

# 基于 POMDP 的主动雷达制导弹干扰措施优化方法

张 翔<sup>1</sup>, 吴 华<sup>1</sup>, 陈 游<sup>1</sup>, 宋海方<sup>2</sup>, 周一鹏<sup>1</sup>

(1. 空军工程大学航空工程学院, 西安, 710038; 2. 92925 部队, 山西长治, 046011)

**摘要** 以主动雷达弹对抗为背景,研究了单平台机载电子对抗资源的调度问题。为解决导弹导引头状态观测不确定性问题,引入部分可观测马尔科夫决策过程(POMDP)理论,分析了 POMDP 模型的要素,建立了针对主动雷达制导弹的干扰资源优化 POMDP 模型,并利用最优策略可达空间的连续近似法(SARSOP)进行求解。将电子支援措施(ESM)的观测值输出与电子干扰措施(ECM)的优化策略相结合,提出了一种基于 POMDP 的电子对抗资源实时调度优化算法。仿真结果表明,该方法能够实现目标状态不确定情况下的干扰资源有效调度,相比现有的调度策略,可以使导引头更多维持在搜索状态,更少停留在稳定跟踪状态。

**关键词** 主动雷达制导弹;干扰资源优化;部分可观测马尔科夫决策过程;最优策略可达空间的连续近似法

**DOI** 10.3969/j.issn.1009-3516.2018.05.016

**中图分类号** TN958.97 **文献标志码** A **文章编号** 1009-3516(2018)04-0090-07

## Optimization of Countermeasure Method Against Active Radar Homing Missiles Based on POMDP

ZHANG Xiang<sup>1</sup>, WU Hua<sup>1</sup>, CHEN You<sup>1</sup>, SONG Haifang<sup>2</sup>, ZHOU Yipeng<sup>1</sup>

(1. Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China;

2. Unit 92925, Changzhi 046011, Shanxi, China)

**Abstract:** Taking an active radar missile countermeasure as background, this paper studies the single-platform airborne electronic countermeasure (ECM) scheduling problem. In order to solve the uncertainty problem of missile seeker state observation, the partially observable Markov decision process (POMDP) theory is introduced. The elements of POMDP model are analyzed, an optimization model of ECM resources is established for the active radar homing missile, and the solution is solved by successive approximations of the reachable space under optimal policies (SARSOP). A new real-time scheduling optimization algorithm for ECM resources based on POMDP is presented, combined the output of electronic support measure (ESM) with the control strategy of ECM together. The simulation results indicate that the proposed algorithm can realize effective scheduling of ECM resources under the situation where the target state is uncertain compared with the existing approach, the radar missile will stay in search state more and stable tracking state less.

**Key words:** active radar homing missile; counter resources scheduling optimization; partially observable Markov decision process; successive approximations of the reachable space under optimal policies

**收稿日期:** 2017-11-20

**基金项目:** 航空科学基金(20152096019)

**作者简介:** 张翔(1994—),男,湖南涟源人,硕士生,主要从事电子对抗理论与技术研究。E-mail:cmaristotle@163.com

**引用格式:** 张翔,吴华,陈游,等:基于 POMDP 的主动雷达制导弹干扰措施优化方法[J].空军工程大学学报(自然科学版),2018,19(5):90-96. ZHANG Xiang, WU Hua, CHEN You, et al: Optimization of Countermeasure Method Against Active Radar Homing Missiles Based on POMDP[J]. Journal of Air Force Engineering University (Natural Science Edition), 2018, 19(5): 90-96.

现代空战中主动雷达制导弹是飞机的致命威胁,运用电子对抗技术有效保护载机安全已成为对抗雷达主动弹特别是导引头的重要手段<sup>[1]</sup>。然而,在实战过程中受各种不确定性因素如战场电磁环境、导引头类型、状态信息等的影响,飞行员需要将精力集中在周围战场态势和飞机驾驶上,导致盲目使用电子干扰,难以起到较好的干扰效果。在这样的情况下,构建高效、智能、自适应的 ECM 调度系统尤为重要。

结合遗传算法、蚁群算法等智能算法提出了不同准则下的干扰资源调度方法,虽然提升了 ECM 系统的智能化水平,但都建立在对敌雷达状态识别的基础上。而作为非合作方,雷达状态识别具有极大的不确定性。文献[3]组建了智能干扰决策系统,干扰机能够根据雷达状态和干扰库实时选择最佳干扰样式。但是仅针对当前状态最佳考虑的干扰样式是一种短视的策略,从长远来看,效果通常不会很好。文献[4~5]将博弈理论运用到雷达对抗过程,考虑不同 ECM 措施的效果,提出了自适应的雷达干扰方法,试图找出最佳 ECM 策略。然而,雷达的一些抗干扰措施会让博弈论模型失效,比如说,雷达在发现干扰存在后,可能转入被动跟踪状态,这时干扰机反而成了暴露的一方<sup>[6]</sup>。以上研究主要针对火控雷达,对于雷达导引头的研究甚少。导弹飞行速度极快,导引头作用距离近,这对电子对抗系统的实时性要求更高,如果干扰不能快速高效的作用于导引头,飞机很可能仍难以摆脱导弹的跟踪。

针对上述不足,本文提出了一种基于部分可观测马尔科夫决策过程(Partially Observable Markov Decision Process, POMDP)的 ECM 调度方法。

## 1 问题分析

考虑典型作战场景:装有自卫电子对抗设备的战斗机在执行某一项任务的过程中受到敌主动雷达制导弹威胁。为了确保战机自身安全及任务完成,战斗机在采取机动的同时会辅以干扰措施以免被击落。主动雷达制导弹和战斗机电子对抗系统交互关系见图 1。

主动雷达制导弹发射后,导引头开机对指定区域辐射雷达脉冲并接收目标回波,以实现目标的搜索、截获、跟踪。而机载自卫电子对抗系统则是根据侦收到的雷达脉冲评估其状态,然后选择合适

的干扰样式发射干扰信号。这样一个交互过程存在着极大的不确定性:

1) 侦收雷达脉冲的不确定性:受各类噪声的影响,ESM 系统本身存在一定虚警概率( $P_f$ ),且截获概率( $P_d$ )也难以达到 100%;

2) 雷达状态的不确定性:雷达脉冲的各特征参数难以准确识别,不同状态下辐射的雷达脉冲串可能完全一致,使得 ESM 系统难以准确评估雷达状态;

3) 干扰效能的不确定性:干扰效能的好坏直接作用在雷达导引头上,作为非合作方的电子对抗系统只能间接通过下一时刻的雷达状态来判断。

本文所提出的基于 POMDP 的调度方法能够很好地消除这些影响。POMDP 理论中提出的信念状态能克服雷达信号侦收及评估带来的不确定性;POMDP 模型将 ESM 系统的输出同 ECM 系统的控制相结合,实时调整干扰措施,优化整体干扰效能。同时,POMDP 的方法考虑的是长期的回报,相对于只考虑当前最大回报的贪婪方法,更加高效。

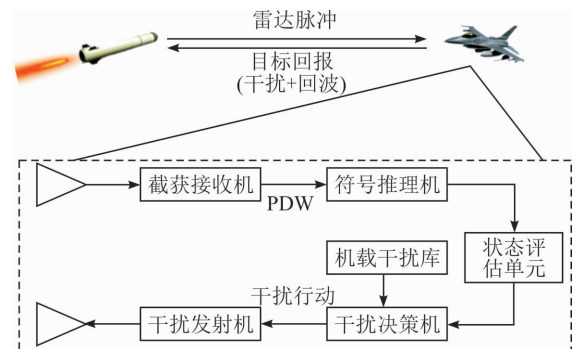


图 1 ECM 与主动雷达制导弹交互关系

Fig. 1 Interactive relationship between ECM system and active radar homing missile

## 2 POMDP 理论

### 2.1 POMDP 模型

POMDP 作为马尔科夫决策过程(Markov Decision Process, MDP)在不确定条件下的延伸,在决策问题中得到了广泛的应用。通常将 POMDP 模型定义为六元组  $\langle S, A, Z, T, O, R \rangle$ :  $S$  为环境的状态集,  $s \in S$ , 其集合元素是部分可观测的, 用信念状态  $b(s)$  表示智能体处于状态  $s$  的概率;  $A$  为智能体的行为集,  $a \in A$ ;  $Z$  表示观测值的集合,  $z \in Z$ ;  $T: S \times A \rightarrow P(s', s, a)$  为状态转移函数, 表示智能体在采取行动  $a$  后环境由状态  $s$  转移到  $s'$  的概率;  $O: S \times A \rightarrow O(z, s', a)$  表示智能体经过行动  $a$  后, 处在状态  $s'$

时得到观测值  $z$  的概率;  $R: S \times A \rightarrow R(s, a)$  为智能体采取行动  $a$  后的奖赏函数。确定 POMDP 模型的各个参数后,即可对模型求解。

## 2.2 值函数与最优策略

POMDP 的解被称为策略,描述为状态集合到行动集合的映射,即  $\pi: S \rightarrow A$ 。对于任何一个策略,都可以用执行这个策略所获得的长期期望回报的大小来评价其优劣,而求解的目的就是要使执行策略所得长期期望回报值最大。定义值函数  $V$ ,智能体在状态  $s$  采取策略  $\pi$  的期望回报为:

$$V^\pi(s) = E \left[ \sum_{t=0}^{\infty} \gamma^t \sum_{s \in S} b_t(s) R(s_t, a_t^\pi) \right] \quad (1)$$

式中:  $\gamma$  为折扣因子,  $0 < \gamma < 1$ 。值函数的本质是期望回报的累积,因此可以用递归的方式描述:

$$V^\pi(s) = \sum_{a \in A} [R(s, a) + \gamma \sum_{s' \in S} b(s') P(s', s, a) V^\pi(s')] \quad (2)$$

定义中间变量,动作值函数  $Q^\pi(s, a)$  表示智能体在  $s$  状态下采取行动  $a$ ,而其他状态采取策略  $\pi$  的值函数:

$$Q^\pi(s, a) = R(s, a) + \gamma \sum_{s' \in S} b(s') P(s', s, a) V^\pi(s') \quad (3)$$

此时,策略  $\pi$  可以计算为:

$$\pi = \arg \max_{a \in A} Q^\pi(s, a) \quad (4)$$

最优策略的求解可以通过对非最优策略改进得到:将策略选择的动作改为当前最优的动作,显然,这样的改进能够改善策略,而值函数对于每个策略点的改进都是单增的<sup>[11]319-339</sup>,求解的过程可以通过这样不断改进完成,直到策略收敛、不再发生改变。此时对于任意  $\epsilon$  有  $|V^*(s) - V^\pi(s)| \leq \epsilon$  成立,记  $V^*(s)$  为最优值函数,求解的最优策略,记为  $\pi^*$ 。

## 2.3 POMDP 求解算法

POMDP 模型的求解一直是限制 POMDP 理论发展、应用的难题。近十几年来,研究人员不断提出各类近似求解算法,这让模型的求解变得可能。在前人算法<sup>[12-14]</sup>的基础上 H. Kurniawati 等<sup>[15]</sup>人提出了最优策略可达空间的连续近似法(Successive Approximations of the Reachable Space under Optimal Policies, SARSOP)。SARSOP 算法主要由 3 步组成:采样(Sampling)、回溯操作(Backup operation)和修剪(Pruning),其求解方法如下:

**Step 1** 初始化  $\alpha^-$  向量的集合  $\Gamma$ ,作为最优值函数  $V^*$  的下边界  $\underline{V}$ ,并初始化  $V^*$  的上边界  $\bar{V}$ ;

**Step 2** 插入初始信念状态点  $b_0$  作为树  $T_R$  的根结点;

**Step 3** repeat

**Step 4** SAMPLE( $T_R, \Gamma$ );

**Step 5** 从  $T_R$  中选择子结点,对每个被选择的结点  $b$ ,BACKUP( $T_R, \Gamma, b$ );

**Step 6** PRUNE( $T_R, \Gamma$ );

**Step 7** until 满足终止条件;

**Step 8** return  $\Gamma$ 。

算法求解得到的是  $\alpha^-$  向量的集合  $\Gamma$ ,每一个  $\alpha$  向量都有相对应的行动。通过内积运算  $\alpha \cdot b$  即可确定信念状态  $b$  下应采取的最佳行动。由于 SARSOP 算法的高效性,本文将采用该算法来求解 POMDP 模型。

## 3 ECM 资源调度模型

### 3.1 雷达状态与信念状态

根据典型的雷达导引头工作流程,以及当干扰持续照射时导引头可能会转入跟踪干扰源模式,共定义 7 种状态,分别为粗搜索、精搜索、检测、截获、距离分辨、稳定跟踪和被动跟踪依次记为  $s_1, s_2, \dots, s_7$ ,其中  $s_{1-6}$  的威胁等级依次升高,而被动跟踪只有角度信息,认为  $s_7$  的威胁等级处于  $s_4$  与  $s_5$  之间。

导引头的信念状态依据观测值按贝叶斯准则更新:

$$b'_i = \frac{O(z, s'_i, a) \sum_{s \in S} P(s', s, a) b(s)}{\sum_{s' \in S} O(z, s', a) \sum_{s \in S} P(s', s, a) b(s)} \quad (5)$$

### 3.2 干扰行动与状态转移函数

考虑 7 种常见干扰样式:密集假目标干扰、多普勒闪烁干扰、噪声重复干扰、箔条干扰、速度距离波门拖引、交叉眼干扰,记为  $a_1, a_2, \dots, a_6$ ,考虑到噪声干扰对多种状态有效,记噪声干扰为  $a_7$ ;同时将不实施干扰也纳入 ECM 的行动,记为  $a_8$ 。

根据最大熵原理,假设无干扰条件下导引头的状态转移概率按照图 2 所示的箭头方向服从均匀分布。用干扰有效因子  $\delta(0 < \delta < 1)$  来评价干扰效能:

$$\delta(s, a_j) = \frac{P(s', s, a_8) - P(s', s, a_j)}{P(s', s, a_8)}, j=1, 2, \dots, 7 \quad (6)$$

式中:  $s'$  的威胁等级高于  $s$ 。干扰因子表示在采用了干扰措施后,导引头转向高威胁状态的概率降低程度。若  $\delta(s, a_i) > \delta(s, a_j)$ ,则对于状态  $s$ ,行动  $a_i$  比行动  $a_j$  更有效。因为  $\sum_s P(s', s, a) = 1$ ,假设干扰行动改变  $P(s', s, a)$  时,  $\delta$  将会平均“分配”给除  $s'$  之外的其他状态。因此,各行动下的状态转移矩阵可由图 2 和式(6)确定。

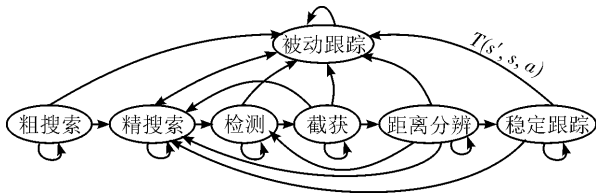


图 2 雷达导引头在 ECM 下各状态可能的转移情况

Fig. 2 The potential transfer of radar seeker under ECM

### 3.3 观测值与观测概率

ESM 系统接收和测量雷达或导弹、导引头工作参数,形成雷达短语,用以识别工作状态<sup>[16]</sup>。假设有 15 种雷达短语,记为  $z_1, z_2, \dots, z_{15}$ ,考虑到 ESM 接收机虚警概率( $P_f$ )的存在和截获概率( $P_d$ )不能达到 100%,将空观测记为  $z_{16}$ ,各观测值与状态以及对应的行动见表 1。导引头状态与雷达短语不是一一对应的,例如, $z_6$  可能是  $s_2, s_3, s_6$  发出的。

表 1 导引头状态及对应行动、观测值

Tab. 1 Seeker state and corresponding action, observed value

序号	雷达状态	对应干扰	观测值
$s_1$	粗搜索	密集假目标干扰	$z_1, z_2, z_3, z_{16}$
$s_2$	精搜索	多普勒闪烁干扰	$z_3, z_4, z_5, z_6, z_{16}$
$s_3$	检测	噪声重复干扰	$z_6, z_7, z_8, z_{16}$
$s_4$	截获	箔条干扰	$z_8, z_9, z_{10}, z_{11}, z_{16}$
$s_5$	距离分辨	速度距离波门拖引	$z_{10}, z_{11}, z_{12}, z_{13}, z_{14}, z_{16}$
$s_6$	稳定跟踪	交叉眼干扰	$z_6, z_{10}, z_{13}, z_{15}, z_{16}$
$s_7$	被动跟踪	关闭干扰	$z_{1\sim 16}$

从表 1 可知,考虑到  $P_d$  和  $P_f$  的存在,同时假设观测概率依旧满足均匀分布,将观测概率定义如下:除了  $s_7$ ,其他状态辐射  $z_{16}$  的概率相同, $s_7$  辐射  $z_{16}$  的概率为  $1 - P_f$ ,其余观测值概率和为  $P_f$ ,则观测概率函数表示为:

$$P(z_k | s_i) = \begin{cases} \frac{P_d}{N(s_i) - 1}, & i = 1, 2, \dots, 6, k \neq 16 \\ 1 - P_d, & i = 1, 2, \dots, 6, k = 16 \\ \frac{P_f}{N(s_i) - 1}, & i = 7, k = 1, 2, \dots, 15 \\ 1 - P_f, & i = 7, k = 16 \end{cases} \quad (7)$$

式中: $N(s_i)$  表示导引头各状态下可能辐射的观测值的数量,且有  $\sum_{z \in Z} P(z | s) = 1$ 。

### 3.4 奖励函数

使用 ECM 的目的是使导引头处于低威胁等级,即搜索状态,这对于飞机是最安全的状态,与其他 ECM 调度方法需要计算检测概率和虚警概率不同<sup>[4]</sup>,本文将奖励函数定义如下:

$$R(s', s, a) = \begin{cases} 1, & g(s') < g(s) \\ 0, & g(s') \sim g(s) \\ -1, & g(s') > g(s) \end{cases} \quad (8)$$

式中: $g(s)$  表示状态  $s$  的威胁等级,若  $s'$  的威胁等级低于  $s$ ,则奖励 +1;反之则惩罚 -1;若两状态威胁等级相同则没有奖惩。需要说明的是,如果 ECM 能使状态停留在低优先级搜索,也奖励 +1,若不能使最高威胁稳定跟踪状态转向其他低威胁状态,也惩罚 -1。

## 4 算法流程

本文采用 POMDP 离线规划算法,即先得到策略集(上文提到的  $\alpha^-$  向量的集合  $\Gamma$ ),再在实际应用中利用该集合确定每次的行动,具体算法步骤为:

**Step 1** 将 ECM 调度模型各参数输入 SAR-SOP 算法中,设置算法终止条件,求得  $\alpha^-$  向量集合  $\Gamma$ ;

**Step 2** 初始化导引头状态  $s_0$ ,及初始信念状态  $b_0$ ,并根据表 2 及式(7)随机生成雷达短语;

**Step 3** 电子对抗系统收到观测值后,根据式(5)更新信念状态  $b$ ;

**Step 4** 将  $\Gamma$  中每个元素同  $b$  内积运算,最大值所对应  $\alpha^-$  向量的行动即为此时应采取的干扰措施;

**Step 5** 受干扰行动的影响,导引头状态依据状态转移矩阵发生变化,并生成新的雷达短语;

**Step 6** 返回 Step 3。

## 5 仿真分析

参数设置:折扣因子  $\gamma = 0.95$ ,截获概率  $P_d = 0.98$ ,虚警概率  $P_f = 0.02$ ,干扰效果因子  $\delta(*, a_i) = 0.25 (i = 1, 2, \dots, 6), \delta(*, a_7) = 0.15$ ,\* 表示任意状态,设初始信念状态满足均匀分布,即  $b_0 = [1/7, 1/7, \dots, 1/7]$ ,同时假设导引头的初始状态为  $s_6$ ,即稳定跟踪状态。SARSOP 算法运行满足精度要求后停止。

在每一步仿真中,真实的雷达导引头状态依据自身工作流程同时受到干扰行动的影响而不断发生变化,并产生每一步的观测值.真实状态和观测值变化见图 3。

由图 3 可以看出通过观测值并不能直接估计出真实的导引头状态。在多个状态下导引头都有可能“辐射” $z_6$ ,收到观测值后 ECM 系统难以准确估计导引头状态。事实上,ECM 系统依据信念状态而非直接通过雷达脉冲来采取行动,而每一步的观测值都将用于更新信念状态,然后自适应选择最优的行动。

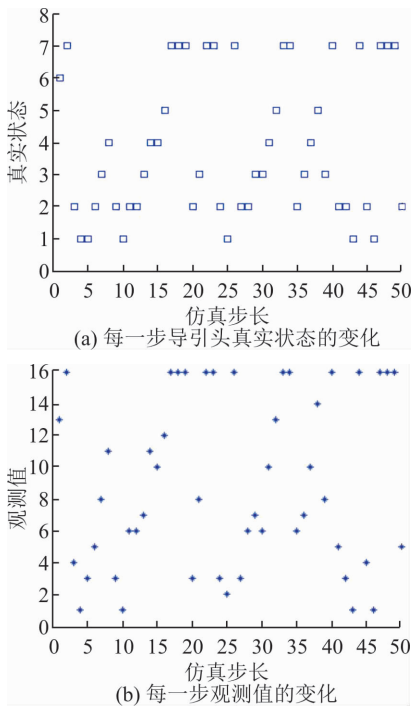


图3 每一步的真实导引头状态及观测值

Fig.3 True seeker states and observations in each step

为了说明 POMDP 策略的优越性,引入现有策略作比较,包括:贪婪策略、随机(盲目)策略、固定策略。贪婪策略只考虑使得当前状态转向低威胁等级概率最大;随机策略在每一步等概率随机选择一种干扰;考虑到噪声干扰对所有状态均有效,将噪声干扰设为固定 ECM 策略。不同策略下导引头真实状态和对应行动见图 4。

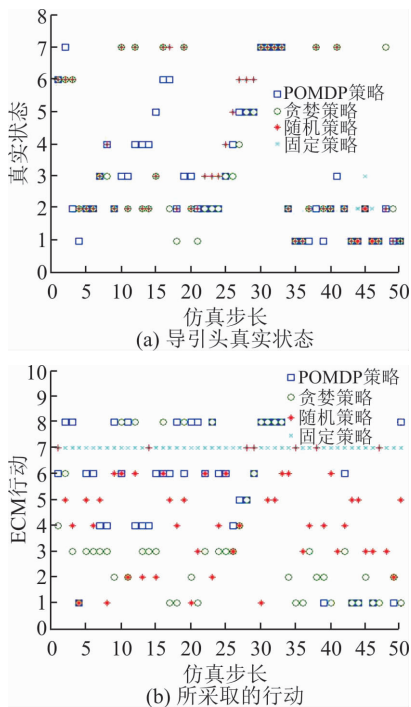


图4 不同策略下真实状态和对应行动

Fig.4 True seeker states and corresponding actions under different strategies

如图 4 所示,在初始状态均为  $s_6$  的情况下,不同的策略采取的行动集是不相同的,且在 POMDP 策略下,导引头状态最快转入最低威胁状态( $s_1$ ),这是因为 POMDP 考虑的是长远利益,所以相比短视的贪婪策略及其他策略,雷达会更快转入  $s_1$  状态,说明了 POMDP 策略的高效性。

使用 ECM 的目的是使导引头的威胁最低,以确保飞机安全。为了衡量所提方法的有效性,用搜索(粗搜索或精搜索)和稳定跟踪状态的平均占有率去定量评估。至当前步长为止搜索和稳定跟踪状态的访问次数与当前步长的比值表示搜索和稳定跟踪状态的占有率,然后通过 1 000 次 Monte Carlo 实验得到平均占有率。不同策略下的搜索和稳定跟踪状态平均占有率见图 5。

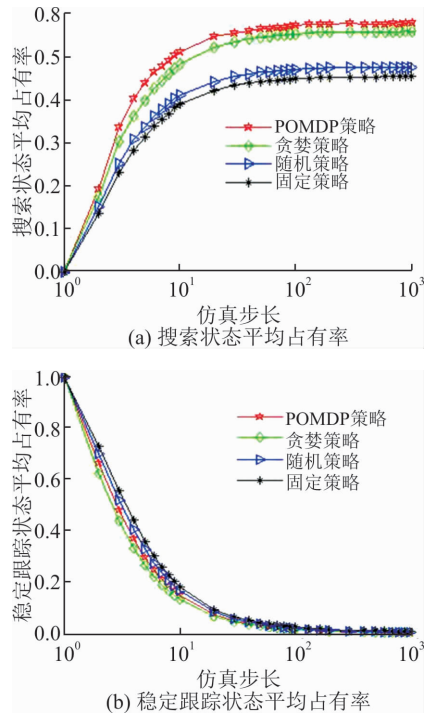


图5 不同策略搜索和稳定跟踪状态平均占有率  
Fig.5 The average occupancy of search and stable tracking state in different strategies

图中,所有策略下搜索和稳定跟踪状态平均占有率都会收敛到稳定值。其中 POMDP 策略下搜索状态平均占有率最高,其次是贪婪策略,最后是固定策略。图 6 展示了搜索和稳定跟踪状态平均占有率的更多细节。

可以看出,模型收敛时,POMDP 策略下搜索状态平均占有率要比贪婪策略下的搜索状态平均占有率高约 4%;同时,POMDP 策略下稳定跟踪状态的平均占有率是所有策略中最低的。这说明 POMDP 策略更能使战机尽可能多的处于低威胁状态,而尽可能少的处于高威胁状态。

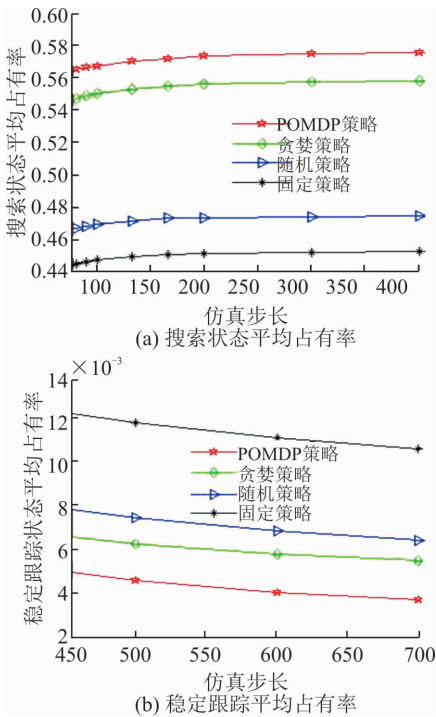


图 6 平均占有率(局部放大图)

Fig. 6 The average occupancy (partial enlarged view)

将初始状态设为截获状态  $s_3$ , 图 7 为这种情况下搜索和稳定跟踪状态平均占有率, 仿真结果同样说明了 POMDP 策略下搜索平均占有率最高, 稳定跟踪平均占有率最低。

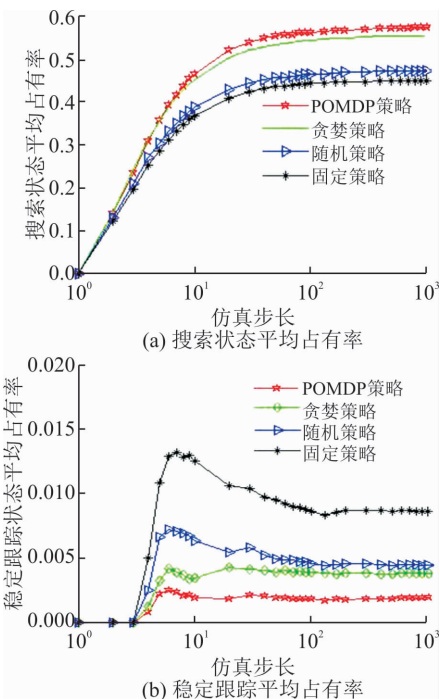


图 7 初始状态为  $s_3$  时搜索与稳定跟踪状态平均占有率

Fig. 7 The average occupancy of search and stable tracking state in initial state  $s_3$

类似的, 图 8 为图 7 的局部放大。对于其他初始状态, 均能得到类似的仿真结果, 限于篇幅, 不做

一一展现。这说明初始状态的选取并不影响仿真实验。

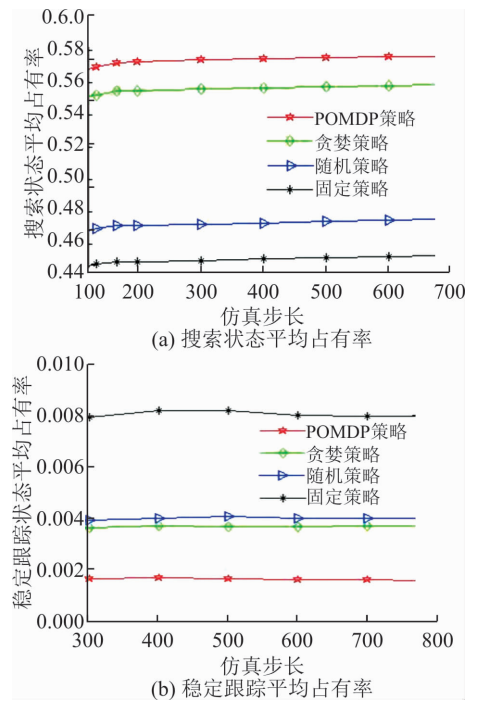


图 8 初始  $s_3$  时平均占有率(局部放大图)

Fig. 8 The average occupancy in initial state  $s_3$  (partial enlarged view)

仿真结果也表明, 随机策略要比固定策略好, 由于随机策略和固定策略都没有任何雷达导引头先验信息, 同样条件下, 随机对抗策略要比单一对抗策略要好。

## 6 结语

针对不确定条件下战斗机主对抗动雷达制导导弹过程中干扰资源调度问题, 本文用 POMDP 理论对 ECM 源调度问题建模, 建立了 ECM 调度问题的统一解决框架。不同于现有的人工盲目选择或者基于专家系统的贪婪方法, 本文所提基于 POMDP 的方法将 ESM 系统的输出和 ECM 系统的控制相结合, 根据观测值自适应的选择最优对抗措施, 提高了 ECM 系统的智能化水平和自适应能力。仿真结果表明: 相对于现有的调度策略, 本文采用的 POMDP 策略能使导引头更快的转入搜索状态, 且该策略下导引头状态更多的处于搜索状态而更少处于稳定跟踪状态, 战机相对更安全。本文仅考虑了单机实时对抗下的场景, 下一步将对多机实时对抗时的 ECM 调度问题进行研究。

## 参考文献(References):

[1] 王星. 航空电子对抗原理[M]. 北京: 国防工业出版社,

- 2008;12-16.
- WANG X. Principles of Avionics Electronic Countermeasures [M]. Beijing: National Defense Industry Press, 2008: 12-16. (in Chinese)
- [2] 陈辉. 雷达干扰辅助决策研究[D]. 西安: 西安电子科技大学, 2014.
- CHEN H. A Study of Assistant Decision-making for Radar Jamming[D]. Xi'an: Xidian University, 2014. (in Chinese)
- [3] 陈凯. 对相控阵雷达的智能干扰决策技术研究[D]. 西安: 西安电子科技大学, 2012.
- CHEN K. Research on Intelligent Jamming Decision-making Technology against Phased-Array Radar[D]. Xi'an: Xidian University, 2012. (in Chinese)
- [4] BROOKS R R, JING E P, GRIFFIN C. Game and Information Theory Analysis of Electronic Countermeasures in Pursuit-evasion Games[C]//IEEE Trans Syst Man Cyber Part A Syst Hum 2008, 38 (6) : 1281-1294.
- [5] NOROUZI T, NOROUZI Y. Scheduling the Usage of Radar and Jammer During Peace and War Time[C]//IET Radar Sonar Navig, 2008, 6 (9) : 929-936.
- [6] ANDREA D. Introduction to Modern EW Systems [M]. Boston: Artech House Inc, 2012.
- [7] 万开方, 高晓光, 李波, 等. 基于部分可观测马尔科夫决策过程的多被动传感器组网协同反隐身探测任务规划[J]. 兵工学报, 2015, 36(4): 731-742. (in Chinese)
- WAN K F, GAO X G, LI B, et al. Mission Planning of Passive Networked Sensors for Cooperative Anti-Stealth Detection Based on POMDP[J]. Acta Armamentarii, 2015, 36(4): 731-742. (in Chinese)
- [8] DANIEL A, BERNHARD W. Closed-Loop Safety Assessment of Uncertain Roadmaps [J]. Autonomous Robots, 2015, 40(2): 267-289.
- [9] ZHANG S H, WANG H J, ZHANG H G, et al. Optimal Spectrum Access Algorithm Based on POMDP in Cognitive Networks[J]. International Journal of Electronic Communication, 2015, 69(6): 943-949.
- [10] QIAO X, WANG Z, CHEN H X, et al. Replenishment Control Policies for a Production/Inventory System with Invisible Stock Loss[J]. IFAC-Papers Online, 2015, 48(3): 1762-1767.
- [11] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 319-339.
- ZHOU Z H. Machine Learning[M]. Beijing: Tsinghua University Press, 2016: 319-339. (in Chinese)
- [12] PINEAU J, GORDON G, THRUN S. Point-Based Value Iteration: An Anytime Algorithm for POMDPs [C]//In Proceedings of the 18th IJCAI, 2003: 1025-1032.
- [13] SMITH T, SIMMONS R. Heuristic Search Value Iteration for POMDPs[C]//In Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI), 2004: 520-527.
- [14] SHANI G, BRAFMAN R L, SHIMONY S E. Forward Search Value Iteration for POMDPs[C]//In Proc Int Conf on Artificial Intelligence, 2007: 2619-2624.
- [15] HANNA K, DAVID H, LEE W S. SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces[C]//In Proc Robotics: Science & Systems, 2008.
- [16] VISNEVSKI N, KRISHNAMURTHY V, HAYKIN S, et al. Multi-Function Radar Emitter Modelling: A Stochastic Discrete Event System Approach [C]//In Proceedings of IEEE Conference Decision and Control, 2003.

(编辑: 徐敏)