

文章编号: 0427-7104(2021)04-0410-10

## 多对多无人机空战的智能决策研究

朱星宇, 艾剑良

(复旦大学 航空航天系, 上海 200433)

**摘 要:** 针对能够进行自主空战的无人机, 设计了一套空战流程. 建立了基于双方位置、姿态和能量等的优势函数评价体系, 通过改进的混合粒子群优化算法实现了多对多空战的目标分配, 结合改进的 Double Q-learning 算法设计了无人机的机动决策策略, 并在 4 对 4 空战场景下进行仿真验证. 仿真结果显示无人机可以很好地追踪并击败敌方目标, 还能够协同其他友方无人机攻击敌方无人机. 同时算法具有很好的时效性, 表明所提出的方法能够有效地解决多对多空战的问题.

**关键词:** 机动决策; 目标分配; 改进的混合粒子群优化算法; 改进的 Double Q-learning 算法

**中图分类号:** V279<sup>+</sup>.3

**文献标志码:** A

DOI:10.15943/j.cnki.fdxh-jns.2021.04.002

近年来, 世界各国纷纷提出六代机的设计理念<sup>[1]</sup>, 其中有人驾驶飞机与无人驾驶飞机 (Unmanned Aerial Vehicle, UAV, 简称无人机) 集群协同作战的观点让人耳目一新. 有人驾驶飞机在后方负责雷达侦查、指挥通信、战术目标制定, 无人机集群负责提供战场信息、对敌攻击以及对有人机的护卫. 这一分工协同理念将大大提升飞机在战场的作战效能, 同时进一步降低人员伤亡的可能性. 自主空战决策一直是研究热点, 它是指: 利用数学优化、人工智能等手段, 对作战态势进行分析, 模拟真实空战状况下飞行员的决策, 消灭对方、保存己方的过程<sup>[2]</sup>. 空战决策可以分为目标分配和机动决策两个部分.

目标分配即考虑各种约束进行敌我的匹配以实现某种尺度下的群体最优, 目前常用的解决办法可分为 3 种: 数学规划法、协商法和群智能算法. 数学规划法, 例如匈牙利算法<sup>[3]</sup>, 具有结构简单、易于实现的特点, 但其实质是对所有可行空间的遍历, 计算量大且运算速度慢. 协商法, 例如分布式算法<sup>[4]</sup>, 具有较强的鲁棒性, 但对无人机的计算通信能力要求更高. 群智能算法<sup>[5-7]</sup>, 例如遗传算法 (Genetic Algorithm, GA)、粒子群优化 (Particle Swarm Optimization, PSO) 算法、模拟退火 (Simulated Annealing, SA) 算法, 仿照生物的进化或自发行为在解空间里不断迭代, 直至找到全局最优解, 具有收敛速度快、算法简单等优点, 但由于初始值设置的随机性, 算法易陷入局部最优, 解的精度有所降低.

机动决策主要有两类方法: 基于对策的机动决策方法, 包括矩阵对策<sup>[8]</sup>、微分对策<sup>[9]</sup>和影响图对策<sup>[10]</sup>等; 基于人工智能的空战决策方法, 包括基于专家系统<sup>[11]</sup>、模糊遗传<sup>[12]</sup>、蒙特卡洛树搜索<sup>[13]</sup>和混合粒子群<sup>[14]</sup>的机动决策算法. 但上述方法仅针对 1 对 1 空战态势下固定或执行单一机动的目标, 或仅考虑某一时刻的机动决策, 计算量大且收敛速度慢, 忽视了空战的连续性和时效性, 不适合多对多环境下连续变化的空战机动决策.

本文针对多对多空战过程的动态与变化, 基于混合粒子群优化算法 (Hybrid particle swarm optimization algorithm) 进行目标分配, 引入局部最优判定参数, 在算法陷入局部最优时加入移民粒子, 从而提高规定迭代次数内的全局搜索能力, 同时在无人机数量发生变化时重新进行目标分配. 基于 Double Q-learning 算法<sup>[15-16]</sup>求解机动决策, 解决了单纯使用 Q-learning 算法带来的正向偏置问题, 同时采用变  $\epsilon$ -贪婪法平衡算法探索与利用的关系, 使无人机的机动决策更加有效. 最后通过 4 对 4 空战的仿真模拟, 验证了本文提出的空战智能决策的有效性与时效性.

收稿日期: 2020-03-13

作者简介: 朱星宇 (1993—), 男, 硕士研究生; 艾剑良, 男, 教授, 博士生导师, 通信作者, E-mail: aijl@fudan.edu.cn

## 1 空战过程的数学建模

假设有敌对的多架无人机,分别属于蓝(我)、红(敌)双方.蓝方多架无人机分别记为  $B_i (i=1,2,3,\dots)$ ,红方多架无人机分别记为  $R_i (i=1,2,3,\dots)$ .以  $B_1$  为例,在  $T=k$  时刻,  $B_1$  在空间坐标系中具有状态:

$$S_k^{B_1} = [x_k^{B_1}, y_k^{B_1}, z_k^{B_1}, \phi_k^{B_1}, \varphi_k^{B_1}, v_k^{B_1}]. \quad (1)$$

其中:  $(x_k^{B_1}, y_k^{B_1}, z_k^{B_1})$  为  $B_1$  的空间坐标;  $\phi_k^{B_1}$  为航迹偏转角,即  $B_1$  的速度矢量方向在  $xy$  平面上的投影与  $y$  轴夹角的大小,顺时针方向为正;  $\varphi_k^{B_1}$  为航迹倾斜角,即  $B_1$  的速度矢量方向与  $xy$  平面之间夹角的大小,向上为正;  $v_k^{B_1}$  为  $B_1$  的速度矢量的大小.

### 1.1 空战中的空间几何关系

考虑空战中红蓝双方的空间几何关系,以 1 对 1 空战中两机近距离格斗的空间几何关系为例,设一个整体状态量  $c$ :

$$c = (\alpha_1^{B_1}(k), \alpha_2^{B_1}(k), d_k^{B_1} | \alpha_1^{B_1}(k), \alpha_2^{B_1}(k) \in [0, \pi]). \quad (2)$$

如图 1 所示,其中:  $\alpha_1^{B_1}$  是  $B_1$  观测的滞后角,即蓝红两机连线与  $B_1$  的速度方向的夹角;  $\alpha_2^{B_1}$  是  $B_1$  观测的超前角,即  $B_1$  和  $R_1$  两机连线与  $R_1$  的速度方向的夹角;  $d_k^{B_1}$  是  $B_1$  到  $R_1$  之间的距离.

由图 1 中的空间几何关系可得

$$\begin{cases} d_k^{B_1} = |\mathbf{d}_k^{B_1}| = \sqrt{(x_k^{R_1} - x_k^{B_1})^2 + (y_k^{R_1} - y_k^{B_1})^2 + (z_k^{R_1} - z_k^{B_1})^2}, \\ v_k^{B_1} = |\mathbf{v}_k^{B_1}|, v_k^{R_1} = |\mathbf{v}_k^{R_1}|, \\ \alpha_1^{B_1}(k) = \arccos\left(\frac{\mathbf{v}_k^{B_1} \cdot \mathbf{d}_k^{B_1}}{v_k^{B_1} \cdot d_k^{B_1}}\right), \\ \alpha_2^{B_1}(k) = \arccos\left(\frac{\mathbf{v}_k^{R_1} \cdot \mathbf{d}_k^{B_1}}{v_k^{R_1} \cdot d_k^{B_1}}\right). \end{cases} \quad (3)$$

其中:  $\mathbf{d}_k^{B_1}$  是  $B_1$  到  $R_1$  之间的距离矢量;  $\mathbf{v}_k^{B_1}$  是  $B_1$  的速度矢量;  $\mathbf{v}_k^{R_1}$  是  $R_1$  的速度矢量.

### 1.2 优势函数

通常飞机通过各种机动来到相对敌机更加优势的位置来达成击毁敌方、保存己方的目的.本文利用优势函数来表示某架无人机对另一架无人机相对优势的大小.优势函数越大,我方无人机生存并击毁敌机的可能性越大.本文采用的优势函数包含角度优势、距离优势和能量优势 3 部分.

**角度优势函数** 由图 1 可知,  $\alpha_1^{B_1}$  是  $B_1$  观测的滞后角,  $\alpha_2^{B_1}$  是  $B_1$  观测的超前角.角度优势函数设置为

$$G_A = 1 - \frac{\alpha_1 + \alpha_2}{180}. \quad (4)$$

**距离优势函数** 无人机空战中优势的大小和距离也紧密相关.距离优势函数设置为

$$G_D = \begin{cases} 1 & d \leq R_0, \\ e^{-\frac{d-R_0}{k}} & d > R_0. \end{cases} \quad (5)$$

其中:  $d$  是我机与敌机之间的距离;  $R_0$  是机载武器的最佳射程;  $k$  是无量纲量.

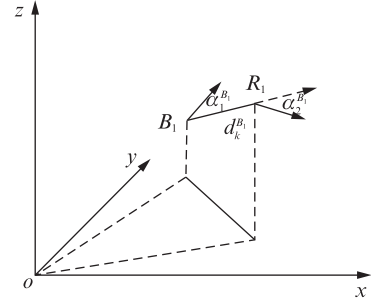


图 1 1 对 1 空战中无人机的空间几何关系

Fig. 1 Space geometry of UAVs in 1 to 1 air combat

能量优势函数 无人机空战中优势也与无人机的速度、所处的高度,即能量有关.定义

$$G_E = \begin{cases} 1 & \eta \geq 2, \\ 0.6\eta - 0.2 & 0.5 \leq \eta < 2, \\ 0.1 & \eta < 0.5. \end{cases} \quad (6)$$

其中:  $\eta = E_B/E_R$ , 表示我方和敌方无人机能量之比. 无人机的能量定义为

$$E = H + \frac{v^2}{2g}. \quad (7)$$

其中:  $H$  为飞机的高度;  $g$  为重力加速度.

综上所述,构建无人机的优势函数和优势矩阵如下:

$$G = \omega_1 G_A \times G_D + \omega_2 G_E, \quad (8)$$

$$G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1n} \\ G_{21} & G_{22} & \cdots & G_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ G_{m1} & G_{m2} & \cdots & G_{mn} \end{bmatrix}. \quad (9)$$

其中:  $\omega_1$  和  $\omega_2$  为各优势函数的权重;  $G_{ij}$  表示我方第  $i$  架无人机对敌方第  $j$  架无人机的优势函数的值;  $m$  为我方无人机数量;  $n$  为敌方无人机数量.

## 2 目标分配

多对多无人机空战的目标分配是指在战场环境、任务需求、无人机性能和敌我无人机相对态势下,为我方无人机分配一个或多个敌方无人机或目标点,使得所有的敌方目标都能被攻击且整体攻击效能达到最优,即

$$\max \left( \sum_{j=1}^n G_{ij} UAV_i^j \right) \quad i = 1, 2, \dots, m \quad (10)$$

$$UAV_i^j \in \{0, 1\}, \quad (11)$$

$$\sum_{i=1}^m UAV_i^j \leq l \quad j = 1, 2, \dots, n, 1 \leq l \leq m, \quad (12)$$

$$\sum_{i=1}^m \sum_{j=1}^n UAV_i^j \geq n. \quad (13)$$

式(10)为适应度函数即目标优化函数,表示搜索能够使整体效能最大的目标分配方案,其中:  $G_{ij}$  表示无人机  $i$  对无人机  $j$  的优势函数的值;式(11)~(13)为约束条件,  $UAV_i^j$  为 1 表示将敌方无人机  $j$  分配给我方无人机  $i$ , 否则为 0;式(12)表示我方最多分配  $l$  架无人机同时攻击敌方无人机  $j$ ;式(13)表示我方无人机至少需要同时攻击  $n$  个目标.

这一数学模型随着无人机数量的增加,复杂度呈指数级增加,是典型的 NP 完全问题. 本文采用改进混合粒子群优化算法进行求解.

### 2.1 混合粒子群优化算法

粒子群优化算法是模拟鸟类觅食行为的一种智能算法,能通过追随个体极值和群体极值完成极值寻优. 如下是粒子群优化算法更新粒子的公式:

$$\begin{aligned} V_{k+1} &= \omega V_k + c_1 r_1 (P_{\text{best},k} - X_k) + c_2 r_2 (G_{\text{best},k} - X_k), \\ X_{k+1} &= X_k + V_{k+1}. \end{aligned} \quad (14)$$

其中:  $\omega$  是惯性权重;  $k$  为当前迭代次数;  $X$  表示粒子的位置;  $V$  为粒子的速度;  $c_1$  和  $c_2$  为学习因子;  $r_1$  和  $r_2$  是分布于  $[0, 1]$  区间的随机数;  $P_{\text{best}}$  和  $G_{\text{best}}$  分别是个体极值和群体极值.

混合粒子群优化算法引入了遗传算法的交叉与变异操作,通过粒子与个体极值和群体极值进行交叉以及粒子本身的变异来搜索最优解。

混合粒子群优化算法中的交叉操作是指:随机选定两个交叉位置,然后将个体和个体极值或群体极值位于交叉位置之间的元素进行互换,然后检查新产生的个体的合法性,对不合法的元素进行调整。对新产生的个体采用保留优秀个体的策略,只有当新粒子适应度好于旧粒子时才更新粒子。混合粒子群优化算法中的变异操作是指:随机选择两个变异位置,将粒子中的两个变异位置的元素互换,同样对新产生的个体采用保留优秀个体的策略。

## 2.2 改进的混合粒子群优化算法

混合粒子群优化算法摒弃了标准粒子群优化算法中的通过跟踪极值来更新粒子位置的方法,避免了人为设置的惯性权重、粒子速度以及学习率对更新粒子群的影响,使得混合算法兼具遗传算法和粒子群优化算法的优点。但混合粒子群优化算法的迭代前期较短,收敛速度快,迭代后期则较长,随着迭代次数的增加,在种群收敛的同时,各粒子也越来越相似,从而求解可能在局部最优解附近无法跳出。

针对此问题,本文提出改进方法,增大迭代后期粒子的变化。引入判定参数 $\lambda$ ,在迭代过程中,如果相邻若干代数的群体极值之差小于 $\lambda$ ,则认为算法陷入局部最优,此时引入移民粒子,计算其适应度,并与现有粒子进行比较,选择适应度高的粒子保留。当相邻相同代数的全局极值之差又小于 $\lambda$ 时,再次引入移民粒子并更新粒子群,直到迭代结束。因为只在算法陷入局部最优时才引入新粒子,所以该算法在提高全局搜索能力的同时,不会额外增加程序的计算量。

改进混合粒子群优化算法的步骤如下:

步骤1 初始化,设定最大迭代次数 $H$ ,设定粒子种群数 $K$ ,随机产生 $K$ 个可行分配方案作为初始粒子,计算各粒子的适应度,记录初始个体极值 $P_{best}$ 与全局极值 $G_{best}$ 。

步骤2 对种群中的每个粒子与 $P_{best}$ 进行交叉操作,计算交叉产生的两个粒子的适应度,选择适应度好的成为新粒子,如果新粒子的适应度优于 $P_{best}$ ,则对 $P_{best}$ 进行更新。

步骤3 对步骤2中得到的新粒子与 $G_{best}$ 进行交叉操作,计算交叉产生的两个粒子的适应度,选择适应度好的成为新粒子,如果新粒子的适应度优于 $P_{best}$ ,则对 $P_{best}$ 进行更新。

步骤4 对步骤3中得到的新粒子进行变异操作,计算变异操作后粒子的适应度,如果新粒子的适应度优于 $P_{best}$ ,则对 $P_{best}$ 进行更新。

步骤5 根据更新后的 $P_{best}$ 确定 $G_{best}$ ,判断是否达到最大迭代次数,如果不是,转到步骤6,如果是,则结束迭代。

步骤6 计算若干代的 $G_{best}$ 之差,如果大于 $\lambda$ ,转到步骤2,如果小于 $\lambda$ ,则引入移民粒子并更新种群,然后转到步骤2。

实际空战中敌我无人机数量往往不相等,而且随着空战的进行,会有双方无人机的加入与击毁,因此双方无人机的数量不相等与变化是一种常态。同时应用混合粒子群优化算法解决实际目标分配问题时,因为约束条件过多,容易遇到粒子群初始化困难的问题。

因此本文建立了一种在不对称、动态空战下的对策方法,用于解决动态空战情况下的不平衡分配问题和粒子群初始化困难的问题。当交战双方无人机数量不相等时,同样建立优势矩阵,当 $m > n$ 时,在矩阵的右边补充 $m - n$ 列常数 $P_1$ :

$$G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1n} & P_1 & \cdots & P_1 \\ G_{21} & G_{22} & \cdots & G_{2n} & P_1 & \cdots & P_1 \\ \vdots & \vdots & \vdots & \vdots & P_1 & \cdots & P_1 \\ G_{m1} & G_{m2} & \cdots & G_{mn} & P_1 & \cdots & P_1 \end{bmatrix}. \quad (15)$$

其中: $P_1$ 为小于所有矩阵元素的常数。当 $m < n$ 时,在矩阵的下方补充 $n - m$ 行常数 $P_2$ , $P_2$ 为大于所有矩阵元素的常数。此时初始粒子群只需要满足是1到新矩阵阶数的正整数随机排列即可。

当  $m > n$  时, 进行分配后会多出  $m - n$  架我方无人机, 此时可以将多余的我方无人机按照任务需求进行灵活再分配. 比如按照防御原则将其分配到对我方目标威胁最大的敌方无人机; 按照盯防原则将其分配到敌方性能最强的无人机; 按照歼灭原则将其分配到对战优势最小的那架敌方无人机.

当有无人机被击落, 新的无人机加入战斗, 或者空战经过一段时间以后, 就再次利用改进混合粒子群优化算法进行目标分配, 以保证目标分配的及时性.

### 3 机动决策

强化学习是机器学习的一个分支, 它可以用于解决序贯决策优化问题. 最典型的强化学习的模型就是马尔可夫决策过程(Markov Decision Process, MDP). MDP 所具有的要素包括状态( $S$ )、行动集( $A$ )、策略( $\pi$ )和奖励( $R$ ). 在本文中, 状态即是敌我双方无人机的位置与速度矢量, 行动集是无人机可以采取的机动动作, 策略即是我方无人机机动选择遵循的原则, 奖励即是优势函数.

#### 3.1 无人机机动

本文将无人机看作 3 自由度质点, 无人机在无侧滑中飞行, 且发动机的推力沿着速度的方向, 则无人机质心运动在坐标系上的投影方程可以简化为

$$\begin{cases} M \frac{dv}{dt} = P - D - Mg \sin \varphi, \\ Mv \frac{d\varphi}{dt} = Y \cos \gamma - Mg \cos \varphi, \\ -Mv \cos \varphi \frac{d\psi}{dt} = Y \sin \gamma. \end{cases} \quad (16)$$

其中:  $M$  为无人机的质量;  $P$  为发动机推力;  $D$  为阻力;  $Y$  为升力;  $g$  为重力加速度;  $\varphi$  为航迹倾斜角;  $\gamma$  为航迹滚转角;  $\psi$  为航迹偏转角. 过载在坐标轴上的投影分量为

$$\begin{cases} n_x = \frac{P - D}{Mg}, \\ n_y = \frac{Y \cos \gamma}{Mg}, \\ n_z = \frac{Y \sin \gamma}{Mg}. \end{cases} \quad (17)$$

式(17)中:  $n_x$  沿着速度方向, 为切向过载;  $n_y$  和  $n_z$  垂直于速度矢量, 其合过载  $n_f = \sqrt{n_y^2 + n_z^2}$  为法向过载.

将式(17)代入式(16), 则运动方程可改写为

$$\begin{cases} \frac{dv}{dt} = g(n_x - \sin \varphi), \\ \frac{d\varphi}{dt} = \frac{g}{v}(n_f \cos \gamma - \cos \varphi), \\ \frac{d\psi}{dt} = -\frac{gn_f \sin \gamma}{v \cos \varphi}. \end{cases} \quad (18)$$

选取具有代表性的 7 种机动作为可选机动:  $A = \{\text{左偏置, 稳定飞行, 最大加速, 最大减速, 右偏置, 爬升, 俯冲}\}$ , 涉及过载的机动均以最大过载进行.

#### 3.2 基于 Q-learning 算法的机动决策

Q-learning 算法是强化学习领域中的一种具有突出优势的时间差分学习算法, 它不依赖于环境模型, 并且可以很好地解决信用分配问题. 如何评价无人机空战中某一机动决策对未来决策乃至将来能否击落敌机的影响, 就是其实际应用的典型例子. Q-learning 算法利用当前决策时刻  $T = k$  往后  $n$  步回报的采样加上  $T = k + n$  步的估计值来评估当前决策的优劣. 以某一敌我无人机配对为例, 算法的决策步骤如下:



步骤1 以敌我双方无人机状态生成当前时刻到其后  $n$  步的优势函数矩阵记为奖励矩阵  $\mathbf{R}$ ; 以任意的方式初始化  $\mathbf{Q}$  矩阵, 设定迭代次数  $H$ .

步骤2 用  $\epsilon$ -贪婪法, 在当前状态  $S$  下选择机动动作, 执行机动动作, 获得奖励  $\mathbf{R}(S, A)$  和下一步的状态  $S'$ . 利用贝尔曼公式更新  $\mathbf{Q}$  矩阵:

$$\mathbf{Q}(S, A) = \mathbf{Q}(S, A) + \alpha [\mathbf{R}(S, A) + \gamma \max_{a \in A} \mathbf{Q}(S', a) - \mathbf{Q}(S, A)]. \quad (19)$$

其中:  $\alpha$  是学习率;  $\gamma$  是折扣率;  $\epsilon$  为一个常数, 当生成的随机数大于  $\epsilon$  时, 选择  $\mathbf{Q}$  矩阵中最优的机动动作, 当生成的随机数小于等于  $\epsilon$  时, 则随机选择动作. 更新当前的状态  $S \rightarrow S'$ . 重复进行  $n$  次, 视为一次迭代.

步骤3 完成设定的迭代次数后, 依照生成的  $\mathbf{Q}$  矩阵, 选择我方无人机当前时刻的下一步机动.

### 3.3 改进的 Double Q-learning 算法

在 3.2 节中利用  $\epsilon$ -贪婪法进行下一步机动动作的选择, 其中  $\epsilon$  是一个常数, 每一步有  $\epsilon$  的概率采用随机机动动作, 有  $1-\epsilon$  的概率采用当前已知最优的机动动作, 也就是未知的探索与已知的利用. 但是将  $\epsilon$  设置为一个常数不够合理, 因为算法前期,  $\mathbf{Q}$  矩阵的大部分处于未探索的状态, 采用已知最优机动会降低  $\mathbf{Q}$  矩阵探索的效率, 而在算法后期,  $\mathbf{Q}$  矩阵大部分已经被探索, 继续选择探索会降低收敛的速度. 因此本文采用变  $\epsilon$ -贪婪法:

$$\epsilon = \begin{cases} 1 - \frac{H'}{H} & 0 \leq H' < 0.95H, \\ 0.05 & 0.95H \leq H' \leq H. \end{cases} \quad (20)$$

其中:  $H'$  为当前正在迭代的轮数, 即已经完成了  $H'-1$  次迭代, 正在进行第  $H'$  次迭代. 算法前期  $\epsilon$  值相对较大, 鼓励算法进行探索; 算法后期  $\epsilon$  逐渐减小, 鼓励算法进行利用, 从而平衡了探索与利用的关系, 使得  $\mathbf{Q}$  矩阵的收敛速度提高.

Q-learning 算法中目标策略基于动作值的最大化, 即  $\max_{a \in A} \mathbf{Q}(S', a)$ . 算法将估计值中的最大值作为预测值, 会带来正向偏置的问题. 例如在某一状态  $S$  下, 无人机的真实动作值  $\mathbf{Q}$  均为零, 但是对这些机动动作的估计值  $\mathbf{Q}$  存在大于零或者小于零的情况, 此时选择最大的估计值会使得最终结果产生正向偏置.

为了避免正向偏置的问题, 本文采用 Double Learning 算法与 Q-learning 算法相结合的 Double Q-learning 算法. 它的核心思想在于交替更新两个独立的预测值:

$$\begin{cases} \mathbf{Q}_1(S, A) = \mathbf{Q}_1(S, A) + \alpha [\mathbf{R}(S, A) + \gamma \cdot \mathbf{Q}_2(S', \arg\max_{a \in A} \mathbf{Q}_1(S', a)) - \mathbf{Q}_1(S, A)], \\ \mathbf{Q}_2(S, A) = \mathbf{Q}_2(S, A) + \alpha [\mathbf{R}(S, A) + \gamma \cdot \mathbf{Q}_1(S', \arg\max_{a \in A} \mathbf{Q}_2(S', a)) - \mathbf{Q}_2(S, A)]. \end{cases} \quad (21)$$

进而解决状态值估计过程中的正向偏置问题. 而且即使有两个估计值, 每次迭代只需要更新其中的一个值, 实际的计算量并不会增加.

## 4 仿真测试

假设红方与蓝方的无人机性能均一致, 最大速度为 500 m/s, 最大法向过载为 3g, 最大切向过载为 8g ( $g$  为重力加速). 空战的决策周期为 0.25 s, 在同一决策周期内执行一种机动动作. 敌方无人机采用简单最大策略, 采取能够使得其优势函数最大的机动; 我方无人机采用最大最小值策略, 采用能够使得敌方最大优势函数值取得最小的机动. 规定: 当一方无人机对方无人机的超前角与滞后角均小于  $30^\circ$  (机载雷达的扫描角度) 且两机距离小于 500 m 时 (机载武器有效射程) 即视为该方无人机击败对方无人机. 一方无人机击败半数及以上对方无人机时, 空战无悬念, 视为该方胜利.

以 12 对 12 的无人机空战的目标分配为例, 验证改进混合粒子群优化算法的性能, 将其与遗传算法和标准混合粒子群优化算法做对比. 表 1 (见第 416 页) 所示是无人机的初始位置信息.

表 1 12 对 12 空战中无人机的初始位置与速度  
Tab. 1 Initial positions and velocities of UAVs in 12 to 12 air combat

UAV	$x/m$	$y/m$	$z/m$	$\phi/(^{\circ})$	$\varphi/(^{\circ})$	$v/(m \cdot s^{-1})$	UAV	$x/m$	$y/m$	$z/m$	$\phi/(^{\circ})$	$\varphi/(^{\circ})$	$v/(m \cdot s^{-1})$
$B_1$	4 500	220	5 800	0	0	270	$R_1$	1 200	500	5 800	10	0	240
$B_2$	0	0	5 600	0	0	280	$R_2$	3 000	500	5 600	10	0	250
$B_3$	2 000	-100	5 700	0	0	290	$R_3$	2 500	1 000	5 100	10	0	280
$B_4$	6 500	0	5 300	0	0	280	$R_4$	6 000	1 000	5 300	10	0	270
$B_5$	4 300	-390	5 400	0	0	270	$R_5$	5 500	2 200	5 800	10	0	290
$B_6$	3 300	-600	5 500	0	0	260	$R_6$	4 500	2 000	5 700	10	0	270
$B_7$	1 300	-300	5 700	0	0	290	$R_7$	1 500	1 500	5 500	10	0	260
$B_8$	5 300	-200	5 400	0	0	270	$R_8$	5 000	2 600	5 300	10	0	290
$B_9$	4 600	-240	5 700	0	0	270	$R_9$	500	2 200	5 800	10	0	290
$B_{10}$	3 300	-60	5 500	0	0	260	$R_{10}$	3 500	2 000	5 700	10	0	270
$B_{11}$	2 300	-590	5 700	0	0	290	$R_{11}$	1 900	500	5 500	10	0	260
$B_{12}$	3 000	-490	5 400	0	0	270	$R_{12}$	3 700	1 600	5 300	10	0	290

算法的参数设置如下:

- 1) 遗传算法 初始个体数为 50,交叉概率为 90%,变异概率为 5%;
- 2) 标准混合粒子群优化算法 初始粒子数为 5;
- 3) 改进混合粒子群优化算法 初始粒子数为 5,连续相邻代数为 15,是否引入移民粒子的判定参数为 0.1.

图 2 是 3 种算法迭代次数为 50 的算法训练图.由图中曲线可以看出,遗传算法、混合粒子群优化算法、改进的混合粒子群优化算法均能有效地求解 12 架蓝方无人机对 12 架红方无人机的目标分配问题.总体上看改进的混合粒子群优化算法的效果最好,能够使目标分配以后,我方无人机的总优势函数最大,对敌方形形成最大的威胁,其次是混合粒子群优化算法,遗传算法的效果最差.图 3~图 5 为 3 种算法的匹配结果.现对每种算法仿真 20 次,每次迭代 50 代,将仿真计算出的结果进行统计,并将其与精确解进行比较,结果如表 2 所示.

从表 2 可以看出,改进的混合遗传算法的平均用时为 0.158 s,与混合粒子群优化算法时间(0.141 s)相当,并且其的最大值、最小值和平均值均不劣于遗传算法和混合粒子群优化算法的,其的最大值与精确解一致,平均值与精确解存在 1.6% 的差异.混合粒子群优化算法的最小值和平均值均劣于改进的混合粒子群优化算法的,平均值与精确解的差异为 2.4%.遗传算法的平均用时最长,为 0.379 s,仿真结果均劣于混合粒子群算法和改进的混合粒子群优化算法的,其最大值与精确解的差异为 2.8%,平均值的差异为 6.4%.同时混合粒子群算法和改进的混合粒子群优化算法的初始样本数仅为遗传算法的  $\frac{1}{10}$ .

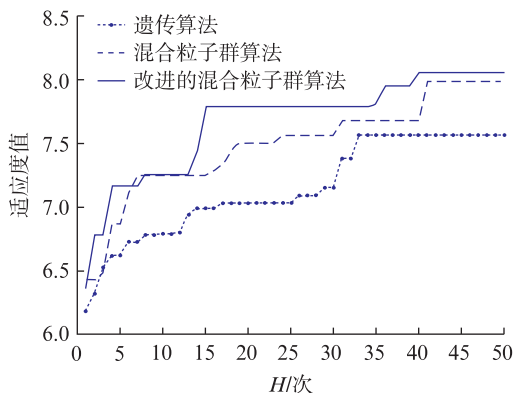


图 2 3 种算法的训练过程

Fig. 2 Training process of three algorithms

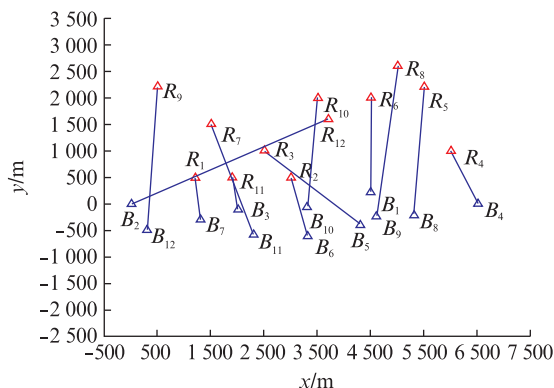


图 3 遗传算法的匹配图

Fig. 3 Matching graph of genetic algorithm

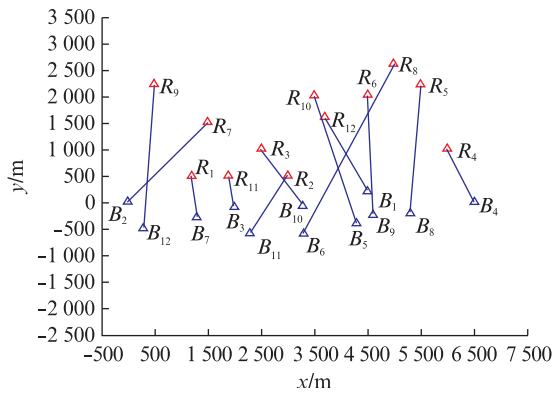


图 4 混合粒子群优化算法的匹配图

Fig. 4 Matching graph of hybrid particle swarm optimization algorithm

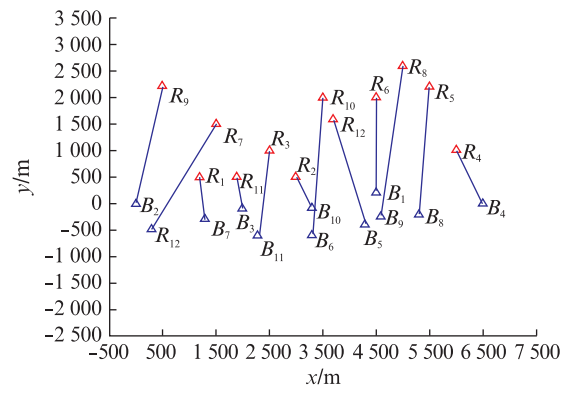


图 5 改进的混合粒子群优化算法的匹配图

Fig. 5 Matching graph of improved hybrid particle swarm optimization algorithm

表 2 3 种算法的训练结果

Tab. 2 Training results of three algorithms

算法(精确解)	最大值	最小值	平均值	标准方差 <sup>1)</sup>	最大值的差异 <sup>1)</sup> /%	均值的差异 <sup>1)</sup> /%	平均用时/s
精确解	8.141 5	8.141 5	8.141 5	/	/	/	/
遗传算法	7.912 5	7.374 9	7.621 2	0.146	2.80	6.40	0.379
混合粒子群算法	8.141 5	7.649 5	7.945 8	0.141	0.00	2.40	0.141
改进的混合粒子群算法	8.141 5	7.807 9	8.013 8	0.093	0.00	1.60	0.158

注：1)表示为算法所得数值与精确解的差异；/表示无此项数值。

选取一组小规模集团作为多对多无人机机动决策的仿真算例,为方便起见对其重新编号,我方和敌方无人机的位置和速度信息如表 3 所示。

表 3 4 对 4 空战中无人机的初始位置与速度

Tab. 3 Initial positions and velocities of UAVs in 4 to 4 air combat

UAV	$x/m$	$y/m$	$z/m$	$\phi/(^{\circ})$	$\varphi/(^{\circ})$	$v/m \cdot s^{-1}$	UAV	$x/m$	$y/m$	$z/m$	$\phi/(^{\circ})$	$\varphi/(^{\circ})$	$v/m \cdot s^{-1}$
$B_1$	4 500	0	5 800	0	0	270	$R_1$	1 200	500	5 800	10	0	270
$B_2$	0	0	5 800	0	0	270	$R_2$	3 000	500	5 800	10	0	270
$B_3$	2 000	0	5 800	0	0	270	$R_3$	2 500	1 000	5 800	10	0	270
$B_4$	6 500	0	5 800	0	0	270	$R_4$	6 000	1 000	5 800	10	0	270

将无人机的位置信息绘制到坐标系中,俯视图如图 6(见第 418 页)所示。图 6 中箭头方向为无人机的速度方向。利用改进的混合粒子群优化算法得到的匹配结果是  $B_1 > R_2, B_2 > R_1, B_3 > R_3, B_4 > R_4$ , 与精确解一致。图 7(见第 418 页)为无人机空战过程的仿真图,由程序模拟可得:在 A 点,  $T=30$ (第 30 次迭代,即进行第 30 次机动决策)时,  $B_4$  击败  $R_4$ ,随后在 B 点按照歼灭原则匹配到新目标  $R_1$ ,与  $B_2$  一同追击  $R_1$ 。最后在 C 点,  $T=35$  时,  $B_3$  击败  $R_3$ 。蓝方获得空战胜利。

图 8(见第 418 页)所示是蓝方无人机的优势函数,蓝方无人机的优势函数始终为正,表明蓝方对红方一直保持优势。在 D 点,  $B_4$  的优势函数达到极大值,对应击败  $R_4$ ,随后  $R_1$  成为目标,此时  $B_4$  的优势函数快速降低。在  $T=35$  时,  $B_3$  的优势函数达到极大值,击败  $R_3$ 。

以  $B_3$  与  $R_3$  配对为例进行机动决策算法的时效性验证:设置学习率  $\alpha=1$ ,折扣率  $\gamma=0.8$ ,迭代次数为 900,决策时长为 0.025 s。本文程序的运行环境为:处理器为 intel-i7 5700,内存 8 G,仿真软件为 Matlab R2016a。实际空战中目标分配和机动决策可由运算能力更强的工作站或者长机进行,因此能够满足空战的要求。



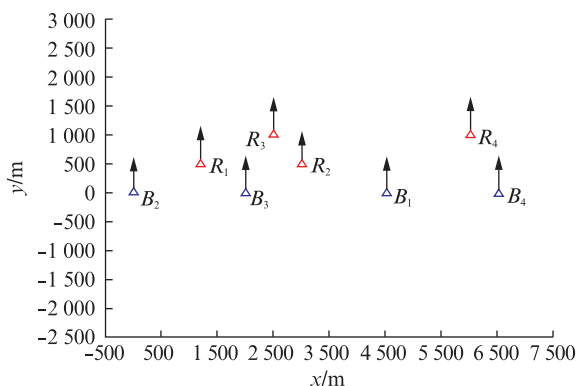


图 6 无人机的初始位置

Fig. 6 Initial position of UAVs

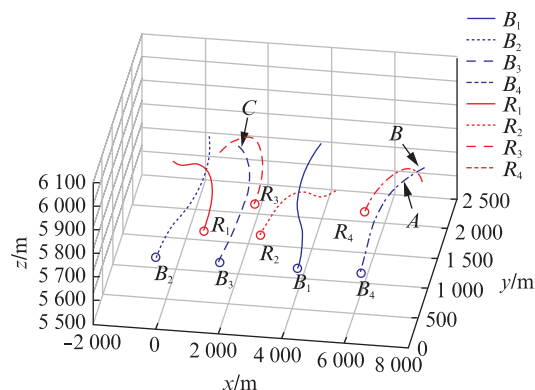


图 7 空战的仿真图

Fig. 7 Simulation chart of air combat

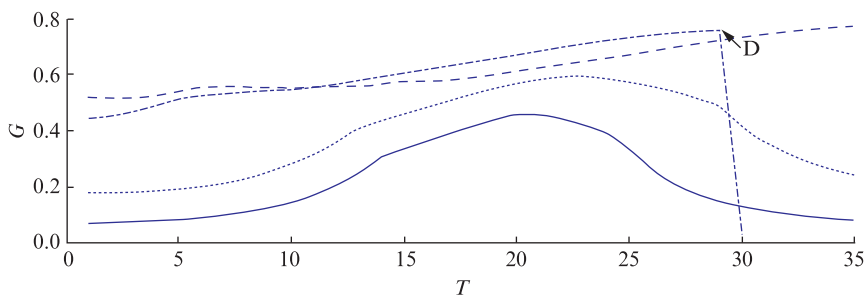


图 8 蓝方的优势函数

Fig. 8 Comprehensive advantage function of Blue

## 5 结 语

本文对多对多无人机空战样式进行了决策仿真,仿真结果表明引入了局部最优判定和加入移民粒子的改进的混合粒子群优化算法的目标分配具有比普通算法更快的收敛速度和更优的分配结果,应用改进的 Double Q-learning 算法的我方无人机能够很好地执行追踪、击败敌机的任务.并且我方无人机在击败敌机后,能够加入其他我方无人机对敌方目标的追击,最终多机协同取得空战的胜利.未来研究将会针对不同性能的无人机,进行更多限制条件下的仿真验证.

### 参考文献:

- [1] 高劲松,陈峭东. 国外六代机发展情况研究[J]. 飞航导弹, 2014(1): 54-59.
- [2] 周思羽,吴文海,张楠,等. 自主空战机动决策方法综述[J]. 航空计算技术, 2012, 42(1): 27-31.
- [3] 柳毅,佟明安. 匈牙利算法在多目标分配中的应用[J]. 火力与指挥控制, 2002, 27(4): 34-37.
- [4] WHITTEN A K, CHOI H L, JOHNSON L B, et al. Decentralized task allocation with coupled constraints in complex missions[C]//American Control Conference. Berlin, Germany: Springer-Verlag, 2011: 1642-1649.
- [5] CAO Y, WEI W, BAI Y, et al. Multi-base multi-UAV cooperative reconnaissance path planning with genetic algorithm[J]. Cluster Computing, 2019, 22(S3): 5175-5184.
- [6] 朱德法,单连平,管莹莹. 基于改进粒子群优化算法的多机协同目标分配[J]. 火力与指挥控制, 2015, 40(8): 38-41.
- [7] SCHWARZROCK, JANAÍNA, ZACARIAS I, BAZZAN A L C, et al. Solving task allocation problem in multi unmanned aerial vehicles systems using swarm intelligence[J]. Engineering Applications of Artificial Intelligence, 2018, 72(C): 10-20.

- [8] PARK H, LEE B Y, TAHK M J, et al. Differential game based air combat maneuver generation using scoring function matrix[J]. *International Journal of Aeronautical and Space Sciences*, 2015, **17**(2): 204-213.
- [9] 傅莉,王晓光. 无人战机近距空战微分对策建模研究[J]. *兵工学报*, 2012, **33**(10): 1210-1216.
- [10] PAN Q, ZHOU D, HUANG J, et al. Maneuver decision for cooperative close-range air combat based on state predicted influence diagram [C]//IEEE International Conference on Information and Automation, China Macao, China: IEEE, 2017: 726-731.
- [11] 王炫,王维嘉,宋科璞,等. 基于进化式专家系统树的无人机空战决策技术[J]. *兵工自动化*, 2019, **38**(1): 48-53.
- [12] ERNEST N, CARROLL D, SCHUMACHER C, et al. Genetic fuzzy based artificial intelligence for unmanned combat aerial vehicle control in simulated air combat missions [J]. *Journal of Defense Management*, 2016, **6**(1): 144.
- [13] 何旭,景小宁,冯超. 基于蒙特卡洛树搜索方法的空战机动决策[J]. *空军工程大学学报(自然科学版)*, 2017, **18**(5): 36-41.
- [14] KARIMI I, POURTAKDOUST S H. Optimal maneuver-based motion planning over terrain and threats using a dynamic hybrid PSO algorithm[J]. *Aerospace Science & Technology*, 2013, **26**(1): 60-71.
- [15] HASSELT H V. Double Q-learning [M]//Advances in Neural Information Processing Systems. San Francisco, USA: Morgan Kaufmann, 2010: 2613-2621.
- [16] SCHULZE C, SCHULZE M. ViZDoom: DRQN with prioritized experience replay, double-Q learning, & snapshot ensembling [M]//Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference(IntelliSys) Volume 1. Berlin, Germany: Springer, 2018: 1-17.

## Research on Intelligent Decision Making of Many to Many Unmanned Aerial Vehicle Air Combat

ZHU Xingyu, AI Jianliang

(Department of Aeronautics and Astronautics, Fudan University, Shanghai 200433, China)

**Abstract:** An air combat process is designed for UAVs capable of autonomous air combat. An evaluation system based on the position, attitude and energy of both sides is established. The target allocation of air combat is realized by improved hybrid particle swarm optimization algorithm. The maneuvering decision-making strategy of UAVs is designed with improved double Q-learning algorithm. Simulation is performed in the form of four-to-four air combat scenario. The simulation results show that UAVs can track and destroy opponent UAVs well, and can also cooperate with other UAVs to attack opponent UAVs. And algorithms have good timeliness. It shows that the proposed method can effectively solve the problem of many to many UAV air combat.

**Keywords:** maneuvering decision-making; target allocation; improved hybrid particle swarm optimization algorithm; improved double Q-learning algorithm