

支持向量聚类联合类型熵识别的雷达信号分选方法

国强^{1,2}, 王常虹², 李峥³

(1. 哈尔滨工程大学信息与通信工程学院, 150001, 哈尔滨; 2. 哈尔滨工业大学空间控制与惯性技术研究中心, 150001, 哈尔滨; 3. 电子对抗国防科技重点实验室, 610036, 成都)

摘要: 针对复杂环境下传统雷达信号分选处理中由于设定容差难以准确分选的问题,提出一种分层互耦支持向量聚类(SVC)联合类型熵识别的多参数雷达信号分选方法. 该方法首先对雷达信号的全脉冲序列进行分层处理,再分别对每个子序列进行 SVC 聚类,然后引入分层耦合的思想,利用所提取子序列的全部质心,重新进行 SVC 聚类,将各分层的全脉冲序列中属于同一雷达信号源的子序列连接起来,从而实现对雷达全脉冲序列信号的分选. 根据类型熵随信号种类数的增加及信号复杂性的增加而增大的特点,对多参数聚类结果进行类型熵的计算,并将类型熵识别用来辅助信号分选. 实验结果表明,对于高脉冲密度环境和复杂的信号形式,提出的方法避免了传统信号分选方法中所面临的容差问题对信号分选的影响,可以实现有效分选.

关键词: 支持向量聚类; 类型熵; 雷达信号; 信号分选

中图分类号: TN974 **文献标志码:** A **文章编号:** 0253-987X(2010)08-0063-05

Support Vector Clustering and Type-Entropy Based Radar
Signal Sorting Method

GUO Qiang^{1,2}, WANG Changhong², LI Zheng³

(1. College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; 2. Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin 150001, China; 3. National Key Laboratory of Electronic Warfare, Chengdu 610036, China)

Abstract: Focusing on the problem that the conventional radar signal sorting methods based on the tolerance are difficult to accurate sorting in complex and dense pulse environment of modern electronic warfare, a new sorting method is presented based on delaminating coupling and support vector clustering (SVC). All pulse sequences of radar signals are determined using delaminating processing and then each sub-sequence is separately processed by SVC clustering. Then the idea of delaminating coupling is introduced. And re-SVC clustering is determined by using all centroids of the extracted sub-sequences so that the sub-sequences which belong to the same radar emitter in all pulse sequences of the various layers are connected to achieve the radar full pulse sequence sorting. The character that type-entropy value is getting bigger with the increasing of categories and complexity of pulse signals is used to calculate type-entropy on the multi-parameters clustering results. The type-entropy recognition is used to assist signal sorting. Experimental results show that the system can efficiently sort radar signals in the complex pulses environment and avoid the impact on signal sorting for the tolerance problem in traditional signal sorting methods.

Keywords: support vector clustering; type-entropy; radar signal; signal sorting

收稿日期: 2010-01-29. 作者简介: 国强(1972—),男,副教授,博士后. 基金项目: 国家自然科学基金资助项目(60872108);第二批中国博士后科学基金特别资助项目(200902411);第四十三批中国博士后科学基金资助项目(20080430903);黑龙江省博士后基金资助项目(LBH-Z08129);哈尔滨科技创新人才研究专项资金资助项目(2008RFQXG030);中央高校基本科研业务费专项资金资助项目(HEUCFZ1015).

信号分选是电子支援系统(ESM)中的重要技术,是实现雷达辐射源识别的重要基础^[1]. 传统的多参数信号分选技术是依次利用脉冲到达角、重频、载波频率及扫描方式等信息对雷达全脉冲序列进行去交错^[2-3],然而在高脉冲密度和大量复杂形式脉冲交叠的情况下,这种分选策略显然是低效的. 因为这种传统处理方式的分选效率是由上述处理过程中每一级预分选的准确率来决定的,其最终的分选结果可能使属于同一辐射源的脉冲序列被分割到不同的集合中,而分属于不同辐射源的部分脉冲被组合到同一个集合当中,即产生了增批和漏批现象. 基于此,文献[4-6]分别提出了基于不同机理的聚类分选算法,这些算法在一定程度上可以解决上述复杂的雷达信号分选问题,但仍然无法避免雷达信号分选领域长期以来一直面临的容差问题. 本文利用数据挖掘中的支持向量聚类分析理论和信息熵的理论,将基于支持向量机和分层互耦的聚类算法(DCSVC)与基于类型熵的识别技术相结合,提出一种新的多参数信号分选方法,可以有效解决容差问题. 通过计算机仿真实验,证明了这种新的雷达信号分选方法的有效性.

1 基于支持向量聚类和分层互耦的分选算法

本文将辐射源脉冲的到达角、载频和脉宽构成一个具有三维属性信息的雷达脉冲描述向量 $\mathbf{v}_i, i = 1, 2, \dots, N$ (N 为总脉冲个数),并将雷达全脉冲序列 $\{\mathbf{v}_i\}$ 截成 n 段长度为 $\text{ent}(N/n)$ 的子序列(ent 表示取整). 分选算法总体上分两步进行:首先分别对 n 段长度为 $\text{ent}(N/n)$ 的子序列进行支持向量聚类,将其聚成 $m_j (j=1, 2, \dots, n)$ 个子类;然后运用分层互耦的思想,将第 1 步得到的 $\sum_{j=1}^n m_j$ 个子类再次运用聚类算法进行组合,得到最终的分选结果.

1.1 支持向量聚类

支持向量聚类算法是由 Ben-Hur 等人提出的^[7-8],它的基本思想是:首先通过非线性变换将数据样本从属性空间变换到一个高维特征空间,然后在这个新空间中求取最优超球面. 这种非线性变换是通过定义核函数非线性映射来构造的,由于经过了核函数的映射,能够较好地分辨、提取并放大有用的特征,使原来没有显现的特征突现出来,从而能够更好地聚类.

设雷达脉冲描述向量 $\mathbf{v}_i$ 的数据空间 $V \subseteq \mathbf{R}^3$, 数

据集 $\{\mathbf{v}_i\} \subseteq V, i = 1, 2, \dots, N$. 由于接收到的雷达脉冲参数的分布特征具有复杂性,因此聚类集合的边界有可能是非常复杂的. 适当地选择一个非线性变换 Φ ,把数据从 V 空间映射到高维特征空间,可使数据点集的聚类特征变得更明显. 在特征空间中可以寻找一个最小半径为 R 的闭凸的超球体,其半径满足以下约束条件

$$\begin{aligned} \|\Phi(\mathbf{v}_j) - \mathbf{a}\|^2 &\leq R^2 + \xi_j \\ \forall j, \xi_j &\geq 0, j = 1, 2, \dots, N \end{aligned} \tag{1}$$

式中: $\mathbf{a}$ 为超球体的球心; ξ_j 为松弛量,作用是使聚类的集合变紧,从而使相交的凸球分离; $\|\cdot\|$ 为 Euclidean 范数.

式(1)的 Lagrangian 函数为

$$\begin{aligned} L = R^2 - \sum_j (R^2 + \xi_j - \|\Phi(\mathbf{v}_j) - \mathbf{a}\|^2) \beta_j - \\ \sum_j \xi_j \mu_j + C \sum_j \xi_j \end{aligned} \tag{2}$$

式中: $\beta_j \geq 0, \mu_j \geq 0$ 为 Lagrangian 乘子; 常数 C 称为惩罚因子, $C \sum_j \xi_j$ 为惩罚项.

根据 Karush-Kuhn-Tucker 条件^[7],可以得出以下结论.

(1)在满足 $\beta_i = C$ 处的特征空间的点 $\Phi(\mathbf{v}_i)$ 位于最小半径为 R 的超球体的外面,我们称这样的向量 $\mathbf{v}_i$ 为野值向量,它位于聚类的边界以外.

(2)在满足 $0 < \beta_i < C$ 处的特征空间的点 $\Phi(\mathbf{v}_i)$ 位于最小半径为 R 的超球体的表面上,我们称这样的向量 $\mathbf{v}_i$ 为支持向量(SV),它位于聚类的边界上.

(3)其余的点处于聚类边界的内部.

式(2)的 Wolfe 对偶形式为

$$W = \sum_j \Phi(\mathbf{v}_j)^2 \beta_j - \sum_{i,j} \beta_i \beta_j \Phi(\mathbf{v}_i) \Phi(\mathbf{v}_j) \tag{3}$$

式(3)的约束条件为

$$0 \leq \beta_j \leq C, \sum_j \beta_j = 1, j = 1, \dots, N \tag{4}$$

引入 Gaussian 核

$$K(\mathbf{v}_i, \mathbf{v}_j) = \Phi(\mathbf{v}_i) \Phi(\mathbf{v}_j) = e^{-q \|\mathbf{v}_i - \mathbf{v}_j\|^2} \tag{5}$$

式中: q 为 Gaussian 核的宽度参数.

式(3)可以改写为

$$W = \sum_j K(\mathbf{v}_i, \mathbf{v}_j) \beta_j - \sum_{i,j} \beta_i \beta_j K(\mathbf{v}_i, \mathbf{v}_j) \tag{6}$$

对于每一个数据向量 $\mathbf{v}$,其映射到特征空间中的像到球心的距离表示为

$$\begin{aligned} R^2(\mathbf{v}) = \|\Phi(\mathbf{v}) - \mathbf{a}\|^2 = \left\| \Phi(\mathbf{v}) - \sum_j \beta_j \Phi(\mathbf{v}_j) \right\|^2 = \\ K(\mathbf{v}, \mathbf{v}) - 2 \sum_j \beta_j K(\mathbf{v}_j, \mathbf{v}) + \end{aligned}$$

$$\sum_{i,j}\beta_i\beta_jK(\mathbf{v}_i,\mathbf{v}_j) \tag{7}$$

满足条件 $\{\mathbf{v}|R(\mathbf{v})=R_0\}$ 的点组成等高面,其中 $R_0=\{R(\mathbf{v}_i)|\mathbf{v}_i\}$ 为支持向量. SV 位于等高面上,它确定了属于同一类雷达辐射源参数的聚类边界.

建立聚类标识关联矩阵:对于 2 个属于不同聚类的数据点,在这 2 点的连线上存在点 $\mathbf{y}$ 使得 $R(\mathbf{y})>R$,据此可以建立关联矩阵 $\mathbf{A}$, $\mathbf{A}$ 中的元素为

$$A_{ij}=\begin{cases}1,&R(\mathbf{y})\leqslant R_0\\0,&\text{其他}\end{cases} \tag{8}$$

聚类可由此定义为 $\mathbf{A}$ 中邻近元素的集合. 本文通过 DFS 算法根据关联矩阵进行聚类分配.

1.2 基于支持向量聚类的分层互耦算法

在雷达信号分选中,所需处理的全脉冲数据量很大,如果把它们全部作为数据样本,就会造成支持向量聚类(SVC)运算的关联矩阵规模庞大,极大地影响其运算的速度. 因此,本文采用基于支持向量机的分层互耦聚类方法,具体步骤如下.

(1)对雷达信号的全脉冲序列进行分层处理,即将全脉冲序列截短,生成 m 个子序列,然后分别对每个子序列进行 SVC 算法聚类,得到的聚类结果分别为 $n_i(i=1,\cdots,m)$ 类,再分别提取第 $i(i=1,\cdots,m)$ 个子序列中第 $j(j=1,\cdots,n_i)$ 个聚类集合的质心 G_{ij} (即求取各集合参数的统计均值).

(2)耦合处理,即利用所提取的全部质心,重新进行 SVC 聚类,将各分层的全脉冲序列中属于同一雷达信号源的序列连接起来,从而实现对雷达全脉冲序列信号的分选. SVC 分层互耦算法的结构如图 1 所示.

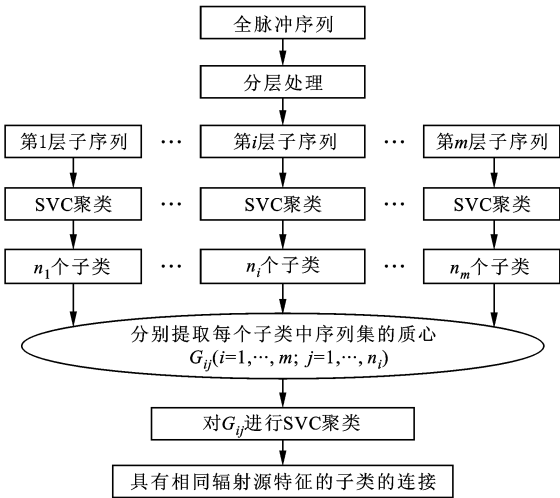


图 1 SVC 分层互耦算法的结构

2 利用熵度量的雷达全脉冲信号识别方法

在电子对抗系统中,目前评估和测度电磁信号环境的复杂程度常用的指标是脉冲密度. 但是,单纯强调外界高的脉冲密度对衡量系统的能力没有太大的意义,脉冲密度相同不能表示同样复杂的信号环境,并且脉冲密度本身缺乏一个公认的可测量的定义. 从这个意义上讲,脉冲密度并非一个衡量电磁信号环境复杂程度的合适量值. 寻找一个合适的物理量作为信号环境复杂程度的量度不但一直为工程实践所迫切需要,也是一个困扰电子对抗领域多年的难题. 采用信息熵的概念,为科学评估信号环境提供了一个可行的依据.

本文将 ESM 系统所面临的雷达全脉冲环境(或此全脉冲环境的子集)看作信源,这样脉冲环境的复杂程度可以用信源的不确定度来表示. 一般情况下,一个信源可以用一个概率空间来描述,而信源的不确定程度可以用这个概率空间的可能状态数目及其概率来描述^[9]. 设信源 $[X]$ 的概率空间为 $\{X,P(x)\}$,其中 $X=\{x_1,x_2,\cdots,x_n\}$ 为信源中随机事件的状态空间, $P(x)=\{P(x_1),P(x_2),\cdots,P(x_n)\}$ 为随机事件各可能状态的概率分布,且 $\sum P(x)=1$.

在信息论中,把信源输出一个消息所提供的平均信息量或者信源的不确定度 $H(x)$ 叫做信源熵,表示为

$$H(X)=-\sum_{i=1}^n P(x_i)\text{lb}P(x_i) \tag{9}$$

从式(9)可以看出,当信源概率空间的概率分布为等概率时,具有最大的不确定度 $H(x)$,其大小与概率空间的可能状态数或概率有关,状态数愈多或概率愈小, $H(x)$ 愈大. 信源熵是信源概率空间的概率分布的函数.

根据信源熵的定义,针对雷达全脉冲信号环境的特点,本文引入类型熵来描述信号环境的复杂程度. 类型熵可以表示为对脉冲种类描述的估计. 假设一个脉冲可以用载频、脉宽和到达角来进行描述,则相同的载频、脉宽和到达角被视为一种脉冲,在这个意义上,类型熵用来描述信号环境中存在脉冲种类的多少,可定义为

$$H_T(P)=-\sum_{n=1}^N P_n\text{lb}P_n \tag{10}$$

式中: P_n 是每种脉冲出现的概率; N 是脉冲的种类.

信号类型不同,各类信号的概率亦不同. 在单位

时间内,概率 P_i 与重频和脉冲类型有关

$$P_i = \frac{S_i}{f_{\text{PRF}}}, \quad i = 1, 2, \dots, N \tag{11}$$

式中: S_i 为单位时间内每种类型脉冲出现的次数; f_{PRF} 为脉冲重复频率.

- 通过以上分析,可以得出如下结论.
- (1)捷变信号、分集信号可以大大加强信号环境的复杂程度,而且一个特殊信号引起的复杂程度远大于一个常规信号.
 - (2)信号数量和类型相同时,重频越接近的情况熵值越大,事实上,这种情况下的分选也更加困难.
 - (3)类型熵与脉冲的最大种类有密切的关系,其最大值为 $\text{lb}N$. 随着信号种类数的增加,环境的熵值也越来越大.

3 利用类型熵调节聚类分选参数

在基于 SVC 的聚类分选过程中,SVC 的聚类边界受 Gaussian 核的宽度参数 q 和 Lagrangian 函数的惩罚因子 C 的控制:随着参数 q 的增加,聚类边界表现出更紧的特性;参数 C 的大小决定了野值点的数量,随着 C 值($C \leq 1$)的减小,野值点的数量会增加,通过 C 值的减小可以平滑聚类边界^[8].

参数 q 和 C 的调整过程采用分裂聚类算法^[8],首先确定参数 q 的初始值

$$q = \frac{1}{\max_{i,j} \| \mathbf{v}_i - \mathbf{v}_j \|^2} \tag{12}$$

在此初始 q 值下,Gaussian 核函数的值会很大,导致聚类分选结果产生一个单一的大的聚类.在这种情况下不需要设置野值点,因此取参数 $C=1$. 然后增大参数 q 的值,这样,单一的聚类开始分裂.随着 q 值的增大,可能引起聚类的碎片过多,聚类的边界过于粗糙,这时可以通过减小 C 值来设置野值点.通常一个好的聚类标准^[9]是用支持向量(SV)的数目来衡量的,即力图使用尽可能少的支持向量来获得平滑的聚类边界.

使用分裂聚类方法所面临的一个重要的问题

是:确定什么时候停止对聚类的分裂.目前采用的方法是用支持向量的数目来作为最终结果是否合理的一个评价^[8],如果支持向量的数目超过了某个门限便停止聚类过程.然而,这个门限在实际应用中很难被合理地确定,原因在于缺乏一个稳定的具有物理意义的量来保证这一门限确定的合理性.引入类型熵恰恰解决了这个问题,根据类型熵随信号种类数的增加及信号复杂性的增加而增大的特点,我们可以对每组多参数聚类结果进行类型熵的计算,通过识别聚类结果的复杂度,对聚类分选结果进行宏观分析,从而对聚类结果进行判断,来确定最终的聚类分选参数 q 和 C .

4 仿真实验结果

为了验证新的分选方法的有效性,本文对上述的多参数聚类分选系统进行了仿真实验.在产生全脉冲仿真数据时,需设置采样间隔,并对同时到达的信号进行丢失处理,得到的仿真数据如表 1 所示.

从表 1 中的雷达数据可以看出,4 部雷达的重频、载频、脉宽和到达角参数均有重叠或部分重叠,利用现有基于设定容差的信号分选方法,必然会存在以下 2 个问题:如果容差选择小了,其最终的分选结果会使属于同一辐射源的脉冲序列被分割到不同的集合中,而如果容差选择大了,会使分属于不同辐射源的部分脉冲被组合到同一个集合当中,即产生了漏批和增批现象.

将得到的雷达全脉冲数据流中的前 5 000 个脉冲先进行参数规一化处理,再利用 DCSVC 聚类分选方法对全脉冲数据流进行分选,通过类型熵识别技术来调节聚类分选参数,调整参数 $q=30$ 、 $C=1$. 聚类分选效果如图 2、图 3 所示.图 2 显示了利用支持向量聚类方法对 50 个数据样本聚类所得到的二维属性参数分布情况,图中到达角、脉宽和载频参数分别为其属性参数的规一化表示.从聚类结果可以看出,支持向量的分布很好地表示了雷达脉冲信号

表 1 雷达参数信息表

雷达序号	重频/kHz	载频/MHz	脉宽/ μs	到达角/($^\circ$)	信号数量/个
1	0.3~0.4(重频滑变)	2 080~2 250(频率捷变)	1.2~1.3	48~60	824
2	0.3~0.4(重频抖动)	2 750~2 850(频率捷变)	1.0~1.1	60~80	823
3	0.8~1.0(重频参差)	2 250~2 350 (频率跳变,跳变数取 10)	1.20~1.25	68~80	2 149
4	0.7~0.9(重频抖动)	2 550~2 750(频率捷变)	1.3~1.4	56~64	1 891

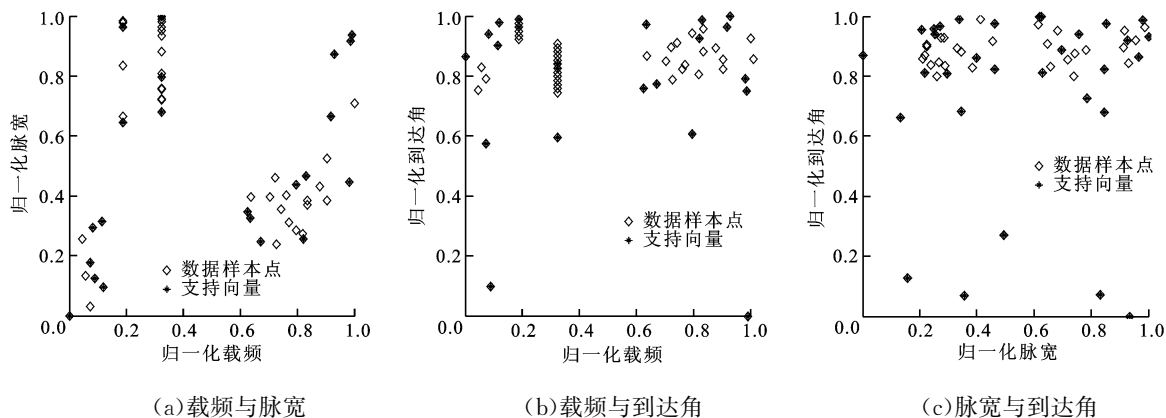


图 2 用 SVC 方法对 50 个数据样本聚类得到的二维属性参数分布图

数据样本聚类边界的轮廓。

图 3 为本文提出的多参数聚类方法分选结果的统计直方图。从图中可以看出,表 1 给出的混合雷达的信号中,4 部雷达信号参数得到了较好的聚类分选结果(分选结果中只有 1 部雷达的少量脉冲被错误地分选到聚类 5 中),对得到的分选结果进行统计,算得分选准确率为 98.06%。

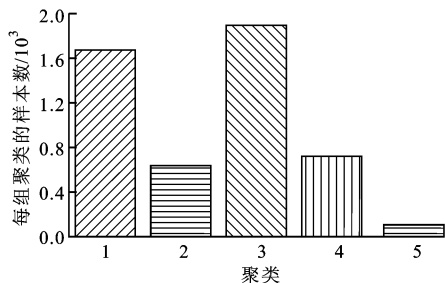


图 3 本文聚类方法的分选结果统计直方图

5 结束语

本文提出了基于支持向量聚类和分层互耦的分选算法,引入类型熵的概念,并将类型熵识别用于辅助信号分选,进而提出了一种新的多参数雷达全脉冲信号分选方法。仿真实验表明,对于高脉冲密度环境和复杂的信号形式,这种分选方法是有效的。

参考文献:

- [1] 张葛祥,金炜东,胡来招. 基于粗集理论的雷达辐射源信号识别[J]. 西安交通大学学报, 2005, 39(8): 871-875.
ZHANG Gexiang, JIN Weidong, HU Laizhao. Radar emitter signal recognition based on rough set theory [J]. Journal of Xi'an Jiaotong University, 2005, 39

(8): 871-875.

- [2] MILOJEVIC D J, POPOVIC B M. Improved algorithm for deinterleaving of radar pulses [J]. IEE Proc, F: Comm, Radar and Signal Processing, 1992, 139(1): 98-104.
- [3] ROGERS J A V. ESM processor system for high pulse density radar environments [J]. IEE Proc, F: Comm, Radar and Signal Process, 1985, 132(7): 621-625.
- [4] 祝正威. 雷达信号的聚类分选方法[J]. 电子对抗, 2005(6): 6-10.
ZHU Zhengwei. Clustering sorting method of radar signals [J]. Electronic Warfare, 2005(6): 6-10.
- [5] ATAA A W, ABDULLAH S N. Deinterleaving of radar signals and PRF identification algorithms [J]. IET Radar, Sonar & Navigation, 2007, 1(5): 340-347.
- [6] 张红昌,阮怀林,龚亮亮,等. 一种新的未知雷达辐射源聚类分选方法[J]. 计算机工程与应用, 2008, 44(27): 200-202.
ZHANG Hongchang, RUAN Huailin, GONG Liangliang, et al. New clustering approach for sorting unknown radar emitter signal [J]. Computer Engineering and Applications, 2008, 44(27): 200-202.
- [7] CRISTIANINI N, SHAW-TAYLOR J. An introduction to support vector machines and other kernel-based learning methods [M]. Cambridge, UK: Cambridge University Press, 2000.
- [8] BEN-HUR A, HORN D, SIEGEL M H T, et al. Support vector clustering [J]. Journal of Machine Learning Research, 2001(2): 125-137.
- [9] 贾世楼. 信息论理论基础 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2001.

(编辑 刘杨)