

# 基于势博弈的聚类算法

王田雨, 徐 勇

河北工业大学理学院, 天津

Email: name\_wangtianyu@qq.com, xuyong@hebut.edu.cn

收稿日期: 2021年1月17日; 录用日期: 2021年2月16日; 发布日期: 2021年2月23日

## 摘 要

针对聚类过程中初始聚类选取方法的不同会影响其最终聚类结果, 且传统的聚类算法很难得到稳定的聚类的问题, 提出一种新颖的以模糊 $c$ 均值作为初始聚类的基于势博弈的聚类算法(简称FS-IBRC)。该算法结合模糊 $c$ 均值算法, 以欧氏距离作为相似性度量, 不断迭代更新隶属度矩阵和聚类中心, 直至目标函数收敛从而得到较优的初始聚类结果。在此基础上, 将数据聚类问题转化为寻找势博弈模型的纯纳什均衡问题, 这对应于稳定的聚类。进而给出了解决这种基于势博弈模型聚类的算法, 即迭代的最佳响应算法。最后, 将FS-IBRC算法和一般初始化算法(S-IBRC)分别在两个不同的人工数据集上测试并实现。

## 关键词

模糊 $c$ 均值, 初始聚类, 势博弈, 纯纳什均衡, 迭代最佳响应

# Clustering Algorithm Based on Initial Clustering

Tianyu Wang, Yong Xu

School of Science, Hebei university of Technology, Tianjin

Email: name\_wangtianyu@qq.com, xuyong@hebut.edu.cn

Received: Jan. 17<sup>th</sup>, 2021; accepted: Feb. 16<sup>th</sup>, 2021; published: Feb. 23<sup>rd</sup>, 2021

## Abstract

Aiming at the problem that the different initial cluster selection methods in the clustering process will affect the final clustering results, and the traditional clustering algorithm is difficult to obtain stable clustering, a novel method using fuzzy  $c$ -means as the initial clustering is proposed. Clustering algorithm is based on potential game (Referred to as FS-IBRC). The algorithm combines the

文章引用: 王田雨, 徐勇. 基于势博弈的聚类算法[J]. 应用数学进展, 2021, 10(2): 461-470.

DOI: 10.12677/aam.2021.102052

fuzzy c-means algorithm with Euclidean distance as the similarity measure, and iteratively updates the membership matrix and clustering center until the objective function converges to obtain a better initial clustering result. On this basis, the data clustering problem is transformed into a pure Nash equilibrium problem of finding a potential game model, which corresponds to stable clustering. Furthermore, an algorithm for solving this clustering based on the potential game model is given, that is, the iterative optimal response algorithm. Finally, the FS-IBRC algorithm and the general initialization algorithm (S-IBRC) were tested and implemented on two different artificial data sets.

## Keywords

Fuzzy C-Means, Initial Clustering, Potential Game, Nash Equilibrium, Iterative Best-Response

Copyright © 2021 by author(s) and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

在处理数据对象时, 将数据按组分类成了最广泛的操作。在此过程中, 可能会出现两种情况: 一种是监督分类, 即每个组(或类)由代表类名的标签标识[1]。另一种则是在没有任何先验知识的情况下得到数据的分组, 这样的无监督分类过程称为聚类。聚类的目的是组内对象的相似性最大, 组间的相似性最小[2]。

聚类分析被广泛用于数据挖掘中, 因此, 大量的聚类算法应运而生[3]。聚类算法大致可分为基于层次的聚类算法[4] [5], 基于划分的聚类算法[6], 基于密度的聚类算法和基于网格的聚类算法[7] [8], 以及其他类型的聚类算法。其中, 模糊  $c$  均值算法[9]是属于基于划分的聚类算法的一种典型算法, 模糊  $c$  均值算法(Fuzzy C-Means, 简称 FCM)是根据隶属度来确定每个数据对象属于某个组别的程度的一种算法, 以欧氏距离作为相似性度量, 不断迭代更新隶属度矩阵和聚类中心, 直至目标函数收敛得到最终的聚类结果。如果把  $k$ -means 算法称为“硬聚类”算法, 那么 FCM 则是与其对应的“软聚类”算法, 其中每个数据样本可以属于多个集群。因其算法设计简单, 易实现, 且能够获得较优的聚类结果, 被广泛应用在生活中的各个方面。

聚类结果的稳定性, 成为了聚类质量所要参考的重要的衡量标准。在这种形式下, 一些基于博弈论的聚类算法不断地被提出, 这些算法提供了一种通过稳定性聚类的概念来解决聚类问题的新方法[10] [11] [12], 这种跨学科的新颖的方法日渐成为众多科研人员关心和研究的对象。基于博弈论的聚类算法可以分为基于合作博弈论[13]和非合作博弈论[14]的聚类算法。势博弈[15]属于非合作博弈的一种特殊形式, 因其始终存在纯纳什均衡[16], 从而被广泛地应用到各个领域。进一步, 我们将势博弈的纯纳什均衡看作是一个稳定的聚类, 那么, 可以解决聚类结果的相关收敛性问题。

随机性选取初始中心可能使聚类结果陷入局部最优且易受到噪点的影响, 进而使所得到的聚类结果并不是稳定的聚类。现在, 越来越多的科学研究者致力于研究如何优化初始聚类中心的选择。与初始中心相同, 本文所提出的算法首先要确定初始聚类的选择方法。该算法结合模糊  $c$  均值作为初始聚类, 形成一种新颖的基于势博弈的聚类算法, 将聚类问题转化为解决势博弈的纯纳什均衡问题。之后, 我们给出一个基于最佳响应策略[17]的迭代算法来计算公式化势博弈的纯纳什均衡。本文以模糊  $c$  均值作为初始

聚类, 以迭代最佳响应算法快速地计算出势博弈的纯纳什均衡点。在该迭代过程中, 我们考虑玩家偏离规则为顺序偏离[18], 由此得到的 FS-IBRC 算法在两个不同的数据集中测试并实现。

## 2. 预备知识

### 2.1. 博弈论

博弈论是现代数学的一个新的重要分支, 其研究的主体(称为玩家)是理性的, 博弈则是研究玩家之间的相互作用, 这种相互作用可能从合作扩展到冲突。玩家的收益仅仅取决于自己所采取的行动策略, 还取决于其他玩家所采取的策略。博弈可以用作建模和解决在多个应用领域中遇到的各种问题的框架, 其中包括经济, 交通, 军事战略和生物学等。

现在, 我们来定义一般形式的博弈, 令  $P$  为参加博弈的所有玩家的集合, 那么  $P = \{1, \dots, n\}, n \geq 2$ 。令  $S_i$  为第  $i$  个玩家的策略集,  $U_i$  为第  $i$  个玩家的效用函数,  $S$  称为局势, 是所有玩家策略集的笛卡尔积。于是, 我们有如下定义,

定义 2.1 [18] 令  $G$  为一般形式的博弈

$$G = P, \{S_i\}_{i \in P}, \{U_i\}_{i \in P} \quad (1)$$

- $P = \{1, \dots, n\}$ : 表示参加博弈的所有玩家的集合;
- $S_i$ : 表示玩家  $i$  的策略集;
- $s_i \in S_i$ : 表示玩家  $i$  所选择的策略;
- $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$ : 表示玩家  $i$  的对手所选择的策略的组合;
- $S_{-i} = \prod_{j \in P, j \neq i} S_j$ : 表示玩家  $i$  的对手的局势的集合;
- $S = S_i \times S_{-i} = \prod_{i \in P} S_i$ : 表示该博弈的所有局势的集合;
- $s = (s_i, s_{-i}) \in S$ : 表示该博弈的一个局势;
- $U_i(\cdot): S \rightarrow \mathbb{R}$ : 玩家  $i$  的效用函数。

定义 2.2 [18] (纳什均衡)局势  $s^* = (s_i^*, s_{-i}^*)$  是博弈(1)的纳什均衡, 如果

$$U_i(s_i^*, s_{-i}^*) \leq U_i(s_i, s_{-i}^*), \forall s_i \in S_i, \forall i \in P \quad (2)$$

纳什均衡还可以定义为满足(3)式的一个局势

$$s_i^* \in BR_i(s_{-i}^*), \forall i \in P, \quad (3)$$

其中,  $BR_i(s_{-i}^*)$  是玩家  $i$  的最佳响应策略集。

定义 2.3 [18] (势博弈 - 势函数)博弈(1)是势博弈, 如果存在一个全局函数  $\psi: S \rightarrow \mathbb{R}$ , 对每个玩家  $i$ , 每个  $s_{-i} \in S_{-i}$ , 对所有的  $s_i, s'_i \in S_i$ , 我们有

$$U_i(s_i, s_{-i}) - U_i(s'_i, s_{-i}) = \psi(s_i, s_{-i}) - \psi(s'_i, s_{-i}). \quad (4)$$

那么, 这个全局函数  $\psi$  就是有限博弈(1)的势函数。

性质 2.1 [18] 势博弈存在以下性质:

- 势博弈至少有一个纯纳什均衡。
- 每一个最佳响应学习过程都趋于一个纯策略纳什均衡。

### 2.2. 聚类

考虑  $n$  个对象的集合  $O = \{o_1, \dots, o_n\}$ , 其中, 每个对象  $o_i \in O$  具有  $r$  个属性, 那么,  $o_i$  可进一步定义

为  $o_i = (o_i^1, \dots, o_i^r)$ 。令  $\pi = (C_1, \dots, C_k)$  是将  $n$  个参与者分为  $k$  个没有交集的群组的一个结果, 即  $O = \bigcup_{j=1}^k C_j$ , 且  $C_i \cap C_j = \emptyset, i, j \in \{1, \dots, k\}, i \neq j$ , 在同一个子集里面的所有对象都是相似性准则  $\phi$  下的相似性最大。接下来, 我们定义所有分组结果的集合  $\Pi$ , 那么我们所研究的数据聚类问题可定义为:

$$\pi^* = \arg \min_{\pi \in \Pi} f(\pi), \tag{5}$$

其中,  $f(\cdot)$  是关于  $\phi$  的一个函数表达。

为了方便接下来关于聚类稳定性的证明, 下面我们将给出稳定的聚类的定义。

定义 2.4 [18] (稳定的聚类) 一个聚类结果  $\pi = (C_1, \dots, C_k)$ ,  $\pi \in \Pi$  是稳定的, 如果  $C_1, \dots, C_k$  都是外部稳定和内部稳定的。

- $C_j, j \in (1, \dots, k)$  内部稳定: 满足

$$\forall o \in C_j, \phi(o, C_j) \geq \phi(o, C_j \setminus \{o\}).$$

- $C_j, j \in (1, \dots, k)$  外部稳定: 满足

$$\forall o \notin C_j, \phi(o, C_j \cup \{o\}) < \phi(o, C_j).$$

近年来, 在聚类分析过程中已经提出众多聚类算法, 最经典的就是属于基于划分的聚类算法[6]中的  $k$  均值算法和 FCM, FCM 算法在运行的精度上优于  $k$  均值算法, 在计算速度上也比  $k$  均值要快, 所以, 本文讨论 FCM 算法与所提出的迭代最佳响应相结合, 快速地得出势博弈的纯纳什均衡。

### 2.3. FCM 算法

模糊  $c$  均值算法(Fuzzy C-Means, 简称 FCM)是根据隶属度来确定每个数据对象属于某个组别的程度的一种算法。FCM 算法的基本思想是: 建立基于隶属度矩阵的目标函数, 通过对聚类中心以及隶属度矩阵的不断迭代优化, 实现目标函数的收敛, 从而实现对样本数据的聚类。

FCM 算法的目标函数定义为

$$J(u_{ij}, c_i) = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m \|o_j - c_i^2\|, \tag{6}$$

其中,  $u_{ij}$  为隶属度矩阵  $U$  的元素, 且  $\sum_{i=1}^k u_{ij} = 1, i = 1, \dots, k; j = 1, \dots, n$ 。  $c_i$  为聚类中心, 且  $i = 1, \dots, k$ 。  $m$  为模糊加权指数, 且  $m \geq 1$ ,  $m$  越大, 聚类的模糊程度越高, 本文默认  $m$  的取值为 2。第  $j$  个样本数据到第  $i$  个聚类中心的欧氏距离为  $d^2 = \|o_j - c_i\|^2$ 。

根据拉格朗日乘数法, 我们可由(6)式构造出  $n$  个限制条件, 引入  $n$  个  $\lambda$  因子, 进而可得到

$$\mathcal{L}(u_{ij}, c_i, \lambda_1, \dots, \lambda_n) = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m \|o_j - c_i^2\| - \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^k u_{ij} - 1 \right), \tag{7}$$

将(7)式分别对  $u_{ij}, c_i$  求偏导, 且令倒数为 0, 可得

$$\frac{\partial \mathcal{L}}{\partial u_{ij}} = m u_{ij}^{m-1} \|o_j - c_i^2\| - \lambda_j = 0 \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial c_i} = (-2) \sum_{j=1}^n u_{ij}^m (o_j - c_i) = 0 \tag{9}$$

由(8)式可得

$$u_{ij} = \frac{1}{\sum_{l=1}^k \left( \frac{\|o_j - c_l\|}{\|o_j - c_i\|} \right)^{\frac{2}{m-1}}} \tag{10}$$

由(9)式可得

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m o_j}{\sum_{j=1}^n u_{ij}^m} \quad (11)$$

FCM 算法的基本步骤如下

1) 初始化参数: 确定聚类的数  $C$ , 最大迭代次数  $M$ , 迭代次数  $t=1$ , 模糊加权指数  $m=2$ , 迭代终止阈值为  $\varepsilon$ ;

2) 初始化隶属度矩阵  $U_0$ , 且令此时  $t=0$ ;

3) 计算聚类中心  $c_i^t$ ,  $c_i = \frac{\sum_{j=1}^n u_{ij}^m o_j}{\sum_{j=1}^n u_{ij}^m}$ ;

4) 计算并更新隶属度矩阵  $U_t$ ,  $u_{ij} = \frac{1}{\sum_{l=1}^k \left( \frac{\|o_j - c_i\|}{\|o_j - c_l\|} \right)^{\frac{2}{m-1}}}$ ;

5) 计算目标函数  $J^{(t)}$ ,  $J^{(t)} = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m \|o_j - c_i\|^2$ ;

6) 直到  $|J^{(t)} - J^{(t-1)}| < \varepsilon$ , 终止。否则, 返回步骤(3)。

FCM 算法是一种简单高效的聚类方法, 已广泛应用于数据分析、图像分割、模式识别等领域。FCM 算法相较于  $k$  均值算法对分割问题的描述更符合, 有更强的灵活性。FCM 将“硬聚类”扩展到“软聚类”, 其中每个数据样本可以属于多个集群, 这样可使数据点尽可能的分配到最佳的类别中, 从而减少了后期对其收敛到最优聚类结果的时效, 也降低了聚类结果收敛的复杂度。同时, 也使得后续的迭代最佳响应算法实现最大的计算价值。这也是本文想要采用 FCM 算法所得聚类结果作为初始聚类的出发点之一。

### 3. 聚类是一个策略势博弈

因为一些聚类算法很难保证聚类结果的稳定性, 所以, 本节提出基于博弈论的聚类模型, 并给出一个效用函数, 用来证明势函数的存在性, 因此保证了纳什均衡的存在, 它是一个稳定的聚类。

#### 3.1. 基于博弈论的聚类模型

首先, 我们给出一个数据集  $O = \{o_1, \dots, o_n\}$ , 每个对象  $o_i \in O$  具有  $r$  个属性, 那么,  $o_i$  可进一步定义为  $o_i = (o_i^1, \dots, o_i^r)$ 。我们的目的是将这些数据分成提前定义的  $k$  个分组, 即  $(C_1, \dots, C_k)$ , 令  $(c_1, \dots, c_k)$  对应这  $k$  个分组的质心。

我们将  $o_i$  看作是博弈中的第  $i$  个玩家。令  $P = (1, \dots, n)$  是  $n$  个玩家的集合,  $P$  中每个玩家都有选择继续留在该群组或者迁移到其余群组的权利, 由此可知, 玩家  $i$  的所有策略的集合就是所有群组的集合, 即  $S_i = C = \{C_1, \dots, C_k\}$ ,  $\forall i \in P$ 。

局势  $s = (s_1, \dots, s_n) = (s_i, s_{-i}) \in S = \prod_{i \in P} S_i$ , 我们定义玩家  $i$  的效用函数为[18]:

$$U_i(s_i, s_{-i}) = \begin{cases} d(o_i, c_{s_i})^2 \left( 1 + \frac{1}{|C_{s_i}|} \right), & \text{if } o_i \in C_{s_i}, \\ d(o_i, c_{s_i})^2 \left( 1 + \frac{1}{|C_{s_i}| + 1} \right), & \text{otherwise.} \end{cases} \quad (6)$$

- $C_{s_i} = (C_{s_i}^1, \dots, C_{s_i}^r)$  为  $C_{s_i}$  的质心。
- $d$  是测量玩家之间相似性的距离函数, 这里我们用欧式距离来度量。

### 3.2. 聚类是势博弈

我们将聚类公式化为非合作博弈  $G(1)$  所示,  $G$  是  $n$  个玩家进行的非合作策略形式的博弈, 它的解叫做纳什均衡。

定理 3.1 [18] 基于非合作博弈模型(1)的聚类是势博弈。

证要证效用函数  $U_i$  满足

$$\psi = (s_i, s_{-i}) = \sum_{i=1}^k \phi(C_i) = \sum_{i=1}^k \sum_{o_j \in C_{s_i}} d(o_j, c_{s_i})^2 = \sum_{i \in P} d(o_i, c_{s_i})^2 \quad (7)$$

为势函数。即证对任意  $i \in P, s_i, s'_i$ , 使得

$$U_i(s_i, s_{-i}) - U_i(s'_i, s_{-i}) = \psi(s_i, s_{-i}) - \psi(s'_i, s_{-i})$$

成立。令  $\Delta_U = U_i(s_i, s_{-i}) - U_i(s'_i, s_{-i}), \Delta_\psi = \psi(s_i, s_{-i}) - \psi(s'_i, s_{-i})$ , 为不失一般性, 我们任意给出其中两个群组  $X = \{X_1, \dots, X_{n_1}, A\}$  和  $Y = \{Y_1, \dots, Y_{n_2}\}$ , 且  $n_1, n_2 \neq 0$ 。令  $o_1, o_2$  分别为  $X, Y$  的质心,  $A$  为  $X$  的玩家, 但通过效用函数(6), 玩家  $A$  会从  $X$  迁徙到  $Y$ , 进而群组将更新为  $X_1 = X \setminus \{A\}, Y_1 = Y \cup \{A\}$ , 质心更新为  $o'_1, o'_2$ 。由此,

$$\Delta_\psi = \sum_{i=1}^{n_1} o'_1 X_i^2 + \sum_{j=1}^{n_2} o'_2 Y_j^2 + o_2 A^2 - \left( \sum_{i=1}^{n_1} o_1 X_i^2 + \sum_{j=1}^{n_2} o_2 Y_j^2 + o_1 A^2 \right),$$

其中,  $o_1 X_i^2 = (o_1 X_i)^2 = o_1 o_1^2 + o'_1 X_i^2 + 2o_1 o'_1 \cdot o'_1 X_i$ ,  $o_2 Y_j^2 = (o_2 Y_j)^2 = o_2 o_2^2 + o'_2 Y_j^2 + 2o_2 o'_2 \cdot o'_2 Y_j$ 。

计算得,

$$\begin{aligned} \Delta_\psi &= n_2^2 o_2 o_2'^2 - n_1 o_1 o_1'^2 - n_2 o_2 o_2'^2 + 2n_2 o_2 o_2'^2 - n_1^2 o_1 o_1'^2 \\ &= o_2'^2 A^2 \left( 1 + \frac{1}{n_2} \right) - o_1 A^2 \left( 1 + \frac{1}{n_1} \right) = \Delta_U \end{aligned}$$

因此, 在效用函数为(6)式的前提下,  $\psi$  为势函数。

性质 3.1 [18] 势博弈  $G$  的每个纯纳什均衡  $s^* \in S$ , 对应于稳定的聚类结果。

## 4. 关于博弈(1)的纳什均衡

本节讨论聚类算法中初始聚类在整体算法过程中有不可忽视的重要性, 并给出博弈(1)的求解过程, 提出迭代最佳响应聚类算法, 将 FCM 与文献[18]提出的顺序迭代最佳响应聚类算法(Sequential iterative best-response clustering, 简称 S-IBRC)相结合, 得出 FS-IBRC 算法, 将其与 S-IBRC 算法在两个不同的数据集上实验并比较。

### 4.1. 初始聚类的重要性

在众多传统聚类算法中, 随机性选取初始中心可能使聚类结果陷入局部最优, 且所得到的聚类结果并不是稳定的聚类。例如传统的 k-means 算法, 所得的聚类结果缺乏稳定性。现在, 越来越多的科学研究者致力于研究如何优化初始聚类中心的选择。与初始中心相同, 本文所提出的算法首先要确定初始聚类的选择方法。结合文献[18]中将一般初始化和 k-means 算法分别作为初始聚类, 进而比较各自的聚类结果和运行时间的思想。那么, 在本文中, 将给出一种将 FCM 作为初始聚类的基于势博弈的算法(算法 4.1), 并与一般初始化作为初始聚类的基于势博弈的聚类算法(算法 4.2)的运行时间相比较, 从而证明 FCM 算



法作为初始聚类在时效上的优越性。

## 4.2. 基于一般初始化的 S-IBRC 算法

基于一般初始化的 S-IBRC 算法可以分为以下几步:

### 算法 4.1

#### 第一步一般初始化

输入包含  $n$  个对象的数据集  $O = \{o_1, \dots, o_n\}$ ,  $o_i = (o_i^1, \dots, o_i^r)$ , 输入  $k$ ,  $k$  为将  $O$  分为  $k$  个群组, 然后从  $O$  中随机选择  $k$  个对象作为质心  $c = \{c_1, \dots, c_k\}$ , 若  $d(o_i, c_i) \leq d(o_i, c_j)$ ,  $i, j \in k$ , 则  $o_i$  属于  $C_i$ 。接着, 根据

$$c_{i'}^h = \frac{1}{|C_{i'}^h|} \sum_{o_i \in C_{i'}^h} o_i^h, \quad \forall h \in \{1, \dots, r\}, \forall i' \in \{1, \dots, k\} \quad (8)$$

更新质心  $c_{i'} = (c_{i'}^1, \dots, c_{i'}^r)$ ,  $i' \in \{1, \dots, k\}$ 。最后输出初始聚类  $C^0$  和质心  $c' = \{c_1, \dots, c_k\}$ 。

#### 第二步初始聚类 $C^0$ 对应局势 $s$

输入包含  $n$  个对象的数据集  $O = \{o_1, \dots, o_n\}$ ,  $o_i = (o_i^1, \dots, o_i^r)$ , 输入  $k$ ,  $k$  为将  $O$  分为  $k$  个群组, 然后输入初始聚类  $C^0$  和第一次更新质心  $c' = \{c_1, \dots, c_k\}$ 。在博弈(1)中,  $\forall i \in P, s_i \in S = \{1, \dots, k\}, o_i \in C_{s_i}$ , 令  $s = (s_i, s_{-i}) \in S$  是博弈(1)的一个局势, 对应于初始聚类  $C^0$ 。

#### 第三步顺序迭代最佳响应聚类

重复以下过程:

初始化状态为 false, 即假设没有玩家  $i$  会迁移到别的群组。从玩家  $i=1$  开始到  $i=n$  结束过程:

若  $s_i \notin BR_i(s_{-i})$ , 则从  $BR(s_{-i})$  中随机选择一个策略  $s_i^*$ ,  $s_i^* \in \{1, \dots, k\}$ , 随后更新局势为  $C_{s_i} \leftarrow C_{s_i} \setminus \{o_i\}$ ,  $C_{s_i^*} \leftarrow C_{s_i^*} \cup \{o_i\}$ , 接着根据关系式(8)更新质心  $c_{s_i}$  和  $c_{s_i^*}$ 。且状态更新为 true, 即代表有玩家发生迁移。

直到初始状态 false 不会改变终止。

输出最终聚类结果  $C$ 。

## 4.3. FS-IBRC 算法

基于 FCM 算法作为初始聚类的 FS-IBRC 算法可以分为以下几步:

### 算法 4.2

#### 第一步初始化

输入包含  $n$  个对象的数据集  $O = \{o_1, \dots, o_n\}$ ,  $o_i = (o_i^1, \dots, o_i^r)$ , 输入  $k$ ,  $k$  为将  $O$  分为  $k$  个群组, 根据 FCM 算法得初始聚类  $C^0$  和质心  $c = \{c_1, \dots, c_k\}$ 。

FCM 算法的基本步骤如下

1) 初始化参数: 确定聚类的数  $C$ , 最大迭代次数  $M$ , 迭代次数  $t=1$ , 模糊加权指数  $m=2$ , 迭代终止阈值为  $\varepsilon$ ;

2) 初始化隶属度矩阵  $U_0$ , 且令此时  $t=0$ ;

3) 计算聚类中心  $c_i^t$ ,  $c_i = \frac{\sum_{j=1}^n u_{ij}^m o_j}{\sum_{j=1}^n u_{ij}^m}$ ;

4) 计算并更新隶属度矩阵  $U_t$ ,  $u_{ij} = \frac{1}{\sum_{l=1}^k \left( \frac{\|o_j - c_i\|}{\|o_j - c_l\|} \right)^{\frac{2}{m-1}}}$ ;

5) 计算目标函数  $J^{(t)}$ ,  $J^{(t)} = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m \|o_j - c_i\|^2$ ;

直到  $|J^{(t)} - J^{(t-1)}| < \varepsilon$ , 终止。否则, 返回步骤(3)。

第二步初始聚类  $C^0$  对应局势  $s$

输入包含  $n$  个对象的数据集  $O = \{o_1, \dots, o_n\}$ ,  $o_i = (o_i^1, \dots, o_i^r)$ , 输入  $k$ ,  $k$  为将  $O$  分为  $k$  个群组, 然后输入初始聚类  $C^0$  和质心  $c = \{c_1, \dots, c_k\}$ 。在博弈(1)中,  $\forall i \in P, s_i \in S = \{1, \dots, k\}, o_i \in C_{s_i}$ , 令  $s = (s_i, s_{-i}) \in S$  是博弈(1)的一个局势, 对应于初始聚类  $C^0$ 。

第三步顺序迭代最佳响应聚类

重复以下过程:

初始化状态为 false, 即假设没有玩家  $i$  会迁移到别的群组。从玩家  $i = 1$  开始到  $i = n$  结束过程:

若  $s_i \notin BR_i(s_{-i})$ , 则从  $BR(s_{-i})$  中随机选择一个策略  $s_i^*$ ,  $s_i^* \in \{1, \dots, k\}$  随后更新局势为  $C_{s_i} \leftarrow C_{s_i} \setminus \{o_i\}$ ,  $C_{s_i^*} \leftarrow C_{s_i^*} \cup \{o_i\}$ , 接着根据关系式(8)更新质心  $c_{s_i}$  和  $c_{s_i^*}$ 。且状态更新为 true, 即代表有玩家发生迁移。

直到初始状态 false 不会改变终止。

输出最终聚类结果  $C$ 。

## 5. 实验评估

为进一步探讨 S-IBRC 和 FS-IBRC 算法的性能, 本节将对两个不同的数据集上进行一系列的实验。并将这两个算法的实验运行时间进行比较。以上实验均以 python 语言实现, 并在 2.50 GHz Intel Core i5-7200U, 4Go RAM 上进行。

### 5.1. 数据集描述

在我们的实验中, 所使用的数据集[19] (Iris, Ecoli)表 1 将给出以上数据集关于数量  $n$ , 属性个数  $r$ , 划分组数  $k$  的描述。

Table 1. Datasets description

表 1. 数据集描述

数据集	$n$	$r$	$k$
Iris	150	4	3
Ecoli	336	8	8

### 5.2. 实验结果描述

为了检查提出的算法 S-IBRC 和 FS-IBRC 的相应计算时间, 我们将算法均在相同的数据集和相同的计算机运行环境中进行。得出的实验结果通过表 2 和图 1 来反映, 通过观察可知, 算法 FS-IBRC 在数据集 Iris 上运行时间与算法 S-IBRC 几乎相同, 而在数据集 Ecoli 上, FS-IBRC 算法的运行时间比算法 S-IBRC 快很多, 这可能与数据集的个数和维度相关。

Table 2. CPU time description (seconds)

表 2. CPU 时间描述(秒)

	S-IBRC	FS-IBRC
Iris	0.0094	0.0091
Ecoli	0.2647	0.1204



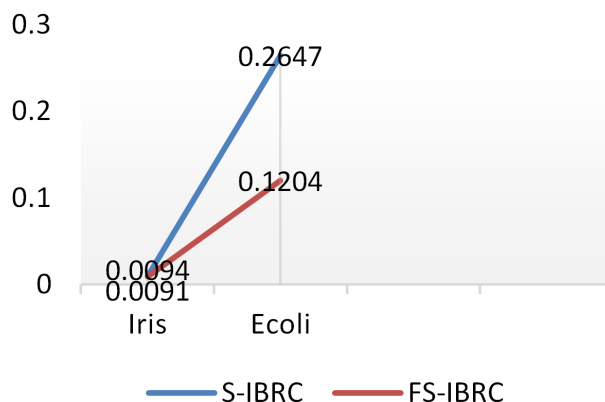


Figure 1. CPU time comparative results

图 1. CPU 时间比较结果

## 6. 结论

本文研究了基于势博弈的聚类算法, 提出了一种新颖的以模糊  $c$  均值作为其初始聚类的基于势博弈的聚类算法, 将聚类任务转化为寻找势博弈模型的纯纳什均衡, 且证明其对应于稳定的聚类。进而给出了解决这种基于势博弈模型的聚类算法, 即迭代的最佳响应算法。将所提算法在两个不同的人工数据集上实现和测试, 将该算法的运行时间与一般初始化的算法相比较, 实验证明了所提出算法的可行性。近年来研究基于博弈论的聚类算法的文章逐渐增多, 除了本文研究的关于势博弈的聚类算法, 认为聚类过程就是势博弈, 进而求得相应的聚类之外, 我们还可以考虑一般的基于非合作博弈的聚类和基于合作博弈的聚类, 这些都是比较新颖的算法, 通过研究博弈论的均衡性, 求得相关的稳定的聚类。在未来的工作中, 我们除了研究改进本文所提出的算法之外, 还可以研究一些相关定理的推论及其证明, 进而获得更多的比较有趣的算法, 并在更多的数据集上实验。

## 参考文献

- [1] Yan, J., Wu, Z., Luo, H., *et al.* (2013) P2P Traffic Identification Based on Host and Flow Behaviour Characteristics. *Cybernetics and Information Technologies*, **13**, 64-76. <https://doi.org/10.2478/cait-2013-0026>
- [2] Marconi, M., Germani, M., Mandolini, M., *et al.* (2019) Applying Data Mining Technique to Disassembly Sequence Planning: A Method to Assess Effective Disassembly Time of Industrial Products. *International Journal of Production Research*, **57**, 599-623. <https://doi.org/10.1080/00207543.2018.1472404>
- [3] Aggarwal, C.C. and Reddy, C.K. (2013) *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, London.
- [4] Sabor, N., Ahmed, S.M., Abo-Zahhad, M., *et al.* (2018) ARBIC: An Adjustable Range Based Immune Hierarchy Clustering Protocol Supporting Mobility of Wireless Sensor Networks. *Pervasive and Mobile Computing*, **43**, 27-48. <https://doi.org/10.1016/j.pmcj.2017.11.003>
- [5] Federica, V., Damiano, F. and Alfonso, N. (2018) A Novel Scale-Invariant, Dynamic Method for Hierarchical Clustering of Data Affected by Measurement Uncertainty. *Journal of Computational & Applied Mathematics*, **344**, 521-531. <https://doi.org/10.1016/j.cam.2018.05.062>
- [6] Tang, R. and Fong, S. (2018) Clustering Big IoT Data by Metaheuristic Optimized Mini-Batch and Parallel Partition-Based DGC in Hadoop. *Future Generation Computer Systems*, **86**, 1395-1412. <https://doi.org/10.1016/j.future.2018.03.006>
- [7] Xue, Q., *et al.* (2018) Automatic Identification of Fractures Using a Density-Based Clustering Algorithm with Time-Spatial Constraints. *Energies*, **11**, 563. <https://doi.org/10.3390/en11030563>
- [8] Zhang, H.Y., Li, X., Wang, X.K. and Zhang, M. (2012) Grid-Based Clustering Algorithm for Sensing. *Applied Mechanics and Materials*, **203**, 257-262. <https://doi.org/10.4028/www.scientific.net/AMM.203.257>
- [9] Bezdek, J.C., Ehrlich, R. and Full, W. (1984) FCM: The Fuzzy C-Means Clustering Algorithm. *Computers & Geos-*

- 
- ciences*, **10**, 191-203. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
- [10] Liu, Q. and Liu, M. (2017) Energy-Efficient Clustering Algorithm Based on Game Theory for Wireless Sensor Networks. *International Journal of Distributed Sensor Networks*, **13**, 12 p. <https://doi.org/10.1177/1550147717743701>
- [11] Heloulou, I., Said, R.M. and Tahar, K.M. (2016) Automatic Multi-Objective Clustering Based on Game Theory. *Expert Systems with Applications*, **67**, 32-48. <https://doi.org/10.1016/j.eswa.2016.09.008>
- [12] Garg, V.K., Narahari, Y. and Narasimha, M.M. (2013) Novel Biobjective Clustering (BiGC) Based on Cooperative Game Theory. *IEEE Transactions on Knowledge & Data Engineering*, **25**, 1070-1082. <https://doi.org/10.1109/TKDE.2012.73>
- [13] Brânzei, R., Dimitrov, D.A. and Tijs, S.H. (2005) Models in Cooperative Game Theory: Crisp, Fuzzy, and Multi-Choice Games (Lecture Notes in Economics and Mathematical Systems). Springer, Berlin.
- [14] Albarelli, A., Torsello, A., *et al.* (2009) Matching as a Non-Cooperative Game. 2009 *IEEE 12th International Conference on Computer Vision*, Kyoto, 27 September-4 October 2009, 1319-1326. <https://doi.org/10.1109/ICCV.2009.5459312>
- [15] Quang, D.L., Chew, Y.H. and Soong, B.H. (2016) Potential Game Theory.
- [16] Liu, F., Bie, Z., Jiang, J., *et al.* (2019) Impact of Microgrid Aggregator on Joint Energy and Reserve Market Based on Pure Strategy Nash Equilibrium. *Energy Procedia*, **159**, 142-147. <https://doi.org/10.1016/j.egypro.2018.12.032>
- [17] Feldman, M., Snappir, Y. and Tamir, T. (2017) The Efficiency of Best-Response Dynamics. *International Symposium on Algorithmic Game Theory*, L'Aquila, 12-14 September 2017, 186-198. [https://doi.org/10.1007/978-3-319-66700-3\\_15](https://doi.org/10.1007/978-3-319-66700-3_15)
- [18] Bouchama, K., Lallouet, A., Radjef, M.S., *et al.* (2019) Solving a Potential Clustering Game. *International Journal on Artificial Intelligence Tools*, **28**, Article ID: 1950006. <https://doi.org/10.1142/S0218213019500064>
- [19] Amarnath, B. and Balamurugan, S.A.A. (2016) Review on Feature Selection Techniques and Its Impact for Effective Data Classification Using UCI Machine Learning Repository Dataset. *Journal of Engineering Science and Technology*, **11**, 1639-1646.