

引用格式:薄宁,李相民,代进进,等.基于MDP的有/无人机组队对地攻击行动方案规划[J].电光与控制,2019,26(2):16-22. BO N, LI X M, DAI J J, et al. MDP based air-to-ground attack operation planning for manned/unmanned aerial vehicle formation[J]. Electronics Optics & Control, 2019, 26(2):16-22.

## 基于MDP的有/无人机组队对地攻击行动方案规划

薄宁<sup>1,2</sup>, 李相民<sup>1</sup>, 代进进<sup>1</sup>, 唐嘉钰<sup>1</sup>, 庞威<sup>3</sup>

(1. 海军航空大学, 山东烟台 264001; 2. 中国人民解放军91213部队, 山东烟台 264001;

3. 中国人民解放军31102部队, 南京 210000)

**摘要:** 针对有/无人机组队对地攻击作战行动方案规划求解问题, 提出一种基于马尔可夫决策过程(MDP)模型的求解策略。首先, 依据有/无人机组队对地攻击兵力组成和行动特点, 给出了MDP模型中状态集与行动集的形成方法, 设计了MDP模型中状态转移概率计算方法和报酬函数, 考虑资源、时间等约束条件, 进一步建立了行动方案生成MDP优化模型; 其次, 对蚁群算法的路径选择概率和信息素更新策略进行了改进, 并应用于MDP优化模型求解; 最后, 进行了仿真分析, 结果表明, 提出的方案能够有效地求解有/无人机组队行动方案生成问题, 求解过程速度快, 解质量较高。

**关键词:** 对地攻击; 有/无人机组队; 行动方案规划; MDP; 蚁群算法

中图分类号: V271.4 文献标志码: A doi:10.3969/j.issn.1671-637X.2019.02.004

## MDP Based Air-to-Ground Attack Operation Planning for Manned/Unmanned Aerial Vehicle Formation

BO Ning<sup>1,2</sup>, LI Xiang-min<sup>1</sup>, DAI Jin-jin<sup>1</sup>, TANG Jia-yu<sup>1</sup>, PANG Wei<sup>3</sup>

(1. Naval Aviation University, Yantai 264001, China; 2. No. 91213 Unit of PLA, Yantai 264001, China;

3. No. 31102 Unit of PLA, Nanjing 210000, China)

**Abstract:** A strategy based on Markov Decision Process (MDP) was proposed to solve the air-to-ground attack operation planning problem for manned/unmanned aerial vehicle formation. Firstly, according to the characteristics of combat operations of manned/unmanned aerial vehicle formation, the forming method of the state set and action set was proposed, and the method for calculating the state transition probability and reward function of MDP model was designed. Then, with consideration of the resources and time constraints, the MDP optimization model was established to execute the operation planning. After that, we made improvements to the Ant Colony Algorithm (ACO) on the path selection probability and pheromone updating method to solve the MDP optimization model. Lastly, a simulation analysis was carried out. The result showed the proposed method can obtain a good solution to the operation planning problem for manned/unmanned aerial vehicle formation with high calculation speed.

**Key words:** air-to-ground attack; manned/unmanned aerial vehicle formation; operation planning; Markov Decision Process (MDP); ant colony algorithm

### 0 引言

无人机当前已被广泛应用于战术压制敌防空系统、侦察搜索、定位跟踪等对地攻击相关任务<sup>[1]</sup>, 但自主能力较差、可靠性不高是其主要不足。有人机与无

人机混合编队协同作战, 充分发挥二者的各自优势, 将成为未来一段时间内无人机作战运用的一种主要作战模式<sup>[2]</sup>。

行动方案规划是有/无人机组队对地攻击作战指挥控制过程的关键环节, 是指根据编队有限的战术资源及战场环境、敌我态势等信息, 选择最佳行动策略, 以有效完成使命任务。其主要确定任务以及任务之间的序列、功能需求等, 一直是军事指挥与控制领域的研究热点问题<sup>[3-6]</sup>。目前已开展的研究主要有基于动态贝叶斯网络(Dynamic Bayesian Network, DBN)、影响

收稿日期: 2018-01-18 修回日期: 2018-03-23

基金项目: 航空科学基金(20135184008)

作者简介: 薄宁(1982—), 男, 山东章丘人, 博士生, 研究方向为武器装备与作战指挥一体化技术。

网、MDP 等建模方法。应用 DBN 和影响网建模过程较为复杂,且需预先给定所有可能影响事件的先验概率,而实际往往难以完全准确估计。与 DBN 等建模方法相比,MDP 应用在中级序贯决策领域时,表达式更为简洁,求解更为方便,近年来得到了广泛应用<sup>[7-8]</sup>,而目前针对有/无人机编队行动方案规划应用 MDP 建模求解的相关研究尚不多见。运用 MDP 时,随着状态集和行动集元素增多出现的模型求解困难也是困扰研究人员的问题之一。文献[9]在处理空军战役方案时,假设任务间相互独立,仅具有总资源数量约束的弱耦合关系,从而将整个方案空间分解为多个以任务为单位的 MDP,以降低求解难度,但其要求的任务独立性条件过于严苛,不能很好地适用于研究的问题。

针对上述问题,提出一种基于 MDP 模型的有/无人机编队对地攻击行动方案,建立适当的状态集和行动集,采用基于作战效果的思想确定状态转移概率和报酬函数,并设计了状态转移图约简方法,建立 MDP 优化问题模型;对经典蚁群算法路径选择概率、信息素更新方案进行改进,用于求解 MDP 问题,改善了算法性能,提高了算法搜索效率,并结合实际算例对设计的整个方案进行了仿真分析。

## 1 问题描述

有/无人机编队作战使命方案生成问题具有如下特点:1) 各类约束条件多,如有/无人机飞机平台、武器等资源约束,时间约束等;2) 优化目标选择多,如可以选择时间最短、效果期望值最大或二者的组合;3) 编队内平台数量较大,可行动种类多;4) 战场环境的不确定性因素多,如战场上态势的变化,行动效果的不确定性,要求方案生成过程迅速有效且具备动态调整适应能力。

借鉴文献[4]中作战行动过程优选问题模型思想,记可行的有/无人机编队行动方案为  $\Pi$ ,则有

$$\Pi = (\pi(S(t_0), S(t_1)), \dots, \pi(S(t_{k-1}), S(t_k)), \dots, \pi(S(t_{f-1}), S(t_f))) \quad (1)$$

$$\text{s. t. } S(t) \in S, \text{re}(\pi(S(t_{k-1}), S(t_k))) \leq Or_t, 0 \leq k \leq e,$$

$$RE(\Pi) \leq Or_g, \text{TIME}(\Pi) \leq T_g$$

式中: $S$  为状态集合; $t_0$  为起始时刻; $t_f$  为终止时刻; $\pi(S(t_{k-1}), S(t_k))$  表示行动过程  $\Pi$  中从  $S(t_{k-1})$  到  $S(t_k)$  的状态转移过程; $\text{re}(\pi(S(t_{k-1}), S(t_k))) \leq Or_t$  为每步转移中的局部资源约束; $RE(\Pi) \leq Or_g$  为整个状态转移过程的全局资源约束; $\text{TIME}(\Pi) \leq T_g$  为全局时限约束。

基于作战效果思想,有/无人机编队对地攻击最优作战行动方案是指使得作战效果期望值最大的行动过

程  $\Pi^*$ , 即

$$\Pi^* = \arg \max_{\Pi \in P_A} \{E(\text{effective}(\Pi))\} \quad (2)$$

式中: $P_A$  为  $\Pi$  的集合; $\text{effective}(\Pi)$  为行动过程  $\Pi$  的期望作战效果。

## 2 MDP 建模

有/无人机编队战场行动方案规划的目的是改变所控制的战场态势,使之向期望的目标状态转移。行动的选择受限于资源状态,行动导致的状态转移具有不确定性,能够达到期望状态的过程是可选过程。根据以上分析,有/无人机编队作战行动过程具有马尔可夫链的特性,期望的目标状态是马尔可夫链的吸收态。方案生成过程具备 MDP 的性质。

### 2.1 MDP 概述

MDP 是解决随机序贯决策问题的重要理论工具,系统的决策者观察到系统当前的状态,根据这个状态选取一个行动,系统以与这个行动有关的概率在下一时刻转移到新的状态。同时,系统获得一个与动作和转移状态有关的报酬,整个过程一直循环下去。运用 MDP 理论进行建模的问题要求具有状态转移的无后效性,即“马尔可夫性”。

离散 MDP 基本模型是一个四元组  $\langle S, A, T, R \rangle$ 。其中: $S$  为状态空间,系统所有可能状态的集合; $A(s)$  为行动空间,指系统状态为  $s$  时,可供决策者选择的行动集合; $T$  为状态转移函数,  $S \times A \times S \rightarrow [0, 1]$ ,  $T(s, a, s')$  表示系统在状态  $s$ , 执行动作  $a$ , 转移到状态  $s'$  的概率; $R$  为报酬函数,  $S \times A \times S \rightarrow \mathbf{R}$ ,  $R(s, a, s')$  表示系统在状态  $s$ , 执行动作  $a$ , 转移到  $s'$  获得的立即报酬。

### 2.2 状态与行动空间建模

有/无人机编队对地攻击状态空间的确定,由作战领域专家确定整个使命任务需要完成的子目标,根据子目标完成与否的排列组合,每一种组合确定一个状态,并根据有/无人机作战过程、使命要求,剔除实际作战过程中不可能出现的状态,剩余状态的集合构成状态空间。

行动由兵力资源基本功能、行动目标两者共同描述。记  $a = \langle f, Tar \rangle$ , 其中: $f$  为行动要执行的资源功能; $Tar$  为目标子集。行动空间  $A = \{a_0, a_1, \dots, a_m\}$ 。

### 2.3 状态转移概率

将有人机、无人机统一抽象为作战智能体(CA)。每个作战智能体均具有一定数量的武器或其他作战载荷资源。假设整个编队共有  $K$  种不同的资源,每种资源的数量为  $L_k (k=1, 2, \dots, K)$ , 对于任一资源  $R_{kl}, l=1, 2, \dots, L_k$ , 当前状态为  $s^x$  时,使用资源  $R_{kl}$  从状态  $s^x$  转移至状态  $s^y$  的概率为  $P_{kl}^{xy}$ , 假设资源之间的效果事件相

互独立,转移概率表示为

$$P_{kl}^{xy} = 1 - \prod_{k=1}^K \prod_{l=1}^{i_k} (1 - P_{kl}^{xy}) \quad (3)$$

式中: $i_k$  为行动中实际使用的该型武器数量; $P_{kl}^{xy}$  指单位攻击武器对敌方目标的毁伤概率。

## 2.4 回报值计算

在对有/无人机编队作战方案生成问题进行 MDP 建模时,定义状态转移的收益值由两部分组成:效果收益与时间成本。其中,效果收益又分为毁伤敌方收益、我方飞机平台被毁伤的损失、武器消耗。时间成本指为完成状态转移而造成的时间消耗。故所建 MDP 模型的回报值为向量形式,即

$$\mathbf{R}(s^x, a, s^y) = (R(s^x, a, s^y), timelen)^T \quad (4)$$

式中: $R(s^x, a, s^y)$  为效果收益值; $timelen$  为时间成本值。

### 2.4.1 效果收益

#### 1) 毁伤敌方收益。

定义从状态  $s^x$  转移至状态  $s^y$ , 假设敌方单位数量为  $N_e$ , 则毁伤敌方收益为

$$R_d(s^x, a, s^y) = \sum_{i=1}^{N_e} tw(t, i) * dam(t, i) * r_i \quad (5)$$

式中: $dam(t, i)$  为毁伤判断函数,若从状态  $s^x$  转移至  $s^y$  后,目标  $i$  的状态从未毁伤变为毁伤,则  $dam(t, i)$  值为 1, 其他情况为 0; $r_i$  为目标  $i$  的价值; $tw(t, i)$  为时间窗判断函数,若目标  $i$  为时敏目标,其时间窗口为  $[t_i^e, t_i^f]$ , 则  $tw(t, i)$  定义为

$$tw(t, i) = \begin{cases} 1 & t_i^e \leq t \leq t_i^f \\ 0 & \text{其他} \end{cases} \quad (6)$$

#### 2) 飞机平台损失成本。

在有/无人机编队执行任务期间,由于 MDP 中定义的行动集元素均由飞机平台执行,存在我方飞机平台被毁伤的可能。定义飞机平台损失成本为

$$C_p(s^x, a, s^y) = \sum_{j=1}^{N_p} Q_j^{xy} * v_j^p \quad (7)$$

式中: $N_p$  为行动中使用的飞机平台数量; $Q_j^{xy}$  为飞机平台  $j$  在状态转移过程中的毁伤概率; $v_j^p$  为飞机平台  $j$  的成本。

由于定义的行动集是以武器载荷使用量来进行区分,无法直接根据当前状态以及行动确定飞机平台的类型和使用量,故采用估算均值的方法对  $C_p(s^x, a, s^y)$  进行计算。首先,估算  $N_p$  为

$$N_p = \left\lfloor \frac{\sum_{k=1}^K v_w^k * i_k}{\sum_{k=1}^K v_w^k * i_{tk}} * T_p \right\rfloor \quad (8)$$

式中: $v_w^k$  为  $k$  型武器载荷的单位成本; $i_{tk}$  为编队中  $k$  型

武器的总数量; $T_p$  为飞机平台总数量; $\lfloor \cdot \rfloor$  为取整运算。飞机平台成本均值为

$$\bar{v}^p = \frac{1}{T_p} \sum_{j=1}^{T_p} v_j^p \quad (9)$$

则飞机平台损失成本估算为

$$C_p(s^x, a, s^y) = N_p * \bar{v}^p * Q^{xy}(N_p) \quad (10)$$

### 3) 武器消耗成本。

武器消耗成本算式为

$$C_w(s^x, a, s^y) = \sum_{k=1}^h v_w^k * i_k \quad (11)$$

### 4) 效果收益。

效果收益为毁伤敌方造成的收益与我方损失成本之差,即

$$R(s^x, a, s^y) = P_{kl}^{xy} * R_d(s^x, a, s^y) - C_p(s^x, a, s^y) - C_w(s^x, a, s^y) \quad (12)$$

### 2.4.2 时间成本

有/无人机编队对地攻击方案生成 MDP 建模中,时间成本是指状态转移造成的时间消耗。对于飞机平台而言,执行某项行动造成的时间消耗分为两部分:从当前位置到达任务执行地点的航线飞行时间和到达任务地点后执行具体行动的时间,即

$$timelen(s^x, a, s^y) = t_{conc}(s^x, a, s^y) + t_{exc}(s^x, a, s^y) \quad (13)$$

式中: $timelen(s^x, a, s^y)$  为时间成本; $t_{conc}(s^x, a, s^y)$  为到达任务地点的航路飞行时间,若是多机协同执行任务,则  $t_{conc}(s^x, a, s^y)$  为所有飞机的航路飞行时间中的最大值; $t_{exc}(s^x, a, s^y)$  为任务执行时间。

## 2.5 状态转移图约简

作战领域专家依据战场态势建立的状态转移图描述了状态间所有可能的转换过程。但初始给定的资源数量可能无法满足状态转移资源需求,特别是对于某一个阶段同时执行多个任务的情况。为减少 MDP 问题求解时的无效路径搜索,采用对状态转移图进行一步转移路径资源约束条件检测方法,实现状态转移图约简,步骤如下。

1) 根据我方资源平台初始参数表计算  $T_k$  以及  $L_k, k=1, 2, \dots, K$ , 其中,  $T_k$  表示初始状态下所有具有  $k$  类型资源的飞机平台数量之和。

2) 从状态转移矩阵中任取一条转移路径,计算该状态转移的飞机平台和资源约束条件,假设为

$$\begin{cases} i_k \geq i_{k0} \\ N_k^{plat} \geq N_{k0}^{plat} \end{cases} \quad k=1, 2, \dots, K \quad (14)$$

式中: $N_k^{plat}$  是指为满足  $k$  类资源约束而需出动的飞机数量; $i_{k0}$  和  $N_{k0}^{plat}$  为常量,表示状态转移过程中所需的  $k$  类资源数量以及具有  $k$  类资源平台数量需求。

3) 检查条件为  $L_k \geq i_{k0}, T_k \geq N_{k0}^{plat}, k=1, 2, \dots, K$  是

否全部满足,若存在某一项不满足,则删除该路径。

4) 若完成状态转移矩阵搜索,则退出;否则转2)。

## 2.6 MDP 优化问题建模

### 2.6.1 目标函数

对于有/无人机编队对地攻击方案生成,按照作战任务要求可选择不同的目标函数。目标函数可选择在资源、时限约束下的作战效果最优。作战效果最优的性能指标为

$$\max J_{ef} = E\left(\sum_{t=0}^{H-1} R(s^t, a(t), s^{t+1})\right)。(15)$$

此外,在满足期望作战效果的前提下,任务完成时间越短越好。任务期望完成时间最短性能指标为

$$\min J_{time} = E\left(\sum_{t=0}^{H-1} timelen(s^t, a(t), s^{t+1})\right)。(16)$$

选择二者的加权和作为 MDP 问题最终的目标函数,通过调整加权系数,可以在作战效果与完成时间之间取得平衡,以适应不同的作战任务需求。目标函数的最终形式为

$$\begin{aligned} \max J &= \alpha_1 J_{ef} + \alpha_2 (-J_{time}) = \\ &\alpha_1 E\left(\sum_{t=0}^{H-1} R(s^t, a(t), s^{t+1})\right) - \\ &\alpha_2 E\left(\sum_{t=0}^{H-1} timelen(s^t, a(t), s^{t+1})\right) \end{aligned} (17)$$

式中, $\alpha_1$  和  $\alpha_2$  分别为作战效果期望与时间的权重系数, $0 \leq \alpha_1, \alpha_2 \leq 1$ 。

### 2.6.2 约束条件

约束条件分为资源约束和行动方案目标约束。行动方案目标即期望作战效果与作战时间,前文已定义,不再赘述。资源约束分为全局资源约束和局部资源约束,仅考虑全局资源约束。任务过程转移全程所使用的每种资源总数不超过该类型最大可用资源数量,同时不超过该类型初始资源数量,即

$$i_k \leq \min\{res_{maxg}, L_k\} \quad k=1, 2, \dots, K (18)$$

式中, $res_{maxg}^k$  为资源类型  $k$  的全局资源使用约束。最终的优化问题模型描述为

$$\max J (19)$$

$$\text{s. t. } E\left(\sum_{t=0}^{H-1} R(s^t, a(t), s^{t+1})\right) \geq V_0, t_0 + E\left(\sum_{t=0}^{H-1} timelen(s^t, a(t), s^{t+1})\right) \leq t_j, i_k \leq \min\{res_{maxg}, L_k\}$$

式中: $V_0$  为要求的最低作战效果期望; $k=1, 2, \dots, K$ 。

## 3 MDP 问题求解

所建 MDP 模型为典型的随机期望值规划模型,文献[10]证明了期望值模型为凸规划模型。同时,所建问题模型随着 MDP 状态集元素数量的增大,搜索复杂度急剧增大,使用确定性搜索算法求解较为困难。仿生智

能计算适用于解决这类复杂的寻优搜索问题。

蚁群算法由 DORIGO 等于 1991 年提出<sup>[11]</sup>,已被广泛运用于解决诸如旅行商问题(TSP)等组合优化问题,但收敛速度较慢、易陷入局部最优解是其主要缺点。因此采用改进路径选择概率及信息素更新策略提高算法收敛速度,改善算法搜索性能。

### 3.1 路径选择概率

经典蚁群算法解决 MDP 优化问题采用下式确定状态转移概率

$$p_k(S_t, a_i) = \frac{[\tau_k(S_t, a_i)]^\alpha \cdot [\eta_k(S_t, a_i)]^\beta}{\sum_{a_i \in A(S_t)} [\tau_k(S_t, a_i)]^\alpha \cdot [\eta_k(S_t, a_i)]^\beta} (20)$$

式中: $p_k(S_t, a_i)$  为第  $k$  只蚂蚁在  $t$  时刻,位于状态  $S_t$  时,选择行动  $a_i$  的概率; $\tau_k(S_t, a_i)$  为路径信息量; $\eta_k(S_t, a_i)$  为路径启发信息; $\alpha$  为信息素启发因子; $\beta$  为期望启发式因子。

由于在初始路径搜索过程中随机性较大,若完全依赖由路径残留信息量和启发信息确定的转移概率进行路径选择,容易过早陷入局部最优解。借鉴模拟退火的思想,在搜索过程的初始阶段,在保证路径原概率值大小顺序不变的情况下,使得路径选择概率较为平均,随着迭代次数逐渐增大,概率值逐步逼近到式(20)所确定的转移概率值。改进后的路径选择概率为

$$p_k(S_t, a_i) = \frac{[\tau_k(S_t, a_i)]^\alpha \cdot [\eta_k(S_t, a_i)]^\beta + \gamma/j_{loop}}{\sum_{a_i \in A(S_t)} [\tau_k(S_t, a_i)]^\alpha \cdot [\eta_k(S_t, a_i)]^\beta + \gamma/j_{loop}} (21)$$

式中: $\gamma$  为迭代次数调节因子; $j_{loop}$  为当前时刻已完成的迭代次数。

### 3.2 信息素更新策略

经典蚁群算法中,路径信息素更新算式为

$$\tau_{i+1}(s, a_i) = (1 - \rho) \cdot \tau_i(s, a_i) + \Delta\tau_i(s, a_i) (22)$$

式中, $\Delta\tau_i(s, a_i)$  为信息素增量,其算式为

$$\Delta\tau_i(s, a_i) = \sum_{k=1}^{N_i} \Delta\tau_k(S_t, a_i) (23)$$

$\Delta\tau_k(S_t, a_i)$  算式为

$$\Delta\tau_k(S_t, a_i) = \frac{Q}{L_k} (24)$$

式中: $Q$  为信息素强度; $L_k$  为转化后的路径长度。

由于蚂蚁均释放信息素,导致较差路径被淘汰较慢。采用精英蚂蚁策略改进此问题,即每一次迭代结束后,将所有蚂蚁按照回报值由高到低进行排序,只有排名靠前的  $N_{max}$  只蚂蚁进行信息素释放,即“精英蚂蚁”信息素更新策略。

### 3.3 算法流程

改进后的蚁群算法流程如图 1 所示。

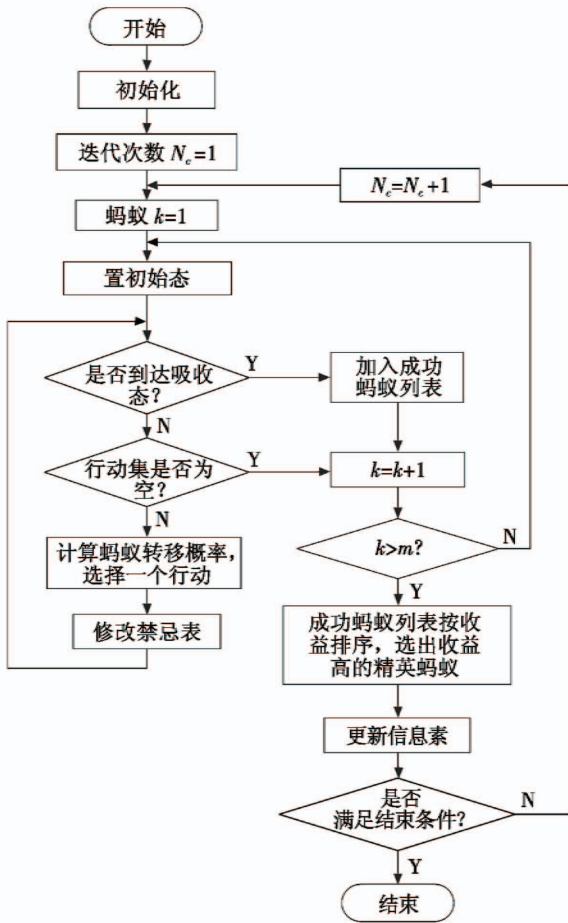


图1 改进的蚁群算法流程图

Fig.1 Flow chart of the improved ACO

算法开始后对蚂蚁数量、迭代次数、启发因子等参数赋初值。对于每次迭代中的每只蚂蚁,使用改进的路径选择概率随机选择路径,并放入禁忌表,不断搜索直至成功结束或不满足约束退出本次迭代。所有蚂蚁均完成一次迭代搜索后按照精英蚂蚁策略进行路径信息素更新,循环进行直至满足搜索退出条件。

### 4 仿真分析

采用计算机仿真验证提出的有/无人机编队行动方案生成方法。计算机基本配置为: Intel Pentium (R) Dual E2180 2 GHz, 2 GiB 内存, WindowsXP 操作系统, 仿真软件平台为 Matlab R2014a。

#### 4.1 初始参数设定

假定战场为 200 km × 200 km 的矩形区域。我方飞机待战集结区中心坐标为 (20 km, 20 km)。共有 8 个待打击敌方战术目标, 编号为 Tar1 ~ Tar8, 目标参数值如表 1 所示。我方有人战斗机、无人机平台载荷资源类型及数量如表 2 所示。表 2 中, MAV1 ~ MAV3 为有人机平台, UAV1 ~ UAV9 为无人机平台。资源与行动

的对应关系如表 3 所示。

表 1 目标参数

Table 1 Target parameters

目标编号	属性	目标价值	威胁概率	半径/km	位置/km
Tar1	防空阵地	1	0.4	80	(110,170)
Tar2	防空阵地	1	0.5	80	(110,130)
Tar3	防空阵地	1	0.6	15	(140,150)
Tar4	机场	5	0	--	(110,190)
Tar5	港口	6	0	--	(100,110)
Tar6	指挥所	10	0	--	(150,150)
Tar7	移动目标	3	0	--	(175,175)
Tar8	移动目标	4	0	--	(175,115)

表 2 飞机平台及载荷参数

Table 2 Parameters of platforms and payloads

无人机编号	资源编号					
	res1	res2	res3	res4	res5	res6
MAV1	0	4	0	0	0	0
MAV2	0	4	0	0	0	0
MAV3	0	0	4	0	0	0
UAV1	2	0	0	1	0	1
UAV2	2	0	0	0	0	1
UAV3	2	0	0	0	1	0
UAV4	2	0	0	0	0	0
UAV5	0	0	2	1	0	1
UAV6	0	0	2	1	0	1
UAV7	0	0	2	0	1	0
UAV8	0	2	0	1	0	1
UAV9	0	2	0	1	0	1

表 3 载荷资源与行动对应关系

Table 3 Corresponding relationship between payload and action

资源编号	资源说明	对应行动	行动说明
res1	反辐射导弹	a1	防空压制行动
res2	对地攻击武器 1	a2	对地攻击行动 1
res3	对地攻击武器 2	a3	对地攻击行动 2
res4	侦察搜索设备 A	a4	侦察搜索行动 A
res5	侦察搜索设备 B	a5	侦察搜索行动 B
res6	目标指示设备	a6	目标指示行动

假定 res4 为阶段 1 的必要资源, res6 为阶段 2 和阶段 3 的必要资源, res5 为阶段 3 的必要资源, 对状态转移概率大小无影响。飞机平台、攻击武器等成本设定如表 4 所示。

表 4 飞机及武器成本

Table 4 Platforms and weapon costs

平台(资源)	MAV	UAV	res1	res2	res3
成本	1	0.2	0.08	0.09	0.1

#### 4.2 状态转移图生成

通过作战领域专家对战场态势的分析, 假设整个任务划分为 3 个阶段: 防空压制阶段、对地攻击阶段 1 (固定目标)、对地攻击阶段 2 (移动目标)。其中, 每一

个阶段对目标既可以同时攻击,也可以顺序攻击。构造状态集合如表 5 所示。

表 5 状态集元素定义

Table 5 Element definition of state set

状态	目标编号							
	Tar1	Tar2	Tar3	Tar4	Tar5	Tar6	Tar7	Tar8
S <sup>1</sup>	0	0	0	0	0	0	0	0
S <sup>2</sup>	1	0	0	0	0	0	0	0
S <sup>3</sup>	0	1	0	0	0	0	0	0
S <sup>4</sup>	1	1	0	0	0	0	0	0
S <sup>5</sup>	1	1	1	0	0	0	0	0
S <sup>6</sup>	1	1	1	1	0	0	0	0
S <sup>7</sup>	1	1	1	0	1	0	0	0
S <sup>8</sup>	1	1	1	0	0	1	0	0
S <sup>9</sup>	1	1	1	1	1	0	0	0
S <sup>10</sup>	1	1	1	0	1	1	0	0
S <sup>11</sup>	1	1	1	1	0	1	0	0
S <sup>12</sup>	1	1	1	1	1	1	0	0
S <sup>13</sup>	1	1	1	1	1	1	1	0
S <sup>14</sup>	1	1	1	1	1	1	0	1
S <sup>15</sup>	1	1	1	1	1	1	1	1

表 5 中,元素值为 1 表示该目标已被攻击,为 0 表示未被攻击。根据状态定义,生成状态转移图并进行约简,如图 2 所示。

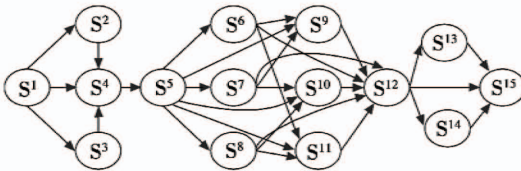


图 2 行动方案状态转移图

Fig. 2 State transition diagram of operation planning

4.3 结果分析

为验证改进蚁群算法的有效性,使用经典蚁群算法和改进蚁群算法对图 2 所示的状态转移图对应的行动过程方案分别进行优化求解。UAV 速度设定为 250 km/h,不考虑地形障碍对航路飞行用时的影响, $t_{enc}(s^i, a, s^j)$ 均设定为 0.25 h。蚂蚁数量为 20,最大迭代次数为 100, $\alpha, \beta, \gamma, \rho$  和  $Q$  分别为 4, 2, 2, 0.1, 0.5。每种算法均运行 10 次,取平均值。设置下列 3 种情景,分别进行仿真验证。

1) 时限约束下目标完成期望最高。时限约束上限为 3.5 h,优化目标为任务完成期望值最大。

2) 目标完成约束下时间最短。任务完成度期望值下限为 0.6,优化目标为时间最短。

3) 时限和目标完成期望约束下消耗最小。取任务整体完成概率下限为 0.6,时间约束上限为 3.5 h,优化目标为我方损失期望值最小。

3 种情况下算法运行结果如图 3 所示。

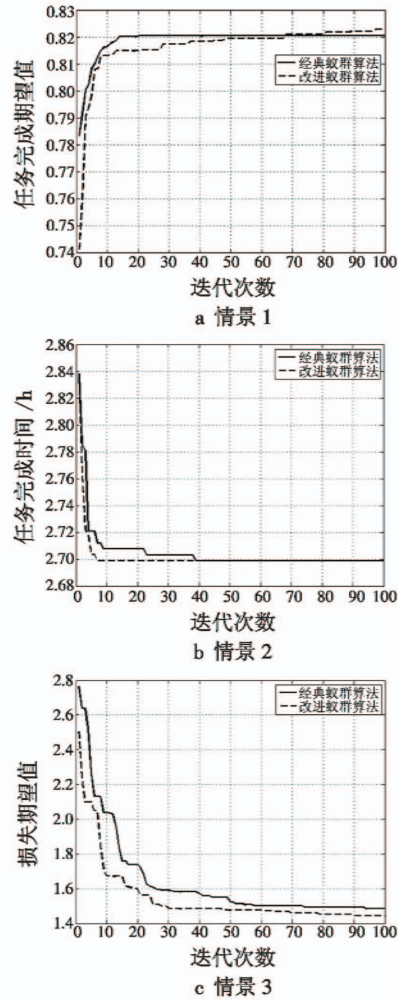


图 3 改进蚁群算法仿真结果

Fig. 3 Simulation results of the improved ACO

由图 3 可以看出,改进后的蚁群算法求得解的质量平均好于经典蚁群算法,3 次实验均在迭代 30 次左右后解质量趋于稳定,由于采用了精英蚂蚁策略,减少了信息素更新时间,实验统计发现,程序运行时间比传统蚁群算法降低约 8.5%,平均单次运行时间为 2.73 s,验证了所采用改进方法的合理性。改进的蚁群算法 3 种情景下得出的最优解如图 4 所示。

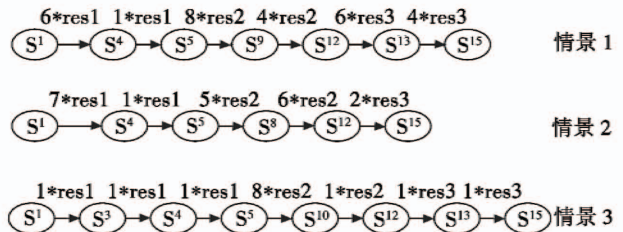


图 4 3 种情况下改进蚁群算法所得最优解

Fig. 4 Optimal solutions of the ACO under 3 situations

与状态转移图比照可以得出,所得解与实际情况相符,验证了模型建立的合理性。

## 5 结 论

1) 将 MDP 模型应用于有/无人机编队对地攻击方案生成问题,设计了模型中状态集、行动集生成及状态转移概率和收益函数的计算方法,提出了状态转移图约简策略。所建模型能够有效解决该问题,对于其他应用领域过程策略选择问题也具有一定的借鉴意义。

2) 通过对经典遗传算法路径选择概率、信息素更新策略进行改进,改善了算法性能,应用于 MDP 优化问题模型求解,可以得到质量较高的解。

3) 在计算状态转移概率时,假定毁伤概率为定值,未考虑目标可能的行动变化对转移概率和收益函数的影响,使得模型建立考虑因素不够全面,下一步将对此开展深入研究,以使模型建立更为合理。

### 参 考 文 献

- [1] 沈林成,牛轶峰,朱华勇. 多无人机自主协同控制理论与方法[M]. 北京: 国防工业出版社,2013.
- [2] 刘跃峰,陈哨东,赵振宇,等. 有人机/UCAV 编队对地攻击指挥控制系统总体研究[J]. 火力与指挥控制, 2013, 38(1):1-5.
- [3] SAJJAD H, LEVIS A H. On finding effective courses of action in a complex situation using evolutionary algorithms [C]//The 10th International Command and Control Research and Technology Symposium, Newport RI, 2004:248-262.
- [4] 彭小宏,阳东升,武云鹏,等. 基于 EAP-GA 的联合作战行动计划[J]. 火力与指挥控制,2009, 34(2):1-9.
- [5] ROSEN J A, SMITH W L. Influence net modeling for strategic planning: a structured approach for information operation[J]. Phalanx, 2000, 33(4):6-7.
- [6] 卜先锦,阳东升,沙基昌,等. 作战过程设计策略及其优选模型[J]. 火力与指挥控制,2006, 31(5):8-12.
- [7] DEAN T, KAELBLING L P, KIRMAN J, et al. Planning with deadlines in stochastic domains [C]//AAAI-93, Washington, D. C, 1993,574-579.
- [8] BOUTILLIER C, DEARDEN R. Using abstractions for decision theoretic planning with time constraints[C]//AAAI-94, Seattle, 1994: 1016-1022.
- [9] MEULEAU N, HAUSKRECHT M, KIM K. Solving very large weakly coupled Markov decision processes [C]//The 15th National Conference on Artificial Intelligence, Madison, 1998:165-172.
- [10] 刘宝碇,赵瑞清,王纲. 不确定规划及应用[M]. 北京:清华大学出版社,2003.
- [11] COLORNI A, DORIGO M, MANIEZZO V, et al. Distributed optimization by ant colonies[C]//The 1st European Conference on Artificial Life, 1991:134-142.
- (上接第 10 页)
- [7] 王永泉,罗建军. 基于多群体改进萤火虫算法的 UCAV 协同多目标分配[J]. 西北工业大学学报,2014, 32(3): 451-456.
- [8] 张涛,于雷,周中良,等. 基于人工势场启发粒子群算法的空战机动决策[J]. 电光与控制,2013, 20(1):77-82.
- [9] ABBEEL P, COATES A, QUIGLEY M, et al. An application of reinforcement learning to aerobatic helicopter flight [C]//Advances in Neural Information Processing Systems, 2007. doi:10.1.1.64.4458.
- [10] 赵振宇,卢广山. 无人机协同空战中目标威胁评估和目标分配算法[J]. 火力与指挥控制,2011, 36(12): 60-71.
- [11] 周德云,李锋. 基于遗传算法的飞机战术飞行动作决策[J]. 西北工业大学学报,2002, 20(1):109-112.
- [12] LUCIAN B, ROBERT B, BART D S. A comprehensive survey of multi-agent reinforcement learning [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 2008, 38(2):156-172.
- [13] HU J, WLLMAN M. Nash Q-learning for general-sum stochastic games [J]. Journal of Machine Learning Research, 2003(4):1039-1069.
- [14] SPECHT D F. Probabilistic neural networks [J]. Neural Networks, 1990, 3(1):109-118.
- [15] 马耀飞,龚光红,彭晓源. 基于强化学习的航空兵认知行为模型[J]. 北京航空航天大学学报,2010, 36(4): 379-383.
- [16] 董彦非,冯惊雷,张恒喜. 多机空战仿真协同战术决策方法[J]. 系统仿真学报,2002, 14(6):723-725.