

一种基于 MDL 原理的信号分选新算法

黄桂根 孟兵 李品

(南京电子技术研究所, 南京, 210039)

摘要: 研究了一种基于脉内调制特征的雷达信号聚类分选算法, 该聚类分选算法根据最小描述长度 (Minimum description length, MDL) 原理判定数据源中存在数据类的数量, 以及任一向量应该归属于哪一类, 从而确定电磁信号环境中雷达辐射源的数量, 并完成脉冲信号的分选处理。该算法将截获系统前端对雷达脉冲信号的采集数据视为脉冲向量, 在对雷达脉冲向量进行必要的预处理后, 进行聚类分选处理。来自不同雷达的脉冲信号由于它们的脉内调制 (有意或无意调制) 特征不同, 聚类处理后将会形成不同的脉冲向量类。该算法的优势在于无需进行信号脉内特征的定量分析、提取和描述等处理。计算机仿真验证了该算法的有效性。

关键词: 电子对抗; 信号分选; 最小描述长度准则; 聚类

中图分类号: TN971 **文献标志码:** A

New Radar Signal Deinterleaving Algorithm Based on MDL Criterion

Huang Guigen, Meng Bing, Li Pin

(Nanjing Research Institute of Electronics Technology, Nanjing, 210039, China)

Abstract: A new radar pulse signal clustering algorithm based on inner-pulse modulation characteristic is studied. The number of clusters in original data and the cluster that data vector belongs to is determined by clustering algorithm using the minimum description length (MDL), thus obtaining the number of radar emitters in the electromagnetic environment. The radar pulse data collected by the interception system front-end is treated as vectors and is sorted after the necessary pre-processing. Pulse signals from different radars form different vector clusters after deinterleaving by using the algorithm based on MDL criterion, because of their different inner-pulse modulation characteristic. The largest advantage of the algorithm is that it is unnecessary for quantitative analysis, extraction and description of the inner-pulse modulation characteristic. Simulation proves its effective.

Key words: electronic countermeasures; signal deinterleaving; minimum description length criterion; clustering

引 言

信号分选是雷达侦察信号处理的核心环节^[1], 也是雷达信号侦察与对抗领域的研究热点问题之一^[2-3]。传统上, 雷达信号分选技术大多基于由 5 大特征参数构成的脉冲描述字 (Pulse description

word, PDW)实现^[4-8],包括:脉冲宽度、脉冲频率、脉冲幅度、脉冲到达时间^[9-12]和脉冲到达角。此类信号分选算法的性能受限制于截获系统的信号检测能力和参数测量精度。

近年来,基于脉内调制特征的信号分选研究越来越受到科研人员的重视^[13]。通过对信号脉内调制方式分类和调制参数估计,形成包含脉内调制方式等特征参数的精细脉冲描述字,依据该脉冲描述字可以获得更为可靠的信号分选性能。针对带有复杂有意调制方式的雷达脉冲信号,此类算法具有明显的性能优势。但是,其缺陷在于不能提取无意调制特征;同时,为了保证脉内调制方式分类识别准确率和参数估计精度,对输入 SNR、系统采样率有较苛刻的要求。

最小描述长度准则是一种信息压缩准则,属于归纳推理工具^[14],能够通用地解决模型选择问题,普遍被应用到选择最优分类模型问题上,被广泛应用于多个领域,包括降噪^[15]、压缩感知^[16]、医学图像处理^[17]、语音识别^[18]等。本文将最小描述长度准则用于脉冲数据聚类处理,不同的雷达辐射源信号由于存在有意或无意调制,将形成不同的向量类,从而实现雷达信号分选。

1 MDL 原理

最小描述长度(Minimum description, MDL)准则源于最优编码。将数据 z 视为信息,对其进行编码并发送给接收者。将模型视为对数据编码的方法,并且选择最节俭的模型,即对于传输,它是短编码。

首先,假设要传输的信息是 z_1, z_2, \dots, z_m ,使用长度为 A 的有限字母表。例如,可以使用长度 $A=2$ 的二进制编码 $\{0,1\}$ 。这是一个有 4 种可能信息的例子,其二进制编码见表 1。

表 1 瞬间前缀码编码规则

Table 1 Coding rule of instant prefix code

信息	z_1	z_2	z_3	z_4
码	0	10	110	111

这是一种所谓的瞬间前缀码,即任何一个码都不是另一个码的前缀,且接收者确切知道何时信息被全部发送。本文的研究就针对这样的瞬间前缀码。

可以使用表 1 中的编码,或者可以排列这些码,例如对于 z_1, z_2, z_3, z_4 使用码 110,10,111,0。决定使用哪一种编码取决于发送信息的频率。例如,如果发送 z_1 最频繁,则对 z_1 使用最短的码 0。使用这种策略平均信息长度将比较短。

假设信息发送的概率为 $\Pr(z_i)$ ($i=1,2,3,4$),则由香农定理可知,应当使用长度为 $L_i = -\log_2(\Pr(z_i))$ 的码,且平均信息长度满足

$$E(\text{长度}) \geq - \sum \Pr(z_i) \log_2(\Pr(z_i)) \quad (1)$$

式(1)右端也称分布 $\Pr(z_i)$ 的熵。当概率满足 $p_i = A^{-L_i}$ 时,式(1)取等号。在前面的例子中,如果有 $\Pr(z_i) = 1/2, 1/4, 1/8, 1/8$,则式(1)中显示的编码是最优的并且能够达到熵的下界。通常情况,下界难以达到,对于无限的信息集,熵可以用 $-\int \Pr(z) \log_2 \Pr(z) dz$ 代替。

由上面的结果可见:为了传递具有概率密度函数 $\Pr(z)$ 的随机变量 z 需要大约 $-\log_2 \Pr(z)$ 位信息。为了方便,通常将 $-\log_2 \Pr(z)$ 记为 $\log \Pr(z) = \log_e \Pr(z)$ 。

假设有一个以 θ 为参数的模型 M 和包括输入、输出的数据 $Z=(X,Y)$ 。令该模型下输出的(条件)概率是 $\Pr(y|\theta, M, X)$,假设接收者知道全部输入,并且希望传送输出。那么传送输出所需要的信息长度是

$$\text{长度} = -\log \Pr(y | \theta, M, X) - \log \Pr(\theta | M) \quad (2)$$

它是给定输入目标值的对数概率。第二项是传送模型参数 θ 的平均码长,而第一项是传送模型和实际目标值之间的偏差的平均码长。例如,假设有单一目标 $y \sim N(\theta, \sigma^2)$, 参数 $\theta \sim N(0, 1)$, 并且无输入(为了简化), 则信息长度是

$$\text{长度} = \text{常量} + \log \sigma + \frac{(y - \theta)^2}{\sigma^2} + \frac{\sigma^2}{2} \quad (3)$$

式中:由于 y 比较集中于 θ 附近,所以较小的 σ 就是较短的信息长度。

MDL 原理表明,应当选择能够使式(2)极小化的模型。将式(2)视为(负的)对数后验分布,因而极小化描述长度等价于极大化后验概率。

2 截获系统接收脉冲信号模型及其预处理

假设环境中,存在 K 个独立的雷达辐射源,接收机一共接收到 N 个来自这些辐射源的脉冲信号。用 $x_n(t; \alpha_n)$ ($n=1, \dots, N$) 表示第 n 个接收到的脉冲,其中“ $\alpha_n = k$ ”表示第 n 个脉冲来自第 k 个雷达辐射源。那么,第 n 个脉冲可表示为

$$x_n(t; \alpha_n) = \eta_n a_{\alpha_n}(t - \tau_n) \exp\{j[\psi_n + \omega_n(t - \tau_n) + \phi_{\alpha_n}(t - \tau_n)]\} + \nu_n(t - \tau_n) \quad (4)$$

式中: η_n 为第 n 个脉冲信号的初始幅度; ψ_n 为第 n 个脉冲信号的初始相位; τ_n 为第 n 个脉冲信号相对于参考时刻的时延; ω_n 为第 n 个脉冲的载频; $a_{\alpha_n}(t)$ 为第 n 个脉冲的原始包络,有 $a_{\alpha_n}(t) \in \{a_1(t), a_2(t), \dots, a_K(t)\}$; $\phi_{\alpha_n}(t)$ 为第 n 个脉冲的原始相位,有 $\phi_{\alpha_n}(t) \in \{\phi_1(t), \dots, \phi_K(t)\}$; $\nu_n(t)$ 为伴随第 n 个脉冲的高斯噪声。

脉冲信号 $x_n(t; \alpha_n)$ 进入截获接收机后,通过对信号的幅度进行门限检测后,可消除时间延迟 τ_n , 那么,式(4)可重写为

$$x_n(t; \alpha_n) = \eta_n a_{\alpha_n}(t) \exp\{j[\psi_n + \omega_n t + \phi_{\alpha_n}(t)]\} + \nu_n(t) \quad (5)$$

就单个脉冲信号而言:参数 η_n, ψ_n 和 ω_n 因单个脉冲信号而异;高斯噪声 $\nu_n(t)$ 在不同脉冲信号之间是独立同分布的;原始包络 $a_{\alpha_n}(t)$ 和原始相位 $\phi_{\alpha_n}(t)$ (非线性)则是雷达辐射源的特征参数,它们取决于雷达脉冲信号的有意和无意脉内调制特征。将式(5)分解,可得到该脉冲信号的幅度表达式和相位表达式分别为式(6)和(7)。

$$s_n^a(t; \alpha_n) = \eta_n a_{\alpha_n}(t) \quad (6)$$

$$s_n^p(t; \alpha_n) = \psi_n + \omega_n t + \phi_{\alpha_n}(t) \quad (7)$$

为了能够顺利应用脉冲信号的聚类处理算法,需要对式(5)所示脉冲信号进行预处理,以便在消除冗余参数 η_n, ψ_n 和 ω_n 时保留特征参数 $a_{\alpha_n}(t)$ 和 $\phi_{\alpha_n}(t)$ 。消除冗余参数包括幅度归一化处理和基于多项式拟合的相位调整。

2.1 幅度归一化处理

令 $S_n^a(\omega; \alpha_n)$ 和 $A_{\alpha_n}(\omega)$ 分别表示 $s_n^a(t; \alpha_n)$ 和 $a_{\alpha_n}(t)$ 的傅里叶变换,那么通过式(8)可以实现消除幅度常数 η_n 和对幅度函数 $s_n^a(t; \alpha_n)$ 归一化的目的。

$$\hat{S}_n^a(\omega; \alpha_n) = \frac{S_n^a(\omega; \alpha_n)}{S_n^a(0; \alpha_n)} = \frac{A_{\alpha_n}(\omega)}{A_{\alpha_n}(0)} \quad (8)$$

然后对 $\hat{S}_n^a(\omega; \alpha_n)$ 进行逆傅里叶变换,得到 $\hat{s}_n^a(t; \alpha_n)$ 即为幅度归一化之后的时域幅度函数,该函数仅与包络 $a_{\alpha_n}(t)$ 有关。

事实上,由于

$$S_n^a(0; \alpha_n) = \int S_n^a(t; \alpha_n) dt \quad (9)$$

即 $S_n^a(0; \alpha_n)$ 为函数 $s_n^a(t; \alpha_n)$ 与时间轴所围图形的面积,实际上,为了减小运算量,幅度归一化处理的过

程直接可用下式表示

$$\hat{s}_n^a(t; \alpha_n) = \frac{s_n^a(t; \alpha_n)}{\int s_n^a(t; \alpha_n) dt} \quad (10)$$

2.2 基于多项式拟合的相位调整

考虑到线性函数可以通过多项式曲线拟合的方法进行估计,可以消除式(6)所示信号相位函数中的线性相位部分。令 $g(t) = \psi_n + \omega_n t$, 通过在最小二乘意义上使得 $g(t)$ 和 $s_n^p(t; \alpha_n)$ 的误差最小, 可以确定系数 ψ_n 和 ω_n 的值。从而, 根据式(10)可以消除 $s_n^p(t; \alpha_n)$ 中的线性相位。

$$s_n^p(t; \alpha_n) = s_n^p(t; \alpha_n) - \psi_n - \omega_n t = \phi_{\alpha_n}(t) \quad (11)$$

事实上, 上面的预处理过程都附着有噪声干扰。通过“幅度归一化处理”和“基于多项式拟合的相位调整”两个预处理步骤, 可以得到消除冗余参数 η_n, ψ_n 和 ω_n 后的带高斯噪声的脉冲信号

$$\hat{y}_n(t; \alpha_n) = \hat{s}_n^a(t; \alpha_n) \exp[j\hat{s}_n^p(t; \alpha_n)] + \hat{v}_n(t) = s_{\alpha_n}(t) + \hat{v}_n(t) \quad (12)$$

式中: $\hat{v}_n(t)$ 为预处理后的噪声, 仍然服从高斯分布; $s_{\alpha_n}(t) = A_{\alpha_n}^{-1}(0) a_{\alpha_n}(t) \exp[j\phi_{\alpha_n}(t)]$ 为理想的预处理后的脉冲波形。

由于前面的预处理都在离散时间上完成, 上式又可以表示为

$$\hat{y}_n(mT; \alpha_n) = s_{\alpha_n}(mT) + \hat{v}_n(mT), \quad m = 1, 2, \dots, M' \quad (13)$$

式中: T 为采样间隔, M' 表示单个脉冲的采样点数。式(13)可用矢量表示为

$$\hat{\mathbf{y}}_n(\alpha_n) = s_{\alpha_n} + \hat{\mathbf{v}}_n \quad (14)$$

3 MDL 雷达脉冲聚类算法

MDL 准则表明, 给定一个数据集合和一组竞争统计模型, 最佳的统计模型将为数据集合产生最小的描述长度。根据 MDL 原理可以将其应用于雷达脉冲信号聚类分选处理。

用 $Y = \{y_1, \dots, y_N\}$ 表示一个包含多个类的数据集合, 集合 $\hat{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$ 为 Y 的关联数据集合, 且有

$$\alpha_i = k \Rightarrow y_i \in \text{cluster } k \quad i = 1, \dots, N; k = 1, \dots, \hat{K} \quad (15)$$

即关联集合 $\hat{\alpha}$ 将数据集 Y 划分为 \hat{K} 个数据类。将属于同一个类的数据元素放置在一起, 那么, 划分后的数据集 Y 可重新表示为

$$Y_{\hat{\alpha}} = \{\{y_1(1), \dots, y_{N_1}(1)\}, \dots, \{y_1(\hat{K}), \dots, y_{N_{\hat{K}}}(\hat{K})\}\} \quad (16)$$

用 θ 表示 \hat{K} 个类的模型参数集合, 且 $\hat{\theta}$ 为 θ 的最大似然估计。那么可以用一个条件密度函数 $f(Y|\hat{\theta}, \hat{\alpha})$ 对数据集合 Y 进行建模, 并定义 Y 在该模型下描述长度为

$$L(Y, \hat{K}) = L(Y|\hat{\theta}, \hat{\alpha}) + L(\hat{\theta}, \hat{\alpha}) \quad (17)$$

式中: $L(Y|\hat{\theta}, \hat{\alpha})$ 为用 $\hat{\theta}$ 和 $\hat{\alpha}$ 对 Y 建模并编码时的码长; $L(\hat{\theta}, \hat{\alpha})$ 为对 $\hat{\theta}$ 和 $\hat{\alpha}$ 进行编码时的码长。这里用负的对数似然函数 $-\log f(Y|\hat{\theta}, \hat{\alpha})$ 表示 $L(Y|\hat{\theta}, \hat{\alpha})$, 该似然函数用于衡量数据集合与模型之间的适合程度; $L(\hat{\theta}, \hat{\alpha})$ 则是作为衡量模型复杂度的代价函数出现。

MDL 准则通过下面的公式确定 Y 中类的数量

$$K^* = \arg(\min_{1 \leq \hat{K} \leq N} L(Y, \hat{K})) \quad (18)$$

在给定数据集合 Y 时, MDL 准则的目的是通过使用关联向量 $\hat{\alpha}$ 找到 Y 中数据类的数量, 根据关联向量 $\hat{\alpha}$ 可以计算最小描述长度 $L(Y, \hat{K})$ 。该准则可以被视为在数学模型的复杂度和数学模型对数据集合的适合度之间求取平衡。

对于雷达脉冲数据而言, 数据集合 $Y = \{y_1, \dots, y_N\}$ 中的每个元素 $y_i (i = 1, \dots, N)$ 均为数据采集后的雷达脉冲数据向量。因为数据采集时, 每个雷达脉冲对应的数据向量可能包含几十、几百甚至上千个采样数据, 考虑到正交相检, 每个采样数据可以为复数, 即 $Y = \{y_1, \dots, y_N\}$ 可为多维复向量集合。

假设雷达脉冲向量中附加的噪声为高斯分布。那么,来自第 k 个雷达辐射源的脉冲向量 $\mathbf{y}_1(k), \dots, \mathbf{y}_{N_k}(k)$ 是服从均值为 μ_k , 协方差矩阵为 Σ_k 的多参数正态分布样本。那么,样本均值的估计量 $\hat{\mu}_k$ 和协方差矩阵的估计量 $\hat{\Sigma}_k$ 可以通过下面的公式计算。

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{y}_n(k) \quad (19)$$

$$\hat{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} (\mathbf{y}_n(k) - \hat{\mu}_k)(\mathbf{y}_n(k) - \hat{\mu}_k)^T \quad (20)$$

数据集合 Y 的条件密度函数为 \hat{K} 维正态分布的联合分布,其参数集为 $\theta = \{\mu_1, \mu_K, \hat{\Sigma}_1, \dots, \hat{\Sigma}_K\}$ 。假设噪声样本之间相互独立,该假设对于雷达辐射源分类处理通常是合理的,则协方差矩阵 Σ_k 为

$$\Sigma_k = \begin{bmatrix} \sigma_{k1}^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_{kM}^2 \end{bmatrix}, \quad k=1, \dots, K \quad (21)$$

式中 M 表示单个脉冲向量的维数。可以定义样本标准偏差 $\hat{\sigma}_k$ 为

$$\hat{\sigma}_k^2 = \frac{1}{N_k} \sum_{n=1}^{N_k} (\mathbf{y}_n(k) - \hat{\mu}_k)^2, \quad k=1, \dots, \hat{K} \quad (22)$$

分别用 μ_{km} 和 σ_{km} 表示第 m 个元素的均值 μ_k 和标准偏差 σ_k 。由于来自同一雷达辐射源的每个脉冲都会出现由电子电路产生的随机噪声,所以,如果采样来自同一雷达辐射源脉冲,则可以假设所有的 μ_{km} 和 σ_{km} 独立同分布。

根据假设,式(17)中右边两项可通过下面的公式计算

$$L(Y | \theta, \alpha) = -\log f(Y | \theta, \alpha) = \frac{1}{2} \sum_{k=1}^K N_k \log |\text{diag}(\hat{\Sigma}_k)| + \frac{MN}{2} + \frac{MN}{2} \log(2\pi) \quad (23)$$

$$L(\theta, \alpha) = \sum_{k=1}^K \log \left(\frac{|\text{diag}(\hat{\Sigma}_0)|}{|\text{diag}(\hat{\Sigma}_k)|} \right) + 2M \sum_{k=1}^K \log(N_k) + KM \log \left(\frac{\sqrt{2}e}{6N} \right) + \log S(N, \alpha) \quad (24)$$

式中: $\text{diag}(\cdot)$ 表示取一个矩阵的对角阵; $|\cdot|$ 表示计算一个矩阵的行列式; $\hat{\Sigma}_0$ 表示把 Y 视为一个类时计算得到的样本协方差矩阵; $S(N, \hat{\alpha})$ 表示不同划分的数量,由式(25)给出计算表达式。

$$S(N, \hat{\alpha}) = \frac{N!}{N_1! \cdots N_K!} \cdot \frac{1}{m_1! \cdots m_N!} \quad (25)$$

式中: N_k 为归属于第 k 个类中的脉冲数据向量的数量(即有 $k=1, \dots, \hat{K}, N_1 + N_2 + \dots + N_K = N$), m_n 表示包含有 n 个脉冲数据向量的类的数量。

4 MDL 雷达脉冲聚类处理流程

为了能够将 MDL 准则合理应用于雷达脉冲信号分选处理,首先需要为每个假定的类数量 \hat{K} , 找到一个适合的划分 $\hat{\alpha}$, 通过该划分可以计算出其对应的最小描述长度 $L(Y, \hat{K})$ 。 \hat{K} 值不同, 则对应的最小描述长度 $L(Y, \hat{K})$ 也不相等, 在 \hat{K} 的可选范围内计算的所有 $L(Y, \hat{K})$ 值中, 最小的 $L(Y, \hat{K})$ 值对应的分类数量 \hat{K} 值即是最佳分类时数据类(雷达辐射源)的数量 K^* 。

为了直观和工程实现的需要, 将上面所述聚类过程用迭代过程描述如下:

首先, 令 $\hat{K}=1$, 即将整个数据集合 Y 视为同一个类。分别计算出这个类的样本均值 $\hat{\mu}_1$ 和标准偏差 $\hat{\sigma}_1$, 并根据式(17, 23, 24) 计算此时的描述长度 $L(Y, \hat{1})$ 。

第 2 次迭代($\hat{K}=2$)时, 数据集合 Y 被划分为两个类, 这两个类的中心根据式(26) 进行初始化。标准偏差向量 $\hat{\sigma}_1$ 给出了数据的散布大小, 因此根据它进行数据类的分裂。然后根据新形成的类中心, 对 Y 中的每个数据向量重新进行分类处理, 分类的依据是数据向量与类中心的距离。对 Y 多次聚类处理,

直到两个类的中心都收敛,然后对 Y 做聚类处理,类的中心不会发生改变。最后,计算出对应的 $L(Y, \hat{\alpha})$ 值。

$$\begin{aligned} \hat{\mu}_1 &\leftarrow \hat{\mu}_1 - \hat{\sigma}_1 \\ \hat{\mu}_2 &\leftarrow \hat{\mu}_1 + \hat{\sigma}_1 \end{aligned} \tag{26}$$

通常,在将 Y 划分为 \hat{K} 个类之后,要完成下一次迭代处理,即需要将 Y 划分为 $\hat{K} + 1$ 个类。通过初始化 $\hat{K} + 1$ 个类的中心值,并对 Y 进行重新聚类处理,便可以实现。对于初始化 $\hat{K} + 1$ 个类的中心值,这里采用的方法是将 \hat{K} 个已有的类中心值中的其中一个分裂为两个类中心,同时保持其他 $\hat{K} - 1$ 个类中心值不变。例如,如果选择第 k 个类进行分裂处理,那么两个新类的中心由式(27)计算。最优的划分将产生最小描述长度 $L(Y|\hat{\theta}, \hat{\alpha})$ 。

$$\begin{aligned} \hat{\mu}_k &= \hat{\mu}_k - \hat{\sigma}_k \\ \hat{\mu}_{k+1} &= \hat{\mu}_k + \hat{\sigma}_k \end{aligned} \tag{27}$$

整个聚类处理过程的流程图如图 1 所示,整个聚类过程由 3 级循环构成。

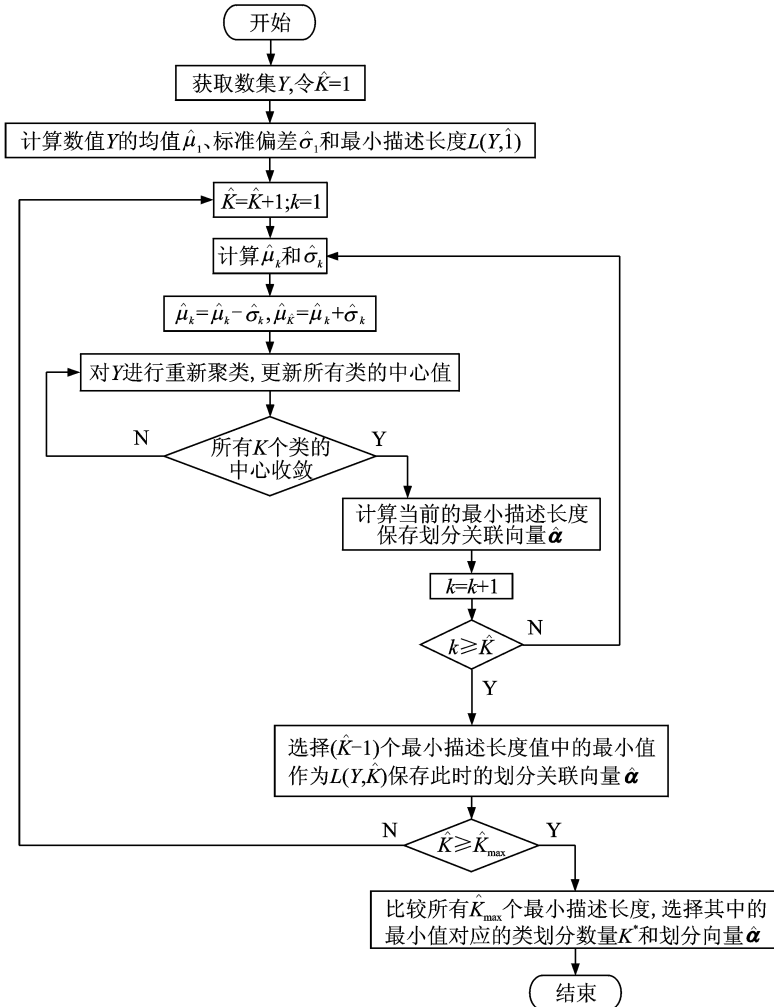


图 1 基于 MDL 原理的雷达脉冲聚类处理流程

Fig. 1 Processing flow of radar pulse clustering based on MDL criterion

5 算法运算量分析

根据图 1 所示处理流程,分析算法的运算量,将整个聚类分选流程分别根据三级循环进行运算量分析。在第 1 级循环中,选择出 $\hat{K}-1$ 个最小描述长度中的最小值,作为当前的最小描述长度 $L(Y, \hat{K})$ 。与第 3 级循环中的运算量相比,第 1 级循环内的运算量可以忽略不计。在第 2 级循环中,首先根据式(19)和(22)计算出第 k 个类的均值和标准偏差,并根据式(27)将第 k 个类的中心分裂为两个类;此外,在新的数据类中心收敛后,计算聚类收敛后的最小描述长度 $L(Y|\hat{\theta}, \hat{\alpha}_k)$ 。与第 3 级循环中的运算量相比,第 2 级循环内的运算量也可以忽略不计。在第 3 级循环中,主要运算可分为两部分。首先,根据欧氏距离,将 Y 中的所有 N 个雷达脉冲数据向量归类至距离最近的数据类中。每计算一个 M 维的雷达脉冲数据向量与一个类中心的欧氏距离包括: M 个浮点数据减法运算、 M 个浮点数据乘法运算和 $(M-1)$ 个浮点加法运算。 Y 中总共有 N 个雷达脉冲数据向量,且当前有 \hat{K} 个数据类,总共需要的运算量有

$$N\hat{K}[M \times \text{Sub} + M \times \text{Mul} + (M-1) \times \text{Add}] \quad (28)$$

式中 Sub, Mul 和 Add 分别表示单次浮点减法、乘法和加法运算。

其次,在对 Y 内的所有雷达脉冲数据一次聚类后,需要更新所有 \hat{K} 个类的中心

$$\sum_{k=1}^{\hat{K}} [M(N_k - 1) \times \text{Add} + M \times \text{Div}] = M[(N - \hat{K}) \times \text{Add} + \hat{K} \times \text{Div}] \quad (29)$$

式中:Div 表示单次浮点除法运算。

如果要确保现有的 \hat{K} 个类中心收敛,第 3 级循环需要执行 N_{itr} 次迭代聚类处理,完成第 3 级循环需要的所有运算量为

$$N_{\text{itr}} \times \{N\hat{K}[M \times \text{Sub} + M \times \text{Mul} + (M-1) \times \text{Add}] + M[(N - \hat{K}) \times \text{Add} + \hat{K} \times \text{Div}]\}$$

通常,可以假设 $M \geq 1$ 且 $N \geq \hat{K}$,对上式简化后可重新记为

$$N_{\text{itr}} \times \{N\hat{K}[M \times \text{Sub} + M \times \text{Mul} + M \times \text{Add}] + MN \times \text{Add}\} = MN N_{\text{itr}} \times [\hat{K}(\text{Sub} + \text{Mul}) + (\hat{K} + 1) \times \text{Add}] \quad (30)$$

因为第 2 级循环需要执行 $\hat{K}-1$ 次,所以,第 2,3 级循环的运算量为

$$MN N_{\text{itr}} \times [\hat{K} \times (\text{Sub} + \text{Mul}) + (\hat{K} + 1) \times \text{Add}] \times (\hat{K} - 1) = MN N_{\text{itr}} \times [\hat{K}^2 \times (\text{Add} + \text{Sub} + \text{Mul}) - \hat{K} \times (\text{Sub} + \text{Mul}) - \text{Add}] \quad (31)$$

在第 1 级循环中, $\hat{K}-1$ 从 2 递增至 \hat{K}_{max} ,完成全部聚类处理的运算量为

$$\begin{aligned} & \sum_{\hat{K}=2}^{\hat{K}_{\text{max}}} MN N_{\text{itr}} \times [\hat{K}^2 \times (\text{Add} + \text{Sub} + \text{Mul}) - \hat{K} \times (\text{Sub} + \text{Mul}) - \text{Add}] = \\ & MN N_{\text{itr}} \times \left[\sum_{\hat{K}=2}^{\hat{K}_{\text{max}}} \hat{K}^2 \times (\text{Add} + \text{Sub} + \text{Mul}) - \sum_{\hat{K}=2}^{\hat{K}_{\text{max}}} \hat{K} \times (\text{Sub} + \text{Mul}) - \sum_{\hat{K}=2}^{\hat{K}_{\text{max}}} \text{Add} \right] = \\ & MN N_{\text{itr}} \times \left[\left(\frac{\hat{K}_{\text{max}}^3}{3} + \frac{\hat{K}_{\text{max}}^2}{2} - \frac{5\hat{K}_{\text{max}}}{6} \right) \times \text{Add} + \frac{1}{3} (\hat{K}_{\text{max}}^3 - \hat{K}_{\text{max}}) \times (\text{Sub} + \text{Mul}) \right] \quad (32) \end{aligned}$$

根据上式,在数据集合 Y 中元素数量 N 确定的情况下,最终运算量与 \hat{K}_{max}^3 成正比关系;如果给定了数据集合 Y 中类的最大数量,那么运算量与数据量 N 成正比关系。

6 计算机仿真及分析

为了验证基于 MDL 准则聚类分选算法的有效性,应用 MATLAB 软件,对两类数据进行了蒙特卡罗仿真,这两类数据分别为二维高斯数据和多维雷达脉冲向量数据。

对于高斯数据的仿真,目的在于验证基于 MDL 准则聚类处理算法的有效性;对于雷达脉冲信号数据的仿真,目的在于考察针对雷达脉冲信号的预处理算法的有效性,同时考察基于 MDL 准则聚类处理算法应用于雷达脉冲信号分选课题的有效性及其分选性能。

6.1 基于二维高斯数据的仿真

对 5 个二维高斯数据类 $[X,Y]$ 进行聚类处理,其中 X 和 Y 为相互独立的一维高斯向量。各个类的均值与方差的设置情况如表 2 所示,每个类的元素数量为 500,生成的 5 个二维高斯数据类的元素分布情况如图 2 所示。

表 2 二维高斯数据类的参数列表

Table 2 Parameters of 2-D Gauss data clusters

	类 1	类 2	类 3	类 4	类 5
均值	(0,0)	(2,7)	(5,0)	(11,7)	(16,9)
方差	[1,1]	[1,1]	[2,2]	[3,3]	[2,2]

表 3 为应用 MDL 聚类处理算法对上述 5 个高斯数据类的分类处理性能,分类结果如图 3 所示。图 4 所示为基于最小描述字长度的聚类检测仿真结果,当将原始数据划分为 5 个数据类时,计算得到的最小描述长度值达到最小。

表 3 基于 MDL 准则的高斯数据聚类处理性能

Table 3 Clustering processing performance for Gauss data based on MDL criterion

算法	分选后每个类中的元素数量	分选正确率/%
MDL 聚类	515/504/484/491/506	98.278

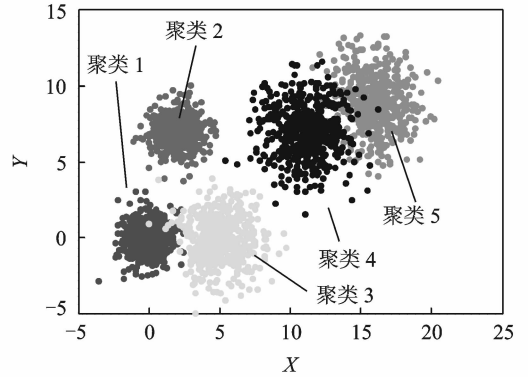


图 2 二维高斯原始数据类元素分布

Fig. 2 Distributing of 2-D original Gauss data

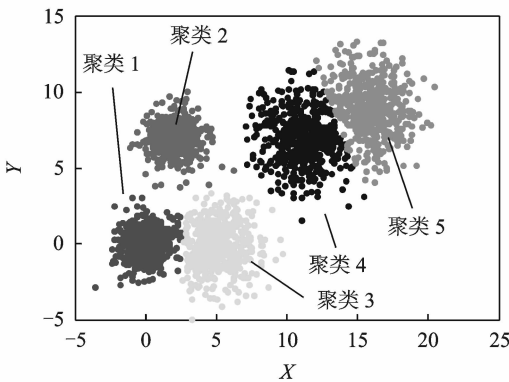


图 3 离线 MDL 聚类处理的分类结果

Fig. 3 Clustering result of offline MDL processing algorithm

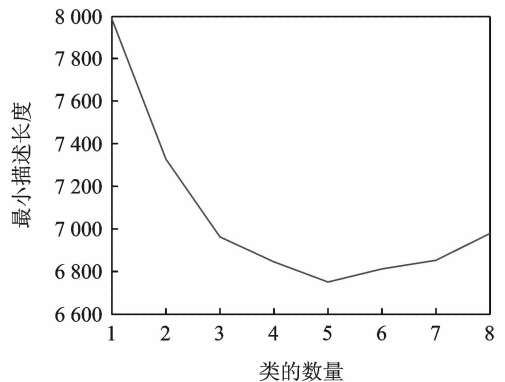


图 4 基于 MDL 聚类处理的检测结果

Fig. 4 Detection result of clustering process based on MDL criterion

6.2 基于雷达脉冲数据的仿真

在仿真过程中,为了尽可能反映外场信号环境的真实情况,根据表 4 的参数设置生成雷达脉冲信号数据。其中每个雷达辐射源产生 20 个脉冲信号,截获采样时间为 $0.1 \mu s$,采样率 1 GHz ,SNR 为 20 dB 。图 5~9 给出了 6 部雷达的幅度和相位特征。

表 4 待分选雷达辐射源的参数

Table 4 Parameters of radar emitters for deinterleaving

序号	$\phi_{a_n}(t)$	$a_{a_n}(t)$	f_n/MHz	η_n	ψ_n
1	图 5 信号 1	图 5 信号 1	150	1.5	$[-\pi, \pi)$
2	图 5 信号 2	图 5 信号 2	180, 200	2.0	
3	BPSK	1	250	2.5	π
4	QPSK	1	300	3.0	$\pi/4$
5	LFM1	1	350	3.5	$[-\pi, \pi)$
6	LMF2	1	400	4.0	

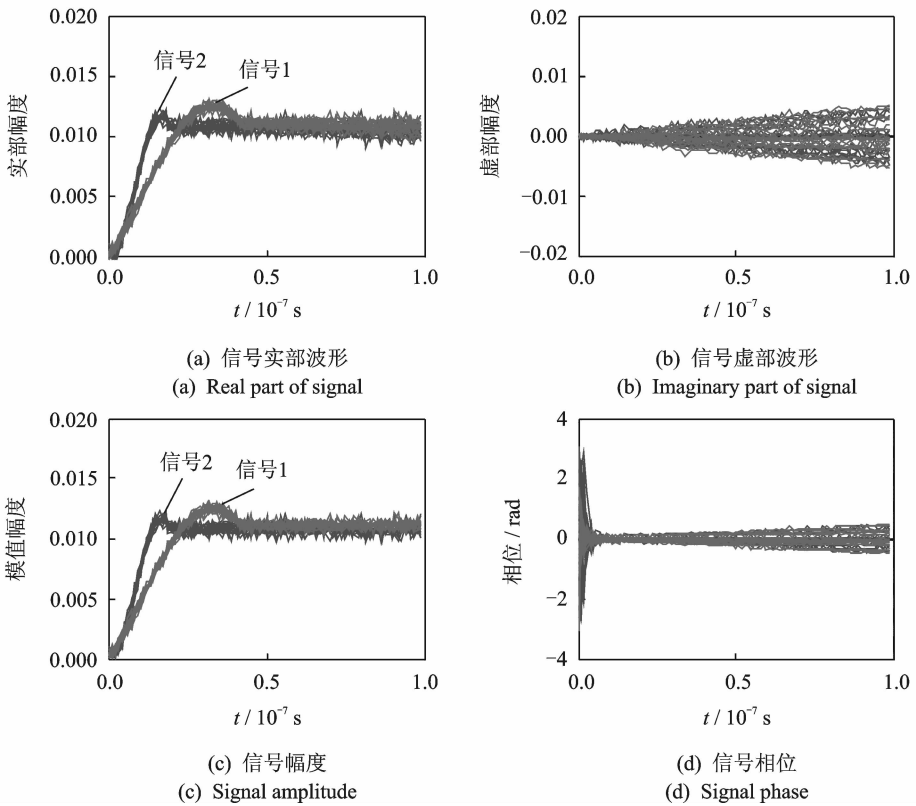


图 5 雷达 1 和 2 脉冲信号特征

Fig. 5 Pulse signal character of Radar 1 and Radar 2

表 5 为针对上述 6 部雷达的脉冲信号分选性能, 雷达脉冲信号得到了很好的分选, 全部脉冲信号都被正确地分选出来, 分选的正确率达 100%。

表 5 基于 MDL 准则聚类算法的雷达脉冲信号分选性能

Table 5 Deinterleaving performance for radar pulse signal based on MDL criterion

算法	分选后每个类中的脉冲数量	分选正确率/%
MDL 聚类	20/20/20/20/20/20	100

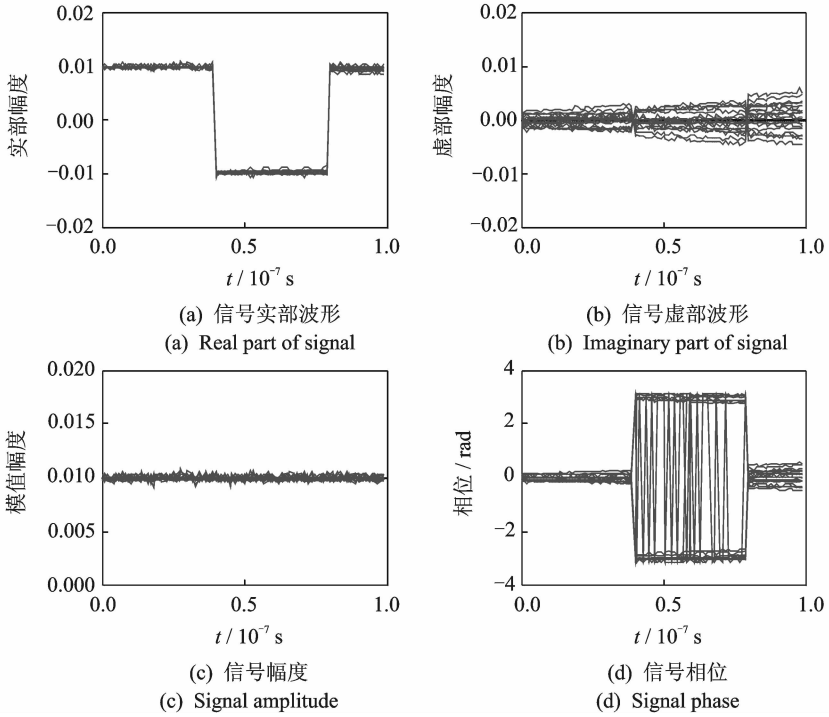


图6 雷达3的信号特征

Fig. 6 Pulse signal character of Radar 3

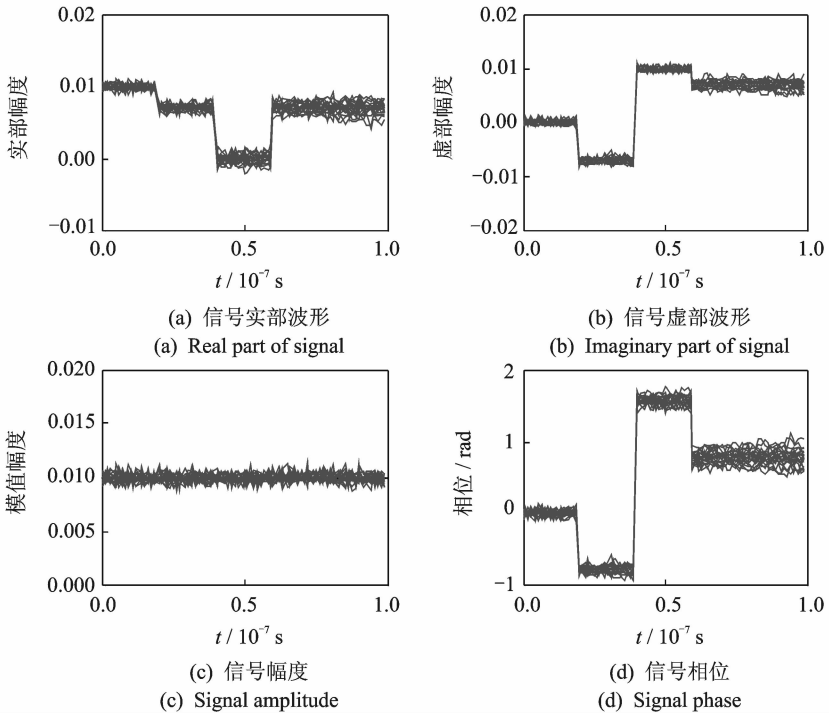


图7 雷达4的信号特征

Fig. 7 Pulse signal character of Radar 4

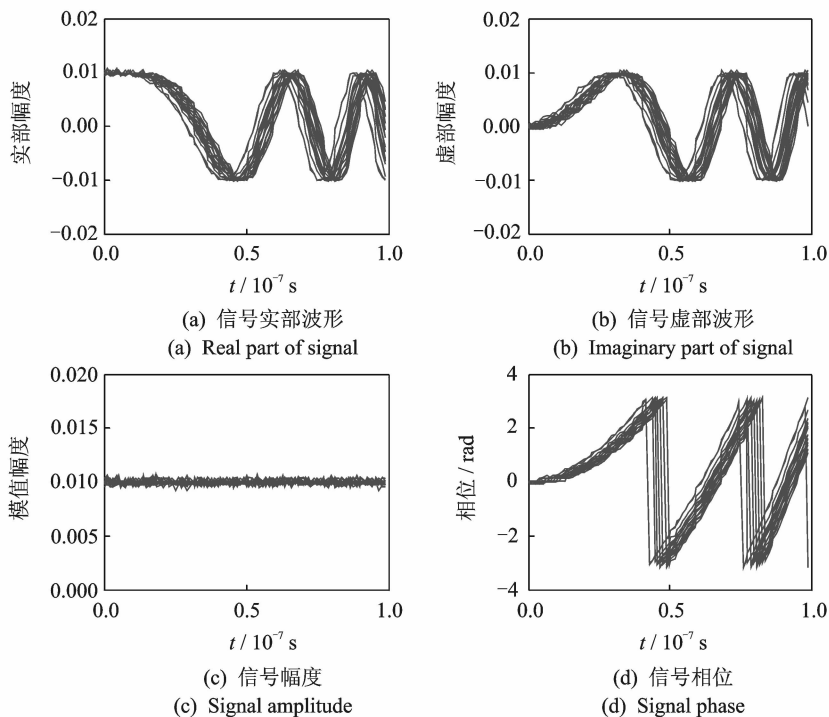


图 8 雷达 5 的信号特征

Fig. 8 Pulse signal character of Radar 5

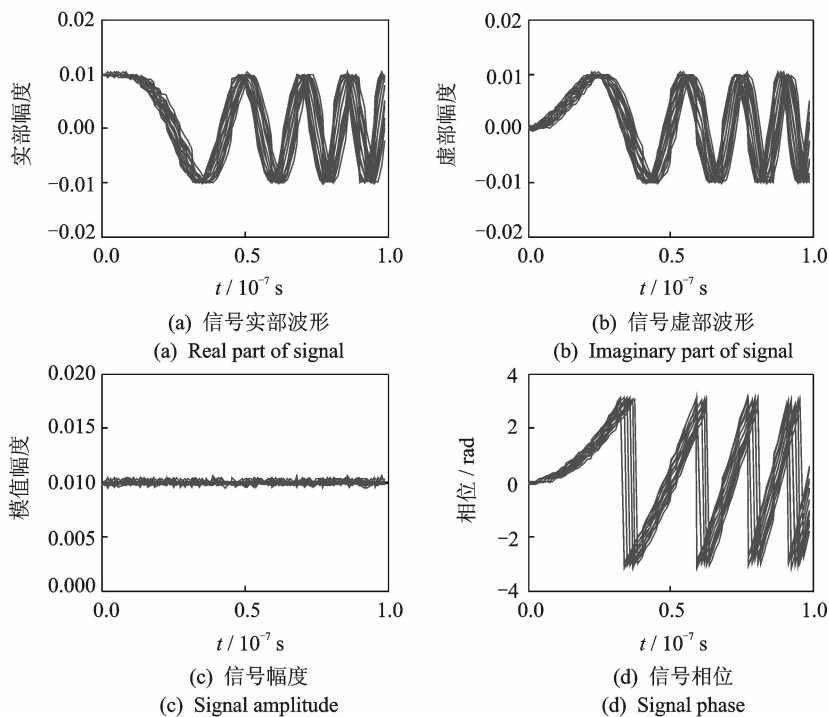


图 9 雷达 6 的信号特征

Fig. 9 Pulse signal character of Radar 6

7 结束语

针对基于 MDL 准则的雷达脉冲信号分选算法,本文开展了原理性研究:与传统的基于脉冲描述字的信号分选算法不同,该算法利用雷达脉冲信号的脉内调制特征差异,实现雷达信号分选;详细论述了算法原理、设计了处理流程、并分析了算法运算量。后续将就以下方面开展深入研究:(1) 研究不同 SNR 情况下的算法性能;(2) 本算法运算量与 \hat{K}_{\max}^3 成正比关系,与一般算法比较,运算量较大,需要根据工程实现,研究并开发实时在线算法;(3) 综合本算法与传统算法的优势,研究基于 PDW、脉间信息和脉内信息的全息信号分选算法。

参考文献:

- [1] 赵国庆. 雷达对抗原理[M]. 西安:西安电子科技大学出版社,1999.
Zhao Guoqing. Radar electronic principle[M]. Xi'an: Xidian University Press, 1999.
- [2] 李合生,韩宇,蔡英武. 雷达信号分选关键技术研究综述[J]. 系统工程与电子技术,2005,27(12):2035-2040.
Li Hesheng, Han Yu, Cai Yingwu. Overview of the crucial technology research for radar signal sorting[J]. Systems Engineering and Electronics, 2005,27(12):2035-2040.
- [3] 杨学永,宋国栋,钱轶,等. 现代雷达信号分选跟踪的几种方法[J]. 现代雷达,2014,36(3):43-48.
Yang Xueyong, Song Guodong, Qian Yi, et al. Several methods of radar signal sorting and tracking[J]. Modern Radar, 2014,36(3):43-48.
- [4] 韩俊,何明浩,朱元清. 基于多参数的雷达辐射源分选新方法[J]. 数据采集与处理,2009,24(1):91-94.
Han Jun, He Minghao, Zhu Yuanqing. Sorting of radar emitter based on multi-parameters[J]. Journal of Data Acquisition and Processing, 2009,24(1):91-94.
- [5] 张西托,饶伟,杨泽刚. 平面变换技术脉冲分选自动实现方法[J]. 数据采集与处理,2012,27(4):495-500.
Zhang Xituo, Rao Wei, Yang Zegang. Automatic fulfillment of deinterleaving pulses based on plane transformation[J]. Journal of Data Acquisition and Processing, 2012,27(4):495-500.
- [6] 杨文华,高梅国. 基于平面变换技术的脉冲信号分选[J]. 北京理工大学学报,2005,25(2):151-154.
Yang Wenhua, Gao Meiguo. The deinterleaving of pulse signal based on plane transformation[J]. Journal of Beijing Institute of Technology, 2005,25(2):151-154.
- [7] Douglas N. Special purpose correlation functions for improved signal detection and parameter estimation[C]//Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 93). [S. l.]: IEEE, 1993:73-76.
- [8] Mardia H K. New techniques for the deinterleaving of repetitive sequences[C]//IEE Proceedings of Radar and Signal Processing. [S. l.]: IEEE, 1989:149-154.
- [9] Milojevic D J, Popovic B M. Improved algorithm for the deinterleaving of radar pulse[J]. IEE Proceedings of Radar and Signal Processing, 1992,139(1):98-104.
- [10] 黄桂根,傅有光,武月婷. 一种改进的基于 DTOA 统计的信号分选算法[J]. 数据采集与处理,2011,26(4):430-435.
Huang Guigen, Fu Youguang, Wu Yueting. Improved radar signal deinterleaving algorithm based on DTOA Histogram[J]. Journal of Data Acquisition and Processing, 2011,26(4):430-435.
- [11] 柴娟芳,司锡才,马晓东. 基于 PRI 谱的双门限雷达信号分选算法及其硬件平台设计[J]. 数据采集与处理,2009,24(1):38-43.
Chai Juanfang, Si Xicai, Ma Xiaodong. Double-threshold PRI sorting algorithm based on PRI spectrum and its sorting platform design[J]. Journal of Data Acquisition and Processing, 2009,24(1):38-43.
- [12] 李英达,肖立志. 一种脉冲重复间隔复杂调制雷达信号分选方法[J]. 电子与信息学报,2013,35(10):2493-2497.
Li Yingda, Xiao Lizhi. A method of signal sorting for radar signal of pulse repetition interval complex modulated[J]. Journal of Electronics & Information Technology, 2013,35(10):2493-2497.

- [13] 张治海, 秦开兵, 张元发. 基于脉内特征的雷达信号分选新方法[J]. 现代防御技术, 2009, 37(2):104-107.
Zhang Zhihai, Qin Kaibing, Zhang Yuanfa. New method for sorting radar signal based on in-pulse characteristics[J]. Modern Defence Technology, 2009, 37(2):104-107.
- [14] 陈爱萍, 范媛媛. MDL 理论的多属性值域划分方法[J]. 计算机工程与应用, 2013, 49(1):167-171.
Chen Aiping, Fan Yuanyuan. Value domain partition method of multiple attributes based on MDL principle[J]. Computer Engineering and Application, 2013, 49(1):167-171.
- [15] 王衍学, 向家伟, 蒋占四, 等. 基于改进标准化最大似然估计的最小描述长度降噪方法[J]. 振动与冲击, 2014, 33(1):137-140.
Wang Yanxue, Xiang Jiawei, Jiang Zhansi, et al. Minimum description length denoising based on improved normalized maximum likelihood[J]. Journal of Vibration and Shock, 2014, 33(1):137-140.
- [16] 麦超云, 孙进平, 崔如心, 等. 压缩感知合成孔径雷达射频干扰抑制处理[J]. 北京航空航天大学学报, 2014, 40(1):59-62.
Mai Chaoyun, Sun Jinping, Cui Ruxin, et al. RFI suppression processing for compressive sensing based SAR imaging[J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(1):59-62.
- [17] 蒋建国, 宣浩, 郝世杰, 等. 最小描述长度优化下的医学图像统计形状建模[J]. 中国图象图形学报, 2011, 16(5):879-885.
Jiang Jianguo, Xuan Hao, Hao Shijie, et al. Statistical shape modeling based on minimum description length optimization in medical images[J]. Journal of Image and Graphics, 2011, 16(5):879-885.
- [18] 徐向华, 朱杰, 郭强. 语音识别中基于最小描述长度准则的决策树动态剪枝算法[J]. 声学学报, 2006, 31(4):370-376.
Xu Xianghua, Zhu Jie, Guo Qiang. Decision tree dynamic pruning method based on minimum description length in speech recognition[J]. Acta Acustica, 2006, 31(4):370-376.

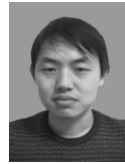
作者简介:



黄桂根 (1978-), 男, 博士, 高级工程师, 研究方向: 雷达侦察信号处理、雷达反干扰技术, E-mail: hguigen@163.com。



孟兵 (1976-), 男, 高级工程师, 研究方向: 雷达侦察信号处理、雷达反干扰技术。



李品 (1982-), 男, 博士, 高级工程师, 研究方向: 雷达侦察信号处理、雷达信号处理跟踪检测一体化、目标分类识别。