



(12) 发明专利申请

(10) 申请公布号 CN 111917508 A

(43) 申请公布日 2020.11.10

(21) 申请号 202010797220.7

G06N 3/08 (2006.01)

(22) 申请日 2020.08.10

(71) 申请人 中国人民解放军陆军工程大学

地址 210000 江苏省南京市秦淮区后标营路88号

(72) 发明人 徐煜华 汪西明 刘典雄 徐逸凡 杨晓琴

(74) 专利代理机构 北京力量专利代理事务所 (特殊普通合伙) 11504

代理人 毛雨田

(51) Int. Cl.

H04K 3/00 (2006.01)

H04B 17/391 (2015.01)

H04B 7/0413 (2017.01)

G06N 3/04 (2006.01)

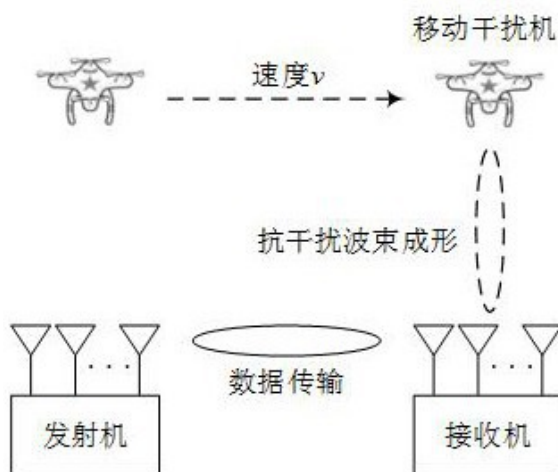
权利要求书4页 说明书9页 附图2页

(54) 发明名称

基于多天线抗干扰通信模型及动态空间谱抗干扰方法

(57) 摘要

本发明公开了一种基于多天线抗干扰通信模型及动态空间谱抗干扰方法,考虑一对具有多天线的发射端和接收端的用户进行通信,一个无人机干扰对用户通信进行干扰,将接收端获取的具有无人机干扰信号到达方向角信息的空间谱图作为学习的状态输入,获取无人机干扰的空间运动特征,并准确地对干扰信号进行消除。算法为:用户通过感知获得干扰信号的到达方向角,并构建空间谱图;根据空间谱图,深度强化学习中的深度神经网络拟合得到对应的Q值函数并输出每个动作的Q值;根据Q值,用户依概率进行信号消除方向的决策,根据该决策的回报值和下一步环境状态对深度神经网络进行训练并更新网络权重,优化决策策略;当达到最大循环次数时,算法结束。



1. 一种基于多天线抗干扰通信模型,其特征在於:对该模型做如下刻画:一个天线数为 N_t 的发射端和一个天线数为 N_r 接收端的用户进行通信,一个天线数为 N_t 的无人机干扰对用户通信进行干扰;在动态未知环境中,将接收端获取的具有无人机干扰信号到达方向角信息的动态空间谱图作为学习算法的输入状态,获取无人机干扰的空间运动特征,并基于抗干扰波束成形方法准确地对干扰信号进行消除。

2. 根据权利要求1所述的基于多天线抗干扰通信模型,其特征在於,考虑用物理信道模型描述多天线信号传输,令 $\varphi_{a,b}$ 和 $\theta_{a,b}$ 分别表示信号从a节点到b节点的信号离去角和信号到达角, $\mu_{a,b}$ 为信号传播的路径损耗,则a节点到b节点之间的信道 $H_{a,b}$ 如公式(1)所示:

$$\mathbf{H}_{a,b} = \mu_{a,b} e^{-j2\pi d_{a,b}/\lambda_c} \mathbf{e}_b(\theta_{a,b}) \mathbf{e}_a^H(\varphi_{a,b}) \quad (1)$$

其中 $d_{a,b}$ 为发送天线到接收天线之间的距离, λ_c 为载波波长, $\mathbf{e}_b(\theta_{a,b})$ 和 $\mathbf{e}_a(\varphi_{a,b})$ 分别为接收机和发送机的天线阵列响应矢量, $\mathbf{e}_a^H(\varphi_{a,b})$ 为 $\mathbf{e}_a(\varphi_{a,b})$ 的共轭转置,天线阵列响应矢量的表达式如公式(2)所示:

$$\begin{aligned} \mathbf{e}_b(\theta_{a,b}) &= \left[1, e^{-j2\pi\Delta_b \cos\theta_{a,b}}, \dots, e^{-j2\pi(N_b-1)\Delta_b \cos\theta_{a,b}} \right]^T \\ \mathbf{e}_a(\varphi_{a,b}) &= \left[1, e^{-j2\pi\Delta_a \cos\varphi_{a,b}}, \dots, e^{-j2\pi(N_a-1)\Delta_a \cos\varphi_{a,b}} \right]^T \end{aligned} \quad (2)$$

其中 j 为虚数因子, e 为自然常数, Δ_a 和 Δ_b 分别为天线之间的物理间隔, N_a 和 N_b 分别为a节点和b节点的天线数, x^T 为对 x 取转置运算;在信号传输阶段,接收机接收到的信号 y 如公式(3)所示:

$$\mathbf{y} = \sqrt{P_t} \mathbf{H}_{t,r} \mathbf{w}_t s_t + \sqrt{P_j} \mathbf{H}_{j,r} \mathbf{w}_j s_j + \mathbf{n} \quad (3)$$

其中 P_t 和 P_j 分别为发射端和干扰机的发送功率, $H_{t,r}$ 和 $H_{j,r}$ 分别表示表示发射端到接收端的和干扰端到接收端的信道系数矩阵, w_t 和 w_j 分别为发送机和干扰机的发送波束成形向量(预编码), s_t 和 s_j 分别为用户和干扰发送的符号,其功率满足 $\mathbb{E}[|s_t|^2] = \mathbb{E}[|s_j|^2] = 1$,其中 $\mathbb{E}[x]$ 为对 x 求期望, \mathbf{n} 表示加性高斯白噪声矢量,其元素服从独立同分布的均值为0、方差为 σ_n^2 的复高斯分布;

接收机通过接受波束成形向量对信号进行处理,获得接收端的信干噪比SINR如公式(4)所示:

$$\lambda = \frac{\mathbb{E} \left[\left| \sqrt{P_t} \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t s_t \right|^2 \right]}{\mathbb{E} \left[\left| \sqrt{P_j} \mathbf{f}^H \mathbf{H}_{j,r} \mathbf{w}_j s_j + \mathbf{f}^H \mathbf{n} \right|^2 \right]} = \frac{P_t \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \left(P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I} \right) \mathbf{f}} \quad (4)$$

其中, $\mathbb{E}[x]$ 为对随机变量 x 求期望, $|x|$ 为对 x 取模运算, x^H 表示对 x 取共轭转置; \mathbf{f} 为接收端的接收波束成形向量(滤波), \mathbf{I} 代表单位对角阵;令 $\mathbf{R}_j = P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I}$,根据公式(4)可知接收到的干扰加噪声的功率为 $\mathbf{f}^H \mathbf{R}_j \mathbf{f}$;定义传输速率 r 如公式(5)所示:

$$r = \begin{cases} 0, \lambda < \lambda_{th} \\ \log_2 \left(1 + \frac{P \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \mathbf{R}_j \mathbf{f}} \right), \lambda \geq \lambda_{th} \end{cases} \quad (5)$$

用 λ_{th} 表示成功解调信号所需要的SINR门限。

3. 根据权利要求1所述的基于多天线抗干扰通信模型,其特征在于,用户接收端通过感知干扰信号并估计干扰加噪声的自相关矩阵,根据估计出的自相关矩阵获得最优的抗干扰波束成形方法,具体如下:

对于抗干扰通信而言,用户的优化目标如公式(6)所示:

$$\begin{aligned} \min_{\mathbf{f}} \mathbf{f}^H \mathbf{R}_j \mathbf{f} \\ \text{s.t. } \mathbf{f}^H \mathbf{e}_r(\theta_{t,r}) = 1 \end{aligned} \quad (6)$$

其中 $\mathbf{e}_r(\theta_{t,r}) = [1, e^{-j2\pi\Delta_r \cos\theta_{t,r}}, \dots, e^{-j2\pi(N_r-1)\Delta_r \cos\theta_{t,r}}]^T$ 为接收端天线阵列响应矢量,

$\theta_{t,r}$ 为发送端发射的信号到接收端的信号到达角, Δ_r 为接收端天线之间的间隔;公式(6)的目标为设计最优的抗干扰接收波束成形向量,使得在保证接收到的有用信号功率一定的条件下,最小化接收到的干扰信号;

当用户发送端不发射信号时,接收端接收到的干扰与噪声叠加的信号为 $\mathbf{y}_j = \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{s}_j + \mathbf{n}$,那么干扰噪声叠加信号的自相关矩阵为 $\mathbf{R}_j = \mathbb{E}[\mathbf{y}_j \mathbf{y}_j^H]$;通过对接收到的信号进行快拍采样,得到估计的自相关矩阵 $\hat{\mathbf{R}}_j$ 为 $\hat{\mathbf{R}}_j = \frac{1}{K} \sum_{k=1}^K \mathbf{y}_j(k) \mathbf{y}_j(k)^H$,其中K为快拍数,k为第k次采样;利用拉格朗日乘数法,对公式(6)进行求解,求得最优的抗干扰接收波束成形向量为如公式(7)所示:

$$\mathbf{f}_{optimal} = \frac{\mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})}{\mathbf{e}_r^H(\theta_{t,r}) \mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})} \quad (7)$$

其中 $\mathbf{e}_r(\theta_{t,r})$ 为接收端天线阵列响应矢量。接收端的波束成形向量确定之后,发送端的发送波束成形向量则由最大比传输方案获得,即如公式(8)所示:

$$\mathbf{w}_t = \mathbf{H}_{t,r}^H \mathbf{f} / \|\mathbf{H}_{t,r}^H \mathbf{f}\| \quad (8)$$

其中 $\|x\|$ 为对x进行欧几里得范数运算。

4. 根据权利要求1所述的基于多天线抗干扰通信模型,其特征在于,无人机干扰的高速运动导致干扰到达用户接收端的信号到达角快速变化,进行了“感知-传输-学习-决策”的通信时隙一体化设计,具体如下:

用户通信以时隙的方式实施,一个通信时隙划分成四个阶段:感知阶段、传输数据阶段、学习阶段、ACK传输阶段,感知阶段,用户发射端保持沉默,接收端通过感知干扰信号来估计干扰加噪声的自相关矩阵;传输数据阶段,发射端向接收端发送数据,接收端利用上一时隙决策的抗干扰接收波束成形向量接收信号;学习阶段,接收端运行抗干扰方法并产生下一时刻接收端的抗干扰接收波束成形向量;ACK传输阶段,接收端向发送端发送信号是否

成功接收的确认信息,并携带接收端的决策信息,以供发送端生成发送波束成形向量。

5. 根据权利要求1所述的基于多天线抗干扰通信模型,其特征在于,具有无人机干扰信号到达方向角信息的动态空间谱图包含了一定历史时间长度内干扰信号的到达角度信息,具体构建方法如下:

对于估计出的信号加噪声自相关矩阵 $\hat{\mathbf{R}}_j$,利用特征值分解可以将该自相关矩阵表示为 $\hat{\mathbf{R}}_j = \hat{\mathbf{U}}_j \mathbf{\Lambda}_j \hat{\mathbf{U}}_j^H + \sigma_n^2 \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H$,其中 $\hat{\mathbf{U}}_j$ 和 $\hat{\mathbf{U}}_n$ 分别表示估计的干扰信号子空间和噪声子空间, $\mathbf{\Lambda}_j$ 为干扰信号的特征值对角矩阵。由此,估计干扰信号到达角的方法如公式(9)所示:

$$\hat{\theta}_{j,r} = \arg \max_{\theta_{j,r}} s(\theta_{j,r}) = \arg \max_{\theta_{j,r}} \frac{\mathbf{e}_r^H(\theta_{j,r}) \mathbf{e}_r(\theta_{j,r})}{\mathbf{e}_r^H(\theta_{j,r}) \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{e}_r(\theta_{j,r})} \quad (9)$$

其中 $s(\theta_{j,r})$ 为当前时刻的干扰信号空间谱函数,该函数最大值对应的角度即为估计的信号到达角;

由于无人机干扰的移动,干扰信号到达角在时间上的变化具有相关性和连续性,因此构建具有时间、空间二维信息的动态空间谱图,即 $\mathbf{S}_k = [\mathbf{S}_k, \mathbf{S}_{k-1}, \dots, \mathbf{S}_{k-L+2}, \mathbf{S}_{k-L+1}]^T$,其中 \mathbf{S}_k 为维度为 $L \times S$ 的动态空间谱矩阵, S 角度分辨率, k 为第 k 个时隙, \mathbf{S}_k 包含 L 个历史时隙的空间谱信息;随着无人机的飞行和接收端的感知与估计,动态空间谱矩阵进行持续动态地更新,记录新估计的角度信息,去除超过 L 个时隙的历史信息。

6. 一种基于权利要求1-5任一项所述的基于多天线抗干扰通信模型的动态空间谱抗干扰方法,其特征在于,包括以下步骤:

步骤1,初始化:以随机权重构建深度强化学习神经网络 Q ,记其权重为 ϕ ,通过完全复制的方式构建与网络 Q 完全相同的目标网络 \hat{Q} ,记其权重为 $\hat{\phi}$;假设干扰信号到达角 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 出现的集合为 Θ (如 $\Theta = \{0^\circ, 1^\circ, \dots, 180^\circ\}$),利用公式(1)、(7)、(8),对所有的 $\theta_{j,r} \in \Theta$ 生成动作集 $\mathcal{A} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$,其中 \mathbf{f}_x 为对应干扰到达角求得的干扰滤波向量,生成容量为 M 的经验池 \mathcal{D} ;

步骤2,依据以下准则选择一个动作 a_k :当迭代次数小于随机探索次数门限 M ,随机选择动作集中的动作;当迭代次数大于 M 小于训练次数门限,以概率 $0 \leq \varepsilon \leq 1$ 随机选择一个动作 $a_k \in \mathcal{A}$,概率 $1-\varepsilon$ 以贪婪准则选择 Q 值最大的动作 $a_k = \arg \max_a Q(\mathbf{S}_k, a; \phi)$;当迭代次数大于训练次数门限,以贪婪准则选择 Q 值最大的动作;

步骤3,接收机感知当前空间谱信息 \mathbf{S}_{k+1} ,用动作 a_k 作为接收波束成形向量接收信号,即 $\mathbf{f} = a_k$ 带入到公式(5)中,获得通信吞吐量作为回报值 r_k ;

步骤4,更新动态空间谱状态 $\mathbf{S}_{k+1} = (\mathbf{S}_{k+1}, \mathbf{S}_k)$,将经验存入经验池 \mathcal{D} 中;

步骤5,判断训练次数是否小于训练门限:如果小于则从经验池 \mathcal{D} 中随机批量采样经验,通过最小化损失函数 $L_k(\phi_k)$ 来更新网络 Q 的权重 ϕ ,且每隔 C 次迭代将网络 Q 的权重复制给目标网络 \hat{Q} ,更新概率 $\varepsilon = \max(0.01, \varepsilon - \Delta \varepsilon)$,其中 $\Delta \varepsilon$ 为步长;如果大于则步骤2,进行步骤2~6循环迭代,直到达到最大循环数,结束算法。

7. 根据权利要求6所述的动态空间谱抗干扰方法,其特征在于,上述步骤1中所述的以

不同的干扰信号到达角 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 生成动作集 $\mathcal{A}=\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$,具体如下:假设干扰信号到达角从 0° 到 180° ,每隔 δ° 生成一个接收波束成形向量作为动作。

8. 根据权利要求6所述的基于深度强化学习的动态空间谱抗干扰方法,步骤2中所述的以贪婪准则选择Q值最大的动作,具体如下:

深度强化学习神经网络Q是用来拟合Q值函数的,该函数的输入为当前获取的环境状态,输出为在当前状态下对每个动作的评价,即每个动作的长期累积折扣回报值(Q值),拟合的Q函数表示如下:

$$Q(\mathbf{S}_k, a_k) = \mathbb{E}\{r_k + \gamma \max_{a_{k+1}} Q(\mathbf{S}_{k+1}, a_{k+1}) | \mathbf{S}_k, a_k\} \quad (10)$$

其中 γ 为折扣因子。

9. 根据权利要求6所述的基于深度强化学习的动态空间谱抗干扰方法,步骤4中所述的将经验存入经验池 \mathcal{D} 中,具体如下:

一个经验定义为一个四元组,包括当前环境状态 S_k 、所做动作 a_k 、获得的回报值 r_k 和更新后的环境状态 S_{k+1} ,即将 k 时刻获得的信息 (S_k, a_k, r_k, S_{k+1}) 定义为一个经验;每次做一次决策,所获得的一个经验存储到经验池中,若经验池存储的经验数量已经达到容量值 M ,则将存储时间超过 M 的经验删除而存储新的经验。

10. 根据权利要求6所述的基于深度强化学习的动态空间谱抗干扰方法,步骤5中所述的从经验池 \mathcal{D} 中随机批量采样经验,通过最小化估计Q值与目标Q值的距离来训练网络Q的权重 ϕ ,具体如下:

训练时,采用经验值回放机制,随机选取一定数量的经验组成一组训练数据;第 k 次迭代时,估计Q值与目标Q值的距离即损失函数表示如下:

$$L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[(y_k - Q(\mathbf{S}_k, a_k; \phi_k))^2 \right] \quad (11)$$

其中 \mathcal{D}_s 为随机选取的存储在经验池中的多个经验, $Q(\mathbf{S}_k, a_k; \phi_k)$ 表示第 k 次迭代神经网络权重为 ϕ_k 时的Q值,为Q网络的输出,目标Q值 y_k 为 $y_k = r_k + \gamma \max_{a'} Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$,其中 $Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$ 为目标网络 \hat{Q} 的输出;

目标是使估计Q值与目标Q值的距离最小,通过对公式(11)进行梯度下降算法,可得参数 ϕ_k 的梯度表示如下:

$$\nabla_{\phi_k} L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[(y_k - Q(\mathbf{S}_k, a_k; \phi_k)) \nabla_{\phi_k} Q(\mathbf{S}_k, a_k; \phi_k) \right] \quad (12)$$

其中 ∇_{ϕ_k} 为梯度运算符。

基于多天线抗干扰通信模型及动态空间谱抗干扰方法

技术领域

[0001] 本发明属于无线通信抗干扰技术领域,包括一种基于多天线抗干扰通信模型和一种基于深度强化学习的动态空间谱抗干扰方法。

背景技术

[0002] 利用天线增益与分集增益,多输入多输出(MIMO)通信在信道状态信息已知的前提下可以有效地抑制干扰信号。随着无人机应用的普及,高机动性的无人机干扰可以在短时间内压制大片区域,对军用和民用无线通信造成严重威胁。然而传统的MIMO抗干扰通信方法大多针对静态干扰。由于无人机干扰的高速移动与信号处理的时延,造成估计的干扰信道的状态信息过时,此时传统的MIMO抗干扰通信方法无法适用。因此,迫切需要研究对抗高机动干扰机的通信方法。

[0003] 人工智能中强化学习方法(参考文献:C.J.Watkins and P.Dayan,“Q-learning,” Machine learning, vol.8, no.3-4, pp.279-292, 1992)可以在动态未知的环境中学习环境变化规律,获取最优的决策策略。通过强化学习方法,在无人机飞行特性(如轨迹、速度、高度等)先验信息未知的前提下,学习干扰信号到达角的变化规律,将是一种可行的方案。然而,由于不同的无人机飞行特性将导致不同的干扰信号变化规律,巨大的状态空间将使得强化学习无法收敛。

[0004] 针对以上问题,本发明创新性地利用深度强化学习算法解决干扰机快速移动场景下的MIMO抗干扰通信。利用深度学习强大的函数拟合能力对巨大状态空间进行特征提取和泛化,通过强化学习在动态未知的环境中获取最优抗干扰策略,有效提升MIMO通信对抗移动干扰的能力。

发明内容

[0005] 本发明的目的在于提供一种基于多天线抗干扰通信模型和一种基于深度强化学习的动态空间谱抗干扰方法,很好地刻画了面向移动干扰的MIMO抗干扰通信场景。

[0006] 实现本发明目的的技术解决方案为:一种基于多天线抗干扰通信模型,对该模型做如下刻画:一种基于多天线抗干扰通信模型,其特征在于,对该模型做如下刻画:一个天线数为 N_t 的发射端和一个天线数为 N_r 接收端的用户进行通信,一个天线数为 N_t 的无人机干扰对用户通信进行干扰;在动态未知环境中,将接收端获取的具有无人机干扰信号到达方向角信息的动态空间谱图作为学习算法的输入状态,获取无人机干扰的空间运动特征,并基于抗干扰波束成形方法准确地对干扰信号进行消除。

[0007] 在用户发射端与干扰机同时发送信号时用户接收端的信干噪比SINR表示为:

$$[0008] \quad \text{SINR} = \frac{\mathbb{E} \left[\left| \sqrt{P_t} \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t s_t \right|^2 \right]}{\mathbb{E} \left[\left| \sqrt{P_j} \mathbf{f}^H \mathbf{H}_{j,r} \mathbf{w}_j s_j + \mathbf{f}^H \mathbf{n} \right|^2 \right]} = \frac{P_t \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \left(P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I} \right) \mathbf{f}} \quad (1)$$

[0009] 公式中 (9) 中, $\mathbb{E}[\cdot]$ 为对随机变量求期望, P_t 和 P_j 分别为发射端和干扰机的发送功率, \mathbf{f} 为接收端的接收波束成形向量 (滤波), $\mathbf{H}_{t,r}$ 和 $\mathbf{H}_{j,r}$ 分别表示表示发射端到接收端的信道系数矩阵, \mathbf{w}_t 和 \mathbf{w}_j 分别为发送机和干扰机的发送波束成形向量 (预编码), s_t 和 s_t 分别为用户和干扰发送的符号, \mathbf{n} 表示加性高斯白噪声矢量, 其元素服从独立同分布的均值为 0、方差为 σ_n^2 的复高斯分布, \mathbf{I} 代表单位对角阵。

[0010] 此外, 用 λ 表示成功解调信号所需要的 SINR 门限, 定义传输速率 r 如公式 (10) 所示:

$$[0011] \quad r = \begin{cases} 0, \text{SINR} < \lambda \\ \log_2 \left(1 + \frac{P_t \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \mathbf{R}_j \mathbf{f}} \right), \text{SINR} \geq \lambda \end{cases} \quad (2)$$

[0012] 用户接收端通过感知干扰信号并估计干扰加噪声的自相关矩阵, 根据估计出的自相关矩阵获得最优的抗干扰波束成形方法。令 $\mathbf{R}_j = P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I}$, 根据公式 (9) 可知接收到的干扰加噪声的功率为 $\mathbf{f}^H \mathbf{R}_j \mathbf{f}$ 。对于抗干扰通信而言, 用户的优化目标如公式 (11) 所示:

$$[0013] \quad \begin{aligned} & \min_{\mathbf{f}} \mathbf{f}^H \mathbf{R}_j \mathbf{f} \\ & \text{s.t. } \mathbf{f}^H \mathbf{e}_r(\theta_{t,r}) = 1 \end{aligned} \quad (3)$$

[0014] 其中 $\mathbf{e}_r(\theta_{t,r}) = [1, e^{-j2\pi\Delta_r \cos\theta_{t,r}}, \dots, e^{-j2\pi(N_r-1)\Delta_r \cos\theta_{t,r}}]^T$ 为接收端天线阵列响应矢量, $\theta_{t,r}$ 为发送端发射的信号到接收端的信号到达角, Δ_r 为接收端天线之间的间隔。公式 (11) 的目标为设计最优的抗干扰接收波束成形向量, 使得在保证接收到的有用信号功率一定的条件下, 最小化接收到的干扰信号。

[0015] 当用户发送端不发射信号时, 接收端接收到的干扰与噪声叠加的信号为 $\mathbf{y}_j = \mathbf{H}_{j,r} \mathbf{w}_j s_j + \mathbf{n}$, 那么干扰噪声叠加信号的自相关矩阵为 $\mathbf{R}_j = \mathbb{E}[\mathbf{y}_j \mathbf{y}_j^H]$ 。通过对接收到的信号进行快拍采样, 可以得到估计的自相关矩阵为 $\hat{\mathbf{R}}_j = \frac{1}{K} \sum_{k=1}^K \mathbf{y}_j(k) \mathbf{y}_j(k)^H$, 其中 K 为快拍数。利用拉格朗日乘数法, 可以对公式 (11) 进行求解, 求得最优的抗干扰接收波束成形向量为如公式 (12) 所示:

$$[0016] \quad \mathbf{f}_{\text{optimal}} = \frac{\mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})}{\mathbf{e}_r^H(\theta_{t,r}) \mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})} \quad (4)$$

[0017] 接收端的波束成形向量确定之后, 发送端的发送波束成形向量则由最大比传输方案获得, 即 $\mathbf{w}_t = \mathbf{H}_{t,r}^H \mathbf{f} / \|\mathbf{H}_{t,r}^H \mathbf{f}\|$ 。

[0018] 进一步地, 针对无人机干扰的高速运动导致干扰到达用户接收端的信号到达角快速变化的问题, 进行了“感知-传输-学习-决策”的通信时隙一体化设计。用户通信以时隙的方式实施, 一个通信时隙划分成四个阶段: 感知、传输数据、学习、ACK 传输。感知阶段, 用户发射端保持沉默, 接收端通过感知干扰信号来估计干扰加噪声的自相关矩阵。传输数据阶段, 发射端向接收端发送数据, 接收端利用上一时隙决策的抗干扰接收波束成形向量接收

信号。学习阶段,接收端运行抗干扰方法并产生下一时刻接收端的抗干扰接收波束成形向量。ACK传输阶段,接收端向发送端发送信号是否成功接收的确认信息,并携带接收端的决策信息,以供发送端生成发送波束成形向量。

[0019] 进一步地,构建具有无人机干扰信号到达方向角信息的动态空间谱图。对于估计出的信号加噪声自相关矩阵 $\hat{\mathbf{R}}_j$, 利用特征值分解可以将该自相关矩阵表示为 $\hat{\mathbf{R}}_j = \hat{\mathbf{U}}_j \mathbf{\Lambda}_j \hat{\mathbf{U}}_j^H + \sigma_n^2 \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H$, 其中 $\hat{\mathbf{U}}_j$ 和 $\hat{\mathbf{U}}_n$ 分别表示估计的干扰信号子空间和噪声子空间, $\mathbf{\Lambda}_j$ 为干扰信号的特征值对角矩阵。由此,估计干扰信号到达角的方法如公式 (13) 所示:

$$[0020] \quad \hat{\theta}_{j,r} = \arg \max_{\theta_{j,r}} s(\theta_{j,r}) = \arg \max_{\theta_{j,r}} \frac{\mathbf{e}_r^H(\theta_{j,r}) \mathbf{e}_r(\theta_{j,r})}{\mathbf{e}_r^H(\theta_{j,r}) \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{e}_r(\theta_{j,r})} \quad (5)$$

[0021] 其中 $s(\theta_{j,r})$ 为当前时刻的干扰信号空间谱函数,该函数最大值对应的角度即为估计的信号到达角。

[0022] 由于无人机干扰的移动,干扰信号到达角在时间上的变化具有相关性和连续性,因此构建具有时间、空间二维信息的动态空间谱图,即 $\mathbf{S}_k = [\mathbf{S}_k, \mathbf{S}_{k-1}, \dots, \mathbf{S}_{k-L+2}, \mathbf{S}_{k-L+1}]^T$, 其中 \mathbf{S}_k 为维度为 $L \times S$ 的动态空间谱矩阵, S 角度分辨率, k 为第 k 个时隙, \mathbf{S}_k 包含 L 个历史时隙的空间谱信息。随着无人机的飞行和接收端的感知与估计,动态空间谱矩阵进行持续动态地更新,记录新估计的角度信息,去除超过 L 个时隙的历史信息。

[0023] 一种基于深度强化学习的动态空间谱抗干扰方法,其特征在于,包括以下步骤:

[0024] 步骤1,初始化:以随机权重 ϕ 构建深度强化学习神经网络 Q , 构建与网络 Q 完全相同的目标网络 \hat{Q} , 权重为 $\hat{\phi}$; 利用公式 (12), 以不同的 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 生成动作集 $\mathcal{A} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$; 生成容量为 M 的经验池 \mathcal{D} 。

[0025] 步骤2,依据以下准则选择一个动作 a_k : 当迭代次数小于随机探索次数门限 M , 随机选择动作集中的动作; 当迭代次数大于 M 小于训练次数门限, 以概率 $0 \leq \epsilon \leq 1$ 随机选择一个动作 $a_k \in \mathcal{A}$, 概率 $1 - \epsilon$ 以贪婪准则选择 Q 值最大的动作 $a_k = \arg \max_a Q(\mathbf{S}_k, a; \phi)$; 当迭代次数大于训练次数门限, 以贪婪准则选择 Q 值最大的动作。

[0026] 步骤3,接收机感知当前空间谱信息 \mathbf{S}_{k+1} , 用动作 a_k 作为接收波束成形向量接收信号, 获得回报值 r_k , 即公式 (10) 所示的通信吞吐量。

[0027] 步骤4,更新动态空间谱状态 $\mathbf{S}_{k+1} = (\mathbf{S}_{k+1}, \mathbf{S}_k)$, 将经验存入经验池 \mathcal{D} 中。

[0028] 步骤5,判断训练次数是否小于训练门限: 如果小于则从经验池 \mathcal{D} 中随机批量采样经验, 通过最小化估计 Q 值与目标 Q 值的距离来训练网络 Q 的权重 ϕ , 且每 C 次迭代将网络 Q 的权重复制给目标网络 \hat{Q} , 更新概率 $\epsilon = \max(0.01, \epsilon - \Delta \epsilon)$, 其中 $\Delta \epsilon$ 为步长; 如果大于则步骤2, 进行步骤2~6循环迭代, 直到达到最大循环数, 结束算法。

[0029] 进一步地,步骤1中所述的以不同的 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 生成动作集 $\mathcal{A} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$, 具体如下: 假设干扰信号到达角从 0° 到 180° , 每隔 δ° 生成一个接收波束成形向量作为动作。

[0030] 进一步地,步骤2中所述的以贪婪准则选择 Q 值最大的动作, 具体为深度强化学习神经网络 Q 是用来拟合 Q 值函数的, 该函数的输入为当前获取的环境状态, 输出为在当前状

态下对每个动作的评价,即每个动作的长期累积折扣回报值(Q值),拟合的Q函数表示如下:

$$[0031] \quad Q(\mathbf{S}_k, a_k) = \mathbb{E} \left\{ r_k + \gamma \max_{a_{k+1}} Q(\mathbf{S}_{k+1}, a_{k+1}) \mid \mathbf{S}_k, a_k \right\} \quad (6)$$

[0032] 其中 γ 为折扣因子。

[0033] 进一步地,步骤4中所述的将经验存入经验池 \mathcal{D} 中的具体方法为:定义一个经验为一个四元组,包括当前环境状态 \mathbf{S}_k 、所做动作 a_k 、获得的回报值 r_k 和更新后的环境状态 \mathbf{S}_{k+1} ,即将 k 时刻获得的信息 $(\mathbf{S}_k, a_k, r_k, \mathbf{S}_{k+1})$ 定义为一个经验。每次做一次决策,所获得的一个经验存储到经验池中,若经验池存储的经验数量已经达到容量值 M ,则将存储时间超过 M 的经验删除而存储新的经验。

[0034] 进一步地,步骤5中所述的从经验池 \mathcal{D} 中随机批量采样经验,通过最小化估计Q值与目标Q值的距离来训练网络Q的权重 ϕ ,具体的方法为:训练时,采用经验值回放机制,随机选取一定数量的经验组成一组训练数据 \mathcal{D}_s 。第 k 次迭代时,估计Q值与目标Q值的距离表示如下:

$$[0035] \quad L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[\left(y_k - Q(\mathbf{S}_k, a_k; \phi_k) \right)^2 \right] \quad (7)$$

[0036] 其中 $Q(\mathbf{S}_k, a_k; \phi_k)$ 表示第 k 次迭代神经网络权重为 ϕ_k 时的Q值,为Q网络的输出, $y_k = r_k + \gamma \max_{a'} Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$ 为目标Q值,其中 $Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$ 为目标网络 \hat{Q} 的输出。

[0037] 目标是使估计Q值与目标Q值的距离最小,通过对公式(15)进行梯度下降算法,可得参数 ϕ_k 的梯度表示如下:

$$[0038] \quad \nabla_{\phi_k} L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[\left(y_k - Q(\mathbf{S}_k, a_k; \phi_k) \right) \nabla_{\phi_k} Q(\mathbf{S}_k, a_k; \phi_k) \right] \quad (8)$$

[0039] 其中 ∇_{ϕ_k} 为梯度运算符。

[0040] 本发明与现有技术相比,其显著优点在于:(1)创新深度强化学习在MIMO抗干扰通信中的应用,解决传统MIMO抗干扰通信方法无法应对移动干扰的不足;(2)模型完备,物理意义清晰,提出的基于深度强化学习的动态空间谱抗干扰方法,实现对提出模型的有效求解,求出MIMO通信的抗干扰波束成形策略;(3)能够有效地应对移动干扰,并很好地刻画基于深度强化学习算法的抗干扰场景。

附图说明

[0041] 图1是本发明多天线抗干扰通信系统模型图。

[0042] 图2是本发明中深度强化学习神经网络模型图。

[0043] 图3是本发明中基于深度强化学习的动态空间谱抗干扰通信时隙结构图。

[0044] 图4是本发明实施例1中平均通信速率随着迭代次数变化的仿真图。

[0045] 图5是本发明实施例2中神经网络输出的平均Q值随着迭代次数变化的仿真图。

具体实施方式

[0046] 本发明所提出的基于多天线抗干扰通信模型和基于深度强化学习的动态空间谱抗干扰方法,旨在提供方案以解决移动干扰下的MIMO抗干扰通信问题。本发明提出的基于深度强化学习的动态空间谱抗干扰方法,将接收机感知得到的动态空间谱图作为算法的

输入,采用深度强化学习神经网络提取动态空间谱特征,获取干扰信号到达角变化规律,对策略函数即Q值函数进行拟合;接着,通过相应的决策算法,更新接收机的接收波束成形选择策略。

[0047] 图1是多天线抗干扰通信系统模型。该模型中,一个发射机向一个接收机发送数据,一个正在飞行的无人机干扰发送干扰信号阻碍通信。

[0048] 图2是深度强化学习神经网络模型。深度强化学习神经网络采用卷积神经网络框架,输入为动态空间谱图,通过两层卷积层提取动态空间谱图特征并降低图维度、两层全连接层拟合Q值函数,然后通过相应的决策算法输出当前空间谱状态下最优的抗干扰波束成形决策。

[0049] 图3是基于深度强化学习的动态空间谱抗干扰通信时隙结构图。用户通信以时隙的方式实施,一个通信时隙划分成四个阶段:感知、传输数据、学习、ACK传输。感知阶段,用户发射端保持沉默,接收端通过感知干扰信号来估计干扰加噪声的自相关矩阵。传输数据阶段,发射端向接收端发送数据,接收端利用上一时隙决策的抗干扰接收波束成形向量接收信号。学习阶段,接收端运行抗干扰方法并产生下一时刻接收端的抗干扰接收波束成形向量。ACK传输阶段,接收端向发送端发送信号是否成功接收的确认信息,并携带接收端的决策信息,以供发送端生成发送波束成形向量。

[0050] 本发明的目的在于提供一种基于多天线抗干扰通信模型和一种基于深度强化学习的动态空间谱抗干扰方法,很好地刻画了面向移动干扰的MIMO抗干扰通信场景。

[0051] 实现本发明目的的技术解决方案为:一种基于多天线抗干扰通信模型,对该模型做如下刻画:一种基于多天线抗干扰通信模型,其特征在于,对该模型做如下刻画:一个天线数为 N_t 的发射端和一个天线数为 N_r 接收端的用户进行通信,一个天线数为 N_t 的无人机干扰对用户通信进行干扰;在动态未知环境中,将接收端获取的具有无人机干扰信号到达方向角信息的动态空间谱图作为学习算法的输入状态,获取无人机干扰的空间运动特征,并基于抗干扰波束成形方法准确地对干扰信号进行消除。

[0052] 在用户发射端与干扰机同时发送信号时用户接收端的信干噪比SINR表示为:

$$[0053] \quad \text{SINR} = \frac{\mathbb{E} \left[\left| \sqrt{P_t} \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t s_t \right|^2 \right]}{\mathbb{E} \left[\left| \sqrt{P_j} \mathbf{f}^H \mathbf{H}_{j,r} \mathbf{w}_j s_j + \mathbf{f}^H \mathbf{n} \right|^2 \right]} = \frac{P_t \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \left(P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I} \right) \mathbf{f}} \quad (9)$$

[0054] 公式中(17)中, $\mathbb{E}[\cdot]$ 为对随机变量求期望, P_t 和 P_j 分别为发射端和干扰机的发送功率, \mathbf{f} 为接收端的接收波束成形向量(滤波), $\mathbf{H}_{t,r}$ 和 $\mathbf{H}_{j,r}$ 分别表示表示发射端到接收端的信道系数矩阵, \mathbf{w}_t 和 \mathbf{w}_j 分别为发送机和干扰机的发送波束成形向量(预编码), s_t 和 s_t 分别为用户和干扰发送的符号, \mathbf{n} 表示加性高斯白噪声矢量,其元素服从独立同分布的均值为0、方差为 σ_n^2 的复高斯分布, \mathbf{I} 代表单位对角阵。

[0055] 此外,用 λ 表示成功解调信号所需要的SINR门限,定义传输速率 r 如公式(18)所示:

$$[0056] \quad r = \begin{cases} 0, & \text{SINR} < \lambda \\ \log_2 \left(1 + \frac{P_t \mathbf{f}^H \mathbf{H}_{t,r} \mathbf{w}_t \mathbf{w}_t^H \mathbf{H}_{t,r}^H \mathbf{f}}{\mathbf{f}^H \mathbf{R}_j \mathbf{f}} \right), & \text{SINR} \geq \lambda \end{cases} \quad (10)$$

[0057] 用户接收端通过感知干扰信号并估计干扰加噪声的自相关矩阵,根据估计出的自相关矩阵获得最优的抗干扰波束成形方法。令 $\mathbf{R}_j = P_j \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_{j,r}^H + \sigma_n^2 \mathbf{I}$,根据公式(17)可知接收到的干扰加噪声的功率为 $\mathbf{f}^H \mathbf{R}_j \mathbf{f}$ 。对于抗干扰通信而言,用户的优化目标如公式(19)所示:

$$[0058] \quad \begin{aligned} & \min_{\mathbf{f}} \mathbf{f}^H \mathbf{R}_j \mathbf{f} \\ & \text{s.t. } \mathbf{f}^H \mathbf{e}_r(\theta_{t,r}) = 1 \end{aligned} \quad (11)$$

[0059] 其中 $\mathbf{e}_r(\theta_{t,r}) = [1, e^{-j2\pi\Delta_r \cos\theta_{t,r}}, \dots, e^{-j2\pi(N_r-1)\Delta_r \cos\theta_{t,r}}]^T$ 为接收端天线阵列响应矢量, $\theta_{t,r}$ 为发送端发射的信号到接收端的信号到达角, Δ_r 为接收端天线之间的间隔。公式(19)的目标为设计最优的抗干扰接收波束成形向量,使得在保证接收到的有用信号功率一定的条件下,最小化接收到的干扰信号。

[0060] 当用户发送端不发射信号时,接收端接收到的干扰与噪声叠加的信号为 $\mathbf{y}_j = \mathbf{H}_{j,r} \mathbf{w}_j \mathbf{s}_j + \mathbf{n}$,那么干扰噪声叠加信号的自相关矩阵为 $\mathbf{R}_j = \mathbb{E}[\mathbf{y}_j \mathbf{y}_j^H]$ 。通过对接收到的信号进行快拍采样,可以得到估计的自相关矩阵为 $\hat{\mathbf{R}}_j = \frac{1}{K} \sum_{k=1}^K \mathbf{y}_j(k) \mathbf{y}_j(k)^H$,其中K为快拍数。利用拉格朗日乘数法,可以对公式(19)进行求解,求得最优的抗干扰接收波束成形向量为如公式(20)所示:

$$[0061] \quad \mathbf{f}_{\text{optimal}} = \frac{\mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})}{\mathbf{e}_r^H(\theta_{t,r}) \mathbf{R}_j^{-1} \mathbf{e}_r(\theta_{t,r})} \quad (12)$$

[0062] 接收端的波束成形向量确定之后,发送端的发送波束成形向量则由最大比传输方案获得,即 $\mathbf{w}_t = \mathbf{H}_{t,r}^H \mathbf{f} / \|\mathbf{H}_{t,r}^H \mathbf{f}\|$ 。

[0063] 进一步地,针对无人机干扰的高速运动导致干扰到达用户接收端的信号到达角快速变化的问题,进行了“感知-传输-学习-决策”的通信时隙一体化设计。用户通信以时隙的方式实施,一个通信时隙划分成四个阶段:感知、传输数据、学习、ACK传输。感知阶段,用户发射端保持沉默,接收端通过感知干扰信号来估计干扰加噪声的自相关矩阵。传输数据阶段,发射端向接收端发送数据,接收端利用上一时隙决策的抗干扰接收波束成形向量接收信号。学习阶段,接收端运行抗干扰方法并产生下一时刻接收端的抗干扰接收波束成形向量。ACK传输阶段,接收端向发送端发送信号是否成功接收的确认信息,并携带接收端的决策信息,以供发送端生成发送波束成形向量。

[0064] 进一步地,构建具有无人机干扰信号到达方向角信息的动态空间谱图。对于估计出的信号加噪声自相关矩阵 $\hat{\mathbf{R}}_j$,利用特征值分解可以将该自相关矩阵表示为 $\hat{\mathbf{R}}_j = \hat{\mathbf{U}}_j \mathbf{\Lambda}_j \hat{\mathbf{U}}_j^H + \sigma_n^2 \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H$,其中 $\hat{\mathbf{U}}_j$ 和 $\hat{\mathbf{U}}_n$ 分别表示估计的干扰信号子空间和噪声子空间, $\mathbf{\Lambda}_j$ 为干扰信号的特征值对角矩阵。由此,估计干扰信号到达角的方法如公式(21)所示:

$$[0065] \quad \hat{\theta}_{j,r} = \arg \max_{\theta_{j,r}} \mathbf{s}(\theta_{j,r}) = \arg \max_{\theta_{j,r}} \frac{\mathbf{e}_r^H(\theta_{j,r}) \mathbf{e}_r(\theta_{j,r})}{\mathbf{e}_r^H(\theta_{j,r}) \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{e}_r(\theta_{j,r})} \quad (13)$$

[0066] 其中 $s(\theta_{j,r})$ 为当前时刻的干扰信号空间谱函数,该函数最大值对应的角度即为估计的信号到达角。

[0067] 由于无人机干扰的移动,干扰信号到达角在时间上的变化具有相关性和连续性,因此构建具有时间、空间二维信息的动态空间谱图,即 $S_k = [S_k, S_{k-1}, \dots, S_{k-L+2}, S_{k-L+1}]^T$,其中 S_k 为维度为 $L \times S$ 的动态空间谱矩阵, S 角度分辨率, k 为第 k 个时隙, S_k 包含 L 个历史时隙的空间谱信息。随着无人机的飞行和接收端的感知与估计,动态空间谱矩阵进行持续动态地更新,记录新估计的角度信息,去除超过 L 个时隙的历史信息。

[0068] 一种基于深度强化学习的动态空间谱抗干扰方法,其特征在于,包括以下步骤:

[0069] 步骤1,初始化:以随机权重 ϕ 构建深度强化学习神经网络 Q ,构建与网络 Q 完全相同的目标网络 \hat{Q} ,权重为 $\hat{\phi}$;利用公式(20),以不同的 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 生成动作集 $\mathcal{A} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$;生成容量为 M 的经验池 \mathcal{D} 。

[0070] 步骤2,依据以下准则选择一个动作 a_k :当迭代次数小于随机探索次数门限 M ,随机选择动作集中的动作;当迭代次数大于 M 小于训练次数门限,以概率 $0 \leq \epsilon \leq 1$ 随机选择一个动作 $a_k \in \mathcal{A}$,概率 $1-\epsilon$ 以贪婪准则选择 Q 值最大的动作 $a_k = \arg \max_a Q(S_k, a; \phi)$;当迭代次数大于训练次数门限,以贪婪准则选择 Q 值最大的动作。

[0071] 步骤3,接收机感知当前空间谱信息 S_{k+1} ,用动作 a_k 作为接收波束成形向量接收信号,获得回报值 r_k ,即公式(18)所示的通信吞吐量。

[0072] 步骤4,更新动态空间谱状态 $S_{k+1} = (S_{k+1}, S_k)$,将经验存入经验池 \mathcal{D} 中。

[0073] 步骤5,判断训练次数是否小于训练门限:如果小于则从经验池 \mathcal{D} 中随机批量采样经验,通过最小化估计 Q 值与目标 Q 值的距离来训练网络 Q 的权重 ϕ ,且每 C 次迭代将网络 Q 的权重复制给目标网络 \hat{Q} ,更新概率 $\epsilon = \max(0.01, \epsilon - \Delta \epsilon)$,其中 $\Delta \epsilon$ 为步长;如果大于则步骤2,进行步骤2~6循环迭代,直到达到最大循环数,结束算法。

[0074] 进一步地,步骤1中所述的以不同的 $\theta_{j,r} \in (0^\circ, 180^\circ]$ 生成动作集 $\mathcal{A} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{|\mathcal{A}|}\}$,具体如下:假设干扰信号到达角从 0° 到 180° ,每隔 δ° 生成一个接收波束成形向量作为动作。

[0075] 进一步地,步骤2中所述的以贪婪准则选择 Q 值最大的动作,具体为深度强化学习神经网络 Q 是用来拟合 Q 值函数的,该函数的输入为当前获取的环境状态,输出为在当前状态下对每个动作的评价,即每个动作的长期累积折扣回报值(Q 值),拟合的 Q 函数表示如下:

$$[0076] \quad Q(S_k, a_k) = \mathbb{E} \left\{ r_k + \gamma \max_{a_{k+1}} Q(S_{k+1}, a_{k+1}) \mid S_k, a_k \right\} \quad (14)$$

[0077] 其中 γ 为折扣因子。

[0078] 进一步地,步骤4中所述的将经验存入经验池 \mathcal{D} 中的具体方法为:定义一个经验为一个四元组,包括当前环境状态 S_k 、所做动作 a_k 、获得的回报值 r_k 和更新后的环境状态 S_{k+1} ,即将 k 时刻获得的信息 (S_k, a_k, r_k, S_{k+1}) 定义为一个经验。每次做一次决策,所获得的一个经验存储到经验池中,若经验池存储的经验数量已经达到容量值 M ,则将存储时间超过 M 的经验删除而存储新的经验。

[0079] 进一步地,步骤5中所述的从经验池 \mathcal{D} 中随机批量采样经验,通过最小化估计 Q 值与目标 Q 值的距离来训练网络 Q 的权重 ϕ ,具体的方法为:训练时,采用经验值回放机制,随

机选取一定数量的经验组成一组训练数据 \mathcal{D}_s 。第k次迭代时,估计Q值与目标Q值的距离表示如下:

$$[0080] \quad L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[(y_k - Q(\mathbf{S}_k, a_k; \phi_k))^2 \right] \quad (15)$$

[0081] 其中 $Q(\mathbf{S}_k, a_k; \phi_k)$ 表示第k次迭代神经网络权重为 ϕ_k 时的Q值,为Q网络的输出, $y_k = r_k + \gamma \max_{a'} Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$ 为目标Q值,其中 $Q(\mathbf{S}_{k+1}, a'; \hat{\phi}_k)$ 为目标网络 \hat{Q} 的输出。

[0082] 目标是使估计Q值与目标Q值的距离最小,通过对公式(23)进行梯度下降算法,可得参数 ϕ_k 的梯度表示如下:

$$[0083] \quad \nabla_{\phi_k} L_k(\phi_k) = \mathbb{E}_{\mathcal{D}_s} \left[(y_k - Q(\mathbf{S}_k, a_k; \phi_k)) \nabla_{\phi_k} Q(\mathbf{S}_k, a_k; \phi_k) \right] \quad (16)$$

[0084] 其中 ∇_{ϕ_k} 为梯度运算符。

[0085] 实施例1

[0086] 本发明的第一个实施例具体描述如下,系统仿真采用Python语言,基于TensorFlow的深度学习框架,参数设定不影响一般性。该实施例验证所提模型与算法的有效性,参数设置为,无人机飞行高度30米,速度为50米每秒,发射机发送的有用信号到接收机的到达角 $\theta_{t,r}$ 固定为30度,无人机干扰和发射机的发射功率分别为20dB和10dB,噪声的方差为1,解调门限为 $\lambda = 5\text{dB}$,另发射机、接收机和干扰的天线数相等。在一个通信时隙中, T_s 、 T_D 、 T_L 和 T_{ACK} 分别设置为0.03秒、0.05秒、0.01秒和0.01秒。算法的折扣因子 $\gamma = 0.8$,动作数为 $|\mathcal{A}| = 60$,状态 \mathbf{S}_k 的维度为 10×180 其中历史长度 $L = 10$,经验池 \mathcal{D} 的容量 $M = 5000$ 。总迭代次数为 1.5×10^5 ,其中前5000次为随机探索动作,最后55000次为贪婪选择动作,中间的90000次为训练神经网络。目标网络的权重每 $C = 100$ 次进行更新。

[0087] 图4是本发明实施例1中在不同的天线数下平均通信速率随着迭代次数变化的仿真图,从图中可以看出,在算法的初始阶段接收机随机探索波束成形向量,积累经验。随着迭代次数的增加,接收机选择获得高回报动作的概率逐渐增加,通信速率也因此上升,直到训练阶段结束,算法趋于收敛,接收机获得稳定的抗干扰波束成形策略。同时,随着天线数的增加,抗干扰性能也提高。该仿真图验证了算法抗干扰的有效性。

[0088] 实施例2

[0089] 本发明的第二个实施例具体描述如下,系统仿真采用Python语言,基于TensorFlow的深度学习框架,参数设定不影响一般性。该实施例验证所提模型与算法的有效性,参数设置为,无人机飞行高度30米,速度为50米每秒,发射机发送的有用信号到接收机的到达角 $\theta_{t,r}$ 固定为30度,无人机干扰和发射机的发射功率分别为20dB和10dB,噪声的方差为1,解调门限为 $\lambda = 5\text{dB}$,另发射机、接收机和干扰的天线数相等。在一个通信时隙中, T_s 、 T_D 、 T_L 和 T_{ACK} 分别设置为0.03秒、0.05秒、0.01秒和0.01秒。算法的折扣因子 $\gamma = 0.8$,动作数为 $|\mathcal{A}| = 60$,状态 \mathbf{S}_k 的维度为 10×180 其中历史长度 $L = 10$,经验池 \mathcal{D} 的容量 $M = 5000$ 。总迭代次数为 1.5×10^5 ,其中前5000次为随机探索动作,最后55000次为贪婪选择动作,中间的90000次为训练神经网络。目标网络的权重每 $C = 100$ 次进行更新。

[0090] 图5是本发明实施例2中神经网络输出的平均Q值随着迭代次数变化的仿真图,随

着迭代次数的增加,神经网络的训练使得其拟合的函数逐渐趋近于最优的Q值函数,最终拟合函数的过程收敛。该仿真图验证了算法的收敛性。

[0091] 综上所述,本发明提出的基于多天线抗干扰通信模型充分考虑了MIMO通信中干扰的移动性、通信环境的动态未知性和状态空间的复杂性等问题,比传统的MIMO抗干扰通信模型更有实际意义;提出的基于深度强化学习的动态空间谱抗干扰方法,能够实现对提出模型的有效求解,求出接收机的抗干扰波束成形向量选择策略,有效地应对高机动性的干扰。

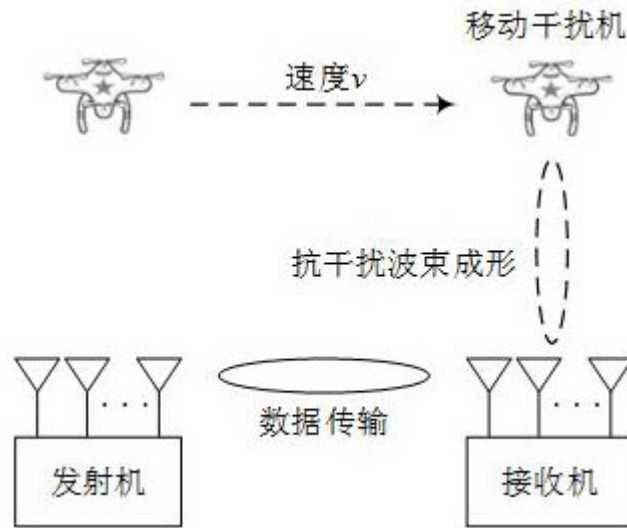


图1

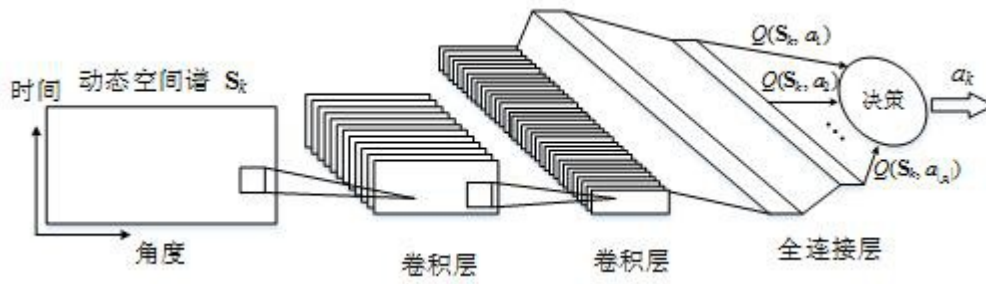


图2

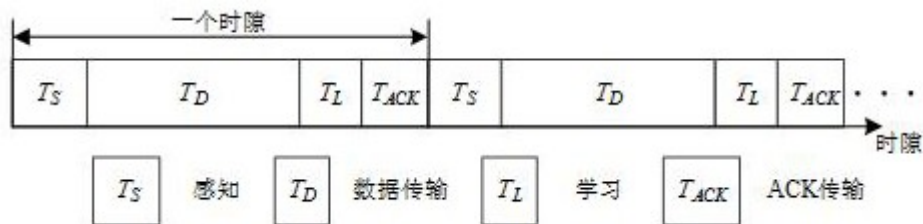


图3

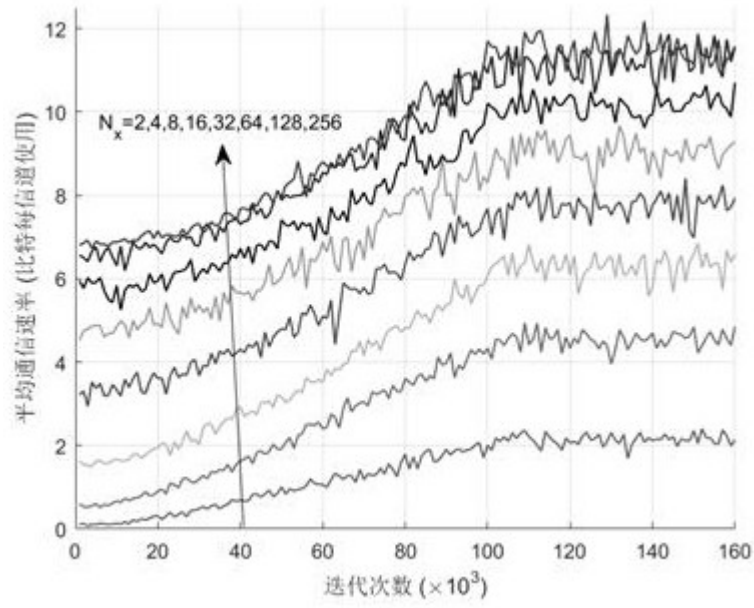


图4

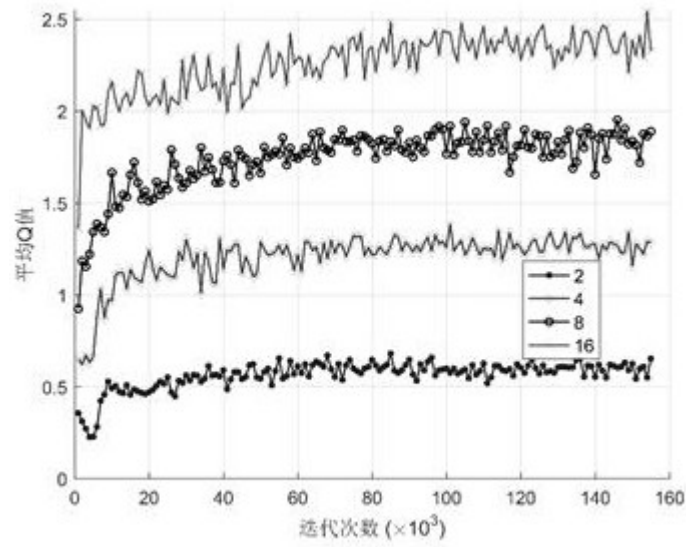


图5