

一种基于改进 DBSCAN 的雷达信号分选算法

王星, 陈相, 周一鹏, 陈游, 肖冰松, 王洪迅

(空军工程大学航空工程学院, 西安, 710038)

摘要 针对传统 DBSCAN 算法参数设置依靠人工经验的不可靠性, 并且对非均匀数据聚类效果差的问题, 基于云模型(Cloud Model)提出了一种 CMDSCAN 算法, 算法首先结合距离曲线倾角突变的特点自适应获得邻域半径, 并根据雷达信号分布密度设置聚类密度点数阈值, 可实现 DBSCAN 算法自适应运行; 同时结合多维云模型理论, 对 DBSCAN 算法分选结果进行有效性评估, 利用判定结果进一步优化参数设置。根据仿真模拟的复杂对抗过程中帧收的雷达信号进行实验, 证明该算法可实现非均匀雷达信号的自适应分选, 同时可有效避免在多功能雷达信号分选中的“增批”问题。

关键词 DBSCAN; 聚类; 云模型; 雷达信号分选

DOI 10.3969/j.issn.1009-3516.2021.03.008

中图分类号 TN97 **文献标志码** A **文章编号** 1009-3516(2021)03-0047-08

A Radar Signal Sorting Algorithm Based on Improved DBSCAN Algorithm

WANG Xing, CHEN Xiang, ZHOU Yipeng, CHEN You, XIAO Bingsong, WANG Hongxun

(Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract Aimed at the problems that the parameter setting of traditional DBSCAN algorithm relies on the unreliability of manual experience, and the heterogeneous data clustering effect is poor, A CMDSCAN model is proposed based on Cloud model. Firstly, according to the characteristics of the distance curve angle mutation, the adaptive neighborhood radius is obtained. And cluster density threshold is set according to the distribution density of radar signal, realizing DBSCAN algorithm run adaptively. And then, in combination with the theory of multidimensional cloud model, the validity of DBSCAN algorithm sorting results is evaluated, and the decision results are used to further optimize parameter setting. The experiment is carried out according to the radar signal received during the complex countermeasure process. The result shows that the algorithm can realize the adaptive sorting of non-uniform radar signals and effectively deal with the “batch” problem in multi-mode radar signal sorting.

Key words DBSCAN; clustering; cloud model; radar signal sorting

信息化条件下, 电磁空间成为继海、陆、空、天之后的第五维战场空间, 电子侦察系统通过截获电磁空间中密集的信息流, 为战场态势感知和辅助干扰

决策提供情报支撑^[1]。雷达信号分选是电子侦察系统的初始环节, 信号分选问题本质上是将在侦察接收机侦收到的不同雷达辐射源信号进行分离的过程,

收稿日期: 2020-12-21

基金项目: 航空科学基金项目(20175596020)

作者简介: 王星(1965—), 男, 辽宁大连人, 教授, 博士生导师, 研究方向: 电子对抗技术。E-mail: 13399289501@189.cn

引用格式: 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): 47-55. WANG Xing, CHEN Xiang, ZHOU Yipeng, et al. A Radar Signal Sorting Algorithm Based on Improved DBSCAN Algorithm[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(3): 47-55.

是电子战得以成功实施的重要前提,对整个电子战系统具有重要的影响。

从机器学习的角度来看,聚类过程是无监督学习过程^[2-3],聚类算法能够在无标记训练样本的条件下将原始数据进行分组,从而发现数据的内在结构。目前常用的无监督聚类算法主要包括:基于层次、基于密度、基于划分等几大类。DBSCAN 算法是一种密度聚类算法,它不需要先验的聚类簇数量,并且能够分选出各种形状的聚类簇^[4],算法复杂度低,适合于雷达信号的自分选。但该算法对聚类参数较敏感,且聚类参数是根据先验知识人为设定的,可靠性不高;同时 DBSCAN 算法适合于处理密度相对比较均匀的数据,当不同簇的密度变化较大时,DBSCAN 分选效果不佳。为了克服传统 DBSCAN 算法的缺陷,李文杰等^[5]提出的 KANN-DBSCAN 算法结合参数寻优策略,利用数据集自身的分布特性寻找最优分选参数,无需人为赋值,便能够实现全过程无监督聚类,但该算法复杂度较高,在处理数据量较大的数据集时效果不佳;OPTICS 算法^[6]是根据 DBSCAN 算法思想演化而来的一种基于层次密度的分选算法,最终得到带有可达距离和核心距离的输出数据,该算法对参数不敏感,但仍然需要聚类参数的人为输入,对于不同的输入,聚类结果有较大差异;胡健等提出 IABC-DBSCAN 算法^[7],利用人工蜂群算法强大的全局和局部优化特征寻找 DBSCAN 算法中的最优聚类参数,可有效克服 DBSCAN 算法对聚类参数敏感问题,但并未完全解决 DBSCAN 算法全局参数设置问题。

随着新体制雷达信号处理能力的不断提高,波形参数捷变、数字波束合成等数字化技术普遍应用,使得新体制雷达信号的变化方式日益复杂,同时雷达与侦察接收机在空间上的距离位置、雷达的不同工作状态、以及雷达的不同工作体制等使得接收机侦收的雷达信号具有非均匀、多密度、脉冲数据量大、脉冲参数交叠严重等特点^[8]。并且先进体制雷达如相控阵雷达的工作模式多样^[9],往往集多种体制于一身,使得侦收到的信号具有很大的不确定性。在实际分选处理中,往往会将同一雷达的不同模式分选成多部雷达,造成分选“增批”问题,给雷达信号分选带来了巨大挑战。

针对上述问题,本文提出了一种 CMDDBSCAN 算法,该算法首先结合距离曲线倾角突变的特点自适应获得邻域半径,并根据雷达信号在参数空间中的分布密度,相应地调整密度点数阈值,进而能够根据空间的密度分布特点自适应生成分选参数;之后

结合多维云模型理论,利用逆向云发生器算法和三维正向高斯云算法^[10],对 DBSCAN 分选结果进行有效性评估,利用反馈的判定结果,进一步优化 DBSCAN 算法邻域参数,提高 DBSCAN 算法的分选准确性和可靠性。

1 改进 DBSCAN 分选算法

1.1 DBSCAN 算法聚类原理

DBSCAN 算法以数据的稠密程度作为划分簇的依据,分类效果主要由一组“邻域”参数(Eps , $MinPts$)决定。给定数据集 $D = \{x_1, x_2, \dots, x_m\}$,给出几个重要概念的定义:

定义 1 Eps 邻域。也称为 ϵ 邻域,对于数据点 $x_j \in D$,其 Eps 邻域定义为样本集 D 中与 x_j 距离不大于 Eps 的样本对象 x_i 的集合,即 $N_{Eps}(x_j) = \{x_i \in D | dist(x_i, x_j) \leq Eps\}$ 。其中, $dist(x_i, x_j)$ 表示 2 个样本点之间的距离刻画。

定义 2 核心对象。给定整数 $MinPts$,若样本 x_j 的 Eps 邻域范围内包含的样本数量不小于 $MinPts$,即满足 $|N_{Eps}(x_j)| \geq MinPts$,则称 x_j 为 $(Eps, MinPts)$ 条件下的核心对象。

定义 3 密度直达。若样本 x_j 在核心对象 x_i 的 Eps 邻域范围内,即满足 $x_j \in N_{Eps}(x_i)$ 且 $N_{Eps}(x_i) \geq MinPts$,则称 x_j 是从 x_i 密度直达的。

定义 4 密度相连。对于对象 x_j 和 x_i ,若存在一个对象 x_k ,使得 x_j 和 x_i 与 x_k 均满足密度可达关系,则称 x_j 和 x_i 是密度相连的。

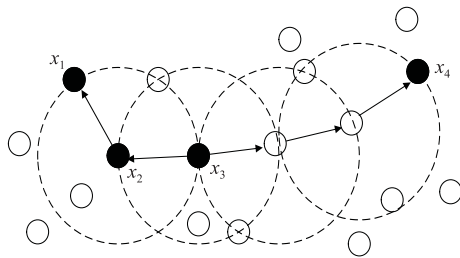


图1 DBSCAN 定义示意图

图1中直观地显示了以上定义,图中虚线圆是 Eps 邻域,设定 $MinPts = 3$ 。DBSCAN 算法首先遍历出样本集 D 中所有核心对象,任意选择其中一个作为初始样本,然后找到这个核心对象最大密度相连样本集作为一个聚类类别,最后剔除已分类的样本,从剩余的核心对象中任选一个生成下一个聚类簇,重复以上步骤,直到核心对象集合为空。

1.2 DBSCAN 算法改进

DBSCAN 算法聚类准确性与邻域参数(Eps , $MinPts$)密切相关,一般情况下邻域参数 Eps 和

$MinPts$ 是根据人工经验设定的,具有不确定性。因此,研究如何自适应获取邻域参数对提高算法准确度和稳定性具有重要意义。

1.2.1 自适应获取邻域参数 Eps

为解决参数设置问题,文献[11]采用 K 距离曲线法确定参数。根据文献[1],实现成功分选雷达信号所需的脉冲数通常取 4~7,首先假定参数 $MinPts = 4$,遍历整个样本集 D ,计算样本 x_i 与周围点的距离,找到最近第 $MinPts$ 个样本点距离 $Dist(x)$,求出序列: $\{Dist(x_1), Dist(x_2), \dots, Dist(x_m)\}$;之后将序列按数值大小排序并绘成曲线,如图 2 所示。将图 2 中曲线开始快速上升处(即 A 点位置)的 $Dist$ 值设置为 Eps 邻域。

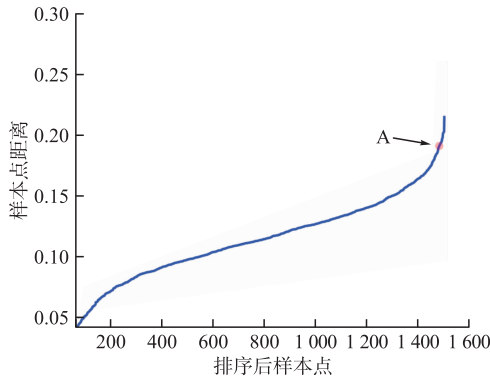


图 2 排序的 $Dist_{min}$ 曲线

上述方法寻找 A 点依然需要人为参与,本文对此方法进行改进,首先将序列 $\{Dist(x_1), Dist(x_2), \dots, Dist(x_m)\}$ 排序后记为 $\{d_i | i=1, 2, \dots, m\}$,对序列 $\{d_i\}$ 进行差分处理,得到差分序列 $\{\Delta d_i\}$:

$$\Delta d_i = d_{i+1} - d_i \quad (1)$$

由于在 A 点距离曲线倾角变化最为显著,因此以倾角的变化作为依据寻找 A 点,进一步求 $\{\Delta d_i\}$ 的反三角函数,求得曲线在各点的倾角:

$$\theta_i = \arctan(e^{\frac{\Delta d_i}{\alpha}}) \quad (2)$$

式中: α 是缩放因子,避免由量级引起的 Δd_i 的数值过大或过小。得到曲线各点的倾角后,再对角度进行差分处理,求出角度差分序列 $\{\Delta \theta_i\}$:

$$\Delta \theta_i = \theta_{i+1} - \theta_i \quad (3)$$

最后寻找差分序列 $\{\Delta \theta_i\}$ 的最大值所在位置 $P_{\max}$, $P_{\max}$ 所在的位置即对应到图 2 距离曲线倾角变化最为显著的 A 点,从而确定出 Eps 值。

1.2.2 设定变密度点数阈值

在战场环境中,侦察接收机侦收的雷达信号在参数空间中的分布具有非均匀、多密度的特点。这种非均匀、多密度分布的形成原因与雷达和侦察接收机的距离、雷达工作状态以及雷达工作体制密切相关。

若采用全局参数 ($Eps, MinPts$),聚类过程中的判别标准相对单一,对于非均匀、多密度的雷达信号分选效果不佳。因此,在自适应求得邻域半径值之后,利用 Eps 值对雷达信号空间密度进行描述,增强算法对多密度雷达信号的适用性。

首先利用 1.2.1 中的方法自适应获得 Eps 值,之后计算数据集 D 中每一个样本点的 Eps 邻域范围包含的雷达样本点个数,将该数值记为样本点邻域密度特征值 K ,按照降序的方式将数据点邻域密度特征值 K 进行重新排序,数据排列的顺序表明了样本点所在簇密度大小。

计算整个数据集邻域密度特征值 K 的均值 $\bar{K}_0$,对于特征值 K 大于 $\bar{K}_0$ 的样本点,以 $\bar{K}_0$ 作为点数阈值进行聚类,并将聚类结果从数据集 D 中清除;对于邻域密度特征值 K 小于 $\bar{K}_0$ 的样本点,计算所有剩余数据点的特征值均值,记为 $\bar{K}_1$,并以 $\bar{K}_1$ 为点数阈值对剩余数据进行聚类,以此类推,由于是在假定 $MinPts=4$ 的情况下求得的 Eps 值,所以当点数阈值 $\bar{K}_i < 4$ 时,停止运算。

1.3 改进 DBSCAN 算法实现步骤

信号参数由脉宽(PW)、载频(RF)和达到角(DOA)3种特征组成。已知样本空间中 n 个雷达信号组成的样本集 $D = \{x_1, x_2, \dots, x_n\}$,其中 $x_i = (PW_i, RF_i, DOA_i)$,基于改进 DBSCAN 聚类算法具体流程如图 3 所示。改进 DBSCAN 算法具体步骤:

输入:雷达信号数据集 D

步骤 1 利用 K 距离曲线法计算 $MinPts = 4$ 时的距离序列 $\{Dist(x_1), Dist(x_2), \dots, Dist(x_m)\}$;

步骤 2 引入差分的方法,利用反三角函数对距离序列预处理,自动获取邻域参数 Eps 值;

步骤 3 计算数据集 D 中每一个样本点的邻域密度特征值 K ,即在 Eps 邻域范围包含的样本点个数,并按照逆序排列;

步骤 4 计算数据集邻域密度特征值 K 的均值 $\bar{K}_0$,对于特征值大于 $\bar{K}_0$ 的样本点,以 $\bar{K}_0$ 作为点数阈值进行聚类,并将聚类结果从数据集 D 中清除;以此类推,计算剩余数据邻域密度特征值均值 $\bar{K}_i$,对于特征值大于 $\bar{K}_i$ 的样本点,以 $\bar{K}_i$ 作为点数阈值进行聚类,直到点数阈值 $\bar{K}_i < 4$ 时,停止运算。

输出:最终的簇划分 $C = \{C_1, C_2, \dots, C_k\}$

2 基于三维云模型的隶属度均值

2.1 云模型基本概念

云模型是一种可以实现在定性概念与定量数据

之间不确定性转换的模型,反映了概念中随机性和模糊性^[12]之间的关联。云的数字特征用3个数值进行表征:期望值 Ex 、熵 En 、超熵 He ,它们反映了定性概念 C 在总体上的定量特征^[14]。其中, Ex 表示云滴在论域空间 U 上分布的期望值,是最能反映该定性概念的点;熵 En 表示对定性概念 C 的不确定性度量,反映总体的离散程度^[15];超熵 He 是对熵的不确定性度量,反映数据的凝聚程度。

设 U 代表用精确数值表示的论域, U 可以是一维或高维的, $C(Ex, En, He)$ 表示 U 上的一个定性概念,若定量数据 $x(x \in U)$ 是定性概念 $C(Ex, En, He)$ 的一次随机实现,则 x 在论域 U 上的分布是一个隶属云,每个 x 是隶属云上的一个云滴^[13]。

2.2 云模型隶属度均值计算

将基于改进 DBSCAN 聚类算法的每个分类簇看作一种定性概念,设第 m 个簇的信号样本 n 为 x_{nmk} , $m=1,2,\dots,M, n=1,2,\dots,N, k=1,2,3$,其中 M 代表分选簇的总数目, N 代表每个簇中信号样本的总数,本文 $k=3$,选取原始信号内的特征参数 PW 、 RF 、 DOA 作为分选向量,并且设 x_{nm1} 为信号样本 n 的 DOA 参数, x_{nm2} 为 RF 参数, x_{nm3} 为 PW 参数。

改进 DBSCAN 算法聚类结果之间的相似度可以根据云模型隶属度进行判定,云模型隶属度均值计算原理如图4所示,主要包括逆向云模型发生器算法和三维正向高斯云算法两部分内容。

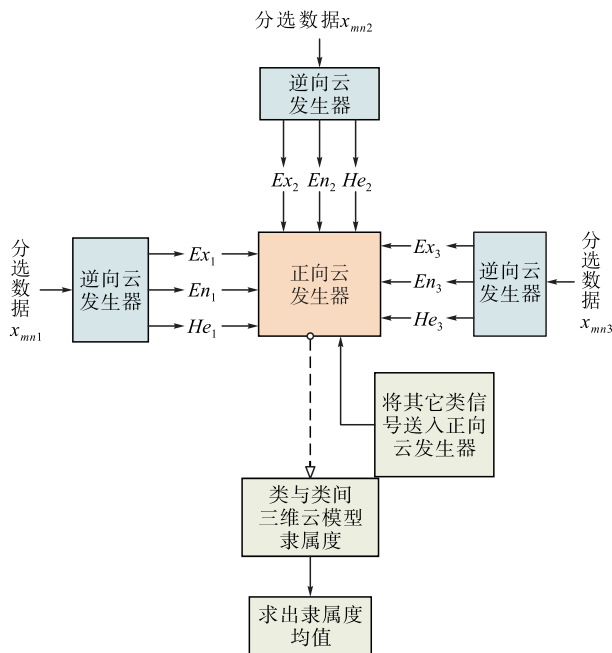


图4 云模型隶属度均值计算流程

2.2.1 逆向云发生器原理

由于输入样本点数据不带有确定度信息^[16],无确定度逆向云发生器算法主要根据高斯云分布的数

学性质求出 Ex_k, En_k 和 He_k 3种数字特征^[17],逆向云发生器如图5所示。

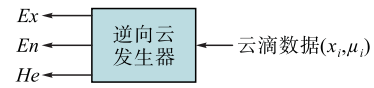


图5 逆向云模型发生器

根据样本数据的一阶绝对中心矩和二阶中心矩逆向云发生器算法^[18],求出样本一阶绝对中心矩和二阶绝对中心矩,如式(5)、(6)所示:

$$Ex_k = \frac{1}{N} \sum_{n=1}^N x_{nmk}, k=1,2,3 \quad (4)$$

$$\frac{1}{N} \sum_{n=1}^N |x_{nmk} - Ex_k|, k=1,2,3 \quad (5)$$

$$S = \frac{1}{N-1} \sum_{n=1}^N (x_{nmk} - Ex_k)^2, k=1,2,3 \quad (6)$$

利用高斯云分布的性质^[19]:在 $0 < He < En/3$ 条件下,高斯云分布一阶绝对中心矩如式(7),求出熵 En_k ,根据方差 $D(X)$ 求出超熵 He ,如式(8)、(10)所示:

$$E\{|X - Ex_k|\} = \sqrt{\frac{2}{\pi}} En_k, k=1,2,3 \quad (7)$$

$$En_k = \sqrt{\frac{\pi}{2}} \times \frac{1}{N} \sum_{n=1}^N |x_{nmk} - Ex_k|, k=1,2,3 \quad (8)$$

$$D(X) = En_k^2 + He_k^2, k=1,2,3 \quad (9)$$

$$He_k = \sqrt{S - En_k^2}, k=1,2,3 \quad (10)$$

2.2.2 基于三维正向高斯云算法的隶属度计算

根据逆向云发生器得到的数字特征,将分选信号的 PW 、 RF 、 DOA 三维特征参数进行整合,每一分选类的总体数字特征可以表示为 $(Ex_1, Ex_2, Ex_3, En_1, En_2, En_3, He_1, He_2, He_3)$,根据云模型理论,每一分选类可以抽象为一个定性概念,分选结果的总体数字特征是每一分选类的定性概念在总体上的定量描述。

根据多维正向云发生器原理^[20],见图6,生成雷达辐射源信号的三维隶属云模型。首先产生一个以三维数据 (En_1, En_2, En_3) 为期望, (He_1^2, He_2^2, He_3^2) 为方差的三维高斯随机数 (En_1', En_2', En_3') ,其中每一维度的 En_k' 均是以 En_k 为期望, He_k^2 为方差生成的高斯随机数,其中 $k=1,2,3$ 。

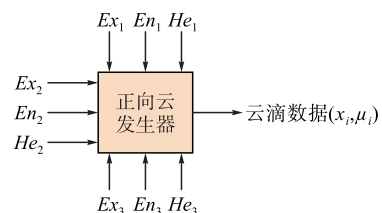


图6 三维正向云模型发生器

根据云模型理论,钟形隶属函数对模糊概念的分析具有普遍性,生成三维隶属云模型:

$$\mu_j(x_{il}) = \exp \left[-\frac{(x_{i1} - Ex_{j1})^2}{2En_{j1}} - \frac{(x_{i2} - Ex_{j2})^2}{2En_{j2}} - \frac{(x_{i3} - Ex_{j3})^2}{2En_{j3}} \right] \quad (11)$$

式中: $\mu_j(x_{il})$ 代表分选簇 i 中的第 l 个信号对分选簇 j 构成的三维云模型的隶属度, 其中 $i, j = 1, 2, \dots, M, l = 1, 2, \dots, N, x_{il}$ 代表第 i 个分选簇中第 l 个信号样本, $Ex_{jk}, k = 1, 2, 3$ 表示第 j 个分选簇信号的第 k 维特征期望值。

2.2.3 聚类结果隶属度均值计算

式(11)表示其他分选簇中的三维信号样本对某一分选簇构成的云模型相似度。利用式(12)求出簇与簇间隶属度均值 $E_{\mu_{ji}}$ 进行分选簇之间的相似度判定:

$$E_{\mu_{ji}} = \frac{1}{N} \sum_{l=1}^N \mu_j(x_{il}) \quad (12)$$

从多维云模型理论分析, 每一个分选簇都被抽象成了三维云模型所代表的定性概念, 隶属度均值 $E_{\mu_{ji}}$ 表示聚类结果 i 中所有三维信号样本隶属于某一定性概念程度的平均度量^[21]。

3 CMDSCAN 算法分选模型

利用基于三维云模型的隶属度均值和目标脉冲到达角参数实现改进 DBSCAN 分选结果有效性评估的 CMDSCAN 分选模型如图 7 所示。

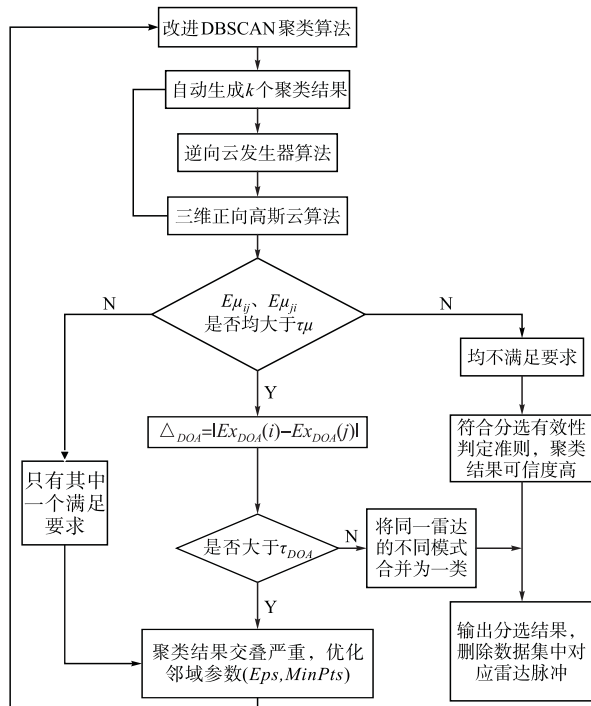


图 7 CMDSCAN 分选模型

具体求解步骤如下:

步骤 1 首先结合距离曲线倾角突变的特点自适应获得邻域半径, 并根据雷达信号在参数空间中的分布密度, 相应地调整密度点数阈值, 进而能够根据空间的密度分布特点自适应生成分选参数;

步骤 2 将基于改进 DBSCAN 算法分选簇抽象成定性概念, 利用逆向云发生器算法和三维正向高斯云算法得到不同分选簇之间的隶属度均值 $E_{\mu_{ji}}$ 。

步骤 3 选取每个聚类结果的脉冲到达角 DOA 均值 Ex_{DOA} 作为该目标辐射源的 DOA 参数。

步骤 4 利用如下规则进行评估判定:

1) 若隶属度均值 $E_{\mu_{ji}}, E_{\mu_{ij}}$ 均小于自适应阈值生成算法^[22]求出的 τ_{μ} , 说明雷达辐射源信号间相似程度低, 聚类分选结果正确度较高, 符合分选结果有效性准则, 直接输出分选结果, 并将对应类的雷达脉冲数据从数据集中删除。

2) 若聚类结果 i 与 j 相互之间的隶属度均值 $E_{\mu_{ji}}, E_{\mu_{ij}}$ 均大于 τ_{μ} 时, 计算聚类结果间脉冲到达角均值 Ex_{DOA} 的差值 $\Delta_{DOA} = |Ex_{DOA}(a) - Ex_{DOA}(b)|$:

a. 若差值 Δ_{DOA} 小于设定阈值 τ_{DOA} , 说明 2 个聚类结果之间有着较大的聚类相似度, 认为第 i, j 个分选簇为一部雷达的多种工作模式, 将 2 个分选簇合并为一类, 输出分选结果, 并将对应类的雷达脉冲数据从数据集中删除;

b. 若差值 Δ_{DOA} 大于设定阈值 τ_{DOA} , 表示虽然 2 个聚类结果之间有着较大的聚类相似度, 但并不属于同一辐射源, 将分选结果保留在数据集中。

3) 当隶属度均值 $E_{\mu_{ji}}, E_{\mu_{ij}}$ 只有其中一个大于设定阈值 τ_{μ} 时, 说明第 i 类与第 j 类之间信号交叠严重, 将分选结果保留在数据集中。

步骤 5 将评估结果反馈到 DBSCAN 算法中, 优化邻域参数 ($Eps, MinPts$), 对数据集中剩余脉冲重新进行计算, 使最终聚类符合分选结果有效性准则。

步骤 6 输出最终分选结果。

4 仿真实验分析

为了模拟实际战场电磁环境, 假设空域内共有 20 个平台参与作战想定, 我方侦察接收机部署于某一飞机平台, 通过仿真实验模拟平台间的复杂对抗过程, 仿真结束后导出侦察接收机帧收的信号全脉冲参数。

在仿真作战过程中, 不同作战平台的雷达装备服务于不同作战任务, 并且不同作战平台在整个作

战过程中的参与阶段和介入时间根据作战过程的演变而变化,其雷达的开关机时间也相应变化,因此帧收到的雷达信号具有非均匀、多密度、数据量大等特点。

首先通过基于载频和脉冲到达角分区方法^[1]对海量脉冲数据进行稀释,选取脉冲数据较丰富的第 2 分钟内的 X 波段,方位在 80°~160°内的 941 312 个脉冲作为实验数据,如图 8~10 所示:

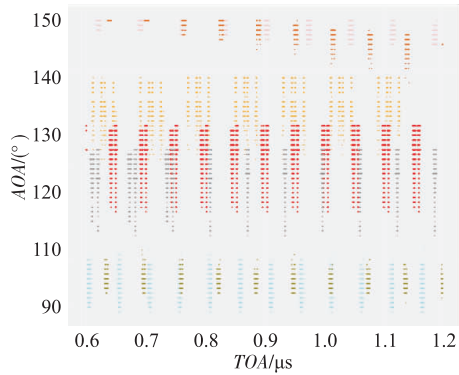


图 8 脉冲到达角-到达时间二维图

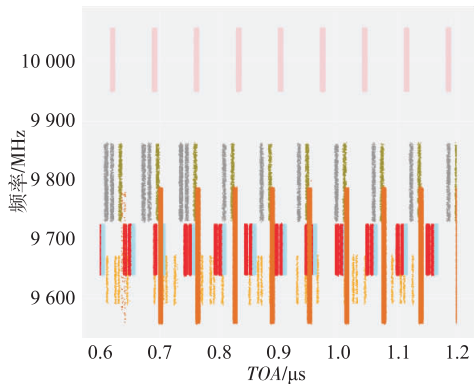


图 9 载频-脉冲到达时间二维图

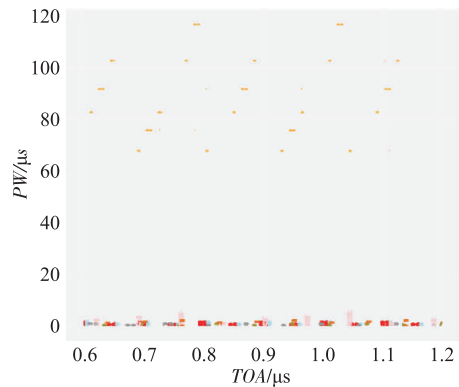


图 10 脉宽-脉冲到达时间二维图

从稀释数据中抽取 10 个信号片段,每个信号片段的脉冲数量为 30 000 条,对比本文算法、原始 DBSCAN 算法、KANN-DBSCAN 算法的分选性能。本文算法的邻域参数 ($Eps, MinPts$) 通过结合距离曲线倾角和邻域密度特征值自适应获得, KANN-DBSCAN 算法的分选参数通过参数寻优策

略确定^[5],原始 DBSCAN 算法参数通过人工赋值。

聚类效果内部指标用 Calinski-Harabasz 分数 (CHI) 表示,计算方法如下:

$$CHI = \frac{\text{tr}(\mathbf{B}_k) \cdot N - k}{\text{tr}(\mathbf{W}_k) \cdot k - 1}$$

(13)

式中: CHI 表示 k 个聚类结果簇内散度与簇间散度之比;其中 $\mathbf{B}_k$ 为分簇之间的协方差矩阵; $\mathbf{W}_k$ 为分簇内部的协方差矩阵。 CHI 指标取值越高表示分簇之间距离越大,簇内越紧密,聚类效果越好。

聚类效果外部指标用 V-测度衡量, V-测度 (VM) 由 Andrew Rosenberg^[23] 等人提出,是基于条件熵衡量聚类结果与真实分簇的同质性和完整性的指标,其定义为:

$$VM = \frac{2HC}{H + C}$$

(14)

本文 V-测度的取值范围为 $[0, 1]$, 取值越大说明分选效果越好。

表 1~3 为 3 种算法在各信号片段上的聚类分选结果,表 4 为 3 种算法聚类分选指标平均值的对比结果。

表 1 DBSCAN 算法实验结果

脉冲 片段	Eps	$MinPts$	VM	CHI	时间/s	分选准 准确率/%
1	0.060 1	5	0.773	27 009	269.14	73.21
2	0.055 9	6	0.856	35 256	285.03	68.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9	0.048 3	5	0.802	31 525	280.23	73.59
10	0.067 3	6	0.759	32 041	282.06	70.43

表 2 KANN-DBSCAN 算法实验结果

脉冲 片段	Eps	$MinPts$	VM	CHI	时间/s	分选准 准确率/%
1	0.055 3	4	0.870	28 569	240.87	83.21
2	0.056 4	6	0.632 6	30 112	239.58	87.52
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9	0.060 2	5	0.901	45 008	240.09	89.03
10	0.049 8	8	0.883	40 158	243.20	87.88

表 3 本文算法实验结果

脉冲 片段	Eps	VM	CHI	时间/s	分选准 准确率/%
1	0.062 3	0.941	109 390	196.71	96.32
2	0.059 2	0.929	41 279	198.00	93.56
⋮	⋮	⋮	⋮	⋮	⋮
9	0.042 1	0.957	68 223	196.38	95.88
10	0.075 5	0.933	58 564	195.20	97.11

表 4 3 种算法实验指标平均值对比				
算法	VM	CHI	时间/s	分选率/%
DBSCAN	0.812	31 641	280.56	72.53
KANN-DBSCAN	0.874	39 553	241.32	87.60
本文算法	0.946	68 381	196.72	94.86

从表 1~4 可以看出,CMDBSCAN 算法的内部指标 CHI 远大于原始 DBSCAN 和 KANN-DBSCAN 算法,因为 DBSCAN 算法结合多维云模型,利用聚类结果间隶属度均值进行相似性判断,可有效提高 DBSCAN 算法对多模雷达信号的分选能力,避免分选结果中的“增批”问题。CMDBSCAN 算法的聚类结果通过云模型进行有效性评估,使得聚类结果簇间分离程度更大,同时簇内数据聚合程度更好,内部指标 CHI 更高,分选效果有较大提高。

外部指标 VM 反映了聚类结果与真实数据之间的相关性,与分选准确率指标具有相似性。利用抽取的 10 个仿真数据片段,从表 1~3 中可以观察到本文算法在外部指标 VM 、分选准确率上都优于原始 DBSCAN 算法和 KANN-DBSCAN 算法,可有效提高分选准确率。

由于 DBSCAN 算法需要检索每个雷达脉冲点邻域内所有数据点,时间复杂度为 $O(n^2)$,其中 n 为全脉冲雷达数据点数,由于 KANN-DBSCAN 算法结合参数寻优策略,迭代计算复杂度与参数 K 密切相关,时间复杂度为 $O(Kn\ln b_n)$,在文中雷达脉冲数据集条件下 $K=62$ 。而本文采用 K - d 树结构,可有效遍历特定点给定距离内的雷达脉冲点,时间复杂度变为 $O(n\ln b_n)$,同时本文引入云模型理论对分选结果进行有效性评估,三维云模型的时间复杂度为 $O(3n)$ 。综上 CMDBSCAN 算法的时间复杂度为 $O(n\ln b_n)+O(3n)$ 。在大数据量条件下,原始 DBSCAN 算法对 2 个分选参数敏感,需要一定时间调试,并且时间复杂度呈指数式增长,CMDBSCAN 算法可有效解决上述问题。

以片段 10 为例,图 11~14 分别展示了片段 10 中原始数据和 3 种算法的聚类结果分布,各参数取值范围为 $[0,1]$,且各参数的单位取值为 1。从图中可以看出:原始的 DBSCAN 聚类算法结果存在明显的聚类误差,将 4 个辐射源的 30 000 条数据归为了 6 类,由于原始的 DBSCAN 算法采用全局固定的邻域参数 ($Eps, MinPts$),对于密度分布不均匀的雷达信号,分选效果较差。KANN-DBSCAN 算法结合参数寻优策略,无需人为赋值,能够实现全过程无监督聚类,如图 13 所示,分选效果有较大改进,能够实现雷达信号的基本分选,但是在处理非均匀、多密

度的雷达信号时仍然存在一定误差,因为 KANN-DBSCAN 参数寻优方法求出的分选参数 ($Eps, MinPts$) 依然是全局参数,未能改变 DBSCAN 算法的固有缺陷。

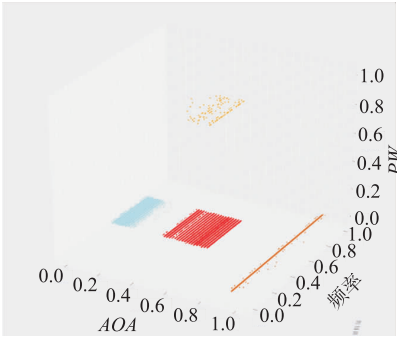


图 11 片段 10 原始数据分布

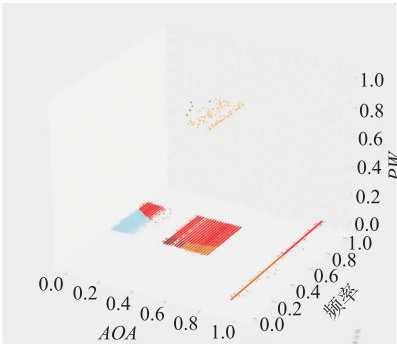


图 12 原始 DBSCAN 聚类分选结果

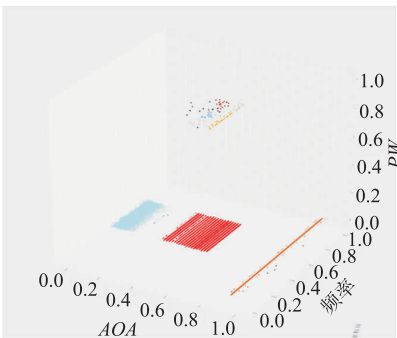


图 13 KANN-DBSCAN 聚类分选结果

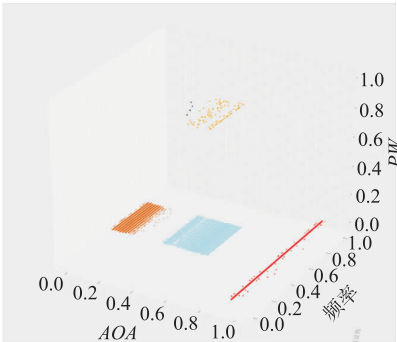


图 14 CMDBSCAN 聚类分选结果

从图 11~14 几种算法分选效果对比可以看出,CMDBSCAN 算法的分选效果表现最好,除了部分离散点以外,其他数据均被正确聚类。由于 CMDBSCAN 算法能够自适应获得邻域半径,并且根据不

同雷达信号在参数空间中的分布密度,相应地调整密度阈值,适用于非均匀雷达信号的分选,克服了传统 DBSCAN 算法的固有缺陷。并且结合云模型理论对分选结果进行有效性评估,可有效避免传统 DBSCAN 算法在多功能雷达信号分选中的“增批”问题,具有更强的信号分选能力。

5 结语

电子对抗侦察中的雷达信号分选是将截获的雷达信号分离的过程,具有非合作性,属于无监督学习问题,DBSCAN 算法是一种无监督的密度聚类算法,可自动分选任意形状的聚类簇,符合实际战场电子侦察需求。本文提出一种 CMDSCAN 算法,首先从自适应获得 Eps 邻域半径和设置变密度点数阈值两个方向针对传统 DBSCAN 算法缺陷进行改进,仿真实验表明,该算法可实现对非均匀、多密度、大数据量的雷达信号进行有效性分选;最后结合多维云模型理论,将改进算法分选结果送入评估模型,结合逆向云发生器算法和三维正向高斯云算法,对分选结果进行有效性评估,利用判定结果进一步优化参数设置,仿真结果表明,DBSCAN 算法结合云模型理论后可有效处理参数交叠严重、工作模式多样的雷达脉冲信号,提高了 DBSCAN 聚类算法的分选准确率。

参考文献

- [1] 王星. 航空电子对抗原理[M]. 北京: 国防工业出版社, 2008: 150-180.
- [2] CHOWDHURY K, CHAUDHURI D, PAL A K, et al. Seed Selection Algorithm Through K-Means on Optimal Number of Clusters, 2019, 78(13):18617-18651.
- [3] LACCETTI G, LAPEGNA M, MELE V, et al. Performance Enhancement of a Dynamic K-Means Algorithm Through a Parallel Adaptive Strategy on Multicore CPUs[J]. Journal of Parallel and Distributed Computing, 2020, 145: 34-41.
- [4] XU S K, WANG C, ZHUANG L H, et al. DBSCAN Clustering Algorithm for the Detection of Nearby Open Clusters Based on Gaia-DR2[J]. Chinese Astronomy and Astrophysics, 2019, 43(2):225-236.
- [5] 李文杰, 闫世强, 蒋莹, 等. 自适应确定 DBSCAN 算法参数的算法研究[J]. 计算机工程与应用, 2019, 55(5): 1-7, 148.
- [6] WANG J, SCHREIBER D K, BAILEY N, et al. The Application of the OPTICS Algorithm to Cluster Analysis in Atom Probe Tomography Data[J]. Microscopy and Microanalysis, 2019, 25(2): 1-11.
- [7] 胡健, 朱海湾, 毛伊敏. 基于自适应蜂群优化的 DBSCAN 聚类算法[J]. 计算机工程与应用, 2019, 55(14): 105-114.
- [8] SKOLNIK M I. Radar Handbook[M]. Beijing: The Defense Industry Press, 2010: 150-180.
- [9] MASUDA H, QI Y, LIU S Y, et al. Reverse Phase Protein Array Identification of Triple-Negative Breast Cancer Subtypes and Comparison with mRNA Molecular Subtypes [J]. Oncotarget, 2017, 8(41): 70481-70495.
- [10] 李海林, 郭崇慧, 邱望仁. 正态云模型相似度计算方法[J]. 电子学报, 2011, 39(11): 2561-2567.
- [11] 夏鲁宁, 荆继武. SA-DBSCAN: 一种自适应基于密度聚类算法[J]. 中国科学院研究生院学报, 2009, 26(4): 530-538.
- [12] 丁玉琦, 邵振峰, 胡石元. 一种云模型和期望最大聚类的遥感影像分割算法[J]. 武汉大学学报(信息科学版), 2015, 40(6): 721-726.
- [13] CHENGJ T, DUAN Z M. Cloud Model Based Sine Cosine Algorithm for Solving Optimization Problems[J]. Evolutionary Intelligence, 2019, 12(4): 503-514.
- [14] YANG M, LIU D Z, CUI Y, et al. Research on Complementary Algorithm of Photovoltaic Power Missing Data Based on Improved Cloud Model[J]. International Transactions on Electrical Energy Systems, 2020, 30(34). Doi:10.1002/2050-7038,12350.
- [15] 梁力, 邢观华, 吴凤元. 基于云理论的评估模型和方法[J]. 东北大学学报(自然科学版), 2019, 40(6): 881-885.
- [16] 李德毅, 杜鹁. 不确定性人工智能[M]. 2 版. 北京: 国防工业出版社, 2016: 68-74.
- [17] LIU J J, ZUO M X, et al. Fuzzy Evaluation Output of Taste Information for Liquor Using Electronic Tongue Based on Cloud Model[J]. Sensors, 2020, 20(3).
- [18] 陈昊, 李兵, 刘常昱. 一种无确定度的逆向云算法[J]. 小型微型计算机系统, 2015, 36(3): 544-549.
- [19] 刘常昱, 冯芒, 戴晓军, 等. 基于云 X 信息的逆向云新算法[J]. 系统仿真学报, 2004, 16(11): 2417-2420.
- [20] CUI J D, ZHENG Q, XIN Y C, et al. Feature Extraction and Classification Method for Switchgear Faults Based on Sample Entropy and Cloud Model[J]. IET Generation, Transmission & Distribution, 2017, 11(11): 2938-2946.
- [21] 国强, 宋文明, 南普龙, 等. 基于数据场与云模型的多模雷达信号分选算法[J]. 哈尔滨工业大学学报, 2015, 47(11): 76-81.
- [22] 张家精, 王焕宝, 倪友聪, 等. 云模型的隶属概念判定中阈值生成[J]. 计算机工程与应用, 2011, 47(24): 125-128.
- [23] ROSENBERG A, HIRSCHBERG J. V-Measure: A Conditional Entropy-Based External Cluster Evaluation [C]//Proceeding of the 2007 Joint Conference on EMNLP-CONLL. Prague, Czech Republic: DBLP, 2007: 410-420.

(编辑:徐楠楠)