

文章编号: 1001-893X(2011)09-0065-04

基于数据流聚类的动态信号分选框架^{*}

张国毅, 王晓峰, 张旭洲

(空军航空大学 航空信息对抗系, 长春 130022)

摘要:为了解决雷达信号分选中准确性与实时性相矛盾的问题,提出了一种基于数据流聚类的动态信号分选框架。该框架分为在线和离线两部分,在线部分利用网格帧保存侦察数据的概要信息;离线部分通过网格聚类算法对网格帧进行聚类分选,并得到分选结果。仿真实验表明,该框架能够分选高密度复杂侦察数据流,对噪声不敏感,且无需先验知识支撑,能够较好地满足信号分选准确性和实时性的需要。

关键词:雷达信号分选;数据流;网格聚类;实时性

中图分类号: TN957.5; TN97 **文献标识码:** A doi: 10.3969/j.issn.1001-893x.2011.09.013

A Dynamic Frame of Signal Sorting Based on Data Stream Clustering

ZHANG Guo-yi, WANG Xiao-feng, ZHANG Xu-zhou

(Department of Aviation Communication Counter, Aviation University of Air Force, Changchun 130022, China)

Abstract: In order to solve the contradiction between the accuracy and the real-time in radar signal sorting, a dynamic signal sorting frame based on data stream clustering is proposed. The frame can be divided into two parts, or online part and offline part. In online part, the reconnaissance data is transformed and saved as grid frame. In offline part, the grid frame is sorted with the algorithm of grid-based clustering. Experimental results show that the proposed frame can sort the high density and complex reconnaissance data stream without any prior knowledge, and is not sensitive to the influence of noise. The requirement of accuracy and real-time in signal sorting can be satisfied with this frame.

Key words: radar signal sorting; data stream; grid-based clustering; real-time

1 引言

雷达信号分选是雷达对抗侦察系统中的关键处理过程,信号分选水平直接制约了雷达对抗侦察系统的整体作战效能^[1]。传统的多参数分选技术依次利用脉冲到达角(DOA)、载波频率(RF)、脉宽(PW)、重频(PRI)和脉幅(PA)对雷达全脉冲序列进行去交错,这种分选方式的分选效率是由上述分选过程中每一级分选准确率来决定的,前一级分选产生的偏差会对后续处理过程造成很大的影响。面对现代高密度复杂信号脉冲流,传统分选技术会产生严重的

增批和漏批,且分选实时性极差^[2]。

针对传统分选技术的缺陷,文献[3]提出一种基于多参数聚类分析的雷达脉冲信号分选方法,提高了分选准确率,但存在容差难以设定的问题。文献[4]中尝试利用 Kohonen 网络模型对雷达信号进行分选,并对参数权值进行了确定,提高了分选的准确率,但算法面临着对初始聚类中心数进行优化的问题。文献[5]将基于支持向量机和分层互耦的聚类算法与基于类型熵的识别技术相结合,可以有效解决信号分选中的容差问题,但该方法计算量大,不能满足雷达信号实时分选的需要。为了解决信号分选中准确性与实时性的矛盾问题,本文提出了一种新

^{*} 收稿日期: 2011-03-30; 修回日期: 2011-06-07

的动态信号分选框架,该框架分为在线和离线两部分,在线部分将侦察数据流进行转换,并在特定的时刻将转换的结果存入暂存数据库,离线部分对指定时间段内的侦察数据进行分选,并给出分选结果。计算机仿真实验验证了这种动态信号分选框架的有效性。

2 问题描述

数据流聚类^[6]是针对连续数据流的聚类分析方法,如果把接收机不断输出的雷达侦察数据看作数据流,则雷达信号分选就转化为一个数据流聚类问题。即将等时间间隔的侦察数据帧序列 $X_1, X_2, \dots, X_k$ 看作数据流,其中, $X_k = (x_{k1}, x_{k2}, \dots, x_{kn})$ 代表 k 时刻数据帧内包含的侦察数据的集合。本文要解决的问题是:给定任意两个时刻 t_a 和 $t_b (t_b > t_a)$,能够快速准确地从数据集 X_{b-a} 中分离出每一部雷达信号,即要形成对集合 X_{b-a} 的一个划分:

$$X_{b-a} = L_1 \cup L_2 \cup \dots \cup L_m \cup S \tag{1}$$

式中,集合 L_i 对应雷达信号,集合 S 对应混叠在雷达信号中的噪声脉冲。

3 基本概念

定义 1 网格单元

设空间 $D = (D_1, D_2, \dots, D_d)$ 为侦察数据形成的特征参数空间,空间的每一维对应侦察数据的一个特征参数,将空间的每一维划分为 m 个相等的小区间,则整个参数空间被分成有限个不相交的矩形(或超矩形)单元,这种划分后的单元称为网格。每一维上的小区间按边界值大小顺序排列,参数空间中的每一个元素都可以对应至数据空间中的某个网格。

定义 2 网格密度

根据网格单元的定义可知,任意一个网格单元的体积都是相等的,因此,设网格单元的密度 ρ 为该网格包含的数据点个数。当 $\rho > \epsilon$ 时,该网格定义为一个密集网格,其中 ϵ 是密度阈值,分选中认为非密集网格内的数据为噪声脉冲数据。

定义 3 网格帧

将某一时刻的网格结构以快照的形式存入暂存数据库,每一幅快照称作一个网格帧,网格帧包含的信息有:当前时刻数据与网格的相互对应关系,以及每个网格的密度。

定义 4 邻接网格

设 d 维空间中两个网格单元为 U_1 和 U_2 ,当且

仅当这两个网格有相邻的边界时,称其为邻接网格。

定义 5 密度可达

如果网格 U_i 是网格 U_j 的邻接网格,且同时满足:

$$\begin{aligned} \rho(U_i) &\geq \epsilon \\ \rho(U_j) &\geq \epsilon \end{aligned} \tag{2}$$

则称网格 U_i 密度可达网格 U_j ,易证网格密度可达是对称的,所以网格 U_i 密度可达 U_j 等价于网格 U_j 密度可达 U_i 。

4 动态信号分选框架

目前公开的信号分选方法都没有考虑侦察数据集随时间变化的问题,本质上都属于静态分选方法,在处理不断涌入的大量高密度脉冲流时,静态方法存在分选时间长、不能保证分选实时性的问题。基于数据流的动态聚类算法能够增量式地跟踪数据流的变化,同时不占用过多的计算资源,对侦察数据集的扫描仅限于单次,因此,数据流聚类方法非常适合对侦察数据进行实时分选处理。本文提出的动态信号分选框架主要分为两部分,如图 1 所示。

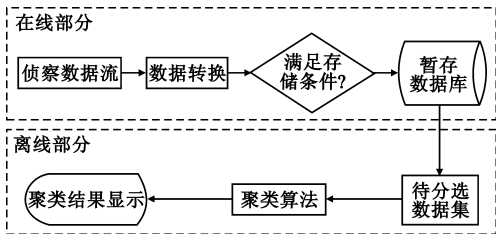


图 1 动态信号分选框架图
Fig.1 Dynamic signal sorting frame

4.1 在线部分

随着侦察数据的不断到达,实时提取其概要信息,通过数据转换得到侦察数据的概要数据结构,在满足一定存储时间条件时将概要数据结构以网格帧快照的形式存入暂存数据库,以备离线部分调用。

4.2 离线部分

从暂存数据库中调用网格帧,计算出指定时间段内的待分选数据集,并利用网格密度聚类算法完成信号分选。

将信号分选框架分为两部分处理的优点在于:一方面,在线部分能够保证实时处理侦察所得的大量数据,获得离线部分所需的概要信息;另一方面,由于离线部分不直接对海量数据进行处理,因而同样可以以较快的速度处理存储于暂存数据库中的概

要信息,保证信号分选的实时性。

5 分选框架的实现

5.1 数据转换

对数据进行归一化处理,并按参数空间位置将数据划分至各个网格,计算每个网格的密度。

5.2 快照存储

随着数据转换的不断进行,当满足一定时间条件时,将网格数据结构以网格帧快照的形式存入暂存数据库。本文采用固定时间间隔保存快照,即每到时间间隔 t_0 的整数倍对网格数据结构进行一次保存,形成一个网格帧。

5.3 寻找待分选数据集

设当前时刻为 t_c ,要分选过去时间范围 h 内的侦察数据,即位于时间段 $t_a \sim t_c$ 范围内的侦察数据,其中 $t_a = t_c - h$ 。通过即时保存和访问暂存数据库,可以得到两个时刻的网格帧,一个是即时保存的 t_c 时刻网格帧,另一个是访问暂存数据库中 t_a 时刻的网格帧。计算这两个时刻的网格帧差,即可以得到待分选的数据集,得到的待分选数据集本质上也是一个网格帧。通过上节分析可知,并不是每个时刻都有网格帧被保存,因此 t_c 和 h 的设定要满足一定条件,即要保证 t_c 和 h 是快照存储时间间隔的整数倍,只有这样才能保证暂存数据库中存有时刻在 t_a 的网格帧,显然,在信号分选中,这个条件是很容易满足的。

5.4 分选实现

利用网格密度聚类算法对指定时间段内的数据集形成的网格帧进行分选。该聚类算法每一次聚类

都首先从数据空间中寻找一个密集网格 U_0 ,按照广度优先搜索原则完成聚类。搜索算法以 U_0 为本次聚类的起始点,依次访问未被访问且与 U_0 密度可达的网格 $U_1, U_2, \dots, U_i$,然后再顺序访问所有与网格 $U_i (i = 1, 2, 3, \dots, i)$ 密度可达且未被访问的网格,当不再有密度可达的网格被发现时,则将所有密度可达网格聚成一类,即完成了一个聚类。算法继续以未被访问的密集网格为初始点重复上述过程,直到所有密集网格均完成聚类。

6 仿真实验及结果分析

实验采用的测试平台为 Intel CPU Q8200, 3GB 内存,操作系统为 Windows XP,仿真软件为 Matlab 7.0。通过计算机形成脉冲描述数字列(PDW),用来模拟接收机输出的侦察数据流。接收机输出的 PDW 描述辐射源类型的特征参数主要有 TOA、DOA、RF、PW、PA 以及 PM 等,针对参数特点及分选需要,本文将侦察数据流中最稳定的 3 个参数 DOA、RF、PW 作为特征参数,构成一个具有三维属性信息的特征空间,以此空间作为信号分选的参数空间。

密度阈值 ϵ 的作用是滤除噪声脉冲,在信号分选过程中,由于噪声脉冲的个数远远少于真实雷达脉冲个数,因此噪声形成的网格密度也会远远小于信号数据形成的网格密度,也就是说密度阈值 ϵ 的设定有很大的浮动空间,本文设定 ϵ 为待分选脉冲个数与网格个数之比,即网格的平均密度。

本文采用动态信号分选框架,对侦察数据流进行分选,实验证明,每一维划分的网格数 m 为 40 时,能够取得比较好的分选效果。实验选用的雷达类型、数目、存在时间以及特征参数值如表 1 所示。

表 1 侦察数据信息表						
Table 1 The information of the reconnaissance data						
信号类型	RF/GHz	PW/ μ s	DOA/($^{\circ}$)	PRI/ μ s	脉冲个数	存在时间/s
雷达 1	2.08 ~ 2.25 频率捷变	1.20 ~ 1.30	38 ~ 41	2 500 ~ 3 300 PRI 滑变	825	0.00 ~ 2.39
雷达 2	2.75 ~ 2.85 频率跳变	1.00 ~ 1.10	36 ~ 39	2 900 ~ 3 300 PRI 抖动	1 200	1.00 ~ 4.72
雷达 3	2.25 ~ 2.35 频率捷变	1.20 ~ 1.25	40 ~ 43	1 000 ~ 1 250 PRI 参差	2 190	1.50 ~ 4.07
雷达 4	2.55 ~ 2.75 频率跳变	1.30 ~ 1.40	42 ~ 45	1 100 ~ 1 400 PRI 抖动	1 500	3.00 ~ 4.72
噪声	2.08 ~ 2.75 随机分布	1.00 ~ 1.40 随机分布	36 ~ 45 随机分布	-	628	0.00 ~ 4.72 随机分布

实验 1:针对表 1 数据,测试快照存储时间间隔 t_0 对在线部分的时间消耗的影响,实验结果如图 2 所示。

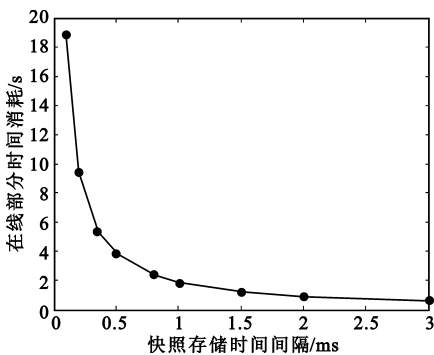


图 2 在线部分时间消耗

Fig.2 The time consumption of the online part

实验结果表明,随着快照存储时间间隔 t_0 的减小,在线部分时间消耗增加,因此为了满足信号分选实时性,应该合理设置快照存储时间间隔。大量实验表明,按照本次实验计算机的性能,快照存储时间间隔取 0.5 ms 时,可以满足雷达信号分选实时性的需要。

实验 2: $t_c = 3\text{ s}$ 、 $h = 2.5\text{ s}$ 、 $t_0 = 0.5\text{ ms}$ 时,采用动态信号分选框架分选表 1 中侦察数据流,聚类结果如表 2 所示。

实验 3: $t_c = 4.5\text{ s}$ 、 $h = 3\text{ s}$ 、 $t_0 = 0.5\text{ ms}$ 时,采用动态信号分选框架分选表 1 中侦察数据流,聚类结果如表 2 所示。

表 2 聚类分选准确率

Table 2 The accuracy of cluster-sorting

类型	雷达 1	雷达 2	雷达 3	雷达 4	平均
实验 2	0.931	0.944	0.925	-	0.933
实验 3	0.920	0.910	0.915	0.919	0.916

实验 2 和 3 结果表明,本文提出的动态信号分选框架能够有效地实现对侦察数据流中任意时间段内的数据进行分选,且无需先验知识,对噪声有很好的屏蔽作用,分选准确率较高。

为了进一步验证动态信号分选框架的有效性,本文分别使用文献[3]和文献[5]中的方法处理实验 3 的分选数据集,其中文献[3]是在选定合适容差的前提下进行的,文献[5]是在获得最佳 SVC 分选参数的条件下进行的。3 种方法的仿真结果对比如表 3 所示。

表 3 分选结果对比

Table 3 The comparison of sorting result

采用方法	分选准确率	时间消耗/s
文献[3]	75.4%	13.60
文献[5]	92.5%	365.40
本文方法	91.6%	4.31

实验结果表明,本文方法的分选准确率比文献[3]高 17.1%,只比文献[5]低 0.9%;而在分选实时性方面,本文提出的动态信号分选框架占有绝对优势。文献[3]方法由于对噪声数据敏感而分选准确率较低,文献[5]方法具有较高的分选准确率,但该方法计算复杂,性能远远达不到实时处理的需要。因此,本文方法的综合性能最优。

7 结束语

雷达侦察接收机是以动态数据流的形式输出侦察数据,而目前的信号分选方法都是建立在静态数据的基础上,难以满足信号分选实时性的需要。本文提出了一种动态信号分选框架,该框架基于数据流聚类算法,能够实现对任意时间段的侦察数据进行聚类分选,分选效率较高,在一定程度上解决了信号分选中准确性与实时性的矛盾问题,是对雷达信号实时分选的一种新探索。

参考文献:

[1] 何明浩. 雷达对抗信息处理[M]. 北京:清华大学出版社, 2010: 8-9.
HE Ming-hao. Radar ECM Information Processing[M]. Beijing: Tsinghua University Press, 2010: 8-9. (in Chinese)

[2] Milojevic D J, Popovic B M. Improved Algorithm for Deinterleaveing of Radar Pulses[J]. IEE Proceedings, 1992, 139 (1): 98-104.

[3] 祝正威. 雷达信号的聚类分选方法[J]. 电子对抗技术, 2005(3): 6-10.
ZHU Zheng-wei. Clustering Sorting Method of Radar Signa[J]. Electronic Warfare Technology, 2005(3): 6-10. (in Chinese)

[4] 郭杰,陈军文. 一种处理未知雷达信号的聚类分选方法[J]. 系统工程与电子技术, 2006, 28(6): 853-856.
GUO Jie, CHEN Wen-jun. Clustering approach for deinterleaving unknown radar signals[J]. Systems Engineering and Electronics, 2006, 28(6): 853-856. (in Chinese)

[5] 国强,王常虹,李峥. 支持向量聚类联合类型熵识别的雷达信号分选方法[J]. 西安交通大学学报, 2010, 44 (8): 63-67.
GUO Qiang, WANG Chang-hong, LI Zheng. Support Vector Clustering and Type-Entropy Based on Radar Signal Sorting Method[J]. Journal of Xi'an Jiaotong University, 2010, 44(8): 63-67. (in Chinese)

[6] Guha S, Mishra N, Motwani R. Clustering data streams [C]//Proceedings of the 41st IEEE Symposium on Foundations of Computer Science. [S.l.]: IEEE, 2000: 359-366.

作者简介:

张国毅(1965—),男,吉林长春人,博士,教授,主要研究领域为信号与信息处理、电子对抗;
ZHANG Guo-yi was born in Changchun, Jilin Province, in 1965. He is now a professor with the Ph.D. degree. His research interests include signal and information processing, electronic countermeasure.

王晓峰(1987—),男,河北承德人,硕士研究生,主要研究方向为雷达信号处理、电子对抗。
WANG Xiao-feng was born in Chengde, Hebei Province, in 1987. He is now a graduate student. His research interests include radar signal processing and electronic countermeasure.
Email: wx870516@126.com