

基于泛化熵随机森林的雷达目标航迹起始方法

刘安宁¹, 史建涛^{1,2}

(1. 南京工业大学 电气工程与控制科学学院, 南京 211816; 2. 中国电子科技集团公司第十四研究所, 南京 210039)

摘要: 文章提出了一种基于泛化熵随机森林的雷达目标航迹起始方法, 通过引入一种泛化熵可调参数, 解决了传统随机森林算法对差异数据集泛化能力和适应性差且难以确定全局最优解的难题。首先利用雷达多周期回波数据构建随机森林算法决策模型, 其次基于设计的泛化熵处理规则对实测数据进行分类处理得到所需测试样本, 最后将测试样本输入完成模型训练的随机森林进行迭代寻优和自主决策以获取雷达目标的航迹起始结果。雷达实测数据验证表明, 文章所提出的方法相比于传统随机森林方法具有更为优越的目标航迹起始性能。

关键词: 目标跟踪; 航迹起始; 随机森林; 泛化熵

中图分类号: V443; TN253

文献标志码: A

文章编号: 1674-7135(2024)01-0007-06

Track initiation based on random forests with generalized entropy

LIU Anning¹, SHI Jiantao^{1,2}

(1. College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing 211816, China;

2. Nanjing Research Institute of Electronic Technology, Nanjing 210039, China)

Abstract: A track initiation approach based on random forests with generalized entropy has been proposed in this paper. In the framework of generalized entropy, the generalization performance of decision tree in random forest is improved on various data sets by introducing an adjustable parameter. Furthermore, it can effectively over-come the problems of poor adaptability to data and local optimization in the process of tree building. In this work, the radar target track initiation problem has been transformed into sample classification with supervised learning. In the first stage, the random forest decision tree model has been constructed based on historical data. Then the measurement data are preprocessed using the generalized entropy adjustable simple rule method to get the test samples. Finally, the test samples will be input the trained random forest to get the target track initiation result. It is verified by radar data that the proposed method has better track initiation performance compared with the traditional random forest method.

Key words: target tracking; track initiation; random forest; generalized entropy

0 引言

目标跟踪是雷达进行探测感知的基本任务, 精确的航迹起始是雷达装备对目标进行稳定跟踪的基础和前提。航迹起始, 是指在雷达未对目标进行稳定跟踪之前所进行的目标初始航迹确立过程。

显然, 航迹起始是雷达完成目标跟踪任务需要解决的首要问题。现代雷达面临的探测环境日益复杂, 特别是强杂波、强干扰的影响, 使雷达量测的不确定性显著增加。雷达目标的航迹起始已成为雷达数据处理研究的主要问题。

收稿日期: 2023-10-30; 修回日期: 2023-12-26

基金项目: 国家自然科学基金项目(编号: U19B2031); 预先研究项目(编号: 61404130210)

引用格式: 刘安宁, 史建涛. 基于泛化熵随机森林的雷达目标航迹起始方法[J]. 空间电子技术, 2024, 21(1): 7-12. LIU A N, SHI J T. Track initiation based on random forests with generalized entropy[J]. Space Electronic Technology, 2024, 21(1): 7-12.

目前,雷达目标跟踪领域经典的航迹起始算法主要包括两种:顺序法和批处理法。其中,顺序法又可以分为逻辑法和直观法两类^[1],此类方法的核心思想是对雷达连续扫描过程中获取的多周期目标电磁回波数据进行序贯处理,通过判断落在跟踪波门内目标点迹的关联结果确定是否触发目标航迹起始流程。显然,上述目标航迹起始算法运算简便、精度较高,但只适用于弱杂波环境中的目标跟踪。目前,常用的批处理航迹起始方法主要基于 Hough 变换处理^[2-3],上述方法通过将雷达多次往复周期性扫描获得的目标数据联合处理,进而利用 Hough 变换实现目标数据的非相干积累,从而提升雷达目标航迹起始性能。目前已有学者指出,批处理方法在强杂波背景下雷达目标航迹起始中具有理想的性能,但面临算法繁琐、信号处理实时性差等突出问题

近年来,兼具数据处理精度高、普适性强、不易过拟合、可处理非线性数据等诸多优点的随机森林算法逐渐受到广泛关注^[4-8]。特别是考虑到随机森林算法在数据并行处理、多目标实时优化、复杂杂波抑制等方面的优越性能,学术界和工业界开始了基于随机森林的雷达目标跟踪技术研究。

随机森林是一种基于构造多棵决策树和集成学习策略相结合的有监督学习算法,所构造的决策树性能优劣会对随机森林的综合性能带来直接影响。近年来广泛使用的决策树分裂准则均存在决策树构造无法全局最优、难以处理异构多源数据等一系列问题,严重制约了雷达目标航迹起始性能^[9]。

针对上述问题,本文提出了一种基于泛化熵随机森林的雷达目标航迹起始方法,论文主要创新点和贡献概述如下:通过引入一种泛化熵可调参数,解决了传统随机森林算法对差异数据集泛化能力和适应性差且难以确定全局最优解的难题,并利用雷达实测数据验证了新方法的有效性。

1 随机森林算法

随机森林算法的基本流程如下:

第一步,利用有放回采样策略从初始数据集中抽取 m 个样本并生成对应的 m 个训练集,进而完成 m 个决策树模型的训练;

第二步,针对每棵决策树,根据训练样本的特征数,每次分裂时基于信息增益系数,选取最优特

征完成分裂;

第三步,利用多棵决策树构造目标随机森林。其使用的基本操作是利用多棵树分类器投票结果确定分类结果,或者利用多棵决策树的预测值均值构造最终的预测结果。

1.1 决策树

决策树是一类具有数据在线分类和预测回归功能的树形结构有监督学习算法,其核心思想是 if-then 规则的集合。建立决策树的主要步骤流程包括:设定根节点,然后将训练数据均置于该节点,并选取最优特征,基于该特征将训练数据集划分为若干个子集,使得在当前条件下各个子集均有一个最好的分类;进行判断分析,若上述划分子集可以完成准确分类,则所设定的根节点即为预期的叶节点,然后把相关子集划分到对应的叶节点中;如果存在无法正确分类的子集,则需要对此类子集进行最优特征重新选取,继续完成子集分割,并构建相应节点,直至所有训练集被正确分类;经过上述步骤的操作,每个子集都被分到相应的节点上,具有了明确标签的子集分类,从而完成决策树的构建^[10-11],如图 1 所示。

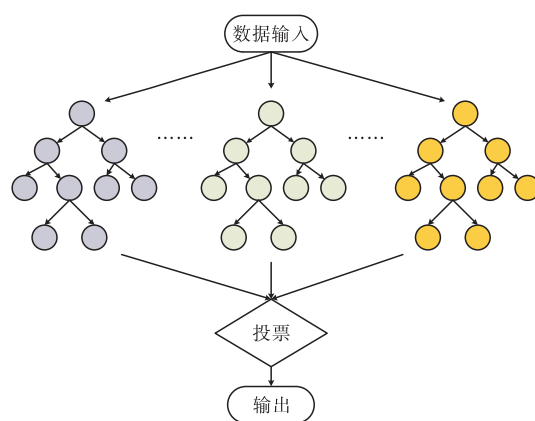


图 1 决策树基本原理图

Fig.1 Decision tree fundamentals

目前,使用最广泛的决策树生成方法是 ID3 算法,其核心思想是以样本属性值对应的香农熵大小作为分裂节点选取和判断的准则;其中,最关键的变量——信息不确定性量度如式(1)所列。

$$H(X) = - \sum_{k=1}^n p(x_k) \ln p(x_k) \quad (1)$$

式(1)中, X 是取值于集合 $\{x_1, x_2, \dots, x_n\}$ 的随机向量, $p(x_k)$ 为变量 x_k 发生的概率。

ID3 算法确定节点分裂属性的基本判据是分裂

前后样本的信息增益大小如式(2)所列;其中,包含信息增益最大的属性划分所构成的子节点中样本类别的纯度最高^[12]。

$$G(M) = H(\Phi) - \sum_{i=1}^n (|\Phi_i|/|\Phi| \times H(\Phi_i)) \quad (2)$$

式(2)中, n 表示属性 M 的个数, $G(M)$ 为根据属性 M 分割样本 Φ 带来的样本信息量减少,定义为信息增益。需要指出的是,ID3 算法的一个重要局限性是难以处理连续变量;此外,信息增益的计算过程中,取值较多的属性会被赋予更高的权值,这显然会增加决策树误差。

1.2 随机森林

随机森林最早由 Leo Breiman 和 Adele Cutler 提出,是一种利用多棵决策树对样本进行训练和预测的新型分类器^[4],分类结果的类型由个别决策树的输出类型的众数决定。经典的随机森林算法流程如图 2 所示。

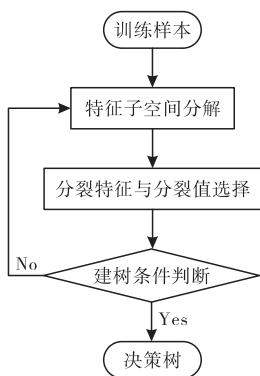


图 2 随机森林算法流程图

Fig. 2 The flow chart of random forest algorithm

随机森林算法包括如下三个部分:

样本抽取:在构建决策树的过程中,以等权重有放回随机抽样方法在特定的数据集中抽取相同数量的数据样本。

特征子空间构建:基于决策树中根节点的功能,从 M 个特征集合中随机抽取 m 个特征($m < M$),并从上述 m 个特征构成的特征子空间中按照最大均方差下降准则选择分裂特征并确定分裂点;循环执行上述过程不断构造决策树节点,直至触发中止条件。

投票表决:每棵决策树首先完成单独预测,基于上述预测结果,随机森林按照多数投票规则确定最终的预测结果。

2 泛化熵随机森林

本项工作采用的泛化熵如式(3)所列。

$$T_k(X) = \frac{1}{1-\lambda} \left(\sum_{i=1}^n p(x_i)^\lambda - 1 \right), \lambda \in \mathbb{R} \quad (3)$$

式(3)中 X 是取值于集合 $\{x_1, x_2, \dots, x_n\}$ 的随机向量, $p(x_k)$ 为变量 x_k 发生的概率。不同于经典的香农熵,此处采用的泛化熵具有可调参数 λ 。显然,上述形式的泛化熵可以看作香农熵的一种变形。

进一步处理上述泛化熵,可以得到如式(4)所列。

$$T_k(X) = - \sum_{i=1}^n p(x_i)^\lambda \frac{(x_i)^{1-\lambda} - 1}{1-\lambda}, \lambda \neq 1, x \geq 0 \quad (4)$$

已有研究表明泛化熵 $T_k(X)$ 具有如下性质^[7]:

(1)当 $\lambda \rightarrow 1$ 时,泛化熵 $T_k(X)$ 可以转化为传统的香农熵 $H(X)$;

(2)当 $\lambda = 2$ 时,泛化熵 $T_k(X)$ 可以转化为如式(5)所列的基尼系数;

$$\sum_{i=1}^n p(x_i)(1 - p(x_i)) \quad (5)$$

(3)信息增益比(Information Gain Ratio, IGR)可以视为在基于传统香农熵的信息增益中增加了归一化项,如果香农熵等价于泛化熵,则 IGR 可看作泛化熵增益比。显然,当 $\lambda \rightarrow 1$ 时,泛化熵增益比可以变为 IGR。

由式(3)可以看出,此处所采用的泛化熵比传统的香农熵增加了一个可调参数 λ ;同时由式(3)的等价形式(4)可以发现,可调参数 λ 的加入可大大提升信息熵标识样本数据一致性和混杂程度的能力。此外,由上述分析过程可以发现,选择不同的参数 λ ,泛化熵可以等价于香农熵、基尼系数和信息增益比,如式(5)所示。显然,决策树的性能可以通过调整参数 λ 来得到提升,同时,泛化熵也可以构建多种不同属性的分裂准则^[13-14]。因此,我们可以利用新的分裂规则重新构造决策树并通过选择合适的可调参数 λ 提升对多种类型数据集的适应性。

3 泛化熵随机森林航迹起始

基于泛化熵随机森林的雷达目标航迹起始算法主要流程如下:

第一步,雷达量测数据预处理:对雷达连续多

次扫描获得的量测数据,首先将明显不满足目标运动特性的量测去除,只有满足下述条件的量测组合才能输入泛化熵随机森林中。测得的速度在目标可能达到的最大速度和最小速度之间;测得的加速度绝对值小于目标的最大加速度;任意两个量测矢量之间的夹角满足目标运动特性约束。

第二步,特征提取:这里选取的特征主要包括雷达 N 次扫描获得的量测的连续两个批次扫描点之间的平均速度;连续 3 个批次扫描点的两段连线的加速度;连续 3 个批次扫描点的两段连线的夹角。

第三步,结果输出:将经过预处理和特征提取后的数据输入到训练完成的泛化熵随机森林中,每个决策树给出决策结果,最后经过投票确定最终的目标航迹起始结果。

4 雷达实测数据验证结果

研究表明,目标跟踪过程中需要重点关注的是杂波问题^[15],杂波抑制效果是衡量目标航迹起始效果的重要指标^[16-17]。因此,本文利用某型 X 波段雷达实测数据对所突出的泛化熵随机森林算法的目标航迹起始性能进行验证。

第一种实验场景是弱杂波下匀速运动目标的目标航迹起始。该场景下,目标为实验小型飞机,飞行速度约为 230 km/h,雷达回波信号的信噪比约为 -10 dB。

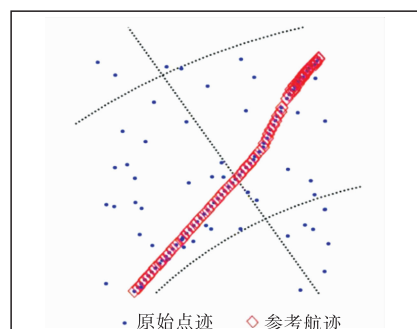
目标航迹起始结果如表 1 所列,可以看出,针对弱杂波背景下的匀速运动目标,基于泛化熵随机森林方法比传统的随机森林方法在目标丢点数、正确杂波点、杂波滤除率等雷达目标航迹起始关键指标方面均具有显著的优势,实现了精确稳定的目标航迹起始。

表 1 匀速目标航迹起始处理结果

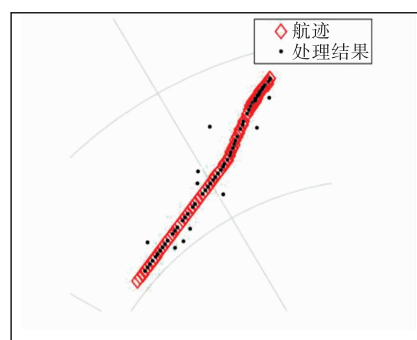
Tab.1 Processing Results of uniform velocity targets

参数类型	泛化熵随机森林方法	随机森林方法
目标点迹数	69	69
杂波点迹数	55	55
目标丢点数	0	11
正确杂波点	53	49
杂波滤除率	96.36%	89.09%

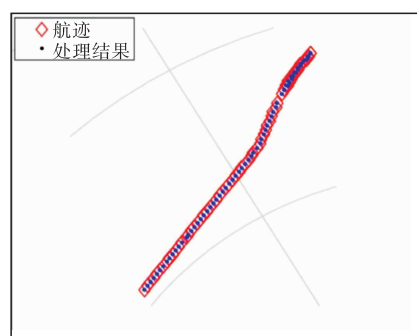
目标跟踪结果如图 3 所示,可以看出基于泛化熵随机森林方法比传统的随机森林方法获得了更为准确的目标跟踪结果。



(a) 原始点迹图



(b) 传统随机森林处理结果



(c) 泛化熵随机森林处理结果

图 3 匀速运动目标跟踪结果

Fig.3 Tracking results of uniform velocity targets

第二种场景是强杂波背景下机动目标的目标航迹起始。该场景下,目标为仍为实验小型飞机,飞行速度在 200 km/h 到 400 km/h 间机动切换,雷达回波信号的信噪比约为 -50 dB。

目标航迹起始结果如表 2 所列,可以看出,针对强杂波背景下的机动目标,基于泛化熵随机森林方法比传统的随机森林方法在目标丢点数、正确杂波点、杂波滤除率等雷达目标航迹起始关键指标方面均具有显著的优势,实现了精确稳定的目标航迹起始。

表 2 机动目标航迹起始处理结果

Tab. 2 Processing Results of maneuvering targets

参数类型	随机森林方法	泛化熵随机森林方法
目标点迹数	102	102
杂波点迹数	255	255
目标丢点数	5	0
正确杂波点	221	242
杂波滤除率	86.67%	94.90%

目标跟踪处理结果如图 4 所示,可以看出基于泛化熵随机森林方法比传统的随机森林方法获得了更为准确的目标跟踪结果。

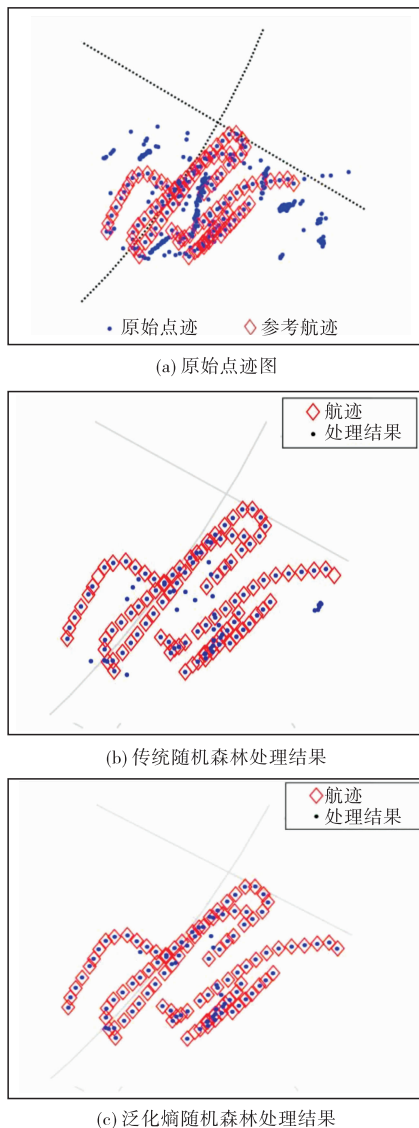


图 4 机动目标跟踪结果

Fig. 4 Tracking results of maneuvering targets

需要说明的是,关于上述实验中“正确杂波点”的获得,文章实验验证中所采用的雷达实测数据为经过事后复盘处理的标准数据集,已根据各方面的

综合信息对杂波点进行了准确标注,从而可以对论文所提出目标航迹起始方法的效果进行验证。

由上述雷达数据处理结果可以看出,无论是对于匀速运动目标还是机动目标,所提出的泛化熵随机森林算法比传统的随机森林算法均具有更好的目标航迹起始和目标跟踪性能。

5 结论

随机森林作为一种性能优良的机器学习分类方法,已经在雷达目标跟踪相关领域得到了应用。但是,传统的决策树分裂准则不可避免会遇到异构数据适应性差、建树难以全局最优等突出问题,会对雷达目标航迹起始性能带来不利影响。针对上述问题,文章提出了一种基于泛化熵随机森林的雷达目标航迹起始方法,通过引入一种泛化熵可调参数,解决了传统随机森林算法对差异数据集泛化能力和适应性差且难以确定全局最优解的难题,从而显著提升雷达目标航迹起始性能。

参考文献:

- [1] 吴顺君,梅晓春. 雷达信号处理和数据处理技术[M]. 北京:电子工业出版社,2008:102-105.
- [2] LIU H W, LIU H L, DAN X D, et al. Cooperative track initiation for distributed radar network based on target track information[J]. IET Radar, Sonar & Navigation, 2016, 10(4):735-741.
- [3] LI X M, YANG, G Q. RFLPP-hough transform for chirp signal detection [J]. Journal of Information Science and Engineering, 2015, 31(6):2125-2135.
- [4] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1):5-32.
- [5] 陈光荣,王军政,郭盛,等. 基于 C4.5 决策树的自主姿态选择算法[J]. 指挥与控制学报, 2021, 7(1):38-45.
- [6] XIA J S, LIAO W Z, CHANUSSOT J, et al. Improving random forest with ensemble of features and semisupervised feature extraction [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(7):1471-1475.
- [7] GHOSH D, CABRERA J. Enriched random forest for high dimensional genomic data [J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022, 19(5):2817-2828.
- [8] 王文竹,李智,来嘉哲,等. 基于数据驱动的空间目标智能识别[J]. 指挥与控制学报, 2019, 5(1):25-30.
- [9] 王奕森. 随机森林和深度神经网络的若干关键技术研究[D]. 北京:清华大学, 2018:45-48.
- [10] BARROS R C, BASGALUPP M P, DE CARVALHO A C

- P L F, et al. A survey of evolutionary algorithms for decision-tree induction [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2012, 42(3):291-312.
- [11] BEN-GAL I, DANA A, SHKOLNIK N, et al. Efficient construction of decision trees by the dual information distance method [J]. Quality Technology & Quantitative Management, 2014, 11(1):133-147.
- [12] QUINLAN J R. Induction of decision trees [J]. Machine Learning, 1986, 1(1):81-106.
- [13] WANG Y S, SONG C B, XIA S T. Improving decision trees by Tsallis Entropy Information Metric method [C]// 2016 International Joint Conference on Neural Networks (IJCNN), 2016:4729-4734.
- [14] WANG Y S, XIA S T, WU J. A less-greedy two-term Tsallis Entropy Information Metric approach for decision tree classification [J]. Knowledge-Based Systems, 2017, 120:34-42.
- [15] ALI MEMON S, KIM M, SHIN M, et al. Extended smoothing joint data association for multi-target tracking in cluttered environments [J]. IET Radar, Sonar & Navigation, 2020, 14(4):564-571.
- [16] LIU H Q, HUANG C, GAN L, et al. Clutter reduction and target tracking in through-the-wall radar [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(1):486-499.
- [17] YAN B, PAOLINI E, XU L P, et al. A target detection and tracking method for multiple radar systems [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-21.

作者简介:刘安宁(2001—),安徽六安人,硕士研究生。主要研究方向分布式协同、目标跟踪。E-mail:Liuaning@njtech.edu.cn

通讯作者:史建涛(1982—),河南开封人,博士、教授、博士生导师。主要研究方向为信号处理目标检测与跟踪、复杂系统协同控制。E-mail:sjt11@tsinghua.org.cn