

基于选择性深度嵌入聚类的复杂未知雷达信号分选方法

丁宸聪

(中国人民解放军92728部队, 上海 200436)

摘要: 随着现代电子战与雷达技术的不断发展, 战场环境中的雷达信号日趋复杂多样, 这给信号分选带来了严峻挑战。传统分选方法通常依赖先验知识并预先设定类别数量, 难以应对实际应用中可能出现的未知信号源与动态变化特征。针对这一问题, 提出一种基于选择性深度嵌入聚类的复杂未知雷达信号分选方法。该方法结合深度学习在特征提取方面的优势与基于密度的聚类算法的灵活性, 且无须依赖特定先验知识, 实现对雷达信号参数的自适应嵌入与动态聚类; 通过神经网络自动学习信号特征, 并与基于密度的聚类算法相结合, 有效克服了人工定义特征泛化能力不足的缺陷。实验结果表明, 在复杂环境下的雷达信号数据集中, 所提方法的聚类纯度达99.46%, 具备良好的适应性与扩展性, 能够准确识别未知或动态变化的信号特征。与传统方法相比, 基于选择性深度嵌入聚类的复杂未知雷达信号分选方法不仅为海量复杂雷达信号处理提供了更具通用性与可靠性的解决方案, 也为其他涉及未知信号的大规模数据挖掘任务提供了有益参考。

关键词: 雷达信号分选; 电子侦察; 聚类; 深度学习

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202529

A complex unknown radar signal deinterleaving method based on selective deep embedding clustering

DING Chencong

92728 Unit of PLA, Shanghai 200436, China

Abstract: With the continuous advancement of electronic warfare and radar technologies, radar signal in battlefield environments has become increasingly complex and diverse, which poses severe challenges to signal deinterleaving. Traditional deinterleaving methods typically rely on prior knowledge and a preset number of categories, making them unsuitable for handling unknown signal sources and dynamically changing features that may arise in real-world applications. To address this issue, a complex unknown radar signal deinterleaving method based on selective deep embedding clustering was proposed. By integrating the feature extraction capabilities of deep learning with the flexibility of a density-based clustering algorithm, this method achieved adaptive embedding and clustering of radar signal parameters without relying on specific prior knowledge. Through automatic feature learning via neural networks coupled with a density-based clustering algorithm, the need for manually defined features with limited generalizability was avoided. Experimental results on radar signal datasets from complex environments indicate that the proposed method achieves a cluster purity of up to 99.46%, demonstrating robust adaptability and extensibility in identifying unknown or dynamically changing signal characteristics. Compared with traditional methods, selective deep embedding clustering provides a more universal and reliable solution for large-scale, complex radar signal processing and offers valuable references for other extensive data mining tasks involving unknown signals.

Key words: radar signal deinterleaving, electronic reconnaissance, clustering, deep learning

0 引言

随着电子战技术的持续发展和雷达系统的规模化部署,现代战场电磁环境中的雷达信号呈多源化和参数复杂化的特征。如何准确、高效地对复杂信号进行分类与识别,已成为雷达信号处理领域亟待解决的问题之一^[1]。雷达信号分选技术的核心任务是从混杂的信号环境中准确提取并区分不同雷达源的独立脉冲序列,这对于电子对抗和战场情报分析至关重要^[2]。高质量雷达信号分选不仅能增强雷达系统的侦察与探测能力,还能为态势感知和干扰对抗策略制定提供关键支撑^[3]。

在雷达技术发展初期,由于雷达种类和数量均较少,信号源单一,且雷达信号通常具有稳定的脉冲重复间隔(pulse repetition interval, PRI)、固定的脉冲宽度(pulse width, PW)和载波频率(carrier frequency, CF),环境简单、干扰有限,因此基于PRI的雷达信号分选方法能够完成侦察与探测任务^[4]。然而,随着现代雷达技术的快速发展,信号环境逐渐复杂化,基于PRI的雷达信号分选方法的局限性日益凸显^[5]。现代雷达系统的信号特征复杂多变,如频率捷变^[6]、多重PRI^[7]和随机PRI^[8],同时雷达数量和种类不断增加,信号间干扰与环境噪声明显增强,传统雷达信号分选方法难以实时适应动态变化的信号特征。为解决上述问题,研究者不断探索新方法,以提升雷达信号分选的精度与鲁棒性。

传统雷达信号分选方法主要基于对脉冲描述字(pulse descriptor word, PDW)参数的分析与统计。脉冲到达时间(time of arrival, TOA)分析法^[9]基于脉冲TOA信号分布和模态来区分雷达信号源的特征,但该方法对多径传播与噪声敏感。PRI变换法^[10]通过分析脉冲重复间隔的频率特性进行分选,常用快速傅里叶变换^[11]和自相关变换^[12]处理周期性PRI信号,但在应对复杂PRI模式雷达信号时识别准确率较差。参数聚类法利用K-means^[13]和层次聚类^[14]等算法对多个参数进行分析,将相似特征脉冲分组实现分选,但对算法和参数选择依赖度高,抗噪声与抗干扰能力不足。脉冲到达角度(angle of arrival, AOA)分析法^[15]根据脉冲AOA进行分选,通常要结合其他参数,容易受测量精度与分辨率的限制。张旗等^[16]提出多帧多级雷达信号分选方法,该方法虽然提升了分选精度,但在实战中难

以获得稳定的PRI值,影响适用性。王易丽等^[17]提出改进谱聚类方法,该方法依赖到达方向(direction of arrival, DOA)、射频(radio frequency, RF)和PW参数,同样面临过度依赖特定参数的问题。

尽管传统雷达信号分选方法在部分场景应用中取得了一定成效,但整体仍存在明显的局限性。首先,传统方法依赖先验知识,需要预先了解雷达信号特性,因此在面对未知或动态变化环境时适应性不足;其次,传统方法的抗噪声与抗干扰能力较弱,易受外界因素影响;最后,传统方法难以有效应对非稳定PRI和频率捷变等复杂特征信号。近年来,深度学习技术在计算机视觉与自然语言处理等领域取得了突破性进展^[18-19]。例如,图像处理领域的ResNet、DenseNet和EfficientNet^[20],自然语言处理领域的BERT与GPT系列模型^[21-22]。深度学习在上述领域的成功案例,促使研究者将其应用于雷达信号分选。施力泉等^[23]提出基于弱化与增强网络的雷达信号分选方法,利用神经网络进行特征提取。然而,该方法将聚类问题转化为分类问题,当面对海量未知信号时,容易出现维度爆炸和计算复杂度剧增问题。

深度嵌入聚类(deep embedded clustering, DEC)^[24]是一种利用神经网络对数据进行特征嵌入和聚类的无监督方法。本文在DEC的基础上,进一步提出选择性深度嵌入聚类(selective DEC, SDEC)方法,通过基于密度的带噪声应用空间聚类(density-based spatial clustering of application with noise, DBSCAN)算法^[25],能够在没有先验知识的情况下对雷达信号进行有效分选。本文旨在系统阐述SDEC的模型原理与实现细节,并通过大规模仿真验证其在复杂战场环境中的分选性能,为下一代实时电子对抗系统提供可行的技术方案。

1 方法设计

1.1 自编码器

自编码器是一种无监督神经网络,用于学习数据的特征表示^[26]。它由编码器和解码器两个部分组成,通过使输入数据与重构数据之间的误差最小化,得到数据的潜在特征表示。

编码器将输入数据映射到特征空间。设输入数据为 x ,编码器的输出 z 定义如式(1)所示:

$$z = f(x; \theta_e) \quad (1)$$

其中, f 为编码器的非线性映射函数, θ_e 为模型参数。

与编码器相反, 解码器需要对特征进行重构。解码器的输出 \hat{x} 定义如式(2)所示:

$$\hat{x} = g(z; \theta_d) \quad (2)$$

其中, g 为解码器的非线性映射函数, θ_d 为模型参数。自编码器的训练目标是使输入数据 x 与输出数据 \hat{x} 之间的重构误差最小化, 通常采用均方误差(mean square error, MSE)作为损失函数。

本文采用的自编码器网络包含编码器和解码器两个部分。编码器由5个线性层构成, 先将原始输入维度映射至16维, 再调整至32维, 随后降至嵌入维度。解码器的结构与编码器相反, 从嵌入维度逐步映射回原始输入维度。每个线性层均采用ReLU激活函数, 以提升网络的非线性表达能力。在训练阶段, 以MSE为自编码器的损失函数 L_{MSE} , 如式(3)所示:

$$L_{\text{MSE}} = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

其中, y_i 和 \hat{y}_i 分别为第 i 个样本的真实值与预测值, n 为样本总数。损失函数 L_{MSE} 用于衡量模型在重构原始输入时的误差大小, 从而引导网络学习更具判别力的特征表示。

1.2 聚类算法

DBSCAN是基于密度的聚类算法, 能够识别任意形状的簇, 并处理含有噪声的数据。该算法根据数据点之间的密度连接形成簇, 可自动识别簇的数量。设距离度量函数为 $\text{dist}(\cdot, \cdot)$, 数据集为 D , 如式(4)所示:

$$\epsilon(p) = \{q \in D \mid \text{dist}(p, q) \leq \epsilon\} \quad (4)$$

若某点 p 的 ϵ 邻域包含至少MinPts个点(包含 p 自身), 则称 p 为核心点; 若某点 p 在其自身 ϵ 邻域内未达到MinPts个点, 但属于某核心点的 ϵ 邻域内, 则称 p 为边界点; 既非核心点又非边界点的点则视为噪声点。

虽然聚类方法种类繁多且效果普遍较好, 但由于本文所用数据集缺少对具体参数的预先定义, 因此本文直接选用DBSCAN作为聚类算法, 以降低对人工特征提取或特定特征处理的依赖, 从而提升在不同应用场景中的泛用性。

1.3 总体设计

针对未知或动态环境下的雷达信号分选任务,

本文改进了SDEC算法。相比传统方法, SDEC充分利用深度学习的特征提取能力与聚类技术的灵活性, 可在缺乏先验知识的情况下自适应识别不同雷达信号的类别。具体而言, 基于DEC框架对输入特征进行选择性地嵌入处理, 并与DBSCAN协同应用, 从而提升在复杂环境中的分选精度和鲁棒性。

在雷达信号分选任务中, 根据接收机捕获的PDW数据区分不同雷达平台的脉冲信号。然而, 实际应用中雷达信号的数量和类型往往难以预先掌握, 且信号特征常处于持续变化的状态。此时, 依赖先验知识或固定类别数的传统算法难以适应。为克服这一局限, 本文将深度学习与聚类算法结合, 先通过自编码器从多维雷达信号参数中自动学习适合分选的高阶特征, 再利用DBSCAN在密度空间中动态检测并划分簇。DBSCAN无须事先指定类别数, 通过邻域半径与最小点数等超参数刻画数据分布的局部密度; 一旦某区域的密度达到阈值, 即可自动判定其为核心簇, 不满足密度要求的点则被标记为噪声或离群点, 使聚类过程具备更高的鲁棒性与适应性。

在特征层面, 不同场景的数据集通常包含若干难以统一定义的特征, 此类参数难以直接归纳为固定维度的手工特征, 却与雷达平台的固有属性紧密相关, 若能加以利用, 将对分选效果产生关键影响。基于此, SDEC采取选择性嵌入策略, 将未知或不定参数直接输入神经网络进行特征映射。同时, TOA与AOA因易受外部环境及平台位置变动的影 响而稳定性不足, 故未纳入自编码器的嵌入过程, 而是在特征提取后再与嵌入向量拼接, 形成最终的PDW特征向量。此方式既能充分保留雷达自有参数的高效表达, 也兼顾TOA与AOA在后续分类与聚类中的辅助作用。

在完成特征融合后, 使用DBSCAN对最终的高维向量进行聚类分析。DBSCAN能够灵活适配任意形状的簇, 并将离散数据点标记为噪声点或孤立点, 对分布多样且噪声较重的雷达信号数据具有显著优势。若数据集中存在异质性强或数量稀少的脉冲, DBSCAN也可通过密度判定将其独立划分为额外簇或视为噪声点, 从而避免强制归类导致的分选精度下降。通过自编码器学习到的嵌入特征与DBSCAN算法的有机结合, SDEC可在复杂战场环境中准确、高效地识别并分离不同雷达源的脉冲信号, 显著提升雷达信号分选的通用性与可靠性。

综上所述，SDEC通过融合深度学习与密度聚类，为未知或动态环境下的雷达信号分选提供了一种高效的解决方案。一方面，减少了对先验知识和手工特征的依赖；另一方面，其动态分簇与噪声识别能力可应对雷达信号数量增长和特征演化的挑战。图1展示了本文方法的总体架构，从自编码器对PW、PRI、CF等关键参数的嵌入学习，到拼接TOA与AOA进行DBSCAN聚类，形成一条逻辑清晰、功能互补的完整分选技术路线。

2 实验结果与分析

2.1 数据集

为全面评估SDEC在类别繁多且特征高度动态场景中的鲁棒性，本文采用覆盖30种雷达辐射源的训练数据集进行实验。与常用公开集相比，该数据集在规模、异质性和噪声水平上更具挑战性。(1) 数据量庞大且类别分布不均衡，既包含海量脉冲样本，也涵盖各类雷达在不同时段、不同方位条件下的发射状态，可有效检验模型处理长尾分布与少样本类别的能力。(2) 参数命名与量纲未统一规范，例如，平均参数1、平均参数2、平均参数3等字段缺乏明确的物理含义，真实反映了实际PDW采集中的多源异构与标准不一致现象，可用于评估算法对未知或部分标注特征的适应性。(3) 数据中混入由电磁干扰、设备误差和信道衰减等因素产生的显著噪声，真实模拟了现代电子战环境。综合而言，该数据集在规模、参数多样性和噪声复杂度方面均对分选算法提出严苛要求，能更准确地验证SDEC的泛化能力与工程应用价值。

在海量数据与参数多样性的共同作用下，该数据集对聚类算法的适应能力提出更高要求，需要自动分辨未知雷达类型，并在高噪声、分布异质的背景中精准提取核心特征。表1列出了各雷达类别的样本数量和主要参数范围，便于直观把握数据整体情况。需要指出的是，由于采集方式、采样时段及雷达类型各不相同，部分参数可能存在缺失或与常规定义不符的情况。这种真实而复杂的属性反映了实际战场环境中面临的挑战，也为本文方法提供了更具代表性的测试条件。在具体应用中，为了调整神经网络的超参数，从数据集中随机抽取10 000条样本作为验证集。实验结果表明，由于神经网络仅用于特征提取而非直接分选，不同超参数设置对最终分选结果的影响不明显。

2.2 实验软硬件配置

表2列出了本文采用的软硬件配置。在软件层面，实验平台运行Ubuntu 20.04操作系统，该系统以内核稳定、驱动完备著称，可长期支撑高负载深度学习训练与大规模数据处理。深度学习框架采用PyTorch 2.3.0^[27]构建并训练本文提出的SDEC模型。传统机器学习与聚类算法依托scikit-learn 1.5.0^[28]实现对比方法（如DBSCAN）的标准化调用，确保实验流程的统一性与可重复性。在硬件层面，计算节点配备32核Intel Xeon Platinum 8336C处理器，提供充足的多线程并行能力，以满足数据预处理和模型推理等CPU密集型任务。图形加速器选用NVIDIA GeForce RTX 4090，该加速器具有16 384个CUDA核心与24 GB GDDR6显存，在大批量矩阵运算和深度网络反向传播阶段，能够显著

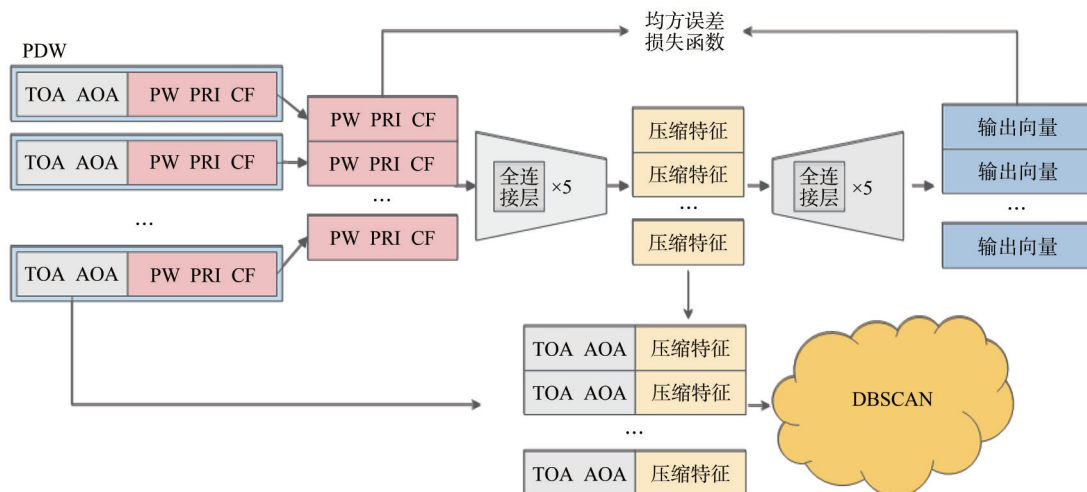


图1 SDEC总体架构

表1 训练样本集信息

雷达编号	样本数量/个	TOA 范围/ns	平均 AOA	平均参数 1	平均参数 2	平均参数 3
1	66 218	[1 559 168 458, 598 014 951 194]	96.88°	5 210.00	12.97	-28.01
2	1 2945 584	[1 322 250 371, 594 925 172 358]	118.61°	9 915.00	0.85	-29.15
3	100 631	[2 927 218 280, 597 766 302 414]	115.27°	9 622.00	2.50	-34.02
4	274 771	[343 616 339, 598 187 200 081]	50.77°	15 111.99	2.02	-30.84
5	135 501	[1 062 473, 597 539 808 192]	84.46°	5 214.99	41.03	-32.38
6	70 529	[117 014 820 219, 599 563 380 134]	233.33°	9 598.19	1.86	-26.71
7	15 115 065	[174 286 295, 599 999 992 051]	192.50°	14 273.58	1.99	-25.82
8	429 494	[313 198 856 508, 595 053 298 685]	157.49°	9 410.00	12.97	-25.60
9	51 111	[292 331 213, 598 868 370 951]	67.41°	3 627.22	20.37	-36.34
10	163 489	[3 491 490 131, 598 252 814 597]	72.02°	3 163.00	32.96	-28.53
11	87 895	[460 108 912, 599 999 502 792]	74.03°	9 417.99	2.16	-28.22
12	68 723	[4 751 172 733, 594 035 640 466]	80.67°	3 295.00	5.57	-27.13
13	43 121	[5 851 290 615, 596 543 440 371]	89.78°	1 227.48	70.99	-27.98
14	73 031	[555 600 225, 598 642 800 385]	297.07°	3 334.51	14.97	-27.47
15	1 337 193	[56 479, 599 999 982 110]	141.45°	3 559.00	2.36	-28.44
16	1 735 357	[400 133, 599 999 737 068]	143.57°	9 254.20	5.79	-12.12
17	104 339	[4 184 353 717, 599 998 801 479]	90.44°	1 264.97	25.96	-24.83
18	38 969	[1 961 568 195, 596 024 283 130]	139.98°	16 612.97	89.12	-35.70
19	135 041	[1 453 165 671, 599 271 612 155]	139.95°	5 499.90	15.19	-17.03
20	1 546 031	[827 164 739, 599 999 966 382]	142.71°	13 673.47	1.81	-33.12
21	999 999	[1 100 210, 599 999 900 125]	229.65°	9 619.49	60.99	-21.34
22	51 858	[1 842 054 607, 599 460 558 385]	96.52°	1 227.45	35.23	-25.07
23	111 637	[6 250 563 672, 596 749 997 375]	105.63°	5 535.52	11.12	-35.32
24	54 666	[1 223 331 012, 598 194 125 504]	92.62°	9 710.00	20.90	-34.18
25	23 433 760	[2 940 396, 599 999 996 721]	203.68°	9 721.40	1.08	-21.65
26	486 390	[670 700 251, 598 386 583 082]	210.73°	9 764.01	4.94	-30.58
27	1 473 254	[93 845 509, 599 525 696 666]	56.79°	9 880.71	3.42	-32.45
28	978 870	[1 430 715 052, 598 901 640 368]	162.85°	13 653.99	3.59	-27.90
29	60 576	[1 189 775 304, 598 914 388 035]	112.48°	5 360.99	0.87	-26.10
30	152 223	[800 472, 599 688 180 427]	128.40°	1 279.39	19.98	-21.38

缩短训练时间。系统内存为251 GB DDR4，为高维数据处理、多进程调度与模型参数缓存提供了充足空间。

表2 软硬件配置

规格	CPU	GPU
型号	Intel Xeon Platinum 8336C	NVIDIA GeForce RTX 4090
频率	2.3 GHz	2.52 GHz
核心数	32	16 384
缓存	L2-40 MB, L3-48 MB	L2-96 MB
内存	251 GB DDR4	24 GB DDR6

2.3 SDEC网络训练过程

表1所示的训练样本集涵盖30类雷达源信息，其中包含大量噪声点与分布不均衡的脉冲样本，且

部分参数缺乏准确定义，全面模拟了真实电子战环境中复杂且动态多变的雷达信号特征。表1中还展示了不同信号源与采集条件下雷达参数的显著差异和不一致性，与战场实况的PDW采集方式相符，为评估算法在复杂环境下的适应性与鲁棒性提供了有利条件。

在模型训练阶段，平均参数1、平均参数2和平均参数3作为输入变量，输入向量维度为3，自编码器的嵌入向量维度同样设定为3，以便在保留主要信息的同时减少冗余。为提升训练效率与收敛速度，本文采用Adam优化器，学习率设置为0.001，批量大小设定为16 384，以兼顾训练稳定性与硬件资源利用率。当训练完成后，将从自编码器中提取

得到的3维嵌入向量与TOA、AOA信息相结合，拼接成4维特征向量，并在该特征空间中应用DBSCAN进行聚类。这种方式既可充分利用深度网络学习到的关键特征，又能保留时间与角度等潜在辅助信息，为后续信号分选与评估奠定坚实基础。

在高噪声、分布不平衡的复杂数据环境下，结合时空信息深入挖掘上述参数之间的潜在关联，对于实现精准、高鲁棒性的聚类分析至关重要。SDEC利用自编码器筛选关键特征，再将关键特征与TOA、AOA拼接，借助DBSCAN完成对不同雷达源信号的分组和辨识，为雷达侦察与对抗提供可靠支撑。

2.4 评价指标

为全面评估SDEC在雷达信号分选任务中的性能，引入多种聚类评价指标，从不同角度衡量聚类结果与真实分组之间的相似度和分簇质量。这些指标在模式识别与机器学习领域应用广泛，可在无监督学习场景下有效反映算法的适用性与精度。具体而言，本文选用调整兰德指数（adjusted rand index, ARI）^[29]、标准互信息（normalized mutual information, NMI）^[30]及聚类纯度^[31]3项指标，分别从样本对一致性、信息熵互相关性以及聚类内部同质性3个层面量化评价结果。

ARI度量聚类结果与真实标签之间的匹配程度，可视作对样本对一致性的修正型评估。与传统兰德指数（rand index, RI）相比，ARI通过修正随机聚类导致的期望一致性，使评分更客观，避免在过度聚类或欠聚类时出现不公平的高分或低分。ARI的定义如式（5）所示：

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \quad (5)$$

其中，RI表示在所有可能的样本对中，聚类结果与真实标签一致的比例； $E[RI]$ 表示在随机分簇假设下RI的期望。由于忽略了随机因素的影响，ARI对不同数据集的聚类稳定性与精确度更具可比性。

NMI用于衡量两个数据分区（如聚类结果与真实标签）之间的依赖程度，反映一个分区对另一个分区不确定性的消减能力，确保评价结果不受数据集规模或类别数量限制。NMI的定义如式（6）所示：

$$NMI(U, V) = \frac{2 \times I(U; V)}{H(U) + H(V)} \quad (6)$$

其中， U 为聚类结果， V 为真实标签， $I(U; V)$ 表征两者之间的互信息， $H(U)$ 和 $H(V)$ 分别为对聚类结果与真实标签的熵度量。NMI越接近1，说明聚类结果与真实标签的匹配度越高；NMI趋近0，则意味着两者几乎没有关联。

聚类纯度关注每个聚类内主导类别所占比例，用于衡量聚类结果的纯净程度，其定义如式（7）所示：

$$\text{Purity} = \frac{1}{N} \sum_{k=1}^K \max_j |c_k \cap t_j| \times 100\% \quad (7)$$

其中，Purity为聚类纯度， N 为样本总数， K 为聚类数， c_k 为第 k 个聚类包含的数据点集合， t_j 为真实标签中第 j 个类别， $|c_k \cap t_j|$ 为第 k 簇与第 j 类的交集大小， $\max(\cdot)$ 为该聚类中占比最高的真实类别。由于聚类纯度只关注簇内部的主要类别比例，而不直接考虑簇数量或分布，若聚类结果过于细分，则纯度值可能偏高，因此，在解读纯度指标时需结合其他指标进行分析。

ARI、NMI与聚类纯度分别从不同维度衡量聚类结果的优劣，可通过准确性、稳定性与可泛化性3个方面，为SDEC在雷达信号分选任务中的表现提供全面、互补的定量评估。综合考量这3种评价指标，能更精准地判断SDEC在应对复杂、多噪声且分布不均衡的雷达信号数据时的聚类质量和实际应用价值。

2.5 测试结果与消融实验

为更好地满足真实电子战环境中的分批处理需求，测试阶段将测试集中所有脉冲先按TOA排序后，再按批次大小16384进行分块测试，从而更准确地模拟实战中对雷达信号的顺序分析与实时分选过程。为验证SDEC的优越性，与下列算法进行对比。

(1) TOA-AOA：基于TOA和AOA计算均值后，将样本分配至最近距离类别。

(2) Statistical：基于统计特征（均值）计算每个类别的均值后，将样本分配至距离最近的类别。

(3) K-means：采用K-means聚类算法分选。

(4) DBSCAN all：对所有维度（如平均参数1、平均参数2、平均参数3、TOA、AOA等）直接进行DBSCAN聚类。

(5) DBSCAN Param1-3：仅用平均参数1、平均参数2、平均参数3进行DBSCAN聚类。

(6) DEC: 将所有维度输入自编码器, 进行深度嵌入聚类。

(7) SDEC: 本文提出的SDEC方法先对平均参数1、平均参数2、平均参数3进行自编码器嵌入至2维, 再与TOA、AOA拼接为4维向量, 最终使用DBSCAN聚类。

数据集中未对平均参数1、平均参数2、平均参数3作具体说明, 因此, 部分研究中针对特定参数进行特殊处理的分选方法在此场景下不适用, 凸显出本文方法的不依赖性与泛化性优势, 可适用于多种场景。表3实验结果表明, SDEC在ARI、NMI与聚类纯度3项指标上均表现最优。其中, 本文方法的ARI达0.992 2, 显著高于DEC (0.978 5)、DBSCAN all (0.968 2) 以及DBSCAN Param1-3 (0.890 9); 本文方法的NMI达0.995 4, 明显优于DEC (0.983 3)、DBSCAN all (0.973 8) 和DBSCAN Param1-3 (0.913 6); 本文方法的聚类纯度为0.994 6, 高于DEC (0.987 4) 和DBSCAN all (0.983 9), 并大幅领先DBSCAN Param1-3 (0.944 7)。整体而言, SDEC在3项指标上均展现出明显优势, 尤其在ARI与NMI方面表现突出, 表明其在未知或动态环境下具备更强的雷达信号区分能力。

表3 多种聚类方法对比实验结果

方法	ARI	NMI	Purity
TOA-AOA	0.500 8	0.510 2	0.757 6
Statistical	0.858 4	0.778 9	0.902 2
K-means	0.880 1	0.810 4	0.895 1
DBSCAN all	0.968 2	0.973 8	0.983 9
DBSCAN Param1-3	0.890 9	0.913 6	0.944 7
DEC	0.978 5	0.983 3	0.987 4
SDEC(本文方法)	0.992 2	0.995 4	0.994 6

为探究模型内部结构对分选效果的影响, 本文进一步开展了消融实验, 主要从自编码器的复杂度和嵌入维度两个角度展开分析。

(1) 自编码器的复杂度

表4对比了不同复杂度的自编码器结构, 其中, (16, 16, 3)表示3层全连接, 输出维度依次为16、16、3; (16, 32, 32, 16, 3)与(16, 64, 64, 16, 3)均为5层全连接, 只是中间层最大维度(32或64)不同。实验结果显示, 采用5层结构且中间层最大维度为32时, ARI可达0.992 2, NMI为0.995 4, 聚类纯度则为0.994 6, 均优于其他方案。与将中间

层最大维度扩展到64的方案相比, 该设置在性能略优的同时有效控制了网络规模与计算开销, 表明特征提取能力与参数规模间保持平衡的重要性。

表4 自编码器对比实验结果

自编码器结构	ARI	NMI	Purity
(16, 16, 3)	0.980 4	0.986 2	0.986 6
(16, 32, 32, 16, 3)	0.992 2	0.995 4	0.994 6
(16, 64, 64, 16, 3)	0.988 7	0.993 2	0.992 4

(2) 嵌入维度

为探讨自编码器输出的特征维度对聚类性能的影响, 在压缩平均参数1、平均参数2、平均参数3的基础上, 依次设置嵌入层维度为1、2、3和4。实验结果表明, 当维度为3时, 本文方法的ARI达0.992 2, NMI为0.995 4, 聚类纯度为0.994 6, 均为最优值。若减少或增加嵌入维度, 则在不同程度上导致分选效果下降。这表明恰当的维度设置有利于保留关键信息, 而过度压缩或扩展会引入信息损失或噪声干扰。

综上所述, SDEC在聚类纯度、ARI、NMI等指标上均表现优异, 充分说明自编码器的深度结构与嵌入维度对分选质量的重要性。通过合理配置层数与中间层维度, SDEC能在高效提取雷达信号判别特征的同时, 兼顾计算量和网络规模的可控性。消融实验结果表明, 本文所用的PDW参数无冗余, 过度降维将显著损害聚类效果, 从而进一步验证了选择性嵌入策略在雷达信号分选中的合理性与必要性。整体而言, 这些结果为SDEC在复杂战场环境中的推广与应用提供了有力支撑, 也为后续深度聚类方法应用于未知信号场景提供了宝贵参考。

3 结束语

针对复杂雷达环境下的信号分选需求, 本文结合深度自编码器与DBSCAN算法提出SDEC方法。与传统方法相比, SDEC借助深度自编码器自动学习雷达信号的潜在特征, 降低对先验知识的依赖。同时, 通过选择性特征嵌入策略, 规避易受环境干扰的TOA与AOA参数, 提取与雷达源本质关联的特征, 并与位置参数拼接形成综合特征向量, 最终利用DBSCAN实现信号的动态聚类。实验结果表明, SDEC在ARI、NMI和聚类纯度等评价指标上均表现优异, 展现出良好的泛化能力与扩展性。

随着电子战装备与雷达系统的快速发展, 信号

多样性与复杂度不断提升。为进一步增强SDEC在高噪声、大规模及复杂信号模式下的分选性能,后续可从以下4个方向开展研究。

(1) 网络结构优化: 探索更深层或多分支结构, 结合注意力机制强化对关键特征的捕捉。

(2) 聚类算法改进: 在DBSCAN的基础上探讨自适应聚类算法(如基于密度峰值), 提升对非均匀分布数据的识别精度。

(3) 多模态信息融合: 考虑将红外或可见光图像等多模态信息与PDW参数结合, 构建更完善的多源数据体系。

(4) 在线更新与增量学习: 针对实战中信号实时流入、特征动态变化的场景, 开发在线更新的训练与聚类方法, 以提升模型的自适应能力。

参考文献:

- [1] DIAO P S, ALVES T, POUSSOT B, et al. A review of radar detection fundamentals[J]. IEEE Aerospace and Electronic Systems Magazine, 2024, 39(9): 4-24.
- [2] 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.
SUI J P, LIU Z, LIU L, et al. Progress in radar emitter signal deinterleaving[J]. Journal of Radars, 2022, 11(3): 418-433.
- [3] 贾金伟, 刘利民, 韩壮志, 等. 低截获雷达抗分选信号设计技术研究综述[J]. 电讯技术, 2023, 63(5): 748-756.
JIA J W, LIU L M, HAN Z Z, et al. Review on anti-sorting signal design technology of low probability intercept radar[J]. Telecommunication Engineering, 2023, 63(5): 748-756.
- [4] REDDY R, SINHA S. State-of-the-art review: electronic warfare against radar systems[J]. IEEE Access, 2025, 13: 57530-57567.
- [5] CHENG W H, ZHANG Q Y, DONG J M, et al. An enhanced algorithm for deinterleaving mixed radar signals[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(6): 3927-3940.
- [6] FANG Y Y, ZHANG L, WEI S P, et al. Online frequency-agile strategy for radar detection based on constrained combinatorial nonstationary bandit[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(2): 1693-1706.
- [7] ZHU M T, LI Y J, WANG S F. Model-based time series clustering and interpulse modulation parameter estimation of multifunction radar pulse sequences[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(6): 3673-3690.
- [8] BLUNT S D, HARNETT L A, RAVENSCROFT B, et al. Implications of diversified Doppler for random PRI radar[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(4): 3811-3834.
- [9] AKAN A, CURA O K. Time-frequency signal processing: today and future[J]. Digital Signal Processing, 2021, 119: 103216.
- [10] WAN M C, ZHANG Y, BAI Y C, et al. A real-time radar signal sorting method under Bayesian framework with dynamic cluster merging[J]. IEEE Sensors Journal, 2024, 24(17): 27859-27869.
- [11] AHMAD Z, CHOWDHURY R, DAS R, et al. A fast algorithm for aperiodic linear stencil computation using fast Fourier transforms[J]. ACM Transactions on Parallel Computing, 2023, 10(4): 1-34.
- [12] 嘉娟. 脉间波形变换雷达信号分选理论研究[D]. 哈尔滨: 哈尔滨工程大学, 2024.
JIA J. Theory research on inter-pulse waveform transform-based radar signal sorting[D]. Harbin: Harbin Engineering University, 2024.
- [13] STEMMER U. Locally private K-means clustering[J]. The Journal of Machine Learning Research, 2021(22): 1-30.
- [14] 王云, 王美蕴, 周健, 等. 基于改进层次聚类和GL-APSO算法的配电网动态重构[J]. 智能科学与技术学报, 2022, 4(3): 410-417.
WANG Y, WANG M Y, ZHOU J, et al. Dynamic configuration of distribution network based on improved hierarchical clustering and GL-APSO algorithm[J]. Chinese Journal of Intelligent Science and Technology, 2022, 4(3): 410-417.
- [15] NEHORAI A, PALDI E. Vector-sensor array processing for electromagnetic source localization[J]. IEEE Transactions on Signal Processing, 1994, 42(2): 376-398.
- [16] 张旗, 董阳阳, 张立东, 等. 基于多帧多级处理的雷达信号分选方法[J]. 电讯技术, 2024, 64(11): 1836-1843.
ZHANG Q, DONG Y Y, ZHANG L D, et al. Radar signal sorting based on multi-frame and multi-stage processing[J]. Telecommunication Engineering, 2024, 64(11): 1836-1843.
- [17] 王易丽, 杨宇明. 一种基于改进谱聚类的雷达信号分选算法[J]. 电讯技术, 2023, 63(9): 1348-1354.
WANG Y L, YANG Y M. A radar signal sorting algorithm based on improved spectral clustering[J]. Telecommunication Engineering, 2023, 63(9): 1348-1354.
- [18] CHOI K, YI J H, PARK C, et al. Deep learning for anomaly detection in time-series data: review, analysis, and guidelines[J]. IEEE Access, 2021, 9: 120043-120065.
- [19] 张钹. 人工智能进入后深度学习时代[J]. 智能科学与技术学报, 2019, 1(1): 4-6.
ZHANG B. Artificial intelligence is entering the post deep-learning era[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 4-6.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [21] GARDAZI N M, DAUD A, MALIK M K, et al. BERT applications in natural language processing: a review[J]. Artificial Intelligence Review, 2025, 58(6): 166.
- [22] 黄峻, 林飞, 杨静, 等. 生成式AI的大模型提示工程: 方法、现状与展望[J]. 智能科学与技术学报, 2024, 6(2): 115-133.
HUANG J, LIN F, YANG J, et al. From prompt engineering to generative artificial intelligence for large models: the state of the art and perspective[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(2): 115-133.
- [23] 施力泉, 张红梅. 基于弱化与增强网络的雷达信号识别[J]. 电讯技术, 2024, 64(11): 1844-1849.
SHI L Q, ZHANG H M. Radar signal recognition based on weakening and strengthening network[J]. Telecommunication Engineering, 2024, 64(11): 1844-1849.
- [24] 尹春勇, 蒋奕阳. 基于个性化时空聚类的差分隐私轨迹保护模型[J]. 信息网络安全, 2024, 24(1): 80-92.
YIN C Y, JIANG Y Y. Differential privacy trajectory protection model based on personalized spatiotemporal clustering[J]. Netinfo Security, 2024, 24(1): 80-92.
- [25] SCHUBERT E, SANDER J, ESTER M, et al. DBSCAN revisited, revisited[J]. ACM Transactions on Database Systems, 2017, 42(3): 1-21.

- [26] KHADER M, AL-NAYMAT G. An overview of various enhancements of DENCLUE algorithm[C]//Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems. New York: ACM, 2019: 1-7.
- [27] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library[J]. arXiv preprint, 2019, arXiv: 1912.01703.
- [28] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in python[J]. Journal of Machine Learning Research, 2011, 12: 2825-2830.
- [29] LIU Y F. Improved top-k threshold estimation in information retrieval[D]. New York: New York University Tandon School of Engineering, 2024.
- [30] KINGMA D P, WELLING M. An introduction to variational autoencoders[J]. Foundations and Trends in Machine Learning, 2019, 12(4): 307-392.
- [31] 李智冈, 吕莉, 谭德坤, 等. 基于加权核密度估计与微簇合并的密度峰值聚类算法[J]. 信息与控制, 2024, 53(3): 302-314.
- LI Z G, LYU L, TAN D K, et al. Density peaks clustering algorithm based on weighted kernel density estimation and micro-cluster merging[J]. Information and Control, 2024, 53(3): 302-314.

[作者简介]



丁宸聪（1981- ），男，中国人民解放军92728部队高级工程师，主要研究方向为航空电子与信息对抗。