

基于SD3算法与轨迹预测的无人机空战决策研究



金文霄,周同乐,陈谋,朱豪杰

南京航空航天大学, 江苏 南京 211106

摘要:无人机自主决策是提升无人机空战自主性的关键技术。针对空战中无人机自主决策问题,本文提出了一种基于深度强化学习的无人机自主决策方法。以无人机对抗的实时态势数据作为状态输入,无人机机动指令为输出。首先,根据导弹攻击区模型建立态势奖励函数,并基于残差门控循环单元(GRU)对目标无人机的轨迹进行预测。然后,引入Softmax深层双确定性策略梯度(SD3)算法,使得无人机能够根据当前空战态势选择合适的机动动作。仿真试验结果表明,该方法能够指导无人机做出准确的机动动作,在空战中获得优势,提高了无人机在决策中的自主性。

关键词:无人机;自主机动决策;轨迹预测;深度强化学习;SD3算法

中图分类号:V279

文献标识码:A

DOI: 10.19452/j.issn1007-5453.2025.07.009

在现代战争中,战场环境越来越复杂多变,无人机需要在较短的时间内根据实时战场态势做出机动,由人工远程控制已经难以满足无人机对抗要求^[1],因此,无人机自主机动决策的研究受到各国的重视,成为无人机发展的趋势。

无人机空战自主机动决策要求无人机基于实时战场态势,利用数学优化、人工智能等技术方法生成适当的机动动作指令。常用的决策方法有专家系统法^[1]、博弈方法^[2-3]、影响图法^[4]和优化算法^[5]等。近年来,深度强化学习的发展为该问题提供了新的方向。采用深度强化学习方法训练得到的决策模型,能够在已知实时战场态势信息的情况下,瞬时给出机动方案,满足空战对抗中决策快速性的要求。欧洋等^[6]采用深度Q网络(DQN)算法训练决策模型,并采用竞争机制与双网络结构来提高算法的收敛性。然而,该模型只能输出有限个机动动作指令。丁维等^[7]以近端策略优化(PPO)算法为基础,通过引入Ornstein-Uhlenbeck随机噪声提升算法在训练过程中的探索能力。李永丰等^[8]通过深度确定性策略梯度(DDPG)算法建立决策模型,并且利用概率神经网络对目标飞机的机动行为进行预测,以提高训练效果。白双霞等^[9]采用双延迟深度确定性策略梯度(TD3)

算法,并通过仿真试验验证了该算法相比于DDPG算法具有更快的收敛速度和优化能力。

本文以无人机空战为背景,首先基于无人机运动模型和空空导弹攻击区建立无人机的空战模型。然后采用基于Softmax深层双确定性策略梯度(SD3)算法训练决策模型。为了协助我方无人机规划合适的攻击点和攻击角度,在决策模型中引入目标轨迹预测模块。最后对该模型进行建模仿真,仿真试验结果证明该模型能够完成预期作战目标。

1 问题描述

1.1 无人机对抗建模

空战对抗中,目标无人机采用指定决策方法,我方无人机决策模块根据当前两机状态和目标无人机下一个轨迹点的预测结果,采用由强化学习方法训练得到的决策模型进行机动决策。空战对抗模型如图1所示^[10]。

对抗决策首先需要收集双方无人机的实时状态,并预测目标无人机下一时刻的轨迹点,然后决策模型对实时态势与预测结果进行评估,并给出合理的机动指令。双方决策系统给出机动指令后,根据运动模型计算得到双方下一

收稿日期: 2024-09-07; 退修日期: 2025-01-07; 录用日期: 2025-03-05

基金项目: 国家自然科学基金(62203217); 江苏省基础 Research 计划自然科学基金(BK20220885); 中国高校产学研创新基金无人集群协同智能项目(2021ZYA05001)

引用格式: Jin Wenxiao, Zhou Tongle, Chen Mou, et al. UAV decision-making in air combat based on SD3 algorithm and trajectory prediction [J]. Aeronautical Science & Technology, 2025, 36(7): 67-76. 金文霄,周同乐,陈谋,等. 基于SD3算法与轨迹预测的无人机空战决策研究[J]. 航空科学技术, 2025, 36(7): 67-76.

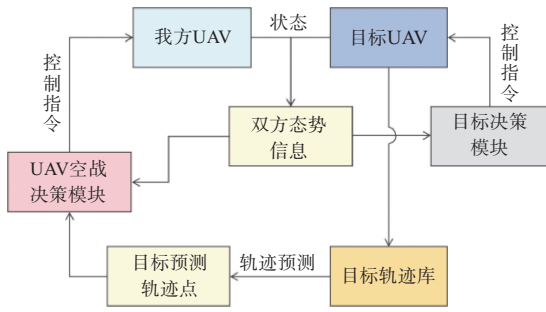


图1 空战对抗模型

Fig.1 Air combat confrontation model

个轨迹点的状态。循环执行以上步骤,直至一方无人机被击败。

1.2 无人机运动模型

为了对无人机机动过程进行建模,将无人机视为一个质点,建立无人机运动学方程和动力学方程为

$$\begin{cases} \dot{x} = v \cos \gamma \cos \psi \\ \dot{y} = v \cos \gamma \sin \psi \\ \dot{z} = v \sin \gamma \end{cases} \quad (1)$$

$$\begin{cases} \dot{v} = g(n_x - \sin \gamma) \\ \dot{\psi} = -\frac{gn_z \sin \phi}{v \cos \gamma} \\ \dot{\gamma} = g(n_z \cos \phi - \cos \gamma)/v \end{cases} \quad (2)$$

式中, x, y, z 为无人机在坐标系中的位置; v, ψ, γ 为无人机的速度、航向角、俯仰角, n_x, n_z 为切向过载和法向过载, ϕ 为绕速度矢量的滚转角; g 为重力加速度。由式(1)可得,在已知初始状态的情况下,通过控制 $[n_x, n_z, \phi]$ 可以控制无人机的机动行为。

1.3 无人机空战态势评估

无人机空战态势如图2所示^[1]。图2中, V_U, V_T 分别是我方无人机和目标无人机的速度矢量。设我机和目标无人

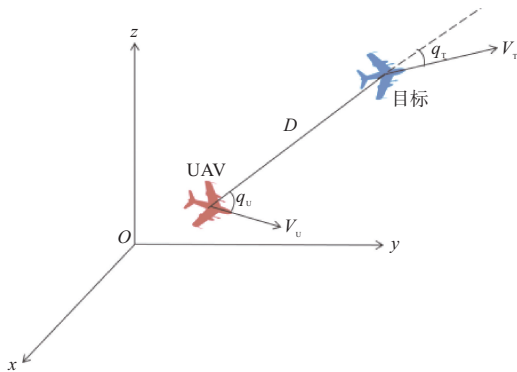


图2 无人机空战态势

Fig.2 Air combat situation

机的位置矢量分别为 $R_U=[x_U, y_U, z_U]$ 和 $R_T=[x_T, y_T, z_T]$, D 是我机和目标无人机的相对位置矢量, q_U 为 V_U 和 D 的夹角, q_T 为 V_T 和 D 的夹角。空战态势可以用 D, d, q_U, q_T 来描述,计算公式如下

$$\begin{cases} D = R_U - R_T \\ d = \|D\|_2 \\ q_U = \arccos((D \times V_U) / (\|D\|_2 \cdot \|V_U\|_2)) \\ q_T = \arccos((D \times V_T) / (\|D\|_2 \cdot \|V_T\|_2)) \end{cases} \quad (3)$$

无人机的作战能力受到空空导弹性能的约束,空空导弹的攻击区由空空导弹的最大射击距离 $D_{\max}$ 、最小射击距离 $D_{\min}$ 和最大离轴发射角 $q_{\max}$ 共同决定。同时,空空导弹需要连续锁定目标一段时间后才能发射。设锁定时间为 $t_{\min}$,目标连续处于无人机空空导弹攻击区的时间为 t_{in} 。当满足如下条件时,可以认为导弹发射成功,目标被摧毁

$$\begin{cases} D_{\min} < d < D_{\max} \\ q_U < q_{\max} \\ t_{\text{in}} > t_{\min} \end{cases} \quad (4)$$

1.4 优势函数设计

结合角度优势、距离优势和空空导弹的攻击区模型^[1]设计无人机的综合态势优势函数 R 公式如下

$$\begin{aligned} R &= \eta_A + \eta_B + \eta_C \\ \eta_A &= \begin{cases} 1 - \frac{q_U + q_T}{2\pi}, d \leq D_{\max} \\ (1 - \frac{q_U + q_T}{2\pi}) e^{-\frac{(d - D_{\max})}{D_{\max}}}, d > D_{\max} \end{cases} \\ \eta_U &= \begin{cases} 10, d < D_{\max}, q_U < q_{\max} \\ 0, \text{其他} \end{cases} \\ \eta_T &= \begin{cases} 10, d < D_{\max}, q_T < q_{\max} \\ 0, \text{其他} \end{cases} \\ \eta_B &= \eta_U - \eta_T \\ \eta_C &= \begin{cases} 0, (1 \text{ km} < h < 12 \text{ km}) \cup (d > 200 \text{ m}) \\ -10, \text{其他} \end{cases} \end{aligned} \quad (5)$$

式中, η_A 为结合角度优势与距离优势的综合优势; η_U 为目标进入我方无人机导弹攻击区的奖励; η_T 为我方无人机进入目标的导弹攻击区时的损失; η_B 为基于导弹攻击区进行胜负判定的综合优势; η_C 为高度与距离限制的惩罚函数。

2 基于残差GRU算法的目标无人机飞行轨迹预测

对目标轨迹的预测,有助于我方无人机评估执行不同机动作后的态势优势,规划下一步的机动方向与攻击位置,故在无人机决策模型中引入无人机轨迹预测模块。由

于无人机飞行轨迹是连续的,在较短时间内运动趋势不会发生突变,因此,无人机轨迹预测问题可以看作时间序列的预测问题。基于残差门控循环单元(GRU)建立目标无人机轨迹预测模型。GRU神经网络结构如图3所示^[12]。

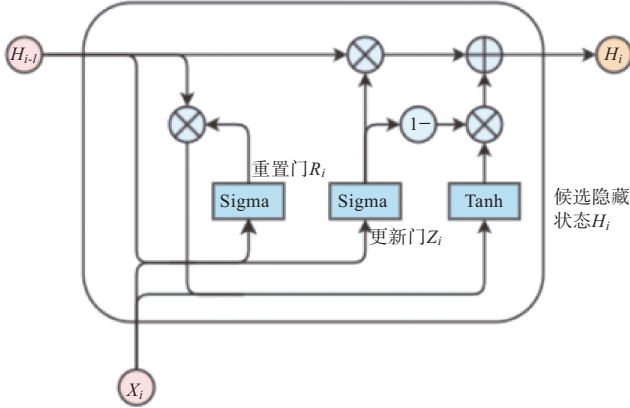


图3 门控循环单元结构

Fig.3 The structure of GRU

GRU包含重置门 R_i 和更新门 Z_i ,计算公式如下

$$R_i = \sigma(X_i W_{xr} + H_{i-1} W_{hr} + b_r) \quad (6)$$

$$Z_i = \sigma(X_i W_{xz} + H_{i-1} W_{hz} + b_z) \quad (7)$$

式中, W_{xr} 、 W_{hr} 、 W_{xz} 、 W_{hz} 为 R_i 和 Z_i 的权重系数; b_r 和 b_z 为 R_i 和 Z_i 的偏置; H_{i-1} 为 $t-1$ 时刻的状态。得到 R_i 和 Z_i 后,计算候选隐藏状态 $\tilde{H}_i$,其公式如下

$$\tilde{H}_i = \tanh(X_i W_{xh} + (R_i \odot H_{i-1}) W_{hh} + b_h) \quad (8)$$

式中, W_{xh} 、 W_{hh} 为 $\tilde{H}_i$ 的权重; b_h 为 $\tilde{H}_i$ 的偏置。 t 时刻状态 H_i 为

$$H_i = Z_i \odot H_{i-1} + (1 - Z_i) \odot \tilde{H}_i \quad (9)$$

残差结构有助于缓解神经网络在训练过程中出现的梯度消失与梯度爆炸问题^[13]。残差结构如图4所示。



图4 残差结构

Fig.4 The structure of residual

残差结构包含两种映射,即恒等映射 x_i 和残差映射 $f(x_i)$,其输出为两者之和,即

$$x_{i+1} = x_i + f(x_i) \quad (10)$$

引入残差的GRU预测模型结构如图5所示。模型包括两个残差GRU层和一个全连接层,残差GRU层将GRU网络的输出与该层输入之和作为该层的输出。

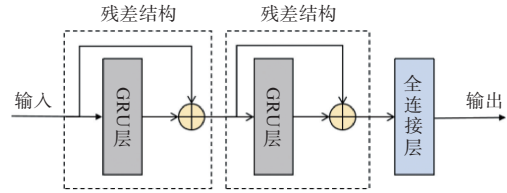


图5 无人机飞行轨迹预测模型

Fig.5 UAV trajectory prediction model

残差GRU轨迹预测模型将目标轨迹的时间序列作为模型输入。由于目标无人机的机动动作不断变化,若时间序列过长,则目标运动趋势发生变化,导致训练误差增大。若时间序列过短,则难以反映目标的运动趋势。综合以上因素,以目标 t 时刻与过去3个时刻的位置组成时间序列,作为模型输入,以目标 $t+1$ 时刻位置 $(x_{t+1}, y_{t+1}, z_{t+1})$ 作为模型输出,相邻轨迹点的时间间隔与机动决策时间间隔相同。

3 基于SD3算法的无人机空战决策

空战对抗中,我方无人机决策模块采用强化学习SD3算法。强化学习将空战决策过程看作一个马尔可夫决策过程,其要素包括状态 S 、动作 A 、策略 P 、奖励 R 和折扣因子 μ 。状态 S 为我方无人机决策模型的输入,包括当前空战态势信息和目标下一时刻的预测坐标。空战态势信息为我方与目标无人机的6个状态量 $(x, y, z, v, \psi, \gamma)$ 。目标下一时刻的预测坐标 $(x_{t+1}, y_{t+1}, z_{t+1})$ 由目标当前和历史坐标组成的时间序列,经残差GRU轨迹预测模型求得。决策模型以空战态势优势 R 作为奖励,如式(5)所示。动作 A 为我方无人机机动行为控制指令 $[n_x, n_z, \phi]$ 。

SD3算法是对DDPG算法和TD3算法的改进,其减少了DDPG算法对 Q 值的高估和TD3算法对 Q 值的低估,其中 Q 值为我机在状态 S 下执行机动动作 A 获得的总收益,即考虑折扣因子 μ 情况下的累计奖励。SD3算法包含6个神经网络,包括2个Actor网络和4个Critic网络,其结构如图6所示,其中Actor网络 A_ϕ 输出我机机动动作控制指令,Critic网络 Q_1 、 Q_2 计算不同的 Q 值,选取两个网络的最小值作为目标 Q 值^[14]。

SD3算法在无人机机动动作中引入随机噪声,加入噪声后的动作 a 为

$$a = A_\phi(s) + \tau, \tau \sim N(0, \sigma) \quad (11)$$

无人机在执行动作 a 后得到奖励 r ,如式(5)所示,并进入新状态 s' ,由此得到经验 $\langle s, a, r, s' \rangle$,并将该经验存入经验池 R 中。随后,SD3算法从经验池随机抽取经验 $\langle s, a, r, s' \rangle$

层节点数均为1000,输出为预测 Q 值。

无人机对抗模型中,空空导弹最大射击距离 $D_{\max}$ 取值3000 m,最小射击距离 $D_{\min}$ 取值0,最大离轴发射角 $q_{\max}$ 取值 $\pi/3$,最小锁定时间 $t_{\min}$ 取值2 s。无人机机动行为控制指令 n_x 取值范围为 $[-1, 1]$, n_z 取值范围为 $[0, 8]$, ϕ 取值范围为 $[-\pi/3, \pi/3]$ 。

4.2 仿真结果分析

4.2.1 无人机轨迹预测仿真试验

采用由强化学习与专家规则库决策模型空战对抗仿真中获得的无人机运动轨迹构成训练样本和测试样本分别训练GRU与残差GRU神经网络模型,训练误差和测试误差分别如图7、图8所示。

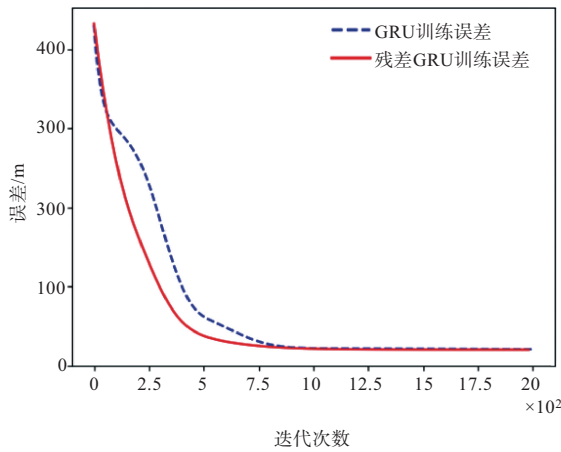


图7 GRU与残差GRU预测模型训练误差

Fig.7 Training error of GRU and residual GRU prediction model

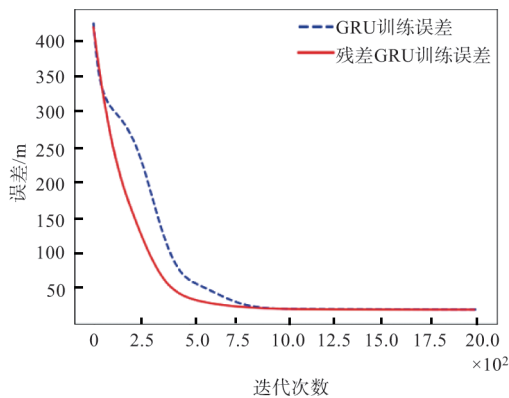


图8 GRU与残差GRU预测模型测试误差

Fig.8 Testing error of GRU and residual GRU prediction model

以上结果显示,引入残差结构的GRU模型收敛速度优于GRU模型。经过2000次的训练, x 、 y 、 z 轴平均误差小于10 m。

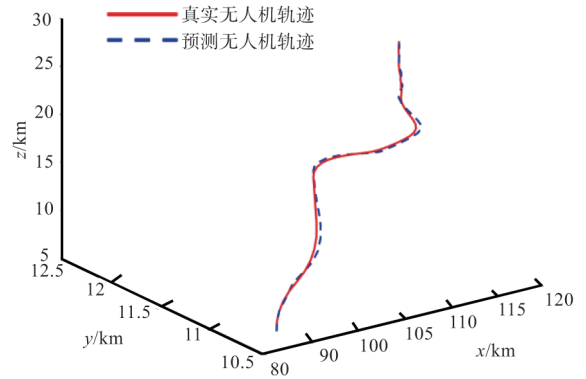


图9 真实轨迹与预测轨迹仿真对比

Fig.9 Comparison between real trajectory and predicted trajectory

采用训练得到的残差GRU预测模型对一组轨迹进行预测,真实轨迹与预测轨迹如图9所示, x 、 y 、 z 轴预测误差与总预测误差如图10所示。结果显示,采用GRU模型预测轨迹的预测误差在15 m之内。

4.2.2 敌机保持直线机动的对抗仿真试验

设目标敌机保持直线飞行,敌我双方初始状态参数设置见表2。分别采用SD3、TD3和DDPG算法训练决策模型,训练中获得的总奖励曲线如图11所示。训练过程中每一次迭代的胜负判定曲线如图12所示,其中1代表我方获得胜利,-1代表我方失利或无人机失控,0代表打平。

从图11、图12中可知,经过2000多次的训练后,我方无人机能够稳定获胜。采用SD3算法训练速度优于TD3算法与DDPG算法,且SD3算法与TD3算法在训练过程中的稳定性优于DDPG算法。无人机对抗轨迹如图13所示。

我机首先向左侧飞行,绕到目标后方,然后掉头从目标的后侧攻击目标,最终目标落入我机的导弹攻击区,战斗结束。对抗中我机与目标无人机的实时态势优势曲线如图14所示。

4.2.3 敌机保持盘旋机动的对抗仿真试验

设目标敌机保持盘旋机动,敌我双方初始状态参数设置见表3。分别采用SD3、TD3和DDPG算法训练决策模型,训练中获得的总奖励曲线如图15所示。训练过程中每一次迭代的胜负判定曲线如图16所示。

由图15、图16可知,采用SD3算法经过2500次的训练后我机能够稳定获胜,TD3算法需要训练3000次以上,而DDPG算法需要6000次以上。SD3算法训练的速度和稳定性都优于TD3算法和DDPG算法。

无人机对抗轨迹如图17所示。我机始终向目标的侧

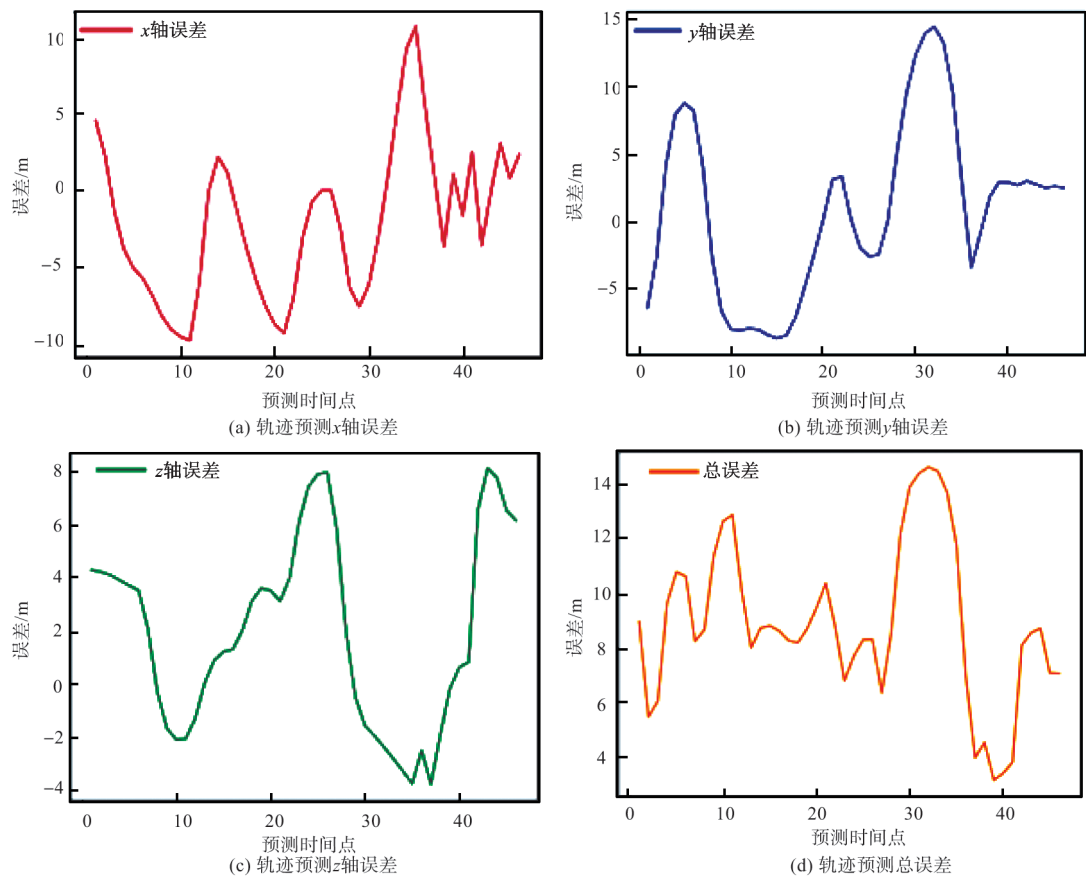


图10 轨迹预测误差

Fig.10 Error of trajectory prediction

表2 场景1初始状态

Table 2 Initial states of scenario 1

参数	我方无人机	目标无人机
x	0	40 000
y	0	40 000
z	5000	5000
v	200	200
ψ	$\pi/4$	$-3\pi/4$
γ	0	0
n_x	0	0
n_z	1	1
φ	0	0

后方飞行,期望从目标的后方向目标发起攻击。最后,我机绕到目标的后面,目标落入我方的导弹攻击区,战斗结束。对抗中我机与目标的实时态势优势如图18所示。

4.2.4 敌机采取规则库机动策略的对抗仿真试验

设目标敌机采用规则库作决策,我机采用SD3算法训练决策模型,敌我双方初始状态参数设置见表4。

分别对引入目标敌机轨迹预测和不包含预测的SD3算

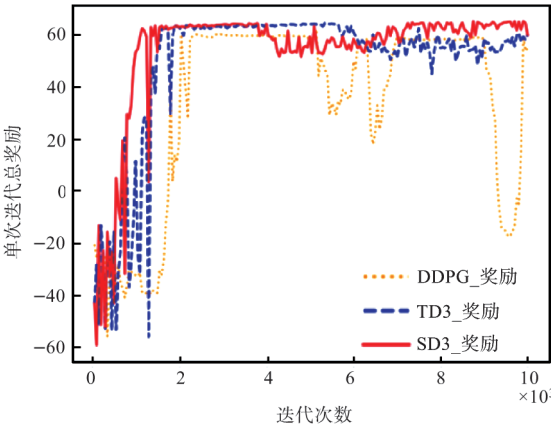


图11 场景1不同算法总奖励

Fig.11 Total rewards of algorithms in scenario 1

法进行训练,训练中获得总奖励曲线如图19所示。图19中,红色曲线表示引入轨迹预测后,每100次训练的平均总奖励,蓝色曲线表示不包含轨迹预测每100次训练的平均总奖励。训练过程中胜负统计曲线如图20所示,红色曲线代表引入轨迹预测后,每1000次训练我方获胜的次数。蓝

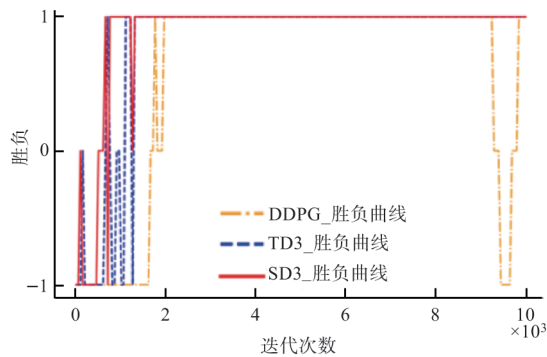


图12 场景1不同算法胜负曲线
Fig.12 Outcome curves of algorithms in scenario 1

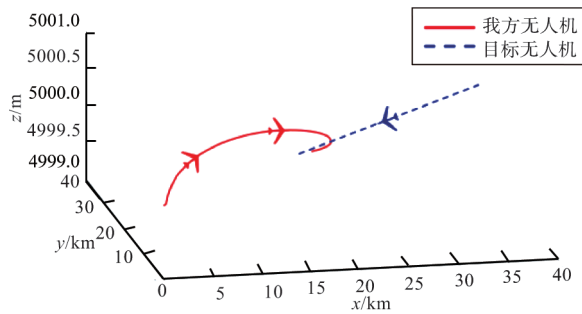


图13 场景1无人机轨迹
Fig.13 Trajectory of UAVs in scenario 1

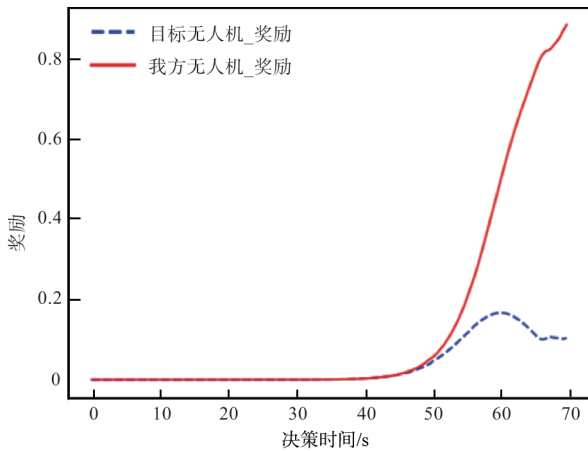


图14 场景1双方奖励曲线
Fig.14 Reward curves of both sides in scenario 1

色曲线表示不包含轨迹预测每 1000 次训练我方获胜的次数。

由图 19、图 20 可知,引入对目标下一个轨迹点的预测提升了SD3算法训练过程中获得的平均奖励,且引入轨迹预测后SD3算法在相同迭代次数情况下击败目标敌机的次数多于不引入轨迹预测的SD3算法。因此,引入轨迹预测提升了模型的训练效果和胜率。

表3 场景2初始状态
Table 3 Initial states of scenario 2

参数	我方无人机	目标无人机
x	0	20 000
y	0	20 000
z	5000	5000
v	200	200
ψ	0	0
γ	0	0
n_x	0	0
n_z	1	1.2
φ	0	$\arccos(1/1.2)$

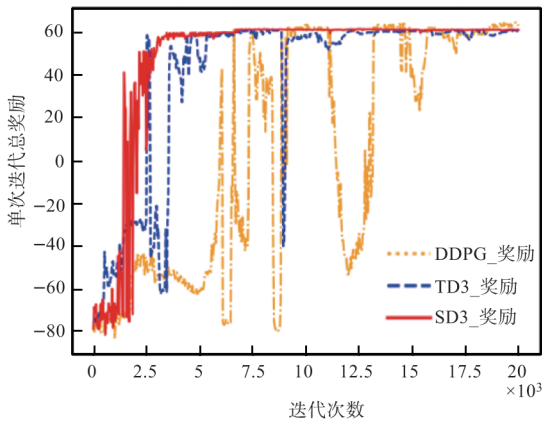


图15 场景2不同算法总奖励
Fig.15 Total rewards of algorithms in scenario 2

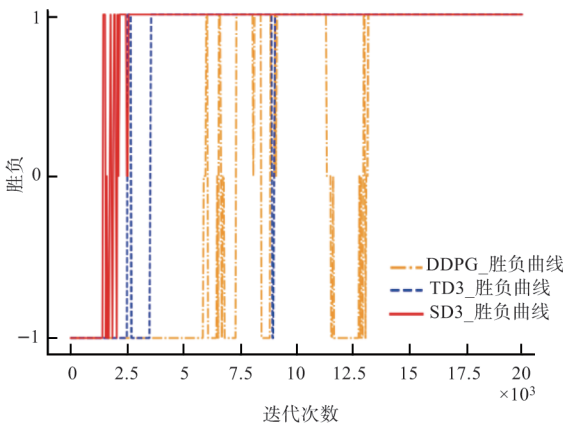


图16 场景2不同算法胜负曲线
Fig.16 Outcome curves of algorithms in scenario 2

无人机对抗轨迹如图 21 所示。我机首先向右上方移动,不断逼近目标,并跟随在目标的侧后方。最后,当我机与目标的距离足够靠近,我机从高处俯冲加速向目标追击,目标落入我方的导弹攻击区,战斗结束。对抗中我机与目标的实时态势优势曲线如图 22 所示。

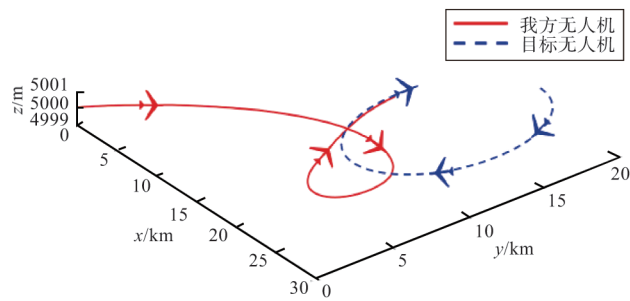


图 17 场景2无人机轨迹
Fig.17 Trajectory of UAVs in scenario 2

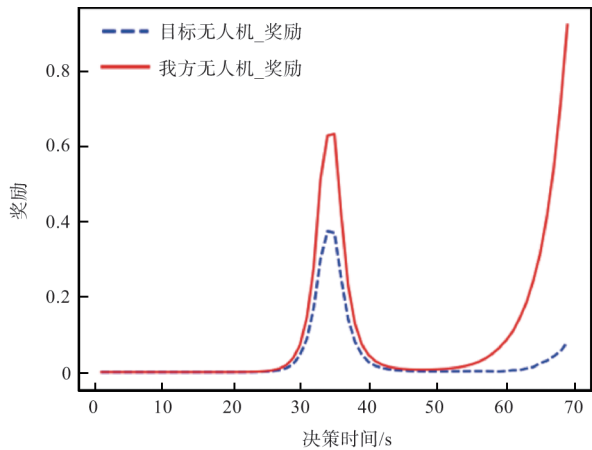


图 18 场景2双方奖励曲线
Fig.18 Reward curves of both sides in scenario 2

表4 场景3初始状态

Table 4 Initial states of scenario 3

参数	我方无人机	目标无人机
x	0	20 000~40 000
y	0	-40 000~40 000
z	5000	3000~7000
v	200	200
ψ	$(-\pi, \pi)$	$(-\pi, \pi)$
γ	0	0
n_x	0	0
n_z	1	1
φ	0	0

5 结论

本文建立了一种基于目标轨迹预测的强化学习SD3无人机机动决策模型。采用经残差结构改进的GRU算法对目标的飞行轨迹进行预测,将预测结果融入决策模型输入状态量中以辅助决策。进一步引入 Softmax 算子以减轻 TD3 算法带来的低估偏差的影响。仿真试验结果表明,SD3 算法在训练收敛的速度上要优于 TD3 算法与 DDPG 算法,并且引入目标轨迹预测提升了无人机机动决策模型在

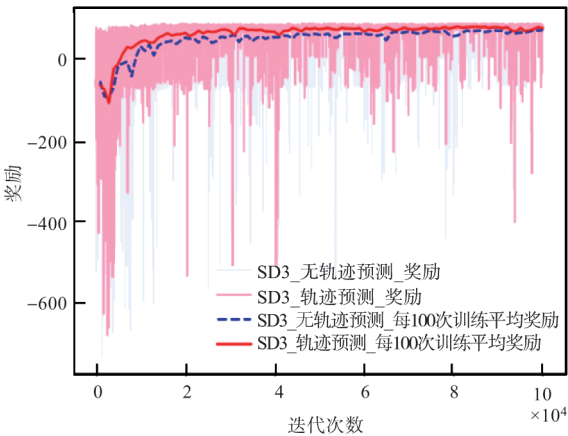


图 19 场景3不同方法总奖励
Fig.19 Total rewards of methods in scenario 3

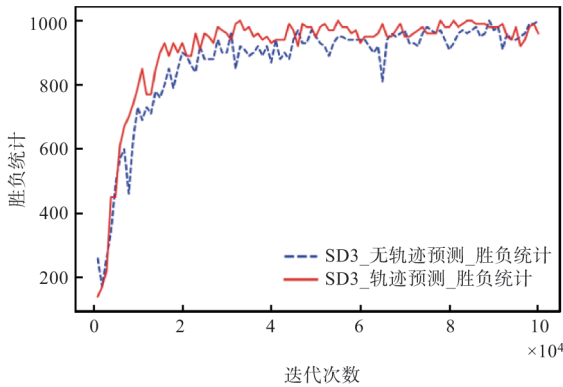


图 20 场景3胜负统计曲线
Fig.20 Outcome curves of algorithms in scenario 3

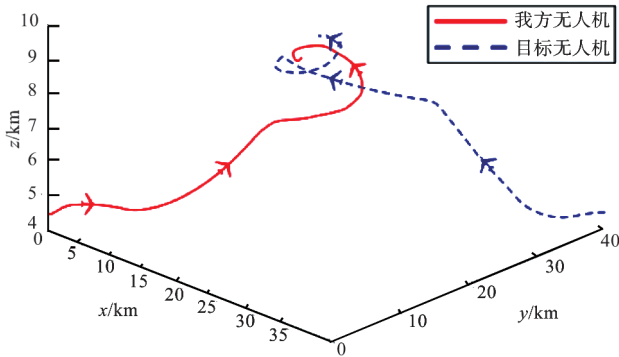


图 21 场景3无人机轨迹
Fig.21 Trajectory of UAVs in scenario 3

训练过程中获得的平均奖励与决策效果。

在未来的研究工作中,考虑将研究拓展至多机协同对抗,以顺应当前无人机空战的发展趋势。

AST

参考文献

[1] Fu Li, Xie Fuhuai, Meng Guanglei, et al. An UAV air-combat

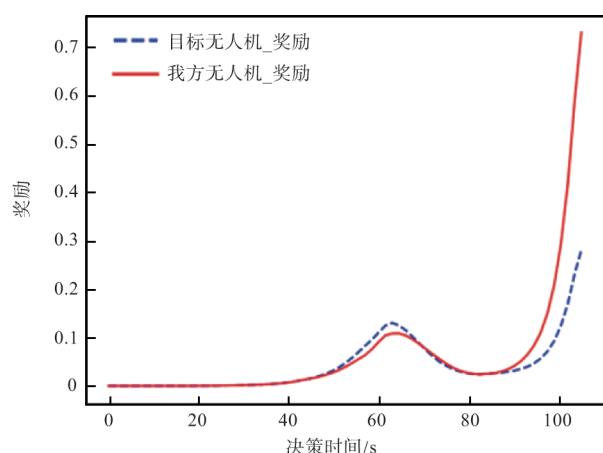


图22 场景3双方奖励曲线

Fig.22 Reward curves of both sides in scenario 3

decision expert system based on receding horizon control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(11): 1994-1999.

- [2] 顾佼佼, 赵建军, 刘卫华. 基于博弈论及Memetic算法求解的空战机动决策框架[J]. 电光与控制, 2015, 22(1): 20-23.

Gu Jiaojiao, Zhao Jianjun, Liu Weihua. Air combat maneuvering decision framework based on game theory and Memetic algorithm[J]. Electronics Optics and Control, 2015, 22(1): 20-23. (in Chinese)

- [3] 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, 2016, 17(2): 204-213.
- [4] Zhong Lin, Tong Ming'an, Zhong Wei, et al. Sequential maneuvering decisions based on multi-stage influence diagram in air combat[J]. Journal of Systems Engineering and Electronics, 2007, 18(3): 551-555.

- [5] 谢建峰, 杨启明, 戴树岭, 等. 基于强化遗传算法的无人机空战机动决策研究[J]. 西北工业大学学报, 2020, 38(6): 1330-1338.

Xie Jianfeng, Yang Qiming, Dai Shuling, et al. Air combat maneuver decision based on reinforcement genetic algorithm [J]. Journal of Northwestern Polytechnical University, 2020, 38 (6): 1330-1338. (in Chinese)

- [6] 欧洋, 徐扬, 张金鹏, 等. 无人机空战的竞争与双重深度强化学习机动对抗决策[J]. 厦门大学学报(自然科学版), 2022, 61 (6): 975-985.

Ou Yang, Xu Yang, Zhang Jinpeng, et al. UAV air combat dueling and double deep reinforcement learning maneuver adversarial decision making[J]. Journal of Xiamen University (Natural Science), 2022, 61(6): 975-985. (in Chinese)

- [7] 丁维, 王渊, 丁达理, 等. 基于LSTM-PPO算法的无人作战飞机近距空战机动决策[J]. 空军工程大学学报(自然科学版), 2022, 23(3): 19-25.

Ding Wei, Wang Yuan, Ding Dali, et al. Maneuvering decision of UCAV in close air combat based on LSTM-PPO algorithm [J]. Journal of Air Force Engineering University(Natural Science Edition), 2022, 23(3): 19-25. (in Chinese)

- [8] 李永丰, 吕永玺, 史静平, 等. 深度确定性策略梯度和预测相结合的无人机空战决策研究[J]. 西北工业大学学报, 2023, 41 (1): 56-64.

Li Yongfeng, Lyu Yongxi, Shi Jingping, et al. UAV's air combat decision-making based on deep deterministic policy gradient and prediction[J]. Journal of Northwestern Polytechnical University, 2023, 41(1): 56-64. (in Chinese)

- [9] Bai Shuangxia, Song Shaomei, Liang Shiyang, et al. UAV maneuvering decision-making algorithm based on twin delayed deep deterministic policy gradient algorithm[J]. Journal of Artificial Intelligence and Technology, 2022, 2(1): 16-22.

- [10] 李波, 白双霞, 孟波波, 等. 基于SAC算法的无人机自主空战决策算法[J]. 指挥控制与仿真, 2022, 44(5): 24-30.

Li Bo, Bai Shuangxia, Meng Bobo, et al. Autonomous air combat decision-making algorithm of UAVs based on SAC algorithm[J]. Command Control and Simulation, 2022, 44(5): 24-30. (in Chinese)

- [11] 王光辉, 徐光达, 谢宇鹏, 等. 无人作战飞机一对一超视距空战态势评估[J]. 指挥控制与仿真, 2018, 40(5): 21-25.

Wang Guanghui, Xu Guangda, Xie Yupeng, et al. Situation assessment of UCAV in One-to-one BVR air combat[J]. Command Control and Simulation, 2018, 40(5): 21-25. (in Chinese)

- [12] 胡玉可, 夏维, 胡笑旋, 等. 基于循环神经网络的船舶航迹预测[J]. 系统工程与电子技术, 2020, 42(4): 871-877.

Hu Yuke, Xia Wei, Hu Xiaoxuan, et al. Vessel trajectory prediction based on recurrent neural network[J]. Systems Engineering and Electronics, 2020, 42(4): 871-877. (in Chinese)

- [13] 郭逸婕, 张君毅, 王鹏. 基于注意力机制的TCN-BiLSTM船舶轨迹预测[J]. 计算机测量与控制, 2024, 32(1): 30-36.
Guo Yijie, Zhang Junyi, Wang Peng. Prediction for TCN-BiLSTM ship trajectory based on attention mechanism[J]. Computer Measurement and Control, 2024, 32(1): 30-36. (in Chinese)
- [14] Ling Pan, Cai Qingpeng, Huang Longbo. Softmax deep double deterministic policy gradients[J]. Advances in Neural Information Processing Systems, 2020, 33: 11767-11777.
- [15] Peng Haoran, Wang Lichun. Energy harvesting reconfigurable intelligent surface for UAV based on robust deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2023, 22(10): 6826-6838.

UAV Decision-making in Air Combat Based on SD3 Algorithm and Trajectory Prediction

Jin Wenxiao, Zhou Tongle, Chen Mou, Zhu Haojie

Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: UAV autonomous decision-making is a key technology to enhance UAV autonomy in air combat. Aiming at the problem of UAV autonomous decision-making in air combat, a UAV decision-making method based on deep reinforcement learning is developed in this paper. The model takes the real-time situation data of air combat as input and outputs UAV maneuver commands. The reward function is established based on the air combat situation and UAV air-to-air missile attack envelop model. The residual gated recurrent unit (GRU) model is introduced to predict the trajectory of the target UAV. The Softmax deep double deterministic policy gradients (SD3) algorithm is cited so that the UAV can choose the appropriate maneuver decision. The simulation experimental demonstrates that the proposed method make accurate maneuvers, gain advantages in air combat, and improve the autonomy of UAV in decision-making.

Key Words: UAV; autonomous maneuver decision; trajectory prediction; deep reinforcement learning; SD3 algorithm

Received: 2024-09-07; **Revised:** 2025-01-07; **Accepted:** 2025-03-05

Foundation item: National Natural Science Foundation of China (62203217); Jiangsu Province Basic Research Program Natural Science Foundation (BK20220885); Unmanned Swarm Cooperation Intelligence Project of China Higher Education Institution Industry-University-Research Innovation Fund (2021ZYA05001)