

基于随机森林的雷达高分辨距离像目标识别新方法

姚莉娜, 吴艳敏, 崔光照

(郑州轻工业学院 电气信息工程学院, 河南 郑州 450002)

摘 要: 高分辨距离像(HRRP)包含了丰富的目标结构信息,被广泛地应用于雷达自动目标识别领域.传统的基于贝叶斯分类器和核分类器的 HRRP 识别方法分别存在模型-数据失配以及计算复杂度大等问题.为了解决以上问题,提出了基于随机森林(RF)的 HRRP 识别方法,详细介绍了随机森林的原理和用于 HRRP 识别的基本步骤.仿真数据的实验结果表明,与贝叶斯分类器和核分类器相比,随机森林在识别性能和计算效率等方面有显著提高.

关键词: 目标识别;高分辨距离像;随机森林

中图分类号: TN959.1 **文献标志码:** A **doi:**10.3969/j.issn.1671-6833.2014.04.025

0 引言

雷达高分辨距离像(High Resolution Range Profile, HRRP)是用宽带雷达信号获取的目标散射点回波在雷达视线上的投影,它包含了目标的尺寸、散射点分布等重要结构信息.与合成孔径雷达像等其它宽带信号相比,HRRP 具有易于获取和处理的优点,且实时性好,因此被广泛地用于雷达自动目标识别领域^[1-2].

分类器设计是构建雷达 HRRP 目标识别系统的一个关键环节,它直接决定了识别系统的性能.目前国内外已有较多工作进行了相关研究,这些工作主要包含以下两类:(1)基于统计建模的贝叶斯分类器^[3-5].该类工作先假设 HRRP 样本服从某种概率分布,对其统计建模,然后在识别阶段计算测试样本的后验概率,并利用贝叶斯分类器进行类别判断.文献[3]和[4]中假设 HRRP 服从高斯分布,分别采用自适应高斯模型和子空间模型对 HRRP 建模;文献[5]中假设 HRRP 服从 Gamma 分布,采用混合 Gamma 模型对其建模.该分类器的优点是能够提供判别置信度,且便于结合附加信息改善性能.但当假设的概率分布与数据真实分布不符时,识别性能将会显著下降;(2)基于核变换的分类器^[6-7],其代表为支持向量机(Support Vector Machine, SVM).文献[6]中使用 SVM 对 HRRP 进行分类.文献[7]中使用核主成分分析对 HRRP 进行识别.核分类器的优点

是能够在较少训练样本的条件下仍保持较好的识别性能.然而它需要多次交叉验证选择合适的核函数与核参数,计算量巨大.

基于此,笔者提出一种基于随机森林(Random Forest, RF)的 HRRP 识别方法.RF 是近年来在模式识别领域中流行的一种组合分类器,其原理是利用自举重采样的方式,从原始训练样本集生成多个训练样本子集,并分别对其进行决策树学习,然后基于多个决策树的判决,采用投票的方式得到最终判决结果.由于 RF 不需要对样本的统计分布进行假设,也不用通过多次实验确定核参数,因此很好地克服了贝叶斯分类器与核分类器的缺陷^[9].

1 高分辨距离像及其敏感性预处理

当雷达工作在较高频段时,目标的部件尺寸远大于雷达信号的波长.此时,可以使用散射点模型^[10]来描述目标特性.由该模型知,目标的 HRRP 可表示为

$$\mathbf{x} = e^{j\omega} [x_1, x_2, \dots, x_n]^T. \quad (1)$$

式中: $\omega = -4\pi\lambda/R$ 为初相, λ 表示波长, R 表示目标-雷达距离; n 表示距离单元数,第 m ($m = 1, 2, \dots, n$) 个距离单元的回波 $x_m = \sum_{i=1}^{I_m} \sigma_{im} e^{j\theta_{im}}$, σ_{im} 表示该距离单元中第 i 个散射点的强度, θ_{im} 表示其相位.

原始的 HRRP 回波具有多种敏感性,在对其

收稿日期:2014-03-30;修订日期:2014-05-19

基金项目:国家自然科学基金资助项目(61070238)

作者简介:姚莉娜(1979-),女,河南息县人,郑州轻工业学院讲师,硕士,主要研究方向为电气工程及信息化技术,
E-mail:ylnsmile@126.com.

识别之前需要进行预处理.下面分别对 HRRP 的各种敏感性及相关预处理办法进行简单介绍.

(1) 初相敏感性及其预处理. 距离像的初相 ω 对于目标距离的变化非常敏感,而当前雷达对于目标的测距精度不足以准确地补偿初相,因此在多数识别工作中都丢掉 HRRP 的相位信息,只使用其幅度进行识别.具体地,令 $\mathbf{x}' = |\mathbf{x}|$,其中 $|\cdot|$ 表示取模操作符;

(2) 强度敏感性及其预处理. 距离像的强度依赖于目标距离远近及雷达发射功率等因素,对于识别来说是无用信息,因此在识别前需对 HRRP 进行归一化预处理,即令 $\mathbf{y} = \mathbf{x}' / \|\mathbf{x}'\|$,其中 $\|\cdot\|$ 表示模 2 范数;

(3) 平移敏感性及其预处理. 目标运动会引起回波信号平移,因此需要对 HRRP 进行平移对齐预处理.文献[11-12]中讨论了多种 HRRP 对齐方法,笔者采用文献[11]中的对齐方法.

2 决策树与随机森林分类器

决策树^[13]是构成 RF 的基础,它是一个树结构分类器,由根节点、内部节点和叶节点构成,如图 1 所示.树中根节点,代表了整个训练样本集.树中的每个内部节点对应一个分裂问题,它将到达该节点的样本按某个特征进行分割,把样本集划分为多块.树的叶节点是带有类别标号的数据集合.从决策树的根节点到叶节点的每一条路径都形成一个分类.决策树分类器不需要样本统计分布的先验知识,并且能够较好地解决数据缺失等问题.但是其分类规则复杂,容易产生过匹配现象.

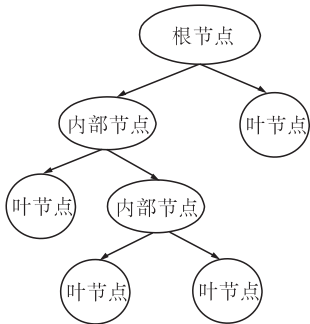


图 1 决策树分类器示意图
Fig.1 Decision tree classifier Schematic

为了解决过匹配问题,Breiman 在决策树的基础上提出了随机森林分类器,其原理是将多个弱分类器集成为一个强分类器.如图 2 所示,一个 RF 分类器由 K 个决策树 $\{h(\mathbf{y}, \boldsymbol{\theta}_k), k = 1, 2, \dots, K\}$ 组成. \mathbf{y} 代表 HRRP 样本, $\{\boldsymbol{\theta}_k\}$ 是一系列统计独立同分布的随机向量,决定了决策树的生长过程. $h(\mathbf{y}, \boldsymbol{\theta}_k)$ 的取值 $t(t = 1, 2, \dots, T)$ 对应于样本 \mathbf{y} 的类别编号, T 表示目标类别总数.最终这 K 个

决策树的投票结果决定了测试样本的类别.

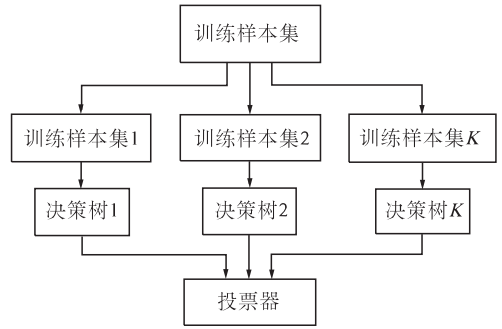


图 2 RF 分类器示意图
Fig.2 RF classifier schematic

给定一组决策树 $h(\mathbf{y}, \boldsymbol{\theta}_1), h(\mathbf{y}, \boldsymbol{\theta}_2), \dots, h(\mathbf{y}, \boldsymbol{\theta}_K)$, 定义如下边缘函数.

$$mg(Y, C) = av_k I(h(Y, \boldsymbol{\theta}_k) = C) - \max_{j \neq C} av_k I(h(Y, \boldsymbol{\theta}_k) = j). \tag{2}$$

式中: Y 表示 HRRP 训练样本集; C 表示与之对应的类别标号集合; $I(\cdot)$ 表示示性函数; $av(\cdot)$ 表示求平均; $\max(\cdot)$ 表示求最大值. 该函数描述了样本集 Y 的平均正确得票数超过最大错误得票数的程度. 其值越大, 表示分类的置信度越高. 基于式 (2), 可将 RF 分类器的推广误差表示为

$$GE = P_{Y,C}(mg(Y, C) < 0). \tag{3}$$

式中: $P_{Y,C}(\cdot)$ 表示该概率测度定义在 Y, C 空间上. 文献[8]中证明随着决策树个数 K 的增加, 对于任意的随机向量序列 $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K$, 推广误差 GE 几乎总是收敛于

$$P_{Y,C}(P_{\boldsymbol{\theta}}(h(Y, \boldsymbol{\theta}) = C) - \max_{j \neq C} P_{\boldsymbol{\theta}}(h(Y, \boldsymbol{\theta}) = j) < 0). \tag{4}$$

式中: $\boldsymbol{\theta}$ 表示单个决策树对应的随机向量. 因此, RF 分类器能够避免决策树分类器的过匹配问题.

下面给出 RF 分类器的训练过程.

(1) 对于给定的 HRRP 训练样本集, 利用 Bagging 算法^[8]产生 K 个有差异的训练样本子集 $\{S_k, k = 1, 2, \dots, K\}$;

(2) 将 S_k 作为第 k 个决策树的输入, 采用如下 C4.5 算法^[13]学习第 k 个决策树;

① 从 HRRP 样本 \mathbf{y} 的 n 维特征中随机选取 L 个作为内部节点的候选分裂特征, 记为 y_1, y_2, \dots, y_L ;

② 按照式(5-8)分别计算 L 个候选特征的信息增益比^[8] $\{GR(y_l), l = 1, 2, \dots, L\}$.

$$GR(y) = \frac{Gain(y)}{Split(y)}, \tag{5}$$

$$Gain(y) = H(Y) - \sum_{i=1}^T \frac{N(Y_i)}{N(Y)} \times H(Y_i), \tag{6}$$

$$Split(y) = - \sum_{i=1}^T \frac{N(Y_i)}{N(Y)} \times \log_2 \left(\frac{N(Y_i)}{N(Y)} \right), \tag{7}$$

$$H(Y) = - \sum_{t=1}^T p(Y,t) \times \log_2(p(Y,t)). \quad (8)$$

式中: $\{Y_1,Y_2,\cdots,Y_T\}$ 表示利用特征 y 对训练样本集 Y 划分得到的子集集合^[8]; $Gain(y)$ 表示特征 y 对应的信息增益; $H(Y)$ 表示样本集 Y 的熵; $H(Y_i)$ 表示样本子集 Y_i 的熵; $N(Y)$ 表示 Y 中的样本数; $Split(y)$ 表示特征 y 的拆分信息; $p(Y,t)$ 表示 Y 中包含的第 $t(t = 1,2,\cdots,T)$ 类目标样本的比例. 最终挑选信息增益比最大的特征对决策树进行分裂;

③ 重复步骤 2.2 直到所有候选分裂特征的信息增益比小于预设的阈值. 文献[14] 中分析并比较了多种阈值计算方法, 这里采用最小化错误率修剪法来确定阈值;

(3) 重复步骤 1 和 2,直到生成 K 个决策树.

在测试阶段,使用 RF 对未知类别的测试样本 y' 进行分类时,其类别标号 $c(c = 1,2,\cdots,T)$ 由投票准则确定,即

$$c = \arg \max_c (\frac{1}{T} \sum_{i=1}^T I(h(y,\theta_i) = c)). \quad (9)$$

3 实验结果和讨论

3.1 数据介绍

本节的识别实验使用电磁仿真软件对三类飞机目标缩比模型计算得到的 HRRP 数据. 雷达信号与飞机的参数如表 1 所示中心频率为 5 520 MHz, 带宽为 400 MHz. 为了测试 RF 分类器的推广性能, 使用 10°俯仰角条件下生成的数据作为训练样本,12°俯仰角条件下生成的数据作为测试样本. 其中,各类目标的训练样本数分别为 7 200, 测试样本数分别为 4 000,各 HRRP 样本为 128 维

的复向量. 训练样本与测试样本均涵盖了目标的所有方位姿态.

表 1 雷达信号和飞机参数
Tab.1 Radar signals and aircraft parameters

飞机类别	长度/m	宽度/m	缩放比例
F15	19.43	13.05	1:1
M21	15.76	7.15	1:1
Y3	36.4	34.9	1:1

3.2 识别结果比较与分析

实验中,分别使用贝叶斯、SVM 和 RF 3 种分类器对以上数据进行识别实验:贝叶斯分类器中,假设 HRRP 样本服从高斯分布,使用文献[3] 中的自适应高斯模型对其统计建模;SVM 采用文献[6]中的形式,核函数为高斯核,且核参数在区间 [1,100]上取值,并通过 5 重交叉验证选取最优核参数;RF 中,根据文献[8]的建议,令 $L = \sqrt{n} \approx 11$,决策树个数 K 取 100. 表 2 中分别给出了各分类器对 3 类目标的识别结果.

从表中可以看到,3 种分类器的识别性能逐渐提高. 其中,贝叶斯分类器的平均识别率最低,只有 94.3%. 这是因为 HRRP 的统计分布比较复杂,而基于高斯分布假设的 AGC 模型与 HRRP 数据之间存在一定程度的失配. SVM 分类器在学习分类界面时引入了监督信息,其平均识别率较贝叶斯分类器提高了 0.9%;RF 分类器集成了多个决策树分类器的识别结果,同时避免了模型-数据失配等问题,因此识别性能最好,平均识别率达到了 97.2%. 值得注意的是 3 种分类器对 Y3 目标的识别率都很高. 这是因为 Y3 的物理尺寸明显大于 F15 和 M21,其 HRRP 可分性较好.

表 2 不同分类器对 3 类目标识别得到的混淆矩阵
Tab.2 The confusion matrix of different classification on three categories target identification %

目标类别	贝叶斯分类器			SVM 分类器			RF 分类器		
	F15	M21	Y3	F15	M21	Y3	F15	M21	Y3
F15	94	1.6	4.4	95.8	1.1	3.1	97.1	0.6	2.3
M21	10.3	89.7	0	7.9	90.2	1.9	2.8	94.7	2.5
Y3	0.9	0	99.1	0.3	0	99.7	0.1	0	99.9
平均识别率	94.3			95.2			97.2		

表 3 中给出了各分类器在训练和测试阶段消耗的时间. 在训练阶段,贝叶斯分类器所用时间最短. SVM 分类器由于需要通过交叉验证选取核参数,其计算复杂度较大,所需时间最长. RF 分类器的训练时间介于二者之间. 在测试阶段,贝叶斯分类器需要将测试样本与多个模板逐一匹配,因此其识别时间最长,SVM 分类器的识别时间略小于贝叶斯分类器,RF 分类器的识别时间最短. 由以上结果可知,采用 RF 分类器对 HRRP 进行识别,能够

在提高识别性能的同时,极大地改善识别效率.

表 3 不同分类器的训练时间与识别时间比较
Tab.3 The comparison of results of training time and recognition time about different classifiers

分类器	训练时间/s	单个样本平均识别时间/ms
贝叶斯分类器	20.0	2.4
SVM 分类器	1 918.5	2.2
RF 分类器	67.2	0.1

4 结论

笔者围绕雷达 HRRP 目标识别问题,研究了基于随机森林的 HRRP 识别方法.该方法通过融合多个决策树分类器的结果来提高识别性能,克服了决策树分类器的过匹配问题.基于仿真数据的实验结果表明,与传统贝叶斯分类器和核分类器相比,随机森林分类器避免了对样本分布假设所引入的失配误差,同时也不需要多次交叉验证来选取分类器参数,因而在平均识别率和识别速度方面均有显著改善.

参考文献:

- [1] 朱劼昊,周建江,吴杰. 基于半参数化概率密度估计的雷达目标识别[J]. 电子与信息学报,2010,32(9): 2161-2166.
- [2] DU Lan, LIU Hong-wei, BAO Zheng. Radar HRRP statistical recognition: parametric model and model selection [J]. IEEE Transactions on Signal Processing, 2008,56(5): 1931-1944.
- [3] JACOBS S P, SULLIVAN J A. Automatic target recognition using sequences of high resolution radar range profiles [J]. IEEE Transactions on Aerospace Electronic Systems, 2000,36(2): 364-380.
- [4] DU Lan, Liu Hong-wei, BAO Zheng. Radar HRRP statistical recognition based on hypersphere model [J]. Signal Processing, 2008,88(5): 1176-1190.
- [5] WEBB A R. Gamma mixture models for target recognition [J]. Pattern Recognition, 2000,33(12): 2045-2054.
- [6] WANG Xiao-dan, Wang Ji-qin. Support vector machine for HRRP classification [C]//Seventh International Symposium on Signal Processing and Its Applications, 2003.
- [7] LIU Hong-wei, Su Hong-tao, Bao Zheng. Radar high range resolution profiles feature extraction based on kernel PCA and kernel ICA [C]//Advances in Neural Networks, 2005.
- [8] BREIMAN L. Random forests [J]. Machine Learning, 2001,45(1): 5-32.
- [9] HAM J, CHEN Y, Crawford M, et al. Investigation of the random forest framework for classification of hyper-spectral data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2005,43(3): 492-501.
- [10] 张爱兵,计科峰,邹焕新,等. 高分辨 SAR 目标散射中心模型分析[J]. 雷达科学与技术, 2009,7(6): 416-424.
- [11] GIL P R, ROSA Z M, JARABO A P, et al. Statistical analysis of the zero-phase method for aligning noisy high-resolution radar signals [J]. IET Radar, Sonar, and Navigation, 2008,3(1): 62-69.
- [12] PORTEGIES J, VAN D H, GELSEMA S, et al. Fast translation invariant classification of HRR range profiles in a zero phase representation [J]. IEE Proceedings-Radar, Sona, and Navigation, 2003,150(6): 411-418.
- [13] POLAT K, GUNES S. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems [J]. Expert Systems with Applications, 2009,36(2): 1587-1592.
- [14] 王黎明. 决策树学习及其剪枝算法研究[D]. 武汉: 武汉理工大学, 2007.

A New Radar HRRP Target Recognition Method Based on Random Forest

YAO Li-na, WU Yan-min, CUI Guang-zhao

(College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China)

Abstract: High resolution range profile (HRRP) contains abundant target structure information, and is widely used in radar automatic target recognition (RATR) area. Traditional Bayes classifier or kernel classifier based recognition method suffers from the model-data mismatch or huge computational burden. To solve this problem, a random forest (RF) based HRRP recognition method is proposed. Both the principle and algorithm procedure of RF are introduced in detail. Experimental results based on simulated data show that RF is superior in recognition performance and computation efficiency compared with Bayes classifier and kernel classifier.

Key words: target recognition; high resolution range profile; random forest