基于复数域深度强化学习的多干扰场景雷达抗干扰方法

解 烽[①] 刘环宇*[①] 胡锡坤[②] 钟 平[②] 李君宝[①]

[①](哈尔滨工业大学计算学部信息对抗技术研究所 哈尔滨 150080)

[②](国防科技大学电子科学学院 长沙 410073)

摘要：在现代电子战中，雷达面临的干扰环境比以前更加复杂，机载干扰机会根据突袭任务与突袭阶段的不同而改变其干扰方式。近年来，基于强化学习的雷达抗干扰方法在单一干扰对抗场景下取得了一定进展，但在实际复杂多干扰场景下的研究仍有不足。为了解决该问题，本文提出了一种基于复数域深度强化学习的多干扰场景雷达抗干扰方法以优化频率捷变雷达的抗干扰策略。首先，针对突袭任务的阶段性特点建立了噪声瞄准干扰、距离假目标欺骗干扰与密集假目标转发干扰3种干扰模型，并设计了3种干扰顺序策略来模拟实际干扰场景。其次，针对多干扰场景模型，构建了一种融合信干噪比与目标航迹完整性的强化学习奖励函数，并针对干扰信号的复数域特征，提出了一种基于复数域深度强化学习的多干扰场景雷达抗干扰方法。最后，基于3种干扰顺序策略设计了雷达抗干扰仿真实验，结果表明，所提方法能够有效解决雷达面临的时序条件下复杂多干扰场景的主瓣干扰问题，与两种经典深度强化学习算法相比该方法抗干扰决策性能大幅提高，平均决策时间降低至405.3 ms。

关键词：复数域；深度强化学习；主瓣干扰；序贯干扰；频率捷变雷达

中图分类号：TN974 **文献标识码**：A **文章编号**：2095-283X(2023)06-1290-15

DOI: 10.12000/JR23139

引用格式：解烽, 刘环宇, 胡锡坤, 等. 基于复数域深度强化学习的多干扰场景雷达抗干扰方法[J]. 雷达学报, 2023, 12(6): 1290–1304. doi: 10.12000/JR23139.

Reference format: XIE Feng, LIU Huanyu, HU Xikun, *et al*. A radar anti-jamming method under multi-jamming scenarios based on deep reinforcement learning in complex domains[J]. *Journal of Radars*, 2023, 12(6): 1290–1304. doi: 10.12000/JR23139.

A Radar Anti-jamming Method under Multi-jamming Scenarios Based on Deep Reinforcement Learning in Complex Domains

XIE Feng[①] LIU Huanyu*[①] HU Xikun[②] ZHONG Ping[②] LI Junbao[①]

[①](*Information Countermeasure Technique Institute*, *Faculty of Computing*, *Harbin Institute of Technology*, *Harbin* 150080, *China*)

[②](*College of Electronic Science and Technology*, *National University of Defense Technology*, *Changsha* 410073, *China*)

Abstract: In modern electronic warfare, the jamming environment of radar is more complex than ever. The airborne jammer adapts its jamming method based on diverse raid missions and stages. Recently, the reinforcement learning–based radar anti-jamming method has made some progress in the confrontation scenario of single jamming; however, the gap with respect to actual complex multi-jamming scenarios is large. To

收稿日期：2023-07-31；改回日期：2023-10-19；网络出版：2023-11-09

*通信作者： 刘环宇 liuhuanyu@hit.edu.cn *Corresponding Author: LIU Huanyu, liuhuanyu@hit.edu.cn

基金项目：国家自然科学基金(62271166)，哈尔滨工业大学医工理交叉基金(IR2021104)

Foundation Items: The National Natural Science Foundation of China (62271166), Interdisciplinary Research Foundation of HIT (IR2021104)

责任主编：全英汇 Corresponding Editor: QUAN Yinghui

address this issue, this paper proposes a multi-jamming scenario radar anti-jamming method based on deep reinforcement learning in the complex domain to optimize the anti-jamming strategy of frequency agile radar. First, according to the stage characteristics of the raid mission, noise spot jamming, range deception jamming , and dense false target forwarding jamming models are established. The three jamming sequence strategies were designed to simulate actual jamming scenarios. Second, a reinforcement learning reward function that integrates the signal-to-noise ratio and target trajectory integrity is constructed for the multi-jamming scenario model. Thus, a multi-jamming scenario radar anti-jamming method based on deep reinforcement learning in a complex domain is proposed, which is based on the complex domain characteristics of the jamming signal. Finally, radar anti-jamming simulation experiments are performed based on the three jamming sequence strategies. The results show that the proposed method can effectively deal with the main-lobe jamming problem of complex multi-jamming scenarios under time-sequence conditions. Moreover, the average decision-making accuracy was improved, and the average decision-making time was reduced to 405.3 ms compared with the two classical reinforcement learning algorithms.

Key words: Complex domain; Deep Reinforcement Learning (DRL); Main-lobe jamming; Sequential jamming; Frequency agile radar

1　引言

认知雷达通过环境和目标的变化情况调整其波形发射策略，实现比传统雷达更好的抗干扰效果[1]。目前认知雷达面临的干扰主要为主瓣干扰。主瓣干扰是电子战领域中最常见的干扰方式，它会显著降低雷达系统的性能。当前抗主瓣干扰的技术分为有源对抗与无源抑制[2]。无源抑制的方法主要针对干扰信号特征不变的场景，研究雷达与干扰机的单次对抗过程。如果雷达和干扰机的博弈持续多个回合，并且干扰机采用灵活多变的干扰形式，那么无源抑制方法的抗干扰性能将大幅降低[3]。有源对抗技术要求认知雷达主动改变抗干扰策略，从根本上降低雷达被干扰的概率。有源对抗方式主要从空、时、频、极化等维度出发进行发射波形设计，并借助自适应滤波等信号处理手段达到抗干扰的目的。如在时域上使用基于压缩感知的抗射频干扰波形[4]、在频域上使用脉冲频率捷变方法[5,6]、在极化域上使用极化和接收极化联合优化的波形[7]、在空域上采用一发多收模式下基于多站波束融合的抗干扰方法[8]和结合时频域特征设计脉间-脉内捷变频雷达抗干扰方法[9]等。对不同的干扰类型(特别是主瓣干扰)和干扰样式而言，频域是一个重要且有效的可分域，因此目前学者大多数从频域出发设计智能抗干扰方法。频率捷变(Frequency Agile, FA)雷达的载波频率可以在每个脉冲中随机跳变，这使得干扰机难以预测雷达的载波频率，无法有效实施干扰。

近年来，人工智能技术，特别是深度学习[10,11]、深度强化学习[12,13]等相关理论的发展和成熟，极大地提高了雷达态势感知、自主学习、自主推理与决策的能力。深度强化学习(Deep Reinforcement Learning, DRL)技术通过建立智能体与环境的交互模型，使智能体充分探索环境信息，提高其行为决策的效果[14]。将DRL引入雷达抗干扰领域，通过采集雷达与干扰机的交互信息，优化雷达抗干扰行为策略，实现雷达抗干扰行为决策能力的提升[15]。使用DRL解决雷达抗干扰问题的研究已经取得了一些进展。文献[16]提出了一种基于强化学习的智能抗干扰方法，在抗干扰策略优化训练过程中，将阵列波束数据与脉冲压缩感知后的干扰状态特征作为模型输入，分别采用Q-learning算法与Sarsa算法对模型的值函数进行计算与迭代，实现抗干扰知识库的智能更新，根据知识库确定最优抗干扰策略。文献[17]使用Q-learning算法对雷达发射功率进行优化分配决策，但只考虑了有限的雷达发射状态。文献[18]以雷达发射行为和回波信号的信干比作为状态输入，使用Q-learning和深度Q网络(Deep Q-Network, DQN)两种算法对认知雷达抗干扰跳频策略进行优化。文献[19]以过去时刻的雷达发射行为作为状态输入，设计了一种基于DQN的FA雷达抗干扰策略。文献[18]与文献[19]中FA雷达面对的是同一种干扰类型的场景，此场景与实际复杂干扰场景差距较大，且文献[16–19]只考虑到信号特征层级，尚未实现从原始回波信号到波形发射的感知决策一体化。

在雷达抗干扰DRL算法中，合理的奖励函数设计可以有效提高DRL算法的收敛速度与最佳性能表现。通常以雷达抗干扰行为是否成功躲避干扰的不同赋值作为奖励函数[20]或以信干噪比作为奖励函数[18]。文献[19]从雷达能否检测到目标的角度出发设计了雷达检测概率作为奖励函数，文献[21]将检测目标的准确度作为奖励函数，文献[22]将雷达信

号被干扰信号遮盖程度和基于信噪比加权算法的检测概率作为奖励函数。

上述文献表明，雷达可以通过DRL算法选择有效的抗干扰行为，且DRL算法在雷达抗干扰领域的应用实现了雷达与干扰机多轮交互博弈的要求，更符合现代电子战的场景假设。

然而，上述雷达抗干扰研究局限于特定单一干扰场景下抗干扰波形参数的寻优问题，并未深入探讨时序条件下复杂多干扰场景的抗干扰行为决策问题。在突防任务中，不同阶段下干扰机选择的干扰方式不同，这意味着不同阶段中雷达所受的干扰策略是非稳定的。因此有必要研究时序条件下复杂多干扰场景的雷达抗干扰行为决策方法。

本文设计了一种工作在多干扰策略模式下的干扰机模型，通过对干扰环境与FA雷达的建模仿真，并结合干扰信号的复数域特征，提出了一种基于复数域深度强化学习的多干扰场景雷达抗干扰方法。本文主要贡献如下：

(1) 在DRL框架下建立了3种干扰模型，包括噪声瞄准干扰、距离假目标欺骗干扰与密集假目标转发干扰。构建了基于上述3种干扰的时序干扰策略，并在该策略的基础上搭建了适用于对抗博弈的复杂电磁环境。

(2) 针对时序多干扰场景，提出了一种基于复数域DRL的雷达抗干扰算法。该算法使用本文提出的针对多干扰场景的奖励函数进行优化，并实现了感知决策一体化的端到端应用。通过对比两种经典DRL算法的抗干扰性能，证明了该算法在雷达抗干扰行为决策上具有较高的准确性与较快的速度。

(3) 在FA雷达与干扰机模型的基础上，设计了一种针对雷达多干扰场景的DRL奖励函数。该奖励函数融合了雷达信干噪比与目标航迹完整性的评价方法，实验验证了该奖励函数在雷达多干扰场景下的有效性。

2 系统模型

本文研究脉冲级FA雷达与自卫式干扰机之间的博弈问题。脉冲级FA雷达可以调控一个相干处理间隔(Coherent Processing Interval, CPI)中每个脉冲发射的频率与持续时间，自卫式干扰机工作在收发分时干扰模式。本节描述了发射线性调频(Linear Frequency Modulated, LFM)信号的FA雷达模型与3种干扰类型的信号特征。

2.1 FA雷达模型

FA雷达指各发射脉冲载频频率在带宽范围内按某种规律快速变化的一种脉冲体制雷达[23]，如图1所示。

与恒定载波频率雷达不同，FA雷达按照频率捷变方式可分为脉内捷变频、脉间捷变频和脉组间捷变频3种方式，本文主要研究脉间FA雷达。现代FA雷达一般采用全相参脉冲体制，雷达的频率综合器可以在雷达载波频率跳变的同时实现各脉冲相位相参，这也保证了目标能够被有效地相参处理与合成。

FA雷达载波频率捷变信号模型可以表示为

$$s_t\left(\hat{t},t_m\right)=u(t)\exp\left(\mathrm{j}2\pi f_m\left(\hat{t}+t_m\right)\right) \tag{1}$$

其中，$u(t)$表示信号的复包络；$t=\hat{t}+t_m$，$\hat{t}$和t_m分别表示快时间和慢时间，慢时间$t_m=mT_\mathrm{r}$，T_r是脉冲重复间隔(Pulse Repetition Interval, PRI)；f_m表示第m个脉冲的载频频率，表示为

$$f_m=f_\mathrm{c}+a(m)\Delta f,\ m=1,2,\cdots,M \tag{2}$$

其中，$a(m)$为随机整数，表示第m个脉冲的频率控制码，$a(m)$取值范围为$[0,\ N-1]$；N为频率带宽内允许的总跳频数，M为脉冲积累数，满足$N>M$，Δf表示最小频率间隔，通常情况下为了保证脉冲信号间的正交性，Δf需要满足

$$\Delta f=n/T_\mathrm{p} \tag{3}$$

其中，n为正整数，T_p为脉冲宽度。

FA雷达一般采用的LFM信号的复包络表示为

$$u(t)=\mathrm{rect}\left(\frac{\hat{t}}{T_\mathrm{p}}\right)\exp\left(\mathrm{j}\pi\gamma\hat{t}^2\right) \tag{4}$$

其中，$\mathrm{rect}\left(\dfrac{\hat{t}}{T_\mathrm{p}}\right)=\begin{cases}1, & |\hat{t}|\le\dfrac{T_\mathrm{p}}{2}\\ 0, & \text{其他}\end{cases}$为窗函数；$\gamma=B/T_\mathrm{p}$为调频斜率，$B$为信号带宽。

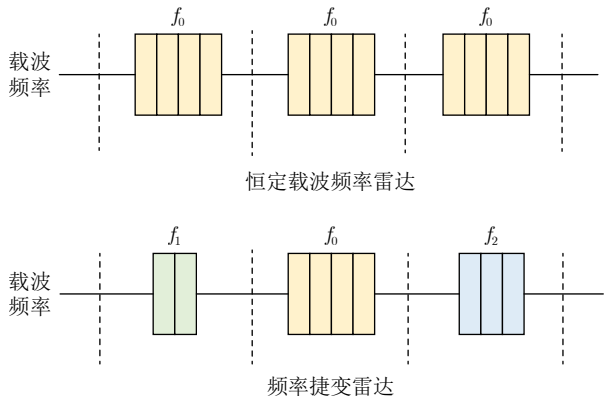


图 1 频率捷变雷达模型

Fig. 1 Frequency agile radar model

2.2 干扰机模型

飞行目标上装有自卫式干扰机装置，本文假设博弈环境中FA雷达主波束与飞行目标在空间上是正对关系，因此干扰机施加的干扰主要从雷达主瓣进入，故FA雷达主要受到主瓣干扰。根据飞行任务与阶段的不同，博弈环境中的干扰机主要发射3种干扰类型，其中每种干扰类型的干扰波形参数都是可变的。

(1) 干扰类型1：噪声瞄准干扰。

在噪声瞄准干扰模式下，干扰机首先截取一个CPI内第1个雷达脉冲，分析该脉冲的载频，选取可用频带宽度内的子频带发射窄带干扰信号。一般情况下，窄带干扰信号的带宽会全覆盖雷达探测波形的频带，使得雷达接收机无法有效分离真实回波信号特征，导致雷达系统的性能严重下降。噪声瞄准干扰中雷达与干扰信号时、频域关系如图2所示。

(2) 干扰类型2：距离假目标欺骗干扰。

在距离假目标欺骗干扰模式下，干扰机首先截取一个CPI内第1个雷达脉冲，分析脉冲载频，然后在CPI剩余时间内多次转发相同频率的干扰波形信号。使得雷达接收机接收到的雷达信号有多组同频分时的雷达峰值，达到混淆真实回波的距离特征维度信息的目的。距离假目标欺骗干扰中雷达与干扰信号时、频域关系如图3所示。

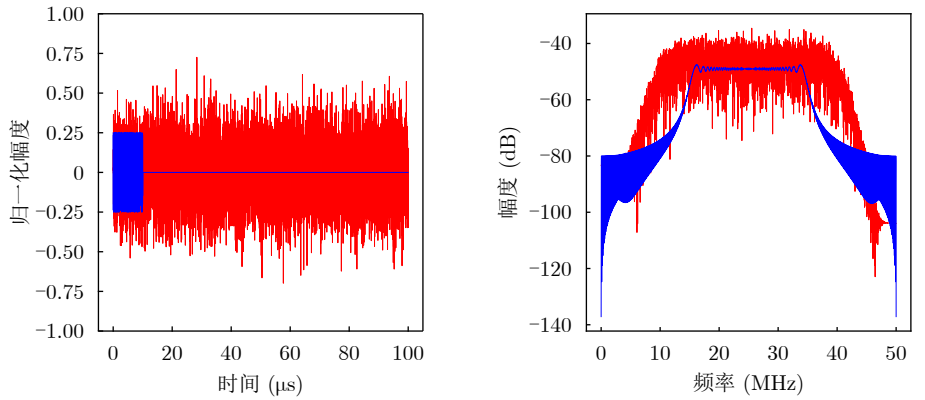
假设在距离假目标欺骗干扰中忽略慢时间t_m的影响，则此时LFM信号可表达为

$$s_m(t)=\mathrm{rect}\left(\frac{t}{T_\mathrm{p}}\right)\exp\left[\mathrm{j}2\pi\left(f_m t+\frac{\gamma}{2}t^2\right)\right] \tag{5}$$

则雷达接收机中截获到的真实目标回波可表示为

$$s_\mathrm{R}(t)=u(t)\exp\left\{\mathrm{j}2\pi\left[f_m\left(t-\tau_\mathrm{R}\right)+\frac{\gamma}{2}\left(t-\tau_\mathrm{R}\right)^2\right]\right\} \tag{6}$$

其中，$u(t)=A_\mathrm{R}\mathrm{rect}\left(\dfrac{t-\tau_\mathrm{R}}{T_\mathrm{p}}\right)$；$A_\mathrm{R}$代表真实目标

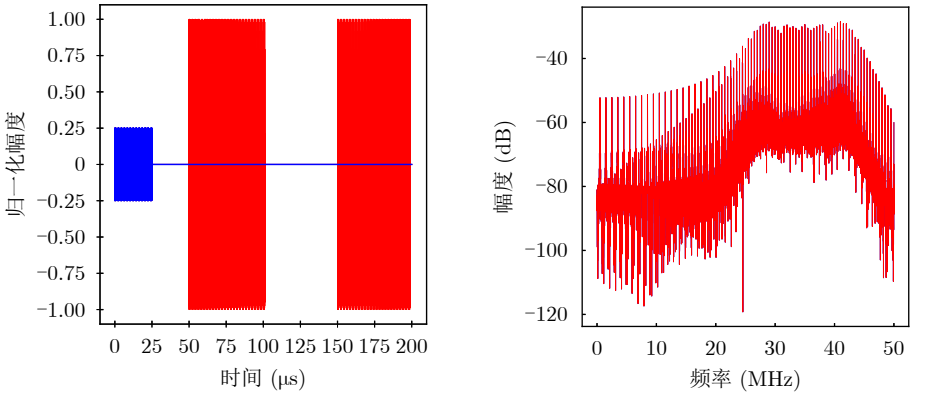


(a) 雷达/干扰机时域
(a) Radar/jammer time domain

(b) 雷达/干扰机频域
(b) Radar/jammer frequency domain

图 2 噪声瞄准干扰仿真图

Fig. 2 Simulation diagram of noise spot jamming



(a) 雷达/干扰机时域
(a) Radar/jammer time domain

(b) 雷达/干扰机频域
(b) Radar/jammer frequency domain

图 3 距离假目标欺骗干扰仿真图

Fig. 3 Simulation diagram of distance false-target deception jamming

回波信号的幅度；$\tau_{\mathrm{R}}=2R_0(t)/\mathrm{c}$表示干扰机的转发时延；$R_0(t)$代表真实目标与雷达接收机的相对距离。

距离假目标欺骗干扰信号的表达式如下，其中干扰机改变了回波的延时参数。

$$s_{\mathrm{J}}(t)=u(t)\exp\left\{\mathrm{j}2\pi\left[f_m(t-\tau_{\mathrm{J}})+\frac{\gamma}{2}(t-\tau_{\mathrm{J}})^2\right]\right\} \tag{7}$$

其中，$u(t)=A_{\mathrm{J}}\mathrm{rect}\left(\dfrac{t-\tau_{\mathrm{J}}}{T_{\mathrm{p}}}\right)$；$A_{\mathrm{J}}$代表干扰回波信号的幅度；$\tau_{\mathrm{J}}=2R_{\mathrm{J}}(t)/\mathrm{c}$代表干扰回波经过调制之后的转发延时；$R_{\mathrm{J}}(t)$代表假目标与接收机的相对距离。

(3) 干扰类型3：密集假目标转发干扰。

在密集假目标转发干扰模式下，干扰机在时序上通常分为侦察窗与干扰窗两个阶段，在侦察窗阶段采样存储，在干扰窗阶段转发生成干扰。雷达接收机收到密集同频信号后难以有效提取真实目标的特征，因此密集假目标转发干扰既能产生压制效果又能产生欺骗干扰效果。

假设在密集假目标转发干扰场景下雷达发射信号$s_{\mathrm{d}}(t)$的表达式为

$$s_{\mathrm{d}}(t)=\mathrm{rect}\left(\frac{t}{T_{\mathrm{p}}}\right)\exp(\mathrm{j}2\pi f_m t) \tag{8}$$

其中，雷达接收机接收Q个距离不同的真实目标回波信号的表达式为

$$s_{\mathrm{R}}(t)=\sum_{q=1}^{Q}A_{\mathrm{R}}^{q}\mathrm{rect}\left(\frac{t-\tau_{\mathrm{R}}^{q}}{T_{\mathrm{p}}}\right)\exp\left[\mathrm{j}2\pi f_m\left(t-\tau_{\mathrm{R}}^{q}\right)\right] \tag{9}$$

其中，A_{R}^{q}代表第q个真实目标的回波幅度；τ_{R}^{q}代表第q个真实目标回波的转发时延。

干扰假目标回波信号的表达式为

$$s_{\mathrm{J}}(t)=\sum_{p=1}^{P}A_{\mathrm{J}}^{p}\mathrm{rect}\left(\frac{t-\tau_{\mathrm{J}}^{p}}{T_{\mathrm{p}}}\right)\exp\left[\mathrm{j}2\pi f_m\left(t-\tau_{\mathrm{J}}^{p}\right)\right] \tag{10}$$

其中，P代表密集假目标转发干扰的假目标个数；A_{J}^{p}代表第p个干扰假目标的幅度；τ_{J}^{p}代表第p个干扰假目标的转发时延。密集假目标转发干扰中雷达与干扰信号时域关系如图4所示。

上述3种干扰类型中的具体波形由波形控制参数决定。在每一轮干扰博弈过程中，干扰机根据概率从可用的类型中选择一个具体的干扰行为。由于每个时刻生成的干扰类型与波形控制参数是不确定的，所以此假设符合雷达面临的动态干扰环境的要求。

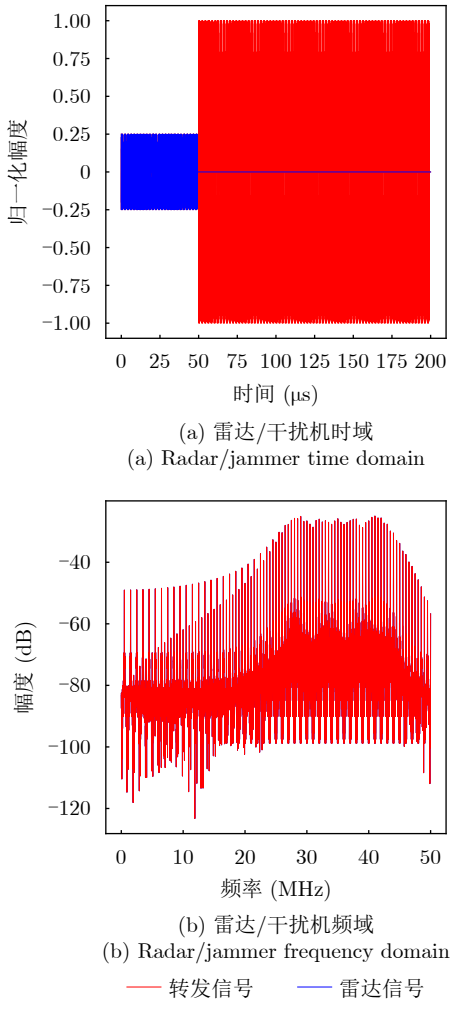
3 雷达与干扰机强化学习要素设计

本文把FA雷达看作DRL要素中的智能体，把干扰看作动态可变的环境特征。此外本文考虑了多种干扰类型下的复杂干扰策略，且每种干扰类型都由不同波形参数所决定。下面将详细描述DRL概念中的各元素设计方法。

3.1 环境状态设计

通常情况下，FA雷达首先处于探测状态，向环境空间发射波形，改变电磁环境s。假设干扰机在时刻t侦测到雷达探测波形，则其将根据态势与专家策略选择相应的干扰类型施加干扰，此时FA雷达会根据态势选择对应的抗干扰行为。本文设计的博弈场景将干扰机与FA雷达各改变一次波形的过程定义为进行了一回合博弈，每一回合博弈结束后，环境状态空间会被重置。

本文将2.2节的3种干扰类型按照一定策略进行编排，如图5所示。



(a) 雷达/干扰机时域
(a) Radar/jammer time domain

(b) 雷达/干扰机频域
(b) Radar/jammer frequency domain

图 4 密集假目标转发干扰仿真图

Fig. 4 Simulation diagram of dense false-target repeater jamming

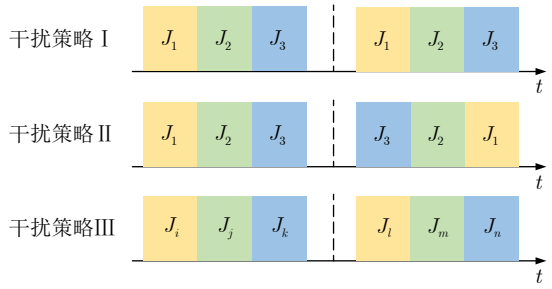


图 5 3种干扰策略顺序

Fig. 5 Order of three jamming strategies

干扰类型J_1表示噪声瞄准干扰，J_2表示距离假目标欺骗干扰，J_3表示密集假目标转发干扰。干扰策略 I 表示3种干扰按照顺序进行切换，每一个CPI切换一种干扰类型。策略 II 表示3种干扰按照回文顺序进行切换，策略III表示3种干扰随机进行切换，每一个CPI后随机切换到下一种干扰类型。

此外，上述3种干扰类型也具有其独立的波形控制参数。

在噪声瞄准干扰中，t时刻干扰机状态可以表示为$s_t=[s_{t,f},s_{t,B}]$，其中，$s_{t,f}$表示干扰机发射信号的下频频率，$s_{t,B}$表示信号带宽。

在距离假目标欺骗干扰中，t时刻干扰机状态为$s_t=[s_{t,f},s_{t,B},s_{t,t_1},...,s_{t,t_N}]$，其中$s_{t,f},s_{t,B}$与类型1中含义相同，$s_{t,t_j}$表示第$j$个假目标的时间标度。$N$表示在$t$时刻产生与原始回波特征相同的假目标个数。

在密集假目标转发干扰中，t时刻干扰机状态可以表示为$s_t=[s_{t,f},s_{t,B},s_{t,t_1},s_{t,t_2}]$，其中$s_{t,f},s_{t,B}$与类型1含义相同，$s_{t,t_1}$表示干扰机侦察窗的持续时间，$s_{t,t_2}$表示干扰窗的持续时间。假设在$t_1$时间内包含有雷达的探测波形，则干扰机将会在$t_2$时间内大量复制雷达波形并持续发射，使得雷达接收机处于资源过饱和状态，无法分辨真实目标的特征信息。

3.2 抗干扰行为设计

FA雷达的抗干扰波形决策对应着DRL中智能体对环境输出的行为a。FA雷达可以决定每一时刻生成波形的载频频率和持续时间等要素。

在噪声瞄准干扰中，FA雷达通常使用频率捷变抗干扰波形，假设t时刻雷达的发射波形可以表示为$a_t=\left[a_{t,f_m},a_{t,\gamma},a_{t,T_\mathrm{p}}\right]$，其中，$a_{t,f_m}$表示$t$时刻雷达信号中第$m$个脉冲的载频频率，$a_{t,\gamma}$表示$t$时刻雷达信号的调频斜率，$a_{t,T_\mathrm{p}}$表示$t$时刻雷达信号中每个脉冲的脉冲宽度。向量中3个元素将唯一确定t时刻发射的雷达脉冲波形，其他参数设定为固定值，忽略其他参数对波形生成的影响。

在距离假目标欺骗干扰中，雷达使用频率正交线性调频信号[24]可以有效对抗干扰。频率正交线性调频信号利用邻近两个发射信号的正交性对脉冲压缩信号幅度的影响对抗假目标欺骗干扰。雷达抗干扰波形可以表示为$a_t=\left[a_{t,f_m},a_{t,\gamma},a_{t,T_\mathrm{p}}\right]$，含义与类型1中相同。

在密集假目标转发干扰中，雷达使用掩护脉冲信号[24]可以有效对抗干扰。其原理是雷达在干扰机的侦察窗时间内发射掩护脉冲，使得干扰机复制转发与掩护脉冲特征相同的信号，并让雷达在干扰机的干扰窗时间内发射真实探测信号，最终，雷达接收端接收到真实目标回波信号与干扰机生成的大量掩护脉冲假信号，信号处理模块根据两种信号特征的不同进而分离出真实目标回波信号[25]。假设t时刻雷达的抗干扰波形可以表示为$a_t=[a_{t,f_{m1}},a_{t,f_{m2}},a_{t,T_{\mathrm{p}1}},a_{t,T_{\mathrm{p}2}}]$，其中$a_{t,f_{m1}}$表示掩护脉冲的载频频率，$a_{t,f_{m2}}$表示真实脉冲的载频频率，$a_{t,T_{\mathrm{p}1}}$表示掩护脉冲的持续时间，$a_{t,T_{\mathrm{p}2}}$表示真实脉冲的持续时间。

3.3 奖励函数设计

在雷达抗干扰场景中，合理的奖励函数设计可以有效提高强化学习算法的收敛速度与最佳性能表现。本文将从两个角度评价FA雷达的抗干扰决策性能：短时行为评价和长时连续性评价。

本文使用雷达获得目标的信干噪比[18](Signal-to-Interference-plus-Noise Ratio, SINR)与干扰机/雷达信号间特征关系两个指标表征短时行为评价体系。SINR是雷达系统中一个重要的性能指标，它表示目标回波信号与干扰和噪声的比值，即在接收端接收到的目标信号功率与干扰加噪声功率之比。本文设计的干扰场景包含一部雷达与一部机载干扰机，如图6所示。

假设雷达同时受到目标携带的自卫式干扰机与环境杂波的影响，且目标此时的雷达散射截面积(Radar Cross Section, RCS)值为σ，从FA雷达到目标干扰机的信道增益为h_s，环境杂波的噪声功率为P_n。假

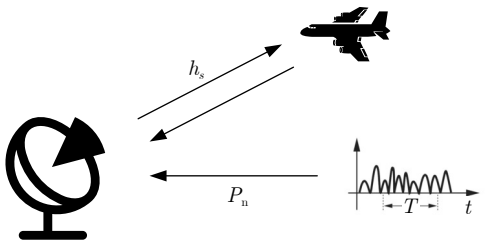


图 6 FA雷达与干扰机

Fig. 6 FA radar and target jammer

设f_n为FA雷达第n个脉冲的载波频率，f_{J}为干扰机的频率，则此时FA雷达的SINR值可表示为

$$\mathrm{SINR}=\frac{P_{\mathrm{s}}h_s^2\sigma}{P_{\mathrm{n}}+P_{\mathrm{J}}h_sI(f_{\mathrm{J}}=f_n)} \tag{11}$$

其中，P_{s}是FA雷达发射功率，P_{n}是FA雷达接收环境噪声的功率，P_{J}是自卫式干扰机功率，如果$f_{\mathrm{J}}=f_n$，则$I(f_{\mathrm{J}}=f_n)$为1，否则为0。SINR越大，表示目标信号越容易被接收机检测到，雷达系统的性能也就越好。

在短时行为评价体系中，针对3种干扰类型分别设计特定的奖励函数。对于噪声瞄准干扰，雷达采用频率捷变抗干扰波形，单步博弈的奖励函数如下所示：

$$R=\begin{cases}30, & \begin{array}{l}\text{if } \mathrm{signal_{high}}<\mathrm{jam_{low}}\\ \text{or } \mathrm{signal_{low}}>\mathrm{jam_{high}}\end{array}\\ \mathrm{SINR}, & \begin{array}{l}\text{if } \mathrm{signal_{low}}<\mathrm{jam_{low}}<\mathrm{signal_{high}}\\ \text{or } \mathrm{signal_{low}}<\mathrm{jam_{high}}<\mathrm{signal_{high}}\end{array}\\ -100, & \text{if } \mathrm{jam_{low}}<\mathrm{signal_{low}}<\mathrm{signal_{high}}<\mathrm{jam_{high}}\end{cases} \tag{12}$$

式(12)中符号如图7所示，当雷达未被干扰时，奖励值为30，雷达被部分噪声信号干扰时，用SINR表示奖励值，当雷达全部被干扰时，用–100表示奖励结果。

针对距离假目标欺骗干扰，雷达采用频率捷变抗干扰波形，如图8所示，动作空间与噪声瞄准干扰相同，但由于距离假目标抗干扰波形的发射频率需具备正交性，故根据文献[26]设计单步奖励函数如下所示：

$$R=\begin{cases}|2n-\max(n)|, & \text{if } f_{\mathrm{signal}}=n\Delta f\\ -\max(n), & \text{else}\end{cases} \tag{13}$$

其中，n为正整数，f_{signal}为发射波形的频率，Δf为最小频差。

如果捷变前后频率满足最小频差，则经过信号处理后的真假目标间存在功率峰值的区别。当相邻发射信号间的频差增大时，互相关函数值会减小，信号间也就更加正交。但频差的增大导致雷达总体带宽的增大，加大了工程实现难度，故设定雷达发射频率的中间频段奖励值最大。

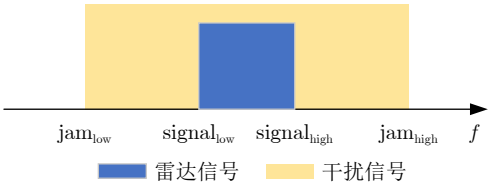


图 7 噪声瞄准干扰频域图

Fig. 7 Frequency domain of noise spot jamming

针对密集假目标转发干扰，雷达通常采用掩护脉冲抗干扰波形，单步博弈的奖励函数如下所示：

$$R=\begin{cases}30, & \text{if } t_{\mathrm{cheat}}>t_{\mathrm{observe}}\\ \mathrm{SINR}, & \text{else}\end{cases} \tag{14}$$

式(14)中符号如图9所示。其中t_{cheat}代表雷达发射掩护脉冲的时间，t_{observe}代表干扰机观察窗口的时间，一般认为当掩护脉冲持续时间完全覆盖干扰机侦察时间时，其所发射的真实脉冲可以不被干扰，这种情况下获得奖励值最大。

在长时连续性评价体系中，主要关注雷达持续探测目标的能力，即目标航迹完整性[27]。FA雷达通过调整抗干扰波形，可在单步博弈中获得较大的SINR，提高单位时间内的目标探测性能，进而提高全过程目标的航迹完整性。

假设一个对抗局存在500个博弈回合，每个博弈回合称为一个干扰元时间，则目标航迹完整性可以用成功博弈回合数与总博弈回合数的比值Pd表示，Pd表达式为：

$$\mathrm{Pd}=\frac{N_{\mathrm{succeed}}}{N_{\mathrm{total}}} \tag{15}$$

其中，N_{succeed}表示3种干扰类型中博弈成功的回合数，N_{total}表示全局博弈的总回合数。Pd值越高，说明雷达探测目标的航迹完整性越高。在3种干扰类型中，类型1跳频后频率在干扰频率外判定成功，类型2跳频后频率是最小频差的正整数倍判定成功，类型3掩护窗时间长于侦察窗时间判定成功。

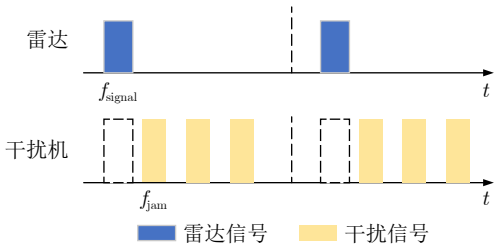


图 8 距离假目标欺骗干扰时域图

Fig. 8 Time domain of distance false-target deception jamming

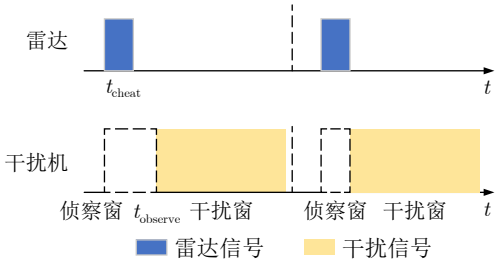


图 9 密集假目标转发干扰时域图

Fig. 9 Time domain of dense false-target repeater jamming

本文研究的雷达与干扰机博弈场景存在多回合与长时效的特点，故设计的奖励函数优势在于融合了单一短时抗干扰与全局长时抗干扰的价值评价方法，并结合了雷达SINR与目标航迹完整性的实际物理意义。

4　雷达抗干扰波形决策优化算法设计

本文提出了一种基于复数域DRL的多干扰场景雷达抗干扰网络(Deep RL based radar Anti-jamming Network under multi-jamming scenes in Complex Domain, DRL-ANCD)，如图10所示。

假设初始环境下雷达发射的LFM信号脉冲重复周期为50 μs，使用100 MHz采样率进行采样得到5000×1的雷达离散波形。将离散波形的复数域特征分为实部和虚部分别进行提取，内部网络图如图11所示。

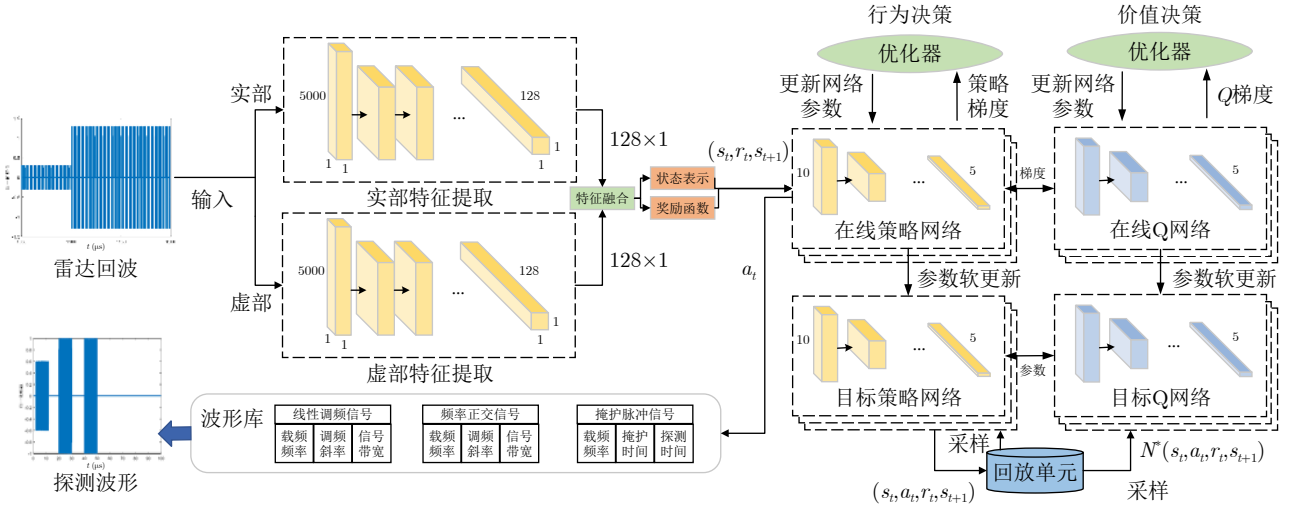


图 10 基于复数域深度强化学习的多干扰场景雷达抗干扰网络

Fig. 10 Deep RL based radar anti-jamming network under multi-jamming scenes in complex domain

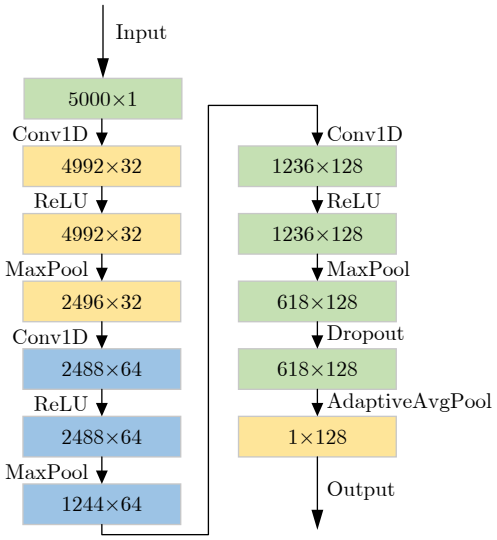


图 11 复数域特征提取网络

Fig. 11 Complex domain feature extraction network

图11中每次一维卷积将通道数翻倍，分别为32, 64, 128，其中一维卷积的卷积核大小为9。每次卷积后使用ReLU激活并进行最大池化操作，卷积核为2，步幅为2。最后将618×128维的特征进行可适应性平均池化，得到1×128维度的向量。将两个1×128向量特征融合，得到256×1向量，经过两个全连接层与一个全局平均池化层后输出感知环境的特征向量。

根据奖励函数的设计，使用此特征向量计算出上一时间节点雷达抗干扰决策的有效性。并将此特征向量与奖励函数值以(s_t, r_t, s_{t+1})形式输入至行为决策网络。

DRL-ANCD行为决策部分集成了3个独立演化的深度确定性策略梯度网络，对应3种干扰类型信号特征。深度确定性策略梯度网络采用策略网络和价值Q网络两组网络并延续DQN[28]中固定目标网络的思想，每组网络再细分为在线网络和目标网络。网络结构图如图12所示，伪代码如算法1所示。

在线策略网络的输入为对抗环境中态势预测的环境特征向量与奖励函数模块计算得到的奖励值，输出为一个确定性的动作$a = \mu_\theta(s)$。以往的策略梯度使用随机的方法在当前策略下进行行为采样，这严重降低了样本的利用效率，也增加了网络的计算负担。本文采用确定性策略，即网络策略可由函数$\mu_\theta(s)$表示，其中θ为神经网络的参数。此外，在线策略网络还有一个相同结构但不同参数的目标策略网络，实现对在线策略网络的软更新。

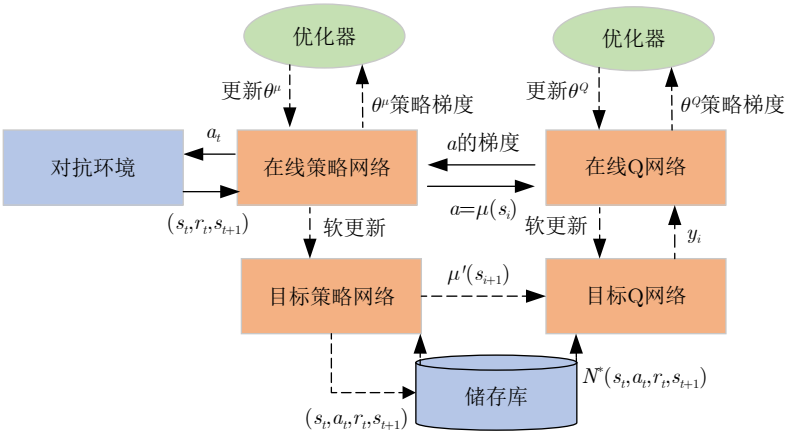


图 12 深度确定性策略梯度网络

Fig. 12 Deep deterministic policy gradient network

算法 1 深度确定性策略梯度算法

Alg. 1 Deep deterministic policy gradient algorithm

1. 使用权重θ^{Q}和θ^{μ}随机初始化Q网络参数$Q\left(s,a\mid\theta^{Q}\right)$和策略网络参数$\mu\left(s\mid\theta^{\mu}\right)$
2. 使用初始化目标网络
3. 使用权重$\theta^{Q'}\leftarrow\theta^{Q}$, $\theta^{\mu'}\leftarrow\theta^{\mu}$初始化目标网络$Q'$和$\mu'$
4. 初始化经验池R
5. for episode=1, 2, ⋯, M，执行：
6. 为行动探索初始化一个随机过程N
7. 获得一个初始化观察状态s_1
8. for $t=1,2,\cdots,T$，执行：
9. 根据当前策略与探索噪声选择行动a_t
10. 执行动作a_t,获得奖励r_t与新的状态s_{t+1}
11. 将样本(s_t,a_t,r_t,s_{t+1})存储至经验池R
12. 从R中随机采样出N个样本(s_i,a_i,r_i,s_{i+1})
13. 设置$y_i=r_i+\gamma Q'\left(s_{i+1},\mu'\left(s_{i+1}\mid\theta^{\mu'}\right)\mid\theta^{Q'}\right)$
14. 使用损失函数L更新Q网络参数
15. 使用采样样本的策略梯度更新行为策略
16. 更新目标网络参数：

$$\theta^{Q'}\leftarrow\tau\theta^{Q}+(1-\tau)\,\theta^{Q'}$$
$$\theta^{\mu'}\leftarrow\tau\theta^{\mu}+(1-\tau)\,\theta^{\mu'}$$

17. end for
18. end for

与策略网络相似，Q网络分为在线Q网络和目标Q网络。在线Q网络的输入是当前状态的观测值和在线策略网络的输出动作，目标Q网络的输入则是当前目标策略网络的输出动作。这两个网络的输出为当前状态的价值Q，在线Q网络目的是拟合价值函数$Q_{\omega}(s,a)$。

Q网络使用TD-error的梯度下降规则进行更新，并结合真实收益r与下一时刻的价值Q_{ω}得到Q_{target}，使用Q_{target}与当前价值Q的均方差作为梯度下降的损失值。策略网络使用梯度上升规则进行更新，梯度上升规则保证了输出的最优动作a的价值Q最大。

DRL-ANCD同时使用了经验回放方法，在训练阶段中将一段时间的序列(s,a,r,s')存储到经验池，每个训练回合需要从经验池随机采样一个批次的样本数据进行训练，提高了样本利用率和训练稳定性。同时其行为决策部分的3个独立演化网络采用分布训练，统一决策的运行模式。

5 仿真实验

本节通过3个仿真实验验证DRL-ANCD网络的抗干扰效果。实验1单独测试特征提取网络，验证特征提取网络识别雷达干扰类型的能力；实验2单独测试决策网络，验证其在单一干扰类型场景下的决策性能；实验3在构建的3个策略序贯多干扰场景下验证DRL-ANCD的抗干扰决策性能。

仿真所使用的计算机硬件参数为：32 GB RAM, Intel i7-12700K CPU, NVIDIA RTX 3090 GPU, Python版本为3.9, Pytorch版本为1.12.0。

5.1 态势预测

假设初始环境下雷达发射信号类型与参数如表1所示，在一个CPI中，雷达波形原始信号被离散采样为5000个点，将波形信号分为实部与虚部，分别传递给态势预测网络的两个通道进行特征提取，将两个通道提取的1×128维特征合并后进行态势预测。

本节设计了一组针对3种干扰类型的态势预测实验，每种干扰类型生成了1000个无序样本，保证了其差异化，并尽量覆盖了雷达工作频段的全部情况，训练样本分布如表2所示。学习率设置为0.005，批样本量为128，损失函数使用MSE损失，优化器使用Adam。

训练过程的损失值和准确度如图13、图14所示。为了准确分析混合类型下的识别精度的影响因素，分别针对3种干扰类型进行测试，随机抽取同一类型下300个不同干扰样本进行对比，识别时间和识别精度如表2所示。针对3种干扰类型的平均识别时间为124 ms，平均预测准确度达到96.8%，结果表明，态势预测网络在3种干扰类型的识别速度和精度上取得了很好的效果，可以为决策网络提供环境特征的支持。

5.2 单一干扰场景下抗干扰行为决策

本节实验将在单一干扰场景下进行3种DRL算法的性能验证。根据5.1节所列的每个干扰类型的动作状态空间和奖励值设定规则，使用DRL-ANCD网络、PPO网络[29]和双延迟深度确定性策略梯度网络(Twin Delayed Deep Deterministic policy gradient, TD3)[30]3个算法进行实验。3个算法的参数设置如表3所示。

表 1　雷达发射信号仿真参数表

Tab. 1　Radar transmit signal simulation parameters

参数类型	数值
信号类型	LFM
采样频率 f_s (MHz)	100
脉冲宽度 T_p (μs)	10
脉冲重复周期 T_r (μs)	50
下变频后的中频频率 f_I (MHz)	25
调频斜率 k (Hz/s)	2×10^{12}
带宽 B (MHz)	20

3种干扰类型下不同DRL的决策性能如图15与表4所示。图15(a)表明3种DRL算法对于噪声瞄准干扰均可以较快地收敛，并且最终达到的最优策略决策性能中，DRL-ANCD获得回合奖励最大，平均决策时间为244 ms；图15(b)表明DRL-ANCD与TD3两个算法对于距离假目标欺骗干扰稳定性较高，但TD3算法的决策时间大于DRL-ANCD与PPO算法；图15(c)表明在密集假目标转发干扰场景下，DRL-ANCD获得的回合奖励值最大，TD3算法训练达到稳定的时间最长。符合预期结果。

以部分博弈回合过程为例对训练所得DRL-ANCD模型的决策行为进行分析，模型对于3种干扰类型的抗干扰行为决策结果如图16所示。图16(a)表明雷达频率捷变后波形频率在干扰机干扰频带外；图16(b)表明雷达对于距离假目标欺骗干扰采取的跳频频率可以满足波形设计的频率正交规则，可以提高后续目标分离的显著性；图16(c)表明在密集假目标干扰时，雷达发射掩护脉冲的时间窗大于干扰机的侦察窗，经信号处理后可以有效分辨真实目标信息。

5.3 时序多干扰场景下抗干扰行为决策

本节实验将在时序多干扰场景下进行3种算法的性能验证。由于3种干扰类型对应的状态空间、动作空间均不同，故将DRL-ANCD的态势预测网络与PPO, TD3算法分别连接，称连接后的网络为PPO-SL与TD3-SL。此时DRL-ANCD, PPO-SL与TD3-SL网络均具备处理复杂多干扰场景的能力，在此基础上比较3种算法的决策性能。

表 2　3种干扰类型下的态势预测性能

Tab. 2　Posture prediction performance under 3 interference types

干扰类型	总体区间	步进	识别时间(ms)	识别精度(%)
噪声瞄准干扰	[3～4 GHz]	1 MHz	96	98.6
距离假目标欺骗干扰	[3～4 GHz]	1 MHz	132	97.4
密集假目标转发干扰	[1～1000 μs]	1 μs	144	94.4

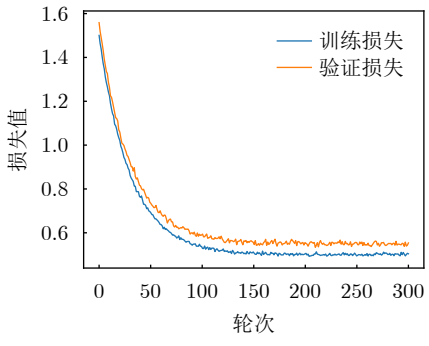


图 13 态势预测过程损失值

Fig. 13 Loss value of situation awareness process

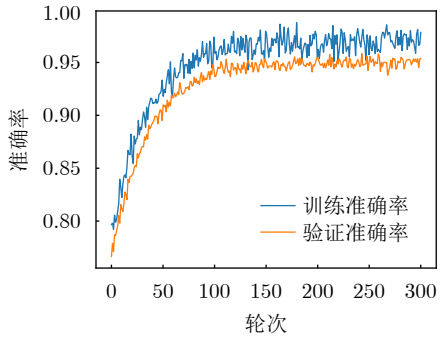


图 14 态势预测过程准确率

Fig. 14 Accuracy value of situation awareness process

表 3 算法参数设置

Tab. 3 Algorithm parameters setting

参数	PPO	TD3	DRL-ANCD
Q网络学习率	10^{-3}	10^{-3}	10^{-3}
策略网络学习率	10^{-3}	10^{-3}	10^{-3}
优化器	Adam	Adam	Adam
目标网络更新率	10^{-3}	5×10^{-3}	5×10^{-3}
批输入	128	128	128
折扣系数	0.99	0.99	0.99
奖励缩放	1.0	1.0	1.0
PPO裁剪参数	0.2	None	None

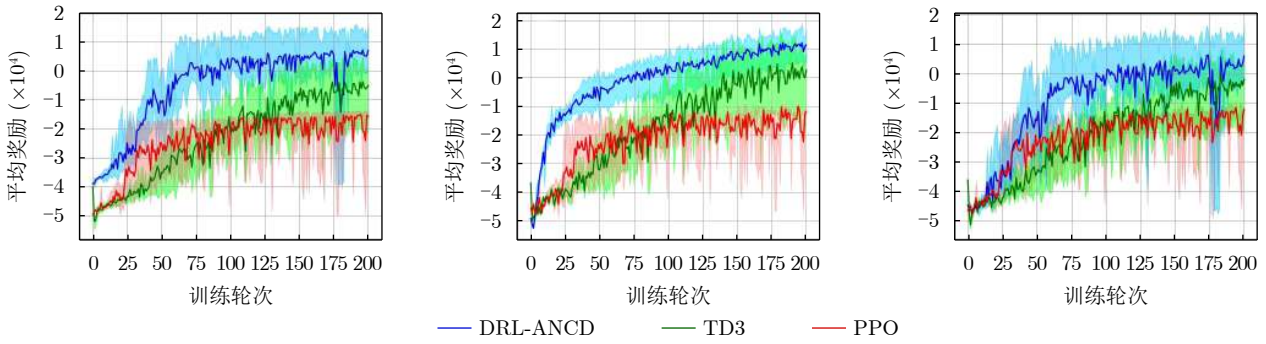
实验网络参数设置如表3所示，态势预测网络的学习率设置为0.005。表5为DRL-ANCD算法中策略网络与Q网络的参数设置，DRL-ANCD网络决策计算一次前向传播占用3.69 Mb的访存空间与0.28GFLOPs的计算量。假设雷达与干扰机在每个对抗局中交互500个回合，每组实验训练200个对抗局，设置5个随机种子。3个干扰策略下的决策性能如图17所示，实验结果如表6所示。

图17(a)表明在干扰策略Ⅰ场景下，DRL-ANCD算法奖励值最高，获得的对抗奖励为3，远大于PPO-SL与TD3-SL，经过75轮次训练模型决策性能趋于稳定。

图17(b)表明，针对回文顺序的干扰策略场景，3种算法的性能表现与策略Ⅰ相似。说明干扰策略Ⅰ与Ⅱ中干扰顺序的改变对雷达抗干扰决策的性能影响较小。DRL-ANCD算法的平均回合奖励值为14，平均决策时间为392 ms。

图17(c)表明在随机干扰情况下，经验池内样本丰富度的增加将极大缩短DRL-ANCD算法探索空间需要的时间，故在相同的训练轮次内，DRL-ANCD算法可以较快地学习更多干扰样本特征，使得算法的收敛速度更快，并拿到更高的博弈场景奖励。DRL-ANCD网络平均对抗奖励为107，决策时间为422 ms。其决策时间相对于干扰策略Ⅰ与Ⅱ分别增加了4.97%与7.65%。

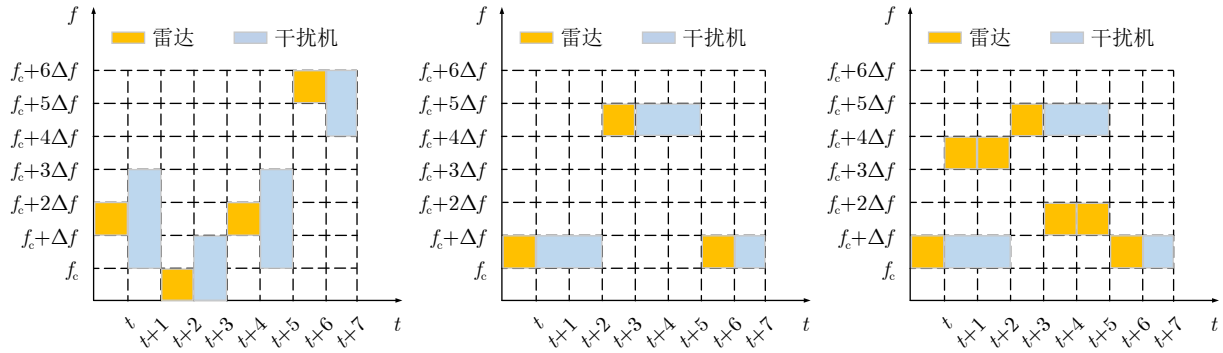
图18、图19、图20表示DRL-ANCD算法在随



(a) 噪声瞄准干扰 (a) Noise spot jamming
(b) 距离假目标欺骗干扰 (b) Distance false-target jamming
(c) 密集假目标转发干扰 (c) Dense false-target jamming

图 15 3种干扰类型下不同强化学习算法的决策性能

Fig. 15 Decision performance of different RL algorithms under three types of interference



(a) 噪声瞄准干扰 (a) Noise spot jamming
(b) 距离假目标欺骗干扰 (b) Distance false-target jamming
(c) 密集假目标转发干扰 (c) Dense false-target jamming

图 16 DRL-ANCD网络对于3种干扰类型的抗干扰行为决策

Fig. 16 Anti-jamming decisions of DRL-ANCD networks for three interference

机选取连续30个时间点内的抗干扰行为决策，其中纵轴的整数部分代表干扰机的干扰类型，小数部分表示雷达与干扰机的决策参数，蓝色折线为干扰机策略，蓝点为干扰机行为，黄点为雷达行为。纵轴的整数部分1, 2, 3表示3种干扰类型，小数部分表示行为编码，不同干扰类型的行为编码区间如表2所示。例如：1.500表示此时干扰机采用干扰类型1(噪声瞄准干扰)中的编码参数为500的干扰行为(噪声瞄准干扰的下频为3.5 GHz)；3.144表示采用干扰类型3(密集假目标转发干扰)中的编码参数为144的干扰行为(密集假目标转发干扰的观察窗为144 μs)。

如图18所示，干扰机工作在顺序干扰类型的实

表 4　单一干扰类型下3种强化学习算法抗干扰性能

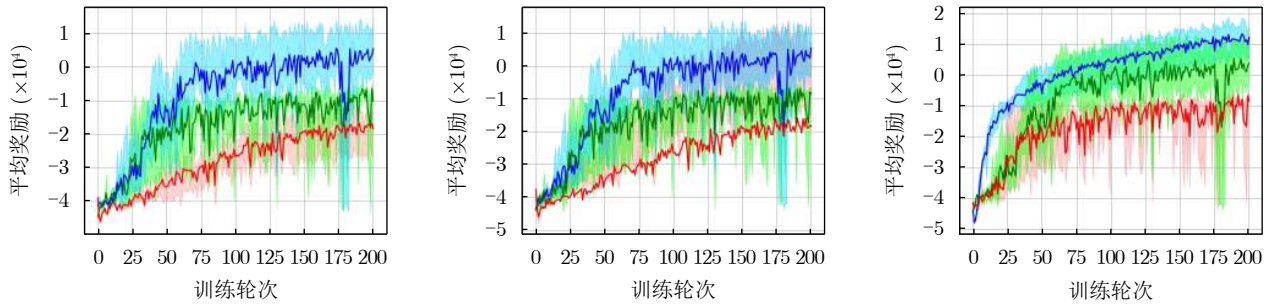
Tab. 4　Performance of 3 RL algorithms for a single jamming type

干扰类型	算法名称	平均奖励	决策时间(ms)
噪声瞄准干扰	PPO	–215	188
	TD3	–51	333
	DRL-ANCD	53	**244**
距离假目标欺骗干扰	PPO	–168	168
	TD3	–25	225
	DRL-ANCD	**94**	**203**
密集假目标转发干扰	PPO	–156	269
	TD3	–45	340
	DRL-ANCD	**24**	**289**

表 5　在线网络参数

Tab. 5　Online net parameters

网络	网络层	输入	输出	激活
策略网络	MLP1	State	256	ReLU
	MLP2	256	256	ReLU
	MLP3	256	128	ReLU
	MLP4	128	1	None
Q网络	MLP1	State+action	256	ReLU
	MLP2	Action+256	256	ReLU
	MLP3	256	128	ReLU
	MLP4	128	1	None



(a) 干扰策略 I　(b) 干扰策略II　(c) 干扰策略III

(a) Jamming strategy I　(b) Jamming strategy II　(c) Jamming strategy III

图 17 3种干扰策略下不同强化学习算法的决策性能

Fig. 17 Decision performance of different RL algorithms under three interference strategies

表 6　多干扰策略下3种强化学习算法抗干扰性能

Tab. 6　Performance of 3 RL algorithms for a multi-jamming strategies

干扰策略	算法名称	对抗奖励	决策时间(ms)
干扰策略 I	PPO-SL	–202	356
	TD3-SL	–125	443
	DRL-ANCD	**3**	**402**
干扰策略 II	PPO-SL	–221	375
	TD3-SL	–122	429
	DRL-ANCD	**14**	**392**
干扰策略III	PPO-SL	–124	386
	TD3-SL	25	463
	DRL-ANCD	**107**	**422**

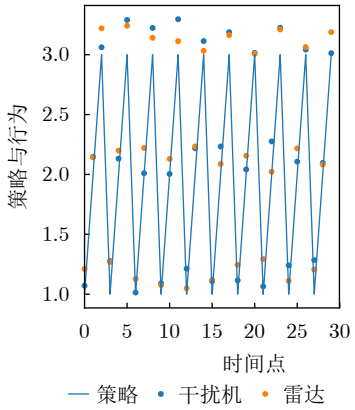


图 18 干扰策略 I 下DRL-ANCD网络的抗干扰行为

Fig. 18 Anti-jamming behaviors of DRL-ANCD networks under interference strategy I

验设定下。每一时间点选择确定干扰类型下的随机波形参数进行干扰。为了便于雷达抗干扰行为与干扰机干扰行为的比较，将雷达的抗干扰行为提前一个时间点进行可视化展示，即在时间t展示的是t时刻干扰机的行为与$t+1$时刻雷达的行为。

图18表示在干扰策略 I 下雷达与干扰机的对抗博弈行为选择，在30个时间片段中雷达均可以主动对抗干扰机的干扰行为。图19为干扰策略 II 下雷达与干扰机的行为选择；图20为干扰策略III下雷达的抗干扰效果，可以看出，在第4, 16, 24, 25个时刻，雷达抗干扰行为选择与干扰行为在时、频域内相近，抗干扰决策准确度降低，体现了算法对于随机复杂干扰场景的决策过程的不稳定性，符合随机策略下的预期效果。

6 结语

基于单一干扰场景下的雷达DRL抗干扰决策方法往往脱离了干扰复杂多样的实际对抗博弈环境，限制了其在实际电子战中的应用。

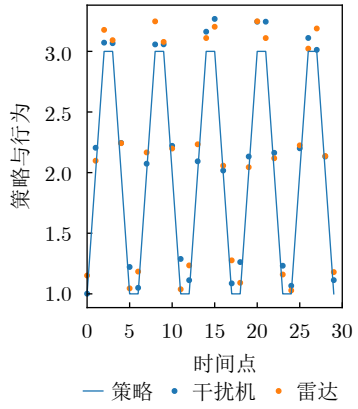


图 19 干扰策略 II 下DRL-ANCD网络的抗干扰行为

Fig. 19 Anti-jamming behaviors of DRL-ANCD networks under interference strategy II

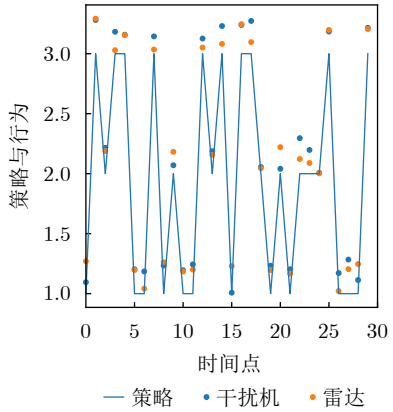


图 20 干扰策略III下DRL-ANCD网络的抗干扰行为

Fig. 20 Anti-jamming behaviors of DRL-ANCD networks under interference strategy III

为了解决该问题，本文提出了一种基于复数域深度强化学习的多干扰场景雷达抗干扰方法(DRL-ANCD)，优化了复杂干扰场景下FA雷达的抗干扰波形选择策略。

本文研究了自卫干扰机的3种典型干扰类型并构建了基于DRL架构的复杂时序干扰环境。同时，为了有效提取雷达接收波形特征信息，构建了一种基于复数域的双通道态势预测网络。基于上述工作设计实验验证了DRL-ANCD网络对于多干扰环境的有效性，证明本文所提算法具有实际意义与应用价值。

实验结果表明，在随机干扰策略下DRL-ANCD网络的决策性能最好，但决策时间略高于PPO-SL算法。对于3种干扰策略，本文提出的DRL-ANCD方法均可以达到较好的决策性能，验证了算法框架的有效性。需要注意的是，固定干扰类型顺序的干扰策略对3种DRL算法的影响较小。

本实验仿真的场景是基于环境完全可观的条件下设置的，然而在实际应用中，由于设备技术的限制与环境噪声的存在，环境态势多为非完全观测状态；此外，构建具有实际物理含义的奖励函数对算法的评估和应用具有较大意义，未来将基于这两项内容深入开展研究。

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参 考 文 献

[1] KOGON S M, HOLDER E J, and WILLIAMS D B. Mainbeam jammer suppression using multipath returns[C]. Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 1997: 279–283. doi: 10.1109/ACSSC.1997.680195.

[2] GRECO M, GINI F, and FARINA A. Radar detection and classification of jamming signals belonging to a cone class[J]. *IEEE Transactions on Signal Processing*, 2008, 56(5): 1984–1993. doi: 10.1109/TSP.2007.909326.

[3] NERI F. Introduction to Electronic Defense Systems[M]. SciTech Publishing, Raleigh, NC, 2006.

[4] 李宇环, 岳显昌, 张兰. 基于压缩感知的时域抗射频干扰方法[J]. 科学技术与工程, 2020, 20(7): 2767–2772. doi: 10.3969/j.issn.671-1815.2020.07.035.

LI Yuhuan, YUE Xianchang, and ZHANG Lan. Time-domain radio frequency interference suppression method based on compressed sensing[J]. *Science Technology and Engineering*, 2020, 20(7): 2767–2772. doi: 10.3969/j.issn.671-1815.2020.07.035.

[5] 杜思予, 刘智星, 吴耀君, 等. 基于SVM的捷变频雷达密集转发干扰智能抑制方法[J]. 雷达学报, 2023, 12(1): 173–185. doi: 10.12000/JR22065.
DU Siyu, LIU Zhixing, WU Yaojun, *et al*. Dense-repeated jamming suppression algorithm based on the support vector machine for frequency agility radar[J]. *Journal of Radars*, 2023, 12(1): 173–185. doi: 10.12000/JR22065.

[6] 董淑仙, 吴耀君, 方文, 等. 频率捷变雷达联合模糊C均值抗间歇采样干扰[J]. 雷达学报, 2022, 11(2): 289–300. doi: 10.12000/JR21205.
DONG Shuxian, WU Yaojun, FANG Wen, *et al*. Anti-interrupted sampling repeater jamming method based on frequency-agile radar joint fuzzy C-means[J]. *Journal of Radars*, 2022, 11(2): 289–300. doi: 10.12000/JR21205.

[7] 施龙飞, 任博, 马佳智, 等. 雷达极化抗干扰技术进展[J]. 现代雷达, 2016, 38(4): 1–7, 29. doi: 10.16592/j.cnki.1004-7859.2016.04.001.
SHI Longfei, REN Bo, MA Jiazhi, *et al*. Recent developments of radar anti-interference techniques with polarimetry[J]. *Modern Radar*, 2016, 38(4): 1–7, 29. doi: 10.16592/j.cnki.1004-7859.2016.04.001.

[8] 陈新竹. 多功能数字阵列雷达空域抗有源干扰方法研究[D]. [博士论文], 上海交通大学, 2022. doi: 10.27307/d.cnki.gsjtu.2020.000627.
CHEN Xinzhu. Research on spatial jamming cancellation in mutifunction digital array radar[D]. [Ph.D. dissertation], Shanghai Jiao Tong University, 2022. doi: 10.27307/d.cnki.gsjtu.2020.000627.

[9] 刘智星, 杜思予, 吴耀君, 等. 脉间-脉内捷变频雷达抗间歇采样干扰方法[J]. 雷达学报, 2022, 11(2): 301–312. doi: 10.12000/JR22001.
LIU Zhixing, DU Siyu, WU Yaojun, *et al*. Anti-interrupted sampling repeater jamming method for interpulse and intrapulse frequency-agile radar[J]. *Journal of Radars*, 2022, 11(2): 301–312. doi: 10.12000/JR22001.

[10] LECUN Y, BENGIO Y, and HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444. doi: 10.1038/nature14539.

[11] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. 计算机应用, 2016, 36(9): 2508–2515, 2565. doi: 10.11772/j.issn.1001-9081.2016.09.2508.
LI Yandong, HAO Zongbo, and LEI Hang. Survey of convolutional neural network[J]. *Journal of Computer Applications*, 2016, 36(9): 2508–2515, 2565. doi: 10.11772/j.issn.1001-9081.2016.09.2508.

[12] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1–27. doi: 10.11897/SP.J.1016.2018.00001.
LIU Quan, ZHAI Jianwei, ZHANG Zongzhang, *et al*. A survey on deep reinforcement learning[J]. *Chinese Journal of Computers*, 2018, 41(1): 1–27. doi: 10.11897/SP.J.1016.2018.00001.

[13] 刘朝阳, 穆朝絮, 孙长银. 深度强化学习算法与应用研究现状综述[J]. 智能科学与技术学报, 2020, 2(4): 312–326. doi: 10.11959/j.issn.2096-6652.202034.
LIU Zhaoyang, MU Chaoxu, and SUN Changyin. An overview on algorithms and applications of deep reinforcement learning[J]. *Chinese Journal of Intelligent Science and Technology*, 2020, 2(4): 312–326. doi: 10.11959/j.issn.2096-6652.202034.

[14] DAYAN P and DAW N D. Decision theory, reinforcement learning, and the brain[J]. *Cognitive, Affective, & Behavioral Neuroscience*, 2008, 8(4): 429–453. doi: 10.3758/CABN.8.4.429.

[15] CAROTENUTO V, DE MAIO A, ORLANDO D, *et al*. Adaptive radar detection using two sets of training data[J]. *IEEE Transactions on Signal Processing*, 2018, 66(7): 1791–1801. doi: 10.1109/TSP.2017.2778684.

[16] 汪浩, 王峰. 强化学习算法在雷达智能抗干扰中的应用[J]. 现代雷达, 2020, 42(3): 40–44, 48. doi: 10.16592/j.cnki.1004-7859.2020.03.009.
WANG Hao and WANG Feng. Application of reinforcement learning algorithms in anti-jamming of intelligent radar[J]. *Modern Radar*, 2020, 42(3): 40–44, 48. doi: 10.16592/j.cnki.1004-7859.2020.03.009.

[17] XING Qiang, ZHU Weigang, and JIA Xin. Research on method of intelligent radar confrontation based on reinforcement learning[C]. 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, China, 2017: 471–475. doi: 10.1109/CIAPP.2017.8167262.

[18] LI Kang, JIU Bo, LIU Hongwei, *et al*. Reinforcement learning based anti-jamming frequency hopping strategies design for cognitive radar[C]. 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, China, 2018: 1–5. doi: 10.1109/ICSPCC.2018.8567751.

[19] LI Kang, JIU Bo, and LIU Hongwei. Deep Q-network based anti-jamming strategy design for frequency agile radar[C]. 2019 International Radar Conference (RADAR), Toulon, France, 2019: 1–5. doi: 10.1109/RADAR41533.2019.171227.

[20] WANG Shanshan, LIU Zheng, XIE Rong, *et al*. Reinforcement learning for compressed-sensing based frequency agile radar in the presence of active interference[J]. *Remote Sensing*, 2022, 14(4): 968. doi: 10.3390/rs14040968.

[21] LI Xinzhi and DONG Shengbo. Research on efficient reinforcement learning for adaptive frequency-agility radar[J]. *Sensors*, 2021, 21(23): 7931. doi: 10.3390/s21237931.

[22] 崔国龙, 余显祥, 魏文强, 等. 认知智能雷达抗干扰技术综述与展望[J]. 雷达学报, 2022, 11(6): 974–1002. doi: 10.12000/

JR22191.

CUI Guolong, YU Xianxiang, WEI Wenqiang, *et al*. An overview of antijamming methods and future works on cognitive intelligent radar[J]. *Journal of Radars*, 2022, 11(6): 974–1002. doi: 10.12000/JR22191.

[23] WATERS W M and LINDE G J. Frequency-agile radar signal processing[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 1979, AES-15(3): 459–464. doi: 10.1109/TAES.1979.308841.

[24] 李尔康. 基于干扰认知的雷达反干扰波形设计与实现[D]. [硕士论文], 电子科技大学, 2022. doi: 10.27005/d.cnki.gdzku.2022.001534.

LI Erkang. Design and implementation of radar anti-jamming waveform based on jamming cognition[D]. [Master dissertation], University of Electronic Science and Technology of China, 2022. doi: 10.27005/d.cnki.gdzku.2022.001534.

[25] 张昭建, 谢军伟, 杨春晓, 等. 掩护脉冲信号抗转发式欺骗干扰性能分析[J]. 弹箭与制导学报, 2016, 36(4): 149–152, 156. doi: 10.15892/j.cnki.djzdxb.2016.04.039.

ZHANG Zhaojian, XIE Junwei, YANG Chunxiao, *et al*. Performance analysis of screening pulse signal confronts to deception jamming[J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2016, 36(4): 149–152, 156. doi: 10.15892/j.cnki.djzdxb.2016.04.039.

[26] 李研. 雷达抗干扰波形设计及仿真分析[D]. [硕士论文], 西安电子科技大学, 2022. doi: 10.27389/d.cnki.gxadu.2022.000930.

LI Yan. Radar anti-jamming waveform design and simulation analysis[D]. [Master dissertation], Xidian University, 2022. doi: 10.27389/d.cnki.gxadu.2022.000930.

[27] 温鹏飞. 基于雷达数据的目标航迹识别和聚类研究[D]. [硕士论文], 合肥工业大学, 2020. doi: 10.27101/d.cnki.ghfgu.2020.001861.

WANG Pengfei. Research on track recognition and clustering based on radar data[D]. [Master dissertation], Hefei University of Technology, 2020. doi: 10.27101/d.cnki.ghfgu.2020.001861.

[28] MNIH V, KAVUKCUOGLU K, SILVER D, *et al*. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533. doi: 10.1038/nature14236.

[29] SCHULMAN J, WOLSKI F, DHARIWAL P, *et al*. Proximal policy optimization algorithms[EB/OL]. https://arxiv.org/abs/1707.06347, 2017. doi: 10.48550/arXiv.1707.06347.

[30] FUJIMOTO S, HOOF H, and MEGER D. Addressing function approximation error in actor-critic methods[C]. 35th International Conference on Machine Learning, Stockholm, Sweden, 2018: 1587–1596.

作 者 简 介

解 烽，博士生，主要研究方向为雷达抗干扰技术、深度强化学习。

刘环宇，讲师，主要研究方向为强化学习、目标识别检测和无人机控制。

胡锡坤，助理研究员，主要研究方向为遥感图像处理和深度学习。

钟 平，研究员，主要研究方向为智能目标识别。

李君宝，教授，主要研究方向为机器学习算法、嵌入式智能系统、图像处理。

(责任编辑：于青)