



基于虚警可控梯度提升树的海面小目标检测

摘要

高维特征检测是提升海面小目标探测性能的一种有效途径,其主要难点在于高维空间分类器设计.本文提出一种基于虚警可控梯度提升树(Gradient Boosting Decision Tree, GBDT)的特征检测方法.首先,从一维长时观测向量中,提取时域、频域、时频域等多个特征,构建高维特征向量,从而将检测问题转换为二分类问题.其次,通过仿真含目标回波,解决两类训练样本非均衡的问题.然后,引入GBDT算法,将高维特征向量凝聚为一维概率预测值,并以预测值作为检测统计量,解决二分类器难以控制虚警的问题.最后,采用IPIX实测数据验证,结果表明:本文所提的检测器充分利用了高维特征的全部信息,性能平均提升13%以上.

关键词

海杂波;小目标检测;梯度提升树;高维特征

中图分类号 TN959.1

文献标志码 A

收稿日期 2021-07-09

资助项目 国家自然科学基金(61901224)

作者简介

刘安邦,男,主要研究方向为海面小目标检测. shilizhi_414@outlook.com

施赛楠(通信作者),女,博士,讲师,主要研究方向为雷达信号处理、微弱目标检测. snshi@nuist.edu.cn

1 南京信息工程大学 江苏省气象探测与信息处理重点实验室,南京,210044

2 南京船舶雷达研究所,南京,210003

0 引言

海面小目标(浮冰、小船、蛙人、飞机残骸等)已成为海洋雷达探测的主要对象^[1-3],海面小目标探测在海上救援、海洋防卫、海上突击等方面具有重要意义.通常,海面小目标的回波较弱,往往被淹没在强海杂波背景下.由于小目标的信杂比(Signal-to-Clutter Ratio, SCR)低,传统的检测器会出现检测概率低、虚警点多等问题.

在雷达信号处理中,长时累积技术是提升小目标 SCR 的有效途径.在秒级长时观测下,海杂波具有复杂的空时变特性^[4-5].同时,小目标的功率不再保持恒定,具有起伏特性.从海杂波幅度非线性建模出发,Hu 等^[6]提出了基于 Hurst 指数的检测器,利用海杂波和目标回波在多尺度分形特性上的差异性,实现了单特征检测.通过对时域数据进行经验模态分解、重构和去噪等预处理,张林等^[7]提出了基于变换域分形特性的检测器.从多特征联合角度出发,Shui 等^[8]提出了基于三特征的检测器,联合一个时域幅度特征和两个频域多普勒几何特征,在三维特征空间中,该检测器采用快速凸包学习算法,获得虚警可控的判决区域.而 Shi 等^[9]提出了基于时频三特征的检测器,充分挖掘了目标和海杂波在时频域的精细化几何特征,包括频谱的波动性和分散性.考虑不同特征对不同数据的敏感性,联合使用更多互补特征成为进一步提高检测性能的有效途径.

在高维特征空间中,目标检测的主要难点是设计虚警可控的分类器.目前,主要有两条途径可以实现高维分类器设计.第一条途径是特征压缩,将高维分类器简化为低维分类器.通过沿着一类最大方差方向投影到低维空间中,Gu^[10]提出了基于主成分分析(Principle Component Analysis, PCA)的三特征检测器.基于两类最大距离的压缩,Shui 等^[11]提出了基于 B 氏距离的特征压缩检测器.但是,特征压缩势必会带来部分性能损失.第二条途径是设计高维分类器.郭子薰等^[12]提出了改进的 KNN 算法,实现高维判决区域划分.通过调节参数实现虚警控制,Zhou 等^[13]提出了基于决策树的特征检测器.这些检测器保证了高维信息量,但高维分类器需要遍历搜索参数以满足给定虚警率,导致计算量大且控制不准确.因此,亟需设计虚警可控且计算复杂度低的高维分类器.

为了满足实际雷达快速检测的需求,本文引入梯度提升树(Gradient Boosting Decision Tree, GBDT)^[14-15]算法,快速获得高维域虚

警可控的二分类器.GBDT 分类器集成了多个弱分类器,具有显著的泛化性能以及更高的分类准确度.首先,将从时域、频域、时频域提取的多维特征作为 GBDT 分类器的输入,获得属于一类的概率预测值.然后,将之作为检测统计量,获得虚警可控的判决门限.最后,基于实测数据的实验结果表明:所提出的检测器能快速获得判决门限,实现海面小目标检测性能提升的目标.

1 高维特征域二分类目标检测

1.1 二元假设检验

假设雷达接收到某个距离单元的 N 个连续脉冲,构成观测向量 $\mathbf{z} = [z(1), z(2), \dots, z(N)]^T$. 若观测向量只包含杂波,则称 H_0 假设;若观测向量由目标回波和杂波组成,则称 H_1 假设.因此,传统的雷达目标检测问题归结为二元假设检验:

$$\begin{cases} H_0: \mathbf{z} = \mathbf{c}, \mathbf{z}_k = \mathbf{c}_k, k = 1, 2, \dots, K; \\ H_1: \mathbf{z} = \mathbf{s} + \mathbf{c}, \mathbf{z}_k = \mathbf{c}_k, k = 1, 2, \dots, K. \end{cases} \quad (1)$$

其中: \mathbf{z} 和 \mathbf{z}_k 分别表示待检测距离单元 (Cell Under Test, CUT) 以及周围 K 个参考距离单元的回波向量,为 CUT 提供杂波信息; \mathbf{c} 表示海杂波; \mathbf{s} 表示目标回波; \mathbf{c}_k 表示第 k 个距离单元为海杂波.

事实上,检测的本质就是判断观测向量属于哪种假设.通常,传统的检测方法基于单个域特性设计统计量,存在小目标检测概率低、虚警点高的问题.基于多特征联合的检测方法主要是利用海杂波和含目标回波在多个域的差异性,并将这些差异性转变为特征,实现高维特征域目标检测.

1.2 高维特征检测问题

为了充分利用观测向量的信息,本文从时域、频域、时频域 3 个域提取 7 个特征,分别记为 $\delta_1, \delta_2, \dots, \delta_7$, 如表 1 所示.在时域,提取 Hurst 指数^[6]和相对平均幅度^[8],分别反映幅度的几何粗糙度和能量特性.在频域,提取相对多普勒峰高^[8]和相对向量熵^[8],分别反映海杂波和含目标回波在频谱宽度、聚集度方面的差异性.在时频域,提取脊能量^[9]、连通区域数目^[9]和最大连通区域尺寸^[9],进一步精细化动态频谱的几何特性.此外,表 1 给出了 2 种假设在不同特征下相对取值的大小,这与海杂波和含目标回波差异性有关.

考虑到不同特征的取值范围不同,对每个特征作归一化预处理,保证所有特征投影到相同的尺度上.对于第 i 个特征,假设获得 H_0 假设下 Q 个样本,

刘安邦,等.基于虚警可控梯度提升树的海面小目标检测.

表 1 7 个特征描述

Table 1 Properties of seven features

符号	描述特性	H_0 假设	H_1 假设
δ_1	幅度粗糙度	小	大
δ_2	时域能量比	小	大
δ_3	频谱几何特性	小	大
δ_4	频谱信息度	大	小
δ_5	时频域能量比	小	大
δ_6	能量聚集度	大	小
δ_7	能量分散度	小	大

归一化定义为

$$\begin{aligned} \bar{\delta}_i &= \frac{\delta_i - \mu_i}{\sqrt{\frac{1}{Q-1} \sum_{q=1}^Q (\delta_i(q) | H_0) - \mu_i)^2}}, \\ \mu_i &= \frac{1}{Q} \sum_{q=1}^Q \delta_i(q) | H_0, i = 1, 2, \dots, 7, \end{aligned} \quad (2)$$

其中, μ_i 表示第 i 个特征的均值.

下面,构建高维特征空间.为了保证信息的充分运用,将每个特征作为一个维度,构建七维特征空间.此时,观测向量转换为归一化特征向量为

$$\bar{\boldsymbol{\delta}}(\mathbf{z}) = [\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_7]^T. \quad (3)$$

在高维特征空间中,式(1)中的目标检测问题转化为二分类检测问题,即:

$$\begin{cases} H_0: \bar{\boldsymbol{\delta}}(\mathbf{z} = \mathbf{c}) = [\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_7]^T; \\ H_1: \bar{\boldsymbol{\delta}}(\mathbf{z} = \mathbf{s} + \mathbf{c}) = [\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_7]^T. \end{cases} \quad (4)$$

第一类为 H_0 假设,观测向量只含杂波;第二类为 H_1 假设,观测向量含有目标回波.因此,高维特征域中分类问题的核心在于设计一个具有虚警可控的二分类器.

2 基于虚警可控 GBDT 的检测器

2.1 检测器结构

基于虚警可控 GBDT (False-Alarm-Controllable GBDT, FAC-GBDT) 检测器的流程如图 1 所示.整个检测结构包括在线检测分支和离线训练分支两部分.在检测分支,首先,CUT 的观测向量通过时域、频域、时频域提取多个特征,并归一化预处理,去除不同尺度的影响;其次,构建高维特征向量,作为 GBDT 分类器的输入;然后,将 GBDT 分类器输出的概率预测值,作为统计量;最后,统计量和门限进行比较,获得最终的判决结果.

在训练分支,两类均衡训练样本用于训练学习,获得 GBDT 模型的最优参数和给定虚警率下的判决门限.通常,由于小目标的空间稀疏性以及军事保密

性,海杂波数据可大量获得但含目标回波数据极其稀少.为了提升二分类器的分类准确度,两类的训练样本数目应当均衡.因此,在 H_1 假设下,观测回波建模^[10,12]为

$$z(n | H_1) = c(n) + \rho(n) 10^{R_{sc}/20} \times \exp\left(j \left(\frac{4\pi}{\lambda} \left(v_0 + a \frac{nT_r}{N} \right) nT_r + \phi_0 \right) \right), n = 1, 2, \dots, N, \quad (5)$$

其中: c 是实测海杂波,来自 H_0 假设下的训练数据; ρ 是均值为 1 的相关序列,描述目标起伏特性; R_{sc} 为信杂比(SCR)的值; λ 是雷达波长; T_r 是雷达的脉冲重复周期; $|a| \leq 2$ 是目标的加速度; $\phi_0 \in [-\pi, \pi]$ 是初始相位; $|v_0| \leq 5$ 是目标的初始速度.

在实际工作中,海洋雷达更关注的是在线检测时间.由于 GBDT 模型参数可离线训练获得,因而不影响雷达实时检测.此外,当探测环境发生变化时,可启动离线训练分支.通过收集当前海杂波数据,重新学习模型的最优参数.FAC-GBDT 检测器具有的这种动态更新模式,可以实现对海洋环境的恒虚警特性.

2.2 虚警可控的 GBDT 分类器

在高维特征空间中,K 近邻、决策树、支持向量机等机器学习方法都可以实现二分类.但是,上述分类器自身无法控制虚警率,只能通过建立某个参数和虚警率的关系,进行全局搜索获得给定虚警率下的参数值^[12-13].这类方法明显计算量大,且参数精度设置对虚警精准度有较大的影响.因此,本文引入 GBDT 算法,无需搜索参数且能精准控制虚警率.

图 2 给出了 GBDT 模型的结构^[14-15],以 Cart 回

归树作为基本的弱分类器,通过多个弱分类器级联,获得鲁棒性更高的分类结果.GBDT 算法核心是一种贪婪最优策略,通过不断减小训练过程产生的残差来达到准确分类.在训练弱分类器的过程中,准则是最小化损失函数.这里,定义损失函数为

$$L(y, f(\mathbf{x})) = y \log(1 + e^{-f(\mathbf{x})}) + (1 - y) (f(\mathbf{x}) + \log(1 + e^{-f(\mathbf{x})})), \quad (6)$$

其中, $y \in \{0, 1\}$ 表示 \mathbf{x} 的真实标签值, $f(\mathbf{x})$ 是 y 的预测值.

在 GBDT 分类器训练时,假设 H_0 假设下有 Q 个向量样本,标签为 0; H_1 假设下有 Q 个向量样本,标签为 1.因此,在 7D 特征空间中,输入训练集 Ω 为

$$\Omega = \{(x_1, y_1), (x_2, y_2), \dots, (x_{2Q}, y_{2Q})\}, \quad \mathbf{x}_q \in \mathbf{R}^7, y_q \in \{0, 1\}. \quad (7)$$

下面,给出 GBDT 算法的具体步骤.

步骤 1.初始化.根据训练集 Ω ,初始预测值为

$$f_0(\mathbf{x}) = \log \left(\frac{\sum_{q=1}^{2Q} y_q}{\sum_{q=1}^{2Q} (1 - y_q)} \right), \mathbf{x} \in \Omega. \quad (8)$$

步骤 2.生成第 m 棵 Cart 回归树.首先,对于第 q 个样本,计算损失函数的负梯度,即残差值

$$g_{m,q} = - \left[\frac{\partial L(y_q, f_{m-1}(\mathbf{x}_q))}{\partial f_{m-1}(\mathbf{x}_q)} \right] = y_q - \frac{1}{1 + e^{-f_{m-1}(\mathbf{x}_q)}}. \quad (9)$$

然后,通过拟合残差值 $g_{m,q}$,得到第 m 棵树的叶子节点区域为 $R_{m,j}, j = 1, 2, \dots, J_m$,其中, J_m 是第 m 棵树叶子节点的个数.通过最小化损失函数,计算第 j 个叶子节点的最佳拟合值 $c_{m,j}$:

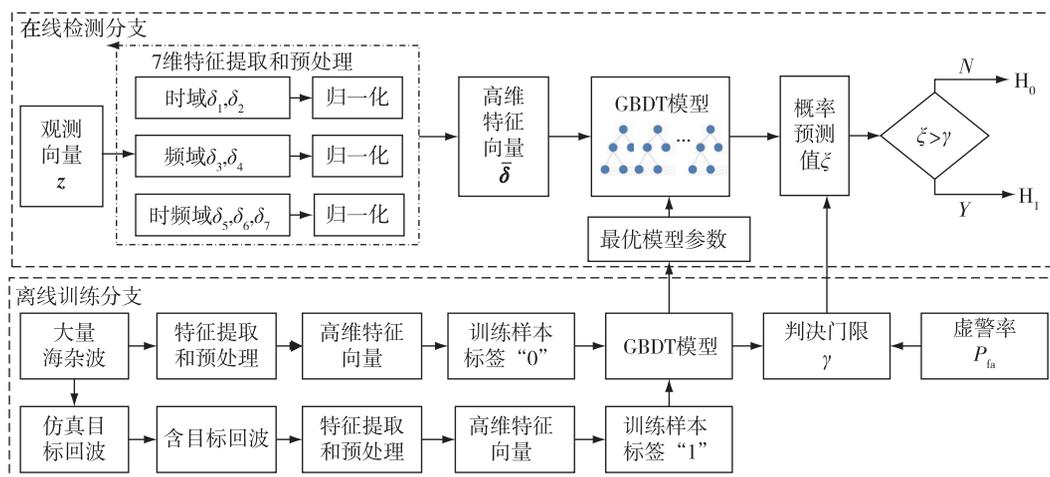


图 1 FAC-GBDT 检测器的流程

Fig. 1 Flowchart of FAC-GBDT detector

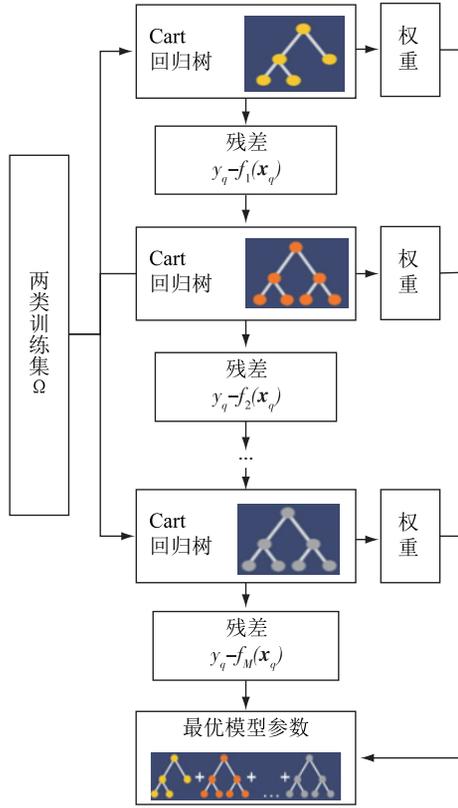


图2 GBDT模型结构

Fig. 2 GBDT model structure

$$c_{m,j} = \frac{\sum_{x_q \in R_{m,j}} g_{m,q}}{\sum_{x_q \in R_{m,j}} (y_q - g_{m,q})(1 - y_q + g_{m,q})}. \quad (10)$$

最后,更新预测值为

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \nu \sum_{j=1}^{J_m} c_{m,j} I(\mathbf{x} \in R_{m,j}), \quad (11)$$

其中, $\nu \in [0, 1]$ 是学习率,用于控制过拟合度.

步骤3.更新迭代.重复步骤2,最终生成 M 棵 Cart 回归树.第 M 棵树的输出,即最终的预测值为

$$f_M(\mathbf{x}) = f_0(\mathbf{x}) + \nu \sum_{m=1}^M \sum_{j=1}^{J_m} c_{m,j} I(\mathbf{x} \in R_{m,j}). \quad (12)$$

至此,可以获得 GBDT 的最优模型参数.

为了获得虚警可控的二分类器,将输入 \mathbf{x} 预测为正样本的概率

$$P(Y = 1 | \mathbf{x}) = \frac{1}{1 + e^{-f_M(\mathbf{x})}} \quad (13)$$

作为统计量 ξ .借助蒙特卡罗实验方法^[8-13],将 H_0 假设下 Q 个样本带入具有最优模型参数的 GBDT 分类器中,按照式(13)计算获得 Q 个统计量,记为 $\xi_1, \xi_2, \dots, \xi_Q$.并且,这些统计量按照值从大到小进行排序.那么,在给定虚警率 P_{fa} 下,判决门限为

刘安邦,等.基于虚警可控梯度提升树的海面小目标检测.

$$\gamma = \xi_{[Q \times P_{fa}]}, \quad (14)$$

其中, $[\]$ 表示取整数.

图3演示了虚警可控门限的获取过程,设置 $P_{fa} = 10^{-3}$,则 11 231 个海杂波样本中存在 11 个虚警点.但是,传统分类器输出以 $\gamma = 0.5$ 为分界线(即黑色线),存在 39 个虚警点,虚警率为 0.003.这就意味着门限 $\gamma = 0.5$ 无法保证分类器获得设定的虚警率.因此,调整门限位置至红色直线位置,精确保证虚警点个数为 11 个.因此,通过调整判决门限,达到设定的虚警率,即虚警可控,实现检测器的恒虚警特性.

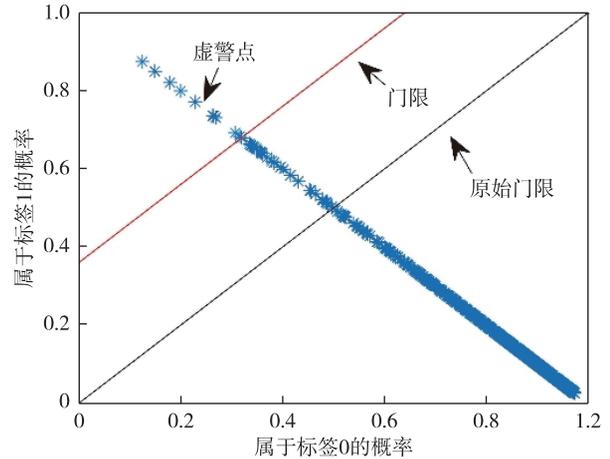


图3 虚警可控判决门限的演示

Fig. 3 Demonstration of false-alarm-controllable decision threshold

3 实验结果与性能分析

实测数据来自 1993 年采集的 IPIX 雷达目标数据库 (<http://soma.mcmaster.ca//ipix.php>, 2012. 10. 11),该数据库是目前公认的小目标测试数据库.实验采用 10 组数据,每组数据包含 14 个距离单元,2¹⁷个脉冲数.雷达工作频率为 9.3 GHz,工作在驻留模式,脉冲重复频率 $f_r = 1\ 000$ Hz,距离分辨率 30 m.测试目标为被金属丝包裹的小球,直径为 1 m,可视为漂浮低速运动小目标.每组数据涉及多种海洋环境且 SCR 起伏较大,具体参照文献[8-10].

3.1 GBDT 分类器性能分析

在高维特征空间中,GBDT 分类器的性能影响着后续的检测结果.因此,有必要对分类器性能进行分析.图4演示了二维特征空间中 GBDT 分类器的区域划分过程.对比第 1 棵 Cart 回归树和第 15 棵 Cart 回归树的结果,判决区域明显具有较好的划分.随着 Cart 回归树的增加,判决区域不断地在优化和更新,

验证了集成多个弱分类器的性能优势.此外,总的判决区域为非线性边界,适用于非线性的两类样本分类情况.

图5讨论了Cart回归树数目 M 和回归树的最大深度 D 对检测概率的影响.在图5a中,随着 M 的增大,检测概率上升,这体现了弱检测器集成的优势.但是,这种性能优势在 $M>12$ 以后,检测概率趋于稳定.在图5b中,当回归树的深度 $D<5$ 时,检测性能波动性大.因此,综合考虑计算量和性能,设置参数 $M=15, D=7$.

3.2 检测性能对比

实验中,设定 $P_{fa}=10^{-3}$,脉冲累积数 $N=512$ (即观测时间0.512 s).由于10组数据都含有目标,因而对每组不同极化下数据进行独立检测.在训练分支中,从纯杂波单元和按式(5)产生的含目标回波中,获取两类均衡训练样本,构建训练集 Ω .在检测分支中,从目标单元中提取特征向量,计算检测概率.图6中给出了4种检测器在10组数据下的检测结果,横轴表示数据名称.可以发现,检测概率最低的是基于Hurst指数的检测器^[6],HH、HV、VH、VV极化下平

均检测概率分别为0.248、0.351、0.351、0.290.这意味着该检测器无法探测到小目标,主要原因在于只利用了幅度信息,且观测时间未达到几秒以上.基于三特征的检测器^[8]联合了时域和频域的特性,4种极化下的平均检测概率分别为0.534、0.622、0.619、0.518,性能得到大幅度提升.对于基于时频三特征的检测器^[9],4种极化下的平均检测概率分别为0.698、0.699、0.698、0.614.相对于三特征的检测器^[8]性能平均提升18%,特别是后5组数据,性能提升明显.对于提出的FAC-GBDT检测器,4种极化下的平均检测概率为0.774、0.777、0.789、0.733,相对于时频三特征的检测器^[9]性能提升13%.此外,可以明显看到,FAC-GBDT检测器的检测概率在不同数据、不同极化下都具有最佳的性能.这主要源于检测器采用了7个特征,充分利用了多个域的信息,而且具有较好的稳健性能.

下面,讨论观测时间对检测性能的影响.表2列出了 $N=512$ 和 $N=1024$ 两种情况下的平均检测概率,对应观测时间为0.512 s和1.024 s.从表2中,得到以下3个结论:

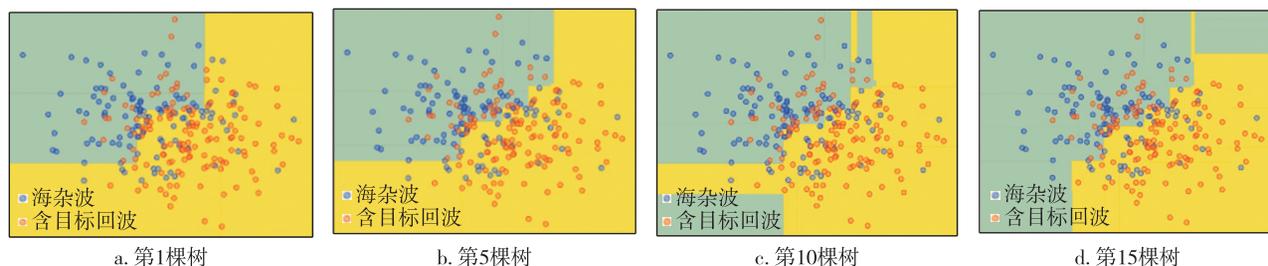


图4 GBDT二分类区域划分过程

Fig. 4 Division process of dichotomy region in GBDT

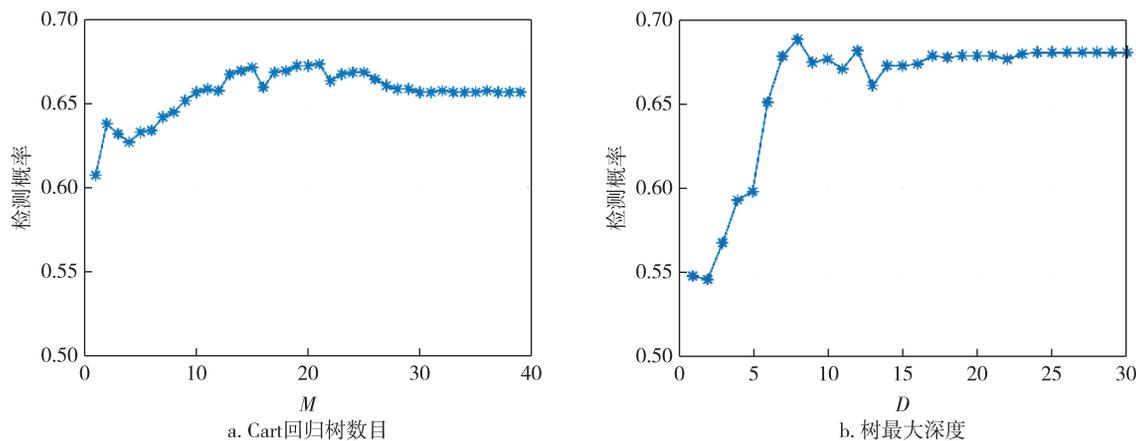


图5 Cart回归树参数对检测性能的影响

Fig. 5 Parameter influence of Cart regression tree on detection performance

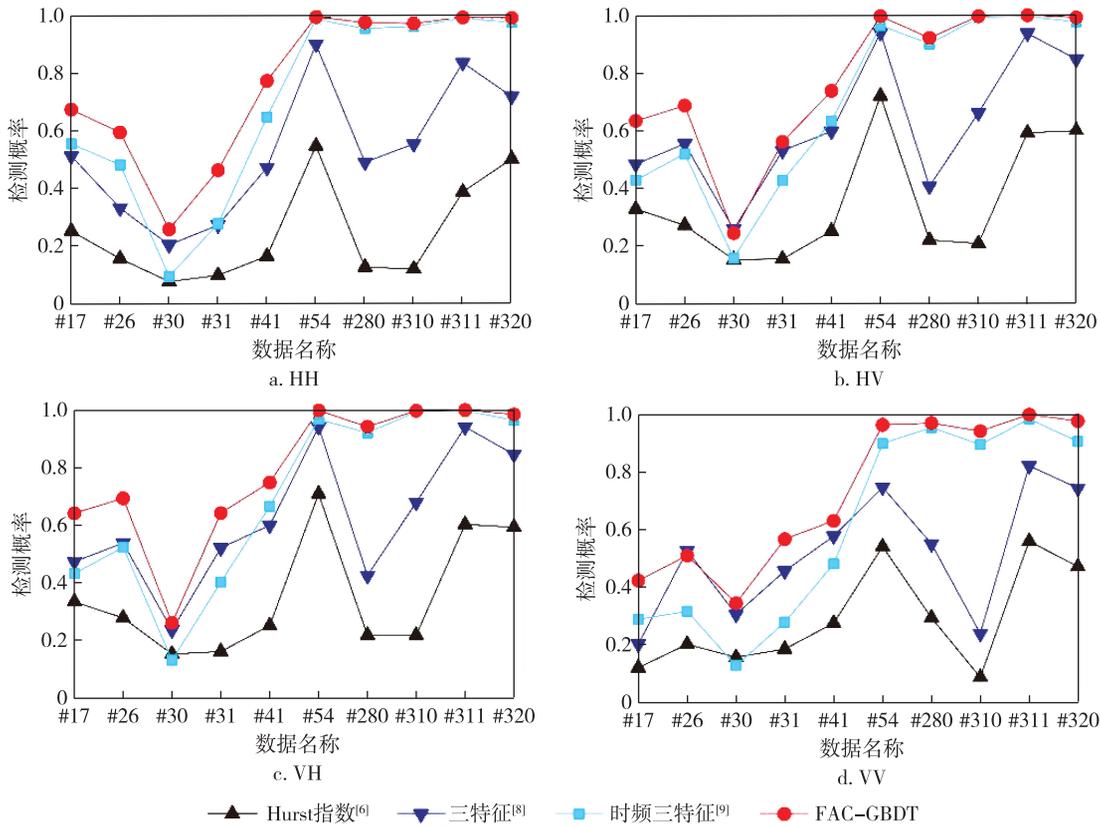


图6 4种检测器的性能对比($N=512, P_{fa}=10^{-3}$)

Fig. 6 Performance comparison between four detectors

1) 所有检测器在更长的观测时间下,检测性能均有所提升.这验证了长时累积是提升小目标的检测性能的有效途径.

2) FAC-GBDT 检测器具有最优的检测结果,这是因为联合了时域、频域以及时频域的7维特征,能够充分利用海杂波和含目标回波在多域的差异性.

表2 不同观测时间下4种检测器性能对比

Table 2 Performances of four detectors under different observation time

检测器	观测时间/s	HH	HV	VH	VV
Hurst 指数检测器 ^[6]	0.512	0.248	0.351	0.351	0.290
	1.024	0.324	0.469	0.469	0.396
三特征检测器 ^[8]	0.512	0.534	0.622	0.619	0.518
	1.024	0.588	0.674	0.667	0.542
时频三特征检测器 ^[9]	0.512	0.698	0.699	0.698	0.614
	1.024	0.757	0.771	0.756	0.702
FAC-GBDT 检测器	0.512	0.774	0.777	0.789	0.733
	1.024	0.843	0.858	0.850	0.817

3) 当观测时间增加1倍后, FAC-GBDT 检测器没有和其他3个检测器性能提升多.这是因为后5组数

据的检测存在“天花板效应”.在观测时间为0.512s时, FAC-GBDT 检测器的检测概率已经接近于1,因此性能提升空间有限.

4 结论

本文提出一种高维特征域 FAC-GBDT 检测器,能充分利用时域、频域、时频域多个域的特征信息.通过将GBDT分类器的一类概率值作为统计量,能快速获得判决门限,解决了高维分类器虚警率难以控制的问题.后续可进一步优化分类器模型参数,满足海洋雷达在复杂海洋环境下稳健且快速探测的需要.

参考文献

References

[1] Zhao W J, Chen Z, Jin M L. Subband maximum eigenvalue detection for radar moving target in sea clutter [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(2): 281-285

[2] 邵夫驰, 行鸿彦. 基于FRFT的多重分形海面小目标检测[J]. 探测与控制学报, 2020, 42(1): 69-74, 80

SHAO Fuchi, XING Hongyan. Small target detection

- based on multi-fractal characteristics of sea clutter FRFT spectrum[J]. Journal of Detection & Control, 2020, 42(1):69-74, 80
- [3] Yan K, Bai Y, Wu H C, et al. Robust target detection within sea clutter based on graphs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(9):7093-7103
- [4] 谢洪森, 邹鲲, 周鹏. 低掠射角海杂波的统计特性分析[J]. 雷达科学与技术, 2011, 9(2):172-179
XIE Hongsen, ZOU Kun, ZHOU Peng. Statistical analysis of sea clutter at low grazing angle[J]. Radar Science and Technology, 2011, 9(2):172-179
- [5] Greco M, Bordonì F, Gini F. X-band sea-clutter nonstationarity: influence of long waves [J]. IEEE Journal of Oceanic Engineering, 2004, 29(2):269-283
- [6] Hu J, Tung W W, Gao J B. Detection of low observable targets within sea clutter by structure function based multifractal analysis[J]. IEEE Transactions on Antennas and Propagation, 2006, 54(1):136-143
- [7] 张林, 李秀友, 刘宁波, 等. 基于分形特性改进的 EMD 目标检测算法[J]. 电子与信息学报, 2016, 38(5):1041-1046
ZHANG Lin, LI Xiuyou, LIU Ningbo, et al. Improved EMD target detection method based on mono fractal characteristics[J]. Journal of Electronics & Information Technology, 2016, 38(5):1041-1046
- [8] Shui P L, Li D C, Xu S W. Tri-feature-based detection of floating small targets in sea clutter[J]. IEEE Transactions on Aerospace and Electronic Systems, 2014, 50(2):1416-1430
- [9] Shi S N, Shui P L. Sea-surface floating small target detection by one-class classifier in time-frequency feature space[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(11):6395-6411
- [10] Gu T C. Detection of small floating targets on the sea surface based on multi-features and principal component analysis[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(5):809-813
- [11] Shui P L, Guo Z X, Shi S N. Feature-compression-based detection of sea-surface small targets[J]. IEEE Access, 2019, 8:8371-8385
- [12] 郭子薰, 水鹏朗, 白晓惠, 等. 海杂波中基于可控虚警 K 近邻的海面小目标检测[J]. 雷达学报, 2020, 9(4):654-663
GUO Zixun, SHUI Penglang, BAI Xiaohui, et al. Sea-surface small target detection based on K-NN with controlled false alarm rate in sea clutter[J]. Journal of Radars, 2020, 9(4):654-663
- [13] Zhou H K, Jiang T. Decision tree based sea-surface weak target detection with false alarm rate controllable [J]. IEEE Signal Processing Letters, 2019, 26(6):793-797
- [14] Gupta A, Gusain K, Popli B. Verifying the value and veracity of extreme gradient boosted decision trees on a variety of datasets [C] // 2016 11th International Conference on Industrial and Information Systems (ICIIS). December 3-4, 2016, Roorkee, India. IEEE, 2016:457-462
- [15] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5):1189-1232

Sea-surface small target detection based on false-alarm-controllable gradient boosting decision tree

LIU Anbang¹ SHI Sainan¹ YANG Jing¹ CAO Ding²

¹ Jiangsu Key Laboratory of Meteorological Detection and Information Processing,

Nanjing University of Information Science & Technology, Nanjing 210044

² Nanjing Marine Radar Institute, Nanjing 210003

Abstract At present, high-dimensional (HD) feature detection is an effective approach to improve the detection performance of sea-surface small targets. The main difficulty lies in the design of classifier in high-dimensional space. Therefore, a feature detection approach based on false-alarm-controllable gradient boosting decision tree (GBDT) is proposed in this paper. First, multiple features are extracted from the 1D long-term observation vector in time, frequency, time-frequency domains to construct an HD feature vector. In this way, the detection problem is converted into a binary classification problem. Second, two types of balanced training samples are solved by simulating returns with target. Third, GBDT algorithm is introduced to condense the HD feature vector into 1D predicted value in probability. The predicted value is used as detection statistics to solve the problem of uncontrollable false alarm rate perplexed the binary classifier. Finally, experimental results are verified by IPIX measured data, which show that the proposed detector can make full use of all the information from the HD characteristics, and the performance is improved by over 13%.

Key words sea clutter; small target detection; gradient boosting decision tree (GBDT); high-dimensional feature