基于无人机群飞行轨迹优化的多辐射源追踪技术

谷江春¹, 王海超¹, 丁国如^{1*}, 邓亦敏²

1. 中国人民解放军陆军工程大学通信工程学院, 南京 210000;

2. 北京航空航天大学自动化科学与电气工程学院仿生自主飞行系统研究组, 北京 100083

* E-mail: guoru_ding@yeah.net

收稿日期: 2021-08-15; 接受日期: 2021-11-04; 网络版发表日期: 2023-02-17

科技创新2030-“新一代人工智能”重大项目(编号: 2018AAA0102303)资助

摘要 无人机具有成本低、机动性强等优势特性, 近年来成为军用和民用领域中备受青睐的对象. 为了使无人机能够快速、稳定、高效地完成不同的任务, 安全、可靠的轨迹控制至关重要. 本文针对无人机群赋能的多辐射源追踪问题, 构建了多约束条件下的无人机群轨迹优化模型, 提出了估计、匹配、定位、追踪框架, 并设计了一种基于深度强化学习的无人机群飞行轨迹优化算法. 首先, 采用深度神经网络估计信道模型, 得到接收信号强度和距离之间的映射关系; 其次, 采用交互式方法生成接收信号强度矩阵, 计算出对应的距离矩阵并得到无人机与辐射源匹配方案; 然后, 采用多球交汇定位方法, 结合接收信号强度和距离之间的映射关系计算出辐射源的参考位置; 进一步, 将原始优化问题转换为马尔科夫决策过程, 并将辐射源定位信息引入强化学习中, 设计高效的无人机飞行轨迹优化算法. 此外, 引入群体熵对所提算法的智能性进行了分析和评估. 最后, 仿真结果对所提算法的平均飞行时间、任务完成率、智能性以及无人机群的轨迹分别进行了分析, 验证了所提方法的有效性.

关键词 无人机, 辐射源, 定位与追踪, 轨迹优化, 强化学习, 群体熵

1 引言

定位和跟踪已经被广泛应用于民用和军用领域, 例如空中交通管制、遥感以及情报、监视和侦察(intelligence, surveillance, and reconnaissance, ISR)^[1~3]. 一般来说, 定位可以分为两类: 主动定位和被动定位. 被动定位由于其固有的简单性和良好的隐蔽性, 具有很大的吸引力和挑战性. 被动定位中有一些典型的技术, 如基于接收信号强度(received signal strength, RSS)^[4]、到达方向(direction of arrival, DOA)^[5]、到达时差(time difference of arrival, TDOA)^[6]、到达频差

(frequency difference of arrival, FDOA)^[7]等.

传统的被动定位站通常部署在地面上, 受地形或建筑物的影响, 精度相对较低. 近年来, 由于无人机具有高机动性、按需部署和低成本等先天性的优势, 其作为移动传感器被广泛应用于定位和追踪系统. 无人机赋能的目标追踪主要包含两个领域的研究: 一方面是基于图像或视觉处理方面的, 另一方面是基于无线电信号的. 本文考虑的是无人机通过感知环境中的无线电信号强度对辐射源进行追踪. 现有文献已经初步开展了一些研究, 例如, 文献[8]提出了一个利用单个无人机感知接收信号强度进行室内单个目标搜索的方

引用格式: 谷江春, 王海超, 丁国如, 等. 基于无人机群飞行轨迹优化的多辐射源追踪技术. 中国科学: 技术科学, 2023

Gu J C, Wang H C, Ding G R, et al. Multiple-radiation source tracking based on the flight trajectory optimization of unmanned aerial vehicles swarm (in Chinese). Sci Sin Tech, 2023, doi: 10.1360/SST-2021-0378

案, 并提出一种基于强化学习的无人机导航算法. 文献[9]利用无人机群协同定位移动射频发射机, 提出了一种基于费舍尔信息矩阵(fisher information matrix, FIM)预测模型的移动时域路径规划算法. 将定位和追踪相结合能够提升追踪的性能. 文献[10]考虑了一个基于无人机的雷达定位问题, 提出了一种基于无源合成孔径天线阵列的解决方法. 为了在复杂环境下精确定位移动信号源, 文献[11]首先建立了一种特殊的无线传感器网络(wireless sensor networks, WSN), 其中每个传感器都安装在无人机上. 然后, 提出了一种结合TDOA方法和FDOA方法定位方案. 文献[12]考虑了采用无人机群对一个间歇性辐射信号进行追踪的问题, 提出了一个定位-追踪框架, 分别比较了基于卡尔曼滤波和贝叶斯估计两种位置估计方法的性能以及基于梯度下降后验克拉美罗下界和生物启发式两种路径规划算法的性能. 文献[13]考虑了采用无人机群对单个移动辐射源进行追踪的问题, 并且以追踪时间最小化为目标, 提出了一个时间约束下的多智能体强化学习追踪算法.

尽管现有工作已经对无人机赋能的辐射源追踪进行了一定的探索, 但是该领域的研究依然存在以下挑战: (1) 现有研究主要集中于对单个辐射源目标的追踪, 而无人机群追踪多辐射源的研究较少, 此场景不仅需要考虑到无人机之间的交互, 还需要考虑到对无人机进行任务分配等问题; (2) 现有研究主要集中于无人机二维轨迹的优化设计, 而三维轨迹的设计与优化更加实际却更具有挑战性, 比如说信道环境的复杂性以及动作和状态空间的庞大性; (3) 现有研究已经尝试应用强化学习方法于辐射源追踪场景, 但是对于无人机群追踪多辐射源目标场景, 如何设计有效的算法提升强化学习的收敛速度是一个亟需研究的问题. 基于以上观察, 本文考虑了一个无人机群追踪多辐射源的场景, 其中, 每一架无人机都携带一个全向天线传感器, 根据其接收信号强度, 对移动的辐射源进行跟踪. 注意到, 此场景中只采用了全向天线接收机, 无其他测量硬件设备. 该场景在实际中具有一定的应用价值, 比如可以用简单的设备对有害的辐射源进行数据采集或目标侦察. 然而, 由于没有任何关于移动辐射源的先验知识, 只能接收到移动辐射源在每个时刻的信号强度, 这是一个具有挑战性的问题. 本质上, 辐射源追踪问题是一种轨迹优化问题, 而强化学习方法适用于求解未

知动态环境下的序贯决策问题, 并在轨迹规划方面表现出了较好的性能. 例如, 文献[14]将无人机应用于城市环境从物联网设备中收集数据, 为了提升时间效率和安全性, 提出了一种基于端到端强化学习的无人机轨迹规划方法. 文献[15]研究了一个无人机辅助的移动终端任务卸载问题, 为了保证移动终端的服务质量, 提出了一种基于深度强化学习的无人机轨迹规划方法, 提升了算法的收敛速度和系统的吞吐量. 受此启发, 本文将强化学习应用于求解多辐射源追踪问题, 提出了一种基于深度强化学习的无人机群飞行轨迹优化方法. 注意到, 深度强化学习在图像或视觉处理方面的追踪问题研究已经成为热点^[16,17]. 与图像和视觉处理不同, 本文的追踪目标主要针对能够发射信号的辐射源. 为了探索无人机群赋能的辐射源追踪方法, 本文构建了多约束条件下的无人机群轨迹优化模型, 提出了估计、匹配、定位、追踪框架, 并设计了一种基于深度强化学习的无人机群飞行轨迹优化算法. 具体地, 本文的主要工作可概括如下.

(1) 提出了一个多约束条件下的无人机群飞行轨迹优化问题. 面向三维位置动态变化的多辐射源, 以无人机群追踪任务完成时间最小化为优化目标, 考虑无人机群的飞行区域、飞行速度以及追踪精度等约束条件, 构建了多约束条件下的无人机群轨迹优化模型.

(2) 提出了估计、匹配、定位、追踪框架. 首先, 采用深度神经网络(deep neural network, DNN)来估计信道模型, 得到接收信号强度和距离之间的映射关系; 其次, 采用交互式方法生成接收信号强度矩阵, 计算出相应的距离矩阵并得到无人机与辐射源匹配方案; 然后, 采用多球交汇定位方法, 结合接收信号强度和距离之间的映射关系计算出辐射源的参考位置; 进一步, 将原始优化问题转换为马尔科夫决策过程, 并将辐射源的位置信息引入强化学习中, 设计高效的无人机群飞行轨迹优化算法. 此外, 引入群体熵对所提算法的智能性进行了分析和评估.

(3) 仿真结果对所提算法分别进行了性能、轨迹以及智能性分析, 验证了所提方法的有效性. 首先, 所提算法在平均追踪时间和任务完成率等指标方面都具有明显的优势; 其次, 无人机群倾向于选择较短的路径对辐射源进行追踪; 然后, 状态群体熵的振荡下降体现出系统激发-汇聚、再激发-汇聚的过程, 说明了系统的智能性.

2 问题建立

2.1 系统模型

如图1所示, 考虑多架巡逻无人机协同对多个辐射源目标进行追踪, 其中, 辐射源的位置是未知且动态变化的. 考虑无人机具有辐射源个体识别能力, 比如识别辐射源的类别以及辐射源的发射功率等^[18,19]. 在此基础上, 无人机之间通过交互进行任务分配, 即一个无人机只选择一个辐射源进行跟随; 随后, 无人机通过接收到的信号强度对辐射源进行搜索, 当接收到的信号强度值达到设定的阈值时, 即完成了追踪任务.

考虑有 K 个辐射源, 在 t 时刻, $0 \leq t \leq T$, 第 k 个辐射源的位置可以表示为 $\mathbf{L}_k^R(t) = (x_k^R(t), y_k^R(t), z_k^R(t))$, $k=1, 2, \dots, K$. 考虑有 $N(N \geq K)$ 个无人机, 第 n 个无人机的位置可以表示为 $\mathbf{L}_n^U(t) = (x_n^U(t), y_n^U(t), z_n^U(t))$, $n=1, 2, \dots, N$. 在进行追踪任务之前, 无人机按照预定的轨迹在空中进行巡查, 第 n 个无人机的初始位置可以表示为

$$\mathbf{L}_n^U(0) = (x_n^U(0), y_n^U(0), z_n^U(0)).$$

采用时间离散法将时间段 T 离散为 M 个等长的时隙 δ_t , $T = M\delta_t$, 时隙 δ_t 需要足够小以保证在一个时隙内无人机的位置可以近似看作不变. 设置集合 $\mathcal{M} = \{1, 2, \dots, M\}$, $\mathcal{N} = \{1, 2, \dots, N\}$, $\mathcal{K} = \{1, 2, \dots, K\}$. 因此, 在第 m 个时隙, 第 k 个辐射源的位置可以表示为 $\mathbf{L}_k^R(m) = (x_k^R(m), y_k^R(m), z_k^R(m))$, 无人机 n 的位置可以表示为 $\mathbf{L}_n^U(m) = (x_n^U(m), y_n^U(m), z_n^U(m))$, $m \in \mathcal{M}$, $n \in \mathcal{N}$. 一般来说, 考虑对指定的区域进行搜索, 即一个三维立方体空间 $\mathbf{R}_{LU} = [x_L, x_U] \times [y_L, y_U] \times [z_L, z_U]$, 其中 L 和 U 分别表示三维空间的最低界限和最高界限. 因此, 无人机的位置需要满足 $\mathbf{L}_n^U(m) \in \mathbf{R}_{LU}$.

2.2 信号传播模型

在第 m 个时隙, 无人机 n 与辐射源 k 之间的距离可以表示为

$$d_{n,k}^{UR}(m) = \sqrt{(x_n^U(m) - x_k^R(m))^2 + (y_n^U(m) - y_k^R(m))^2 + (z_n^U(m) - z_k^R(m))^2}, \quad m \in \mathcal{M}, n \in \mathcal{N}, k \in \mathcal{K}. \quad (1)$$

对于辐射源 k , 第 n 个无人机在第 m 个时隙接收到的功率可以表示为^[13]

$$p_{n,k}^U(m) = p_k(m) - PL_{n,k}(m) + \zeta_n, \quad m \in \mathcal{M}, n \in \mathcal{N}, k \in \mathcal{K}, \quad (2)$$

式中, $p_k(m)$ 表示辐射源 k 在第 m 个时隙的发射功率, 由于空地传播链路受到障碍物的遮挡, $PL_{n,k}(m)$ 表示在第 m 个时隙第 n 个无人机与辐射源 k 之间链路的路径损耗, ζ_n 表示一个与瑞利衰落有关的服从指数分布的单位均值随机变量.

如图1所示, 空地链路同时存在视距和非视距链路, 因此路径损耗 $PL_{n,k}(m)$ 可以表示为^[20]

$$PL_{n,k}(m) = P_{\text{LoS},n,k}(m) \times PL_{\text{LoS},n,k}(m) + P_{\text{NLoS},n,k}(m) \times PL_{\text{NLoS},n,k}(m), \quad (3)$$

式中, $P_{\text{LoS},n,k}(m)$ 和 $P_{\text{NLoS},n,k}(m)$ 分别表示视距链路和非视距链路存在的概率. 如图1所示, $\theta_{n,k}(m)$ 表示在第 m 个时隙辐射源 k 与第 n 个无人机之间的仰角, 满足

$$\theta_{n,k}(m) = \frac{180}{\pi} \times \sin^{-1} \left(\left(z_n^U(m) - z_k^R(m) \right) / d_{n,k}^{UR}(m) \right).$$

$P_{\text{LoS},n,k}(m)$ 和 $P_{\text{NLoS},n,k}(m)$ 都是关于变量 $\theta_{n,k}(m)$ 的函数, 可以表示为^[20]

图1 基于无人机群的多辐射源追踪场景

Figure 1 Multi-radiation source tracking based on UAV swarm.

$$P_{\text{LoS}}(\theta_{n,k}(m)) = \frac{1}{1 + a \exp(-b(\theta_{n,k}(m) - a))}, \quad (4)$$

$$P_{\text{NLoS}}(\theta_{n,k}(m)) = 1 - P_{\text{LoS}}(\theta_{n,k}(m)), \quad (5)$$

式中, a 和 b 都是由空地传播环境决定的环境参数. $PL_{\text{LoS},n,k}(m)$ 和 $PL_{\text{NLoS},n,k}(m)$ 分别表示在第 m 个时隙辐射源 k 与第 n 个无人机视距链路和非视距链路的路径损耗, 可以表示为^[20]

$$PL_{\text{LoS},n,k}(m) = 20 \lg d_{n,k}^{\text{UR}}(m) + 20 \lg f + 20 \lg \left(\frac{4\pi}{c} \right) + \eta_{\text{LoS}}, \quad (6)$$

$$PL_{\text{NLoS},n,k}(m) = 20 \lg d_{n,k}^{\text{UR}}(m) + 20 \lg f + 20 \lg \left(\frac{4\pi}{c} \right) + \eta_{\text{NLoS}}, \quad (7)$$

式中 f 是系统频率, c 是光速, η_{LoS} 表示视距传播衰落参数以及 η_{NLoS} 表示非视距传播衰落参数.

2.3 优化问题建立

精准性和快速性是定位与追踪性能评估的两大关键指标, 基于此, 本文的研究目标为所有无人机完成追踪任务的时间最小化, 即从任务开始到最后一架无人机完成任务这一时间段, 同时考虑以一定的追踪精度为约束条件. 具体思路如下: 首先, 无人机之间通过交互进行任务分配, 即一个无人机至多选择一个辐射源进行跟随; 随后, 基于接收信号强度, 多无人机协作对多辐射源进行定位和追踪, 当一个无人机接收到的信号强度值达到设定的阈值时, 则此无人机完成了追踪任务. 综上, 建立以下优化问题一:

$$\begin{aligned} & \min_{a_{n,k}, \mathbf{v}_n^U(m)} \max \{T_1, T_2, \dots, T_N\} \\ C1: T_n &= \underset{I: p_{n,k}^U(I) \geq \gamma}{\operatorname{argmin}} I, n \in \mathcal{N} \\ C2: a_{n,k} &\in \{0, 1\}, n \in \mathcal{N}, k \in \mathcal{K}, \\ C3: \sum_{k=1}^K \sum_{n=1}^N a_{n,k} &= K, m \in \mathcal{M}, \\ C4: \mathbf{L}_n^U(m) &\in \mathbf{R}_{LU}, m \in \mathcal{M}, n \in \mathcal{N}, \\ C5: E_{n,\text{total}} &\leq E_{\max}, n \in \mathcal{N}, \\ C6: \mathbf{L}_n^U(m+1) &= \mathbf{L}_n^U(m) + \delta_t \mathbf{v}_n^U(m), \\ & m = 0, 1, \dots, M-1, n \in \mathcal{N}, \\ C7: \|\mathbf{v}_n^U(m)\| &\leq V_{\max}, m \in \mathcal{M}, n \in \mathcal{N}. \end{aligned} \quad (8)$$

式中, 优化目标为所有无人机追踪时间最小化. 优化变量一为无人机与辐射源的匹配系数 $a_{n,k}$, 即一个无人机至多对一个辐射源进行追踪; 优化变量二为无人机的飞行速度矢量 $\mathbf{v}_n^U(m)$, 包含飞行的方向和飞行的步长. 约束条件 C1 给出了第 n 架无人机的飞行时间表达式, 如果在第 I 时刻, 无人机首次接收到的信号强度大于某一设定的阈值 γ , 则表示无人机任务完成, 计时结束; 约束条件 C2 表示匹配系数的取值范围为 0 或者 1; 约束条件 C3 表示一个无人机至多与一个辐射源进行匹配, 即存在 K 个匹配对; 约束条件 C4 表示无人机的飞行范围; 约束条件 C5 表示无人机携带的能量有限, 其中 $E_{n,\text{total}}$ 表示第 n 架无人机飞行消耗的能量, 是一个与速度和步长成正相关的函数, $E_{\max}$ 表示每架无人机携带的最大能量; 约束条件 C6 和 C7 表示无人机应该满足速度约束, 其中 $V_{\max}$ 为无人机的最大飞行速度.

3 多辐射源追踪方法设计

由于辐射源的位置是未知且动态变化的, 并且空间中的信道环境对于无人机来说也是未知的, 因此, 传统的优化方法难以对问题一进行直接求解. 本文采用一种智能的方法给出优化问题一的解, 并对所提方法的智能性进行评估.

注意到, 定位-追踪的框架已经在文献[12]中被提出, 但是上述框架难以解决本文提出的无人机群追踪多辐射源这一问题, 主要包含以下因素: 一方面, 本文利用无人机感知到的接收信号强度对辐射源进行定位, 首先需要估计出无人机与辐射源之间的距离, 因此, 接收信号强度与距离之间的映射关系需要进行估计; 另一方面, 无人机选择哪一个辐射源进行追踪需要在飞行之前进行任务分配, 即无人机与辐射源的匹配关系需要优化. 因此, 本文提出估计、匹配、定位和追踪框架, 如图2所示. 总体来看, 步骤一获得的 RSS-Distance 映射关系作为步骤二和步骤三的先验知识; 步骤二输出的匹配结果将无人机群追踪多辐射源简化为单无人机追踪单无人机场景; 步骤三求解的辐射源参考坐标给步骤四中的追踪提供了参考依据, 不仅降低了搜索空间, 还可以作为模型预训练的数据, 从而提升模型训练的效率. 具体来讲, 步骤一采用深度神经网络估计信道模型, 得到接收信号强度和距离之间的映射关系; 步骤二采用交互式方法生成信号强

图2 多辐射源追踪框架

Figure 2 Multi-radiation source tracking framework.

度矩阵, 结合步骤一计算距离矩阵并得到无人机与辐射源匹配方案; 步骤三采用多球交汇定位方法, 结合步骤一计算辐射源的参考位置。注意到, 由于考虑的环境包含建筑、树木等障碍物, 导致信号传播中存在视距和非视距链路, 因此利用RSS进行定位只是粗略地给出了辐射源的坐标。此外, 辐射源的位置是动态变化的, 仅仅依据一次粗略的定位无法实现实时的追踪。因此, 步骤四首先将原始优化问题转换为马尔科夫决策过程, 然后将定位信息引入强化学习中, 并设计高效的无人机群飞行轨迹优化算法。接下来, 将分别对每一步骤进行详细的介绍和分析。

3.1 基于DNN的信道模型估计方法

定位能够减少追踪的动作空间, 从而减小追踪的时间。目标定位与追踪都基于接收到的信号强度, 而信道环境直接影响接收信号强度。因此, 在进行追踪之前, 首先需要对信道模型进行估计。在进行追踪之前, 真实环境是未知的, 即环境中的信号传播模型也是未知的, 为了估计RSS-Distance映射关系, 本文采用DNN模型进行回归。采用DNN模型主要基于以下考虑: 由于RSS与距离之间存在非线性关系, 而DNN可以表达十分复杂的非线性关系, 并且随着数据量增加, DNN的性能往往也会有有一定的提升。基于DNN的信道模型估计框架如图3所示, 其主要分为三个阶段: 分别为训练、测试和利用阶段。训练和测试阶段都采用自由空间信道传播模型生成一系列RSS-Distance模拟数据, 并先后对DNN进行训练和测试。注意到训练和测试样本不同, 训练样本是对RSS-Distance数据进行了90%的随机采样, 而测试样本是对RSS-Distance数据进行了10%的随机采样。利用阶段将无人机接收信号强

图3 基于DNN的信道模型估计框架

Figure 3 DNN-based channel model estimation framework.

度输入训练好的DNN模型, 即可输出无人机与辐射源之间的距离。

上述数据生成过程也可以在真实环境中实现, 具体步骤如下: 首先构建虚拟辐射源来模拟真实辐射源。考虑通过辐射源识别技术可以识别出辐射源的类型、调制方式以及发射功率等^[18,19], 因此可在真实环境中采用具有相同参数设置的设备模拟真实的辐射源发射信号并随机运动。然后, 无人机在空间中随机飞行并接收信号, 同时记录模拟辐射源的位置、无人机位置以及无人机处接收到的信号强度大小, 通过模拟辐射源和无人机的位置可以计算出无人机与辐射源之间的距离, 因此, 可以生成一系列的RSS-Distance数据对。

所采用的DNN结构: 包含输入层、隐藏层和输出层, 其中隐藏层包含两个具有50个神经元的Dense层

以及两个Dropout层。Dense层采用经典的Relu激活函数, 设置Dropout层是为了防止训练中出现过拟合现象, 其丢弃率设置为0.01。优化器采用adam, 学习速率设置为0.001。

随机生成150000个RSS-Distance模拟数据对, 其中随机采样135000个训练样本以及15000个测试样本分别对DNN模型进行训练和测试, 得到的误差曲线如图4所示。训练误差和测试误差随着回合次数的增加逐渐减小, 并且最后趋于稳定。此外, 训练误差和测试误差之间保持一个较为平稳的差值, 并未随着训练次数的增加而增加, 说明DNN模型具有一定的泛化能力。另外, 图5给出了模拟数据的分布和DNN的回归曲线。模拟数据包含300000个RSS-Distance数据对, 可以很直观地看到, DNN能够很好地估计RSS与距离之间的映射关系, 回归与模拟数据之间的偏差很小, 也进一步说明了DNN的泛化能力。

3.2 基于距离矩阵的无人机-辐射源匹配方法

在进行追踪之前, 一个无人机需要选择一个辐射源进行追踪。如图6(a)所示, 如果无人机的数量等于辐射源的数量, 即 $N=K$, 则每架无人机刚好匹配一个辐射源; 如图6(b)所示, 如果无人机的数量大于辐射源的数量, 即 $N>K$, 则有 $(N-K)$ 架无人机不需要参与追踪任务, 因此, 无人机如何选择辐射源进行匹配是一个值得研究的问题。直观地, 无人机作为单独的智能体, 可以采取自主决策的方式进行辐射源的匹配, 如图6(a)所示, 单个无人机首先根据感知到的辐射源信号强度进行排

图4 网络训练和测试阶段的误差曲线

Figure 4 Error curves of network training and testing phases, respectively.

图5 RSS-距离关系(模拟数据和DNN回归结果)

Figure 5 RSS-distance relationship (simulation data and DNN regression result).

序, 每架无人机都倾向于选择接收信号强度最大对应的辐射源。然而, 可以发现无人机3和4同时都选择了辐射源4, 导致无人机之间发生了冲突。为了避免上述情况, 无人机可以采取交互式的方法将接收信号强度序列与邻居无人机进行共享, 经过一定的交互后, 每个无人机都会得到一个RSS矩阵, 此时再结合RSS和距离的映射关系, 计算出距离矩阵, 并依据此选择无人机与辐射源匹配对, 完成匹配任务。根据距离矩阵选择无人机与辐射源匹配对的规则如下: 首先, 选择出矩阵每一行中的最小值, 然后观察最小值中是否有在同一列的值, 若否, 则按照取出最大值的结果进行匹配; 若是, 同列中较小者进行匹配, 较大者选择剩下的辐射源匹配($N=K$, 如场景一所示)或者不选择($N>K$, 如场景二所示)。

3.3 基于多球交汇的辐射源定位方法

典型的三球交汇定位方法如图7所示。具体而言, 由于多个无人机的位置以及接收到的信号强度已知, 首先, 根据步骤一得到的RSS-距离的映射关系来估计无人机与辐射源之间的距离; 随后通过三球交汇方法对辐射源进行定位。注意到, 由于步骤一中DNN模型估计的精度问题以及真实环境中非视距链路的影响, 此处只是粗略地给出一个辐射源的参考坐标。尽管如此, 此参考坐标依然给步骤四的追踪提供了先验信息, 并加速了网络的训练过程。在本文中, 多个无人机并不需要一直对辐射源定位, 只需要在追踪任务开始之前

图 6 无人机与辐射源典型匹配场景。(a) 场景一($N=K$; 交互前, 发生冲突; 交互后, 依据距离矩阵, 避免冲突); (b) 场景二($N>K$, 交互前, 发生冲突; 交互后, 避免冲突, 节省无人机资源)

Figure 6 Typical matching scenes of UAVs and radiation sources. (a) Scene 1 ($N=K$; before interaction, conflict occurs; after interaction, conflicts are avoided according to the distance; (b) scene 2 ($N>K$; before interaction, conflict occurs; after interaction, conflicts are avoided and UAV resources can be saved).

协同定位即可。无人机间可以通过通信将所在位置信息和接收到的信号强度共享, 实现一架无人机获取其匹配辐射源的参考位置信息。为了求解辐射源的位置, 可以建立以下三元二次方程组:

$$\begin{cases} (x_1^U - x)^2 + (y_1^U - y)^2 + (z_1^U - z)^2 = d_1^2, \\ (x_2^U - x)^2 + (y_2^U - y)^2 + (z_2^U - z)^2 = d_2^2, \\ (x_3^U - x)^2 + (y_3^U - y)^2 + (z_3^U - z)^2 = d_3^2, \end{cases} \quad (9)$$

上述方法也可以扩展到 N 个无人机的场景, 即可建立以下方程组:

$$\begin{cases} (x_1^U - x)^2 + (y_1^U - y)^2 + (z_1^U - z)^2 = d_1^2, \\ (x_2^U - x)^2 + (y_2^U - y)^2 + (z_2^U - z)^2 = d_2^2, \\ \vdots \\ (x_N^U - x)^2 + (y_N^U - y)^2 + (z_N^U - z)^2 = d_N^2, \end{cases} \quad (10)$$

式中, (x, y, z) 为所求辐射源的参考位置坐标。

3.4 基于 D^3 QN-PT的多辐射源追踪方法及评估

3.4.1 MDP构建

优化问题一可以映射为马尔科夫决策过程 $\langle S, A, P, R \rangle$, 对于单个无人机而言, 具体如下:

图 7 三球交汇定位示意图

Figure 7 Schematic diagram of three-sphere intersection localization.

S: 定义状态空间为在可行空间内第 n 个无人机的所有可能的位置, 即 $S = \{s_n(m) = L_n^U(m)\}, n \in \mathcal{N}, m \in \mathcal{M}$. 此外, 对一些特殊的状态进行定义. “出界”表示无人机飞出来规定的可行空间, “位置重复”表示无人机多次到达同一个位置, “终点”表示无人机接收信号强度达到设定的阈值.

A: 定义动作空间为第 n 个无人机所有可能的飞行速度矢量, 即 $A = \{v_n(m) : \|v_n(m)\| = 10\}, n \in \mathcal{N}, m \in [0, 1, \dots, M-1]$, 飞行方向集合 $\bar{A} = \{\text{东, 西, 南, 北, 上, 下}\}$.

P: 状态转移函数取决于约束C6: $L_n^U(m+1) = L_n^U(m) + \delta_t v_n^U(m)$.

R: 定义实时奖励为一个关于飞行时间和接收信号强度的函数, 具体如下:

$$\bar{p}(m-1) = \sum_{i=1}^{m-1} p(i), i = 1, 2, \dots, m-1, \quad (11)$$

$$r_1(m) = \begin{cases} p_{\text{out}}, & \text{if } s = \text{出界}, \\ p_{\text{again}}, & \text{else if } s = \text{位置重复}, \\ r_{\text{final}}, & \text{else if } s = \text{终点}, \\ \mu(p(m) - \bar{p}(m-1)), & \text{else}, \end{cases} \quad (12)$$

$$r_2 = p_{\text{step}}, \quad (13)$$

$$r(m) = r_1(m) + r_2, \quad (14)$$

式中, $p(i)$ 表示在第 i 个时隙无人机接收的信号强度, $\bar{p}(m-1)$ 表示前 $m-1$ 个时隙无人机接收信号强度的平均值, p_{out} 表示无人机出界后环境给出的惩罚, p_{again} 表示无人机位置重复后环境给出的惩罚, r_{final} 表示无人机达到终点环境给出的奖励, u 为给定的权重因子, p_{step} 表示无人机飞行时间的成本. 奖励函数的设计有以下两点考虑: 一是鼓励无人机尽可能靠近移动辐射源. 由于无人机与辐射源的距离越近, 无人机接收到的信号强度就越高, 因此越接近辐射源的位置奖励越高; 二是尽量减少无人机飞行的时间. 这与问题一的优化目标相对应, 智能体飞行的步长数越少, 惩罚越小.

将优化问题一转换为马尔科夫决策过程后, 智能体的目标转换为在离散时间 M 内累计获得的奖励,

$$\text{即 } R_{\text{total}} = \sum_{m=1}^M r(m). \text{ 解决马尔科夫决策过程的强化学习}$$

算法有多种, 其中Q学习是一种基于表格的基本强化学习算法. 注意到, 在所提出的场景中, 由于状态和动作空间很大, 因此Q学习不适合解决所提出的问题. 接下来, 提出一种基于竞争和双层深度Q网络(dueling-double deep Q network, D³QN)和预训练(preliminary training, PT)的无人机群轨迹优化算法来解决多辐射源追踪问题.

3.4.2 基于D³QN-PT的多辐射源追踪算法设计

注意到, 所提场景中存在多个动作对应的值函数相近这一问题, 因此, 此部分利用基于竞争架构的网络来估计动作值函数^[21]. 此外, 为了防止训练出现过拟合, 采用深度双Q网络的方法^[22]. 综上, 提出一种基于D³QN-PT的辐射源追踪框架, 如图8所示. 具体来说, 无人机首先获取自身的状态, 并根据策略对环境做出动作, 随后环境给无人机一个反馈, 并且无人机的状态转移到下一个状态. 深度双Q网络中具有两个网络, 分别是评估网络和目标网络. 两者都采用了竞争网络结构, 并且有相同的参数设置, 不同的是参数更新的时间. 无人机飞行的策略是由评估网络给出的. 此外, 无人机与环境交互的数据存储于经验池中, 每一次将在其中采样一小批次数据作为评估网络的训练样本, 评估网络将得到的状态值函数用于计算损失函数, 再通过反向传播不断改进评估网络, 并且设定一定的回合数后对目标网络进行一次参数更新. 此外, 此框架

图 8 基于D³QN-PT的多辐射源追踪框架

Figure 8 D³QN-PT based multi-radiation tracking framework.

中还引入了步骤三计算出的辐射源参考位置信息,一方面可以对网络进行初始化预训练,提升收敛的速度;另一方面,基于辐射源参考位置,减小无人机的状态空间,避免无效的搜索和试错。

具体来说,竞争网络结构的核心思想是将优势函数和值函数进行聚合,然后输出不同动作的 Q 值,具体可以表示如下:

$$\begin{aligned} \widehat{Q}(\mathbf{s}, \mathbf{a}; \boldsymbol{\theta}, \alpha, \beta) &= V(\mathbf{s}; \boldsymbol{\theta}, \beta) \\ &+ \left(A(\mathbf{s}, \mathbf{a}; \boldsymbol{\theta}, \alpha_0) - \frac{1}{K} \sum_a A(\mathbf{s}, \mathbf{a}'; \boldsymbol{\theta}, \alpha_0) \right), \end{aligned} \quad (15)$$

式中, V 表示值函数, A 表示优势函数, $\boldsymbol{\theta}$ 、 α_0 和 β 分别代表隐藏层以及两个隐藏层流的参数。

给定一个状态 $\mathbf{s}(m) = \mathbf{L}^U(m)$, $m \in \mathcal{M}$, 对应的未来 M_1 步奖励和可以表示为

$$R(m : m + M_1) = \sum_{i=0}^{M_1-1} \gamma^i R(m + i + 1), \quad (16)$$

因此, M_1 步深度Q网络的损失函数可表示为

$$\begin{aligned} \text{loss}_2 &= \left(R(m : (m + M_1)) \right. \\ &+ \gamma^{M_1} \max_{\mathbf{v} \in \bar{\mathcal{A}}} \widehat{Q}(\mathbf{s}(m + M_1), \mathbf{v}'; \boldsymbol{\theta}^-) \\ &\left. - \widehat{Q}(\mathbf{s}(m), \mathbf{v}(m); \boldsymbol{\theta}) \right)^2. \end{aligned} \quad (17)$$

与单步学习相比, M_1 步学习通常会有更好的表现, 由于它加强了序列的最后 M_1 个动作, 而前者只加强了最后一个动作. M_1 步深度双Q网络的损失函数可以表示为^[23]

$$\begin{aligned} \text{loss}_3 &= \left(R(m : (m + M_1)) \right. \\ &+ \gamma^{M_1} \max_{\mathbf{v} \in \bar{\mathcal{A}}} \widehat{Q}(\mathbf{s}(m + M_1), \mathbf{v}^*; \boldsymbol{\theta}^-) \\ &\left. - \widehat{Q}(\mathbf{s}(m), \mathbf{v}(m); \boldsymbol{\theta}) \right)^2, \end{aligned} \quad (18)$$

式中,

$$\mathbf{v}^* = \underset{\mathbf{v} \in \bar{\mathcal{A}}}{\operatorname{argmax}} \widehat{Q}(\mathbf{s}(m + M_1), \mathbf{v}'; \boldsymbol{\theta}). \quad (19)$$

表1中的算法1给出了基于D³QN-PT的轨迹优化算法设计的全部过程. 具体地, 步骤1对参数进行了初始化; 步骤2对网络进行了初始化预训练; 步骤3~14表示一个回合的训练过程, 其中步骤6给出了基于 ϵ -greedy

表 1 基于D³QN-PT的多辐射源追踪算法**Table 1** D³QN-PT based multi-radiation source tracking algorithm算法1: 基于D³QN-PT的多辐射源追踪算法

- 1 初始化最大回合数 M_{epi} , 每回合最大步数 M_{step} , 探索率 ε_0 , 衰减率 α_1 , 出界惩罚 p_{out} , 位置重复惩罚 p_{again} , 任务完成门限 γ , 容量为 C 的回放记忆序列 D .
- 2 根据式(10), 获取辐射源参考坐标, 并计算无人机与参考坐标距离, 对网络进行初始训练, 更新网络参数 $\theta^- \leftarrow \theta$ 以及 $\varepsilon \leftarrow \varepsilon_0$.
- 3 重复(步骤4~17).
- 4 初始化一个容量为 M_1 的滑动窗序列 W , 设置时间步数 $m \leftarrow 0$.
- 5 重复(步骤6~14).
- 6 根据 ε -greedy从动作空间 $\hat{A}$ 中选择动作 $\mathbf{v}_{j^*}$, 其中

$$j^* = \begin{cases} \text{randi}(j), & \text{概率 } \varepsilon \\ \text{argmax}_{\mathbf{v} \in \hat{A}} \hat{Q}(\mathbf{s}(m), \mathbf{v}(m); \theta), & \text{概率 } 1 - \varepsilon \end{cases}, j = 1, 2, \dots, J; m \in [0, 1, \dots, M-1]$$
- 7 执行动作 $\mathbf{v}(m) = \mathbf{v}_{j^*}$, 根据约束C6得到智能体的下一个状态 $\mathbf{s}(m+1)$, 以及式(14)得到实时奖励 $R(m)$.
- 8 将 $(\mathbf{s}(m), \mathbf{v}(m), R(m), \mathbf{s}(m+1))$ 保存在滑动窗序列 W 中.
- 9 如果 $m \geq M_1$, 根据式(16), 计算 M_1 步累计奖励并且将 $(\mathbf{s}(m-M_1), \mathbf{v}(m-M_1), R((m-M_1):m), \mathbf{s}(m))$ 储存在回放记忆池 D 中.
- 10 从回放记忆池 D 中随机采样 $(\mathbf{s}(j), \mathbf{v}(j), R(j:j+M_1), \mathbf{s}(j+M_1))$.
- 11 令

$$y(j) = R(j:j+M_1) + \gamma^{M_1} Q(\mathbf{s}(j+M_1), \mathbf{v}^*; \theta^-)$$
 式中 $\mathbf{v}^* = \text{argmax}_{\mathbf{v} \in \hat{A}} Q(\mathbf{s}(j+M_1), \mathbf{v}; \theta^-)$.
- 12 在 $(y(j) - \hat{Q}(\mathbf{s}(j), \mathbf{v}(j); \theta))^2$ 上关于网络参数 θ 执行梯度下降.
- 13 更新 $m \leftarrow m+1$, $\varepsilon \leftarrow \varepsilon \alpha_1$.
- 14 直到 $p(m) \geq \gamma$ 或者 $m = M_{\text{step}}$.
- 15 $m_{\text{epi}} \leftarrow m_{\text{epi}} + 1$.
- 16 每 N_{epi} 个回合之后更新一次目标网络的参数 $\theta^- \leftarrow \theta$.
- 17 直到 $m_{\text{epi}} = M_{\text{epi}}$.

的动作选择策略, 步骤9计算了 M_1 步的累计奖励, 步骤11~13对评估网络进行训练并更新参数. 步骤16表示每间隔 N_{epi} 个回合更新一次目标网络的参数, 即 $\theta^- \leftarrow \theta$. 注意到, 强化学习的应用从虚拟环境转向实际环境一直是一个热点和难点问题, 特别是如何保证算法的实时性和稳定性. 本文将强化学习应用于无人机的路径规划, 其主要采用“线下学习、线上利用+学习”的模式. 首先, 从真实环境中采集一定的数据生成逼真的虚拟环境, 采集的数据包含辐射源的发射功率、无人机接收信号强度(RSS)-无人机与辐射源距离(Distance)数据对等. 由于无人机与辐射源的距离是未知的, 因此可以采用具有相同参数设置的设备模拟辐射源, 从而产生Distance数据. 当虚拟的环境生成之后, 所提的辐射源追踪算法便可以在该环境中进行训练, 直到算法收敛. 其次, 虚拟环境到真实环境中的过渡. 由于真实环境存在很多不确定因素, 包含环境的动态变化、辐射源的移动等. 因此, 不可将训练好的算法直接在真实环境中运用. 此处提出一种“线上利用+学习”的思

路, 一方面, “线上利用”充分利用“线下学习”的结果, 提升了真实环境中无人机追踪的时效性; 另外一方面, “线上学习”克服了真实环境不确定性带来的问题, 提升了算法的稳定性. 此外, 当环境变化较大但新旧环境之间存在一定的相似性时, 可以采用迁移学习^[24,25]等方法将训练好的模型迁移到新的环境中, 然后再进行训练, 可以减少模型训练的时间, 从而提升算法的时效性.

3.4.3 基于状态群体熵的多辐射源追踪算法评估

强化学习通过与环境交互来不断学习, 被认为是走向通用人工智能关键技术之一. 然而, 强化学习的智能性如何评估仍然是一个值得探索的问题. 信息论之父香农认为智能就是获取信息、处理信息和利用信息解决问题的能力^[26]. 由于信息熵是量化信息的常用方式, 因此, 本小节尝试引入信息熵对强化学习的智能性进行初步的分析.

首先, 香农对信息熵的定义为自信息的数学期望,

离散信息熵具体可以表示如下^[27]:

$$E(x) = - \sum_{i=1}^q p(x_i) \log(p(x_i)), \quad (20)$$

式中, $p(x_i)$ 表示 x_i 事件发生的概率. 可以看出, 当 $p(x_i)=1$ 时, 其余事件发生的概率 $p(x_1)=\dots=p(x_{i-1})=p(x_{i+1})=\dots=p(x_q)=0$, 其离散信息熵 $E(x)=0$, 即事件发生的不确定性为0. 当 $p(x_1)=\dots=p(x_i)=\dots=p(x_q)=1/q$ 时, 其离散信息熵 $E(x)=\log q$, 事件发生的不确定性最大. 对于单个智能体, 离散信息熵越大, 表明其行为的不确定性越大; 反之, 其行为的不确定性越小.

对于强化学习而言, 每个智能体都存在许多状态, 在每个状态下都会选择一个动作, 在状态 s 下选择动作 a 可以称为一个事件 s_a . 那么状态 s 下的离散信息熵可以表示为

$$E(s) = - \sum_{a \in A} p(s_a) \log(p(s_a)). \quad (21)$$

大规模复杂群体的群体熵被定义为系统结构演化熵与个体相互作用信息熵的复合函数, 本文针对强化学习独有的多状态、多回合特点, 定义状态群体熵为一个回合所有状态下的离散信息熵平方和, 表达式如下:

$$\begin{aligned} E(S) &= E^2(s(1)) + E^2(s(2)) + \dots + E^2(s(M_{\text{step}})) \\ &= \sum_{i=1}^{M_{\text{step}}} E^2(s(i)). \end{aligned} \quad (22)$$

从上式可以看出, $E(S)$ 的取值都是非负的, 其大小不仅与某个状态下离散信息熵的大小成正相关, 也与一个回合下的状态数目成正相关. 随着回合数的增加, 无人机在某个状态下的探索率减小, 动作选择的确定性逐渐增加, 其离散信息熵也逐渐减小; 同时, 为了获取更大的回报, 无人机倾向于以更少的步数, 即状态数目更少, 来完成任务. 因此, 一个具有智能性的算法, 随着回合数目的增加, $E(S)$ 将呈现振荡下降, 并逐渐稳定的趋势.

4 仿真分析

4.1 参数设置

此部分给出了仿真结果来验证所提追踪框架和算法的有效性. 首先, 没有其他说明的情况下, 考虑3架无

人机追踪3个移动辐射源的场景, 飞行区域空间大小为 $300 \text{ m} \times 200 \text{ m} \times 100 \text{ m}$. 无人机的初始位置随机给定, 分别为 $L_1^U(0)=(53, 162, 88)$, $L_2^U(0)=(150, 102, 67)$, 以及 $L_3^U(0)=(235, 166, 77)$. 辐射源1, 2 和3的位置坐标分别表示为 $(x_1, 50, 50)$, $(x_2, 40, 10)$ 以及 $(x_3, 80, 22)$, 其中

$$x_k = \begin{cases} X_k, & \text{if } t = 0, \\ X_k + 0.5 \times 10, & \text{if } t \neq 0 \text{ and } t \% 10 = 0, \\ X_k + 0.5 \times (t \% 10), & \text{else,} \end{cases} \quad (23)$$

式中, $k=\{1, 2, 3\}$ 表示辐射源的类别, $X_1=20$, $X_2=160$, $X_3=250$. $t=\{1, 2, \dots, T\}$ 表示某一时刻, 从 $t=0$ 开始计时, 辐射源以10为周期做周期运动. 因此, 辐射源1起始点和终点分别为 $(20, 50, 50)$ 和 $(25, 50, 50)$, 辐射源2的起始点和终点分别为 $(160, 40, 10)$ 和 $(165, 40, 10)$, 辐射源3的起始点和终点分别为 $(250, 80, 22)$ 和 $(255, 80, 22)$. 当无人机数量显著增加时, 主要分为两种情况: 第一种是辐射源的数量也显著增加, 这时在“匹配”阶段需分配更多的无人机去进行追踪任务, 并需在“追踪”阶段对更多无人机的飞行轨迹进行优化; 另一种是辐射源的数量保持不变, 在“匹配”阶段之后, 多余的无人机将不执行“追踪”任务. 显然, 由于第一种情况还需要考虑更多无人机的飞行轨迹优化, 因此其复杂度也明显高于第二种情况. 在“匹配”阶段, 本文采用了基于距离矩阵的无人机-辐射源匹配方法, 无人机的数量增加仅仅增加了矩阵的维度, 对算法的复杂度增加较小; “匹配”阶段之后, 多无人机追踪多辐射源问题转换为多个单无人机追踪单辐射源问题, 尽管无人机数量显著增加, 算法的复杂度仅仅呈现倍数的增加, 因此, 所提算法依然可以对该问题进行有效的求解.

表2中给出了一些关键的参数设置, 其中, 设置最大步数是考虑到无人机的能量有限, 假定无人机每飞行一步消耗的能量相同, 因此步数越多, 能耗越大. 此处分别给出初始探索率、训练时间以及奖励函数设计相关参数选取的原则. 首先是初始探索率设置为0.5, 这是由于本文所提算法具有预训练的步骤, 因而在选择动作时不需要完全随机, 可以有1/2的概率利用预训练的结果, 从而提升网络的收敛速度. 其次是训练时间设置为1000个回合, 这是综合考虑了初始探索率和探索率衰减因子两个参数, 尽可能保证在训练前期可以有足够的时间探索, 也考虑到训练后期可以尽可能地利用, 确保网络输出的稳定性. 然后是奖励函数的设

表 2 关键参数设置

Table 2 Key parameter settings

参数	数值	定义
M_{epi}	1000	最大回合数目
M_{step}	200	每回合最大时间步数
ε_0	0.5	初始探索率
α_1	0.99	探索率衰减因子
C	100000	回放记忆容量
M_1	30	多步学习步数
δ_t	1 s	时隙长度
V_{max}	10 m/s	无人机的最大速度
p_{out}	-10	出界的惩罚
p_{again}	-10	位置重复的惩罚
p_{step}	-3	每走一步的惩罚
u	1	权重因子
γ	-10 dBm	任务完成门限

计, 为了使无人机尽快追踪到辐射源, 设计了两层“奖励”: 一个是当前接收信号强度与之前接收信号强度平均值的差值, 另外一个为每飞行一步的惩罚。除此之外, 还包含无人机出界和位置重复的惩罚。在奖励参数设置方面, 主要遵循的原则为无人机出界和位置重复的惩罚最大, 其次, 每飞行一步的惩罚在数值上大于接收信号强度的差值。这样设计可以保证奖励值始终为负值, 避免奖励值正负交替的现象, 使得算法具有一定的稳定性。此外, 辐射源的发射功率相同且都为 $p_k(m)=30$ dbm。信道相关参数设置如下: $a=10$, $b=0.6$, $f=200$ MHz, $c=3.0 \times 10^8$ m/s, $\eta_{\text{LoS}}=0.1$ 以及 $\eta_{\text{NLoS}}=21$ ^[18]。深度网络相关参数设置如下: 每个无人机使用相同的神经网络结构, 具体而言, 单个网络是由5个隐层组成的全连通前馈神经网络, 前4个隐层分别包含512, 256, 128, 128个神经元, 最后一个层叫做竞争层, 包含 $K+1$ 个神经元, 其中一个神经元用于价值函数, 另外 K 个神经元用于优势函数^[20]。仿真环境设置如下: Python 3.7.6, TensorFlow 1.4和Keras。

4.2 性能分析

此部分对所提算法和基准算法在平均飞行时间和任务完成率等指标上进行了比较。本文所提算法为 $D^3\text{QN-PT}$, 基准算法为 $D^3\text{QN}$ 算法、DDQN(double deep Q network)算法、DQN(deep Q network)算法和Random

算法, 其中, $D^3\text{QN}$ 算法相比于 $D^3\text{QN-PT}$ 算法缺少了网络的预训练, 此外, 其初始化探索率 ε_0 。DDQN算法和DQN算法采用了传统的网络架构, Random算法中选择动作的方式为随机策略。图9对比了5种算法中各无人机的平均飞行时间。为了避免一次飞行时间在性能方面的偶然性, 此处采取训练中最后200回合无人机的飞行时间的平均值作为飞行时间性能方面的指标, 即平均飞行时间。例如, 对于无人机1而言, 平均飞行时间 $T_{\text{aver}} = \sum_{m_{\text{epi}}=M_{\text{epi}}-200}^{M_{\text{epi}}} T_{m_{\text{epi}}} / 200$ 。观察图9, 可以看出,

对于所有的无人机, 所提算法 $D^3\text{QN-PT}$ 的平均飞行时间都是最短的, 算法 $D^3\text{QN}$ 的性能次之, 算法DDQN和DQN的性能紧随其后, Random算法的平均飞行时间最长, 这也从飞行时间指标层面验证了所提算法的高效性。

图10对比了五种算法的任务完成率。此任务完成率分析了训练中最后200回合的任务完成情况。可以观察到, 所提算法 $D^3\text{QN-PT}$ 的任务完成率都是最高的, 并且接近于1。 $D^3\text{QN}$ 算法性能次之, DDQN和DQN算法的性能紧随其后, 并且不够稳定, 另外, Random算法的任务完成率最低。

图11给出了所提算法 $D^3\text{QN-PT}$ 中无人机与辐射源距离随着训练回合的变化曲线。可以看出, 无人机与辐射源的距离随着训练回合数的增加逐渐减小, 最终的平均距离在20 m以内。追踪的距离精度与设置的任务完成门限有关, 如果门限值设置越高, 对应的距离精度也相应越高, 但是这将会提高训练的次数。此外, 由

图 9 不同算法的平均飞行时间

Figure 9 Average flight time of different algorithms.

图 10 不同算法的任务完成率

Figure 10 Task completion rate of different algorithms.

图 11 无人机与辐射源距离随着训练回合的变化曲线

Figure 11 The curve of the distance between the drone and the radiation source with episode.

于接收信号强度不仅仅与距离有关, 还与两者之间的俯仰角有关, 因此, 任务完成门限设置越高时, 训练的困难就会越大。

4.3 轨迹分析

此小节分别给出了所提算法D³QN-PT在一定迭代次数后每个无人机的三维轨迹图以及二维平面轨迹图。从图12可以看出, 每个无人机都选择了距离较近的辐射源进行追踪, 并且为了尽快完成任务, 都尽可能选择一条距离较短的路径去追踪辐射源, 当接收信号强度达到任务完成门限后, 无人机到达路径的终点。

4.4 智能性分析

图13给出了所提算法D³QN-PT中回报值和状态群

体熵随回合数的变化情况。首先, 回报值随着训练回合的增加而逐渐增加, 并在一定次数后趋于振荡收敛。另外, 状态群体熵随着训练回合数的增加呈现振荡减小的趋势, 这与回报值的趋势正好相反。当状态群体熵很大时, 回报值处于振荡上升, 这是由于智能体的探索率较大, 选择动作的不确定性较大, 呈现激发的态势。随着训练回合的增加, 智能体探索的概率逐渐减小, 状态群体熵减小, 选择动作的不确定性减小, 呈现汇聚的态势, 回报值趋于稳定。从系统理论上分析, 整个系统呈现出激发-汇聚, 再激发-汇聚, 最后趋于稳定, 这说明所采用的强化学习算法能够较好地探索环境(获取信息)、训练网络(处理信息)、做出决策(利用信息), 最终获得较高的增益。此外, 每个个体通过改变自己的动作(学习策略、经验), 提升系统的整体增益, 表

图 12 无人机的飞行轨迹图. (a) 三维轨迹; (b) 水平轨迹

Figure 12 The flight trajectories of the UAVs. (a) Three-dimensional trajectories; (b) horizontal trajectories.

图 13 回报值和状态群体熵随回合数的变化曲线
Figure 13 The curves of reward and state group entropy with episode.

现出一定的智能性^[28].

5 结论与展望

针对无人机群追踪多辐射源问题, 本文构建了多

约束条件下的无人机群轨迹优化模型, 提出了估计、匹配、定位、追踪框架, 并设计了一种基于深度强化学习的无人机群飞行轨迹优化算法. 仿真得到以下结论: (1) 多无人机利用各自接收的信号强度可实现协同定位并完成较高精度的追踪任务; (2) 利用定位信息对评估网络进行预训练能够提高算法的收敛速度和无人机的追踪性能; (3) 状态群体熵是无人机在所有状态下的离散信息熵平方和, 可以被用来评估所提方法的智能性. 本文给出了一种无人机群追踪多辐射源问题的解决思路, 未来还可以在所提框架上进一步改进, 比如说多无人机在“追踪”阶段同时对辐射源进行协同定位, 从而提升追踪的精度, 但也会相应地增加交互的成本以及算法的复杂度, 如何在追踪精度、成本以及算法复杂度之间寻找到一个平衡点是未来值得研究的问题. 此外, 无人机由于飞行能力, 还存在加速度、转向角^[29]等约束, 因此, 在未来的工作中, 更加实际的无人机运动模型值得被考虑.

参考文献

- 1 Liu Z H, Liu H B, Xu W Y, et al. An error-minimizing framework for localizing jammers in wireless networks. *IEEE Trans Parallel Distrib Syst*, 2014, 25: 508–517
- 2 Orfanus D, Freitas E P de, Eliassen F, et al. Self-organization as a supporting paradigm for military UAV relay networks. *IEEE Wirel Commun*, 2016, 20: 804–807
- 3 Laoudias C, Moreira A, Kim S, et al. A survey of enabling technologies for network localization, tracking, and navigation. *IEEE Commun Surv Tutorials*, 2018, 20: 3607–3644
- 4 Yassin A, Nasser Y, Awad M, et al. Recent advances in indoor localization: A survey on theoretical approaches and applications. *IEEE Commun Surv Tutorials*, 2017, 19: 1327–1346
- 5 Lin B, Liu J, Xie M, et al. Direction-of-arrival tracking via low-rank plus sparse matrix decomposition. *Antennas Wirel Propag Lett*, 2015, 14: 1302–1305
- 6 Makki A, Siddig A, Saad M, et al. Indoor localization using 802.11 time differences of arrival. *IEEE Trans Instrum Meas*, 2016, 65: 614–623
- 7 Kim Y, Kim D, Kim H. Two-step estimator for moving-emitter geolocation using time difference of arrival/frequency-difference of arrival measurements. *IET Radar Sonar Navig*, 2015, 9: 881–887
- 8 Chowdhury M M U, Erden F, Guvenc I. RSS-based Q-learning for indoor UAV navigation. In: *IEEE Military Communications Conference (MILCOM)*, 2019. 121–126
- 9 Koohifar F, Kumbhar A, Guvenc I. Receding horizon multi-UAV cooperative tracking of moving RF source. *IEEE Commun Lett*, 2017, 21: 1433–1436
- 10 Hao B, Zhu J, Li Z, et al. Passive radar source localization based on PSAAA using single small size aircraft. In: *IEEE Global Communications Conference Workshops*, 2015. 1–6
- 11 Wan P, Huang Q, Lu G, et al. Passive localization of signal source based on UAVs in complex environment. *China Commun*, 2020, 17: 107–116
- 12 Koohifar F, Guvenc I, Sichitiu M L. Autonomous tracking of intermittent RF source using a UAV swarm. *IEEE Access*, 2018, 6: 15884–15897
- 13 Chen Y J, Chang D K, Zhang C. Autonomous tracking using a swarm of UAVs: A constrained multi-agent reinforcement learning approach. *IEEE Trans Veh Technol*, 2020, 69: 13702–13717
- 14 Bayerlein H, Theile M, Caccamo M, et al. UAV path planning for wireless data harvesting: A deep reinforcement learning approach. In: *IEEE*

- Global Communications Conference (Globecom), 2020. 1–6
- 15 Liu Q, Shi L, Sun L, et al. Path planning for UAV-mounted mobile edge computing with deep reinforcement learning. *IEEE Trans Veh Technol*, 2020, 69: 5723–5728
 - 16 Yun S, Choi J, Yoo Y, et al. Action-decision networks for visual tracking with deep reinforcement learning. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2017. 1349–1358
 - 17 Zhang D, Maei H, Wang X, et al. Deep reinforcement learning for visual object tracking in videos. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2017. 1–6
 - 18 Ezuma M, Erden F, Kumar Anjinappa C, et al. Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and bluetooth interference. *IEEE Open J Commun Soc*, 2020, 1: 60–76
 - 19 Al-Sa'd M F, Al-Ali A, Mohamed A. RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database. *Future Gener Comput Syst*, 2019, 100: 86–97
 - 20 Al-Hourani A, Kandeepan S, Lardner S. Optimal lap altitude for maximum coverage. *IEEE Wireless Commun Lett*, 2014, 3: 569–572
 - 21 Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning. In: *The 33rd International Conference on Machine Learning (ICML)*. 2016. 1995–2003
 - 22 Zeng Y, Xu X, Jin S, et al. Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning. *IEEE Trans Wireless Commun*, 2021, 20: 4205–4220
 - 23 Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. *Comput Sci*, 2015, 8: A187
 - 24 Pan S J, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*, 2010, 22: 1345–1359
 - 25 Lin F, Chen J, Sun J, et al. Cross-band spectrum prediction based on deep transfer learning. *China Commun*, 2020, 17: 66–80
 - 26 Zhong Y X. *Principles of Information Science* (in Chinese). Beijing: Beijing University of Posts and Telecommunication Press, 2013. 12 [钟义信. 信息科学原理. 北京: 北京邮电大学出版社, 2013. 12]
 - 27 Núñez J A, Cincotta P M, Wachlin F C. Information entropy. *Celest Mech Dynam Astron*, 1996, 64: 43–53
 - 28 Duan H B, Qiu H X. *Unmanned Aerial Vehicle Swarm Autonomous Control Based on Swarm Intelligence* (in Chinese). Beijing: Science Press, 2018. 1–25 [段海滨, 邱华鑫. 基于群体智能的无人机集群自主控制. 北京: 科学出版社, 2018. 1–25]
 - 29 Wang H, Wang J, Ding G, et al. Completion time minimization for turning angle-constrained UAV-to-UAV communications. *IEEE Trans Veh Technol*, 2020, 69: 4569–4574

Multiple-radiation source tracking based on the flight trajectory optimization of unmanned aerial vehicles swarm

GU JiangChun¹, WANG HaiChao¹, DING GuoRu¹ & DENG YiMin²

¹ School of Communication Engineering, Army Engineering University of PLA, Nanjing 210000, China;

² Bio-inspired Autonomous Flight Systems Research Group, School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China

In recent years, unmanned aerial vehicles (UAVs) have become popular objects in the military and civil fields due to their advantages of low cost and mobility. Safe and reliable trajectory control is very important in enabling UAVs to quickly, stably, and efficiently accomplish different tasks. First, this paper constructs a trajectory optimization model of the UAV swarm under multiple constraints and proposes the framework of estimation, matching, localization, and tracking. Next, a flight trajectory optimization algorithm of the UAV swarm based on deep reinforcement learning is designed. A deep neural network is specifically used to estimate the channel model, and the mapping relationship between the received signal strength and the distance is obtained. The received signal strength matrix is generated by the interactive method. Moreover, the corresponding distance matrix is calculated, and the matching results between the UAV and the radiation source are obtained. The multisphere intersection method is used to calculate the reference location of the radiation source by combining the mapping relationship between the received signal strength and the distance. Furthermore, the original optimization problem is transformed into the Markov decision process, and the radiation source location information is introduced into reinforcement learning to design an efficient UAV swarm flight trajectory optimization algorithm. In addition, group entropy is introduced to analyze and evaluate the intelligence of the proposed algorithm. Finally, the performance and the intelligence of the proposed algorithm and the UAV swarm trajectories are analyzed using the simulation results, and the effectiveness of the proposed method is verified.

UAVs, radiation sources, localization and tracking, trajectory optimization, reinforcement learning, swarm entropy

doi: [10.1360/SST-2021-0378](https://doi.org/10.1360/SST-2021-0378)