

模拟退火聚类算法在结构面产状分组中的应用

王述红, 朱宝强, 王鹏宇

(东北大学 资源与土木工程学院, 辽宁 沈阳 110819)

摘 要: 鉴于以往的结构面产状分组方法常存在算法复杂、聚类精度差及分组效率低的不足,提出了一种新型的融合模拟退火算法及 K-means 聚类(SAK)的结构面分组算法,该算法简单易实现.利用模拟退火算法的退火原理,对 K-means 算法聚类的结构面分组结果进行优化,以期克服 K-means 算法易受初始聚类中心影响的缺陷.计算机模拟生成的结构面数据的分析表明,所提方法相较于传统 K-means 算法具有明显优势.将该方法应用于重庆市三环高速公路兴隆隧道实测结构面的分组中,并与已有方法进行对比.结果表明:该方法不仅聚类精度高,而且迭代速度也较快,具有较强的工程实用性.

关 键 词: 岩体;结构面产状;优势分组;模拟退火算法;K-means 聚类

中图分类号: TU 45 **文献标志码:** A **文章编号:** 1005-3026(2020)09-1328-06

Application of Simulated Annealing Clustering Algorithm in Grouping of Discontinuity Orientation

WANG Shu-hong, ZHU Bao-qiang, WANG Peng-yu

(School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: ZHU Bao-qiang, E-mail: zbq2289675237@163.com)

Abstract: Aiming at the complexity, poor clustering accuracy and low grouping efficiency of the previous methods of dominant grouping of discontinuity orientation, a new method of dominant grouping of discontinuity orientation based on simulated annealing algorithm and K-means clustering (SAK) was proposed. The algorithm is simple and easy to implement. Based on the annealing principle of simulated annealing algorithm, the grouping results of K-means algorithm were optimized, which aimed to overcome the shortcoming of the K-means algorithm's susceptibility to the initial clustering center. The analysis of discontinuities generated by computer simulation showed that the proposed method is superior to the traditional K-means algorithm. The method was applied to the grouping of measured discontinuity orientation of Xinglong Tunnel on Third Ring Expressway of Chongqing City, and compared with the existing methods. The results showed that this method not only has high clustering accuracy, but also has fast iteration speed and strong engineering practicability.

Key words: rock mass; discontinuity orientation; dominant grouping; simulated annealing algorithm; K-means clustering

岩体是由结构面和岩石组成的复杂块体,它作为一种非均质材料,破坏时往往是沿着结构面断裂,因此,结构面决定了岩体的强度及稳定性^[1].然而,由于野外环境的复杂性,结构面的很多特征参数获取较为困难,并且在利用数学手段对结构面进行分组时,分组参数的增加则使得计

算的维度大大增加,数据之间的差异性使得其相似性也难以度量,因此目前最常用的仍然是依据结构面的产状数据(即倾向和倾角)进行分组.传统的产状分组方法多是为人为的由玫瑰花图和极点图进行直观判断,主观性很大,无法准确定量地给出客观性分组结果.

为了避免这种主观性,在 1976 年,Shanley 等^[2]首次提出了一种客观的聚类分组方法,该方法有着严格的数学理论依据,可以得到较为理想的结构面分组结果,然而在寻找密度点时小球半径的合理确定问题一直未能得到解决.后来,Harrison 等^[3]和 Hammah 等^[4]在总结前人的研究成果后又提出并发展了模糊 C 均值(FCM)聚类算法在结构面分组中的应用,均取得了不错的进展.但上述方法本质上都属于局部寻优的聚类方法,当分组边界不明确时,极易陷入局部极小值.随着计算机技术的发展,智能算法等机器学习方法迅速兴起,越来越多的学者将智能算法应用于结构面的分组中.李宁等^[5]将改进遗传算法引入支持向量机的分类中,以此对结构面进行分组,分组结果较为客观,但由于遗传算法过程较为复杂,因此该方法应用时存在一定的缺陷.后来,Li 等^[6]、王述红等^[7]、Li 等^[8]分别将蚁群算法、鱼群算法、粒子群算法引入岩体结构面分组的 K-means 聚类算法中,均得到较为满意的分组结果,但是上述算法的复杂性使得这些方法的分组效率较低,不利于实际工程应用.

鉴于此,本文提出了一种新型的融合模拟退火算法与 K-means 聚类(simulated annealing algorithm and K-means clustering, SAK)的结构面产状优势分组方法.该方法通过对 K-means 聚类算法进行优化,克服了 K-means 算法对初始值敏感,从而影响聚类结果的缺陷^[9],该算法简单易实现,最终可搜索到全局最优的结构面分组结果.通过对计算机模拟生成的结构面数据及现场实测结构面数据进行分析,并与已有方法进行对比,验证了该方法的合理性、高效性和工程实用性.

1 岩体结构面分组数学模型的建立

1.1 结构面产状的空间表示法

对结构面产状进行分组,先要对结构面倾向 α 和倾角 β 数据进行归一化处理,常用的方法是将结构面视作无厚度的无限延伸平面,然后用其单位法向量来表示^[10].由数学知识可知,这可以由空间中的单位球体表示,其在各坐标轴上的分量如下:

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

(1)

则结构面的单位法向量坐标可表示为

$$\boldsymbol{p} = (n_1, n_2, n_3).$$
(2)

1.2 结构面产状数据之间的相似性度量

为了避免倾向相差约 180° 的两组高陡倾角结构面分组时出错的情况,本文采用任意两结构面 \boldsymbol{p}_1 和 \boldsymbol{p}_2 之间所夹的锐角 γ 的正弦值作为两结构面之间产状的相似性度量^[11],即

$$\gamma = \arccos |\boldsymbol{p}_1 \cdot \boldsymbol{p}_2^T|.$$

(3)

则两结构面单位法向量之间的距离为

$$d(\boldsymbol{p}_1, \boldsymbol{p}_2) = \sin\gamma = \sqrt{1 - (\boldsymbol{p}_1 \cdot \boldsymbol{p}_2^T)^2}.$$

(4)

1.3 结构面分组的目标函数

假设有 n 个结构面 $F_i (i=1, 2, \dots, n)$, 其单位法向量分别为 $\boldsymbol{p}_i (i=1, 2, \dots, n)$, 可划分为 k 组, 每组聚类中心为 $\boldsymbol{c}_j (j=1, 2, \dots, k)$, 定义 u_{ij} 为第 i 个结构面属于第 j 个分组的隶属度, 定义目标函数 J 为所有结构面单位法向量 \boldsymbol{p} 与各分组中心 \boldsymbol{c} 之间的距离和(即结构面分组的总类间离散度), 即

$$u_{ij} = 1 / \sum_{s=1}^k \left(\frac{d(\boldsymbol{p}_i, \boldsymbol{c}_j)}{d(\boldsymbol{p}_i, \boldsymbol{c}_s)} \right)^{2/(m-1)},$$

(5)

$$J = \sum_{j=1}^k \sum_{i=1}^n d(\boldsymbol{p}_i, \boldsymbol{c}_j).$$

(6)

式中: \boldsymbol{p}_i 分别为第 i 个结构面的单位法向量坐标; \boldsymbol{c}_j 和 \boldsymbol{c}_s 为各分组中心的法向量坐标; m 为权值分配系数, 一般取 $1 \sim 2$, 本文取 $m=2$; $d(\boldsymbol{p}_i, \boldsymbol{c}_j)$ 和 $d(\boldsymbol{p}_i, \boldsymbol{c}_s)$ 分别为第 i 个结构面到第 j 个和第 s 个分组中心之间的距离, 采用式(4)计算. 由式(6)可知: 结构面参数到各结构面分组中心的距离越小, 聚类的误差越小, 分组的结果也更精确.

2 基于 SAK 的岩体结构面产状分组

2.1 算法基本原理

模拟退火算法^[12](SA)在 20 世纪 80 年代由 Metropolis 首次提出的一种启发式随机搜索算法,其基本原理来源于固体的退火过程.在 SA 算法中,目标函数由内能模拟,控制参数为温度和温度冷却参数,通过对初始解重复执行“扰动产生新解—计算目标函数差—由 Metropolis 准则判断是否接受新解”的过程,逐步进行优化,算法迭代终止时的当前解即为近似的最优解.该算法具有渐近收敛性和并行性,已在理论上被证明是一种以概率 1 收敛于全局最优解的全局优化算法^[13].

本文将其引入结构面分组中,组成全局寻优能力强的 SAK 算法,通过对 K-means 算法聚类

结果进行优化,从而有效改进 K - means 算法易受初始聚类中心影响的缺陷.

2.2 SAK 算法的实现过程

在 SAK 算法中,初始温度 t_0 和温度冷却参数 q 的选取是非常重要的,它们对算法的收敛速度和全局最优性有很大的影响. 当 q 取值较大时,算法的收敛速度大大降低;当 q 取值较小时,温度下降过快,此时则不易获得全局最优解. 为了使最初产生的初始解能够按照 Metropolis 准则被接受,算法一开始就应达到准平衡状态,因此选取初始温度为 t_0 时的聚类结果 $t_0 = J_0$ 作为算法的初始解,温度冷却参数由优化效果适当选取,降温过程如下:

$$t(b+1) = t(b)q. \tag{7}$$

式中: $t(b)$ 为当前循环次数下的温度值; $t(b+1)$ 为进一步循环后的温度值; q 为冷却参数,略小于 1.

另外,SAK 算法中新解的产生是通过对当前解随机扰动得到,扰动公式如下:

$$r = \text{fix}[\text{rand}() \times n + 1], \tag{8}$$

$$h = \text{fix}[\text{rand}() \times k + 1], \tag{9}$$

$$s'_r = \begin{cases} s_r + h - k, & s_r + h > k; \\ s_r + t, & s_r + h \leq k. \end{cases} \tag{10}$$

式中: r 为随机选择的一个结构面样本; $\text{rand}()$ 为 $[0,1]$ 之间的随机数; n 为结构面个数; fix 为取整函数; h 为因扰动生成的一个介于 $[1,k]$ 的整数; k 为分组数; s_r 为随机抽取的一个结构面所属的分类(即当前解); s'_r 为扰动产生的结构面所属的分类(即新解).

基于 SAK 算法的结构面优势分组流程如下:

- 1) 首先对待分组的结构面数据集进行归一化处理并执行 K - means 聚类,将聚类分组的结果作为初始解 s ,由式(6)计算出初始目标函数值 J_s ;
- 2) 初始化温度 $t_0 = J_s$ 、最大迭代次数 L 、温度冷却参数 q 和每个温度 t 下的循环次数 l (即 Metropolis 链长);
- 3) 通过随机扰动产生新解 s' ,这代表随机产生了一个新的结构面分组方式,计算此时对应的目标函数值 $J_{s'}$;
- 4) 判断此时的目标函数值 $J_{s'}$ 是否为最优解,如果是则保存此时的结构面分组方式为最优分组、 $J_{s'}$ 为最优目标函数值,否则执行步骤 6);
- 5) 计算目标函数差 $\Delta J = J_{s'} - J_s$,并判断其是否小于 0,若小于 0 则接受新解 s' 作为下一个当前解,否则按照 Metropolis 准则(即以概率 $\exp(-\Delta J/t)$)接受新解 s' ;
- 6) 判断当前迭代次数是否达到最大迭代次

数,若是,则执行步骤 7),若否,则继续执行步骤 3) ~ 5);

7) 判断是否达到终止温度,若是,则算法结束,输出当前聚类划分结果作为结构面优势组分分类结果;若否,则降低温度,继续执行步骤 3) ~ 6).

2.3 聚类有效性评价

为了避免对最优分组结果判断的单一性,本文采用模糊分类系数 F 和分类熵指标 H 进行聚类有效性评价,这两类指标也是最常用的衡量聚类有效性的函数^[5],其计算公式分别为

$$F = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k u_{ij}^2, \tag{11}$$

$$H = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k u_{ij} \log_a(u_{ij}). \tag{12}$$

式中: n 为结构面数据集的个数; u_{ij} 为第 i 个结构面属于第 j 个分组的隶属度,由式(5)计算所得; a 为对数的底数, $a \in (1, +\infty)$,规定当 $u_{ij} = 0$ 时 $u_{ij} \log_a(u_{ij}) = 0$,本文取 $a = 10$. 当 F 越大, H 越小,表明分类的模糊度越小,聚类效果越好. 因此,对于同一方法,当 F 较大、 H 较小时,结构面分组结果较好;而对于不同方法, F 相对较大、 H 相对较小的方法为较好的方法.

3 算法准确性验证

为了验证 SAK 算法应用于结构面产状优势分组中的准确性,采用计算机随机模拟生成了 3 组界限并不明显的 120 个结构面数据(服从正态分布). 表 1 为这三组结构面数据的详细参数及两种算法聚类后分组中心的对比结果;表 2 为聚类有效性评价结果;图 1 为这 3 组结构面数据的

表 1 结构面数据参数及分组中心
Table 1 Discontinuity data parameters and grouping centers

组号	类别	产状		结构面数量/组
		倾向/(°)	倾角/(°)	
第一组	已知中心	130	79	30
	SAK	128. 63	77. 86	30
	K - means	122. 58	75. 77	45
第二组	已知中心	308	46	40
	SAK	311. 01	46. 28	42
	K - means	301. 93	54. 34	30
第三组	已知中心	48	49	50
	SAK	46. 94	50. 16	48
	K - means	25. 01	30. 51	45

极点及密度等值线图;图 2a 和图 2b 分别为采用 K-means 算法和 SAK 算法分组的结果. SAK 算法相关参数设置为:初始温度 $t_{\text{begin}} = 20\text{ }^{\circ}\text{C}$,终止温度 $t_{\text{end}} = 0.1\text{ }^{\circ}\text{C}$,温度冷却参数 $q = 0.92$,每个温度 t 下的循环次数(即 Metropolis 链长) $l = 400$,最大迭代次数 $L = 500$.

表 2 随机数据聚类有效性检验			
Table 2 Clustering validity test of random data			
算法	分组数	F	H
K-means 算法	$k = 2$	0.425 4	0.433 4
	$k = 3$	0.434 0	0.414 9
	$k = 4$	0.406 2	0.485 0
SAK 算法	$k = 2$	0.460 9	0.403 7
	$k = 3$	0.470 1	0.392 1
	$k = 4$	0.426 8	0.483 2

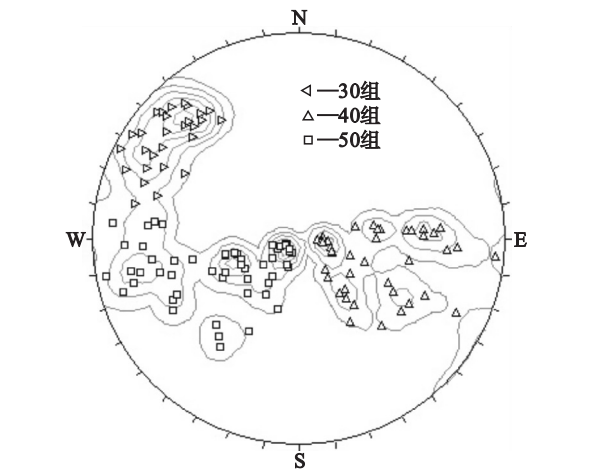


图 1 随机生成的结构面产状极点及密度等值线图
Fig. 1 Pole and contour plot of randomly generated discontinuity orientation

由表 1 中分组结果与已知结果的对比中可以看出,SAK 算法的分组中心与已知中心更为接近;表 2 显示了两种算法的聚类有效性结果,当分

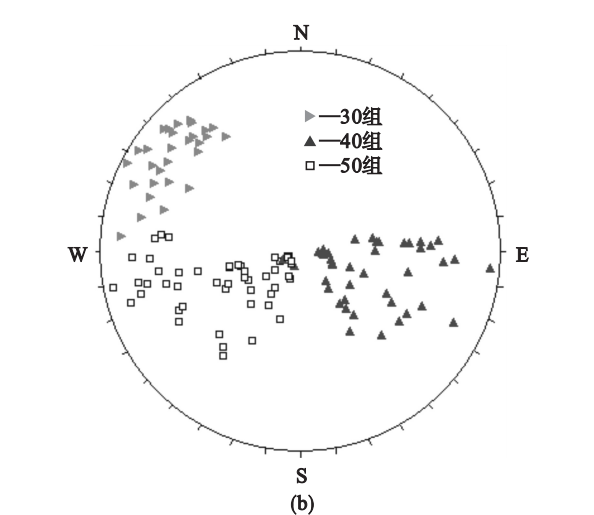
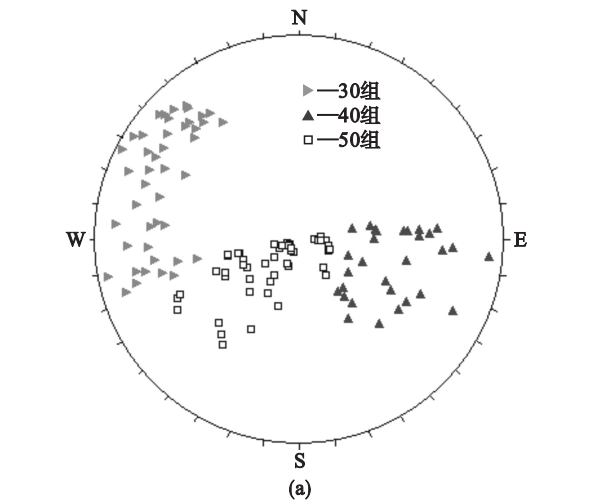


图 2 结构面分组结果
Fig. 2 Results of cluster analysis of discontinuities
(a)—K-means; (b)—SAK.

组数为 2 和 3 时,结果较为接近,但综合比较 F 和 H 两指标后可知最优分组数为 3. 此时,从图 1、图 2a 和图 2b 的对比中,也可以很明显地看出 SAK 算法准确性更高,分组结果更符合图 1 中实际的情况.

4 工程实例

重庆市三环高速公路合川至长寿段兴隆隧道位于重庆市渝北区木耳镇,隧址区属构造侵蚀丘陵地貌,隧道大体沿垂直构造线方向布设,与岩层走向呈大角度相交,穿越地层主要为侏罗系上沙溪庙组地层,地层分布连续. 围岩岩性主要为侏罗系中统上沙溪庙组泥岩和砂岩,中风化岩体较完整,发育高倾角的构造裂隙. 隧址区无活动性断裂、泥石流、滑坡等不良地质现象,地下水类型主要为第四系松散土层孔隙水及基岩裂隙水.

本文以隧道洞口段现场实测的 118 个基岩露头结构面产状数据为例,进一步验证所提 SAK 模型在结构面分组中的合理性和工程实用性. 赤平投影下结构面产状的极点及密度等值线图如图 3 所示.

由图 3 可看出,各组结构面之间的界限并不明显,根据该图可大致判断出分组数为 2~4,因此对结构面数据分别采用 K-means 和 SAK 算法进行聚类分析,并与文献[8]中的 KPSO 算法进行对比,利用 F 和 H 两指标进行聚类有效性评价,对比结果如表 3 所示. 由表中结果可以看出,当分组数为 2 时,聚类效果最好,此时三种算法聚类结果分别如图 4a、图 4b 和图 4c 所示.

K-means算法、SAK 算法和文献[8]中 KPSO 算法迭代速度对比结果如图 5 所示;结构面优势产状统计结果见表 4.

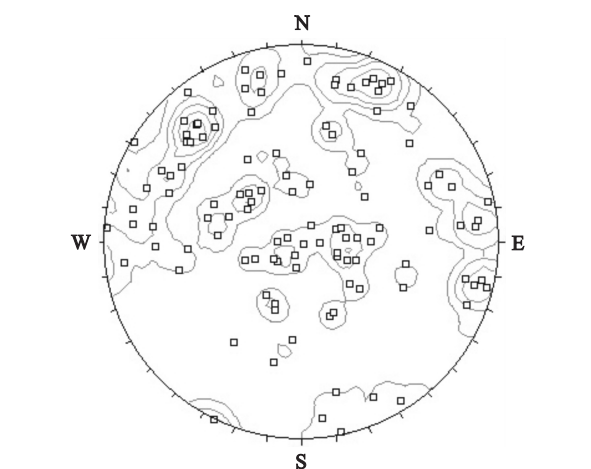


图 3 实测结构面产状极点及密度等值线图
Fig. 3 Pole and contour plot of measured discontinuity orientation

表 3 实测数据聚类有效性检验 Table 3 Clustering validity test of measured data			
算法	分组数	<i>F</i>	<i>H</i>
K-means 算法	<i>k</i> = 2	0. 518 6	0. 292 8
	<i>k</i> = 3	0. 372 6	0. 452 9
	<i>k</i> = 4	0. 301 1	0. 562 5
SAK 算法	<i>k</i> = 2	0. 543 0	0. 281 5
	<i>k</i> = 3	0. 400 7	0. 435 9
	<i>k</i> = 4	0. 322 5	0. 546 3
KPSO 算法	<i>k</i> = 2	0. 537 8	0. 283 6
	<i>k</i> = 3	0. 395 9	0. 438 2
	<i>k</i> = 4	0. 314 7	0. 553 7

对比图 3、图 4a 和图 4b,不难看出,SAK 算法的结构面优势分组结果更符合密度等值线图所反映的分组区域,且与图 4c 中 KPSO 算法的分组结果几乎一致. 原因主要是 K-means 算法通过随机生成初始聚类中心进行聚类,因此其聚类结果容易不理想,而采用全局寻优能力强的模拟退火算法可以对 K-means 算法聚类结果进行优化,从而改善原始 K-means 聚类算法由于初始聚类中心选择不当而使聚类效果不理想的缺陷. 根据表 3 中聚类有效性检验结果的对比也可以看出,无论分组数为几组,SAK 算法及 KPSO 算法的聚类的结果都要大大优于 K-means 算法的聚类结果,并且本文所提方法略优于文献[8]中的 KPSO 算法,进一步验证了所提算法具有较高的精度.

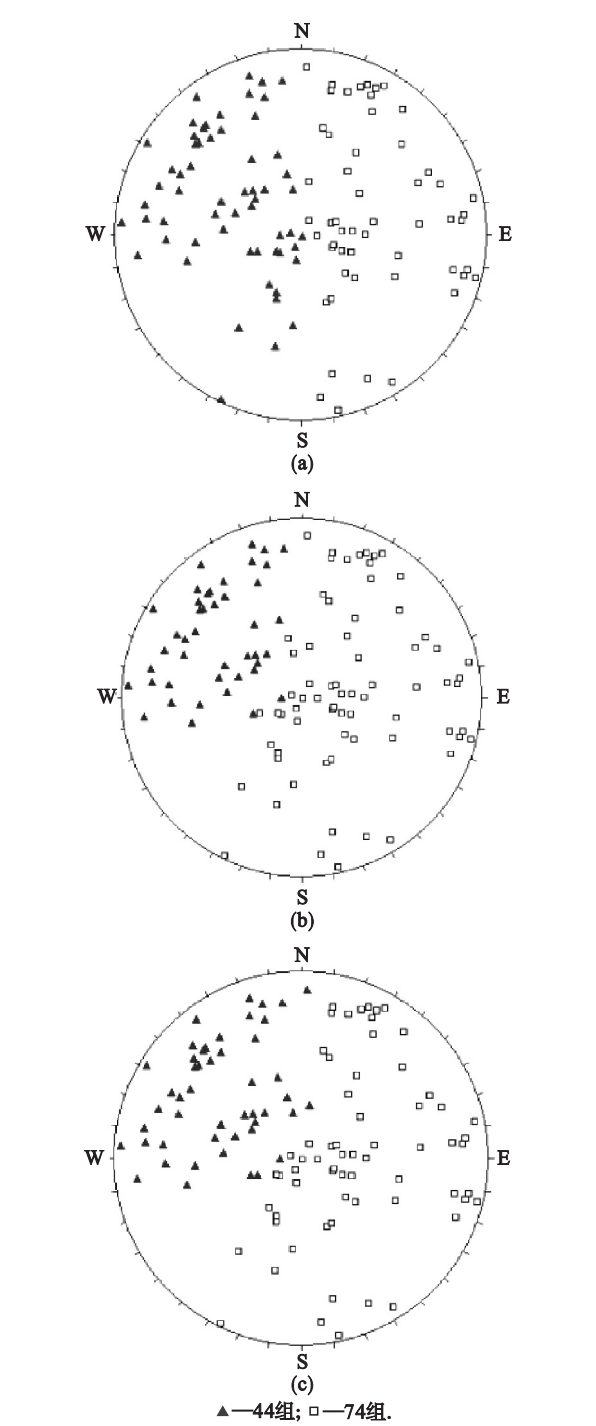


图 4 结构面分组结果 (*k* = 2)
Fig. 4 Grouping results of discontinuities (*k* = 2)
(a)—K-means; (b)—SAK; (c)—KPSO.

另外,从图 5 中可以很明显的看出,同样都是对 K-means 算法进行优化,文献[8]中算法整体迭代速度大大降低,而本文提出的 SAK 算法的迭代速度则降低较少,本文方法的迭代速度要大大优于文献[8]中 KPSO 算法(大约提高 50%),表明了算法的高效性. 因此,基于模拟退火算法 K-means 聚类(SAK)的结构面优势分组方法更适合应用于结构面的分组中,具有较强的合理性和工

程实用性.

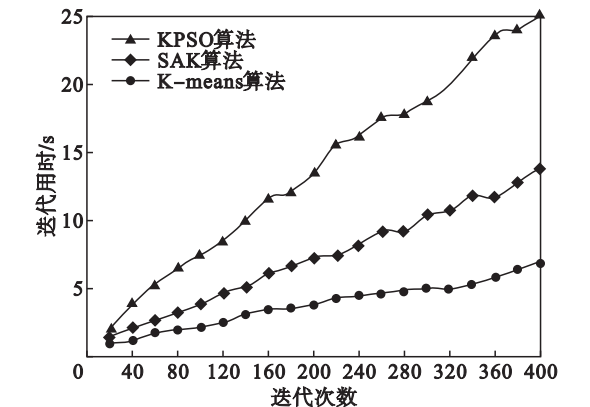


图 5 迭代速度对比结果

Fig. 5 Comparison results of iteration speed

表 4 结构面优势产状分组结果

Table 4 Grouping results of discontinuity dominant orientation

组号	优势产状		结构面数量/组
	倾向/(°)	倾角/(°)	
第一组	201. 64	61. 01	74
第二组	133. 17	65. 89	44

5 结 语

本文将模拟退火算法引入了结构面的优势分组中,提出了一种新型的基于模拟退火算法 K - means 聚类(SAK)的结构面优势分组算法.该算法通过逐次迭代后进行最优解的精确搜索,最终搜索到全局最优的结构面分组结果,有效克服了 K - means 算法对初始值敏感,从而影响聚类效果的缺陷,避免了人为划定分组方式的主观性.结合计算机模拟生成的结构面数据及重庆市三环高速公路兴隆隧道洞口段现场实测的结构面数据,将该算法与文献[8]中提出的 KPSO 算法进行对比,证明了 SAK 算法在聚类准确性、聚类精度及迭代速度上均较优,可以得到较为合理的结构面分组结果,有一定的推广应用价值.

参考文献:

[1] Wang S H,Huang R Q,Ni P P P,et al. Advanced discretization of rock slope using block theory within the framework of discontinuous deformation analysis [J]. *Geomechanics & Engineering*, 2017,12(4):723 - 738.

[2] Shanley R J, Mahtab M A. Delineation and analysis of clusters in orientation data[J]. *Mathematical Geology*, 1976, 8(1):9 - 23.

[3] Harrison J P. Fuzzy objective functions applied to the analysis of discontinuity orientation data[C]//Rock Characterization: ISRM Symposium. London: Thomas Telford, 1992: 25 - 30.

[4] Hammah R E, Curran J H. Fuzzy cluster algorithm for the automatic identification of joint sets [J]. *International Journal of Rock Mechanics & Mining Sciences*, 1998, 35 (7):889 - 905.

[5] 李宁,王李管,贾明涛,等.改进遗传算法和支持向量机的岩体结构面聚类分析[J]. *岩土力学*, 2014, 35 (sup2): 405 - 411.

(Li Ning, Wang Li-guan, Jia Ming-tao, et al. Application of improved genetic algorithm and support vector machine to clustering analysis of rock mass structural plane[J]. *Rock and Soil Mechanics*, 2014, 35 (sup2): 405 - 411.)

[6] 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.

[7] 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.)

[8] 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.

[9] Wang X, Sun Q. The study of K-means based on hybrid SA-PSO algorithm [C]//International Symposium on Computational Intelligence and Design (ISCID). Los Alamitos: IEEE, 2016: 211 - 214.

[10] 秦胜伍,陈骏骏,陈剑平,等.基于粗糙集理论的岩体结构面模糊 C 均值聚类分析[J]. *中南大学学报(自然科学版)*, 2016, 47(9): 3125 - 3130.

(Qin Sheng-wu, Chen Jun-jun, Chen Jian-ping, et al. Fuzzy C-means cluster analysis based on rough set for grouping of discontinuities [J]. *Journal of Central South University (Science and Technology)*, 2016, 47 (9): 3125 - 3130.)

[11] 王俊杰,冯登,柴贺军,等.基于赤平极射投影和 K - 均值聚类算法的优势结构面分析[J]. *岩土工程学报*, 2018, 40 (1): 74 - 81.

(Wang Jun-jie, Feng Deng, Chai He-jun, et al. Dominant discontinuities based on stereographic projection and K-means clustering algorithm [J]. *Chinese Journal of Geotechnical Engineering*, 2018, 40 (1): 74 - 81.)

[12] Selim S Z, Alsultan K. A simulated annealing algorithm for the clustering problem [J]. *Pattern Recognition*, 1991, 24 (10): 1003 - 1008.

[13] Nguyen L T, Nestorovic T. Unscented hybrid simulated annealing for fast inversion of tunnel seismic waves[J]. *Computer Methods in Applied Mechanics and Engineering*, 2016, 301: 281 - 299.