引用格式: 毛梦月,张安,周鼎,等. 基于机动预测的强化学习无人机空中格斗研究[J]. 电光与控制,2019,26(2):5-10,22. MAO M Y, ZHANG A, ZHOU D, et al. Reinforcement learning of UCAV air combat based on maneuver prediction[J]. Electronics Optics & Control, 2019, 26(2):5-10, 22.

基于机动预测的强化学习无人机空中格斗研究

毛梦月, 张 安, 周 鼎, 毕文豪
(西北工业大学,西安 710086)

摘 要: 在无人机空中格斗过程中,由于无人机自身状态以及空战态势,敌我双方机动动作及行为策略的选择具有极强的不确定性。针对这个问题,将强化学习方法引入无人机空中格斗过程,建立无人机机动模型及动作集;将空战态势评估函数作为强化学习中的信号函数;采用概率神经网络(PNN)作为对敌机动预测单元;在敌我双方战场信息完全感知条件下,该算法能够不断学习,使无人机通过与环境的交互来掌握其最佳机动行为策略,实现无人机的一对一空中对抗。

关键词: 无人机; 空中格斗; 机动预测; 态势评估; 强化学习; 概率神经网络
中图分类号: V271.4 **文献标志码:** A **doi:** 10.3969/j.issn.1671-637X.2019.02.002

Reinforcement Learning of UCAV Air Combat Based on Maneuver Prediction

MAO Meng-yue, ZHANG An, ZHOU Ding, BI Wen-hao
(Northwestern Polytechnical University, Xi'an 710086, China)

Abstract: Due to the status of the Unmanned Combat Aerial Vehicle (UCAV) itself and the situation of the air combat, there are many uncertainties in UCAV aerial combat, including the maneuvering and choosing of behavioral strategy of the two sides. To solve this problem, the reinforcement learning method is introduced into UCAV air combat, and the UCAV maneuvering model and action set are established. The function of air combat situation assessment is taken as the signal function for reinforcement learning, and the Probabilistic Neural Network (PNN) is used as the unit for enemy maneuver prediction. Under the condition of thorough perception of battlefield information of both sides, the UCAV can grasp the best maneuvering strategy through interacting with the environment by reinforcement learning, thus to implement one-on-one aerial combat.

Key words: UCAV; air combat; maneuver prediction; situation assessment; reinforcement learning; probabilistic neural network

0 引言

无人机空中格斗作为未来空战的理想形态,一直倍受国内外学者关注,而如何利用无人机信息化、智能化的攻击能力,保证无人机能够自主完成复杂的空战过程是研究的焦点,目前已有大量成果涌现[1-9]。然而,无人机自主空战过程因具有复杂的、动态的作战过程以及极强的不确定性,使得基于专家知识等方法所设计的决策系统缺乏足够完备性和灵活性,在空战机动决策、态势评估和目标攻击过程中具有较强的主观性[10-11]。

收稿日期: 2017-11-24 修回日期: 2018-03-27
基金项目: 国家自然科学基金(61573283)
作者简介: 毛梦月(1993—),男,陕西西安人,硕士生,研究方向为无人机自主空战、路径规划。

本文研究了一种基于机动预测的强化学习无人机空中格斗方法,将强化学习与无人机空中格斗过程相结合,建立无人机机动动作库以及战场双方威胁评估函数;采用概率神经网络(PNN)方法对敌机机动动作进行预测,根据预测结果,强化学习单元将结合无人机状态、空战态势等选择"收益"最高的对敌机动策略,实现无人机一对一空中对抗。

1 基于机动预测的智能体强化学习

1.1 智能体强化学习

随机策略可作为智能体的一般化马尔可夫决策过程[12],用元组 $\langle S; A^1 - A^n; P^1 - P^n; R^1 - R^n \rangle$ 表示。其中: n 为系统中的智能体个数; S 表示状态集; A^i 为智能体可选择的动作集合; $P^i(s; a; s') \in [0,1]$ 为状态转移函数; R_i 为强化信号函数,即回报函数奖励值。在多

智能体系统中，状态转移是系统中所有智能体动作选择的结果，故将状态到动作的映射称为策略 π，而获得最大的奖励值是强化学习的目标，其对应的策略为最优策略 $\pi^{\#}$；因此，多智能体强化学习方法转化为在策略 π 下的状态空间到动作空间的映射学习。

Q 学习是强化学习算法之一，其中的函数依赖于所有智能体的执行动作，基于以上定义，根据 HU[13] 提出的基于一般和对策学习方法，智能体在 t 时刻的函数更新规则可表示为

$$Q_t^i(s_t^i,a^t)=(1-\alpha_t)Q_{t-1}^i(s_t^i,a^t)+\alpha_t[r_t^i+\pi^1(s_{t+1})\cdots\pi^n(s_{t+1})Q_{t-1}^i(s_{t+1}^i)] \tag{1}$$

$$\pi^1(s_{t+1})\cdots\pi^n(s_{t+1})Q_{t-1}^i(s_{t+1}^i)=\sum_{a^1\in A}\cdots\sum_{a^n\in A}f_t^1(s_{t+1},a^1)\cdots f_t^n(s_{t+1},a^n)Q_{t-1}^i(s_{t+1}^i,a^1,\cdots,a^n) \tag{2}$$

式中：s_t^i 为智能体 i 的状态变量；$a^t=\{a^1,\cdots,a^n\}$ 表示智能体基本动作集；s_{t+1}^i 为下一时刻智能体的状态，其状态转移根据函数 $s_{t+1}=P_t^i(s_t,a_t)$ 确定。智能体机动策略用其动作集合的概率分布 π^i 表示，$f_t^i(s_{t+1},a^i)$ 为智能体 i 在状态 s_{t+1} 下选择动作 a^i 的概率。

对于一对一 UCAV 自主空战而言，由于双方 UCAV 都是同时选择机动动作，这样红方 UCAV 是无法得知蓝方 UCAV 将要选择执行什么机动动作；然而对于大多数学习问题而言，蓝方 UCAV 的机动动作并不是随意发生的，而是根据一定概率分布的机动选择策略，由此可通过预测蓝方 UCAV 的机动行为来改进空战决策算法。因此，红方 UCAV 的机动决策算法由机动预测单元和强化学习单元两部分构成，其结构见图 1。

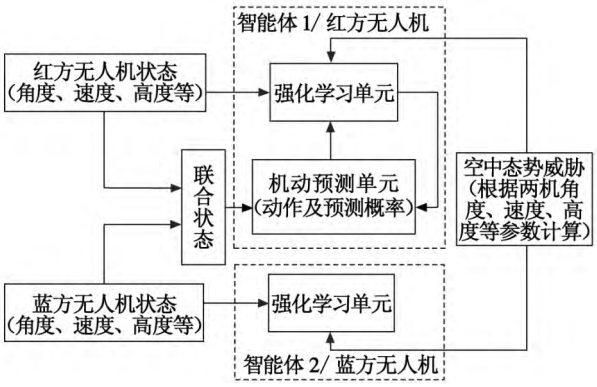


图 1 红蓝双方无人机机动决策算法结构框图

Fig. 1 Algorithm structure of maneuvering decision-making for UCAVs of two sides

在该系统中：红方无人机的机动决策算法包括强化学习单元和机动预测单元两部分，机动预测单元使用基于概率的方法来预测蓝方无人机的机动动作，并向学习单元提供蓝方无人机可能选择的动作及预测概率，完成强化学习算法，然后强化学习单元将机动策略选择结果返回给动作预测单元来更新预测模型；而蓝方无人机的机动决策算法则只有强化学习单元。

1.2 机动预测单元

在 UCAV 一对一空战过程中，红方 UCAV 采取行动时需要考虑敌机的机动动作及机动行为，因此，采用概率神经网络(PNN)作为对敌机动动作预测单元，对敌机机动动作选择进行预测，预测结果为蓝方 UCAV 可能选取的机动动作所对应的概率；将此概率作为我机机动的参考，使得我机在强化学习单元能够快速得到更高的奖励回报。

概率神经网络(PNN)[14] 是一种常用于模式分类的神经网络，具有如下优点：学习过程简单、训练速度快；分类更准确，容错性好等；通过 PNN 对样本进行分类从而预测敌机机动动作。其网络结构如图 2 所示。

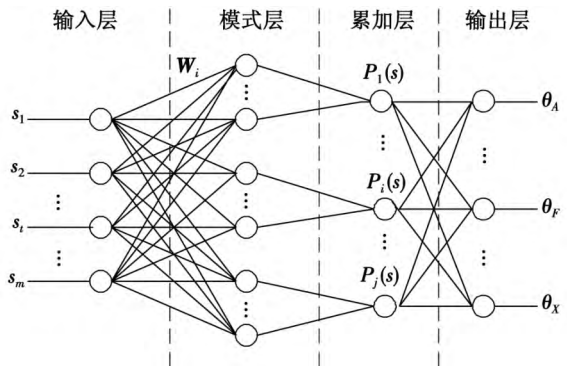


图 2 PNN 网络结构

Fig. 2 PNN structure

对蓝方无人机的机动预测方法是将红蓝双方的联合状态 S 作为网络输入矢量，蓝机的候选机动动作作为网络输出的决策类别 $\theta_A,\cdots,\theta_F,\cdots,\theta_X$，然后通过网络的推理得到在联合状态下蓝机选择各机动动作的概率。PNN 的 4 层前向网络结构分别为输入层、模式层、累加层和输出层。输入层的作用是将输入矢量 S 直接传给模式层的各节点；模式层完成输入模式矢量 S 与给定类的权矢量 W_i 的加权求和，并将结果进行非线性操作后传递给累加层；累加层是将输入矢量属于同一类别的概率进行累计，并将结果传递给输出层；输出层是一种竞争神经元，它接收来自累加层的各类概率密度函数。

PNN 概率密度函数算式为

$$P_i(s)=\frac{1}{(2\pi)^{m/2}\sigma^m}\cdot\frac{1}{n_F}\sum_{j=1}^{n_F}\exp\left[-\frac{(s-s_{Fj})^{\mathrm{T}}(s-s_{Fj})}{2\sigma^2}\right] \tag{3}$$

式中：m 为输入模式矢量的维数；s_{Fj} 是 K 类的第 j 个训练样本矢量；n_F 是 K 类训练的样本矢量个数；σ 是光滑系数，用来调整密度函数。

在 PNN 决策层中,采用贝叶斯准则来判断所属类别 $\theta \in \theta_X$ 的状态,选择具有最小“风险”的类别,即最大后验概率:

当 $h_X l_X P_X(s) > h_Y l_Y P_Y(s)$ 时

$$d(s) \in \theta_X \tag{4}$$

其中: h_X,h_Y 分别为 $\theta \in \theta_X$,$\theta \in \theta_Y$ 的先验概率; l_X 为 θ 本属于 θ_X 时却被错分为其他类的损失。

由上式分析可知,可以使用 PNN 来预测智能体所要选择的动作。在多智能体强化学习中,联合状态 S 作为输入的模式矢量,输入层节点个数由 S 分量个数确定,智能体的动作空间作为决策类别,候选动作个数决定累加层的节点个数。所以智能体动作预测就相当于使用 PNN 将输入的联合状态矢量分类给它所属的动作。动作预测单元和强化学习单元在学习过程中同时进行,最终实现完善的动作预测和动作选择策略。PNN 模式层节点根据决策类别进行分组,每组中节点的权值表示训练样本矢量。在强化学习过程中,每一步的学习样例被不断补充到模式层的相应分组中,同时更新属于各类的样本个数。根据式(3)可以计算输入的联合状态矢量 S 输入 K 类的概率 $P_i(s)$,最终智能体在联合状态 S 下选择动作 D_K 的条件概率可以表示为

$$P(D_K \mid s) \propto h_K l_K P_K(s) \tag{5}$$

其中,选择动作 D_K 的先验概率 h_K 可以从学习过程中动作 D_K 出现的频率来估计,即

$$h_K = R_K / R \tag{6}$$

式中: R_K 为输入联合状态矢量集合选择动作中属于类别 D_K 的样本个数; R 为训练样本总数。

因此,在上文所述系统中,蓝方无人机动作由 PNN 进行预测,并将预测结果返回至红方无人机强化学习单元中,然后通过式(1)与式(2)更新 Q 值,从而实现 Q 学习。

2　无人机空中格斗环境建模

针对无人机一对一空中格斗,双方在有限的三维空间内展开对抗,两架无人机对彼此位置、速度等基本信息完全感知,其根据强化单元各自选择最佳机动动作,并计算相应空中态势,使得自身到达最佳攻击位置,获得最佳攻击角度。

2.1　无人机状态模型

三维战场空间中,红蓝双方共两架无人机,根据空战状态的评估参数[9],选取参与格斗的无人机自身状态信息及双方相对位置、速度关系等作为格斗的初始状态,用向量 S 表示,$S = [q_r, q_b, d, \beta, \Delta h, \Delta v^2, v_r^2, h]$,其相互关系可用图 3 表示。

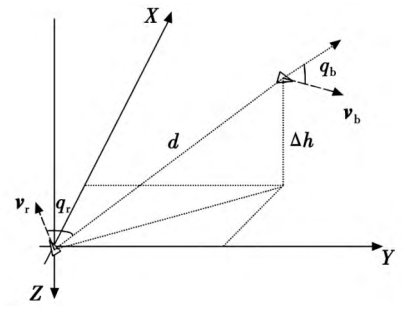


图 3　红蓝两机位置状态关系

Fig. 3　State of the UCAVs of two sides

以红方飞机为中心,取水平面 X–Y 为平面,其中,X 轴与红方速度 v_r 在水平面的投影一致,X 轴顺时针转 90°为 Y 轴方向,Z 轴则由右手准则确定。在向量 S 中: q_r 为偏离角,表示红方无人机速度矢量方向与红蓝无人机质心方向的夹角; q_b 为脱离角,表示蓝方无人机速度矢量和蓝红无人机质心方向的夹角; d 为相对距离,表示红蓝无人机的质心距离; β 表示红蓝无人机速度矢量的夹角; Δh 表示红蓝无人机质心的高度差; Δv^2 表示红蓝无人机速度的平方差; v_r^2 表示红方无人机速度的平方; h 表示红方无人机飞行高度。

根据双方状态参数可计算出 S 中的各参数

$$\begin{cases} q_r = \arccos\{[(x_b - x_r)\cos\varphi_r\cos\rho_r + (y_b - y_r)\sin\varphi_r\cos\rho_r + (h_b - h_r)\sin\rho_r]/r\} \\ q_b = \arccos\{[(x_r - x_b)\cos\varphi_b\cos\rho_b + (y_r - y_b)\sin\varphi_b\cos\rho_b + (h_r - h_b)\sin\rho_b]/r\} \\ d = \sqrt{(x_r - x_b)^2 + (y_r - y_b)^2 + (h_r - h_b)^2} \\ \beta = \arccos(\cos\varphi_r\cos\rho_r\cos\varphi_b\cos\rho_b + \sin\rho_r\cos\rho_r\sin\rho_b\cos\rho_b + \sin\rho_r\sin\rho_b) \\ \Delta h = h_r - h_b \\ \Delta v^2 = v_r^2 - v_b^2 \end{cases} \tag{7}$$

式中: (x_r, y_r, h_r) 为红方无人机在地面坐标系下的坐标; $\varphi_r \in [-\pi, \pi]$ 为航向角,是与地面坐标系下 X 轴的夹角,左偏为正,右偏为负; ρ_r 为航迹倾斜角; 相应地,下标为 b 的参数为蓝方无人机的状态信息。

2.2　无人机机动动作集合

对于无人机的状态信息 S,本文采用地面坐标系描述无人机的坐标,表示无人机在空战过程中的路径轨迹; 用图 3 中的坐标系来描述无人机的受力情况。研究重点在于空中机动决策算法,因此忽略在动作转换过程中的力矩不平衡,以简化模型; 同时选用三自由度无人机质心运动模型(将无人机视为质点),分析其受力特征,坐标变换后可以得到简化的无人机动力学方程为

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{g}(\eta_x - \sin\rho) \\ \dfrac{\mathrm{d}\rho}{\mathrm{d}t} = \dfrac{\boldsymbol{g}}{\boldsymbol{v}}(\eta_f\cos\gamma_x - \cos\rho) \\ \dfrac{\mathrm{d}\varphi}{\mathrm{d}t} = \dfrac{\boldsymbol{g}}{\boldsymbol{v}\cos\rho}\eta_f\sin\gamma_x \end{cases} \tag{8}$$

式中：η_x 为切向过载；η_f 为法向过载；γ_x 为速度滚转角。η_f 和 γ_x 决定了航向角、航迹倾斜角的变化率，能够改变飞行方向和飞行高度。

根据简化的无人机动力学方程，可以求出速度、航迹倾斜角、航向角随时间的变化，并得出无人机的坐标为

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=v\cos\rho\cos\varphi\\ \dfrac{\mathrm{d}y}{\mathrm{d}t}=v\cos\rho\sin\varphi\\ \dfrac{\mathrm{d}z}{\mathrm{d}t}=v\sin\rho\end{cases}。\tag{9}$$

由式(9)可看出，在给定无人机初始速度 v、爬升角 ρ 和航向角 φ 时，通过积分运算可得到其他参数变化规律以及无人机在地面坐标系下的坐标。

在空中格斗过程中，无人机的基本机动动作包括定常飞行、加速、减速、左右转弯、俯冲、上升等。由于本文研究的重点是基于强化学习的无人机空战机动决策算法，所以在空战过程中不考虑两机型号、机动性能等差距，只对两机机动决策算法的优缺点进行分析研究。故本文选用基本机动动作作为可选的动作集合，并且每次机动动作执行均使用最大过载，以保证在两机机动性能相当的前提下对其两种机动决策算法的优缺点进行分析。

2.3 两机空中态势威胁

如图1所示，对于无人机空中格斗机动决策来说，环境信号反馈就是对每一次执行机动动作的“好坏”进行评价，本文将双方无人机瞬时空中态势映射为反馈信号，将无人机获得攻击占位机会作为目标状态，故此时给予一个“奖励”；反之则给予“惩罚”。将无人机攻击奖励函数定义为[15]

$$R=\begin{cases}1 & r<10\ \mathrm{km},q_r<30^\circ,q_b>30^\circ,\beta<40^\circ\\ -1 & r<10\ \mathrm{km},q_b<30^\circ\\ -1 & v_r<70\ \mathrm{m/s}\ \text{或}\ v_r>300\ \mathrm{m/s}\\ -1 & 150\ \mathrm{m}<h<15\ 000\ \mathrm{m}\\ 0 & \text{其他}\end{cases}。\tag{10}$$

当红蓝双方距离小于 10 km，红方速度矢量与红蓝质心夹角 q_r 小于 30°，蓝方速度矢量与红蓝质心夹角 q_b 大于 30°，且双方速度矢量夹角 β 小于 40°时，红方获得攻击机会，得到奖励 +1。而红方处于不利的情况有红方处于蓝方攻击位置、无人机失速或超速、飞行高度过低或过高等，其得到负回报 −1。以此作为空战结束条件。

由于双方无人机空中格斗过程中，每一时刻的空中态势的威胁程度是不同的，因此需要建立空中态势威胁评估模型。结合图3并根据相关的空中态势威胁指数[16]，建立由方向、距离、速度、相对高度组成的威胁评估函数。

1) 角度威胁函数为

$$T_a=\frac{|q_r|+|q_b|}{2\pi}。\tag{11}$$

2) 速度威胁函数为

$$T_v=\begin{cases}0.1 & v_b<0.6v_r\\ -0.5+v_b/v_r & 0.6v_r\leqslant v_b\leqslant 1.5v_r\\ 1.0 & v_b>1.5v_r\end{cases}\tag{12}$$

式中，v_r 和 v_b 分别为红方与蓝方无人机的飞行速度。

3) 距离威胁函数为

$$T_r=\begin{cases}0.5 & r\leqslant r_r,r\leqslant r_b\\ 0.5-0.2\dfrac{r-r_b}{r-r_r} & r_b<r\leqslant r_r\\ 1 & r_b>r>r_r\\ 0.8 & \max(r_r,r_b)<r<rr\end{cases}\tag{13}$$

式中：r_r 和 r_b 分别表示红方、蓝方无人机的攻击范围；rr 为红方雷达探测范围。

4) 高度威胁函数为

$$T_h=\begin{cases}1 & \Delta h<-5\ \mathrm{km}\\ 1-|\Delta h| & -5\ \mathrm{km}\leqslant\Delta h<5\ \mathrm{km}\\ 0.1 & \Delta h\geqslant 5\ \mathrm{km}\end{cases}\tag{14}$$

式中，Δh 表示红蓝两机质心高度差。

基于以上空中态势威胁指数的分析，用以上4个方面的威胁指数加权求得总的威胁指数 T，本文假设敌我双方无人机具备的空战能力相当，不考虑武器数量性能、机动性能带来的威胁，则总的威胁指数算式为

$$T=aT_a+bT_r+cT_v+dT_h\tag{15}$$

式中，$a+b+c+d=1$。

在上述内容中，式(10)是作为空战是否结束的判别条件，R 根据相应条件可分别取值 −1，0 和 1；而式(15)是作为两机空战过程中瞬时态势的反馈信号，T 值的范围是 [0,1]，所以令

$$K=R+T\tag{16}$$

故由此可知，当 $K\leqslant 0$ 或 $K\geqslant 2$ 时结束该空战；而当 $0<K<1$ 时，空战继续进行，同时输出 K 值即两机的空中态势威胁值。

3 验证及分析

本文研究的重点是基于强化学习的 UCAV 空战机动决策算法，参战的红蓝两机机动决策算法流程如图4所示。在该系统中：红方 UCAV 的机动决策算法包括强化学习单元和机动预测单元两部分，机动预测单元使用基于概率的方法来预测蓝方 UCAV 的机动动作，并向学习单元提供蓝方 UCAV 可能选择的机动动

作及对应的预测概率,强化学习单元根据结果选择相应的机动动作;随后,强化学习单元将红方 UCAV 机动动作的选择结果返回给机动预测单元,此时机动预测单元更新预测模型;而蓝方 UCAV 的机动决策算法则只有强化学习单元。

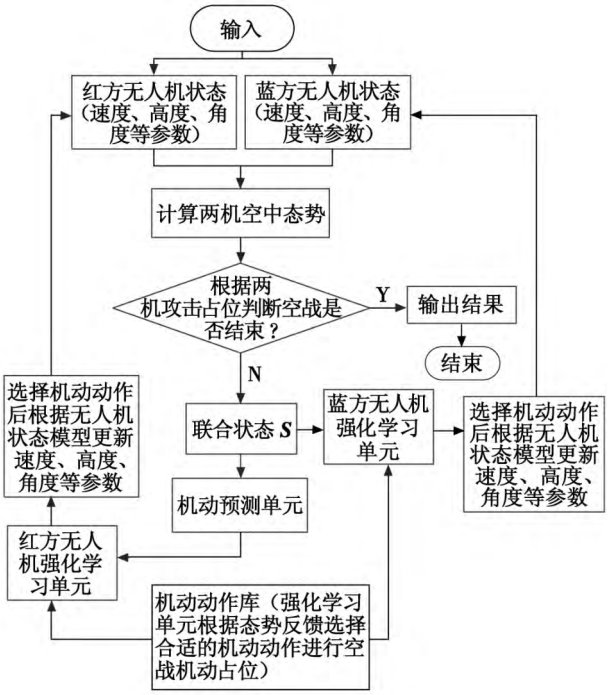


图 4　红蓝双方空战机动决策仿真流程

Fig. 4　Flow chart of maneuvering decision-making for UCAVs of two sides in air combat

在两种算法中,参数设置为:学习速度 $\alpha = 0.2$,折扣率 $\gamma = 0.9$;ε-greedy 策略中 $\varepsilon = 0.1$。假定红蓝双方 UCAV 格斗在 200 km × 250 km × 20 km 的区域内进行。两种机动决策算法分别是:红方采用的基于机动预测的强化学习机动算法(算例一)和蓝方采用的基于 ε-greedy 策略的 Q 强化学习算法(算例二)。

选择算例一并分析其结果,双方 UCAV 初始状态见表 1。

表 1　算例一中双方无人机初始状态

Table 1　Initial states of UCAVs of two sides(Example 1)

	红方	蓝方
位置/km	(60,195,15)	(35,140,2)
速度/(m·s^{-1})	100	280
ρ/(°)	27	-30
φ/(°)	-148	-57

算例一中的红蓝双方无人机空战航迹如图 5 所示。对航迹分析可得:在空战开始阶段,两机距离较近,由于红方 UCAV 高度高,所以对蓝方 UCAV 的空中威胁极大;因此在空战开始阶段蓝方 UCAV 不断远离敌方并且试图降低高度来躲避威胁;但红方 UCAV 预测准确,及时地转向蓝方 UCAV 可能出现的机动空域,此时两机距离不断减小,对蓝方 UCAV 威胁不断增大;最终红方 UCAV 通过提前转向机动,并且成功地攻击占位,从而抢得了攻击的主动权。

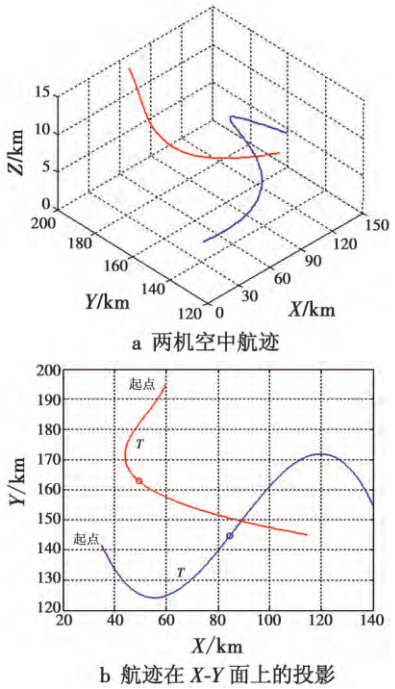


图 5　算例一中红蓝双方无人机的航迹图以及航迹在 X-Y 面上的投影

Fig. 5　Flight paths of UCAVs of two sides and their projections in the X-Y surface(Example 1)

算例一中,两机空战时间为 672 s,统计空战过程中每相邻的 5 个步长(时间总共为 5 s)对应的空中威胁值。因为威胁值数据点较多,且一定步长区间内数值的大小变化很小,所以为了图表描述方便简洁,本文选取在一定步长区间内威胁值变化较大的点来描述两机的空中威胁过程,结果如图 6 所示。

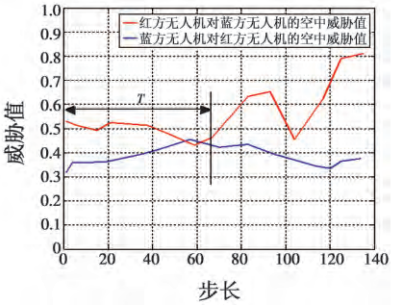


图 6　算例一中两机空战对敌威胁值

Fig. 6　The threat values of two UCAVs(Example 1)

从图 6 可看出,具有机动预测的红方 UCAV 在空战过程中具有较大优势,能够对蓝方 UCAV 进行持续有效的空中压制。

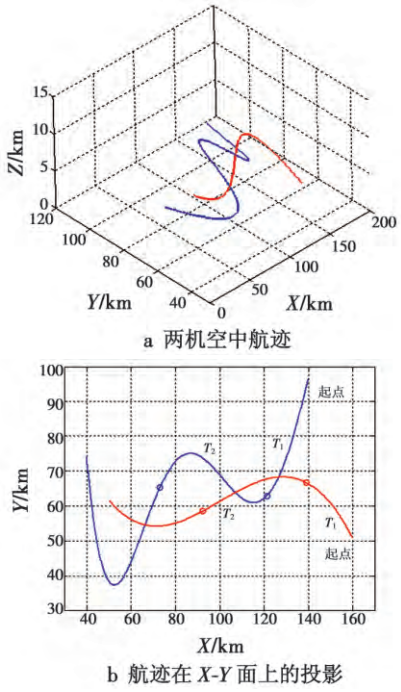
算例二中双方无人机初始状态如表 2 所示。

表 2 算例二中双方无人机初始状态

Table 2 Initial states of UCAVs of two sides(Example 2)

	红方	蓝方
位置/km	(160,50,3.3)	(140,96,6)
速度/(m·s^{-1})	120	230
ρ/(°)	24	38
φ/(°)	107	-132

算例二中的红蓝两机空战航迹如图 7 所示。



a 两机空中航迹

b 航迹在 X-Y 面上的投影

图 7 算例二中红蓝双方无人机的航迹图以及航迹在 X-Y 面上的投影

Fig. 7 Flight paths of UCAVs of two sides and their projections in the X-Y surface(Example 2)

对航迹分析可得: 两机在空战开始 T_1 阶段时都在持续地主动进攻,但由于红方 UCAV 初速较小,因此对蓝方 UCAV 威胁不大; 随后,在 T_2 阶段,蓝方 UCAV 试图掉头转向重新占据有利位置,但是经过对蓝方 UCAV 机动预测,红方 UCAV 继续飞行并及时转向,保持两机距离、高度,给予蓝方 UCAV 持续威胁,最终获得攻击机会。

算例二中空战总时长 989 s,可分为 3 个阶段。第一阶段中,蓝方 UCAV 空中占优并主动攻击,但是空中威胁较小; 随后在 T_2 阶段,红方 UCAV 通过预测结果,并针对蓝方 UCAV 飞行趋势进行机动响应,主动占据着空中优势,并最终获得攻击机会,如图 8 所示。

算例一和算例二中,红方 UCAV 对敌机动动作预测的成功率分别为 81.2% 和 84.3%。从该项统计结果可以看出,机动预测单元可以很好地预测蓝方 UCAV 的机动动作以及飞行趋势,预测成功率较高。在两个算例中,红方 UCAV 均成功预测到蓝方的飞行趋势并成功机动最终获得攻击机会。

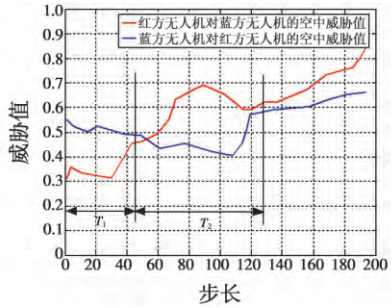


图 8 算例二中两机空战对敌威胁值

Fig. 8 The threat values of two UCAVs(Example 2)

4 总结

本文研究了一种基于机动预测的强化学习无人机空中格斗方法。采用概率神经网络(PNN)构成机动预测模块,将两机联合状态作为该模块输入量,从而预测敌机机动动作。建立无人机基本动作集合,将两机空中态势评估函数作为强化学习的信号函数。实验证明,该算法具有良好的收敛性及空战可靠性,能够使无人机通过与环境的交互学习逐渐掌握行为能力和决策能力,对无人机空中格斗的仿真研究具有一定参考意义。

参 考 文 献

[1] UALATI D U, SIMAAN M A. Effectiveness of the Nash strategies in competitive multi-team target assignment problems [J]. Transaction of Aerospace and Electronic Systems, 2007, 43(1): 126-134.

[2] LI Y, DONG Y N. Weapon-target assignment based on simulated annealing and discrete particle swarm optimization in cooperative air combat [J]. Acta Aeronautic et Astronautica Sinica, 2010, 31(3): 626-631.

[3] LIU B, QIN Z, SHAO L P. Air combat decision making for coordinated multiple target attack using collective intelligence [J]. Acta Aeronautic et Astronautica Sinica, 2010, 31(7): 1727-1739.

[4] LIU B, ZHANG X P, WANG R. Air combat decision making for coordinated multiple target attack using combinatorial auction [J]. Acta Aeronautic et Astronautica Sinica, 2010, 31(7): 1434-1444.

[5] LIU X, LIU Z, HOU W S. Improved MOPSO algorithm for multi-objective programming model of weapon-target assignment [J]. Journal of Systems Engineering and Electronics, 2013, 36(2): 326-330.

[6] 姚敏,王绪芝,赵敏. 无人机群协同作战任务分配方法研究 [J]. 电子科技大学学报,2013, 42(5): 723-727.

(下转第 22 页)

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,对于其他应用领域的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] BOUTILIER 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.