

文章编号: 1671-0576(2024)01-0038-05

基于 CART 算法的空间目标本体反演方法

史珮颖¹, 孟夏莹¹, 朱凌轩¹, 杜君¹, 王月娟²

(1. 散射辐射全国重点实验室, 上海 200438; 2. 上海黎明瑞达电子科技有限公司, 上海 200090)

摘要: 针对预警中空间群目标反演精度不高、鲁棒性较差的问题, 提出了一种基于分类回归树(classification and regression tree, CART)算法的空间目标本体反演方法。首先构建分类与回归决策树, 并将空间目标的入射角、光谱辐射亮度、温度等特征数据作为决策树的输入; 再基于基尼系数评估数据纯度, 对目标在不同温度、不同入射角下的光谱辐射亮度进行分割; 最后通过在每个节点处对数据集进行划分, 实现对空间目标本体的有效反演。数值对比实验表明, 3种典型空间目标的决策树模型反演结果与实际情况一致, 验证了所提方法的有效性。

关键词: 空间目标; 本体反演; 分类回归树算法; 基尼系数

中图分类号: TP75

文献标志码: A

DOI: 10.3969/j.issn.1671-0576.2024.01.007

Spatial Object Ontology Inversion Method Based on CART Algorithm

SHI Peiyong¹, MENG Xiaying¹, ZHU Lingxuan¹, DU Jun¹, WANG Yuejuan²

(1. National Key Laboratory of Scattering and Radiation, Shanghai 200438, China;

2. Shanghai LimRadar Electronic Technology Co., Ltd., Shanghai 200090, China)

Abstract: Aiming at the problems of low precision and poor robustness of spatial group object inversion in early warning process, a spatial object ontology inversion method was proposed based on the classification and regression tree (CART) algorithm. Firstly, the classification and regression decision tree was built, the feature data such as incidence angles, spectral radiance, temperature of the spatial object were extracted as the input of the decision tree. Then the data purity was evaluated according to Gini coefficient, the spectral radiance of the object at different temperatures and incidence angles was divided. Finally, the effective inversion of the spatial object ontology was realized by dividing the data set at each node. Numerical experiments results show that the prediction results of the decision tree model for three typical spatial objects are consistent with the truth, which verifies the effectiveness of the proposed method.

Key words: spatial object; ontology inversion; CART algorithm; Gini coefficient

收稿日期: 2023-06-15

基金项目: 国家自然科学基金(U22B2045)

作者简介: 史珮颖(1998—), 女, 硕士, 助理工程师, 主要从事目标特性仿真技术研究。

0 引言

近年来,随着对太空开发利用的不断深入,空间目标的数量、类型、功能日益增多,带来了轨道拥挤、碰撞风险等问题。空间目标可分为在役卫星、退役卫星、空间碎片、宇宙飞行器和深空天体等,在飞行过程中均可能会发生分离、调姿、抛洒、轨道机动等关键事件^[1-2]。通过对这些事件进行监测和分析,可以对目标的运动状态、潜在风险等作出判断,再根据目标的运动状态采用不同的识别方法,以减轻目标识别系统的负担,辅助目标识别过程。根据目标识别结果,判断目标的类型或功能,这对于了解空间目标的性质、意图和威胁,以及制定相应的应对策略都有重要意义。

当前,国内外针对舰船及飞机编队等群目标的分类关联算法已有很多研究成果^[3-4]。文献[5]提出采用多假设跟踪方法处理复杂数据关联问题,提高了空间群目标跟踪的稳定性。但是多假设跟踪方法很难快速找出合适的假设,而且计算量大,实时处理能力不强。文献[6]基于实测数据,提出了一种结合辐射相对量和熵特征的目标分类方法,但该方法对于飞机俯仰和方位信息的鲁棒性较差,且对数据源的要求较高。在对空间目标进行中段预警的过程中,对群目标的识别存在背景复杂、目标辐射强度弱、成像像素少、探测数据不足等问题,且目前国内外对群目标中段预警的研究公开报道得较少。

决策树算法^[7-8]是在信息论基础上建立的数据分类方法,具有分类精度高、鲁棒性好、模型可解释性强等特点。本文在空间目标特性建模的基础上,以目标的几何形状、温度、入射角、光谱辐射亮度等特征作为输入,采用基于决策树的分类回归树(classification and regression tree, CART)算法实现目标本体特征反演,并通过3种典型空间目标的决策树模型预测数值实验结果与实际数据结果的对比分析,验证所提算法的有效性。

1 反演方法

1.1 空间目标特征建模

在实际观测过程中,天基传感器观测到的空

间目标通常是一个点辐射源,只能利用点目标的光谱辐射信息进行识别^[9-12]。光谱辐射特性与物体大小、物体形状、物体材质、热物参数、太阳位置、运动状态、运动姿态、观测点位置等因素均有关。通过对中段弹头和诱饵红外辐射特性差异的分析可知,天基传感器能够从点目标光谱辐射信息中提取出的目标本体特征主要有温度特征、谱分布特征、姿态特征等。本文将采用黑体辐射定律计算在不同入射角、不同温度下的目标本体光谱辐射亮度,实现目标本体特征的建模。建模过程共包含投影面积计算、光谱辐射亮度计算、空间目标温度计算等3个方面。

(1) 投影面积计算

由于空间目标形状复杂多样,进行目标投影面积计算时,通常采用三角面元分割法将目标表面分割为一系列小三角面片,并尽量保证每个面片内部各处的表面法矢量近似一致。剖分后的面片信息通常分两个文件存储:一个存储小三角面片顶点的顺序,可按存储顺序连成一个三角形面片;另一个存储按小三角面片顺序出现的每个顶点的 x 轴、 y 轴和 z 轴坐标,以此表示整个目标模型。

对探测视场范围内所有三角面片的投影面积累积求和,并减去被遮掩部分的面积,可得目标的总投影面积。定义描述凸体目标表面的参数化坐标系,如图1所示。图中, $\mathbf{r}(x, y)$ 为半径矢量,在点 (x, y) 处的 x, y 方向正切矢量分别为 \mathbf{r}_x 和 \mathbf{r}_y ,这两个正切矢量的外积垂直于目标表面该点处的单位法矢量 \mathbf{n} 。

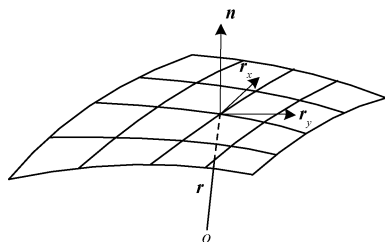


图1 目标表面参数化坐标系示意图

点 (x, y) 处的单位法矢量

$$\mathbf{n} = (\mathbf{r}_x \times \mathbf{r}_y) / |\mathbf{r}_x \times \mathbf{r}_y| \quad (1)$$

式中: $|\cdot|$ 为取模运算符。

将目标表面按 dx 和 dy 划分为无数小面片,则点 (x, y) 位置处的面片面积

$$dS = |\mathbf{r}_x \times \mathbf{r}_y| dx dy \quad (2)$$

采用上述三角面元分割方法,可以将凸体表面剖分为 N 个小三角面片。设探测视线矢量为 \mathbf{n}_{LOS} ,第 i ($i=1,2,\dots,N$) 个面片的面积和法矢量分别为 S_i 和 \mathbf{n}_i 。由于凸体表面上任意三角面片之间不存在遮掩现象,则目标投影面积即为探测视场内所有三角面片投影面积的累积和,即目标投影面积

$$S_{\text{proj}} = \sum_{i=1}^N S_i \max(\cos(-\mathbf{n}'_{\text{LOS}}, \mathbf{n}_i), 0) \quad (3)$$

式中: $\max(\cdot)$ 为最大值取值函数; \mathbf{n}'_{LOS} 为目标局部坐标系下的探测视线矢量; $\cos(-\mathbf{n}'_{\text{LOS}}, \mathbf{n}_i)$ 为 $-\mathbf{n}'_{\text{LOS}}$ 和 \mathbf{n}_i 之间夹角的余弦。当 $\cos(-\mathbf{n}'_{\text{LOS}}, \mathbf{n}_i)$ 大于 0 时,该面片在探测视场范围内,投影面积有效,为此引入表达式 $\max(\cos(-\mathbf{n}'_{\text{LOS}}, \mathbf{n}_i), 0)$ 。

(2) 光谱辐射亮度计算

根据普朗克(Planck)辐射定律,温度为 T_i 的面片的光谱辐射亮度 $L_{b,\lambda}(T_i)$ [13] 可以表示为

$$L_{b,\lambda}(T_i) = \frac{c_1}{\lambda^5} (\exp(c_2/(\lambda T_i)) - 1)^{-1} \quad (4)$$

式中: $c_1 = 2hc^2$ 为第一辐射常数,其中 $h = 6.626\ 176 \times 10^{-34}$ J·s 为普朗克常量, c 为真空中的光速; λ 为波长; $c_2 = hc/k$ 为第二辐射常数,其中 $k = 1.380\ 6 \times 10^{-23}$ J/K 为玻耳兹曼常数。 $L_{b,\lambda}(T_i)$ 也称为黑体光谱辐射亮度,其单位为 $\text{W}/(\text{m}^2 \cdot \text{sr} \cdot \text{cm})$ 。

(3) 空间目标温度计算

空间目标的自身辐射是红外目标探测时需要分析的重要方面。空间目标表面的温度场是决定其自身辐射的主要因素,而温度场又由目标所处环境及目标材料等因素决定。空间目标表面温度主要受到的影响因素有太阳直接辐射、地球系统直接辐射、经地球反射的太阳辐射、目标向宇宙空间发射的辐射及内能的增加量。把空间目标表面划分为若干相对等温且热力学特性基本相同的面片,分析每个面片与环境的热交换,可以建立各面片的瞬时热平衡方程,则第 i 个面片的内能增加量 $Q_{5,i}$ 满足热平衡方程

$$Q_{5,i} = Q_{1,i} + Q_{2,i} + Q_{3,i} - Q_{4,i} \quad (5)$$

其中

$$\begin{cases} Q_{1,i} = E_s S_i F_{s,i} \alpha_{s,i} \\ Q_{2,i} = E_e S_i F_{e,i} \alpha_{e,i} \\ Q_{3,i} = E_s \rho_{se} S_i F_{se,i} \alpha_{s,i} \\ Q_{4,i} = \varepsilon \sigma T_i^4(t) S_i \\ Q_{5,i} = C_{m,i} M_{m,i} dT_i(t)/dt \end{cases} \quad (6)$$

式中: $Q_{1,i}$ 为太阳直接照射到面片上的辐射; $Q_{2,i}$ 为地球系统直接入射到面片上的辐射; $Q_{3,i}$ 为经过地球系统反射并入射到面片上的太阳辐射; $Q_{4,i}$ 为面片向宇宙空间发射出的辐射; E_s 为太阳辐射常数; $F_{s,i}$ 为面片接收到的太阳直接辐射的角系数; $\alpha_{s,i}$ 为面片对太阳直接辐射的吸收系数; $E_e = 237 \text{ W}/\text{m}^2$ 为地球自身总辐射照度; $F_{e,i}$ 为面片接收到的地球辐射的角系数; $\alpha_{e,i}$ 为面片对地球辐射的吸收系数; ρ_{se} 为地球系统对太阳辐射的平均反射率,取值为 0.3; $F_{se,i}$ 为面片接收到的经地球系统反射的太阳辐射的角系数; ε 为目标外表面材料的发射率; σ 为斯特藩-玻耳兹曼常数; $T_i(t)$ 为面片在 t 时刻的温度; $C_{m,i}$ 为面片表面材料的比热容; $M_{m,i}$ 为面片的质量。

1.2 基于 CART 算法的反演流程

CART 算法是一种经典的处理分类和回归问题的机器学习方法,对离散型、连续型变量的分类问题均能较好地解决。本文中的空间目标本体特征为离散型变量,采用 CART 算法可实现对目标本体类型的有效分类。基于 CART 算法的目标本体反演是先采用基尼(Gini)系数最小化准则进行特征选择,再采用递归方法构建二叉分类决策树分类,最终实现对目标本体的反演。

基尼系数是一种度量数据纯度的参数,具有计算速度快、容易理解的特点,常用于表征样本集合中一个随机选中的样本被错误分类的概率。基尼系数具体定义为样本被正确分类的概率与样本被错误分类的概率的乘积,其表达式为

$$\text{Gini}(p_k) = \sum_{k=1}^K p_k (1 - p_k) = 1 - \sum_{k=1}^K p_k^2 \quad (7)$$

式中: p_k 为选中的样本属于 k 类别的正确分类概率,则这个样本被错误分类的概率是 $(1 - p_k)$; K 为样本的总类别数。基尼系数的取值越小表示集合中被选中的样本被分错的概率越小,即集合的纯度越高,反之,集合越不纯。

除基尼系数外,数据纯度还可用信息熵、信息增益、信息增益比等信息论中的常用概念来衡量。鉴于基尼系数计算速度快、更易理解,本文在 CART 算法中使用基尼系数作为数据纯度的衡量指标。

CART 算法的基本流程如下。

步骤 1,选择一个合适的属性作为根节点,将训练数据集按照该属性的不同取值分成若干子集。

步骤 2,如果子集中所有样本属于同一类别,则将该子集标记为叶节点,返回该类;否则,重复步骤 1,选择一个合适的属性作为该子集的分支节点,继续划分。

步骤 3,重复步骤 2,直到所有子集都被标记为叶节点或者达到预设的停止条件。

本论文将 CART 算法应用于目标本体反演预测,通过递归学习,基于目标温度、入射角、光谱辐射亮度等特征推测目标决策分类基准,构建空间目标多特征决策树分类反演模型。基于 CART 算法的目标本体反演流程如图 2 所示。

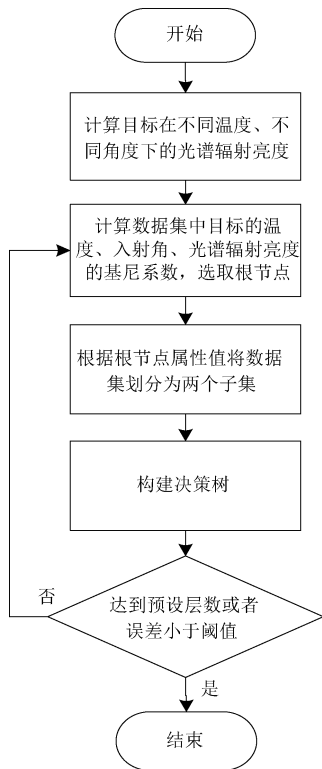


图 2 基于 CART 算法的目标本体反演流程图

2 实验验证

采用 CART 算法,选用基尼系数作为数据纯度的衡量指标,进行空间目标本体反演数值验证实验。选用的 3 个空间目标模型如图 3 所示,分别标记为模型 1、模型 2、模型 3。

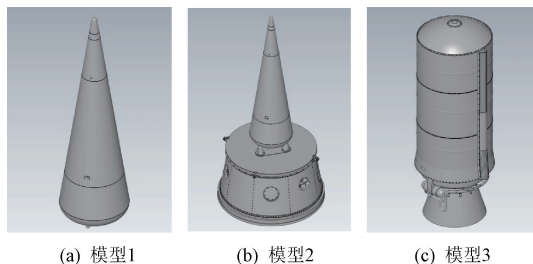


图 3 空间目标模型

建立 3 类典型目标本体反演的决策树模型,将 3 个模型的温度、入射角、光谱辐射亮度特征数据作为 CART 算法的决策树输入,进行分类性能验证。为确保实验结果真实可靠,分别开展 4 次实验,分类结果如表 1 所示。

表 1 采用 CART 算法的分类结果

实验序号	决策树层数	分类精度/%
1	7	98.81
2	8	99.40
3	8	99.10
4	8	97.31

由表 1 可以看出:本文算法对测试集数据的分类精度平均为 98.66%,其中最高可达到 99.40%。这验证了采用决策树方法可以获得较好的空间目标本体分类结果。

第 1 次实验的可视化结果如图 4 所示。该决策树的叶子表示实际输出类别标签,其中橙色、绿色、紫色分别对应 3 类目标,颜色越深代表基尼系数越小,即数据越纯。

由图 4 可看出:本次实验采取 7 层决策树结构,且决策树最终层的颜色最深,这表明实验获得了较好的判别结果。但是,决策树层数越多不一定分类效果越好,层数过多可能会存在过拟合。

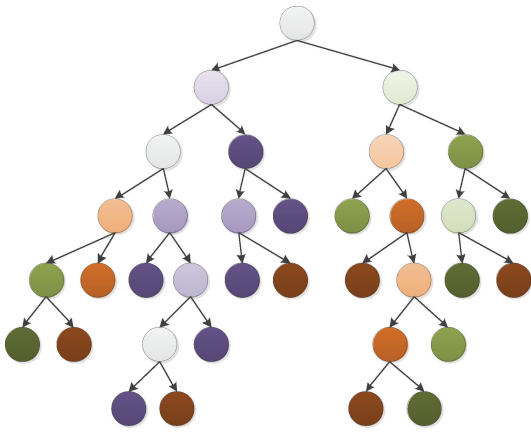


图 4 第 1 次实验的可视化结果

一般可以通过修剪等方法来避免过拟合问题的产生。

3 结论

空间目标在飞行过程中会发生分离、调姿、抛洒、轨道机动等关键事件,针对关键事件导致群目标难以区分的问题,本文提出了基于 CART 算法的空间目标本体的反演方法。该方法以仿真计算得到的目标的温度、入射角、光谱辐射亮度等特征数据为输入,利用 CART 算法,实现目标的本体反演预测。实验结果验证了该方法的有效性。后续继续深入开展空间目标本体反演方法的研究对空间群目标预警探测具有重要意义。

参考文献

- [1] 汤绍勋,易先清,罗雪山. 面向目标跟踪与探测的天基弹道导弹防御系统仿真[J]. 系统工程, 2011, 29(3): 103-107.
- [2] 杨少春,吴林峰,王刚,等. 弹道导弹中段轨迹预

- 测研究[J]. 空军工程大学学报(自然科学版), 2012, 13(4): 31-35.
- [3] ZHU Y Q, ZHOU S L, GAO G. Emitter group targets tracking using GM-P HD filter combined with clustering [J]. Systems Engineering and Electronics, 2015, 37(9): 1967-1973.
- [4] 汪云,胡国平,刘进忙,等. 群目标跟踪自适应 IMM 算法[J]. 哈尔滨工业大学学报, 2016, 48(10): 103-109.
- [5] 张伟. 空间群目标下多假设跟踪方法研究[D]. 成都: 电子科技大学, 2014.
- [6] 李明,吴娇娇,左磊,等. 基于实测数据的空中目标分类识别算法[J]. 电子与信息学报, 2018, 40(11): 2606-2613.
- [7] 汪靖翔. 决策树算法的原理研究和实际应用[J]. 电脑编程技巧与维护, 2022(8): 54-56, 72.
- [8] 王文霞. 数据挖掘中改进的 C4.5 决策树分类算法[J]. 吉林大学学报(理学版), 2017, 55 (5): 1274-1277.
- [9] 丁小峰. 基于窄带信息的弹道中段目标特性反演技术的研究[D]. 长沙: 国防科学技术大学, 2011.
- [10] 樊志鹏. 基于时序光度信号的空间目标物性参数反演方法[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [11] 毋亚北. 基于红外辐射强度序列的空间目标姿态和形状快速反演技术研究[D]. 长沙: 国防科技大学, 2017.
- [12] KAASALAINEN M, TORPPA J. Optimization methods for asteroid lightcurve inversion: I. shape determination[J]. Icarus, 2001, 153(1): 24-36.
- [13] 李亨,李劲东,王玉莹,等. 中段飞行弹道导弹表面温度与辐射特性计算[J]. 红外技术, 2022, 44(2): 134-139.