伴随压制干扰与组网雷达功率分配的深度博弈研究

王跃东[①②] 顾以静[①②] 梁 彦*[①②] 王增福[①②] 张会霞[①②]

[①](西北工业大学自动化学院 西安 710072)

[②](信息融合技术教育部重点实验室 西安 710072)

摘要：传统的组网雷达功率分配一般在干扰模型给定的情况下进行优化，而干扰机资源优化是在雷达功率分配方式给定情况下，这样的研究缺乏博弈和交互。考虑到日益严重的雷达和干扰机相互博弈的作战场景，该文提出了伴随压制干扰下组网雷达功率分配深度博弈问题，其中智能化的目标压制干扰采用深度强化学习(DRL)训练。首先在该问题中干扰机和组网雷达被映射为两个智能体，根据干扰模型和雷达检测模型建立了压制干扰下组网雷达的目标检测模型和最大化目标检测概率优化目标函数。在组网雷达智能体方面，由近端策略优化(PPO)策略网络生成雷达功率分配向量；在干扰机智能体方面，设计了混合策略网络来同时生成波束选择动作和功率分配动作；引入领域知识构建更加有效的奖励函数，目标检测模型、等功率分配策略和贪婪干扰功率分配策略3种领域知识分别用于生成组网雷达智能体和干扰机智能体的导向奖励，从而提高智能体的学习效率和性能。最后采用交替训练方法来学习两个智能体的策略网络参数。实验结果表明；当干扰机采用基于DRL的资源分配策略时，采用基于DRL的组网雷达功率分配在目标检测概率和运行速度两种指标上明显优于基于粒子群的组网雷达功率分配和基于人工鱼群的组网雷达功率分配。

关键词：雷达资源管理；伴随压制干扰；深度强化学习；检测概率；深度博弈；领域知识辅助学习

中图分类号：TN974 **文献标识码**：A **文章编号**：2095-283X(2023)03-0642-15

DOI: 10.12000/JR23023

引用格式：王跃东, 顾以静, 梁彦, 等. 伴随压制干扰与组网雷达功率分配的深度博弈研究[J]. 雷达学报, 2023, 12(3): 642–656. doi: 10.12000/JR23023.

Reference format: WANG Yuedong, GU Yijing, LIANG Yan, *et al*. Deep game of escorting suppressive jamming and networked radar power allocation[J]. *Journal of Radars*, 2023, 12(3): 642–656. doi: 10.12000/JR23023.

Deep Game of Escorting Suppressive Jamming and Networked Radar Power Allocation

WANG Yuedong[①②] GU Yijing[①②] LIANG Yan*[①②] WANG Zengfu[①②]
ZHANG Huixia[①②]

[①](*School of Automation*, *Northwestern Polytechnical University*, *Xi'an* 710072, *China*)

[②](*Key Laboratory of Information Fusion Technology*, *Ministry of Education*, *Xi'an* 710072, *China*)

Abstract: The traditional networked radar power allocation is typically optimized with a given jamming model, while the jammer resource allocation is optimized with a given radar power allocation method; such research lack gaming and interaction. Given the rising seriousness of combat scenarios in which radars and jammers compete, this study suggests a deep game problem of networked radar power allocation under escort suppression jamming, in which intelligent target jamming is trained using Deep Reinforcement Learning (DRL). First, the jammer and the networked radar are mapped as two agents in this problem. Based on the jamming model and the radar detection model, the target detection model of the networked radar under suppressed

收稿日期：2023-02-17；改回日期：2023-03-29；网络出版：2023-04-18

*通信作者： 梁彦 liangyan@nwpu.edu.cn *Corresponding Author: LIANG Yan, liangyan@nwpu.edu.cn

基金项目：国家自然科学基金(61873205)

Foundation Item: The National Natural Science Foundation of China (61873205)

责任主编：易伟 Corresponding Editor: YI Wei

jamming and the optimized objective function for maximizing the target detection probability are established. In terms of the networked radar agent, the radar power allocation vector is generated by the Proximal Policy Optimization (PPO) policy network. In terms of the jammer agent, a hybrid policy network is designed to simultaneously create beam selection and power allocation actions. Domain knowledge is introduced to construct more effective reward functions. Three kinds of domain knowledge, namely target detection model, equal power allocation strategy, and greedy interference power allocation strategy, are employed to produce guided rewards for the networked radar agent and the jammer agent, respectively. Consequently, the learning efficiency and performance of the agent are improved. Lastly, alternating training is used to learn the policy network parameters of both agents. The experimental results show that when the jammer adopts the DRL-based resource allocation strategy, the DRL-based networked radar power allocation is significantly better than the particle swarm-based and the artificial fish swarm-based networked radar power allocation in both target detection probability and run time metrics.

Key words: Radar resource management; Escort suppression jamming; Deep Reinforcement Learning (DRL); Detection probability; Deep game; Domain knowledge assisted learning

1　引言

组网雷达(Networked Radar, NR)因具有资源共享、协同探测、空间覆盖范围大和抗干扰等优势，已经受到广大学者和机构的关注[1–8]。组网雷达资源管理在提升信息融合系统的探测、跟踪性能中扮演着至关重要的角色。然而，干扰技术向智能化方向发展[9–13]，给雷达系统资源管理带来新的挑战和任务需求。如何在时间、能量和计算等软硬件资源限制下，降低干扰带来的不利影响，是实现组网雷达探测性能提升的关键。

现有的组网雷达资源分配方法主要分为3类：基于启发式优化方法、基于博弈论方法和基于强化学习方法。基于启发式优化方法通常利用最优化方法或者群智能优化方法求解某一探测性能指标下的最优解。文献[6]以最小化多输入多输出雷达的发射功率为目标，通过推导了各个目标定位误差的克拉美罗界建立机会约束模型，并通过等效变换将机会约束问题变为非线性方程求解问题。文献[14]将目标的后验克拉美罗下界作为优化目标函数，提出一种同时优化雷达功率和带宽的改进型麻雀搜索算法对目标函数进行求解。启发式优化方法是资源优化的有效手段，然而最优化方法需要在每一个资源分配时刻沿着目标函数的负梯度方向寻找最优值，这个过程耗费大量时间且要求目标函数具有可导性。群体智能体方法在高维场景下其性能受到严重影响，导致算法搜索能力下降。

博弈论方法将组网雷达中的雷达节点视为博弈参与者，利用决策理论进行雷达资源分配。文献[15]将雷达功率分配问题建立为合作博弈模型，提出一种基于合作博弈的分布式功率分配算法，利用一种基于shapley值的求解算法得到功率分配结果。文献[16]针对组网雷达的抗截获问题，将信干噪比(Signal to Interference plus Noise Ratio, SINR)和各雷达的发射功率作为约束条件，提出了一种基于非合作博弈的迭代功率控制方法，该方法可以快速收敛至纳什均衡解。文献[17]提出基于纳什均衡的弹载雷达波形设计方法，根据最大化SINR准则分别设计了雷达和干扰的波形策略。博弈论方法无法提供资源分配的唯一解，而且需要每一时刻计算博弈双方的收益矩阵，具有较大的计算复杂度。

近年来，随着深度强化学习(Deep Reinforcement Learning, DRL)在资源分配和控制决策方面的成功应用，已经有基于DRL的雷达资源优化技术被提出。DRL具有利用智能体与环境交互来学习状态到动作最优映射策略的能力。将组网雷达作为智能体，文献[3]提出基于领域知识辅助强化学习的多输入多输出雷达功率方法，其利用领域知识来设计导向奖励，从而增加策略网络收敛性和收敛速度。文献[18]考虑目标信息感知和平台安全的情况下获得传感器目标探测分配序列，提出一种基于DRL的机载传感器任务分配方法。文献[19]考虑无线通信系统中的功率分配问题，提出一种近似SARSA[20]功率分配算法，其通过线性近似避免了SARSA功率分配策略中可能出现的“维数灾难”问题。毫无疑问DRL已经成功的运用于组网雷达资源分配问题。

然而，上述组网雷达资源分配方法都是建立在没有干扰或者干扰模型已知的基础上，缺少干扰机和雷达的博弈与交互。随着干扰技术的发展，干扰机在干扰时间、干扰功率控制方面具有更强的对抗能力。在干扰机资源调度方面，文献[21]提出一种鲁棒的干扰波束选择和功率调度策略来协同压制NR系统，其中多个目标的后验克拉美罗下界之和

用来评估干扰性能。文献[22]考虑在干扰资源有限的情况下的干扰波束和功率的分配问题，建立了一种基于改进遗传算法的干扰资源分配模型，推导了压制干扰下NR系统的探测概率，并将其作为评价干扰性能指标，提出一种基于粒子群算法的两步求解方法。文献[23]采用模糊综合评价方法对影响辐射源威胁水平和干扰效率的综合因素进行量化，提出了一种基于改进萤火虫算法的干扰资源分配方法。文献[11]提出一种基于双Q学习算法的干扰资源分配策略。文献[24]提出基于DRL的智能频谱干扰方法，其对不同种类的跳频通信信号具有很好的干扰效果。

综上所述，DRL已经被用于组网雷达或者干扰机的资源分配任务，但是同时考虑伴随压制干扰与组网雷达功率分配的深度博弈仍然是一个开放性问题。由于以下因素，应用DRL解决上述问题颇具挑战：组网雷达功率分配动作属于连续动作，因此智能体探索空间很大，导致策略难以收敛；组网雷达和干扰机博弈过程中环境动态性增强，进一步增加智能体的策略学习难度。

考虑DRL在处理动态环境下的资源分配的优势，本文首先将干扰机和组网雷达映射为智能体，根据雷达目标检测模型和干扰模型建立了压制干扰下组网雷达目标检测模型和检测概率最大化优化目标函数。然后，采用PPO策略网络生成组网雷达功率分配动作；引入目标检测模型和等功率分配策略两类领域知识构建导向奖励以辅助智能体探索。其次，设计混合策略网络生成干扰机智能体的波束选择和功率分配动作；同样引入领域知识(贪婪干扰资源分配策略)生成干扰机智能体的导向奖励。最后，通过交替训练更新两种智能体的策略网络参数。实验结果表明：当干扰机采用基于DRL的资源分配策略时，采用基于DRL的组网雷达功率分配在目标检测概率和运行速度两种指标上明显优于基于粒子群的组网雷达功率分配和基于人工鱼群的组网雷达功率分配。

2 问题描述

本文目的是在智能化压制干扰下通过调度组网雷达的功率资源以提升雷达的探测性能。为此，首先提出干扰机掩护目标穿越组网雷达探测区域的任务想定。其次，根据干扰模型和雷达检测模型建立压制干扰下的组网雷达目标检测模型，进而提出最大化目标检测概率优化目标函数。

2.1 任务想定

图1给出干扰机掩护目标穿越组网雷达防区的资源分配任务的示例。由一架干扰机伴随一架飞机(目标)试图穿越由N部雷达组成的组网雷达探测区域。在此过程中，干扰机生成电磁噪声干扰雷达的探测信号来掩护目标，这种噪声干扰被称为压制式干扰。在该任务想定中，干扰方希望尽可能地使组网雷达探测不到目标，而我方组网雷达则期望最大化目标的检测性能。

如图2所示，上述干扰机和组网雷达的博弈过程被进一步细化为干扰机智能体和组网雷达智能体资源分配策略的博弈。

(1) 假设干扰机在k时刻能够发射$L < N$个干扰波束。干扰机智能体需要完成以下任务，即：在k时刻选择干扰哪几部雷达？被选中的雷达的干扰功率分配多少才能使组网雷达探测目标的概率最小？

(2) 假设组网雷达每个节点都工作单波束模式，在各个探测时刻，所有雷达节点均发射波束，即每个探测时刻有N个雷达波束探测目标。组网雷达智能体需要怎么为每个雷达-目标分配合理的发射功率使得目标检测概率最大化？

与无干扰情况下的组网雷达功率分配不同，在干扰机干扰下，组网雷达需要考虑干扰机对资源分配和目标检测的影响，因此需要引入干扰机的干扰特性和模型来优化组网雷达功率分配。同时，由于干扰机的干扰波束和功率分配具有不确定性和动态性，因此需要动态地调整组网雷达功率分配策略，以实现最优的干扰抑制和探测性能的平衡。在组网雷达功率分配策略求解方面，传统的方法通常采用全局优化算法对问题求解，如遗传算法、粒子群算法等，这些方法都需要较高的计算成本，难以在大规模优化问题中保证优化的时效性和可靠性，因此需要探测探索具有大规模优化空间搜索能力的DRL分配策略。

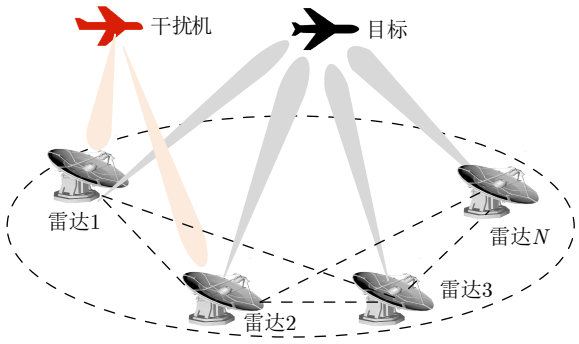


图 1 压制干扰机掩护目标穿越组网雷达防区的示例

Fig. 1 An example of a suppression jammer protecting a target through the networked radar defense area

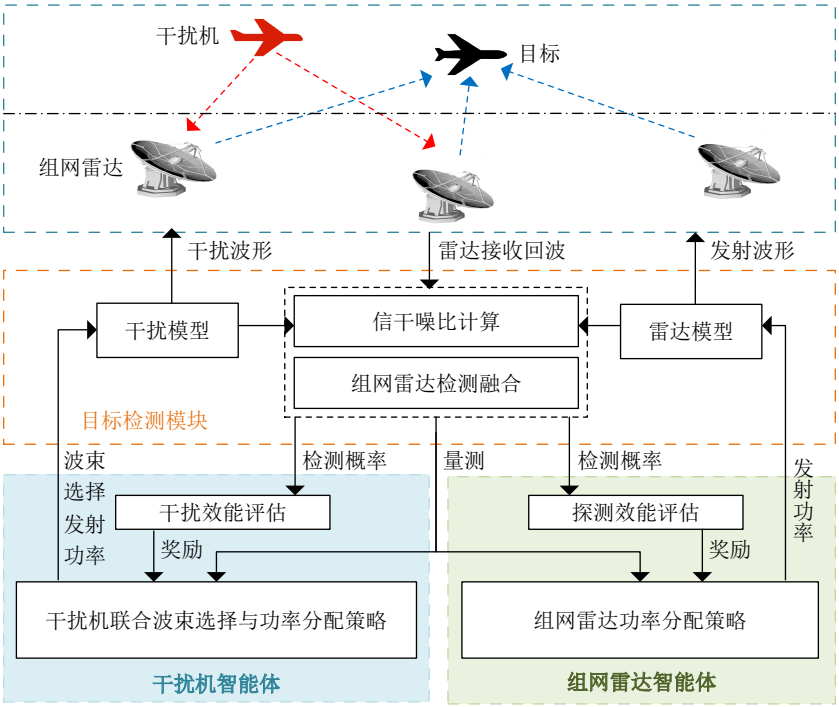


图 2 干扰机智能体和组网雷达智能体的博弈流程图

Fig. 2 The game closed-loop process of the jammer agent and the networked radar agent

2.2 目标检测模块

2.2.1 干扰模型

压制干扰是一种噪声干扰手段，干扰机发射强干扰信号进入雷达接收机，进而形成对雷达的回波的掩盖和压制，使雷达对目标的检测性能下降。本文采用噪声调频干扰信号进行干扰信号建模，假设干扰机向敌方雷达n施加噪声调频干扰信号[10,21,23]，即

$$j_k^n(t)=\sqrt{p_k^n}\exp\left\{-\mathrm{j}2\pi\left[\omega t+2\pi K_{\mathrm{FM}}\int_0^t u(t')\,\mathrm{d}t'+\varphi\right]\right\} \tag{1}$$

其中，$\sqrt{p_k^n}$为噪声调频信号的幅度；ω和K_{FM}分别为中心频率和调频斜率；$u(t')$为零均值平稳随机分布的调制噪声；φ为干扰信号的初始相位，且在$[0,2\pi)$内满足均匀分布。

2.2.2 压制干扰下单雷达目标检测模型

在无干扰情况下，目标的检测概率与雷达接收天线处的信噪比(Signal Noise Ratio, SNR)相关。SNR的大小由目标回波功率y_{signal}和接收机输入噪声P_{n}共同决定[10,21,23]。

雷达n接收到的目标回波信号功率y_{signal}可表示为

$$y_{\mathrm{signal}}=\frac{P_{\mathrm{r},k}G_{\mathrm{r}}^2\lambda^2\sigma}{(4\pi)^3(R_{\mathrm{r},k}^n)^4} \tag{2}$$

其中，$P_{\mathrm{r},k}$为雷达的发射功率，G_{r}为雷达天线主瓣方向上的增益，σ为目标有效反射面积，λ为雷达的工作波长，$R_{\mathrm{r},k}^n$为k时刻目标与探测雷达n之间的距离。

雷达接收机的内部噪声P_{n}可表示为

$$P_{\mathrm{n}}=kT_0B_{\mathrm{n}}F_{\mathrm{n}} \tag{3}$$

其中，$k=1.38\times10^{-23}$ J/K为玻尔兹曼常数，B_{n}为接收机带宽，T_0为接收机内部有效热噪声温度，F_{n}为接收机噪声系数。

因此，雷达n接收端的SNR表示为

$$\mathrm{SNR}_n=\frac{y_{\mathrm{signal}}}{P_{\mathrm{n}}} \tag{4}$$

在噪声压制干扰下，雷达接收端的信号由目标回波功率y_{signal}、内部噪声P_{n}以及干扰信号功率y_{interf}3部分组成。根据干扰方程[10,21,23]，雷达n接收到来自干扰机发射的干扰信号功率为

$$y_{\mathrm{interf}}=c_k^nP_{\mathrm{j},k}^l\frac{G_{\mathrm{j}}G_{\mathrm{j}}'(\theta_k)\lambda^2\gamma_{\mathrm{j}}}{(4\pi)^2(R_{\mathrm{j},k}^n)^2} \tag{5}$$

其中，$c_k^n\in\{0,1\}$为二元变量，用来指示k时刻雷达n是否受到干扰，$c_k^n=1$表示雷达n受到干扰，$c_k^n=0$表示雷达n未受干扰；$P_{\mathrm{j},k}^l$为干扰机波束的发射功率；$R_{\mathrm{j},k}^n$为干扰机和雷达n之间的距离；γ_{j}为两者天线之间的极化损失；λ为干扰机的工作波长，其与雷达工作波长相等；G_{j}为干扰机天线主瓣方向

上的增益；$G'_{\mathrm{j}}(\theta_k)$为雷达在干扰机主瓣方向上的天线增益，其与干扰机波束主瓣与雷达天线主瓣之间的夹角θ_k有关：

$$G'_{\mathrm{j}}(\theta_k)=\begin{cases}G_{\mathrm{j}}, & 0\le|\theta_k|\le\dfrac{\theta_{0.5}}{2}\\ \beta\left(\dfrac{\theta_{0.5}}{\theta_k}\right)^2 G_{\mathrm{j}}, & \dfrac{\theta_{0.5}}{2}<|\theta_k|\le 90^\circ\\ \beta\left(\dfrac{\theta_{0.5}}{90}\right)^2 G_{\mathrm{j}}, & 90^\circ<|\theta_k|\le 180^\circ\end{cases}\tag{6}$$

其中，$\theta_{0.5}$为雷达天线波瓣宽度；β为常数。

如图3所示，θ_k取决于干扰机、目标机和雷达三者之间的相对位置关系。根据干扰信号进入雷达的角度，压制干扰划分为伴随干扰和支援干扰两种类型。当干扰信号从雷达天线主瓣进入接收机时为伴随干扰；当$\theta_k>\theta_{0.5}/2$时干扰信号主要从雷达天线旁瓣进入，干扰方式为支援干扰。

压制干扰下，雷达n接收机接收到关于目标的SINR为[10,23]

$$\mathrm{SINR}_n=\frac{y_{\mathrm{signal}}}{y_{\mathrm{interf}}+P_{\mathrm{n}}}\tag{7}$$

假设目标的起伏特性为Swerling I型，雷达累积脉冲数为1，则雷达n对目标的检测概率可表示为[10,23,25]

$$P_{\mathrm{d},k}^n=\mathrm{e}^{-V_{\mathrm{T}}/(1+\mathrm{SINR}_n)}\tag{8}$$

其中，V_{T}为检测门限。将式(7)代入式(8)可得

$$P_{\mathrm{d},k}^n\propto\frac{y_{\mathrm{signal}}}{c_k^n P_{\mathrm{j},k}^l\dfrac{G_{\mathrm{j}}G'_{\mathrm{j}}(\theta_k)\lambda^2\gamma_{\mathrm{j}}}{(4\pi)^2(R_{\mathrm{j},k}^n)^2}+P_{\mathrm{n}}}\tag{9}$$

由式(9)可以发现雷达对目标的检测概率与干扰资源分配变量以及干扰机、目标机和雷达间的空间位置有关。

2.2.3 组网雷达检测融合

组网雷达采用K-N融合规则来实现信息融合[10,23,26]。假设雷达n的局部判决为$d_n\in\{0,1\}$，其中$d_n=1$或0表示发现目标与否。融合中心根据这些局部判据产生全局判决向量$\boldsymbol{D}=[d_1\,d_2\cdots d_N]$，共有$2^N$种可能。定义全局判决规则为$R(\boldsymbol{D})$，当组网雷达中发现目标的雷达数超过检测门限$K(1\le K\le N)$时，判定发现目标，否则判定为未发现目标，即

$$R(\boldsymbol{D})=\begin{cases}1, & \displaystyle\sum_{n=1}^{N}d_n\ge K\\ 0, & \displaystyle\sum_{n=1}^{N}d_n<K\end{cases}\tag{10}$$

根据秩K融合准则可以得到k时刻组网雷达对目标的检测概率$P_{\mathrm{d},k}$为

$$P_{\mathrm{d},k}=\sum_{\boldsymbol{D}}\left[R(\boldsymbol{D})\prod_{d_n\in\boldsymbol{S}_0}\left(1-P_{\mathrm{d},k}^n\right)\prod_{d_n\in\boldsymbol{S}_1}P_{\mathrm{d},k}^n\right]\tag{11}$$

其中，$P_{\mathrm{d},k}^n$为雷达节点n对目标的检测概率，$\boldsymbol{S}_0$是$\boldsymbol{D}_i(i=1,2,\cdots,2^N)$中判决为未发现目标的判决集合，$\boldsymbol{S}_1$是$\boldsymbol{D}_i$中判决为发现目标的判决集合。

2.3 优化函数设计

组网雷达探测任务的要求是在统计意义下探测到目标的次数越多越好。该指标可进一步量化为组网雷达对目标的检测概率$P_{\mathrm{d},k}$，其值越大说明目标越容易被发现。根据任务需求，本文的优化目标函数为

$$\max_{\boldsymbol{P}_{\mathrm{r},k}}P_{\mathrm{d},k}(\boldsymbol{P}_{\mathrm{r},k}|\boldsymbol{z}_k,\boldsymbol{P}_{\mathrm{j},k})$$
$$\text{s.t.}\begin{cases}\sum_{n=1}^{N}P_{\mathrm{r},k}^n\le P_{\mathrm{r}}^{\mathrm{total}}\\ P_{\mathrm{r}}^{\min}\le P_{\mathrm{r},k}^n\le P_{\mathrm{r}}^{\max}\\ \sum_{n=1}^{N}z_k^n=L\\ \sum_{l=1}^{L}P_{\mathrm{j},k}^l\le P_{\mathrm{j}}^{\mathrm{total}}\\ P_{\mathrm{j}}^{\min}\le P_{\mathrm{j},k}^l\le P_{\mathrm{j}}^{\max}\end{cases}\tag{12}$$

其中，$\boldsymbol{P}_{\mathrm{r},k}=[P_{\mathrm{r},k}^1\,P_{\mathrm{r},k}^2\cdots P_{\mathrm{r},k}^N]$表示组网雷达发射功率分配向量；$\boldsymbol{z}_k=[z_k^1\,z_k^2\cdots z_k^N]$表示干扰机的波束选择向量；$\boldsymbol{P}_{\mathrm{j},k}=[P_{\mathrm{j},k}^1\,P_{\mathrm{j},k}^2\cdots P_{\mathrm{j},k}^L]$表示干扰功率分配向量，$\boldsymbol{P}_{\mathrm{j},k}$中的元素与$\boldsymbol{z}_k$中为1的元素依次对应；$P_{\mathrm{r}}^{\mathrm{total}}$为$k$时刻组网雷达总的发射功率，$P_{\mathrm{r}}^{\min}$和$P_{\mathrm{r}}^{\max}$分别表示单雷达的最小和最大发射功率；$P_{\mathrm{j}}^{\mathrm{total}}$为$k$时刻干扰机总的干扰功率，$P_{\mathrm{j}}^{\min}$和$P_{\mathrm{j}}^{\max}$分别表示每个干扰波束的最小和最大功率。

传统的组网雷达功率方法一般先通过干扰性能评估建立优化目标函数，然后利用启发式搜索算法进行策略求解。这些方法通常是在假定探测环境没有干扰或者干扰模型给定的情况下进行方案设计，缺少干扰机和组网雷达相互博弈，不符合实际作战

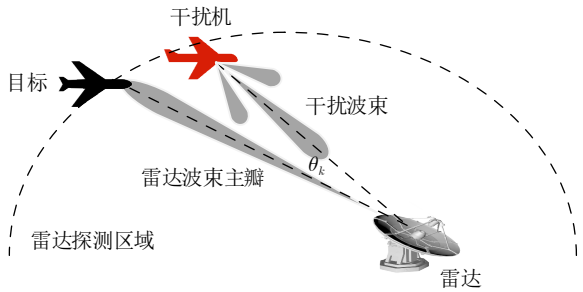


图 3 干扰机、雷达和目标的相对空间位置

Fig. 3 The relative spatial position of the jammer, radar and target

需求。同时启发式搜索方法存在计算成本高、搜索速度慢的缺点，难以保证优化的有效性。与这些方法不同，本文考虑到体系协同作战下干扰机与组网雷达的博弈，提出基于DRL的干扰机波束和功率分配条件下的组网雷达功率分配问题。在策略求解方面，结合了人工智能方法，干扰机和组网雷达被映射为智能体，利用DRL的交互试错学习机制生成从环境状态到组网雷达功率分配向量的映射。由于采用离线训练的方式进行策略探索，因此DRL相较于一般方法具有更快的在线运行速度。

3　组网雷达智能体的MDP建模

本节首先将组网雷达智能体功率分配模型化为马尔可夫决策过程(Markov Decision Process, MDP)[27]。一个MDP通常采用元组(S, A, P, r)表示，其中S为环境状态，它是智能体的环境观测；A为动作，它是执行器的输出；P为状态的转移概率。值得注意的是，在无模型强化学习中P是未知的。r是由环境产生的单步奖励。

图4显示组网雷达策略网络同干扰机与雷达博弈环境的交互过程。首先组网雷达智能体的策略网络根据环境状态生成一个功率分配动作，并将该动作传递给组网雷达。然后雷达执行探测动作获取目标量测，并提取下一时刻的环境状态。智能体的状态、动作和奖励被存入经验池用于组网雷达智能体的策略网络参数更新。

(1) 组网雷达智能体的状态

组网雷达能够获取的环境信息包括k时刻雷达n与目标的距离$R_{\mathrm{r},k}^{n}$和雷达被干扰指示c_k^n。通过对组网雷达的观测进行预处理生成策略网络的输入状态。预处理过程包括标准化和连接操作。标准化是为了将不同量纲的雷达观测统一到$[0,1]$。定义距离标准化函数为

$$\bar{R}_{\mathrm{r},k}^{n}=\frac{R_{\mathrm{r},k}^{n}-R^{\min}}{R^{\max}-R^{\min}} \tag{13}$$

其中，$R^{\min}$, $R^{\max}$分别表示雷达的最小和最大观测距离。

连接操作是在数据标准化后将不同类型的雷达观测组合成策略网络的输入状态。首先将任意雷达的观测按照被干扰指示和雷达与目标的距离组合，即$\boldsymbol{o}_n^{\mathrm{Radar}}=[c_k^n\ \bar{R}_{\mathrm{r},k}^{n}]$。然后将所有雷达的观测按照雷达编号组合，即

$$\boldsymbol{S}_k^{\mathrm{Radar}}=\left[\boldsymbol{o}_1^{\mathrm{Radar}}\ \boldsymbol{o}_2^{\mathrm{Radar}}\cdots\boldsymbol{o}_N^{\mathrm{Radar}}\right] \tag{14}$$

其中，$\boldsymbol{S}_k^{\mathrm{Radar}}$表示组网雷达智能体的输入状态。

(2) 组网雷达智能体的动作

组网雷达智能体的动作定义为$\boldsymbol{a}_{\mathrm{r},k}=[P_{\mathrm{r},k}^{n}]_{N\times 1}$ $(n=1,2,\cdots,N)$，其中$P_{\mathrm{r},k}^{n}$表示k时刻雷达节点n的发射功率。

(3) 知识辅助的组网雷达智能体奖励函数设计

强化学习模拟人类奖惩机制，利用智能体与环境交互试错改进策略，本质上是选择奖励大的动作。然而，强化学习的试错过程仍然是随机探索，对于压制干扰下组网雷达功率分配任务，干扰机和组网雷达的博弈与目标运动使得探测环境的动态性显著增加，进而导致智能体策略学习困难。为了辅助智能体的探索，有必要引入人的认知模型和知识，提出知识辅助下的奖励设计。通过专家知识和模型知识设计导向奖励，以引导智能体向人类认知方向探索，最终生成符合任务想定的资源分配策略。如图5所示，本文给出知识辅助的组网雷达智能体奖励函数设计框图。

根据压制干扰下组网雷达目标探测模型(模型知识)可知，雷达发现目标的概率和雷达与目标的距离以及雷达的发射功率相关。对于使用相同发射功率的雷达，目标距离雷达越近，雷达接收端获得的SINR越大，意味着发现目标的概率越大。因此定义评价函数为

$$J_{\mathrm{r}}\left(\boldsymbol{S}_k^{\mathrm{Radar}},\boldsymbol{a}_{\mathrm{r},k}\right)=\sum_{n=1}^{N}\left(1-\bar{R}_{\mathrm{r},k}^{n}\right)P_{\mathrm{r},k}^{n} \tag{15}$$

其中，$J_{\mathrm{r}}\left(\boldsymbol{S}_k^{\mathrm{Radar}},\boldsymbol{a}_{\mathrm{r},k}\right)$表示$k$时刻组网雷达的探测收益。

在组网雷达功率分配任务中等功率分配策略(专家知识)被用来作为判断智能体分配动作好坏的基准策略。如果k时刻智能体的探测收益大于基准

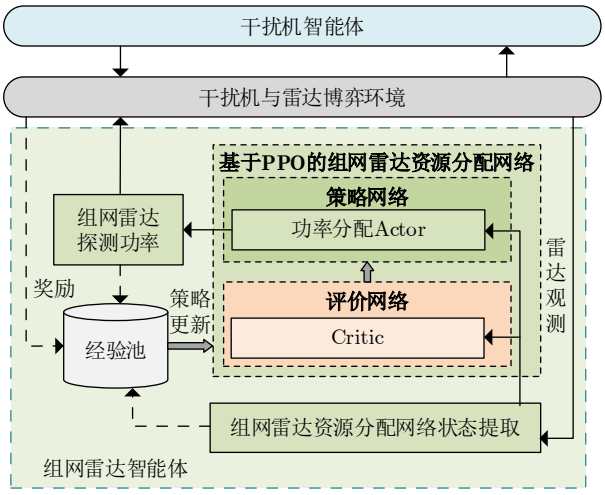


图 4 组网雷达智能体与环境交互图

Fig. 4 The networked radar agent and environment interaction diagram

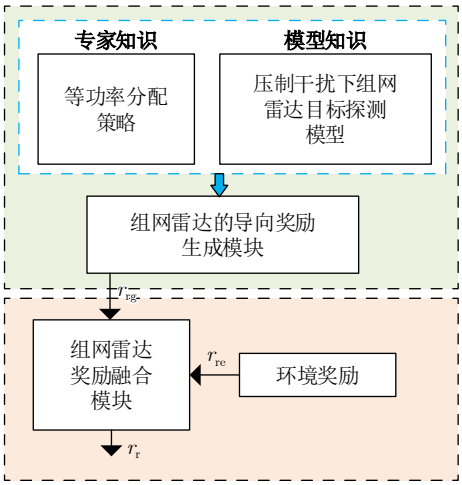


图 5 知识辅助的组网雷达智能体奖励模块

Fig. 5 The knowledge-assisted reward module for the networked radar agent

策略的探测收益，那么给予智能体正的导向奖励，否则做出适当的惩罚。具体的规则定义如下：

$$r_{\mathrm{rg},k}=\begin{cases}b_1, & \text{如果} J_{\mathrm{r}}\left(\boldsymbol{S}_k^{\mathrm{Radar}},\boldsymbol{a}_{\mathrm{r},k}\right)>J_{\mathrm{r}}\left(\boldsymbol{S}_k^{\mathrm{Radar}},\bar{\boldsymbol{a}}_{\mathrm{r},k}\right)\\ -b_2, & \text{其他}\end{cases} \tag{16}$$

其中，$r_{\mathrm{rg},k}$表示组网雷达智能体的导向奖励；b_1和b_2是正实数；$\bar{\boldsymbol{a}}_{\mathrm{r},k}$为等功率分配动作。

组网雷达智能体的环境奖励是根据优化目标给出的。组网雷达期望发现目标的概率越大越好，即目标的检测概率越接近1给予的奖励越大。因此，组网雷达的环境奖励定义为

$$r_{\mathrm{re},k}=-\left(1-P_{\mathrm{d},k}\right) \tag{17}$$

组网雷达智能体的导向奖励和环境奖励共同用于改进组网雷达智能体的策略。考虑到随着训练次数的增加智能体的策略将超越基准策略，此时导向奖励起到促进策略探索作用，相反会影响智能体向最优策略探索。因此，本文设计导向奖励衰减奖励融合模块，由知识产生的导向奖励随着训练幕数的增加逐渐减小，即

$$r_{\mathrm{r},k}=\left(1-\beta^t\right)r_{\mathrm{re},k}+\beta^t r_{\mathrm{rg},k} \tag{18}$$

其中，$r_{\mathrm{r},k}$为融合后组网雷达智能体的奖励；β为衰减因子；t为训练幕数。

注意，上述设计过程中使用等功率分配策略作为专家知识来生成导向奖励，事实上可以引入更加先进的分配策略辅助智能体探索。

(4) 组网雷达智能体的策略网络

如图6，组网雷达智能体的策略网络采用演员-评论家(Actor-Critic, AC)框架，由一个Actor和一个Critic组成，其中Actor策略网络用于产生功率分配动作，Critic策略网络用来评估动作的好坏。Actor策略网络采用3层全连接神经网络(Neural Network, NN)搭建，中间层采用ReLU激活函数激活，输出层采用Tanh激活函数激活。Critic策略网络同样采用3层全连接NN搭建并使用Tanh激活。采用PPO算法进行策略学习[28]。

4 干扰机智能体的MDP建模

图7显示了干扰机策略网络同干扰机与雷达博弈环境的交互过程。首先由基于混合强化学习的干扰资源分配策略网络生成干扰机智能体的波束选择动作和波束功率分配动作。然后，干扰机执行该动作对被选中雷达发射干扰波束。干扰机获取环境观

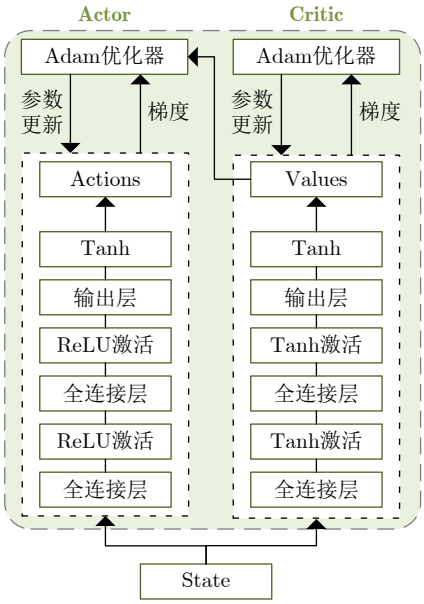


图 6 组网雷达智能体的策略网络

Fig. 6 The policy network of the networked radar agent

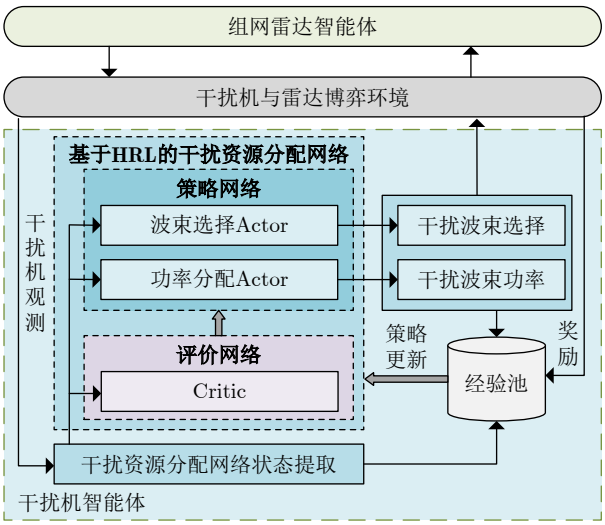


图 7 干扰机智能体与环境交互图

Fig. 7 The jammer agent and environment interaction diagram

察并提取下一时刻状态。干扰机智能体的状态、动作和奖励被存入经验池，这些样本用于混合策略网络的参数更新。

(1) 干扰机智能体的状态

影响干扰资源分配的主要因素有干扰机与雷达的距离$R_{\mathrm{j},k}^n$和干扰机、雷达与目标的夹角θ_k^n。因此选择$R_{\mathrm{j},k}^n$和θ_k^n作为干扰机智能体的原始观测。该原始观测经过标准化和连接操作生成干扰机智能体的输入状态，即

$$\boldsymbol{S}_k^{\mathrm{Jammer}} = \left[\boldsymbol{o}_1^{\mathrm{Jammer}}\, \boldsymbol{o}_2^{\mathrm{Jammer}} \cdots \boldsymbol{o}_N^{\mathrm{Jammer}}\right] \tag{19}$$

其中，$\boldsymbol{o}_n^{\mathrm{Jammer}} = \left[\bar{R}_{\mathrm{j},k}^n\, \bar{\theta}_k^n\right]$，$\bar{R}_{\mathrm{j},k}^n$和$\bar{\theta}_k^n$分别表示归一化后的距离和夹角。

(2) 干扰机智能体的动作

二元离散变量$z_k^n \in \{0,1\}$用来指示k时刻干扰机是否对雷达n施加干扰，那么干扰机的干扰波束分配结果可以采用向量$\boldsymbol{z}_k = [z_k^1\, z_k^2 \cdots z_k^N]$表示。例如，对于$N=5$的组网雷达，$\boldsymbol{z}_k = [0\,1\,0\,1\,1]$表示$k$时刻干扰机选择干扰雷达2、雷达4和雷达5。根据约束条件$\sum_{n=1}^{N} z_k^n = L$，干扰机的波束选择动作空间大小为$N!/(L!\,(N-L)!)$。将干扰机波束选择任务映射为一个$N!/(L!\,(N-L)!)$分类问题，定义干扰机智能体的波束选择动作$a_{\mathrm{j},k}^1 \in \{0,1,\cdots,N!/(L!\,(N-L)!)-1\}$。

干扰机智能体的波束功率分配动作定义为$\boldsymbol{a}_{\mathrm{j},k}^2 = [P_{\mathrm{j},k}^l]_{L\times 1}(l=1,2,\cdots,L)$，其中$P_{\mathrm{j},k}^l$表示$k$时刻干扰机发射的第$l$个波束的功率。

(3) 干扰机智能体的奖励

干扰机的波束和功率联合分配具有由离散动作和连续动作组成的混合动作空间，这比其他的资源分配任务更加复杂。其中，混合动作空间增加了智能体的探索难度，更少的最优动作被遍历，这意味着最优动作下的环境奖励是稀疏的，这导致DRL的策略难以改进。因此引入模型知识和专家知识设计导向奖励辅助智能体探索，如图8所示。

将贪婪干扰资源分配策略视作评价干扰机智能体的资源分配动作的基准。当采用干扰机智能体的波束选择和功率分配动作下组网雷达发现目标的概率小于使用基准干扰资源分配策略时，给予正的导向奖励，否则惩罚，即

$$r_{\mathrm{jg},k} = \begin{cases} b_1, & \text{如果} P_{\mathrm{d},k} < \bar{P}_{\mathrm{d},k} \\ -b_2, & \text{其他} \end{cases} \tag{20}$$

其中，$r_{\mathrm{jg},k}$为干扰机智能体的导向奖励；$\bar{P}_{\mathrm{d},k}$表示基准干扰资源分配策略下组网雷达发现目标的概率。

干扰机的优化目标与组网雷达的优化目标相反，目标的发现概率越小越好。因此干扰机智能体的环境奖励表示为

$$r_{\mathrm{je},k} = -P_{\mathrm{d},k} \tag{21}$$

与组网雷达的导向奖励和环境奖励融合的方法相同，干扰机智能体的奖励融合模块定义为

$$r_{\mathrm{j},k} = \left(1-\beta^t\right) r_{\mathrm{je},k} + \beta^t r_{\mathrm{jg},k} \tag{22}$$

其中，$r_{\mathrm{j},k}$表示融合后干扰机智能体的奖励。

(4) 干扰机智能体的混合策略网络

干扰机智能体需要同时产生两种不同质的混合动作，即离散的干扰波束选择动作和连续的波束功率分配动作。因此本文设计一种混合策略网络，如图9所示，用来表示两种分配动作，其中利用具有分类分布输出的离散Actor来表示干扰波束选择动作，采用具有高斯分布的连续Actor来表示干扰波束功率分配动作。

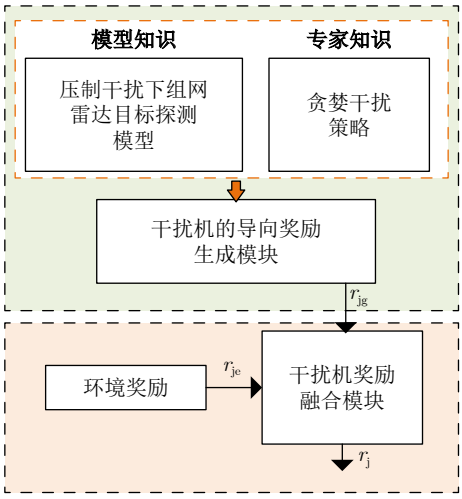


图 8 知识辅助的干扰机智能体奖励函数模块

Fig. 8 The knowledge-assisted reward function module for the jammer agent

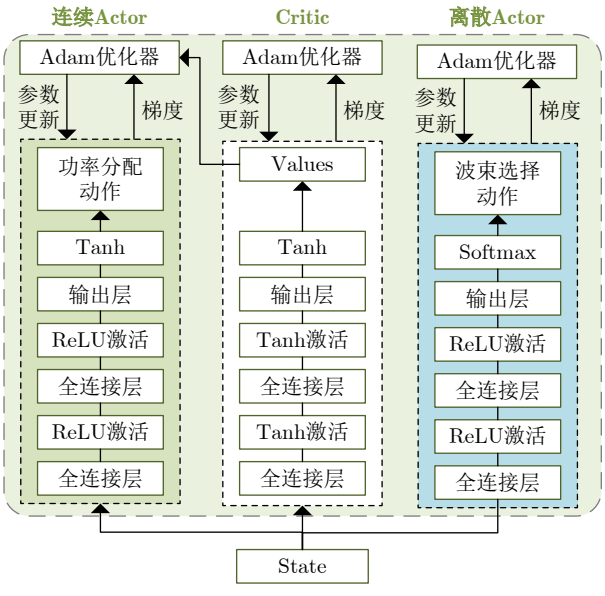


图 9 干扰机智能体的混合策略网络

Fig. 9 The hybrid policy network of the jammer agent

5 基于交替训练的干扰机与组网雷达资源分配策略学习

如图2所示，由组网雷达和干扰机组成的资源对抗优化问题由于以下困难使得资源分配策略很难收敛：(1)组网雷达和干扰的功率分配都是连续变量，因此策略学习的状态-动作空间维度很大，难以收敛；(2)干扰机波束分配为非凸优化并与功率分配耦合，这进一步增加策略搜索空间；(3)组网雷达和干扰机博弈过程中资源分配环境动态性增加。

为此本文提出基于交替训练的多步求解方法，设置最大迭代次数为M。具体步骤为：

步骤1 固定组网雷达的功率分配策略，训练干扰机的联合波束与功率分配策略。

当迭代次数为$m=1$时组网雷达智能体使用等功率分配策略Π^{m_0}，即$\boldsymbol{a}_{\mathrm{r},k}=[P_{\mathrm{r},k}^1 P_{\mathrm{r},k}^2\cdots P_{\mathrm{r},k}^N]_{\Pi^{m_0}}$。当$m>1$时组网雷达智能体使用基于PPO的组网雷达功率分配策略Π_{RL}^{m-1}，即$\boldsymbol{a}_{\mathrm{r},k}=[P_{\mathrm{r},k}^1 P_{\mathrm{r},k}^2\cdots P_{\mathrm{r},k}^N]_{\Pi_{\mathrm{RL}}^{m-1}}$。干扰机智能体感知环境状态$\boldsymbol{S}_k^{\mathrm{Jammer}}$并由图9所示的资源分配网络产生波束选择动作和功率分配动作$\boldsymbol{a}_{\mathrm{j},k}=[a_{\mathrm{j},k}^1 a_{\mathrm{j},k}^2]_{\Pi_{\mathrm{RL}}^{m-1}}$。在干扰机执行动作$\boldsymbol{a}_{\mathrm{j},k}$后，转移到下一个状态$\boldsymbol{S}_{k+1}^{\mathrm{Jammer}}$，由环境返回奖励$r_{\mathrm{je},k}$，并由干扰机导向奖励生成模块返回导向奖励$r_{\mathrm{jg},k}$。两种奖励由干扰机奖励融合模块生成干扰机智能体的奖励$r_{\mathrm{j},k}$。以上数据被组合为一个元组$(\boldsymbol{S}_k^{\mathrm{Jammer}},\boldsymbol{a}_{\mathrm{j},k},r_{\mathrm{j},k},\boldsymbol{S}_{k+1}^{\mathrm{Jammer}})$然后存入干扰机经验池$D_{\mathrm{j}}$。每隔一定的训练幕，从$D_{\mathrm{j}}$中采样小批次的样本训练基于混合强化学习(Hybrid Reinforcement Learning, HRL)的干扰资源分配网络。

步骤2 固定干扰机的资源分配策略，训练组网雷达的功率分配策略。

在步骤2中，干扰机使用基于HRL的干扰资源分配策略Π_{RL}^{m-1}，即$\boldsymbol{a}_{\mathrm{j},k}=[a_{\mathrm{j},k}^1 a_{\mathrm{j},k}^2]_{\Pi_{\mathrm{RL}}^{m-1}}$。组网雷达智能体感知环境状态$\boldsymbol{S}_k^{\mathrm{Radar}}$，并由图6所示的资源分配网络产生组网雷达功率分配动作$\boldsymbol{a}_{\mathrm{r},k}=[P_{\mathrm{r},k}^1 P_{\mathrm{r},k}^2\cdots P_{\mathrm{r},k}^N]_{\Pi_{\mathrm{RL}}^{m-1}}$。在雷达执行动作$\boldsymbol{a}_{\mathrm{r},k}$后，转移到下一个状态$\boldsymbol{S}_{k+1}^{\mathrm{Radar}}$，并由环境返回奖励$r_{\mathrm{re},k}$，由组网雷达导向奖励生成模块返回导向奖励$r_{\mathrm{rg},k}$。两种奖励由组网雷达奖励融合模块生成组网雷达智能体的奖励$r_{\mathrm{r},k}$。将$(\boldsymbol{S}_k^{\mathrm{Radar}},\boldsymbol{a}_{\mathrm{r},k},r_{\mathrm{r},k},\boldsymbol{S}_{k+1}^{\mathrm{Radar}})$存入组网雷达经验池$D_{\mathrm{r}}$。每隔$T$个训练幕，从$D_{\mathrm{r}}$中采样小批次的样本来训练组网雷达功率分配网络。

进行下一次迭代$m\leftarrow m+1$，重复执行步骤1和步骤2，直到迭代训练次数$m>M$。此时得到训练后的组网雷达功率分配策略Π_{RL}^{M}。

6 仿真实验

6.1 场景描述与参数设置

6.1.1 任务场景描述

如图10所示，代表一种典型的部署方式，各部雷达以扇形方式部署到作战区域，探测范围相互重叠，这种部署有效地增加了雷达发现目标的能力。目标由西北方向朝向东南方向匀速运动，并且逐渐靠近雷达4和雷达5所在区域。

值得注意的是，在测试场景干扰机和目标飞行轨迹的趋势与训练场景中轨迹的飞行样式相同，但每一次运行目标和干扰机的位置与速度都在一个区间内随机生成。在实际作战过程中，如果测试场景与训练场景的匹配度很低，训练好的参数可能不再有效。因为DRL是通过训练阶段不断地与环境交互学习最优策略。如果测试环境改变较大，可能会导致性能下降。此时，需要通过在线训练方式对模型进行微调，以适应新的环境。同时，本文通过对训练数据进行随机扰动，即每一个训练幕干扰机和目标的位置和速度随机产生，来增加模型对新情况的鲁棒性。

6.1.2 仿真参数设置

仿真实验在10 km×10 km的二维作战平面进行。组网雷达由$N=5$部广泛分布的单站雷达组成，融合中心采用K-N准则$(K=2)$。假设组网雷达和干扰机的工作频率相等，基于文献[9,10,23,29,30]提供的数据，每部雷达的工作参数设置如表1所示，干扰机的工作参数设置如表2所示。雷达的工作带宽为300 MHz，干扰机带宽为雷达带宽的2倍，各场景中目标的有效反射面积均设置为5 m^2。

本文算法的参数设置如表3所示，其中组网雷达智能体中Actor网络的层数和节点数与干扰机智能体中连续Actor的参数设置相同。仿真所使用的

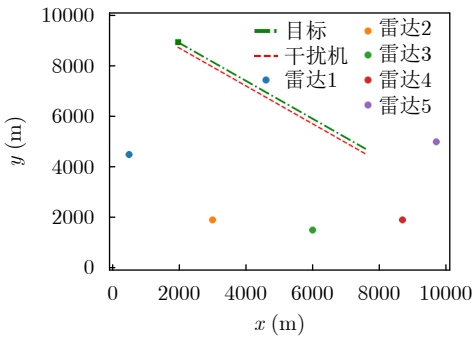


图 10 组网雷达部署和目标编队轨迹

Fig. 10 The deployment of the networked radar and the trajectory of the target formation

计算机硬件参数为：Intel i5-10400F CPU, 8 GB RAM, NVIDIA GTX 1650显示适配器，python版本为3.6，tensorflow版本为1.14.1。

算法的计算复杂度分析：算法的计算复杂度包含时间复杂度和空间复杂度[3]。前者由NN中乘法和加法的数量来衡量，后者由NN中带优化的参数数量决定，即

$$\begin{cases} O_{\text{time}} = \sum\limits_{m=1}^{M} \text{FC}_{\text{in}}(m)\,\text{FC}_{\text{out}}(m) \\ O_{\text{space}} = \sum\limits_{m=1}^{M} \{\text{FC}_{\text{in}}(m)\,\text{FC}_{\text{out}}(m) + \text{FC}_{\text{out}}(m)\} \end{cases} \tag{23}$$

其中，M是NN的层数(隐藏层数+1)，m表示NN层编号，$\text{FC}_{\text{in}}(m)$和$\text{FC}_{\text{out}}(m)$分别表示第$m$层NN的输入节点数和输出节点数。根据表3所示的Actor网络的参数设置，本文算法干扰机策略网络中离散Actor的时间复杂度是17792，空间复杂度是18049， 连续Actor的时间复杂度是18048，空间复杂度是18307。组网雷达智能体策略网络的时间复杂度为18304，空间复杂度为18565。

表 1 雷达工作参数

Tab. 1 The working parameters of the radars

参数	数值	参数	数值
发射总功率$P_{\text{r}}^{\text{total}}$	10 mW	工作频率	100 GHz
最小发射功率$P_{\text{r}}^{\min}$	0	最大发射功率$P_{\text{r}}^{\max}$	2 mW
天线增益G_{r}	45 dB	虚警概率P_{f}	10^{-6}

表 2 干扰机工作参数

Tab. 2 The working parameters of the jammer

参数	数值	参数	数值
干扰总功率$P_{\text{j}}^{\text{total}}$	60 W	干扰天线增益G_{j}	10 dB
最小发射功率$P_{\text{j}}^{\min}$	0	最大发射功率$P_{\text{j}}^{\max}$	60 W
干扰波束个数L	3	天线波瓣宽度$\theta_{0.5}$	3°
工作频率	100 GHz	极化失配损失γ_{j}	0.5

6.2 对比策略和评价指标

为验证所提方法的有效性，在干扰机使用基于DRL的干扰策略时，将基于DRL的组网雷达功率分配算法与如下2种组网雷达功率分配策略进行对比：

基于粒子群(Particle Swarm Optimization, PSO)算法的组网雷达分配策略：该方法采用粒子群算法作为雷达功率资源分配策略，在使用时设计参数较少、粒子群规模较小，所以收敛速度相对较快。

基于人工鱼群算法(Artificial Fish Swarms Algorithm, AFSA)的组网雷达分配策略：应用人工鱼群算法进行功率资源的分配，该方法通过模拟鱼群的觅食行为进行策略寻优，具有较好的全局最优解的求解能力，对初始值和参数要求较低、鲁棒性强。

组网雷达功率资源分配的目的是最大化目标的检测概率，因此选取目标检测概率以及资源调度运行时间(Scheduling Run Time, SRT)作为性能评估指标。

6.3 训练过程

根据6.1.2节设置的参数和第5节的训练方法学习组网雷达功率分配策略。每隔50步对目标运动状态进行初始化，称为一幕。干扰机策略训练的总幕数设置为3000幕，组网雷达智能体的训练总幕数设置为10000幕。图11显示了不同训练幕下奖励收敛情况。从图11(a)可以看出，随着训练幕数的增加组网雷达智能体的奖励逐渐收敛，表明训练是有效的。由图11(b)可以发现，随着训练幕数的增加干扰机智能体的奖励也逐渐收敛，表明干扰机的策略训练是有效的。

6.4 测试结果

将训练好的组网雷达功率分配策略参数和干扰机资源分配策略参数加载到测试环境。图12显示了单次运行下的干扰资源分配结果。可以发现，在初始阶段距离干扰机较近的雷达1、雷达2和雷达3受

表 3 算法参数设置

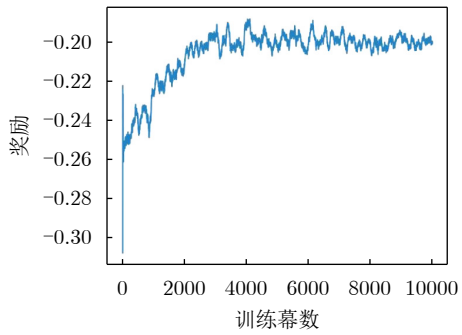
Tab. 3 The algorithm parameters setting

参数	取值	参数	取值
离散Actor的NN层数/节点数	3/128	离散Actor网络学习率	$3e^{-3}$
连续Actor的NN层数/节点数	3/128	连续Actor网络学习率	$3e^{-3}$
批量训练样本大小	128	经验回放区大小	1024
Critic网络学习率	$3e^{-3}$	执行性奖励权重	0.15
PPO裁剪参数	0.2	检测概率奖励权重	0.85
折扣因子	0.998	优化器	Adam
衰减因子β	0.9999	导向奖励参数b_1, b_2	0.5, 0.1

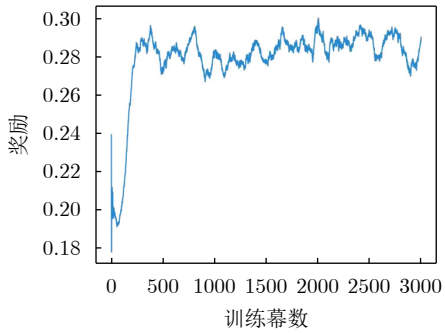
到的干扰较大；在运行到中间时刻时干扰机分配更多的干扰功率给雷达2；随着目标编队逐渐靠近雷达，干扰机选择对距离近的雷达4和雷达5施加干扰。

通过50次蒙特卡罗仿真测试了3种组网雷达功率分配策略在干扰机采用基于DRL的压制干扰下的目标检测性能。图13显示了几种组网雷达功率分配策略在基于DRL干扰下的目标检测概率，可以发现基于DRL的组网雷达功率分配方法可以有效地提升压制干扰下的目标检测性能，相较于其他两种策略，目标检测概率最多提升了大约11%，这是由于DRL通过智能体与环境交互学习，因此DRL分配策略考虑了干扰机带来的不确定性。

为了验证本文算法在时变干扰条件下的优势，



(a) 组网雷达智能体的奖励
(a) The reward of the networked radar agent



(b) 干扰机智能体的奖励
(b) The reward of the jammer agent

图 11 奖励变化曲线

Fig. 11 The rewards convergence curve

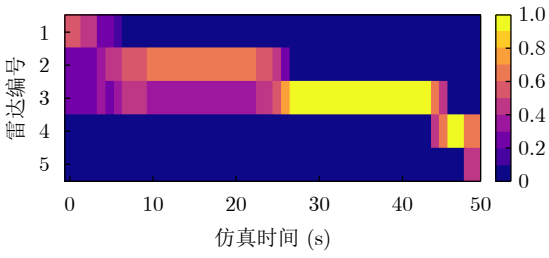


图 12 干扰资源分配结果

Fig. 12 The interference resource allocation result

本文分别采用3种不同的策略进行测试，包括基于DRL干扰下训练的组网雷达功率分配策略(DRL-TI)、在无干扰下训练的组网雷达分配策略(DRL-NI)以及固定干扰情况下训练的组网雷达功率分配策略(DRL-FI)。其中，固定干扰设置为干扰机在所有时刻采用均等功率干扰雷达3、雷达4和雷达5。本文测试了这3种策略的目标检测性能，结果如图14所示。从图14可以看出，DRL-TI组网雷达功率分配策略的目标检测概率要比DRL-NI和DRL-FI策略的目标检测概率高，最多提升了约15%。这是因为，DRL-TI分配策略在训练过程中考虑了干扰机与组网雷达的资源博弈，能够适应时变干扰带来的不确定性，从而具有更好的目标检测性能。

图15对比了单次仿真测试下3种组网雷达功率分配策略雷达功率分配结果。图16显示了各雷达节点受干扰压制干扰情况。图17显示了干扰机和组网雷达的距离变化。由图15(a)—图17可以发现基于DRL的组网雷达功率分配方法具有以下现象：

在1～25步，干扰机与雷达1、雷达2和雷达3的距离最近，因此干扰资源偏向于分配给这3部雷达，以达到最佳的干扰效果。为了对抗上述干扰策

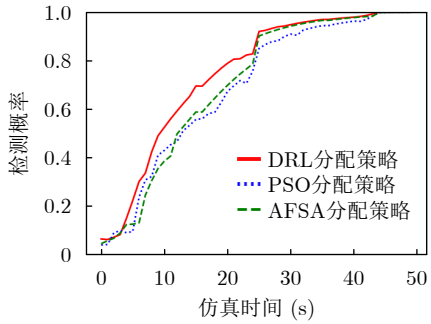


图 13 3种组网雷达功率分配策略的目标检测概率

Fig. 13 The target detection probability of three networked radar power allocation strategies

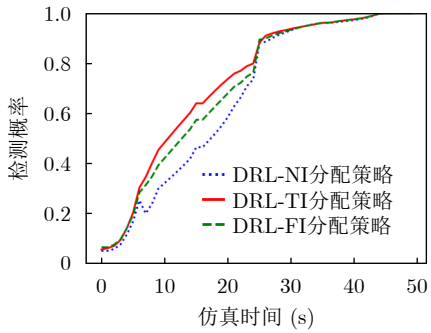
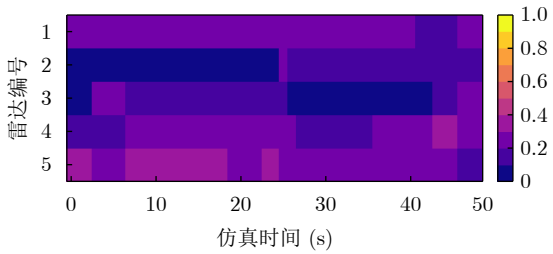
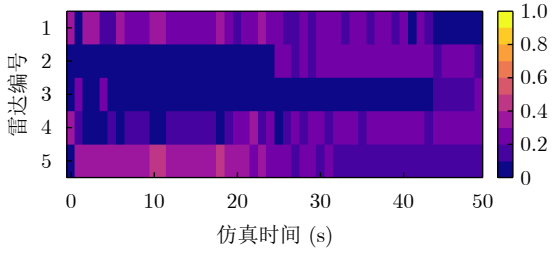


图 14 不同干扰模式下基于DRL组网雷达功率分配策略的目标检测概率

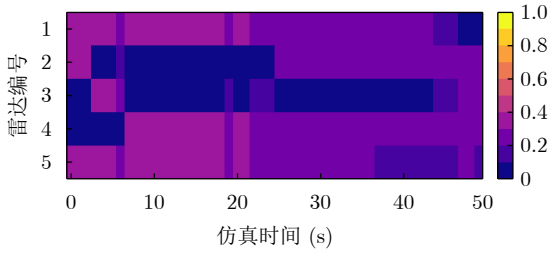
Fig. 14 The target detection probability of the DRL-based networked radar power allocation strategy under different interference models



(a) 基于DRL策略的组网雷达功率分配结果
(a) The result of the DRL-based networked radar power allocation



(b) 基于PSO策略的组网雷达功率分配结果
(b) The result of the PSO-based networked radar power allocation



(c) 基于AFSA策略的组网雷达功率分配结果
(c) The result of the AFSA-based networked radar power allocation

图 15 组网雷达功率分配结果

Fig. 15 The networked radar power allocation results

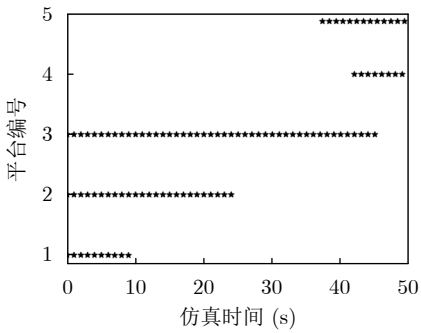


图 16 各雷达节点受压制干扰情况

Fig. 16 The indication that each radar node is interfered

略，组网雷达分配资源大部分资源给雷达1、雷达4和雷达5，其能够提升未被干扰且距离较远的雷达4和雷达5检测概率，同时能够提升受干扰最严重的雷达1的检测概率。采用该策略保证在K-N融合准

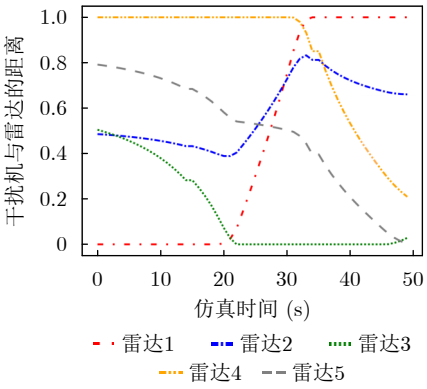


图 17 干扰机和组网雷达的距离变化

Fig. 17 The distance variation of the jammer and the networked radar

则下对的检测概率最大。在26～38步，干扰机所有的功率都用于干扰雷达3。在这种情况下组网雷达3的探测性能受到极大限制，所以在系统资源有限的情况下，几乎不分配资源于此节点，以保证对突防目标的及时探测。总体来看，基于DRL的组网雷达功率分配方法能够随着压制干扰强度以及目标运动实时动态调整每个雷达节点的功率，从而提高资源的利用率，进而提高压制干扰下目标的发现概率。

从图15(b)可以发现，基于PSO的组网雷达功率分配在整个仿真时刻呈现出交替分配较大功率给部分雷达，这种各个分配时刻交替分配功率的分配策略与雷达受干扰情况和雷达-目标的距离曲线的变化不符。原因在于基于PSO的组网雷达功率分配不能保证各个时刻的分配结果都是最优。

从图15(c)可以发现，基于AFSA的组网雷达功率分配在1～10步的功率分配结果变化较大，选择为部分雷达分配均等的发射功率；在6～25步组网雷达的能量主要分配给雷达1、雷达4和雷达5，而此时雷达2和雷达3受到干扰，可以发现AFSA将组网雷达功率分配给未受干扰的雷达，这种分配结果使雷达2和雷达3对目标的检测概率较低，进而导致K-N融合准则下目标的检测性能降低。在26～45步具有类似的结果，但是仅雷达3被干扰，因此融合后的目标检测性能下降小。总体来看，AFSA将组网雷达功率均匀地分配给未受到干扰的雷达节点，这种分配方式是一种保守的分配策略，在被干扰雷达节点较少时有较好的目标检测性能。

表4对比了50次蒙特卡罗仿真下各分配策略的资源调度运行时间。其中PSO算法和AFSA算法的种群规模数和最大迭代次数均相同，所有算法均在相同的仿真平台上运行。从运行时间来看，所提方法的资源调度运行时间能够达到0.01 s以下，相对于PSO优化方法和AFSA优化方法有显著提升，完

表 4 各策略的资源调度运行时间

Tab. 4 The resource scheduling running time of each strategy

调度策略	资源调度运行时间(s)
基于DRL的分配策略	0.009237
基于PSO的分配策略	5.227107
基于AFSA的分配策略	67.365983

全能够满足高动态博弈场景下雷达功率资源调度的实时性要求。

7 结语

考虑到干扰机与雷达相互博弈作战场景，本文提出了一种基于DRL的伴随压制干扰下组网雷达功率分配问题的解决方案。在该问题中，干扰机和组网雷达被映射为智能体。基于DRL的策略网络被用来训练组网雷达的功率分配策略，同时采用DRL生成干扰机智能体的波束选择和功率分配动作。此外，引入模型知识和专家知识，以协助两类智能体的策略探索。在仿真测试中，干扰机采用了基于DRL的干扰策略，而组网雷达分别采用了基于DRL的功率分配以及其他两种启发式组网雷达功率分配方法。比较了3种组网雷达功率分配方法在目标检测概率和运行时间两个指标下的表现。结果表明，当干扰机采用DRL资源分配策略时，组网雷达采用基于DRL的功率分配策略在两个指标上都优于其他方法。这是因为DRL采用离线训练生成策略模型，因此在线功率分配的运行时间相比PSO和AFSA更快。其次由于干扰机的干扰波束和功率具有不确定性和动态性，基于启发式搜索的组网雷达功率分配策略难以在这种环境下求得最优解，而DRL的分配策略是从智能体与环境交互的训练样本得到的，这些样本中包含了干扰机带来的不确定性，因此基于DRL组网雷达功率分配具有更好的目标检测性能。

在未来的工作中，我们将探究在多干扰机协同干扰下的组网雷达资源分配，并且拓展当前的组网雷达资源分配算法，使其能够适应分布式学习结构，以应对集群体系的对抗场景。同时，我们也会考虑其他针对雷达的抗干扰措施，如波束置零和干扰滤除等。

参 考 文 献

[1] 郝宇航, 蒋威, 王增福, 等. 分布式MIMO体制天波超视距雷达仿真系统[J/OL]. 系统工程与电子技术. https://kns.cnki.net/kcms/detail/11.2422.TN.20220625.1328.008.html, 2022.
HAO Yuhang, JIANG Wei, WANG Zengfu, *et al*. A distributed MIMO sky-wave over-the-horizon-radar simulation system[J/OL]. *Systems Engineering and Electronics*. https://kns.cnki.net/kcms/detail/11.2422.TN.20220625.1328.008.html, 2022.

[2] 潘泉, 王增福, 梁彦, 等. 信息融合理论的基本方法与进展(II)[J]. 控制理论与应用, 2012, 29(10): 1233–1244. doi: 10.7641/j.issn.1000-8152.2012.10.CCTA111336.
PAN Quan, WANG Zengfu, LIANG Yan, *et al*. Basic methods and progress of information fusion (II)[J]. *Control Theory & Applications*, 2012, 29(10): 1233–1244. doi: 10.7641/j.issn.1000-8152.2012.10.CCTA111336.

[3] WANG Yuedong, LIANG Yan, ZHANG Huixia, *et al*. Domain knowledge-assisted deep reinforcement learning power allocation for MIMO radar detection[J]. *IEEE Sensors Journal*, 2022, 22(23): 23117–23128. doi: 10.1109/JSEN.2022.3211606.

[4] 闫实, 贺静, 王跃东, 等. 基于强化学习的多机协同传感器管理[J]. 系统工程与电子技术, 2020, 42(8): 1726–1733. doi: 10.3969/j.issn.1001-506X.2020.08.12.
YAN Shi, HE Jing, WANG Yuedong *et al*. Multi-airborne cooperative sensor management based on reinforcement learning[J]. *Systems Engineering and Electronics*, 2020, 42(8): 1726–1733. doi: 10.3969/j.issn.1001-506X.2020.08.12.

[5] YAN Junkun, JIAO Hao, PU Wenqiang, *et al*. Radar sensor network resource allocation for fused target tracking: a brief review[J]. *Information Fusion*, 2022, 86/87: 104–115. doi: 10.1016/j.inffus.2022.06.009.

[6] 严俊坤, 陈林, 刘宏伟, 等. 基于机会约束的MIMO雷达多波束稳健功率分配算法[J]. 电子学报, 2019, 47(6): 1230–1235. doi: 10.3969/j.issn.0372-2112.2019.06.007.
YAN Junkun, CHEN Lin, LIU Hongwei, *et al*. Chance constrained based robust multibeam power allocation algorithm for MIMO radar[J]. *Acta Electronica Sinica*, 2019, 47(6): 1230–1235. doi: 10.3969/j.issn.0372-2112.2019.06.007.

[7] 时晨光, 董璟, 周建江. 频谱共存下面向多目标跟踪的组网雷达功率时间联合优化算法[J]. 雷达学报, 2023, 12(3): 590–601. doi: 10.12000/JR22146.
SHI Chenguang, DONG Jing, and ZHOU Jianjiang. Joint transmit power and dwell time allocation for multitarget tracking in radar networks under spectral coexistence[J]. *Journal of Radars*, 2023, 12(3): 590–601. doi: 10.12000/JR22146.

[8] 程婷, 恒思宇, 李中柱. 基于脉冲交错的分布式雷达组网系统波束驻留调度[J]. 雷达学报, 2023, 12(3): 616–628. doi: 10.12000/JR22211.
CHENG Ting, HENG Siyu, and LI Zhongzhu. Real-time dwell scheduling algorithm for distributed phased array radar network based on pulse interleaving[J]. *Journal of Radars*, 2023, 12(3): 616–628. doi: 10.12000/JR22211.

[9] 孙俊, 张大琳, 易伟. 多机协同干扰组网雷达的资源调度方法[J]. 雷达科学与技术, 2022, 20(3): 237–244, 254. doi: 10.3969/j.

issn.1672-2337.2022.03.001.

SUN Jun, ZHANG Dalin, and YI Wei. Resource allocation for multi-Jammer cooperatively jamming netted radar systems[J]. *Radar Science and Technology*, 2022, 20(3): 237–244, 254. doi: 10.3969/j.issn.1672-2337.2022.03.001.

[10] 张大琳, 易伟, 孔令讲. 面向组网雷达干扰任务的多干扰机资源联合优化分配方法[J]. 雷达学报, 2021, 10(4): 595–606. doi: 10.12000/JR21071.

ZHANG Dalin, YI Wei, and KONG Lingjiang. Optimal joint allocation of multijammer resources for jamming netted radar system[J]. *Journal of Radars*, 2021, 10(4): 595–606. doi: 10.12000/JR21071.

[11] 黄星源, 李岩屹. 基于双Q学习算法的干扰资源分配策略[J]. 系统仿真学报, 2021, 33(8): 1801–1808. doi: 10.16182/j.issn1004731x.joss.20-0253.

HUANG Xingyuan and LI Yanyi. The allocation of jamming resources based on double Q-learning algorithm[J]. *Journal of System Simulation*, 2021, 33(8): 1801–1808. doi: 10.16182/j.issn1004731x.joss.20-0253.

[12] 段燕辉. 雷达智能抗干扰决策方法研究[D]. [硕士论文], 西安电子科技大学, 2021. doi: 10.27389/d.cnki.gxadu.2021.001639.

DUAN Yanhui. Research on radar intelligent anti-jamming decision method[D]. [Master dissertation], Xidian University, 2021. doi: 10.27389/d.cnki.gxadu.2021.001639.

[13] 宋佰霖, 许华, 齐子森, 等. 一种基于深度强化学习的协同通信干扰决策算法[J]. 电子学报, 2022, 50(6): 1301–1309. doi: 10.12263/DZXB.20210814.

SONG Bailin, XU Hua, QI Zisen, *et al*. A collaborative communication jamming decision algorithm based on deep reinforcement learning[J]. *Acta Electronica Sinica*, 2022, 50(6): 1301–1309. doi: 10.12263/DZXB.20210814.

[14] 肖悦, 张贞凯, 杜聪. 基于改进麻雀搜索算法的雷达功率与带宽联合分配算法[J]. 战术导弹技术, 2022(5): 38–43, 92. doi: 10.16358/j.issn.1009-1300.20220077.

XIAO Yue, ZHANG Zhenkai, and DU Cong. Joint power and bandwidth allocation of radar based on improved sparrow search algorithm[J]. *Tactical Missile Technology*, 2022(5): 38–43, 92. doi: 10.16358/j.issn.1009-1300.20220077.

[15] 靳标, 邝晓飞, 彭宇, 等. 基于合作博弈的组网雷达分布式功率分配方法[J]. 航空学报, 2022, 43(1): 324776. doi: 10.7527/S1000-6893.2020.24776.

JIN Biao, KUANG Xiaofei, PENG Yu, *et al*. Distributed power allocation method for netted radar based on cooperative game theory[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(1): 324776. doi: 10.7527/S1000-6893.2020.24776.

[16] SHI Chenguang, WANG Fei, SELLATHURAI M, *et al*. Non-cooperative game-theoretic distributed power control technique for radar network based on low probability of intercept[J]. *IET Signal Processing*, 2018, 12(8): 983–991. doi: 10.1049/iet-spr.2017.0355.

[17] 李伟, 王泓霖, 郑家毅, 等. 博弈条件下雷达波形设计策略研究[J]. 电子与信息学报, 2019, 41(11): 2654–2660. doi: 10.11999/JEIT190114.

LI Wei, WANG Honglin, ZHENG Jiayi, *et al*. Research on radar waveform design strategy under game condition[J]. *Journal of Electronics & Information Technology*, 2019, 41(11): 2654–2660. doi: 10.11999/JEIT190114.

[18] HE Jin, WANG Yuedong, LIANG Yan, *et al*. Learning-based airborne sensor task assignment in unknown dynamic environments[J]. *Engineering Applications of Artificial Intelligence*, 2022, 111: 104747. doi: 10.1016/j.engappai.2022.104747.

[19] MU Xingchi, ZHAO Xiaohui, and LIANG Hui. Power allocation based on reinforcement learning for MIMO system with energy harvesting[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(7): 7622–7633. doi: 10.1109/TVT.2020.2993275.

[20] RUMMERY G A and NIRANJAN M. On-Line Q-learning Using Connectionist Systems[M]. Cambridge, UK: Cambridge University, 1994: 6–7.

[21] LI Jun and SHEN Xiaofeng. Robust jamming resource allocation for cooperatively suppressing multi-station radar systems in multi-jammer systems[C]. 2022 25th International Conference on Information Fusion (FUSION), Linköping, Sweden, 2022: 1–8. doi: 10.23919/FUSION49751.2022.9841340.

[22] YAO Zekun, TANG Chuanbin, WANG Chao, *et al*. Cooperative jamming resource allocation model and algorithm for netted radar[J]. *Electronics Letters*, 2022, 58(22): 834–836. doi: 10.1049/ell2.12611.

[23] ZHANG Dalin, SUN Jun, YI Wei, *et al*. Joint jamming beam and power scheduling for suppressing netted radar system[C]. 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 2021: 1–6. doi: 10.1109/RadarConf2147009.2021.9455337.

[24] 夏成龙, 李祥, 刘辰烨, 等. 基于深度强化学习的智能干扰方法研究[J]. 电声技术, 2022, 46(5): 144–149. doi: 10.16311/j.audioe.2022.05.035.

XIA Chenglong, LI Xiang, LIU Chenye, *et al*. Reserch of intelligent interference methods based on deep reinforcement learning[J]. *Audio Engineering*, 2022, 46(5): 144–149. doi: 10.16311/j.audioe.2022.05.035.

[25] LIU Weijian, WANG Yongliang, LIU Jun, *et al*. Performance analysis of adaptive detectors for point targets in subspace interference and Gaussian noise[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2018,

54(1): 429–441. doi: 10.1109/TAES.2017.2760718.

[26] 王国良, 申绪涧, 汪连栋, 等. 基于秩K融合规则的组网雷达系统干扰效果评估[J]. 系统仿真学报, 2009, 21(23): 7678–7680. doi: 10.16182/j.cnki.joss.2009.23.017.
WANG Guoliang, SHEN Xujian, WANG Liandong, *et al*. Effect evaluation for noise blanket jamming against netted radars Based on Rank-K information fusion rules[J]. *Journal of System Simulation*, 2009, 21(23): 7678–7680. doi: 10.16182/j.cnki.joss.2009.23.017.

[27] SUTTON R S and BARTO A G. Reinforcement Learning: An Introduction[M]. Cambridge: MIT Press, 2018: 327–331.

[28] SCHULMAN J, WOLSKI F, DHARIWAL P, *et al*. Proximal policy optimization algorithms[EB/OL]. https://arxiv.53yu.com/abs/1707.06347, 2017.

[29] WU Zhaodong, HU Shengliang, LUO Yasong, *et al*. Optimal distributed cooperative jamming resource allocation for multi-missile threat scenario[J]. *IET Radar, Sonar & Navigation*, 2022, 16(1): 113–128. doi: 10.1049/rsn2.12168.

[30] BARTON D K. Radar System Analysis and Modeling[M]. Boston: Artech House, 2004: 88–89.

作 者 简 介

王跃东，博士生，主要研究方向为传感器管理、目标跟踪、深度强化学习等。

顾以静，硕士生，主要研究方向为雷达资源调度、深度强化学习等。

梁 彦，博士，教授，主要研究方向为多源信息融合，复杂系统建模、估计与控制等。

王增福，博士，副教授，主要研究方向为天波雷达数据处理、信息融合、传感器管理等。

张会霞，博士生，主要研究方向为集群意图推理识别、智能体控制等。

(责任编辑：于青)