

采用混合策略联合优化的模糊C-均值聚类信息熵点云简化算法

黄鹤^{1,2}, 黄佳慧^{1,2}, 刘国权¹, 王会峰², 高涛³

(1. 长安大学西安市智慧高速公路信息融合与控制重点实验室, 710064, 西安;

2. 长安大学电子与控制工程学院, 710064, 西安; 3. 长安大学数据科学与人工智能研究院, 710064, 西安)

摘要: 针对传统聚类算法处理点云简化问题时精度低、耗时长且易丢失特征信息等问题,提出了一种基于动态精英自适应混合策略的鸚鵡算法(DEAMPOA)与加权熵法联合优化的模糊C-均值聚类(FCM)信息熵点云简化算法。采用动态自适应种群混合策略,同时融合了精英反向化思路,显著提升了鸚鵡优化算法(POA)的收敛趋势和全局寻优能力,提高了寻找FCM最优聚类中心的成功率;利用DEAMPOA结合加权熵法对FCM进行优化,提高鲁棒性的同时增强了搜索精度,得到较好的聚类结果;在8种UCI标准数据集上与4种算法对比进行聚类性能评估实验,验证了所提方法综合性能优越;将所提方法与信息熵融合,并应用在三维点云KITTI数据集简化中。实验结果表明:与包围框简化法、随机采样简化法和特征选择简化法对比,所提方法全局误差简化前后点集之间平均欧式距离(MED)指标分别降低了2.25%、6.93%、5.74%,点云简化效果最优且运行速度满足要求。

关键词: C-均值聚类;鸚鵡优化算法;点云简化;信息熵

中图分类号: U666.1 **文献标志码:** A

DOI: 10.7652/xjtuxb202407020 **文章编号:** 0253-987X(2024)07-0214-13

Fuzzy C-means Clustering Information Entropy Point Cloud Simplification Using Mixed Strategy Joint Optimization

HUANG He^{1,2}, HUANG Jiahui^{1,2}, LIU Guoquan¹, Wang Huifeng², GAO Tao³

(1. Xi'an Key Laboratory of Intelligent Expressway Information Fusion and Control, Chang'an University, Xi'an 710064,

China; 2. School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China;

3. Institute of Data Science and Artificial Intelligence, Chang'an University, Xi'an 710064, China)

Abstract: To solve the problems of low accuracy, long time consumption, and easy loss of feature information of the traditional clustering algorithm in conducting point cloud simplification, a point cloud simplification method for FCM information entropy jointly optimized by DEAMPOA and weighted entropy method is proposed. Firstly, a dynamic adaptive population mixing strategy is proposed, which integrates the elite reverse idea. This strategy significantly improves the convergence trend and global optimization ability of POA and increases the success rate of finding the optimal clustering center of FCM. Secondly, the optimization of FCM using DEAMPOA combined with weighted entropy method improves the robustness while enhancing

收稿日期: 2023-10-05。 作者简介: 黄鹤(1979—),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(62341301);西安市智慧高速公路信息融合与控制重点实验室(长安大学)开放基金资助项目(300102323502);中央高校基本科研业务费资助项目(300102324501)。

网络出版时间: 2024-04-16

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20240415.1134.006>

the search accuracy, yielding better clustering results. Thirdly, clustering performance evaluation experiments are carried out through comparison with four comparison algorithms in eight UCI standard datasets, which verifies that the proposed method has superior comprehensive performance. Finally, the proposed method is fused with information entropy and applied to the KITTI point cloud dataset simplification. The experimental results show that the global error MED index of the proposed method is reduced by 2.25%, 6.93%, and 5.74% respectively compared with that of the bounding frame simplification method, random sampling simplification method, and feature selection simplification method. Furthermore, this method generates the optimal point cloud simplification effect and the running speed meets the requirements.

Keywords: fuzzy C-means; pelican optimization algorithm; point cloud simplification; information entropy

随着数字化激光扫描技术的不断发展,三维点云数据在自动驾驶领域的应用愈发广泛,但激光扫描采集到的点云数据往往受光照、设备或人为原因影响,包含大量冗余点以及噪声点。因此,在保证原始纹理特征的前提下,对点云进行去噪和简化非常必要。近年来,国内外学者对三维点云简化算法的研究主要分为两类:基于深度学习^[1]和基于聚类^[2]。基于深度学习的简化算法通过输入大量的点云数据进行训练从而学习点云的高级特征,能够捕捉点云的语义信息和结构信息,但是训练过程计算复杂度高,需要大量的标注数据。聚类是一种无监督的机器学习算法,因其具有算法简单、可控性强与计算效率高特点,已成为点云数据简化处理的关键算法之一。

点云聚类方法主要有 K 均值聚类^[3] (KMC)、基于密度的聚类^[4] (DBSCAN) 和模糊 C -均值聚类 (FCM)^[5] 等, KMC 简单但对初始聚类中心依赖性过高^[6]; DBSCAN 按点的密度大小进行聚类,耗时长、计算量大^[7]; FCM 引入了迭代优化,处理效率高且稳定性强,但在点云简化时存在对初始聚类中心敏感、易陷入局部最优等问题,需要进一步优化。Yang 等^[8] 提出一种基于特征保留改进 FCM 聚类的点云精简算法,通过引入引力搜索算法优化 FCM,根据不同特征区域的简化参数与百分比进行简化,简化率较高但特征丢失严重。张佳琦等^[9] 提出了一种基于曲率约束的改进狼群算法优化模糊 C -均值聚类的混合算法,利用改进狼群算法得到的聚类中心作为模糊聚类的初始值迭代,并基于点云的法矢量和曲率定义点云之间距离,完成点云数据简化,但可能导致点云模型部分边界点丢失。以上算法针对 FCM 在点云简化方面的不足做了一些改进,但简化效率和简化精度无法同时兼顾,存在细节特征保留差、易出现关键信息丢失等问题,仍有提升空间。鹈鹕优化算法 (POA) 是 2022 年由 Pavel

Trojovský 提出的新型元启发式群体智能优化算法^[10], 具有搜索精度高、适应性强等特点, 为利用 FCM 聚类进行点云简化提供了新思路。本文提出了一种基于动态精英自适应混合策略的鹈鹕算法 (DEAMPOA) 与加权熵法联合优化的 FCM 聚类算法, 并与信息熵结合, 构成三维点云简化模型, 利用 POA 优化算法的寻优能力提升聚类效率, 同时利用信息熵的引入保留轮廓特征和细节特征, 有效提高了点云简化精度和效率。

1 FCM 数学模型

1.1 聚类原理

样本集 $X = \{x_1, x_2, \dots, x_n\}$, 每个样本 x_l 可以包含多个特征, 即 $x_l = \{a_1, a_2, \dots, a_k\}$, 分类数为 c , FCM 聚类目标函数计算公式为

$$\left\{ \begin{aligned} J(\mathbf{X}, \mathbf{U}, \mathbf{V}) &= \sum_{m=1}^c \sum_{l=1}^n u_{lm}^\sigma d_{lm}^2 \\ d_{lm} &= \|x_l - v_m\| \\ \text{s. t. } \begin{cases} 0 \leq u_{lm} \leq 1 \\ \sum_{m=1}^c u_{lm} = 1 \\ c \leq n \end{cases} \end{aligned} \right. \quad (1)$$

式中: $\mathbf{U}$ 为各样本属于不同类别的隶属度矩阵; $\mathbf{V}$ 为所有聚类中心构成的矩阵; n 为样本总数; u_{lm} 为样本 x_l 属于类别 m 的隶属度; σ 为加权指数, $\sigma \geq 1$ 。 d_{lm} 为样本 x_l 与聚类中心 v_m 的欧几里得距离。

1.2 FCM 算法存在的不足

首先, FCM 算法中, c 需要根据经验随机给定^[11], 随机性较大, 影响聚类结果, 故设计聚类有效性函数来确定 c

$$G(\mathbf{U}, c) = \max_{g=1}^c (\max_{m=1, m \neq g}^c G_{gm}(\mathbf{U}, c)) \quad (2)$$

式中: G_{gm} 是第 g 类与第 m 类的模糊相关度, 表达式如下

$$G_{gm} =$$

$$\frac{\sum_{i=1}^n u_{ig}^{\sigma/2} u_{im}^{\sigma/2} \|x_i - v_g\| \|x_i - v_m\|}{(\sum_{i=1}^n u_{ig}^{\sigma} \|x_i - v_g\|^2)^{1/2} (\sum_{i=1}^n u_{im}^{\sigma} \|x_i - v_m\|^2)^{1/2}} \quad (3)$$

其中 $\|x_i - v_g\|$ 和 $\|x_i - v_m\|$ 分别为样本点 x_i 与聚类中心 v_g, v_m 的距离。若想得到尽可能明显的分类效果, 要求不同类别之间的模糊相关度尽可能小。令 Ω_c 表示 $\mathbf{U}$ 的最优有限集合, 则最优有效聚类 $(\mathbf{U}^*, c^*)$ 应满足下式

$$G(\mathbf{U}^*, c^*) = \min_c \{ \min_c G(\mathbf{U}, c) \} \quad (4)$$

式中: c^* 为最佳分类数。

此外, FCM 还存在如下问题: ①未考虑样本和指标对聚类结果的差异影响^[12], 导致聚类结果不理想; ②若初始隶属度矩阵靠近局部最优值, 算法会陷入局部最小^[13], 无法跳出则导致提前收敛, 因此需要进一步对 FCM 进行优化。

2 POA 算法

2.1 POA 算法流程

2.1.1 初始化

$$X_{i,j} = l_j + r(u_j - l_j)$$

$$i = 1, 2, \dots, N, j = 1, 2, \dots, w \quad (5)$$

式中: $X_{i,j}$ 为第 i 个鸬鹚个体第 j 维位置; N 为种群中个体数; w 为求解问题维度; r 是 $[0, 1]$ 范围内的随机数, u_j 和 l_j 是搜索空间的上、下边界。

2.1.2 逼近猎物

本阶段中鸬鹚确定猎物位置, 然后向此区域逼近。数学建模如下

$$X_{i,j}^{P_1} = \begin{cases} X_{i,j} + r(P_j - IX_{i,j}), & F_p < F_i \\ X_{i,j} + r(X_{i,j} - P_j), & F_p \geq F_i \end{cases} \quad (6)$$

式中: $X_{i,j}^{P_1}$ 为第 1 阶段更新后鸬鹚的位置; I 为随机整数 1 或 2; P_j 为猎物位置, 在搜索空间中随机生成; F_p 为猎物目标函数值; F_i 为第 i 个鸬鹚的目标函数值。

2.1.3 水面飞行

本阶段鸬鹚到达水面后展开翅膀, 令猎物向上移动, 然后将猎物放在喉咙袋里进行捕获。数学建模如下

$$X_{i,j}^{P_2} = X_{i,j} + R \left(1 - \frac{\tau}{T} \right) (2r - 1) X_{i,j} \quad (7)$$

式中: $X_{i,j}^{P_2}$ 为基于第 2 阶段更新后的位置; R 为常数, 取为 0.2; τ 为当前迭代次数; T 为最大迭代次数。

2.2 基于动态精英自适应混合策略的 POA 优化策略

鸬鹚能快速适应不同生态环境并有效捕食猎物。因此, 为使算法能收敛到狩猎区域更好的位置, 应将 POA 搜索方式设计为动态的, 最大限度发挥 POA 在搜索空间中不同区域的勘探能力, 有利于后续寻找最优 FCM 聚类中心。同时, POA 中鸬鹚位置更新具有向原点收敛的趋势, 这增加了求解非原点问题的局限性。因此, 设计动态精英自适应种群博弈策略对 POA 算法作进一步改进。

2.2.1 初始精英反向化

鸬鹚在捕食猎物时, 通过更新式(5)初始化种群位置, 因上、下界的限制, 其可选界范围有限, 故设计精英反向策略进行修正。透过透镜成像反向学习法计算反向解, 由此扩大可选解的范围, 增加选取更优解的概率, 从而实现对初始鸬鹚个体质量的提升。透镜成像反向后生成反向种群 X^* , 鸬鹚个体位置为

$$X_{i,j}^* = \frac{S_{\max,j} + S_{\min,j}}{2} + \frac{S_{\max,j} + S_{\min,j}}{2\lambda} - \frac{X_{i,j}}{\lambda} \quad (8)$$

式中: $S_{\max,j}, S_{\min,j}$ 分别为种群内所有个体中第 j 维的最大值和最小值; λ 是透镜的缩放系数, 经过大量的实验测试, 在 $\lambda = 8\,000$ 时种群产生的反向解最优, 故本文设定 λ 为 8 000。种群 X^* 和 X 内鸬鹚个体根据适应度排序, 选择前 N 个较优个体, 构成新初始精英种群 X 来增强 POA 的勘探能力。

2.2.2 动态自适应种群博弈策略

设计动态自适应种群博弈策略旨在通过动态调整种群结构, 提高算法搜索能力并避免过早收敛。在寻优过程中引入多个采用不同搜索策略的子种群。在迭代过程中, 子种群之间依据适应度进行博弈, 动态调整其规模, 并令优秀个体在种群之间进行迁移, 实现全局和局部搜索动态平衡, 提高 POA 的寻优能力。最后, 通过 POA 与 FCM 的交叉迭代优化聚类算法, 进一步提升效率与精度, 具体步骤如下。

步骤 1 初始化。生成 3 个种群, 定义为 S_1, S_2 和 S_3 , 每个子种群有相同初始数量规模, 设定最大迭代次数 T 和适应度评价函数 f 。

步骤 2 生成初始解。对于每个子种群, 通过精英反向策略生成 N 个初始解。

步骤 3 分配搜索策略。为每个子种群分配不同的搜索策略。 S_1 采用具有较高全局搜索能力的原始

POA 搜索策略; S_2 在 POA 基础上引入了变异操作,使其更注重局部搜索,增强局部搜索能力; P_3 使用麻雀算法(SSA)搜索策略,在全局搜索和局部搜索之间取得平衡。 S_2 中个体执行变异操作后的位置 x'_i 由下式计算

$$x'_i(t+1) = x_i(t+1) + \alpha(u_j - l_j)r \tag{9}$$

式中: α 为变异强度参数。

步骤 4 种群演化。计算适应度值,并根据各自的搜索策略完成迭代演化。

步骤 5 计算子种群的贡献值。根据种群适应度值计算子种群的贡献值,通过种群博弈按照比例动态调整各个种群数。一般而言,个体适应度值越高,相应贡献值越大,将所有个体贡献值累加得到子种群的最终贡献值。个体适应度对应的贡献值定义如下

$$F_{i,q} = \frac{f_{i,q}}{\sigma_q} e^{-\frac{d_{i,q}}{\Delta}} \tag{10}$$

式中: $F_{i,q}$ 表示第 i 个鸚鵡个体对第 q 个子种群的适应度贡献值; $f_{i,q}$ 表示个体 i 在第 q 个子种群中的适应度值; σ_q 表示第 q 个子种群中适应度值的标准差; $d_{i,q}$ 表示第 i 个鸚鵡个体和第 q 个子种群局部最优解之间的欧几里得距离; Δ 表示缩放因子。根据式(10)计算子种群每个个体的适应度贡献值,然后计算所有个体的适应度贡献值之和,记为 $f_{t,q}$ 。对于每个子种群,计算所有个体的相对适应度值之和

$$f_{r,t,q} = \sum_{i=1}^N f_{r,i} \tag{11}$$

式中: $f_{r,i}$ 表示个体的相对适应度贡献,计算公式如下

$$f_{r,i} = \frac{f_{i,q}}{f_{t,q}} \tag{12}$$

最后,计算各个子种群对整个种群的贡献值

$$C_q = \frac{f_{r,t,q}}{f_{t,t}} \tag{13}$$

式中: $f_{t,t}$ 表示 3 个子种群所有个体的适应度贡献值之和。

步骤 6 动态自适应种群。依据上述博弈过程得到

的 C_q 动态调整规模。
步骤 7 判断是否符合结束条件,如果不符合则循环步骤 4~6;如果符合则跳出循环,输出全局最优解。

DEAMPOA 算法的具体流程如图 1 所示。

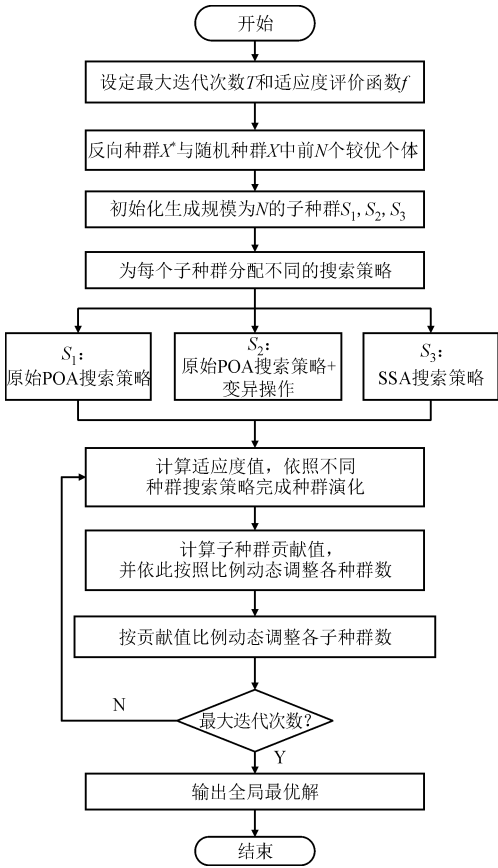


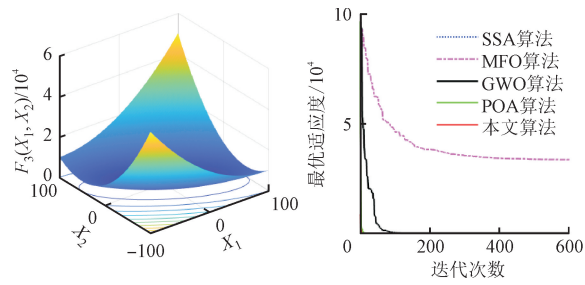
图 1 动态精英自适应 POA 算法流程图
Fig. 1 Dynamic elite adaptive POA algorithm flow chart

2.3 优化策略的测试

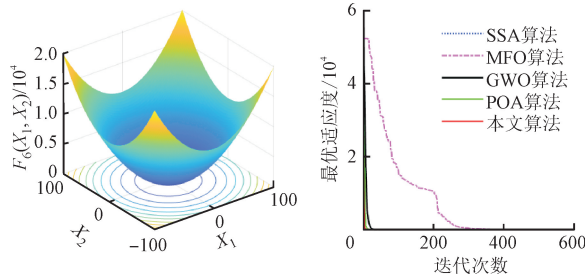
为了验证改进算法性能,将本文算法 DEAMPOA 与 POA、GWO^[14]、MFO^[15]和 SSA^[16]算法在 5 种基准测试函数上进行比较。设置各群智能算法种群规模为 50,迭代次数为 600。测试中,每种算法运行 30 次,得到的适应度均值与标准差见表 1,测试函数空间表示以及适应度收敛曲线如图 2 所示。

表 1 测试函数结果
Table 1 Test results of the function

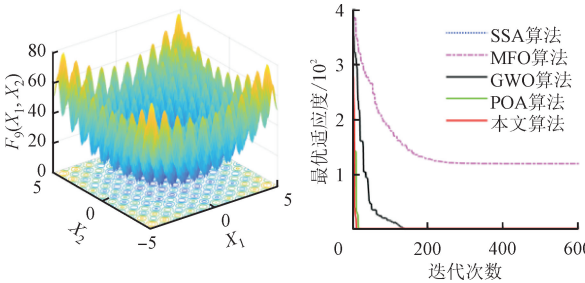
测试函数	适应度均值					适应度标准差				
	本文算法	POA 算法	SSA 算法	MFO 算法	GWO 算法	本文算法	POA 算法	SSA 算法	MFO 算法	GWO 算法
F_3	5.42×10^{-116}	2.76×10^{-68}	0×10^0	3.34×10^2	1.10×10^{-9}	0×10^0	0×10^0	0×10^0	2.84×10^2	4.60×10^{-9}
F_6	0.53×10^{-7}	0.48×10^{-7}	2.82×10^{-7}	1.53×10^{-3}	0.69×10^0	2.63×10^{-7}	0.17×10^{-6}	3.45×10^{-7}	1.13×10^{-3}	1.46×10^{-2}
F_9	0×10^0	4.23×10^{-9}	4.72×10^{-10}	1.18×10^2	5.67×10^{-9}	0×10^0	5.34×10^{-9}	7.94×10^{-11}	3.28×10^2	9.84×10^{-9}
F_{11}	3.94×10^{-9}	8.92×10^{-9}	5.27×10^{-9}	9.12×10^{-1}	1.69×10^{-2}	2.36×10^{-9}	5.36×10^{-9}	3.68×10^{-9}	6.35×10^{-1}	5.18×10^{-2}
F_{20}	2.17×10^{-1}	2.76×10^{-1}	3.32×10^0	3.20×10^0	3.21×10^0	4.02×10^{-1}	6.27×10^{-1}	4.09×10^0	3.61×10^0	5.17×10^0



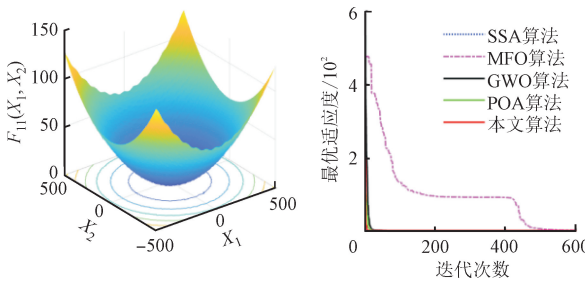
(a) Schwefel 1.2



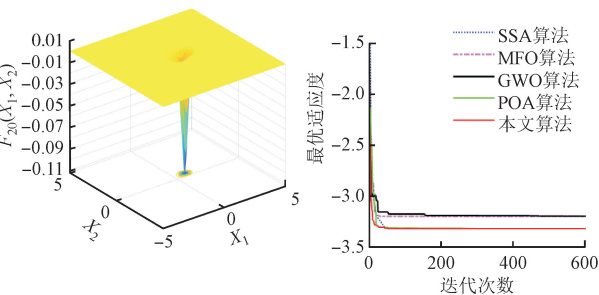
(b) Step



(c) Rastrigin



(d) Griewank



(e) Hartman

图 2 测试函数空间表示以及适应度收敛曲线

Fig. 2 Convergence curves of test functions

由表 1 和图 2 可以看出,与其他算法相比,DEAMPOA 收敛速度更快、效率更高,总体性能较

优。设计精英反向策略和动态自适应混合种群策略显著改善了 POA 的收敛趋势和全局寻优能力,实现了算法在全局搜索和局部搜索之间的动态平衡。

3 POA 优化更新路径的改进 FCM 算法

3.1 基于加权熵法优化的改进 FCM 算法

针对样本存在影响差异问题,设计加权熵法来优化 FCM 算法。采用熵权法确定不同维度特征的权重,并通过设置样本加权的方法定义各个样本对聚类结果的影响程度。以三维点云图像为例进行分析,给定三维 n 个样本的点云数据集如下

$$H = \begin{bmatrix} h_{1x} & h_{1y} & h_{1z} \\ h_{2x} & h_{2y} & h_{2z} \\ \vdots & \vdots & \vdots \\ h_{nx} & h_{ny} & h_{nz} \end{bmatrix} \quad (14)$$

熵权法的计算步骤如下。

步骤 1 计算 x 轴维度下第 l 个样本所占的比

$$p_{lx} = \frac{h_{lx}}{\sum_{l=1}^n h_{lx}} \quad (15)$$

步骤 2 计算 x 轴维度的信息熵值

$$E_x = -\frac{1}{\ln n} \sum_{l=1}^n p_{lx} \ln p_{lx} \quad (16)$$

步骤 3 重复步骤 1~2,计算 y, z 维度信息熵值。

步骤 4 计算各个维度特征的权重(以 x 维度为例)

$$w_x = \frac{1 - E_x}{(1 - E_x)(1 - E_y)(1 - E_z)} \quad (17)$$

各样本的权重计算如下

$$e_l = \exp\left(-\sum_{m=1}^c u_{lm} d_{lm}^w\right) \quad (18)$$

$$d_{lm}^w = w_x w_y w_z \|x_l - v_m\| \quad (19)$$

式中: e_l 是样本 l 的权重; d_{lm} 是样本 l 与第 m 个聚类中心的加权欧式距离。

综上,改进 FCM 的目标函数为

$$J(\mathbf{X}, \mathbf{U}, \mathbf{V}) = \sum_{m=1}^c \sum_{l=1}^n e_l u_{lm}^\sigma (d_{lm}^w)^2 \quad (20)$$

3.2 引入动态精英自适应 POA 的改进 FCM 算法

针对 FCM 算法对初值敏感和存在局部极小问题,利用 DEAMPOA 的全局寻优能力平衡 FCM 算法的局部最优解与聚类速度、精度之间的矛盾,克服 FCM 算法的单一性。

3.2.1 适应度函数设计

采用加权模糊均值聚类目标函数 J_P 作为适应度函数

$$J_P = \sum_{m=1}^c \sum_{l=1}^n e_l u_{lm}^\sigma (d_{lm}^w)^2 \quad (21)$$

表 3 各聚类算法对标准数据集的 S_{DBI} 测试结果

Table 3 S_{DBI} index test results of each clustering algorithm on the standard dataset

数据集	S_{DBI}				
	本文算法	POA-FCM 算法	multi-core FCM 算法	EWFCM 算法	FCM 算法
Bupa	0.485 5±4.6%	0.704 9±5.0%	0.959 4±7.7%	0.911 5±9.9%	0.961 7±10%
WBC	0.609 1±1.7%	0.703 0±3.1%	0.704 8±4.2%	0.754 8±3.2%	0.796 0±4.6%
Diabetes	0.419 5±0.1%	0.682 9±0.2%	0.703 1±0.6%	0.523 2±0.5%	0.758 8±0.6%
Wine	0.271 6±1.3%	0.482 7±4.9%	0.516 3±4.1%	0.841 7±4.7%	0.989 3±4.4%
Iris	0.396 0±1.1. %	0.685 7±1.6%	0.729 8±2.2%	0.701 8±1.4%	0.751 5±3.0%
Vowel	0.635 9±2.3%	0.810 9±3.0%	0.859 5±3.7%	0.691 5±3.3%	0.949 4±3.1%
Aggregation	0.517 9±3.0%	0.681 9±6.0%	0.726 4±5.7%	0.714 7±8.9%	0.972 9±10.3%
Ecoli	0.594 9±3.5%	0.817 4±5.7%	1.071 8±5.0%	1.184 5±6.3%	1.225 0±6.3%

表 4 各聚类算法对标准数据集的 T_{ARI} 测试结果

Table 4 T_{ARI} index test results of each clustering algorithm on the standard dataset

数据集	T_{ARI}				
	本文算法	POA-FCM 算法	multi-core FCM 算法	EWFCM 算法	FCM 算法
Bupa	0.867 1±0.1%	0.714 5±0.2%	0.634 6±0.4%	0.541 9±0.6%	0.011 4±0.9%
WBC	0.983 0±0.2%	0.504 9±3.1%	0.881 6±3.4%	0.618 4±3.3%	0.830 0±4.1%
Diabetes	0.871 4±0.1%	0.409 1±0.4%	0.510 2±0.5%	0.597 6±0.5%	0.080 4±0.6%
Wine	0.947 1±0.1%	0.279 0±2.3%	0.617 9±6.9%	0.351 7±5.5%	0.353 9±4.8%
Iris	0.892 9±0.1%	0.390 1±2.0%	0.708 4±1.8%	0.613 5±2.1%	0.420 6±2.5%
Vowel	0.800 1±3.8%	0.456 6±4.0%	0.487 1±4.6%	0.487 2±7.0%	0.315 3±7.1%
Aggregation	0.926 0±2.4%	0.884 9±6.2%	0.412 6±7.4%	0.651 9±11.0%	0.743 3±11.9%
Ecoli	0.784 6±3.3%	0.607 8±5.5%	0.587 5±6.4%	0.432 8±7.0%	0.364 3±7.5%

表 5 各聚类算法对标准数据集的 B_{NMI} 测试结果

Table 5 B_{NMI} index test results of each clustering algorithm on the standard dataset

数据集	B_{NMI}				
	本文算法	POA-FCM 算法	multi-core FCM 算法	EWFCM 算法	FCM 算法
Bupa	0.796 1±0.1%	0.4903±0.2%	0.2076±0.4%	0.3198±0.5%	0.0070±1.1%
WBC	0.9901±1.3%	0.8306±1.9%	0.5599±2.7%	0.3209±2.6%	0.7300±3.3%
Diabetes	0.8130±0.1%	0.6612±0.2%	0.49444±0.2%	0.7016±0.2%	0.2337±0.3%
Wine	0.8844±1.2%	0.6015±3.7%	0.7913±5.4%	0.4497±5.0%	0.4168±4.4%
Iris	0.9579±0.2%	0.7088±4.1%	0.6112±2.2%	0.5690±1.2%	0.2931±3.0%
Vowel	0.9146±3.0%	0.8046±3.3%	0.6898±5.8%	0.4901±4.1%	0.4717±8.5%
Aggregation	0.9989±2.8%	0.8947±3.9%	0.6991±7.5%	0.8029±10.3%	0.8438±11.4%
Ecoli	0.9063±3.0%	0.8846±4.5%	0.7394±6.0%	0.6273±6.7%	0.5670±7.8%

表 6 各聚类算法对标准数据集的 L_{ACC} 测试结果

Table 6 L_{ACC} index test results of each clustering algorithm on the standard dataset

数据集	L_{ACC}				
	本文算法	POA-FCM 算法	multi-core FCM 算法	EWFCM 算法	FCM 算法
Bupa	0.946 9±0.1%	0.756 1±0.1%	0.709 2±1.0%	0.604 9±1.6%	0.379 7±1.5%
WBC	0.906 7±1.2%	0.799 1±1.3%	0.701 4±2.3%	0.769 4±1.5%	0.556 1±2.8%
Diabetes	0.989 8±0.1%	0.746 6±0.2%	0.849 6±0.2%	0.694 1±0.3%	0.658 9±0.6%
Wine	0.917 9±2.7%	0.804 9±4.9%	0.569 4±4.2%	0.552 3±4.5%	0.385 4±5.0%
Iris	0.967 8±1.5%	0.800 8±6.2%	0.926 8±4.5%	0.784 6±2.3%	0.693 3±3.6%
Vowel	0.999 9±1.2%	0.881 5±2.0%	0.874 4±4.1%	0.865 9±3.0%	0.464 9±5.1%
Aggregation	0.999 9±2.6%	0.884 5±4.4%	0.806 9±6.9%	0.294 7±7.0%	0.731 5±9.0%
Ecoli	0.911 3±2.2%	0.869 5±3.4%	0.741 2±4.5%	0.804 6±5.2%	0.706 5±6.2%

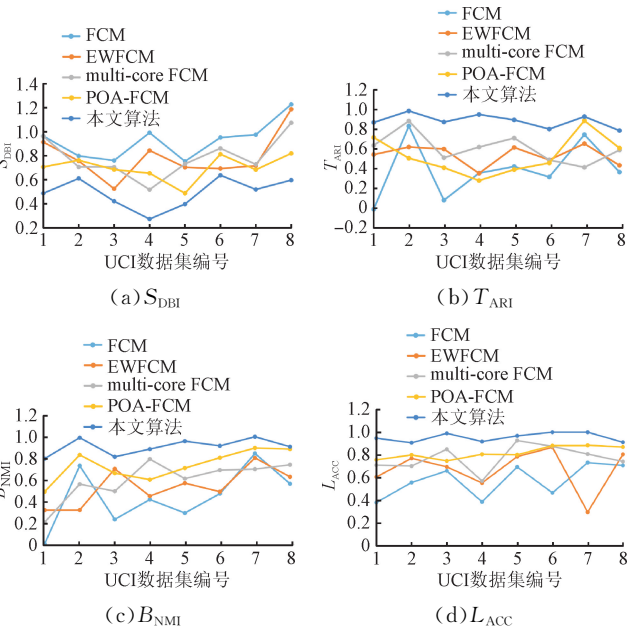


图 4 各算法在各数据集上的聚类评价指标可视化结果
Fig. 4 Visualization results of clustering evaluation indicators of each algorithm on each dataset

由表 3~表 6 和图 4 可以看出,对于 T_{ARI} ,算法结果按从优到劣排序是本文算法、multi-core FCM、POA-FCM、EWFCM、FCM, B_{NMI} 和 L_{ACC} 指标亦如此。同时,本文算法的 S_{DBI} 指标最低,说明使用该算法进行数据分类的适确性最佳。原始 FCM 算法由于初始化隶属度矩阵时随机性过大,分类的结果受其影响,各评价指标均最差; multi-core FCM 和 EWFCM 由于分别结合多个核函数和引入熵权法优化原始算法的迭代搜索过程,克服了随机初始化带来的影响,一定程度上增加了准确率 L_{ACC} ,其中 multi-core FCM 结合多个核函数对数据的多种特征进行同步描述,并且自适应地调整核函数权重对

多个核函数进行凸组合,找到最佳的隶属度以及核之间的最优结合方式,可以提高计算效率,但是易受噪声点和异常点的干扰,而熵权法在针对大数据集时计算复杂度较高,造成两个算法的结果均不够稳定,因此 S_{DBI} 、 T_{ARI} 和 B_{NMI} 指标均较差; POA-FCM 算法由于混合了群智能算法进行聚类交叉迭代,增强了聚类的适确性和鲁棒性, S_{DBI} 和 B_{NMI} 指标优于 multi-core FCM 和 EWFCM 算法的,但 POA 算法的寻优能力有限,易陷入局部最优,因此 S_{DBI} 和 B_{NMI} 以及 L_{ACC} 指标均比本文算法差, T_{ARI} 指标的差异更大。此外,通过各指标标准差可知,本文算法相较于其他 4 种对比算法,各指标在应用于各标准数据集时的标准差均最小,算法稳定性最高。

以上各指标对比结果证明本文算法比其他算法具有更强的寻优能力,通过多个评价指标的综合对比可知,本文算法在对标准数据集进行分类时,鲁棒性更好,精度也更高,综合表现能力最佳。

4 基于改进 FCM 聚类融合信息熵的点云精简算法

4.1 融合算法

点云数据受周遭物体反射、光照和采集装备自身固有噪声等因素影响,包含大量冗余点、重复点以及噪声点,因此结合信息熵,利用 DEAMPOA 与熵权法联合优化的 FCM 算法(简称改进的 FCM 算法),对 KITTI 数据集的点云数据进行去噪、简化,从而实现对数据集应用于自动驾驶场景时的辅助效果。具体实现步骤如下。

步骤 1 获取交通场景下的非均匀分布三维点云数据集,依据式(4)确定聚类的类别数 c ,利用改进的 FCM 算法将点云划分为若干个簇。

步骤2 分别计算点云簇中每个点的贡献信息熵,并累加得到整个点云簇的信息熵,根据所得熵对聚类簇进行降序排列。其中,信息熵表示点分布的不确定度,计算公式如下

$$H(X) = - \sum_{l=1}^n P(X = x_l) \lg P(X = x_l) \quad (22)$$

式中: $P(X = x_l)$ 表示 x_l 的密度估计值,采用基于高斯核函数的 KNN 密度估计得到, d 维 KNN 估计的定义^[23]如下

$$P_{\text{KNN}}(x) = N^{-1} R_k(x)^{-d} \sum_{l=1}^n K(R_k(x)^{-1}(x - x_l)) \quad (23)$$

式中: $R_k(x)$ 是样本对象 x 与第 k 个近邻点间距离,样本对象 x 与其余点之间的距离关系为 $R_1(x) < \dots$

$< R_{k-1}(x) < R_k(x) < \dots < R_N(x)$, $K(u)$ 为高斯核函数^[24], 定义如下

$$K(u) = (2\pi)^{-(d/2)} \exp\left(-\frac{1}{2} u^T u\right) \quad (24)$$

将式(24)回代至式(23)得到

$$P_{\text{KNN}}(x) = N^{-1} C_d^{-1} R_k(x)^{-1} \sum_{l=1}^n \exp\left(-\frac{1}{2} (R_k(x)^{-1}(x - x_l))^2\right) \quad (25)$$

式中: C_d 为 d 维空间中单位球体的体积。对于孤立的噪声点云而言,计算得到其信息熵为 0。

步骤3 将信息熵低于阈值的簇删除,保留点云的轮廓和细节特征。本文改进的点云简化算法流程图如图 5 所示。

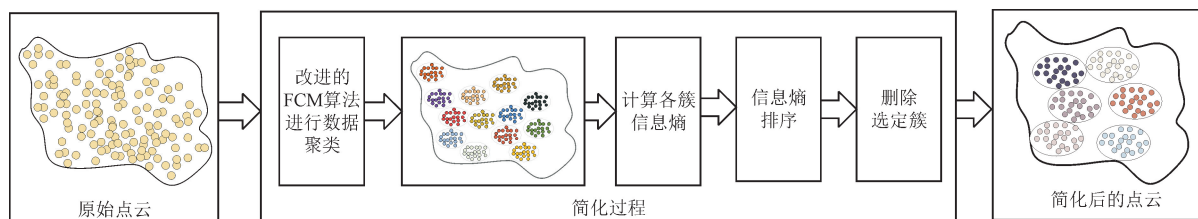


图5 改进的点云简化算法流程

Fig. 5 Flow chart of the improved point cloud simplification algorithm

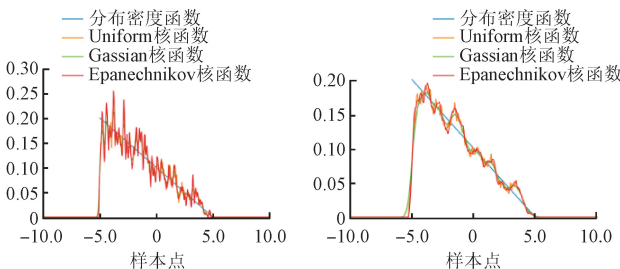
4.2 实验结果与分析

4.2.1 不同带宽下核函数对比实验

为了平滑拟合点云数据的概率密度函数,使用基于 Gaussian 函数的 KNN 算法对其进行密度估计,在不同带宽 h 时与 Uniform 核函数和 Epanechnikov 核函数进行拟合实验,分别取 h 为 0.1、0.3、0.7、1.4,各个核函数拟合结果如下。

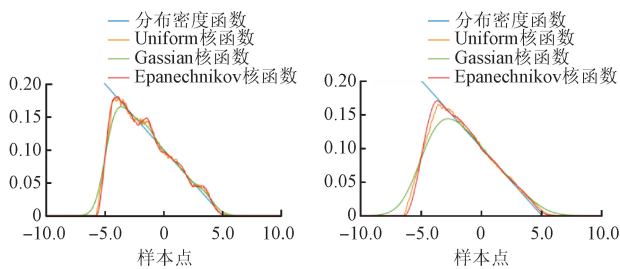
由图 6 可知,在 $h=0.1$ 时,核函数拟合曲线波动剧烈,细节处理过于粗糙; $h=0.3$ 时,曲线抖动减少,但包含大量噪声点,无法拟合真实分布曲线;随着带宽 h 的增加,各条核函数拟合曲线与概率密度曲线之间的方差都增大,因此 $h=1.4$ 时,虽然拟合曲线平滑线得到保障,方差却显著增大,对真实曲线进行拟合时,细节点丢失严重,拟合效果不佳;当 $h=0.7$ 时,核函数拟合曲线在平滑性与准确性之间取得了良好的平衡。与此同时, Gaussian 核函数相比 Uniform 核函数和 Epanechnikov 核函数具有更好的平滑性,拟合更加接近真实概率分布曲线。考虑到点云中的局部结构通常是连续的,即相邻的点之间存在较小的距离差异, Gaussian 核函数的平滑性可以很好地适应这种连续的局部结构,因此选择

在带宽 $h=0.7$ 时结合 Gaussian 核函数对点云数据进行密度估计与分析。



(a) $h=0.1$

(b) $h=0.3$



(c) $h=0.7$

(d) $h=1.4$

图6 不同带宽时各个核函数拟合结果

Fig. 6 Fitting results of each kernel function under different bandwidths

4.2.2 三维点云简化性能分析实验

将本文提出的点云简化算法与特征选择简化法^[20]、包围框简化法^[21]和随机采样简化法^[22]在三维点云 KITTI 数据集上进行定性分析与定量分析,特征点云法需要提前设定简化比例,包围框法和随机采样法需

要设定简化后点云数^[25],故将简化后的点云数设置为原始点云数的 80%,以保证对比实验的严谨性。

图 7、8 分别为本文算法和对比模型对住宅区环境、校园环境、道路环境和市区环境下采集到的点云数据进行简化处理后得到的可视化结果。

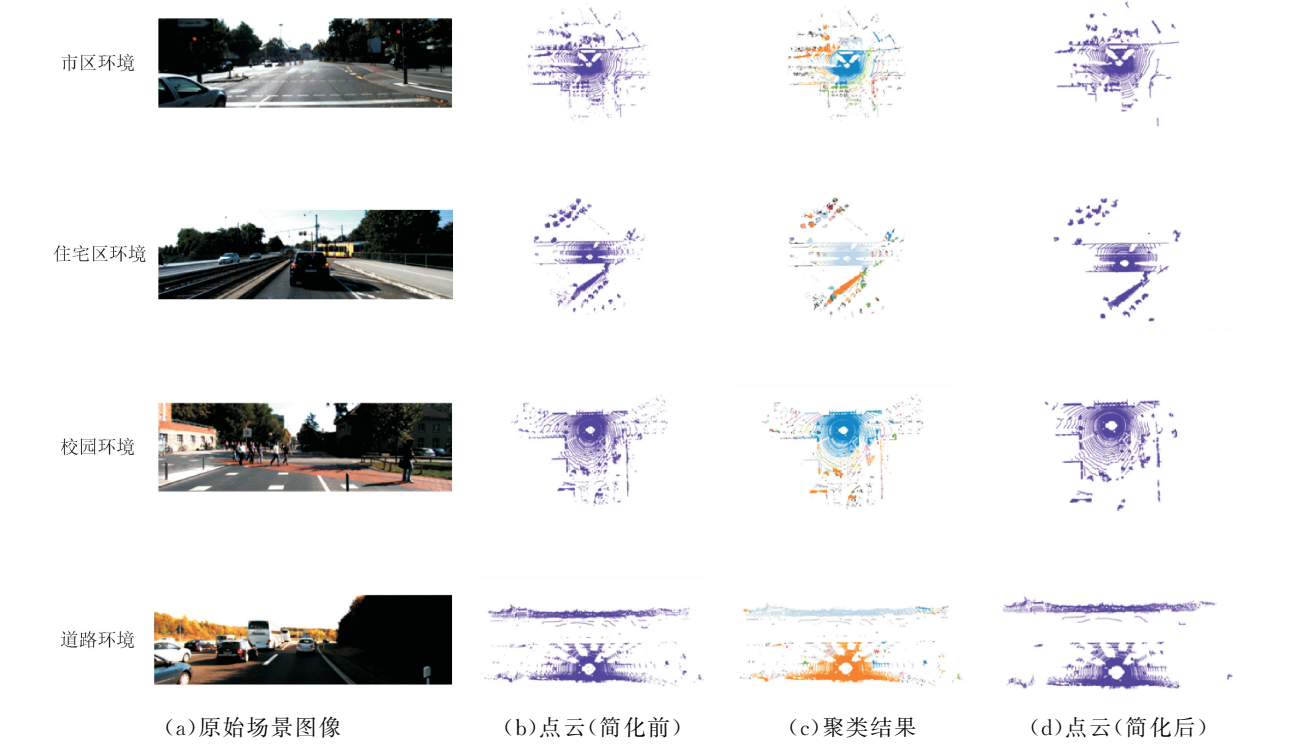


图 7 本文算法在不同环境下的点云简化效果

Fig. 7 Point cloud simplification effect of the algorithm in different environments in this paper

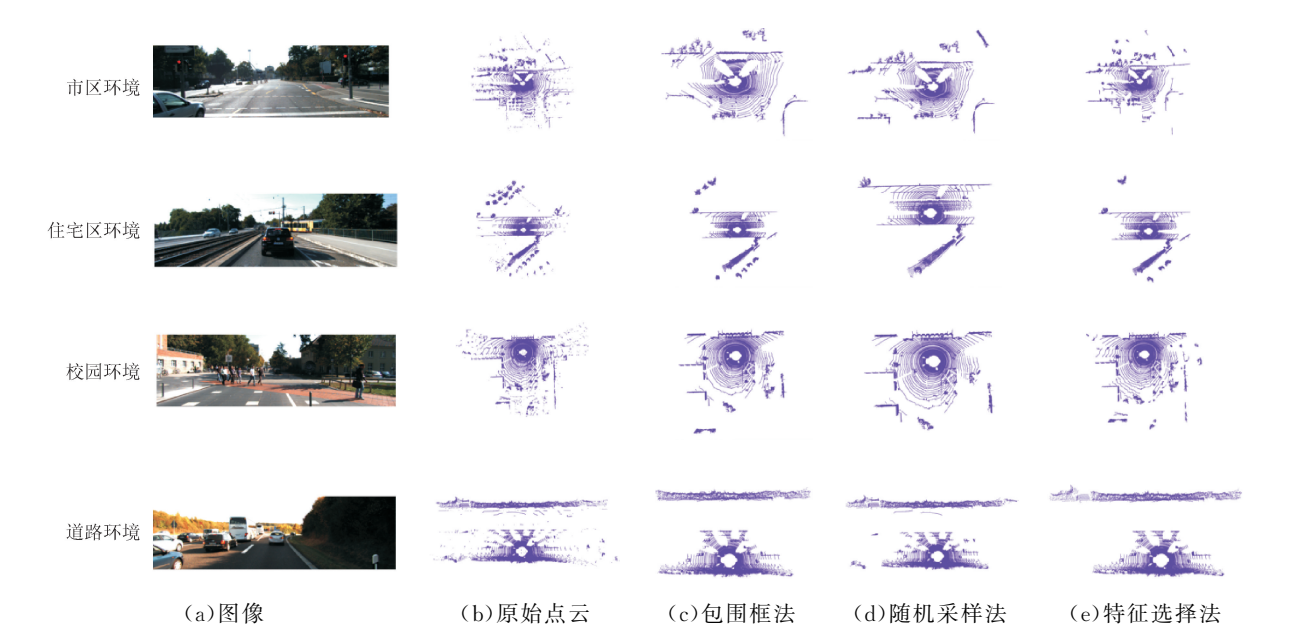


图 8 对比算法在不同环境下的点云简化效果

Fig. 8 Comparing the simplification effect of the algorithms on point clouds in different environments

由图 8 可知:①市区环境的点云数据呈现干扰严重、噪声点多的特点,点云密度较高,包含了大量的建筑物、道路、交通标志等细节,包含了多种复杂的结构和纹理信息。图 7 聚类结果中黑色数据表示噪声点云,本文算法简化后的点云数据表现出高内聚的特点,且保留了较高的完整性,避免了关键特征丢失的问题;②住宅区环境的点云数据具有相对简单、噪声较小、密度适中的特点,包含了一些低层建筑物、小型道路等,在简化点云时,本文算法在保留关键特征和细节轮廓的前提下,对噪声干扰点进行了去除;③校园环境下的点云具有视野开阔、结构简单、密度较小的特点,包含了一些教学楼、行人、自行车等特征,因此在简化过程中本文算法删除小目标的情况较少,只删除了孤立的无效数据点,从而减少点云数;④道路环境下的点云数据呈现视野开阔,障碍物较少、密度较高的特点,包含了道路的结构、车辆、行人等信息。由于视野较为开阔,雷达采集到大量远距离稀疏点云,这类点云无法在后期提供有效特征,且容易成为误检目标,由本文算法简化后的点云可以看出,因稀疏造成的无效点云信息熵较低,从而作了删除处理,保留的点云均具有足够的特征信息。

综上所述,基于随机采样的简化算法并不能很好地识别到离群数据,稳定性相对较差;基于包围框的简化算法只考虑了点云的空间位置,没有考虑点云的形状和特征信息,在点云的形状比较复杂或者噪声较多时,算法会产生较大的误差,导致简化后的点云与原始点云的差异较大。相较于特征法,本文算法可以更有效地删除由于采样设备带来的噪声数据和离散点数据,这些数据分布范围广,但是包含信息量少。综合来看,相较于对比算法,本文算法在点云简化方面的效果更佳,有效避免了传统精简算法易误删特征点、出现孔洞、边界点丢失的问题。

最后进行定量分析实验,更加直接地展示本文

算法与对比算法的点云简化效果。

为证明算法的有效性性与先进性,采用简化前、后点集之间平均欧式距离(M)、Hausdorff 距离(H)和算法运行时间(t)3 个指标来评价算法的性能。其中, M 用来衡量算法简化的全局误差, H 用来衡量简化前后的两个点集之间的局部误差,运行时间反映了算法的简化效率。

上述指标的计算方法分别如下

$$M = \frac{1}{m} \sum_{l=1}^m \min_{j=1}^n |x_l - y_j| + \frac{1}{n} \sum_{j=1}^n \min_{l=1}^m |x_l - y_j| \tag{26}$$

$$H = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \tag{27}$$

式中: x_l, y_j 分别为点集 $X = \{x_1, x_2, \dots, x_m\}, Y = \{y_1, y_2, \dots, y_n\}$ 中的样本点; $\sup(\cdot)$ 是上确界操作; $\inf(\cdot)$ 是下确界操作。

将本文算法与对比算法在不同环境下的简化结果进行统计,分别进行 50 次独立实验取平均值,结果如表 7 和图 9 所示。

可以看出,本文算法在以 M 为指标的全局误差方面优于其他 3 种点云简化算法,且以 H 为指标的局部误差仅在道路环境下略低于基于特征的点云简化算法,在其他环境下均远高于另外 3 种简化算法;在运行时间方面,本文算法相对包围框法和特征法,运行时间均有明显的减少,尽管与随机采样法相比,运行速度在校园环境和市区环境下较慢,但随机采样法在点云数据简化过程中 M 全局误差和 H 局部误差无法得到有效的控制,时高时低不稳定,且简化精度无法保证。综上所述,本文提出的 DEAMPOA 与加权熵法联合优化的 FCM 信息熵点云简化算法综合简化效果最优,且运行速度较快,满足简化要求。

表 7 不同算法点云简化统计数据
Table 7 Different algorithms' point clouds simplification statistics

算法	M				H				t/s			
	市区	住宅区	校园	道路	市区	住宅区	校园	道路	市区	住宅区	校园	道路
本文算法	0.021 0	0.001 1	0.016 4	0.005 1	0.000 3	0.001 5	0.002 7	0.002 3	0.317	0.326	0.372	0.438
包围框法	0.043 5	0.023 1	0.068 9	0.055 6	0.007 2	0.004 6	0.003 2	0.006 6	1.497	1.449	1.431	1.364
随机采样法	0.090 3	0.097 9	0.098 7	0.069 1	0.007 2	0.004 6	0.003 2	0.006 6	0.243	0.199	0.276	0.274
特征法	0.078 3	0.019 5	0.065 2	0.009 3	0.001 0	0.009 5	0.001 7	0.002 4	0.737	0.889	0.798	0.712

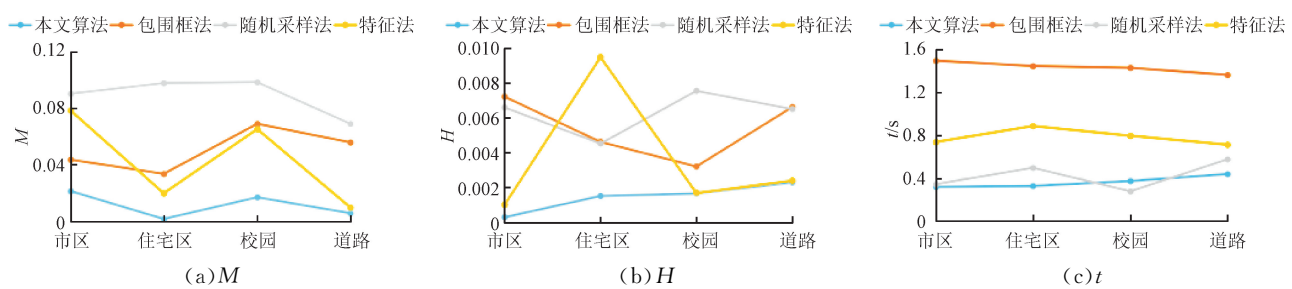


图 9 本文算法与对比算法在不同环境下的简化结果

Fig. 9 Simplification results of the proposed algorithm and the comparative algorithms in different environments

5 结 论

本文提出了一种基于 DEAMPOA 与加权熵法联合优化的 FCM 信息熵点云简化算法。利用动态精英自适应混合策略提升了 POA 的收敛趋势和跳出局部最优能力,并与加权熵法联合,提高了寻找 FCM 最优聚类中心的成功率,提高鲁棒性的同时能够增强搜索精度,得到较好的聚类结果。融合信息熵进行去噪,在 KITTI 点云数据集上的简化实验结果可知,本文算法可以更有效地实现三维点云数据的简化,具有一定的应用价值。但是,目前只针对道路信息点云的简化结果进行了分析,面对实际应用场景复杂多变的环境未做针对性提升,在后续研究中应扩大样本数据,对算法在复杂场景下的鲁棒性做进一步的强化提升,研究多场景通用、适用性更佳 的点云简化算法。

参考文献:

[1] 赵佳琦,周勇,何欣,等. 基于深度学习的点云分割研究进展分析 [J]. 电子与信息学报, 2022, 44(12): 4426-4440.
ZHAO Jiaqi, ZHOU Yong, HE Xin, et al. Research progress analysis of point cloud segmentation based on deep learning [J]. Journal of Electronics & Information Technology, 2022, 44(12): 4426-4440.

[2] 王子洋,李琼琼,张子蕴,等. 应用于无人驾驶车辆的点云聚类算法研究进展 [J]. 世界科技研究与发展, 2021, 43(3): 274-285.
WANG Ziyang, LI Qionqiong, ZHANG Ziyun, et al. Research progress of unmanned vehicle point cloud clustering algorithm [J]. World Sci-Tech R&D, 2021, 43(3): 274-285.

[3] 鲁斌,范晓明. 基于改进自适应 k 均值聚类的三维点云骨架提取的研究 [J]. 自动化学报, 2022, 48(8): 1994-2006.
LU Bin, FAN Xiaoming. Research on 3D point cloud

skeleton extraction based on improved adaptive k -means clustering [J]. Acta Automatica Sinica, 2022, 48(8): 1994-2006.

[4] CHENG Fang, NIU Guofeng, ZHANG Zhizhong, et al. Improved CNN-based indoor localization by using RGB images and DBSCAN algorithm [J]. Sensors, 2022, 22(23): 9531.

[5] SHI Yuqing. Application of FCM clustering algorithm in digital library management system [J]. Electronics, 2022, 11(23): 3916.

[6] 黄鹤,李昕芮,吴琨,等. 引入改进飞蛾扑火的 K 均值交叉迭代聚类算法 [J]. 西安交通大学学报, 2020, 54(9): 32-39.
HUANG He, LI Xinrui, WU Kun, et al. Hybrid iterative K -means clustering with improved Moth-Flame optimization [J]. Journal of Xi'an Jiaotong University, 2020, 54(9): 32-39.

[7] CHEN Yewang, ZHOU Lida, BOUGUILA N, et al. BLOCK-DBSCAN: fast clustering for large scale data [J]. Pattern Recognition, 2021, 109: 107624.

[8] YANG Yang, LI Ming, MA Xie. A point cloud simplification method based on modified fuzzy C -means clustering algorithm with feature information reserved [J]. Mathematical Problems in Engineering, 2020, 2020: 5713137.

[9] 张佳琦,王建民. 应用改进狼群算法优化模糊聚类实现点云数据的区域分割 [J]. 科学技术与工程, 2023, 23(30): 13002-13013.
ZHANG Jiaqi, WANG Jianmin. Region segmentation of point cloud data based on improved wolf pack algorithm optimization fuzzy clustering [J]. Science Technology and Engineering, 2023, 23(30): 13002-13013.

[10] TROJOVSKÝ P, DEGHANI M. Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications [J]. Sensors, 2022, 22(3): 855.

[11] 张建华,刘艺琳,郭启迪,等. 融合改进 FCM 与 PFS 的知识供需匹配 [J]. 计算机工程与设计, 2023, 44

- (1): 99-107.
- ZHANG Jianhua, LIU Yilin, GUO Qidi, et al. Knowledge supply and demand matching based on improved FCM and PFS [J]. Computer Engineering and Design, 2023, 44(1): 99-107.
- [12] 宋燕, 李元昊, 李明. 融合稀疏自表示和残差驱动自适应模糊C均值聚类 [J]. 控制与决策, 2024, 39(4): 1333-1341.
- SONG Yan, LI Yuanhao, LI Ming. Sparse self-representation incorporated and residual driven adaptive fuzzy C-means clustering [J]. Control and Decision, 2024, 39(4): 1333-1341.
- [13] 吴辰文, 王莎莎, 曹雪同. 结合柯西分布和蚁狮算法改进的模糊聚类算法 [J]. 计算机工程与应用, 2023, 59(17): 91-98.
- WU Chenwen, WANG Shasha, CAO Xuetong. Fuzzy clustering algorithm combined with Cauchy distribution and ant lion algorithm [J]. Computer Engineering and Applications, 2023, 59(17): 91-98.
- [14] EDWIN DHAS P, SANKARA GOMATHI B. A novel clustering algorithm by clubbing GHFCM and GWO for microarray gene data [J]. The Journal of Supercomputing, 2020, 76(8): 5679-5693.
- [15] 黄鹤, 李文龙, 吴琨, 等. 动态自适应特征融合的MFOPA跟踪器 [J]. 电子学报, 2023, 51(5): 1350-1358.
- HUANG He, LI Wenlong, WU Kun, et al. MFOPA tracker with dynamic adaptive feature fusion [J]. Acta Electronica Sinica, 2023, 51(5): 1350-1358.
- [16] XUE Jiankai, SHEN Bo. A novel swarm intelligence optimization approach: sparrow search algorithm [J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.
- [17] BEZDEK J C, EHRLICH R, FULL W. FCM: the fuzzy c -means clustering algorithm [J]. Computers & Geosciences, 1984, 10(2/3): 191-203.
- [18] 余庆, 胡尧. 基于改进FCM聚类算法的高速公路交通状态识别 [J]. 交通运输研究, 2021, 7(2): 47-54.
- YU Qing, HU Yao. Identification on traffic state of expressway based on improved FCM clustering algorithm [J]. Transport Research, 2021, 7(2): 47-54.
- [19] 刘张桥, 王成良, 焦晓军. 多核环境下的图像分割并行算法研究 [J]. 计算机工程, 2011, 37(15): 197-200.
- LIU Zhangqiao, WANG Chengliang, JIAO Xiaojun. Research of parallel algorithm for image segmentation under multi-core environment [J]. Computer Engineering, 2011, 37(15): 197-200.
- [20] 赵夫群, 李映莹. 基于点重要性判断的点云简化 [J]. 计算机系统应用, 2023, 32(9): 197-202.
- ZHAO Fuqun, LI Yingxuan. Point cloud simplification based on point importance judgment [J]. Computer Systems & Applications, 2023, 32(9): 197-202.
- [21] HE Siwei, LIU Baolong. Review of bounding box algorithm based on 3D point cloud [J]. International Journal of Advanced Network Monitoring and Controls, 2021, 6(1): 18-23.
- [22] 孔利, 王延存, 周茂伦, 等. 基于随机抽样与特征值法的点云平面稳健拟合方法 [J]. 测绘与空间地理信息, 2019, 42(3): 43-46.
- KONG Li, WANG Yancun, ZHOU Maolun, et al. A robust point cloud planar fitting method based on random sampling and eigenvalue [J]. Geomatics & Spatial Information Technology, 2019, 42(3): 43-46.
- [23] MAHDAOUI A, SBAI E H. 3D point cloud simplification based on k -nearest neighbor and clustering [J]. Advances in Multimedia, 2020, 2020: 8825205.
- [24] THARWAT A. Parameter investigation of support vector machine classifier with kernel functions [J]. Knowledge and Information Systems, 2019, 61(3): 1269-1302.
- [25] 陈辉, 黄晓铭, 刘万泉. 基于动态网格 k 邻域搜索的激光点云精简算法 [J]. 控制与决策, 2020, 35(12): 2986-2992.
- CHEN Hui, HUANG Xiaoming, LIU Wanquan. Laser point cloud simplification algorithm based on dynamic grid k -nearest neighbors searching [J]. Control and Decision, 2020, 35(12): 2986-2992.

(编辑 武红江)