

DOI:10.3969/j.issn.1003-5060.2026.05.007

基于改进模糊 C 均值聚类的雷达信号分选算法

张振宇, 陈 磬

(电子科技大学 数学科学学院, 四川 成都 611731)

摘 要:在雷达信号分选中,针对模糊 C 均值(fuzzy C-means, FCM)聚类算法需预先指定聚类数目、对初值敏感以及数据数量分布不均时分选效果不佳等问题,文章提出一种基于数据场和类中心约束的改进 FCM 算法。该算法首先利用数据场理论确定势心,再将这些势心作为约束点进行聚类分选;改进后的算法无需指定聚类数目,对初值不敏感,且在数据数量分布不均时仍能取得很好的分选效果。仿真实验结果表明,改进算法的分选准确率高达 95% 以上,与传统的聚类算法相比,改进算法的分选准确率更高,结果更稳定。

关键词:雷达信号分选;模糊 C 均值(FCM)聚类算法;数据场;脉冲描述字

中图分类号:TN957.51

文献标志码:A

文章编号:1003-5060(2026)05-0621-08

Radar signal sorting algorithm based on improved fuzzy C-means clustering

ZHANG Zhenyu, CHEN Die

(School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: In radar signal sorting, in response to the challenges faced by the fuzzy C-means(FCM) clustering algorithm, including the requirement to specify the number of clusters, sensitivity to the initial value, and unsatisfactory sorting performance in cases of uneven data distributions, this paper proposes an improved FCM clustering algorithm based on data field and cluster center constraints. The algorithm firstly determines the potential center by using the theory of the data field, and then the potential center is taken as a constraint point of the algorithm for clustering and sorting. The improved algorithm does not need to specify the number of clusters, is insensitive to the initial value, and achieves good sorting results when the data is unevenly distributed. Simulation experiments show that the sorting accuracy of the improved algorithm can reach more than 95%. Compared with the conventional clustering algorithms, the improved algorithm exhibits higher sorting accuracy and enhanced stability.

Key words: radar signal sorting; fuzzy C-means(FCM) clustering algorithm; data field; pulse description word

雷达信号分选是指从随机交错的雷达脉冲信号中将各个雷达的脉冲序列分离出来的过程^[1]。雷达信号分选是信号处理和雷达侦察的重要基础,在现代电子战中,雷达信号分选有着举足轻重的作用。而随着雷达技术的发展,电磁环境复杂程度也与日俱增,面对日益增多的脉冲密度、愈发

多变的脉冲信号,雷达信号分选也面临着前所未有的挑战。

传统的雷达信号分选算法主要利用脉冲重复间隔(pulse repetition interval, PRI)对雷达信号进行分选,有相关函数法、直方图法、PRI 变换法等分选算法^[2]。

收稿日期:2023-12-15;修回日期:2024-05-30

基金项目:国家自然科学基金资助项目(71871046)

作者简介:张振宇(1998—),男,四川成都人,电子科技大学硕士生;

陈 磬(1984—),男,四川成都人,博士,电子科技大学副教授,硕士生导师,通信作者,E-mail:chendie@uestc.edu.cn.

雷达信号的多参数分选主要采用雷达脉冲信号的载频(radio frequency, RF)、脉冲宽度(pulse width, PW)、脉冲幅度(pulse amplitude, PA)、脉冲到达方向(direction of arrival, DOA)等参数进行分选。早期多参数分选主要采用模板匹配法,将接收到的雷达信号参数与特征数据库中的特征进行比对,但该方法只适用于雷达信号较为简单的情况,随着雷达信号不断发展,其适用性越来越差。

随着计算机性能的提升以及机器学习的发展,利用机器学习来进行雷达多参数分选也成为学者们关注的重点,但在实际使用中,很难得到雷达信号的先验知识,因此聚类算法是目前多参数分选研究的首选方法。文献[3-4]利用改进的DBSCAN算法对雷达信号进行分选,DBSCAN算法无需预设聚类数量,而是通过邻域半径 ϵ 和最小点数目 N 这2个关键参数来界定数据点的密集程度,这2个参数共同定义了“密集”区域的标准,在这种标准下,若某点的 ϵ 邻域内包含的点数不少于 N ,则该区域被认定为密集区域;尽管这一机制具有灵活性,但准确选取合适的邻域半径和最小点数仍具有一定挑战性。文献[5-6]使用了改进 K -means算法对雷达信号进行分选, K -means算法实现简单,但是需要提前指定聚类的数目。文献[7-8]则使用了支持向量聚类对雷达信号进行分选。上述算法中, K -means算法的一个核心难题是,在应用于雷达信号分选这样的复杂场景中时,往往难以预先获知精确的类别数量,使得 K -means这类必须预先设定聚类数量的算法难以有效实施。相较之下,DBSCAN和支持向量聚类等算法虽无需预设聚类数目,但同样要求设置其他参数,而这些参数的最优选择同样难度较大。

本文聚焦于模糊 C 均值(fuzzy C -means, FCM)聚类算法,模糊聚类的内在优势使得本研究能够对聚类结果开展更为细致深入的解读。然而,传统的FCM算法存在若干问题,如需要设置聚类数目以及在处理数据分布不均衡的情形时聚类结果可能不尽人意。为此,本文借助数据场理论的思想对算法进行优化,创新地提出一种改进的FCM算法。改进后的算法无需预设聚类数目,并能克服传统FCM算法在处理非均匀数据分布时的局限性,提高聚类准确度;在雷达信号分选任务上,改进FCM算法展现出良好的适用性。

1 模糊 C 均值聚类算法

FCM聚类算法是一种基于划分的聚类算法,该算法将数据点划分为 c 个类,其中类的数目需要事先给出,通过计算划分的中心 v_i 以及每个点的隶属度 u_{ij} 使得下面的目标函数达到最小,即

$$J_m(u, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d(x_k, v_i) \quad (1)$$

其中

$$1 < m < \infty; \quad \sum_{i=1}^c u_{ik} = 1, u_{ik} \in [0, 1];$$

$$\forall k, d(x_k, v_i) = \|x_k - v_i\|^2.$$

通过目标函数及约束条件,可以得到聚类中心和隶属度的更新公式如下:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m}, \quad i = 1, 2, \dots, c \quad (2)$$

$$u_{ik} = 1 / \sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}},$$

$$i = 1, 2, \dots, c; k = 1, 2, \dots, n \quad (3)$$

FCM算法的实现简单且运算速度快,并且隶属度的存在可以对聚类的结果进行更细致的分析。

通过观察类中心点 v_i 的迭代公式发现, v_i 的迭代会受到所有数据点的影响,虽然 u_{ik} 可以对影响的权重进行控制,但是当数据类分布不均匀时,若一个类中数据的数量远大于其他类,或存在干扰点或孤立点等情况时,则 v_i 就会发生偏移,继而导致聚类结果发生偏差。而FCM算法的另一个缺陷是需要提前指定聚类数目,这使得基于FCM聚类的分选算法在未知环境中很难使用。

由上可知,FCM算法的缺陷主要来自于未知的聚类簇数目和聚类中心,而数据场理论可以解决这些问题。

2 数据场理论

数据场理论是将物理学中场的概念引入数据科学^[9]中,在数据场理论中数据之间存在一种作用力,这种力的大小与数据间的距离呈负相关,利用数据场的性质可以剔除数据的噪声点,找到数据的聚集中,从而改进FCM算法。下面介绍本文使用的数据场概念。

2.1 场强函数

场强函数用于衡量数据点之间的相互作用力大小,数据点 x 在场中 y 点产生的作用力为:

$$f_y(x) = \rho e^{-[d^2(x, y)/(2\sigma^2)]} \quad (4)$$

2.2 势值函数

场强函数用于衡量两点之间的作用力,而某一数据点受到的所有场强函数作用的累加和被称为势值,某一点 x 的势值为:

$$F(x) = \sum_{i=1}^n f_x(y_i) = \sum_{i=1}^n e^{-[d^2(x, y_i)/(2\sigma^2)]} \quad (5)$$

其中, σ 为辐射因子,辐射因子与场强函数的大小成正比。辐射因子分别取 0.025、0.100、0.150、0.200 时产生的等势线,如图 1 所示。

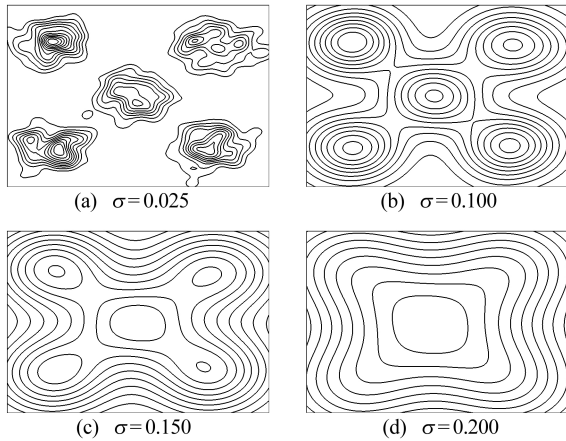


图1 不同辐射因子产生的等势线

从图 1 可以看出:当辐射因子较小时,数据场中的势心会呈现出过度细分的现象,导致在整个数据集内部形成大量的局部势心;随着辐射因子逐渐增大,这些势心之间开始显现融合趋势,它们之间的边界变得模糊不清;当辐射因子达到某个阈值后,所有数据点可能会被归纳至单个全局势心之下。无论是辐射因子设置过低导致的过分割,还是设置过高引起的过度凝聚,都将对数据分析的有效性和准确性产生显著影响。因此,选择一个恰当的辐射因子对于后续的数据分析至关重要。为了寻找到最适宜的辐射因子值,可以引入势熵这一概念。

在数据场理论中,数据之间分布的规律主要是根据数据分布的不确定性来判断的,而通常采用势熵描述数据对象之间的不确定性。因此较小的势熵表明当前的数据对象之间具有更强的关联性与确定性^[10]。设每个数据对象的势值为 F_i ($i=1,2,\dots,n$),则势熵的表达式为:

$$H = - \sum_{i=1}^n \frac{F_i}{T} \ln\left(\frac{F_i}{T}\right) \quad (6)$$

其中, $T = \sum_{i=1}^n F_i$ 为标准化因子。

由于势值受到辐射因子 σ 的影响,使得 H 最

小的 σ 就是最优 σ ,即

$$\min_{\sigma>0} H(\sigma) = \min_{\sigma>0} \left(- \sum_{i=1}^n \frac{F_i}{T} \ln \frac{F_i}{T} \right) \quad (7)$$

2.3 势心

势心是势值的极大值点,从数据场理论的角度,势心代表了数据的聚集中心,因此可以用势心来判断数据的聚类簇数目。势心的数学描述如下:

$$F(x) \geq F(y), \quad y \in U_\delta(x) \quad (8)$$

其中, $U_\delta(x)$ 为 x 的 δ 邻域。

通过数据场理论,可以找到当前数据下的所有势心,并将势心的数目作为 FCM 算法的参数,这样就解决了 FCM 算法需要预设聚类数目的问题,但是并不能解决 FCM 算法初值敏感性以及数据分布不均衡时分选效果不佳等问题。因此,有必要对 FCM 算法进一步改进。

3 改进的 FCM 算法

针对上述 FCM 算法的问题,根据数据场理论,采用类中心约束的思想对算法进行改进。对原 FCM 算法的目标函数加入惩罚项 $\alpha \sum_{i=1}^c d(p_i, v_i)$, 得到改进算法的目标函数为:

$$J = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d(x_k, v_i) + \alpha \sum_{i=1}^c d(p_i, v_i) \quad (9)$$

从惩罚项可以看出,当 v_i 到 p_i 的距离增大时,惩罚项也随之增大,导致目标函数 J 增大。若使目标函数最小化,则 v_i 不能距离 p_i 过远。这样便将 v_i 限制在 p_i 附近,称 p_i 为约束点。

此外, α 控制了惩罚项对目标函数的影响大小。当 α 增大时,惩罚项增大,相应地,为使目标函数最小化, p_i 到 v_i 的距离会减小,因此称 α 为约束力度。

改进 FCM 算法的 v_i 和 u_{ik} 的迭代公式为:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k + \alpha p_i}{\sum_{k=1}^n u_{ik}^m + \alpha}, \quad i = 1, 2, \dots, c \quad (10)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}}}, \quad i = 1, 2, \dots, c \quad (11)$$

可以看出,改进后的算法 v_i 受到 p_i 的控制,因此,如果 p_i 距离真实类中心很远,那么 p_i 会将 v_i 约束到错误的地方。本文算法将数据场的势心作为约束点,利用势心的性质,可以很好地地将 v_i 的迭代限制在 p_i 附近,如图 2 所示。

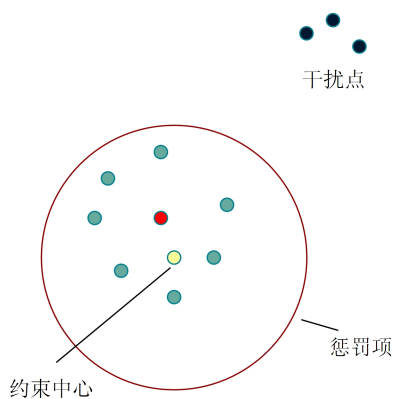


图 2 改进 FCM 算法原理示意图

从图 2 可以看出,惩罚项的作用就是将 v_i 的迭代约束在以 p_i 为中心的邻域内,而邻域半径的大小由约束力度 α 控制。

4 改进 FCM 算法的收敛性证明

文献[11-12]给出了 FCM 算法收敛性的证明,同理可以证明本文改进 FCM 算法的收敛性。

为方便证明,定义一些符号,设 $X = \{x_1, \dots, x_m\}$, $X \subset \mathbf{R}^s$, 是一个包含至少 $n(n > c)$ 个点的有限数据集,并且设 $\mathbf{R}^{c \times n}$ 表示所有 $c \times n$ 的实数矩阵,设矩阵 $U \in \mathbf{R}^{c \times n}$, $U = (u_{ik})$, u_{ik} 表示一个 X 的模糊划分。其中 U 满足如下条件:

$$u_{ik} \in [0, 1], \quad 1 \leq i \leq c, 1 \leq k \leq n \quad (12)$$

$$\sum_{i=1}^c u_{ik} = 1, \quad 1 \leq k \leq n \quad (13)$$

$\mathbf{R}^{c \times n}$ 中所有满足上述描述的矩阵用集合 M_{fc} 表示,矩阵 $U \in M_{fc}$ 被用于描述 X 的一个划分。设类中心为 $v_i (v_i \in \mathbf{R}^s)$, $v = [v_1 \ v_2 \ \dots \ v_c]^T$, $v \in \mathbf{R}^{c \times s}$, 约束点 p_i 也同 v_i 一样, $p = [p_1 \ p_2 \ \dots \ p_c]^T$, $p \in \mathbf{R}^{c \times s}$ 。

设 $G: M_{fc} \rightarrow \mathbf{R}^{c \times s}$ 是一个函数,其中 $G(U) = v = [v_1 \ v_2 \ \dots \ v_c]^T$, v_i 由式(10)来计算;同时设 $F: \mathbf{R}^{c \times s} \rightarrow P(M_{fc})$ 表示点到集合的映射, $F(v) = \{U \in M_{fc} | (U, v) \text{ 满足式(11)}\}$ 。改进 FCM 算法的迭代可以被表示为点到集合的映射 $T_m: M_{fc} \times \mathbf{R}^{c \times s} \rightarrow P(M_{fc} \times \mathbf{R}^{c \times s})$, 定义式如下:

$$T_m = A_2 \circ A_1 \quad (14)$$

其中

$$A_1: M_{fc} \times \mathbf{R}^{c \times s} \rightarrow \mathbf{R}^{c \times s};$$

$$A_1(U, v) = G(U) \quad (15)$$

$$A_2: \mathbf{R}^{c \times s} \rightarrow P(M_{fc} \times \mathbf{R}^{c \times s});$$

$$A_2(v) = \{(U, v) | U \in F(v)\} \quad (16)$$

因此 $T_m(U, v) = \{(\hat{U}, \hat{v}) | \hat{v} = G(\hat{U}), \hat{U} \in F(\hat{v})\}$, 由

此得到改进 FCM 算法的迭代序列为:

$$(U^{(1)}, v^{(1)}) \in M_{fc} \times \mathbf{R}^{c \times s},$$

$$(U^{(k)}, v^{(k)}) \in T_m(U^{(k-1)}, v^{(k-1)}),$$

$$k = 2, 3, \dots \quad (17)$$

为了证明算法的收敛性,下面将证明几个引理。

引理 1 令 $\phi: \mathbf{R}^{c \times s} \rightarrow \mathbf{R}$, $\phi(v) = J_m(U, v)$, 将 U 固定,那么 v^* 是 ϕ 的严格局部最小值当且仅当 $v_i^* = G(U)$ 。

证明 由于 ϕ 在 $\mathbf{R}^{c \times s}$ 上的最小值是一个无约束问题,方程的必要性需要证明 $\nabla_{v_i} \phi(v^*)$ 对所有 i 都为 0。

等价于 ϕ 关于 v_i 的方向导数 $\phi'(v_i^*, y) = 0$ 在任意方向 $y \in \mathbf{R}^s, y \neq 0$ 令 $t \in \mathbf{R}$, 并且对 $\forall i$ 有:

$$h_i(t) = \phi(v_i^* + ty) =$$

$$\sum_{k=1}^n (u_{ik})^m ||x_k - (v_i^* + ty)||^2 +$$

$$\alpha ||p_i - (v_i^* + ty)||^2 =$$

$$\sum_{k=1}^n (u_{ik})^m \langle x_k - v_i^* - ty, x_k - v_i^* - ty \rangle +$$

$$\alpha \langle p_i - (v_i^* + ty), p_i - (v_i^* + ty) \rangle \quad (18)$$

其中, $\langle z, z \rangle = ||z||^2$ 是 $\mathbf{R}^s$ 内积。

$$\frac{dh_i(t)}{dt} = \sum_{k=1}^n (u_{ik})^m \langle -y, x_k - v_i^* - ty \rangle +$$

$$\alpha \langle -y, p_i - v_i^* - ty \rangle = -2 \times$$

$$\left[\sum_{k=1}^n (u_{ik})^m \langle y, x_k - v_i^* - ty \rangle + \right.$$

$$\left. \alpha \langle y, p_i - v_i^* - ty \rangle \right] \quad (19)$$

由式(19)可以得到:

$$\frac{dh_i(0)}{dt} = \langle y, \sum_{k=1}^n (u_{ik})^m (x_k - v_i^*) +$$

$$\alpha (p_i - v_i^*) \rangle = 0, \quad \forall y \in \mathbf{R}^s, y \neq 0 \quad (20)$$

如式(20)所示,方程对所有 $y \neq 0$ 成立的条件是当且仅当其第 2 项是零向量,即

$$v_i^* = \left[\sum_{k=1}^n (u_{ik})^m x_k + \alpha p_i \right] / \left[\sum_{k=1}^n (u_{ik})^m + \alpha \right] \quad (21)$$

于是引理 1 的必要性得证。

设 $y = (y_1, y_2, \dots, y_c)$, $y_i \in \mathbf{R}^s, t \in \mathbf{R}$, 定义:

$$h(t) = \phi(v^* + ty) =$$

$$\sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m ||x_k - v_i^* - ty_i||^2 +$$

$$\sum_{j=1}^c \alpha ||p_j - v_j^* - ty_j||^2 \quad (22)$$

那么

$$h'(t) = \sum_{k=1}^n \sum_{i=1}^c -2(u_{ik})^m \langle \mathbf{y}_i, \mathbf{x}_k - \mathbf{v}_i^* - t\mathbf{y}_i \rangle - 2 \sum_{j=1}^c \alpha \langle \mathbf{y}_j, \mathbf{x}_k - \mathbf{v}_j^* - t\mathbf{y}_j \rangle \quad (23)$$

并且

$$h''(t) = \mathbf{y}^T [H_\psi(\mathbf{v}^* + t\mathbf{y})] \mathbf{y} = \sum_{k=1}^n 2(u_{ik})^m \langle \mathbf{y}_i, \mathbf{y}_i \rangle + \sum_{j=1}^c 2\alpha \langle \mathbf{y}_j, \mathbf{y}_j \rangle \quad (24)$$

当 $t=0$ 时,对 $\forall \mathbf{y} \in \mathbf{R}^{c \times s}$ 有:

$$h''(0) = \mathbf{y}^T [H_\psi(\mathbf{v}^*)] \mathbf{y} = 2 \left[\sum_{i=1}^n ||\mathbf{y}_i||^2 \sum_{k=1}^n (u_{ik})^m \right] + 2 \sum_{j=1}^c \alpha ||\mathbf{y}_j||^2 = 2 \sum_{i=1}^n ||\mathbf{y}_i||^2 \left[\sum_{k=1}^n (u_{ik})^m + \alpha \right] \quad (25)$$

由此可知,对于 $\mathbf{y} \neq \mathbf{0}$, $h''(0) > 0$, 因此 $H_\psi(\mathbf{v}^*)$ 是正定的。证明完毕。

引理 2 设 $[\text{conv}(X)]^c$ 是 X 的笛卡尔积的凸包,令 $(\mathbf{U}^0, G(\mathbf{U}^0))$ 为 J_m 的初始迭代点,其中 $\mathbf{U}^0 \in M_{fc}$ 并且 $\mathbf{v}^0 = G(\mathbf{U}^0)$, 那么

$$(\mathbf{U}^k, \mathbf{v}^k) \in M_{fc} \times [\text{conv}(X)]^c \quad (26)$$

$M_{fc} \times [\text{conv}(X)]^c$ 是被包含于 $M_{fc} \times \mathbf{R}^{c \times s}$ 的紧集。

证明 令 $\mathbf{U}^0 \in M_{fc}$, 则 $\mathbf{v}^0 = G(\mathbf{U}^0)$ 通过下式计算得到:

$$\mathbf{v}_i^0 = \left[\sum_{k=1}^n (u_{ik}^0)^m \mathbf{x}_k + \alpha \mathbf{p}_i \right] / \left[\sum_{k=1}^n (u_{ik}^0)^m + \alpha \right] \quad (27)$$

设

$$\rho_{ik} = (u_{ik}^0)^m / \left[\sum_{l=1}^n (u_{il}^0)^m + \alpha \right], \quad 1 \leq k \leq n; \quad \rho_\alpha = \frac{\alpha}{\sum_{k=1}^n (u_{ik}^0)^m + \alpha} \quad (28)$$

可以看出: 当 $\forall i, k$ 时, $0 < \rho_{ik} < 1$; 当 $\alpha > 0$ 时, $0 < \rho_\alpha < 1$ 。则对 $\forall i$ 有:

$$\mathbf{v}_i^0 = \sum_{k=1}^n \rho_{ik} \mathbf{x}_k + \rho_\alpha \mathbf{p}_i \quad (29)$$

其中

$$\sum_{k=1}^n \rho_{ik} + \rho_\alpha = \frac{\sum_{k=1}^n (u_{ik}^0)^m}{\sum_{l=1}^n (u_{il}^0)^m + \alpha} + \frac{\alpha}{\sum_{l=1}^n (u_{il}^0)^m + \alpha} = \left[\sum_{k=1}^n (u_{ik}^0)^m + \alpha \right] / \left[\sum_{l=1}^n (u_{il}^0)^m + \alpha \right] = 1 \quad (30)$$

因此,对 $\forall i, \mathbf{v}_i^0 \in \text{conv}(X)$, 即 $\mathbf{v}^0 \in [\text{conv}(X)]^c$ 。

$\mathbf{U}^1 = F(\mathbf{v}^0) \in M_{fc}$, $\mathbf{v}^1 = G(\mathbf{U}^1) \in [\text{conv}(X)]^c$, 因此所有关于 J_m 的迭代结果都属于 $M_{fc} \times [\text{conv}(X)]^c$ 。

为了证明 $M_{fc} \times [\text{conv}(X)]^c$ 是紧的,注意到 X 是有限的,因此 X 的直径等于 $\text{conv}(X)$ 的直径,该直径是有界的。因为 $\text{conv}(X)$ 是由 $\mathbf{x}_k$ 生成的凸包,它是闭的,所以 $\text{conv}(X)$ 在 $\mathbf{R}^s$ 上是闭的 $\Rightarrow \text{conv}(X)$ 是紧的,根据 Heine-Borel 定理,得到的 $[\text{conv}(X)]^c$ 也是紧的。文献[13]也证明了 M_{fc} 是紧的。

因此 $M_{fc} \times [\text{conv}(X)]^c$ 是紧的。

证明完毕。

由引理 1、引理 2 以及文献[11]可以得到下述定理。

定理 设 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 包含至少 $n(n > c)$ 个点,

且 $J_m(\mathbf{U}, \mathbf{v}) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d(\mathbf{x}_k, \mathbf{v}_i) + \alpha \sum_{j=1}^c d(\mathbf{v}_j, \mathbf{p}_j)$, 其中 $1 < m < \infty, \alpha > 0, \mathbf{U} \in M_{fc}$; 且有 $\mathbf{v} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_c]^T$, 其中 $\mathbf{v}_i \in \mathbf{R}^s (\forall i)$; 设 $(\mathbf{U}^0, G(\mathbf{U}^0))$ 是关于 T_m 迭代序列的起始点,其中 $\mathbf{U}^0 \in M_{fc}$ 并且 $\mathbf{v}^0 = G(\mathbf{U}^0)$, 则迭代序列 $\{\mathbf{U}^r, \mathbf{v}^r\} (r=1, \dots)$ 收敛于一个在集合 Ω 中的点 $(\mathbf{U}^*, \mathbf{v}^*)$, 或其一个子序列收敛到 Ω 中的一个点。

5 改进聚类的分选流程

在雷达信号的各项参数中,载频 RF、脉冲宽度 PW、到达方向 DOA 等参数变化范围有限且特征明显,因此成为雷达信号分选中比较常用的分选参数。改进的 FCM 算法分选流程如下。

5.1 数据预处理

1) 归一化数据。由于 3 个参数的数值差异较大,为了消除量纲差异,使用归一化技术对数据进行处理,将数据转化到 $[0, 1]$ 之间,即

$$x_{\text{std}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (31)$$

2) 去除干扰点。由于在雷达数据的接收过程中可能会受到各种干扰,在正式聚类之前需要对干扰点进行处理。从数据场理论来看,干扰点的势值远小于数据场势值的平均值,因此,采用 $F(x) < \mu \bar{F}$ 来筛选干扰点, $F(x)$ 是干扰点的势值, μ 一般取值小于 1。

3) 确定初始聚类数。采用数据场来找到数据点的势值并且搜索势值的局部极大值,并将势心作为改进算法的约束点。

5.2 聚类流程

本文聚类算法的具体流程如下:

- 1) 将数据 X 归一化处理。

2) 使用数据场计算势值 $F(X)$ 。

3) 通过势值的平均值来筛选干扰点。

4) 找到势值的势心 $\hat{F}(X)$ 。

5) 将势心作为改进 FCM 聚类算法的约束点,进行聚类。

6) 输出聚类结果。

6 仿真实验

利用仿真实验来验证改进后 FCM 算法的有效性和可靠性,为此模拟了实验数据,本文用兰德系数(Rand index, RI)作为聚类算法的准确率评价指标。计算机仿真的实验环境为 Intel(R) Core(TM) i5-11400H@2.70 GHz, 2.69 GHz, Windows 10, Python 3.8。

6.1 仿真实验 1

实验 1 采用脉冲信号较为单一且脉冲数量少的 4 部雷达仿真数据,详细参数配置见表 1 所列。数据点的三维图如图 3 所示。

表 1 雷达仿真参数

雷达序号	DOA/(°)	RF/MHz	PW/ μ s	脉冲数
1	38.0~40.0	3 270~3 530 (捷变)	8.40~12.60 (抖动)	300
2	40.3~42.0	2 935~3 565 (捷变)	9.75~16.25 (抖动)	380
3	42.2~44.0	3 310~3 750 (捷变)	16.65~20.35 (抖动)	290
4	38.0~46.5	3 050~3 250 (捷变)	18.90~23.10 (抖动)	305

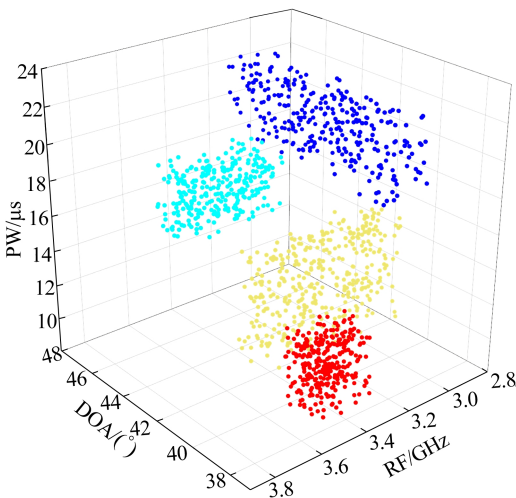


图 3 仿真数据示意图

该数据集的 DOA-PW 二维等势线示意图如

图 4 所示,从图 4 可以看出,等势线环绕的地方就是势心所在,与数据的真实类中心的位置较为接近。

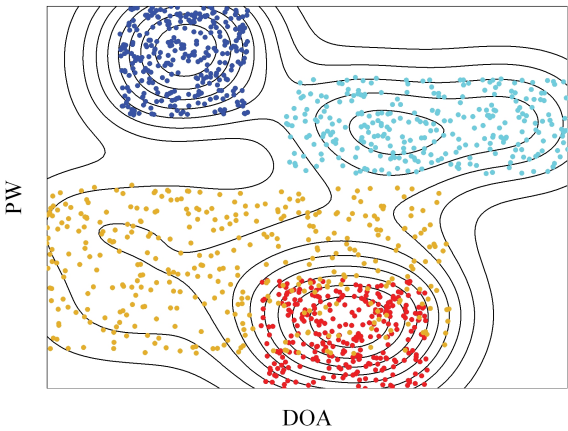


图 4 等势线示意图

根据数据场和式(8)找到 4 个势心,因此设置聚类数目为 4,并且将这些势心作为聚类的初始类中心。

利用 K-means、DBSCAN、FCM 算法与本文改进后的 FCM 算法对仿真数据进行分选,并计算聚类准确率进行对比,结果见表 2 所列。

表 2 实验 1 各仿真算法的准确率

算法	准确率/%
K-means	88.55
DBSCAN	76.47
FCM	86.04
改进 FCM	96.55

6.2 仿真实验 2

实验 2 模拟了脉冲丢失的情况,仿真数据见表 3 所列,其中雷达 3 的脉冲数减少了 50% 以上。各算法的聚类结果见表 4 所列。

表 3 脉冲丢失的仿真数据

雷达序号	DOA/(°)	RF/MHz	PW/ μ s	脉冲数
1	38.0~40.0	3 270~3 530 (捷变)	8.40~12.60 (抖动)	300
2	40.3~42.0	2 935~3 565 (捷变)	9.75~16.25 (抖动)	380
3	42.2~44.0	3 310~3 750 (捷变)	16.65~20.35 (抖动)	120
4	38.0~46.5	3 050~3 250 (捷变)	18.90~23.10 (抖动)	305

表 4 实验 2 各仿真算法的准确率

算法	准确率/%
K-means	76.19
DBSCAN	72.85
FCM	77.47
改进 FCM	96.65

6.3 复杂雷达环境下的仿真实验

上文采用了几个简单的雷达数据进行仿真实验,结果表明改进 FCM 算法具有较高的准确率。但在实际场景中,雷达信号更加复杂,可能出现雷达数量更多、信号数量更大以及各部雷达数量分布不均衡等情况。

为了验证本文改进 FCM 算法在这些复杂情况下的适用性以及分选准确率,从 3 个不同方向设置子实验,分别测试改进 FCM 算法在不同雷达数量、不同雷达信号数量以及不同雷达信号分布情况下的适用性与准确性。

1) 雷达数量。本实验采用 8 部雷达测试算法在不同的雷达数量下的表现,雷达的具体参数设置见表 5 所列。

表 5 多雷达仿真数据

雷达序号	DOA/(°)	RF/MHz	PW/ μ s	脉冲数
1	38.0~40.0	3 270~3 530	8.40~12.60	300
2	40.3~42.0	2 935~3 565	9.75~16.25	380
3	42.2~44.0	3 310~3 750	16.65~20.35	290
4	38.0~46.5	3 050~3 250	18.9~23.10	305
5	39.0~43.2	3 620~3 820	33.9~38.10	280
6	45.2~48.1	2 650~3 150	24.3~32.10	330
7	35.2~37.9	3 550~3 850	21.6~25.40	350
8	41.7~42.2	3 180~3 450	18.1~23.50	410

从前 3 个雷达开始,依次增加雷达信号,使用 3 种算法对雷达信号进行分选,得到的结果见表 6 所列。

表 6 不同雷达数量下 3 种算法的分选结果

雷达数量	准确率/%		
	改进 FCM	FCM	K-means
3	97.63	92.78	93.81
4	96.39	82.22	88.55
5	88.94	80.00	79.49
6	91.09	75.65	83.08
7	93.37	74.33	81.59
8	86.43	74.33	81.59

从表 6 可以看出,随着雷达数量的增加,各算法都受到了一定的影响,其中本文改进 FCM 算

法受到的影响相对较小,且比 FCM 和 K-means 算法具有更高的分选准确率。

2) 雷达信号数量。本实验使用仿真实验 1 的雷达信号参数,为了测试算法在不同雷达信号数下的表现,从每部雷达 400 信号数开始,不断增加雷达信号数,得到的结果见表 7 所列。

表 7 不同雷达信号数下 3 种算法的分选结果

雷达信号数	准确率/%		
	改进 FCM	FCM	K-means
400	96.00	86.94	74.63
500	95.10	89.55	90.50
600	94.29	88.58	90.50
700	96.61	84.50	90.89
800	95.31	88.25	91.31
900	94.05	88.44	90.38
1 000	95.90	73.97	89.15

从表 7 可以看出,雷达信号数量变化对各算法的影响较小,而本文改进 FCM 算法的准确率在不同的雷达信号数量下均有着最高的分选准确率。

3) 雷达信号分布。本实验使用仿真实验 1 的雷达信号参数,将每部雷达的脉冲信号数设置为 500。为了测试在雷达信号分布不均衡情况下算法的适用性,将其中的一些雷达信号数设为 150,称这样的雷达为缺失雷达。不同的缺失雷达下 3 种算法的分选结果见表 8 所列。

表 8 不同缺失雷达下 3 种算法的分选结果

缺失雷达 序号	准确率/%		
	改进 FCM	FCM	K-means
1	92.57	82.86	87.54
2	97.54	88.56	89.60
3	87.98	81.93	82.84
4	94.21	86.80	89.31
1/2	93.45	85.83	86.77
1/3	91.25	76.09	79.28
1/4	93.97	79.56	84.60
2/3	97.05	96.27	84.79
2/4	96.33	89.18	96.01
3/4	93.24	87.81	84.10
1/2/3	91.15	79.87	80.21
1/3/4	87.20	77.56	81.83
1/2/4	93.89	77.08	79.21
2/3/4	94.18	86.88	93.68

表 8 中:第 1 列表示缺失雷达的序号,若缺失雷达序号为 1,则表示将雷达 1 的信号数量设为 150,其余雷达的信号数量保持为 500,从而构造

出雷达 1 信号数量远低于其他雷达的情况;同理,缺失雷达序号为 1/2,则表示将雷达 1 和雷达 2 的信号数量均设为 150,依次类推。

通过设置不同的缺失雷达组合,可以生成多种雷达信号数量的分布情况,进而测试算法在不同分布下的分选表现。实验考虑了所有可能的缺失雷达组合。结果表明,信号分布的不均衡对各部雷达均有一定影响,但综合来看,本文改进 FCM 算法受到的影响较小,且比 FCM 和 K-means 算法具有更高的分选准确率。

6.4 实验总结

通过上述实验发现,改进 FCM 算法在雷达分选中拥有较好的适用性和准确性,并且在复杂环境中,如较多的雷达数、较多的雷达信号数以及雷达信号分布不均匀的情况,该算法依然保持较高的准确性。由此可以认为,在复杂环境下,本文改进 FCM 算法能够保持较好的适用性和准确性,相较于改进前的 FCM 算法,性能有较大的提高。

7 结 论

本文针对雷达信号的分选问题,提出了一种基于数据场和 FCM 聚类的分选算法。该算法采用了类中心约束点的思想,将数据场的势心作为约束点,将聚类中心迭代限制在约束点附近,从而减少干扰。仿真实验验证了该算法的有效性,实验结果表明改进 FCM 算法的性能相较于原算法有显著提升。

本文改进 FCM 算法的优点如下:① 克服了 FCM 算法需要预先设置聚类数目的缺点,通过数据场自动设定聚类数目;② 相较于经典算法能够得到更好的分选结果;③ 面对雷达信号数量不均衡的情况,改进后的算法能发挥很好的效果。

但是改进后的 FCM 算法引入了约束力度 α ,加大了调参的难度。虽然通过实验验证了改进 FCM 算法在复杂环境下依然有着较好的适用性,但是现实中的雷达环境更加复杂,并且因为多参

数分选使用的分选参数并不包含太多的雷达时域参数,所以对于当前雷达环境中雷达数目的变化不太敏感。改进 FCM 算法在现实环境中的适用情况,还需要进行更多的研究与实验。

[参 考 文 献]

- [1] 隋金坪,刘振,刘丽,等. 雷达辐射源信号分选研究进展[J]. 雷达学报,2022,11(3):418-433.
- [2] 何炜. 雷达信号分选关键算法研究[D]. 成都:电子科技大学,2007.
- [3] 刘鲁涛,王璐璐,李品,等. DSets-DBSCAN 无参数聚类的雷达信号分选算法[J]. 国防科技大学学报,2022,44(4):158-163.
- [4] 王星,陈相,周一鹏,等. 一种基于改进 DBSCAN 的雷达信号分选算法[J]. 空军工程大学学报(自然科学版),2021,22(3):47-54.
- [5] 赵贵喜,骆鲁秦,陈彬. 基于蚁群算法的 K-Means 聚类雷达信号分选算法[J]. 雷达科学与技术,2009,7(2):142-146.
- [6] 张冉,夏厚培. 一种新的 k-means 聚类雷达信号分选算法[J]. 现代防御技术,2015,43(6):136-141.
- [7] 黎蓉,刘以安. 基于支持向量聚类改进的雷达信号分选方法[J]. 电子信息对抗技术,2014,29(5):40-43,54.
- [8] 李文君,鱼佳欣,周春来. 一种快速的支持向量聚类雷达信号分选方法[J]. 现代电子技术,2013,36(17):65-67.
- [9] 李德毅,杜鹏. 不确定性人工智能[M]. 北京:国防工业出版社,2005:193-207.
- [10] 陈涛. 基于数据场与云模型的雷达信号分选算法研究[D]. 哈尔滨:哈尔滨工程大学,2020.
- [11] HATHAWAY R J, BEZDEK J C, TUCKER W T. An improved convergence theory for the fuzzy c-means clustering algorithms[J]. Analysis of Fuzzy Information, 1987, 3:123-131.
- [12] BEZDEK J C. A convergence theorem for the fuzzy ISO-DATA clustering algorithms[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1980, 2(1):1-8.
- [13] BALL G H, HALL D J. A clustering technique for summarizing multivariate data[J]. Behavioral Science, 1967, 12(2):153-155.

(责任编辑 胡亚敏)