi} 指标中权重参数 s 进行优化,寻找出聚类准确率指标最大时所对应的权重参数 s 以及截断距离 d_c .该方法有效避免了截断距离 d_c 较为敏感以及聚类结果易受边缘点和异常点影响的问题.通过与 3 种传统的聚类分析方法进行对比,结果表明改进后的方法具有良好的实用性.

3) 应用改进后的 DPC 算法对京西露天矿边坡实测 118 组结构面进行聚类分析,所提方法有效避免了边缘点以及异常点的影响,将实测结构面分为 4 组,分组结果符合实际情况,进一步验证了该方法的有效性及其工程应用价值.

参考文献:

- [1] 陈剑平,石丙飞,王清.工程岩体随机结构面优势方向的表示法初探[J].岩石力学与工程学报,2005,24(2): 241-245.
(Chen Jian-ping, Shi Bing-fei, Wang Qing. Study on the dominant orientations of random fractures of fractured rock masses [J]. Chinese Journal of Rock Mechanics and Engineering, 2005, 24(2): 241-245.)
- [2] Liu J, Zhao X D, Xu Z H. Identification of rock discontinuity sets based on a modified affinity propagation algorithm [J].

- International Journal of Rock Mechanics and Mining Sciences*, 2017, 94: 32-42.
- [3] 张化进,吴顺川,韩龙强.基于DBSCAN选择性聚类集成的岩体结构面优势产状分组方法[J].*岩土力学*, 2022, 43(6): 1585-1595.
(Zhang Hua-jin, Wu Shun-chuan, Han Long-qiang. Dominant partitioning method of rock mass discontinuity based on DBSCAN selective clustering ensemble[J].*Rock and Soil Mechanics*, 2022, 43(6): 1585-1595.)
- [4] 葛云峰,钟鹏,唐辉明,等.基于钻孔图像的岩体结构面几何信息智能测量[J].*岩土力学*, 2019, 40(11): 4467-4476.
(Ge Yun-feng, Zhong Peng, Tang Hui-ming, et al. Intelligent measurement on geometric information of rock discontinuities based on borehole image[J].*Rock and Soil Mechanics*, 2019, 40(11): 4467-4476.)
- [5] Park H J, West T R, Woo I. Probabilistic analysis of rock slope stability and random properties of discontinuity parameters, Interstate Highway 40, Western North Carolina, USA[J].*Engineering Geology*, 2005, 79(3/4): 230-250.
- [6] Shanley R J, Mahtab M A. Delineation and analysis of clusters in orientation data[J].*Journal of the International Association for Mathematical Geology*, 1976, 8(1): 9-23.
- [7] 孙宪春,万力,蒋小伟.节理产状分组的 k 均值聚类分析及其分组结果的费歇尔分布验证法[J].*岩土力学*, 2008, 29(sup1): 533-537.
(Sun Xian-chun, Wan Li, Jiang Xiao-wei. Effective categorization of joints by k -means and cluster analysis its verification by Fisher distribution [J].*Rock and Soil Mechanics*, 2008, 29(sup1): 533-537.)
- [8] Boomija M D, Phil M. Comparison of partition based clustering algorithms[J].*Journal of Computer Applications*, 2008, 1(4): 18-21.
- [9] Zhang T, Ramakrishnan R, Livny M. Birch: an efficient data clustering method for very large databases [J].*Acm Sigmod Record*, 1996, 25(2): 103-114.
- [10] Saini S, Rani P. A survey on sting and clique grid based clustering methods[J].*International Journal of Advanced Research in Computer Science*, 2017, 8(5): 1510-1512.
- [11] Hammah R E, Curran J H. Fuzzy cluster algorithm for the automatic identification of joint sets [J].*International Journal of Rock Mechanics and Mining Sciences*, 1998, 35(7): 889-905.
- [12] Zhang J H, Yin Z, Wang R B. Pattern classification of instantaneous cognitive task-load through GMM clustering, laplacian eigenmap, and ensemble SVMs [J].*IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017, 14(4): 947-965.
- [13] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J].*Science*, 2014, 344(6191): 1492-1496.
- [14] Li X B, Wang Z W, Peng K, et al. Ant colony ATTA clustering algorithm of rock mass structural plane in groups [J].*Journal of Central South University*, 2014, 21(2): 709-714.
- [15] 王述红,任艺鹏,陈俊智,等.一种改进鱼群聚类算法在结构面分组中的应用[J].*东北大学学报(自然科学版)*, 2019, 40(3): 420-424.
(Wang Shu-hong, Ren Yi-peng, Chen Jun-zhi, et al. An improved fish swarm clustering algorithm for structural grouping [J].*Journal of Northeastern University (Natural Science)*, 2019, 40(3): 420-424.)
- [16] 蔡美峰,王鹏,赵奎,等.基于遗传算法的岩体结构面的模糊 C 均值聚类方法[J].*岩石力学与工程学报*, 2005, 24(3): 371-376.
(Cai Mei-feng, Wang Peng, Zhao Kui, et al. Fuzzy C -means cluster analysis based on genetic algorithm for automatic identification of joint sets [J].*Chinese Journal of Rock Mechanics and Engineering*, 2005, 24(3): 371-376.)
- [17] Li Y Y, Wang Q, Chen J P, et al. K -means algorithm based on particle swarm optimization for the identification of rock discontinuity sets [J].*Rock Mechanics and Rock Engineering*, 2015, 48(1): 375-385.
- [18] Jimenez-Rodriguez R, Sitar N. A spectral method for clustering of rock discontinuity sets [J].*International Journal of Rock Mechanics and Mining Sciences*, 2006, 43: 1052-1061.
- [19] 侯钦宽,雍睿,杜时贵,等.基于编网算法的岩体结构面优势产状聚类分析[J].*岩石力学与工程学报*, 2020, 39(sup1): 2871-2881.
(Hou Qin-kuan, Yong Rui, Du Shi-gui, et al. A method for clustering orientation data of discontinuities of rock mass based on netting algorithm [J].*Chinese Journal of Rock Mechanics and Engineering*, 2020, 39(sup1): 2871-2881.)
- [20] 张奇,王清,阙金声,等.基于凝聚层次聚类分析法的岩体随机结构面产状优势分组[J].*岩土工程学报*, 2014, 36(8): 1432-1437.
(Zhang Qi, Wang Qing, Que Jin-sheng, et al. Dominant partitioning of discontinuities of rock masses based on agnes [J].*Chinese Journal of Geotechnical Engineering*, 2014, 36(8): 1432-1437.)
- [21] 刘铁新,邓建辉,陈滔,等.考虑结构面间距的双因素分组方法研究[J].*岩土力学*, 2017, 38(sup1): 219-224.
(Liu Tie-xin, Deng Jian-hui, Chen Tao, et al. Double-factor clustering method considering spacing of discontinuities [J].*Rock and Soil Mechanics*, 2017, 38(sup1): 219-224.)
- [22] Xie X J, Wang Y, Zhong L. A new fuzzy clustering validity index with strong robustness[M/OL]//*Fuzzy Information & Engineering and Operations Research & Management*. Heidelberg: Springer Berlin Heidelberg, 2013: 317-327.
- [23] Askarzadeh A. A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm[J].*Computers & Structures*, 2016, 169: 1-12.
- [24] Jimenez R. Fuzzy spectral clustering for identification of rock discontinuity sets [J].*Rock Mechanics and Rock Engineering*, 2008, 41: 929-939.
- [25] Wang L M, Li M Y, Han X M, et al. Improved density peak clustering algorithm based on choosing strategy automatically for cut-off distance and cluster centre [J].*Tehnicky Vjesnik-Technical Gazette*, 2018, 25(2): 536-545.
- [26] Mehmooda R, Bie R F, Jiao L B, et al. Adaptive cutoff distance: clustering by fast search and find of density peaks [J].*Journal of Intelligent & Fuzzy Systems*, 2016, 31(5): 2619-2628.
- [27] 刘来权,陈燕,雷燕瑞.模糊 C 均值聚类算法的有效性检验研究[J].*软件*, 2017, 38(2): 16-18.
(Liu Lai-quan, Chen Yan, Lei Yan-rui. Research on the validity of fuzzy C mean clustering algorithm [J].*Computer Engineering & Software*, 2017, 38(2): 16-18.)
- [28] Xie X L, Beni G. A validity measure for fuzzy clustering [J].*IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(8): 841-847.