

考虑综合性能最优的非短视快速天基雷达多目标跟踪资源调度算法

王增福*^① 杨广宇^① 金术玲^②

^①(西北工业大学自动化学院 西安 710072)

^②(中国电子科技集团公司第三十八研究所 合肥 230088)

摘要: 合理有效的资源调度是天基雷达效能得以充分发挥的关键。针对天基雷达多目标跟踪资源调度问题,建立了综合考虑目标威胁度、跟踪精度与低截获概率(LPI)的代价函数;考虑目标的不确定、天基平台约束以及长期期望代价,建立了多约束下的基于部分可观测的马尔可夫决策过程(POMDP)的资源调度模型;采用拉格朗日松弛法将多约束下的多目标跟踪资源调度问题转换分解为多个无约束的子问题;针对连续状态空间、连续动作空间及连续观测空间引起的维数灾难问题,采用基于蒙特卡罗树搜索(MCTS)的在线POMDP算法——POMCPOW算法进行求解,最终提出了一种综合多指标性能的非短视快速天基雷达多目标跟踪资源调度算法。仿真表明,与已有调度算法相比,所提算法资源分配更合理,系统性能更优。

关键词: 天基雷达; 资源调度; 多目标跟踪; 部分可观测的马尔可夫决策过程; 蒙特卡罗树搜索(MCTS)

中图分类号: TN95

文献标识码: A

文章编号: 2095-283X(2024)01-0253-17

DOI: 10.12000/JR23162

引用格式: 王增福, 杨广宇, 金术玲. 考虑综合性能最优的非短视快速天基雷达多目标跟踪资源调度算法[J]. 雷达学报(中英文), 2024, 13(1): 253–269. doi: 10.12000/JR23162.

Reference format: WANG Zengfu, YANG Guangyu, and JIN Shuling. A non-myopic and fast resource scheduling algorithm for multi-target tracking of space-based radar considering optimal integrated performance[J]. *Journal of Radars*, 2024, 13(1): 253–269. doi: 10.12000/JR23162.

A Non-myopic and Fast Resource Scheduling Algorithm for Multi-target Tracking of Space-based Radar Considering Optimal Integrated Performance

WANG Zengfu*^① YANG Guangyu^① JIN Shuling^②

^①(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

^②(The 38th Research Institute of China Electronics Technology Group Corporation, Hefei 230088, China)

Abstract: Appropriate and effective resource scheduling is the key to achieving the best performance for a space-based radar. Considering the resource scheduling problem of multi-target tracking in a space-based radar system, we establish a cost function that considers target threat, tracking accuracy, and Low Probability of Interception (LPI). Considering target uncertainty and constraints of the space-based platform and long-term expected cost, we establish a resource scheduling model based on the Partially Observable Markov Decision Process (POMDP) with multiple constraints. To transform and decompose the resource scheduling problem of multi-target tracking with multiple constraints into multiple unconstrained sub-problems, we use the Lagrangian relaxation method. To deal with the curse of dimensionality caused by the continuous state space,

收稿日期: 2023-09-07; 改回日期: 2023-11-18; 网络出版: 2023-12-20

*通信作者: 王增福 wangzengfu@nwpu.edu.cn *Corresponding Author: WANG Zengfu, wangzengfu@nwpu.edu.cn

基金项目: 国家自然科学基金(U21B2008)

Foundation Item: The National Natural Science Foundation of China (U21B2008)

责任主编: 严俊坤 Corresponding Editor: YAN Junkun

©The Author(s) 2023. This is an open access article under the CC-BY 4.0 License
(https://creativecommons.org/licenses/by/4.0/)

continuous action space and continuous observation space, we use the online POMDP algorithm based on the Monte Carlo Tree Search (MCTS) and partially observable Monte Carlo planning with observation widening algorithm. Finally, a non-myopic and fast resource scheduling algorithm with comprehensive performance indices for multi-target tracking in a space-based radar system is proposed. Simulation results show that the proposed algorithm, when compared with the existing scheduling algorithms, allocates resources more appropriately and shows better performance.

Key words: Space-based radar; Resource scheduling; Multi-target tracking; Partially observable Markov decision process; Monte Carlo Tree Search (MCTS)

1 引言

天基雷达具有全天时、全天候的陆、海、空、天目标探测能力,可有效弥补现有预警系统的不足。但对于目标搜索、跟踪等复杂多任务需求,天基雷达系统资源十分有限。合理高效的资源调度是提升天基雷达性能的关键技术之一,面临着任务复杂、资源约束复杂、目标不确定等问题。如何统筹分配天基雷达系统资源,如辐射功率、驻留时间等,以最大限度满足系统任务需求,成为目前亟需解决的问题。

雷达资源调度问题是系统资源配置约束下的优化问题。其中,目标函数的构建至关重要。根据不同应用场景选择合理的目标函数可以保证资源调度的充分性,其通常考虑两方面准则:跟踪性能最大化准则和低截获概率(Low Probability of Intercept, LPI)性能最大化准则。文献[1]以预测条件克拉默-拉奥下界(Predicted Conditional-Cramér-Rao Lower Bound, PC-CRLB)为优化性能指标,基于两步半正定规划方法,提出了一种联合节点选择和功率分配策略。文献[2]面向相控阵雷达组网多目标跟踪问题,为最小化雷达资源消耗,采用后验克拉默-拉奥下界(Posterior Cramér-Rao Lower Bound, PCRLB)量化跟踪性能,提出了一种优化的资源分配策略。与统一分配资源方法相比,所提策略可以利用更少的资源实现预期性能。文献[3]将PCRLB作为目标跟踪性能指标,利用两步法求解建立的非凸优化问题,提出了一种在非理想检测环境下用于多基地雷达系统多目标跟踪的联合节点选择和功率分配策略。文献[4]针对共址多输入多输出(Multiple Input Multiple Output, MIMO)雷达中的机动目标跟踪任务,以PC-CRLB为优化性能指标,提出了一种功率分配策略,以提高机动目标跟踪性能。文献[5]以多目标跟踪误差效用函数为优化目标,利用基于分区的三阶段法和梯度投影法,提出了一种联合在线航路规划和资源优化策略,以提高多目标跟踪能力。文献[6]针对雷达组网目标跟踪,以LPI为优化准则,通过优化雷达网络中的重访问

隔、驻留时间和发射功率,提出了一种自适应资源管理方法。文献[7]为解决分布式相控阵雷达网络目标跟踪问题,分别采用PC-CRLB和拦截概率作为目标跟踪精度和LPI性能指标,利用优化技术协同各雷达节点的发射功率、驻留时间、波形带宽和脉冲宽度,提出了一种联合发射资源管理和波形选择策略,以提高雷达网络的目标跟踪精度和LPI性能。

在面向动态系统的优化控制中,通过仅考虑一个预测步骤优化目标函数的方法称为短视或贪婪策略,而考虑未来多步的方法称为非短视策略^[8,9]。文献[1-7]的调度策略均为短视策略。考虑到雷达多目标跟踪中大部分目标为非合作目标,量测不确定,以及当前决策对更长远未来的影响,可引入有限长时折扣的序贯决策模型框架,将雷达资源调度问题建模为部分可观测的马尔可夫决策过程(Partially Observable Markov Decision Process, POMDP),实现非短视雷达资源分配。与短视策略相同,非短视策略通过计算动作对目标函数的影响来评估每个可用动作;不同的是,其目标函数综合了未来一系列动作,具有提高资源分配效能的优势^[10]。

文献[11]考虑单平台多传感器目标感知,采用POMDP对问题建模,通过序贯贝叶斯推理计算目标信念状态,使用贪心算法优化雷达资源配置。文献[12]将多模式传感器管理问题转换为多变量POMDP的雷达调度问题,提出了用于估计最佳多线性阈值策略的随机近似算法,计算出调度策略以确定被选目标以及持续时间使得代价函数最小化。文献[13]考虑了存在干扰机时跟踪多目标的认知雷达资源管理问题,将其表述为基于混合POMDP的博弈模型,提出了一种低复杂度的联合优化算法计算雷达最优抗干扰策略。

实际问题中,资源分配可能会受到多种约束限制,其中约束代表所有任务的可用资源限度,需要采用受约束的POMDP对资源调度问题进行建模。文献[14]提出了一种在跟踪精度约束下最小化辐射代价的非短视调度,其采用POMDP建模资源调度问题,通过PCRLB和隐马尔可夫模型滤波器分别

预测未来有限时间范围内的跟踪精度和辐射代价，最后利用分支定界剪枝算法实现资源调度优化。在基于POMDP的多目标跟踪的资源调度问题中，通常面临着维数爆炸问题^[15]。利用拉格朗日松弛法(Lagrangian Relaxation, LR)可将此类问题解耦为更易解决的子问题^[16,17]。例如，文献^[18]使用LR将整数规划形式的多目标跟踪调度问题解耦为多个可以快速求解的单目标POMDP问题，提出了基于拉格朗日对偶问题的可行解构造方法。

鉴于POMDP的广泛应用，其求解方法得到广泛关注。中小规模POMDP问题可通过众多离线算法求解^[19]。对于大规模POMDP问题，目前已提出基于离线和在线采样的方法。文献^[20]提出的基于点的值迭代(Point-Based Value Iteration, PBVI)算法在具有数百个状态的问题上表现出良好性能。后续的Perseus、启发式搜索值迭代(Heuristic Search Value Iteration, HSVI)、SARSOP (Successive Approximations of the Reachable Space under Optimal Policies)等离线策略在速度以及最优性上逐步提升^[19,21,22]。在线方法通常由前向搜索组成，以找到各时间步内可执行的最佳动作，通过计算较好的局部策略来减轻计算复杂度，主要方法为分支定界剪枝，蒙特卡罗采样以及启发式搜索算法等^[23]。其中，POMCP (Partially Observable Monte Carlo Planning)^[24]是一种基于蒙特卡罗采样的在线算法，使用广泛，其基本框架为蒙特卡罗树搜索(Monte Carlo Tree Search, MCTS)。MCTS依赖于信念的粒子表示，采用粒子滤波来更新信念，可使POMCP类算法应用于具有非常大甚至连续状态空间的POMDP问题。虽然已提出DESPOT (DEterminized Sparse Partially Observable Tree)^[25]，ABT (Adaptive Belief Tree)^[26]等算法，但仍然需要通用的在线POMDP求解算法来解决连续空间，尤其是连续观测空间下的POMDP问题。对此，带有观测加宽的部分可观测的蒙特卡罗规划(Partially Observable Monte Carlo Planning with Observation Widening, POMCPOW)算法^[27]使用双渐进加宽(Double Progressive Widening, DPW)来逐步增加需要考虑的观测集合，本质上利

$$\mathbf{Q}_{i,k} = \begin{bmatrix} \sigma_{i,x}^2 T^4/4 & 0 & 0 & \sigma_{i,x}^2 T^3/2 & 0 & 0 \\ 0 & \sigma_{i,y}^2 T^4/4 & 0 & 0 & \sigma_{i,y}^2 T^3/2 & 0 \\ 0 & 0 & \sigma_{i,z}^2 T^4/4 & 0 & 0 & \sigma_{i,z}^2 T^3/2 \\ \sigma_{i,x}^2 T^3/2 & 0 & 0 & \sigma_{i,x}^2 T^2 & 0 & 0 \\ 0 & \sigma_{i,y}^2 T^3/2 & 0 & 0 & \sigma_{i,y}^2 T^2 & 0 \\ 0 & 0 & \sigma_{i,z}^2 T^3/2 & 0 & 0 & \sigma_{i,z}^2 T^2 \end{bmatrix} \quad (3)$$

其中， $\sigma_{i,x}^2$ ， $\sigma_{i,y}^2$ ， $\sigma_{i,z}^2$ 表示目标*i*在*x*，*y*，*z*这3个方向上的机动强度。

用观测采样逐步离散化观测空间，是一种有效的求解算法。

上述研究成果为解决天基雷达资源调度问题奠定了良好基础。然而，现有研究主要考虑地基、空基雷达，面向天基雷达的资源调度问题鲜有研究；同时，建立POMDP框架后的解决方法多以离线算法为主，计算量大，难以适应当下和未来的资源调度问题；鉴于天基雷达系统存在约束复杂、状态与动作、观测空间连续等特性，需研究更合理的求解方法。如图1所示，本文面向天基雷达多目标跟踪任务，基于POMDP，综合考虑目标威胁度、跟踪精度、LPI性能指标，以平均辐射功率、相参积累时间为待优化变量，建立合理准确的雷达资源调度模型；在此基础上，通过LR将带有约束的多目标POMDP问题分解为多个单目标的POMDP问题；针对连续的状态、动作以及观测空间，基于MCTS，采用POMCPOW算法近似求解，最终提出了基于LR-POMCPOW的天基雷达多目标跟踪资源调度方法，并通过仿真验证了所提方法的有效性。

2 建模与问题描述

2.1 目标运动模型

考虑单基地天基雷达多目标跟踪应用场景。设定雷达可视区域内共有 \mathcal{I} 个运动目标。记离散时刻为 $k(k=1,2,\dots,\mathcal{U})$ ，跟踪采样间隔为 T 。记 k 时刻目标*i*在地心地固(Earth-Centered, Earth-Fixed, ECEF)坐标系下的运动状态向量为 $\mathbf{x}_{i,k}=[x_{i,k}, y_{i,k}, z_{i,k}, \dot{x}_{i,k}, \dot{y}_{i,k}, \dot{z}_{i,k}]$ 。不失一般性，假设目标做匀速直线(Constant Velocity, CV)运动。目标*i*的运动模型为

$$\mathbf{x}_{i,k+1} = \mathbf{F}_{i,k} \mathbf{x}_{i,k} + \mathbf{w}_{i,k} \quad (1)$$

其中，状态转移矩阵 $\mathbf{F}_{i,k}$ 为

$$\mathbf{F}_{i,k} = \begin{bmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

式(1)中， $\mathbf{w}_{i,k}$ 为零均值高斯白噪声，其协方差矩阵为

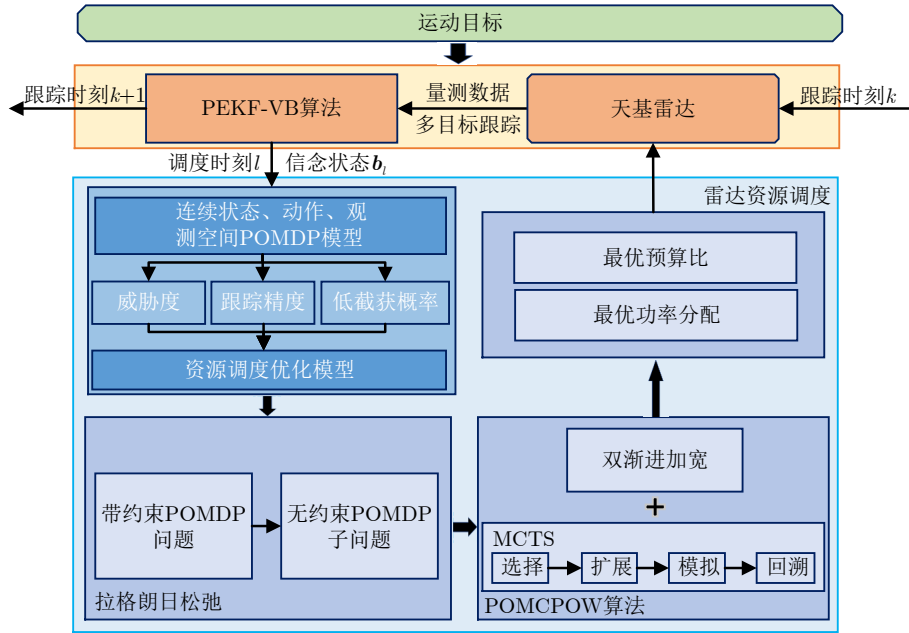


图1 基于LR-POMCPOW的天基雷达多目标跟踪资源调度方法框图

Fig. 1 Schematic diagram of the proposed LR-POMCPOW for resource scheduling of space-based radar multi-target tracking

2.2 雷达量测模型

考虑天基雷达采用共址MIMO体制, 执行多目标跟踪任务时可利用同时多波束机制, 使不同发射波束指向不同目标, 实现同一区域内多目标同时跟踪。雷达量测包括径向距、方位角与俯仰角。记 $k+1$ 时刻关于目标 i 的量测向量为 $\mathbf{y}_{i,k+1} = [r_{i,k+1} \ \theta_{i,k+1} \ \phi_{i,k+1}]^T$ 。量测方程为

$$\mathbf{y}_{i,k+1} = h(\mathbf{x}_{i,k+1}) + \mathbf{v}_{i,k+1} \quad (4)$$

其中, $h(\cdot)$ 的具体形式涉及多次坐标转换, 可参见GB/T 32296—2015; $\mathbf{v}_{i,k+1}$ 为零均值高斯白噪声, 其协方差矩阵为 $\mathbf{R}_{i,k+1}$, 可表示为^[28,29]

$$\mathbf{R}_{i,k+1} = \text{diag} [\sigma_{i,k+1,r}^2 \ \sigma_{i,k+1,\theta}^2 \ \sigma_{i,k+1,\phi}^2]$$

其中, $\sigma_{i,k+1,r}^2$, $\sigma_{i,k+1,\theta}^2$, $\sigma_{i,k+1,\phi}^2$ 分别是目标 i 在 $k+1$ 时刻的径向距、方位角和俯仰角的量测方差, 可进一步表示为

$$\begin{aligned} \sigma_{i,k+1,r}^2 &= \frac{c_1^2 v_{k+1}^2}{2 \cdot \text{SNR}_{i,k+1}}, \\ \sigma_{i,k+1,\theta}^2 &= \frac{\rho_\theta^2}{2c_2^2 \cdot \text{SNR}_{i,k+1}}, \\ \sigma_{i,k+1,\phi}^2 &= \frac{\rho_\phi^2}{2c_2^2 \cdot \text{SNR}_{i,k+1}} \end{aligned}$$

其中, c_1 为电磁波传播速度, v_{k+1} 为 $k+1$ 时刻的脉冲宽度, c_2 为给定常数值, 一般可取为1.57^[30], ρ_θ , ρ_ϕ 分别表示方位向、俯仰向的往返波束宽度, $\text{SNR}_{i,k+1}$ 为 $k+1$ 时刻目标 i 的信噪比, 建模方法如下。

相参体制下, 单基地雷达方程为

$$R_{\max}^4 = \frac{p_{\text{av}} \tau G^2 \bar{\lambda}^2 \sigma}{(4\pi)^3 L_\Sigma k T_0 F_n (\text{SNR})_{\min}} \quad (5)$$

其中, R_{\max} 为雷达最大作用距离, p_{av} 为平均辐射功率, G 为天线增益, $\bar{\lambda}$ 为雷达工作波长, σ 为雷达目标横截面积(Radar Cross Section, RCS), L_Σ 为雷达系统总损耗, 主要包括发射支路损耗、大气损耗、电离层损耗、天线防护损耗、波束形状损耗、处理损耗等, k 为玻尔兹曼常数, T_0 为标准温度(290 K), $kT_0=4 \times 10^{-21}$ W/Hz, F_n 为接收机噪声系数, τ 为相参积累时间, $(\text{SNR})_{\min}$ 为多普勒滤波器输出的最小可检测信噪比。

由式(5)可看到, 信噪比与雷达平均辐射功率、目标RCS、相参积累时间成正相关。由文献[31]可得, $k+1$ 时刻目标 i 的期望信噪比 $\text{SNR}_{i,k+1}$ 与雷达、目标参数的关系式为

$$\begin{aligned} \text{SNR}_{i,k+1}(p_{\text{av},i}, \sigma_i, \tau_i, \mathbf{y}_{i,k+1}) &= \frac{\text{SNR}_0}{p_{\text{av},0} \cdot \sigma_0 \cdot \tau_0 \cdot r_0^{-4}} \cdot p_{\text{av},i} \cdot \sigma_i \cdot \tau_i \cdot (r_{i,k+1})^{-4} \\ &\cdot \exp \left(-2 \left(\frac{(\hat{\theta}_{i,k+1} - \theta_{i,k+1})^2}{\rho_\theta^2} + \frac{(\hat{\phi}_{i,k+1} - \phi_{i,k+1})^2}{\rho_\phi^2} \right) \right) \end{aligned} \quad (6)$$

其中, SNR_0 , $p_{\text{av},0}$, σ_0 , τ_0 , r_0 分别为参考目标的信噪比、发射机平均辐射功率、RCS、相参积累时间、雷达与参考目标的相对距离, $\hat{\theta}_{i,k+1}$, $\hat{\phi}_{i,k+1}$ 分别为 $k+1$ 时刻目标 i 的预测方位角和俯仰角。

2.3 POMDP模型构建

受目标运动系统噪声、雷达量测噪声等多种随机噪声以及量测方程式(4)的非线性等因素的影响,天基雷达多目标跟踪过程是典型的部分可观的随机动态系统,其资源调度是一类典型的非完美状态信息下的动态规划问题。POMDP为该类问题提供了强有力的建模框架。为此,本节采用POMDP框架,对天基雷达多目标跟踪资源调度进行建模。

本文考虑在两个不同时间尺度上完成天基雷达的资源调度。在微观尺度上(即一段调度间隔内)进行雷达对多目标的跟踪,在宏观尺度上完成雷达在有限时长上的资源调度。具体的,离散跟踪时刻为 k ,调度间隔为 $u(T \leq u \leq \mathcal{U})$,其为跟踪采样间隔 T 的整数倍,每经过一次调度间隔 u ,系统进行雷达资源的调度。在执行第 $l(l = 1, 2, \dots, \mathcal{L})$ 次调度时,经调度算法求取的动作在接下来的微观尺度内保持不变,用于此调度间隔内的目标跟踪,直到下一次调度时刻到来时进行更新。从而,天基雷达多目标跟踪资源调度的POMDP模型可表示为7元组 $\langle \mathcal{X}, \mathcal{A}, \mathcal{Z}, \mathcal{Y}, \mathcal{W}, c, \gamma \rangle$,具体含义如下:

(1) \mathcal{X} : 状态空间,系统所有的可能状态集合。在执行第 l 次调度时,目标运动状态为 \mathbf{x}_l 。基于目标运动状态和雷达量测,信念状态定义了目标运动状态的概率分布,表示为 \mathbf{b}_l 。

(2) \mathcal{A} : 动作空间,对系统可以实施的动作集合。在执行第 l 次调度时,选取连续动作向量为 $\mathbf{a}_l = [\tau_l \ p_{av,l}]^T$,其中, τ_l 为雷达波束在目标处的相参积累时间,考虑到天基平台能量受限, $p_{av,l}$ 为单位时间内雷达用于目标跟踪的平均辐射功率。

(3) \mathcal{Z} : $\mathcal{X} \times \mathcal{A} \in \Pi(\mathcal{X})$,状态转移函数,即给定系统状态和动作后关于系统状态转移的概率分布函数。在执行第 l 次调度时,当系统状态处于 \mathbf{x} 时采取动作 \mathbf{a}_l 后转移到状态 \mathbf{x}_{l+1} 的概率为 $\Pr(\mathbf{x}_{l+1} | \mathbf{x}_l, \mathbf{a}_l)$,其中 $\mathbf{x}_l, \mathbf{x}_{l+1} \in \mathcal{X}$ 。假设状态具有马尔可夫性。

(4) \mathcal{Y} : 观测空间,系统量测值的集合。在执行第 l 次调度时,获取的量测值定义为 \mathbf{y}_l 。

(5) \mathcal{W} : $\mathcal{X} \times \mathcal{A} \in \Pi(\mathcal{Y})$,量测分布函数,采取动作 \mathbf{a}_l 后状态转移至 \mathbf{x}_{l+1} 后的量测值为 \mathbf{y}_{l+1} ,其分布函数记为 $\Pr(\mathbf{y}_{l+1} | \mathbf{x}_{l+1}, \mathbf{a}_l)$ 。

(6) c : $\mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$,代价函数,状态 \mathbf{x}_l 下选择动作 \mathbf{a}_l 后的代价函数值为 $c(\mathbf{x}_l, \mathbf{a}_l)$, $\mathbf{x}_l \in \mathcal{X}$, $\mathbf{a}_l \in \mathcal{A}$ 。

(7) γ : 折扣因子。

在执行第 $l(l = 1, 2, \dots, \mathcal{L})$ 次调度后,需要对目标信念状态 \mathbf{b}_l 进行更新。当执行动作 \mathbf{a}_l ,并得到量测 \mathbf{y}_{l+1} 时,采用贝叶斯准则更新信念状态,更新后的信念状态为

$$\begin{aligned} & \mathbf{b}_{l+1}(\mathbf{x}_{l+1}) \\ &= \Pr(\mathbf{x}_{l+1} | \mathbf{b}_l(\mathbf{x}_l), \mathbf{a}_l, \mathbf{y}_{l+1}) \\ &= \frac{\Pr(\mathbf{y}_{l+1} | \mathbf{x}_{l+1}, \mathbf{a}_l) \sum_{\mathbf{x}_l \in \mathcal{X}} \Pr(\mathbf{x}_{l+1} | \mathbf{x}_l, \mathbf{a}_l) \mathbf{b}_l(\mathbf{x}_l)}{\Pr(\mathbf{y}_{l+1} | \mathbf{b}_l(\mathbf{x}_l), \mathbf{a}_l)} \end{aligned} \quad (7)$$

其中, $\Pr(\mathbf{y}_{l+1} | \mathbf{b}_l(\mathbf{x}_l), \mathbf{a}_l)$ 为归一化因子。

2.4 约束条件

根据天基雷达工作特性,考虑如下约束:

(1) 由于卫星平台的定轨和高速运动特性,天基雷达需运行在目标附近的星下点轨迹段以开展目标跟踪任务。从而,对一定区域内的目标,只在固定时段存在可见关系。由此,可设定可见时间窗口为 $[t_{start}, t_{end}]$ 。假设天基雷达运行至窗口起始时刻 t_{start} 开始工作,但跟踪任务结束时刻应不晚于 t_{end} 。在可见时间窗口内,调度结束时间应早于可见时间窗口结束时间,即:

$$\mathcal{L}u \leq t_{end} - t_{start} \quad (8)$$

其中, $t_{end} - t_{start}$ 为卫星与目标的可见时间间隔,雷达资源总调度时长为 $\mathcal{L}u$ 。

(2) 在第 l 次调度时,用于不同目标跟踪的平均辐射功率 $p_{av,i,l}$ 之和不大于雷达跟踪消耗的总能量 E 与可见时间窗口内跟踪总时长 $\mathcal{U}(\mathcal{U} \leq (t_{end} - t_{start})/T)$ 的比值,即:

$$\sum_{i=1}^I p_{av,i,l} \leq \frac{E}{\mathcal{U}}, \forall l, 1 \leq l \leq \mathcal{L} \quad (9)$$

(3) 在第 l 次调度时,各跟踪任务的相参积累时间与采样时间间隔比值之和不大于规定的预算比,便于雷达在采样时间间隔 T 内可执行其他任务,即:

$$\sum_{i=1}^I \frac{\tau_{i,l}}{T} \leq \eta, \forall l, 1 \leq l \leq \mathcal{L} \quad (10)$$

2.5 目标函数

在设计目标函数时考虑如下3个方面:雷达对目标进行有效的威胁度区分;最小化目标跟踪误差;雷达在可见时间窗口内用于目标跟踪的能量最小化,从而降低截获概率。为此,第 l 次调度时,定义目标 i 的综合代价函数 $c(\mathbf{x}_{i,l}, \mathbf{a}_{i,l})$ 为

$$\begin{aligned} & c(\mathbf{x}_{i,l}, \mathbf{a}_{i,l}) \\ &= \varphi^2(\mathbf{x}_{i,l}) \cdot (\omega_1 \cdot \text{Tr}(\mathbf{P}_{i,l+1}(\mathbf{a}_{i,l})) + \omega_2 \cdot u \cdot p_{av,i,l}) \end{aligned} \quad (11)$$

其中, ω_1, ω_2 分别是跟踪预测误差代价函数以及能量代价函数的权重。令 $c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l})$ 表示信念状态 $\mathbf{b}_{i,l}$ 下的期望代价。接下来详细介绍式(11)的计算方法。

令 $\varphi(\mathbf{x}_{i,l})$ 表示目标*i*的威胁度函数,取值范围为 $[0,1]$,不失一般性,假设其与目标*i*在第*l*次调度时的状态 $\mathbf{x}_{i,l}$ 有关, φ 值越大,目标威胁度越高。对高威胁度目标,雷达可分配更多资源用于收集目标信息。本文采用目标*i*相对于受保护对象*j*时间、距离的最接近点(Closest Point of Approach, CPA)计算其威胁度^[32]。目标*i*对受保护对象*j*构成的威胁取决于目标*i*可以多快以及多近地接近受保护对象*j*。受保护对象*j*在第*l*次调度时状态为 $\mathbf{x}_{j,l} = [x_{j,l} \ y_{j,l} \ z_{j,l} \ \dot{x}_{j,l} \ \dot{y}_{j,l} \ \dot{z}_{j,l}]^T$,相对运动状态 $\bar{\mathbf{x}}_{ij,l} = [x_{i,l} \ y_{i,l} \ z_{i,l}]^T - [x_{j,l} \ y_{j,l} \ z_{j,l}]^T$, $\bar{\mathbf{v}}_{ij,l} = [\dot{x}_{i,l} \ \dot{y}_{i,l} \ \dot{z}_{i,l}]^T - [\dot{x}_{j,l} \ \dot{y}_{j,l} \ \dot{z}_{j,l}]^T$,则定义时间最接近点为^[32]

$$t_{\text{CPA}} = -\frac{\bar{\mathbf{x}}_{ij,l}^T \bar{\mathbf{v}}_{ij,l}}{\|\bar{\mathbf{v}}_{ij,l}\|_2} \quad (12)$$

距离最接近点为^[32]

$$d_{\text{CPA}} = \|\bar{\mathbf{x}}_{ij,l} + t_{\text{CPA}} \cdot \bar{\mathbf{v}}_{ij,l}\|_2 \quad (13)$$

利用sigmoid函数,目标*i*对对象*j*的威胁度可计算为^[32]

$$\varphi_{\text{time}}(\mathbf{x}_{i,l}; \mathbf{x}_{j,l}) = \begin{cases} 1, & |t_{\text{CPA}}| \leq t_1 \\ 1 - 2 \left(\frac{|t_{\text{CPA}}| - t_1}{t_0 - t_1} \right)^2, & t_1 < |t_{\text{CPA}}| \leq t_{0.5} \\ 2 \left(\frac{|t_{\text{CPA}}| - t_0}{t_0 - t_1} \right)^2, & t_{0.5} < |t_{\text{CPA}}| \leq t_0 \\ 0, & t_0 < |t_{\text{CPA}}| \end{cases} \quad (14)$$

$$\varphi_{\text{distance}}(\mathbf{x}_{i,l}; \mathbf{x}_{j,l}) = \begin{cases} 1, & d_{\text{CPA}} \leq d_1 \\ 1 - 2 \left(\frac{d_{\text{CPA}} - d_1}{d_0 - d_1} \right)^2, & d_1 < d_{\text{CPA}} \leq d_{0.5} \\ 2 \left(\frac{d_{\text{CPA}} - d_0}{d_0 - d_1} \right)^2, & d_{0.5} < d_{\text{CPA}} \leq d_0 \\ 0, & d_0 < d_{\text{CPA}} \end{cases} \quad (15)$$

其中, $t_1 < t_{0.5} < t_0$, $d_1 < d_{0.5} < d_0$,下标表示威胁度值 φ 分别为1, 0.5, 0。令 $\varphi(\mathbf{x}_{i,l}) = g_{\text{time}}\varphi_{\text{time}}(\mathbf{x}_{i,l}; \cdot) + g_{\text{distance}}\varphi_{\text{distance}}(\mathbf{x}_{i,l}; \cdot)$,其中, $g_{\text{time}} \geq 0$, $g_{\text{distance}} \geq 0$, $g_{\text{time}} + g_{\text{distance}} = 1$ 。

$\text{Tr}(\mathbf{P}_{i,l+1}(\mathbf{a}_{i,l}))$ 是*l*+1时刻跟踪目标*i*的(预测的)状态误差协方差矩阵的迹,选用该指标可以衡量雷达跟踪精度性能。具体计算可参见PEKF-VB (Proximal Extended Kalman Filter-Variational Bayes)算法的相应步骤^[33]。

$u \cdot p_{\text{av},i,l}$ 是雷达在单次资源调度间隔*u*内消耗在跟踪目标*i*上的能量。考虑*u*不变,则系统消耗能量与辐射功率成正相关。本文选用系统消耗能量作为LPI

性能指标^[34]。一般利用截获概率表示雷达LPI性能,目标*i*对雷达发射信号的截获概率可表示为^[6]

$$p_{i,l}^I(p_{\text{av},i,l}) = \frac{1}{2} \text{erfc} \left(\sqrt{-\ln p_{\text{fa}}} - \sqrt{\frac{p_{\text{av},i,l} G_t G_1 \lambda^2 G_{\text{IP}}}{(4\pi)^2 r_{i,l}^2 k T_0 B_1 F_1} + 0.5} \right) \quad (16)$$

其中, p_{fa} 为拦截接收器的虚警概率, G_t 为雷达发射天线在拦截器方向的增益, G_1 为拦截器天线增益, G_{IP} 为拦截器处理增益, B_1 表示拦截接收器的带宽, F_1 表示拦截接收器噪声系数,函数 $\text{erfc}(x)$ 为

$$\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$$

式(16)中,径向距离 $r_{i,l}$ 为量测分量,无需进行优化,则辐射功率 $p_{\text{av},i,l}$ 与截获概率 $p_{i,l}^I$ 成正相关,从而LPI性能与功率分配密切相关。每经过一次调度间隔*u*,可将雷达的目标跟踪信息通过中继卫星或者直接传送至地面站中心,进行资源调度分析。

2.6 问题声明

在上述目标运动模型、雷达量测模型、POMDP模型、代价函数的基础上,保持微观尺度下的目标跟踪,则宏观尺度下的天基雷达多目标跟踪资源调度可描述为:给定系统初始状态,在满足资源约束条件的基础上,确定最优的可容许确定性策略,使得有限时间下期望的累积多目标总代价函数最小,即:

$$V^*(\mathbf{B}_1) = \min_{\pi \in \Pi} \mathbb{E}^\pi \left[\sum_{l=1}^{\mathcal{L}} \sum_{i=1}^{\mathcal{I}} \gamma^l \cdot c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}) \mid \mathbf{B}_1 \right] \quad (17)$$

s.t. 式(8), 式(9), 式(10)

其中, $V^*(\mathbf{B}_1)$ 表示最优值函数, π 表示可行策略,所有目标的初始信念状态集为 $\mathbf{B} = [\mathbf{b}_{1,1} \ \mathbf{b}_{2,1} \ \dots \ \mathbf{b}_{\mathcal{L},1}] \in \mathcal{B}^{\mathcal{I}}$, $\mathcal{B}^{\mathcal{I}}$ 为*I*个目标的信念状态空间, Π 是可行策略集,最优策略为 $\pi^* = [[\mathbf{a}_{1,1}^* \ \mathbf{a}_{2,1}^* \ \dots \ \mathbf{a}_{\mathcal{L},1}^*]^T \ [\mathbf{a}_{1,2}^* \ \mathbf{a}_{2,2}^* \ \dots \ \mathbf{a}_{\mathcal{L},2}^*]^T \ \dots \ [\mathbf{a}_{1,\mathcal{L}}^* \ \mathbf{a}_{2,\mathcal{L}}^* \ \dots \ \mathbf{a}_{\mathcal{L},\mathcal{L}}^*]^T]$,其满足 $V_{\pi^*}(\mathbf{B}_1) = V^*(\mathbf{B}_1)$ 。不失一般性,本文接下来令折扣因子 $\gamma = 1$ 。式(17)是一个PSPACE-hard问题,最优解求解困难,因此,只能寻找近似解法。

3 基于LR-POMCPOW的天基雷达多目标跟踪资源调度

针对式(17),本节采用LR方法将带有约束的动态规划问题转换为无约束的动态规划问题,然后将由*I*个目标构成的高维无约束动态规划问题分解为*I*个由单目标构成的一维无约束动态规划子问题。考虑到连续状态空间、连续动作空间及连续观测空

间引起的维数灾难问题，采用基于MCTS的POM-CPOW算法，最终给出了一种综合多指标性能的非短视快速天基雷达多目标跟踪资源调度算法。

3.1 拉格朗日松弛

优化问题式(17)带有多个约束，因此需要在最小化累积多目标总代价函数的同时满足约束条件。首先，引入拉格朗日乘子向量 $\mathbf{A} = [\lambda_1 \lambda_2 \cdots \lambda_L]^T$ 对式(17)进行拉格朗日松弛，其中， $\lambda_l = [\lambda_{1,l} \lambda_{2,l}] \geq \mathbf{0}$ (注意到，式(8)并不包含动作向量 \mathbf{a} 的任一分量，因此不需要对该约束构造拉格朗日乘子)。然后，构建拉格朗日对偶函数 $J^L(\mathbf{B}_1, \mathbf{A})$ ，即：

$$\begin{aligned} J^L(\mathbf{B}_1, \mathbf{A}) &= \min_{\pi \in \bar{\Pi}} \mathbb{E}^\pi \left[\sum_{l=1}^L \sum_{i=1}^{\mathcal{I}} (c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}) \right. \\ &\quad \left. + \lambda_{1,l} p_{av,i,l} + \lambda_{2,l} \tau_{i,l}/T) \right. \\ &\quad \left. - \sum_{l=1}^L (\lambda_{1,l} E/\Omega + \lambda_{2,l} \eta) \mid \mathbf{B}_1 \right] \\ &= \sum_{i=1}^{\mathcal{I}} \min_{\pi_i \in \bar{\Pi}_i} \mathbb{E}^{\pi_i} \left[\sum_{l=1}^L (c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}) \right. \\ &\quad \left. + \lambda_{1,l} p_{av,i,l} + \lambda_{2,l} \tau_{i,l}/T) \mid \mathbf{b}_{i,1} \right] \\ &\quad \left. - \sum_{l=1}^L (\lambda_{1,l} E/\Omega + \lambda_{2,l} \cdot \eta) \right] \quad (18) \end{aligned}$$

其中， $\bar{\Pi}$ 为LR后的策略集。拉格朗日对偶函数 $J^L(\mathbf{B}_1, \mathbf{A})$ 具有如下性质：

(1) 对于任意 $\mathbf{A} \geq \mathbf{0}$ ， $\mathbf{B}_1 \in \mathcal{B}^{\mathcal{I}}$ ， $J^L(\mathbf{B}_1, \mathbf{A}) \leq V^*(\mathbf{B}_1)$ ；

(2) 当 $\mathbf{A} \geq \mathbf{0}$ ， $J^L(\mathbf{B}_1, \mathbf{A})$ 是分段线性凹的；

因此， $J^L(\mathbf{B}_1, \mathbf{A})$ 可作为性能下界来评估原始优化式(17)可行策略的质量。可计算得到拉格朗日对偶函数最大值，其与 $V^*(\mathbf{B}_1)$ 间隙最小。即求解下列拉格朗日对偶问题，

$$\max_{\mathbf{A} \geq \mathbf{0}} J^L(\mathbf{B}_1, \mathbf{A}) \quad (19)$$

注意到，拉格朗日对偶问题为凸优化问题。

给定拉格朗日乘子 λ_l ，结合式(11)重新定义第 l 次调度时目标 i 的综合代价函数为 $C(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}, \lambda_l) = c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}) + \lambda_{1,l} p_{av,i,l} + \lambda_{2,l} \tau_{i,l}/T$ ，则对于 $\forall \mathbf{B}_1 \in \mathcal{B}^{\mathcal{I}}$ ， $J^L(\mathbf{B}_1, \mathbf{A})$ 可写成：

$$J^L(\mathbf{B}_1, \mathbf{A}) = \sum_{i=1}^{\mathcal{I}} J_i^L(\mathbf{b}_{i,1}, \mathbf{A}) - \sum_{l=1}^L (\lambda_{1,l} E/\Omega + \lambda_{2,l} \eta) \quad (20)$$

其中，各目标的累积总代价函数 $J_i^L(\mathbf{b}_{i,1}, \mathbf{A}) =$

$\max_{\pi_i \in \bar{\Pi}_i} \mathbb{E}^{\pi_i} \left[\sum_{l=1}^L (c(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}) + \lambda_{1,l} p_{av,i,l} + \lambda_{2,l} \tau_{i,l}/T) \mid \mathbf{b}_{i,1} \right]$
 $= \min_{\pi_i \in \bar{\Pi}_i} \mathbb{E}^{\pi_i} \left[\sum_{l=1}^L C(\mathbf{b}_{i,l}, \mathbf{a}_{i,l}, \lambda_l) \mid \mathbf{b}_{i,1} \right]$ 。通过式(20)看到，可通过求解 \mathcal{I} 个单个目标下的动态规划子问题求解拉格朗日对偶问题。

3.2 基于LR-POMCPOW的非短视快速天基雷达多目标跟踪资源调度算法

建立LR后，可基于次梯度法求解拉格朗日对偶问题式(19)。其中， $J_i^L(\mathbf{b}_{i,1}, \mathbf{A})$ 的计算是关键。由2.3节可知，各目标的累积总代价函数优化模型表现出状态空间、动作空间、观测空间连续的特点。但即便是有限时间的POMDP问题也是PSPACE-hard的，这表明离线精确算法不可用于会出现维数灾难的问题，而基于MCTS的在线近似算法已被证明可解决大规模状态空间的POMDP问题^[24]。因此，接下来采用MCTS优化计算 $J_i^L(\mathbf{b}_{i,1}, \mathbf{A})$ ，过程如图2所示。为方便起见，省略区分各目标的下标 i 。具体过程说明如下。

首先在一次蒙特卡罗模拟中，给定初始信念状态 \mathbf{b}_1 作为树的根节点，往下延伸出的子节点表示可选动作 \mathbf{a} ，利用UCB (Upper Confidence Bound) 公式计算各动作节点的UCB值。注意到优化目标是 minimize 累积总代价函数，找出UCB值最小的动作节点并由此节点出发继续向下扩展。但如果信念状态节点存在未探索的动作节点，则从未探索动作节点中随机选取一个节点。在选定的动作节点下获取量测 \mathbf{y} 作为分支，并根据信念状态更新式(7)得到下一时刻的信念状态节点，图中显示为节点 \mathbf{b}_2 。在多目标跟踪中，可通过执行 \mathbf{b}_2 步预测得到下一调度时刻的信念状态。此时若节点 \mathbf{b}_2 存在子节点，则继续重复上述过程直至树的层数达到最大深度 d ；若节点 \mathbf{b}_2 是新扩展的节点，则利用rollout算法，以默认策略向下计算 d 步。模拟结束后，该节点的父节点

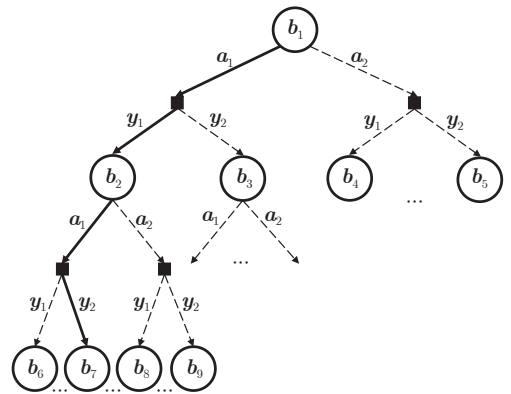


图2 蒙特卡罗树搜索算法的过程

Fig. 2 The process of the MCTS algorithm

直至根节点的路径上所有节点都会根据本次模拟的结果重新计算代价估计值。当达到最大迭代次数后,选择根节点 \mathbf{b}_1 下最优的动作节点 \mathbf{a}^* 作为本次调度结果,同时求出对应量测以及更新的信念状态,并对根节点以下的其余节点及分支进行裁剪。由于优化时长为 \mathcal{L} ,以本次调度的最优动作下更新的信念状态节点为新的根节点,重复 \mathcal{L} 次后结束,则可行策略 π^* 可由根节点通往路径末尾处的叶节点的最优动作集表示。

基于MCTS的常见在线近似算法虽然可行,但分析可知,如果观测空间连续且量测分布函数是给定的,则两次采样得到相同量测的概率为零,因此MCTS的蒙特卡罗模拟永远不会两次通过同一个信念状态节点,并且永远不会构建第一层信念状态节点以下的树。所以,在动作和观测空间很大或连续的情况下,基于MCTS的常见算法将生成非常浅的树,使得渐进最优性能较差。POMCPOW算法利用双渐进加宽可有效解决此问题,双渐进加宽指的是动作空间和观测空间的渐进加宽。与考虑所有动作(量测)的基于MCTS的常见算法相比,POMCPOW算法通过控制有限但逐渐增加的动作(量测)数量以更多地关注后续计算,能够更加深入地向下搜索树。

现具体介绍基于MCTS的POMCPOW算法^[27],其算法结构如**算法1**所示。对应的 h 代表直至当前调度时刻以 \mathbf{b}_1 为初始节点的一段历史记录 $(\mathbf{a}_1, \mathbf{y}_1, \mathbf{a}_2, \mathbf{y}_2, \dots, \mathbf{a}_l, \mathbf{y}_l)$,即已经确定的部分搜索路径。 $h\mathbf{a}$ 和 $h\mathbf{a}\mathbf{y}$ 分别表示在末尾附加了新生成的 \mathbf{a} 和 (\mathbf{a}, \mathbf{y}) 的历史。设定 $d \leq \mathcal{L}$ 。 \mathcal{C} 是节点的子节点集, $\mathcal{C}(h)$ 为信念状态节点的动作子节点集, $\mathcal{C}(h\mathbf{a})$ 为动作节点的量测子节点集, N 是节点访问次数的计数值,对应的 $N(h)$ 为信念状态节点计数值, $N(h\mathbf{a})$ 为动作节点计数值, M 是生成不同量测的计数值。与信念状态节点关联的采样状态集为 X , W 是对应于采样状态集 X 的权重集,而 $Q(h\mathbf{a})$ 是在历史 h 后选取动作 \mathbf{a} 的代价估计值, \mathcal{C} , N , M , X , W 和 Q 都初始化为0或空集。其中,渐进加宽分别体现在**算法1**的14~17行、20~24行。在处理连续动作空间问题时,渐进加宽使用15行的NEXTACTION(h)运算进行处理。具体为,在动作空间 \mathcal{A} 中采用均匀分布采样方法选择一个动作,但选择的动作不与 $\mathcal{C}(h)$ 内的已采样动作重复,直至动作采样空间 $\mathcal{C}(h)$ 大小达到 $\delta_{\mathbf{a}}N(h)^{\alpha_{\mathbf{a}}}$ 。观测空间的渐进加宽同理。 $\mathcal{G}(\cdot)$ 是默认的生成函数,在本文中目标运动模型、雷达量测模型以及综合代价函数的计算, $\Lambda(\cdot)$ 单指具体的综合代价函数计算,所得的 C 即为综合代价值 $C(\mathbf{b}, \mathbf{a}, \lambda)$ 。**算法1**第

算法1 POMCPOW算法

Alg. 1 POMCPOW algorithm

Input: 信念状态 \mathbf{b}_1 , 搜索深度 d , 拉格朗日算子向量 Λ^e , 模拟次数 Γ , 动作空间 \mathcal{A}
Output: 最优策略 π^e

- 1: for $l=1:\mathcal{L}$ do
- 2: for $n=1:\Gamma$ do
- 3: $\mathbf{x} \leftarrow$ 从 \mathbf{b}_l 中采样
- 4: SIMULATE($\mathbf{x}, h, \lambda_l^e, d$)
- 5: end for
- 6: $\mathbf{a}_l^e \leftarrow \arg \min_{\mathbf{a}_l} Q(\mathbf{b}_l, \mathbf{a}_l)$
- 7: 预测 u 步得到下一调度时刻的信念状态 \mathbf{b}_{l+1}
- 8: end for
- 9: return $\pi^e = [\mathbf{a}_1^e \mathbf{a}_2^e \dots \mathbf{a}_{\mathcal{L}}^e]$

10: procedure SIMULATE ($\mathbf{x}, h, \lambda, d$)

- 11: if $d=0$ then
- 12: return 0
- 13: end if
- 14: if $|\mathcal{C}(h)| \leq \delta_{\mathbf{a}}N(h)^{\alpha_{\mathbf{a}}}$ then
- 15: $\mathbf{a} \leftarrow$ NEXTACTION(h)
- 16: $\mathcal{C}(h) \leftarrow \mathcal{C}(h) \cup \{\mathbf{a}\}$
- 17: end if
- 18: $\mathbf{a} \leftarrow \arg \min_{\mathbf{a} \in \mathcal{C}(h)} Q(h\mathbf{a}) - \mu \sqrt{\frac{\log N(h)}{N(h\mathbf{a})}}$
- 19: $\mathbf{x}', \mathbf{y}, C \leftarrow \mathcal{G}(\mathbf{x}, \mathbf{a}, \lambda)$
- 20: if $|\mathcal{C}(h\mathbf{a})| \leq \delta_{\mathbf{y}}N(h\mathbf{a})^{\alpha_{\mathbf{y}}}$ then
- 21: $M(h\mathbf{a}\mathbf{y}) \leftarrow M(h\mathbf{a}\mathbf{y}) + 1$
- 22: else
- 23: 选择 $\mathbf{y} \in \mathcal{C}(h\mathbf{a})$ w.p. $\frac{M(h\mathbf{a}\mathbf{y})}{\sum_{\mathbf{y}} M(h\mathbf{a}\mathbf{y})}$
- 24: end if
- 25: 增加 \mathbf{x}' 至 $X(h\mathbf{a}\mathbf{y})$
- 26: 增加 $\Pr(\mathbf{y}|\mathbf{x}', \mathbf{a})$ 至 $W(h\mathbf{a}\mathbf{y})$
- 27: if $\mathbf{y} \notin \mathcal{C}(h\mathbf{a})$ then
- 28: $\mathcal{C}(h\mathbf{a}) \leftarrow \mathcal{C}(h\mathbf{a}) \cup \{\mathbf{y}\}$
- 29: $C_{\text{total}} \leftarrow \text{ROLLOUT}(\mathbf{x}, h, \lambda, d)$
- 30: else
- 31: 选择 $\mathbf{x}' \in X(h\mathbf{a}\mathbf{y})$ w.p. $\frac{W(h\mathbf{a}\mathbf{y}[i])}{\sum_{j=1}^m W(h\mathbf{a}\mathbf{y}[j])}$
- 32: $C \leftarrow \Lambda(\mathbf{x}, \mathbf{a})$
- 33: $C_{\text{total}} \leftarrow C + \gamma \text{SIMULATE}(\mathbf{x}, h\mathbf{a}\mathbf{y}, \lambda, d-1)$
- 34: end if
- 35: $N(h) \leftarrow N(h) + 1$
- 36: $N(h\mathbf{a}) \leftarrow N(h\mathbf{a}) + 1$
- 37: $Q(h\mathbf{a}) \leftarrow Q(h\mathbf{a}) + \frac{C_{\text{total}} - Q(h\mathbf{a})}{N(h\mathbf{a})}$
- 38: end procedure

29行Rollout算法^[24]部分如**算法2**所示,其中,使用默认的策略 $\pi_{\text{rollout}}(h, \cdot)$ 在整个动作空间 \mathcal{A} 中随机选择一个动作,递归 d 步后返回累积折扣代价值 C_{total} 。

算法2 Rollout算法
Alg. 2 Rollout algorithm

```

1: procedure ROLLOUT( $\mathbf{x}, h, \lambda, d$ )
2:   if  $d=0$  then
3:     return 0
4:   end if
5:    $\mathbf{a} \leftarrow \pi_{\text{rollout}}(h, \cdot)$ 
6:    $\mathbf{x}', \mathbf{y}, C \leftarrow \mathcal{G}(\mathbf{x}, \mathbf{a}, \lambda)$ 
7:   return  $C + \gamma \text{ROLLOUT}(\mathbf{x}', h, \lambda, d-1)$ 
8: end procedure

```

算法1第31行中 m 指采样状态集 X 包含的样本数。与常见MCTS算法不同的是，UCB为

$$\arg \min_{\mathbf{a} \in \mathcal{C}(h)} \mathcal{Q}(h\mathbf{a}) - \mu \sqrt{\frac{\log N(h)}{N(h\mathbf{a})}}$$

在POMCPOW算法中，每一次模拟时转移采样状态 \mathbf{x}' 都会被插入到代表信念状态的加权采样状态集 $X(h\mathbf{a})$ 中，当动作节点的量测子节点数 $|\mathcal{C}(h\mathbf{a})|$ 超出对应渐进加宽上限 $\delta_y N(h\mathbf{a})^{\alpha_y}$ 后，其选择概率取决于权重集 $W(h\mathbf{a})$ 中该状态的量测分布函数之和，如算法1中第31行所示。注意到，信念状态节点包含的样本数与节点被访问的次数有关，

因此，一个信念状态节点信念表示越丰富，最优策略访问经过的可能性就越大^[27]。转移采样状态的加权方法被证明是合理的^[35]，这些采样状态集会随着树的搜索而逐渐改进。随着转移采样状态 \mathbf{x}' 的不断加入，采样状态集 $X(h\mathbf{a})$ 会逐渐扩展，由此改善已有在线算法的性能。

对于给定的天基雷达多目标跟踪任务，给定调度间隔 u 后可确定实际调度总次数为 $K = \lfloor \mathcal{U}/u \rfloor$ 。在第 κ ($\kappa = 1, 2, \dots, K$)次调度时，经过LR并利用POMCPOW分别求得各目标的最优策略 π_i^e 后，需进行拉格朗日乘子向量的更新直至迭代结束。此时选取各目标最优策略的首个动作值，构成本次调度的最优动作向量 $\bar{\pi}_\kappa = [\mathbf{a}_{1,1}^* \ \mathbf{a}_{2,1}^* \ \dots \ \mathbf{a}_{T,1}^*]^\top$ 。对于信念状态的更新，尽管粒子滤波算法比PEKF-VB算法更通用，但其需要更大的计算量，并可能受样本贫化等影响。本文应用PEKF-VB算法得到各目标的信念状态 $\mathbf{b}_{i,\kappa+1}$ 。在经历实际 K 次调度后，可求得多目标跟踪的最优策略 π^* 以及近似最优解 $V^*(\mathbf{B}_1)$ ，因此可将最优策略的分量分别应用于持续时间为 u 的多目标跟踪过程中。基于LR-POMCPOW的非短视快速天基雷达多目标跟踪资源调度算法的整体流程总结如算法3所示。

算法3 基于LR-POMCPOW的天基雷达多目标跟踪资源调度算法
Alg. 3 LR-POMCPOW-based resource scheduling algorithm for multi-target tracking of space-based radar

```

Input: 动作空间 $\mathcal{A}$ ，初始信念状态 $\mathbf{B}_1$ ，最大迭代次数 $e_m$ ，初始迭代步长 $\gamma_{\text{LR}}$ ，模拟次数 $\Gamma$ ，搜索深度 $d$ 
Output: 最优策略 $\pi^*$ ，最优累积多目标总代价值 $V^*(\mathbf{B}_1)$ 
1: 调度次数 $\kappa = 1$ 
2: while  $\kappa \leq K$  do
3:   迭代次数 $e=0$ ，拉格朗日乘子向量初始值设定为 $\Lambda^0 = [\lambda_1^0 \ \lambda_2^0 \ \dots \ \lambda_{\mathcal{L}}^0]^\top$ 
4:   while  $e \leq e_m$  do
5:     for  $i=1:\mathcal{I}$  do
6:       给定信念状态 $\mathbf{b}_{i,\kappa}$ ，搜索深度 $d$ ，拉格朗日算子向量 $\Lambda^e$ ，模拟次数 $\Gamma$ ，动作空间 $\mathcal{A}$ ，转至算法1进行求解，得到目标 $i$ 的最优策略
        $\pi_i^e = [\mathbf{a}_{i,1}^e \ \mathbf{a}_{i,2}^e \ \dots \ \mathbf{a}_{i,\mathcal{L}}^e]$ 
7:     end for
8:     分别计算次梯度 $s_{1,l} = \sum_{i=1}^{\mathcal{I}} p_{\text{av},i,l} - E/\mathcal{U}$ ， $s_{2,l} = \sum_{i=1}^{\mathcal{I}} \tau_{i,l}/T - \eta$ ， $\forall l, 1 \leq l \leq \mathcal{L}$ 
9:     对于 $\forall l, 1 \leq l \leq \mathcal{L}$ ， $s_{1,l}, s_{2,l}$ 等于0或小于给定误差阈值 $\epsilon$ ，则迭代结束，并保存对应的策略 $\bar{\pi} = [\pi_1^* \ \pi_2^* \ \dots \ \pi_{\mathcal{I}}^*]^\top$ ，转至步骤13
10:    更新拉格朗日乘子向量 $\Lambda^e$ ，令 $\lambda_{1,l}^{e+1} = \max\{0, \lambda_{1,l}^e + \gamma_{\text{LR}} \cdot s_{1,l}\}$ ， $\lambda_{2,l}^{e+1} = \max\{0, \lambda_{2,l}^e + \gamma_{\text{LR}} \cdot s_{2,l}\}$ ， $\forall l, 1 \leq l \leq \mathcal{L}$ 
11:    令 $e=e+1$ ，返回至步骤4
12:   end while
13:   选取本次调度各目标策略的首个动作，构成动作向量 $\bar{\pi}_\kappa = [\mathbf{a}_{1,1}^* \ \mathbf{a}_{2,1}^* \ \dots \ \mathbf{a}_{\mathcal{I},1}^*]^\top$ 
14:   当 $\kappa$ 大于 $K$ 时结束迭代，利用PEKF-VB算法执行完剩余更新步，并转至步骤20
15:   for  $i=1:\mathcal{I}$  do
16:     利用PEKF-VB算法执行 $u$ 步更新，得到信念状态 $\mathbf{b}_{i,\kappa+1}$ 
17:   end for
18:   令 $\kappa = \kappa + 1$ 
19: end while
20: 根据最优策略 $\bar{\pi}^* = [\bar{\pi}_1 \ \bar{\pi}_2 \ \dots \ \bar{\pi}_K]$ 计算式(17)的最优累积多目标总代价值 $V^*(\mathbf{B}_1)$ 

```

在计算复杂度方面,单独取第 κ 次调度进行分析,则LR-POMCPOW算法中最为耗时的部分是各子问题最优策略计算以及拉格朗日松弛。首先计算单次迭代求解单个目标的最优策略的复杂度,此处由于算法1第31行状态选取为随机采样,其复杂度为 $\mathcal{O}(\Gamma)$,则各子问题最优策略计算的复杂度为 $\mathcal{O}(d\Gamma^2)$,从而算法3第5-7行计算复杂度为 $\mathcal{O}(d\mathcal{L}\Gamma^2)$ 。而LR的次梯度计算和乘子更新计算复杂度为 $\mathcal{O}(\mathcal{L})$,因此第 κ 次调度的计算复杂度为 $\mathcal{O}(\mathcal{L}e_m + d\mathcal{L}\Gamma^2)$ 。在同样使用LR后,同属于MCTS类型的POMCP算法在求解最优策略时的复杂度为 $\mathcal{O}(d\mathcal{L}(|\mathcal{A}| + |\mathcal{Y}|))$,在动作空间、观测空间很大甚至连续的情况下,空间大小 $|\mathcal{A}|, |\mathcal{Y}|$ 使得POMCP算法计算复杂度远高于POMCPOW算法。而离线算法以PBVI为代表,其计算复杂度为 $\mathcal{O}(\mathcal{L}|\mathcal{A}||\mathcal{Y}|\Gamma)$,相较在线近似算法,其在连续空间问题下需要更多计算资源^[36]。

4 仿真与验证

本节通过仿真验证所提算法的有效性。首先,构建各目标威胁度区分度较高的场景,分析威胁度对天基雷达资源分配的影响,验证所构建优化函数中威胁度的有效性。然后,分析了目标与天基雷达的相对距离对资源分配的影响程度。最后,将现有适用于连续状态空间、连续动作空间、连续观测空间POMDP问题的几种算法与本文算法进行对比,验证本文所用算法的优越性。

4.1 威胁度对雷达资源分配影响分析

考虑运行于圆形轨道的天基雷达卫星,其轨道6根数包括:轨道半长轴、偏心率、轨道倾角、升交点赤经、近地点幅角以及真近点角。为保证坐标转换的准确性需给定格林尼治恒星时角(Greenwich Hour Angle, GHA),雷达天线阵面的偏航角、俯仰角、滚动角都设置为0 rad,对应参数设置如表1所示。雷达在同一时刻需跟踪 $\mathcal{I} = 2$ 个飞机目标,采样间隔设为1 s,各飞机目标以及受保护飞机都处于匀速运动状态,其在STK (Satellite Tool Kit)中运动轨迹如图3所示(数据来源为AGI),图3中箭头所指为各目标运动方向以及天基雷达运动方

向,雷达对目标持续跟踪的轨迹段用白色标记。各目标和受保护飞机在ECEF坐标系下的初始位置和速度,计算雷达回波信噪比 $\text{SNR}_{i,k}$ 时所需的RCS,相对距离、相参积累时间以及发射机平均辐射功率参数如表2所示,参考目标的信噪比为 $\text{SNR}_0 = 15$,方位角波束宽度 ρ_θ 为0.002 rad,俯仰角波束宽度 ρ_ϕ 为0.001 rad。

表1 仿真基本参数设置

Tab. 1 Basic parameter settings of simulation

参数	数值
搜索深度 d	6
模拟次数 Γ	600
状态粒子数 $N_{\text{particles}}$	600
折扣因子 γ	1
脉冲宽度 ν	1 μs
第 k 次调度时初始拉格朗日算子 λ_i^0	[50, 50]
LR最大迭代次数 e_m	50
LR初始迭代步长 γ_{LR}	20
LR误差阈值 ϵ	0.01
最大时间预算比 η	0.5
轨道6根数 ¹	[7400 km, 0, 0.61 rad, 0 rad, 0 rad, 0.84 rad]
格林尼治恒星时角(GHA)	4.98 rad
窗口起始时间 t_{start} (UTCG)	4 May 2023 04:14:43.000
窗口结束时间 t_{end} (UTCG)	4 May 2023 04:19:42.000

¹轨道高度指圆形轨道下的半长轴,即地心与天基雷达卫星之间的距离。

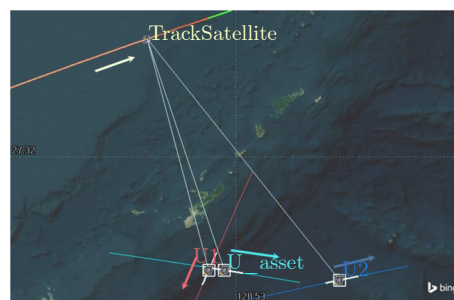


图3 威胁度对雷达资源分配影响分析STK仿真图示

Fig. 3 STK-based demonstration for impact of distinct target threat levels on radar resource allocation

表2 场景1初始时刻目标相关参数

Tab. 2 Parameters related to target initialization of scenario 1

区域内目标	初始位置(km)	初始速度(km/s)	σ (m^2)	r (km)	τ (s)	p_{av} (W)
参考目标	—	—	11	1250.00	0.20	1×10^4
目标1	[-3563.04, 4533.712, 2741.618]	[0.008, 0.109, -0.168]	14	1570.05	0.20	1×10^4
目标2	[-3728.76, 4427.435, 2695.038]	[-0.123, -0.126, 0.037]	15	1572.54	0.20	1×10^4
我方飞机	[-3560.39, 4547.62, 2722.091]	[-0.164, -0.110, -0.031]	—	—	—	—

为简化调度过程，令雷达调度的连续动作量仅选取相参积累时间，其取值范围为 $[0.01, 0.40]$ s，其余分量与参考目标对应值保持一致。采用LR-POMCPOW算法进行资源调度。雷达从窗口起始时刻开始跟踪，调度间隔 $u=70$ s，至窗口结束时刻可进行 $\mathfrak{U} = 300$ 次跟踪，实际调度终止时刻为 $K \cdot u = 280$ s，从而，此仿真场景中雷达调度资源4次。

在PEKF-VB跟踪算法中，飞机目标 $i(i=1, 2)$ 在3个方向上的机动噪声方差 $\sigma_{i,x}^2, \sigma_{i,y}^2, \sigma_{i,z}^2$ 都设定为 $6 \times 10^{-5} \text{ km}^2/\text{s}^4$ ，迭代次数为4，惩罚因子 $\beta = 1$ 。威胁度相关参数为 $g_{\text{time}}=0.3, g_{\text{distance}}=0.7, t_0, t_{0.5}, t_1$ 取值为60, 55, 50, $d_0, d_{0.5}, d_1$ 取值为160, 110, 60，仿真中可添加0.01到威胁度评估函数的0威胁度处，避免不合理取值。另外，LR所用参数及调度算法的部分参数也如表1所示。

为验证威胁度对雷达资源调度的影响，图4给出了整个调度时期两目标的威胁度值。可以看到，与图3中受保护飞机有交叉轨迹的目标1(运动方向为红色箭头)威胁度接近1，而较远处的目标2威胁度保持在较低水平。跟踪过程中分配给两目标的相参积累时间与采样间隔比(即预算比)结果如图5所示。可以看到，目标的威胁度与雷达分配的预算比具有相关性，在满足不大于总预算比 $\eta = 0.5$ 的条件下，需要获得更多关注的目标1始终可分配到较多的相参积累时间资源。若构建优化目标函数时只考虑量测和目标运动的不确定性，会导致雷达将更多资源分配给距离受保护飞机更远的目标，而综合了威胁度的优化目标函数会更为合理地反映真实场景，当

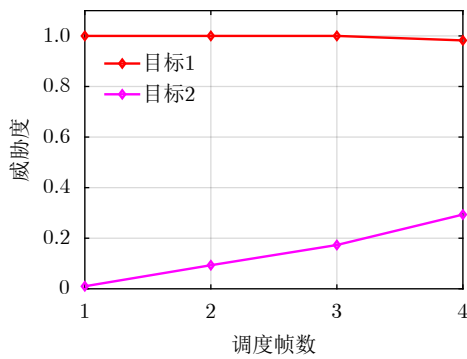


图4 目标威胁度结果

Fig. 4 Target threat level results

目标相对较快或较近地向受保护飞机移动时，雷达资源便会倾向于此类目标。注意到图5给出的资源分配结果呈现出阶跃式现象，这是由于每次资源优化调度计算完成后，优化参数立即输入天基雷达，应用于下一调度间隔的多目标跟踪过程中。

4.2 相对距离对雷达资源分配影响分析

在该仿真场景中，主要研究目标与雷达的相对距离大小以及其变化对雷达资源分配的影响。由于常见天基雷达多目标跟踪场景中目标与沿轨高速运行的天基雷达的相对距离变化并不明显，无法有效体现其对资源分配的影响程度，则在图6中令目标沿特殊轨迹运行。具体参数如下：调度间隔为15 s，跟踪总时长为 $\mathfrak{U} = 124$ s，连续动作量选定为相参积累时间，其取值空间为 $[0.01, 0.40]$ s，初始位置和速度如表3所示，威胁度相关参数 $t_0, t_{0.5}, t_1$ 取值为80, 55, 30； $d_0, d_{0.5}, d_1$ 取值为110, 75, 40；窗口起

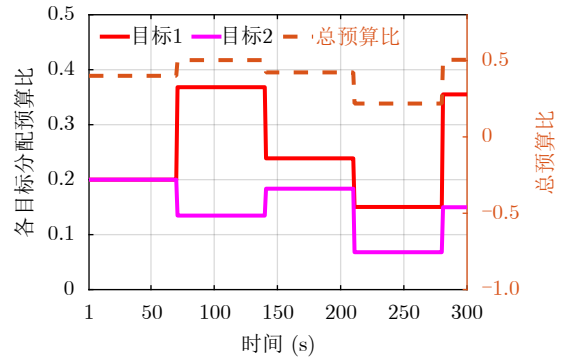


图5 各目标的预算比(τ/T)分配结果

Fig. 5 Budget ratio (τ/T) allocation results for each target

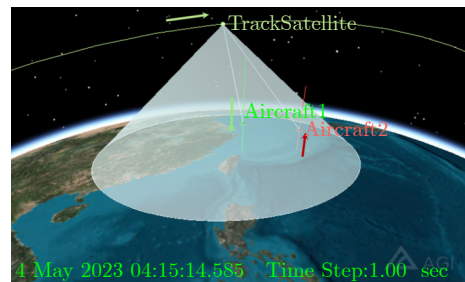


图6 相对距离对雷达资源分配影响分析STK仿真图示

Fig. 6 STK-based demonstration for impact of distinct relative distances on radar resource allocation

表3 场景2初始时刻目标相关参数

Tab. 3 Parameters related to target initialization of scenario 2

区域内目标	初始位置(km)	初始速度(km/s)	σ (m^2)	r (km)	τ (s)	p_{av} (W)
参考目标	—	—	11	900.00	0.20	1×10^4
目标1	$[-3579.26, 4512.89, 2754.72]$	$[0.139, 0.115, -0.008]$	10	750.44	0.25	1×10^4
目标2	$[-3596.33, 4574.503, 2628.813]$	$[-0.055, 0.051, -0.164]$	15	1413.93	0.25	1×10^4

始时间为4 May 2023 04:14:09.500, 窗口结束时间为4 May 2023 04:16:12.000, 其余参数同场景1一致。

目标径向距离变化如图7所示。可以看到, 跟踪阶段初期两目标逐渐靠近雷达, 两者与雷达的相对距离差别大, 但变化程度较为缓慢。为突出相对距离对雷达资源分配的影响, 该场景中的优化目标函数中不考虑威胁度以及能量代价, 只存在量测和目标运动不确定性的作用。通过图8中分配给两目标的相参积累时间与采样间隔比可以看到, 开始阶段两目标在向雷达方向运动, 相对距离都在减小, 但目标2比目标1更远, 对应的信噪比明显小于目标1,

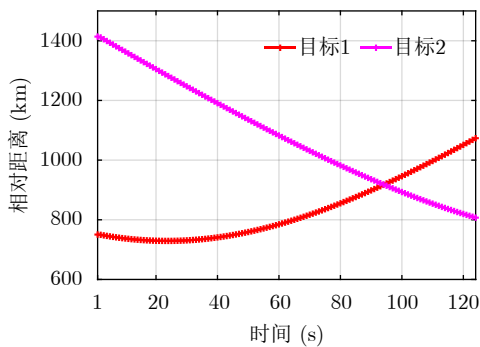


图7 目标径向距离

Fig. 7 The slant range of targets

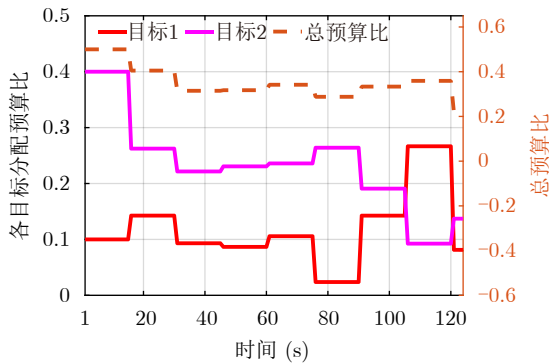


图8 相对距离影响下预算比(τ/T)分配结果

Fig. 8 Budget ratio (τ/T) allocation results influenced by relative distance

经过算法调度可以使目标2获得持续增加的资源。但随着目标在第80~100 s时段内距雷达的相对距离大小发生转变后, 可以发现资源会偏向于目标1, 但变化程度较为缓慢。由于总预算比的限制, 雷达会不断减少目标2的资源配置, 但总体保持分配资源多于目标1的状态。由此, 可以说明目标与雷达的相对距离越大, 雷达分配的资源越多, 但目标与雷达的相对距离呈现增长趋势, 雷达资源则会逐渐倾向于分配给此类目标。

4.3 算法性能比较分析

为说明LR-POMCPOW算法的优越性, 首先构建包括 $\mathcal{I} = 4$ 个飞机目标以及受保护舰船的跟踪场景, 如图9所示。在保持多数参数设置与场景1相同的条件下, 其余参数设置如下: 调度间隔为40 s, 跟踪总时长为 $\mathcal{U} = 244$ s, 分配至各目标的相参积累时间取值范围为 $[0.01, 0.30]$ s, 针对每个目标的雷达平均辐射功率取值范围为 $[5 \times 10^3, 2 \times 10^4]$ W, 最大时间预算比为 $\eta = 1$, 最大消耗能量值为 $E = 1.5 \times 10^7$ J, 各目标平均辐射功率之和的阈值为 $E/\mathcal{U} = 6.15 \times 10^4$ W; $t_0, t_{0.5}, t_1$ 取值为180, 140, 100; $d_0, d_{0.5}, d_1$ 取值为160, 130, 100; GHA为4.97 rad; 窗口起始时间为4 May 2023 04:12:26.000, 窗口结束时间为4 May 2023 04:16:28.185, 目标初始位置和速度如表4所示。综合代价函数中, $\omega_1 = 0.7$,

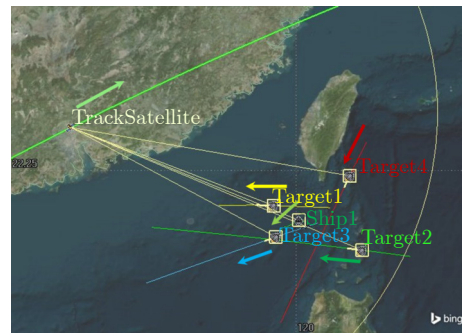


图9 多目标跟踪的STK仿真图示

Fig. 9 STK-based demonstration for multi-target tracking

表4 场景3初始时刻目标相关参数

Tab. 4 Parameters related to target initialization of scenario 3

区域内目标	初始位置(km)	初始速度(km/s)	σ (m ²)	r (km)	τ (s)	p_{av} (W)
参考目标	—	—	11	1300.00	0.20	1.00×10^4
目标1	$[-2924.49, 5193.69, 2294.53]$	$[0.201, 0.111, 0.008]$	10	1325.83	0.15	1.42×10^4
目标2	$[-3192.74, 5100.13, 2155.14]$	$[0.173, 0.098, 0.025]$	15	1559.51	0.12	1.60×10^4
目标3	$[-2947.68, 5222.76, 2197.083]$	$[0.137, 0.102, -0.057]$	15	1353.37	0.20	1.10×10^4
目标4	$[-3109.56, 5044.01, 2392.48]$	$[0.028, 0.081, -0.135]$	12	1480.35	0.23	1.50×10^4
我方舰船	$[-2992.27, 5162.81, 2244.70]$	$[0.010, 0.009, -0.007]$	—	—	—	—

$\omega_2 = 0.3$ ，避免资源分配过于倾向量级较大的能量代价。

图10给出了经过雷达资源调度的各目标跟踪轨迹对比情况，其中，各目标起始位置以空心圆标注，红色箭头所指为目标运动方向。可以看到，估计轨迹与真实轨迹经过一段时间后基本重合，经过资源分配后天基雷达能够保持良好的跟踪效果。相参积累时间(预算比)和平均辐射功率分配结果分别如图11、图12所示。可以看到，各目标的资源总量可以保持在约束值以内，总预算比在0.4~0.9，未

超出最大总预算比 $\eta = 1$ ；总平均辐射功率在 $2 \times 10^4 \sim 6 \times 10^4$ W，小于各目标平均辐射功率之和的阈值 $E/\mu = 6.15 \times 10^4$ W。由于优化目标函数综合考虑了威胁度，而且各目标与雷达的相对距离变化较小，图中并未出现某一目标长期占据大多数资源的情况，处于合理的取值范围内。

将LR-POMCPOW算法与其他3种结合了LR的算法进行比较。蒙特卡罗仿真次数为100次，仿真相关的超参数参见表5。

(1) POMCPDPW^[27]：此算法与本文所用算法

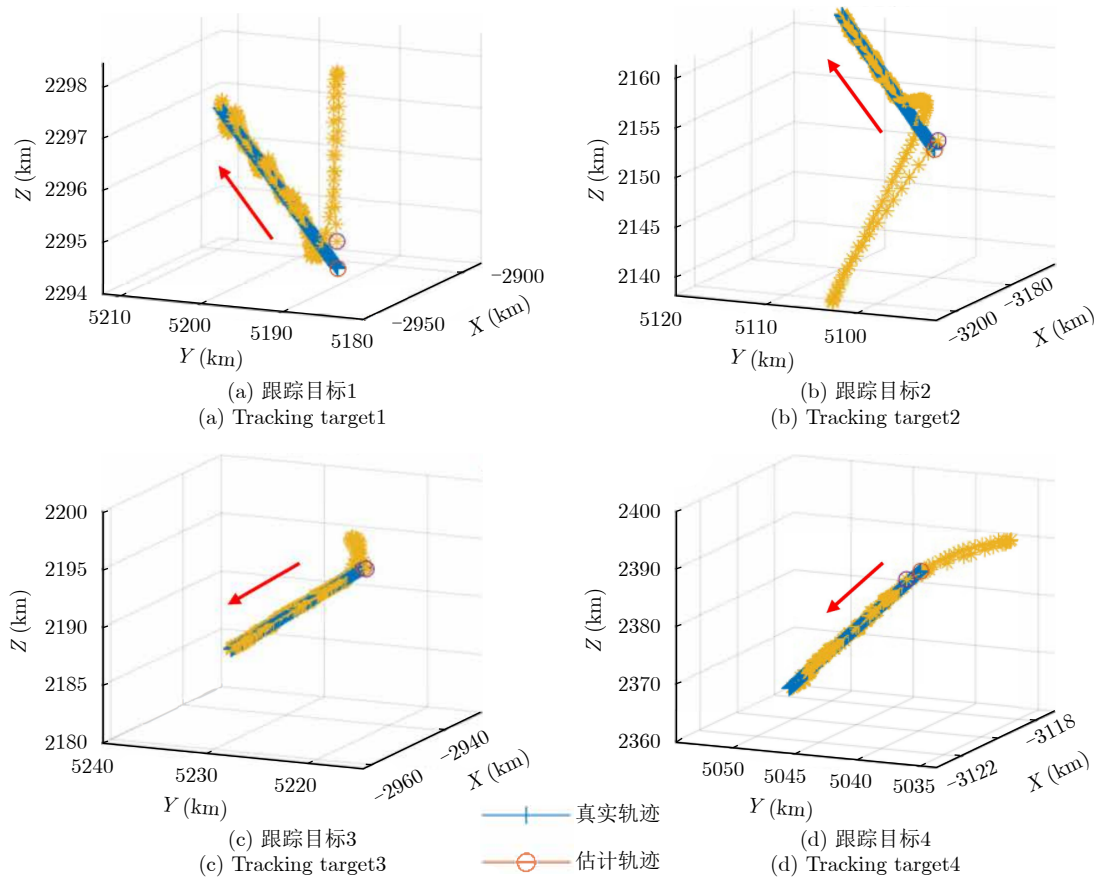


图 10 目标跟踪轨迹

Fig. 10 Target tracking trajectory

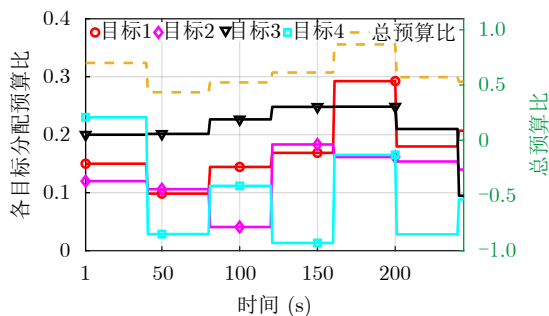


图 11 多目标跟踪下各目标的预算比(τ/T)分配结果

Fig. 11 Budget ratio (τ/T) allocation results of each target under multi-target tracking

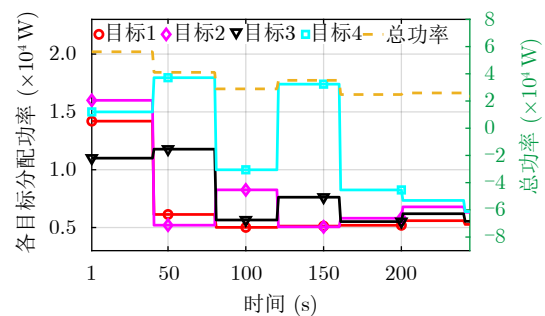


图 12 各目标分配平均辐射功率

Fig. 12 The average radiation power allocated to each target

类似,但区别是,对于每个生成的历史,只有一个状态插入采样状态集中。

(2) POMCP^[24]: 此算法不考虑连续动作以及观测值的渐进加宽,仿真中需先离散化观测空间。

(3) Rollout^[37]: 此算法采用预期未来的蒙特卡罗样本,随机探索可能的未来动作和相应的代价,候选动作在第1步选取,后续步的动作通过给定的基本策略选取。

现以目标1为分析对象,图13、图14给出了LR-POMCPOW与上述3种算法在跟踪目标1时的均方根误差(Root Mean Squared Error, RMSE)对比图。由于代价函数综合考虑了威胁度以及能量代价的因素,而RMSE仅为目标量测和机动性的度量指标,在此发现, RMSE并非与算法性能有直接关

联性,但可以看到在应用各算法后目标跟踪的RMSE值基本保持在较低水平,相差并不明显。

图15、图16给出了不同算法的多目标总平均辐射功率和总预算比的对比结果。LR-POMCPOW算法随着资源调度的进行会不断减少平均辐射功率的消耗, POMCPDPW算法次之, POMCP, Rollout算法使消耗功率维持在较高水平。各算法得到的总预算比与RMSE结果类似,其通过影响SNR的计算进而影响状态误差协方差矩阵迹的求取,此并非代价函数的唯一考虑指标,可以看到,所提算法并非一直保持总预算比最大,各算法结果并无显著差异。各算法在资源调度期间的期望累积多目标总折扣代价如图17所示。所提算法可以得到最小期望累积多目标总折扣代价值,而其他3种

表 5 各算法超参数
Tab. 5 Algorithm hyperparameters

比较的算法	μ	δ_a	α_a	δ_y	α_y	M_r
POMCPOW	100	35	1/100	8	1/120	—
POMCPDPW	30	3	1/30	5	1/55	—
POMCP	70	—	—	—	—	—
Rollout	—	—	—	—	—	30

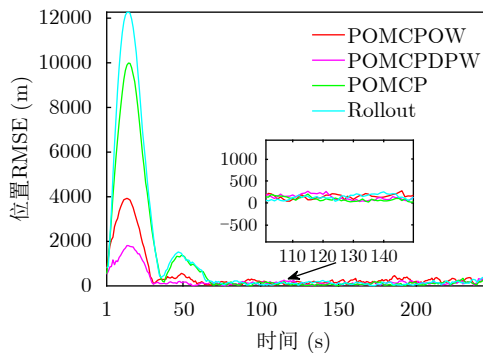


图 13 目标1跟踪位置RMSE
Fig. 13 RMSE on position of target1

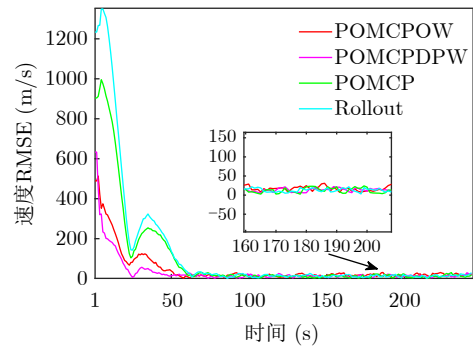


图 14 目标1跟踪速度RMSE
Fig. 14 RMSE on speed of target1

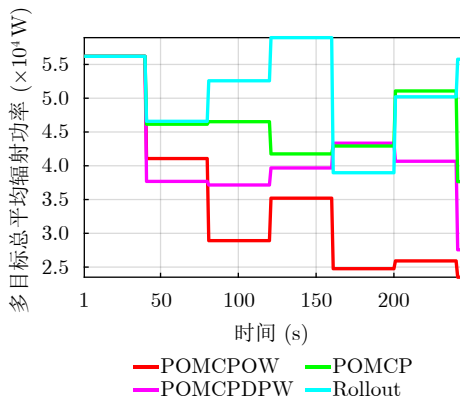


图 15 总平均辐射功率对比
Fig. 15 Comparison of total average radiation power

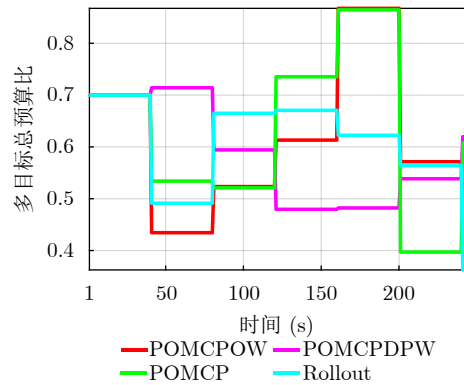


图 16 总预算比对比
Fig. 16 Comparison of the total budget ratio (τ/T)

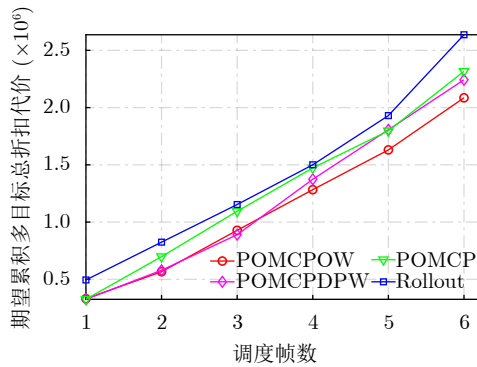


图 17 各算法期望累积多目标总折扣代价对比

Fig. 17 Comparison of the expected cumulative multi-target discount cost

算法的期望累积多目标总折扣代价值的增长趋势类似，其中Rollout算法始终保持最大总折扣代价，而POMCP与POMCPDPW算法略优于此算法，但均高出所提算法，即所提算法可达到资源调度的近似最优。综合考虑选用LR-POMCPOW算法可有效解决天基雷达资源调度问题，其性能在所比较的算法中也最为优越。

5 结语

本文考虑了天基雷达多目标跟踪过程中连续状态空间、连续动作空间以及连续观测空间下的资源调度问题，给出了基于LR-POMCPOW的综合多指标性能的非短视快速天基雷达多目标跟踪资源调度算法。针对天基雷达多目标跟踪问题，以最小化综合考虑了目标威胁度、跟踪精度与低截获概率的代价函数为优化目标，在满足雷达平台运动特性，雷达时间和能量资源的约束条件下，通过优化相参积累时间和平均辐射功率，有效保证雷达在多目标跟踪下资源分配的近似最优性。仿真结果表明，所设置的目标函数适用于天基雷达多目标跟踪场景，与已有几种调度算法相比，所提算法性能更好。

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参考文献

- [1] XIE Mingchi, YI Wei, KIRUBARAJAN T, *et al.* Joint node selection and power allocation strategy for multitarget tracking in decentralized radar networks[J]. *IEEE Transactions on Signal Processing*, 2018, 66(3): 729–743. doi: [10.1109/TSP.2017.2777394](https://doi.org/10.1109/TSP.2017.2777394).
- [2] DAI Jinhui, YAN Junkun, WANG Penghui, *et al.* Optimal resource allocation for multiple target tracking in phased array radar network[C]. 2019 International Conference on Control, Automation and Information Sciences (ICCAIS), Chengdu, China, 2019: 1–4. doi: [10.1109/ICCAIS46528.2019.9074602](https://doi.org/10.1109/ICCAIS46528.2019.9074602).
- [3] SUN Jun, LU Xiujuan, YUAN Ye, *et al.* Resource allocation for multi-target tracking in multi-static radar systems with imperfect detection performance[C]. 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 2020: 1–6. doi: [10.1109/RadarConf2043947.2020.9266340](https://doi.org/10.1109/RadarConf2043947.2020.9266340).
- [4] ZHANG Haowei, LIU Weijian, ZONG Binfeng, *et al.* An efficient power allocation strategy for maneuvering target tracking in cognitive MIMO radar[J]. *IEEE Transactions on Signal Processing*, 2021, 69: 1591–1602. doi: [10.1109/TSP.2020.3047227](https://doi.org/10.1109/TSP.2020.3047227).
- [5] LU Xiujuan, YI Wei, and KONG Lingjiang. Joint online route planning and resource optimization for multitarget tracking in airborne radar systems[J]. *IEEE Systems Journal*, 2022, 16(3): 4198–4209. doi: [10.1109/JSYST.2021.3116020](https://doi.org/10.1109/JSYST.2021.3116020).
- [6] SHI Chenguang, ZHOU Jianjiang, and WANG Fei. Adaptive resource management algorithm for target tracking in radar network based on low probability of intercept[J]. *Multidimensional Systems and Signal Processing*, 2018, 29(4): 1203–1226. doi: [10.1007/s11045-017-0494-8](https://doi.org/10.1007/s11045-017-0494-8).
- [7] SHI Chenguang, WANG Yijie, SALOUS S, *et al.* Joint transmit resource management and waveform selection strategy for target tracking in distributed phased array radar network[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2022, 58(4): 2762–2778. doi: [10.1109/TAES.2021.3138869](https://doi.org/10.1109/TAES.2021.3138869).
- [8] CHHETRI A S, MORRELL D, and PAPANDREOU-SUPPAPPOLA A. Energy efficient target tracking in a sensor network using non-myopic sensor scheduling[C]. 2005 7th International Conference on Information Fusion, Philadelphia, USA, 2005: 558–565. doi: [10.1109/ICIF.2005.1591904](https://doi.org/10.1109/ICIF.2005.1591904).
- [9] HERO A O and COCHRAN D. Sensor management: Past, present, and future[J]. *IEEE Sensors Journal*, 2011, 11(12): 3064–3075. doi: [10.1109/JSEN.2011.2167964](https://doi.org/10.1109/JSEN.2011.2167964).
- [10] FERRI G, MUNAFÒ A, GOLDHAHN R, *et al.* A non-myopic, receding horizon control strategy for an AUV to track an underwater target in a bistatic sonar scenario[C]. 53rd IEEE Conference on Decision and Control, Los Angeles, USA, 2014: 5352–5358. doi: [10.1109/CDC.2014.7040226](https://doi.org/10.1109/CDC.2014.7040226).
- [11] JI Shihao, PARR R, and CARIN L. Nonmyopic multiaspect sensing with partially observable Markov decision processes[J]. *IEEE Transactions on Signal Processing*, 2007, 55(6): 2720–2730. doi: [10.1109/TSP.2007.893747](https://doi.org/10.1109/TSP.2007.893747).
- [12] KRISHNAMURTHY V and DJONIN D V. Optimal threshold policies for multivariate POMDPs in radar resource management[J]. *IEEE Transactions on Signal*

- Processing*, 2009, 57(10): 3954–3969. doi: [10.1109/TSP.2009.2022915](https://doi.org/10.1109/TSP.2009.2022915).
- [13] JIANG Xiaofeng, ZHOU Feng, JIAN Yang, *et al.* An optimal POMDP-based anti-jamming policy for cognitive radar[C]. 2017 13th IEEE Conference on Automation Science and Engineering (CASE), Xi'an, China, 2017: 938–943. doi: [10.1109/COASE.2017.8256224](https://doi.org/10.1109/COASE.2017.8256224).
- [14] SHAN Ganlin, XU Gongguo, and QIAO Chenglin. A non-myopic scheduling method of radar sensors for maneuvering target tracking and radiation control[J]. *Defence Technology*, 2020, 16(1): 242–250. doi: [10.1016/j.dt.2019.10.001](https://doi.org/10.1016/j.dt.2019.10.001).
- [15] SCHÖPE M I, DRIESSEN H, and YAROVY A. A constrained POMDP formulation and algorithmic solution for radar resource management in multi-target tracking[J]. *ISIF Journal of Advances in Information Fusion*, 2021, 16(1): 31–47.
- [16] HAWKINS J T. A Lagrangian decomposition approach to weakly coupled dynamic optimization problems and its applications[D]. [Ph.D. dissertation], Massachusetts Institute of Technology, 2003.
- [17] CASTANON D A. Approximate dynamic programming for sensor management[C]. The 36th IEEE Conference on Decision and Control, San Diego, USA, 1997: 1202–1207. doi: [10.1109/CDC.1997.657615](https://doi.org/10.1109/CDC.1997.657615).
- [18] LI Yuan, ZHU Huayong, and SHEN Lincheng. The Lagrangian relaxation based resources allocation methods for air-to-ground operations under uncertainty circumstances[C]. 2009 Chinese Control and Decision Conference, Guilin, China, 2009: 5609–5614. doi: [10.1109/CCDC.2009.5195197](https://doi.org/10.1109/CCDC.2009.5195197).
- [19] KURNIAWATI H, HSU D, and LEE W S. SARSOP: Efficient Point-based POMDP Planning by Approximating Optimally Reachable Belief Spaces[M]. BROCK O, TRINKLE J, and RAMOS F. Robotics: Science and Systems. Cambridge: MIT Press, 2009: 1–8. doi: [10.15607/RSS.2008.IV.009](https://doi.org/10.15607/RSS.2008.IV.009).
- [20] PINEAU J, GORDON G, and THRUN S. Point-based value iteration: An anytime algorithm for POMDPs[C]. The 18th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003: 1025–1030.
- [21] SPAAN M T J and VLASSIS N. Perseus: Randomized point-based value iteration for POMDPs[J]. *Journal of Artificial Intelligence Research*, 2005, 24: 195–220. doi: [10.1613/jair.1659](https://doi.org/10.1613/jair.1659).
- [22] SMITH T and SIMMONS R. Heuristic search value iteration for POMDPs[C]. The 20th Conference on Uncertainty in Artificial Intelligence, Banff, Canada, 2004: 520–527.
- [23] ROSS S, PINEAU J, PAQUET S, *et al.* Online planning algorithms for POMDPs[J]. *Journal of Artificial Intelligence Research*, 2008, 32: 663–704. doi: [10.1613/jair.2567](https://doi.org/10.1613/jair.2567).
- [24] SILVER D and VENESS J. Monte-Carlo planning in large POMDPs[C]. The 23rd International Conference on Neural Information Processing Systems, Vancouver, British, 2010: 2164–2172.
- [25] YE Nan, SOMANI A, HSU D, *et al.* DESPOT: Online POMDP planning with regularization[J]. *Journal of Artificial Intelligence Research*, 2017, 58: 231–266. doi: [10.1613/jair.5328](https://doi.org/10.1613/jair.5328).
- [26] KURNIAWATI H and YADAV V. An online POMDP Solver for Uncertainty Planning in Dynamic Environment[M]. INABA M and CORKE P. Robotics Research: The 16th International Symposium ISRR. Cham, Switzerland: Springer, 2016: 611–629. doi: [10.1007/978-3-319-28872-7_35](https://doi.org/10.1007/978-3-319-28872-7_35).
- [27] SUNBERG Z and KOCHENDERFER M. Online algorithms for POMDPs with continuous state, action, and observation spaces[C]. The Thirty-Third International Conference on Automated Planning and Scheduling, Delft, The Netherlands, 2018: 259–263. doi: [10.1609/icaps.v28i1.13882](https://doi.org/10.1609/icaps.v28i1.13882).
- [28] KERSHAW D J and EVANS R J. Optimal waveform selection for tracking systems[J]. *IEEE Transactions on Information Theory*, 1994, 40(5): 1536–1550. doi: [10.1109/18.333866](https://doi.org/10.1109/18.333866).
- [29] SIRA S P, PAPANDREOU-SUPPAPPOLA A, and MORRELL D. Dynamic configuration of time-varying waveforms for agile sensing and tracking in clutter[J]. *IEEE Transactions on Signal Processing*, 2007, 55(7): 3207–3217. doi: [10.1109/TSP.2007.894418](https://doi.org/10.1109/TSP.2007.894418).
- [30] LI Xi, CHENG Ting, SU Yang, *et al.* Joint time-space resource allocation and waveform selection for the collocated MIMO radar in multiple targets tracking[J]. *Signal Processing*, 2020, 176: 107650. doi: [10.1016/j.sigpro.2020.107650](https://doi.org/10.1016/j.sigpro.2020.107650).
- [31] KOCH W. Adaptive parameter control for phased-array tracking[C]. SPIE 3809, Signal and Data Processing of Small Targets 1999, Denver, USA, 1999: 444–455. doi: [10.1117/12.364042](https://doi.org/10.1117/12.364042).
- [32] KATSILIERIS F, DRIESSEN H, and YAROVY A. Threat-based sensor management for target tracking[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2015, 51(4): 2772–2785. doi: [10.1109/TAES.2015.140052](https://doi.org/10.1109/TAES.2015.140052).
- [33] HU Yumei, WANG Xuezi, LAN Hua, *et al.* An iterative nonlinear filter using variational Bayesian optimization[J]. *Sensors*, 2018, 18(12): 4222. doi: [10.3390/s18124222](https://doi.org/10.3390/s18124222).
- [34] 何子述, 程子扬, 李军, 等. 集中式MIMO雷达研究综述[J]. *雷达学报*, 2022, 11(5): 805–829. doi: [10.12000/JR22128](https://doi.org/10.12000/JR22128).
HE Zishu, CHENG Ziyang, LI Jun, *et al.* A survey of collocated MIMO radar[J]. *Journal of Radars*, 2022, 11(5): 805–829. doi: [10.12000/JR22128](https://doi.org/10.12000/JR22128).
- [35] LIM M H, TOMLIN C J, and SUNBERG Z N. Sparse tree search optimality guarantees in POMDPs with continuous

- observation spaces[C]. Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 2020: 1–16. doi: [10.24963/ijcai.2020/572](https://doi.org/10.24963/ijcai.2020/572).
- [36] JI Shihao, PARR R, LI Hui, *et al.* Point-based policy iteration[C]. The Twenty-Second National Conference on Artificial Intelligence, Vancouver, British, 2007: 1243–1249.
- [37] SCHÖPE M I, DRIESSEN H, and YAROVY A. Multi-task sensor resource balancing using Lagrangian relaxation and policy rollout[C]. 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 2020: 1–8. doi: [10.23919/FUSION45008.2020.9190546](https://doi.org/10.23919/FUSION45008.2020.9190546).

作者简介

王增福，博士，副教授，主要研究方向为信息融合、传感器管理、路径规划等。

杨广宇，硕士生，主要研究方向为天基雷达资源调度。

金术玲，博士，高级工程师，主要研究方向为雷达总体设计等。

(责任编辑：高山流水)