

doi:10.16112/j.cnki.53-1223/n.2026.01.202408190005

引用格式:孙林,薛红科,吕娟.改进鲸鱼优化的模糊C-均值聚类算法[J].昆明理工大学学报(自然科学版),2026,51(1):112-125.

Citation:SUN Lin,XUE Hongke,LYU Juan.Improved Whale Optimization-Based Fuzzy C-means Clustering Algorithm[J].Journal of Kunming University of Science and Technology(Natural Science),2026,51(1):112-125.

## 改进鲸鱼优化的模糊C-均值聚类算法

孙林,薛红科,吕娟\*

(天津科技大学人工智能学院,天津300457)

**摘要:** 由于模糊C-均值聚类(Fuzzy C-Means Clustering, FCM)算法对初始化聚类中心的选取具有高度敏感性,以及在迭代优化过程中易陷入局部最优的困境,提出了一种改进鲸鱼优化的FCM算法.首先,为了解决鲸鱼优化算法(Whale Optimization Algorithm, WOA)对初始解依赖较强的问题,在WOA初始化环节,融入Logistics-Tent混沌序列,扩大了WOA的搜索范围;其次,在WOA的位置更新公式构建中,引入了自适应权重机制,通过调整自适应权重的大小,控制个体的搜索范围,平衡了WOA的全局搜索的搜索能力与局部开发能力;然后,在位置更新结束后,引入邻域扰动策略,对最优解附近的邻域进行扰动,避免WOA陷入局部最优和出现过早收敛的情况;最后,为解决FCM算法对初始聚类中心过度敏感的问题,将FCM的目标函数作为WOA的适应度函数,使FCM算法获得与自适应优化算法相同的特性.10个基准测试函数和10个公开的数据集上的实验结果表明,所提算法是有效的.

**关键词:** 模糊C-均值聚类;鲸鱼优化;混沌映射;自适应权重;邻域扰动

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 1007-855X(2026)01-0112-14

### Improved Whale Optimization-Based Fuzzy C-means Clustering Algorithm

SUN Lin, XUE Hongke, LYU Juan\*

(College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China)

**Abstract:** Due to the high sensitivity of the Fuzzy C-Means Clustering (FCM) algorithm to the selection of initial clustering centers and its tendency to get stuck in local optima during the iterative optimization process, an improved FCM algorithm based on Whale Optimization Algorithm (WOA) is proposed. Firstly, to address the strong dependence of the WOA on initial solutions, the Logistics-Tent chaotic sequence is integrated into the initialization phase of WOA, ensuring that the initialized population of WOA is evenly distributed within the solution space. Secondly, in the construction of the position update formula for WOA, an adaptive weight mechanism is introduced. By adjusting the size of the adaptive weight, the search range of individuals is controlled, balancing the global exploration and local exploitation capabilities of WOA. Then, after the position update, a neighborhood perturbation strategy is introduced to perturb the neighborhood near the optimal solution, preventing WOA from getting stuck in local optima and solving the problem of premature convergence. Finally, to address the issue of FCM algorithm being overly sensitive to initial clustering centers, the objective function of FCM is used as the fitness function of WOA, endowing the FCM algorithm with the same characteristics as adaptive optimization algorithms.

**收稿日期:** 2024-08-19. **基金项目:** 国家自然科学基金项目(62576245).

**作者简介:** 孙林(1979-),男,博士,教授.主要研究方向:数据挖掘,深度学习. **E-mail:** sunlin@tust.edu.cn

**通信作者:** 吕娟(1991-),女,博士,副教授.主要研究方向:深度学习,机器学习. **E-mail:** lvjuan@tust.edu.cn

Experimental results on 10 benchmark test functions and 10 publicly available datasets demonstrate the effectiveness of the proposed algorithm.

**Keywords:** fuzzy C - means clustering; whale optimization; chaotic map; adaptive weight; neighborhood perturbation

## 0 引言

近年来,聚类分析作为无监督机器学习技术的主要代表之一,其本质是将数据划分为若干类簇,将相似度较高的数据聚集在同一个簇中<sup>[1-2]</sup>.随着数据复杂性的增加,聚类算法研究从硬聚类逐步转移到模糊聚类上,以FCM为代表的软聚类方法更具优势<sup>[3]</sup>.FCM通过引入隶属度的概念,对数据点与各类簇之间的归属关联程度进行量化衡量,能够深度反映现实世界数据中复杂且模糊的信息结构.

由于FCM的初始聚类中心的选择是随机的,不同的初始值可能导致截然不同的聚类效果.同时,FCM算法在聚类过程中易陷入局部最优的问题,导致聚类效果不佳,严重影响了聚类算法的稳定性<sup>[4]</sup>.何庆龄等<sup>[5]</sup>采用犹豫模糊理论和改进麻雀优化算法确定FCM的模糊度指数,进而提高求解高维复杂问题的优化性能.但随着环境的变化,该算法的寻优稳定性会下降.Zhang等<sup>[6]</sup>从多目标优化问题出发,将多目标优化问题中的SVM和FCM目标函数融合,但混合算法的聚类效果受初始化参数设置影响较大.针对FCM算法对初始化聚类中心过度敏感问题,学者们提出将群体智能优化算法与聚类分析相结合,取代原始聚类中的随机初始化机制,减少了FCM初始化参数对聚类性能的影响,提高了聚类质量,进而克服了传统聚类算法在优化过程中易陷入局部最优的局限.Sun等<sup>[7]</sup>将多策略金枪鱼群优化算法与FCM结合,引入全局搜索能力和自适应参数调整机制,提升了聚类精度,但该算法仍存在对初始聚类中心过度敏感的问题.孟学尧等<sup>[8]</sup>采用改进的萤火虫算法优化FCM聚类中心,增强了FCM的准确率和稳定性,但该算法需要进一步调整陷入局部最优的问题.李清<sup>[9]</sup>将FCM与粒子群优化相结合,提高了电力数据异常值检测的准确率,但该算法聚类效果对初始参数设置敏感.上述群体智能优化算法<sup>[10]</sup>仍存在位置更新局限性的问题,导致优化算法求解容易陷入局部最优,使得FCM效果受限.作为一种新型的群体智能优化算法,WOA<sup>[11]</sup>因其参数少、寻优规则简单、算法鲁棒性强等特点备受关注.该算法以随机解为起点,通过迭代方式求得最优解,因此具有较高的求解效率,有助于FCM找到全局最优解,提高聚类的准确性和效率.Zhang等<sup>[12]</sup>构建了基于动态搜索策略WOA的FCM算法,使其获得与自适应优化算法相同的自动调整策略,从而提高算法的收敛速度和精度,但该算法参数调整困难.

综上所述,群体智能优化算法具备高求解效率与强适应性等显著特性<sup>[13]</sup>,在针对FCM聚类算法的优化过程中,能有效提升该算法的聚类精度.而WOA因全局搜索能力强、参数设置简单、收敛速度快、鲁棒性强等特点在群体智能优化算法中展现出独特的优势.然而,WOA仍存在陷入局部最优和全局搜索能力与局部开发能力不平衡等问题.因此,本文采用改进WOA对FCM算法进行优化和改进.主要贡献包括如下4个方面:

- 1) 为了避免初始化种群中个体重复导致WOA陷入局部最优解,提高WOA的全局搜索能力,采用Logistics混沌映射和Tent混沌映射结合后的Logistics-Tent混沌映射来替代WOA中的随机初始化过程.
- 2) 为了提高WOA的收敛速度,平衡WOA的全局搜索能力和局部开发能力,提出自适应权重来限制领头鲸鱼的空间搜索范围;为了避免WOA陷入局部最优解,解决WOA早熟现象,提出了最优邻域扰动策略.
- 3) 为了克服FCM算法对初始聚类中心过度敏感的问题,基于改进WOA优化FCM的初始聚类中心,使FCM算法获得与自适应优化算法相同的特性,提高FCM的聚类效果.
- 4) 为了提高FCM算法的全局搜索能力,将FCM的隶属度矩阵进行重构,用于对WOA种群的初始化,引导算法的搜索过程.

## 1 鲸鱼优化

WOA是以鲸鱼在捕食和迁徙过程中的群体行为为基础,通过模拟鲸鱼群体的寻找和追逐行为,从而寻

找全局最优解<sup>[12]</sup>. 其主要步骤包括初始化种群、随机搜索和位置更新三个阶段. 首个阶段被定义为猎物包围阶段. 在鲸鱼探测到猎物存在后, 开始实施对猎物的包围策略. 在此阶段, 鲸鱼依据目标猎物的具体位置信息或是距离目标猎物最近的点位, 对其自身位置进行更新调整. 鲸鱼位置更新<sup>[11]</sup>的相关公式为:

$$D_1 = |CX^*(n) - X(n)| \quad (1)$$

$$X(n+1) = X^*(n) - AD_1 \quad (2)$$

式中:  $D_1$  表示包围过程中的步长参数,  $X(n)$  表示当前鲸鱼个体在搜索空间中的位置,  $X^*(n)$  表示目前迭代阶段获得的最优位置,  $n$  表示当前迭代次数,  $A$  和  $C$  均为系数向量;  $|\cdot|$  表示绝对值操作. 两个系数向量  $A$  和  $C$ <sup>[14]</sup> 的计算公式为:

$$A = 2ar - a \quad (3)$$

$$C = 2r \quad (4)$$

式中: 参照文献[14],  $a$  为收敛因子, 在迭代中从 2 到 0 线性递减;  $r$  为  $[0, 1]$  中的随机向量.

当座头鲸识别到猎物后, 围绕猎物进行收缩环绕, 逐渐缩小包围圈. 收缩环绕机制依赖于参数  $A$  和  $a$ . 向量  $A$  的值根据迭代次数和最大迭代次数来计算, 其范围在  $[-2, 2]$  之间波动, 在迭代过程中线性递减到 0, 使得  $A$  的取值范围逐渐缩小.

第二阶段为捕猎阶段. 该阶段中, 鲸鱼与猎物之间的距离通过螺旋数学方法确定, 鲸鱼依据趋向最佳位置的螺旋更新机制, 对其自身位置进行动态调整. 其螺旋更新公式<sup>[11]</sup>为:

$$X(n+1) = e^{bk} \cos(2\pi z) D_2 + X^*(n) \quad (5)$$

$$D_2 = |X^*(n) - X(n)| \quad (6)$$

式中: 常数  $b$  为螺旋的形状特征,  $z$  为取值范围处于  $[-1, 1]$  之间的任意数,  $k$  为  $[-1, 1]$  之间的随机数,  $D_2$  为第  $i$  条鲸鱼与猎物之间的距离.

为了精确模拟鲸鱼在逐步缩小的圆形区域内围绕猎物游动, 同时沿着螺旋路径移动的行为, 参照文献[14], 设定在收缩包围策略与螺旋模型之间进行选择的概率为 0.5, 对鲸鱼的位置进行更新. 收缩包围和螺旋模型结合公式为:

$$X(n+1) = \begin{cases} X^*(n) - AD_1, & p < 0.5 \\ D_2 e^{bk} \cos(2\pi z) + X^*(n), & p \geq 0.5 \end{cases} \quad (7)$$

第三阶段是探索阶段, 座头鲸根据彼此位置随机搜索. 探索阶段公式<sup>[14]</sup>为:

$$D^* = |CX_{\text{rand}} - X| \quad (8)$$

$$X(n+1) = X_{\text{rand}} - AD^* \quad (9)$$

式中:  $X_{\text{rand}}$  表示随机位置向量,  $D^*$  表示搜索位置与随机个体的距离.

## 2 基于改进 WOA 的 FCM 算法

### 2.1 基于 Logistics - Tent 混沌映射的初始化

在 WOA 中, 其初始种群生成采用随机选择策略, 导致初始种群分布不均衡, 种群在可行域内覆盖范围受限, 降低了算法的寻优精度<sup>[11]</sup>. Logistics 混沌系统生成的映射点分布不均匀, 影响了算法的收敛效率; 而 Tent 混沌系统具有结构简单、自相关性强的特点, 但其映射区间受限<sup>[4]</sup>. 将 Logistics 和 Tent 映射的优势相结合, 构建的 Logistics - Tent 混沌系统, 具有更广的映射空间, 能够使初始种群分布更加均匀, 从而优化 WOA 的性能. 由此给出 Logistics - Tent 混沌映射的定义:

$$x_{n+1} = \begin{cases} \left[ r_2 x_n (1 - x_n) + \frac{(4 - r_2)}{2} x_n \right] \bmod 1, & x_n < 0.5 \\ \left[ r_2 x_n (1 - x_n) + \frac{(4 - r_2)(1 - x_n)}{2} \right] \bmod 1, & x_n \geq 0.5 \end{cases} \quad (10)$$

式中:  $x_n$  表示系统变量,  $0 < x_0 < 1$ ;  $r$  表示控制参数, 依据文献[14]的参数设置,  $0 < r_2 \leq 4$ .

## 2.2 基于自适应权重的位置更新

在传统 WOA 中,领导者负责引导鲸鱼群朝食物源移动,但缺乏对领导者搜索范围的限制,导致搜索可能超出固定区域,进而影响算法求解的准确性<sup>[12]</sup>.此外,领导者过度开展搜索行为,会致使 WOA 迭代速度降低.为了解决这些问题,本文构造基于鲸鱼个体的适应度,通过动态调整权重,帮助 WOA 在搜索过程中平衡全局搜索和局部搜索.于是,适应度更好的鲸鱼个体获得更大权重,以便在搜索空间中进一步探索;而适应度较差的鲸鱼个体则获得较小权重,以便减少在搜索空间中的影响.这种动态调整策略使 WOA 在初期有较强的全局搜索能力,而在后期则更加注重局部搜索和收敛速度.鲸鱼个体自适应权重公式为:

$$w = w_{\max} - \frac{(w_{\max} - w_{\min})F}{Leader} \quad (11)$$

式中: $w$ 表示自适应权重, $w_{\max}$ 为最大权重, $w_{\min}$ 为最小权重, $F$ 为 WOA 适应度, $Leader$ 为全局最优适应度.由此,基于自适应权重改进后的鲸鱼位置更新公式为:

$$\vec{X}(n+1) = e^{bk} \cos(2\pi z) \vec{D} + w \vec{X}^*(n) \quad (12)$$

## 2.3 最优邻域扰动策略

WOA 在求解过程中,会过早收敛到局部最优解,而产生早熟现象<sup>[14]</sup>.为了解决该问题,本文引入最优邻域扰动策略改进 WOA.该策略在 WOA 接近最优解时,对最优解附近的邻域进行扰动.这种扰动的引入打破了原有的局部最优结构,为 WOA 提供了新的搜索方向,从而帮助 WOA 跳出局部最优,继续在全局范围内搜索更好的解.扰动策略公式为:

$$position = Leader + vector(ub - lb)q \quad (13)$$

式中: $position$ 是鲸鱼算法产生的新位置; $Leader$ 是当前找到的最优位置向量; $vector$ 是一个  $n$  维的随机向量; $ub$ 和  $lb$ 分别是搜索空间的上界和下界,通过调整系数  $q(0 \leq q \leq 1)$ 控制扰动的幅度,从而平衡算法的局部搜索和全局搜索能力.

为了解决 WOA 过早收敛、提高 WOA 的全局搜索能力、平衡 WOA 的全局搜索能力和局部开发能力,基于 Logistics - Tent 混沌映射、自适应权重和最优邻域扰动策略三个改进点设计改进的 WOA (Improved WOA, IWOA),该算法的伪代码步骤如算法 1 所示.

## 2.4 FCM 的初始聚类中心

FCM<sup>[5]</sup>的目标函数与数据的聚类特性紧密相关,能够有效地反映数据样本之间的相似性和差异性.在 WOA 中,适应度函数是用来评估解的优劣的重要指标,它对于指导搜索过程和衡量解的质量起着至关重要的作用<sup>[11]</sup>.将 FCM 的目标函数作为改进 WOA 的目标函数,来优化 FCM 的初始聚类中心,使改进的 WOA 能够更加具有针对性地搜索与问题相关的解空间.通过利用目标函数进行迭代优化,可以选出更好的初始聚类中心,从而提高 FCM 的聚类效果.因此,IWOA 的目标函数公式为:

$$f = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - v_j\|_2^2 \quad (14)$$

### 算法 1 IWOA

输入:数据集、初始化鲸鱼种群  $pop$ 、最大迭代次数  $Max\_iter$ 、搜索空间上界  $ub$  和下界  $lb$ .

输出:最优适应度  $f$ .

1. 计算鲸鱼个体适应度  $f^*$
2.  $X^*$  = 最佳个体适应度
3. while ( $t < Max\_iter$ )
4.   for 每个搜索个体
5.     更新参数  $a$ 、 $A$ 、 $C$ 、 $l$  和  $p$
6.     if  $p < 0.5$
7.       if  $|A| < 1$
8.         根据式(1)更新鲸鱼位置
9.       else if  $|A| \geq 1$
10.         随机选择鲸鱼  $X_{rand}$
11.         根据式(2)更新鲸鱼位置
12.       end if
13.       else if  $p \geq 0.5$
14.         根据式(5)更新当前搜索位置
15.         根据式(16)调用邻域扰动策略
16.       end if
17.   end for
18. 检查鲸鱼位置
19. 计算每个鲸鱼个体的适应度
20. 如果有更好位置则更新  $X^*$
21.  $t = t + 1$
22. end while
23. return  $X^*$

### 2.5 重构 FCM 隶属度矩阵

为了解决 FCM 算法全局搜索能力差的问题,本文提出将 FCM 的隶属度矩阵进行重构,使其具有与 WOA 种群等维度的数据结构,实现对 WOA 种群的初始化,并通过 WOA 的随机搜索和位置更新两个阶段来更新种群. FCM 算法在迭代过程中主要依赖于当前的聚类中心和样本的隶属度进行更新,这种更新方式导致算法在搜索过程中过早地收敛到局部最优解. 而 WOA 通过包围猎物、气泡网攻击等策略,在搜索空间中进行广泛的探索,使得 WOA 能够在较大的搜索范围内寻找最优解,从而避免陷入局部最优. 因此,本文利用 WOA 的全局搜索能力,提出利用重构的隶属度矩阵对 WOA 种群进行初始化,从而使 FCM 算法具备更强的全局搜索和位置更新能力. 该方法既利用了 WOA 的全局搜索能力,又保留了 FCM 在模糊数学中柔性划分与隶属度的思想,能够使 FCM 算法具备更强的全局搜索和位置更新能力,可以有效解决 FCM 算法全局搜索性差,易陷入局部最优的问题. 所提出的隶属度矩阵重构方法如图 1 所示. 假设隶属度矩阵为  $j \times i$  维,其中  $j$  表示聚类样本个数, $i$  表示聚类中心个数,矩阵中每个元素表示对应的隶属度值. 具体重构方法为:首先选择与 WOA 特征数相同的样本,将样本对各聚类中心的隶属度从隶属度矩阵中提取出来,然后将每行元素按原有顺序串接,最终重构为与 WOA 种群相同的维度.

重构后 WOA 的种群数为  $j$  个样本与  $i$  个聚类中心的乘积,其表达式为:

$$dim = i \times j \quad (15)$$

式中: $dim$  表示重构后 WOA 的种群数量.

在 FCM 算法中,每个样本对各个聚类的隶属度被表示为 0 到 1 之间的数值,这些数值之和为 1,表示了样本属于各个聚类的程度,对应为隶属度矩阵中的每一行. 保持这一约束使每个样本对各聚类中心的归属程度归一化,有助于更准确地解释聚类结果,减少由于 WOA 随机搜索过程中的随机性导致的聚类结果波动,增强了算法的鲁棒性. 因此,在对 WOA 进行种群初始化时仍保留隶属度矩阵这一特性,通过对重构后的种群进行约束,其约束公式为:

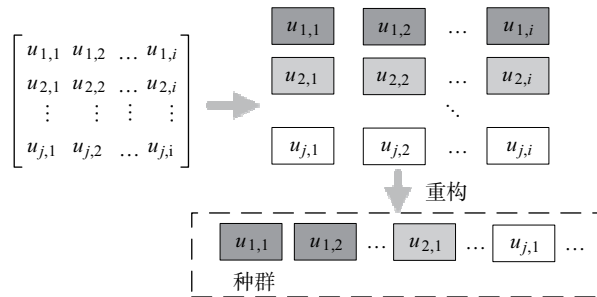


图 1 隶属度矩阵重构图

Fig. 1 Reconstruction diagram of membership matrix

$$\sum_{j=1}^i u_{ij} = 1 \quad (16)$$

式中: $j$  为样本个数, $i$  为聚类中心个数,使第  $j$  个样本点对  $i$  个聚类中心隶属度值的和为 1.

### 2.6 IWOAFCM 算法流程

为了清晰地展示基于 IWOA 的 FCM 算法的整体流程,其流程图如图 2 所示. 利用 IWOA 构建 FCM 算法的具体步骤如下:首先,针对 FCM 确定初始聚类中心,通过随机选取的方式确定初始的  $c$  个聚类中心,并将这些聚类中心编码为 WOA 中鲸鱼个体所对应的位置信息;其次,利用 IWOA 对给定的数据集开展寻优操作,生成最优的鲸鱼位置作为 FCM 算法的初始聚类中心;最后,由 FCM 算法输出隶属度矩阵以及聚类中心. 由此,可以得到基于 IWOA 的 FCM 算法(FCM Algorithm Based on IWOA, IWOAFCM),其伪代码如算法 2 所示.

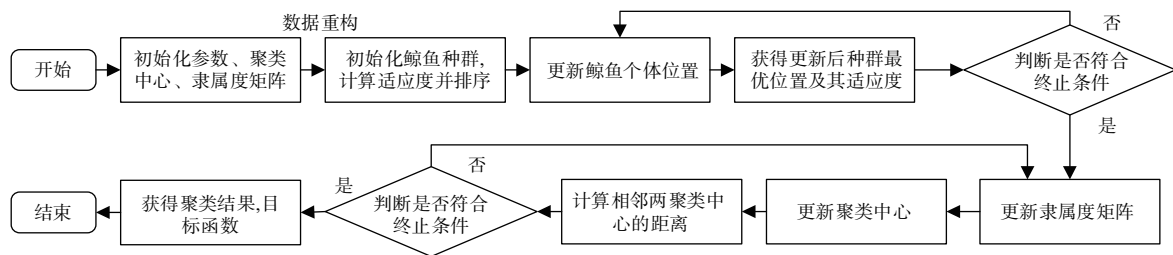


图 2 算法流程图

Fig. 2 Flowchart of the algorithm

### 3 实验结果与分析

本实验硬件环境为 12th Gen Intel(R) Core (TM) i5 - 12500 3.00 GHz 和 16G RAM, 运行平台为 MATLAB R2016b 和 Python 3.9. 为了评估所提算法的效果, 实验分为两部分: 第一部分选择 10 个基准测试函数对 IWOA 的优化性能进行评估; 第二部分实验选择普遍采用的 5 个人工数据集和 5 个 UCI 数据集, 运用消融实验精准分析算法各部分对整体性能的影响, 同时开展聚类效果对比实验, 验证 IWOAFCM 算法在聚类任务中的有效性.

#### 3.1 IWOA 的优化性能分析

为对 IWOA 算法的寻优性能展开验证, 参照文献[15], 选取 10 个常用的基准测试函数进行实验验证. 其中,  $f_1 - f_5$  是单峰函数, 此类函数仅存在一个局部极值点, 运用这些函数能够精准评估算法的迭代效率;  $f_6 - f_{10}$  是多峰函数, 具有多个极值点, 可以验证算法的全局搜索与局部开发能力.

为有效验证 IWOA 算法中 3 种改进策略的有效性, 设计了 3 种 IWOA 算法的变体形式: (1) 将 Logistics - Tent 混沌映射策略融入 WOA 算法(IWOA1); (2) 融合自适应权重的 WOA 算法(IWOA2); (3) 融合最优邻域扰动策略的 WOA 算法(IWOA3). 表 1 呈现了标准 WOA 算法、3 种 IWOA 算法的变体和本文提出的 IWOA 算法在 10 个基准测试函数上运行 20 次的平均值. 这 5 种算法均在 10 个基准测试函数上进行寻优操作, 每种算法种群规模设为 30, 种群维度设为 30, 迭代次数 Max\_iter 设为 500.

为了测试上述 5 种算法的优化性能, 采用 4 个指标: 最优值 (Best)、最差值 (Worst)、平均值 (Mean) 和标准方差 (Standard Deviation, STD), 评判所有对比算法的寻优精度. 其中, 最优值反映算法最好的寻优精度, 最差值反映最差的寻优精度, 平均值反映算法的平均寻优精度, 标准方差反映算法的寻优稳定性. 所有实验结果中的最优结果均用粗体表示.

表 1 5 种优化算法在 10 个基准测试函数上的 4 种指标结果对比

Tab. 1 Comparison results of five optimization algorithms on ten benchmark functions in four metrics

函数	指标	WOA	IWOA1	IWOA2	IWOA3	IWOA
$f_1$ (Sphere)	Best	6.53E - 87	<b>0.00E + 00</b>	2.69E - 224	1.03E - 234	<b>0.00E + 00</b>
	Worst	3.67E - 73	<b>0.00E + 00</b>	2.28E - 193	1.54E - 189	<b>0.00E + 00</b>
	Mean	2.05E - 74	<b>0.00E + 00</b>	9.12E - 195	5.63E - 191	<b>0.00E + 00</b>
	std	7.50E - 74	<b>0.00E + 00</b>	<b>0.00E + 00</b>	<b>0.00E + 00</b>	<b>0.00E + 00</b>
$f_2$ (Schwefel2.22)	Best	9.94E - 60	1.01E - 298	1.03E - 234	1.80E - 148	<b>0.00E + 00</b>
	Worst	1.09E - 50	5.81E - 254	1.54E - 189	3.80E - 129	<b>0.00E + 00</b>
	Mean	8.27E - 52	1.93E - 255	5.63E - 191	1.49E - 130	<b>0.00E + 00</b>
	std	2.47E - 51	<b>0.00E + 00</b>	<b>0.00E + 00</b>	6.96E - 130	<b>0.00E + 00</b>
$f_3$ (Schwefel2.21)	Best	8.27E - 01	1.75E - 120	5.04E - 150	2.76E - 127	<b>8.55E - 154</b>
	Worst	8.79E + 01	1.58E - 92	4.82E - 129	3.13E - 119	<b>2.83E - 133</b>
	Mean	5.46E + 01	6.96E - 94	1.77E - 130	1.28E - 122	<b>1.35E - 134</b>
	std	2.51E + 01	3.00E - 93	8.80E - 130	5.72E - 124	<b>5.31E - 134</b>
$f_4$ (Schwefel1.2)	Best	9.37E + 03	0.00E + 00	2.94E - 223	8.19E - 230	<b>0.00E + 00</b>
	Worst	6.13E + 04	9.87E + 04	3.26E - 192	1.01E - 196	<b>2.96E - 247</b>
	Mean	4.44E + 04	3.43E + 03	1.09E - 193	4.88E - 198	<b>1.65E - 249</b>
	std	1.13E + 04	1.80E + 04	<b>0.00E + 00</b>	<b>0.00E + 00</b>	<b>0.00E + 00</b>

#### 算法 2 IWOAFCM 算法

输入: 数据集

输出: 目标函数  $f$ 、隶属度矩阵  $u$  和聚类中心  $Q$

1. 设定聚类数目  $c$ 、迭代次数  $T$ 、模糊指数  $m$  和阈值  $\theta$
2. 利用算法 1 初始化 FCM 中心
3. while( $t < T$ )
4. for 任意样本
5. 根据式 (11) 更新隶属度矩阵  $u$
6. 根据式 (12) 更新聚类中心  $Q$
7. 根据式 (10) 计算目标函数  $f$
8. if  $\|f(t) - f(t-1)\| \leq \theta$
9. break /\* 跳出循环结束聚类 \*/
10. end while
11. return 隶属度矩阵  $u$  和聚类中心  $Q$

续表 1

函数	指标	WOA	IWOA1	IWOA2	IWOA3	IWOA
$f_5$ (Quartic)	Best	2.56E-04	6.57E-06	1.30E-05	2.21E-05	<b>8.58E-07</b>
	Worst	8.60E-03	9.86E-04	1.30E-03	3.00E-03	<b>3.88E-04</b>
	Mean	2.70E-03	2.84E-04	2.88E-04	4.82E-04	<b>8.81E-05</b>
	std	2.50E-03	2.51E-04	3.38E-04	6.80E-04	<b>1.06E-04</b>
$f_6$ (Rastrigin)	Best	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	Worst	5.68E-14	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	Mean	1.89E-15	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	std	1.03E-14	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
$f_7$ (Ackley)	Best	<b>8.88E-16</b>	<b>8.88E-16</b>	<b>8.88E-16</b>	<b>8.88E-16</b>	<b>8.88E-16</b>
	Worst	7.99E-15	<b>8.88E-16</b>	7.36E-15	<b>8.88E-16</b>	<b>8.88E-16</b>
	Mean	4.79E-15	<b>8.88E-16</b>	3.73E-15	<b>8.88E-16</b>	<b>8.88E-16</b>
	std	2.35E-15	<b>0.00E+00</b>	2.16E-15	<b>0.00E+00</b>	<b>0.00E+00</b>
$f_8$ (Griewank)	Best	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	Worst	1.28E-01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	Mean	4.30E-03	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	std	2.35E-02	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
$f_9$ (Rosenbrock)	Best	2.69E+01	2.69E+01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>
	Worst	2.87E+01	2.87E+01	2.88E+01	2.88E+01	<b>2.87E+01</b>
	Mean	2.80E+01	2.84E+01	1.01E+01	1.33E+01	<b>9.53E+00</b>
	std	4.79E-01	6.68E-01	2.33E+01	1.87E+01	<b>1.37E+01</b>
$f_{10}$ (Goldstein - Price Function)	Best	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>
	Worst	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>
	Mean	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>
	std	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>

由表 1 可以看出,3 种改进策略均可以提高标准 WOA 算法的性能. 具体来说,IWOA1 算法因 Logistics-Tent 混沌映射策略使种群均匀分布,在  $f_1$   $f_2$   $f_6$   $f_7$   $f_8$  和  $f_{10}$  6 个基准测试函数上的 4 种指标结果优于 WOA,提高了 WOA 的全局搜索性能;IWOA2 算法因自适应权重限制了鲸鱼搜索方向,相较于 WOA,除在  $f_3$  和  $f_5$  两个基准测试函数上 4 个指标提升不明显外,其他 8 个基准测试函数均优于 WOA,使在搜索过程中能更好地平衡全局搜索和局部搜索;IWOA3 算法因最优扰动策略在算法搜索完毕后加入扰动公式,相较于 WOA,除在  $f_2$   $f_3$  和  $f_5$  这 3 个基准测试函数上 4 个指标提升不明显外,其他 7 个基准测试函数均优于 WOA,相较于 WOA 避免了算法早熟问题;而 IWOA 算法为三者改进结合,所以在表 1 的 10 个基准测试函数上均为最优,3 种改进策略的结合增强了算法的寻优精度和鲁棒性.

接下来,选择 8 种最新的优化算法与 IWOA 算法进行优化性能的对比分析,主要包括:蜉蝣优化算法 (Mayfly Optimization Algorithm, MA)<sup>[16]</sup>、人工蜂鸟算法 (Artificial Hummingbird Algorithm, AHA)<sup>[17]</sup>、阿基米德优化算法 (Arithmetic Optimization Algorithm, AOA)<sup>[18]</sup>、海洋捕食者算法 (Marine Predators Algorithm, MPA)<sup>[19]</sup>、哈里斯鹰算法 (Harris Hawks Optimization, HHO)<sup>[20]</sup>、酶作用优化算法 (Enzyme Action Optimizer, EAO)<sup>[21]</sup>、座头鲸优化算法 (Whale Migration Algorithm, WMA)<sup>[22]</sup> 和 WOA 算法<sup>[14]</sup>. 这 8 种对比算法的参数均取自其发表的论文中,本文参数参考 WOA 算法<sup>[14]</sup> 设置. 表 2 为上述 9 种算法在 10 个基准测试函数上的 4 种指标的结果对比. 由表 2 可知,IWOA 算法与其他 8 种算法相比,其优化性能较好,在  $f_1 \sim f_3$   $f_6 \sim f_8$  和  $f_{10}$  这 7 种函数上,其寻优性能均优于其他 8 种算法;在  $f_4$  函数上,IWOA 算法的最优值虽然最好,但在最差值、平均值和标准差上低于 AHA 算法,说明 IWOA 算法稳定性较差,容易陷入局部最优;在  $f_5$  函数上,IWOA 算法的最差值和方差略低于 AOA 算法;在  $f_9$  函数上,IWOA 算法最优值最好,但在最差值、平均值和标准差上低于 HHO 算法,算法仍存在早熟现象. 这些结果说明 IWOA 算法的寻优精度有了显著提高,但在基准函数  $f_5$  和  $f_9$  上的寻优效果不稳定,IWOA 算法仍然存在陷入局部最优的问题. 但从整体上来看,IWOA 算法寻优精度和稳定性均高于其他算法.

表 2 9 种优化算法在 10 个基准测试函数上的 4 种指标结果对比

Tab. 2 Comparison results of nine optimization algorithms on ten benchmark functions in four metrics

函数	指标	MA	AHA	AOA	MPA	HHO	EAO	WMA	WOA	IWOA
$f_1$	Best	3.50E-10	3.98E-161	5.97E-139	3.72E-25	6.56E-121	<b>0.00E+00</b>	3.62E-11	6.53E-87	<b>0.00E+00</b>
	Worst	8.31E-06	1.39E-134	1.15E-27	4.58E-22	9.35E-93	<b>0.00E+00</b>	4.00E-06	3.67E-73	<b>0.00E+00</b>
	Mean	3.45E-07	4.65E-136	3.84E-29	5.29E-23	3.18E-94	<b>0.00E+00</b>	2.34E-07	2.05E-74	<b>0.00E+00</b>
	std	1.46E-06	2.21E-135	2.24E-28	7.49E-23	1.58E-93	<b>0.00E+00</b>	6.69E-07	7.50E-74	<b>0.00E+00</b>
$f_2$	Best	3.01E-07	1.63E-83	<b>0.00E+00</b>	1.69E-14	9.61E-58	3.13E-263	5.03E-07	9.94E-60	<b>0.00E+00</b>
	Worst	5.40E-03	2.57E-72	<b>0.00E+00</b>	6.09E-13	9.68E-48	3.08E-213	3.76E-04	1.09E-50	<b>0.00E+00</b>
	Mean	4.28E-04	1.91E-73	<b>0.00E+00</b>	2.31E-13	6.25E-49	1.06E-214	3.33E-05	8.27E-52	<b>0.00E+00</b>
	std	1.20E-03	5.93E-73	<b>0.00E+00</b>	1.67E-13	2.37E-48	<b>0.00E+00</b>	1.05E-04	2.47E-51	<b>0.00E+00</b>
$f_3$	Best	5.54E-01	1.18E-74	2.77E-60	2.99E-10	1.96E-56	4.33E-315	1.18E-06	8.27E-01	<b>0.00E+00</b>
	Worst	4.32E+00	5.69E-63	4.70E-02	6.73E-09	8.25E-48	524E-292	1.59E-04	8.79E+01	<b>0.00E+00</b>
	Mean	1.73E+00	3.75E-64	2.88E-02	3.44E-09	3.06E-49	174E-293	3.63E-05	5.46E+01	<b>0.00E+00</b>
	std	9.31E-01	1.17E-63	1.90E-02	1.84E-09	1.50E-48	<b>0.00E+00</b>	4.00E-05	2.51E+01	<b>0.00E+00</b>
$f_4$	Best	1.70E-02	5.42E-147	1.16E-159	5.30E-08	8.28E-101	<b>0.00E+00</b>	8.30E-05	9.37E+03	<b>0.00E+00</b>
	Worst	9.28E-01	1.15E-119	6.24E-02	5.01E-04	1.12E-75	2.71E-188	2.66E-01	6.13E+04	<b>2.96E-247</b>
	Mean	2.26E-01	3.83E-121	6.10E-03	1.13E-04	3.73E-77	9.04E-190	1.28E-02	4.44E+04	<b>1.65E-249</b>
	std	2.05E-01	2.09E-120	1.42E-02	1.55E-04	2.04E-76	<b>0.00E+00</b>	4.84E-02	1.13E+04	<b>0.00E+00</b>
$f_5$	Best	1.05E-02	3.67E-05	5.86E-06	3.00E-04	1.47E-06	5.25E-05	1.40E-03	2.56E-04	<b>8.58E-07</b>
	Worst	3.32E-02	5.80E-04	<b>1.51E-04</b>	3.00E-03	4.87E-04	8.37E-04	1.57E-02	8.60E-03	3.88E-04
	Mean	1.90E-02	2.52E-04	6.20E-05	1.00E-03	1.19E-04	3.17E-04	9.20E-03	2.70E-03	<b>8.81E-05</b>
	std	5.80E-03	1.46E-04	<b>4.72E-05</b>	6.32E-04	1.00E-04	2.20E-04	3.60E-03	2.50E-03	1.06E-04
$f_6$	Best	5.98E+00	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	1.19E+01	<b>0.00E+00</b>	<b>0.00E+00</b>
	Worst	2.89E+01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	1.51E+02	5.68E-14	<b>0.00E+00</b>
	Mean	1.54E+01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	1.22E+01	1.89E-15	<b>0.00E+00</b>
	std	5.70E+00	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	3.71E+01	1.03E-14	<b>0.00E+00</b>
$f_7$	Best	2.31E+00	<b>8.88E-16</b>	<b>8.88E-16</b>	3.49E-13	<b>8.88E-16</b>	<b>8.88E-16</b>	3.99E-15	<b>8.88E-16</b>	<b>8.88E-16</b>
	Worst	5.22E+00	<b>8.88E-16</b>	<b>8.88E-16</b>	4.24E-12	<b>8.88E-16</b>	<b>8.88E-16</b>	1.34E+00	7.99E-15	<b>8.88E-16</b>
	Mean	3.56E+00	<b>8.88E-16</b>	<b>8.88E-16</b>	1.64E-12	<b>8.88E-16</b>	<b>8.88E-16</b>	1.82E-01	4.79E-15	<b>8.88E-16</b>
	std	8.02E-01	<b>0.00E+00</b>	<b>0.00E+00</b>	9.58E-13	<b>0.00E+00</b>	<b>0.00E+00</b>	4.23E-01	2.35E-15	<b>0.00E+00</b>
$f_8$	Best	1.34E-07	<b>0.00E+00</b>	4.32E-02	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	0.00E+00	<b>0.00E+00</b>	<b>0.00E+00</b>
	Worst	3.20E-02	<b>0.00E+00</b>	4.60E-01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	3.43E-02	1.28E-01	<b>0.00E+00</b>
	Mean	1.34E-02	<b>0.00E+00</b>	1.85E-01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	5.60E-03	4.30E-03	<b>0.00E+00</b>
	std	9.90E-03	<b>0.00E+00</b>	1.20E-01	<b>0.00E+00</b>	<b>0.00E+00</b>	<b>0.00E+00</b>	9.20E-03	2.35E-02	<b>0.00E+00</b>
$f_9$	Best	1.72E+01	2.46E+01	2.79E+01	2.46E+01	2.23E-04	2.50E+01	3.22E+00	2.69E+01	<b>0.00E+00</b>
	Worst	1.58E+02	2.61E+01	2.89E+01	2.62E+01	<b>1.05E-01</b>	3.69E+01	1.03E+02	2.87E+01	2.87E+01
	Mean	4.23E+01	2.67E+01	2.84E+01	2.53E+01	<b>1.45E-02</b>	2.64E+01	1.62E+01	2.80E+01	9.53E+00
	std	3.59E+01	4.74E-01	2.68E-01	3.78E-01	<b>2.34E-02</b>	5.79E-01	2.14E+01	4.79E-01	1.37E+01
$f_{10}$	Best	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	<b>3.00E+00</b>	3.00E+00	3.00E+00	<b>3.00E+00</b>	<b>3.00E+00</b>
	Worst	8.40E+01	<b>3.00E+00</b>	7.29E+01	<b>3.00E+00</b>	<b>3.00E+00</b>	3.00E+00	3.00E+00	<b>3.00E+00</b>	<b>3.00E+00</b>
	Mean	5.70E+00	<b>3.00E+00</b>	2.78E+01	<b>3.00E+00</b>	<b>3.00E+00</b>	3.00E+00	3.00E+00	<b>3.00E+00</b>	<b>3.00E+00</b>
	std	1.47E+01	<b>3.00E+00</b>	2.10E+01	<b>3.00E+00</b>	<b>3.00E+00</b>	3.00E+00	3.00E+00	<b>3.00E+00</b>	<b>3.00E+00</b>

为了验证 IWOA 算法的收敛性,图 3 展示了 10 个基准函数在 9 种优化算法上的寻优收敛曲线,其中横轴为迭代次数,纵轴为算法的最佳适应度值.由图 3 可知,在  $f_1 \sim f_7$  和  $f_9$  这 8 个基准函数上,IWOA 算法收敛速度较快,而且寻优效果比其他 5 种算法好,有效减少了迭代次数;在  $f_8$  和  $f_{10}$  上,IWOA 算法会陷入局部最优,分别在迭代 200 次和 400 次时达到最优值;在  $f_4 \sim f_6$  函数上,IWOA 算法收敛速度更快,分别在迭代 10 次、20 次和 40 次时达到最优值;在  $f_1$  和  $f_2$  函数上,IWOA 算法的收敛速度基本一致,在迭代 450 次时可以找到基准函数的最优值;在  $f_5$  函数上,IWOA 算法的收敛速度虽然快于其他优化算法,但是最后寻找到的最优值却与 AHA 和 HHO 这 2 种算法持平,这是由于搜索范围扩大造成的.从整体上来看,IWOA 算

法的收敛速度优于其他算法,而且能有效减少迭代次数.

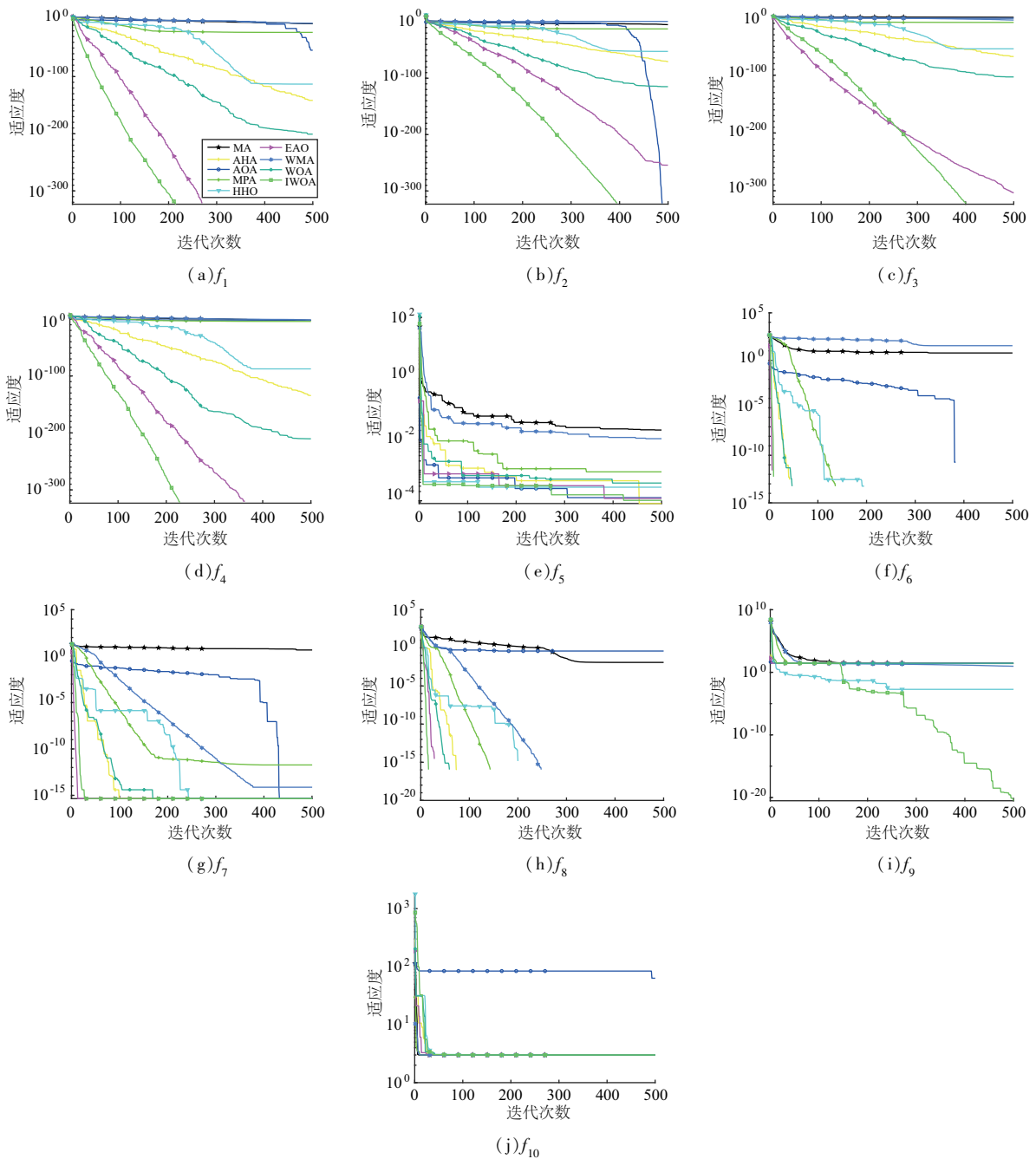


图 3 9 种优化算法在 10 个基准函数上的收敛曲线

Fig. 3 Convergence curves of 9 optimization algorithms on 10 benchmark functions

### 3.2 人工数据集上的聚类效果实验

为更全面地探讨 IWOAFCM 算法的聚类性能,参照文献[14],实验选取 5 个具有不同特征的二维人工数据集作为实验数据集.其中,Aggregation 数据集样本量 788、类数为 7;R15 数据集样本量 1500、类数为 15;Compound 数据集样本量 499、类数为 6;Jain 数据集样本量 373、类数为 2;Flame 数据集样本量 240、类数为 2.在聚类效果对比实验中,采用的对比算法包括 K-means 算法<sup>[23]</sup>、FCM 算法<sup>[5]</sup>、基于动态惯性权重改进粒子群算法优化的 FCM 算法(Fuzzy C-means with Dynamic Particle Swarm Optimization algorithm, DP-SOFCM)<sup>[24]</sup>、改进樽海鞘群优化的模糊 C-均值聚类算法(Fuzzy C-mean Clustering Algorithm Based on

Improved Salp Swarm Optimization, ISSAFCM)<sup>[25]</sup>、自适应遗传优化的模糊 C-均值聚类算法(A GA-Optimized Fuzzy Clustering Approach, GAAFCM)<sup>[26]</sup>以及结合标准的 WOA<sup>[11]</sup>和 FCM<sup>[5]</sup>的聚类算法(Fuzzy C-means clustering algorithms optimized by WOA, WOAFCM)。

参照文献[26-28],本实验采用准确率(Accuracy, ACC)<sup>[26]</sup>、标准化互信息(Normalized Mutual Information, NMI)<sup>[27]</sup>和调整兰德指数(Adjusted Rand Index, ARI)<sup>[28]</sup>作为衡量聚类效果的指标.其中,ACC 用于衡量聚类正确分类的样本数与样本总数的比例关系,该指标数值越高,表明聚类算法的精度水平越高;NMI 通过精准计算聚类结果和真实类别信息之间的相似性来评估聚类效果,其数值越大,意味着聚类效果越理想;ARI 通过比较聚类结果与真实类别信息之间的一致性来评价聚类效果,其数值越大,算法的性能表现越好.7 种聚类算法在 5 个人工数据集上的聚类效果如表 3 所示.

表 3 7 种聚类算法在 5 个人工数据集上的聚类结果

Tab. 3 Clustering results of seven clustering algorithms on five artificial datasets

聚类指标	算法	Aggregation/%	R15/%	Compound/%	Jain/%	Flame/%
ACC	K-means	74.57	88.47	74.12	80.37	82.37
	FCM	74.82	89.35	67.49	88.41	89.63
	DPSOFCM	87.65	92.75	82.36	92.63	91.42
	ISSAFCM	<b>99.24</b>	99.62	<b>88.47</b>	94.37	95.43
	GAAFCM	87.68	96.53	78.48	88.65	90.86
	WOAFCM	88.46	90.12	70.22	88.95	90.36
	IWOAFCM	98.42	<b>100.00</b>	84.74	<b>97.24</b>	<b>97.32</b>
NMI	K-means	75.71	96.41	72.41	43.32	55.47
	FCM	77.63	94.47	71.52	55.46	62.38
	DPSOFCM	85.71	96.85	82.32	62.17	62.57
	ISSAFCM	<b>99.64</b>	99.42	<b>86.18</b>	65.08	72.86
	GAAFCM	86.17	96.42	72.31	63.16	62.37
	WOAFCM	92.07	96.44	71.53	60.12	62.36
	IWOAFCM	93.79	<b>99.83</b>	81.57	<b>84.33</b>	<b>85.65</b>
ARI	K-means	84.53	97.53	85.64	59.37	79.63
	FCM	88.64	98.62	85.57	74.63	80.76
	DPSOFCM	90.24	99.13	90.32	88.43	83.57
	ISSAFCM	<b>99.87</b>	99.91	<b>93.51</b>	89.35	91.24
	GAAFCM	89.79	97.48	86.26	79.25	81.63
	WOAFCM	93.95	97.48	85.67	79.06	84.37
	IWOAFCM	95.64	<b>99.93</b>	89.52	<b>89.72</b>	<b>93.69</b>

通过表 3 可以看出,IWOAFCM 算法在 5 个数据集上的各项指标均处于较为领先的位置,尤其在 R15、Jain 和 Flame 这 3 个数据集上,该算法的聚类效果提升明显,展现了该算法在全局探索和局部开发的优越性,提高了聚类的聚类准确率和效率,在所有对比算法中综合性能最好;在 Compound 和 Aggregation 这 2 个数据集上,IWOAFCM 算法在 3 个评价指标上略低于 ISSAFCM 算法,IWOAFCM 算法在处理复杂的人工数据结构时泛化能力略低于 ISSAFCM 算法. Aggregation 数据集密度不均且噪声点较多,IWOAFCM 的鲁棒性低于 ISSAFCM,导致 IWOAFCM 在 Aggregation 数据集上的评价指标略低于 ISSAFCM. Compound 数据集包含多个非球形聚类,WOA 主要依赖螺旋更新和气泡网攻击,在边界重叠区域反复震荡,导致聚类中心偏离.综合比较 7 种算法在 5 个数据集上的聚类结果可知,在处理复杂的人工数据集时,IWOAFCM 算法的聚类性能优于其他算法,能够取得更好的聚类效果.

为了更直观地对比 7 种算法的聚类效果,选取 Aggregation 数据集进行可视化展示.图 4 呈现了这 7 种聚类算法在 Aggregation 数据集上的聚类结果.IWOAFCM 算法由于 IWOA 的引入缓解了 FCM 算法对初始聚类中心选取的影响,有效减少了样本点的错误识别,提高了算法的准确率和效率.相比之下,其他 5 种对比算法均无法实现类簇边界的精准界定以及同类样本点的正确识别.综合看来,IWOAFCM 算法在聚类效果和聚类稳定性上均优于其他 6 种对比算法,聚类结果更接近样本点的真实分布.

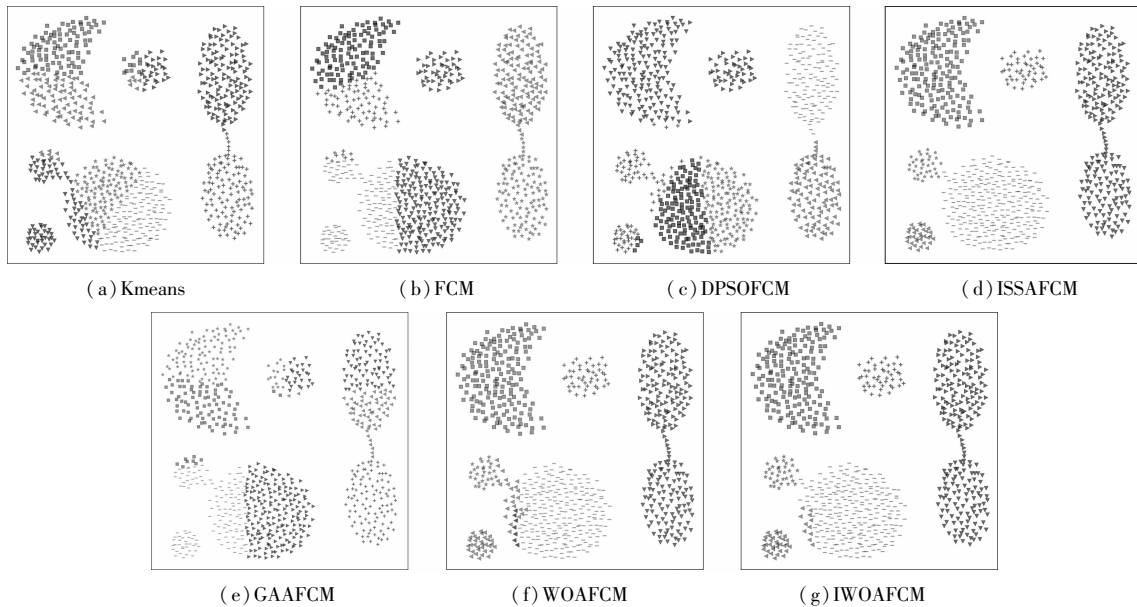


图 4 7 种聚类算法在 Aggregation 数据集上的聚类效果

Fig. 4 Clustering results of 7 clustering algorithms on the Aggregation dataset

### 3.3 UCI 数据集上的聚类效果实验

本次实验选取 5 个 UCI 数据集,采用 ACC、NMI 和 ARI 作为衡量聚类质量的评价指标,进一步验证 IWOAFCM 算法的聚类性能.采用的对比算法包括 K-means 算法<sup>[23]</sup>、FCM 算法<sup>[5]</sup>、DPSOFCM 算法<sup>[24]</sup>、ISSAFCM 算法<sup>[25]</sup>、GAAFCM 算法<sup>[26]</sup>和 WOAFCM 算法<sup>[14]</sup>.表 4 给出了 5 个 UCI 数据的信息描述,表 5 为 7 种聚类算法在 5 个 UCI 数据集上的聚类结果.

由表 5 可知,在 Heart Disease 数据集上,IWOAFCM 算法在 ACC、NMI 和 ARI 3 种指标上运行效果均最优.在 Iris 数据集上,IWOAFCM 算法的 ACC 取得了最优;在 Glass 数据集上,IWOAFCM 算法的 ACC 和 ARI 都取得了最优,而在 NMI 指标上 IWOAFCM 算法表现为次优,算法捕捉到的信息低于真实类别的信息复杂度,对数据内在结构的理解弱于 ISSAFCM;在 Seeds 数据集上,IWOAFCM 算法因其优异的全局搜索和局部开发能力使算法在 ACC 和 ARI 指标上显著优于其他算法;在 Lung cancer 数据集上,IWOAFCM 算法的 ACC 表现不佳,其原因为 IWOAFCM 算法的全局探测能力略弱于 ISSAFCM 算法,但在 NMI 和 ARI 指标上都取得了最优.综上,在这 5 个数据集上,与另外 6 种聚类算法相比,IWOAFCM 的聚类结果基本上都能达到最优.通过实验和文献分析可以看出,本文算法平衡了 WOA 的全局搜索能力和局部开发能力,缓解了算法早熟和 FCM 对聚类中心过于敏感的问题,提高了 FCM 的聚类效果.

### 3.4 IWOAFCM 算法的聚类分析结果

为验证 IWOAFCM 算法的有效性,在 5 个人工数据集和 5 个 UCI 数据集上,将 K-means 算法<sup>[23]</sup>、FCM 算法<sup>[5]</sup>、DPSOFCM 算法<sup>[24]</sup>、ISSAFCM 算法<sup>[25]</sup>、GAAFCM 算法<sup>[26]</sup>、WOAFCM 算法<sup>[14]</sup>和 IWOAFCM 算法分别以 ACC、NMI 和 ARI 为评价指标,进行聚类效果测试,测试效果如图 5 所示.

由图 5 可知,图 5(a)中的 IWOAFCM 算法除 Compound、Lung cancer 和 Aggregation 这 3 个数据集外,在其他 7 个数据集上的 ACC 均是最优的,IWOAFCM 算法在 3 个数据集上未实现正确的聚类划分,算法在数据分布不均的数据集上仍存在过早收敛的情况.在图 5(b)的 NMI 效果图和图 5(c)的 ARI 效果图中,除在 Compound、Iris 和 Aggregation 3 个数据集上,IWOAFCM 算法的 NMI 和 ARI 指标效果提升不明显外,

表 4 5 个 UCI 数据集的信息描述

Tab. 4 Information description of 5 UCI datasets

数据集	样本数量	特征维度	类别数量
Iris	150	4	3
Glass	214	9	6
Seeds	210	7	3
Lung cancer	32	56	3
Heart Disease	270	13	2

其他 7 个数据集上 IWOAFCM 算法的 NMI 和 ARI 均是最优的. 其原因是 IWOAFCM 算法在数据分布不均的数据集上聚类效果弱于 ISSAFCM 算法, 导致 IWOAFCM 算法在这 3 个数据集上的 NMI 和 ARI 表现不佳. 从综合比较结果可以看出, IWOAFCM 算法的聚类效果和聚类稳定性均是最优的.

表 5 7 种聚类算法在 5 个 UCI 数据集上的聚类结果

Tab. 5 Clustering results of seven clustering algorithms on five UCI datasets

聚类指标	算法	Iris/%	Glass/%	Seeds/%	Lung cancer/%	Heart Disease/%
ACC	K - means	73.35	47.15	80.33	65.16	30.17
	FCM	78.65	50.43	78.21	68.46	32.46
	DPSOFCM	90.01	63.37	90.16	74.26	39.42
	ISSAFCM	96.67	68.69	93.48	<b>81.25</b>	44.07
	GAAFCM	92.00	53.68	81.56	70.13	35.68
	WOAFCM	89.33	49.40	89.88	73.50	41.66
	IWOAFCM	<b>96.73</b>	<b>70.13</b>	<b>93.74</b>	77.75	<b>50.82</b>
NMI	K - means	58.66	45.17	66.37	10.62	14.06
	FCM	63.84	50.32	62.51	14.82	15.27
	DPSOFCM	79.63	55.25	74.17	18.26	15.95
	ISSAFCM	<b>88.01</b>	<b>64.34</b>	75.29	30.73	18.98
	GAAFCM	70.36	50.13	67.49	15.97	15.04
	WOAFCM	77.00	51.42	68.18	18.39	15.11
	IWOAFCM	81.26	61.32	<b>76.86</b>	<b>31.53</b>	<b>19.45</b>
ARI	K - means	78.86	56.73	83.57	59.29	57.14
	FCM	80.15	75.37	84.36	59.78	60.33
	DPSOFCM	84.32	77.93	90.11	60.32	61.96
	ISSAFCM	<b>95.76</b>	83.79	<b>91.02</b>	68.55	63.40
	GAAFCM	81.36	77.45	85.31	60.72	61.06
	WOAFCM	77.31	76.22	86.62	60.13	61.83
	IWOAFCM	85.61	<b>83.81</b>	88.90	<b>71.53</b>	<b>63.90</b>

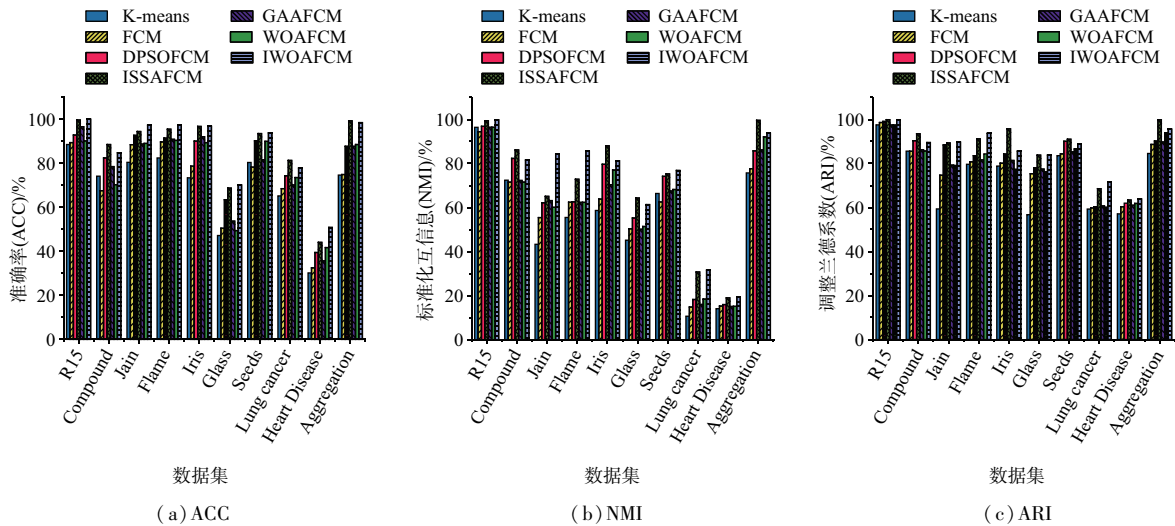


图 5 7 种聚类算法在 3 个指标下的聚类效果

Fig. 5 Clustering effect test of 7 clustering algorithms under three metrics

综上所述,通过采用 Logistics - Tent 混沌映射初始化种群,加入自适应权重来限制领头鲸鱼的空间搜索范围和最优邻域扰动策略改进的 WOA 与 8 种优化算法在 10 个基准测试函数上的 4 种指标结果对比可知,除  $f_0$  在基准测试函数上不是最优外,其他 9 个基准测试函数均是最优. 利用改进的 WOA 优化 FCM 的初始聚类中心,并将 FCM 的隶属度矩阵进行重构后的 IWOAFCM 与另外 6 种聚类算法分别在 5 个人工数据集和 5 个 UCI 数据集上进行聚类实验,从实验结果可以看出, IWOAFCM 除了在 Aggregation 和 Com-

pound 数据集上不是最优外,在其他数据集上均是最优. IWOAFCM 虽然在部分数据集上不是最优,算法仍存在过早收敛的问题,但综合对比来看, IWOAFCM 的全局搜索能力、局部开发能力和稳定性是最优的.

## 4 结论

本文提出了一种基于改进 WOA 的 FCM 算法. 该算法通过引入 Logistics - Tent 混沌映射、自适应权重和邻域扰动策略,全面提升了 WOA 的迭代速度与寻优精度,将隶属度矩阵进行重构用于对 WOA 种群的初始化,并生成 FCM 算法的初始聚类中心,通过多个单峰和多峰函数进行测试. 测试结果表明:相较于改进前的 WOA, IWOA 在收敛精度和速度上均取得显著提升;在 5 个人工数据集和 5 个 UCI 数据集上,通过对比多种先进的聚类算法的聚类效果, IWOAFCM 算法在多项指标上的表现均最优. 未来将持续探索如何提升算法在高维数据集上的聚类效果,进而设计鲁棒性更强、泛化能力更高的聚类算法.

## 参考文献:

- [1] 阮永芬,李鹏辉,王必顺,等. 基于主成分与聚类分析滇池湖相粉土物理力学参数[J]. 昆明理工大学学报(自然科学版),2023,48(3):41-52.  
RUAN Y F, LI P H, WANG B S, et al. Investigation of physical and mechanical parameters of lacustrine facies silty soils in Dianchi Lake based on principal component and cluster[J]. Journal of Kunming University of Science and Technology (Natural Science), 2023, 48(3): 41-52.
- [2] 薛占熬,杨梦丽,辛现伟,等. 基于优势关系的多粒度 Pythagorean 模糊粗糙集的粒度选择方法[J]. 河南师范大学学报(自然科学版),2025,53(3):79-88.  
XUE Z A, YANG M L, XIN X W, et al. Granularity selection method of multi-granularity Pythagorean fuzzy rough set based on dominance relationship[J]. Journal of Henan Normal University (Natural Science Edition), 2025, 53(3): 79-88.
- [3] 张玉杰,高晗. 基于改进 FCM 的冲件缺陷图像分割算法[J]. 计算机工程,2024,50(10):342-351.  
ZHANG Y J, GAO H. Image segmentation algorithm for stamping defects based on improved FCM[J]. Computer Engineering, 2024, 50(10): 342-351.
- [4] 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 J Q, WANG J M. 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.
- [5] 何庆龄,刘静,王昌锋,等. 基于混合策略改进 SSA - FCM 的快速路交通流数据修复研究[J]. 重庆交通大学学报(自然科学版),2025,44(12):53-61+71.  
HE Q L, LIU J, WANG C F, et al. Expressway traffic flow data repair based on hybrid strategy improved SSA - FCM[J]. Journal of Chongqing Jiaotong University (Natural Science), 2025, 44(12): 53-61+71.
- [6] ZHANG T, TAO Q, YU L J, et al. A new prediction strategy for dynamic multi-objective optimization using hybrid Fuzzy C - Means and support vector machine[J]. Neurocomputing, 2025, 621: 129291.
- [7] SUN C K, SHAO Q L, ZHOU Z Q, et al. An enhanced FCM clustering method based on multi-strategy tuna swarm optimization[J]. Mathematics, 2024, 12(3): 453.
- [8] 孟学尧,郭倩倩,郭海儒. 一种改进萤火虫算法的模糊聚类方法[J]. 小型微型计算机系统,2021,42(6):1165-1170.  
MENG X Y, GUO Q Q, GUO H R. Fuzzy clustering method based on improved firefly algorithm[J]. Journal of Chinese Computer Systems, 2021, 42(6): 1165-1170.
- [9] 李清. 基于改进 PSO - PFCM 聚类算法的电力大数据异常检测方法[J]. 电力系统保护与控制,2021,49(18):161-166.  
LI Q. Power big data anomaly detection method based on an improved PSO - PFCM clustering algorithm[J]. Power System Protection and Control, 2021, 49(18): 161-166.
- [10] 王梦绚,万仁霞,苗夺谦,等. 基于三支决策的蚁群聚类算法[J]. 昆明理工大学学报(自然科学版),2024,49(1):83-97.  
WANG M X, WAN R X, MIAO D Q, et al. An ant colony clustering algorithm based on three-way decision[J]. Journal of Kunming University of Science and Technology (Natural Science), 2024, 49(1): 83-97.
- [11] 夏正龙,陈宇,陆良帅,等. 基于多目标鲸鱼算法的配电网动态无功优化研究[J]. 河南师范大学学报(自然科学版),2025,53(1):116-126.  
XIA Z L, CHEN Y, LU L S, et al. Research on dynamic reactive power optimization of distribution network based on multi-objective

- whale optimization algorithm[J]. Journal of Henan Normal University (Natural Science Edition), 2025, 53(1): 116 - 126.
- [12] ZHANG B Y, ZHANG C X, YANG Y, et al. A scenario - driven bi - level programming model for optimal distributed generation planning: Integrating fuzzy C - means (FCM) clustering and enhanced whale optimization (EWOA) [C]//2024 4th International Conference on Smart City and Green Energy (ICSCGE), 10 - 13 December 2024, Sydney, Australia. IEEE, 2024: 207 - 212.
- [13] 杜晓昕,牛翔慧,王波,等. 基于多策略改进的金豺优化算法[J]. 河南师范大学学报(自然科学版), 2025, 53(4): 39 - 50.  
DU X X, NIU X H, WANG B, et al. A modified golden jackal optimization algorithm based on multiple strategies[J]. Journal of Henan Normal University (Natural Science Edition), 2025, 53(4): 39 - 50.
- [14] 冯琼,谢晓扬,王鹏辉,等. 基于鲸鱼优化算法 - 反向传播神经网络的钢筋混凝土耐久性预测[J]. 吉林大学学报(工学版), 2025, 55(7): 2276 - 2285.  
FENG Q, XIE X Y, WANG P H, et al. Prediction of reinforced concrete durability based on whale optimization algorithm - back propagation neural network[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 55(7): 2276 - 2285.
- [15] 孙林,蔡怡文. 卷积神经网络与人工水母搜索的图特征选择方法[J]. 南京大学学报(自然科学), 2023(5): 759 - 769.  
SUN L, CAI Y W. Convolutional Neural Network and Artificial Jellyfish Search - based graph feature selection method[J]. Journal of Nanjing University (Natural Science), 2023(5): 759 - 769.
- [16] ZERVOUDAKIS K, TSAFARAKIS S. A mayfly optimization algorithm [J]. Computers & Industrial Engineering, 2020, 145: 106559.
- [17] ZHAO W G, WANG L Y, MIRJALILI S. Artificial hummingbird algorithm: A new bio - inspired optimizer with its engineering applications[J]. Computer Methods in Applied Mechanics and Engineering, 2022, 388: 114194.
- [18] ABUALIGAH L, DIABAT A, MIRJALILI S, et al. The arithmetic optimization algorithm [J]. Computer Methods in Applied Mechanics and Engineering, 2021, 376: 113609.
- [19] FARAMARZI A, HEIDARINEJAD M, MIRJALILI S, et al. Marine predators algorithm: A nature - inspired metaheuristic [J]. Expert Systems with Applications, 2020, 152: 113377.
- [20] HEIDARI A A, MIRJALILI S, FARIS H, et al. Harris Hawks optimization: Algorithm and applications [J]. Future Generation Computer Systems, 2019, 97: 849 - 872.
- [21] RODAN A, AL - TAMIMI A K, AL - ALNEMER L, et al. Enzyme action optimizer: A novel bio - inspired optimization algorithm [J]. The Journal of Supercomputing, 2025, 81(5): 686.
- [22] LU H G, CHENG S H, ZHANG X S. An improved whale migration algorithm for global optimization of collaborative symmetric balanced learning and cloud task scheduling [J]. Symmetry, 2025, 17(6): 841.
- [23] 贾鹤鸣,张棕淇,姜子超,等. 基于混合身份搜索黏菌优化的模糊 C - 均值聚类算法[J]. 智能系统学报, 2022, 17(5): 999 - 1011.  
JIA H M, ZHANG Z Q, JIANG Z C, et al. An optimization fuzzy C - means clustering algorithm based on the hybrid identity search and slime mold algorithms [J]. CAAI Transactions on Intelligent Systems, 2022, 17(5): 999 - 1011.
- [24] DHANACHANDRA N, CHANU Y J. An image segmentation approach based on fuzzy c - means and dynamic particle swarm optimization algorithm [J]. Multimedia Tools and Applications, 2020, 79(25): 18839 - 18858.
- [25] 张艳菊,陈泽荣,吴俊. 基于改进樽海鞘群优化的模糊 C - 均值聚类算法 [J/OL]. 控制工程, 1 - 16 [2026 - 01 - 12]. <https://doi.org/10.14107/j.cnki.kzgc.20230841>.  
ZHANG Y J, CHEN Z R, WU J. Fuzzy C - mean clustering algorithm based on improved salp swarm optimization [J/OL]. Control Engineering of China, 1 - 16 [2026 - 01 - 12]. <https://doi.org/10.14107/j.cnki.kzgc.20230841>.
- [26] XU T G, MA C, SU Z Z, et al. Enhancing cybersecurity in the big data era: A GA - optimized fuzzy clustering approach [J]. International Journal of Uncertainty, Fuzziness and Knowledge - Based Systems, 2025, 33(3): 353 - 377.
- [27] VINH N X, EPPS J, BAILEY J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance [J]. Journal of Machine Learning Research, 2010, 11(95): 2837 - 2854.
- [28] HE Z Y, XU X F, DENG S C. K - ANMI: A mutual information based clustering algorithm for categorical data [J]. Information Fusion, 2008, 9(2): 223 - 233.

( 编 辑:朱银周)