基于强化学习的频控阵-多输入多输出雷达发射功率分配方法

丁梓航* 谢军伟 齐 铖
(空军工程大学防空反导学院 西安 710051)

摘 要：当前电磁环境日益复杂多变，新式干扰手段层出不穷，对雷达系统带来了极大的挑战和威胁。该文引入频谱干扰模型并提出了一种在频控阵-多输入多输出(FDA-MIMO)雷达与干扰机动态博弈框架下基于强化学习(RL)的发射功率分配优化方法，使雷达系统能够获得最大的信干噪比(SINR)。在此基础上，构造了频谱干扰模型。其次，雷达和干扰机之间存在一种Stackelberg博弈关系，且将雷达作为领导者，干扰机作为跟随者，建立动态博弈框架下的发射功率分配优化模型。采用深度确定性策略梯度(DDPG)算法，结合功率约束设计了奖赏函数，对雷达发射功率进行实时分配来获得最大的输出SINR。最后，仿真结果表明，在雷达与干扰机博弈的框架下，所提优化算法能够有效地对雷达发射功率进行优化，使雷达具备较好的抗干扰性能。

关键词：频控阵；强化学习；博弈论；功率分配
中图分类号：TN958.5 **文献标识码**：A **文章编号**：1009-5896(2023)02-0550-08
DOI: 10.11999/JEIT211555

Transmit Power Allocation Method of Frequency Diverse Array-Multi Input and Multi Output Radar Based on Reinforcement Learning

DING Zihang XIE Junwei QI Cheng
(*Air and Missile Defense College*, *Aire Force Engineering University*, *Xi'an* 710051, *China*)

Abstract: In recent years, the electromagnetic environment has been becoming increasingly complex and changeable, and new jamming methods emerge one after another, which brings great challenges and threats to the radar system. In this paper, the spectrum interference model is introduced and a transmit power allocation optimization method based on Reinforcement Learning (RL) under the dynamic game framework of Frequency Diverse Array Multi Input and Multi Output (FDA-MIMO) radar and the spectrum interference is proposed, so that the radar system can obtain the maximum output Signal-to-Interference plus Noise Ratio (SINR). Firstly, the mathematical model of FDA-MIMO radar is established, and on this basis, the spectrum interference model is constructed. Secondly, there is a Stackelberg game relationship between radar and jammer. Taking radar as the leader and jammer as the follower, the transmit power allocation optimization model under the framework of dynamic game is established. Using the Deep Deterministic Policy Gradient (DDPG) algorithm and power constraints, a reward function is designed to allocate the radar transmit power in real time to obtain the maximum output SINR. Finally, the simulation results show that under the framework of the game between radar and interference, the proposed optimization algorithm can effectively optimize the radar transmit power and make the radar have better anti-jamming performance.

Key words: Frequency Diverse Array (FDA); Reinforcement Learning (RL); Game theory; Power allocation

1 引言

雷达系统位于复杂多变的电磁环境中，在敌方干扰机和其他干扰源会对雷达正常工作带来巨大的影响。因此，如何抑制环境中的干扰，提高雷达接收端的信干噪比(Signal-to-Interference-plus-Noise Ratios, SINR)，对于雷达系统是至关重要的。

频控阵(Frequency Diverse Array, FDA)这一概念于2006年被提出[1]。相较于传统的相控阵雷达，FDA雷达的每个发射阵元间存在一个远小于载波频率的频率偏移量，这一频偏量使其能够获得角度-距离2维相关的波束方向图[2–4]。FDA波束因具有角度-距离相关这一特性，使其被广泛应用于包括目标角度-距离定位[5]，2维波束形成技术和波束方向图设计等领域[6,7]。

多输入多输出(Multi-Input and Multi-Output,

收稿日期：2021-12-22；改回日期：2022-02-24；网络出版：2022-03-07
*通信作者： 丁梓航 dingzihang0831@163.com

MIMO)雷达因其与传统相控阵(Phase Array, PA)雷达相比所具有的独特优势而得到了广泛的研究。文献[8]将FDA与MIMO雷达相结合，并提出了FDA-MIMO雷达接收处理模型。FDA-MIMO雷达同时具有FDA雷达距离-角度相关的波束方向图和MIMO雷达所拥有的多自由度的特点，由此可以被用于欺骗干扰压制[9–11]、联合角度-距离估计[5]和空时自适应杂波抑制[12,13]等。

近年来，雷达与干扰的博弈现象受到广泛关注。文献[14]对雷达对抗中的博弈论问题进行了系统的分析与梳理。文献[15]对博弈论思想在雷达系统设计中的应用进行了综述，主要集中于雷达对抗、雷达资源管理、雷达波形设计、雷达射频隐身等方面。文献[16]提出基于回波间互信息量(Mutual Information, MI)准则的Stackelberg博弈波形设计。文献[17]对多基地分布式MIMO雷达组网的功率进行了纳什均衡分析，提出了一种以SINR为约束的雷达功率分配优化方法。上述文献建立的博弈模型用于雷达与干扰的对抗分析，而针对频谱式干扰的研究还很少。

在博弈的阶段中，实际上是一个动态优化的过程。若干扰信号发生变化，雷达系统就需要立即调整发射功率分配模式，以获得较高的SINR。传统的优化方法普遍存在计算复杂度高的问题，而对抗过程是一个高实时性问题，因此亟需一种处理速度快的优化方法。近年来，深度学习(Deep Learning, DL)成为研究热点，而强化学习可以实现离线学习、在线寻优。对于已经离线训练好的网络，将当前状态输入到网络中，可以实时获取优化的结果。文献[18]利用凸优化方法对MIMO雷达发射功率进行优化以获得最优的检测性能。

在此基础上，本文建立了FDA-MIMO雷达与频谱干扰机的Stackelberg博弈模型。在两者动态博弈的过程中，利用强化学习中的DDPG算法对采集的干扰信号状态进行离线训练，获得演员和评论家网络的参数，然后根据雷达当前侦测到的频谱干扰样式对发射功率进行在线动态优化，使雷达在工作时间段内获得最优的输出SINR性能，达到对抗频谱干扰的效果。

2　数据模型

2.1　FDA-MIMO雷达

考虑一个发射和接收阵列均为均匀线性阵列的FDA-MIMO雷达。其中，雷达发射阵列含有M个发射阵元，阵元间隔为$d=\lambda/2$(λ为波长)。在接收阵列中，接收阵元数为N_r，阵元间隔为$d=\lambda/2$。假设该雷达发射信号类型为脉冲信号，则第m个发射阵元发射信号的表达式为

$$s_m(t)=w_m\varphi_m(t)\,\mathrm{e}^{\mathrm{j}2\pi(f_0+\Delta f_m)t},\ 0\le t\le T_\mathrm{d} \tag{1}$$

其中，$\Delta f_m=(m-1)\Delta f$表示第m个发射阵元的频偏量，f_0表示发射信号的载波频率，T_d为发射脉冲信号的脉冲持续时间。由于阵元间Δf_m的存在，FDA-MIMO雷达能够同时工作在多个频率上，也使其具有精准频谱干扰的抗干扰能力。w_m, $\varphi_m(t)$分别是第m个发射阵元的发射信号功率值和基带波形且$\varphi_m(t)$满足关系式为

$$\int\varphi_{m_1}(t)\,\varphi_{m_2}^*(t-\tau)\mathrm{d}t=0,\ m_1\ne m_2 \tag{2}$$

假设空间中一个远场目标位于空间位置(θ,r)，经过目标反射，第n个接收阵元接收到来自第m个发射的信号可以表示为

$$\begin{aligned}s_{n,m}(t)=&w_m\varphi_m(t-\tau_{m,n})\,\mathrm{e}^{\mathrm{j}2\pi(f_0+\Delta f_m)(t-\tau_{m,n})},\\&\tau_{m,n}\le t\le\tau_{m,n}+T_\mathrm{d}\end{aligned} \tag{3}$$

$\tau_{m,n}$为信号在空间中传播的时延，其表达式为

$$\tau_{m,n}=2r/\mathrm{c}-(m-1)\,d\sin\theta/\mathrm{c}-(n-1)\,d\sin\theta/\mathrm{c} \tag{4}$$

c表示光速。在窄带信号假设下，式(3)可以近似改写为

$$\begin{aligned}s_{n,m}(t)\approx&w_m\varphi_m(t-\tau_0)\,\mathrm{e}^{\mathrm{j}\psi(t)},\\&\tau_0\le t\le\tau_0+T_\mathrm{d},\tau_0=2r/\mathrm{c}\end{aligned} \tag{5}$$

$\psi(t)$为信号传播带来的相位变化量，且可以表示为

$$\begin{aligned}\psi(t)=&2\pi\left[f_0t+\Delta f_mt+\frac{(m-1)\sin\theta}{2}\right.\\&\left.+\frac{(n-1)\sin\theta}{2}-2r\Delta f_m/\mathrm{c}\right]\end{aligned} \tag{6}$$

当信号被雷达接收系统接收后，会经过一系列的信号处理过程。文献[19]提出了一种多匹配滤波器的FDA-MIMO雷达的接收处理系统，本文也采用该接收处理方法。根据发射信号的相互正交性，经过匹配滤波器处理后的信号可以表示为

$$\begin{aligned}\boldsymbol{b}(\theta,r)=&\left[s_{1,1}^{\mathrm{output}},\cdots,s_{1,m}^{\mathrm{output}},\cdots,s_{n,m}^{\mathrm{output}}\right]^\mathrm{T}\\=&\gamma\boldsymbol{a}_r(\theta)\otimes[\boldsymbol{w}\odot\boldsymbol{a}_t(\theta,r)]\end{aligned} \tag{7}$$

$$\begin{aligned}\boldsymbol{a}_t(\theta,r)=&\left[1,\mathrm{e}^{\mathrm{j}\pi(\sin\theta-4r\Delta f/\mathrm{c})},\cdots,\right.\\&\left.\mathrm{e}^{\mathrm{j}\pi(M-1)(\sin\theta-4r\Delta f/\mathrm{c})}\right]^\mathrm{T}\end{aligned} \tag{8a}$$

$$\boldsymbol{a}_r(\theta)=\left[1,\mathrm{e}^{\mathrm{j}\pi\sin\theta},\cdots,\mathrm{e}^{\mathrm{j}\pi(N_r-1)\sin\theta}\right]^\mathrm{T} \tag{8b}$$

$$\boldsymbol{w}=[w_1,w_2,\cdots,w_{N_t}]^\mathrm{T} \tag{8c}$$

其中，$\otimes$表示克罗内克积，$\odot$表示哈达玛积，$(\cdot)^\mathrm{T}$表

示转置操作。$\boldsymbol{a}_t(\theta,r)$, $\boldsymbol{a}_r(\theta)$分别为发射、接收导向矢量，$\boldsymbol{w}$为发射功率向量，γ为目标反射系数。

2.2 干扰模型

考虑雷达系统处于频谱干扰环境，该干扰可能来自敌方的干扰机和其他与雷达共享频段的无线电。假设干扰信号可以表示为$s(t)$，为了方便分析，考虑从第1个接收阵元。由2.1节，FDA-MIMO雷达将接收到的干扰信号$s(t)$通过信号接收处理过程后，在第m个通道采集到的信号为

$$s_m(t)=\int_{-\infty}^{+\infty}s(t)\,\mathrm{e}^{-\mathrm{j}2\pi f_m t}\varphi^*(t-\tau)\,\mathrm{d}t \tag{9}$$

其中，$(\cdot)^*$表示共轭操作，τ表示采样时延，对应目标所处的距离单元。

将经过M个通道处理后的干扰信号表示为矢量形式$\boldsymbol{s}=[s_1,s_2,\cdots,s_M]^{\mathrm{T}}$。经过接收处理的频谱干扰信号$\boldsymbol{s}$服从均值为0，协方差矩阵为$\boldsymbol{P}$的复高斯分布，其中

$$\boldsymbol{P}=\begin{bmatrix}P_1&0&\cdots&0\\0&P_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&P_M\end{bmatrix} \tag{10}$$

当N_r个接收阵元接收到干扰信号，每个阵元中都有M个处理通道，假设存在K个远场干扰信号，信号方位角为$\{\theta_k\}_{k=1}^K$，接收阵列采集到的干扰加噪声信号向量为

$$\boldsymbol{h}=\sum_{k=1}^{K}\boldsymbol{s}_k\otimes\boldsymbol{a}_r(\theta_k)+\boldsymbol{n} \tag{11}$$

其中，$\boldsymbol{n}\in\mathbb{C}^{MN_r\times 1}$为噪声矢量，且服从均值为0，协方差矩阵为$\sigma_n^2\boldsymbol{I}_{MN_r}$的复高斯分布($\sigma_n^2$为接收噪声的功率，$\boldsymbol{I}_{MN_r}$表示$MN_r$的单位矩阵)。

干扰加噪声的协方差矩阵可以表示为

$$\boldsymbol{Q}=\sum_{k=1}^{K}\boldsymbol{P}_k\otimes\left[\boldsymbol{a}_r(\theta_k)\boldsymbol{a}_r^{\mathrm{H}}(\theta_k)\right]+\sigma_n^2\boldsymbol{I}_{MN_r} \tag{12}$$

则雷达接收到目标、干扰、噪声的总信号可以表示为

$$\begin{aligned}\boldsymbol{y}=&\boldsymbol{b}+\boldsymbol{h}=\gamma\boldsymbol{a}_r(\theta)\otimes[\boldsymbol{w}\odot\boldsymbol{a}_t(\theta,r)]\\&+\sum_{k=1}^{K}\boldsymbol{s}_k\otimes\boldsymbol{a}_r(\theta_k)+\boldsymbol{n}\end{aligned} \tag{13}$$

雷达系统的抗干扰能力可以用接收获得的信干噪比(SINR)来表征。当雷达系统工作时间位于$t(t=1,2,\cdots,T)$时刻，基于最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)的线性检测器

$$\boldsymbol{W}_t=\boldsymbol{Q}_t^{-1}\boldsymbol{b}_t \tag{14}$$

具有最高的输出SINR，在该检测器下获得的SINR由式(15)给出

$$\begin{aligned}\mathrm{SINR}=&\frac{\mathbb{E}\left\{\left|\gamma_t\boldsymbol{W}_t^{\mathrm{H}}\boldsymbol{b}(\theta_t,r_t)\right|^2\right\}}{\mathbb{E}\left\{\left|\boldsymbol{W}_t^{\mathrm{H}}h\right|^2\right\}}\\=&|\gamma_t|^2\boldsymbol{b}^{\mathrm{H}}(\theta_t,r_t)\boldsymbol{Q}^{-1}\boldsymbol{b}(\theta_t,r_t)\end{aligned} \tag{15}$$

其中，$(\cdot)^{\mathrm{H}}$表示共轭转置操作。

从式(15)可以看出，不同的发射阵元功率分配模式，在MVDR线性检测器下，可以获得不同的SINR。因此，通过对发射阵元的功率分配优化，可以获得最高的SINR。

3 基于强化学习的雷达抗干扰博弈的功率分配算法

3.1 博弈的基本模型

一种基于FDA-MIMO雷达的功率分配博弈论框架被建立。在雷达系统工作时，环境中的频谱干扰对其产生了极大的影响，且雷达系统与干扰信号之间没有合作关系，因此两者建立非合作博弈关系，前者控制发射阵元功率矢量$\boldsymbol{w}$，后者控制干扰信号的发射功率矩阵$\boldsymbol{P}$。雷达和干扰之间是一种零和博弈，即一个参与者的增益是另一个参与者的损失。雷达和干扰的博弈框架可以表示如式(16)的形式

$$G=\{P,\ S,\ U\} \tag{16}$$

(1) 参与者集：$P=\langle$雷达,干扰机$\rangle$表示博弈的参与者。

(2) 策略集：$S=\Big\{(\boldsymbol{w},\boldsymbol{P})\,|\boldsymbol{w}=[w_1,w_2,\cdots,w_{N_t}]^{\mathrm{T}},\ \boldsymbol{P}=\mathrm{diag}\{P_1,P_2,\cdots,P_{N_t}\}\Big\}$，其中，$\boldsymbol{w}$为雷达的行动策略，$\boldsymbol{P}$为干扰机的行动策略。

(3) 效用函数：$U=\{U_r,U_i\}$，其中，$U_r=\max\{\mathrm{SINR}\}$为雷达的效用函数，雷达通过调整$\boldsymbol{w}$获得输出的最大SINR，$U_i=\min\{\mathrm{SINR}\}$为干扰机的效用函数，干扰机通过改变$\boldsymbol{P}$来获得输出的最大SINR。

在实际环境中，雷达与干扰机间存在一种Stackelberg博弈关系。Stackelberg博弈是一种完全信息动态博弈，跟随者根据主导者的行为制定自己的行为策略，然后主导者再根据跟随者的行为策略调整更新自己的策略以获得最大效用。在本文中，考虑我方雷达需要在变化的干扰环境中保持较好的抗干扰性能，因此雷达是主导者，干扰机是跟随者。在Stackelberg博弈框架下，该问题转化为两阶段的优化问题，该优化问题如式(17)所示

$$\max_{\boldsymbol{w}}\min_{\boldsymbol{P}}\ |\gamma_t|^2\boldsymbol{b}^{\mathrm{H}}(\theta_t,r_t)\boldsymbol{Q}^{-1}\boldsymbol{b}(\theta_t,r_t) \tag{17}$$

$$\text{s.t.}\ \sum_{i=1}^{N_t} w_i \le P_{\text{total}} \tag{17a}$$

$$P_{\min} \le w_i \le P_{\max} \tag{17b}$$

其中，$P_{\min}$, $P_{\max}$分别表示每个发射阵元的功率最小值和最大值，P_{total}为发射阵元的总功率。

需要说明的是，在整个博弈时间T中，干扰机对雷达频谱功率的感知需要一定的时间，即在雷达根据干扰信号动态调整功率分配策略后，干扰机只能在多个时刻后才能对发射的干扰信号进行调整，使得雷达获得最小输出SINR。雷达系统可通过外部的频谱感知模块和辅助传感器阵列，感知环境中的干扰噪声信号，可以实时地估计出干扰信号的频谱以及协方差矩阵，便于雷达在博弈过程中获得最佳的抗干扰性能。

3.2 基于DDPG算法的优化求解

当雷达接收到干扰机发射的干扰信号后，需要自适应地调整发射阵元的功率来避开干扰信号。在整个博弈的过程中，本文采用强化学习中的DDPG算法对雷达发射功率矢量$\boldsymbol{w}$进行优化。将雷达各个阵元设置为智能体，负责收集环境中的干扰信号，并对发射阵元功率进行控制。

DDPG算法对智能体神经网络进行训练，得到网络参数。在第t个雷达工作时刻，雷达阵元根据所接收到的状态$\boldsymbol{s}_t^*$，利用行为网络输出功率分配行为$\boldsymbol{a}_t^*$，并获得对应的奖赏$\boldsymbol{r}_t^*$和下一步的状态$\boldsymbol{s}_{t+1}^*$，进入第t+1个工作时刻。将每个工作时刻获得的经验$(\boldsymbol{s}_t, \boldsymbol{a}_t, r_t, \boldsymbol{s}_{t+1})$存储在经验池中。接下来，本文将对强化学习网络中的状态、行为、奖赏和行为-评论家网络进行说明。

(1)状态。在第t个雷达工作时刻，强化学习中的状态是一个向量$\boldsymbol{s}_t = [x_t, y_t, \boldsymbol{p}_t]$，其中，$x_t$,$y_t$分别表示目标的空间位置，$\boldsymbol{p}_t$是第$t$个工作时刻下，干扰信号的功率矩阵的对角线元素组成的行向量。

(2)行为。在深度强化学习框架中，智能体的行为向量为$\boldsymbol{a}_t = [w_{1,t}, w_{2,t}, \cdots, w_{N_t,t}]$，其中每一个元素代表在每一个确定的工作时刻下雷达系统发射阵元的功率分配情况。

经过sigmoid函数输出的$\boldsymbol{a}_t$范围为[0, 1]，为保证行为$\boldsymbol{a}_t$中元素满足约束条件式(17a)$P_{\min} \le w_i \le P_{\max}$，其作用于环境时再通过线性变换映射至真实范围。设输出的某一个行为取值为$a_{t,i}$，将其从[0, 1]映射到$[P_{\min}, P_{\max}]$范围上的线性变换为$a_{t,i}\,(P_{\max} - P_{\min}) + P_{\min}$。

(3)奖赏函数。根据式(17)中的目标函数，奖赏函数被定义为雷达输出的SINR，即在第t个工作时刻下的奖赏函数为

$$r_t = |\gamma_t|^2 \boldsymbol{b}^{\mathrm{H}}(\theta_t, r_t)\, \boldsymbol{Q}^{-1} \boldsymbol{b}(\theta_t, r_t) \tag{18}$$

由于行为向量存在约束条件，为使强化学习网络能够满足行为的约束条件，本文提出一种新的奖赏函数来实现对行为的约束。该奖赏函数更新为

$$r_t = \begin{cases} |\gamma_t|^2 b^{\mathrm{H}}(\theta_t, r_t)\, \boldsymbol{Q}^{-1} \boldsymbol{b}(\theta_t, r_t), & \sum_{i=1}^{N_t} w_i \le P_{\text{total}} \\ -3, & \sum_{i=1}^{N_t} w_i > P_{\text{total}} \end{cases} \tag{19}$$

通过重构奖赏函数，可以将行为的约束引入到深度学习网络中。

3.2.1 功率优化方法的整体框架

在雷达侦测到干扰机发射干扰信号的先验信息的条件下，DDPG网络被用来求解雷达发射阵元的功率分配问题。在每一雷达工作时刻，雷达侦测到的先验信息被存储到记忆回放池并将其作为强化学习网络训练的输入。对于传统方法，在每一个工作时刻都需要对雷达的发射功率进行优化求解。经过记忆回放池中采样数据对DDPG网络训练完成后，可以直接获得雷达当前工作时刻优化后的功率分配结果。图1为雷达-干扰机博弈下的功率优化方法的整体框架示意图。

3.2.2 DDPG算法流程

采用DDPG算法对上述雷达-干扰机博弈功率优化模型进行离线训练，算法的整体流程如算法1所示。

在网络更新迭代训练过程中，先积累经验回放到经验池中，根据经验池中随机抽取的样本分别更新评论家和演员网络。首先通过损失函数更新评论家网络参数$\theta^{\boldsymbol{Q}}$。接下来通过评论家网络得到的$\boldsymbol{Q}$函数相对于动作的策略梯度，将梯度传递到演员网络中对其参数θ^{μ}进行更新。最后通过得到的$\theta^{\boldsymbol{Q}}$，θ^{μ}通过参数τ，按照设定的比例更新各自所属的目标网络中，其目标网络会在下一步的训练中用来预测行为和$\boldsymbol{Q}$值。

4 仿真分析

FDA-MIMO雷达发射接收阵列均为均匀线性阵列且阵元数分别为$M = 6$, $N_r = 5$，阵元间距均为$d = \lambda/2$。发射阵列阵元间频偏量$\Delta f = 10$ kHz，载波频率$f_0 = 1$ GHz。雷达工作时间最小间隔为2 s，雷达工作总时间设置为20帧(40 s)。每个发射阵元的功率最小值和最大值分别为$P_{\min} = 0$和$P_{\max} = P_{\text{total}}$。

在雷达工作初始时间内，空间中的目标位于(5 km, 5 km)，在整个雷达工作时间内，假设目标沿着45°方向，以速度$v = 100$ m/s做匀速直线运

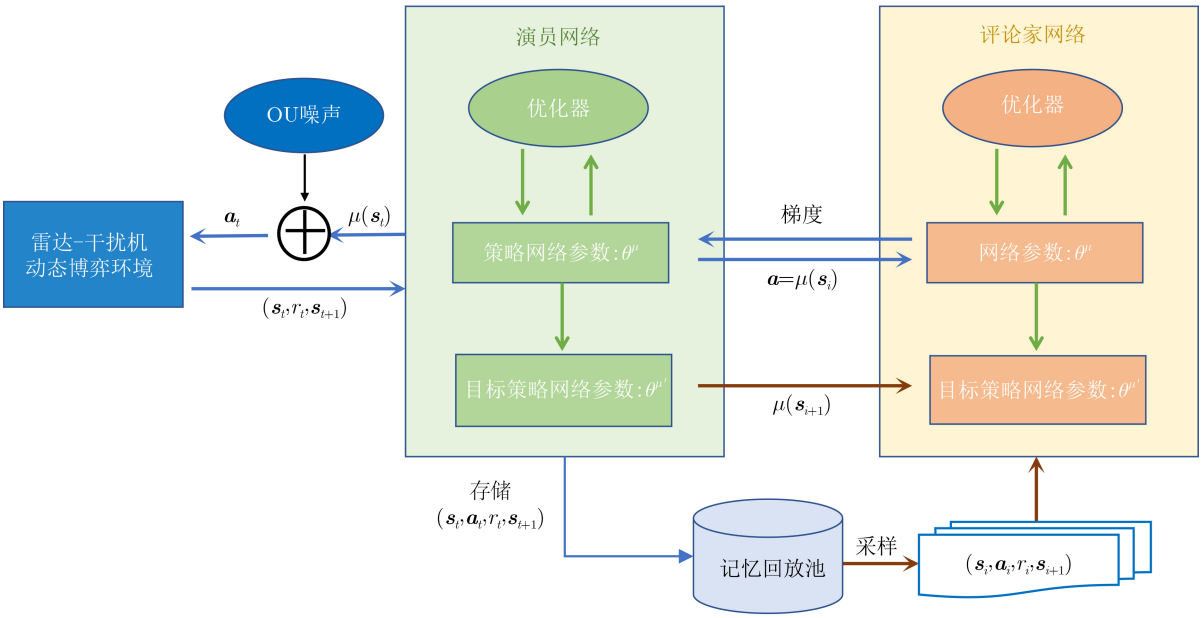


图 1 发射功率优化方法整体框架

算法1 DDPG算法

随机初始化评论家网络$Q(\cdot|\theta^Q)$和演员网络$\mu(\cdot|\theta^\mu)$的网络参数θ^Q, θ^μ

初始化目标评论家和演员网络的参数$\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$

初始化回放记忆池B

FOR 回合数=1:L do

在动作探索策略中初始化随机过程$\mathcal{O}$

接收初始观测状态$\boldsymbol{s}_1$

FOR t=1:T do

根据当前策略和随机噪声选择动作$a_t=\mu(\boldsymbol{s}_t|\theta^\mu)+\mathcal{O}_t$

执行动作a_t并且获得奖赏值r_t，得到新状态$\boldsymbol{s}_{t+1}$，

保存传递样本组合$(\boldsymbol{s}_t,\boldsymbol{a}_t,r_t,\boldsymbol{s}_{t+1})$到回放记忆池$B$

从回放记忆池B中随机采样生成H维数据库$(\boldsymbol{s}_t,\boldsymbol{a}_t,r_t,\boldsymbol{s}_{t+1})$

根据评论家网络$Q(\cdot|\theta^Q)$，计算目标值

$y_i=r_i+\varepsilon Q'\left(\boldsymbol{s}_{i+1},\mu'\left(\boldsymbol{s}_{i+1}|\theta^\mu\right)|\theta^Q\right)$

通过最小化损失函数更新评论家网络：

$\dfrac{1}{H}\sum\limits_{i=1}^{H}\left(y_i-Q\left(\boldsymbol{s}_i,\boldsymbol{a}_i|\theta^Q\right)\right)^2$

计算评论家网络的策略梯度：

$\nabla_{\boldsymbol{a}}Q\left(\boldsymbol{s},\boldsymbol{a}|\theta^Q\right)|_{a=\mu(\boldsymbol{s}_{i+1}|\theta^\mu),\boldsymbol{s}=\boldsymbol{s}_j}$

使用样本的策略梯度更新演员网络参数θ^μ：

$\dfrac{1}{H}\sum\limits_{i=1}^{H}\nabla_a Q\left(\boldsymbol{s},\boldsymbol{a}|\theta^Q\right)|_{a=\mu(\boldsymbol{s}_{i+1}|\theta^\mu),\boldsymbol{s}=\boldsymbol{s}_i}\cdot\nabla_{\theta^\mu}\mu\left(\boldsymbol{s}|\theta^\mu\right)|_{\boldsymbol{s}=\boldsymbol{s}_i}$

评论家和演员目标网络参数更新：

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

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

其中，$\tau(0<\tau<1)$为参数更新速率

END FOR

END FOR

动，雷达系统通过跟踪算法在每个工作时刻对目标的位置进行实时更新。

假设干扰机的干扰频段数量小于等于3。在初始工作时刻，干扰信号从方位角为45°进入雷达接收机系统，以30 dB, 25 dB和30 dB的干扰功率，干扰第1、第4和第6个发射阵元的频段。根据所提博弈准则，雷达作为领导者会根据干扰机释放的干扰信号通过自适应控制发射阵元的功率来提高输出的SINR。在雷达调整阵元功率后，经过一段时间，干扰机侦测到雷达发射信号的变化，通过调整干扰信号使得雷达系统接收端获得的SINR最小。雷达再根据新的干扰信号调整发射阵元的功率，形成博弈态势。

干扰信号的功率分布计算公式为

$$G_k(z,\theta)=\frac{1}{\boldsymbol{a}_z^{\mathrm{H}}(\theta)\boldsymbol{Q}_k^{-1}\boldsymbol{a}_z(\theta)} \tag{20}$$

其中，$\boldsymbol{a}_z(\theta)=\boldsymbol{e}_z\otimes\boldsymbol{a}_r(\theta)$, $\boldsymbol{e}_z$表示第z个元素为1，其余元素都为0的单位向量。z表示干扰频谱的位置索引。

(4)DDPG中的参数设置。智能体的状态表示为8维数组向量，动作表示为6维数组向量。演员网络包含2个隐含层，每个隐含层的神经元个数为32，所有隐含层都采用tanh激活函数。评论家网络包含3个隐含层，每个隐含层的神经元个数为32，所有隐含层都采用tanh激活函数。演员网络和评论家网络的学习率分别为0.001,0.01。折扣因子为$\kappa=1$，超参数$\varepsilon=0.1$。

本文程序是基于keras框架编写的，计算机硬件条件为Core i5-10210CPU, 3.60 GHz, 8 GB内存，设置回合数为800，每一个回合中的迭代步数为20。

在各个回合中，雷达与干扰机智能博弈的完整工作过程的累计奖赏值和平均SINR值如图2(a)和图2(b)所示，雷达的目的是通过不断地学习提升奖赏值，来获得最大的奖赏，即最大SINR值。累计奖赏值和SINR值越大表明雷达的功率分配优化结果越好，网络的学习效果越好。由仿真结果可以看出累计奖赏值与平均SINR值整体的变化趋势是逐步增加的。在回合数大于400时，累计奖赏值和平均SINR值均达到最大且基本保持稳定。

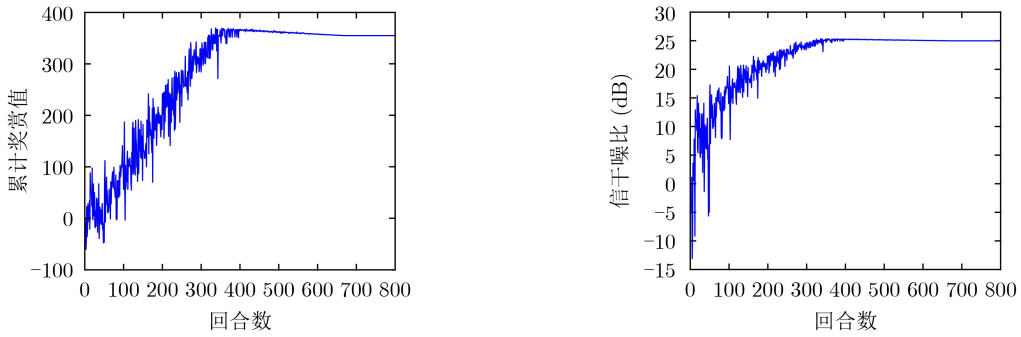
图3(a)为通过所提算法对FDA-MIMO雷达发射功率分配结果。从结果中可以看出，当$t=1$时，在当前工作时刻，雷达没有感知到外界的干扰信号，因此保留基本的功率均匀分配策略。当$2\le t\le 10$时，雷达侦测到干扰信号后，使发射功率集中在第2个发射阵元上；当$11\le t\le 15$时，由于干扰机调整了频谱干扰策略，雷达在侦测到干扰信号后立即对发射功率进行优化，使功率集中到第1个阵元；同样地，当$16\le t\le 20$时，雷达将发射功率集中到第6个阵元，以获得最优的SINR。图3(b)为通过内点法对雷达发射功率优化分配的结果。由仿真结果可以看出，内点法得到的功率分配策略与所提算法类似，在雷达工作时间段中，功率都集中在某一个发射阵元上，且均避开了频谱干扰信号所在的频点。由此，验证了本文所提算法能够实现与传统优化算法相似的优化效果，验证了其有效性。

为了直观地展现出内定法和DDPG算法的性能，两种算法得到的SINR值随雷达工作时间的变化情况如图4所示。从该仿真结果可以看出，当$t=1$时，雷达采用功率均匀分配策略，此时雷达已经受到干扰，因此SINR较低。当$t>1$时，雷达系统开始根据干扰信号对发射功率进行优化，得到的SINR较高。相较于内点法，文中所提算法对发射功率优化后得到的SINR有一定的波动，这是因为所提算法在训练时所用到的目标位置信息和干扰信息与仿真中设置的信息有一定差异。但通过两种算法得到的SINR基本一致，因此验证了所提算法的有效性。

干扰在整个工作时间段的干扰信号参数变化情况如表1所示。

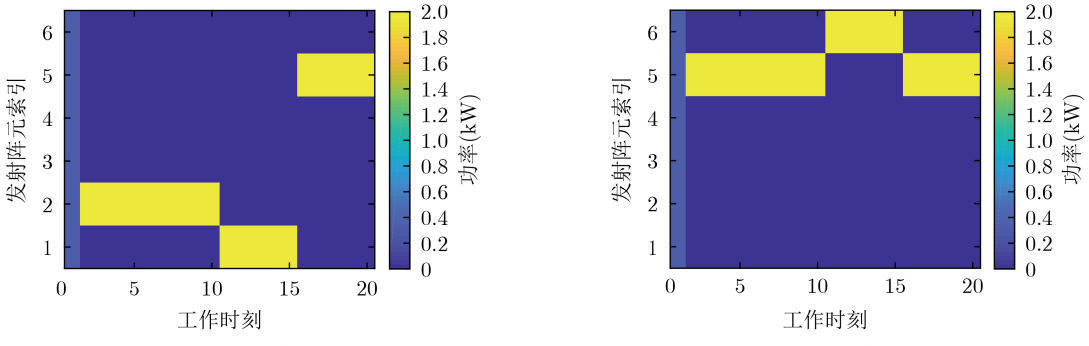
各个时刻干扰信号在角度-频率2维平面上的功率分布如图5所示。

强化学习网络中的演员网络为全连接网络，为了将本文所提算法与传统的优化算法复杂度性能进行对比，表2给出了所提算法和内点法的计算复杂度。其中N_{input}, N_1, N_2, N_{output}分别表示输入层神经元



(a) 累计奖赏值　　(b) 平均SINR值

图 2 累计奖赏值和SINR随回合数的变化情况



(a) 所提算法(帧)　　(b) 内点法(帧)

图 3 发射功率分配情况

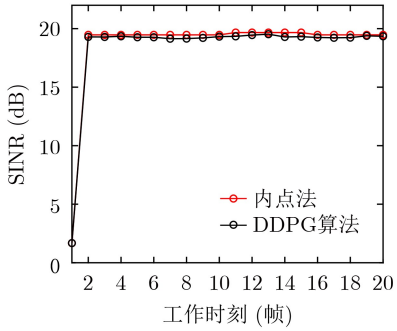
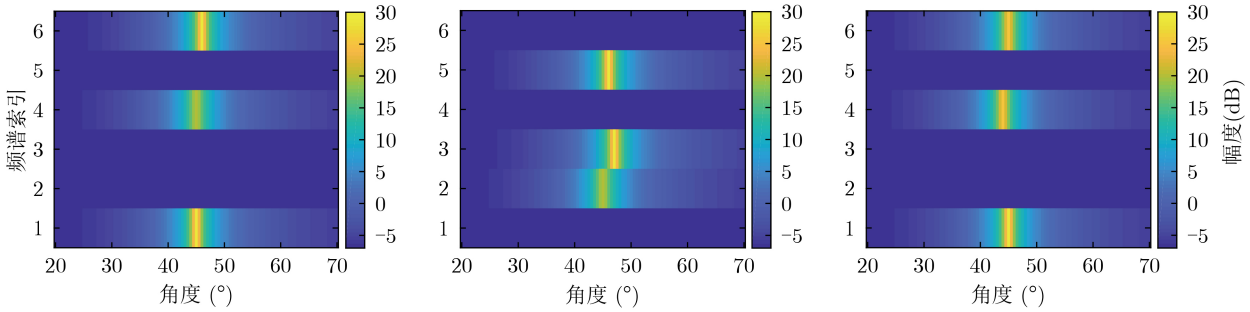


图 4 SINR值变化情况

个数，第1个隐含层神经元个数，第2个隐含层神经元个数和输出层神经元个数。N_{input}等于发射阵元数目。而传统优化算法，内点法的计算复杂度为：$\mathcal{O}\left((N_{\text{input}})^{3.5}\lg(1/\varepsilon)\right)$，$\varepsilon$表示求解精度。图6显示了所提算法和内点法的计算复杂度随发射阵元个数的变化情况，其中精度设置为$\varepsilon=0.01$。从该仿真结果可以看出当发射阵元数目较小时，两种算法的计算复杂度相近，但随着发射阵元数目的增加，内点法的计算复杂度增加迅速，而所提方法增加较少。该结果证明了所提算法优化的时间优势。

表 1 频谱干扰信号在工作时间段内的参数变化情况

参数	$1\le t\le 10$	$11\le t\le 15$	$16\le t\le 20$
干扰功率 (dB)	30, 20, 30	20, 30, 30	30, 25, 25
干扰频谱索引	1, 4, 6	2, 3, 5	1, 4, 6
干扰角度 (°)	45, 45, 46	45, 47, 46	45, 44, 45



(a) $1\le t\le 10$时干扰功率分布 (b) $11\le t\le 15$时干扰功率分布 (c) $16\le t\le 20$时干扰功率分布

图 5 干扰信号在频率-角度的功率分布情况

表 2 算法复杂度

	所提算法	内点法[20]
计算复杂度	$\mathcal{O}\left(N_{\text{input}}N_1+N_1N_2+N_2N_{\text{output}}\right)$	$\mathcal{O}\left((N_{\text{input}})^{3.5}\lg(1/\varepsilon)\right)$

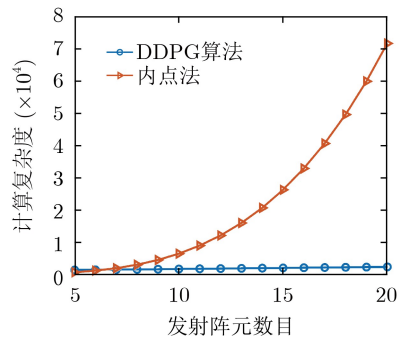


图 6 计算复杂度随发射阵元数目变化情况

5 结论

本文建立了FDA-MIMO雷达与释放频谱干扰信号的干扰机的Stackelberg博弈关系，并将雷达作为领导者，干扰机作为跟随者。为了使雷达获得最大的抗干扰效果，将深度确定策略梯度算法应用于雷达发射阵列的功率优化分配中，使得雷达在干扰信号产生变化过程中，能够动态调整阵元功率分配来获得最优SINR。算法中考虑了单个阵元功率约束和阵列总功率约束，并在约束下输出动作。仿真结果表明，通过DDPG算法的多个回合的训练，使雷达能够很好地感知干扰信号的变化并合理地分配发射阵元的功率，达到最优的SINR，实现抗干扰的效果。

参 考 文 献

[1] ANTONIK P, WICKS M C, GRIFFITHS H D, *et al*. Frequency diverse array radars[C]. The 2006 IEEE Conference on Radar, Verona, Italy, 2006: 215–217. doi: 10.1109/RADAR.2006.1631800.

[2] WANG Wenqing. Overview of frequency diverse array in radar and navigation applications[J]. *IET Radar, Sonar & Navigation*, 2016, 10(6): 1001–1012. doi: 10.1049/iet-rsn.2015.0464.

[3] WANG Wenqing and SHAO Huaizong. Range-angle

localization of targets by a double-pulse frequency diverse array radar[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2014, 8(1): 106–114. doi: 10.1109/JSTSP.2013.2285528.

[4] DING Zihang, XIE Junwei, WANG Bo, *et al*. Robust adaptive null broadening method based on FDA-MIMO radar[J]. *IEEE Access*, 2020, 8: 177976–177983. doi: 10.1109/ACCESS.2020.3025602.

[5] XU Jingwei, LIAO Guisheng, ZHU Shengqi, *et al*. Joint range and angle estimation using MIMO radar with frequency diverse array[J]. *IEEE Transactions on Signal Processing*, 2015, 63(13): 3396–3410. doi: 10.1109/TSP.2015.2422680.

[6] WANG Bo, XIE Junwei, ZHANG Jing, *et al*. Dot-shaped beamforming analysis of subarray-based sin-FDA[J]. *Frontiers of Information Technology & Electronic Engineering*, 2019, 20(10): 1429–1444. doi: 10.1631/FITEE.1800722.

[7] XIONG Jie, WANG Wenqing, SHAO Huaizong, *et al*. Frequency diverse array transmit beampattern optimization with genetic algorithm[J]. *IEEE Antennas Wireless Propagation Letters*, 2016, 16: 469–472. doi: 10.1109/LAWP.2016.2584078.

[8] SAMMARTINO P F, BAKER C J, and GRIGGITHS H D. Frequency diverse MIMO techniques for radar[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2013, 49(1): 201–222. doi: 10.1109/TAES.2013.6404099.

[9] XU Jingwei, LIAO Guisheng, ZHU Shengqi, *et al*. Deceptive jamming suppression with frequency diverse MIMO radar[J]. *Signal Processing*, 2015, 113: 9–17. doi: 10.1016/j.sigpro.2015.01.014.

[10] LAN Lan, XU Jingwei, LIAO Guisheng, *et al*. Suppression of mainbeam deceptive jammer with FDA-MIMO radar[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(10): 11584–11598. doi: 10.1109/TVT.2020.3014689.

[11] LAN Lan, LIAO Guisheng, XU Jingwei, *et al*. Transceive beamforming with accurate nulling in FDA-MIMO radar for imaging[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(6): 4145–4159. doi: 10.1109/TGRS.2019.2961324.

[12] XU Jingwei, ZHU Shengqi, and LIAO Guisheng. Space-time-range adaptive processing for airborne radar systems[J]. *IEEE Sensors Journal*, 2015, 15(3): 1602–1610. doi: 10.1109/JSEN.2014.2364594.

[13] XU Jingwei, LIAO Guisheng, HUANG Lei, *et al*. Robust adaptive beamforming for fast-moving target detection with FDA-STAP radar[J]. *IEEE Transactions on Signal Processing*, 2017, 65(4): 973–984. doi: 10.1109/TSP.2016.2628340.

[14] 赫彬, 苏洪涛. 认知雷达抗干扰中的博弈论分析综述[J]. 电子与信息学报, 2021, 43(5): 1199–1211. doi: 10.11999/JEIT200843.
HE Bin and SU Hongtao. A review of game theory analysis in cognitive radar anti-jamming[J]. *Journal of Electronics & Information Technology*, 2021, 43(5): 1199–1211. doi: 10.11999/JEIT200843.

[15] 吴家乐, 时晨光, 周建江. 博弈论在雷达系统中的应用研究综述[J]. 飞航导弹, 2021(9): 59–66.
WU Jiale, SHI Chenguang, and ZHOU Jianjiang. A review of game theory application in radar system[J]. *Aerodynamic Missile Journal*, 2021(9): 59–66.

[16] SONG Xiufeng, WILLETT P, ZHOU Shengli, *et al*. The MIMO radar and jammer games[J]. *IEEE Transactions on Signal Process*, 2012, 6(2): 687–699. doi: 10.1109/TSP.2011.2169251.

[17] DELIGIANNIS A, PANOUI A, LAMBOTHARAN S, *et al*. Game-theoretic power allocation and the NASH equilibrium analysis for a multistatic MIMO radar network[J]. *IEEE Transactions on Signal Processing*, 2017, 65(24): 6397–6408. doi: 10.1109/TSP.2017.2755591.

[18] GODRICH H, PETROPULU A P, and POOR H V. Power allocation strategies for target localization in distributed multiple-radar architectures[J]. *IEEE Transactions on Signal Processing*, 2011, 59(7): 3226–3240. doi: 10.1109/TSP.2011.2144976.

[19] DING Zihang and XIE Junwei. Joint transmit and receive beamforming for cognitive FDA-MIMO radar with moving target[J]. *IEEE Sensors Journal*, 2021, 21(18): 20878–20885. doi: 10.1109/JSEN.2021.3100332.

[20] LUO Zhiquan, MA W K, SO A M C, *et al*. Semidefinite relaxation of quadratic optimization problems[J]. *IEEE Signal Processing Magazine*, 2010, 27(3): 20–34. doi: 10.1109/MSP.2010.936019.

丁梓航：男，博士生，研究方向为频控阵阵列优化设计.
谢军伟：男，教授，研究方向为雷达干扰与抗干扰技术.
齐　铖：男，硕士生，研究方向为雷达资源管理与阵列优化设计.

责任编辑：余　蓉