

doi:10.19306/j.cnki.2095-8110.2022.05.009

Q-learning 强化学习协同拦截制导律

王金强, 苏日新, 刘莉, 刘玉祥, 龙永松

(江南机电设计研究所, 贵阳 550025)

摘要:为实现多枚导弹协同拦截机动目标,提升拦截效能,提出了一种 Q-learning 强化学习协同拦截制导律。首先,基于逃逸域覆盖理论,建立了非线性多弹协同拦截模型。其次,以视线角速率为状态,依据脱靶量构造奖励函数,通过离线训练生成强化学习智能体,并结合传统比例制导控制方法,构建基于强化学习的变导引系数制导律,实时生成实现协同拦截的制导指令。最终,通过数值仿真验证了所提算法的有效性和优越性。

关键词:协同拦截;强化学习;机动目标;逃逸域;制导律

中图分类号:TJ765.3+1

文献标志码:A

文章编号:2095-8110(2022)05-0084-07

Cooperative Interception Guidance Law Based on Reinforcement Learning of Q-learning

WANG Jin-qiang, SU Ri-xin, LIU Li, LIU Yu-xiang, LONG Yong-song

(Jiangnan Institute of Mechanical and Electrical Design, Guiyang 550025, China)

Abstract: To achieve the cooperative interception of multiple missiles against a maneuvering target and improve the interception effectiveness, a cooperative interception guidance law is proposed through Q-learning technology. Firstly, based on escape domain covering theory, a nonlinear cooperative interception model is established. Then, a reward function is constructed by using miss distance with taking line-of-sight rate as the state, and a reinforcement learning agent is generated by offline training. At the same time, a variable coefficient guidance law based on reinforcement learning algorithm is designed by combining proportional navigation guidance law to generate guidance commands in real time. Finally, the effectiveness and superiority of the proposed algorithm are verified based on numerical simulation.

Key words: Cooperative interception; Reinforcement learning; Maneuvering target; Escape domain; Guidance law

0 引言

随着飞行器技术的不断发展,现代战争呈现出智能化、信息化、多样化的特点,体系与体系的对抗将贯穿战争始终,尤其是以精确制导武器为主的攻击体系和以地空舰导弹为主的防御体系之间的对

抗,在上述情况下,传统的单导弹作战模式已难以满足实际作战需求^[1]。弹群协同作战是将所有参战导弹组成一个作战网络,在指挥中心的调控下,实现弹间信息通信和共享,具有更高的作战效能,是未来智能导弹的重点发展方向^[2]。

I. Jeon 等^[3]在制导律设计中引入时间约束,提

收稿日期:2022-04-10;修订日期:2022-05-16

基金项目:国防科工委重点基础研究项目(2019-JCJQ-ZD-049)

作者简介:王金强(1995-),男,博士,工程师,主要研究方向为导弹智能控制。

出了一种可变攻击时间的协同制导律,并通过数值仿真验证了算法的有效性。Chen Y. 等^[4]在时间控制的基础上,进一步考虑了存在末端攻击角度约束的情况。李强^[5]针对协同制导问题,分别在视线方向和视线法向设计了有限时间收敛滑模制导律。H. B. Oza 等^[6]为提高运算效率,设计了考虑末端多约束的模型预测静态规划制导律。Liu X. 等^[7]为处理制导过程中存在的不确定性,基于李雅普诺夫稳定性理论,提出了一种自适应滑模协同制导律,但该方法存在系统抖振的问题。宋俊红等^[8]基于超螺旋滑模控制算法,设计了一种双层协同制导律,有效改善了制导控制系统的暂态特性。肖惟等^[9]研究了多枚过载受限的弱机动导弹拦截强机动目标的协同拦截问题,提出了基于标准弹道的分布式协同拦截策略设计方法。Zhai C. 等^[10]为提高协同拦截的成功率,设计了一种基于覆盖的拦截算法。虽然上述算法具有良好的控制效果,但在设计过程中均需预先指定期望攻击时间,各枚导弹间没有信息交互,并没有实现真正意义的智能协同作战^[11]。

随着人工智能领域的迅猛发展,强化学习算法作为一种智能决策算法,在导弹制导控制、智能任务规划和故障诊断等方面取得了显著成果。B. Gaudet 等^[12]为提高制导律鲁棒性,基于神经网络设计了一种强化元学习制导律。张秦浩等^[13]基于 Q-learning 强化学习算法设计了最优拦截制导律。南英等^[14]则对传统 Q 网络进行改进,提出了一种基于 Markov 决策过程的制导律,且不需要训练样本,可自主搜索奖励值最大的动作并完成训练。陈中原等^[15]提出了一种基于深度确定性策略梯度的强化学习协同制导律,引入 Actor 和 Critic 网络选取动作和奖励值的逼近。上述算法虽然使导弹具有自主决策能力,但运算量大,现有的弹载计算机难以满足要求。

为解决上述问题,本文以传统比例制导律为基础,引入智能决策,提出了一种 Q-learning 强化学习协同拦截制导律,并通过数值仿真验证了算法的有效性和优越性。

1 协同拦截模型

图 1 给出了导弹平面拦截几何,其中 OXY 为惯性系, M 和 T 分别代表导弹和目标, V 表示速度, θ 表示弹道倾角, ϵ 表示视线角, η 表示前置角, a 表示法向角速度, R 表示弹目相对距离。

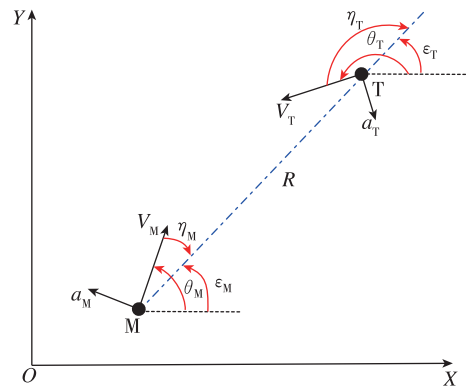


图 1 导弹拦截平面几何

Fig. 1 Planar interception geometry of missile

基于坐标转换得到导弹与目标的非线性相对运动方程为^[9]

$$\begin{cases} \dot{R} = V_T \cos(\eta_T) - V_M \cos(\eta_M) \\ R\dot{\epsilon}_M = V_M \sin(\eta_M) - V_T \sin(\eta_T) \\ \epsilon_M = \theta_M + \eta_M \\ \epsilon_T = \theta_T + \eta_T \\ \dot{\theta}_M = a_M / V_M \\ \dot{\theta}_T = a_T / V_T \end{cases} \quad (1)$$

随后,建立导弹非线性协同拦截模型,其示意图如图 2 所示,其中 $S_{M,i}$ 表示第 i 枚导弹的最大机动区域, S_T 为目标的最大机动逃逸区域,记为逃逸域, $a_{M,max}$ 为导弹最大机动过载, a_T 表示目标的逃逸加速度,定义为 $a_T = a_{T,s} + \delta$,其中 $a_{T,s}$ 为标准拦截弹道下目标期望逃逸加速度, δ 为小量,且 $|\delta|$ 越大,拦截弹道越弯曲。为简化非线性模型,便于数学处理,此处假设 a_T 为常值。因此,基于文献^[16]中标准弹道的思想和逃逸域理论,弹群协同拦截模型的构建过程如下:

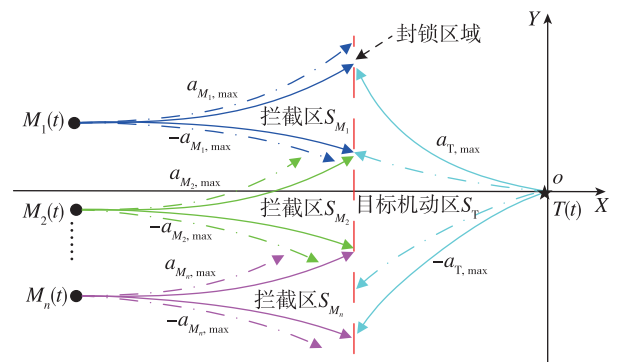


图 2 弹群协同拦截策略

Fig. 2 Cooperative interception strategy of multiple missiles

1) $\forall a_{T,s} \in S_T = [-a_{T,\text{Max}}, a_{T,\text{Max}}]$, 结合式(1)确定初始位置 $\Omega_{i,s} = [\eta_{M,i,0,s}, \eta_{T,i,0,s}, R_{i,0,s}]^T$ 和制导律 $a_{M,i}$, 以满足拦截过程的评价函数 $J = a_{M,i,0}^2 + a_{M,i,f}^2$ 最小, 其中下标 0 和 f 分别表示拦截初始时刻和终止时刻对应的值, 下标 s 表示目标以 $a_{T,s}$ 逃逸所对应的值。

2) 确定导弹覆盖区域 $S_{M,i}$, 其中 $S_{M,i}$ 是 $\hat{S}_{M,i}$ 的闭包, 且 $\hat{S}_{M,i}$ 满足以下 2 个条件: ① $\hat{S}_{M,i}$ 是 a_T 的邻域; ② $\forall a_T \in \hat{S}_{M,i}$ 机动的目标, 导弹的拦截过载不饱和, 即满足 $|a_{M,i}| < a_{M,\text{Max}} (\forall 0 \leq t < t_f)$ 。

3) 分配弹群拦截区域 $S_{M,i}$, 以均匀覆盖目标的逃逸区域 S_T , 即满足 $S_T \subseteq \sum_{i=1}^n S_{M,i}$, 其中 n 表示导弹数量。

2 协同拦截制导律

本章将结合 Q-learning 强化学习算法进行协同制导律设计。首先, 基于标准弹道的思想, 以导弹的最大机动区域 $S_{M,i}$ 中的标准弹道拦截 $a_{T,s}$ 机动的目标, 以非标准弹道拦截 $a_{T,s} + \delta$ 机动的目标, 则协同制导律 $a_{M,i}$ 可设计为

$$a_{M,i} = N(|\dot{R}_i| |\dot{\epsilon}_i - \Phi_i) \quad (2)$$

式中, N 为有效导航比; Φ_i 为偏置项。当目标以 $a_{T,s}$ 机动时, 偏置项 Φ_i 与比例项 $|\dot{R}_i| |\dot{\epsilon}_i$ 相互抵消; 当目标以 $a_{T,s} + \delta$ 机动时, 导弹在比例制导律作用下对其进行拦截。偏置项 Φ_i 可通过弹间网络进行信息交互, 以实现导弹 M_i 的制导参数趋近弹群中需用过载最小的导弹。

随后, 定义导弹与目标飞行过程中的零控脱靶量 Π_i 为

$$\Pi_i = R_i \sqrt{\frac{(R_i \dot{\epsilon}_i)^2}{\dot{R}_i^2 + (R_i \dot{\epsilon}_i)^2}} \quad (3)$$

同理, 导弹 M_i 以标准弹道拦截机动目标过程中的零控脱靶量 $\Pi_{i,s}$ 定义为

$$\Pi_{i,s} = \begin{cases} R_{i,s} \sqrt{\frac{(R_{i,s} \dot{\epsilon}_{i,s})^2}{\dot{R}_{i,s}^2 + (R_{i,s} \dot{\epsilon}_{i,s})^2}}, & R_{i,s} \geq \mu_R \\ 0, & R_{i,s} < \mu_R \end{cases} \quad (4)$$

式中, μ_R 为一个小量, 且 $\mu_R > 0$, 用于避免求解式(5)中 $R_{i,s}, \epsilon_{i,s}$ 和 $\theta_{M,i,s}$ 时发生奇异。

$$\begin{cases} \dot{R}_{i,s} = V_T \cos((\epsilon_{i,s} - \epsilon_{i,s,0}) - (t_{T,i,s}/V_T - \eta_{T,i,0,s})) - V_M \cos(\epsilon_{i,s} - \theta_{M,i,s}) \\ R_{i,s} \dot{\epsilon}_{i,s} = V_M \sin(\epsilon_{i,s} - \theta_{M,i,s}) - V_T \sin(\epsilon_{i,s} - \epsilon_{i,s,0}) - (t_{T,i,s}/V_T - \eta_{T,i,0,s}) \\ \dot{\theta}_{M,i,s} = N(|\dot{R}_{i,s}| |\dot{\epsilon}_{i,s} - \Phi_i) / V_M \end{cases} \quad (5)$$

则偏置项 Φ_i 自适应调节律可设为

$$\Phi_i(t_{k+1}) = \begin{cases} k_i \Phi_{i,0} + (1 - k_i) \Phi_i(t_k), & \Delta \geq \mu_\Delta \\ \Phi_i(t_k), & \Delta < \mu_\Delta \end{cases} \quad (6)$$

式中, $\Phi_{i,0}$ 为预先设定偏置项; k_i 为偏置系数, 定义为 $k_i = \exp(\Pi_i - \Pi_{i,s})$; μ_Δ 为一个小量, 且 $\mu_\Delta > 0, \Delta = \exp(\Pi_i - \Pi_{i,s})$ 。

进一步, 采用 Q-learning 强化学习算法对协同制导律(2)中的有效导航比 N 进行智能在线调节。协同拦截问题状态空间设为 $S_t = \dot{\epsilon}_i$, 考虑过载对弹体本身特性的约束, $\dot{\epsilon}_i$ 取值范围为 $[-0.5, 0.5]$ rad/s, 则离散化后的状态空间取为 $S_t = [-0.5, -0.3, -0.2, -0.15, -0.1, -0.05, -0.04, -0.03, -0.02, -0.01, 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2, 0.3, 0.5]$ 。动作空间设为 $A_t = N$, 由经验可知, 比例制导律的有效导航比通常取 2~4 之间, 即离散化的动作空间可选为 $A_t = [2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4]$ 。奖励函数 L_t 设为

$$L_t = \begin{cases} 1000, & R_t \leq R_h \\ 1/R_t, & R_t > R_h \end{cases} \quad (7)$$

式中, R_t 为导弹终止时刻弹目相对距离; R_h 为导弹命中目标所需最小弹目距离, 常取 $R_h = 1$, 即表示在拦截过程中, 奖励值随弹目距离的减小而增大, 若最终命中目标, 则得到一个更大的奖励, 若没有命中目标则奖励值为 0。

综上, 基于 Q-learning 强化学习算法的目标策略设为

$$Q(S_{t+1}, A_t) \leftarrow \lambda [L_t + \gamma \max_A Q(S_{t+1}, A) - Q(S_t, A_t)] + Q(S_t, A_t) \quad (8)$$

行为策略为 ϵ -greedy 策略, 即

$$A_t = \begin{cases} A \text{ 中的随机行为}, & P(\bar{\omega}) \\ \arg \max(Q(S_t, A_t)), & P(1 - \bar{\omega}) \end{cases} \quad (9)$$

式中, λ 为学习效率参数; γ 为折扣率参数; $\bar{\omega}$ 为策略参数, 即导弹以 $\bar{\omega}$ 的概率在动作空间中进行随机选择, 则以 $1 - \bar{\omega}$ 的概率会选择得到最大 Q 值的动作^[13]。Q-learning 强化学习算法流程如图 3 所示。

3 拦截区域分配

本章基于逃逸域覆盖理论进行多弹拦截区域分配算法设计^[16]。为方便推导, 定义归一化的目标加速度为 $\alpha = a_T/a_{T,\text{Max}}$, 导弹覆盖区域 $S_{M,i}$ 和目标逃逸域 S_T 可分别归一化为 $[\alpha_{i,\text{low}}, \alpha_{i,\text{up}}]$ 和 $[-1, 1]$ 。

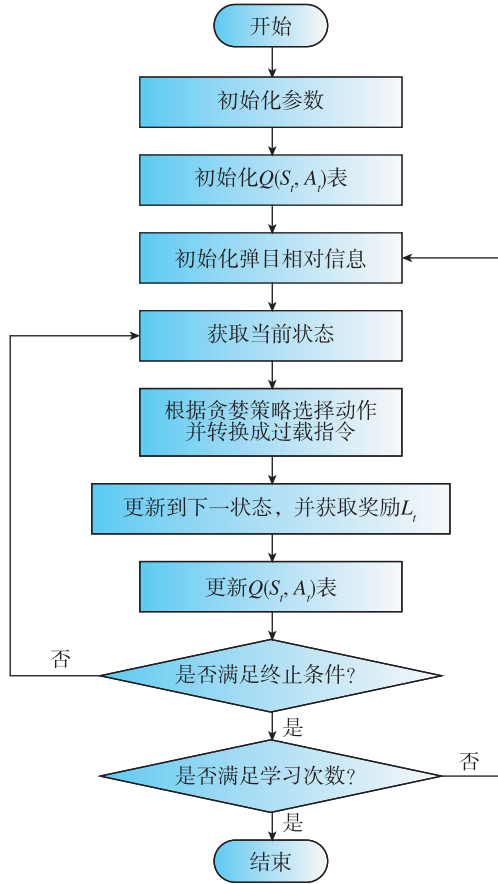


图 3 Q-learning 强化学习算法流程

Fig. 3 Flow chart of Q-learning algorithm

所设计拦截区域分配策略如图 4 所示, 导弹 1 的拦截区域 $S_{M,1}$ 左边界与目标逃逸域左边界对齐, 导弹 n 拦截区域 $S_{M,n}$ 右边界与目标逃逸域右边界对齐, 每枚导弹覆盖范围相同。

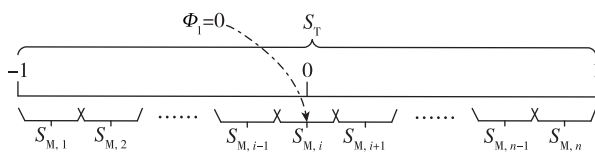


图 4 拦截区域分配模式

Fig. 4 Allocation modes of intercept area

为实现上述分配策略, 首先求解 n 个导弹的拦截区域 $\{S_{M,i} | i = 1, 2, 3, \dots, n\}$, 其中 $S_{M,1}$ 的左边界与 -1 对齐, $S_{M,n}$ 的右边界与 1 对齐, $S_{M,i-1}$ 和 $S_{M,i}$ 不重叠相接, $S_{M,n-1}$ 与 $S_{M,n}$ 可重叠相交, 即 $\alpha_{1,low} = -1, \alpha_{n,up} = 1, \alpha_{n-1,up} = \alpha_{n,low}, \alpha_{n-1,up} - \alpha_{n,low} \geq 0$ 。同时, 为使每枚导弹拦截覆盖区域均匀分布, 此处将重叠区域 $[\alpha_{n,low}, \alpha_{n-1,up}]$ 均匀分配到其余子区域上, 即将覆盖区域 $S_{M,2} \sim S_{M,n-1}$ 分别向左移动 $(i -$

$1)/(n - 1) \times (\alpha_{n-1,up} - \alpha_{n,low})$ 长度, 算法具体伪代码如表 1 所示。

表 1 拦截区域分配的实现算法

Tab. 1 Algorithm for intercept area allocation

1	输入: 拦截弹、目标参数、初始阵位约束
2	求解 $[\Phi_1, \Omega_{1,s}, \alpha_{T,1}, S_{M,1}]^T$, 并令 $\alpha_{1,low} = -1$
3	令 $i = 2$
4	WHILE $\alpha_{i-1,up} < 1$
5	求解 $[\Phi_i, \Omega_{i,s}, \alpha_{T,i}, S_{M,i}]^T$ 并令 $\alpha_{i,low} = \alpha_{i-1,up}$
6	IF $\alpha_{i,up} > 1$
7	求解 $[\Phi_i, \Omega_{i,s}, \alpha_{T,i}, S_{M,i}]^T$ 并令 $\alpha_{i,up} = 1$
8	END
9	$i = i + 1$
10	END
11	令 $i = 2, t = (\alpha_{n-1,up} - \alpha_{n,low}) / (n - 1)$
12	WHILE $i < n - 1$
13	$\alpha_{T,s} = \alpha_{T,s} - (i - 1)t$
14	以更新的 $\alpha_{T,s}$ 求解 $[\Phi_i, \Omega_{i,s}, S_{M,i}]^T$
15	$i = i + 1$
16	END
17	输出: $\{\Phi_i, \Omega_{i,s}, \alpha_{T,s}, S_{M,i} i = 1, 2, 3, \dots, n\}$

4 数值仿真分析

本章分别针对多弹齐射(模式 1)和子母弹分离发射(模式 2)两种作战模式, 对上述协同制导律的有效性进行数值仿真验证。在多弹齐射作战模式下, 假设导弹速度方向与轴线重合, 即攻角、侧滑角和前置角均为 0, 因此其初始阵位约束为

$$\begin{cases} \eta_{M,i,0,s} = 0 \\ R_{i,0,s} = R_0 \end{cases} \quad (10)$$

子母弹分离作战模式下, 忽略子弹和母弹间的动态过程, 并假设初始时刻目标前置角和弹目的距离相同, 因此其初始阵位约束为

$$\begin{cases} \eta_{T,i,0,s} = -\pi \\ R_{i,0,s} = R_0 \end{cases} \quad (11)$$

仿真环境下假设导弹数目为 3, 分别记为 M_1, M_2 和 M_3 , 弹目初始相对距离设为 60km, 导弹速度为 $7Ma$, 最大机动过载 $3g$, 目标速度为 $6Ma$, 最大机动过载 $5g$, 有效导航比 N 由 Q-learning 算法在线计算, 学习率参数 λ 设为 0.01, 折扣率参数 γ 则设为 0.99。

针对目标最大正机动 ($\alpha = 1$)、不机动 ($\alpha = 0$) 和

最大负机动($\alpha = -1$)的协同拦截仿真结果如图5~图10所示。图5和图7所示分别为两种作战模式下的拦截轨迹,从中可知,针对上述三种目标机动形式,本文所提协同制导律可确保至少有一枚导弹成功命中目标,证明了算法的有效性。图6和图8所示分别为两种作战模式下的导弹过载曲线。图9和图10所示分别为导弹1最大负机动($\alpha = -1$)条件下的有效导航比曲线和 Q 均值曲线,从中可知,在制导过程中有效导航比可进行自适应调节,且随着训练的进行, Q 均值逐渐收敛。

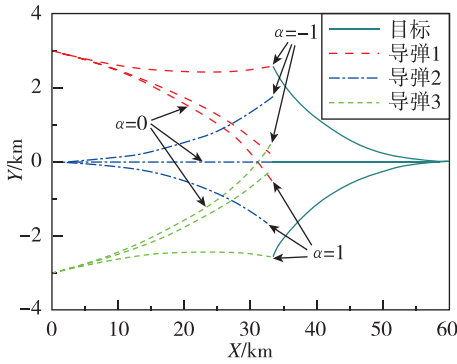


图5 作战模式1的拦截弹道

Fig. 5 Interception trajectory under mode 1

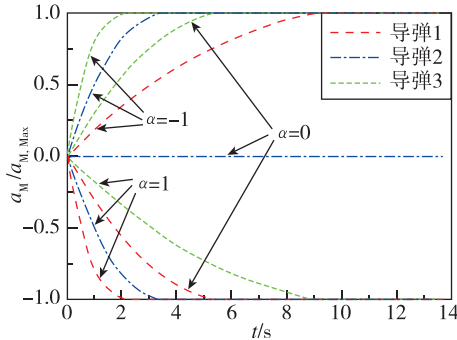


图6 作战模式1的导弹过载

Fig. 6 Acceleration of missile under mode 1

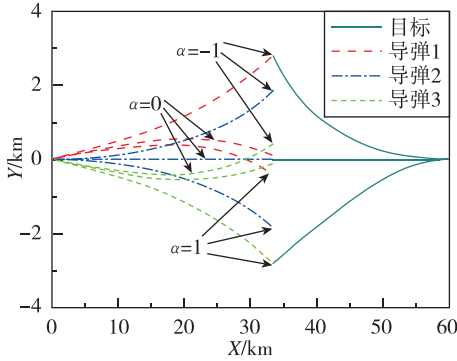


图7 作战模式2的拦截弹道

Fig. 7 Interception trajectory under mode 2

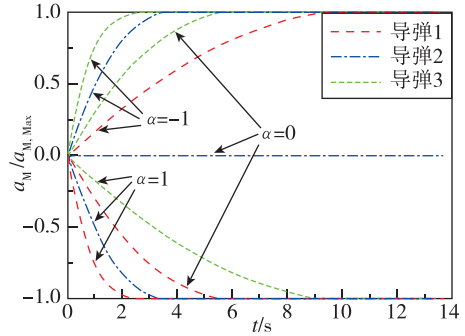


图8 作战模式2的导弹过载

Fig. 8 Acceleration of missile under mode 2

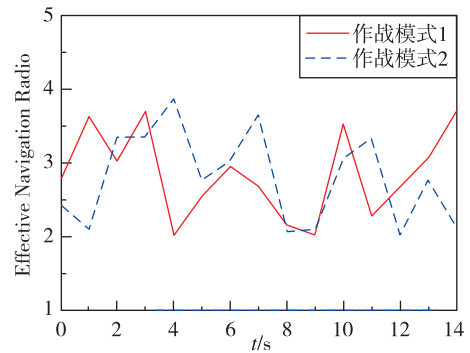


图9 有效导航比N曲线

Fig. 9 Curve of effective navigation ratio N

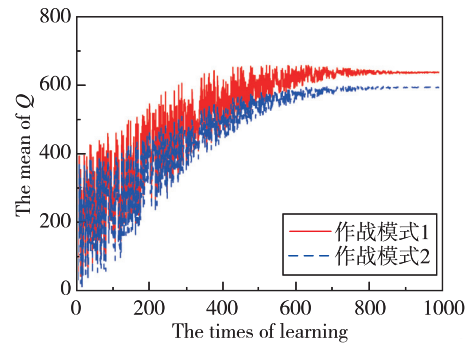


图10 Q均值收敛曲线

Fig. 10 Convergence curve of the mean of Q

针对目标做 $a_T = -a_{T,Max} \text{sign}(\sin(\pi t/2))$ 蛇形机动的仿真结果如图11~图14所示。图11和图12所示分别为两种作战模式下的拦截轨迹,从中可知,引入偏置项 Φ_i 可使导弹在拦截过程中更加接近目标,有效提升了拦截效果。图13和图14所示为零控脱靶量曲线,可以看出,与传统比例制导律相比,本文设计的协同制导律零控脱靶量更低,具有更强的工程实用价值。

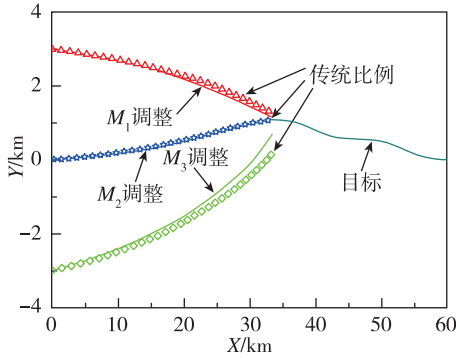


图 11 作战模式 1 的拦截弹道

Fig. 11 Interception trajectory under mode 1

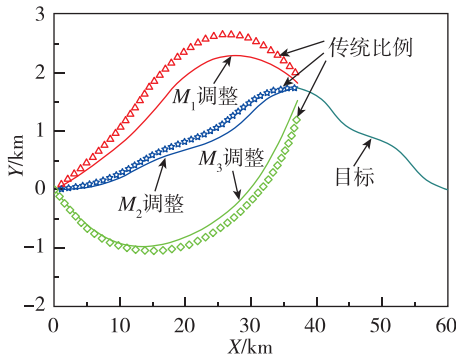


图 12 作战模式 2 的拦截弹道

Fig. 12 Interception trajectory under mode 2

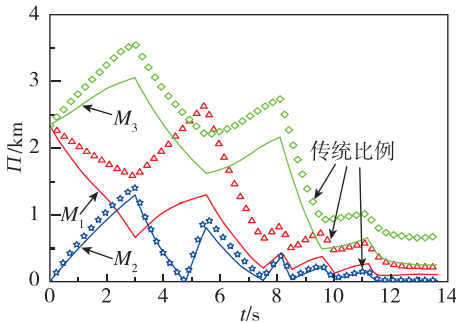


图 13 作战模式 1 的零控脱靶量

Fig. 13 Zero effort miss under mode 1

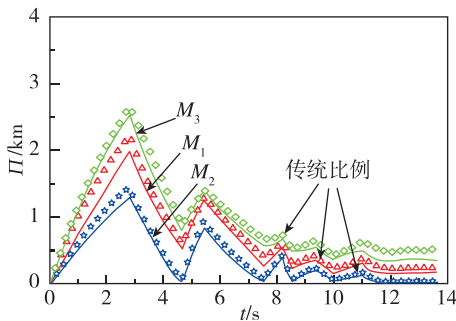


图 14 作战模式 2 的零控脱靶量

Fig. 14 Zero effort miss under mode 2

5 结论

本文研究了多弹协同拦截机动目标问题, 具体结论如下:

1) 基于逃逸域覆盖理论、比例制导律和 Q-learning 算法提出了一种强化学习协同制导律。

2) 针对多弹齐射和子母弹分离发射两种作战模式进行了数值仿真, 结果验证了所提算法的有效性和优越性。

3) 后续工作可进一步研究三维空间中存在攻击角约束、能量约束、避障和避撞等因素的协同拦截问题, 为实现多约束条件下的智能协同制导奠定基础。

参考文献

- [1] 槐泽鹏, 梁雪超, 王洪波, 等. 多弹协同及其智能化发展研究[J]. 战术导弹技术, 2019, 40(5): 77-85.
Huai Zepeng, Liang Xuechao, Wang Hongbo, et al. Research on multi-missile collaborative and its intelligence development[J]. Tactical Missile Technology, 2019, 40(5): 77-85(in Chinese).
- [2] 徐胜利, 陈意芬, 范晋祥, 等. 多弹协同技术在防空导弹发展中的应用探讨[J]. 电光与控制, 2017, 24(2): 55-59.
Xu Shengli, Chen Yifen, Fan Jinxiang, et al. Application of multi-missile cooperative technology in development of air-defense missiles[J]. Electronics Optics and Control, 2017, 24(2): 55-59(in Chinese).
- [3] Jeon I, Lee J, Tahk M. Homing guidance law for cooperative attack of multiple missiles[J]. Journal of Guidance, Control and Dynamics, 2010, 33(1): 275-280.
- [4] Chen Y D, Wang J N, Wang C Y, et al. Impact time and angle control based on constrained optimal solutions[J]. Journal of Guidance, Control, and Dynamics, 2016, 39(10): 2445-2451.
- [5] 李强. 多导弹协同作战目标分配和制导律研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
Li Qiang. Research on target assignment and guidance law for cooperative engagement of multi-missile[D]. Harbin: Harbin Institute of Technology, 2017(in Chinese).
- [6] Oza H B, Padhi R. Impact-angle-constrained suboptimal model predictive static programming guidance of air-to-ground missiles[J]. Journal of Guidance, Control, and Dynamics, 2012, 35(1): 153-164.
- [7] Liu X D, Li G F. Adaptive sliding mode guidance

- with impact time and angle constraints[J]. *IEEE Access*, 2020, 8: 26926-26932.
- [8] 宋俊红, 宋申民, 徐胜利. 一种拦截机动目标的多导弹协同制导律[J]. *宇航学报*, 2016, 37(12): 1306-1314.
Song Junhong, Song Shenmin, Xu Shengli. A cooperative guidance law for multiple missiles to intercept maneuvering target [J]. *Journal of Astronautics*, 2016, 37(12): 1306-1314(in Chinese).
- [9] 肖惟, 于江龙, 董希旺, 等. 过载约束下的大机动目标协同拦截[J]. *航空学报*, 2020, 41(S1): 184-194.
Xiao Wei, Yu Jianglong, Dong Xiwang, et al. Cooperative interception against highly maneuvering target with acceleration constraints[J]. *Acta Aeronautica et Astronautica Sinica*, 2020, 41(S1): 184-194(in Chinese).
- [10] Zhai C, He F H, Hong Y G, et al. Coverage-based interception algorithm of multiple interceptors against the target involving decoys[J]. *Journal of Guidance, Control, and Dynamics*, 2016, 39(7): 1-7.
- [11] 张春妍, 宋建梅, 侯博, 等. 带落角和时间约束的网络化导弹协同制导律[J]. *兵工学报*, 2016, 37(3): 431-438.
Zhang Chunyan, Song Jianmei, Hou Bo, et al. Cooperative guidance law with impact angle and impact time constraints for networked missiles[J]. *Acta Armamentarii*, 2016, 37(3): 431-438(in Chinese).
- [12] Gaudet B, Furfaro R, Linares R. Reinforcement learning for angle-only intercept guidance of maneuvering targets[J]. *Aerospace Science and Technology*, 2020, 99: 105746.
- [13] 张秦浩, 敖百强, 张秦雪. Q-learning 强化学习制导律[J]. *系统工程与电子技术*, 2020, 42(2): 414-419.
Zhang Qin hao, Ao Baiqiang, Zhang Qin xue. Reinforcement learning guidance law of Q-learning[J]. *Systems Engineering and Electronics*, 2020, 42(2): 414-419(in Chinese).
- [14] 南英, 蒋亮. 基于深度强化学习的弹道导弹中段突防控制[J]. *指挥信息系统与技术*, 2020, 11(4): 1-9, 27.
Nan Ying, Jiang Liang. Midcourse penetration and control of ballistic missile based on deep reinforcement learning[J]. *Command Information System and Technology*, 2020, 11(4): 1-9, 27(in Chinese).
- [15] 陈中原, 韦文书, 陈万春. 基于强化学习的多发导弹协同攻击智能制导律[J]. *兵工学报*, 2021, 42(8): 1638-1647.
Chen Zhongyuan, Wei Wenshu, Chen Wanchun. Reinforcement learning-based intelligent guidance law for cooperative attack of multiple missiles[J]. *Acta Armamentarii*, 2021, 42(8): 1638-1647(in Chinese).
- [16] Su W S, Li K B, Chen L. Coverage-based cooperative guidance strategy against highly maneuvering target [J]. *Aerospace Science and Technology*, 2017, 71: 147-155.

(编辑:李瑾)